bioRχiv

ModrRNA1: A tool for rapid general identification of possible RNA modification sites in bacterial rRNA and prediction of the associated putative enzymes

D. W. U. Thilanka,[a,*] P. S. Warushavithana,[a]
[a]Department of Chemistry, University of Colombo
*Correspondence should be addressed to (uviwick@gmail.com)

Abstract

Background

Ribosomal RNA modifications are pivotal in bacteria. A method for locating rRNA modification sites and associated enzymes is crucial for understanding ribosome formation, bacterial functionality, and antibiotic resistance correlated with rRNA modifications. ModrRNA1 introduces a bioinformatics approach to streamline this identification process by replacing manual methods and combining all the required computational steps in a user-friendly web tool. It may aid in locating potential uncharacterized modification sites and previously unknown RNA modification enzymes across various bacteria, enhancing the comprehension of their biological importance.

Availability and implementation

ModrRNA1 employs a curated database of known bacterial rRNA modification sequences and associated enzymes. Sequence alignment, with a customized scoring system, enables accurate general identification of modification sites. High-confidence alignments generate annotated plots over the RNA sequence. Subsequent enzyme detection involves structural comparisons, phylogenetic assessments, and BLAST searches, aiding in enzyme identification. The tool boasts a user-friendly web interface and is hosted on Azure, offering accessible rRNA modification analysis. It is available as a web tool on https://modrrna.biotechie.org and the source code is freely available at https://github.com/sciguysl/ModrRNA1 for any customization purposes.

Conclusion

The results demonstrate the effectiveness of ModrRNA1 in identifying potential non-species-specific rRNA modification sites in different bacteria and their associated enzymes specific to the query bacteria. It can be utilized to identify uncharacterized rRNA modification sites and undiscovered proteins with RNA modifying activity.

Introduction

Nucleotide modifications of ribosomal RNA (rRNA) can occur post-transcriptionally through three main pathways—namely, conversion of uridine to pseudouridine (5-ribosyl uracil), methylation of 2' hydroxyls, and base alteration; most of which are followed by methylation at various positions [1]. The distribution of all Escherichia coli rRNA modifications in the ribosome has been identified and localized [1]. Modifications mostly occur in the interior of the RNA mass of the ribosomes, oriented towards the faces of the ribosomal subunits, and they are distributed mostly in conserved regions of both the large and small subunits. Hence, it is apparent that modifications affect not only the structure but also the ribosomal functions [1]. In contrast, since modified nucleotides are absent from regions where ribosomal proteins are abundant, modifications do not directly impact RNA-protein interactions.

rRNA modifications commonly seen in *E. coli* are carried out by site-specific or region-specific protein-only enzymes. Almost all rRNA modifications of the 70S ribosome of *E. coli* are known and there are 36 modified nucleotides in total [2]. The number of methylated and pseudouridylated nucleotides found in 16S rRNA of *E. coli* are 10 and 1, respectively. Modified nucleotides of 23S rRNA in *E. coli* are 13 methylated ones, 9 pseudouridines, 1 methyl pseudouridine, 1 dihydrouridine, and 1 hydroxycytidine. Positions of modified nucleotides in the *E. coli* 16S and 23S rRNA and the enzymes responsible for their formation are shown in Table 1 and 2 [3].

In bacteria, the most prominent type of nucleotide modification is methylation. Additionally, most methylated nucleotides are found exposed on the surface of ligand-free small subunits to be readily accessible to methyltransferases. Some enzymes need ribosomal proteins to carry out modifications while other enzymes only modify while rRNA is protein free. In vitro, sub-ribosomal particle analysis has shown that most methylations occur during the early or late steps of ribosome assembly [4]. Various Pseudouridine synthases carry out pseudouridylation.

Pseudouridylation leads to increased structural rigidity of the sugar-phosphate backbone, alteration of base stacking

interactions [5], steric properties, and different H-bonding patterns [1]. Both pseudouridylation and methylation of 2' hydroxyls cause enhanced structural rigidity for single-stranded and double-stranded regions [1].

Table 1: Modified nucleotides of 16S rRNA of *E. coli* and associated modifying enzymes.

Position in the 16S rRNA	Modified nucleoside	Modification enzyme
516	Ψ	RsuA
527	m^7G	RsmG, GidB
966	m^2G	RsmD, YhhF
967	m^5C	RsmB, YhdB
1207	m^2G	RsmC, YjjT
1402	m^4Cm	RsmH, RsmL
1407	m^5C	RsmF, YebU
1498	m^3U	RsmE, YggJ
1516	m^2G	RsmJ, YhiQ
1518	$m^6{}_2A$	RsmA, KsgA
1519	$m^6{}_2A$	RsmA, KsgA

Table 2: Modified nucleotides of 23S rRNA of *E. coli* and associated modifying enzymes.

Position in the 23S rRNA	Modified nucleoside	Modification enzyme
745	m^1G	RrmA, YebH
746	Ψ	RluA
747	m^5U	YbjF, RumB
955	Ψ	RluC
1618	m^6A	YbiN
1835	m^2G	RlmG, YgjO
1911	Ψ	RluD
1915	$m^3Ψ$	RluD, YbeA
1917	Ψ	RluD
1939	m^5U	RlmD
1962	m^5C	RlmL
2030	m^6A	YhiR, RlmJ
2069	m^7G	RlmKL
2251	Gm	RlmB
2445	m^2G	RlmL
2449	hU	
2457	Ψ	RluE
2498	Cm	RlmM
2501	oh^5C	
2503	m^2A	RlmN
2504	Ψ	RluC
2552	Um	RlmE, RrmJ
2580	Ψ	RluC
2604	Ψ	RluF
2605	Ψ	RluB

Methylation of 2′-OH prevents hydrolysis of inter-nucleotide bonds and could influence base stacking [3]. Such effects on the chemical properties in turn help RNA folding by resolving misfolded structures or by preventing their formation and change the assembly and activity of the ribosome. Methyltransferase-mediated rRNA modifications play a vital role in the generation of functional ribosomes. A unique modified cytosine found in almost all bacterial 16S rRNA which is methylated at the $m^4Cm1402$, is important for P-site structure formation and enhances the accuracy of initiation codon recognition [6]. Another highly conserved modification is the dimethylation of two adenosines carried out by Methyltransferase KsgA. X-ray crystal structure of the *Thermus thermophilus* 30S ribosomal subunit shows that defective KsgA methyltransferase results in the absence of methylation, which causes translational mistakes in the initiation and elongation stages of protein synthesis and confers resistance to the antibiotic kasugamycin, due to a conformation change that affect the antibiotic binding [7]. Moreover, mutations in the gene that codes for methyltransferase RsmG lead to streptomycin resistance [8]. Therefore, it is likely that the mechanisms of a significant number of antibiotic resistances arise from rRNA methylations at specific nucleotides.

Deletions of *rsmB* and *rsmD* genes have been shown to cause disruption of transcription, and misregulation of attenuation control of tryptophan operon in *E. coli* [9]. Mutations in methyltransferase genes have also been proven to cause moderate to drastic cell growth retardation respectively in the cases of methyltransferase RlmF [10] and RrmJ [11]. Methyltransferase orthologs may possess functions in addition to their rRNA methyltransferase activity such as acting as RNA chaperones, as seen in KsgA. This enzyme also plays a role in the suppression of coldsensitivity and may be involved in the acid shock response of *E. coli* [12].

In summary, these rRNA modifications play a pivotal role in bacteria. Therefore, A tool for general identification of modification sites and the enzymes associated with each modification can be important in studying the formation and functionality of a ribosome of a given bacteria species, and in gaining clinically relevant insight into what confers antibiotic resistance in bacteria by better understanding rRNA modifications. The precise localization of these sites and the identification of potential enzymes are essential for deciphering their biological significance. Existing tools for this purpose are lacking, leading researchers to rely on manual methods. ModrRNA1 integrates tasks from the general identification of potential rRNA modification sites to shortlisting putative enzymes and performing BLAST analysis [13] on the query bacterial genome, thereby streamlining and simplifying the process for researchers in a user-friendly manner. This tool can be utilized in finding uncharacterized probable rRNA modification sites and previously unacknowledged RNA modification enzymes in a variety of bacteria and to find proteins that have RNA modification functionality.

Method

Database Preparation

ModrRNA1 is an open-source web application developed in Python 3. It utilizes a database of bacterial rRNA modifications, which consists of conserved sequences located both

downstream and upstream of each modification site, along with the details on the respective modification enzymes. This database was curated by collating data from the *RMBase* [14] bacterial RNA modification site datasets and cross-referencing it with information from *MODOMICS* [15] for annotation with the associated enzymes. In addition, a protein structure library was created using *ESMFold* [16] for all the enzymes involved in these reference modifications. (Figure 1)

Sequence Alignment

In ModrRNA1, firstly the modification site sequences are duplicated and transformed to their reverse complements using the *Bio.Seq* module in *Biopython* [17]. Then it aligns these sequences to the input query sequence using the *Bio.pairwise2* module. We customized the scoring function of the aligner to give 2 points for an identical character, -1 point for a non-identical character, -4 points for a gap opening, and -1 point for a gap extension to minimize gap formation when aligning. An accuracy score is calculated as a percentage to assess the quality of the alignment using the simple equation below.

$$Accuracy\ score = \frac{Alignment\ score\ \times 100}{Length\ of\ the\ site\ sequence\ \times\ 2}$$

Site Annotation

The tool removes redundant sequences and obtains alignments of high confidence by applying a threshold of an accuracy score greater than or equal to 60. Based on this data, it generates a single-axis plot with the query sequence on the x-axis using the *DnaFeaturesViewer* library [18]. The detected RNA modifications are annotated visually and are distinguishable at their respective sites on the plot. These annotations include the type of modification, nucleotide location, and the unique identifier used to find the enzyme(s) in subsequent steps. The *Bio.SeqIO* package, in conjunction with the *Bio.SeqFeature* module is employed to tag these annotations and output them into GenBank and CSV format, which can be downloaded by the user and utilized in external analyses.

Enzyme Detection

If the source bacterium of the query sequence is known, ModrRNA1 allows the user to find the most likely enzymes involved in a particular modification. The specific enzyme structures are screened using the *Protein Data Bank (PDB)* [19] and virtual protein structure databases (*AlphaFold*) [20] to screen structurally similar enzymes, employing the *FoldSeek* API [21]. Thresholds greater than or equal to 50% of query coverage ($50 \leq$ qscore) and $60 \leq$ TM-score normalized by the query length (prob) are applied to filter and shortlist the results.

The evolutionary distance is calculated based on the topology (branching structure of the tree) between the query bacterium and the source bacteria of these enzymes using the *NCBITaxa* module in the *ete3* library [22]. The most phylogenetically related enzymes are selected and sorted by comparing the evolutionary distance. The results are outputted as entries to the Protein Data Bank (PDB) and the AlphaFold Protein Structure Database (AFDB) [20]. The user can compare these calculated distances, query coverage scores, and TM-scores (prob) to locate probable enzymes and then perform a *blastp* or a *tblastn* analysis [13] within this tool itself on the query bacterial genome. In cases where high-scoring alignments are found in the BLAST search, it suggests the presence of the specific rRNA modification-associated enzyme and the possible modification of the respective site in the queried bacteria. (Figure 1)

Web Application

Streamlit, which is an open-source Python-based application framework was used to create the web application. A user-friendly UI with clear instructions was implemented, ensuring accessibility to a broad user base. It is hosted in an Azure app service instance in Linux with Python 3.8 version.

Testing

The tool was tested to assess its operational functionality, employing 16S ribosomal RNA (rRNA) sequences derived from *Escherichia sp.* strain Esraa 4 (MT647245.1) and *Thermus thermophilus* HB8 (NR_037066.1). The potential modification sites identified in ModrRNA1 were compared with known rRNA modification sites obtained from MODOMICS. As the positions are formatted in *E. coli* numbering, the *T. thermophilus* 16S rRNA annotated Genbank file obtained from ModrRNA1 was aligned with the *E. coli* 16S rRNA sequence using Benchling to locate the positions.

For tool evaluation, it was applied to 23S rRNA sequences derived from organisms significantly divergent from the source bacteria from which the initial conserved site sequences were obtained. The query sources were *Deinococcus radiodurans* strain R1 (NR_076078.2) and *Haloarcula marismortui* ATCC 43049 (NR_076318.1). Since the positions are *E. coli*-numbered, the query sequences were aligned with the 23S rRNA of *E. coli*. The results were tabulated in Table 4. To evaluate the Enzyme Explorer in ModrRNA1, six randomly selected potential modification sites in *T. Thermophilus* were processed, specifically: Ψ (499), $m^2{}_2G$ (943), m^5C (1378), $m^6{}_2A$ (1491), m^4Cm (1380), and m^3U (1471). A blastp search was done using the top match which was selected out of FoldSeek shortlisted hits based on evolutionary distance. The BLAST alignment yielding the highest score was postulated as the putative enzyme, and the results for each site were tabulated in Table 5.

Results and discussion

The results in Table 3 indicated that all the known sites are identified in the tool as intended. There were several extra potential modification sites identified by the tool as in this step the code aligns all the conserved sequences from the site database. These sites may or may not represent actual modification sites, and their confirmation may require subsequent enzyme search steps.

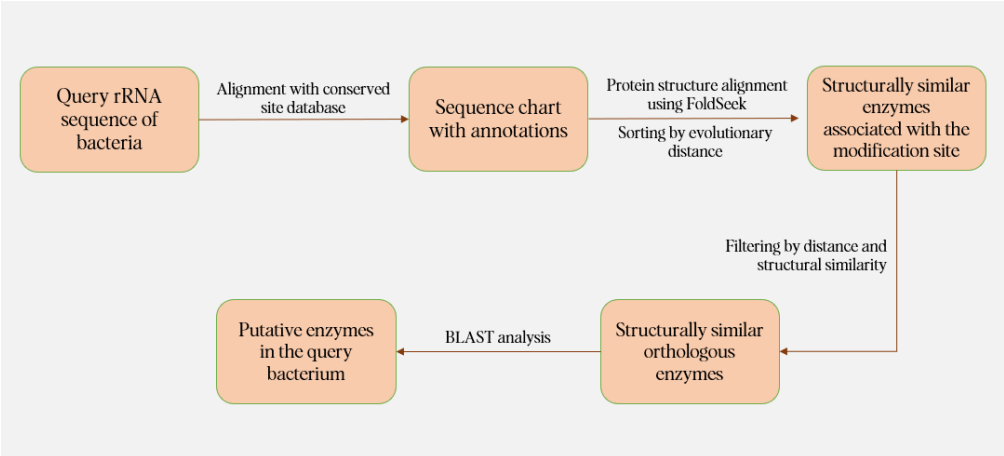

Figure 1: A flowchart depicting the workflow of ModrRNA1.

In the evaluation results presented in Table 4, we compared the modifications detected by the tool with those reported in a previous study. All pseudouridylation sites were successfully identified, with the exception of the one located at position 2586 (*E. coli* numbering). This discrepancy might be attributed to variations in the query sequence of *Haloarcula marismortui* used during the evaluation, as the exact 23S rRNA sequence employed had not been disclosed in this study. It should be noted that the absence of certain missing sites may be due to their non-inclusion of the specific conserved site sequences in the ModrRNA database. This database exclusively comprises sequences that have been confirmed through multiple studies. Therefore, if a new modification site is identified in a single study, it may not necessarily appear as a potential modification site in this tool. To address this limitation, our objective is to maintain an up-to-date database by continuously incorporating newly confirmed (through more than one study) site sequences.

Table 1: Known 16S rRNA modification sites and ModrRNA1 fetched sites comparison to assess functionality.
Here the known modification sites from *E. coli* and *T. Thermophilus* is compared with potential modification sites located by the tool. Modification positions in the first column are in *E. coli* numbering.

Modification position	Modification (known) in *E. coli*	Modification (ModrRNA1) in *E. coli*	Modification (known) in *T. Thermophilus*	Modification (ModrRNA1) in *T. Thermophilus*
516	Ψ	Ψ	Ψ	Ψ (499)
527	m^7G	m^7G	m^7G	m^7G (510)
966	m^2G	m^2G	$m^2{}_2G$	$m^2{}_2G$ (943)
967	m^5C	m^5C	m^5C	m^5C (944)
1207	m^2G	m^2G	m^2G	m^2G (1186)
1400	-	-	m^5C	m^5C (1378)
1402	m^4Cm	m^4Cm	m^4Cm	m^4Cm (1380)
1404	-	-	m^5C	m^5C (1382)
1407	m^5C	m^5C	m^5C	m^5C (1385)
1498	m^3U	m^3U	m^3U	m^3U (1471)
1516	m^2G	m^2G	-	-
1518	$m^6{}_2A$	$m^6{}_2A$	$m^6{}_2A$	$m^6{}_2A$ (1491)
1519	$m^6{}_2A$	$m^6{}_2A$	$m^6{}_2A$	$m^6{}_2A$ (1492)
1540	-	-	Ψ	Ψ (1513)
1541	-	-	Ψ	Ψ (1514)

Table 4: Known 16S rRNA modification sites and ModrRNA1 fetched sites comparison for evaluation.
Here the known modification sites from *Haloarcula marismortui* and *Deinococcus radiodurans* is compared with potential modification sites located by the tool. Modification positions in the second column are in *E. coli* numbering.

Organism	Modification position	Modification (known)	Modification identified by ModrRNA1
Haloarcula marismortui	1915	Ψ	Ψ (1955)
	1917	Ψ	Ψ (1957)
	2586	Ψ	-
Deinococcus radiodurans	1911	Ψ	Ψ (1891)
	1915	$m^3Ψ$	$m^3Ψ$ (1895)
	1917	Ψ	Ψ (1897)
	2605	Ψ	Ψ (2581)

In the enzyme evaluation presented in Table 5, it is clear that the putative enzymes identified exhibit a consistent correlation with the associated modification types, as evidenced by the comparisons with Table 1 and Table 2 and further literature. The methyltransferase RsmF of *T. thermophilus* has been identified by multiple studies as responsible for multi-site-specific modifications: m^5C1407 (1385), m^5C1400 (1378), and m^5C1404 (1382) [23], [24]. Methyltransferase RsmH which is responsible for the N^4-methylation of C1402 (1380) in *E. coli*[25] 16S rRNA and RsmH is also conserved in *T. Thermophilus* [26]. Dimethyltransferase RsmA (KsgA) carries out the most highly conserved rRNA modifications; A1518 (1491) and A1519 (1492) of 16S rRNA [27]. One study found the crystal structure of *Thermus thermophilus* hypothetical RNA 2'-O-ribose methyltransferase that seems corresponds with m^3U (1471) modification [28].

Table 2: ModrRNA1 BLAST results of several modification sites in *T. Thermophilus* 16S rRNA

Modification (ModrRNA1) in *T. Thermophilus*	Putative enzyme
Ψ (499)	Ribosomal small subunit pseudouridine synthase A (NCBI: WP_124105079.1)
$m^2{}_2G$ (943)	RsmD family RNA methyltransferase [Thermus] (NCBI: WP_024119826.1)
m^5C (1378)	16S rRNA (cytosine(1407)-C(5))-methyltransferase RsmF [Thermus] (NCBI: WP_011228642.1)
$m^6{}_2A$ (1491)	16S rRNA (adenine(1518)-N(6)/adenine(1519)-N(6))-dimethyltransferase RsmA [Thermus] (WP_011227698.1)
m^4Cm (1380)	16S rRNA (cytosine(1402)-N(4))-methyltransferase RsmH [Thermus] (NCBI: WP_011228422.1)
m^3U (1471)	RNA 2′-O-RIBOSE METHYLTRANSFERASE [Thermus thermophilus] (NCBI: 1IPA_A)

The ability of ModrRNA1 to find potential modification sites and associated enzymes within bacterial rRNA sequences as demonstrated through the aforementioned results has wide-reaching implications. It can streamline the preliminary bioinformatics steps involved in rRNA modification enzyme research, providing a user-friendly platform for researchers to explore uncharacterized modification sites and enzymes across diverse bacteria, eliminating the need for manual methods that researchers usually rely upon. The open-source codebase, available on GitHub without restrictions, can be utilized to develop novel tools built upon this pipeline. These tools can be designed to target enzymes that operate on conserved sequences, similar to the methodology employed in this study.

Future directions involve expanding this tool to organisms beyond bacteria and incorporating other RNA types. Additionally, the database of conserved modification sequences is to be improved continuously.

Conclusion

ModrRNA1 appears to be effective in identifying potential rRNA modification sites in a non-species-specific manner and associated enzymes specific to the query bacteria. It streamlines bioinformatics processes, offering a user-friendly alternative to manual methods. This tool may aid in research that advances our understanding of rRNA modifications and their functional implications in various biological contexts.

References

[1] W. A. Decatur and M. J. Fournier, 'rRNA modifications and ribosome function', *Trends Biochem Sci*, vol. 27, no. 7, pp. 344–351, 2002, doi: 10.1016/S0968-0004(02)02109-6.

[2] S. A. Woodson, 'RNA folding and ribosome assembly', *Curr Opin Chem Biol*, vol. 12, no. 6, pp. 667–673, Dec. 2008, doi: 10.1016/J.CBPA.2008.09.024.

[3] O. V. Sergeeva, A. A. Bogdanov, and P. V. Sergiev, 'What do we know about ribosomal RNA methylation in Escherichia coli?', *Biochimie*, vol. 117, no. December, pp. 110–118, 2015, doi: 10.1016/j.biochi.2014.11.019.

[4] T. Siibak and J. Remme, 'Subribosomal particle analysis reveals the stages of bacterial ribosome assembly at which rRNA nucleotides are modified', doi: 10.1261/rna.2160010.

[5] M. Mueller, 'Critical review', *Critical Review*, vol. 4, no. 1–2, p. ebi, 1990, doi: 10.1080/08913819008459588.

[6] S. Kimura and T. Suzuki, 'Fine-tuning of the ribosomal decoding center by conserved methyl-modifications in the Escherichia coli 16S rRNA', *Nucleic Acids Res*, vol. 38, no. 4, pp. 1341–1352, Mar. 2010, doi: 10.1093/NAR/GKP1073.

[7] H. Demirci *et al.*, 'Modification of 16S ribosomal RNA by the KsgA methyltransferase restructures the 30S subunit to optimize ribosome function', *RNA*, vol. 16, no. 12, pp. 2319–2324, Dec. 2010, doi: 10.1261/RNA.2357210.

[8] S. Okamoto *et al.*, 'Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria', *Mol Microbiol*, vol. 63, no. 4, pp. 1096–1106, Feb. 2007, doi: 10.1111/J.1365-2958.2006.05585.X.

[9] I. V. Prokhorova *et al.*, 'Modified nucleotides m2G966/m5C967 of Escherichia coli 16S rRNA are required for attenuation of tryptophan operon', *Scientific Reports 2013 3:1*, vol. 3, no. 1, pp. 1–6, Nov. 2013, doi: 10.1038/srep03236.

[10] P. V. Sergiev, M. V. Serebryakova, A. A. Bogdanov, and O. A. Dontsova, '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*, vol. 375, no. 1, pp. 291–300, Jan. 2008, doi: 10.1016/J.JMB.2007.10.051.

[11] J. Tan, U. Jakob, and J. C. A. Bardwell, 'Overexpression of two different GTPases rescues a null mutation in a heat-induced rRNA methyltransferase', *J Bacteriol*, vol. 184, no. 10, pp. 2692–2698, 2002, doi: 10.1128/JB.184.10.2692-2698.2002/ASSET/4E724E1B-C65A-4993-9ECA-639640E99667/ASSETS/GRAPHIC/JB1021051006.JPEG.

[12] K. Inoue, S. Basu, and M. Inouye, 'Dissection of 16S rRNA methyltransferase (KsgA) function in Escherichia coli', *J Bacteriol*, vol. 189, no. 23, pp. 8510–8518, Dec. 2007, doi: 10.1128/JB.01259-07/ASSET/98CDF2E6-D79B-47B3-94C5-47C81A8999FB/ASSETS/GRAPHIC/ZJB0230773230008.JPEG.

[13] S. McGinnis and T. L. Madden, 'BLAST: at the core of a powerful and diverse set of sequence analysis tools', *Nucleic Acids Res*, vol. 32, no. Web Server issue, p. W20, Jul. 2004, doi: 10.1093/NAR/GKH435.

[14] J.-J. Xuan *et al.*, 'RMBase v2.0: deciphering the map of RNA modifications from epitranscriptome sequencing data', *Nucleic Acids Res*, vol. 46, pp. 327–334, 2017, doi: 10.1093/nar/gkx934.

[15] P. Boccaletto *et al.*, 'MODOMICS: a database of RNA modification pathways. 2021 update', *Nucleic Acids Res*, vol. 50, no. D1, pp. D231–D235, Jan. 2022, doi: 10.1093/NAR/GKAB1083.

[16] Z. Lin *et al.*, 'Evolutionary-scale prediction of atomic-level protein structure with a language model', *Science (1979)*, vol. 379, no. 6637, pp. 1123–1130, Mar. 2023, doi: 10.1126/SCIENCE.ADE2574/SUPPL_FILE/SCIENCE.ADE2574_SM.PDF.

[17] P. J. A. Cock *et al.*, 'Biopython: freely available Python tools for computational molecular biology and bioinformatics', *Bioinformatics*, vol. 25, no. 11, p. 1422, Jun. 2009, doi: 10.1093/BIOINFORMATICS/BTP163.

[18] V. Zulkower and S. Rosser, 'DNA Features Viewer: a sequence annotation formatting and plotting library for Python', *Bioinformatics*, vol. 36, no. 15, pp. 4350–4352, Aug. 2020, doi: 10.1093/BIOINFORMATICS/BTAA213.

[19] H. M. Berman *et al.*, 'The Protein Data Bank', *Nucleic Acids Res*, vol. 28, no. 1, p. 235, Jan. 2000, doi: 10.1093/NAR/28.1.235.

[20] M. Varadi *et al.*, 'AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with

high-accuracy models', *Nucleic Acids Res*, vol. 50, no. D1, pp. D439–D444, Jan. 2022, doi: 10.1093/NAR/GKAB1061.

[21] M. van Kempen *et al.*, 'Fast and accurate protein structure search with Foldseek', *Nature Biotechnology 2023*, pp. 1–4, May 2023, doi: 10.1038/s41587-023-01773-0.

[22] J. Huerta-Cepas, F. Serra, and P. Bork, 'ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data', *Mol Biol Evol*, vol. 33, no. 6, p. 1635, Jun. 2016, doi: 10.1093/MOLBEV/MSW046.

[23] R. Guymon, S. C. Pomerantz, P. F. Crain, and J. A. McCloskey, 'Influence of phylogeny on posttranscriptional modification of rRNA in thermophilic prokaryotes: The complete modification map of 16S rRNA of Thermus thermophilus', *Biochemistry*, vol. 45, no. 15, pp. 4888–4899, Apr. 2006, doi: 10.1021/BI052579P/SUPPL_FILE/BI052579PSI20060322_014224.PDF.

[24] H. Demirci *et al.*, 'Multi-site-specific 16S rRNA methyltransferase RsmF from Thermus thermophilus', *Rna*, vol. 16, no. 8, pp. 1584–1596, 2010, doi: 10.1261/rna.2088310.

[25] S. Kimura and T. Suzuki, 'Fine-tuning of the ribosomal decoding center by conserved methyl-modifications in the Escherichia coli 16S rRNA', *Nucleic Acids Res*, vol. 38, no. 4, pp. 1341–1352, 2009, doi: 10.1093/nar/gkp1073.

[26] Y. Wei *et al.*, 'Crystal and solution structures of methyltransferase RsmH provide basis for methylation of C1402 in 16S rRNA', *J Struct Biol*, vol. 179, no. 1, pp. 29–40, 2012, doi: 10.1016/j.jsb.2012.04.011.

[27] G. Jogl *et al.*, 'Structure of the Thermus thermophilus 16 S rRNA methyltransferase KsgA', *The FASEB Journal*, vol. 23, no. S1, pp. 839.2-839.2, Apr. 2009, doi: 10.1096/FASEBJ.23.1_SUPPLEMENT.839.2.

[28] O. Nureki *et al.*, 'An enzyme with a deep trefoil knot for the active-site architecture', *Acta Crystallogr D Biol Crystallogr*, vol. 58, no. Pt 7, pp. 1129–1137, 2002, doi: 10.1107/S0907444902006601.

Competing interest statement

The authors declare that they have no competing financial or any other interests that could influence the work reported in this paper.