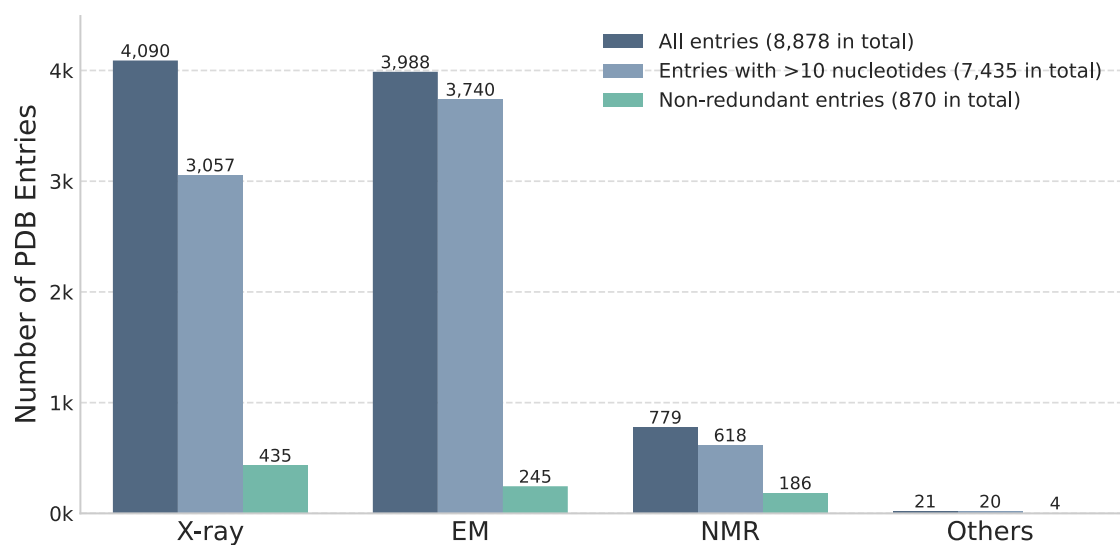

Integrated experimental and AI innovations for RNA structure determination

In the format provided by the
authors and unedited

Supplementary Information



Supplementary Fig. 1. Breakdown of RNA structures in the PDB by determination method (data as of July 2025). The non-redundant entries are derived by removing sequences with over 80% identity using the CD-HIT-EST program ¹, which by default excludes short fragments (≤ 10 nucleotides) for improved cluster stability.

Supplementary Table 1. Protein-free RNA structures solved by cryo-EM at sub-3.0 Å resolution and deposited before July 2025. The table includes 42 total entries, many of which correspond to the same RNAs (e.g., the *Tetrahymena* ribozyme and group IIC intron). After applying an 80% sequence identity cutoff, 20 non-redundant entries remain.

Entry ID	Structure Title	Resolution (Å)	Residues	Release Date
7UR5	allo-tRNA ^U Tu1 in the A, P, and E sites of the E. coli ribosome	2.6	270	2022/8/10
7URI	allo-tRNA ^U Tu1A in the A site of the E. coli ribosome	2.6	90	2022/8/10
7URM	allo-tRNA ^U Tu1A in the P site of the E. coli ribosome	3	90	2022/8/10
7XD5	The Tet-S2 state of wild-type <i>Tetrahymena</i> group I intron with 30nt 3'/5'-exon	2.84	406	2023/4/5
7XD6	The Tet-S2 state with a pseudoknotted 4-way junction of wild-type <i>Tetrahymena</i> group I intron with 30nt 3'/5'-exon	2.84	459	2023/4/5
7YCI	Cryo-EM structure of <i>Tetrahymena</i> ribozyme conformation 4 undergoing the first-step self-splicing	2.98	399	2023/7/5
7YG8	Cryo-EM structure of <i>Tetrahymena</i> ribozyme conformation 5 undergoing the second-step self-splicing	2.97	402	2023/3/29
7YG9	Cryo-EM structure of <i>Tetrahymena</i> ribozyme conformation 1 undergoing the second-step self-splicing	2.68	407	2023/3/29
7YGA	Cryo-EM structure of <i>Tetrahymena</i> ribozyme conformation 2 undergoing the second-step self-splicing	2.35	407	2023/3/29
7YGB	Cryo-EM structure of <i>Tetrahymena</i> ribozyme conformation 3 undergoing the second-step self-splicing	2.62	407	2023/3/29
7YGC	Cryo-EM structure of <i>Tetrahymena</i> ribozyme conformation 4 undergoing the second-step self-splicing	2.65	407	2023/3/29
8I7N	The Tet-S1 state of G264A mutated <i>Tetrahymena</i> group I intron with 6nt 3'/5'-exon and 2-aminopurine nucleoside	2.98	418	2023/3/29
8K0P	The <i>Anoxybacillus pushchinoensis</i> ORF-less Group IIC Intron HYER1 at symmetric apo state	2.99	1266	2024/5/15
8K15	The <i>Streptococcus azizii</i> ORF-less Group IIC intron HYER2 at apo state	2.58	621	2024/5/15
8SA3	Adenosylcobalamin-bound riboswitch dimer, form 2	3	420	2023/7/26
8T5O	Cryo-EM structure of RNA device 43, holo state	2.98	124	2025/2/19
8TJV	<i>Tetrahymena</i> Ribozyme scaffolded Fluoride riboswitch	2.99	417	2023/11/1
8TJX	<i>Tetrahymena</i> Ribozyme cryo-EM scaffold	2.44	440	2023/11/1
8UPT	<i>Candidatus Methanomethylophilus alvus</i> tRNA ^{Pyl} in A-site of ribosome	2.8	71	2024/1/10
8UPY	<i>Methanosarcine mazei</i> tRNA ^{Pyl} in A-site of ribosome	2.8	72	2024/9/4
8X9K	Cte_Branching,Ca,inactive	2.63	808	2025/3/12

Entry ID	Structure Title	Resolution (Å)	Residues	Release Date
8X9M	Cte-Branch Product, Mg, post	2.78	808	2025/3/12
8XTP	Comamonas testosteroni KF-1 circularly permuted group II intron Post-2S state	2.9	728	2025/2/12
8XTQ	Comamonas testosteroni KF-1 circularly permuted group II intron 1S state	2.96	808	2025/2/12
8XTR	Comamonas testosteroni KF-1 circularly permuted group II intron 2S state	2.67	814	2025/2/12
8XTS	Comamonas testosteroni KF-1 circularly permuted group II intron Pre-1S state	2.54	814	2025/2/12
9C6I	Cryo-EM structure of a group IIC intron RNA	2.56	417	2024/12/4
9C6J	Group IIC intron embedded with the TPP riboswitch	2.41	495	2024/12/4
9C6K	Cryo-EM structure of the TPP riboswitch embedded in an RNA scaffold bound to thiamine pyrophosphate	2.94	497	2024/12/4
9CBU	Tetrahymena ribozyme with consensus water and magnesium ions	2.2	387	2024/11/20
9CBW	Tetrahymena ribozyme with consensus water and magnesium ions	2.3	387	2024/11/20
9CBX	Tetrahymena ribozyme with automatically identified water and magnesium ions	2.2	387	2024/11/20
9CBY	Tetrahymena ribozyme with automatically identified water and magnesium ions	2.3	387	2024/11/20
9ELY	raiA RNA motif (Clostridium acetobutylicum)	2.9	205	2025/4/23
9IS7	Paracandidimonas lactea CP group II intron 2S state	2.87	735	2025/2/12
9LCR	Clostridium botulinum OLE RNA dimer	2.6	1156	2025/3/19
9MCW	OLE RNA dimer (Clostridium acetobutylicum)	2.9	1154	2025/4/23
9MEE	GOLLD RNA nanocage (env-38)	3	11662	2025/4/30
9MM6	ROOLefa-octamer-wild type	2.94	4640	2025/6/4
9MME	ROOLfirm-octamer-wild type	2.93	4640	2025/6/4
9MQT	CryoEM Structure of the Candida albicans Group I Intron-Compound 11 Complex under Magnesium Condition	2.43	332	2025/4/30
9MQU	CryoEM Structure of the Candida albicans Group I Intron-Compound 11 Complex under Calcium Condition	2.59	332	2025/4/30

References

- (1) Li, W.; Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658-1659.