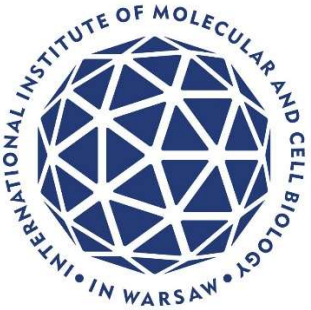


Development of MODOMICS and NACDDB databases and their application to studying the impact of RNA modifications on structure

by

Andrea Cappannini

Doctoral dissertation
Laboratory of Bioinformatics and Protein Engineering
International Institute of Molecular and Cell Biology in Warsaw

Supervisor: Professor Janusz Marek Bujnicki

Warsaw, July, 2025

Abstract

Post-transcriptional modifications shape RNA structure and function across all domains of life. These chemical alterations influence folding, molecular interactions, and stability, thereby affecting translational efficiency and fidelity. However, their distribution, conservation, and structural impact remain incompletely characterised. In addition to natural modifications, a wide variety of artificial chemical modifications have been introduced into nucleic acids for use in biophysical studies, structural probing, and therapeutic applications. A major part of the work presented in this thesis involves the development of databases that collect and organise diverse types of data related to nucleic acid modifications. First, a substantial expansion of the MODOMICS database was carried out, focusing on RNA modifications and their structural context. Second, a new database, NACDDB, was developed for the collection and analysis of circular dichroism (CD) spectra of nucleic acids, encompassing both naturally occurring and synthetically modified molecules. Beyond the development of these resources, the data contained within them were analysed to investigate how chemical modifications exert position- and context-dependent effects on RNA structure. A comparative analysis of RNA sequence alignments and three-dimensional structures in MODOMICS was performed to identify evolutionary signatures and structural preferences of modified nucleotides. Domain-specific analysis of ribosomal rRNA showed evolutionary conservation of key modifications, especially within the peptidyl transferase centre. The structural contexts of uridine and its modified derivatives — pseudouridine (Ψ) and dihydrouridine (D) — were examined with respect to base stacking interactions. The analysis showed that dihydrouridine residues typically occur outside of stacking interactions, while pseudouridine residues are more frequently involved in stacking, particularly with their 5′-neighbouring nucleotides, compared to uridine. This demonstrates the link between the altered stacking preferences of Ψ and D and their selection at positions where the presence or absence of stacking interactions is important for the RNA structure. Further, circular dichroism data from NACDDB were used to evaluate the structural impact of m^1A and m^6A modifications. The data indicated that methylated adenosines alter local conformational stability, with thermal destabilisation observed at elevated temperatures. Future research integrating diverse experimental data on the positions of modified residues in RNA sequences, together with structural data for RNAs containing such modifications, will be essential to fully understand how RNA modifications influence RNA structure. Toward this goal, a collaboration has been initiated between the MODOMICS team and the developers of the Sci-ModoM database, which aims to extract modified transcript sequences from high-throughput sequencing datasets and to analyse their structural preferences.

Streszczenie

Modyfikacje posttranskrypcyjne kształtują strukturę i funkcję RNA we wszystkich organizmach. Te zmiany chemiczne wpływają na tworzenie się struktury, odziaływania molekularne i stabilność, wpływając tym samym na wydajność i dokładność biosyntezy białek. Jednak ich dystrybucja, zachowawczość w ewolucji i wpływ strukturalny są niekompletnie scharakteryzowane. Oprócz modyfikacji naturalnych, do kwasów nukleinowych wprowadzono szeroką gamę sztucznych modyfikacji chemicznych, które wykorzystuje się m.in. w badaniach biofizycznych, w chemicznym sondowaniu struktur i w zastosowaniach terapeutycznych. Główna część pracy badawczej przedstawionej w tej rozprawie obejmuje rozwój baz danych, które gromadzą i organizują różne typy danych związanych z modyfikacjami kwasów nukleinowych. Po pierwsze, przeprowadzono znaczną rozbudowę bazy danych MODOMICS, skupiając się na modyfikacjach RNA i ich kontekście strukturalnym. Po drugie, opracowano nową bazę danych, NACDDB, do gromadzenia i analizy widm dichroizmu kołowego (CD) kwasów nukleinowych, obejmujących zarówno cząsteczki występujące naturalnie, jak i syntetycznie zmodyfikowane. Oprócz rozwoju tych narzędzi bioinformatycznych, dane w nich zawarte zostały przeanalizowane w celu zbadania, w jaki sposób modyfikacje chemiczne wywierają zależne od położenia i kontekstu efekty na strukturę RNA. Przeprowadzono analizę porównawczą przyrównań sekwencji RNA i struktur trójwymiarowych w MODOMICS w celu zidentyfikowania sygnatur ewolucyjnych i preferencji strukturalnych modyfikowanych nukleotydów. Analiza rybosomalnego rRNA wykazała wzory zachowania kluczowych modyfikacji w procesie ewolucji dywergentnej sekwencji rRNA, szczególnie w obrębie centrum peptydylotransferazy. Konteksty strukturalne urydyny i jej zmodyfikowanych pochodnych pseudourydyny (Ψ) i dihydrourydyny (D) zostały zbadane pod kątem oddziaływania warstwowego z rybonukleotydami sąsiadującymi w sekwencji RNA. Analiza wykazała, że reszty dihydrourydyny zazwyczaj występują w kontekście pozbawionym oddziaływań warstwowych, podczas gdy reszty pseudourydyny częściej niż niezmodyfikowane reszty urydyny biorą udział w tworzeniu tych oddziaływań, szczególnie z rybonukleotydami sąsiadującymi z nimi od strony 5'. Wyniki te wskazują na związek między zmienionymi preferencjami Ψ i D względem oddziaływań warstwowych a ich preferencjami do znajdowania się w takim kontekście strukturalnym, w którym obecność lub brak oddziaływań warstwowych jest ważna dla poprawnego utworzenia struktury RNA. Kolejnym elementem prac była analiza danych dichroizmu kołowego z NACDDB pod kątem oceny wpływu strukturalnego modyfikacji m^1A i m^6A. Wyniki wskazują, że metylowane adenozyny zmieniają lokalną stabilność konformacyjną oraz wywołują destabilizację w podwyższonych temperaturach. Przyszłebadania integrujące różne dane doświadczalne dotyczące pozycji zmodyfikowanych reszt w sekwencjach RNA, wraz z danymi strukturalnymi dla RNA zawierających takie modyfikacje, będą niezbędne do pełnego zrozumienia, w jaki sposób modyfikacje RNA wpływają na strukturę RNA. W tym celu zainicjowano współpracę między zespołem MODOMICS a twórcami bazy danych Sci-ModoM, której celem jest wyodrębnienie

zmodyfikowanych sekwencji transkryptów z zestawów danych sekwencjonowania o wysokiej przepustowości i ustalenie ich preferencji strukturalnych.

In Memoriam

Before any acknowledgement or appreciation, this thesis, the effort made for its realisation, and the outcome are dedicated to the eternal memory of my sweetest Grandparents, **Mario Nicoziani** (**1/02/1934 – 7/04/2024**) and **Claretta De Pau** (**9/06/1937 – 9/07/2024)**, who passed away peacefully but with everyone's great sadness during the last few months of my doctorate. My thoughts and my heart go to you.

Acknowledgements

We made it!

My parents, **Gino Cappannini** and **Manola Nicoziani**, are the first to be thanked. You were the first faithful supporters of my academic journey. You represent the pillars of my moral convictions and attitudes, the primordial source of my resilience. Every no for a brighter yes is because of what you have been and still are.

To my wife **Immacolata** and my little "mouse" **Nevio**, my son. You are the cornerstone of my very essence and my Valinor, as splendent and serene as the undying lands where I always find peace and reinvigorated strength. You are the foundation upon which this work stands, a place of beauty and strength akin to the towering peaks of Taniquetil and the tranquil gardens of Lórien. Your love and support have been my light throughout my PhD journey and more.

To my colleagues at IIMCB: **Dr. Pietro Boccaletto**, to whom I am deeply grateful for the rapid growth of my skills in database development; **Dr. Filip Stefaniak**, with whom I collaborated on the analysis of tautomers in MODOMICS and from whom I learned so much about handling biochemical data; and last but not least, to the "**Iranian gang**"—**Dr. Farhang Jaryani**, **Masoud Amiri Farsani**, and **Naeim Moafinejad**—you truly made me laugh out loud. Thank you!

Sincere thanks to **Professor Janusz Marek Bujnicki.** Your precision, selectivity, and profound knowledge of science are truly remarkable. I could never have mastered the ability to explore and understand RNA structures alone—it is through your guidance and example, I have learned so much. Here's to many more electrifying adventures in structural biology ahead!

To my former mentor, Professor **Andrea Giansanti**, your immense culture, knowledge, and kindness is a beacon that enriches my way of doing science. I'm looking forward to many other research adventures with disordered proteins and molecular dynamics!

I dedicate this work to Poland — an extraordinary Nation, welcoming and full of kind, hard-working people, eager to embrace you and draw you into their vibrant and beautiful culture.

Poland is a land of strong traditions and deep-rooted values. Its splendid conservatism is rooted in principles and dignity, guided by a spirit that honours both heritage and humanity, with remarkably openhearted, generous, and warm people.

A safe country — a place where respect, hospitality, and dignity still hold great importance.

I will always be grateful for the support, serenity, and inspiration I found here.

Thank you!

Table of Contents

List of the most important abbreviations

Ψ/PSU: Pseudouridine

D: Dihydrouridine

m^{6}A: *N^6-methyladenosine*

m^{1}A: *N^1-methyladenosine*

DNA: deoxyribonucleic acid

RNA: Ribonucleic acid

ncRNA: non-coding RNAs

snRNA: small-nuclear RNAs

snoRNA: small nucleolar RNAs

RCSB: Research Collaboratory for Structural Bioinformatics

DNA: Deoxyribonucleic Acid

ncRNA: non-coding RNAs

tRNA: transfer RNA

rRNA: ribosomal RNAs

LSU: Long Subunit

SSU: Short Subunit

snRNA: small nuclear RNAs

snoRNA: small nuclear RNAs

SAXS: Small Angle Scattering X-Ray Crystallography

NMR: Nuclear Magnetic Resonance spectroscopy

ONT: Oxford Nanopore Technologies

RNAMDB: RNA Modification Database

DirectRMDB: Direct RNA Modification Database

DRS: Direct RNA Sequencing

SQL: Structured Query Language

JS: Javascript

HTML: HyperText Markup Language

DRY: Don’t Repeat Yourself

CG: Coarse-grained

CD: Circular Dichroism

MS: Mass Spectrometry

NACDDB: Nucleic Acid Circular Dichroism Database

SRCD: Synchrotron Radiation Circular Dichroism

IIMCB: International Institute of Molecular and Cell Biology in Warsaw

Papers covered in this thesis

- (*) **Cappannini A***, Mosca K*, Mukherjee S, Moafinejad SN, Sinden RR, Arluison V, Bujnicki J, Wien F. NACDDB: Nucleic Acid Circular Dichroism Database. Nucleic Acids Res. 2023 Jan 6;51(D1):D226-D231. doi: 10.1093/nar/gkac829. PMID: 36280237; PMCID: PMC9825466.
- **Cappannini A**, Ray A, Purta E, Mukherjee S, Boccaletto P, Moafinejad SN, Lechner A, Barchet C, Klaholz BP, Stefaniak F, Bujnicki JM. MODOMICS: a database of RNA modifications and related information. 2023 update. Nucleic Acids Res. 2024 Jan 5;52(D1):D239-D244. doi: 10.1093/nar/gkad1083. PMID: 38015436; PMCID: PMC10767930.

Coauthored papers during this Ph.D.

- Eliana Destefanis, Denise Sighel, Davide Dalfovo, Riccardo Gilmozzi, Francesca Broso, **Andrea Cappannini**, Janusz M Bujnicki, Alessandro Romanel, Erik Dassi, Alessandro Quattrone, The three YTHDF paralogs and VIRMA are strong cross-histotype tumor driver candidates among m6A core genes, NAR Cancer, Volume 6, Issue 4, December 2024, zcae040, https://doi.org/10.1093/narcan/zcae040
- Boccaletto P, Stefaniak F, Ray A, **Cappannini A**, Mukherjee S, Purta E, Kurkowska M, Shirvanizadeh N, Destefanis E, Groza P, Avşar G, Romitelli A, Pir P, Dassi E, Conticello SG, Aguilo F, Bujnicki JM. MODOMICS: a database of RNA modification pathways. 2021 update. Nucleic Acids Res. 2022 Jan 7;50(D1):D231-D235. doi: 10.1093/nar/gkab1083. PMID: 34893873; PMCID: PMC8728126.
- Etienne Boileau, Harald Wilhelmi, Anne Busch, **Andrea Cappannini**, Andreas Hildebrand, Janusz M Bujnicki, Christoph Dieterich, Sci-ModoM: a quantitative database of transcriptome-wide high-throughput RNA modification sites, Nucleic Acids Research, 2024;, gkae972, https://doi.org/10.1093/nar/gkae972
- Szulc NA, Stefaniak F, Piechota M, Soszyńska A, Piórkowska G, **Cappannini A**, Bujnicki JM, Maniaci C, Pokrzywa W. DEGRONOPEDIA: a web server for proteome-wide inspection of degrons. Nucleic Acids Res. 2024 Jul 5;52(W1):W221-W232. doi: 10.1093/nar/gkae238. PMID: 38567734; PMCID: PMC11223883.
- (**) Elsa Balduzzi, Frédéric Geinguenaud, Dominik Sordyl, Satyabrata Maiti, Masoud Amiri Farsani, Grigory Nikolaev, Véronique Arluison, Janusz M Bujnicki, NAIRDB: a database of Fourier transform infrared (FTIR) data for nucleic acids, Nucleic Acids Research, 2024;, gkae885, https://doi.org/10.1093/nar/gkae885
- (*) Joint first authorship; (**) Acknowledged in this paper

Other coauthored papers not related to the Ph.D.

- **Cappannini A**, Forcelloni S, Giansanti A. Evolutionary pressures and codon bias in low complexity regions of plasmodia. Genetica. 2021 Aug;149(4):217-237. doi: 10.1007/s10709-021-00126-6. Epub 2021 Jul 12. PMID: 34254217.
- Buccarelli M, D'Alessandris QG, Matarrese P, Mollinari C, Signore M, **Cappannini A**, Martini M, D'Aliberti P, De Luca G, Pedini F, Boe A, Biffoni M, Pallini R, Ricci-Vitiani L. Elesclomol-induced increase of mitochondrial reactive oxygen species impairs glioblastoma stem-like cell survival and tumor growth. J Exp Clin Cancer Res. 2021 Jul 12;40(1):228. doi: 10.1186/s13046-021-02031-4. PMID: 34253243; PMCID: PMC8273992.
- Castellani G, Buccarelli M, D'Alessandris QG, Ilari R, **Cappannini A**, Pedini F, Boe A, Lulli V, Parolini I, Giannetti S, Biffoni M, Zappavigna V, Marziali G, Pallini R, Ricci-Vitiani L. Extracellular

vesicles produced by irradiated endothelial or Glioblastoma stem cells promote tumor growth and vascularization modulating tumor microenvironment. Cancer Cell Int. 2024 Feb 12;24(1):72. doi: 10.1186/s12935-024-03253-0. PMID: 38347567; PMCID: PMC10863174.

Grants and fundings

ROPES

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956810

1 Introduction

1.1 Overview of the structure and function of RNA molecules

Ribonucleic acid (RNA) is a ubiquitous biomolecule found in all living organisms. It participates in processes such as catalysis, gene regulation, and the transfer of genetic information. RNA has also emerged as a biotechnological tool in drug development (Mondragon et al., 2020; Ishigaki et al., 2018; Cheng-Kai Huang et al., 2020; Abramson et al., 2024; Laurents et al., 2022).

RNA families are classified primarily by biochemical function, often reflecting unique three-dimensional (3D) structures (Ontiveros et al., 2025). These structures support specific molecular roles. For instance, catalytic RNAs like hammerhead and hairpin ribozymes show how folding enables self-cleavage, though such RNAs are rare (Scott et al., 2013). Many others act within ribonucleoprotein complexes. Ribonuclease P (RNase P), for example, relies on RNA–protein contacts to process precursor tRNAs (Guerrier-Takada et al., 1983). While correct folding is necessary for function, not all RNAs adopt one stable structure. Long non-coding RNAs (lncRNAs) often form dynamic structures that regulate gene expression (Statello et al., 2021; Zampetaki et al., 2018).

Knowledge of RNA structural behaviour and interaction patterns guides the formulation of RNA-based therapeutic approaches. These approaches aim to correct defects caused by misfolding or abnormal modifications. In neurological disorders, mutations in the RNA-binding protein FUS compromise its role in stabilising mRNAs involved in synaptic plasticity, leading to enhanced mRNA degradation and neuronal morphological changes (Ishigaki et al., 2018). A comparable example is found in MELAS syndrome (Mitochondrial Encephalomyopathy, Lactic Acidosis, and Stroke-like Episodes), where point mutations in mitochondrial tRNA genes impair mitochondrial protein synthesis. The most common mutation, m.3243A>G in the MT-TL1 gene encoding $tRNA^{Leu}$(UUR), disrupts the dihydrouridine loop, affecting both aminoacylation and ribosome interaction. This structural alteration reduces translation efficiency of mitochondrial-encoded proteins, particularly those forming the electron transport chain. The resulting complex I deficiency compromises ATP production, predominantly affecting high-energy tissues such as the brain and skeletal muscles. Clinically, this leads to stroke-like episodes, muscle weakness, lactic acidosis, and progressive neurological decline. Other mutations, such as m.5541C>T in the MT-TW gene ($tRNA^{Trp}$), have been described with similar consequences. These cases underline the impact of local RNA structural perturbations on translation fidelity and cellular

energy metabolism, reinforcing the importance of RNA–protein interactions in maintaining organelle function (Goto et al., 1990).

The role of RNA structure is central to therapeutic strategies. Recent developments use both structure- and sequence-specific recognition. Small molecules bind folded regions, while oligonucleotides target unstructured areas. These strategies show promise in treating Huntington's disease, Amyotrophic Lateral Sclerosis, and cardiovascular conditions (Cheng-Kai Huang et al., 2020). RNA structure–function principles also underpin RNA-based vaccine platforms. The structure of RNA affects its resistance to nucleases, a key challenge in therapeutic applications. For example, mRNA vaccines developed by Pfizer-BioNTech and Moderna include modified nucleotides and lipid nanoparticles (LNPs) to stabilise the RNA and foster efficient protein translation (EMA_COVID19: https://bit.ly/4d9ypXc; Harcourt et al., 2017).

RNA structure prediction is limited by both experimental and computational challenges. While AlphaFold has advanced protein modelling, high-resolution RNA structures remain scarce (Varadi et al., 2024). RNA crystallization is particularly difficult. One reason is the molecule's polyanionic backbone. The phosphate groups carry negative charges that repel each other, especially near the 5′ end. These electrostatic forces destabilise tertiary folding (Jackson et al., 2023; Shen et al., 2022; Miao et al., 2015).

To support proper folding, metal ions such as magnesium (Mg^{2+}) are added during sample preparation. These ions reduce repulsion along the phosphate backbone and help stabilise tertiary structures (Yamagami et al., 2021). In addition to metal binding, RNA structure is influenced by chemical modifications. Over 150 such modifications have been identified, collectively called the epitranscriptome (Cappannini et al., 2023). These marks regulate stability, folding dynamics, and molecular interactions (Morais et al., 2021).

Detecting RNA modifications remains technically challenging. Experimental artefacts, such as misincorporation by reverse transcriptase, may lead to overestimated modification levels. These issues affect reproducibility and accuracy (Sas-Chen et al., 2019; Potapov et al., 2018). Despite progress in RNome databases and Deep Learning approaches (Boileau et al., 2024; Shen et al., 2024), the *"crusade"* for RNA structure prediction and accurate modification mapping is far from over. Methodological gaps and experimental limitations still hinder progress.

Reliable and rigorously curated resources are, therefore, necessary to overcome existing hurdles. The subsequent sections discuss RNA structural principles, methodologies for determining RNA architecture, and challenges associated with

identifying post-transcriptional modifications, providing a foundation to examine RNA structure–function relationships in detail.

1.2 RNA Structure

RNA folding results from a dynamic equilibrium among multiple forces, including hydrogen bonding, base stacking interactions, and electrostatic repulsions arising from the negatively charged phosphate backbone of RNA with metal ions, particularly divalent cations such as Mg^{2+}, neutralising electrostatic repulsions and promoting the stabilisation of RNA tertiary structures. The physicochemical properties and structural characteristics of RNA emerge from interactions among its constituent components: phosphate groups, ribose sugars, and nitrogenous bases. The following sections provide a detailed description of these building blocks.

1.2.1 Nucleotides

RNA is a nucleotide polymer composed of adenosine (A), uridine (U), guanosine (G), and cytidine (C) nucleotide monomers. Each nucleotide consists of a nitrogenous base linked to a ribose sugar by a glycosidic bond at the 1′ carbon (Lehninger et al., 2017). The presence of a hydroxyl group at the 2′ carbon distinguishes RNA structurally from DNA.

The 2′-hydroxyl group contributes to structural adaptability, chemical reactivity, and catalytic behaviour. Its nucleophilic oxygen, positioned near the phosphate backbone, facilitates intramolecular reactions (Tzokov et al., 2002; Sood et al., 2002; Ayadi et al., 2019). It also participates in hydrogen bonding that stabilises folding, helix packing, and both local and long-range interactions (Sood et al., 2002). These interactions support the formation of structural motifs such as hairpins, pseudoknots, and tetraloops (Lilley, 2007).

The phosphate backbone, composed of phosphodiester bonds between the 3′ carbon of one nucleotide and the 5′ carbon of the next, carries negative charges. These charges affect folding through electrostatic repulsion and interactions with water. Each phosphate binds from 2 to 6 water molecules, contributing at least –1.8 kcal/mol in hydration-free energy for the first two molecules. During folding, the release of these waters promotes compaction. In the adenine riboswitch, about 170 water molecules—2.4 per nucleotide—are released (Herschlag, 1995). Backbone flexibility also imposes an entropic cost, especially at loops and tertiary contacts. In the M-box riboswitch, a regulatory element typically found in the 5′ untranslated regions (5′ UTR) of Gram-positive bacteria, this cost is estimated at 0.72 kcal/mol per nucleotide.

Metal ions, especially magnesium (Mg^{2+}), reduce electrostatic repulsion by neutralising phosphate charges. In RNA, two potassium ions (K^{+}) are typically replaced by one Mg^{2+}, releasing a net ion and promoting folding. Mg^{2+} binds specific sites—such as the ten anionic oxygens in the M-box riboswitch—orienting the backbone for noncanonical base pairing. It also binds with high affinity to structurally strained regions (Mak et al., 2014).

The physicochemical properties of the nitrogenous bases—adenine, guanine, cytosine, and uracil—modulate RNA folding dynamics through hydrogen bonding and stacking interactions. Differences in electron distribution and steric properties shape local secondary structures, such as those formed by A–U and G–C base pairs. **Table 1** summarises the chemical and structural characteristics of the canonical RNA nucleotides.

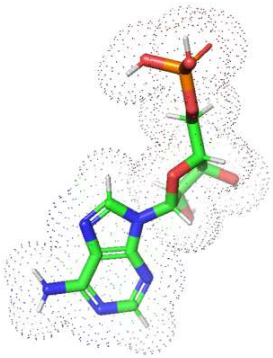
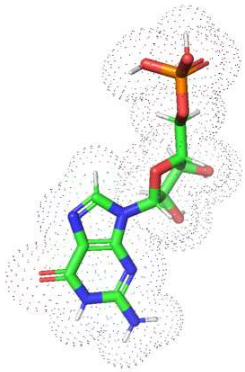
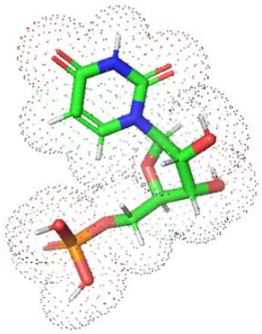
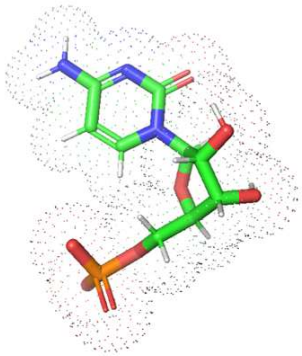
					
	RCSB: AMP	RCSB: 5GP	PUBCHEM: UMP	PUBCHEM: CMP	DATABASE NAMES
	A	G	U	C	ABBREVIATION
	6-aminopurine	2-amino-6-oxopurine	2,4-dioxopyrimidine	2-oxo-4-aminopyrimidine	NITROGENOUS BASE
	ADENOSINE	GUANOSINE	URIDINE	CYTIDINE	NUCLEOSIDE
STANDARD NUCLEOTIDES					

Table 1 Schematic representation of the four standard nucleotides. Nucleotides are rendered with **PyMol**. Following, links to the database pages: **AMP**: https://www.rcsb.org/ligand/AMP **5GP**: https://www.rcsb.org/ligand/5GP **U5P**: https://pubchem.ncbi.nlm.nih.gov/compound/Uridine-Monophosphate **CMP**: https://pubchem.ncbi.nlm.nih.gov/compound/Cytidine-monophosphate

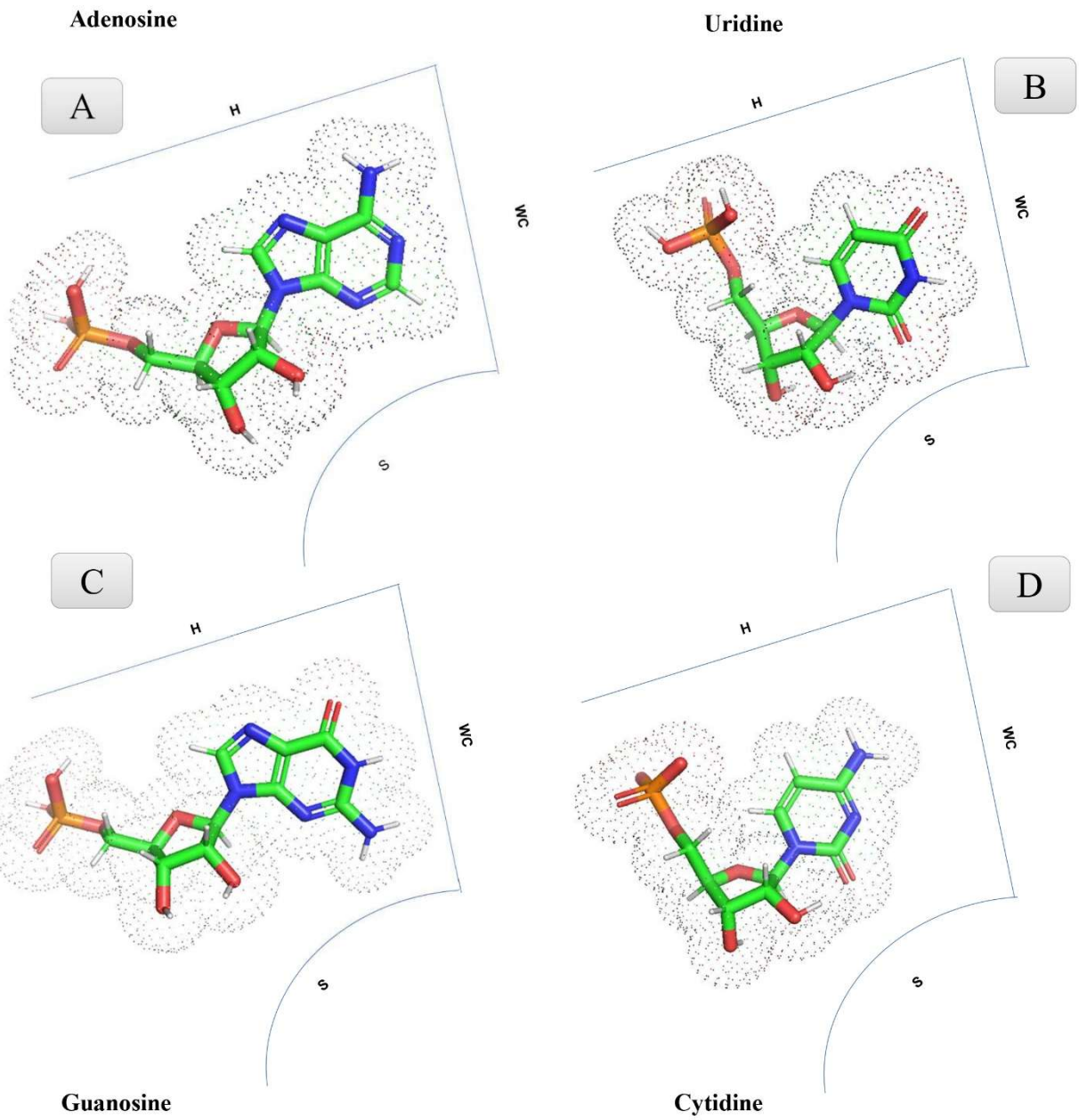


Figure 1 Schematic illustration of the four standard nucleotides emphasizing their three interactive edges. From the top left corner in clockwise direction: Adenosine, Uridine, Guanosine, Cytosine. Picture inspired by Lescoute et al. 2006 .

1.2.2 Nucleotide edges

Watson–Crick (WC), Hoogsteen (HE), and Sugar (SE) edges are the hydrogen-bonding interfaces of nucleotides, containing atoms that act as hydrogen donors or acceptors (**Fig. 1**). The Watson–Crick edge enables canonical base pairing in the double helix. In adenine, hydrogen bonding involves donation from N^6 and acceptance via N^1. Uracil (or thymine) donates through N^3 and accepts via O^4. Guanine accepts through O^6 and

donates via N^1 and N^2, while cytosine donates via N^4 and accepts through N^3 and O^2. The Hoogsteen edge provides an additional hydrogen-bonding interface that supports non-canonical interactions, such as the formation of tetrads in G-quadruplexes (Sengar et al., 2014) or the binding of a third RNA strand to the major groove of a Watson–Crick duplex in triple helices (Brown, 2020). In purines, the Hoogsteen edge comprises the N^7 and C^8 atoms, while in pyrimidines it involves the C^5 and C^6 atoms, forming weaker C–H hydrogen bonds (Halder et al., 2019). For instance, uracil forms C^5–$H\cdots O^6$ bonds with guanine in non-canonical G:U W:H trans base pairs (see ClaRNA server supplementary materials, Walen et al., 2014). The sugar edge in RNA corresponds to the base region oriented toward the sugar moiety, including the 2′-hydroxyl group of ribose. Alongside the Watson–Crick and Hoogsteen edges, it constitutes one of the three primary interaction regions in the Leontis–Westhof classification (2001). The sugar edge stabilises non-canonical base pairs and RNA tertiary structures that cannot occur in DNA, such as ribose zippers and A-minor motifs (Leslie & Grover, 2020).

1.2.3 Nucleotide pairing families

Nucleotide pairing families are grouped into twelve distinct classes. These classes are defined by the nucleotide edges involved in forming hydrogen bonds (**Tables 2** and **3**). Leontis et al. (2001) introduced a triangular representation to describe the spatial orientation of these edges, where each side corresponds to one of the three interaction interfaces. The hypotenuse represents the Hoogsteen edge (C–H in pyrimidines), the minor leg corresponds to the Watson–Crick edge, and the remaining cathetus represents the Sugar edge.

Each nucleotide within the pair contributes an edge, forming a doublet. The relative orientation of the edges defines the pairing class, resulting in twelve families. The Watson–Crick (W), Hoogsteen (H), and Sugar (S) edges interact in two configurations: cis, where the ribose sugars lie on the same side of the hydrogen bond vector, and trans, where they lie on opposite sides. These configurations divide the twelve families into two groups based on ribose orientation.

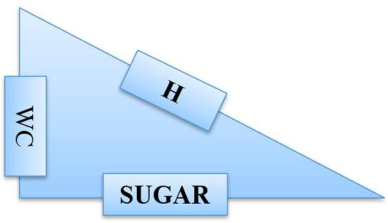
Cis Configuration: The ribose moieties of both nucleotides are on the same side of the hydrogen bond vector.

- cis Watson-Crick/Watson-Crick (cWW)
- cis Watson-Crick/Hoogsteen (cWH)
- cis Watson-Crick/Sugar edge (cWS)

- cis Hoogsteen/Hoogsteen (cHH)
- cis Hoogsteen/Sugar edge (cHS)
- cis Sugar edge/Sugar edge (cSS)

Glycosidic Bond Orientation	Abbreviation	Interacting Edges		Symbol	Triangle Abstraction Representation
		Nu1	Nu2		
CIS	cWW	Watson-Crick	Watson-Crick		
	cWH	Watson-Crick	Hoogsteen		
	cWS	Watson-Crick	Sugar		
	cHH	Hoogsteen	Hoogsteen		
	cHS	Hoogsteen	Sugar		
	cSS	Sugar	Sugar		

Table 2 Schematic triangle representation of nucleotide edges for nucleotide cis- canonical and non-canonical interactions. WC : Watson-Crick edge; H: Hoogsteen-edge; SUGAR : Sugar-edge; This representation is inspired from Leontis et al. 2001.



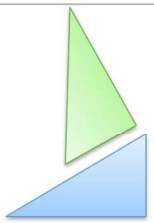
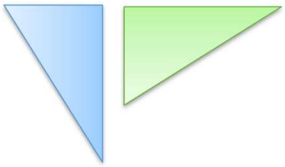
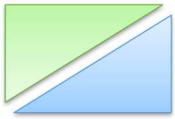
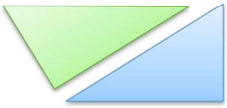
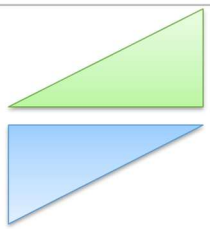
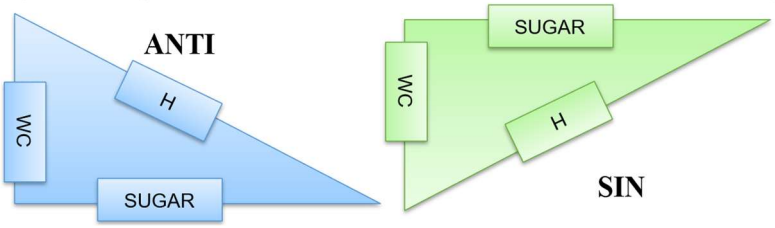
Glycosidic Bond Orientation	Abbreviation	Interacting Edges		Symbol	Triangle Abstraction Representation
		Nu1	Nu2		
TRANS	cWW	Watson-Crick	Watson-Crick		
	cWH	Watson-Crick	Hoogsteen		
	cWS	Watson-Crick	Sugar		
	cHH	Hoogsteen	Hoogsteen		
	cHS	Hoogsteen	Sugar		
	cSS	Sugar	Sugar		

Table3 Schematic triangle representation of nucleotide edges for nucleotide trans-non-canonical interactions. This representation is inspired from Leontis et al. 2001.



1. **Trans Configuration:** The ribose moieties are on opposite sides of the hydrogen bond vector.
    - trans Watson-Crick/Watson-Crick (tWW)
    - trans Watson-Crick/Hoogsteen (tWH)
    - trans Watson-Crick/Sugar edge (tWS)
    - trans Hoogsteen/Hoogsteen (tHH)
    - trans Hoogsteen/Sugar edge (tHS)
    - trans Sugar edge/Sugar edge (tSS)

Nucleotide pairing involves 18 configurations, combining cis and trans orientations (Almakarem et al., 2011). The cWW (cis Watson–Crick/Watson–Crick) interaction is canonical, underlying standard base pairing in DNA and RNA helices. All other pairings are non-canonical and contribute to the formation of RNA tertiary structures, such as loops, junctions, and ribosomal RNAs.

The nucleotide edges contribute distinct chemical groups, differing between purines (adenine, guanine) and pyrimidines (cytosine, uracil in RNA; thymine in DNA). These differences expand the range of base pairing geometries and interaction modes, contributing to structural diversity beyond the double helix.

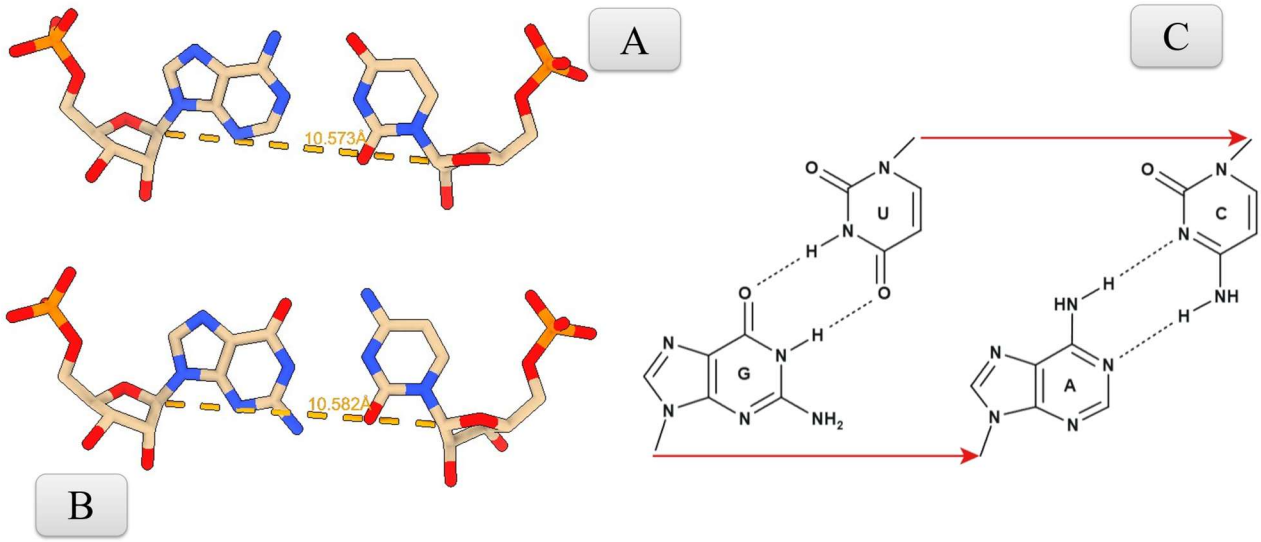


Figure 2 Isosteric base pairs. (A) cAU. (B) cGC. (C)tGU and tAC.Four known isosteric base pairs (Westhof, 2014). Canonical nucleotide pairs have been taken from ClaRNA website (Walen et al. 2014). Noncanonical nucleotide pairs have been depicted with Marvin Sketch

1.2.4 Isosteric nucleotide pairs

Isosteric base pairs share equivalent $C^{1\prime}$ atom positions and distances. Substituting one canonical doublet with another isosteric pair does not alter the overall geometry of the RNA helix. This geometric compatibility maintains a consistent helical architecture. The $C^{1\prime}$–$C^{1\prime}$ distance typically measures ~10.5 Å, supporting helical regularity during base substitutions. Non-complementary WC/WC pairs, such as G·U and A·C in the trans configuration, are isosteric with each other (**Fig. 2**).

1.2.5 Stacking interactions

Nucleotide stackings are non-covalent interactions that involve the planar juxtaposition of aromatic bases (Taghavi et al., 2022; Condon et al., 2015; Brown et al., 2015). In aqueous environments, base stacking minimises hydrophobic surface exposure (Friedman & Honig, 1995). The hydrophobic effect excludes water from stacked interfaces, thereby decreasing base–solvent interaction energy. Stacking interactions involve additional contributions beyond hydrophobicity. The aromaticity of nucleotide bases enables π-π interactions via delocalised π-electron clouds, producing favourable electrostatic forces (Carter-Fenk & Herbert, 2020), whereas London dispersion forces

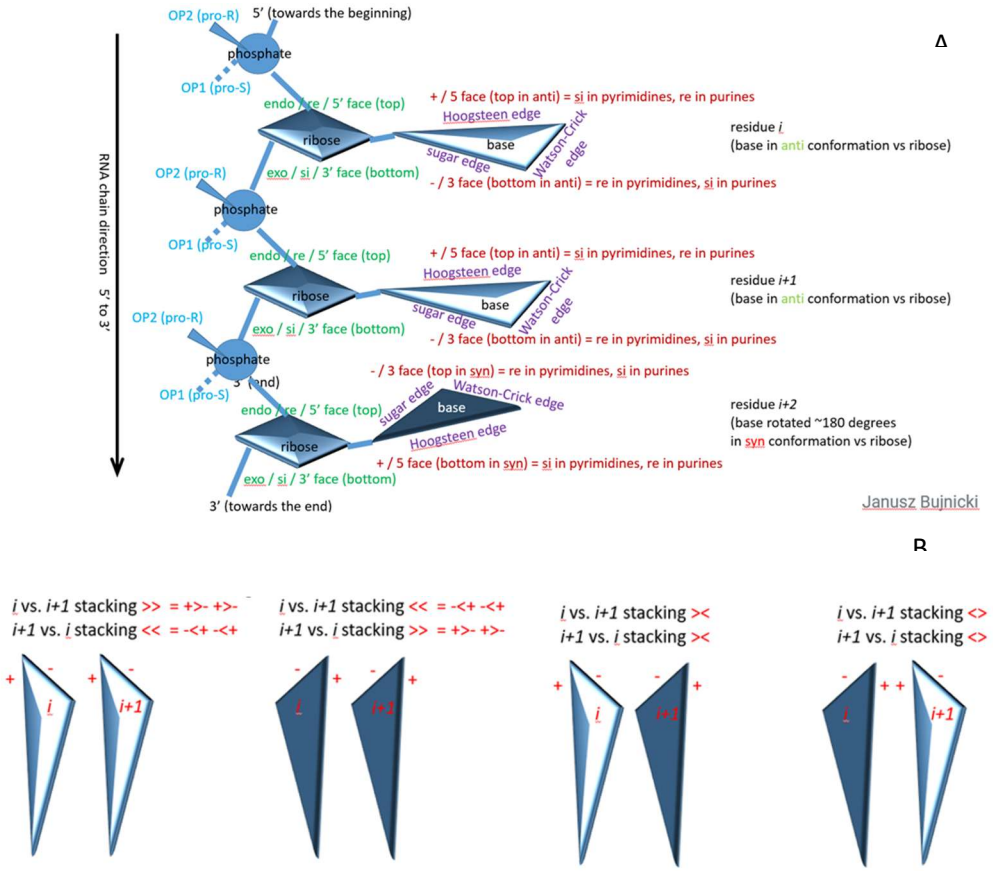


Figure 3 Geometrical representation of nucleotide stacking interactions. (**A**) RNA chain and visual configuration of nucleotide stacking partners. (**B**) Nucleotide faces schematic representation and list of possible stacking interaction (specular) classes. **This image is a modified version of a picture by Janusz M. Bujnicki, used with permission of the author.**

(LDF), a type of van der Waals interaction, arise from transient dipoles resulting from fluctuations in electron density, as shown by symmetry-adapted perturbation theory (SAPT).

LDFs favour optimal base overlap at interplanar distances of ~3.3 Å. They act with electrostatic and induction forces to offset repulsive exchange interactions. This interplay of quantum mechanical forces supports the structural cohesion and conformational adaptability of nucleic acids (Matta et al., 2006; Parker et al., 2013). The physical requirement to maximise stabilising interactions, particularly London dispersion forces, imposes constraints on nucleotide geometry. These constraints result in discrete stacking arrangements, which can be systematically classified based on the orientation of the base face.

RNA stacking interactions between two nucleotides (n_{th} and $n_{(th+i)}$) are classified into four categories (++, +-, --, -+) based on the nucleotide face orientation. The "+" face aligns with the side facing the 3'-end in the anti-conformation, while the "-" face is the opposite. Because stacking interactions depend on edge geometry rather than absolute orientation, rotation of the glycosidic bond to the syn-conformation does not alter classification. This rotational invariance reflects the physical symmetry that governs stacking.

The glycosidic bond links the base to the sugar via the $N–C^{1\prime}$ atom. It can rotate into a syn ($0° \pm 90°$) or anti ($180° \pm 90°$) conformation. This rotation does not modify the nucleotide edge geometry and thus preserves stacking classification. Dinucleotide conformational analysis defines 96 classes describing stacked, unstacked, intercalated, or open conformations, using torsional angles such as χ_1 and χ_2. This invariance illustrates how stacking interactions remain conserved despite local conformational variability, maintaining stability through LDFs and other contributing forces (NT Standards: https://nakb.org/basics/nucleotides.html).

The ClaRNA classification formalises these principles, allowing consistent identification of stacking types across diverse conformations. Its descriptors are topology-driven and align with physicochemical criteria that define RNA structure. As shown in **Fig. 3A**, rotation to the syn-conformation does not alter nucleotide edge geometry, preserving stacking classification. ClaRNA nomenclature captures these equivalences regardless of viewing orientation (Walen et al., 2014). Figure 3B illustrates equivalent stacking interactions, highlighting how classification depends on the interacting edges.

1.2.6 Dihedral angles

Compared to proteins, the RNA backbone exhibits greater flexibility, densely populating conformational niches and transiently exploring others during functional state transitions (Tzokov et al., 2003). The sugar-phosphate backbone of RNA consists of rotatable torsion angles (**α**, **β**, **γ**, **δ**, **ε**, and **ζ**), which define its conformational space.

Table 4

Torsional Angle	Involved Atoms	Plane 1	Plane 2
Alpha (α)	O3'-P-O5'-C5'	O3', P, O5'	P, O5', C5'
Beta (β)	P-O5'-C5'-C4'	P, O5', C5'	O5', C5', C4'
Gamma (γ)	O5'-C5'-C4'-C3'	O5', C5', C4'	C5', C4', C3'
Delta (δ)	C5'-C4'-C3'-O3'	C5', C4', C3'	C4', C3', O3'
Epsilon (ε)	C4'-C3'-O3'-P	C4', C3', O3'	C3', O3', P
Zeta (ζ)	C3'-O3'-P-O5'	C3', O3', P	O3', P, O5'
Chi (χ) Purines	O4'-C1'-N9-C4	O4', C1', N9	C1', N9, C4
Chi (χ) Pyrimidines	O4'-C1'-N1-C2	O4', C1', N1	C1', N1, C2

Table 4 Torsional angles in nucleic acid structures, specifying the involved atoms and the planes these atoms form. For each torsional angle, the specific atoms contributing to the dihedral angle are listed, as well as the two planes that intersect to define the angle. The torsional angles include alpha (α), beta (β), gamma (γ), delta (δ), epsilon (ε), zeta (ζ), and chi (χ), with chi angles listed separately for purines and pyrimidines. Colours are associated to **Figure 4**

Figure 5

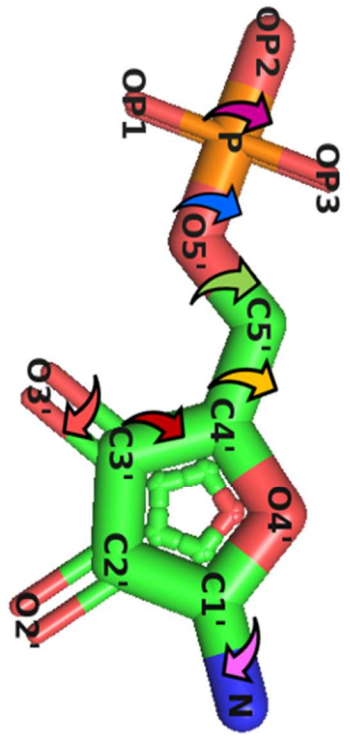


Figure 4 2D rendering of the ribose sugar with PyMol. AMP SMILES code has been utilized to generate a 2D pdb with OBABEL from which ribose's atoms have been extracted. Rows' colours are associated to **Table 4**

Torsional or dihedral angles describe the rotation around a bond. Four atoms define two intersecting planes in three-dimensional space whose angle defines the dihedral angle. The first and fourth atoms form the endpoints, and the positions of the second and third atoms determine the angle. The glycosidic torsion angle (**χ**) expands the RNA conformational landscape by allowing base rotation around the C^{1}'–N^{9} (purines) or C^{1}'–N^{1} (pyrimidines) bond. Nucleotides adopt either the syn or anti-conformation, with the

latter being predominant as it minimises steric clashes and positions the base away from the ribose, reducing electrostatic repulsion between the electron-dense aromatic ring and the negatively charged sugar (Leslie & Grover, 2020). **Table 4** reports the atomic planes that generate the torsional angles in a nucleotide. **Figure 4** provides a visual representation of the directions of these torsional angles (Hershkovitz et al., 2006).

1.2.7 Sugar Puckering

Sugar puckering arises from the out-of-plane displacement of atoms within the furanose ring. This distortion is driven by torsional strain in *C–C* and *C–O* bonds and serves to reduce steric and torsional repulsion. The displacement of a single atom from the ring plane relieves eclipsing interactions between adjacent bond electrons, stabilising the conformation. In RNA, the sugar ring typically adopts a *C³′-endo* (north) conformation (**Fig. 5A**), while in DNA, the absence of the 2′-hydroxyl group leads to a preference for the *C²′-endo* conformation (**Fig. 5B**), stabilised in the A- and B-form helices, respectively.

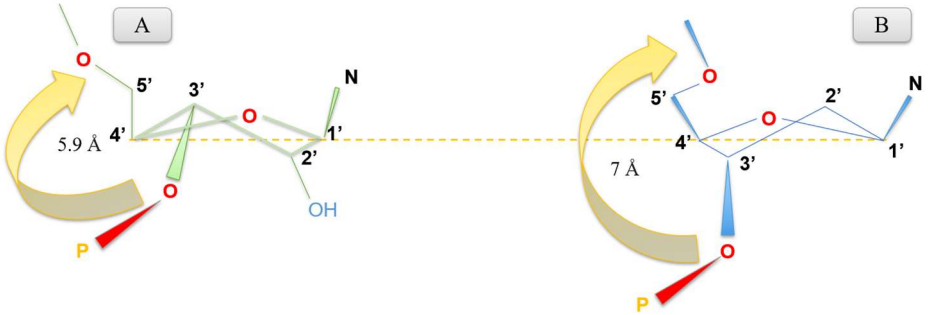


Figure 5 Schematic representation of ribo- (**A**) and 2'-deoxyribo-(**B**) 5'-monophopshate adenosine and the different sugar puckering. The picture is inspired from Leslie & Grover (2020)

The *C³′-endo* sugar pucker imposes a more compact geometry on the phosphate backbone. The distance between adjacent phosphate groups is reduced to approximately 5.9 Å, compared to ~7.0 Å in the *C²′-endo* (south) conformation (Egli, 2010). This reduced spacing brings the phosphate strands closer together and displaces the base pairs away from the helical axis. The resulting A-form helix exhibits a cylindrical cross-section, a deep and narrow major groove, and base pairs inclined ~20° relative to the helix axis (Dickerson et al., 1982). These features also produce a wider minor groove and a helical pitch of ~11 base pairs per turn.

In contrast, the B-form helix retains a more extended geometry due to the flatter C^2'-*endo* sugar conformation. This arrangement increases the inter-phosphate spacing and allows the base pairs to align nearly perpendicular to the helical axis. The helix thus presents a narrower minor groove, a wider major groove, and a helical pitch of ~10.5 base pairs per turn. The overall shape is more elongated, with reduced base pair inclination and greater structural flexibility in aqueous environments (Egli, 2010; Dickerson et al., 1982).

The structural differences between these two conformations affect not only helical shape and groove dimensions but also the physicochemical properties of the nucleic acid surface. Sugar puckering determines the exposure of functional groups, the hydration pattern, and the accessibility of the grooves to proteins and ions. It acts as a conformational regulator that links sugar stereochemistry to backbone geometry and ultimately to recognition by macromolecular partners. **Table 5** outlines the different structural features between A- and B- conformations, highlighting how variations in helical geometry, sugar pucker, and groove dimensions influence the overall architecture and biological recognition properties of nucleic acid helices (Brown, 2020).

Feature	A-form	B-form
Feature	A-form RNA	B-form DNA
Helix conformation	Right-handed A-form	Right-handed B-form
Sugar pucker	C^3'-*endo* (north)	C^2'-*endo* (south)
2′-substitution	2′-OH present	2′-H (deoxy)
Base pairs per turn	~11	~10
Helical rise per base pair (Å)	~2.6	~3.4
Helical pitch (Å)	~27.5	~34
Helical twist (°/bp)	~33°	~36°
Base-pair inclination (°)	~20°	~0° (nearly perpendicular)
Helix diameter (Å)	~23	~20
Major groove	Narrow, deep (bases inaccessible)	Wide, deep (bases accessible)
Minor groove	Wide, shallow	Narrow, shallow
Groove accessibility	Minor groove accessible	Major groove accessible
Hydration spine	Lacks ordered spine	Structured minor groove hydration spine
Backbone stability	Susceptible to cleavage via in-line *2′-OH* nucleophilic attack	Chemically stable phosphodiester backbone
Thermodynamic stability (ΔG°)	More negative per base pair (stronger base stacking)	Less negative per base pair
Persistence length (nm)	~62–63 (stiffer, more rigid)	~45–50 (more flexible)

Table 5. Schematical comparison of RNA and DNA double helix structural characteristics. Reviewed in Brown (2020)

1.2.8 RNA primary, secondary and tertiary structure

RNA structure is a hierarchical molecular organisation with three levels of increasing complexity: primary (nucleotide sequence), secondary (base pairing and folding), and tertiary (3D arrangement). The primary structure consists of an ordered array of

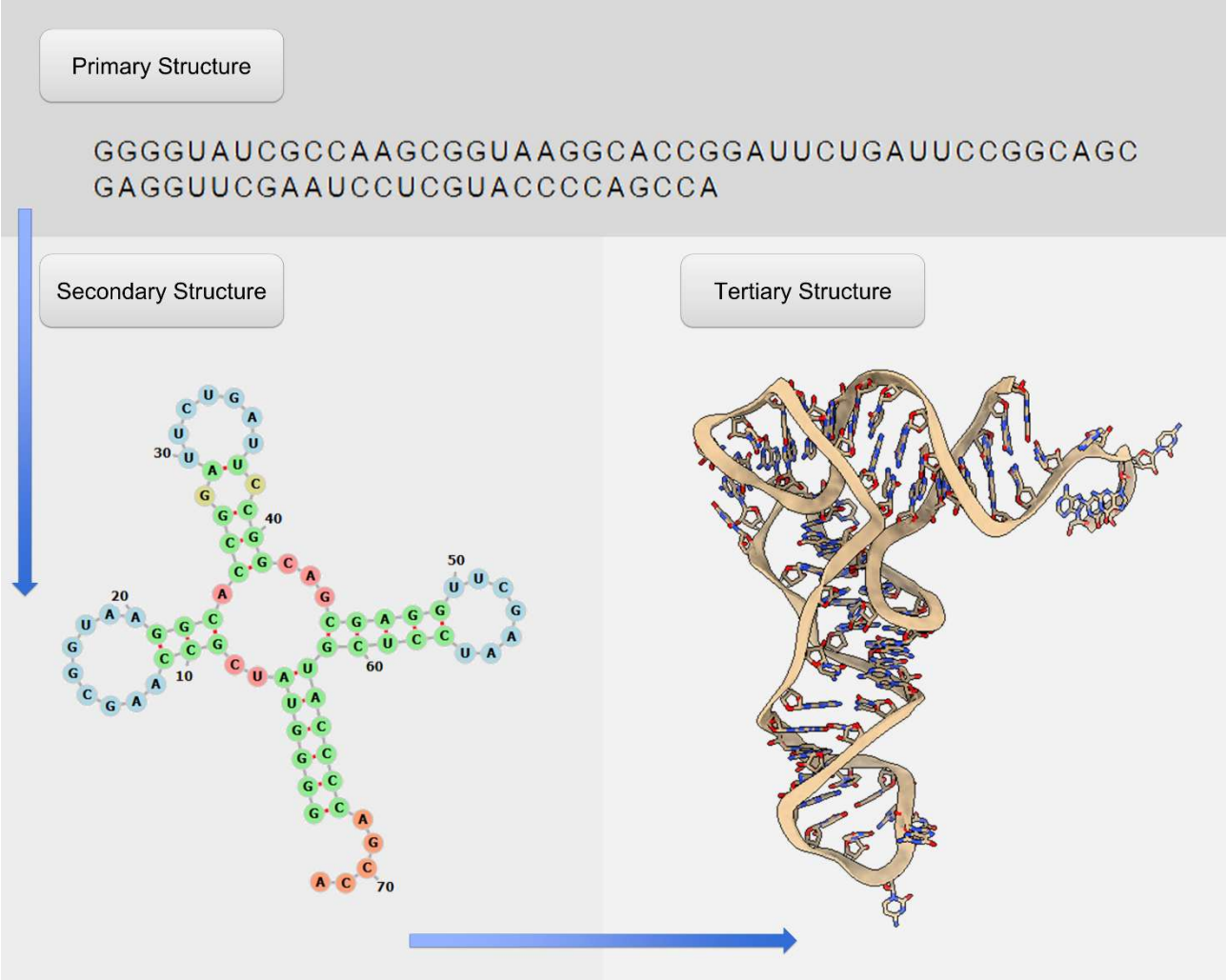


Figure 6. The picture highlights primary, secondary, and tertiary structure of glutamine-tRNA extracted from 1EUQ RCSB entry. The arrows follow the hierarchical folding flow attributed to RNA molecules. Secondary structure has been manually derived from visual nucleotide inspection and then confirmed and plotted with FoRNA server (Kerpedjiev et al., 2015). 3D structure is rendered with ChimeraX (Pettersen et al., 2021)

nucleotides joined by phosphodiester bonds between the 3′-oxygen of one nucleotide and the 5′-phosphate of the next. The resulting sequence is oriented in the 5′ to 3′ direction. The secondary structure includes canonical base-pairing interactions within a single RNA molecule, stabilised by hydrogen bonds between complementary bases. It contains elements such as stems (helices), hairpin loops, bulges, internal loops, multiloops (junctions), and long-range contacts such as pseudoknots and kissing loops (Macke et al., 2001).

RNA tertiary structure refers to its three-dimensional spatial conformation, which reflects the arrangement of all atoms within the molecule (Boniecki et al., 2016). Describing RNA tertiary structure as a single, well-defined conformation oversimplifies its folding behaviour, which differs fundamentally from that of proteins. In proteins, the native state typically corresponds to a global energy minimum on a smooth, funnel-shaped energy landscape. A single amino acid sequence folds into one predominant functional conformation (Anfinsen, 1973; Frauenfelder et al., 1991; Englander & Mayne, 2017). By contrast, RNA molecules exhibit rugged energy landscapes with multiple low-energy conformations coexisting in thermodynamic equilibrium. This feature allows RNA to occupy diverse conformational states in response to the cellular milieu (Chen et al., 2000). A representative illustration of the three levels is provided in **Figure 6**.

1.3 RNA modifications

The four canonical nucleotides constitute the chemical basis of RNA, and their permutations in quantity and order dictate its physicochemical properties and functional family. However, the RNA alphabet is significantly expanded by post- or co-transcriptional modifications that alter its physicochemical characteristics (Cappannini et al., 2023). RNA modifications decorate the nitrogenous base, ribose sugar, or phosphate backbone and collectively form the epitranscriptome. Over 150 naturally occurring RNA modifications have been identified (Cappannini et al., 2023), modulating RNA structure, stability, conformational dynamics, and interactions under physiological conditions.

Methylation exemplifies a frequent RNA modification. Despite its simple chemical nature, it influences RNA conformation through stereochemical constraints that shift electron density around nucleobases and generate stereoelectronic effects within the sugar–phosphate backbone. Specifically, 2′-*O*-methylation induces the anomeric effect through electron donation from the $O^{4'}$ lone pair into the σ* orbital of the glycosidic bond ($C^{1'}$–*N*) and promotes the 3′-gauche effect by altering torsion angles ($O^{2'}$–$C^{2'}$–$C^{1'}$–*N*) (Das et al., 2022).

Sugar 2′-*O* methylations stabilise the $C^{3'}$-*endo* pucker via hydrogen bonding interactions. Methylation restricts conformational space by reversing hydrogen-bond polarity, shifting from $O^{2'}H\cdots O^{3'}$ in RNA to $O^{2'}\cdots HO^{3'}$ in 2′-*O*-methylated nucleosides (He et al., 2018), highlighting the relevance of stereoelectronic effects induced by nucleotide modifications in synthetic biology and drug development (Das et al., 2022; He et al., 2018; Liebl et al., 2018; Yang et al., 2015).

RNA species undergo evolutionary modifications, driven by Darwinian selection, introducing post-transcriptional changes at structurally and functionally relevant sites. Transfer RNAs (tRNAs) undergo extensive modifications at conserved nucleotide positions, which are necessary for their proper biological function (extensively reviewed in Suzuki, 2021). Modification patterns vary across species and among tRNA isoacceptors within the same organism, reflecting selective pressures enhancing translational efficiency and codon-anticodon recognition (Berg et al., 2021; Nilsson et al., 2019).

Ribosomal RNA (rRNA) also undergoes modifications, particularly within functional domains such as tRNA binding sites. These modifications, introduced by snoRNA-guided or enzyme-catalysed processes, optimise mRNA translation (Sloan et al., 2017).

Post-transcriptional modifications enable biotechnological strategies to regulate RNA stability, modulate translation, and engineer nucleic acids for therapeutic or industrial purposes. During the COVID-19 pandemic, pseudouridine (Ψ) addressed limitations in *in vitro* transcription (IVT) technology by reducing RNA-mediated immunogenicity. Incorporation of Ψ into IVT mRNA suppresses Toll-like receptor (TLR) activation, thereby reducing innate immune responses and increasing RNA stability and translational efficiency. Ψ also alters RNA secondary structure and reduces RNase-mediated degradation. Substituting uridine with N^1-methyl-pseudouridine further enhanced these effects, contributing substantially to the efficacy of Pfizer–BioNTech and Moderna COVID-19 vaccines (reviewed in Morais et al., 2021).

Nucleotide modifications also possess applications in nanotechnology, differentially influencing RNA and DNA molecules. For instance, 5-halogenation enhances base pairing within DNA i-motifs but generally weakens these interactions in RNA. Specifically, 5-fluorination and 5-iodination stabilise DNA i-motifs, whereas in RNA, only 5-iodination exhibits a moderate stabilising effect (Rodgers et al., 2022).

Given their substantial impact on RNA structure, stability, and function, accurate methods for detecting and characterising modifications are essential to understanding their biological roles and therapeutic potential.

1.3.1 Pseudouridine (Ψ/PSU)

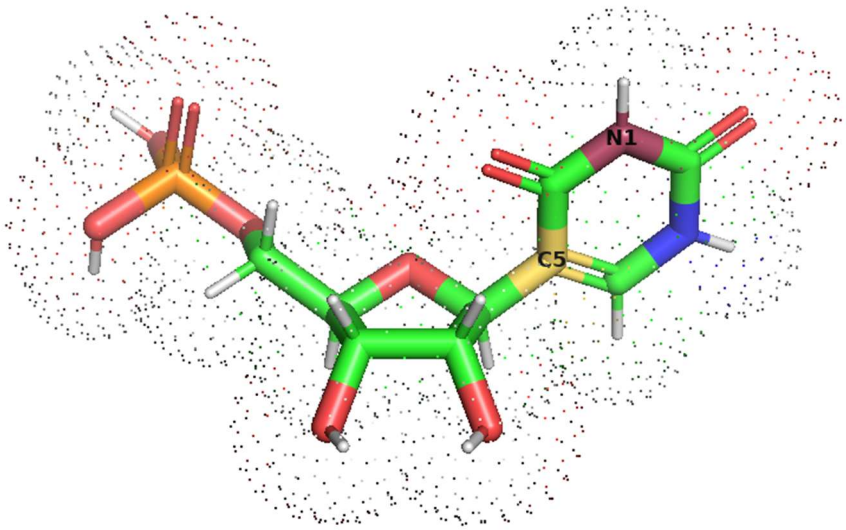


Figure 7 3D-rendering of the Pseudouridine nucleotide. RCSB PSU ideal *sdf* entry has been utilized and rendered with PyMol. N1 and C5 atoms are rendered with Bordeaux and yellow colour, respectively.

Pseudouridine (**Ψ**, **Fig. 7**) is one of the most abundant post-transcriptional RNA modifications. Ψ occurs in all major classes of functional RNAs, including tRNA, rRNA, snRNA, and mRNA (Voegele et al., 2023; Carlile et al., 2019; Yu Zhao et al., 2023; Davis et al., 1995; Charette et al., 2000).

Ψ results from the isomerisation of uridine, catalysed by pseudouridine synthase. This reaction replaces the N^1 β-glycosidic bond with a C–C glycosidic bond at C^5, flipping the uracil base (Huang et al., 2012). The resulting structure increases chemical stability and introduces an additional hydrogen bond donor on the CH edge, distinguishing Ψ from uridine and enhancing its structural contribution to RNA.

Ψ improves base stacking and promotes helix integrity, stabilising RNA duplexes in Ψ–A, Ψ–G, Ψ–U, and Ψ–C pairs. The degree of stabilisation depends on sequence context, base pair identity, and neighbouring Watson–Crick interactions (Voegele et al., 2023). In the neomycin-sensing riboswitch, Ψ exerts position-dependent effects, either stabilising local and global structures or destabilising specific local interactions. Among these effects, Ψ-A is the most stabilising doublet (Kierzek et al., 2013), where the N^1-H group of pseudouridine coordinates structural water molecules, rigidifying the sugar-phosphate backbone and reinforcing base stacking (Davis et al., 1995; Charette et al., 2000).

Ψ influences RNA folding and molecular interactions, modulating both intra- and intermolecular contacts that affect gene expression and stability. In mRNA, it reduces translational efficiency by slowing elongation and promoting amino acid substitutions. Ψ also affects codon–anticodon recognition at the ribosomal A-site, increasing the acceptance of near-cognate tRNAs (Eyler et al., 2019). In rRNA, Ψ enhances structural stability, translational fidelity, and catalytic activity. Ribosomes lacking these modifications exhibit impaired decoding centre geometry, disrupted intersubunit bridges, and defective translation fidelity (Yu Zhao et al., 2023; King et al., 2003).

Together, these findings establish pseudouridine as a structurally and functionally versatile modification present across all major RNA classes. Its influence spans from fine-tuning RNA architecture and dynamics to modulating decoding fidelity, gene expression, and ribosome performance. Given the context-dependent nature of its effects and its prevalence in both natural and engineered RNAs, pseudouridine continues to represent a critical focus in studies of RNA structure, function, and therapeutic design.

1.3.2 Dihydrouridine (D)

Dihydrouridine (D, **Fig. 8**) is an RNA modification formed by the enzymatic reduction of the double bond in uridine, catalysed by NADPH-dependent dihydrouridine synthases (DUS). It predominantly localises in the D-loop of tRNAs but also occurs in mRNA, snoRNA, and other RNA classes (Finet et al., 2022; Draycott et al., 2022).

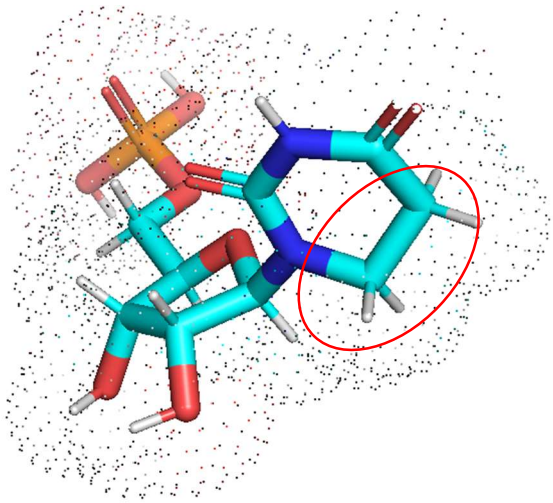


Figure 8 3D-rendering of the Dihydrouridine nucleotide. RCSB H2U ideal SDF entry has been utilized and rendered with PyMol. Additional hydrogens placed on the 5th and 6th carbons are highlighted with a red eclipse.

Dihydrouridine alters RNA sugar conformation by destabilising the C^{3}'-*endo* pucker typical of A-form helices and promoting the C^{2}'-*endo* conformation. This shift increases conformational flexibility and structural plasticity, facilitating tertiary interactions and functional adaptations (Dalluge et al., 1996; Deb et al., 2014).

In mRNAs, dihydrouridine influences interactions with the eukaryotic ribosome, modulating translation efficiency through changes in RNA dynamics (Finet et al., 2022; Draycott et al., 2022; Dixit et al., 2022).

In tRNAs, D primarily occurs at conserved positions in the D-loop (16, 17, 20, and 20A), where it promotes C^{2}'-*endo* puckering and increases local flexibility. This enhancement facilitates contacts between the D-loop and T-loop, stabilising the canonical L-shaped tRNA conformation (Dalluge et al., 1996; Deb et al., 2014).

The effects of dihydrouridine are also temperature-dependent: the C^{2}'-*endo* sugar pucker becomes more thermodynamically favourable at lower temperatures, suggesting a role in RNA adaptation to environmental conditions (Dalluge et al., 1996; Deb et al., 2014).

Together, these features define dihydrouridine as a context-dependent regulator of RNA conformation, with implications for both molecular flexibility and translational output.

1.3.3 N^6-methyladenosine (m^6A)

N^6-methyladenosine (**m^6A**, **Fig. 9**) is an internal modification found in eukaryotic RNA, influencing RNA metabolism and function by altering structure, stability, processing, and RNA–protein interactions. This alkylation is catalysed by a multicomponent methyltransferase complex. METTL3 acts as the catalytic core, while

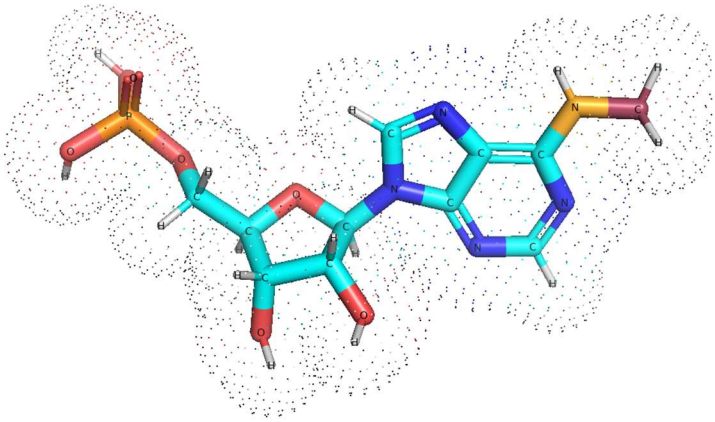


Figure 9 3D-rendering of the $N^6-methyladenosine$ nucleotide. RCSB 6MZ ideal *sdf* entry has been utilized and rendered with PyMol. Additional carbon on the N6 atom (light orange) is evidenced (bordeaux colour)

METTL14 forms a stable heterodimer with METTL3, enhancing RNA-binding specificity and methylation efficiency. Additionally, auxiliary proteins such as Wilms Tumour 1 Associated Protein (WTAP), KIAA1429, RBM15/RBM15B, and ZC3H13 assist in region-selective methylation of target RNAs.

Demethylases, or "erasers," dynamically regulate m^6A levels**.** Enzymes such as FTO (fat mass and obesity-associated protein) and ALKBH5 remove m^6A marks from RNA. This reversible modification influences RNA metabolism by recruiting specific "reader" proteins that recognise m^6A marks. Reader proteins, including members of the YTH domain-containing protein family such as YTHDF1, YTHDF2, YTHDF3, YTHDC1, and YTHDC2, as well as additional RNA-binding proteins like HNRNPA2B1 and IGF2BP, mediate effects on RNA processing, stability, splicing, nuclear export, translation efficiency, and degradation (Su et al., 2022).

N^6-methyladenosine influences RNA metabolism, impacting splicing regulation by modulating exon inclusion or exclusion, altering RNA secondary structure, and modulating RNA–protein interactions. Through these mechanisms, m^6A modification regulates RNA stability, translational efficiency, and the overall cellular RNA landscape, contributing to cell differentiation, embryonic development, and disease states, including cancer progression and metabolic disorders (Yang et al., 2018).

Apart from its metabolic and pathological implications, the structural impact of m^6A can be quantified by thermodynamical considerations. When incorporated into RNA duplexes, m^6A introduces energetic perturbations that reduce thermodynamic stability by 0.5–1.7 kcal/mol. The methylamino group rotates into the major groove, weakening base pairing (Roost et al., 2015; Shi et al., 2019).

N^6-methyladenosine enhances base stacking in single-stranded regions and stabilises RNA structure. This stabilisation is pronounced when m^6A occupies a helix end or lies adjacent to loop structures. It reinforces single-stranded dangling ends more effectively than unmethylated adenosine. Optical melting experiments show that it provides –0.43 ± 0.15 kcal/mol greater stability than analogous A dangling ends (Kierzek et al., 2021).

In addition to thermodynamic destabilisation, m^6A modifies the structural topology of RNA in ways that affect its accessibility to enzymatic and protein recognition. These structural changes influence the recruitment of RNA-binding proteins, regulate splicing outcomes, and interact functionally with other modifications, such as A-to-I editing.

N^6-methyladenosine alters RNA accessibility to heterogeneous nuclear ribonucleoproteins G (HNRNPG) and C (HNRNPC), promoting RNA-protein

interactions. These interactions induce structural changes in RNA, exposing previously buried RNA-binding motifs. These structural switches regulate RNA stability, splicing, and processing by facilitating or hindering the recruitment of specific proteins (Kierzek et al., 2021).

Moreover, m^6A in intronic regions near alternatively spliced exons correlates with exon inclusion or skipping, depending on the cellular context and splicing factors. The demethylase FTO mediates these effects, linking dynamic RNA methylation to the regulation of exon usage (Bartošovič et al., 2017; Zhu et al., 2022).

Notably, N^6-methyladenosine and adenosine-to-inosine (A-to-I) RNA editing are inversely correlated. A-to-I editing is enriched in transcripts lacking m^6A. This antagonism arises because m^6A reduces the accessibility of the ADAR enzyme to double-stranded RNA regions by enhancing local flexibility (Tassinari et al., 2021; Xiang et al., 2018).

Together, these observations establish m^6A as a chemically simple yet structurally versatile modification that reshapes RNA behaviour at multiple levels—altering stability, guiding protein recognition, modulating splicing, and interacting with other regulatory marks. Its widespread distribution and reversible nature underscore its relevance as a dynamic regulator of RNA structure and function.

1.3.4 N^1-methyladenosine (m^1A)

N^1-methyladenosine (**m^1A**, **Fig.10**) is a post-transcriptional modification that is present in different RNA classes, including mRNA and non-coding RNA. It involves methylation of the N^1 position on the purine ring of adenosine. The primary protein complex identified for catalysing m^1A in mRNA is TRMT6/61A. This complex is a heterodimer comprising TRMT6, the non-catalytic subunit responsible for substrate

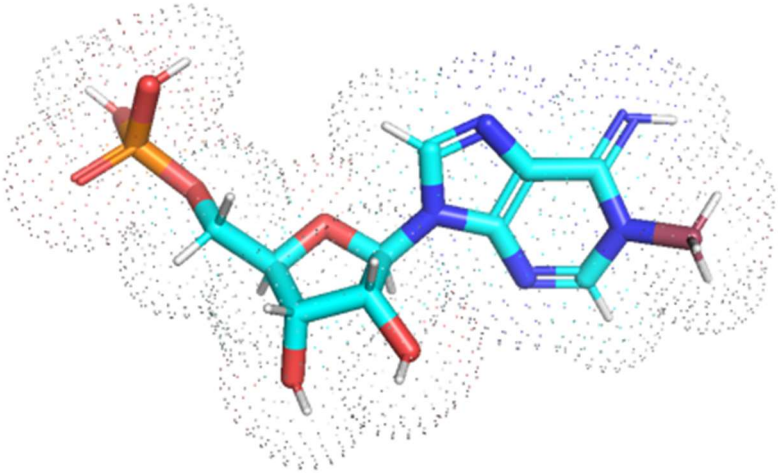


Figure 10 3D-rendering of the $N^1-methyladenosine$ nucleotide. RCSB 1MA ideal *sdf* entry has been utilized and rendered with PyMol. Additional carbon on the N1 atom is evidenced (Bordeaux colour)

recognition, and TRMT61A, the catalytic subunit containing the methyltransferase domain. Early research on yeast homologs, Gcd10p and Gcd14p, established the role of this complex in tRNA modification, later extended to human mRNA (Anderson et al., 1998; Ozanick et al., 2007).

Studies have shown that TRMT6/61A recognises specific sequence and structural motifs in mRNA, particularly the GUUCRA motif, which resembles the T-loop structure found in tRNA. This recognition determines substrate specificity, ensuring that m^{1}A is installed at appropriate sites, often enriched around the start codon and upstream of the first splice site (Li et al., 2017).

Methylation at the N^1 position disrupts Watson–Crick base pairing between adenine and uracil. The methyl group blocks the N^1 nitrogen from donating a hydrogen atom, thereby interfering with hydrogen bond formation and base pairing dynamics. This steric hindrance forces the base to adopt an alternative conformation and prevents canonical pairing.

Instead of Watson–Crick pairing, m^{1}A forms Hoogsteen base pairs. The methyl group introduces steric interference that forces the base into a T(anti)·m^{1}A(syn) configuration, enabling non-canonical interactions (Yang et al., 2008).

N^1-methyladenosine is frequently enriched near start codons and upstream of the first splice site in mRNAs. It accumulates in structured regions around translation initiation sites, and its presence correlates positively with protein production (Dominissini et al., 2016).

These structural and positional properties define N^1-methyladenosine as a conformation-altering modification that interferes with Watson–Crick pairing, promotes Hoogsteen geometry, and modulates translation efficiency through selective enrichment near start codons and splice junctions.

1.3.5 Synthetically altered nucleotide: unnatural modifications

Synthetic RNA modifications serve as tools to study and manipulate RNA structure, enabling the analysis of their biological functions and therapeutic applications. Targeting specific functional groups allows interrogation of how site-specific chemical modifications influence RNA three-dimensional architecture and intermolecular interactions (Grasby & Gait, 1994).

In addition to structural probing, modified nucleotides have been applied in biophysical studies, particularly as spectroscopic labels for real-time analysis of RNA dynamics. For example, fluorine-modified nucleotides have been used in high-resolution NMR

spectroscopy to analyse secondary structure and folding equilibria (Wachowius & Höbartner, 2010).

Halogenated and thiolated RNA analogues have been investigated as radio-sensitising agents in cancer therapy. Modified uridine derivatives, such as 5-bromo-4-thiouracil and 5-bromo-4-thio-2′-deoxyuridine, exhibit synergistic radio sensitising effects. Bromination at the C^5 position enhances radiosensitivity by increasing electron affinity, promoting dissociative electron attachment (DEA), and generating reactive radicals that induce DNA strand breaks under ionising radiation. Thiolation at the C^4 position amplifies this effect by stabilising electron-rich intermediates and enhancing radiation reactivity. This combination selectively enhances activity in tumour microenvironments, such as hypoxia, increasing therapeutic efficacy and illustrating the applicability of synthetic RNA modifications in radiotherapy (Izadi et al., 2023).

Altogether, synthetic RNA modifications provide a controlled framework to dissect structure–function relationships, optimise spectroscopic analysis, and expand therapeutic strategies through rational chemical design.

1.4 Experimental approaches for RNA structural determination and associated challenges

RNA physicochemical properties limit structural determination, as reflected by significantly lower RNA structures deposited annually in the Research Collaboratory

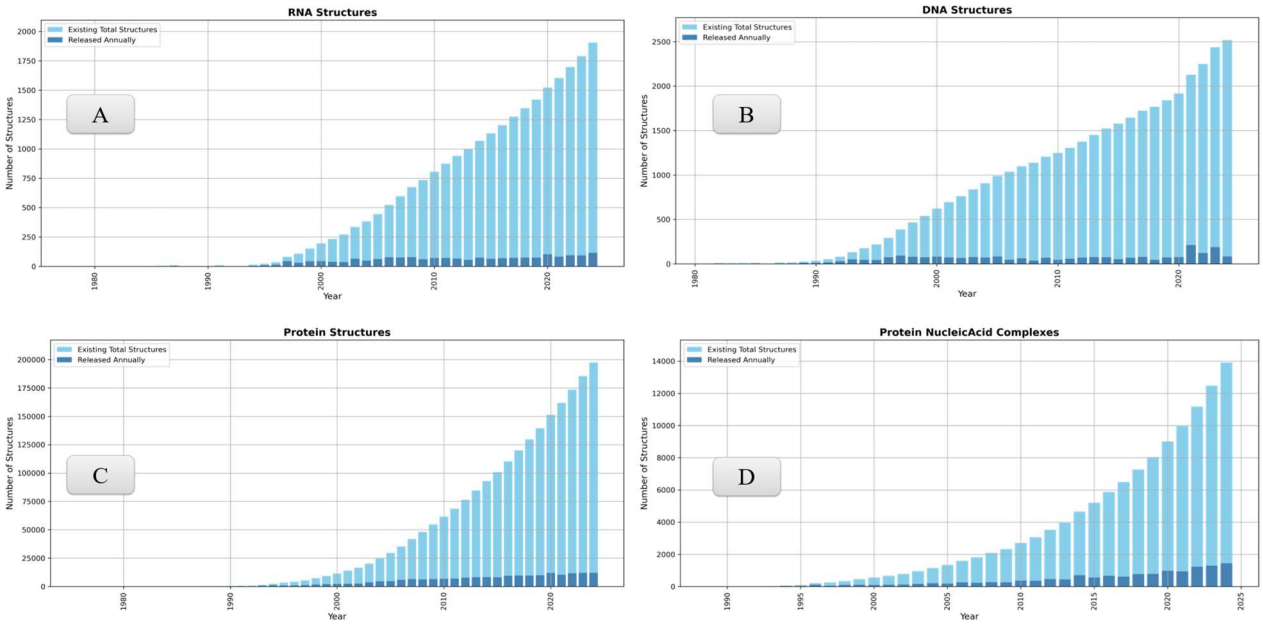


Figure 11 RCSB available structures. (**A**) RNA only structures. (**B**) DNA only structures. (**C**) Protein only structures. (**D**) Protein Nucleic acids complexes. Data are publicly available at PDB Statistics: https://www.rcsb.org/stats/

for Structural Bioinformatics (RCSB) Protein Data Bank compared to proteins (**Fig. 11**, 10^2 vs 10^4 in 2024).

The primary structure of RNA may initially appear simple, as it consists of only four nucleotide monomers. Although RNA is composed of only four nucleotide monomers, its limited chemical diversity reduces its ability to form strong lattice contacts. The resulting crystals are loosely packed, with solvent contents often exceeding 50%. Such high solvent levels hinder diffraction resolution, further complicating structural determination by X-ray crystallography (Jackson et al., 2023).

The phosphate backbone of RNA carries a negative charge, leading to significant electrostatic repulsion among phosphate groups. This contrasts with protein side chains, which provide a broader range of charges, polarities, and hydrophobicity. Divalent and trivalent metal ions (e.g., Mg^{2+}, Mn^{2+}, and Tb^{3+}) reduce electrostatic repulsion, stabilising RNA three-dimensional structure through charge neutralisation and supporting proper folding (Fingerhut et al., 2021).

RNA generates stronger electrostatic fields than DNA, attracting more cations and retaining fewer anions. RNA helices adopt the A-form conformation, where phosphate groups are positioned closer together along and across the major groove. The arrangement in A-form helices results in shorter phosphate-phosphate distances (5.65 Å in RNA versus 6.62 (~7) Å in DNA), contributing to a higher linear charge density in RNA (Gebala et al., 2019). Therefore, fine-tuning Mg^{2+} concentrations in crystallisation buffers supports lattice stability and minimises non-specific aggregation (Jackson et al., 2023).

Due to the limitations of crystallographic approaches, solution-based techniques offer complementary means to investigate RNA folding dynamics, intermediate states, and conformational heterogeneity. In these contexts, molecular interactions often dominate over ionic effects. RNA folding proceeds through metastable intermediate states that lack extensive long-range interactions and are primarily stabilised by local contacts rather than electrostatic neutralisation.

Small-angle X-ray scattering (SAXS) has shown that increased Mg^{2+} concentration minimally affects intermediate compaction, underscoring the role of molecular interactions in shaping the folding landscape (Baird et al., 2010).

Single-molecule fluorescence resonance energy transfer (smFRET) addresses these limitations by enabling real-time visualisation of RNA folding pathways. smFRET reveals transient intermediates and conformational heterogeneity within individual molecules (Roy et al., 2008; Bokinsky et al., 2003; Zhuang et al., 2000). It also tracks

catalytic activity alongside structural transitions, linking RNA dynamics to function (Helm et al., 2009). Therefore, the coexistence of multiple RNA conformations during crystallisation may lead to disordered or poorly diffracting crystals.

NMR spectroscopy provides an alternative for studying RNA structures in solution, enabling the observation of folding dynamics under near-physiological conditions. Unlike X-ray crystallography, NMR does not require crystallisation, making it well-suited for small RNAs. However, the technique is limited to RNAs shorter than ~50 nucleotides due to signal overlap and chemical shift degeneracy (Barnwal et al., 2017).

Cryo-electron microscopy (cryo-EM) has become valuable for visualising large RNA-protein complexes and long RNAs. By capturing molecules in vitrified ice, cryo-EM avoids crystallisation artefacts and preserves near-native conformations. However, RNA's low contrast, small size, and structural flexibility often require associated proteins or ligands to enhance stability and facilitate reconstruction (Langeberg et al., 2023).

1.4.1 X-Ray crystallography

X-ray crystallography measures the angles and intensities of X-ray diffraction patterns in crystals to generate three-dimensional electron density maps, revealing atom positions, chemical bonds, and structural features. It is widely applied to determine the three-dimensional structures of biological macromolecules, including proteins, nucleic acids, and their complexes, with comparatively fewer size limitations than solution-based methods (Yonath et al., 2011; Ennifar et al., 2013; Atalay et al., 2022).

RNA structure determination by X-ray crystallography poses specific challenges, including low yields of properly folded RNA, limited sequence diversity, and difficulty in forming stable crystal contacts. In addition, the scarcity of effective phasing methods makes RNA crystallisation a rate-limiting step in structural analysis (Jackson et al., 2023).

Approaches to address these challenges include native RNA purification, engineered crystallisation modules, and protein-assisted phasing methods. For example, the incorporation of GAAA tetraloop/tetraloop receptor interactions has been shown to improve crystal growth and diffraction quality (Ferré-D'Amaré et al., 1998).

Directed heavy atom soaking strategies enhance RNA phasing and facilitate high-resolution structural interpretation (Keel et al., 2007). Recent techniques, such as serial femtosecond crystallography and 3D electron crystallography, broaden the use of X-ray crystallography by enabling high-resolution data collection from microcrystals,

thereby reducing the requirement for large, well-ordered crystals (Espinosa et al., 2017).

Together, these strategies refine RNA crystallographic workflows, mitigating limitations associated with molecular yield, folding, and phase determination, and expanding the applicability of X-ray crystallography to structurally diverse RNA targets.

1.4.2 Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is widely applied to investigate RNA structures in solution. It provides atomic-resolution information on RNA dynamics and conformational flexibility, complementing crystallographic data.

NMR detects signals from *nuclei* with non-zero magnetic moments, such as ^{1}H, ^{13}C, ^{15}N, and ^{31}P. When exposed to a static magnetic field, these nuclei establish an equilibrium spin population required for signal detection. Radiofrequency pulses excite nuclei, which transition between spin states and emit electromagnetic radiation upon relaxation. This emitted signal forms the basis of NMR spectroscopy.

While NMR provides atomic-resolution data in aqueous environments, RNA flexibility and the size of RNA-protein complexes introduce challenges for traditional solution-state NMR. Large RNA complexes exhibit spectral crowding due to the high number of NMR-active *nuclei,* reducing resolution and complicating resonance assignments. Slow tumbling in solution enhances transverse relaxation (T_2), broadening linewidths and reducing sensitivity, particularly for *nuclei* with low natural abundance. Conformational heterogeneity further broadens signals, complicating the detection of distinct structural features. Due to these limitations, solution-state NMR techniques typically struggle to resolve structures beyond ~30–50 kDa effectively.

Solid-state NMR (SSNMR) addresses these limitations, improving resolution and extending applicability to larger and more rigid RNA complexes. It can be integrated with cryo-electron microscopy (cryo-EM) and small-angle X-ray scattering (SAXS) to link high-resolution structure with solution-phase dynamics. While cryo-EM provides static high-resolution structures, SSNMR complements it by capturing molecular flexibility, enabling a combined strategy for RNA structure characterisation (Daubner et al., 2015; He et al., 2024).

Altogether, solid-state and solution-state NMR complement crystallographic and cryo-EM approaches, expanding the toolkit available for characterising RNA conformational landscapes under near-physiological conditions.

1.4.3 Cryo-electron microscopy (cryo-EM)

Cryo-electron microscopy (cryo-EM) enables the visualisation of biomolecules at sub-nanometre resolution and has become increasingly applied in RNA structural biology. The method involves vitrifying RNA molecules or RNA–protein complexes in amorphous ice to preserve native structure under cryogenic conditions, without the need for staining or chemical fixation. A thin sample layer is applied to a grid and plunge-frozen in liquid ethane cooled by liquid nitrogen, preventing ice crystal formation. Subsequent imaging at cryogenic temperatures produces 2D projections, which are computationally aligned to reconstruct 3D structures (Zhang et al., 2019).

Cryo-EM allows structural analysis under near-native conditions, avoiding crystallisation requirements. It is particularly useful for large RNA–protein assemblies such as the ribosome, where molecular mass enhances contrast. However, small RNAs remain challenging targets due to low electron density, conformational flexibility, and low molecular weight. These factors limit particle alignment and result in blurred reconstructions through ensemble averaging. The conformational heterogeneity originates from the rotational freedom of the ribose–phosphate backbone and weak tertiary contacts, resulting in multiple coexisting structural states (Ma et al., 2022).

Strategies such as RNA oligomerisation via engineered motifs (e.g., kissing loops) have improved resolution. In one example, the ROCK method enabled cryo-EM analysis of the Tetrahymena group I intron at 2.98 Å resolution (Liu et al., 2022).

Integrating cryo-EM with NMR spectroscopy and molecular dynamics simulations improves structure determination for small RNAs, such as the 30 kDa HIV-1 dimerisation signal. These molecules are difficult to resolve due to conformational mobility and low mass. NMR provides high-resolution local structural data, while simulations model folding and conformational transitions. This approach revealed stabilising base-pairing interactions and conformational shifts relevant to genome packaging. In specific cases, it has enabled sub-nanometre reconstructions of small RNA structures, contributing to therapeutic model development (Zhang et al., 2019).

Cryo-EM-informed molecular dynamics simulations model tRNA translocation across the ribosome. By combining medium-resolution cryo-EM maps with atomic-resolution X-ray structures, intermediate states can be reconstructed. This approach captures intersubunit rotations and domain rearrangements, mapping conformational transitions beyond equilibrium timescales. It enables detailed analysis of ribosomal mechanochemistry and provides a transferable framework for studying other dynamic molecular machines (Vaiana et al., 2014).

Despite methodological progress, challenges in resolution and structural heterogeneity persist. Recent improvements include phase plates and machine learning-based reconstruction. Phase plates increase image contrast and signal-to-noise ratio, allowing visualisation of smaller motifs and subtle features. Reported resolutions range from 2.6 to 3.3 Å for replication–transcription complexes, with local resolution at the active site reaching 2.2 Å.

Machine learning algorithms improve cryo-EM workflows by automating particle picking, enhancing 3D reconstruction, and refining resolution. These tools support the localisation of functionally relevant RNA domains and are increasingly applied to RNA–protein assemblies in therapeutic discovery (Malone et al., 2023).

Altogether, the combination of cryo-EM, computational modelling, and machine learning has expanded the applicability of structural analysis to RNA molecules previously inaccessible to high-resolution techniques.

1.4.4 Circular dichroism (CD) spectroscopy

Circular dichroism (CD) is an optical phenomenon in which chiral molecules differentially absorb left- and right-circularly polarised light. In nucleic acids, CD is used to investigate secondary structure, helicity, and base stacking by monitoring changes in electronic transitions and their coupling.

The underlying mechanism of CD signals in nucleic acids is exciton coupling between stacked chromophores. This phenomenon arises from Coulombic interactions between electronic transition dipoles, producing excitonic energy splitting into two states: a lower-energy (red-shifted) and a higher-energy (blue-shifted) transition. The $\pi \rightarrow \pi^*$ transitions of nitrogenous bases serve as the optical basis for this effect, involving excitation from bonding π orbitals to antibonding π^* orbitals. The π^* orbitals contain nodes and represent higher potential energy states, thereby reducing electronic stability relative to π orbitals.

Exciton splitting is analogous to molecular orbital formation: interaction of chromophore wavefunctions yields symmetric (bonding-like) and antisymmetric (antibonding-like) states. The magnitude and nature of this splitting depend on the relative orientation, distance, and transition energies of the interacting chromophores. When chromophores possess identical transition energies, degenerate exciton coupling produces a bisignate CD couplet with symmetric energy splitting—such as in G-quadruplexes. In contrast, non-degenerate coupling, as seen in DNA-ligand complexes,

arises when chromophores differ in transition energy, leading to asymmetric spectral profiles.

The resulting CD spectrum reflects excitonic energy splitting as a bisignate signal. According to the exciton chirality rule, a positive band at shorter wavelengths followed by a negative band at longer wavelengths indicates right-handed helicity, as in B-DNA. The inverse pattern identifies left-handed structures, such as Z-DNA. CD spectral features encode information about stacking geometry and helical handedness. The bisignate pattern reflects wavefunction symmetry and electrostatic interactions between stacked bases, allowing indirect inference of helical organisation and conformational changes (Šmidlehner et al., 2018).

RNA CD spectra show characteristic maxima and minima that reflect conformational responses to environmental factors and molecular interactions. Right-handed helices, such as A-form and B-form, display a positive maximum in the vacuum ultraviolet (VUV) near 185 nm and a negative band in the far-ultraviolet (FUV) around 200–210 nm. A-form helices show increased negative ellipticity and a stronger positive band near 265 nm, which is flat or diminished in B-DNA. RNA primarily adopts the A-form. Spectral changes between 260–268 nm reflect alterations in base stacking and hydrogen bonding, including transitions between Watson–Crick and Hoogsteen base pairs. Temperature-dependent variations in this band indicate helix stability, with decreasing ellipticity corresponding to thermal denaturation (Le Brun et al., 2020).

Z-DNA, a left-handed DNA conformation, displays distinct CD signatures. In this structure, purines adopt a syn-conformation while pyrimidines remain in anti-conformation. The alternation in glycosidic bond angles generates a zigzag backbone, producing a strong negative band at 195 nm and a positive maximum near 180 nm (Holm et al., 2010; Riazance et al., 1985). These spectral features also serve to identify left-handed RNA helices with Z-like topology (Le Brun et al., 2020). CD spectroscopy is highly sensitive to environmental parameters such as temperature, pH, and ionic strength. These factors influence nucleic acid folding, altering helicity and base stacking. Exciton coupling remains the primary physical mechanism underlying CD spectra, with energy splitting due to $\pi \rightarrow \pi^*$ transitions between stacked bases giving rise to bisignate signals. Changes in ellipticity under varying conditions reflect modifications in secondary structure and hydrogen bonding networks (Dijk et al., 2010).

The application of CD spectroscopy extends to nucleic acid–ligand interactions. For instance, intercalating agents such as ethidium bromide (EB) provide structural

information on nucleolar DNA and RNA. EB intercalates between base pairs in double-helical regions, generating a characteristic positive CD band in the 300–360 nm region, peaking near 308 nm. While electrostatic interactions with the phosphate backbone also occur, they contribute less to EB binding.

Reduced EB-induced ellipticity in nucleolar RNA and DNA reflects their compact organisation and strong protein associations, which limit helical folding and intercalation. Enzymatic digestion experiments indicate that DNA contributes most to EB-induced CD signals, whereas RNA produces a weaker ellipticity. These results suggest reduced base pairing and altered secondary structures in nucleolar RNA (Huang & Baserga, 1976). Despite minor artefacts such as light scattering, CD spectroscopy reliably detects structural transitions and accessibility changes under physiological and pathological conditions.

Unlike crystallography or NMR, CD spectroscopy does not resolve atomic-level structure. However, it remains a valuable tool for rapid structural characterisation in solution. Its ability to detect conformational shifts makes it particularly suited for screening RNA structural responses to chemical or physical perturbations.

Although traditional CD instrumentation is limited in the far-UV region, this restriction is overcome by synchrotron radiation circular dichroism (SRCD). SRCD enables measurements below 190 nm, a range critical for detecting nucleic acid secondary structures. It offers enhanced signal-to-noise ratios and requires smaller sample volumes and concentrations compared to conventional sources (Nielsen et al., 2005). SRCD supports the structural analysis of rare, unstable, or poorly folded nucleic acid species under diverse experimental conditions. This capability has proven particularly useful in characterising modified RNAs, low-abundance regulatory elements, and non-canonical structures (Cappannini et al., 2022).

1.5 Detection of RNA modification

Just as RNA structure determination is hindered by its physicochemical properties, the detection of RNA modifications presents additional challenges due to their chemical diversity, dynamic regulation, and often sub-stoichiometric occurrence (Helm & Motorin, 2017). Many detection methods are susceptible to artefacts and side products introduced during sample processing. For example, reverse transcriptase may misincorporate nucleotides at modified positions, leading to overestimated modification levels and compromising detection reliability (Sas-Chen et al., 2019; Potapov et al., 2018).

High-throughput sequencing technologies have expanded the capacity to detect RNA modifications but remain prone to artefacts. Misreads and dropouts complicate the identification of post-transcriptional changes, highlighting the need for orthogonal validation techniques. Mass spectrometry and chemical-based methods address these limitations by providing independent confirmation of specific modifications (Helm & Motorin, 2017; Addepalli & Limbach, 2011). These approaches allow localisation of modified nucleotides at single-nucleotide resolution, including pseudouridine and N^6-methyladenosine (m^6A), and complement sequencing-based data (Jora et al., 2018).

Some apparent sequencing errors may instead reflect true biological modifications. For example, nucleotide substitutions and 3′-end extensions frequently signal post-transcriptional changes in small RNAs such as tRNAs and miRNAs (Ebhardt et al., 2009; Bacusmo et al., 2018). Antibody-based detection methods, including immuno-northern blotting, are also commonly used but can suffer from cross-reactivity. Antibodies designed for specific modifications may bind non-selectively to chemically similar groups, producing false-positive signals. Ensuring detection specificity therefore requires rigorous validation (Mishima et al., 2015).

To overcome resolution and cross-reactivity constraints, alternative platforms have been developed. In contrast to reverse transcription- or antibody-based approaches, which infer modifications indirectly through signal patterns or binding elements, mass spectrometry and nanopore sequencing detect modifications more directly. Nanopore sequencing, especially when integrated with machine learning algorithms, has shown improved accuracy and single-base resolution. Recent implementations report accuracy rates above 90% for certain modifications, including m^6A and pseudouridine (Liu Q. et al., 2019). CRISPR-based detection platforms similarly reduce reliance on chemical conversion steps, minimising artefactual signals and expanding modification detection capabilities (Furlan et al., 2021; Liu Q. et al., 2019).

To assess methodological advances and current limitations, the following section reviews strategies for detecting RNA modifications, focusing on pseudouridine, N^6-methyladenosine (m^6A), N^1-methyladenosine (m^1A), and dihydrouridine—each discussed in detail throughout this thesis.

1.5.2 Detection of N^6-methyladenosine

The identification of N^6-methyladenosine (m^6A) relies on diverse approaches differing in resolution, technical complexity, and chemical principles. One of the earliest methods, m^6A-seq, combines immunoprecipitation using m^6A-specific antibodies with next-generation sequencing (NGS). This technique enriches for methylated RNA

fragments, enabling transcriptome-wide mapping at moderate resolution. However, it does not permit single-nucleotide localisation (Dominissini et al., 2012).

To achieve nucleotide-level precision, m^{6}A-REF-seq employs m^{6}A-sensitive endoribonucleases such as MazF and ChpBK, which cleave unmethylated RNA at specific motifs. MazF, for example, cuts RNA at unmethylated ACA sites, leaving methylated motifs intact. Sequencing the resulting fragments enables the identification of m^{6}A sites based on cleavage resistance. Treatment with demethylases such as FTO further enhances mapping specificity. This antibody-free method avoids batch effects and low reproducibility and has revealed conserved m^{6}A enrichment near stop codons across tissues and species (Zhang et al., 2019).

Alternative chemical labelling approaches also provide base-resolution detection. m^{6}A-label-seq substitutes m^{6}A with *N*^{6}-allyladenosine (a^{6}A) using a modified S-adenosylmethionine donor derived from Se-allyl-L-selenohomocysteine. Iodination converts a^{6}A into a cyclised form (cycA), which induces misincorporation events during reverse transcription. These events are then detected through sequencing. While this method avoids antibody-based enrichment and achieves single-nucleotide resolution, limitations include low labelling efficiency and mild cellular stress. Ongoing efforts aim to engineer more efficient methyltransferases to overcome these constraints (Shu et al., 2022).

DART-seq achieves single-nucleotide mapping by fusing the m^{6}A-binding YTH domain to APOBEC1 cytidine deaminase. The YTH domain binds m^{6}A sites, positioning APOBEC1 to catalyse cytidine-to-uridine conversions at neighbouring residues. These mutations are subsequently identified through sequencing. DART-seq does not require immunoprecipitation, reducing variability and RNA input requirements. Variants such as the D422N mutant in the YTH domain improve detection sensitivity. An *in vitro* adaptation extends this approach to low-abundance RNA samples, expanding its utility in studying m^{6}A dynamics (Zhu et al., 2022).

NOseq offers a distinct chemical strategy that bypasses antibody or protein-based recognition. Nitrous acid selectively deaminates unmethylated adenosines to inosines, which are read as guanosines during reverse transcription, while m^{6}A residues remain unmodified. This differential conversion provides single-nucleotide resolution. NOseq has validated known m^{6}A sites, such as those in 18S rRNA and MALAT1, and revealed new m^{6}A residues in *Drosophila* mRNAs, enabling site-specific analyses and method validation (Werner et al., 2020).

While the methods described above focus on localised detection, global m^{6}A abundance can be assessed using immune-enzymatic quantification. m^{6}A-ELISA

detects global levels by immobilising mRNA on microplates, applying anti-m^{6}A antibodies, and measuring a colorimetric signal proportional to methylation content. Optimisation strategies, including the use of m^{6}A-deficient RNA as a blocking control, reduce non-specific binding. The method has been applied to dynamic m^{6}A profiling in systems such as yeast and embryonic stem cells. Although it lacks nucleotide or transcript specificity, m^{6}A-ELISA offers a rapid and cost-effective solution for estimating global methylation changes (Ensinck et al., 2023).

The ongoing development of m^{6}A detection techniques reflects the necessity of selecting methods based on the research objective, resolution requirements, and available resources. Each approach provides complementary insights, from transcriptome-wide mapping to site-specific and quantitative analyses.

1.5.3 Detection of N^1-methyladenosine

Like *N^6-methyladenosine*, techniques for targeting *N^1-methyladenosine* have been developed exploiting the unique chemical properties of this post-transcriptional modification.

Ligation-assisted differentiation enables single-base resolution detection of *N^1-methyladenosine* in mRNA by exploiting its reduced ligation efficiency. Specific oligonucleotides hybridise to regions flanking the m^{1}A site, serving as templates for ligation. T3 DNA ligase discriminates between m^{1}A and unmodified adenosine, enabling precise site-specific detection. Quantification of ligation products via real-time PCR enables the evaluation of m^{1}A levels. The method has been validated by detecting m^{1}A at position 1674 of BRD2 mRNA in HEK293T cells, showing reduced m^{1}A levels in lung carcinoma compared to normal tissues, suggesting a link between m^{1}A and cancer pathogenesis (Ding et al., 2020).

The RNA-targeting CRISPR-Cas13a system is a sensitive and specific method to detect m^{1}A modifications. Cas13a's "collateral cleavage" activity is reduced by mismatches caused by m^{1}A, leading to decreased fluorescence compared to unmodified RNA. The method involves crRNAs (CRISPR RNA) complementary to target RNA regions containing m^{1}A, guiding Cas13a to the site. Activation triggers cleavage of a fluorescent reporter, with lower fluorescence indicating the presence of m^{1}A. This approach has been used to quantify m^{1}A levels and to analyse demethylation dynamics using AlkB demethylase. The fluorescence-based platform is time-efficient and can be applied in portable formats, complementing existing RNA modification analysis techniques (Chen et al., 2019).

In parallel with synthetic biology approaches, traditional immunochemical techniques remain relevant for m^{1}A detection at the RNA species level. Immuno-Northern Blotting

(INB) detects RNA modifications, including m^1A, using nucleoside-specific antibodies. RNA is separated by denaturing electrophoresis, transferred to a nylon membrane, cross-linked with UV light, and probed with primary antibodies specific to modifications such as m^1A, m^6A, or pseudouridine. A chemiluminescent secondary antibody enables visualisation. INB preserves RNA integrity without requiring fragmentation, distinguishing modified RNA species across samples and conditions. Applications include detecting m^1A-modified tRNA and stress-induced tRNA fragments in mammalian cells (Mishima et al., 2015).

RNA bisulfite sequencing (RBS-Seq) enables transcriptome-wide detection of RNA modifications, such as m^1A, 5-methylcytidine (m^5C), and pseudouridine at single-nucleotide resolution. For m^1A, bisulfite treatment induces a Dimroth rearrangement, converting m^1A to m^6A under alkaline conditions and elevated temperatures. RBS-Seq enables the detection of multiple RNA modifications within the same molecule, facilitating comprehensive analyses of modification landscapes. It has been used in human cell lines to identify novel modification sites and gain insights into RNA modification patterns (Khoddami et al., 2019).

In contrast to chemically induced rearrangements, enzymatic demethylation enables an alternative strategy for resolving RNA methylation marks. AlkB-Facilitated RNA Demethylation Sequencing (ARM-Seq) uses AlkB enzymes to demethylate RNA modifications, including m^1A, m^3C, and m^1G, which disrupts the Watson-Crick base pairing. Demethylation ensures a broader representation of modified RNA sequences in small RNA libraries. The method involves purifying AlkB from *E. coli*, incubating RNA samples with AlkB under optimised conditions, and applying standard small RNA library preparation protocols. Comparing sequencing data from AlkB-treated and untreated samples enables the identification of methylated sites based on differential read abundance. ARM-Seq has been validated for mapping methylation in tRNAs and tRNA-derived fragments, allowing the detection of changes in RNA methylation status under different cellular conditions (Hrabeta-Robinson et al., 2017).

Recent developments have expanded the toolkit for detecting N^1-methyladenosine with increased precision and throughput. Demethylase-Assisted Detection of m^1A (DA-m^1A) utilises demethylase treatment to remove m^1A modifications prior to downstream analysis selectively. The approach simplifies detection by generating differential signals between treated and untreated RNA, allowing for efficient and cost-effective site-specific analysis (Xiong et al., 2022).

Nanopore direct RNA sequencing provides a single-molecule platform for m^1A detection, eliminating the need for chemical conversion or amplification. Recent

studies have introduced computational pipelines capable of identifying m^1A-induced current signal deviations, enabling detection at single-base resolution within full-length transcripts (Chen et al., 2024). This technique enables real-time, label-free analysis and complements existing approaches that rely on reverse transcription or antibody enrichment.

Together, these methods represent complementary strategies for detecting N^1-methyladenosine, each with distinct advantages in resolution, throughput, and technical complexity. The choice of method depends on the biological question, the RNA context, and the required precision, underscoring the importance of methodological awareness in the study of RNA modifications.

1.5.4 Detection of Pseudouridine (Ψ)

Pseudouridine (Ψ) detection has advanced significantly through a range of innovative techniques. A qPCR-based method facilitates rapid and radiolabel-free detection of pseudouridine at specific sites. This approach involves chemical labelling and reverse transcription, producing distinct melting curves during qPCR analysis (Lei & Yi, 2017).

While qPCR-based detection provides site-specific information through melting curve analysis, mass spectrometry-based strategies offer chemical-level resolution through direct analysis of modified nucleotides. Mass spectrometry-based approaches, such as LC-MS/MS, enable the quantification and localisation of pseudouridine by detecting fragment ions unique to Ψ. Pseudouridine, although mass-silent in its unmodified form, exhibits unique glycosidic bond stability due to its *C-C* bond, unlike the typical *C-N* bond. The detection and localisation of pseudouridine can be achieved using LC-MS/MS combined with chemical derivatisation or specific fragmentation pathways, as the stability deriving from its glycosidic bond produces distinct fragmentation patterns during tandem MS. Beyond identifying the presence of modifications, MS pinpoints their exact positions within nucleic acid sequences. Techniques such as RNase mapping, which involve enzymatic digestion of RNA into smaller fragments, enable LC-MS/MS analysis to determine the location of modifications. Diagnostic ions or mass shifts in the spectra reveal the specific nucleotide carrying the modification and its sequence context (Addepalli & Limbach, 2011).

A method combining 10–23 DNAzyme activity with rolling circle amplification (RCA) enables the detection of pseudouridine with nucleotide-level discrimination, eliminating the need for chemical or radioactive labelling. The DNAzyme selectively cleaves uridine (U) but spares pseudouridine due to its unique structure. RNA hybridises to DNAzymes immobilised on streptavidin-coated beads, and cleavage

occurs in the presence of Mg^{2+} ions. The resulting fragments are ligated with padlock probes, and rolling circle amplification (RCA) amplifies the signal for fluorescence-based detection, enabling the differentiation of uridine-pseudouridine. Optimisation of complementary regions, use of SplintR Ligase for efficient ligation, and refined reaction conditions allow the detection of pseudouridine at RNA concentrations as low as 5 nM. This method is label-free and supports transcriptome-wide studies of RNA structure and function (Wei Li et al., 2019).

HydraPsiSeq maps pseudouridines at single-nucleotide resolution using a chemical cleavage strategy. Hydrazine selectively modifies unmodified uridines by opening their ring structure, while pseudouridines remain intact due to their stability. Following aniline treatment, subsequent cleavage of RNA occurs at the abasic sites, leaving the pseudouridine sites protected and identifiable during sequencing. RNA fragments from hydrazine/aniline treatment are processed through adapter ligation, PCR, and sequencing. Pseudouridines are distinguished by their resistance to cleavage, with background signals normalised and quantified using scoring systems like PsiScore. HydraPsiSeq detects pseudouridine with high sensitivity from RNA inputs as low as 10–50 ng. Applied to human rRNAs and tRNAs, HydraPsiSeq has revealed dynamic changes in pseudouridylation during processes such as chondrogenic differentiation and uncovered conserved and variable pseudouridylation patterns across different cell types, supporting investigations into the structural and regulatory roles of pseudouridine in RNA biology (Marchand et al., 2020).

In contrast to cleavage-based sequencing approaches, derivatisation coupled with high-resolution mass spectrometry (MS) enables pseudouridine identification based on mass shifts. MALDI-MS combined with CMC (N-cyclohexyl-N'-β-(4-methylmorpholinium) ethylcarbodiimide) derivatisation detects pseudouridine by exploiting its N^3 reactivity. CMC forms a stable adduct at pseudouridine sites, resistant to alkaline conditions. RNA is enzymatically digested into smaller fragments, selectively modifying pseudouridine, while unmodified nucleotides lose the adduct. Derivatised fragments show a mass shift of 252 Da, enabling pseudouridine identification via MALDI-MS. Validated on bacterial RNA, such as *E. coli* tRNA, the method is sensitive enough for small RNA samples. However, challenges such as incomplete derivatisation and interference from other modifications require careful optimisation. Despite these limitations, this method combines chemical selectivity with the mass-based resolution to characterise pseudouridine and its structural context (Durairaj & Limbach, 2008).

Recent strategies have expanded pseudouridine detection beyond derivatisation and sequencing-coupled cleavage approaches. Nanopore direct RNA sequencing enables

the detection of pseudouridine by analysing characteristic U-to-C base-calling errors and disruptions in ionic current as RNA molecules pass through the nanopore sensor. This method allows detection without chemical treatment, maintaining the native structure and full-length context of transcripts. By preserving isoform-level information, nanopore-based analysis supports the study of pseudouridylation in transcript processing and gene expression regulation (Rodell et al., 2023).

A complementary strategy involves *in vivo* deuterium labelling of uridine residues, followed by mass spectrometric analysis. When cells are cultured in media containing deuterated uridine, enzymatic pseudouridylation results in specific mass shifts that are detectable via liquid chromatography-tandem mass spectrometry (LC-MS). This method enables the site-specific determination of Ψ content without derivatisation. It allows for the metabolic tracing of pseudouridylation events in cellular RNA (Yamaki et al., 2020).

Together, these detection strategies utilise chemical specificity, enzymatic activity, and direct sequencing to identify pseudouridine with varying degrees of resolution, throughput, and input requirements, thereby enabling the exploration of RNA modification dynamics across biological systems.

1.5.5 Detection of Dihydrouridine

Dihydrouridine (D) is a post-transcriptional modification characterised by the saturation of the fifth and sixth carbon atoms of uridine. Traditional nucleoside analysis methods, such as high-performance liquid chromatography (HPLC) coupled with ultraviolet (UV) detection, are not suitable for detecting D due to its non-aromatic structure and lack of a chromophore, which limits detection sensitivity. To overcome these constraints, fluorescence-based approaches have been developed.

Boronic acid-based fluorescent sensors exploit the reactivity of the *2′, 3′*-cis-diol group in D. Upon interaction, dihydrouridine forms boronate esters, resulting in measurable fluorescence. This behaviour distinguishes D from nucleosides like uridine, which typically reduce fluorescence through stacking interactions. The application of various boronic acid derivatives has produced fluorescence intensity changes sufficient for D quantification in biological samples. Such approaches have also been investigated for their potential in diagnostic applications, including studies that correlate urinary D levels with disease states, such as cancer (Luvino et al., 2006).

D-seq enables the detection of D at single-nucleotide resolution. The method is based on sodium borohydride-mediated reduction of D to dihydrouridine, which stalls reverse transcriptase (RT) at the nucleotide immediately downstream of the modified site. These termination events are then detected through next-generation sequencing. D-seq

has been applied to tRNAs, small nucleolar RNAs (snoRNAs), and messenger RNAs (mRNAs). It has confirmed known D sites, identified additional positions linked to dihydrouridine synthase (DUS) activity, and revealed modification patterns in conserved RNA structures. In mRNA, D-seq has mapped over 100 dihydrouridine sites in both coding and non-coding regions (Draycott et al., 2022).

Rho-seq follows a similar reduction principle but introduces a rhodamine-based chemical tag to label the reduced D residues. The bulky adduct causes steric hindrance, generating RT arrest sites, which are then analysed by sequencing. Rho-seq has been applied in yeast and mammalian systems, identifying dihydrouridine at single-nucleotide resolution in various RNA classes, including mRNA. Modifications have been detected in transcripts encoding cytoskeletal proteins and meiosis-related genes. Perturbation of DUS enzymes or mutations at D-modified positions alters translation efficiency, supporting a role in translation regulation (Dixit et al., 2022).

D-seq and Rho-seq are complementary. Their combined application enables the characterisation of D distribution across RNA species, improving the interpretation of its biological function in RNA structure, processing, and translation.

1.5.6 Oxford Nanopore Technologies

The increasing role of nanopore-based approaches in detecting RNA modifications at single-nucleotide resolution warrants a focused analysis of Oxford Nanopore Technologies (ONT). Oxford Nanopore Technologies detects changes in ionic current caused by a molecule's passage through a nanopore, associating these disruptions with its identity. The approach is applied in genomic analysis to perform real-time, long-read sequencing, yielding structural and chemical information about RNA molecules. ONT offers portability and scalability, making it suitable for large-scale genomic studies. The method applies an electric potential across a nanopore-embedded membrane, driving single-stranded RNA through the pore. As the RNA translates, characteristic disruptions in ionic current occur, reflecting specific nucleotide groups.

During ONT Direct RNA Sequencing (DRS), variations in ionic current are detected and recorded as raw signal data. These signal changes correspond to nucleotide identities and are converted into sequences in real-time using specialised algorithms. ONT allows full-length RNA sequencing while capturing information about chemical modifications. However, challenges remain, including systematic errors, lower base-calling accuracy, and error rates of approximately 10% per read. Cytosine- and uracil-rich regions exhibit higher error rates, primarily characterised by deletions and mismatches, compared to guanine- and adenine-rich regions. Such errors stem from difficulties in distinguishing signal deviations caused by modifications from noise,

necessitating robust analytical tools to differentiate real modifications from artefacts (Kovaka et al., 2025; Liu et al., 2024).

Despite these challenges, ONT has been successfully applied to identify RNA modifications, such as m^6A**.** Modifications such as m^6A introduce systematic signal deviations and base-calling errors, particularly at sequence motifs like DRACH, enabling computational detection through comparison with unmodified controls. Specific base modifications often produce context-dependent signal deviations, such as mismatches in guanine-to-adenine transitions or deletions in cytosine-rich regions. The analysis of dwell times, signal intensity, and deviation patterns reveals modifications such as m^6A, m^5C, and pseudouridine. Current limitations in nanopore devices and base-calling algorithms affect performance, but machine learning-based callers are expected to reduce systematic errors and improve detection accuracy (Liu et al., 2024).

To improve detection robustness, multi-modal strategies combining ONT with orthogonal validation tools have been developed. A multi-modal approach has been applied to profile tRNA T-loop base modifications across 42 yeast isoacceptors by combining ONT RNA sequencing, genetic methods, and mass spectrometry. Sequence-specific miscalls induced by modifications were validated using enzyme knockouts, such as Pus4 for Ψ_{55} and Trm6-Trm61 for m^1A_{58}, in conjunction with liquid chromatography-tandem mass spectrometry (LC-MS/MS). These orthogonal techniques confirmed modification-specific disruptions in ionic current and base-calling errors, revealing distinct patterns of modifications. For instance, Ψ_{55} was shown to influence the addition of m^1A_{58} in certain isoacceptors, underscoring a modification circuit critical for tRNA stability and function.

Global ribonucleoside profiling via LC-MS/MS quantified modification stoichiometry and confirmed modification sites, detecting rare and dynamic changes under physiological conditions, such as nutrient depletion. Integrating orthogonal methods enhanced sensitivity and mitigated systematic sequencing errors, thereby establishing a reliable pipeline for high-throughput mapping of tRNA modifications. Collectively, these studies demonstrate the applicability of ONT in detecting RNA modifications, resolving their functional roles, and achieving transcriptome-wide resolution (Shaw et al., 2024).

1.6 Computational approaches for RNA 3D structure prediction

RNA remains a challenging subject in structural biology, with many aspects still unresolved. The limited availability of high-resolution structural data, combined with ongoing experimental efforts and the integration of orthogonal techniques, highlights

the complexity of studying RNA architecture. Unlike proteins, no existing computational method achieves the level of accuracy obtained by AlphaFold for tertiary structure prediction. Nonetheless, recent developments have produced algorithms that offer measurable progress toward this goal.

1.6.1 Physics-based and coarse-grained (CG) methods

Several computational strategies have been developed to predict RNA three-dimensional structures, ranging from simplified coarse-grained models to atomic-resolution simulations.

Physics-based methods contribute to RNA folding and structural modelling. The Vfold model, for instance, integrates statistical mechanics with a coarse-grained representation of RNA. Nucleotides are represented as pseudoatoms connected by virtual bonds, simplifying secondary and tertiary structures while maintaining relevant geometric constraints (Xu & Chen, 2015). Among coarse-grained approaches, SimRNA uses a reduced nucleotide representation combined with a Monte Carlo algorithm to explore RNA conformational space. This strategy improves sampling efficiency and reduces computational demands (Boniecki et al., 2016).

While coarse-grained methods enable efficient exploration, atomic-resolution techniques provide detailed descriptions of local interactions. All-atom molecular dynamics (MD) simulations refine experimentally derived RNA structures from techniques such as NMR spectroscopy and X-ray crystallography. Despite their resolution, MD simulations remain limited in de novo prediction from sequence due to the complex energetic landscape of RNA folding (Dawson et al., 2016). To improve prediction accuracy, multiscale models combine 2D secondary structure information with 3D scaffold generation and atomic-level refinement (Cao & Chen, 2011).

More recently, machine learning-assisted approaches have been introduced. These techniques combine data-driven predictions with physics-based constraints, improving the modelling of large RNA molecules in specific contexts (Xiujuan Ou et al., 2022). Hybrid models that integrate statistical and physical principles have expanded the applicability of computational RNA structure prediction.

Together, these computational strategies highlight the complementary roles of coarse-grained, atomic-resolution, and machine learning-assisted models in addressing the complexity of RNA three-dimensional structure prediction.

1.6.2 Deep learning-based methods

Deep learning approaches have increased both the accuracy and scalability of RNA structure prediction compared to traditional methods. Geometric deep learning techniques have enabled predictions of RNA three-dimensional conformations, even in settings with limited training data. The Atomic Rotationally Equivariant Scorer (ARES), which relies on atomic coordinates, outperforms earlier models trained on larger datasets (Townshend et al., 2021).

For RNA secondary structure prediction, RNAformer applies axial attention and latent space recycling within a deep learning framework, achieving results consistent with existing benchmarks without requiring external datasets (Franke et al., 2023). The 2dRNA method combines bidirectional long short-term memory (LSTM) networks with U-Net architectures to predict secondary structures, including pseudoknots, by incorporating sequence homology (Mao et al., 2022). For the prediction of three-dimensional RNA structures, E2Efold employs an end-to-end deep learning architecture to generate conformational models with competitive performance (Shen et al., 2022).

More recent models have combined pretrained sequence embeddings with structural learning pipelines. RhoFold+ represents an extension of this direction, utilizing a deep learning framework based on RNA-FM, a language model pretrained on more than 23.7 million RNA sequences. RhoFold+ incorporates invariant point attention and structural recycling to iteratively refine predicted structures. The model performs consistently across benchmarking datasets such as RNA-Puzzles and CASP15 and generalizes to RNA families not included in the training set. Secondary structure outputs are generated in parallel, supporting applications in structure–function analyses and RNA nanotechnology (Shen et al., 2024).

AlphaFold 3 (AF3) extends this progression by introducing a general-purpose framework for biomolecular structure prediction based on a diffusion model. AF3 directly infers atomic coordinates using a generative denoising strategy, eliminating the need for torsion-angle parametrizations or frame-based representations. The model accommodates a wide range of molecular entities, including modified RNA bases and covalently bonded ligands, without requiring task-specific retraining. Benchmarks show that AF3 outperforms nucleic acid-specific tools such as RoseTTAFold2NA in RNA structure prediction, including tasks involving protein–RNA complexes and CASP15 RNA targets. Structural models generated by AF3 include both unmodified and chemically modified RNA residues, supporting its use in the analysis of modification-induced structural changes (Abramson et al., 2024).

The integration of structural priors with pretrained sequence representations, alongside diffusion-based generative architectures, is currently broadening the scope of computational RNA structure prediction.

1.6.3 RNA puzzles

To systematically evaluate RNA structure prediction methods, community-wide benchmarking efforts such as RNA-Puzzles have been established. The RNA-Puzzles initiative advances RNA 3D structure prediction through blind prediction rounds, where groups model structures before experimental data are published.

Its goal is to evaluate and improve computational methods for predicting RNA structures. Results show accuracy sufficient for topological assessment but highlight limitations in modelling non-Watson-Crick interactions and ligand-binding sites. Human predictions often surpass automated methods for complex structures (Miao et al., 2020).

Template-based approaches outperform *de novo* methods, especially for larger RNAs. The initiative has developed tools, such as the RNA-Puzzles toolkit (Magnus et al., 2019), which offers decoy sets and metrics for structural comparison. Feedback from experimental results and community efforts, such as EteRNA, has informed algorithm development (Koodli et al., 2019).

These efforts contribute to the iterative refinement of computational methods by providing objective performance metrics, reference datasets, and experimental validation.

1.7 RNA and RNA modification databases

Computational approaches to RNA structure and modification rely on curated reference resources, including dedicated databases that compile sequences, structures, and chemical annotations. RNA and RNA modification databases contribute to advancing RNA biology, cataloguing sequences, structures, and post-transcriptional modifications across species. Resources like MODOMICS provide information on RNA modifications and biosynthetic pathways (Cappannini et al., 2023), RMBase links RNA modifications to RNA-protein interactions (Xuan et al., 2024), and RNAcentral integrates non-coding RNA sequences (RNAcentral Consortium, 2021). These platforms support the analysis of modification patterns, regulatory roles, and RNA conformational stability. Cross-species comparisons support computational predictions and facilitate studies in epitranscriptomics, developmental biology, disease mechanisms, and the development of RNA therapeutics. The following sections examine the databases most relevant to this thesis. The following sections examine the

databases used in this thesis and those relevant to the broader scientific understanding of RNA modifications and structure.

1.7.1 Rfam database

Rfam is a database that catalogues RNA families. It offers multiple sequence alignments and covariance models to represent ncRNA families. It supports the identification, classification, and annotation of ncRNAs across diverse genomes (Ontiveros et al., 2025).

1.7.2 MODOMICS

MODOMICS is a curated database focused on RNA modifications. It provides structured information on modified ribonucleotides, their biosynthetic pathways, RNA-modifying enzymes, and the positions of modified residues within RNA sequences. First introduced in 2006 (Horkawicz et al., 2005), the resource has been regularly updated to integrate developments in RNA biology.

Initial expansions included data on tRNAs, rRNAs, and RNAs with known three-dimensional structures, along with newly identified RNA-modifying enzymes in model organisms such as *E. coli* and *S. cerevisiae* (Czerwoniec et al., 2008). Later updates incorporated a census of small nucleolar RNAs (snoRNAs), which are responsible for RNA-guided modifications, and introduced a section on the 5′-end capping process. Additional sections described the chemical building blocks for the synthesis of modified nucleosides and provided extended datasets of RNA modifications and the corresponding enzymes. These included experimentally mapped modifications in small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs), as well as updated sequences of rRNA and tRNA (Machnicka et al., 2013).

In 2017, the database integrated mass spectrometry and liquid chromatography data for modified nucleosides, added links to MINTbase for human tRNA sequences (Pliatsika et al., 2018), and standardised abbreviations for modified nucleosides. Updates also included bacterial tRNA modifications, expanded mammalian tRNA entries, newly characterised nucleosides, and functionally validated RNA-modifying proteins (Boccaletto et al., 2017).

More recent revisions have introduced links between RNA modifications and human diseases, restored sequence alignments with annotations to assess the reliability of modification data, and introduced synthetically engineered RNA modifications collected from RCSB-deposited structures. These additions increased the relevance of the database for research in RNA biology (Cappannini et al., 2023; Boccaletto et al., 2021).

1.7.3 Nucleic acid circular dichroism database

Complementing MODOMICS, which focuses on the annotation of chemical RNA modifications, a parallel initiative was developed to provide spectroscopic insights into RNA conformational states. The Nucleic Acid Circular Dichroism Database (NACDDB) is a collaborative project involving research groups at SOLEIL Synchrotron, Université Paris Cité, Université Paris-Saclay, and the International Institute of Molecular and Cell Biology in Warsaw (IIMCB). NACDDB offers a curated collection of circular dichroisms (CD) and synchrotron radiation CD (SRCD) spectral data for various nucleic acid molecules, including DNA, RNA, DNA/RNA hybrids, and chemically modified derivatives. The database provides access to standardised CD data, including spectral profiles, experimental conditions, 3D structural models, and references to associated publications.

The current content of NACDDB includes over 135 entries, combining newly collected SRCD data generated at the DISCO beamline at SOLEIL Synchrotron with digitised spectra extracted from published sources. Each entry contains metadata such as nucleic acid sequences, spectral band ranges, measurement intervals, and parameters specific to SRCD measurements, including detector high tension (HT), pseudo-absorbance, dwell time, and sample pathlength.

To support structural comparisons, NACDDB allows searches based on sequence similarity, spectral features, or structural characteristics. Eigen-spectra, derived from single-value decomposition (SVD) of representative datasets, serve as reference spectra for identifying recurrent structural motifs. These tools support the analysis of correlations between CD spectra and folding patterns, which are not readily captured by high-resolution structural methods.

NACDDB includes information on chemically modified nucleic acids, with cross-references to the MODOMICS database (Cappannini et al., 2023) for residue-level annotations. Applications of CD spectroscopy documented in the database include analyses of conformational dynamics, folding transitions, and sequence-dependent structural responses to mutations or chemical modifications.

By consolidating and standardising CD spectral data, NACDDB offers a structured platform for investigating nucleic acid conformations, interactions, and folding behaviour. It is used in both fundamental and applied contexts, such as the quality assessment of nucleic acid therapeutics and the identification of structural motifs across diverse biological systems (Cappannini et al., 2022). Together, these resources consolidate structural and functional data across molecular, spectroscopic, and computational dimensions, supporting comprehensive RNA research.

2 Aim of this work

This thesis investigates the hypothesis that post-transcriptional RNA modifications influence RNA structure. To address this hypothesis, the research followed three specific steps:

1. Assessment and refinement of RNA modification data: Critical evaluation and validation of RNA modification data deposited in MODOMICS, with a focus on confirming the accuracy and quality of annotations related to experimentally determined RNA structures.

2. Structural and evolutionary analyses of RNA modifications: Analysis of RNA sequence alignments and experimentally determined structures stored within MODOMICS, investigating the evolutionary conservation of modified positions in tRNA and rRNA, and assessing their structural implications.

3. Integration of Circular Dichroism spectroscopy data: Establishment and expansion of the Nucleic Acid Circular Dichroism Database (NACDDB) to provide circular dichroism spectra for RNA molecules, including modified nucleic-acid oligos. Development of computational tools based on Cosine Similarity and Root Mean Square Deviation algorithms enabled quantitative comparisons of nucleic acid structures, further elucidating structural effects induced by specific RNA modifications.

By integrating evolutionary insights, structural annotations from MODOMICS, and spectral analyses from NACDDB, this research provides a comprehensive understanding of how chemical modifications modulate RNA structure and function, offering new tools and refined datasets to support future RNA studies.

3 Results

This section presents an integrated analysis of RNA modifications, combining evolutionary conservation, structural alignment, and spectroscopic data to investigate their functional relevance. Alignments of tRNA and rRNA sequences across domains of life revealed conserved modification patterns at structurally constrained positions. Structural mapping using experimentally resolved RNA structures and quantitative CD spectroscopy supported the hypothesis that chemical modifications contribute to RNA stability, folding, and functional adaptability. Comparative metrics, including RMSD, cosine similarity, and Kendall's τ, were applied to CD spectra to evaluate structural effects of specific modifications. Together, the results support a model in which nucleotide modifications are preserved at functionally critical sites, reflecting selective pressure on RNA structural integrity and adaptability.

3.1 RNA Modification Data Analysis

This section addresses the central hypothesis of the thesis: that post-transcriptional modifications influence RNA structure in a position- and context-dependent manner. Comparative analyses of MODOMICS alignments were performed to identify conserved patterns suggestive of Darwinian selection acting on modified nucleotides. These findings provide a sequence-level context for the evolutionary relevance of specific modifications. To complement this evolutionary perspective, circular dichroism data from modified RNA sequences were used to assess the structural impact of m^1A and m^6A relative to their unmodified counterparts. The analysis focused on how the position of each modification affected the resulting spectral profile, providing experimental support for the hypothesis. Finally, stacking interactions were analysed by comparing dihydrouridine, pseudouridine, and uridine, focusing on the distribution of stacking classes. In this structural framework, pseudouridine exhibited a higher frequency of stacking-class interactions compared to unmodified uridine, suggesting a stabilising contribution at the structural level.

3.1.1 Analysis of tRNA-aligned sequences

To investigate how RNA modifications contribute to structural conservation and metabolic adaptation, domain-specific alignments of transfer RNA (tRNA) sequences from MODOMICS were analysed. The objective was to detect modification patterns reflecting evolutionary constraints within conserved tRNA domains. Therefore, modification frequencies were quantified across Eukarya, Bacteria, and Archaea, and visualised to identify conserved evolutionary signatures (**Fig. 12A**, **B**).

A general statistical analysis was performed to obtain a quantitative overview of RNA modifications present within the tRNA pool annotated in MODOMICS. To this end, a set of Python scripts was implemented to process the data and generate a pie chart representing the relative proportions of modified nucleotides, alongside a bar plot providing more granular quantitative distributions. Pseudouridine (**Ψ**), dihydrouridine (**D**), and ribothymidine (**T**) emerged as the most prevalent modifications, together accounting for ~45% of all annotated tRNA modifications. Specifically, Ψ constituted 22.14% (n = 1495), D 16.43% (n = 1109), and T 7.29% (n = 492). Additionally, 2′-O-methylations made up over 4%.

Although the global analysis provided an initial overview of tRNA modification frequencies across organisms, it lacked the positional resolution needed to explore evolutionary constraints. To address this, each nucleotide modification was quantified according to the specific position it occupied within the tRNA sequences. Counts were then stratified across the three domains of life—Eukarya, Bacteria, and Archaea—yielding three separate heatmaps. This representation allowed a domain-wise comparison of modification retention, revealing how evolutionary pressures have shaped the positional landscape of tRNA modifications in a lineage-specific manner. Domain-specific heatmaps (**Fig. 13**) revealed differential distributions, with Eukaryotic tRNAs showing broader modification patterns, possibly reflecting increased metabolic complexity. Among the identified patterns, certain position-specific modifications exhibited domain-dependent conservation. For example, Ψ_{22} was shared across Bacteria and Archaea, but not Eukarya. In thermophilic Archaea such as *H. volcanii* and *S. acidocaldarius*, this position may contribute to thermal adaptation by enhancing base stacking (Wolff et al., 2020). In Bacteria, position 22 is instead occupied by N^1-methyladenosine (m^1A_{22}), hypothesised to support tRNA folding and ribosomal binding. The essentiality of TrmK, the methyltransferase responsible for m^1A_{22}, reinforces its physiological relevance (Oerum et al., 2017).

Another highly conserved site, position 37, carries N^6-threonylcarbamoyladenosine (t^6A_{37}) across all three domains. This modification stabilises the anticodon stem-loop and optimises codon–anticodon interactions, thereby reducing frameshifting and enhancing translational fidelity (Su et al., 2022; Bacusmo et al., 2018). Its universal retention suggests evolutionary pressure to preserve decoding accuracy.

Dihydrouridine (D) was systematically mapped to the D-loop, consistent with its role in increasing local structural flexibility required for aminoacylation and ribosomal accommodation (Finet et al., 2022; Suzuki et al., 2021). Methylation at positions 9 and 58 (m^1A_9 and m^1A_{58}) showed partial conservation. In Eukarya and Bacteria, m^1A_{58} increases the melting temperature of tRNA, stabilising the molecule under thermal

stress. In contrast, m^1A_9 was absent in Archaea. Archaeal tRNAs retained m^1A_{58} but not m^1A_9, consistent with selective retention of thermostabilising modifications under extreme conditions (Oerum et al., 2017).

To further highlight shared versus domain-specific patterns, Venn diagrams (**Fig. 13D**, **E**) and intersection tables (**Tables 6** and **7**) were generated. These visualizations highlight both the overlap in modification types and the divergence in positional conservation across lineages. **Table 7** further resolves these positions by mapping them onto the tRNA consensus secondary structure using domain-specific colour codes. Positions specific to individual domains frequently localize within structurally and functionally distinct tRNA regions. For example, Bacteria-only modifications include position 5 in the acceptor stem (lobster red), whereas Archaea-only modifications such as 45D and 15 localize to the TΨC and D domains, respectively. Positions shared by Bacteria and Archaea, including 22 and 17, are concentrated in the D-domain (orange) and anticodon arm (light blue), regions associated with tertiary structure stabilization and codon decoding.

Modifications conserved across all three domains of life, such as positions 32, 34, 38, and 39, are predominantly located in the TΨC-loop and anticodon domain. These regions are functionally constrained due to their involvement in decoding fidelity and ribosomal accommodation. The functional relevance of these structural domains is well-supported by previous studies, which have shown that anticodon modifications such as those at positions 34 and 37 regulate translational accuracy and prevent frameshifting errors, while D-loop and T-loop modifications promote proper folding and facilitate tRNA-ribosome interactions (Suzuki et al., 2021; Motorin & Helm, 2011; Finet et al., 2022).

The integration of structural domain mapping and cross-domain conservation data supports the view that modification positions are not distributed randomly but instead reflect convergent selective pressures. This positional constraint operates in conjunction with chemical identity selection, reinforcing the dual-level model of Darwinian selection acting on both modification type and structural location within the tRNA molecule.

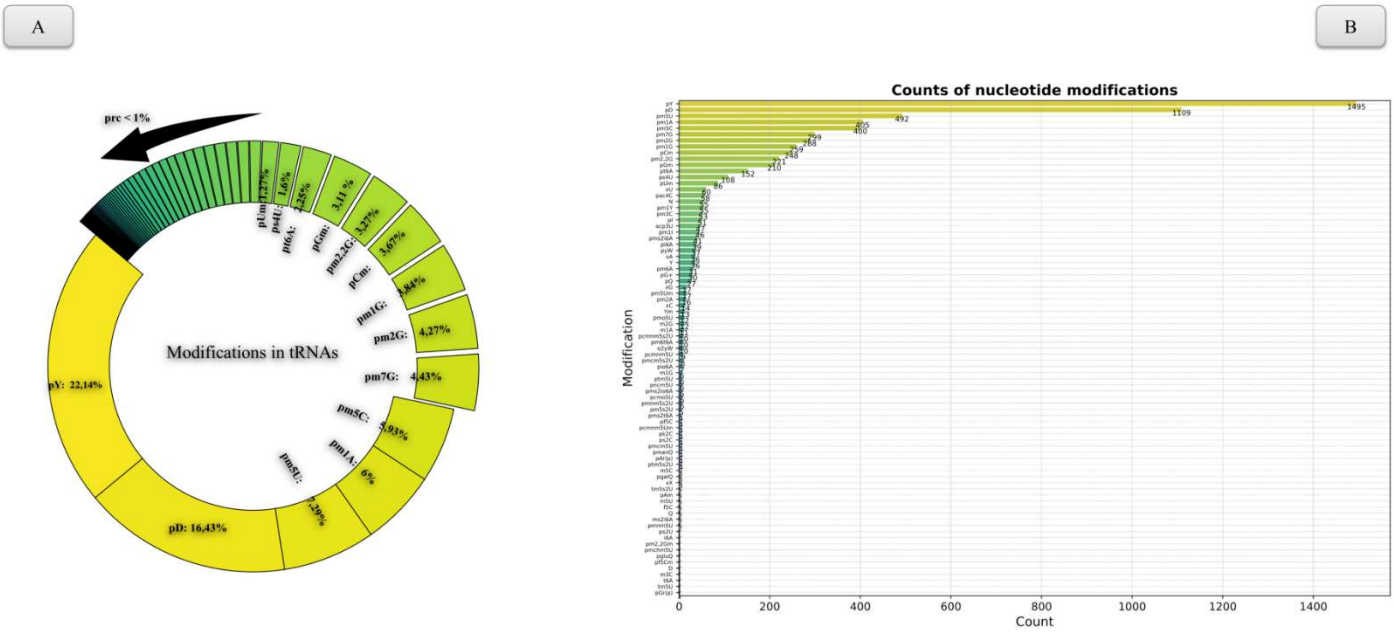


Figure 12 Descriptive Statistical Analysis of tRNA modifications present within the MODOMICS's alignments. (**A**) Percentage of the most prevalent modifications. (**B**) Absolute Numbers of the tRNA modifications.

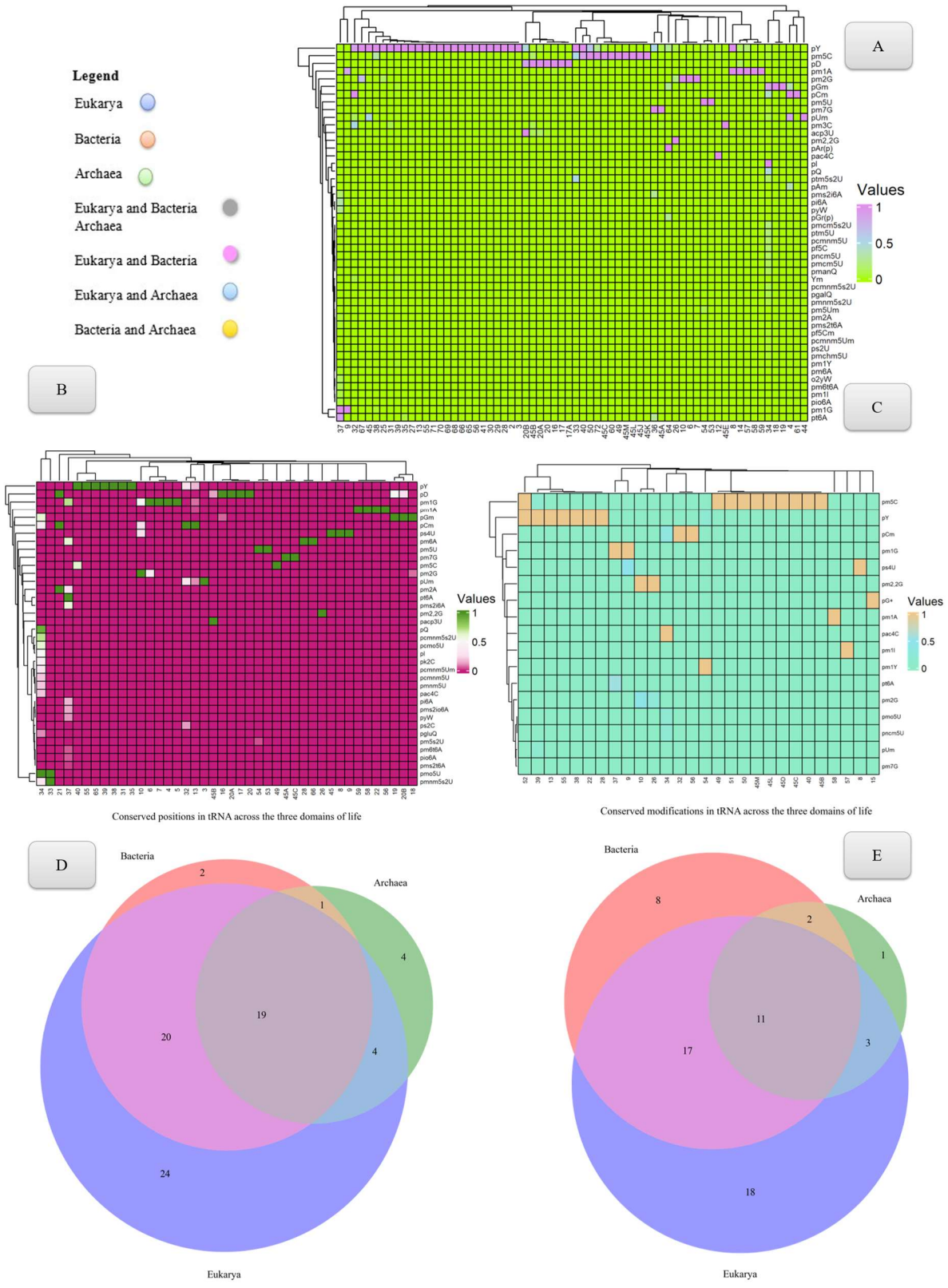


Figure 13 Comparative Statistical Analysis of tRNA modifications present within the MODOMICS's alignments. (**A**) Eukarya cluster map. (**B**) Bacteria cluster map. (**C**) Archaea cluster map. (**D**) Venn Diagram of the post-transcriptional modification positions of Eukarya, Bacteria, and Archaea tRNA sequences. (**E**) Venn Diagram of the post-transcriptional modifications of Eukarya, Bacteria, and Archaea tRNA sequences.

Bacteria Only	Archaea Only	Eukarya Only	Bacteria & Archaea	Bacteria & Eukarya	Archaea & Eukarya	All Three
pGluQ	pG+	pAm	pmo5U	pcmnm5s2U	pm1Y	pm5C
pcmo5U		pmanQ	ps4U	pcmnm5Um	pm1I	pm2G
pk2C		pmchm5U		pyW	pncm5U	pac4C
pacp3U		pGalQ		pQ		pm1G
pmnm5U		o2yW		pm5U		pt6A
pms2io6A		pf5C		pm6t6A		pY
pm5s2U		pm3C		pm2A		pUm
ps2C		pGr(p)		pi6A		pm7G
		ps2U		pms2t6A		pm2,2G
		pmcm5U		pD		pm1A
		pmcm5s2U		pcmnm5U		pCm
		ptm5s2U		pm6A		
		pAr(p)		pGm		
		Ym		pio6A		
		pm5Um		pms2i6A		
		acp3U		pmnm5s2U		
		ptm5U		pI		
		pf5Cm				

Table 6 Modification Venn diagram's set

Bacteria Only	Archaea Only	Eukarya Only	Bacteria & Archaea	Bacteria & Eukarya	Archaea & Eukarya	All Three
21	15	25	22	7	45L	32
5	45D	45K		59	45M	45B
	51	17A		6	57	34
	52	12		19	50	38
		14		16		13
		30		4		39
		44		45		45C
		45J		65		26
		64		17		28
		41		18		10
		36		31		54
		67		53		58
		69		20B		55
		1		35		9
		45E		20A		56
		68		66		8
		70		45A		40
		2		33		37
		60		20		49
		27		3		
		71				
		72				
		29				
		61				

Table 7 Modification position Venn diagram's set. The colour code represents the domain in which the modification position falls in the consensus secondary structure. Lobster red: Acceptor stem. Grey: Not a specific domain. Orange: D-domain. Light blue: anticodon-domain. Light green: variable-region. Purple: TPC-domain

3.1.2 Analysis of rRNA-aligned sequences: domain I

To investigate the modification landscape of early-assembling ribosomal domains, domain I of the large subunit (LSU) was analysed with respect to its post-transcriptional modification profile. Domain-I contributes to the structural organisation of the pre-60S ribosomal subunit and interacts with domains II and VI during early stages of nucleolar assembly, forming the solvent-exposed surface of the LSU. Although these domains adopt near-final conformations early in biogenesis, the functional significance of their chemical decoration remains unclear. Therefore, a heatmap and bar plot composite (**HWB**, **Fig. 14A**) was generated using R scripts to quantify the distribution of RNA modifications across Domain I in representative species.

The resulting profile indicated that domain I exhibits low modification frequency, with most events restricted to higher eukaryotes. Among these, *2′-O-methyladenosine* (A_m) was the predominant modification, consistently detected in *H. sapiens* and *O. cuniculus*, and less prominently in *S. lycopersicum*. In contrast, other 2′-*O*-methylations (N_m)—including G_m, C_m, U_m—and pseudouridine (Ψ) were identified exclusively in *S. lycopersicum*, each occurring at low relative abundance. This species-specific clustering is visible in the bar plot, where the full modification burden of domain I is concentrated in a subset of taxa, particularly within plant lineages.

To determine whether conserved positions are shared across organisms, modification sites were aligned using *H. sapiens* 28S rRNA (MODOMICS ID 1223) as the reference. Donut plots (**Fig. 14B**) illustrate that A_m is the only modification detected at conserved positions in both mammals and plants. The remaining modifications identified in *S. lycopersicum* correspond to non-conserved, lineage-specific sites, consistent with the annotations reported by Cottilli et al. (2022).

These findings suggest that domain I does not undergo extensive chemical modification and that its few modifications are preferentially retained in higher eukaryotes. Rather than reflecting universal structural constraints, the observed decoration pattern may instead indicate lineage-specific modulation of ribosomal function during early subunit maturation. Methylation restricts the conformational space by altering hydrogen bond polarity, shifting from $O^{2'}H\cdots O^{3'}$ in canonical ribose to $O^{2'}\cdots HO^{3'}$ in *2′-O*-methylated residues (He et al., 2018). These stereo-electronic effects modulate backbone flexibility and solvent accessibility, features particularly relevant in the crowded, electrostatically regulated environment of the nucleolus where early LSU assembly initiates. Given that domain-I engages in early contacts with domains II and IV during nucleolar maturation, it is plausible that N_m modifications

contribute to fine-tuning of the local geometry, stabilising interdomain interactions under the physicochemical constraints of phase-separated ribonucleoprotein condensates (Feric et al., 2016; Lafontaine et al., 2019). Such context-dependent modulation may reflect lineage-specific adaptations in ribosome biogenesis.

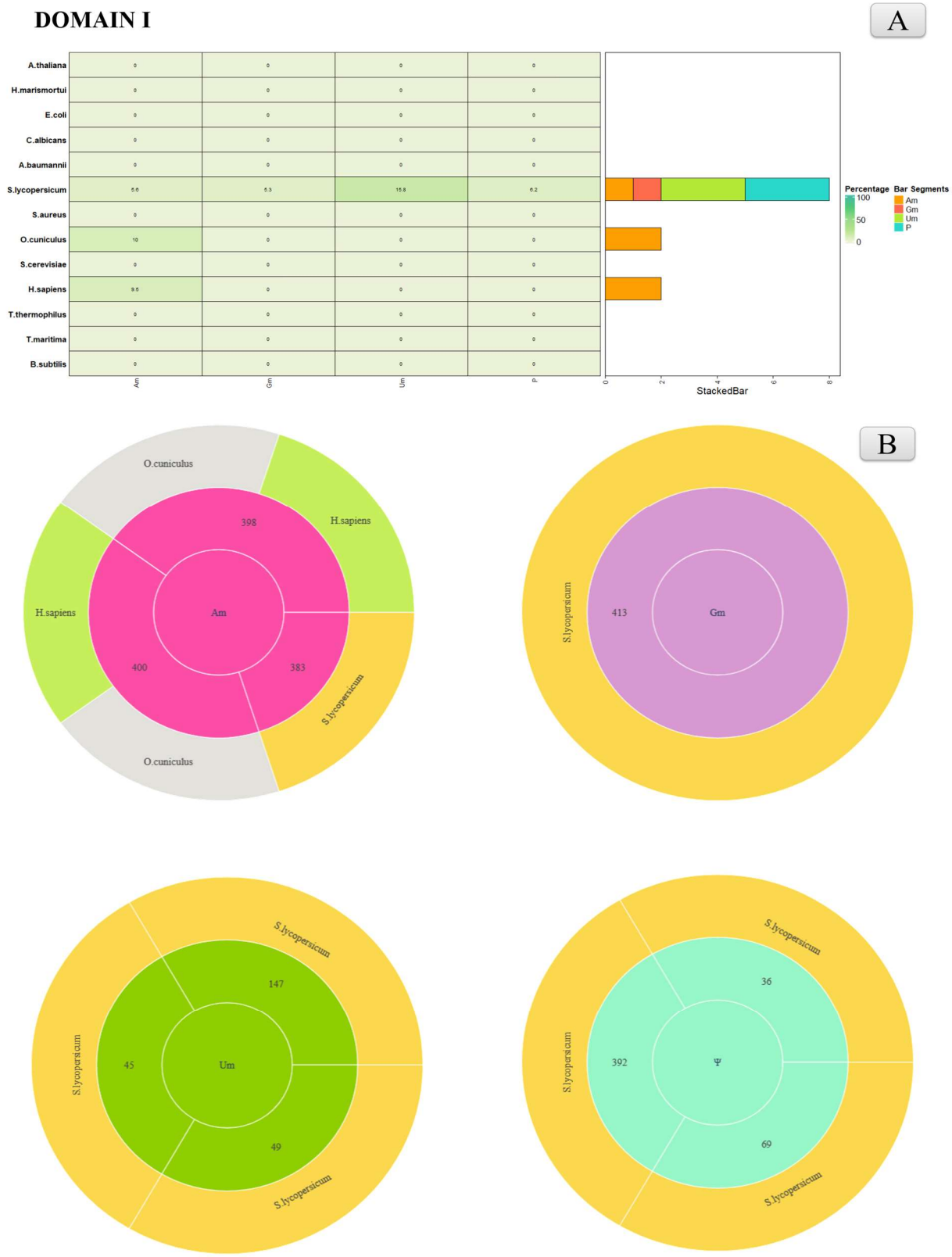


Figure 14 Distribution of RNA modifications in DOMAIN-I. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S.

3.1.3 Analysis of rRNA-aligned sequences: domain II

As for the previous LSU domain, R scripts were employed to characterise the post-transcriptional modification landscape of domain II, aiming at identifying evolutionarily conserved patterns in both biochemical composition and positional distribution.

Domain II of the large ribosomal subunit (LSU) was analysed to investigate the taxonomic distribution and structural positioning of post-transcriptional RNA modifications. This region contributes to ribosome recycling by mediating interactions with elongation factor G (EF-G) during translation termination, enabling dissociation and reuse of ribosomal subunits (Guo et al., 2006). A heatmap and stacked bar plot (**Fig. 15A**) were generated to visualise the modification profiles across 14 representative species of the MODOMICS rRNA sequence pool.

The results indicated that *2′-O*-ribose methylations (N_m) are largely restricted to eukaryotes. Specifically, *A. thaliana*, *S. cerevisiae*, and *H. sapiens* exhibited 15–35% A_m within domain II, with additional detection of G_m and C_m. No U_m events were observed in *O. cuniculus*. In contrast, pseudouridine (Ψ) was widely distributed across eukaryotic species with a few pseudourydilation events specific to *E.coli* and *A.baumannii*. Two additional methylated nucleotides—*N^1*-methylguanosine (m^1G) and *5-methyluridine* (m^5U)—were restricted to *E. coli*, *S. aureus*, and *B. subtilis*.

To determine whether these modifications occurred at conserved *loci*, sunburst plots were produced (**Fig. 15B**) by aligning all positions to the human 28S rRNA reference (MODOMICS ID 1223). The resulting patterns showed that N_m events were present across multiple eukaryotic clades, with subsets restricted to mammals or shared among plants and fungi. Most N_m modifications mapped near ribosomal protein 60S L39, which is located adjacent to the peptide exit tunnel and the peptidyl transferase centre (PTC) (Ayers et al., 2024). Their placement suggests a structural role in fine-tuning interfacial geometry between rRNA and protein elements involved in elongation and exit channel architecture.

Pseudouridines appeared sparsely but consistently distributed across domain II, mapping to evolutionarily retained positions. These residues are likely to contribute to local tertiary structure formation and stabilisation of ribosomal RNA–protein interfaces, consistent with previous findings on pseudouridine's structural role (Yu Zhao et al., 2021; King et al., 2003).

Two bacterial modifications appear conserved at functionally relevant positions. Human h_{1596} corresponds to *E. coli* position 747, which carries a *5*-methyluridine (m^5U) located in the peptide exit tunnel. The persistence of this modification across *E. coli*,

S. aureus, and *B.subtilis* suggests functional relevance for nascent peptide positioning or translational fidelity, with possible implications for antibiotic sensitivity (Madsen et al., 2003). Similarly, position h_{1594} aligns with *E. coli* m^1G_{745}, a methylation introduced by the RrmA methyltransferase. Its absence has been linked to reduced translational efficiency, slower growth rates, and increased viomycin resistance (Gustafsson et al., 1998).

Taken together, the results show that domain II modification patterns are taxonomically diverse, with dispersed positioning. While N_m events appear restricted to eukaryotic lineages, bacterial methylations such as m^5U and m^1G exhibit conservation at structurally strategic sites. These observations support a model in which lineage-specific modifications reinforce domain architecture while conserved marks anchor functional ribosomal features.

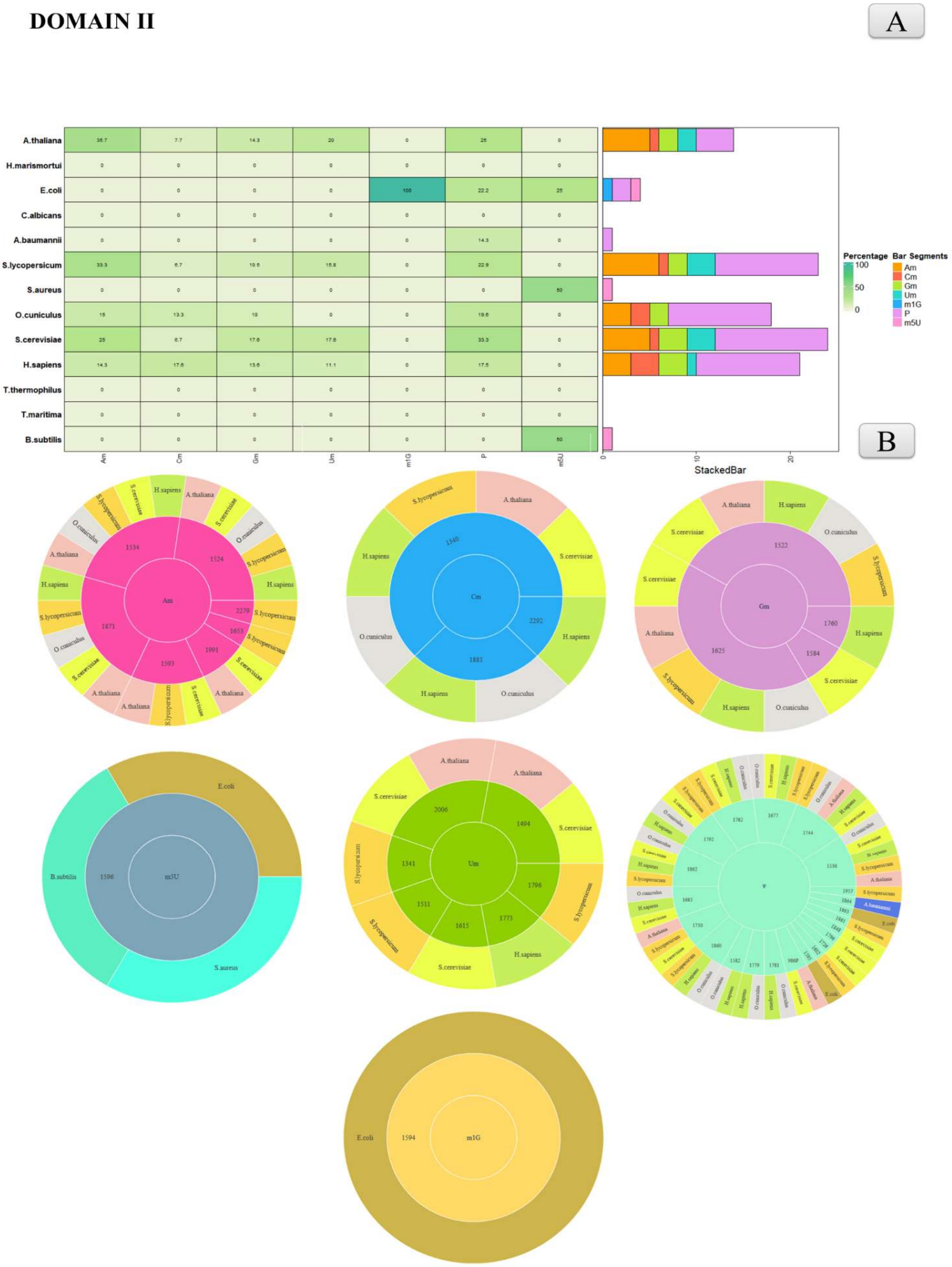


Figure 15 Distribution of RNA modifications in DOMAIN-II. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S.

3.1.4 Analysis of rRNA-aligned sequences: domain III

Domain III of the large ribosomal subunit (LSU rRNA) folds autonomously and contributes to ribosomal assembly and structural organisation. Its conformational independence has led to the hypothesis that it may have been integrated as a discrete functional unit during ribosomal evolution (NASA Astrobiology Institute: NASA Astrobiology Institute: https://bit.ly/3CMo065). To evaluate its post-transcriptional landscape, an initial statistical analysis was conducted. A heatmap with bar plot (HWB, **Fig. 16A**) was generated using R scripts to assess the distribution of RNA modifications in domain III across representative taxa.

The resulting profile resembled that of domain II, with 2′-*O*-ribose methylations (N_m) predominating among eukaryotic species. A_m was absent in *S. lycopersicum*, while Ψ residues were detected in *S. lycopersicum*, *O. cuniculus*, and *H. sapiens*. In *E. coli*, approximately half of the *N*⁶-methyladenosine (m^6A) modifications identified within the LSU were in domain III.

To assess positional conservation, modification sites were mapped to the *H. sapiens* 28S rRNA reference (HS28; **Fig. 16B**; MODOMICS 1223). Among the N_m class, C_m showed the highest degree of conservation, followed by A_m and G_m, all confined to eukaryotic lineages. A single m^6A site at position h_{2798} (corresponding to *E. coli* m^6A_{1619}) was uniquely observed in *E. coli*. This modification is positioned near the nascent peptide exit tunnel and is absent from all other surveyed organisms. The methyltransferase YbiN, later renamed RlmF, catalyses this modification. Deletion of ybiN results in the loss of m^6A_{1618}, while complementation restores its presence, confirming enzymatic specificity. YbiN modifies partially assembled 50S ribosomal subunits but is inactive on fully assembled subunits or fully deproteinized 23S rRNA, indicating a requirement for structural intermediates during catalysis. Loss or overexpression of ybiN impairs cellular growth, suggesting that m^6A_{1618} contributes to ribosomal function and fitness. This modification is located within a cluster of nucleotides near ribosomal proteins L22 and L34, two regions implicated in regulatory peptide interactions and antibiotic binding (Sergiev et al., 2008).

Overall, domain III exhibits a post-transcriptional modification profile that is predominantly restricted to eukaryotic species. Although the domain shows only minimal length variation between *H. sapiens* and *E. coli* (Ontiveros et al., 2024; Petrov et al., 2014), the absence of conserved modifications in bacterial and archaeal sequences indicates that no structural barrier exists to prevent their presence. The available reference positions are aligned and accessible yet remain unutilised in non-eukaryotic taxa. This observation, combined with the presence of minor insertion

events during evolution, suggests that the observed modifications represent a eukaryotic innovation rather than a universally retained ancestral feature. The fixation of these chemical marks at specific positions in eukaryotic lineages reflects Darwinian selection acting not only at the level of domain structure but also at the resolution of individual nucleotide modification sites.

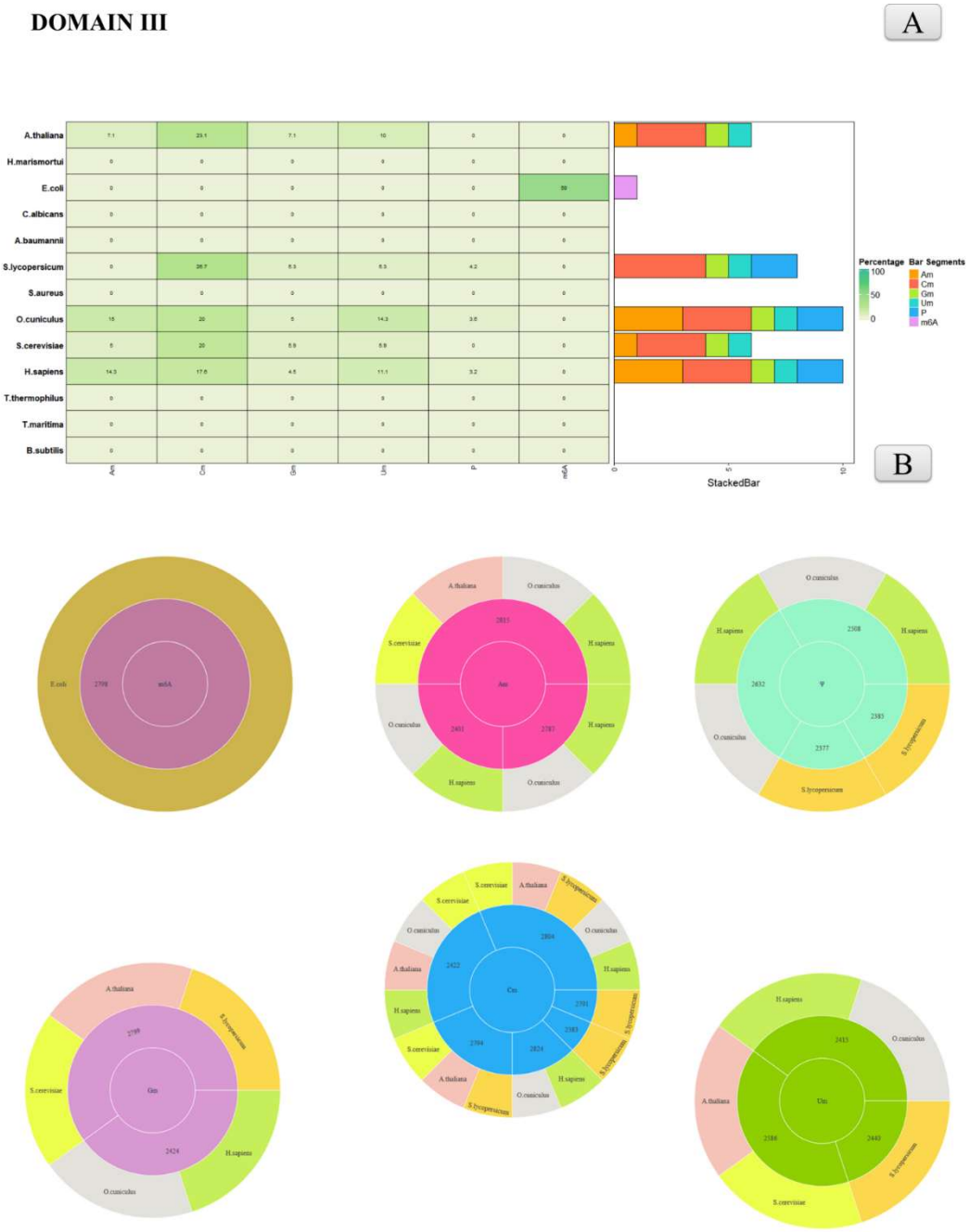


Figure 16 Distribution of RNA modifications in DOMAIN-III. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S.

3.1.5 Analysis of rRNA-aligned sequences: domain IV

Domain IV of the large ribosomal subunit (LSU rRNA) is highly conserved, with over 50% of its nucleotide sequence retained across Bacteria, Archaea, and Eukarya. This conservation reflects its structural and catalytic significance, particularly in mediating peptidyl transferase activity and intersubunit coordination. In eukaryotes, domain IV contributes to reading frame maintenance through its interaction with elongation factor eEF2 (Leviev et al., 1995; Djumagulov et al., 2021).

To analyse the post-transcriptional landscape and assess the evolutionary conservation of nucleotide modifications in this region, a statistical analysis was performed. The heatmap with bar plot (HWB; **Fig. 17A**) displays modification percentages for each species within domain IV, along with the frequency and class of each event. In combination with domain III, domain IV accounts for over half of the total 2′-*O*-ribose methylations (N_m) detected across LSU domains. Pseudouridylation (Ψ) appears broadly conserved across taxa, while other modifications—such as m^5U, m^5C, and m^2G—exhibit more restricted phylogenetic patterns.

To evaluate conservation at the single-nucleotide level, all modification positions were mapped to the *H. sapiens* 28S rRNA reference (**Fig. 17B–C**). N_m modifications (A_m, G_m, C_m, U_m, Ψ_m) cluster adjacent to the peptidyl transferase centre (PTC), located in domain V (purple in **Fig. 17B**), with many falling within domain IV (highlighted in yellow). This spatial co-localisation suggests a functional *continuum* between domains IV and V. Deletion of N_m sites in this region has been shown to disrupt reading frame maintenance and promote premature termination (Yang et al., 2022), supporting a structural role in translation fidelity. Pseudouridines cluster in the same region, indicating a convergent evolutionary strategy to stabilise RNA helices and support tRNA positioning during catalysis.

A conserved modification of interest is ribothymidine (m^5U), which is present in both gram-positive and gram-negative bacteria. Using *E. coli* as a reference, position h_{3786} (*E. coli* 1939) maps to the PTC-adjacent region. Studies have shown that loss or alteration of this modification leads to translational defects, implicating it in ribosomal integrity and assembly (Agarwalla et al., 2002). In *A. baumannii*, m^5U_{1935} is located 11.3 Å from Arg_{238} of protein L2 (encoded by *rplB*), a structural hub for PTC formation. m^5U_{1935} engages in a base stacking (>< ClaRNA class) interaction with A_{1933} and is integrated into a hydrogen bond network that stabilises the surrounding helices. This arrangement may support interdomain communication between domains IV and V and contribute to subunit assembly.

Another conserved editing is m^5C_{1964}, found in *T. thermophilus* and *E. coli* (position 1962 in the original paper by Larsen and coworkers of 2012, 1964 in MODOMICS based alignment). Human position h_{3789} corresponds to *T. thermophilus* C_{1964} and *E. coli*. The RlmO (TTHA1493) methyltransferase catalyses this modification early in biogenesis, acting on unassembled 23S rRNA but not on mature subunits (Larsen et al., 2012). The adjacent C_{1492}, although surface-exposed in the 50S, lies near the A-site tRNA in the functional ribosome. m^5C_{1964} contributes to stacking interactions with neighbouring nucleotides (C_{1941} and U_{1943}), with potential consequences for A-site geometry and mRNA-tRNA dynamics.

Contrary to Larsen et al. (2012), MODOMICS currently does not report this modification in *A. thaliana*, suggesting either a lineage-specific absence or a gap in annotation. Additional methylation at position h_{3809} (reference to *E. coli* C_{1936}) is catalysed by RlmI in *E. coli* and TTHA1280 in *T. thermophilus*. This site is located within helices 69/71 of 23S rRNA, which form intersubunit bridge B2a. These bridges are essential for subunit association and integrity during translation. Deletion of *rlmI* reduces bacterial growth and alters ribosomal fitness, consistent with a role in translation efficiency (Larsen et al., 2012).

Finally, G_{3698} is conserved across all available species and corresponds to *E. coli* m^2G_{1835}. This residue is methylated by RlmG and lies in helix 68 near intersubunit bridges B2a, B2b, and B2c. This region contains seven additional modifications within a 120-nt window, suggesting local enrichment of chemical signals supporting translational precision (Sergiev et al., 2006). The accumulation of modifications in this region may contribute to conformational flexibility and stabilisation of tRNA accommodation sites.

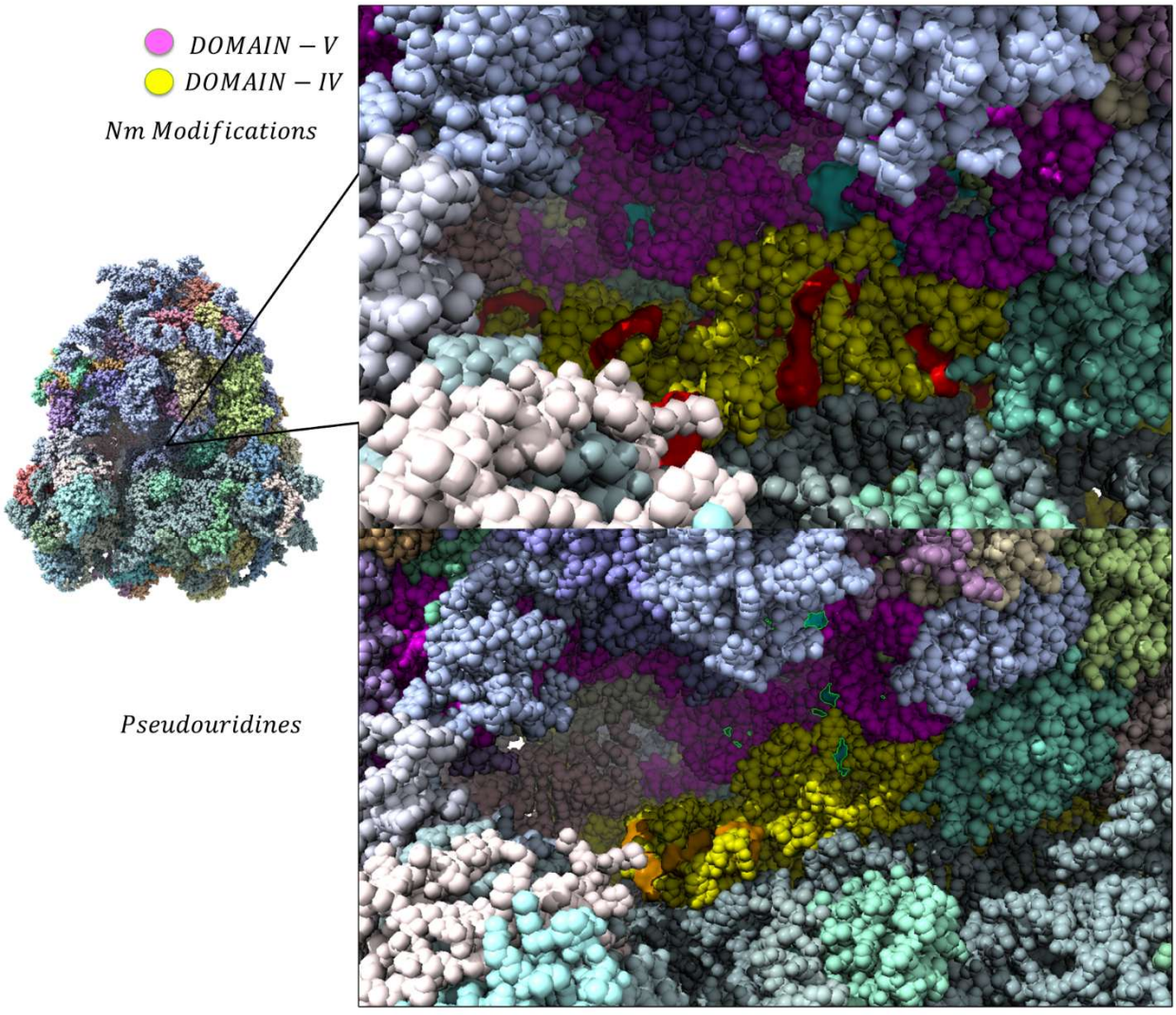
Taken together, domain IV displays a dense and conserved modification landscape, particularly around the PTC and intersubunit bridges. These features are consistent with roles in ribosome stability, catalytic coordination, and evolutionary adaptation of translational mechanisms. Moreover, the presence of pseudouridines conserved across all three domains of life underscores their evolutionary retention as structural stabilisers within the ribosome. Their transversal conservation suggests that, unlike other modifications that exhibit lineage-specific fixation, pseudouridylation contributes to fundamental ribosomal functions such as RNA helix packing, tRNA accommodation, and maintenance of catalytic architecture. These observations indicate that pseudouridines represent a core set of modifications subjected to universal evolutionary constraints, likely due to their biophysical role in ensuring ribosome integrity and function.

DOMAIN IV

A


A.thaliana
H.marismortui
E.coli
C.albicans
A.baumannii
S.lycopersicum
S.aureus
O.cuniculus
S.cerevisiae
H.sapiens
T.thermophilus
T.maritima
B.subtilis
Am
Cm
m3Y
Gm
Um
m2G
m5C
P
m5U
Ym
StackedBar
Percentage
Bar Segments

B


DOMAIN − V
DOMAIN − IV
Nm Modifications
Pseudouridines

DOMAIN IV

C

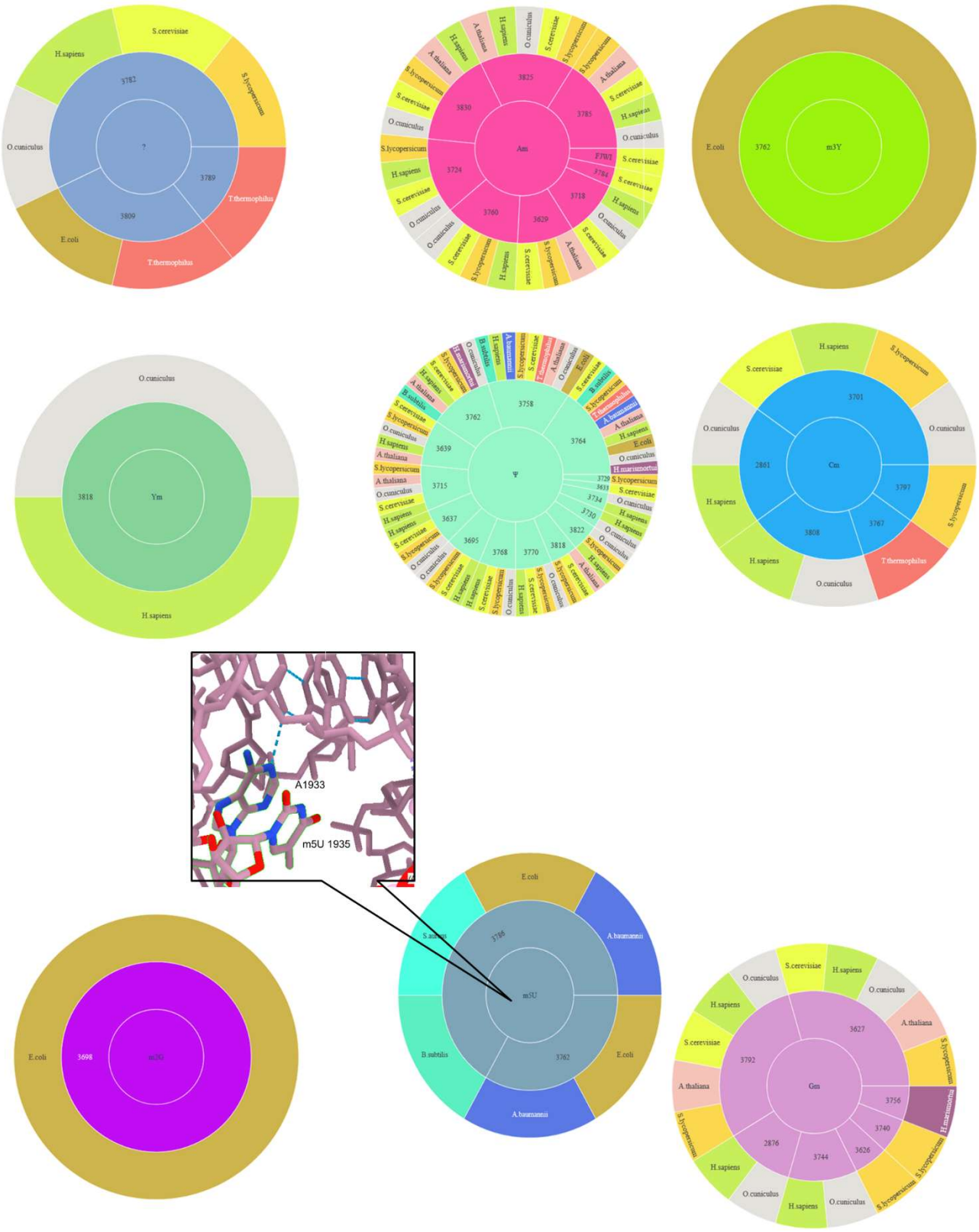


Figure 17 Distribution of RNA modifications in DOMAIN-IV. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Zoom over Peptdyl Transferase Centre: 4UG0 from RCSB has been used. (**C**) Mapping of RNA modifications to Human rRNA 28S.

3.1.6 Analysis of rRNA-aligned sequences: domain V

Domain V of the 23S rRNA harbours the peptidyl transferase centre (PTC), the catalytic site responsible for peptide bond formation. Beyond catalysis, domain V also accommodates multiple antibiotics targeting translation and forms the recognition substrate for ErmE methyltransferase, which methylates A_{2058} to confer macrolide resistance. The primary and secondary structures of this domain are sufficient to enable methyltransferase recognition in isolation, highlighting its autonomous biochemical identity (Vester et al., 1994). Moreover, the PTC modulates nascent peptide folding, underscoring the multifunctionality of domain V beyond enzymatic activity (Samanta et al., 2008).

To assess the post-transcriptional modification landscape, heatmaps and bar plots (**Fig. 18A**) were generated using R scripts. The data show that 2′-*O*-ribose methylation (N_m) is the dominant modification in domain V, especially in bacteria, where it comprises 100% of the annotated modifications in *E. coli*, *T. thermophilus*, and *A. baumannii*. In contrast, eukaryotic *taxa* such as *S. cerevisiae* and *H. sapiens* exhibit a broader modification palette, including substantial fractions of 5-methylcytidine (m^5C) and *N*^{6}-methyladenosine (m^6A), the latter observed in *H. sapiens* and *O. cuniculus*.

To identify conserved nucleotide positions, all sites were mapped to the *H. sapiens* 28S rRNA reference (**Fig. 18B**). Most N_m modifications fall directly within the PTC, where they participate in nucleotide interactions rather than protein contacts, suggesting their contribution to RNA helix packing and local conformational regulation (Yang et al., 2022). Pseudouridines follow a similar distribution, reinforcing the continuity between domains IV and V and indicating conserved mechanisms for stabilising the catalytic core.

An interspecies conserved nucleotide is G_{4196} (*E. coli* G_{2251}), which carries a 2′-*O*-methylguanosine installed by RlmB before ribosome assembly. This residue is conserved in *H. sapiens*, *S. cerevisiae*, *T. thermophilus*, *A. baumannii*, and other lineages. Located in helix 80 near the PTC, $G_{(m\text{-}2251)}$ contributes to structural stabilisation, particularly under thermal stress, as inferred from its conservation in *T. thermophilus* (Lövgren et al., 2001;Petrov et al., 2014)

Another conserved modification is $U_{(m\text{-}4498)}$ (*E. coli* U_{2552}), observed across *H. sapiens*, *S. cerevisiae*, *O. cuniculus*, *T. thermophilus*, and *A. baumannii*. In *E. coli*, it is introduced by the RrmJ methyltransferase during ribosome maturation (Caldas et al., 2000). Although its deletion does not affect viability under standard conditions, its evolutionary conservation suggests a structural role in refining the PTC geometry and protecting catalytic RNA from degradation.

G_{4391} (*E. coli* G_{2445}) carries an N^2-methylguanosine introduced by RlmL (YcbY), positioned within helix 74. Although its direct role in translation remains unclear, this GC-rich region contributes to ribosome integrity, and the conservation of m^2G_{2445} supports a function in LSU structural stability (Kimura et al., 2011). Similarly, A_{4449} (*E. coli* A_{2503}) carries an m^2A installed by RlmN. This modification resides within the PTC and has been implicated in fine-tuning translation and modulating antibiotic susceptibility (Toh et al., 2008).

ho^5C_{2501} (h_{4451} in *H. sapiens*) is uniquely present in *E. coli* and is catalysed by RlhA. The modification increases during stationary phase and oxidative stress, reducing protein synthesis while enhancing survival, suggesting a stress-responsive role (Fasnacht et al., 2022).

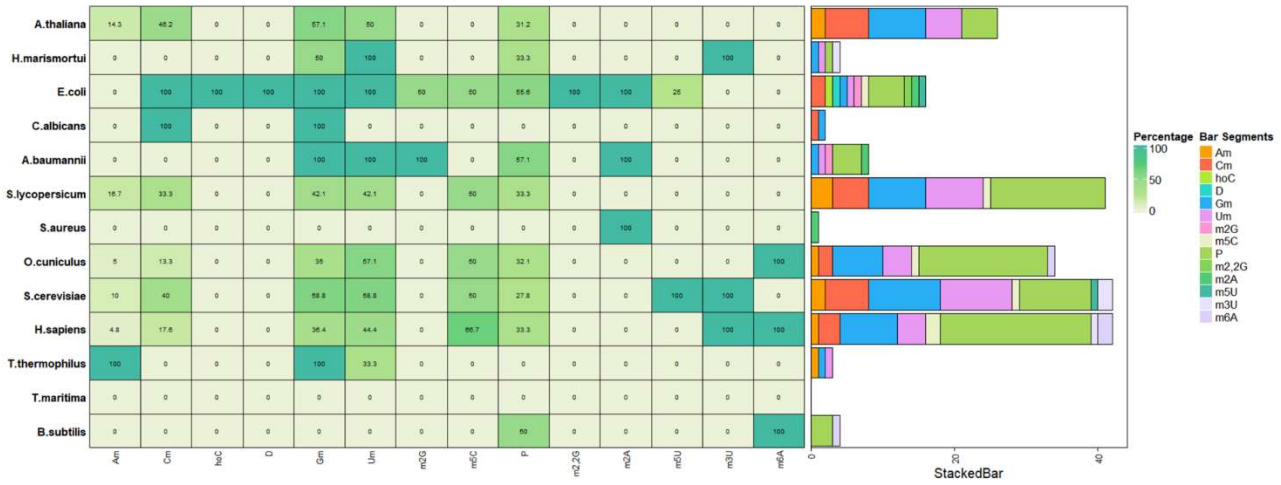
The 3-methyluridine (m^3U) at *H. marismortui* 2619 (human position 4500) is conserved in *O. cuniculus* and *S. lycopersicum*. It localises to the PTC and may stabilise tertiary interactions important for ribosome function (Petrov et al., 2014).

As a specific hallmark, *E. coli* shows a dihydrouridine at position 2450 (human position h_{4395}), located within the large central loop of domain V and adjacent to the PTC (Toubdji et al., 2024). Dihydrouridine is known to introduce conformational flexibility by disrupting base stacking and reducing local rigidity. Given that the large central loop anchors the 3′ ends of A- and P-site tRNAs and undergoes conformational rearrangements during peptide bond formation, the presence of dihydrouridine at this site may facilitate the necessary plasticity for tRNA accommodation and catalytic alignment. By softening the RNA backbone, the modification may support dynamic transitions within the PTC, contributing to efficient substrate positioning, ribosome processivity, or transitions into translocation.

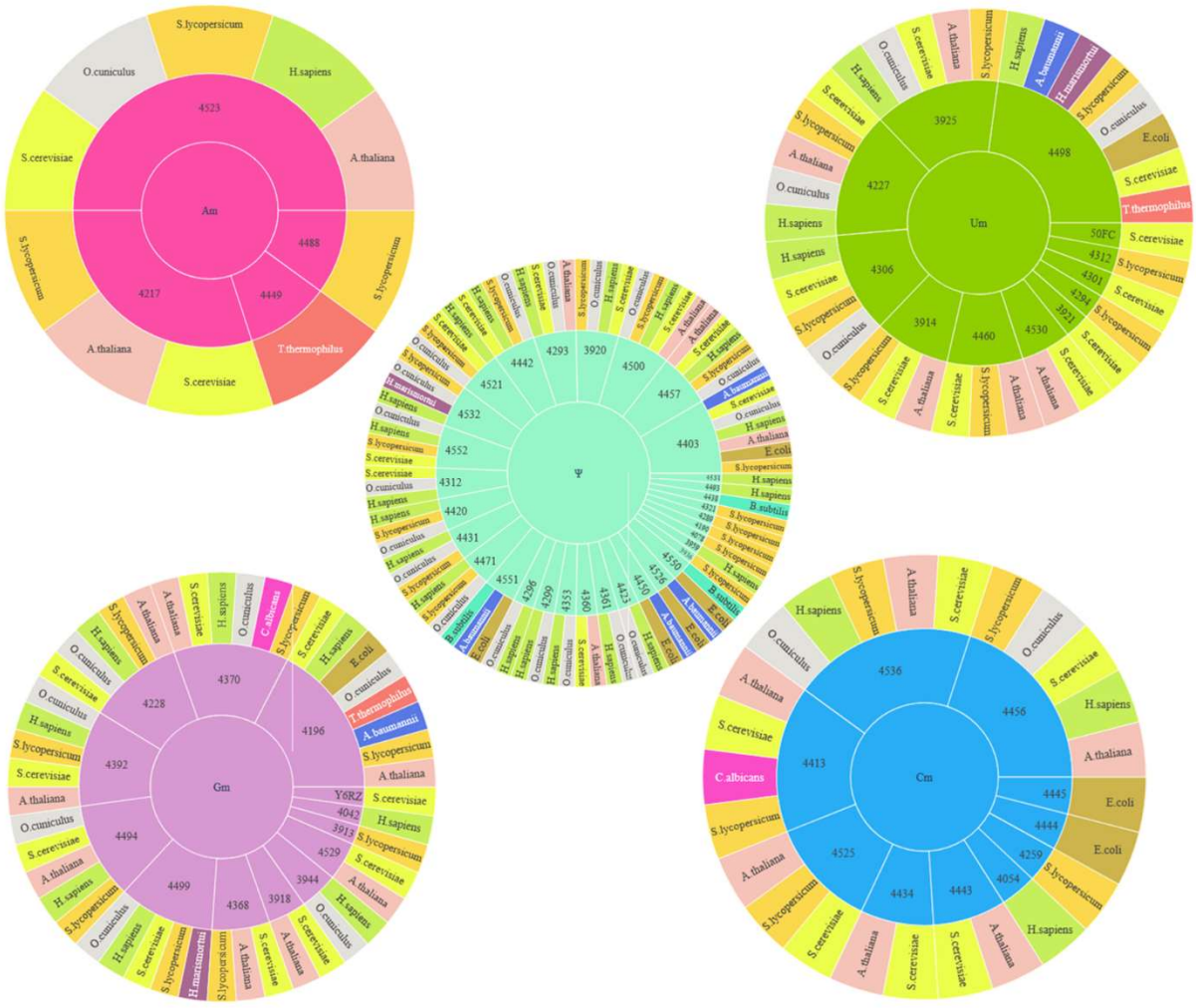
Altogether, domain V exhibits a densely modified landscape centred on the PTC and adjacent helices. Transversally conserved marks such as $G_{(m\text{-}4196)}$ and $U_{(m\text{-}4498)}$ suggest evolutionary selection for structural robustness under diverse conditions. Their persistence across phylogenetic boundaries reflects the conserved necessity of local rRNA stabilisation to sustain efficient and accurate translation.

DOMAIN V

A


A.thaliana
H.marismortui
E.coli
C.albicans
A.baumannii
S.lycopersicum
S.aureus
O.cuniculus
S.cerevisiae
H.sapiens
T.thermophilus
T.maritima
B.subtilis
Am
Cm
hoC
D
Gm
Um
m2G
m5C
P
m2,2G
m2A
m5U
m3U
m6A
StackedBar
Percentage
Bar Segments

B


Am
Ψ
Um
Gm
Cm

DOMAIN V

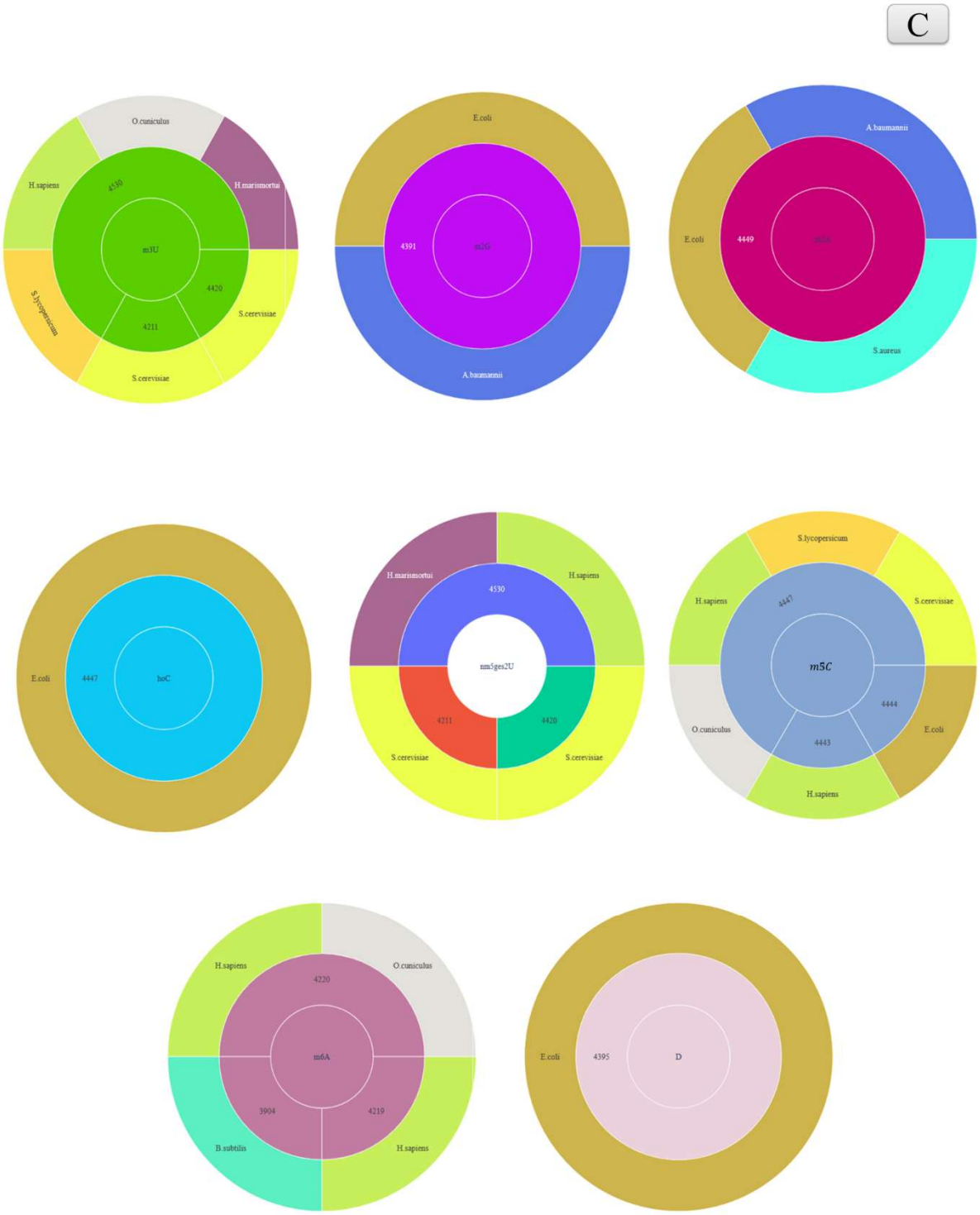


Figure 18 Distribution of RNA modifications in DOMAIN-V. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S, first half. (**C**) Mapping of RNA modifications to Human rRNA 28S, second half.

3.1.7 Analysis of rRNA-aligned sequences: domain VI

To investigate the post-transcriptional landscape of domain VI, R scripts were used to quantify and visualise modification patterns across taxa. Domain VI marks the terminal region of the large ribosomal subunit in Bacteria, Archaea, and Eukarya, and includes helix 95, which hosts the sarcin–ricin loop (SRL) in *E. coli*. The SRL is essential for GTPase activation during elongation and remains structurally conserved. However, in terms of RNA modifications, domain VI exhibits sparse conservation. **Figure 19**(**A**) reveals that chemical modifications are largely restricted to eukaryotic *taxa—H. sapiens*, *O. cuniculus*, and *S. lycopersicum*—with bacterial and archaeal sequences lacking detectable events in this region.

A_m modifications at positions h_{4571} and h_{4590} are observed in *H. sapiens* and *O. cuniculus*, while G_m occurs at h_{4618}, h_{4623}, and h_{4637} in the same two species. Two U_m modifications, located at h_{4620} and h_{4984}, are shared across *H. sapiens*, *O. cuniculus*, and *S. lycopersicum*, suggesting a broader phylogenetic retention. Pseudouridine positions (Ψ) were mapped to *H. sapiens* 28S coordinates and include Ψ_{4628}, Ψ_{4579}, Ψ_{4576}, Ψ_{4689}, Ψ_{4973}, Ψ_{5010}, Ψ_{5001}, Ψ_{4972}, Ψ_{4673}, and Ψ_{4636}. Among these, Ψ_{4689} is conserved across *S. lycopersicum*, *O. cuniculus*, and *H. sapiens*; Ψ_{4973} appears restricted to *H. sapiens*; while the remaining positions are shared between mammals.

Structural mapping in **Figure 19**(**B**) shows that pseudouridines localise to the external ribosomal surface near L3 and L29, without forming direct contacts with ribosomal proteins. Their distribution suggests a role in maintaining the local RNA topology or modulating peripheral rRNA–protein recognition. In contrast, N_m modifications reside near the peptide exit tunnel, adjacent to ribosomal protein L23 (Contreras-Martinez et al., 2011). These sites form dense hydrogen-bond networks with neighbouring residues, possibly contributing to local helix stabilisation or preserving tunnel geometry.

Despite the relatively low number of modifications, the phylogenetic retention of specific U_m and Ψ positions indicates evolutionary pressures to maintain structural constraints in select *taxa*. Notably, across the large subunit, domain IV exhibits the most pronounced expansion in eukaryotes relative to bacterial and archaeal homologues, highlighting a possible link between structural enlargement and the emergence of lineage-specific functional innovations. The consistent placement of N_m sites in structurally sensitive regions, including the exit tunnel interface, further supports a role for chemical modifications in reinforcing ribosomal conformation during dynamic stages of translation. Domain VI thus exemplifies a scenario where

limited but strategically positioned modifications contribute to architectural integrity and the phylogenetic fine-tuning of ribosome function.

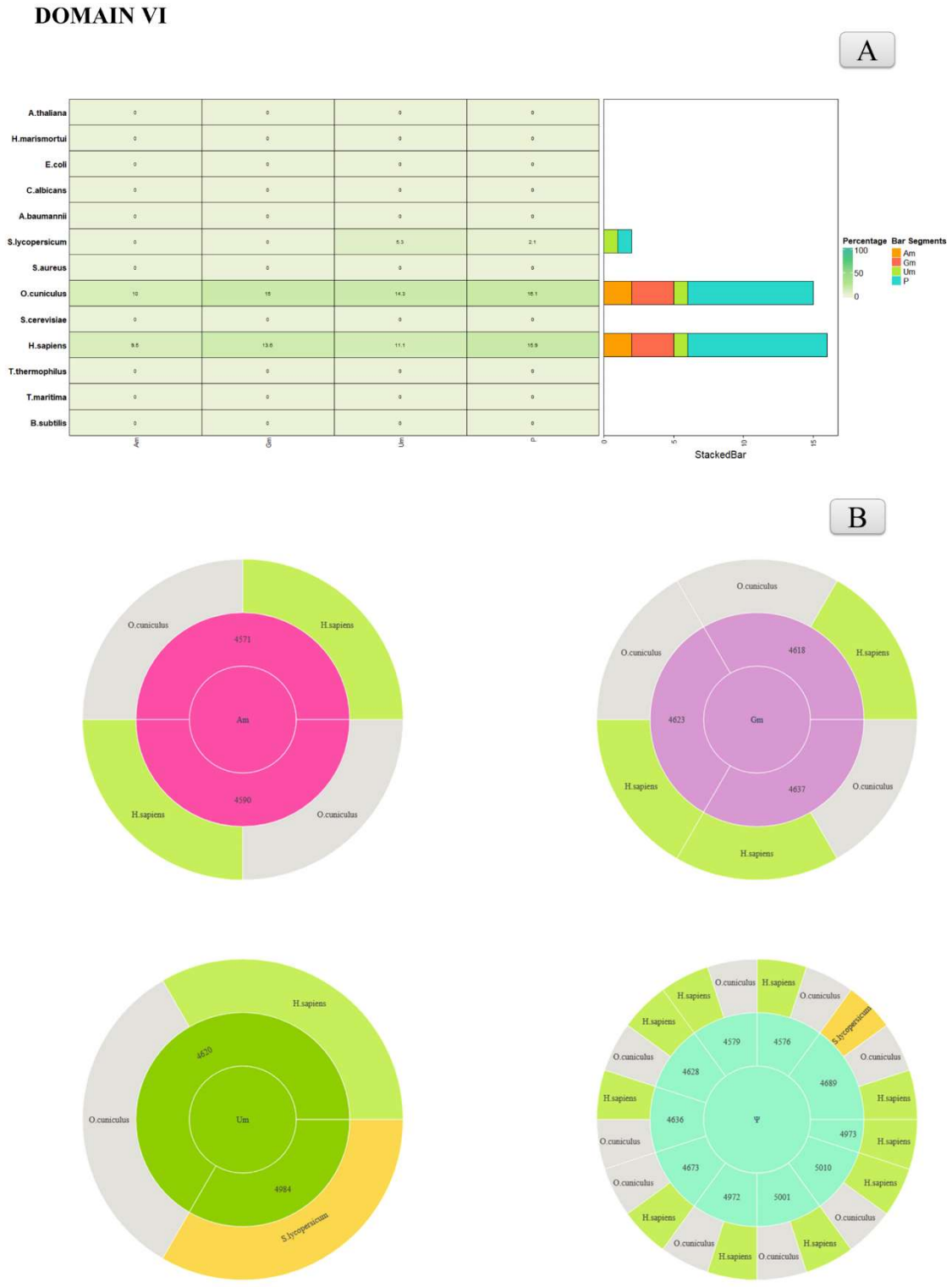


Figure 19 Distribution of RNA modifications in DOMAIN-VI. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S.

3.1.8 Analysis of rRNA-aligned sequences: domain 0

To investigate the post-transcriptional landscape of domain 0, R scripts were employed to generate heatmaps and stacked bar plots quantifying the percentage and type of RNA modifications across the selected species (**Fig. 20A**). Domain 0 serves as the scaffold from which the six structural domains of the large ribosomal subunit radiate, anchoring the core rRNA framework. This central domain is highly conserved across all domains of life, consistent with a primary role in early ribosome assembly and structural coordination.

The heatmap analysis reveals that 2′-*O*-methylated nucleotides (N_m) dominate the modification profile in eukaryotic species such as *H. sapiens*, *O. cuniculus*, and *S. cerevisiae*, where A_m, C_m, and G_m are the predominant chemical forms. These modifications appear absent in archaeal and bacterial taxa, suggesting a eukaryote-specific enrichment in structural complexity. Notably, *S. lycopersicum* shows a reduced density of modifications, with Ψ and C_m being the most represented events.

To visualise and quantify the conservation of individual modified positions, each modification site was mapped to human 28S rRNA coordinates using Python scripts, and the results were displayed using sunburst charts (**Fig. 20B**). One of the most conserved marks is *N*¹-methyladenosine (m^1A) at position 1322, observed in *H. sapiens*, *S. cerevisiae*, *H. marismortui*, *O. cuniculus*, and *S. lycopersicum*. This modification is catalysed by nucleomethylin (NML) in humans and Rrp8 in yeast. Its role is structurally validated: m^1A_{1322} stabilises the interface between eL32 and adjacent rRNA, with knockout mutants showing imbalanced translation and metabolic dysregulation (Sharma et al., 2018).

Interestingly, N_m residues cluster around this site within Helix 25.1, including G_{m1316}, A_{m1323}, A_{m1326}, and C_{m1327}. Several of these are conserved across Animalia and Plantae, indicating a potential methylation hotspot that supports the stability of the interdomain interface between domains 0 and V.

In addition, *N*⁶-methyladenosine (m^6A) is detected at human position 3876, corresponding to *E. coli* 2031 and *A. baumannii* 1945. Although this mark lies within the linear boundaries of domain 0, its structural location places it inside domain V, close to the peptidyl transferase centre (PTC). The RlmF methyltransferase installs m^6A_{2031} early during 50S subunit maturation. This modification modulates resistance to erythromycin and stabilises local RNA architecture, with deletion mutants showing impaired initiation and stress sensitivity (Deng et al., 2015).

DOMAIN 0

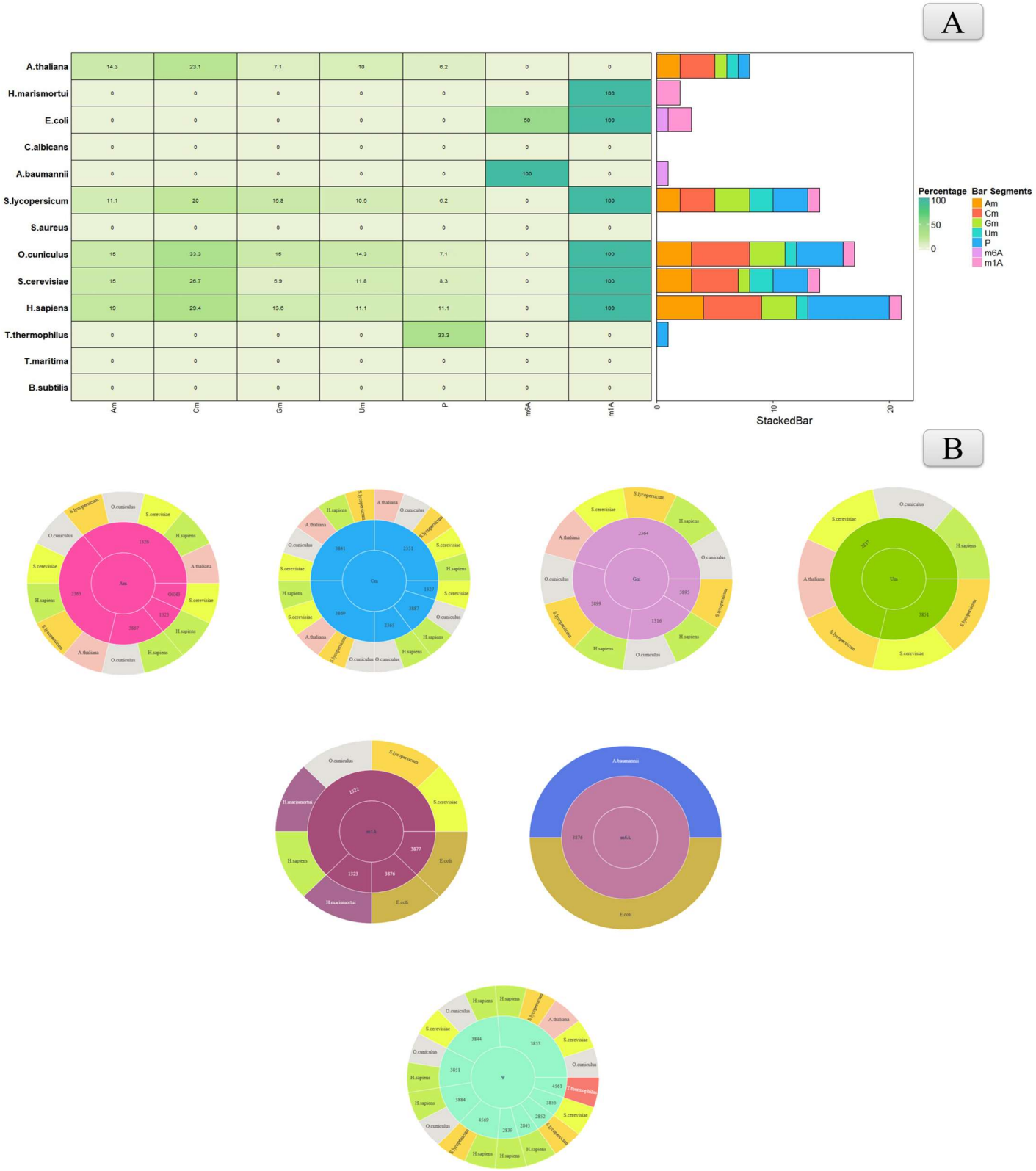


Figure 20 Distribution of RNA modifications in DOMAIN-V. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S.

Additional evidence for structural relevance arises from a constellation of N_m sites—$C_{(m\text{-}3841)}$, $U_{(m\text{-}3851)}$, $C_{(m\text{-}3869)}$, and $C_{(m\text{-}3899)}$—clustered near m^6A_{3876} and conserved across *S. cerevisiae*, *H. sapiens*, *O. cuniculus*, *S. lycopersicum*, and *A. thaliana*. This pattern suggests a coordinated network of methylations, despite being linearly positioned in domain 0, that may reinforce structural stability at the PTC entry site, aligning with the hypothesis that domain 0 contributes to fine-tuning the ribosome's catalytic core.

Taken together, domain 0 exhibits a species-specific and topologically strategic RNA modification pattern. Eukaryotic sequences show enhanced density and chemical diversity, including functionally validated marks like m^1A_{1322} and m^6A_{3876}, as well as flanking N_m residues that may act cooperatively to stabilise essential ribosomal domains. These findings point to architectural and regulatory roles for domain 0 modifications, both at the evolutionary and biochemical levels.

3.1.9 Analysis of rRNA-aligned sequences: short subunit 5′ domain

To evaluate the landscape of rRNA modifications among the small subunits (SSU) deposited in the MODOMICS database (Cappannini et al., 2023), R scripts were used to extract and quantify modification distributions for each SSU domain across the available species, starting from the 5′ domain. The 5′ domain of the 16S/18S rRNA contributes to ribosome biogenesis and structural organisation. During nucleolar maturation, this domain undergoes conformational transitions, aligning with the central domain to reach its final architecture, while the 3′ Major domain remains embedded in the SSU processome core (Oborská-Oplová et al., 2022).

The heatmap and bar plots (HWB, **Fig. 21A**) display an enrichment of RNA modifications predominantly in the 5′ domain of eukaryotic species and *P.abyssi*. Notably, 2′-*O*-methylated nucleotides (N_m) cluster within this region, particularly in accessible, solvent-exposed helices. Indeed, sunburst visualisations (**Fig. 21B**) mapping modification positions to human 18S rRNA coordinates revealed a scattered but functionally patterned distribution of pseudouridines. While most sites are species-specific, one conserved pseudouridine corresponds to *E. coli* Ψ_{516} (h_{612}), introduced by the enzyme RsuA. This modification contributes to translation fidelity by stabilising the 530-loop (helix 18 in human), a structural motif involved in codon recognition and tRNA accommodation during protein synthesis. Despite these roles, bacterial ribosomes show resilience to the deletion of Ψ residues, suggesting lower reliance on these modifications compared to eukaryotes (Petrov et al., 2014; O'Connor et al., 2018).

N_m modifications follow a similar distribution, contributing to local and global rRNA stability. For instance, A_m stabilises ribose conformation in the 3′-*endo* geometry,

thereby influencing ribosomal folding and interactions with translation machinery (Khoshnevis et al., 2022). A key example is $A_{(m-436)}$ (h_{484}), located in helix h15. Helix 15 is buried and part of the inner core of the 5′ domain and contributes to maintaining the compact structure of rRNA backbone and acting as a scaffold for the binding of ribosomal proteins (Woodson, 2008). Using *S.cerevisiae* as a template, conserved A_m sites include positions 28 (h_{27}), 100 (h_{99}) (Buchhaupt et al., 2014), 436 (h_{484}), 464 (h_{512}), and 541(h_{590}), spanning helices helix-5, helix-9, helix-15, helix-17, and helix-19, respectively (Petrov et al., 2014). Several are guided by conserved snoRNAs such as snR51 and snR87 and are annotated in curated modification databases (FOURNIER LAB: https://bit.ly/4lgxxDR).

Conserved 2′-*O*-methylated residues such as $C_{(m-517)}$ (helix 4), $C_{(m-462)}$ (helix 14), $C_{(m-174)}$ (helix 8), $G_{(m-388)}$ (helix 11), $G_{(m-644)}$ (helix 3), $G_{(m-436)}$ (helix 13), $G_{(m-509)}$ (helix 15), and $G_{(m-601)}$ (helix 18) are distributed across helices that constitute the compact structural core of the 5′ domain of the small subunit. These helices provide architectural stability for the ribosomal body and contribute to the correct spatial arrangement of functional domains, particularly the platform. Methylation at these positions may enhance local base stacking, reduce conformational fluctuations, and support folding geometry required during pre-initiation complex formation. In solvent-exposed helices such as h3 and h18, 2′-*O*-methylation may also reduce ribose–solvent interactions, limiting local reactivity with water and enhancing structural persistence (Leroy et al., 2025; Höfler et al., 2020). Similarly, 2′-*O*-methyluridine residues—$U_{(m-627)}$ (helix 18), $U_{(m-428)}$ (helix 12), $U_{(m-354)}$ (helix 4), $U_{(m-306)}$ (helix 9), $U_{(m-172)}$ (helix 8), and $U_{(m-121)}$ (helix 7)—map to loop or peripheral regions where they may fine-tune local RNA dynamics or facilitate subunit assembly through modulation of RNA flexibility and molecular interactions.

Pseudouridines such as (helix 3) Ψ_{651}, (helix 18) Ψ_{612}, (helix 11) Ψ_{406}, (helix 7) Ψ_{119}, and (helix 6a) Ψ_{109} further support the structural integrity of the 5′ domain, with Ψ_{612} directly positioned at the body–platform junction. This residue lies adjacent to several 2′-*O*-methylated sites and may contribute to the spatial organisation required for platform alignment and subunit compaction (Gilbert et al., 2007). Ψ_{651} and Ψ_{109}, located in solvent-exposed regions of helices h3 and h6a, may stabilise external curvature and maintain surface topology important for ribosomal protein contacts. In contrast, Ψ_{406} and Ψ_{119} are buried within internal helices h11 and h7, where they likely enhance local RNA folding stability through improved base stacking and conformational rigidity. Collectively, these pseudouridines contribute to the mechanical resilience of the ribosomal body and its readiness for initiation complex assembly (Petrov et al., 2014).

Among Archaea, modifications appear sporadically. Nevertheless, their sequence conservation suggests functional retention. For example, A_m residues unique to *P. abyssi* are conserved in aligned positions (see Supplementary Sheet SSU_ACROSS_DOMAINS). Additionally, ac^4C modifications detected in *P. abyssi*—reported in 34 sites with a preference for CCG helices—enhance Watson–Crick base pairing and stacking interactions, promoting hyper-thermostability of the ribosome (Coureux et al., 2020).

Overall, the 5′ domain of the 18S rRNA contains a dense cluster of conserved 2′-*O*-methylated and pseudouridylated residues across eukaryotes and archaea. These modifications localise to helices involved in early subunit folding, structural compaction, and the formation of the body-platform architecture. Their distribution suggests a role in maintaining RNA stability, controlling local flexibility, and preserving the geometry of rRNA domains critical for ribosome assembly and initiation.

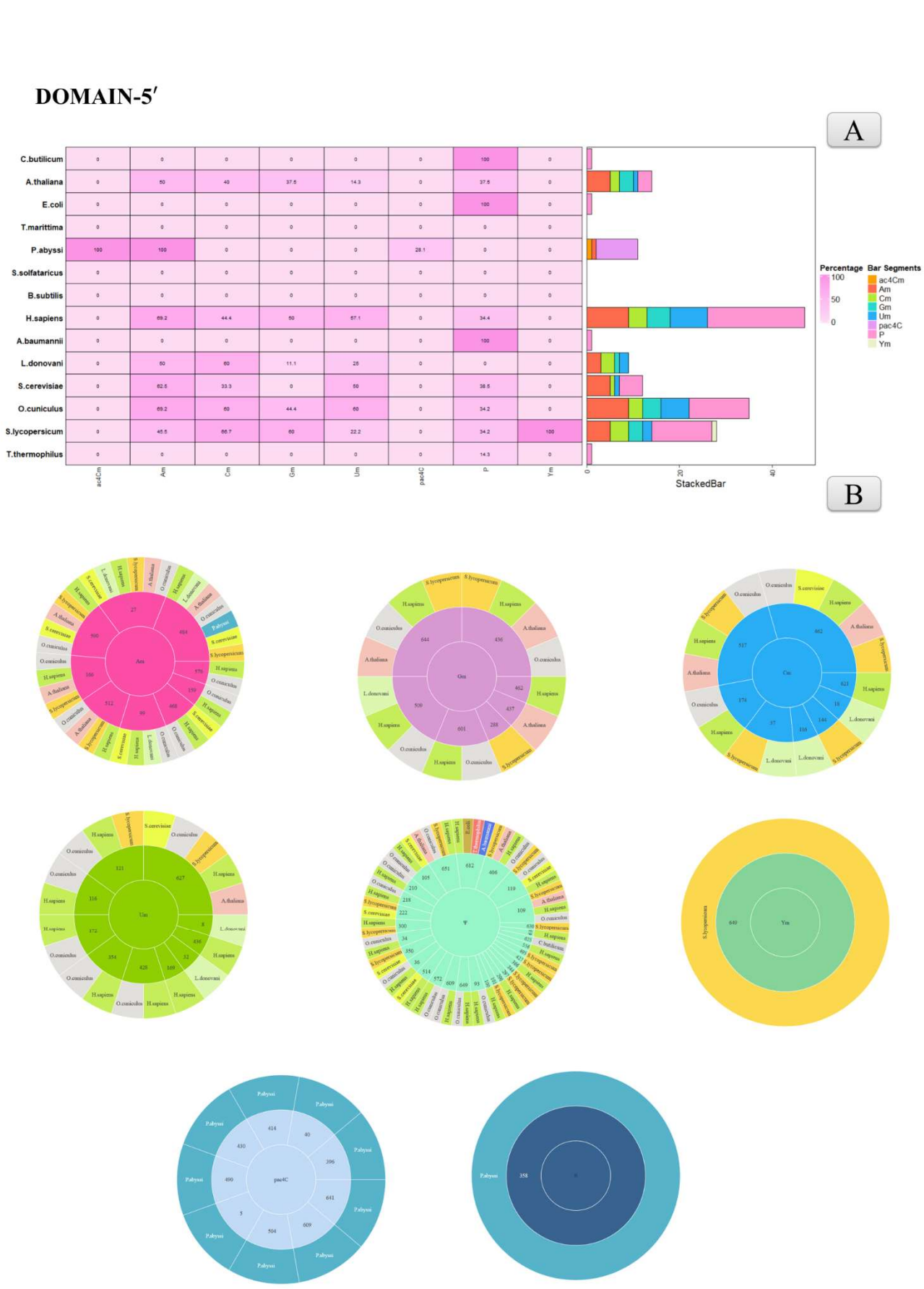


Figure 21 Distribution of RNA modifications in DOMAIN-5′. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S.

3.1.10 Analysis of rRNA-aligned sequences: short subunit central domain

Domain-C is a structurally distinct, curved element composed of nine helices. It folds into the platform region of the SSU and includes the junction of helices 20, 21, and 22 as its architectural pivot (Sun et al., 2017). During ribosome biogenesis, domain-C assembles independently via Mg^{2+} coordination and hierarchical protein binding. Together with the 5′ and 3′ domains, it contributes to the modular architecture of the 30S subunit, with ribosomal proteins binding in ordered temporal phases (Xu et al., 2010).

To explore the distribution of RNA modifications within the central domain (domain-C) of the 16S/18S rRNA, R scripts were used to extract species-wise counts and compute relative percentages across aligned SSU sequences. The resulting heatmap and stacked bar plot (HWB, **Fig. 22A**) display the frequency and compositional diversity of post-transcriptional modifications across *taxa*. In parallel, modification positions were mapped to the human 18S reference sequence and visualised using Python-generated sunburst plots (**Fig. 22B**), allowing phylogenetic tracking of site-specific events.

The HWB analysis reveals that RNA modifications in domain-C are predominantly restricted to Eukarya. The archaeon *P. abyssi* displays a distinct repertoire dominated by ac^4C but lacking modifications detectable pseudouridylation or 2-*O*-methylation events in this domain. In contrast, eukaryotic *taxa* show extensive N_m and Ψ modifications, many of which appear at conserved *loci*. The sunburst charts confirm that A_m, G_m, C_m, and Ψ are consistently installed across eukaryotic species, suggesting a role in stabilising the rRNA platform architecture and preserving inter-helix geometry required for subunit assembly and mRNA positioning.

For instance, $A_{(m\text{-}854)}$ (*S.cerevisiae*, 796) and $A_{(m\text{-}1032)}$ **(***S.cerevisiae*, 974**)** are located in the SSU body and decoding centre, respectively. These 2′-*O*-methylations promote structural stability by restricting ribose flexibility and enforcing the 3′-*endo* sugar pucker conformation, thereby strengthening RNA helix packing and improving decoding fidelity (Khoshnevis et al., 2022). Although $A_{(m\text{-}854)}$ is not annotated in the MODOMICS sequence, its relevance is experimentally established in human ribosomes. The aligned sequences of *H.sapiens* to *S.cerevisiae* reveal that $A_{(m\text{-}668)}$ (*S.cerevisiae*, 619) lies near the base of helix 21 (Ben-Shem et al., 2011) and is retained across plants, fungi, and metazoans. Its conservation at the junction of helices 20–22 suggests a structural role in maintaining the three-way junction geometry of the platform (Ben-Shem et al., 2011).

Additional N_m sites include $G_{(m-683)}$, $G_{(m-867)}$, and $G_{(m-1184)}$ (not conserved in human) which are positioned in or near helices 21 and 24. These regions contact ribosomal proteins such as uS13 and uS19 and contribute to shaping the neck and platform interfaces of the small subunit. Given their placement in structured helices adjacent to ribosomal proteins, these G_m residues may stabilise RNA–protein interfaces and preserve domain topology under translation-coupled stress (Taoka et al., 2018).

Pseudouridine residues such as Ψ_{681}, Ψ_{814}, Ψ_{815}, Ψ_{822}, Ψ_{863}, Ψ_{966}, Ψ_{1005}, Ψ_{1057}, and Ψ_{1175} are distributed across helices h20 to h27 in the central domain. These positions cluster within the structural platform, a region that anchors tertiary contacts and connects the body of the small subunit to the neck and head (Culver et al., 2008). The modification sites localise to helices involved in scaffold formation (h21–h23), surface exposure (h24), and interdomain transitions (h25–h27). Their distribution suggests a role in stabilising local RNA geometry and maintaining compactness required for pre-initiation complex assembly. For instance, Ψ_{1075} and Ψ_{1082} lie near the h24–h25 junction, where they may reinforce RNA folding across the platform–neck interface, while Ψ_{681} and Ψ_{814} are positioned in h21, a core helix that coordinates early protein interactions and domain organisation (Ofengand & Bakin, 1995; Taoka et al., 2018).

Other clustered N_m sites are found near ribosomal protein S23. While these sites do not form direct contacts with S23, their proximity suggests a role in sculpting the surrounding 3D RNA framework, especially in regions critical for decoding factor coordination. This aligns with the view that N_m modifications contribute not only to stabilisation but also to translational efficiency and elongation control (Khoshnevis et al., 2022).

Species-specific N_m and Ψ events—such as $G_{(m-1184)}$ or Ψ_{1082}—may reflect functional adaptation to distinct translational environments, particularly in metazoan or plant lineages. Their recurrent appearance near the platform's outer surface or at the edge of conserved helices suggests roles in ribosome plasticity and context-specific modulation.

Overall, domain-C represents a chemically diverse region in eukaryotic SSUs. The combination of conserved A_m and G_m sites with selectively retained pseudouridines underlines an evolutionary pressure to reinforce structural integrity in decoding-relevant regions. Prokaryotic sequences lack this diversity, suggesting a lower dependency on fine-tuned rRNA modification for platform stability.

DOMAIN-C

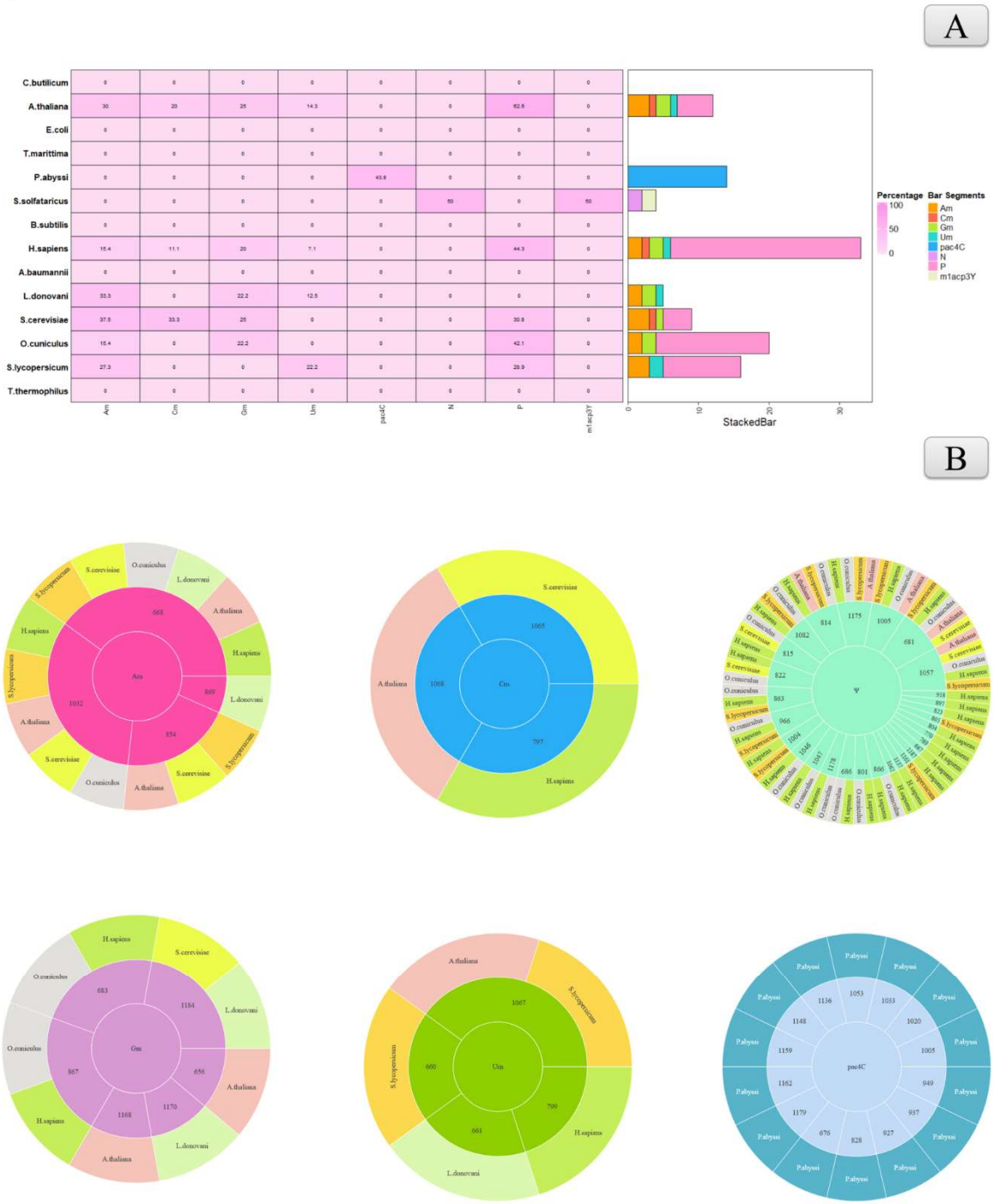


Figure 22 Distribution of RNA modifications in DOMAIN-C. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 18S.

3.1.11 Analysis of rRNA-aligned sequences: short subunit 3′ major domain

The 3′ Major (3′ M) domain of the 18S rRNA comprises 15 helices and forms the structural head of the small ribosomal subunit (BIOCHEM EDU: https://bit.ly/4aToDHI). To assess the post-transcriptional landscape of this domain, a heatmap was generated using R (**Fig. 23A**), quantifying modification distribution across species. Modifications were subsequently mapped to human 18S rRNA positions (h28-h43, 16 helices) via Python, and conserved sites were visualised using sunburst charts (**Fig. 23B–C**).

G_m and U_m modifications are distributed across both the 3′ Major and central domains, accounting for a substantial proportion of the total marks in several *taxa*. These methylations localise near ribosomal proteins S27A, S19, S18, and S9, in regions consistent with protein–RNA anchoring sites. Pseudouridines show a similar structural arrangement, clustering at helical junctions likely involved in stabilising protein interfaces in eukaryotic ribosomes.

Mapping of conserved N_m residues in the 3′ major domain reveals a recurring pattern of modification across distant eukaryotic *taxa*, including *H. sapiens*, *O. cuniculus*, *S. cerevisiae*, *A. thaliana*, *S. lycopersicum*, and *L. donovani*. $A_{(m\text{-}1384)}$ (helix 37) and $A_{(m\text{-}1679)}$ (helix 43) are conserved in plants and mammals and reside at the head-body interface and domain periphery (Montpetit et al., 1998), respectively, where they may stabilise rotational alignment and maintain peripheral scaffold integrity. $C_{(m\text{-}1273)}$ (helix 33) and $C_{(m\text{-}1392)}$ (helix 38), found in *H. sapiens*, *O. cuniculus*, and *S. lycopersicum*, localise within internal helices of the head, suggesting a function in limiting local fluctuations under mechanical strain. $G_{(m\text{-}1329)}$, conserved in six species including yeast and kinetoplastids, is situated in helix 33 and aligns with a structurally conserved core that supports decoding-centre positioning. $G_{(m\text{-}1448)}$ (helix 39) and $G_{(m\text{-}1491)}$ (helix 34), present in *H. sapiens* and *O. cuniculus*, occur at decoding-adjacent helices, reinforcing regions known to affect decoding fidelity in *E. coli* (O'Connor et al., 1997).

However, several U_m residues such as $U_{(m\text{-}1443)}$, $U_{(m\text{-}1327)}$, $U_{(m\text{-}1320)}$, and $U_{(m\text{-}1505)}$ (helices 34 and 39) are more broadly conserved, spanning five or more species, and cluster at decoding-proximal positions where RNA backbone orientation is critical. $U_{(m\text{-}1289)}$ and $C_{(m\text{-}1273)}$, found in helix 33, are positioned internally and may contribute instead to intra-domain rigidity, especially during head swivelling. Therefore, the conservation and spatial restriction of these N_m modifications suggest selective pressure to maintain local rRNA conformation at key structural and functional interfaces. Their correspondence with regions sensitive to mutational decoding errors (O'Connor et al., 1997) and high stoichiometry in human ribosomes (Taoka et al., 2018) further supports a role in fine-tuning translation through structural stabilisation rather than regulatory flexibility.

Rare hyper-modifications were also detected. The central domain includes $m^1acp^3\Psi$ in *S. lycopersicum* and m^3C in *H. sapiens* (see HWB). Ac^4C_{1338}, found in helix 34 and conserved in *S. cerevisiae*, *H. sapiens*, *S. lycopersicum*, and *O. cuniculus*, stabilises the *C3′-endo* ribose and enhances G–C base pairing. In *S. cerevisiae*, this mark contributes to decoding accuracy and translation factor binding (Yang et al., 2022).

The bacterial site h_{1250} hosts m^5C_{967} in *E. coli*, *A. baumannii*, *T. maritima*, *T. thermophilus*, *P. abyssi*, and *L. donovani*. In *E. coli*, this site is modified by RsmB and lies within a loop near the P-site tRNA. It cooperates with m^2G_{966} to modulate helix 31 stability and regulate start codon recognition (Tscherne et al., 1998; Burakovsky et al., 2012; Pletnev et al., 2020; Guymon et al., 2007). In *T. thermophilus*, the equivalent residue carries $m^{2,2}G$, supporting initiation fidelity under thermophilic conditions (McCown et al., 2020). Together, these data define the 3′ Major domain as a focal region for structural reinforcement and translational tuning. Conservation of 2′-*O*-methylations and pseudouridines across distant *taxa* reflects selective pressures preserving decoding accuracy and subunit communication. The spatial co-occurrence of modifications with decoding helices and ribosomal protein contacts supports a structural logic in which chemical marks reinforce the architecture of the small subunit head. Lineage-specific and universally retained modifications follow a dual evolutionary principle: diversification for environmental responsiveness and conservation for functional integrity.

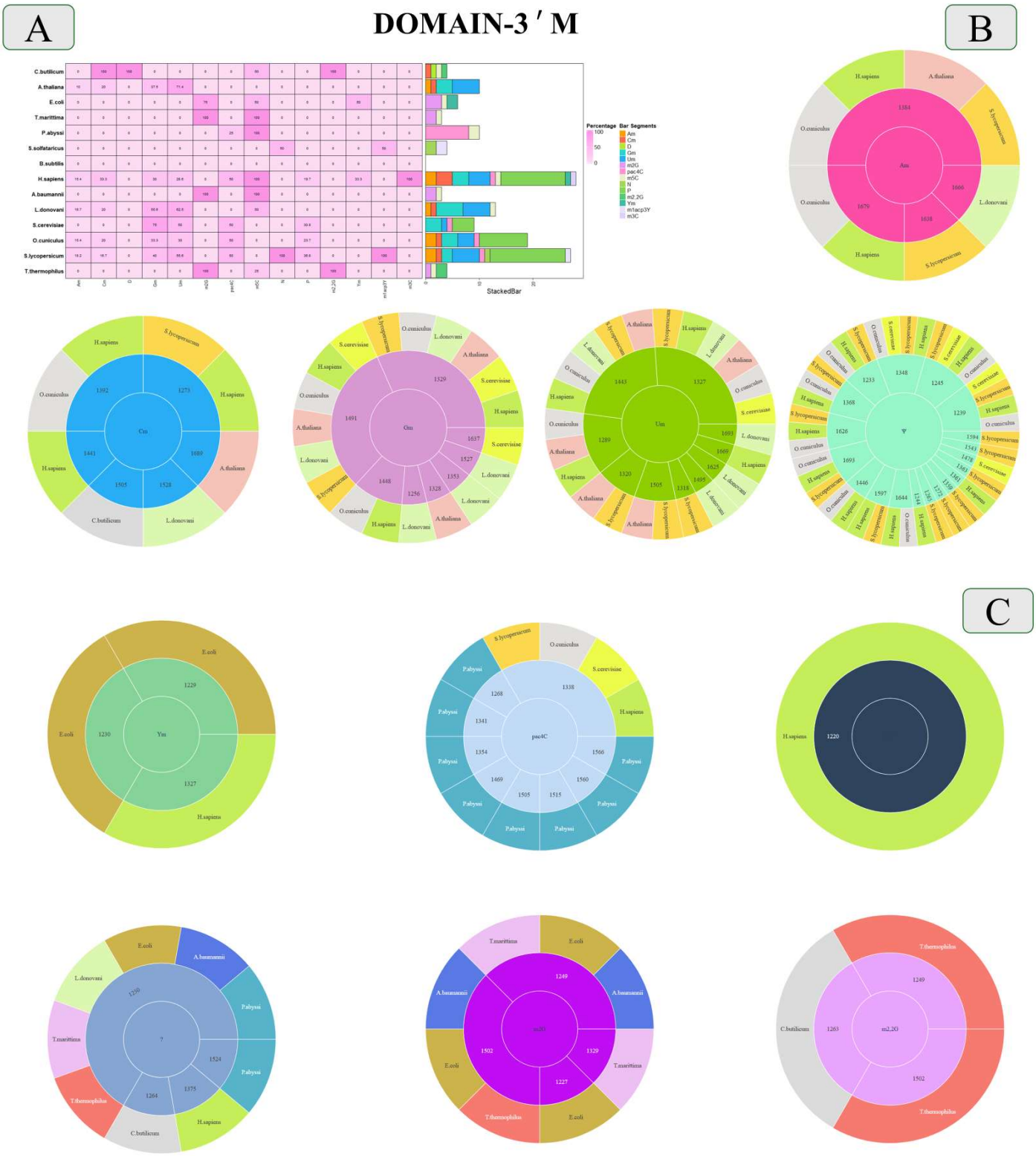


Figure 23 Distribution of RNA modifications in DOMAIN-3′M. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S, first half. (**C**) Mapping of RNA modifications to Human rRNA 28S, second half.

3.1.12 Analysis of rRNA-aligned sequences: short subunit 3′ minor domain

The 3′ minor domain of the 16/18S rRNA comprises two helices and extends from the small ribosomal subunit to contact the 50S subunit (BIOCHEM EDU: https://bit.ly/4aToDHI). Ribosomal protein S20 binds independently to both the 5′ and 3′ minor domains. Its interaction with the 5′ domain is extensive, while contacts with the 3′ minor domain are more limited. These two domains are organized relative to S20 at different stages of 30S subunit assembly (Dutca et al., 2008). Mutagenesis studies have shown that point mutations in the 3′ minor domain do not prevent assembly into 30S subunits or interaction with the 50S subunit, indicating resilience in ribosome assembly (Jemiolo et al., 1985).

A heatmap summarising RNA modifications across species (**Fig. 24A**) indicates that the 3′ minor domain contains fewer modifications than other SSU domains yet includes chemically distinct residues absent elsewhere. Most positions mapped to human 18S rRNA show limited conservation. However, a conserved modification was identified at h_{1704} —housed in helix 44 which forms a key structural element of the A site (aminoacyl-tRNA binding site) within the decoding centre (Maksimova et al., 2021; Maksimova et al., 2022; Qin et al., 2012)— corresponding to position 1402 in *E. coli*, where N^4, 2′-*O*-dimethylcytidine ($m^4C_{(m\text{-}1402)}$) is installed. Its biosynthesis involves sequential action by RsmH, which methylates the base, and RsmI, which catalyses ribose methylation. Noteworthy, C_{1704} is conserved across of the life domains where the eukaryotic species harbour a 2'-*O*-methylcytidine modification, suggesting some degree of evolutionary preservation and similar role in stabilising the decoding centre.

Near the end of helix 44, a discrepancy was noted in MODOMICS at position h_{1831}. Several organisms, including *T. thermophilus* and *E. coli*, carry 5-aminomethyl-2-geranylthiouridine (nm^5ges^2U) at this site, yet isolates of the same species contain 3-methyluridine. In *E. coli*, this mark is introduced by the RsmE enzyme and influences decoding centre regulation, potentially affecting translation fidelity (McCown et al., 2020). Current MODOMICS pathways do not list m^3U as an intermediate in the biosynthesis of nm^5ges^2U, and the presence of the latter at this site remains unverified.

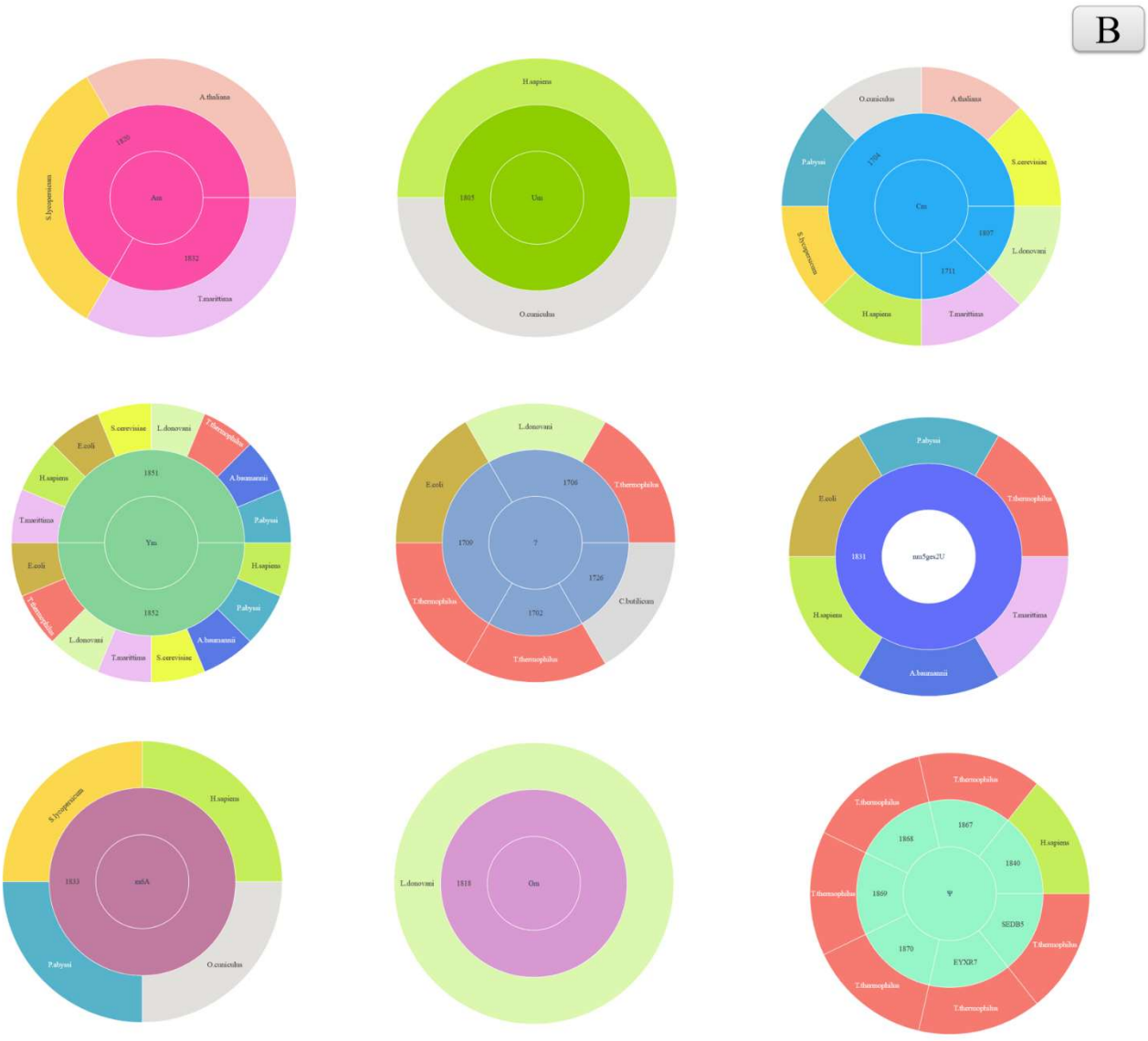
Both $m^4C_{(m\text{-}1402)}$ and m^3U_{1831} are in helix 44, a structural component of the ribosome's decoding centre. $m^4C_{(m\text{-}1402)}$ contributes to P-site stabilisation and enhances translation initiation accuracy by preserving the reading frame (Kimura et al., 2009). This modification is conserved in *T. thermophilus*, *T. maritima*, *A. baumannii*, and *L. donovani* (Polikanov et al., 2015), indicating selective retention across diverse *taxa*.

Taken together, the low density of conserved modifications in the 3′ minor domain contrasts with the chemically diverse marks observed at critical decoding-centre

positions. The localisation of these marks within helix 44, a decoding-centre element, supports a functional model in which rare but strategically placed modifications preserve reading frame fidelity, fine-tune initiation, and adjust decoding stringency. The presence of alternative modifications at the same site across *taxa* suggests adaptive flexibility under selective constraints rather than strict conservation, reinforcing the modularity of post-transcriptional regulation within ribosome evolution.

DOMAIN-3 ′ m


A
C.butilicum
A.thaliana
E.coli
T.marittima
P.abyssi
S.solfataricus
B.subtilis
H.sapiens
A.baumannii
L.donovani
S.cerevisiae
O.cuniculus
S.lycopersicum
T.thermophilus
Am
Cm
Gm
Um
m2G
pac4C
m5C
hm5C
P
Ym
m1acp3Y
m3U
m4Cm
m6A
StackedBar
Percentage
Bar Segments


B

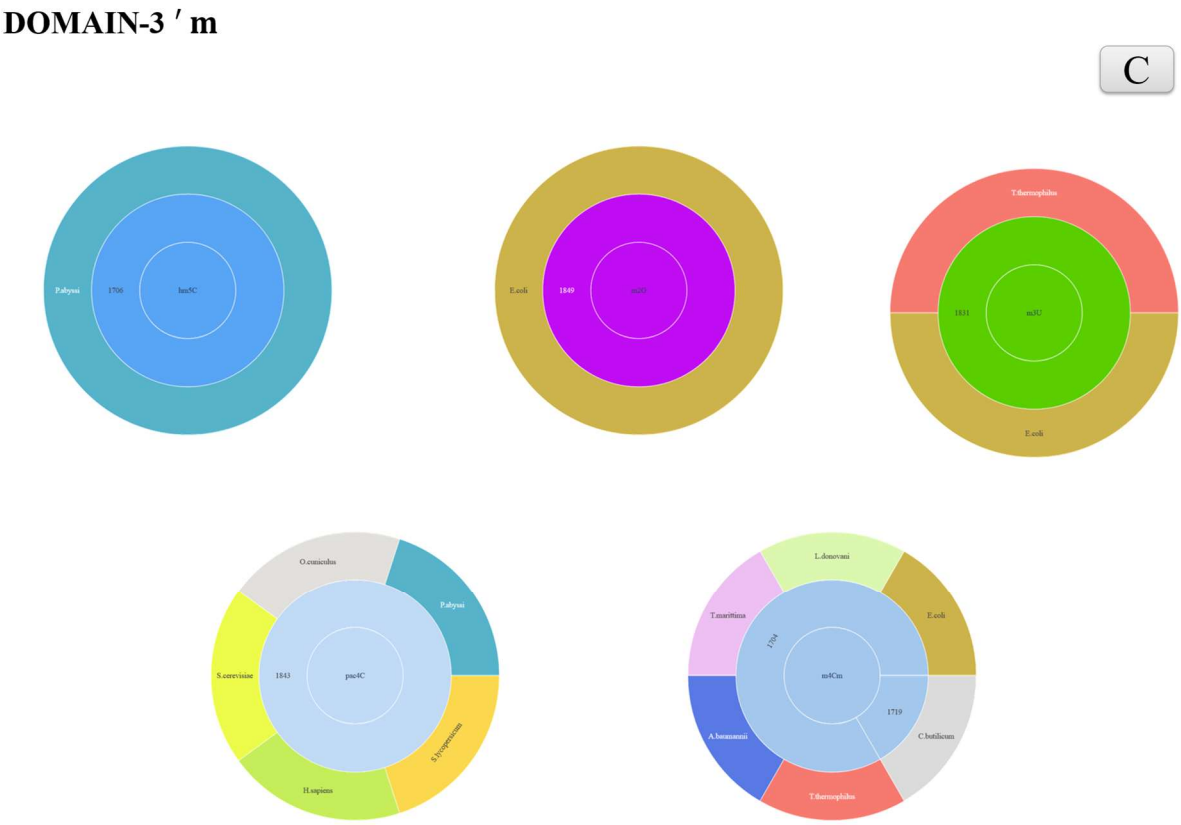


Figure 24 Distribution of RNA modifications in DOMAIN-3′m. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S, first half.

3.1.13 P1 stem-loop of the guanidine II riboswitch data analysis

The functional relevance of localised RNA folding motifs extends beyond the ribosome. Stem-loops such as helices 28 and 44 in the 16/18S rRNA contribute to subunit interaction and decoding fidelity, modulated by site-specific chemical modifications (Ontiveros et al., 2024; Petrov et al., 2014). In regulatory RNAs like riboswitches, structurally analogous helices act as ligand-responsive elements that control folding transitions (Scott et al., 2013). To explore this convergence, circular dichroism was used to characterise the folding behaviour of the P1 stem-loop from the guanidine II riboswitch, a regulatory RNA element whose ligand-induced conformational shift mirrors the structural logic of chemically modified ribosomal stem-loops.

Figure 25 presents the thermal-dependent circular dichroism (CD) profiles of three RNA variants: P1_WT (wild type), P1_N_6 (bearing an N^6-methyladenosine at position 16), and P1_N_1 (bearing an N^1-methyladenosine at the same site). CD measurements

were conducted across vacuum ultraviolet (VUV) and far ultraviolet (FUV) regions to assess conformational stability under thermal stress.

Panels A–C show averaged ellipticity values across three key absorption bands. Temperature increases resulted in changes in CD amplitude, interpreted as structural transitions. P1_WT maintained a stable positive signal in the 180–190 nm region at lower temperatures, consistent with preservation of A-form helicity. Denaturation onset was observed around 40°C, marked by a decline in ellipticity, suggesting partial backbone destabilization.

P1_N_6 exhibited higher ellipticity values across the VUV region, indicating enhanced chiral organization relative to the wild type. The m^6A_{16} modification slightly delayed structural destabilization, with unfolding evident only beyond 60°C. This trend supports the hypothesis that m^6A promotes local stacking interactions and/or conformational compactness, consistent with previous structural findings (Kierzek et al., 2021).

P1_N_1 displayed an initial increase in ellipticity between 20–30°C, reflecting early stabilization. However, the signal plateaued between 30–40°C, then sharply declined around 40°C, indicating an abrupt loss of structural integrity. Compared to P1_N_6 and P1_WT, P1_N_1 showed lower thermal resilience, with complete denaturation evident above 60°C. The m^1A_{16} modification disrupts canonical base pairing (Panel H) with uridine, preventing regular helix formation and impairing the alignment required for stable base stacking. This structural incompatibility likely contributes to destabilization of the helical fold.

In the FUV band (200–210 nm), P1_WT retained near-zero ellipticity values, consistent with stable helicity. P1_N_6 displayed more negative and stable values, consistent with increased structural order (lowering and increasing values of this band are critically evaluated and justified in Discussion). P1_N_1, in contrast, exhibited greater deviation across the thermal range, reflecting increased instability. The upward trend toward non-positive ellipticity at higher temperatures suggests progressive loss of helical structure.

All variants retained features of A-form RNA, as indicated by the persistence of the 185 nm peak and characteristic negative signal in the 200–210 nm range, typical of A-conformation helices (Holm et al., 2010; Le Brun et al., 2020). In the 260–268 nm region (**Panel C**), which reflects base pairing and higher-order interactions, both P1_N_6 and P1_N_1 exhibited pronounced ellipticity declines. These trends imply disruption of Watson–Crick and Hoogsteen interactions, likely due to steric interference by the

methyl groups. P1_WT retained greater stability in this region, despite late-stage fluctuations.

To assess the interdependence between the three absorption bands across temperature conditions, the Python-based analysis was complemented with correlation matrix visualisation in R using the *corrplot* package. Correlation matrices (**Panels D–F**) quantified the covariance between spectral bands. All variants showed a positive correlation between the VUV and base-pairing bands (185–190 nm and 260–268 nm), indicating coordinated unfolding of helices and interaction networks. P1_WT and P1_N_6 also showed a weak positive correlation between FUV and base-pairing regions, consistent with structural coherence. In contrast, P1_N_1 displayed a strong negative correlation between 200–210 nm and 260–268 nm (negative but weaker in the other two variants), suggesting that N^1-methylation impairs coordination between helical and base interaction domains, accelerating denaturation.

These results demonstrate that chemical modifications at position 16 influence the thermal stability and folding behaviour of the P1 stem-loop in distinct ways. The N^6-methyladenosine modification confers increased structural resilience, delaying denaturation and enhancing helical stability across VUV and FUV spectral regions (Kierzek et al., 2021). In contrast, the N^1-methyladenosine variant exhibits early stabilization followed by an abrupt collapse, reflecting limited structural adaptability and impaired coordination between helical and base-pairing domains.

The correlation between spectral bands indicates that unfolding events involve coordinated disruptions of both backbone and base-pairing elements. The persistence of the A-form signature at 185 nm and its complementary negative signal in the 200–210 nm band further supports this interpretation (Holm et al., 2010; Le Brun et al., 2020). The differential effects of methylation highlight how even single-atom modifications can modulate RNA conformational plasticity. These findings reinforce the broader principle that local chemical alterations—whether enzyme-installed or ligand-induced—directly affect RNA structural dynamics, a phenomenon that operates both in riboswitches and in functionally critical regions of the ribosome.

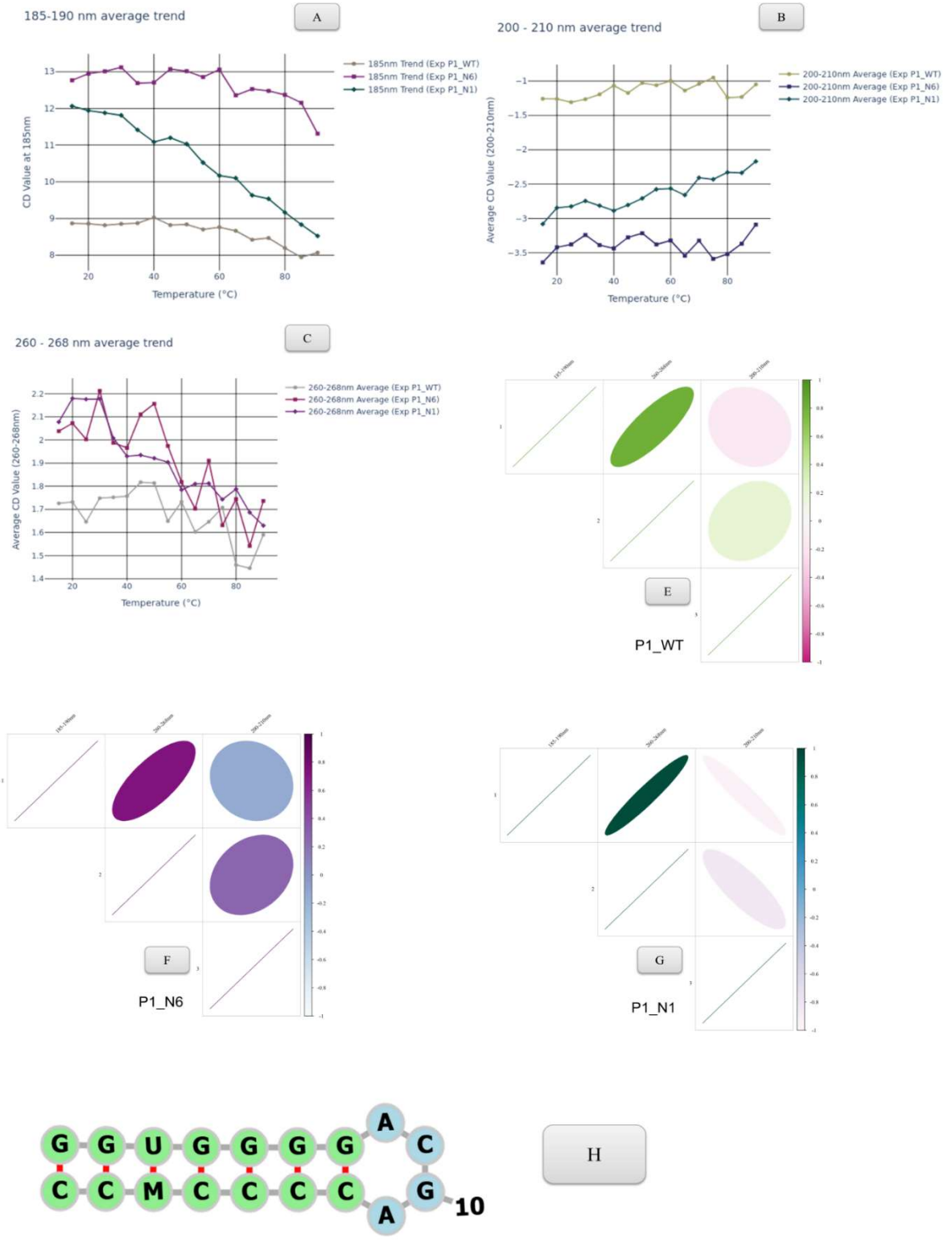


Figure 25 Circular Dichroism Data Analysis of P1 stem-loop of the guanidine II riboswitch. (**A**) 185nm absorption band. (**B**) 200-210 absorption band. (**C**) 260-268 nm absorption band. (**D**) Correlation plot among absorption bands of the $P1\_WT$ RNA variant. (**E**) Correlation plot among absorption bands of the $P1\_WT$ RNA variant. (**F**) Correlation plot among absorption bands of the $P1\_N_6$ RNA variant. Lateral bar display correlation intensity.(**G**) Correlation plot among absorption bands of the $P1\_N_1$ RNA variant. Lateral bar display correlation intensity. (**H**) Secondary structure of the P1 WT riboswitch obtained from visual inspection of the 5NE0

3.1.14 Sarcin Ricin loop data analysis

The Sarcin–Ricin Loop (SRL) is a universally conserved RNA motif located within the large ribosomal subunit, positioned near the GTPase-associated centre and functionally linked to translational factor recruitment and GTP hydrolysis. Despite its short length, the SRL adopts a defined secondary and tertiary architecture that is critical for efficient interaction with elongation factors and the fidelity of translation. Previous structural analyses mapped the SRL within functionally important regions of the ribosome, suggesting its involvement in long-range communication between the decoding centre and the catalytic core (Grela et al., 2019; Huang et al., 2020).

To extend the structural profiling of ribosomal RNA elements, circular dichroism was applied to examine the folding behaviour of the SRL under varying thermal conditions. The analysis aimed to characterize helicity, stacking interactions, and thermal stability, drawing parallels with other functional rRNA segments such as helices 28 and 44 in the SSU and stem-loop motifs in regulatory RNAs (Qin et al., 2012; Maksimova et al., 2022).

Figure 26 illustrates the temperature-dependent changes in ellipticity across six SRL variants: SRL_WT (wild type), SRL_8N_6, SRL_11N_6, SRL_19N_6, SRL_23N_6, and SRL_23N_1, each bearing either N^6- or N^1-methyladenosine modifications at indicated positions. CD amplitude was measured across the vacuum and far-ultraviolet spectral range, focusing on the ribose backbone region (185–190 nm), the helicity region (200–210 nm), and the base stacking region (260–268 nm).

Panel A of **Figure 26** shows that all variants exhibit a comparable rate of decline in ellipticity within the 180–190 nm band, suggesting a shared intrinsic sensitivity to thermal unfolding in terms of backbone chirality. SRL_23N_1 displays the highest initial ellipticity values, consistent with enhanced chiroptical intensity conferred by the m^1A modification. Across all variants, methylation appears to increase baseline CD amplitude without significantly altering the slope of denaturation. SRL_19N_6 exhibits a trend closest to wild type, indicating minimal impact of this modification position on overall backbone rigidity.

In **Panel B**, ellipticity trends within the 200–210 nm band reflect the helical content of each variant. SRL_23N_1 and SRL_11N_6 show trajectories most closely aligned with SRL_WT, whereas SRL_8N_6, SRL_19N_6, and SRL_23N_6 retain positive ellipticity throughout thermal exposure, a feature consistent with partial denaturation and departure from canonical A-form conformation (Le Brun et al., 2020).

Panel C explores the 260–268 nm region, associated with base stacking and hydrogen bonding (Le Brun et al., 2020). Here, all methylated variants exhibit higher ellipticity

than SRL_WT, with SRL_23N_1 displaying the strongest signal. While this increase may reflect enhanced local order, it is accompanied by greater fluctuations under thermal stress, particularly in the N^1-modified variant, likely due to the steric interference introduced at the N^1 position, which disrupts canonical base pairing.

To investigate how temperature-induced structural transitions affect the chiroptical response of the Sarcin–Ricin Loop (SRL), correlation plots were generated from thermal denaturation curves across three CD absorption bands (**Fig. 27**). The analysis includes the wild-type SRL (SRL_WT) and five chemically modified variants bearing N^6- or N^1-methyladenosine substitutions at specific nucleotide positions: SRL_8N_6, SRL_11N_6, SRL_19N_6, SRL_23N_6, and SRL_23N_1 (**Panels A–F**).

In SRL_WT (**Panel A**), a strong positive correlation is observed between the 185–190 nm and 260–268 nm regions, indicating coordinated structural destabilization involving both helical backbone and base interaction domains. Notably, the 200–210 nm band displays a negative correlation with both flanking regions, suggesting a divergence in its thermal trajectory. This decoupling reflects a flattening of chiral signals toward the abscissa, consistent with a departure from canonical A-form helicity. The same inverse pattern is present across most variants, supporting the notion that thermal denaturation unfolds the ribose conformation and nucleotide interactions along distinct physical axes. This interpretation is expanded in the Discussion section, where the independence of this region is further justified.

In SRL_8N_6 and SRL_19N_6 (**Panels B and D**), the correlation between 200–210 nm and the other bands weakens or disappears entirely. The more circular geometry of the correlation ellipses and the loss of anticorrelation suggest that methylation at positions 8 and 19 alters the coupling between backbone geometry and stacking behaviour. This positional effect indicates that not all N^6-methylations equally disrupt the thermal progression of CD-active signals.

SRL_11N_6 and SRL_23N_6 (**Panels C** and **E**) retain elongated correlation ellipses between 185–190 nm and 260–268 nm, alongside persistent negative correlation with the 200–210 nm band. These profiles mirror the wild-type, indicating that methylation at these *loci* does not fundamentally alter the thermodynamic interplay between backbone distortion and base-pair rearrangement.

In SRL_23N_1 (**Panel F**), the correlation network is substantially disrupted. While a weak positive relationship remains between VUV and base-pairing regions, the signal in the 200–210 nm band becomes largely uncorrelated. This pattern further supports the idea that N^1-methylation at position 23 decouples helicity from stacking integrity, aligning with its known disruption of canonical base-pairing geometry.

These results indicate that the folding behaviour of the Sarcin–Ricin Loop is modulated by site-specific methylation, with structurally distinct outcomes depending on both the position and the chemical identity of the modification. N^6-methyladenosine increases CD signal amplitude and preserves interband correlation, consistent with stabilisation of stacking and base interactions. In contrast, N^1-methylation at position 23 disrupts canonical base pairing, impairs stacking coherence, and reduces the structural synchrony of thermal unfolding. The persistent negative correlation between the 200–210 nm region and the other spectral bands across most variants supports the interpretation that helical transitions follow an uncoupled trajectory. This decoupling reflects the independent evolution of helicity under thermal stress and highlights the modular nature of RNA structural responses. Together, these observations underscore how individual chemical modifications regulate both local conformation and global folding pathways in conserved ribosomal RNA elements.

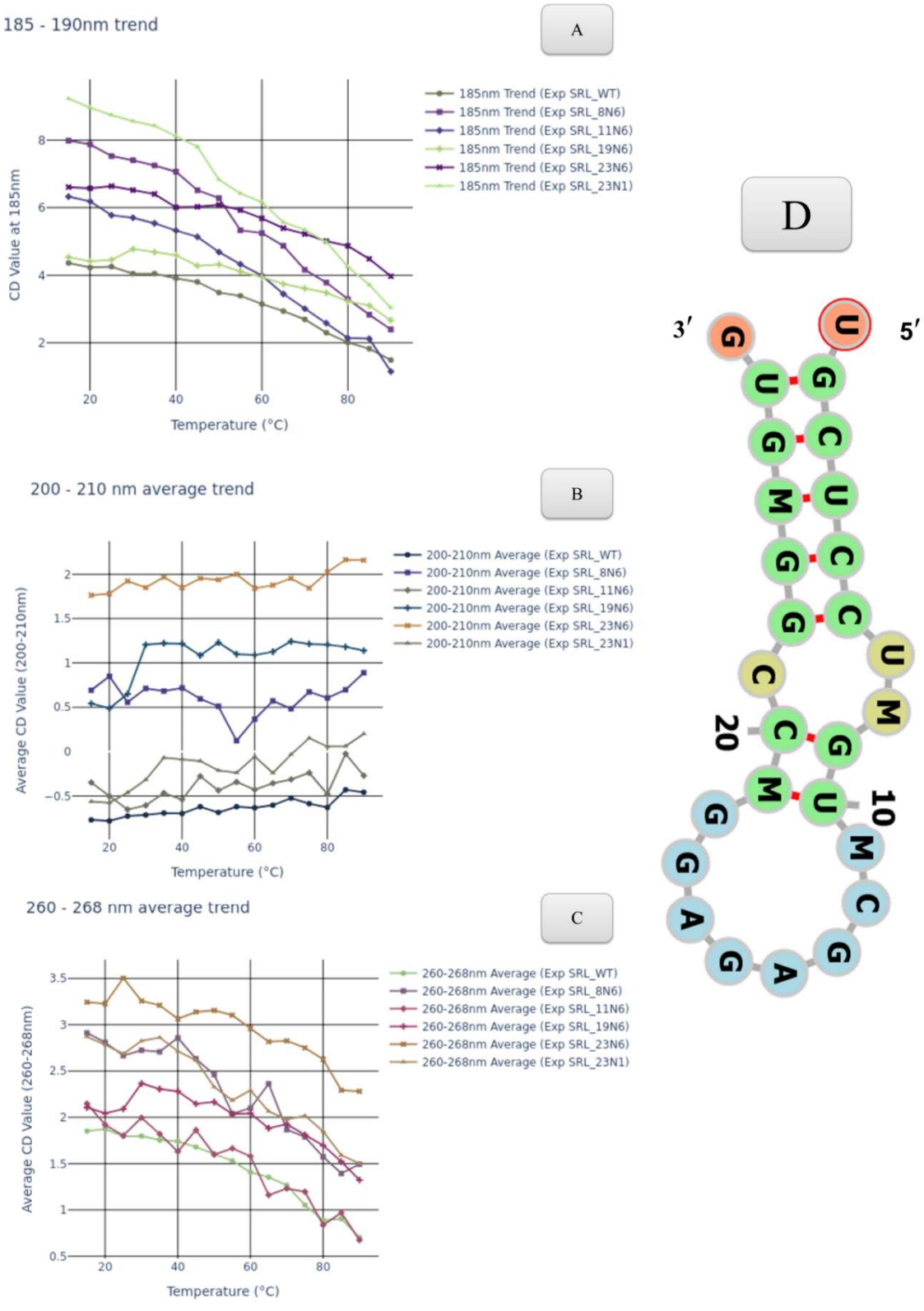


Figure 26 Circular Dichroism Data Analysis of the Sarcin Ricin Loop. (**A**) 185nm absorption band. (**B**) 200-210nm absorption band. (**C**) 260-268 nm absorption band. (**D**) Secondary structure derived from 1Q9A visual inspection. M indicates chemically modified nucleotides.

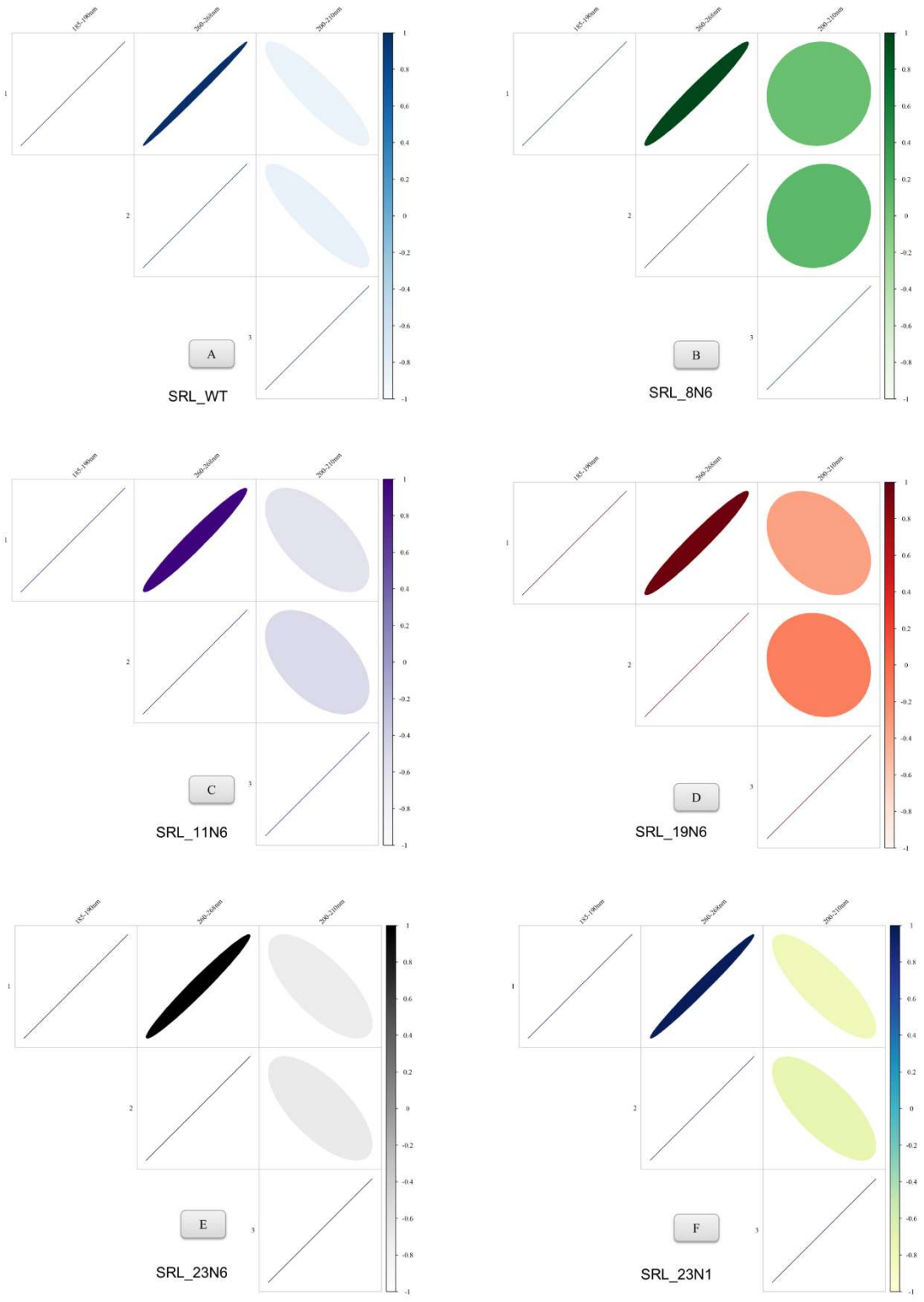


Figure 27 Correlation plots of the three absorption bands related to Sarcin Ricin Loop. (**A**) SRL_WT. (**B**) SRL_8N_6. (**C**) SRL_11N_6. (**D**) SRL_19N_6. (**E**) SRL_23N_6 (**F**) SRL_23N_1. Lateral bar display correlation intensity.

3.1.15 Analysis of Base Stacking Interactions on Uridine and its Pseudouridine and Dihydrouridine Derivatives

To investigate the stacking preferences of uridine and its derivatives, a comparative analysis was performed using a dataset of experimentally determined RNA structures. The dataset includes multiple RNA classes from diverse species, enabling the assessment of base-stacking behaviour across phylogenetic contexts. This approach provides insight into the structural logic underlying the differential distribution of uridine, pseudouridine (Ψ), and dihydrouridine (D), and their potential evolutionary selection based on conformational stability and stacking behaviour.

Figure 28 displays the distribution of stacking interaction classes involving uridine (U), pseudouridine (Ψ), and dihydrouridine (D), with numerical summaries presented in **Table 8**. Uridine is the most represented nucleotide in terms of raw frequency, with 29,918 observed stacking events, compared to 90 for pseudouridine and 50 for dihydrouridine. As a reminder of what specified in the introduction, stacking interactions are evaluated considering the central nucleotide among the two, left and right, stacking partners.

Percentage-based comparisons provide a more accurate measure of stacking preference. Uridine shows a marked tendency toward the “>>>” class, corresponding to anti–anti conformations of both stacking partners. This configuration accounts for 11,689 occurrences, representing 39.07% of uridine-involved stacking interactions. Uridine also participates frequently in the “>>.” class (anti–unstacked), with 5,089 instances or 17.01%, indicating structural adaptability in maintaining stacking even when only one face is engaged. In contrast, uridine contributes less frequently to other classes such as “>><” (anti–syn; 3.82%) and “<><” (syn–syn; 0.023%), consistent with a lower propensity for stacking in less favourable conformations. In all cases, the nucleotide under analysis is positioned between the directional indicators.

Pseudouridine, although far less abundant, exhibits a relatively higher proportion of stacking in the anti–anti class. Within its 90 stacking events, 39 belong to the “>>>” class (48.75%), exceeding the relative contribution of uridine in this category. The “>>.” class accounts for 35 interactions (38.88%), also indicating an ability to engage in partial stacking. Involvement in the “>><” (2.22%) and “<><” (0%) classes remains minimal. These distributions suggest a strong geometric preference for anti–anti conformations, likely attributable to the increased base-stacking surface area and altered glycosidic orientation of pseudouridine (Huang et al., 2012).

Dihydrouridine is rarely observed in stacking contexts. Only two interactions are recorded: one in the “>>.” class and one in the “.>>” class, each representing 0.02% of

total observed stacking events. This low incidence is consistent with its known conformational flexibility and tendency to disrupt base stacking.

These data suggest that pseudouridine may be preferentially retained in structured RNAs due to its enhanced ability to stabilize anti–anti (>>>) stacking interactions, reinforcing local architecture in functionally constrained domains. In contrast, the minimal stacking involvement of dihydrouridine, particularly in canonical conformations, aligns with its proposed role in destabilizing helical elements and promoting local flexibility. The distinct stacking profiles of these uridine derivatives, observed across diverse RNA classes, support a model in which evolutionary pressures selectively reinforce or suppress stacking to modulate RNA folding, structural plasticity, and functional specificity.

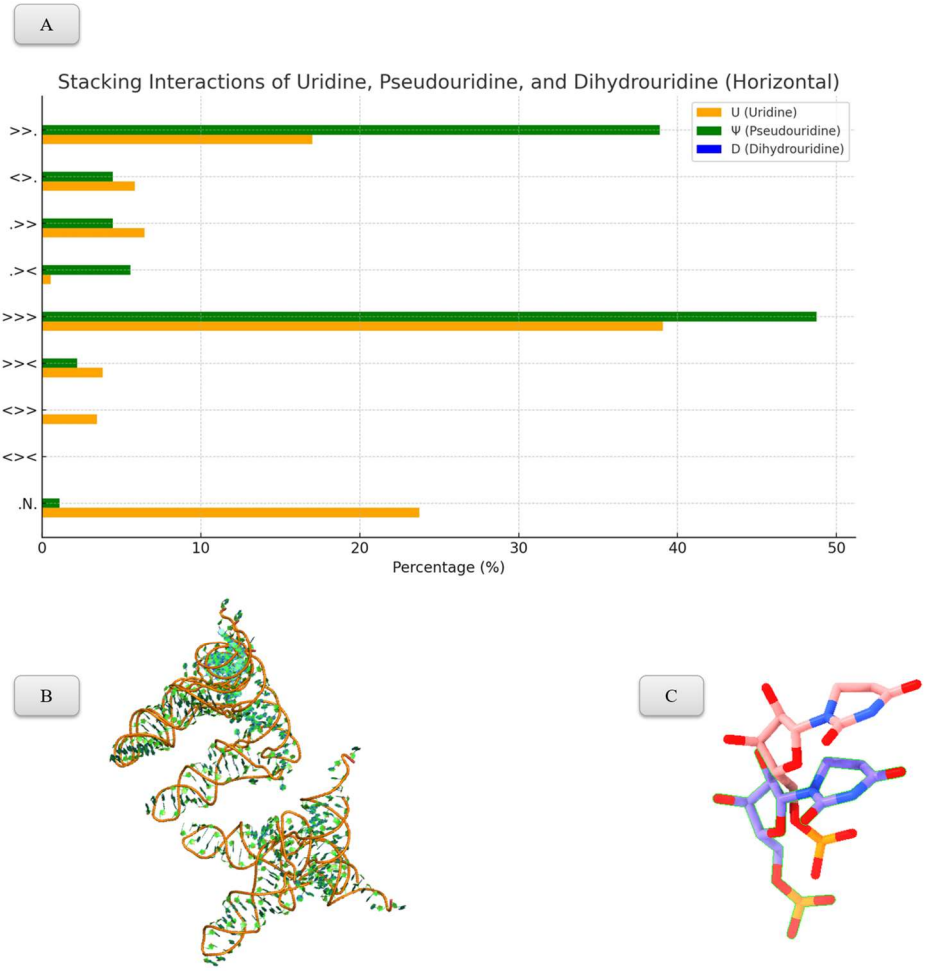


Figure 28 Analysis of the stacking interactions of the uridine derivatives. (**A**) Bar chart. (**B**) 1FCW experimentally determined structure rendered with PyMol. (**C**) Unique stacking interactions coming from 1FCW (Agrawal et al., 2000)

Stacking Classes	Uridine (U)	Pseudouridine (Ψ)	Dihydrouridine (D)
.N.	23.75 %	1.11 %	0 %
<><	0.023 %	0 %	0 %
<>>	3.45 %	0 %	0 %
>><	3.82 %	2.22 %	0 %
>>>	39.07 %	48.75 %	0 %
.><	0.541 %	5.55 %	0 %
.>>	6.45 %	4.44 %	0 %
<>.	5.85 %	4.44 %	0 %
>>.	17 %	38.88 %	0.02 %
Absolute values	29918	90	50

Table 8 Percentages of stacking interactions classes in Uridine, Pseudouridine, and Dihydrouridine.

3.1.16 Limitations of the analyses

The analyses presented are subject to limitations stemming from the restricted availability of sequence data. Approximately 1,000 non-redundant tRNA sequences are currently available, with uneven representation across phylogenetic domains. Ribosomal RNA sequences are even more limited. This scarcity constrains alignment quality and restricts the accuracy of modification site annotation. Increasing the number of sequences stored in MODOMICS is expected to improve alignment depth and support more comprehensive reference-based annotations.

While percentage-based comparisons provide insight into the stacking preferences of uridine and its derivatives, interpretation is affected by sample size disparity. The dataset contains 29,918 uridine residues, but only 90 pseudouridines and 50 dihydrouridines. This imbalance introduces sampling bias and limits the robustness of comparative analyses. Expanding the dataset—particularly for pseudouridine and dihydrouridine—will allow for a more statistically reliable assessment of their stacking behaviour and reduce the impact of current sampling constraints.

3.2 Technical Improvements of NACDDB and MODOMICS

This section describes the technical implementation of NACDDB and outlines the functionality integrated during the development phase. Key features include the dynamic search tool, the BLAST-based sequence query system, and the interface for accessing circular dichroism experiments. Together, these components establish NACDDB as a structured platform for retrieving and navigating nucleic acid spectra. In parallel, updates to MODOMICS involved restructuring RNA alignments and restoring access to deprecated data layers. These developments improved the interoperability between the two databases and expanded the accessibility of structured data for nucleic acid research.

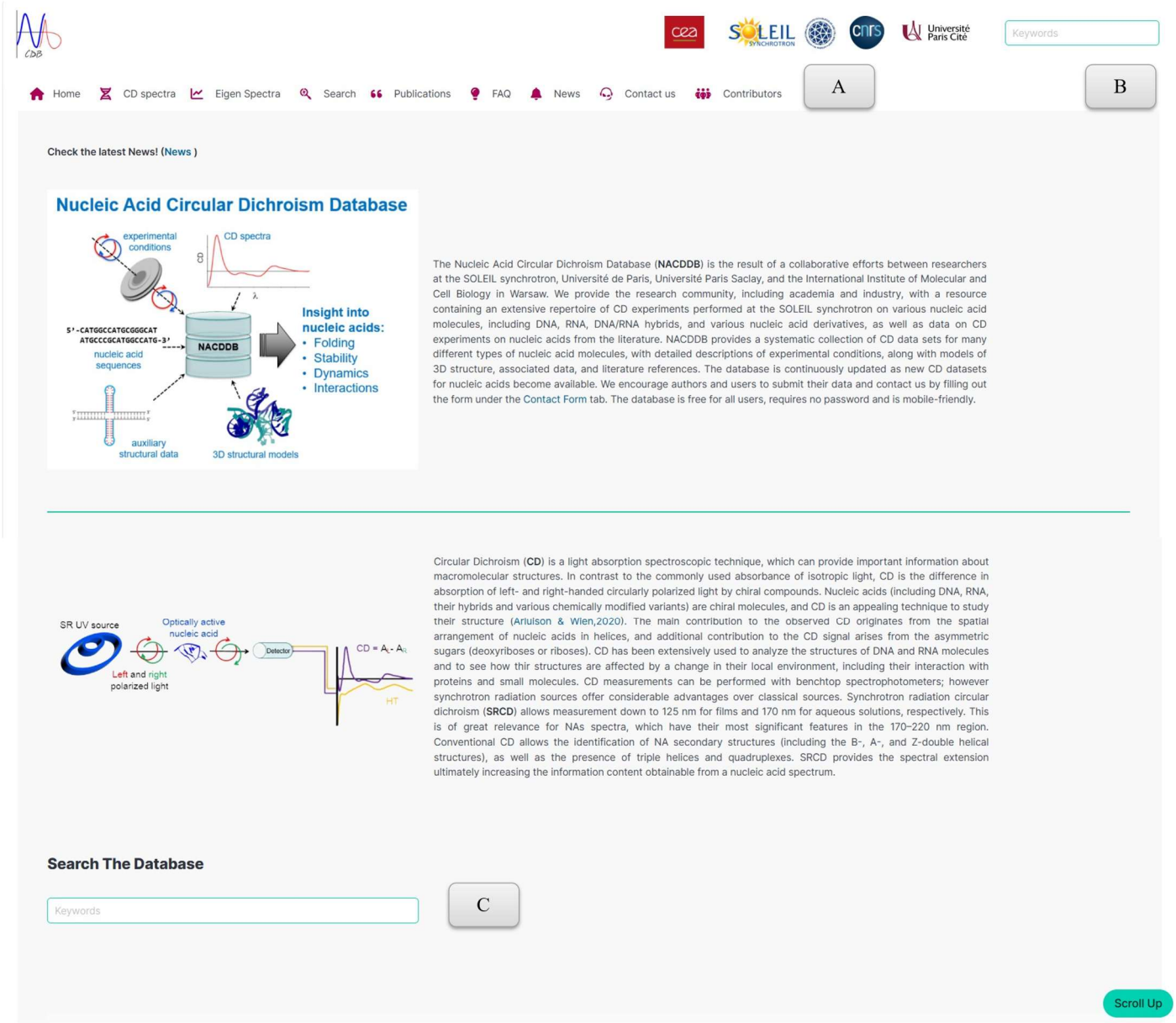


Figure 30 NACDDB Homepage. (**A**) Navigation Bar. (**B**) and (**C**) Search engine

3.2.1 General NACDDB functions and services

Figure 29 presents the NACDDB homepage and user interface. The top navigation bar grants access to the platform's core functionalities (**Fig. 30A**), including a dynamic search tool designed to facilitate the retrieval of experiments, publications, and associated metadata (**Fig. 30B–C**). **Figure 31** outlines the database's main features. The search system returns endpoints matching user-defined keywords, enabling targeted access to data stored within the repository.

The database includes circular dichroism experiments involving DNA, RNA, and hybrid nucleic acid molecules. Search results are organised by molecular type and displayed in dedicated tables (**Fig. 31A–C**), while associated publications are listed in a separate panel (**Fig. 31D**). The search interface is accessible via the "Search" tab, marked by a magnifying glass icon on the navigation bar.

Additional query tools are available for sequence-based retrieval. A BLAST search interface is integrated into NACDDB using BLAST 2.12.0+, deployed on an IIMCB Linux Server virtual machine. This feature enables identification of similar sequences by aligning user-submitted oligonucleotides against database entries. Customisable parameters, such as the expected-value (E-value), can be specified during submission.

Figure 32 depicts the BLAST workflow. Users enter input sequences and define query settings via the dedicated BLAST tab (**Fig. 32A**). Upon submission, a loading page is displayed while results are generated (Fig. 32B). The output includes a results table listing sequence matches (**Fig. 32C**). Two download formats are provided: one containing full alignment data and another summarising overall query statistics (Fig. 32D).

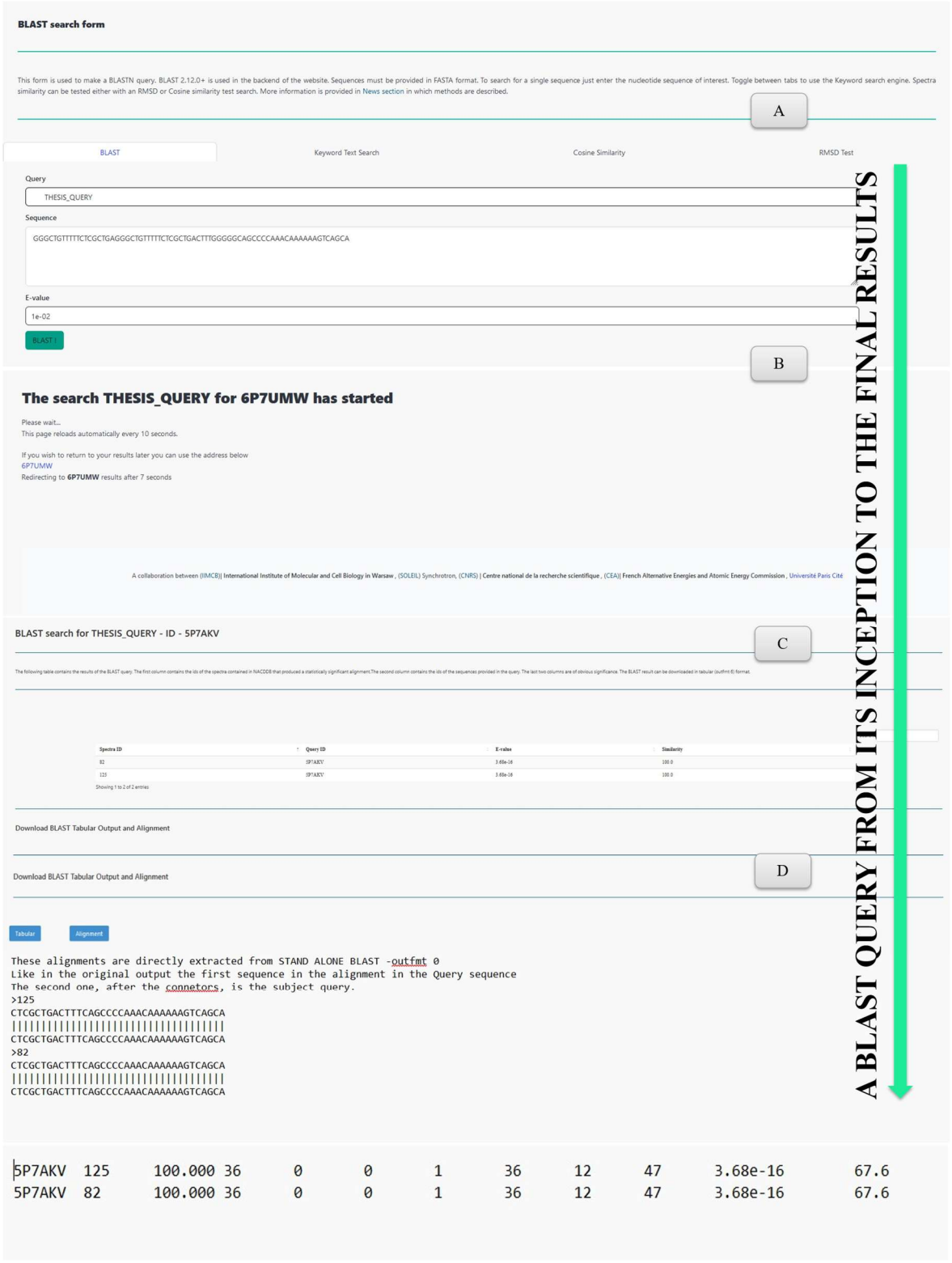


Figure 32 BLAST query results. (**A**) BLAST Tab on the search page. (**B**) Waiting Page. (**C**) Results page. (**D**) Downloaded results

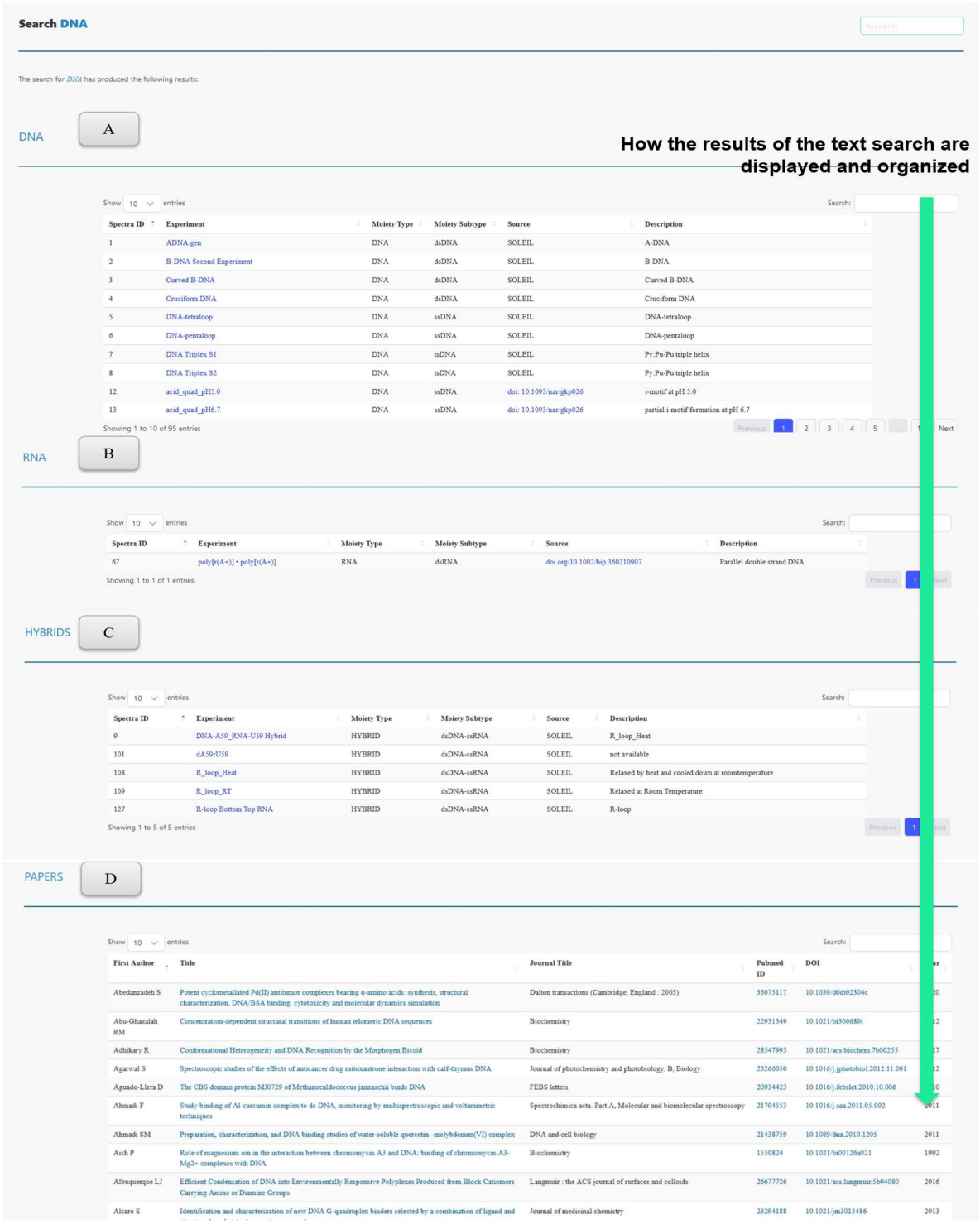


Figure 31 Utilization of the search engine with a DNA keyword. (**A**) DNA only CD experiments. (**B**) Hybrid only CD experiments. (**C**) RNA only CD experiments. (**D**) List of related papers

3.2.2 Circular dichroism experiments

The CD Spectra page, accessible via the DNA double helix icon on the navigation bar, provides access to circular dichroism experiments stored in NACDDB. **Figure 33** illustrates the layout of this page, which displays available experiments in a seven-column table. The first column contains direct links to individual experiment pages. Subsequent columns report metadata, including nucleic acid type, acquisition method (e.g., JASCO, synchrotron radiation), and oligonucleotide description. The final column uses green exclamation marks and red "X" symbols to indicate the presence or absence of thermal acquisition data. Filtering options are available above the table, allowing users to refine results according to selected criteria. Each row links to a detailed experiment-specific page.

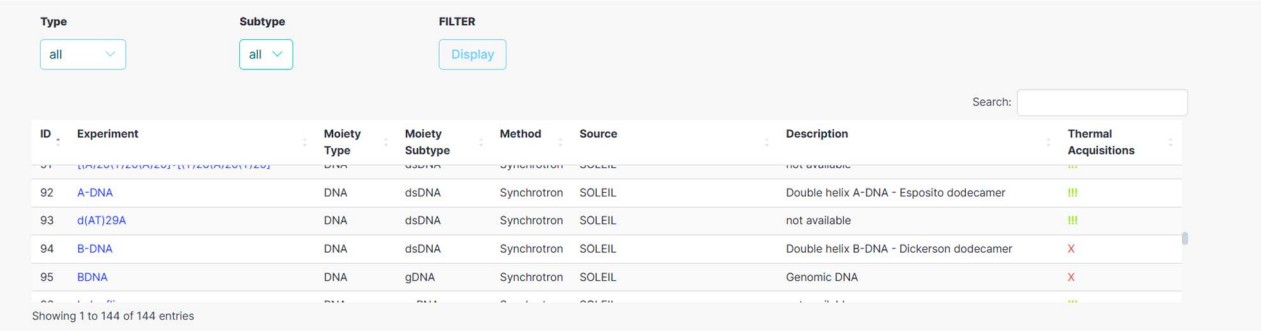


Figure 33 CD Spectra Experiment Page.

Figure 34 outlines the structure of individual experiment pages. CD spectra are displayed at the top of each page as interactive plots. When available, absorbance and high-voltage data can be toggled from the same graphical interface. Circular dichroism curves are shown in green, with absorbance and high-voltage traces rendered in red and blue, respectively (**Figs. 34A** and **34B**).

Below the plot area, a metadata table summarizes experimental parameters (**Fig. 34C**). Adjacent buttons enable data download (**Fig. 34C.1**) and provide access to three comparison modes: cosine similarity (**Fig. 34C.2**), RMSD (**Fig. 34C.3**), and a combined dual-metric approach (**Fig. 34C.4**).

If thermal acquisition data are present, additional panels become available. **Figure 34D.1** shows interactive thermal plots generated using the Plotly library. Melting curves can be rendered by averaging spectral values around 186 nm (**Fig. 34D.2**). Additional tools include the ability to select a specific thermal acquisition for targeted comparison (**Fig. 34E.1**). The selected curve can then be evaluated against other database entries using one of the three similarity metrics, selectable via a dropdown menu (**Fig. 34E.2**).

3.2.3 Methods to compare spectra

The integration of circular dichroism (CD) analysis into the NACCDB platform addresses the need for systematic, quantitative tools capable of comparing spectral profiles across modified and unmodified RNA structures. This implementation provides the computational framework necessary to evaluate nucleic acid conformation through metrics such as RMSD and cosine similarity, supporting comparisons across diverse experimental conditions and structural classes. When paired with structural datasets from MODOMICS, which stores experimentally determined RNA sequences containing site-specific modifications, this approach enables a new analytical strategy: the spectral profiling of modification permutations within homologous RNA regions.

By selecting conserved ribosomal RNA segments from MODOMICS and applying targeted computational depletion of chemical modifications, it becomes possible to generate multiple RNA variants differing only in their modification status. CD spectra obtained from such constructs can be compared using the NACCDB metric suite to evaluate the structural influence of individual modifications or their combinations. This framework establishes a pipeline to systematically interrogate how chemical modifications alter local helicity, base stacking, or global folding, as reflected in measurable spectral changes.

Circular dichroism (CD) spectra were compared using multiple similarity metrics to identify structural features of nucleic acids, including handedness, helicity, and nucleotide interactions such as base pairing and stacking (Le Brun et al., 2020). The NACCDB system accommodates spectra acquired from both synchrotron radiation sources and laboratory setups. Due to variation in available wavelength ranges across instruments, comparison algorithms operate within overlapping spectral regions. Root mean square deviation (RMSD), cosine similarity (CS), or combined metrics are applied to ensure consistent and interpretable comparisons.

Figure 35 summarises the results of this comparative framework. **Figures 35A** and **35B** report RMSD and cosine similarity values, respectively, across NACCDB-stored datasets, visualising how spectral similarity varies with temperature and experiment identifiers. These metrics capture distinct aspects of similarity: RMSD reflects amplitude-based dissimilarity, while cosine similarity evaluates directional alignment, including spectral inversions.

Figure 35C illustrates this distinction through a comparison between an RNA i-motif at 15 °C and a $(G4T2)_4$ quadruplex at 91 °C. A high RMSD (~11.90; magenta box) indicates large-scale dissimilarity, while the corresponding cosine similarity of −0.51 identifies an inverse pattern not reflected in RMSD alone. Figure 35D evaluates the

relationship between these two metrics using Kendall's tau correlation. The absence of a linear relationship supports their complementarity: RMSD captures global intensity differences, while cosine similarity detects pattern reversals and peak orientation shifts.

Spectral clustering using the k-means algorithm (k = 10) further illustrates this complementarity. Spectra assigned to the same cluster based on cosine similarity often differ in RMSD, and vice versa. RMSD-based dissimilarity does not preclude cosine-based similarity, indicating the presence of anti-aligned spectral profiles within structurally similar RNA folds.

Figure 35E presents additional comparisons from the NACCDB, including cases with high RMSD but conserved cosine similarity, underscoring the interpretive limitations of single-metric approaches. In **Figure 35F**, an RNA i-motif and an acid-quad structure at pH 7.2 exhibit low RMSD and high cosine similarity across overlapping wavelengths, indicating similarity in stacking or base-pairing features as inferred from shared absorption signatures (Le Brun et al., 2020).

The combined use of RMSD and cosine similarity enables a more complete characterisation of nucleic acid CD spectra. RMSD captures magnitude-based dissimilarity, while cosine similarity identifies directional consistency and spectral inversion. This dual approach is particularly suited for datasets with heterogeneous spectral ranges, enabling detection of structural similarities that are masked when relying on a single metric.

Together, the coupling of MODOMICS sequence data with NACCDB spectral comparison methods provides a foundation for high-resolution functional studies of RNA modifications. The capacity to evaluate conformational consequences across chemically distinct RNA variants, using consistent and scalable CD-based metrics, supports the rational interpretation of modification-driven structural dynamics. As more modified sequences become available and modular depletion strategies are implemented, this synergy will enable controlled, spectrum-guided dissection of RNA structure–function relationships.

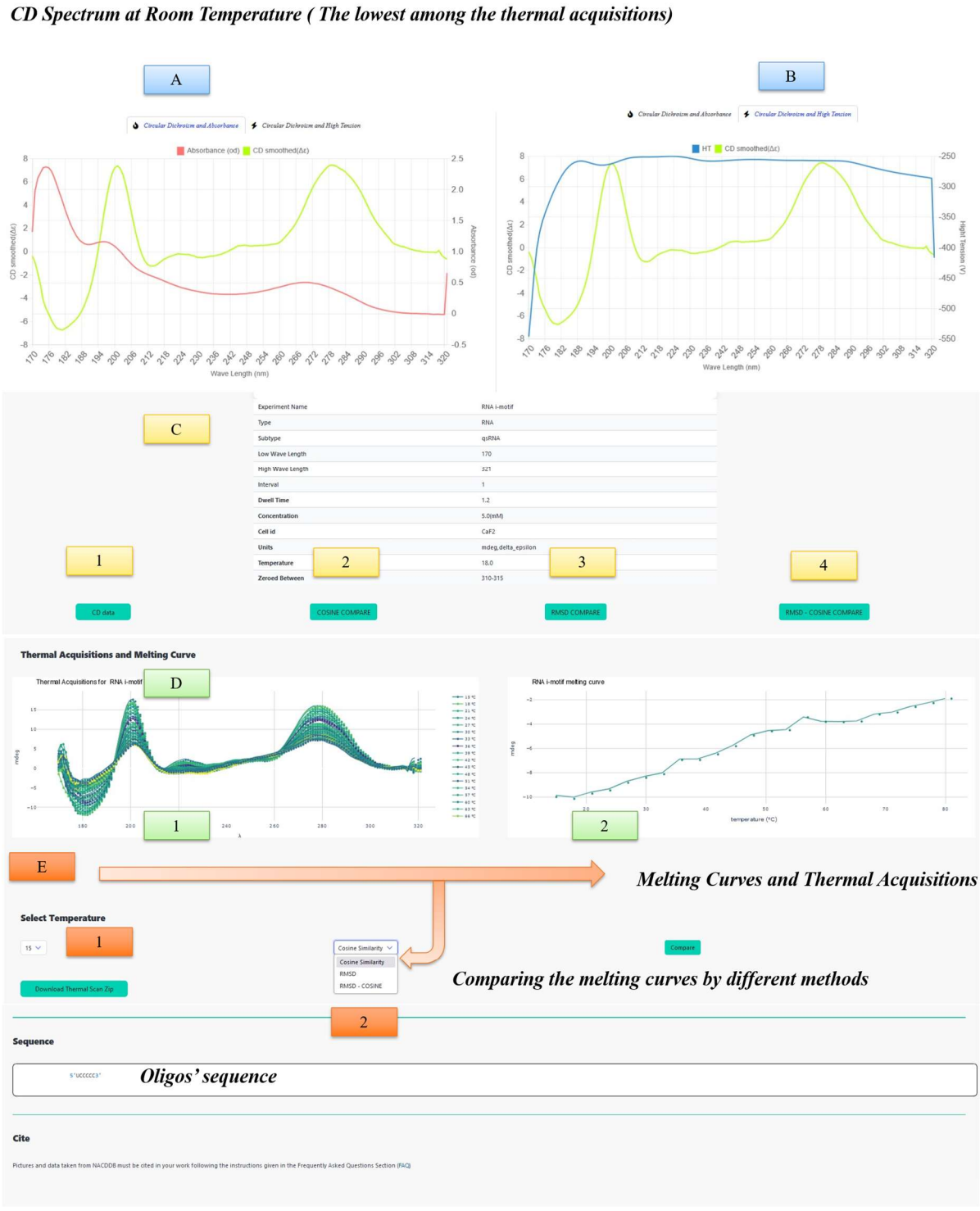


Figure 34 Experiment summary page of RNA i-motif experiment, experiment number 137. (A) Circular Dichroism plus absorbance chart. (B) Circular Dichroism plus high voltage chart. (C) Metadata table and (1) Download button, (2) Cosine Similarity (CS) Comparison, (3) RMSD comparison, (4) RMSD and CS comparison. (D) Melting curves. (E) Functions to compare melting curves and to download thermal acquisitions.

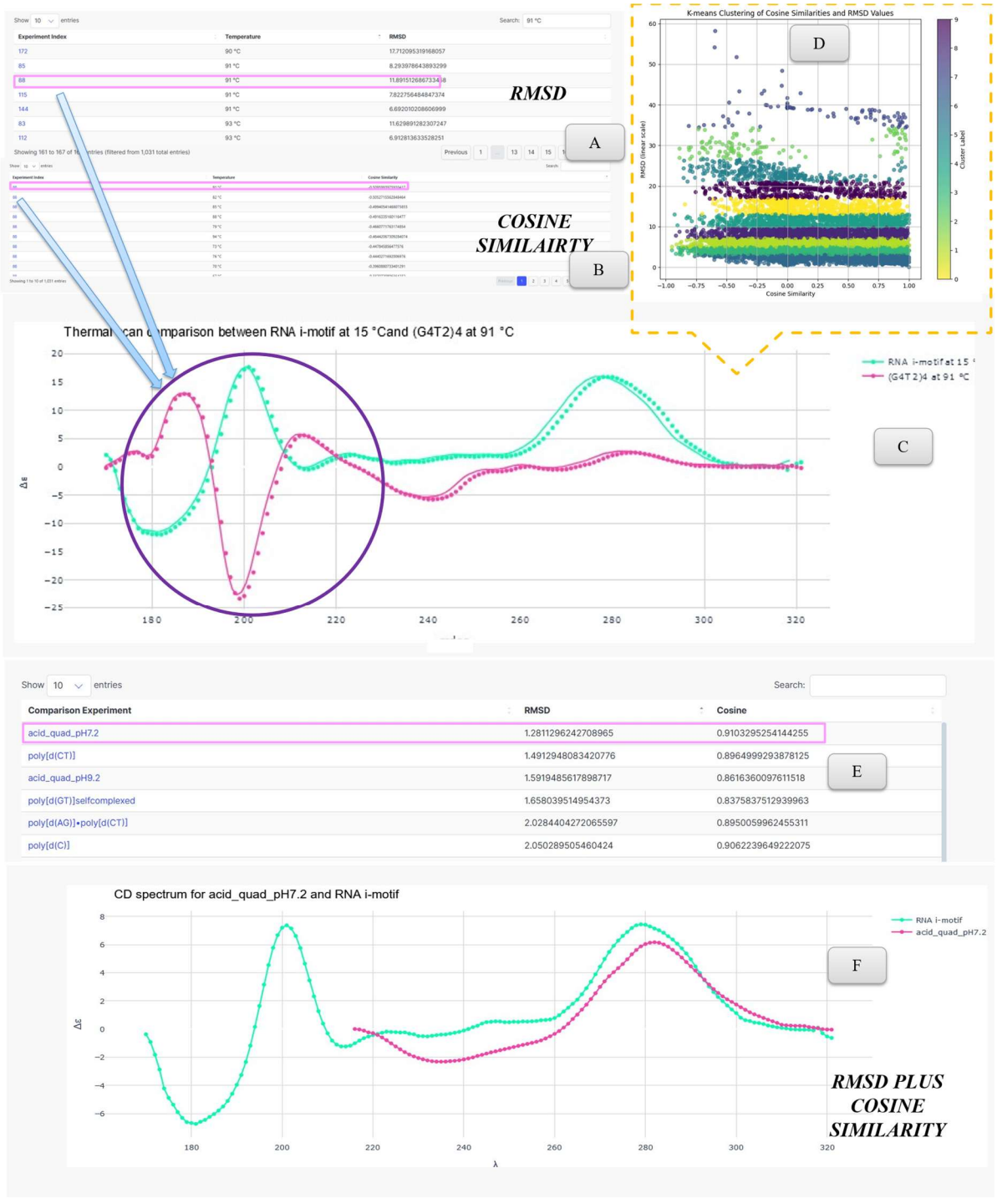


Figure 35 Utilization of mathematical comparison metrics between spectra. (**A**) RMSD results and (**B**) Cosine Similarity results, displayed in sequence. (**C**) Joint RMSD and Cosine Similarity results illustrating the comparison between spectra. (**D**) Kendall-tau analysis and K-means clustering distribution, highlighting the complementary relationship between RMSD and cosine similarity. (**E**) Extended dataset view for spectral comparison using RMSD and Cosine Similarity metrics. (**F**) CD spectrum comparison for acid_quad_pH7.2 and RNA i-motif, integrating RMSD and Cosine Similarity measures to provide a comprehensive view of alignment and dissimilarity.

3.2.4 RNA alignment section

To investigate how evolutionary pressures shape the post-transcriptional landscape of structured RNAs, the RNA alignment section was reimplemented to enable comparative analyses of sequence conservation and chemical modification across multiple RNA classes. The updated system facilitates the identification of conserved positions, domains, and structural contexts targeted by RNA modifications, thereby supporting hypotheses about their functional retention, evolutionary constraint, or structural divergence.

RNA families are grouped under their respective CLANs, following the classification system proposed by Ontiveros et al. (2024). Each Rfam ID is hyperlinked to its corresponding MODOMICS RNA alignment page. The disambiguation interface, illustrated in Figure 36, allows users to select among homologous RNA families within the same clan.

Transfer RNA (tRNA) sequences were aligned by integrating sequence and secondary structure data, matching each entry to a consensus structure defined in Rfam. The canonical tRNA domains, including the acceptor stem, D-arm, anticodon loop, variable loop, and TΨC arm, are annotated and visually differentiated using distinct colours (Quigley et al., 1976; **Fig. 37A**). Detected modifications are hyperlinked to their respective MODOMICS nucleotide entries, providing direct access to chemical and structural annotation. A legend is included to clarify domain boundaries.

The tRNA alignment interface includes multiple download options. In addition to standard sequence export, users may retrieve tabular (TSV) files in which each column corresponds to a consensus structural element. This format supports both modified and unmodified versions of each sequence. Modifications may be encoded using UNICHARACTERS, SHORT_NAMES, or MODOMICS-specific identifiers, allowing compatibility with diverse pipelines for computational or comparative analyses.

Ribosomal RNA (rRNA) alignments were also restructured. The small subunit (SSU) and large subunit (LSU) are handled independently, with domain annotations based on the framework described by Petrov et al. (2004), as illustrated in **Figures 31B** and **31C**. These alignments preserve known secondary structure domains and enable positional mapping of modifications relative to evolutionarily conserved elements.

Together, these alignment features establish a unified framework for analysing RNA structure, sequence conservation, and modification topology. By linking Rfam-derived structural contexts with MODOMICS annotations and providing flexible export

options, the system supports evolutionary comparisons, modification mapping, and structural classification across RNA types.

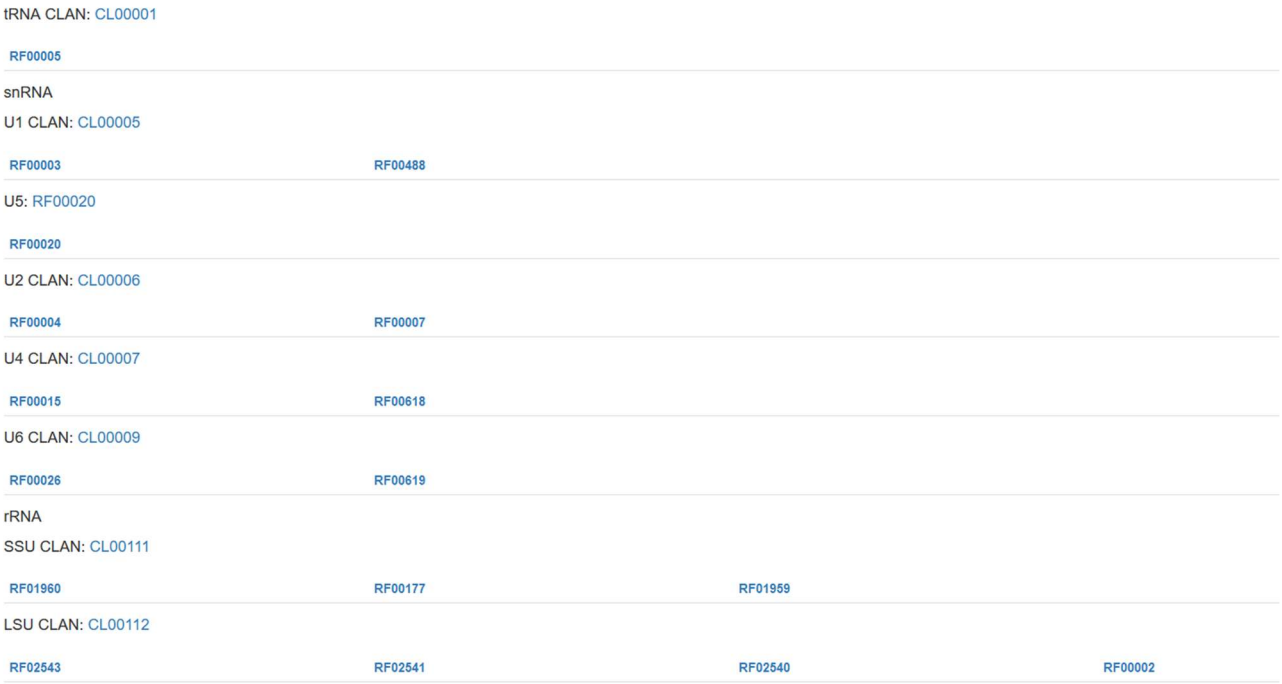


Figure 36 MODOMICS Alignment disambiguation page.

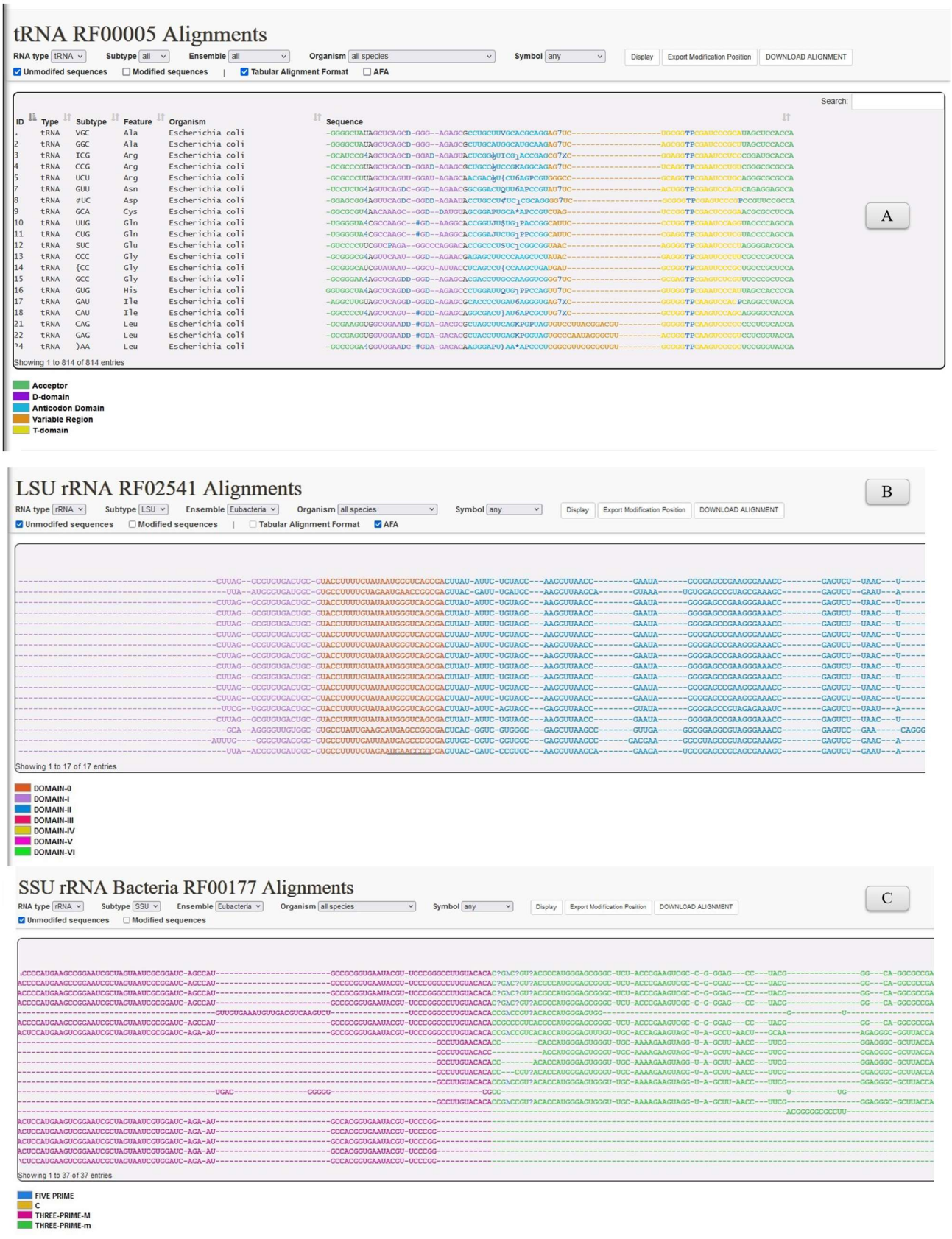
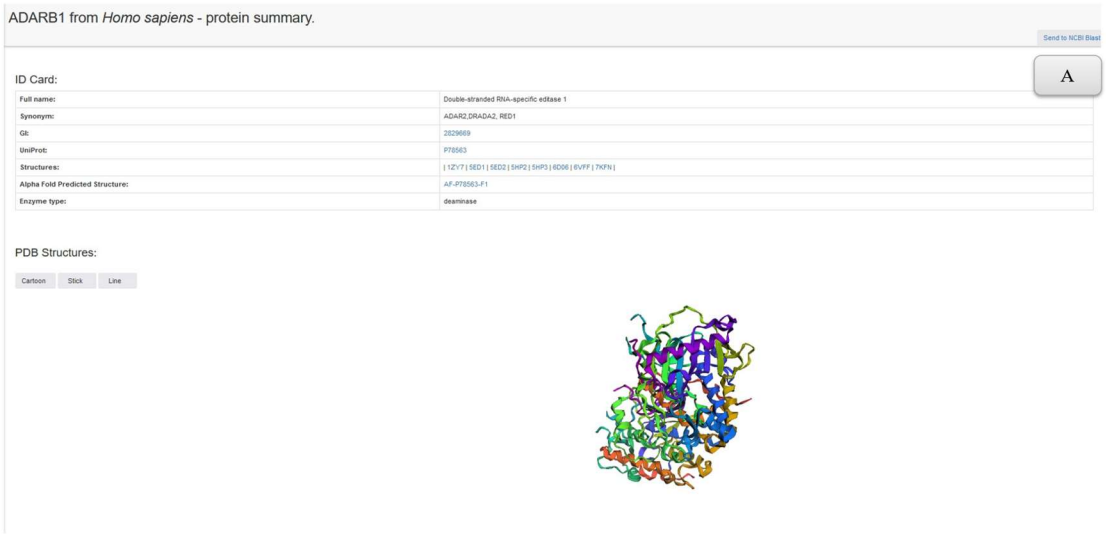


Figure 36 tRNA RF00005; (B) RF02541 LSU; (C) SSU RF00177

3.2.7 Improvements in the MODOMICS protein section

Enzymes such as methyltransferases and pseudouridine synthases install chemical modifications along RNA chains. Each enzyme targets specific nucleotides or structural contexts, contributing to the final modification pattern. Variation in enzyme presence and activity across species and conditions affects RNA processing and function. An updated protein section is required to correctly annotate the enzymatic machinery responsible for modification pathways. MODOMICS has undergone a comprehensive update to the protein section, enhancing both functionality and data accessibility. Major revisions were applied to the protein information pages, which now display protein specificities with over one thousand experimentally determined structures available for download in both MMCIF and PDB formats (**Fig. 37(B)**).



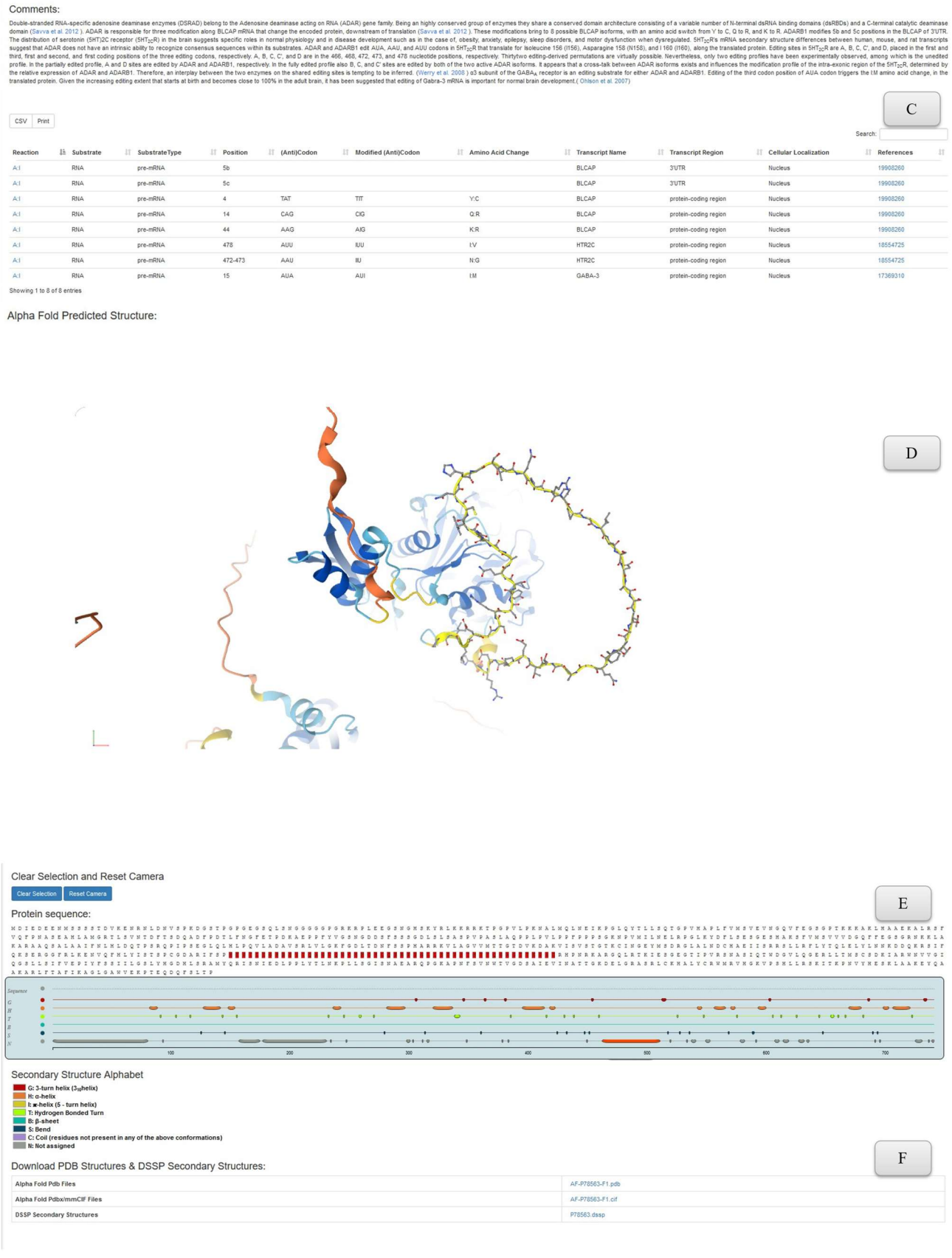


Figure 37 3DJMol Experimentally Determined Structure Visualization Tool (**A**); Comments Section (**B**); Enzymatic Reaction Sections (**C**); Mol* Protein Visualization Tool for AlphaFold protein Visualization (D); Interactive HTML sequence tool (**E**); Refactored Calipho Feature Viewer (**F**). ADARB1 protein has been selected to render this picture

Dynamic visualization of protein structures is facilitated through the integration of **3Dmol.js** (Rego et al., 2014), a JavaScript library designed for rendering and manipulating 3D molecular models (**Fig. 37**(**A**)). This system allows for real-time loading and visualization of protein structures in PDB format, offering multiple molecular representations, such as cartoon, stick, and line accessible via interactive buttons that trigger render-switching functions (RSF). Users can select different PDB structures for a given protein through a dropdown menu, which updates the molecular visualization and the associated information table, accordingly. The metadata table displays key details, including the determination technique (e.g., X-ray diffraction or NMR), resolution, and other structural information, populated by custom JavaScript functions that retrieve data from the backend RCSB Django class.

The revision and annotation of the modification data belonging to RNA sequences prompted to introduce new literature data in the protein section, better linking the information contained within the MODOMICS database. A literature review was conducted to enhance the enzymatic reactions subsection. The protein section now includes annotated comments and descriptions, with references directing users to original publications. Over 1,400 enzymatic reactions from the literature have been catalogued and re-annotated in MODOMICS, presented in tables on the protein detail pages. These tables highlight critical information such as modification positions, RNA domains, and modification types (**Fig. 37(C)**).

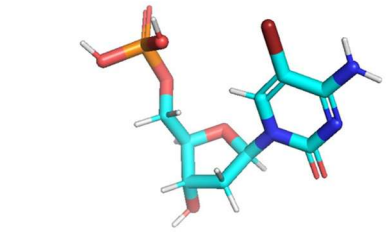
AlphaFold-predicted structures have been incorporated into MODOMICS (Jumper et al. 2021; Varadi et al. 2024) (**Fig. 37**(**D**)). These predictions, available in both CIF and PDB formats from the AlphaFold Protein Structure Database, provide comprehensive structural representations where experimental data may be incomplete. A custom visualization tool, developed by refactoring the Calipho Feature Viewer (https://github.com/calipho-sib/feature-viewer), enables the exploration of protein sequences and their secondary structural elements. AlphaFold structures were analysed using the Dictionary of Secondary Structure in Proteins (DSSP) (Kabsch & Sander 1983) to extract and map secondary structure motifs onto the predicted models.

Structural domains such as the Rossmann-like fold, conserved across multiple *taxa,* underpin many SAM-dependent methyltransferases (Su et al., 2024; Chouhan et al., 2018). Their conservation enables comparative studies linking enzyme architecture to catalytic mechanisms, evolutionary origin, and potential mutational effects. With recent updates of tools like VEP—now integrating AlphaFold-predicted structures—such analyses gain additional relevance for the interpretation of human variants in RNA-modifying enzymes (New predictions of genetic variant pathogenicity using AlphaFold protein structures | EMBL-EBI).

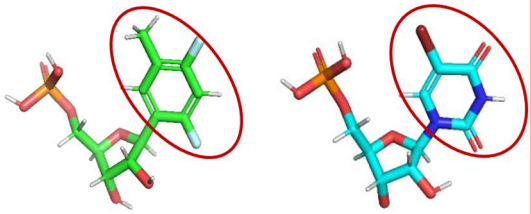
Therefore, the MODOMICS protein section has been updated with **Mol*** (Sehnal et al., 2021) allowing users to interact with the protein's secondary structure. Selecting a secondary structure segment highlights the corresponding region in the 3D model and zooms into that area for detailed inspection. Additionally, a clickable HTML-based protein sequence tool was developed, where each amino acid is highlighted in response to interactions with the feature viewer, ensuring synchronized exploration between the sequence and the 3D structure (**Fig. 37**(**E**) and **Fig. 37**(**F**)).

The integration of experimental and predicted structural data, combined with interactive tools for exploration and annotation, establishes the updated protein section of MODOMICS as a functional resource for analysing the structural and mechanistic properties of RNA-modifying enzymes. This enables cross-comparison of structural conformations, facilitates hypothesis generation regarding substrate specificity, and supports ongoing studies into the architecture and evolution of RNA modification systems.

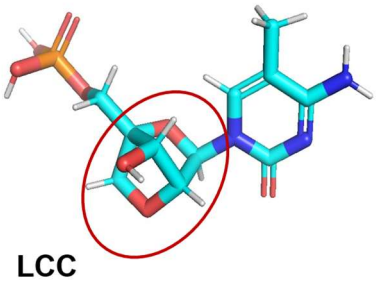
Figure 38 Nonnatural modifications in MODOMICS (**A**) Disambiguation page; Modified Nucleotide rendering (**B**); Chemical information sub table (**C**); Tautomers (**D**); CYP Metabolic Sites (**E**).



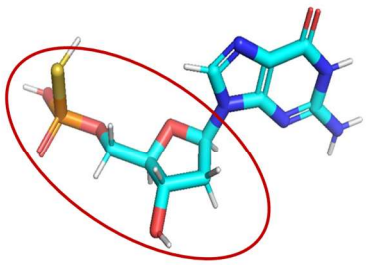
CBR



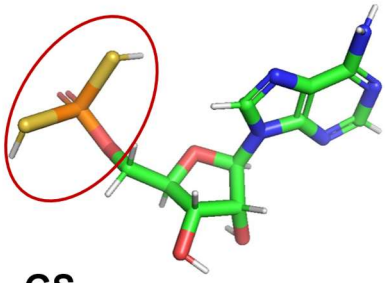
NF2 5BU



LCC



ADS



GS

Methods

MODOMICS codes have been updated to include a 10-digit barcoding system followed by a trailing letter that identifies the original moiety from which the chemically altered nucleotide is derived. Six classes have been devised to categorize all chemical alterations, with the addition of the 'X' trailing letter for modifications whose origin-moiety is unclear or unclassifiable. The classification system is determined by the first digit of the barcode. Six examples have been provided to illustrate the implementation of these classes.

1st **CBR (1000000030C)**:
The first category collects all the entries whose sugar is a D-deoxyribose or its derivatives and whose phosphate is natural or absent. This class can be used also for DNA nucleosides. CBR is a DNA linking nucleotide with a D-deoxyribose as sugar and a natural phosphate on its 5' end.

2nd 5BU (**2000000010U**) and NF2 (**2000000053X**):
The second category collects all the entries whose sugar is a D-ribose and whose phoshate is natural or absent. This class can be used for RNA nucleosides. Both 5BU and NF2 have a D-ribose and a natural phosphate. X has been used for NF2 as its base is a modified benzene with two fluorine residues attached.

3rd LCC (3000000046C):
The third category collects all the entries whose sugar is unnatural, as for instance arabinoses, and the phosphate is natural or absent. The LCC's byciclic sugar is not present within natural occurring nucleo(s)tides or wasn't found in any natural occurring pathway and the phosphate is a standard one.

4th GS (4000000041G): The fourth category collects all the entries whose sugar is a D-deoxyribose with an unnatural phosphate. GS has a thiophopshate with a D-deoxyribose.

5th GS (5000000022A): The fifth category collects all the entries whose sugar is a D-ribose with an unnatural phopshate. ADS is D-deoxyribonucleotide with double thiolation on its phophate.

6th : The sixth category collects all the entries whose sugar and phosphate is unnatural. MODOMICS does not have any entries classified in this category, as of yet.

General remarks
The determination of whether a part of a chemical alteration is natural or unnatural is based on its presence in naturally occurring pathways. To categorize sugars or phosphates as unnatural, searches against the MODOMICS database for similar molecules were conducted utilizing the MODOMICS SMILES search system (Boccaletto et al. 2021). If a characteristic was not found within the pool of natural modifications, it was classified as unnatural.

Figure 39 Examples of synthetic modified nucleotides and associated MODOMICS codes.

3.2.8 Synthetic RNA modifications

The focus on synthetic RNA modifications during this work stems from their increasing relevance in drug development and molecular design. These chemically engineered nucleotides serve as tools for probing RNA structure, guiding therapeutic strategies, and supporting the rational modulation of RNA function. Applications range from structural interrogation—via targeted substitution of functional groups—to real-time conformational analysis using spectroscopic labels. Fluorinated, halogenated, or thiolated RNA analogues have been deployed in biophysical and clinical contexts, including high-resolution NMR studies and radio-sensitisation in cancer therapy (Grasby & Gait, 1994; Wachowius & Höbartner, 2010; Izadi et al., 2023). Collectively, synthetic modifications offer a controlled framework for decoding structure–function relationships and enhancing pharmacological specificity through chemical fine-tuning.

MODOMICS functions as a centralised resource for experimentally validated RNA modifications, serving as a structured repository for data that would otherwise remain dispersed across the literature. By consolidating structural, chemical, and biological annotations from multiple sources, the database provides direct access to curated information that is not readily retrievable through conventional literature searches. Given the emerging applications of chemically synthesised modifications in drug discovery and medical research, this section was implemented to systematically catalogue synthetic entries and extend MODOMICS coverage to engineered nucleotide chemistries.

The updated catalogue of RNA modifications includes 94 entries (Cappannini et al., 2023). This expansion introduces synthetic modifications comprising RNA and DNA-linking monomers not yet observed in natural post-transcriptional pathways. These entries are classified as synthetic due to the absence of evidence for their biological occurrence; their designation may be revised if future data confirm their presence in natural systems.

Synthetic nucleotides are represented using Chinese UNICODE characters and are accessible on the MODOMICS Modification Page by enabling the full modification table (**Fig. 38A**). The accompanying disambiguation table includes essential features such as short names and corresponding RCSB Ligand IDs, which link directly to their entries in the RCSB PDB database.

A new nomenclature system was implemented to ensure compatibility with both human-readable formats and automated pipelines. This system, initially designed for synthetic entries, has been extended to natural modifications. Each modified nucleotide was assigned a reference nucleobase based on comparative analysis of atomic

composition and stereochemistry. When possible, unmodified sequences or oligonucleotides containing the modification were used as references. Modifications without a clear reference base were labelled as "unknown" and represented by the symbol "X" in MODOMICS codes.

Structural visualisation resources were also generated. Each modification is accompanied by a 2D structural image and downloadable SDF file for molecular modelling (**Fig. 38B**). Annotated chemical metadata (**Fig. 38C**) include tautomeric states and predicted Cytochrome P450 (CYP) metabolic sites (**Fig. 38D**). These catalytic hotspots were predicted using the SMARTCyp algorithm (Rydberg et al., 2010) and are visualised in Figure 38E.

The inclusion of synthetic modifications in MODOMICS enables future integration with experimental and computational pipelines, particularly those assessing structure–activity relationships, metabolic stability, or therapeutic potential. By curating these non-natural entries with the same level of structural granularity as biological modifications, the database offers a foundation for systematic exploration of synthetic nucleotide chemistries across structural biology, chemical biology, and pharmacology.

3.2.9 New human/machine-friendly nomenclature

The MODOMICS nucleotide barcoding system was significantly revised to accommodate chemically diverse, including synthetic, nucleotide residues. The updated format employs a standardised 10-digit code followed by a single letter, forming a unique identifier for each chemical moiety. **Figure 39** provides representative examples. This format ensures compatibility for both human interpretation and automated processing. The first digit serves as a barcode indicating the sugar moiety and phosphorylation status, distinguishing between D-deoxyribose, D-ribose, unnatural sugars, and whether a phosphate group is present or chemically modified. Table 1 details the interpretation of this encoding.

The terminal letter of the identifier represents the base moiety and is assigned according to the following logic:

- A, U, C, or G denotes canonical bases (adenine, uracil, cytosine, guanine), regardless of whether they originate from direct synthesis or result from enzymatic editing (e.g., C-to-U transitions are labelled as U).
- A, U, C, or G is also used for chemically modified derivatives—natural or synthetic—that retain all non-hydrogen atoms of the parent base.
- Modifications that preserve the core structure of the base (e.g., inosine as A; queuosine as G) are classified accordingly.

- X is assigned to residues that diverge from canonical base atom composition, such as 2-amino-9-[2-deoxyribofuranosyl]-9H-purine-5'-monophosphate or d-ribofuranosyl-benzene-5′-monophosphate.

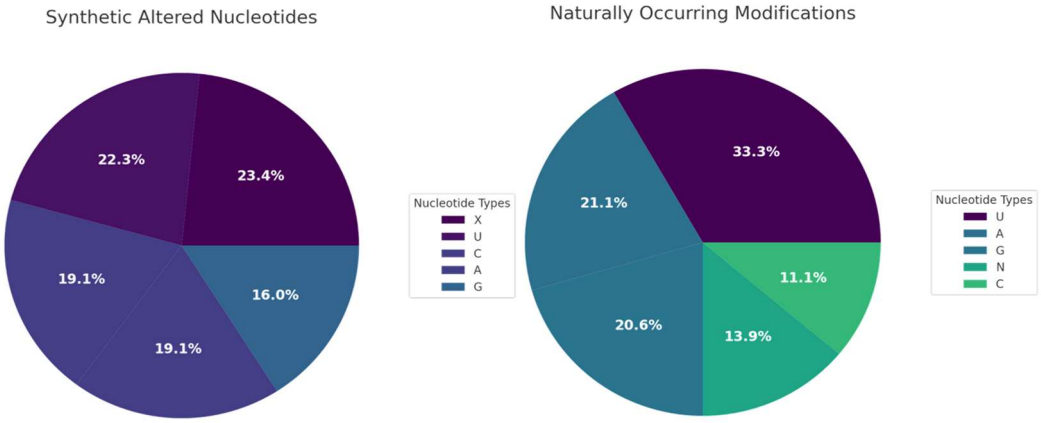
The combined 10-digit string and terminal base letter define each chemical entity in MODOMICS as a barcode-level identifier, supporting classification, retrieval, and downstream applications. **Figure 39** illustrates several examples of this updated nomenclature.

To summarise the current content of MODOMICS (Cappannini et al., 2023), descriptive statistics were compiled for both synthetic and natural nucleotide modifications. **Figure 40** displays two pie charts illustrating the distribution of base types. Among synthetic modifications, uridine derivatives constitute 23.4%, cytosine and guanosine each account for 19.1%, and adenosine derivatives contribute 16.0%. In the set of naturally occurring modifications, uridine again dominates at 33.3%, followed by adenosine (21.1%), cytosine (20.6%), guanosine (13.9%), and non-standard entries grouped as "N" (11.1%).

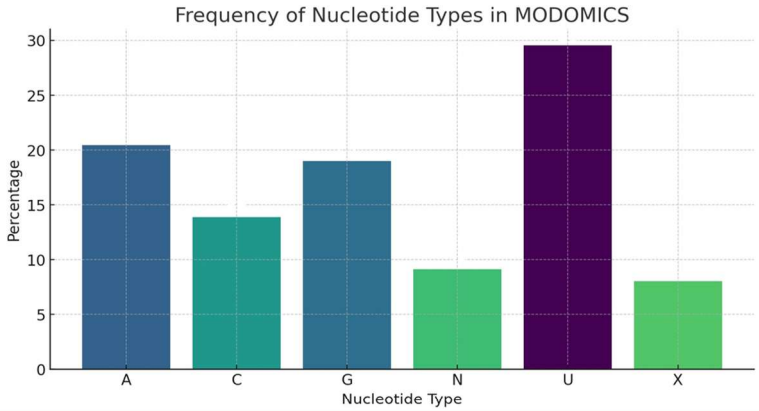
A bar plot showing overall modification frequencies highlights uridine as the most frequently modified base, followed by adenosine. Cytosine and guanosine appear at intermediate levels, while synthetic or non-canonical bases (denoted as N and X) are also represented. The accompanying legend distinguishes canonical from synthetic bases and clarifies abbreviations used throughout the dataset.

Proportion of chemically synthesized decoration applied to different nucleotide bases

Proportion of natural occurring chemical decoration applied to different nucleotide bases



Distribution of modifications across standard nucleotides in MODOMICS



MODOMICS currently catalogs approximately 230 modifications, encompassing both natural and unnatural types. Uridine is the most frequently modified nucleotide, followed by adenosine. Among the unnatural modifications, the most common original moiety does not belong to any of the four standard nucleotides.

A – Adenine: - Originating from Adenosine

U – Uridine: - Originating from uridine

C – Cytosine: - Originating from cytosine

G – Guanosine: Originating from guanosine

X – Unknown: The orginal base cannot be classified clearly

N – Any : Chemical decoration that can be applied to any nucleotide such as 5' cap

Figure 40 Descriptive statistics related to modified residues in MODOMICS

4 Discussion

RNA modifications were shown to exert sequence- and position-dependent effects on structural stability, thermodynamic behaviour, and spectral properties across conserved *loci*. These effects were observed in chemically modified oligonucleotides and correlated with conserved nucleotide positions identified through comparative sequence alignments. Variations in melting profiles and changes in spectral ellipticity revealed how specific chemical marks influence RNA conformation, providing mechanistic insight into modification-driven folding behaviour.

In MODOMICS (Cappannini et al., 2023), sequence alignment functionalities were restored, synthetic modifications were curated, and the conservation status of modified residues was re-evaluated. The nucleotide barcoding system was extended with machine- and human-readable identifiers. The protein section was reorganised to improve accessibility to enzyme annotation tools. These updates enabled annotation of newly introduced entries based on structural features and alignment-derived conservation metrics.

NACDDB (Cappannini et al., 2022) was developed to analyse the structural consequences of RNA modifications using circular dichroism (CD) spectra. Modified RNA sequences from the Genesilico Lab were included, and spectra were acquired using synchrotron light at SOLEIL. Structural effects of modifications such as N^6-methyladenosine, N^1-methyladenosine, pseudouridine, and dihydrouridine were quantified. Shifts in CD profiles consistently aligned with conserved modification positions identified through MODOMICS, supporting the structural relevance of these chemically marked sites.

A unified annotation system was introduced to support comparative analysis across MODOMICS, NACDDB, and NAIRDB (Balduzzi et al., 2024). These platforms were linked through shared identifiers and structural metadata, enabling the integration of curated spectral datasets with literature-derived sequence information. Synthetic modifications and structural inference tools developed in this study expanded the scope of database entries, allowing for systematic analysis of structure–function relationships in modified RNA.

The analytical results presented in this thesis were built upon these technical developments, which enabled the extraction, comparison, and interpretation of structural and evolutionary patterns across diverse datasets. The following sections contextualise these findings within existing structural models and recent advances in RNA modification research.

4.1 MODOMICS and analyses of RNA modification data

The expansion of MODOMICS enabled a detailed analysis of Darwinian selection acting on post-transcriptional modifications in tRNA and rRNA. Sequence alignments across species revealed conserved modification patterns that localise in structurally and functionally constrained regions. These conserved positions often coincide with known ribosomal domains that support critical processes such as tRNA accommodation, decoding accuracy, and subunit association. Thermophilic species exhibit patterns consistent with structural reinforcement—particularly through 2′-*O*-methylations in helices undergoing thermal strain—while mesophilic *taxa* retain modifications at functionally relevant sites for translational fidelity.

Ribosomal RNA was subdivided by functional domain (Petrov et al., 2014), allowing the identification of distinct conservation profiles. 2′-*O*-methylations and pseudouridines cluster around the peptidyl transferase centre (PTC), the decoding site, and intersubunit bridges. These modifications modulate local flexibility and stacking geometry. N_m residues contribute to steric reinforcement in the core, while pseudouridines stabilise junctions between helices, such as in helix 44 of the SSU or domain V of the LSU (Yang et al., 2022; Yu Zhao et al., 2021; King et al., 2003;Maksimova et al., 2022; Qin et al., 2012). The presence of these marks across divergent phylogenetic groups supports a model of convergent structural adaptation, selectively preserving modifications that stabilise folding cores or facilitate long-range communication, supporting A-form helix formation and suppressing nucleophilic activity of the 2′-hydroxyl group, thus making RNA less susceptible to hydrolysis and enzymatic degradation (Leroy et al., 2025; Höfler et al., 2020).

Further evidence of selective constraint was found in the evaluation of stacking tendencies. Uridine emerged as the dominant contributor to global stacking geometry, engaging broadly in the anti–anti and anti–unstacked stacking classes. Its widespread involvement across these configurations explains its ubiquity in regions where extended base stacking is required. Pseudouridine, despite its lower abundance, exhibited a higher relative preference for anti–anti conformations. This behaviour supports a model in which pseudouridine enhances stacking only at structurally restricted sites where its specific hydrogen bonding potential confers local architectural reinforcement (Huang et al., 2012; Kierzek et al., 2013; Charette et al., 2000). Helices flanking decoding centres and the ribosomal core are examples where this mode of local reinforcement is both observed and phylogenetically conserved.

Dihydrouridine showed near-complete exclusion from stacking interactions, which aligns with its known destabilising effect and conformational flexibility (Dalluge et al.,

1996; Deb et al., 2014). Its occurrence in looped regions of tRNA and in thermodynamically pliable elements of domain V (Toubdji et al., 2024) points to a role in promoting local disorder, essential for structural transitions such as A- and P-site accommodation. This pattern is consistent with evolutionary selection for backbone flexibility at dynamic junctions (Yu Zhao et al., 2021).

The contrasting stacking behaviours of uridine, pseudouridine, and dihydrouridine reflect a domain-specific logic of structural constraint. Pseudouridine appears as selectively retained where enhanced stacking and rigidity are advantageous, while dihydrouridine is tolerated or enriched in positions requiring local destabilisation or increased motion. These stacking-derived features, derived from experimentally determined structures, substantiate the thesis that modification retention is not random, but rather constrained by local geometry and conformational compatibility.

Collectively, these findings clarify how RNA modifications are distributed under evolutionary pressure, not only by biochemical type or species, but also by structural context and stacking behaviour. MODOMICS now offers a framework that links sequence conservation, chemical specificity, and conformation-resolved data, supporting the integration of experimental observations with evolutionary models.

The forthcoming integration with SciModoM (Boileau et al., 2025) is expected to enrich the dataset and introduce a conceptual shift in how RNA modifications are structured, contextualised, and analysed within the database framework. Unlike existing repositories, SciModoM offers quantitative, per-site data for over six million modifications across 156 transcriptomic datasets, allowing users to compare modification profiles across experimental technologies, cell types, and conditions. Its bedRMod format supports interoperability with genome browsers and facilitates cross-dataset analysis by embedding stoichiometry, confidence scores, and sequence coverage at single-nucleotide resolution. Importantly, it links each modification to its MODOMICS entry and offers a secure upload function for custom data comparison, enabling researchers to independently validate or extend published results.

The initial contribution focused on the systematic enumeration of modification sites deposited in SciModoM. Although transcript annotations can be retrieved using tools such as ENSEMBL's API, which provide mappings between genomic and transcriptomic coordinates, SciModoM aggregates data from multiple assemblies and sequencing platforms. Genome assemblies may differ due to the inclusion of newly resolved exons, corrected splicing boundaries, updated UTR definitions, or refined repeat regions based on improved sequencing technologies and long-read scaffolding. These biological updates directly impact the sequence context and structure of

transcribed genes, particularly in non-coding RNAs and transcript isoforms. As a result, the same gene may exhibit substantial differences in sequence or length depending on the reference build, complicating coordinate resolution for deposited modification sites (https://www.ensembl.org/info/genome/genebuild/assembly.html; https://www.ncbi.nlm.nih.gov/assembly/help/).

This heterogeneity limits the immediate usability of SciModoM's genomic-format data for integrative transcript-level analyses. A practical solution would involve aligning homologous genes across assemblies to a unified reference transcriptome. Genomic coordinates could then be converted into transcriptomic space, producing per-transcript maps of post-transcriptional modifications. This would allow the recovery of full-length, modification-annotated transcripts for each species in SciModoM, with chemical marks validated by multiple, independent detection technologies for each modification type.

Establishing such a coordinate harmonisation strategy would bridge MODOMICS's structure-based annotation (Cappannini et al., 2023) with SciModoM's high-throughput evidence (Boileau et al., 2024), enabling joint evolutionary and structural analyses. This integration would ultimately expand the interpretability of conserved modification sites and support comparative models across species, conditions, and detection pipelines.

Overall, this doctoral work aggregates and refines existing knowledge on how post-transcriptional RNA modifications relate to structural context and evolutionary conservation. By combining large-scale sequence alignments with structural domain partitioning and stacking analyses from experimentally resolved RNA molecules, it adds quantitative and domain-specific detail to previously proposed models. MODOMICS was updated and reorganised to facilitate this investigation, enabling a more systematic comparison of conserved modification sites across species and structural classes. The forthcoming integration with SciModoM introduces a quantitative layer, offering stoichiometry-resolved datasets and contextual annotations across transcriptomic conditions. While the functional implications of RNA modifications have long been recognised, this work contributes a framework for examining their conservation and structure-function associations with higher granularity. In doing so, it builds on the foundations laid by previous research, aiming to support future studies with interoperable tools and curated data that extend the visibility and interpretability of RNA chemical diversity.

4.2 NACDDB and analysis of circular dichroism data

NACDDB marked the foundational stage of a broader initiative aimed at consolidating nucleic acid structural data derived from circular dichroism (CD) and other spectroscopic techniques. This effort targeted the fragmentation of CD datasets in the literature by centralising information on RNA and DNA modifications within a unified platform (Cappannini et al., 2022; Balduzzi et al., 2024).

The backend development of NACDDB established a scalable infrastructure for curating, managing, and analysing CD data from both natural and chemically modified nucleic acids. In the final phase of this doctoral work, this framework was extended with analytical algorithms designed to perform cross-spectral comparisons. Specifically, cosine similarity (CS) and root mean square deviation (RMSD) metrics were implemented to assess convergence and divergence between spectral datasets. These functions enable reliable comparison of spectral features even across experiments acquired using different instruments or setups, thus facilitating a broader structural interpretation.

NACDDB accommodates not only canonical sequences but also post-transcriptionally or synthetically modified variants, including oligonucleotides subjected to environmental perturbations such as thermal stress or pH shifts. The combined use of CS and RMSD enhances the resolution of pattern detection, offering more nuanced insight into spectral variations. A user-friendly interface supports database exploration through keyword-based queries and sequence similarity searches via a built-in BLAST engine.

Collectively, these tools position NACDDB as a structural repository and analytical hub for nucleic acid research, offering an integrated platform to support RNA/DNA molecular design and its downstream applications in synthetic biology and biotechnology (Kloczewiak et al., 2022).

The investigation of CD spectra extended beyond platform development, with a focus on the structural effects of site-specific RNA modifications. Particular attention was devoted to understanding how methylated adenosine variants, such as N^6-methyladenosine (m^6A) and N^1-methyladenosine (m^1A), influence RNA stability, folding, and spectral response.

Three distinct spectral regions (*i.e.*, 185–190 nm, 200–210 nm, 260–268 nm) were selected for circular dichroism (CD) analysis, following reference standards for nucleic acid chiroptical properties (Le Brun et al., 2020). A revised interpretative approach is proposed for the 200–210 nm region, which is typically evaluated alongside the 185 nm

VUV peak. In this work, the 200–210 nm region is assessed independently to investigate its specific contribution to structural interpretation. The analysis highlights that flattening of ellipticity in this band—manifested as values approaching zero or becoming less negative—may indicate deviations from the canonical A-form conformation. This spectral behaviour is interpreted as a loss of helical order or partial unfolding, reflecting a transition toward increased molecular flexibility.

Earlier works on DNA conformations support this interpretation. Sutherland et al. (1986) and Zimmer et al. (1982) showed that B- to Z-form transitions in DNA produce an increase in amplitude within this region, including the appearance of a distinct positive peak in Z-form duplexes. Vorlíčková et al. (1982) further documented that overlapping A- and Z-structures result in spectral transitions from negative to positive values within the same wavelength band. However, these analyses were limited to DNA duplexes and quadruplexes. Application of these insights to single-stranded RNA requires caution, as RNA differs fundamentally from duplex DNA in structural dynamics. Single-stranded RNA molecules possess greater backbone flexibility and are more exposed to solvent interactions, including extensive hydration and ion coordination (Šponer et al., 2014; Šponer et al., 2018). These properties introduce higher conformational variability, reducing the reliability of extrapolations from duplex-based spectral models. Moreover, single-stranded Z-form peaks are spectrally shifted by ~5 nm relative to those of A- and B-DNA, limiting direct comparisons (Holm et al., 2010).

More consistent benchmarks arise from the contrast between A- and B-form geometries (Holm et al., 2010). The A-form—predominant in RNA—displays enhanced base stacking, compact helical rise (~2.8 Å), and $C^{3'}$-*endo* sugar puckering (Shakked et al., 1983). This contrasts with the B-form's longer rise (~3.4 Å), $C^{2'}$-*endo* pucker, and reduced packing density (Leslie and Grover, 2020). A-form helices exhibit narrow major grooves and wide, shallow minor grooves, in contrast to the broader groove profiles of B-form structures. These topological differences influence hydration patterns, ion accessibility, and spatial orientation of functional groups, all of which contribute to the chiral optical response observed in CD spectra. The compact groove architecture of A-form helices enhances base stacking and spatial regularity, leading to more distinct chiroptical signatures in the 200–210 nm region compared to the more flexible B-form helices. Intramolecular stacking dominates in A-form helices, while B-form conformations rely on inter-strand stacking (Aida, 1988). These differences are reflected in CD signals. A-form helices produce weakly negative ellipticity in the 200–210 nm band. In contrast, B-form or more disordered conformations yield flatter trends

approaching zero, as demonstrated in the findings of Holm et al. (2010), which analysed single-stranded DNA exhibiting A-like conformations.

Therefore, consistent flattening of the 200–210 nm signal under thermal stress is interpreted as a loss of A-form chiral signature. This interpretation is further supported by the observation that the FUV region (185–190 nm) and the nucleotide interaction region (260–268 nm) exhibit a positive correlation in their thermal response, while both show a negative correlation with the 200–210 nm region. This opposing behaviour suggests that the 200–210 nm band tracks a distinct conformational axis, likely reflecting a loosening of A-form helicity. The signal is further modulated by the presence of site-specific modifications, which alter local stacking or hydrogen bonding networks. These changes reshape the electronic environment and base orientation, reinforcing or disrupting the A-form architecture depending on the chemical nature and position of the modification. whose effects depend on local RNA geometry and stacking environment.

In addition to proposing a revised interpretation of CD spectra, the analysis highlighted notable structural features of the molecular variants under investigation. In all cases, post-transcriptional modification introduced signal broadening, increasing intensity across key CD-active regions. Additionally, experiments on the Sarcin Ricin Loop demonstrated positional dependence, as spectral variations were observed based on the specific location of the introduced modification. The observed correlation among the absorption bands suggests a connection between molecular denaturation and conformational changes in the sugar-phosphate backbone.

In general, methylated oligos displayed an overall improvement of chiroptical signatures, as evidenced by the increased CD signals with respect to the wild-type conformation. However, the thermal behaviour of the methylated variants often contrasts with the native form, which under thermal stress remains close to CD values displayed at room temperature. Conversely, most modified sequences exhibit steeper decreases in their CD trend as temperature rises, indicating a reduced capability to cope with an increasing thermal stress.

A computational analysis by $AlphaFold_3$ (Abramson et al., 2024, data not shown but provided in supplementary materials) complements the experimental observations, showing that the introduction of N^6-methyladenosine does not disrupt canonical Watson–Crick base pairing or stacking interaction patterns, found in original structures. Their preservation implies that the fundamental scaffold of RNA helices remains intact.

As either in P1 or SRL molecules m^6A maintains the shape of their trend profiles, the overall data suggest that N^6-methyladenosine primarily alters surface characteristics by

increasing local hydrophobicity. The introduction of such methylations may promote tighter local nucleobase packing and reduced solvent exposure in the shallow groove, sharpening CD signals at room temperature. However, reduced hydrogen-bonding potential or altered solvation around the methyl group may diminish stabilising water interactions that normally confer resilience against thermal unfolding (Šponer et al., 2014; Šponer et al., 2018).

While m^6A creates a more compact and hydrophobic *locus*, enhancing chiroptical properties at moderate conditions, it may reduce thermal adaptability. Indeed, m^6A destabilises RNA duplexes thermodynamically (by 0.5–1.7 kcal/mol) and slows hybridisation kinetics by raising the annealing energy barrier. The N^6-methyl group favours the syn conformation when unpaired, altering the RNA folding pathway. N^6-methyladenosine also increases the structural accessibility of modified regions; its destabilising impact becomes more pronounced at elevated temperatures (Shi et al. 2019), consistent with the steeper CD signal decay compared to the native form. This may explain while the CD region relative to nucleotide interactions (260-268 nm) tend to collapse as temperature raises up.

The *in vitro* analysis of the Sarcin-Ricin Loop revealed position-dependent effects of chemical modifications, reflecting structural constraints that appear to have been shaped by evolutionary selection, as inferred from RNA sequence alignments. Two representative cases, SRL_$11N_6$ and SRL_$23N_1$, exhibit reduced spectral divergence relative to the native form. In SRL_$11N_6$, m^6A is in a loop, illustrating the position-dependent behaviour of N^6-methyladenosine (Kierzek et al. 2021), likely with a limited effect on chiroptical properties of the molecule. Structural analysis shows that the methylated adenosine participates in a hydrogen-bonding network. Both N^6 hydrogens interact with the sugar rim of G_{18}. Indeed, the AlphaFold predictions display the same orientations between the m^6A_{11} and G_{18} , maintaining the trans-hoogsteen interaction between the two nucleotides. However, N^6-methyladenosine introduces steric interference in AlphaFold-generated models that is not supported by the thermal stability observed in CD spectra, indicating the need for further refinement in RNA structure prediction algorithms, particularly in the context of chemical modifications.

In SRL_$23N_1$, m^1A interferes with a Watson–Crick interaction with U_4, which is reported in the original RCSB structure. Steric hindrance from methylation on the N^1 atom disrupts this pairing. Terminal bases, including those forming AU doublets, are generally less stable than core helical regions (Zuber et al. 2022). The effects of m^1A are also context-dependent, as m^1A can alter local RNA stability and folding based on its location (Vaidyanathan et al. 2017). Although canonical base pairing is unlikely for m^1A due to altered chemical properties, N^1-methyladenosine can still form structured

hydrogen-bond networks. These interactions may help stabilise L-shaped tRNA conformations, as observed for m^1A at position 58 (Oerum et al. 2017). This may explain the limited deviation from the native CD spectra at similar temperatures.

Nonetheless, CD spectra alone do not yield direct insights into the behaviour of individual modifications. The observed trends provide general information on molecular characteristics but limit interpretability. The study should be expanded with NMR experiments under thermal stress to allow direct observation of stacking interactions, hydrogen bonding, and conformational changes at specific nucleotide positions. Temperature-dependent NMR could validate and extend CD-based findings, clarifying how modifications like m^6A and m^1A affect local and global RNA stability during denaturation and annealing. Additionally, real-time NMR monitoring could reveal how modifications contribute to folding kinetics and the formation of nucleotide interaction networks (Shi et al. 2019). Finally, increasing the number of experimentally resolved structures with associated CD spectra would allow a statistically robust correlation between spectral features and structural variation.

Apart from the specific observations discussed above, the combined interpretation of CD spectra and *in silico* predicted structures requires general methodological considerations. $AlphaFold_3$ (Abramson et al., 2024) enables the generation of RNA structural models that include a limited set of post-transcriptional modifications, such as m^6A, $m^{6,6}A$ and A_m. However, no current model supports the full set of chemical modifications relevant to the sequences examined in this study.

The ModeRNA algorithm (Rother et al., 2011) allows the insertion of m^1A at predefined positions by aligning shared atomic coordinates between the reference and the modified nucleotide. This procedure introduced the methyl group at N^1 by replacing the corresponding hydrogen but does not reposition the nucleotide or adjust the local backbone geometry. Consequently, this rigid substitution can lead to steric conflicts that are not reconciled structurally and are inconsistent with the CD spectra, which indicate some degree of thermal stability.

As an alternative to NMR spectroscopy, which provides atomic-resolution data but is limited by experimental cost and scalability, molecular dynamics (MD) simulations using the AMBER software package (Case et al., 2020; ambermd.org) represent a viable strategy. This approach would involve parameterizing the modified nucleotides—such as m^1A and m^6A —using tools like ANTECHAMBER and RESP charge fitting to generate force-field-compatible topologies. Once incorporated into oligonucleotide structures, these modified systems can be subjected to energy minimization, solvation, equilibration, and long-timescale MD simulations. The

resulting trajectories allow for the evaluation of steric compatibility, conformational fluctuations, hydrogen bonding networks, and solvent interactions at atomic detail.

In contrast to static models, MD provides temperature-resolved structural ensembles that can be correlated with CD signal evolution. This would facilitate validation against experimental spectra while enabling predictive modelling for uncharacterized sequences. The ability to simulate thousands of conformations across modified RNA sequences opens the possibility of constructing a curated structural database. Such a dataset could serve as a reference for benchmarking CD-based predictions, guiding folding pathway analysis, and training future machine learning models aimed at simulating RNA with arbitrary chemical modifications.

5 Conclusions

This thesis demonstrated that post-transcriptional RNA modifications exert position- and context-dependent effects on structural stability, folding behaviour, and thermodynamic adaptability. Through the integration of MODOMICS, NACDDB, and spectral alignment tools, it was possible to quantify and compare the impact of individual modifications on local and global RNA structure.

The first achievement involved the comparative analysis of conserved modification *loci* across tRNA and rRNA alignments. Uridine derivatives, particularly pseudouridine (Ψ) and dihydrouridine (D), were analysed in relation to base stacking interactions, revealing divergent structural outcomes. Ψ consistently reinforced stacking-class interactions, especially in constrained ribosomal regions, supporting its role in stabilising the RNA backbone. In contrast, D preferentially disrupted helical structure, promoting local flexibility.

Secondly, the spectroscopic characterisation of methylated RNAs using synchrotron-based circular dichroism allowed the identification of modification-induced spectral shifts. For instance, N^6-methyladenosine (m^6A) increased CD signal intensity under room conditions but was associated with thermal destabilisation, consistent with a loss of solvent hydrogen-bonding and increased hydrophobicity. Conversely, N^1-methyladenosine (m^1A) exhibited context-dependent behaviour, disrupting Watson–Crick interactions when placed in structured helices, but being tolerated or even stabilising in loop regions such as tRNA position 58.

Importantly, these *in vitro* findings align with evolutionary constraints. Modification sites conserved across phylogenetic domains—especially in thermophilic archaea—were shown to coincide with spectroscopically stable configurations, reinforcing the

hypothesis that evolution selectively retains modifications that contribute to structural resilience.

The technical improvements to MODOMICS and the construction of NACDDB represent a further contribution. MODOMICS now enables position-specific evolutionary comparisons of RNA modifications, while NACDDB permits quantitative comparisons of CD profiles. These tools collectively enable multi-modal correlation of sequence conservation, structural impact, and functional relevance.

However, current structure prediction models are limited in their ability to accommodate chemical modifications. While AlphaFold$_3$ permits modelling of a small subset of methylations, most modifications remain unsupported. ModeRNA's rigid replacement protocol does not account for steric repacking, often generating models inconsistent with experimental observations.

To bridge this gap, molecular dynamics simulations using the AMBER suite are proposed as a viable alternative to costly NMR approaches. Such simulations would involve parameterisation of modified residues using tools like ANTECHAMBER and RESP, followed by energy minimisation, solvation, and temperature-controlled dynamics. This approach would enable evaluation of solvent accessibility, stacking disruption, hydrogen-bond networks, and conformational sampling, aligning MD-derived ensembles with CD signal evolution. Over time, this strategy could produce a curated simulation-based dataset of modified RNAs, ultimately guiding folding pathway prediction and machine learning efforts in RNA modification modelling.

In summary:

- RNA modifications are structurally and evolutionarily conserved.
- Ψ reinforces base stacking; D enhances flexibility; m^{6}A and m^{1}A induce thermal and conformational variation in a context-specific manner.
- The 200–210 nm CD band is an independent helicity read-out—track $\Delta\varepsilon_{205}$; its drift towards zero signals A-form loss.
- NACDDB enables spectral comparison; MODOMICS provides evolutionary annotation; MD may soon offer predictive modelling.
- The next frontier lies in developing predictive models of RNA 3D structure that account for chemical modifications, expanding our understanding of structure–function relationships in non-coding RNAs and ribonucleoprotein complexes.

6 Future Perspectives

The landscape of RNA modifications and nucleic acid spectroscopy is rapidly evolving, and both MODOMICS and NACDDB must expand accordingly. Over 150 distinct RNA modifications are now known (Cappannini et al., 2024), and continued discovery by high-resolution structural and sequencing methods will further broaden this repertoire. The MODOMICS update of 2023 already expanded its catalogue to include both natural and synthetic modified nucleosides and introduced new nomenclature to manage them (Cappannini et al., 2024). MODOMICS should continue to ingest newly identified modifications, for example, bases revealed by cryo-EM mapping of ribosomes and mass spectrometry and link them to the enzyme pathways that produce them (Boccaletto et al., 2022). Similarly, NACDDB can extend its sequence space by including modified RNAs; it already uses MODOMICS notation for modified residues and links each to the corresponding MODOMICS entry (Cappannini et al., 2023). In practice, NACDDB curators and users should prioritize depositing and annotating CD spectra of RNAs containing novel modifications, so that the databases remain comprehensive. For instance, high-resolution structures of the ribosome have identified specific modified nucleotides (e.g. 2′-*O*-methylguanosine and acetylcytidine) that were absent from earlier annotations. These have already been added to MODOMICS (Cappannini et al., 2024), and in future NACDDB could similarly catalogue their CD signatures. In short, both resources will need continuous updating: MODOMICS to enumerate new modifications and their synthetic variants, and NACDDB to map those modifications onto spectral data.

Cryo-electron microscopy has provided structural maps that pinpoint the location of modified nucleotides in rRNA, validated by matching mass spectrometry signals. For example, modified nucleotides such as $G_{m\text{-}436}$ and ac^4C_{1842} have been detected through electron density and subsequently confirmed chemically (Boccaletto et al., 2022). These new sites of modification should be incorporated into MODOMICS, while NACDDB could aim to collect and annotate the CD spectral features of synthetic oligonucleotides bearing the same modifications. This cross-linking between structure and spectroscopy enhances the biochemical context of each modification and allows researchers to trace how it manifests in biophysical data.

Modern experimental and computational methods will transform how data are generated and analysed. Deep sequencing and direct RNA nanopore reads now detect modifications transcriptome-wide. For example, direct RNA sequencing captures raw signal shifts caused by modified bases, allowing machine-learning algorithms to

identify modifications (e.g. m^6A) in each molecule (Hendra et al., 2022). In parallel, high-throughput mass spectrometry and structural proteomics rapidly characterize enzyme–RNA interactions and modification sites (Wang et al., 2022). To leverage these data, MODOMICS and NACDDB should plan interfaces for high-throughput data submission and validation. Automated pipelines (e.g. modification-calling software or deep learning predictors) could feed results into the databases, with flagging for manual curation of novel entries. Recent advances in AI-driven structure prediction will also be relevant: for example, RASP v2.0 now collects hundreds of transcriptome-wide SHAPE/DMS probing datasets and uses deep learning to impute missing signals (Mu et al., 2025), and RhoFold+ (a language-model approach) can accurately predict RNA 3D folds from sequence (Shen et al., 2024). MODOMICS might incorporate such predictions to suggest unknown modification sites or to model the impact of modifications on RNA structure. Likewise, NACDDB could adopt machine learning tools to predict a CD spectrum from a given sequence or secondary structure (or vice versa), enhancing its analytical capabilities. In summary, the databases should not only catalogue experimental data but also integrate outputs from computational pipelines and high-throughput assays to remain current.

Scalable infrastructure and interoperability will be key to handle these advances. Both databases should implement robust, well-documented APIs and adopt FAIR data principles, so that data are machine-findable and reusable (Wilkinson et al., 2016). For example, the BioThings framework demonstrates how multiple bio-databases can be made accessible through high-performance REST APIs (Moore-Kelly et al., 2019). Similar strategies could allow MODOMICS and NACDDB to serve large queries (e.g. retrieving all entries for a given modification or structural motif) and to interoperate with other resources. Cloud-based and microservice architectures will support on-demand scaling as datasets grow. Embedding ontology links and standardized metadata (consistent with FAIR) will enable users and algorithms to navigate between entries easily. Finally, enhanced bulk-download options and data mirrors will ensure that researchers can access complete datasets for offline analysis. By investing in flexible, scalable technology (following guidelines for bioinformatics infrastructure), MODOMICS and NACDDB can accommodate the "big data" nature of epitranscriptomics and structural biology.

The experimental repertoire for probing RNA and nucleic acids is expanding, and the databases should reflect this. High-throughput chemical probing methods (SHAPE, DMS, etc.) and ligation-based assays generate transcriptome-wide secondary structure maps that are increasingly abundant. RASP v2.0 alone added hundreds of structure-probing datasets and even provides RNA–RNA interaction profiles (Mu et al., 2025).

Both databases could consider linking such structural information to modification data: for example, seeing how a modification influences local flexibility in SHAPE data or alters an interaction footprint. New modalities like single-cell structure probing and cryo-electron tomography of RNA complexes will also emerge; capturing metadata from these experiments could reveal context-specific modification patterns. Beyond purely structural data, emerging biophysical methods (e.g. time-resolved spectroscopy, single-molecule FRET on RNAs) could be integrated by NACDDB or sister resources to give a multi-faceted view of conformation. As RNA biology embraces integrative and dynamic assays, the databases should provide mechanisms for users to upload and link new experiment types.

Circular dichroism spectroscopy will continue to grow as a tool for nucleic acid characterization, and NACDDB should support this trend. The database already stores curated CD and SRCD spectra for DNA, RNA, hybrids, and derivatives (Cappannini et al., 2023), but future development could make it even more useful. For instance, NACDDB should facilitate high-throughput CD measurements, such as those performed using automated capillary systems that analyze 96-well plates (Moore-Kelly et al., 2019). These setups can profile large libraries of sequences and experimental conditions. Machine learning could then learn from these data to predict structural features; for proteins, analogous tools already link CD signals to secondary structure classification. NACDDB could implement similar algorithms for nucleic acids, helping users infer helix versus loop content or unusual conformations from a spectrum. In addition, integrating CD data with other spectroscopies, such as infrared or Raman, would enrich biophysical annotation. Since CD requires minimal sample and is non-destructive, it is well suited for validating structures of modified RNAs or complexes. Thus, expanded NACDDB coverage will directly support studies of modified nucleic acids under diverse conditions.

In summary, the complementary futures of MODOMICS and NACDDB lie in integration and expansion. Both resources should continue cross-linking—for example, NACDDB entries already point to MODOMICS pages for each modified base (Cappannini et al., 2023)—so that a modification's chemical, structural, and spectral information are unified. Encouraging community submissions (new spectra for NACDDB, new modifications for MODOMICS) and collaborations with experimental consortia will keep the data current. Together, these enhancements will ensure that MODOMICS and NACDDB remain indispensable, reflecting the latest trends in post-transcriptional and post-translational biology, RNA structure, and spectroscopy.

7 Materials and methods

The following sections describe the methods used to implement the databases and analyse their content. Programming languages and mathematical approaches are presented to establish the computational and analytical framework of the study.

7.1Computational resources and hardware

MODOMICS and NACDDB are hosted on dedicated virtual machines within the server infrastructure of the International Institute of Molecular and Cell Biology in Warsaw (IIMCB). Data analysis, RNA alignments, statistical processing, and manuscript preparation were carried out on a 2023 ACER laptop (Ryzen 9, 16 GB RAM) using NitroSense and CoolBoost.

7.2 Programming languages

The development and analysis described in this thesis required the integration of multiple programming languages to support database implementation, data processing, and spectral comparison tools. Each language was selected based on its suitability for a specific class of tasks: Python for data processing and backend development, JavaScript for client-side interactivity and structural visualisation, and R for statistical analysis and data visualisation, which are further described in the following sections.

7.2.1 Python

Python was used for data processing and database implementation. It is an interpreted, high-level language designed for general-purpose tasks. The default implementation, CPython, is written in C and functions as the reference interpreter. CPython translates Python code into bytecode, which is executed by the Python virtual machine (PVM). It supports dynamic typing and automatic memory management via a garbage collector based on reference counting and cycle detection.

Python syntax is processed through lexical analysis and parsing, producing abstract syntax trees (AST). These trees are then compiled into platform-independent bytecode, which the interpreter executes. This architecture underlies Python's classification as an interpreted language.

Python 3.12 was used for MODOMICS maintenance and NACDDB implementation. A functional, module-based approach was adopted to ensure maintainability and readability. Extended procedures were encapsulated into reusable functions executed within broader workflows. Object-oriented design was applied where appropriate, notably in Django SQL model definitions and in scripts handling specific data components.

Datasets for statistical analysis were also prepared using Python. Best practices defined in PEP 8 (https://peps.python.org/pep-0008/) were followed to maintain consistency across the codebase.

7.2.2 Javascript

JavaScript is a high-level, interpreted programming language that conforms to the ECMAScript specification maintained by Ecma International. It is used for both client-side and server-side web development. The language is dynamically typed, event-driven, and single-threaded.

JavaScript code is executed by engines embedded in browsers or standalone environments. The V8 engine, developed by Google and written in C++, is the most widely adopted implementation. JavaScript engines convert source code into abstract syntax trees (ASTs) via a parser. These ASTs are either interpreted as bytecode or compiled into machine code using Just-In-Time (JIT) techniques to improve execution speed.

Memory management is handled by a garbage collector that automatically releases memory from unreachable objects. JavaScript's concurrency model is based on an event loop, which manages asynchronous operations through callbacks, Promises, and async/await constructs.

Integration with runtime environments allows interaction with platform-specific APIs. In browsers, JavaScript interfaces with the Document Object Model (DOM) and network components, while in Node.js, it accesses file systems and system-level modules.

7.3 Database Backend

The backend defines the server-side logic and infrastructure that support data storage, retrieval, and processing. It operates independently of the user interface and enables interaction between applications and databases. The following sections describe the frameworks and implementations used to construct the MODOMICS and NACDDB backends.

7.3.1 Django

Within the scope of this study, Django version 3.0.5 was used for both MODOMICS and NACDDB (Cappannini et al., 2023; Cappannini et al., 2022), ensuring consistency across backend implementations **(https://djangoproject.com/).** Django is a high-level Python web framework intended for structured, database-driven application development. It follows the Model-View-Controller (MVC) architectural pattern,

separating application logic, user interface, and data storage. This separation simplifies implementation and maintenance.

The framework promotes the DRY principle, supporting code reusability and modularity. Django includes a set of integrated components such as an object-relational mapper (ORM) for database interaction, an administrative interface for content management, and tools for user authentication, URL routing, and template rendering.

Security features are embedded in the framework, including protections against SQL injection, cross-site scripting (XSS), and cross-site request forgery (CSRF). These functionalities allow the development of scalable, secure web applications.

7.3.2 Nginx

To serve Django-based applications in a production environment, a web server is required to manage client connections and route requests. Nginx (**https://www.nginx.com/**) is an open-source web server used for this purpose in both MODOMICS and NACDDB. It employs an event-driven, asynchronous architecture that enables efficient handling of multiple concurrent connections.

Nginx is used to deliver static files, forward client requests to the Django backend via reverse proxying and manage SSL termination where applicable. Its modular design supports load balancing and caching, reducing the computational load on application servers. These features ensure stable and scalable operation in web-facing deployments.

7.3.3 Gunicorn

Gunicorn (Green Unicorn, https://gunicorn.org/) is a WSGI-compliant server that bridges web servers like Nginx with Python applications, such as those built using Django or Flask. Gunicorn employs a pre-fork worker model that allows for multiple worker processes to handle incoming requests concurrently, making it highly scalable.

Gunicorn efficiently delegates the task of executing Python code, enabling seamless interaction between the web server and the application. Its robust configuration options allow fine-tuning for better performance, particularly in handling multiple requests while ensuring the database remains responsive under heavy loads.

7.3.4 Operating system

Data analyses and databases have been carried out on Linux Ubuntu/Windows Subsystems for Linux (WSL). Linux is a Unix-like operating system kernel used as the base for open-source distributions. It manages core functions such as process scheduling, memory allocation, file system control, and hardware abstraction.

The kernel operates in conjunction with user-space utilities, libraries, and package managers to form complete systems. These systems support multitasking, multi-user environments, and access control mechanisms.

The MODOMICS and NACDDB servers operate on Ubuntu, a Debian-based Linux distribution commonly used in scientific and web server environments.

7.3.5 R programming language

R is a programming language and software environment designed for statistical computing. It provides built-in support for linear and nonlinear modelling, statistical tests, time series analysis, and graphical representation of data. R includes a large collection of packages, many of which are developed for specialised data analysis workflows. In this study, R was used for data visualisation and clustering, specifically through the pheatmap package.

7.3.6 SQL backend

SQL (Structured Query Language) is a standardised programming language used for managing and querying relational databases. It operates on structured data organised in tables, where columns represent attributes and rows represent records. SQL supports operations such as data selection, aggregation, joins across multiple tables, and the creation of views, which are virtual tables defined by query results.

In this work, SQL was used through the SQLite engine, which supports local data storage without requiring a separate database server. NACDDB and MODOMICS use SQLite to manage relational data structures, including RNA modification entries, sequence annotations, and spectrum metadata.

SQLite is a self-contained, embedded database engine that implements a subset of the ANSI SQL standard. It does not support advanced features such as full outer joins or concurrent write access but provides sufficient functionality for small to medium-scale applications. This architecture simplifies deployment and reduces external dependencies, which is consistent with the design of both platforms. The choice of SQLite reflects its compatibility with Python-based frameworks and its ability to support structured queries while operating under limited system resources.

7.3.7 Large language models

OpenAI's ChatGPT was employed exclusively to standardize the linguistic and stylistic features of the text. The model was prompted with explicit constraints to maintain a scientific register defined by formal tone, concise syntax, absence of subjective qualifiers, and adherence to evidentiary neutrality. Demonstratives, anthropomorphic phrasing, and non-informative adjectives were excluded to ensure impersonal

expression and semantic precision. This protocol was implemented to mitigate a common linguistic bias in large language models: the tendency to introduce descriptors lacking scientific function—such as “transformative,” “crucial,” or “essential.” These elements were systematically removed to preserve clarity and information density, following a compositional standard herein referred to as *“CHaD-style”* (Context, Hypothesis, and Data). The resulting text was subsequently verified and edited to eliminate any deviation from the original empirical content, interpretative intent, or scientific formulation.

7.4 Database frontend

The frontend represents the client-facing component of the database platforms. It handles the visual presentation of data and supports user interaction through responsive web interfaces. In both MODOMICS and NACDDB, frontend elements were designed to facilitate structured access to RNA-related content, including annotations, alignments, and spectral data.

HTML, CSS, and JavaScript were used to build the interfaces, with additional support from CSS frameworks for maintaining layout consistency and rendering components. The MODOMICS frontend was implemented using Bootstrap, while NACDDB uses the BULMA framework. Both frameworks simplify layout construction through pre-defined grid systems and allow consistent visual formatting across devices.

These frontend components enable dynamic display of database entries, search outputs, and visualisations, and are integrated with backend services to allow efficient data retrieval and user-driven queries. The following sections describe the library used for the MODOMICS and NACDDB databases.

7.4.1 Bootstrap frontend

The MODOMICS frontend is implemented using the Bootstrap CSS framework. Bootstrap is an open-source library that provides a standardised set of components for building responsive web interfaces. It includes pre-defined classes for layout, typography, forms, buttons, and navigation elements, which support consistent styling across browsers and devices.

The framework incorporates a grid system that enables adaptive layouts, allowing the interface to adjust to different screen resolutions. This supports usability on both desktop and mobile platforms. Bootstrap was originally developed at Twitter and remains under active maintenance (https://getbootstrap.com).

7.4.2 BULMA CSS

The NACDDB frontend is implemented using the BULMA CSS framework. BULMA is an open-source library based on the Flexbox layout model. It provides a collection of pre-defined classes for building responsive interfaces, including components for layout, typography, forms, and buttons.

BULMA enables the development of adaptable page structures that function across different screen sizes. Its use of semantic HTML and Flexbox supports layout control without extensive custom styling. The framework is suitable for projects requiring maintainable and accessible web interfaces. BULMA is used in NACDDB to ensure consistency in frontend behaviour across devices (https://bulma.io/).

7.5 Methods used in MODOMICS

The following sections describe the computational procedures used to support MODOMICS. These include RNA secondary structure alignment, modification annotation, data classification, and clustering. The methods were applied to tRNAs and rRNAs using both structural and sequence-based approaches to identify conserved modification positions. All procedures were implemented to enable consistent mapping, reproducibility, and integration within the MODOMICS system.

7.5.1 Methods for tRNA alignments

When available, secondary structures were obtained by visual inspection of experimentally resolved 3D structures or sourced from original publications and databases (Sprinzl et al., 1996). The dot-bracket notation included only Watson–Crick base pairs. Square brackets were used to indicate kissing loop interactions between the D-loop and TΨC loop.

For sequences without available structural references, secondary structures were assigned by identifying similar sequences within the MODOMICS tRNA dataset. The resulting structures were aligned to the consensus secondary structure (CS).

Alignment to the consensus required adjusting secondary structures to minimise unpaired gaps in stem regions. Gaps were accepted only if the complementary pairing position also contained a gap, defined as a Paired Gap (PG). These were restricted to the termini of stems. Loops could contain uneven gaps, provided nucleotide positions corresponded with those in available PDB structures.

Within stem regions, non-canonical interactions were tolerated under the following conditions:

- The anticodon position conformed to both the secondary structure and the consensus.
- At least two stems contained either perfect Watson–Crick matches or quasi-perfect matches, where the latter permitted GU wobble pairs but excluded other non-canonical interactions.

Once alignment between the secondary structure (SS) and the consensus structure (CS) was validated, the nucleotide sequence was aligned to the SS. The Varna applet was used to confirm alignment accuracy.

7.5.2 Methods for other family alignments

Infernal software (INFERence of RNA Alignments) (Nawrocki et al., 2013) was used to align all RNA sequences except tRNAs. Infernal performs RNA alignments using covariance models (CMs), which incorporate both sequence and secondary structure information derived from multiple sequence alignments.

The alignment process consisted of the following steps:

i. Covariance models for each RNA family were downloaded from Rfam.
ii. Seed alignments were collected from Rfam for each RNA family.
iii. RNA sequences from MODOMICS were processed in their unmodified form.
iv. Infernal was used to align the sequences.
v. Manual curation was applied to refine the resulting alignments.

Ribosomal RNA (rRNA) alignments presented in this work include newly integrated sequences from recent studies (Piekna-Przybylska, 2006; Marchand et al., 2020; Incarnato et al., 2017; Krogh et al., 2020; Motorin et al., 2021). These alignments differ from those included in the MODOMICS 2023 release and are available in the supplementary file (THESIS_ALIGNMENT.xlsx).

7.5.3 Memory optimization for RNA alignments using swap file allocation

To cope with demanding memory resource requirements for RNA alignments temporarily allocation of virtual memory was employed. The following steps have been implemented:

i. Creation of a swap file by the command sudo dd if=/dev/zero of=/swapfile bs=1M count=12288 to create a 12GB swap file, where bs=1M specifies a block size of 1MB, and count = 12288 allocates 12GB of disk space

Permissions were set by ensuring the swap file was secured by restricting access to root using sudo chmod 600/swapfile. Swap file was initialized using the command sudo mkswap/swapfile. The swap file was temporarily activated using sudo swapon/swapfile

7.5.4 Structural alignment to map rRNA modifications

Experimentally resolved rRNA structures were used to map RNA modifications and determine their structural positions. The following RCSB PDB entries were selected: 7PZY.cif, 4UG0.cif, 6QZP.cif, 7BHP.cif, 7QIZ.cif, and 7O7Y.cif. All structures were aligned to the human reference structure 4UG0.cif using the ChimeraX MatchMaker tool. To simplify visual representation, 4UG0.cif was used as the structural background in all figures.

7.5.5 Annotation of RNA modification in MODOMICS

A systematic review of the literature was conducted, with a focus on recovering as many publications related to MODOMICS RNA sequences as possible.

The reliability of the modification data was assessed using two classification systems: level of experimental evidence and level of estimated reliability.

Transfer RNA sequences were further annotated with their secondary structures, which were either obtained directly from the publications or manually derived from 3D structures. When neither of these sources were available, the tRNA structures were inferred by identifying similarities within the MODOMICS tRNA pool.

7.5.6 Level of experimental evidence

The level of experimental evidence of RNA modification was classified (Cappannini et al., 2023). Experimental evidence indicates the level of experimental or computational support for a given data feature, along with its source (e.g. specific publication, database import, method used, computational prediction or human inference from metadata).

The level of experimental evidence has been annotated as described in **Tab. 9**. As for instance, let's consider an experimentally determined structure with 4 Angstrom (Å) resolution whose modifications have been confirmed by also MS. Not all the atoms would be clearly distinguishable yet confirmed by another direct technique. The modification would have a score of 1.

If identified by one direct technique alone, the score would be 2. As per another illustration, indirect methods for RNA modification identification, such as antibody-dependent techniques, face significant challenges, including off-target binding and

cross-reactivity, which often lead to false positives (Zhang et al., 2019), as well as limitations in enrichment efficiency and library preparation that amplify noise and undermine data reliability (Dominissini et al., 2012).

Contamination of rRNA in mRNA samples often arises due to limitations in poly(A)-based purification protocols. Such contamination can interfere with N^6-methyladenosine (m^6A) profiling, leading to inaccuracies in modification mapping (Garcia-Campos et al., 2019). This issue highlights the potential for artificial detection of RNA modifications, where substantial errors may persist. The associated score should be set to 3, as significant inaccuracies can affect downstream results.

To explain the fourth category, one possible scenario involves the identification of DRACH motifs. The DRACH sequence motif (D = A, G, or U; R = A or G; H = A, C, or U) is a well-characterized feature associated with N^6-methyladenosine RNA modifications. A significant proportion of m^6A modifications have been shown to occur within DRACH motifs, reflecting their specificity and enrichment as preferred target sites for m6A methyltransferases, including METTL3 and METTL14.

This motif is predominantly identified in messenger RNA (mRNA), where m6A marks are enriched near stop codons and within 3′ untranslated regions (3' UTRs). These modifications impact RNA stability, splicing, and translation efficiency. DRACH motifs are also prominent in long non-coding RNAs (lncRNAs), where m^6A facilitates structural modulation and functional interactions critical for gene regulation (Wang et al., 2020**;** Shachar et al., 2024).

Additionally, DRACH motifs have been observed in viral RNAs, including SARS-CoV-2, where m^6A modifications influence viral replication and host immune responses (Campos et al., 2021). The sequence specificity of DRACH motifs allows their identification using computational tools, such as algorithms based on regular expressions, enabling a predictive approach to locate potential m^6A methylation sites.

This type of prediction, independent of experimental validation, would fall under category 4. The same classification applies to other computational analyses, such as identifying conserved domains in aligned RNA sequences where strict nucleotide conservation is observed. In contrast, a score of 5 indicates that no information is available for the modification at the given position.

7.5.7 Level of estimated reliability

The estimated reliability scores in MODOMICS offer a structured evaluation of the confidence associated with the provided data. These scores are based on the available evidence and potential sources of error. The levels of estimated reliability of the

MODOMICS RNA modifications were assessed relying on six different qualification classes as described in **Tab. 10**.

7.5.8 Methods classification for modification annotation

The quality classification of the MODOMICS RNA modification data is based on the detection methods employed and, where applicable, the availability of RNA structural data. For RNA structures, a visual assessment was performed to evaluate the resolution and level of detail in the models. The highest reliability scores were assigned to

Level	Description	Explanation
1	Multiple Direct Evidence	Identified and characterized experimentally by at least two direct method
2	Direct Experimental Evidence	Identified and characterized by one direct method
3	Inferred Based on indirect experimental evidence	Inferred from experimental data based on indirect method e.g. mutation patterns in cDNA sequencing
4	Predicted Computationally	e.g., predicted from sequence, without experimental data
5	Unknown	Experimental evidence not yet annotated
6	Irrelevant	Experimental evidence cannot be estimated for a given data type

Table 9 Classes of Experimental Evidence

modifications identified through direct experimental methods, which offer higher confidence. Indirect methods, while informative, received lower scores due to their inherent limitations.

Level	Description	Explanation
1	Rock Solid	It would be surprising to see it proven wrong
2	Solid	Large errors are unlikely, but cannot be completely excluded, minor errors should not undermine the meaning of the data
3	Speculative	More likely to be correct than incorrect, major errors may happen and minor errors are quite likely
4	Questionable	e.g., different results in publications and/or database entries
5	Unknown	Reliability assessment has not yet been performed
6	Irrelevant	Reliability cannot be estimated for a given data type

Table 10 Classes of Experimental Reliability

The full scoring criteria are outlined in **Tables 9 and 10**, which describe the ranking system for data quality. **Table 11** details the rationale for classifying methods as either direct or indirect, ensuring transparency in the evaluation process. Direct methods include techniques that unequivocally pinpoint modifications, such as high-resolution mass spectrometry or X-ray crystallography. In contrast, indirect methods rely on inference or computational predictions, introducing greater uncertainty into the data.

7.5.9 tRNA data analysis and clustering

The analysis of tRNA alignments identified and quantified nucleotide modifications at conserved positions using a systematic approach. A numeric vector was generated for each modification, with its elements corresponding to the conserved positions in the tRNA alignment. When a modification was present at a specific position, the corresponding element was incremented by 1. The value of each vector element therefore indicates how frequently the position is edited by that modification.

Method Name	General/Specific	Direct/Indirect	Method Type	
High-resolution X-ray crystallography	General	Direct	Biophysical	PUBMEDID: 31619810
High-resolution NMR	General	Direct	Biophysical	PUBMEDID: 30997719
High-resolution cryo-EM	General	Direct	Biophysical	PUBMEDID: 29143818
MS/MS	General	Direct	Biophysical	PUBMEDID: 1706062
RiboMethSeq	Specific (2''-O-Me)	Indirect	Chemical probing + sequencing (CPS)	PUBMEDID: 27302133
ICE-Seq	Specific (I)	Indirect	CPS	PUBMEDID: 25855956
M1A-Seq	Specific (m1A)	Indirect	Immunoprecipitation + sequencing (IS)	PUBMEDID: 29072297
ARM-Seq; DM-tRNASeq	Specific (m1A, m3C, m1G, m,2,2G)	Indirect	RNA de-modification + sequencing	PUBMEDID: 26237225
meCLIP	Specific (m6A)	Indirect	IS	PUBMEDID: 33376190
AlkAnilineSeq	Specific (m7G and m3C)	Indirect	CPS	PUBMEDID: 30370969
JACUSA2	Specific (Q)	Indirect	Biophysical + sequencing (BS)	PUBMEDID: 37811872
HydraPsiSeq	Specific (Ψ)	Indirect	CS	PUBMEDID: 32976574
hmC-CATCH	Specific (hm5C)	Indirect	Chemical treatment + sequencing (CTS)	PUBMEDID: 30278133
HAC-Seq	Specific (m3C)	Indirect	CTS	PUBMEDID: 33313824
bisulfite RNA sequencing	Specific (m5C)	Indirect	CTS	PUBMEDID: 19059995
m7G-MaP-seq	Specific (m7G)	Indirect	CTS	PUBMEDID: 31504776
TRAC-seq	Specific (m7G)	Indirect	CTS	PUBMEDID: 31619810

Table 11 Examples of modification detection methods and their classification as direct or non-direct. This table is also used in MODOMICS to explain the rationale beyond modification classification (Cappannini et al. 2023)

Cluster heatmaps were normalized along columns to a range between 0 and 1, ensuring that each cell value represented the relative selective pressure acting on a modification at a given position. Columns containing only zero values, reflecting the absence of modifications, were excluded from calculations to maintain clarity and focus. This method established a quantitative framework for assessing the conservation and evolutionary significance of tRNA modifications across the alignment.

7.6 Methodologies used in NACDDB

The following sections describe the computational procedures used to create and expand NACDDB. These include its frontend and backend, together with the analysis implemented on modified RNA sequences.

7.6.1 Treating modified RNA spectra

RNA CD spectra and thermal acquisitions were treated either with CDToolX (Miles et al., 2018) or with homemade python scripts. The following pipeline was followed:

i. Averaging baselines in relation to each experiment
ii. Averaging of the camphor sulfonic acid (CSA) acquisitions
iii. Averaging of the thermal scans of each experiment for a specific temperature
iv. Baseline subtraction to the thermal acquisitions
v. CSA calibration

Best practices for CDToolX data processing were adhered to, as described in the original supplementary materials and demonstrated in the provided reference (https://www.youtube.com/watch?v=ajNkfi9OzBU&t=225s).

7.6.2 Converting rotational values to molar circular dichroism

In circular dichroism (CD) spectroscopy, ellipticity (θ) and delta epsilon ($\Delta\epsilon$) are parameters used to describe the interaction between circularly polarized light and chiral molecules. Ellipticity (θ) measures the rotation of polarized light as it interacts with a chiral sample, expressed in degrees (°) or millidegrees (mdeg). It reflects the extent to which the plane of polarized light is altered due to molecular chirality and is commonly used to evaluate the secondary structures of proteins and nucleic acids (Kelly et al., 2005).

Delta epsilon ($\Delta\epsilon$) provides a quantitative measure of CD, representing the difference in molar absorption coefficients between left- and right-circularly polarized light (ϵL and ϵR, respectively). $\Delta\epsilon$ is expressed in molar absorption units (L mol^{-1} cm^{-1}) and is calculated as:

$$\Delta\varepsilon = {\vartheta \cdot m.r.w.}/{l \cdot c \cdot \mu}$$

Where:

- θ is the ellipticity (in degrees),
- m.r.w. is the mean residue weight (g/mol),
- l is the optical path length (cm),
- c is the concentration (mol/L),
- μ is the conversion factor (32982, Miles et al., 2018).

The Δϵ parameter allows for concentration-independent comparisons of CD signals across different samples and experimental conditions, providing precise quantification of chiral interactions. For modified RNAs, converting CD spectra from ellipticity to Δϵ facilitates standardized data analysis, enabling more precise and comparable assessments of structural changes and molecular interactions. This conversion connects qualitative and quantitative analysis, making Δϵ a key parameter for evaluating chiral properties in nucleic acid studies (Greenfield, 2006).

7.6.3 Root mean square deviation and comparison inherent algorithm

Root mean square deviation (RMSD) was used as one of the algorithms to identify similarities between CD spectra. The algorithm identifies common wavelength ranges between the selected spectra and extracts the corresponding delta epsilon values for these ranges. The RMSD value between the two vectors is then calculated using the following formula:

$$RMSD = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (v1_i - v2_i)^2}$$

where:

- N is the number of common wavelength points
- and are $v1 - 2_i$ the delta epsilons values for the i^{th} common wavelength values.

7.6.4 Cosine similarity

Cosine similarity (CS) was used to implement another algorithm for identifying similarities between CD spectra. Similar to the previous algorithm, it identifies common wavelength ranges between the selected spectra and extracts the

corresponding delta epsilon values. Cosine similarity quantitatively assesses the angular relationship between the two spectral vectors using the following formula:

$$\vartheta(A,B) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

where:

- $A \cdot B$ represents the dot product of vectors A and B, calculated as sum of their i_{th} components:

$$A \cdot B = \sum_{N}^{1} A_i B_i$$

- and $\|A\|, \|B\|$ are the Euclidean norms (magnitudes) of vectors A and B, computed as:

$\|A\| = \sqrt{\sum_N^1 A_i^2}$, and $\|B\| = \sqrt{\sum_N^1 B_i^2}$, respectively

7.6.5 Pearson correlation coefficient

The Pearson correlation coefficient (ρ) is a statistical measure that quantifies the strength and direction of the linear relationship between two variables, X and Y. It is denoted by the formula:

$$\rho = \frac{\sum_{i=1}^{N}(X_i - \bar{X}) \sum_{i=1}^{N}(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2}}$$

- X_i, Y_i are the individual data points of variables X and Y, respectively.

- $\bar{X}$ and $\bar{Y}$ are the means (averages) of the X and Y data sets:

$\bar{X} = \frac{1}{N}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{N}\sum_{i=1}^{n} Y_i$

n is the number of data points, and ρ ranges from -1 to 1. Person correlation coefficient has been used as a metric in the corrplots in the spectra data analyses

7.6.6 Kendall's Tau Correlation

Kendall's tau correlation coefficient (τ) is a non-parametric measure used to assess the strength and direction of association between two ranked variables. Unlike other correlation coefficients that operate on the actual values of the data, Kendall's tau focuses on the order of observations, making it particularly robust against non-normality and outliers.

This method is used to compare RMSD and CS similarity, specifically when providing complementary information on CD spectrum alignment. Kendall's tau is calculated by comparing pairs of observations involves (x_i, y_i) from two variables. The formula for Kendall's tau is given by:

$$\tau = {(C - D)} \Big/ \left({}^{1}/_{2} * n * (n-1)\right)$$

where:

- C is the number of concordant pairs (*i.e.,* pairs where the ranks for both variables move in the same direction).

- D is the number of discordant pairs (*i.e.,* pairs where the ranks for the variables move in opposite

- n is the total number of observations

A concordant pair occurs when both $x_i < (>)\ x_j$ and $y_i < (>)\ y_j$. Conversely, a discordant pair occurs when the comparisons are discordant.

The value of τ ranges between -1 and 1:

- τ = 1 indicates a perfect positive correlation, where all pairs are concordant.
- τ = -1 indicates a perfect negative correlation, where all pairs are discordant.
- τ = 0 suggests no correlation, implying the ranks of one variable do not predict the ranks of the other.

7.6.7 K-means clustering

K-means clustering is an unsupervised machine learning algorithm that partitions a dataset into 'k' clusters based on feature similarity. It assigns each data point to one of the clusters by minimizing the within-cluster variance. The number of clusters, 'k', is predetermined. In this analysis, K-means was applied to RMSD and cosine similarity values from spectral comparisons, grouping spectra with similar features, including anti-aligned spectra sharing inversion patterns. The K-means algorithm operates iteratively through the following steps:

i. Initialization: The algorithm randomly selects 'k' initial centroids from the dataset.

ii. Assignment: Each data point is assigned to the nearest centroid, forming 'k' clusters.

iii. Update: The centroids are recalculated as the mean of all data points within each cluster.

iv. Repeat: Steps 2 and 3 are repeated until the centroids stabilize, or the maximum number of iterations is reached.

Mathematically, the K-means objective is to minimize the following cost function:

$$J = sum\left(sum\left(\left\|x_i - \mu_j\right\|^2\right)\right)$$

where:

- J is the objective function (total within-cluster variance)
- $x_i$represents a data point
- $\mu_i$is the centroid of the cluster j
- $\left\|x_i - \mu_j\right\|^2$is the squared Euclidean distance between a data point and its centroid

7.6.8 Analysis of RNA modified sequences

Nucleotides oligos were extracted from 5NEO and 1Q9A RCSB structure and modified in laboratory. CD spectra were obtained at SRCD SOLEIL Synchrotron. Three primary CD absorption bands were analysed to assess the circular dichroism data of modified RNA sequences:

i. **180–190 nm and 200–210 nm bands**: These regions were examined to detect conformational transitions that deviate from the canonical A-form RNA conformation.

ii. **260–268 nm band**: This range was analysed to evaluate contributions from hydrogen-bonding and base-stacking interactions, which reflect the stability and structural organisation of RNA (Le Brun et al., 2020).

7.6.9 ClaRNA

Uridine-modified derivatives (PSU and D) were analysed to identify differences in stacking interaction tendencies. In this analysis, ClaRNA was used to detect stacking interactions (Ray et al., unpublished) within a curated dataset of experimentally resolved RNA structures and their modified counterparts. ClaRNA software identifies nucleotide interactions by analysing geometric features of nucleotide pairs relative to

a centroid model (Walen et al., 2014). Initially, ClaRNA identified stacking interactions involving unmodified nucleotides. The presence of modifications was then verified through visual inspection to determine whether the stacking interactions were preserved or altered.

Table of figures

Figure 11 RCSB available structures. (**A**) RNA only structures. (**B**) DNA only structures. (**C**) Protein only structures. (**D**) Protein Nucleic acids complexes. Data are publicly available at PDB Statistics: https://www.rcsb.org/stats/

Figure 12 Descriptive Statistical Analysis of tRNA modifications present within the MODOMICS's alignments. (**A**) Percentage of the most prevalent modifications. (**B**) Absolute Numbers of the tRNA modifications.

Figure 13 Comparative Statistical Analysis of tRNA modifications present within the MODOMICS's alignments. (**A**) Eukarya cluster map. (**B**) Bacteria cluster map. (**C**) Archaea cluster map. (**D**) Venn Diagram of the post-transcriptional modification positions of Eukarya, Bacteria, and Archaea tRNA sequences. (**E**) Venn Diagram of the post-transcriptional modifications of Eukarya, Bacteria, and Archaea tRNA sequences.

Figure 14 Distribution of RNA modifications in DOMAIN-I. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S.

Figure 15 Distribution of RNA modifications in DOMAIN-II. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S.

Figure 16 Distribution of RNA modifications in DOMAIN-III. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S.

Figure 17 Distribution of RNA modifications in DOMAIN-IV. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Zoom over Peptdyl Transferase Centre: 4UG0 from RCSB has been used. (**C**) Mapping of RNA modifications to Human rRNA 28S.

Figure 18 Distribution of RNA modifications in DOMAIN-V. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S, first half. (**C**) Mapping of RNA modifications to Human rRNA 28S, second half.

Figure 19 Distribution of RNA modifications in DOMAIN-VI. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S.

Figure 20 Distribution of RNA modifications in DOMAIN-V. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S.

Figure 21 Distribution of RNA modifications in DOMAIN-5′. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S.

Figure 22 Distribution of RNA modifications in DOMAIN-C. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S.

Figure 23 Distribution of RNA modifications in DOMAIN-3′M. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S, first half. (**C**) Mapping of RNA modifications to Human rRNA 28S, second half.

Figure 24 Distribution of RNA modifications in DOMAIN-3′m. (**A**) Heatmap of RNA modification percentages Occurrence Stacked barchart. (**B**) Mapping of RNA modifications to Human rRNA 28S, first half.

Figure 25 Circular Dichroism Data Analysis of P1 stem-loop of the guanidine II riboswitch. (**A**) 185nm absorption band. (**B**) 200-210 absorption band. (**C**) 260-268 nm absorption band. (**D**) Correlation plot among absorption bands of the P1_WT RNA variant. (**E**) Correlation plot among absorption bands of the P1_N6 RNA variant. (**F**) Correlation plot among absorption bands of the P1_N1 RNA variant.

Figure 26 Circular Dichroism Data Analysis of the Sarcin Ricin Loop. (**A**) 185nm absorption band. (**B**) 200-210 absorption band. (**C**) 260-268 nm absorption band.

Figure 27 Correlation plots of the three absorption bands related to Sarcin Ricin Loop. (**A**) SRL_WT. (**B**) SRL_8N6. (**C**) SRL_11N6. (**D**) SRL_19N6. (**E**) SRL_23N6. (**F**) SRL_23N1.

Figure 28 Analysis of the stacking interactions of the uridine derivatives. (A) Bar chart. (B) 1FCW experimentally determined structure rendered with PyMol. (C) Unique stacking interactions coming from 1FCW

Figure 29 MODOMICS Alignment disambiguation page.

Figure 30 NACDDB Homepage. (**A**) Navigation Bar. (**B**) and (**C**) Search engine

Figure 31 Utilization of the search engine with a DNA keyword. (**A**) DNA only CD experiments. (**B**) Hybrid only CD experiments. (**C**) RNA only CD experiments. (**D**) List of related papers

Figure 32 BLAST query results. (**A**) BLAST Tab on the search page. (**B**) Waiting Page. (**C**) Results page. (**D**) Downloaded results

Figure 33 CD Spectra Experiment Page.

Figure 34 Experiment summary page of RNA i-motif experiment, experiment number 137. (A) Circular Dichroism plus absorbance chart. (B) Circular Dichroism plus high voltage chart. (C) Metadata table and (1) Download button, (2) Cosine Similarity (CS) Comparison, (3) RMSD comparison, (4) RMSD and CS comparison. (D) Melting curves. (E) Functions to compare melting curves and to download thermal acquisitions.

Figure 35 Utilization of mathematical comparison metrics between spectra. (**A**) RMSD results and (**B**) Cosine Similarity results, displayed in sequence. (**C**) Joint RMSD and Cosine Similarity results illustrating the comparison between spectra. (**D**) Kendall-tau analysis and K-means clustering distribution, highlighting the complementary relationship between RMSD and cosine similarity. (**E**) Extended dataset view for spectral comparison using RMSD and Cosine Similarity metrics. (**F**) CD spectrum comparison for acid_quad_pH7.2 and RNA i-motif, integrating RMSD and Cosine Similarity measures to provide a comprehensive view of alignment and dissimilarity.

Figure 36 tRNA RF00005; (B) RF02541 LSU; (C) SSU RF00177

Figure 37 3DJMol Experimentally Determined Structure Visualization Tool (**A**); Comments Section (**B**); Enzymatic Reaction Sections (**C**); Mol* Protein Visualization Tool for AlphaFold protein Visualization (D); Interactive HTML sequence tool (**E**); Refactored Calipho Feature Viewer (**F**). ADARB1 protein has been selected to render this picture

Figure 38 Nonnatural modifications in MODOMICS (**A**) Disambiguation page; Modified Nucleotide rendering (**B**); Chemical information sub table (**C**); Tautomers (**D**); CYP Metabolic Sites (**E**).

Figure 39 Examples of synthetic modified nucleotides and associated MODOMICS codes.

Figure 40 Descriptive statistics related to modified residues in MODOMICS

Table of tables

Bibliography

- A. Ferré-D'Amaré, Kaihong Zhou, J. Doudna, A general module for RNA crystallization., Journal of molecular biology, Q1.0 SJR, 1998, doi: 10.1006/JMBI.1998.1789
- A. Vaiana, Carsten Kutzner, L. Bock, C. Blau, H. Grubmüller, Refining Crystal Structures Against Cryo-EM Data using Molecular Dynamics Simulations to Obtain a Complete Atomistic Pathway of Transfer RNA Translocation, Biophysical Journal, Q1.0 SJR, 2014, doi: 10.1016/J.BPJ.2013.11.1405
- Abramson, J., Adler, J., Dunger, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024). https://doi.org/10.1038/s41586-024-07487-w
- Abu Almakarem AS, Petrov AI, Stombaugh J, Zirbel CL, Leontis NB. Comprehensive survey and geometric classification of base triples in RNA structures. Nucleic Acids Res. 2012 Feb;40(4):1407-23. doi: 10.1093/nar/gkr810. Epub 2011 Nov 3. PMID: 22053086; PMCID: PMC3287178.
- Addepalli, B., & Limbach, P. (2011). Mass Spectrometry-Based Quantification of Pseudouridine in RNA. Journal of The American Society for Mass Spectrometry, 22, 1363-1372. https://doi.org/10.1007/s13361-011-0137-5.
- Agarwalla S, Kealey JT, Santi DV, Stroud RM. Characterization of the 23 S ribosomal RNA m5U1939 methyltransferase from Escherichia coli. J Biol Chem. 2002 Mar 15;277(11):8835-40. doi: 10.1074/jbc.M111825200. Epub 2002 Jan 4. PMID: 11779873.
- Agrawal RK, Spahn CM, Penczek P, Grassucci RA, Nierhaus KH, Frank J. Visualization of tRNA movements on the Escherichia coli 70S ribosome during the elongation cycle. J Cell Biol. 2000 Aug 7;150(3):447-60. doi: 10.1083/jcb.150.3.447. PMID: 10931859; PMCID: PMC2175196.
- Aida, M. (1988). An ab initio molecular orbital study on the sequence-dependency of DNA conformation: an evaluation of intra- and inter-strand stacking interaction energy. *Journal of theoretical biology*, 130 3, 327-35 . https://doi.org/10.1016/S0022-5193(88)80032-8 .
- Alonso D, Mondragón A. Mechanisms of catalytic RNA molecules. Biochem Soc Trans. 2021 Aug 27;49(4):1529-1535. doi: 10.1042/BST20200465. PMID: 34415304; PMCID: PMC10583251.
- Anderson, J., Phan, L., Cuesta, R., Carlson, B. A., Pak, M., Asano, K., Björk, G. R., Tamame, M., & Hinnebusch, A. G. (1998). The essential Gcd10p-Gcd14p protein complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. Genes & Development, 12(23),3650–3662. https://doi.org/10.1101/gad.12.23.3650
- Anfinsen CB. Principles that govern the folding of protein chains. Science. 1973 Jul 20;181(4096):223-30. doi: 10.1126/science.181.4096.223. PMID: 4124164.
- Atalay, N., Akcan, E., Gül, M., Ayan, E., Destan, E., Ertem, F., Tokay, N., Çakilkaya, B., Nergiz, Z., Karakadioğlu, G., Kepceoğlu, A., Yapici, I., Tosun, B., Baldir, N., Yildirim, G., Johnson, J., Güven, Ö., Shafiei, A., Arslan, N., Yılmaz, M., Kulakman, C., Paydos, S., Çinal, Z., Şabanoğlu, K., Pazarçeviren, A., Yılmaz, A., Canbay, B., Aşci, B., Kartal, E., Tavli, S., Çaliseki, M., Göç, G., Mermer, A., Yeşilay, G., Altuntaş, S., Tateishi, H., Otsuka, M., Fujita, M., Tekin, Ş., Ciftci, H., Durdağı, S., Doğanay, G., Karaca, E., Türköz, B., Kabasakal, B., Katı, A., & Demirci, H. (2022). Cryogenic X-ray crystallographic studies of biomacromolecules at Turkish Light Source "Turkish DeLight". Turkish Journal of Biology, 47, 1 - 13. https://doi.org/10.55730/1300-0152.2637.
- Austin S Draycott, Cassandra Schaening-Burgos, Maria F Rojas-Duran, Loren Wilson, Leonard Schärfen, K. Neugebauer, S. Nachtergaele, Wendy V. Gilbert, Transcriptome-wide mapping reveals a diverse dihydrouridine landscape including mRNA, PLoS Biology, Q1.0 SJR, 2022, doi: 10.1371/journal.pbio.3001622
- Ayadi, L., Galvanin, A., Pichot, F., Marchand, V., & Motorin, Y. (2019). RNA ribose methylation (2'-O-methylation): Occurrence, biosynthesis and biological functions.. *Biochimica et biophysica acta. Gene regulatory mechanisms*, 1862 3, 253-269 . https://doi.org/10.1016/j.bbagrm.2018.11.009.

- Ayers TN, Woolford JL. Putting It All Together: The Roles of Ribosomal Proteins in Nucleolar Stages of 60S Ribosomal Assembly in the Yeast *Saccharomyces cerevisiae*. Biomolecules. 2024 Aug 9;14(8):975. doi: 10.3390/biom14080975. PMID: 39199362; PMCID: PMC11353139.
- Bacusmo JM, Orsini SS, Hu J, DeMott M, Thiaville PC, Elfarash A, Paulines MJ, Rojas-Benítez D, Meineke B, Deutsch C, Iwata-Reuyl D, Limbach PA, Dedon PC, Rice KC, Shuman S, Crécy-Lagard V. The t^6A modification acts as a positive determinant for the anticodon nuclease PrrC, and is distinctively nonessential in Streptococcus mutans. RNA Biol. 2018;15(4-5):508-517. doi: 10.1080/15476286.2017.1353861. Epub 2017 Sep 13. PMID: 28726545; PMCID: PMC6103680.
- Baird, N., Gong, H., Zaheer, S., Freed, K., Pan, T., & Sosnick, T. (2010). Extended structures in RNA folding intermediates are due to nonnative interactions rather than electrostatic repulsion. *Journal of molecular biology*, 397 5, 1298-306 . https://doi.org/10.1016/j.jmb.2010.02.025.
- Balduzzi E, Geinguenaud F, Sordyl D, Maiti S, Farsani MA, Nikolaev G, Arluison V, Bujnicki JM. NAIRDB: a database of Fourier transform infrared (FTIR) data for nucleic acids. Nucleic Acids Res. 2024 Oct 16:gkae885. doi: 10.1093/nar/gkae885. Epub ahead of print. PMID: 39413200.
- Barnwal, R., Yang, F., & Varani, G. (2017). Applications of NMR to structure determination of RNAs large and small. *Archives of biochemistry and biophysics*, 628, 42-56. https://doi.org/10.1016/j.abb.2017.06.003.
- Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M. The structure of the eukaryotic ribosome at 3.0 Å resolution. Science. 2011 Dec 16;334(6062):1524-9. doi: 10.1126/science.1212642. Epub 2011 Nov 17. PMID: 22096102.
- Berg MD, Brandl CJ. Transfer RNAs: diversity in form and function. RNA Biol. 2021 Mar;18(3):316-339. doi: 10.1080/15476286.2020.1809197. Epub 2020 Sep 9. PMID: 32900285; PMCID: PMC7954030.
- Bo Yang, A. R. Moehlig, C. E. Frieler, M. Rodgers, Base-pairing energies of protonated nucleobase pairs and proton affinities of 1-methylated cytosines: model systems for the effects of the sugar moiety on the stability of DNA i-motif conformations., The journal of physical chemistry. B, Q1.0 SJR, 2015, doi: 10.1021/acs.jpcb.5b00035
- Boccaletto P, Machnicka MA, Purta E, Piatkowski P, Baginski B, Wirecki TK, de Crécy-Lagard V, Ross R, Limbach PA, Kotter A, Helm M, Bujnicki JM. MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Res. 2018 Jan 4;46(D1):D303-D307. doi: 10.1093/nar/gkx1030. PMID: 29106616; PMCID: PMC5753262.
- Boccaletto P, Stefaniak F, Ray A, Cappannini A, Mukherjee S, Purta E, Kurkowska M, Shirvanizadeh N, Destefanis E, Groza P, Avşar G, Romitelli A, Pir P, Dassi E, Conticello SG, Aguilo F, Bujnicki JM. MODOMICS: a database of RNA modification pathways. 2021 update. Nucleic Acids Res. 2022 Jan 7;50(D1):D231-D235. doi: 10.1093/nar/gkab1083. PMID: 34893873; PMCID: PMC8728126.
- Boileau E, Wilhelmi H, Busch A, Cappannini A, Hildebrand A, Bujnicki JM, Dieterich C. Sci-ModoM: a quantitative database of transcriptome-wide high-throughput RNA modification sites. Nucleic Acids Res. 2024 Nov 5:gkae972. doi: 10.1093/nar/gkae972. Epub ahead of print. PMID: 39498498.
- Bokinsky, G., et al. (2003). Single-molecule FRET studies highlight conformational heterogeneity in the equilibrium folding of the P4-P6 domain of the Tetrahymena ribozyme.
- Boniecki MJ, Lach G, Dawson WK, Tomala K, Lukasz P, Soltysinski T, Rother KM, Bujnicki JM. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. Nucleic Acids Res. 2016 Apr 20;44(7):e63. doi: 10.1093/nar/gkv1479. Epub 2015 Dec 19. PMID: 26687716; PMCID: PMC4838351.
- Brown JA. Unraveling the structure and biological functions of RNA triple helices. Wiley Interdiscip Rev RNA. 2020 Nov;11(6):e1598. doi: 10.1002/wrna.1598. Epub 2020 May 22. PMID: 32441456; PMCID: PMC7583470.

- Brown RF, Andrews CT, Elcock AH. Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised AMBER Parameters and Compared with Experiment. J Chem Theory Comput. 2015 May 12;11(5):2315-28. doi: 10.1021/ct501170h. Epub 2015 Apr 7. Erratum in: J Chem Theory Comput. 2018 Dec 11;14(12):6742-6743. doi: 10.1021/acs.jctc.8b00594. PMID: 26574427; PMCID: PMC4651843.
- Buchhaupt M, Sharma S, Kellner S, Oswald S, Paetzold M, Peifer C, Watzinger P, Schrader J, Helm M, Entian KD. Partial methylation at Am100 in 18S rRNA of baker's yeast reveals ribosome heterogeneity on the level of eukaryotic rRNA modification. PLoS One. 2014 Feb 28;9(2):e89640. doi: 10.1371/journal.pone.0089640. PMID: 24586927; PMCID: PMC3938493.
- Burakovsky, D., Prokhorova, I., Sergiev, P., Milón, P., Sergeeva, O., Bogdanov, A.,et al. (2012). Impact of methylations of m2G966/m5C967 in 16S rRNA on bacterial fitness and translation initiation. Nucleic Acids Res. 40, 7885–7895. doi: 10.1093/nar/gks508
- C. He, L. Hamlow, Zachary J. Devereaux, Yanlong Zhu, Y.-w. Nei, Lin Fan, Christopher P. McNary, Philippe Maître, Vincent Steinmetz, B. Schindler, Isabelle Compagnon, P. Armentrout, M. Rodgers, Structural and Energetic Effects of O2'-Ribose Methylation of Protonated Purine Nucleosides., The journal of physical chemistry. B, Q1.0 SJR, 2018, doi: 10.1021/acs.jpcb.8b07687
- Caldas T, Binet E, Bouloc P, Costa A, Desgres J, Richarme G. The FtsJ/RrmJ heat shock protein of Escherichia coli is a 23 S ribosomal RNA methyltransferase. J Biol Chem. 2000 Jun 2;275(22):16414-9. doi: 10.1074/jbc.M001854200. PMID: 10748051.
- Campos JHC, Maricato JT, Braconi CT, Antoneli F, Janini LMR, Briones MRS. Direct RNA Sequencing Reveals SARS-CoV-2 m6A Sites and Possible Differential DRACH Motif Methylation among Variants. Viruses. 2021 Oct 20;13(11):2108. doi: 10.3390/v13112108. PMID: 34834915; PMCID: PMC8620083.
- Cao, S., & Chen, S. (2011). Physics-based de novo prediction of RNA 3D structures.. The journal of physical chemistry. B, 115 14, 4216-26 . https://doi.org/10.1021/jp112059y.
- Cappannini A, Mosca K, Mukherjee S, Moafinejad SN, Sinden RR, Arluison V, Bujnicki J, Wien F. NACDDB: Nucleic Acid Circular Dichroism Database. Nucleic Acids Res. 2023 Jan 6;51(D1):D226-D231. doi: 10.1093/nar/gkac829. PMID: 36280237; PMCID: PMC9825466.
- Cappannini A, Ray A, Purta E, Mukherjee S, Boccaletto P, Moafinejad SN, Lechner A, Barchet C, Klaholz BP, Stefaniak F, Bujnicki JM. MODOMICS: a database of RNA modifications and related information. 2023 update. Nucleic Acids Res. 2024 Jan 5;52(D1):D239-D244. doi: 10.1093/nar/gkad1083. PMID: 38015436; PMCID: PMC10767930.
- Carter-Fenk K , Herbert JM . Reinterpreting π-stacking. Phys Chem Chem Phys. 2020 Nov 21;22(43):24870-24886. doi: 10.1039/d0cp05039c. Epub 2020 Oct 27. PMID: 33107520.
- Case, D. A., Aktulga, H. M., Belfon, K. A. A., Ben-Shalom, I. Y., Berryman, J. T., Brozell, S. R., Cerutti, D. S., Cheatham III, T. E., Cisneros, G. A., Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Giambaşu, G. M., Gilson, M. K., Gohlke, H., Götz, A. W., Harris, R. C., Izadi, S., Izmailov, S. A., ... Kollman, P. A. (2022). *Amber 2022 Reference Manual: (Covers Amber22 and AmberTools22)*. University of California, San Francisco. https://doi.org/10.13140/RG.2.2.31337.77924
- Chen S, Meng J, Zhang Y. Quantitative profiling N1-methyladenosine (m1A) RNA methylation from Oxford nanopore direct RNA sequencing data. Methods. 2024 Aug;228:30-37. doi: 10.1016/j.ymeth.2024.05.009. Epub 2024 May 18. PMID: 38768930.
- Chen SJ, Dill KA. RNA folding energy landscapes. Proc Natl Acad Sci U S A. 2000 Jan 18;97(2):646-51. doi: 10.1073/pnas.97.2.646. PMID: 10639133; PMCID: PMC15384.
- Chen, S.-J., & Dill, K. A. (2000). RNA folding energy landscapes. Proceedings of the National Academy of Sciences, 97(2), 646-651. PubMed.

- Chen, Y., Yang, S., Peng, S., Li, W., Wu, F., Yao, Q., Wang, F., Weng, X., & Zhou, X. (2019). N1-Methyladenosine detection with CRISPR-Cas13a/C2c2. Chemical Science, 10, 2975 - 2979. https://doi.org/10.1039/c8sc03408g.
- Cheng-Kai Huang, S. Kafert-Kasting, T. Thum, Preclinical and Clinical Development of Noncoding RNA Therapeutics for Cardiovascular Disease, Circulation Research, Q1.0 SJR, 2020, doi: 10.1161/CIRCRESAHA.119.315856
- Chouhan, B. P. S., Maimaiti, S., Gade, M., & Laurino, P. (2018). Rossmann-Fold Methyltransferases: Taking a "β-Turn" around Their Cofactor, S-Adenosylmethionine. Biochemistry, 58(3), 330–335. https://doi.org/10.1021/acs.biochem.8b00994
- Christian Toft Madsen, Jonas Mengel-Jørgensen, Finn Kirpekar, Stephen Douthwaite, Identifying the methyltransferases for m^5U747 and m^5U1939 in 23S rRNA using MALDI mass spectrometry, *Nucleic Acids Research*, Volume 31, Issue 16, 15 August 2003, Pages 4738–4746, https://doi.org/10.1093/nar/gkg657
- Condon DE, Kennedy SD, Mort BC, Kierzek R, Yildirim I, Turner DH. Stacking in RNA: NMR of Four Tetramers Benchmark Molecular Dynamics. J Chem Theory Comput. 2015 Jun 9;11(6):2729-2742. doi: 10.1021/ct501025q. Epub 2015 Apr 16. PMID: 26082675; PMCID: PMC4463549.
- Contreras-Martinez LM, Boock JT, Kostecki JS, DeLisa MP. The ribosomal exit tunnel as a target for optimizing protein expression in Escherichia coli. Biotechnol J. 2012 Mar;7(3):354-60. doi: 10.1002/biot.201100198. PMID: 22076828; PMCID: PMC3382190.
- Cottilli P, Itoh Y, Nobe Y, Petrov AS, Lisón P, Taoka M, Amunts A. Cryo-EM structure and rRNA modification sites of a plant ribosome. Plant Commun. 2022 Sep 12;3(5):100342. doi: 10.1016/j.xplc.2022.100342. Epub 2022 May 27. PMID: 35643637; PMCID: PMC9483110.
- Coureux PD, Lazennec-Schurdevin C, Bourcier S, Mechulam Y, Schmitt E. Cryo-EM study of an archaeal 30S initiation complex gives insights into evolution of translation initiation. Commun Biol. 2020 Feb 6;3(1):58. doi: 10.1038/s42003-020-0780-0. PMID: 32029867; PMCID: PMC7005279.
- Culver GM.Kirthi N.2008.Assembly of the 30S Ribosomal Subunit. 3:10.1128/ecosalplus.2.5.3.https://doi.org/10.1128/ecosalplus.2.5.3
- Czerwoniec A, Dunin-Horkawicz S, Purta E, Kaminska KH, Kasprzak JM, Bujnicki JM, Grosjean H, Rother K. MODOMICS: a database of RNA modification pathways. 2008 update. Nucleic Acids Res. 2009 Jan;37(Database issue):D118-21. doi: 10.1093/nar/gkn710. Epub 2008 Oct 14. PMID: 18854352; PMCID: PMC2686465.
- D. Davis, Stabilization of RNA stacking by pseudouridine., Nucleic acids research, Q1.0 SJR, 1995, doi: 10.1093/NAR/23.24.5020
- D. Laurents, AlphaFold 2 and NMR Spectroscopy: Partners to Understand Protein Structure, Dynamics and Function, Frontiers in Molecular Biosciences, Q1.0 SJR, 2022, doi: 10.3389/fmolb.2022.906437
- Dawson WK, Maciejczyk M, Jankowska EJ, Bujnicki JM. Coarse-grained modeling of RNA 3D structure. Methods. 2016 Jul 1;103:138-56. doi: 10.1016/j.ymeth.2016.04.026. Epub 2016 Apr 25. PMID: 27125734.
- Deng X, Chen K, Luo GZ, Weng X, Ji Q, Zhou T, He C. Widespread occurrence of N6-methyladenosine in bacterial mRNA. Nucleic Acids Res. 2015 Jul 27;43(13):6557-67. doi: 10.1093/nar/gkv596. Epub 2015 Jun 11. PMID: 26068471; PMCID: PMC4513869.
- Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, R. M., Fratini, A. V., & Kopka, M. L. (1982). The anatomy of A-, B-, and Z-DNA. *Science*, 216(4545), 475–485.
- Dijk, L., Bobbert, P., & Spano, F. (2010). Extreme sensitivity of circular dichroism to long-range excitonic couplings in helical supramolecular assemblies.. The journal of physical chemistry. B, 114 2, 817-25 . https://doi.org/10.1021/jp911081b.

- Ding, J., Ma, C., Chen, M., Chen, B., Yuan, B., & Feng, Y. (2020). Quantification and single-base resolution analysis of N1-methyladenosine in mRNA by ligation-assisted differentiation.. Analytical chemistry. https://doi.org/10.1021/acs.analchem.9b04454.
- Djumagulov M, Demeshkina N, Jenner L, Rozov A, Yusupov M, Yusupova G. Accuracy mechanism of eukaryotic ribosome translocation. Nature. 2021 Dec;600(7889):543-546. doi: 10.1038/s41586-021-04131-9. Epub 2021 Dec 1. PMID: 34853469; PMCID: PMC8674143.
- Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, Sorek R, Rechavi G. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature. 2012 Apr 29;485(7397):201-6. doi: 10.1038/nature11112. PMID: 22575960.
- Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, Peer E, Kol N, Ben-Haim MS, Dai Q, Di Segni A, Salmon-Divon M, Clark WC, Zheng G, Pan T, Solomon O, Eyal E, Hershkovitz V, Han D, Doré LC, Amariglio N, Rechavi G, He C. The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. Nature. 2016 Feb 25;530(7591):441-6. doi: 10.1038/nature16998. Epub 2016 Feb 10. PMID: 26863196; PMCID: PMC4842015.
- Dunin-Horkawicz S, Czerwoniec A, Gajda MJ, Feder M, Grosjean H, Bujnicki JM. MODOMICS: a database of RNA modification pathways. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D145-9. doi: 10.1093/nar/gkj084. PMID: 16381833; PMCID: PMC1347447.
- Durairaj, A., & Limbach, P. (2008). Improving CMC-derivatization of pseudouridine in RNA for mass spectrometric detection. Analytica chimica acta, 612 2, 173-81. https://doi.org/10.1016/j.aca.2008.02.026.
- Dutca LM, Culver GM. Assembly of the 5' and 3' minor domains of 16S ribosomal RNA as monitored by tethered probing from ribosomal protein S20. J Mol Biol. 2008 Feb 8;376(1):92-108. doi: 10.1016/j.jmb.2007.10.083. Epub 2007 Nov 6. PMID: 18155048; PMCID: PMC2443948.
- E. Kierzek, Magdalena Małgowska, Jolanta Lisowiec, D. Turner, Z. Gdaniec, R. Kierzek, The contribution of pseudouridine to stabilities and structure of RNAs, Nucleic Acids Research, Q1.0 SJR, 2013, doi: 10.1093/nar/gkt1330
- E. Kierzek, Xiaoju Zhang, R. M. Watson, S. Kennedy, M. Szabat, R. Kierzek, D. Mathews, Secondary structure prediction for RNA sequences including N6-methyladenosine, Nature Communications, Q1.0 SJR, 2021, doi: 10.1038/s41467-022-28817-4
- E. Nilsson, R. Alexander, Bacterial wobble modifications of NNA-decoding tRNAs, IUBMB Life, Q2.0 SJR, 2019, doi: 10.1002/iub.2120
- Egli, M. (2010). Sugar Pucker and Nucleic Acid Structure. *International Journal of Molecular Sciences*, *11*(3), 309–334.
- Eliana Destefanis, Denise Sighel, Davide Dalfovo, Riccardo Gilmozzi, Francesca Broso, Andrea Cappannini, Janusz M Bujnicki, Alessandro Romanel, Erik Dassi, Alessandro Quattrone, The three YTHDF paralogs and VIRMA are strong cross-histotype tumor driver candidates among m6A core genes, NAR Cancer, Volume 6, Issue 4, December 2024, zcae040, https://doi.org/10.1093/narcan/zcae040
- Emily M. Harcourt, A. Kietrys, E. Kool, Chemical and structural effects of base modifications in messenger RNA, Nature, Q1.0 SJR, 2017, doi: 10.1038/nature21351
- Englander SW, Mayne L. The case for defined protein folding pathways. Proc Natl Acad Sci U S A. 2017 Aug 1;114(31):8253-8258. doi: 10.1073/pnas.1706196114. Epub 2017 Jun 19. PMID: 28630329; PMCID: PMC5547639.
- Ennifar, E. (2013). X-ray crystallography as a tool for mechanism-of-action studies and drug discovery.. Current pharmaceutical biotechnology, 14 5, 537-50. https://doi.org/10.2174/13892010140513111104824.

- Ensinck I, Sideri T, Modic M, Capitanchik C, Vivori C, Toolan-Kerr P, van Werven FJ. m6A-ELISA, a simple method for quantifying N6-methyladenosine from mRNA populations. RNA. 2023 May;29(5):705-712. doi: 10.1261/rna.079554.122. Epub 2023 Feb 9. PMID: 36759126; PMCID: PMC10159001.
- Eyler DE, Franco MK, Batool Z, Wu MZ, Dubuke ML, Dobosz-Bartoszek M, Jones JD, Polikanov YS, Roy B, Koutmou KS. Pseudouridinylation of mRNA coding sequences alters translation. Proc Natl Acad Sci U S A. 2019 Nov 12;116(46):23068-23074. doi: 10.1073/pnas.1821754116. Epub 2019 Oct 31. PMID: 31672910; PMCID: PMC6859337.
- Feric M, Vaidya N, Harmon TS, Mitrea DM, Zhu L, Richardson TM, Kriwacki RW, Pappu RV, Brangwynne CP. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. Cell. 2016 Jun 16;165(7):1686-1697. doi: 10.1016/j.cell.2016.04.047. Epub 2016 May 19. PMID: 27212236; PMCID: PMC5127388.
- Finet O, Yague-Sanz C, Marchand F, Hermand D. The Dihydrouridine landscape from tRNA to mRNA: a perspective on synthesis, structural impact and function. RNA Biol. 2022 Jan;19(1):735-750. doi: 10.1080/15476286.2022.2078094. PMID: 35638108; PMCID: PMC9176250.
- Fingerhut BP, Schauss J, Kundu A, Elsaesser T. Contact pairs of RNA with magnesium ions-electrostatics beyond the Poisson-Boltzmann equation. Biophys J. 2021 Dec 7;120(23):5322-5332. doi: 10.1016/j.bpj.2021.10.029. Epub 2021 Oct 27. PMID: 34715079; PMCID: PMC8715182.
- Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. Science. 1991 Dec 13;254(5038):1598-603. doi: 10.1126/science.1749933. PMID: 1749933.
- Friedman RA, Honig B. A free energy analysis of nucleic acid base stacking in aqueous solution. Biophys J. 1995 Oct;69(4):1528-35. doi: 10.1016/S0006-3495(95)80023-8. PMID: 8534823; PMCID: PMC1236383.
- G. M. Daubner, A. Cléry, F. Allain, RRM-RNA recognition: NMR or crystallography…and new findings., Current opinion in structural biology, Q1.0 SJR, 2013, doi: 10.1016/j.sbi.2012.11.006
- Garcia-Campos MA, Edelheit S, Toth U, Safra M, Shachar R, Viukov S, Winkler R, Nir R, Lasman L, Brandis A, Hanna JH, Rossmanith W, Schwartz S. Deciphering the "m^{6}A Code" via Antibody-Independent Quantitative Profiling. Cell. 2019 Jul 25;178(3):731-747.e16. doi: 10.1016/j.cell.2019.06.013. Epub 2019 Jun 27. PMID: 31257032.
- Gebala, M., & Herschlag, D. (2019). Quantitative studies of an RNA duplex electrostatics by ion counting. *bioRxiv*. https://doi.org/10.1101/645697.
- Goto Y, Nonaka I, Horai S. A mutation in the tRNA(Leu)(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. Nature. 1990 Dec 13;348(6302):651-3. doi: 10.1038/348651a0. PMID: 2102678.
- Gourav Das, S. Harikrishna, Kiran R. Gore, Influence of Sugar Modifications on the Nucleoside Conformation and Oligonucleotide Stability: A Critical Review, The Chemical Record, Q1.0 SJR, 2022, doi: 10.1002/tcr.202200174
- Grasby JA, Gait MJ. Synthetic oligoribonucleotides carrying site-specific modifications for RNA structure-function analysis. Biochimie. 1994;76(12):1223-34. doi: 10.1016/0300-9084(94)90053-1. PMID: 7538326.
- Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. Nature Protocols, 1(6), 2876–2890. https://doi.org/10.1038/nprot.2006.202
- Grela P, Szajwaj M, Horbowicz-Drożdżal P, Tchórzewski M. How Ricin Damages the Ribosome. Toxins (Basel). 2019 Apr 27;11(5):241. doi: 10.3390/toxins11050241. PMID: 31035546; PMCID: PMC6562825.
- Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell. 1983 Dec;35(3 Pt 2):849-57. doi: 10.1016/0092-8674(83)90117-4. PMID: 6197186.

- Guo P, Zhang L, Zhang H, Feng Y, Jing G. Domain II plays a crucial role in the function of ribosome recycling factor. Biochem J. 2006 Feb 1;393(Pt 3):767-77. doi: 10.1042/BJ20050780. PMID: 16262604; PMCID: PMC1360730.
- Gustafsson C, Persson BC. Identification of the rrmA gene encoding the 23S rRNA m1G745 methyltransferase in Escherichia coli and characterization of an m1G745-deficient mutant. J Bacteriol. 1998 Jan;180(2):359-65. doi: 10.1128/JB.180.2.359-365.1998. PMID: 9440525; PMCID: PMC106891.
- Guymon R, Pomerantz SC, Ison JN, Crain PF, McCloskey JA. Post-transcriptional modifications in the small subunit ribosomal RNA from Thermotoga maritima, including presence of a novel modified cytidine. RNA. 2007 Mar;13(3):396-403. doi: 10.1261/rna.361607. Epub 2007 Jan 25. PMID: 17255199; PMCID: PMC1800508.
- H. Ebhardt, Herbert H. Tsang, D. C. Dai, Yifeng Liu, Babak Bostan, R. Fahlman, Meta-analysis of small RNA-sequencing errors reveals ubiquitous post-transcriptional RNA modifications, Nucleic Acids Research, Q1.0 SJR, 2009, doi: 10.1093/nar/gkp093
- Halder, A., Data, D., Seelam, P. P., Bhattacharyya, D., & Mitra, A. (2019). Estimating strengths of individual hydrogen bonds in RNA base pairs: Toward a consensus between different computational approaches. *ACS Omega, 4*(5), 7354–7368. https://doi.org/10.1021/acsomega.8b03689
- He, J., Liu, X., & Wang, S. (2024). Advances in solid-state NMR methods for studying RNA structures and dynamics. *Magnetic Resonance Letters*. https://doi.org/10.1016/j.mrl.2024.200133.
- Helm M, Kobitski AY, Nienhaus GU. Single-molecule Förster resonance energy transfer studies of RNA structure, dynamics and function. Biophys Rev. 2009 Dec;1(4):161. doi: 10.1007/s12551-009-0018-3. Epub 2009 Nov 10. PMID: 28510027; PMCID: PMC5418384.
- Helm M, Motorin Y. Detecting RNA modifications in the epitranscriptome: predict and validate. Nat Rev Genet. 2017 May;18(5):275-291. doi: 10.1038/nrg.2016.169. Epub 2017 Feb 20. PMID: 28216634.
- Hendra, C., Pratanwanich, P. N., Wan, Y. K., Goh, W. S., Thiery, A., & Göke, J. (2022). *Detection of m6A from direct RNA sequencing using a multiple instance learning framework. Nature Methods, 19*, 1590–1598.
- Herschlag D. RNA chaperones and the RNA folding problem. J Biol Chem. 1995 Sep 8;270(36):20871-4. doi: 10.1074/jbc.270.36.20871. PMID: 7545662.
- Hershkovitz E, Sapiro G, Tannenbaum A, Williams LD. Statistical analysis of RNA backbone. IEEE/ACM Trans Comput Biol Bioinform. 2006 Jan-Mar;3(1):33-46. doi: 10.1109/TCBB.2006.13. PMID: 17048391; PMCID: PMC2811324.
- Holm AI, Nielsen LM, Hoffmann SV, Nielsen SB. Vacuum-ultraviolet circular dichroism spectroscopy of DNA: a valuable tool to elucidate topology and electronic coupling in DNA. Phys Chem Chem Phys. 2010 Sep 7;12(33):9581-96. doi: 10.1039/c003446k. Epub 2010 Jul 7. PMID: 20607185.
- Hrabeta-Robinson, E., Marcus, E., Cozen, A., Phizicky, E., & Lowe, T. (2017). High-Throughput Small RNA Sequencing Enhanced by AlkB-Facilitated RNA de-Methylation (ARM-Seq). Methods in molecular biology, 1562, 231-243 https://doi.org/10.1007/978-1-4939-6807-7_15.
- Huang BY, Fernández IS. Long-range interdomain communications in eIF5B regulate GTP hydrolysis and translation initiation. Proc Natl Acad Sci U S A. 2020 Jan 21;117(3):1429-1437. doi: 10.1073/pnas.1916436117. Epub 2020 Jan 3. PMID: 31900355; PMCID: PMC6983393.
- Huang CH, Baserga R. Circular dichroism studies of ethidium bromide binding to the isolated nucleolus. Nucleic Acids Res. 1976 Aug;3(8):1857-73. doi: 10.1093/nar/3.8.1857. PMID: 967680; PMCID: PMC343045.
- Huang S, Mahanta N, Begley TP, Ealick SE. Pseudouridine monophosphate glycosidase: a new glycosidase mechanism. Biochemistry. 2012 Nov 13;51(45):9245-55. doi: 10.1021/bi3006829. Epub 2012 Oct 30. PMID: 23066817; PMCID: PMC3526674.

- I. Deb, J. Sarzynska, L. Nilsson, A. Lahiri, Rapid communication capturing the destabilizing effect of dihydrouridine through molecular simulations., Biopolymers, Q2.0 SJR, 2014, doi: 10.1002/bip.22495
- Incarnato D, Anselmi F, Morandi E, Neri F, Maldotti M, Rapelli S, Parlato C, Basile G, Oliviero S. High-throughput single-base resolution mapping of RNA 2′-O-methylated residues. Nucleic Acids Res. 2017 Feb 17;45(3):1433-1441. doi: 10.1093/nar/gkw810. PMID: 28180324; PMCID: PMC5388417.
- Ishigaki S, Sobue G. Importance of Functional Loss of FUS in FTLD/ALS. Front Mol Biosci. 2018 May 3;5:44. doi: 10.3389/fmolb.2018.00044. PMID: 29774215; PMCID: PMC5943504.
- Izadi, F.; Szczyrba, A.; Datta, M.; Ciupak, O.; Demkowicz, S.; Rak, J.; Denifl, S. Electron-Induced Decomposition of 5-Bromo-4-thiouracil and 5-Bromo-4-thio-2′-deoxyuridine: The Effect of the Deoxyribose Moiety on Dissociative Electron Attachment. Int. J. Mol. Sci. 2023, 24, 8706. https://doi.org/10.3390/ijms24108706
- J. Dalluge, T. Hashizume, A. Sopchik, J. McCloskey, D. Davis, Conformational flexibility in RNA: the role of dihydrouridine., Nucleic acids research, Q1.0 SJR, 1996, doi: 10.1093/NAR/24.6.1073
- J. H. Riazance, W. Baase, W. Johnson, K. Hall, P. Cruz, I. Tinoco, Evidence for Z-form RNA by vacuum UV circular dichroism, Nucleic acids research, Q1.0 SJR, 1985, doi: 10.1093/NAR/13.13.4983
- J. Jumper, Richard Evans, A. Pritzel, Tim Green, Michael Figurnov, O. Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A A Kohl, Andy Ballard, A. Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, J. Adler, T. Back, Stig Petersen, D. Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, S. Bodenstein, David Silver, O. Vinyals, A. Senior, K. Kavukcuoglu, Pushmeet Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature, Q1.0 SJR, 2021, doi: 10.1038/s41586-021-03819-2
- Jennifer Voegele, E. Duchardt-Ferner, H. Kruse, Zhengyue Zhang, J. Šponer, Miroslav Krepl, J. Wohnert, Structural and dynamic effects of pseudouridine modifications on non-canonical interactions in RNA., RNA, Q1.0 SJR, 2023, doi: 10.1261/rna.079506.122
- Jiajia Xuan, Lifan Chen, Zhirong Chen, Junjie Pang, Junhong Huang, Jinran Lin, Lingling Zheng, Bin Li, Lianghu Qu, Jianhua Yang, RMBase v3.0: decode the landscape, mechanisms and functions of RNA modifications, *Nucleic Acids Research*, Volume 52, Issue D1, 5 January 2024, Pages D273–D284, https://doi.org/10.1093/nar/gkad1070
- Jian-Feng Xiang, Qin Yang, Chu-Xiao Liu, Man Wu, Ling-Ling Chen, Li Yang, N6-Methyladenosines Modulate A-to-I RNA Editing., Molecular cell, Q1.0 SJR, 2018, doi: 10.1016/j.molcel.2017.12.006
- Jora M, Lobue PA, Ross RL, Williams B, Addepalli B. Detection of ribonucleoside modifications by liquid chromatography coupled with mass spectrometry. Biochim Biophys Acta Gene Regul Mech. 2019 Mar;1862(3):280-290. doi: 10.1016/j.bbagrm.2018.10.012. Epub 2018 Nov 7. PMID: 30414470; PMCID: PMC6401287.
- Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983 Dec;22(12):2577-637. doi: 10.1002/bip.360221211. PMID: 6667333.
- Kelly, S. M., Jess, T. J., & Price, N. C. (2005). How to study proteins by circular dichroism. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1751(2), 119-139. https://doi.org/10.1016/j.bbapap.2005.06.005
- Kerpedjiev P, Hammer S, Hofacker IL. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. Bioinformatics. 2015 Oct 15;31(20):3377-9. doi: 10.1093/bioinformatics/btv372. Epub 2015 Jun 22. PMID: 26099263; PMCID: PMC4595900.
- Khoddami, V., Yerra, A., Mosbruger, T., Fleming, A., Burrows, C., & Cairns, B. (2019). Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution.

Proceedings of the National Academy of Sciences of the United States of America, 116, 6784 - 6789. https://doi.org/10.1073/pnas.1817334116.

- Khoshnevis S, Dreggors-Walker RE, Marchand V, Motorin Y, Ghalei H. Ribosomal RNA 2'-O-methylations regulate translation by impacting ribosome dynamics. Proc Natl Acad Sci U S A. 2022 Mar 22;119(12):e2117334119. doi: 10.1073/pnas.2117334119. Epub 2022 Mar 16. PMID: 35294285; PMCID: PMC8944910.
- Kimura S, Suzuki T. Fine-tuning of the ribosomal decoding center by conserved methyl-modifications in the Escherichia coli 16S rRNA. Nucleic Acids Res. 2010 Mar;38(4):1341-52. doi: 10.1093/nar/gkp1073. Epub 2009 Dec 3. PMID: 19965768; PMCID: PMC2831307.
- Kimura, S., Ikeuchi, Y., Kitahara, K., Sakaguchi, Y., Suzuki, T., & Suzuki, T. (2011). Base methylations in the double-stranded RNA by a fused methyltransferase bearing unwinding activity. *Nucleic Acids Research*, 40, 4071 - 4085. https://doi.org/10.1093/nar/gkr1287.
- King TH, Liu B, McCully RR, Fournier MJ. Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. Mol Cell. 2003 Feb;11(2):425-35. doi: 10.1016/s1097-2765(03)00040-6. PMID: 12620230.
- Kloczewiak M, Banks JM, Jin L, Brader ML. A Biopharmaceutical Perspective on Higher-Order Structure and Thermal Stability of mRNA Vaccines. Mol Pharm. 2022 Jul 4;19(7):2022-2031. doi: 10.1021/acs.molpharmaceut.2c00092. Epub 2022 Jun 17. PMID: 35715255; PMCID: PMC9257798.
- Koodli RV, Keep B, Coppess KR, Portela F; Eterna participants; Das R. EternaBrain: Automated RNA design through move sets and strategies from an Internet-scale RNA videogame. PLoS Comput Biol. 2019 Jun 27;15(6):e1007059. doi: 10.1371/journal.pcbi.1007059. PMID: 31247029; PMCID: PMC6597038.
- Korbinian Liebl, M. Zacharias, How methyl–sugar interactions determine DNA structure and flexibility, Nucleic Acids Research, Q1.0 SJR, 2018, doi: 10.1093/nar/gky1237
- Kovaka S, Hook PW, Jenike KM, Shivakumar V, Morina LB, Razaghi R, Timp W, Schatz MC. Uncalled4 improves nanopore DNA and RNA modification detection via fast and accurate signal alignment. bioRxiv [Preprint]. 2024 Mar 10:2024.03.05.583511. doi: 10.1101/2024.03.05.583511. Update in: Nat Methods. 2025 Apr;22(4):681-691. doi: 10.1038/s41592-025-02631-4. PMID: 38496646; PMCID: PMC10942365.
- Krogh N, Asmar F, Côme C, Munch-Petersen HF, Grønbæk K, Nielsen H. Profiling of ribose methylations in ribosomal RNA from diffuse large B-cell lymphoma patients for evaluation of ribosomes as drug targets. NAR Cancer. 2020 Dec 22;2(4):zcaa035. doi: 10.1093/narcan/zcaa035. PMID: 34316692; PMCID: PMC8210301.
- Lafontaine, D.L.J. (2019). Birth of nucleolar compartments: Phase separation-driven ribosomal RNA sorting and processing. *Molecular Cell*, 76(5), 767–783. https://doi.org/10.1016/j.molcel.2019.11.015
- Langeberg CJ, Kieft JS. A generalizable scaffold-based approach for structure determination of RNAs by cryo-EM. Nucleic Acids Res. 2023 Nov 10;51(20):e100. doi: 10.1093/nar/gkad784. PMID: 37791881; PMCID: PMC10639074.
- Larsen LH, Rasmussen A, Giessing AM, Jogl G, Kirpekar F. Identification and characterization of the Thermus thermophilus 5-methylcytidine (m5C) methyltransferase modifying 23 S ribosomal RNA (rRNA) base C1942. J Biol Chem. 2012 Aug 10;287(33):27593-600. doi: 10.1074/jbc.M112.376160. Epub 2012 Jun 18. PMID: 22711535; PMCID: PMC3431712.
- Le Brun, E., Arluison, V., Wien, F. (2020). Application of Synchrotron Radiation Circular Dichroism for RNA Structural Analysis. In: Arluison, V., Wien, F. (eds) RNA Spectroscopy. Methods in Molecular Biology, vol 2113. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-0278-2_11
- Lei, Z., & Yi, C. (2017). A Radiolabeling-Free, qPCR-Based Method for Locus-Specific Pseudouridine Detection.. Angewandte Chemie, 56 47, 14878-14882 . https://doi.org/10.1002/anie.201708276.

- Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. RNA. 2001 Apr;7(4):499-512. doi: 10.1017/s1355838201002515. PMID: 11345429; PMCID: PMC1370104.
- **Leroy, E.**, Challal, D., Pelletier, S., Goncalves, C., Menant, A., Marchand, V., Jaszczyszyn, Y., van Dijk, E., Naquin, D., Andreani, J., Motorin, Y., Palancade, B., & Rougemaille, M. (2025). *A bifunctional snoRNA with separable activities in guiding rRNA 2′-O-methylation and scaffolding gametogenesis effectors*. **Nature Communications**, 16, Article 3250. https://doi.org/10.1038/s41467-025-58664-y
- Lescoute A, Westhof E. The interaction networks of structured RNAs. Nucleic Acids Res. 2006;34(22):6587-604. doi: 10.1093/nar/gkl963. Epub 2006 Nov 28. PMID: 17135184; PMCID: PMC1747187.
- Leslie, E., Grover, N. (2022). RNA: Composition and Base Pairing. In: Grover, N. (eds) Fundamentals of RNA Structure and Function. Learning Materials in Biosciences. Springer, Cham. https://doi.org/10.1007/978-3-030-90214-8_1
- Leviev I, Levieva S, Garrett RA. Role for the highly conserved region of domain IV of 23S-like rRNA in subunit-subunit interactions at the peptidyl transferase centre. Nucleic Acids Res. 1995 May 11;23(9):1512-7. doi: 10.1093/nar/23.9.1512. PMID: 7784204; PMCID: PMC306890.
- Li X, Xiong X, Zhang M, Wang K, Chen Y, Zhou J, Mao Y, Lv J, Yi D, Chen XW, Wang C, Qian SB, Yi C. Base-Resolution Mapping Reveals Distinct m^1A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. Mol Cell. 2017 Dec 7;68(5):993-1005.e9. doi: 10.1016/j.molcel.2017.10.019. Epub 2017 Nov 5. PMID: 29107537; PMCID: PMC5722686.
- Li, W., Wang, F., Chen, Y., Weng, X., & Zhou, X. (2019). A sensitive and radiolabeling-free method for pseudouridine detection.. Analytical biochemistry, 113350. https://doi.org/10.1016/j.ab.2019.113350.
- Lilley D. M. and Eckstein F., in Ribozymes and RNA Catalysis, ed. D. M. J. Lilley and F. Eckstein, The Royal Society of Chemistry, 2007, ch. 1, pp. 1-10.
- Liu D, Thélot FA, Piccirilli JA, Liao M, Yin P. Sub-3-Å cryo-EM structure of RNA enabled by engineered homomeric self-assembly. Nat Methods. 2022 May;19(5):576-585. doi: 10.1038/s41592-022-01455-w. Epub 2022 May 2. PMID: 35501384.
- Liu Q, Fang L, Yu G, Wang D, Xiao CL, Wang K. Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. Nat Commun. 2019 Jun 4;10(1):2449. doi: 10.1038/s41467-019-10168-2. PMID: 31164644; PMCID: PMC6547721.
- Liu-Wei W, van der Toorn W, Bohn P, Hölzer M, Smyth RP, von Kleist M. Sequencing accuracy and systematic errors of nanopore direct RNA sequencing. BMC Genomics. 2024 May 28;25(1):528. doi: 10.1186/s12864-024-10440-w. PMID: 38807060; PMCID: PMC11134706.
- Lövgren JM, Wikström PM. The rlmB gene is essential for formation of Gm2251 in 23S rRNA but not for ribosome maturation in Escherichia coli. J Bacteriol. 2001 Dec;183(23):6957-60. doi: 10.1128/JB.183.23.6957-6960.2001. PMID: 11698387; PMCID: PMC95539.
- Luvino, D., Smietana, M., & Vasseur, J.-J. (2006). Selective fluorescence-based detection of dihydrouridine with boronic acids. *Tetrahedron Letters, 47*(52), 9253–9256. https://doi.org/10.1016/j.tetlet.2006.10.150
- M. Charette, M. Gray, Pseudouridine in RNA: What, Where, How, and Why, IUBMB Life, Q2.0 SJR, 2000, doi: 10.1080/152165400410182
- M. Rodgers, Yakubu S Seidu, E. Israel, Influence of 5-Halogenation on the Base-Pairing Energies of Protonated Cytidine Nucleoside Analogue Base Pairs: Implications for the Stabilities of Synthetic i-Motif Structures for DNA Nanotechnology Applications., Journal of the American Society for Mass Spectrometry, Q2.0 SJR, 2022, doi: 10.1021/jasms.2c00137

- Ma H, Jia X, Zhang K, Su Z. Cryo-EM advances in RNA structure determination. Signal Transduct Target Ther. 2022 Feb 23;7(1):58. doi: 10.1038/s41392-022-00916-0. PMID: 35197441; PMCID: PMC8864457.
- Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM, Helm M, Bujnicki JM, Grosjean H. MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res. 2013 Jan;41(Database issue):D262-7. doi: 10.1093/nar/gks1007. Epub 2012 Oct 30. PMID: 23118484; PMCID: PMC3531130.
- Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R. RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Res. 2001 Nov 15;29(22):4724-35. doi: 10.1093/nar/29.22.4724. PMID: 11713323; PMCID: PMC92549.
- Magnus M, Antczak M, Zok T, Wiedemann J, Lukasiak P, Cao Y, Bujnicki JM, Westhof E, Szachniuk M, Miao Z. RNA-Puzzles toolkit: a computational resource of RNA 3D structure benchmark datasets, structure manipulation, and evaluation tools. Nucleic Acids Res. 2020 Jan 24;48(2):576-588. doi: 10.1093/nar/gkz1108. PMID: 31799609; PMCID: PMC7145511.
- Mak CH, Matossian T, Chung WY. Conformational entropy of the RNA phosphate backbone and its contribution to the folding free energy. Biophys J. 2014 Apr 1;106(7):1497-507. doi: 10.1016/j.bpj.2014.02.015. PMID: 24703311; PMCID: PMC3976523.
- Maksimova E, Kravchenko O, Korepanov A, Stolboushkina E. Protein Assistants of Small Ribosomal Subunit Biogenesis in Bacteria. Microorganisms. 2022 Mar 30;10(4):747. doi: 10.3390/microorganisms10040747. PMID: 35456798; PMCID: PMC9032327.
- Maksimova EM, Korepanov AP, Kravchenko OV, Baymukhametov TN, Myasnikov AG, Vassilenko KS, Afonina ZA, Stolboushkina EA. RbfA Is Involved in Two Important Stages of 30S Subunit Assembly: Formation of the Central Pseudoknot and Docking of Helix 44 to the Decoding Center. Int J Mol Sci. 2021 Jun 7;22(11):6140. doi: 10.3390/ijms22116140. PMID: 34200244; PMCID: PMC8201178.
- Malone BF, Perry JK, Olinares PDB, Lee HW, Chen J, Appleby TC, Feng JY, Bilello JP, Ng H, Sotiris J, Ebrahim M, Chua EYD, Mendez JH, Eng ET, Landick R, Götte M, Chait BT, Campbell EA, Darst SA. Structural basis for substrate selection by the SARS-CoV-2 replicase. Nature. 2023 Feb;614(7949):781-787. doi: 10.1038/s41586-022-05664-3. Epub 2023 Feb 1. PMID: 36725929; PMCID: PMC9891196.
- Mao K, Wang J, Xiao Y. Length-Dependent Deep Learning Model for RNA Secondary Structure Prediction. Molecules. 2022 Feb 2;27(3):1030. doi: 10.3390/molecules27031030. PMID: 35164295; PMCID: PMC8838716.
- Marchand, V., Pichot, F., Neybecker, P., Ayadi, L., Bourguignon-Igel, V., Wacheul, L., Lafontaine, D., Pinzano, A., Helm, M., & Motorin, Y. (2020). HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. Nucleic Acids Research, 48, e110 - e110. https://doi.org/10.1093/nar/gkaa769.
- Marek Bartošovič, Helena Covelo Molares, P. Gregorova, Dominika Hroššová, Grzegorz Kudla, Š. Vaňáčová, N6-methyladenosine demethylase FTO targets pre-mRNAs and regulates alternative splicing and 3′-end processing, Nucleic Acids Research, Q1.0 SJR, 2017, doi: 10.1093/nar/gkx778
- Mathias Sprinzl, Clemens Steegborn, Frank Hübel, Sergey Steinberg, Compilation of tRNA Sequences and Sequences of tRNA Genes, Nucleic Acids Research, Volume 24, Issue 1, 1 January 1996, Pages 68–72, https://doi.org/10.1093/nar/24.1.68
- Matta CF, Castillo N, Boyd RJ. Extended weak bonding interactions in DNA: pi-stacking (base-base), base-backbone, and backbone-backbone interactions. J Phys Chem B. 2006 Jan 12;110(1):563-78. doi: 10.1021/jp054986g. PMID: 16471569.

- Mattia Furlan, Anna Delgado-Tejedor, Logan Mulroney, Mattia Pelizzola, E. Novoa, T. Leonardi, Computational methods for RNA modification detection from nanopore direct RNA sequencing data, RNA Biology, Q1.0 SJR, 2021, doi: 10.1080/15476286.2021.1978215
- Miao Z, Adamiak RW, Antczak M, Boniecki MJ, Bujnicki J, Chen SJ, Cheng CY, Cheng Y, Chou FC, Das R, Dokholyan NV, Ding F, Geniesse C, Jiang Y, Joshi A, Krokhotin A, Magnus M, Mailhot O, Major F, Mann TH, Piątkowski P, Pluta R, Popenda M, Sarzynska J, Sun L, Szachniuk M, Tian S, Wang J, Wang J, Watkins AM, Wiedemann J, Xiao Y, Xu X, Yesselman JD, Zhang D, Zhang Y, Zhang Z, Zhao C, Zhao P, Zhou Y, Zok T, Żyła A, Ren A, Batey RT, Golden BL, Huang L, Lilley DM, Liu Y, Patel DJ, Westhof E. RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. RNA. 2020 Aug;26(8):982-995. doi: 10.1261/rna.075341.120. Epub 2020 May 5. PMID: 32371455; PMCID: PMC7373991.
- Miao Z, Adamiak RW, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cheng C, Chojnowski G, Chou FC, Cordero P, Cruz JA, Ferré-D'Amaré AR, Das R, Ding F, Dokholyan NV, Dunin-Horkawicz S, Kladwang W, Krokhotin A, Lach G, Magnus M, Major F, Mann TH, Masquida B, Matelska D, Meyer M, Peselis A, Popenda M, Purzycka KJ, Serganov A, Stasiewicz J, Szachniuk M, Tandon A, Tian S, Wang J, Xiao Y, Xu X, Zhang J, Zhao P, Zok T, Westhof E. RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. RNA. 2015 Jun;21(6):1066-84. doi: 10.1261/rna.049502.114. Epub 2015 Apr 16. PMID: 25883046; PMCID: PMC4436661.
- Miao Z, Westhof E. RNA Structure: Advances and Assessment of 3D Structure Prediction. Annu Rev Biophys. 2017 May 22;46:483-503. doi: 10.1146/annurev-biophys-070816-034125. Epub 2017 Mar 30. PMID: 28375730.
- Michael O'Connor, Cheryl L. Thomas, Robert A. Zimmermann, Albert E. Dahlberg, Decoding Fidelity at the Ribosomal A and P Sites: Influence of Mutations in three Different Regions of the Decoding Domain in 16S rRNA, *Nucleic Acids Research*, Volume 25, Issue 6, 1 March 1997, Pages 1185–1193, https://doi.org/10.1093/nar/25.6.1185
- Michel Fasnacht, Stefano Gallo, Puneet Sharma, Maximilian Himmelstoß, Patrick A Limbach, Jessica Willi, Norbert Polacek, Dynamic 23S rRNA modification ho^{5}C2501 benefits *Escherichia coli* under oxidative stress, *Nucleic Acids Research*, Volume 50, Issue 1, 11 January 2022, Pages 473–489, https://doi.org/10.1093/nar/gkab1224
- Mihaly Varadi, Damian Bertoni, Paulyna Magana, Urmila Paramval, Ivanna Pidruchna, Malarvizhi Radhakrishnan, Maxim Tsenkov, Sreenath Nair, Milot Mirdita, Jingi Yeo, Oleg Kovalevskiy, Kathryn Tunyasuvunakool, Agata Laydon, Augustin Žídek, Hamish Tomlinson, Dhavanthi Hariharan, Josh Abrahamson, Tim Green, John Jumper, Ewan Birney, Martin Steinegger, Demis Hassabis, Sameer Velankar, AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences, Nucleic Acids Research, Volume 52, Issue D1, 5 January 2024, Pages D368–D375, https://doi.org/10.1093/nar/gkad1011
- Miles AJ, Wallace BA. CDtoolX, a downloadable software package for processing and analyses of circular dichroism spectroscopic data. Protein Sci. 2018 Sep;27(9):1717-1722. doi: 10.1002/pro.3474. PMID: 30168221; PMCID: PMC6194270.
- Mishima, E., Jinno, D., Akiyama, Y., Itoh, K., Nankumo, S., Shima, H., Kikuchi, K., Takeuchi, Y., Elkordy, A., Suzuki, T., Niizuma, K., Ito, S., Tomioka, Y., & Abe, T. (2015). Immuno-Northern Blotting: Detection of RNA Modifications by Using Antibodies against Modified Nucleosides. PLoS ONE, 10. https://doi.org/10.1371/journal.pone.0143756.
- Montpetit A, Payant C, Nolan JM, Brakier-Gingras L. Analysis of the conformation of the 3' major domain of Escherichia coli16S ribosomal RNA using site-directed photoaffinity crosslinking. RNA. 1998 Nov;4(11):1455-66. doi: 10.1017/s1355838298981079. PMID: 9814765; PMCID: PMC1369717.

- Moore-Kelly, C., Welsh, J., Rodger, A., Dafforn, T. R., & Thomas, O. R. T. (2019). *Automated high-throughput capillary circular dichroism and intrinsic fluorescence spectroscopy for rapid determination of protein structure. Analytical Chemistry, 91*(21), 13794–13802.
- Morais P, Adachi H, Yu YT. The Critical Contribution of Pseudouridine to mRNA COVID-19 Vaccines. Front Cell Dev Biol. 2021 Nov 4;9:789427. doi: 10.3389/fcell.2021.789427. PMID: 34805188; PMCID: PMC8600071.
- Mu, K., Fei, Y., Xu, Y., & Zhang, Q. C. (2025). *RASP v2.0: an updated atlas for RNA structure probing data. Nucleic Acids Research, 53*(D1), D211–D219.
- Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013 Nov 15;29(22):2933-5. doi: 10.1093/bioinformatics/btt509. Epub 2013 Sep 4. PMID: 24008419; PMCID: PMC3810854.
- Nelson, D.L. and Cox, M.M. (2017) Lehninger Principles of Biochemistry. 7th Edition, W.H. Freeman, New York, 1328.
- Nielsen, S., Chakraborty, T., & Hoffmann, S. (2005). Synchrotron radiation circular dichroism spectroscopy of ribose and deoxyribose sugars, adenosine, AMP and dAMP nucleotides.. Chemphyschem : a European journal of chemical physics and physical chemistry, 6 12, 2619-24 . https://doi.org/10.1002/CPHC.200500236.
- Oborská-Oplová, M., Fischer, U., Altvater, M., Panse, V.G. (2022). Eukaryotic Ribosome assembly and Nucleocytoplasmic Transport. In: Entian, KD. (eds) Ribosome Biogenesis. Methods in Molecular Biology, vol 2533. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-2501-9_7
- O'Connor M, Leppik M, Remme J. Pseudouridine-Free Escherichia coli Ribosomes. J Bacteriol. 2018 Jan 24;200(4):e00540-17. doi: 10.1128/JB.00540-17. PMID: 29180357; PMCID: PMC5786706.
- Oerum S, Dégut C, Barraud P, Tisné C. m1A Post-Transcriptional Modification in tRNAs. Biomolecules. 2017 Feb 21;7(1):20. doi: 10.3390/biom7010020. PMID: 28230814; PMCID: PMC5372732.
- Ofengand J, Bakin A, Wrzesinski J, Nurse K, Lane BG. The pseudouridine residues of ribosomal RNA. Biochem Cell Biol. 1995 Nov-Dec;73(11-12):915-24. doi: 10.1139/o95-099. PMID: 8722007.
- Ontiveros N, Cooke E, Nawrocki EP, Triebel S, Marz M, Rivas E, Griffiths-Jones S, Petrov AI, Bateman A, Sweeney B. Rfam 15: RNA families database in 2025. bioRxiv [Preprint]. 2024 Sep 24:2024.09.23.614430. doi: 10.1101/2024.09.23.614430. Update in: Nucleic Acids Res. 2024 Nov 11:gkae1023. doi: 10.1093/nar/gkae1023. PMID: 39372780; PMCID: PMC11451735.
- Ozanick SG, Bujnicki JM, Sem DS, Anderson JT. Conserved amino acids in each subunit of the heteroligomeric tRNA m1A58 Mtase from Saccharomyces cerevisiae contribute to tRNA binding. Nucleic Acids Res. 2007;35(20):6808-19. doi: 10.1093/nar/gkm574. Epub 2007 Oct 10. PMID: 17932071; PMCID: PMC2175304.
- P. J. McCown, A. Ruszkowska, Charlotte N Kunkler, Kurtis Breger, Jacob P Hulewicz, Matthew C Wang, Noah A Springer, Jessica A. Brown, Naturally occurring modified ribonucleosides, Wiley Interdisciplinary Reviews. RNA, Q1.0 SJR, 2020, doi: 10.1002/wrna.1595
- P. Pletnev, E. Guseva, A. Zanina, S. Evfratov, Margarita M. Dzama, Vsevolod Treshin, A. Pogorel'skaya, I. Osterman, A. Golovina, M. Rubtsova, M. Serebryakova, O. Pobeguts, V. Govorun, A. Bogdanov, O. Dontsova, P. Sergiev, Comprehensive Functional Analysis of Escherichia coli Ribosomal RNA Methyltransferases, Frontiers in Genetics, Q2.0 SJR, 2020, doi: 10.3389/fgene.2020.00097
- Parker, T., Hohenstein, E., Parrish, R., Hud, N., & Sherrill, C. (2013). Quantum-mechanical analysis of the energetic contributions to π stacking in nucleic acids versus rise, twist, and slide.. *Journal of the American Chemical Society*, 135 4, 1306-16 . https://doi.org/10.1021/ja3063309.
- Petrov AS, Bernier CR, Gulen B, Waterbury CC, Hershkovits E, Hsiao C, Harvey SC, Hud NV, Fox GE, Wartell RM, Williams LD. Secondary structures of rRNAs from all three domains of life. PLoS

One. 2014 Feb 5;9(2):e88222. doi: 10.1371/journal.pone.0088222. PMID: 24505437; PMCID: PMC3914948.

- Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 2021 Jan;30(1):70-82. doi: 10.1002/pro.3943. Epub 2020 Oct 22. PMID: 32881101; PMCID: PMC7737788.
- Piekna-Przybylska D, Decatur WA, Fournier MJ. The 3D rRNA modification maps database: with interactive tools for ribosome analysis. Nucleic Acids Res. 2008 Jan;36(Database issue):D178-83. doi: 10.1093/nar/gkm855. Epub 2007 Oct 18. PMID: 17947322; PMCID: PMC2238946.
- Pliatsika V, Loher P, Magee R, Telonis AG, Londin E, Shigematsu M, Kirino Y, Rigoutsos I. MINTbase v2.0: a comprehensive database for tRNA-derived fragments that includes nuclear and mitochondrial fragments from all The Cancer Genome Atlas projects. Nucleic Acids Res. 2018 Jan 4;46(D1):D152-D159. doi: 10.1093/nar/gkx1075. PMID: 29186503; PMCID: PMC5753276.
- Polikanov YS, Melnikov SV, Söll D, Steitz TA. Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. Nat Struct Mol Biol. 2015 Apr;22(4):342-344. doi: 10.1038/nsmb.2992. Epub 2015 Mar 16. PMID: 25775268; PMCID: PMC4401423.
- Qin D, Liu Q, Devaraj A, Fredrick K. Role of helix 44 of 16S rRNA in the fidelity of translation initiation. RNA. 2012 Mar;18(3):485-95. doi: 10.1261/rna.031203.111. Epub 2012 Jan 25. PMID: 22279149; PMCID: PMC3285936.
- Quigley GJ, Rich A. Structural domains of transfer RNA molecules. Science. 1976 Nov 19;194(4267):796-806. doi: 10.1126/science.790568. PMID: 790568.
- R. Yamagami, Jacob P Sieg, P. Bevilacqua, Functional Roles of Chelated Magnesium Ions in RNA Folding and Function., Biochemistry, Q1.0 SJR, 2021, doi: 10.1021/acs.biochem.1c00012
- R.J.C. Gilbert, Y. Gordiyenko, T. von der Haar, A.F. Sonnen, G. Hofmann, M. Nardelli, D.I. Stuart, & J.E.G. McCarthy, Reconfiguration of yeast 40S ribosomal subunit domains by the translation initiation multifactor complex, Proc. Natl. Acad. Sci. U.S.A. 104 (14) 5788-5793, https://doi.org/10.1073/pnas.0606880104 (2007).
- Rego N, Koes D. 3Dmol.js: molecular visualization with WebGL. Bioinformatics. 2015 Apr 15;31(8):1322-4. doi: 10.1093/bioinformatics/btu829. Epub 2014 Dec 12. PMID: 25505090; PMCID: PMC4393526.
- Rodell R, Robalin N, Martinez NM. Why U matters: detection and functions of pseudouridine modifications in mRNAs. Trends Biochem Sci. 2024 Jan;49(1):12-27. doi: 10.1016/j.tibs.2023.10.008. Epub 2023 Dec 14. PMID: 38097411; PMCID: PMC10976346.
- Roost C, Lynch SR, Batista PJ, Qu K, Chang HY, Kool ET. Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. J Am Chem Soc. 2015 Feb 11;137(5):2107-15. doi: 10.1021/ja513080v. Epub 2015 Feb 2. Erratum in: J Am Chem Soc. 2015 Jul 1;137(25):8308. doi: 10.1021/jacs.5b05858. PMID: 25611135; PMCID: PMC4405242.
- Rother M, Milanowska K, Puton T, Jeleniewicz J, Rother K, Bujnicki JM. ModeRNA server: an online tool for modeling RNA 3D structures. Bioinformatics. 2011 Sep 1;27(17):2441-2. doi: 10.1093/bioinformatics/btr400. Epub 2011 Jul 4. PMID: 21727140.
- Roy R, Hohng S, Ha T. A practical guide to single-molecule FRET. Nat Methods. 2008 Jun;5(6):507-16. doi: 10.1038/nmeth.1208. PMID: 18511918; PMCID: PMC3769523.
- Rydberg P, Gloriam DE, Zaretzki J, Breneman C, Olsen L. SMARTCyp: A 2D Method for Prediction of Cytochrome P450-Mediated Drug Metabolism. ACS Med Chem Lett. 2010 Mar 15;1(3):96-100. doi: 10.1021/ml100016x. PMID: 24936230; PMCID: PMC4055970.
- Ryland W. Jackson, Claire M Smathers, A. Robart, General Strategies for RNA X-ray Crystallography, Molecules, Q1.0 SJR, 2023, doi: 10.3390/molecules28052111

- S. Dixit, S. Jaffrey, Expanding the epitranscriptome: Dihydrouridine in mRNA, PLoS Biology, Q1.0 SJR, 2022, doi: 10.1371/journal.pbio.3001720
- Samanta D, Mukhopadhyay D, Chowdhury S, Ghosh J, Pal S, Basu A, Bhattacharya A, Das A, Das D, DasGupta C. Protein folding by domain V of Escherichia coli 23S rRNA: specificity of RNA-protein interactions. J Bacteriol. 2008 May;190(9):3344-52. doi: 10.1128/JB.01800-07. Epub 2008 Feb 29. PMID: 18310328; PMCID: PMC2347393.
- Sara Espinosa, Lingdi Zhang, Xueni Li, R. Zhao, Understanding pre-mRNA splicing through crystallography., Methods, Q1.0 SJR, 2017, doi: 10.1016/j.ymeth.2017.04.023
- Sas-Chen A, Schwartz S. Misincorporation signatures for detecting modifications in mRNA: Not as simple as it sounds. Methods. 2019 Mar 1;156:53-59. doi: 10.1016/j.ymeth.2018.10.011. Epub 2018 Oct 23. PMID: 30359724.
- Scott WG, Horan LH, Martick M. The hammerhead ribozyme: structure, catalysis, and gene regulation. Prog Mol Biol Transl Sci. 2013;120:1-23. doi: 10.1016/B978-0-12-381286-5.00001-9. PMID: 24156940; PMCID: PMC4008931.
- Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, Velankar S, Burley SK, Koča J, Rose AS. Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. Nucleic Acids Res. 2021 Jul 2;49(W1):W431-W437. doi: 10.1093/nar/gkab314. PMID: 33956157; PMCID: PMC8262734.
- Sengar A, Heddi B, Phan AT. Formation of G-quadruplexes in poly-G sequences: structure of a propeller-type parallel-stranded G-quadruplex formed by a G_{15} stretch. Biochemistry. 2014 Dec 16;53(49):7718-23. doi: 10.1021/bi500990v. Epub 2014 Dec 1. PMID: 25375976.
- Sergiev PV, Lesnyak DV, Bogdanov AA, Dontsova OA. Identification of Escherichia coli m2G methyltransferases: II. The ygjO gene encodes a methyltransferase specific for G1835 of the 23 S rRNA. J Mol Biol. 2006 Nov 17;364(1):26-31. doi: 10.1016/j.jmb.2006.09.008. Epub 2006 Sep 8. PMID: 17010380.
- Sergiev PV, Serebryakova MV, Bogdanov AA, Dontsova OA. The ybiN gene of Escherichia coli encodes adenine-N6 methyltransferase specific for modification of A1618 of 23 S ribosomal RNA, a methylated residue located close to the ribosomal exit tunnel. J Mol Biol. 2008 Jan 4;375(1):291-300. doi: 10.1016/j.jmb.2007.10.051. Epub 2007 Oct 23. PMID: 18021804.
- Shachar R, Dierks D, Garcia-Campos MA, Uzonyi A, Toth U, Rossmanith W, Schwartz S. Dissecting the sequence and structural determinants guiding m6A deposition and evolution via inter- and intra-species hybrids. Genome Biol. 2024 Feb 15;25(1):48. doi: 10.1186/s13059-024-03182-1. PMID: 38360609; PMCID: PMC10870504.
- Shakked, Z., Rabinovich, D., Singh, T., Cruse, W., Salisbury, S., & Viswamitra, M. (1983). Sequence-dependent conformation of an A-DNA double helix. The crystal structure of the octamer d(G-G-T-A-T-A-C-C).. *Journal of molecular biology*, 166 2, 183-201 . https://doi.org/10.1016/s0022-2836(83)80005-9.
- Sharma, S., Hartmann, J.D., Watzinger, P. *et al.* A single N^1-methyladenosine on the large ribosomal subunit rRNA impacts locally its structure and the translation of key metabolic enzymes. *Sci Rep* **8**, 11904 (2018). https://doi.org/10.1038/s41598-018-30383-z
- Shaw EA, Thomas NK, Jones JD, Abu-Shumays RL, Vaaler AL, Akeson M, Koutmou KS, Jain M, Garcia DM. Combining Nanopore direct RNA sequencing with genetics and mass spectrometry for analysis of T-loop base modifications across 42 yeast tRNA isoacceptors. Nucleic Acids Res. 2024 Oct 28;52(19):12074-12092. doi: 10.1093/nar/gkae796. PMID: 39340295; PMCID: PMC11514469.
- Shen, T., Hu, Z., Sun, S. *et al.* Accurate RNA 3D structure prediction using a language model-based deep learning approach. *Nat Methods* (2024). https://doi.org/10.1038/s41592-024-02487-0
- Shi H, Liu B, Nussbaumer F, Rangadurai A, Kreutz C, Al-Hashimi HM. NMR Chemical Exchange Measurements Reveal That N6-Methyladenosine Slows RNA Annealing. J Am Chem Soc. 2019 Dec

26;141(51):19988-19993. doi: 10.1021/jacs.9b10939. Epub 2019 Dec 16. PMID: 31826614; PMCID: PMC7807175.

- Shu X, Cao J, Liu J. m6A-label-seq: A metabolic labeling protocol to detect transcriptome-wide mRNA N6-methyladenosine (m6A) at base resolution. STAR Protoc. 2022 Jan 13;3(1):101096. doi: 10.1016/j.xpro.2021.101096. PMID: 35059657; PMCID: PMC8760545.
- Simone Höfler, T. Carlomagno, Structural and functional roles of 2'-O-ribose methylations and their enzymatic machinery across multiple classes of RNAs., Current opinion in structural biology, Q1.0 SJR, 2020, doi: 10.1016/j.sbi.2020.05.008
- Sloan KE, Warda AS, Sharma S, Entian KD, Lafontaine DLJ, Bohnsack MT. Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. RNA Biol. 2017 Sep 2;14(9):1138-1152. doi: 10.1080/15476286.2016.1259781. Epub 2016 Dec 2. PMID: 27911188; PMCID: PMC5699541.
- Šmidlehner T, Piantanida I, Pescitelli G. Polarization spectroscopy methods in the determination of interactions of small molecules with nucleic acids - tutorial. Beilstein J Org Chem. 2018 Jan 8;14:84-105. doi: 10.3762/bjoc.14.5. PMID: 29441133; PMCID: PMC5789433.
- Sood, V., Yekta, S., & Collins, R. (2002). The contribution of 2'-hydroxyls to the cleavage activity of the Neurospora VS ribozyme. *Nucleic acids research*, 30 5, 1132-8 . https://doi.org/10.1093/NAR/30.5.1132.
- Šponer J, Banáš P, Jurečka P, Zgarbová M, Kührová P, Havrila M, Krepl M, Stadlbauer P, Otyepka M. Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. J Phys Chem Lett. 2014 May 15;5(10):1771-82. doi: 10.1021/jz500557y. Epub 2014 May 7. PMID: 26270382.
- Šponer J, Bussi G, Krepl M, Banáš P, Bottaro S, Cunha RA, Gil-Ley A, Pinamonti G, Poblete S, Jurečka P, Walter NG, Otyepka M. RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. Chem Rev. 2018 Apr 25;118(8):4177-4338. doi: 10.1021/acs.chemrev.7b00427. Epub 2018 Jan 3. PMID: 29297679; PMCID: PMC5920944.
- Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021 Feb;22(2):96-118. doi: 10.1038/s41580-020-00315-9. Epub 2020 Dec 22. Erratum in: Nat Rev Mol Cell Biol. 2021 Feb;22(2):159. doi: 10.1038/s41580-021-00330-4. PMID: 33353982; PMCID: PMC7754182.
- Su C, Jin M, Zhang W. Conservation and Diversification of tRNA t^6A-Modifying Enzymes across the Three Domains of Life. Int J Mol Sci. 2022 Nov 6;23(21):13600. doi: 10.3390/ijms232113600. PMID: 36362385; PMCID: PMC9654439.
- Su S, Li S, Deng T, Gao M, Yin Y, Wu B, Peng C, Liu J, Ma J, Zhang K. Cryo-EM structures of human m^6A writer complexes. Cell Res. 2022 Nov;32(11):982-994. doi: 10.1038/s41422-022-00725-8. Epub 2022 Sep 27. PMID: 36167981; PMCID: PMC9652331.
- Su, L., Huber, E. M., Westphalen, M., Gellner, J., Bode, E., Köbel, T., Grün, P., Alanjary, M. M., Glatter, T., Cirnski, K., Müller, R., Schindler, D., Groll, M., & Bode, H. B. (2024). Isofunctional but Structurally Different Methyltransferases for Dithiolopyrrolone Diversification. *Angewandte Chemie International Edition*, 63, e202410799. https://doi.org/10.1002/anie.202410799
- Sun Q, Zhu X, Qi J, An W, Lan P, Tan D, Chen R, Wang B, Zheng S, Zhang C, Chen X, Zhang W, Chen J, Dong MQ, Ye K. Molecular architecture of the 90S small subunit pre-ribosome. Elife. 2017 Feb 28;6:e22086. doi: 10.7554/eLife.22086. Erratum in: Elife. 2017 Jun 26;6:e29876. doi: 10.7554/eLife.29876. PMID: 28244370; PMCID: PMC5354517.
- Sutherland, John C., Bohai Lin, Jo Ann Mugavero, John Trunk, Maria Tomasz, Regina Santella, Luis Marky, and Kenneth J. Breslauer. "Vacuum ultraviolet circular dichroism of double stranded nucleic acids." *Photochemistry and photobiology* 44, no. 3 (1986): 295-301.

- Taghavi A, Riveros I, Wales DJ, Yildirim I. Evaluating Geometric Definitions of Stacking for RNA Dinucleoside Monophosphates Using Molecular Mechanics Calculations. J Chem Theory Comput. 2022 Jun 14;18(6):3637-3653. doi: 10.1021/acs.jctc.2c00178. Epub 2022 Jun 2. PMID: 35652685; PMCID: PMC9715528.
- Tao Shen, Zhihang Hu, Zhangzhi Peng, Jiayang Chen, Peng Xiong, Liang Hong, Liangzhen Zheng, Yixuan Wang, Irwin King, Sheng Wang, Siqi Sun, Yu Li, E2Efold-3D: End-to-End Deep Learning Method for accurate de novo RNA 3D Structure Prediction, https://doi.org/10.48550/arXiv.2207.01586
- Taoka M, Nobe Y, Yamaki Y, Sato K, Ishikawa H, Izumikawa K, Yamauchi Y, Hirota K, Nakayama H, Takahashi N, Isobe T. Landscape of the complete RNA chemical modifications in the human 80S ribosome. Nucleic Acids Res. 2018 Oct 12;46(18):9289-9298. doi: 10.1093/nar/gky811. PMID: 30202881; PMCID: PMC6182160.
- Tassinari V, Cesarini V, Tomaselli S, Ianniello Z, Silvestris DA, Ginistrelli LC, Martini M, De Angelis B, De Luca G, Vitiani LR, Fatica A, Locatelli F, Gallo A. ADAR1 is a new target of METTL3 and plays a pro-oncogenic role in glioblastoma by an editing-independent mechanism. Genome Biol. 2021 Jan 28;22(1):51. doi: 10.1186/s13059-021-02271-9. PMID: 33509238; PMCID: PMC7842030.
- The RNAcentral Consortium , RNAcentral: a hub of information for non-coding RNA sequences, *Nucleic Acids Research*, Volume 47, Issue D1, 08 January 2019, Pages D221–D229, https://doi.org/10.1093/nar/gky1034
- Thomas M. Carlile, Nicole M. Martinez, Cassandra Schaening, Amanda Su, Tristan A. Bell, Boris Zinshteyn, Wendy V. Gilbert, mRNA structure determines modification by pseudouridine synthase 1, Nature chemical biology, Q1.0 SJR, 2019, doi: 10.1038/s41589-019-0353-z
- Toh SM, Xiong L, Bae T, Mankin AS. The methyltransferase YfgB/RlmN is responsible for modification of adenosine 2503 in 23S rRNA. RNA. 2008 Jan;14(1):98-106. doi: 10.1261/rna.814408. Epub 2007 Nov 19. PMID: 18025251; PMCID: PMC2151032.
- Toubdji S, Thullier Q, Kilz LM, Marchand V, Yuan Y, Sudol C, Goyenvalle C, Jean-Jean O, Rose S, Douthwaite S, Hardy L, Baharoglu Z, de Crécy-Lagard V, Helm M, Motorin Y, Hamdane D, Brégeon D. Exploring a unique class of flavoenzymes: Identification and biochemical characterization of ribosomal RNA dihydrouridine synthase. Proc Natl Acad Sci U S A. 2024 Aug 6;121(32):e2401981121. doi: 10.1073/pnas.2401981121. Epub 2024 Jul 30. Erratum in: Proc Natl Acad Sci U S A. 2025 Jan 28;122(4):e2424888121. doi: 10.1073/pnas.2424888121. PMID: 39078675; PMCID: PMC11317573.
- Townshend RJL, Eismann S, Watkins AM, Rangan R, Karelina M, Das R, Dror RO. Geometric deep learning of RNA structure. Science. 2021 Aug 27;373(6558):1047-1051. doi: 10.1126/science.abe5650. Erratum in: Science. 2023 Jan 27;379(6630):eadg6616. doi: 10.1126/science.adg6616. PMID: 34446608; PMCID: PMC9829186.
- Tscherne JS, Nurse K, Popienick P, Michel H, Sochacki M, Ofengand J. Purification, cloning, and characterization of the 16S RNA m5C967 methyltransferase from Escherichia coli. Biochemistry. 1999 Feb 9;38(6):1884-92. doi: 10.1021/bi981880l. PMID: 10026269.
- Tsutomu Suzuki, The expanding world of tRNA modifications and their disease relevance, Nature Reviews Molecular Cell Biology, Q1.0 SJR, 2021, doi: 10.1038/s41580-021-00342-0
- Tzokov, S., Murray, I., & Grasby, J. (2002). The role of magnesium ions and 2'-hydroxyl groups in the VS ribozyme-substrate interaction. *Journal of molecular biology*, 324 2, 215-26 . https://doi.org/10.1016/S0022-2836(02)01063-X.
- V. Potapov, Xiaoqing Fu, Nan Dai, Ivan R. Corrêa, N. Tanner, J. L. Ong, Base modifications affecting RNA polymerase and reverse transcriptase fidelity, Nucleic Acids Research, Q1.0 SJR, 2018, doi: 10.1093/nar/gky341

- Vaidyanathan PP, AlSadhan I, Merriman DK, Al-Hashimi HM, Herschlag D. Pseudouridine and N6-methyladenosine modifications weaken PUF protein/RNA interactions. RNA. 2017 May;23(5):611-618. doi: 10.1261/rna.060053.116. Epub 2017 Jan 30. PMID: 28138061; PMCID: PMC5393172.
- Vester B, Douthwaite S. Domain V of 23S rRNA contains all the structural elements necessary for recognition by the ErmE methyltransferase. J Bacteriol. 1994 Nov;176(22):6999-7004. doi: 10.1128/jb.176.22.6999-7004.1994. PMID: 7961464; PMCID: PMC197073.
- Vorlickova, M., J. Kypr, S. Stokrova and J. Sponar (1982a) A Z-like form of poly(dA- dC)'poly(dG-dT) in solution? Nucleic Acids Res. 10, 107 1- 1080.
- Wachowius F, Höbartner C. Chemical RNA modifications for studies of RNA structure and dynamics. Chembiochem. 2010 Mar 1;11(4):469-80. doi: 10.1002/cbic.200900697. PMID: 20135663.
- Waleń T, Chojnowski G, Gierski P, Bujnicki JM. ClaRNA: a classifier of contacts in RNA 3D structures based on a comparative analysis of various classification schemes. Nucleic Acids Res. 2014 Oct 29;42(19):e151. doi: 10.1093/nar/gku765. Epub 2014 Aug 26. PMID: 25159614; PMCID: PMC4231730.
- Wang J, Wang L. Deep analysis of RNA N^6-adenosine methylation (m^6A) patterns in human cells. NAR Genom Bioinform. 2020 Feb 7;2(1):lqaa007. doi: 10.1093/nargab/lqaa007. PMID: 33575554; PMCID: PMC7671394.
- Wang, J., Chen, S., Baldi, P., Flynn, R. A., & Spitale, R. C. (2022). *An atlas of posttranslational modifications on RNA binding proteins. Nucleic Acids Research, 50*(18), 10373–10390.
- Werner S, Galliot A, Pichot F, Kemmer T, Marchand V, Sednev MV, Lence T, Roignant JY, König J, Höbartner C, Motorin Y, Hildebrandt A, Helm M. NOseq: amplicon sequencing evaluation method for RNA m6A sites after chemical deamination. Nucleic Acids Res. 2021 Feb 26;49(4):e23. doi: 10.1093/nar/gkaa1173. PMID: 33313868; PMCID: PMC7913672.
- Westhof E. Isostericity and tautomerism of base pairs in nucleic acids. FEBS Lett. 2014 Aug 1;588(15):2464-9. doi: 10.1016/j.febslet.2014.06.031. Epub 2014 Jun 17. PMID: 24950426.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., … & Mons, B. (2016). *The FAIR guiding principles for scientific data management and stewardship. Scientific Data, 3*, 160018.
- Williamson, M. (2011). How Proteins Work (1st ed.). Garland Science. https://doi.org/10.1201/9781136665493
- Wolff P, Villette C, Zumsteg J, Heintz D, Antoine L, Chane-Woon-Ming B, Droogmans L, Grosjean H, Westhof E. Comparative patterns of modified nucleotides in individual tRNA species from a mesophilic and two thermophilic archaea. RNA. 2020 Dec;26(12):1957-1975. doi: 10.1261/rna.077537.120. Epub 2020 Sep 29. PMID: 32994183; PMCID: PMC7668247.
- Woodson SA. RNA folding and ribosome assembly. Curr Opin Chem Biol. 2008 Dec;12(6):667-73. doi: 10.1016/j.cbpa.2008.09.024. Epub 2008 Oct 18. PMID: 18935976; PMCID: PMC2651837.
- Xiong W, Zhao Y, Wei Z, Li C, Zhao R, Ge J, Shi B. N1-methyladenosine formation, gene regulation, biological functions, and clinical relevance. Mol Ther. 2023 Feb 1;31(2):308–30. doi: 10.1016/j.ymthe.2022.10.015. Epub 2022 Oct 29. PMCID: PMC9931621.
- Xiujuan Ou, Yi Zhang, Yi Xiong, Yi Xiao, Advances in RNA 3D Structure Prediction, Journal of chemical information and modeling, Q1.0 SJR, 2022, doi: 10.1021/acs.jcim.2c00939
- Xu X, Chen SJ. Physics-based RNA structure prediction. Biophys Rep. 2015;1:2-13. doi: 10.1007/s41048-015-0001-4. Epub 2015 Jul 9. PMID: 26942214; PMCID: PMC4762127.
- Xu Z, Culver GM. Differential assembly of 16S rRNA domains during 30S subunit formation. RNA. 2010 Oct;16(10):1990-2001. doi: 10.1261/rna.2246710. Epub 2010 Aug 24. PMID: 20736336; PMCID: PMC2941107.

- Y. Motorin, M. Helm, tRNA stabilization by modified nucleotides., Biochemistry, Q1.0 SJR, 2010, doi: 10.1021/bi100408z
- Y. Motorin, M. Quinternet, Wassim Rhalloussi, V. Marchand, Constitutive and variable 2'-O-methylation (Nm) in human ribosomal RNA, RNA Biology, Q1.0 SJR, 2021, doi: 10.1080/15476286.2021.1974750
- Yamaki Y, Nobe Y, Koike M, Yamauchi Y, Hirota K, Takahashi N, Nakayama H, Isobe T, Taoka M. Direct Determination of Pseudouridine in RNA by Mass Spectrometry Coupled with Stable Isotope Labeling. Anal Chem. 2020 Aug 18;92(16):11349-11356. doi: 10.1021/acs.analchem.0c02122. Epub 2020 Jul 28. PMID: 32662983.
- Yang H, Zhan Y, Fenn D, Chi LM, Lam SL. Effect of 1-methyladenine on double-helical DNA structures. FEBS Lett. 2008 May 14;582(11):1629-33. doi: 10.1016/j.febslet.2008.04.013. Epub 2008 Apr 22. PMID: 18435925.
- Yang, J., Watzinger, P., Sharma, S. (2022). Mapping of the Chemical Modifications of rRNAs. In: Entian, KD. (eds) Ribosome Biogenesis. Methods in Molecular Biology, vol 2533. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-2501-9_11
- Yang, Y., Hsu, P.J., Chen, YS. *et al.* Dynamic transcriptomic m^6A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Res* **28**, 616–624 (2018). https://doi.org/10.1038/s41422-018-0040-8
- Yonath, A. (2011). X-ray crystallography at the heart of life science.. Current opinion in structural biology, 21 5, 622-6 . https://doi.org/10.1016/j.sbi.2011.07.005.
- Yu Zhao, Jay Rai, Hong Li, Regulation of translation by ribosomal RNA pseudouridylation, Science Advances, Q1.0 SJR, 2023, doi: 10.1126/sciadv.adg8190
- Zampetaki A, Albrecht A, Steinhofel K. Long Non-coding RNA Structure and Function: Is There a Link? Front Physiol. 2018 Aug 24;9:1201. doi: 10.3389/fphys.2018.01201. Erratum in: Front Physiol. 2019 Sep 03;10:1127. doi: 10.3389/fphys.2019.01127. PMID: 30197605; PMCID: PMC6117379.
- Zhang Z, Chen LQ, Zhao YL, Yang CG, Roundtree IA, Zhang Z, Ren J, Xie W, He C, Luo GZ. Single-base mapping of m6A by an antibody-independent method. Sci Adv. 2019 Jul 3;5(7):eaax0250. doi: 10.1126/sciadv.aax0250. PMID: 31281898; PMCID: PMC6609220.
- Zhang, K., Li, S., Kappel, K. et al. Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. Nat Commun 10, 5511 (2019). https://doi.org/10.1038/s41467-019-13494-7
- Zhao, Y., Rai, J., Yu, H., & Li, H. (2021). *Pseudouridine-free ribosome exhibits distinct inter-subunit movements*. bioRxiv. https://doi.org/10.1101/2021.06.02.446812
- Zhu H, Yin X, Holley CL, Meyer KD. Improved Methods for Deamination-Based m6A Detection. Front Cell Dev Biol. 2022 Apr 27;10:888279. doi: 10.3389/fcell.2022.888279. PMID: 35573664; PMCID: PMC9092492.
- Zhuang, X., et al. (2000). A Single-Molecule Study of RNA Catalysis and Folding. *Science.*
- Zimmer, C., S. Tymen, C. March and W. Guschlbauer (1982) Conformational transitions of poly(dA-dC).poly(dG-dT) induced by high salt or in ethanolic solution. Nucleic Acids Res. 10, 1081-1091
- Zuber J, Schroeder SJ, Sun H, Turner DH, Mathews DH. Nearest neighbor rules for RNA helix folding thermodynamics: improved end effects. Nucleic Acids Res. 2022 May 20;50(9):5251-5262. doi: 10.1093/nar/gkac261. PMID: 35524574; PMCID: PMC9122537.