frontiers in
GENETICS

ORIGINAL RESEARCH ARTICLE
published: 18 April 2012
doi: 10.3389/fgene.2012.00059

HD-RNAS: an automated hierarchical database of RNA structures

Shubhra Sankar Ray[1]†, Sukanya Halder[2]†, Stephanie Kaypee[2] and Dhananjay Bhattacharyya[2]*

[1] Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India
[2] Biophysics Division, Saha Institute of Nuclear Physics, Kolkata, India

Edited by:
Fengfeng Zhou, Shenzhen Institutes of Advanced Technology, China

Reviewed by:
Peng Jiang, University of Iowa, USA
Gaurav Sablok, Huazhong Agricultural University, China
Ao Li, University of Science and Technology of China, China
Yi Zhao, Institute of Computing Technology,C.A.S., China
Bin Song, Oracle, USA

Correspondence:
Dhananjay Bhattacharyya, Biophysics Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700 064, India.
e-mail: dhananjay.bhattacharyya@saha.ac.in

†They contributed equally and can be considered as joint first authors.

One of the important goals of most biological investigations is to classify and organize the experimental findings so that they are readily useful for deriving generalized rules. Although there is a huge amount of information on RNA structures in PDB, there are redundant files, ambiguous synthetic sequences etc. Moreover, a systematic hierarchical organization, reflecting RNA classification, is missing in PDB. In this investigation, we have classified all the available RNA structures from PDB through a programmatic approach. Hence, it would be now a simple assignment to regularly update the classification as and when new structures are released. The classification can further determine (i) a non-redundant set of RNA structures and (ii) if available, a set of structures of identical sequence and function, which can highlight structural polymorphism, ligand-induced conformational alterations etc. Presently, we have classified the available structures (2095 PDB entries having RNA chain longer than nine nucleotides solved by X-ray crystallography or NMR spectroscopy) into nine functional classes. The structures of same function and same source are mostly seen to be similar with subtle differences depending on their functional complexation. The web-server is available online at http://www.saha.ac.in/biop/www/HD-RNAS.html and is updated regularly.

Keywords: RNA classification, RNA crystal structures, RNA database, functional RNA, structure prediction, functional annotation

INTRODUCTION

Keeping pace with advancement in the field of RNA functions, the number of RNA structures whose coordinates are available in the Protein Data Bank (PDB; Berman et al., 2000) is growing rapidly. The total number of structures of RNA with oligomeric or polymeric length, as available in July 2011, is 2095 and the number is increasing at a pace of about 100 per year. The determination of various RNA structures, such as the hammerhead ribozyme (Scott et al., 1995), SRP-RNA (Zwieb et al., 2005), and the 5S, 16S and 23S RNAs of ribosome has greatly increased our knowledge of RNA folds and the three-dimensional organization of RNA chains (Batey et al., 1999; Ferre-d'Amare and Doudna, 1999; Hermann and Patel, 1999). Collectively, these structures provide a large amount of information about RNA structural motifs (Moore, 1999). Similar exponential growth in number of crystal structures of proteins is also taking place in the PDB. Considering the need of classification of these proteins, there are a number of methods available, such as SCOP (Murzin et al., 1995; Hubbard et al., 1997), FSSP (Holm and Sander, 1997), Pisces (Wang and Dunbrack, 2005), BIPA (Lee and Blundell, 2009) etc. These methods can classify a protein structure based on its structural class, source organism, secondary structure content, resolution, etc. One can further determine a set of non-redundant structures of proteins, which are not evolutionarily related, for a statistical analysis in an unbiased method. In a similar manner, it is also necessary to organize the available RNA structures to determine different structure–function relationships. Furthermore, it is often important to compare several structures of RNA of same function and from same source, which have identical sequence, to understand effect of ligand binding, crystallization environments etc., on the three-dimensional folding. Such sets of structures could reveal significant information on structural flexibility, binding thermodynamics etc., of the biological macromolecules (Halder and Bhattacharyya, 2010; Samanta et al., 2010). They carry signatures that may indicate variations introduced in the molecular structure due to ligand binding or alteration of crystallization conditions. In our recent study, we also found that structural variability of double-helical RNA as observed in molecular dynamics simulation studies mimic that of the crystallographic ensembles (Halder and Bhattacharyya, 2010; Halder and Bhattacharyya, manuscript in preparation). Likewise, the differences in RNA structural organization among various species can be studied if a classification is available. A non-redundant set of RNA structures is also necessary to analyze the local environments at basepairing level, which provided important information in recent analyses of structure and energetics of different non-canonical basepairs (Panigrahi et al., 2011). Databases like RNABase (Murthy and Rose, 2003) and SCOR (Klosterman et al., 2002) attempted to classify the available RNA structures but failed to regularly update these only by manual curation of the RNA structures, as the number of structures is increasing quite fast. Any PDB structure released after 2004 is not classified by SCOR and RNABase database can no longer be accessed at the published address www.rnabase.org. Also, there are more activities toward classification of RNA structures on the basis of secondary structure (Tamura et al., 2004; Sarver et al., 2008), canonical as well as unusual base pairing (Lu and Olson, 2003; Das et al., 2006; Roy et al., 2008), isosteric base pairs (Leontis et al., 2002), etc., but the determination of non-redundant set of structures is only done partially by a few groups (Leontis and Westhof, 2001; Stombaugh et al., 2009).

In order to organize and classify the information of RNA structures in PDB files and make it available to the general users, we have developed a web-server, called Hierarchical Database of RNA Structures (HD-RNAS; http://www.saha.ac.in/biop/www/HD-RNAS.html, see **Figure 1**). Keeping in mind that some of the earlier attempts to classify RNA structures failed to keep pace with the structure determination speed, we have adapted an automated programmatic scheme with minimal or no manual intervention for the classification procedure. With the number of RNA structures increasing rapidly, there is a constant pressure of regularly updating this database. As the classification and database creation processes are done by an Octave program, our automated tool is capable of frequently classifying the newly released structures. Hence we expect that HD-RNAS can remain dynamic and would not phase out like the earlier attempts. Some manual curation is obviously involved in this automated procedure to avoid erroneous results. Whenever new structures are released in PDB, they are classified accordingly and verified manually for any inaccuracy. The programmatic scheme is flexible enough to be modified to ensure proper classification of all the RNA structures in case of discrepancies.

CLASSIFICATION METHODOLOGY

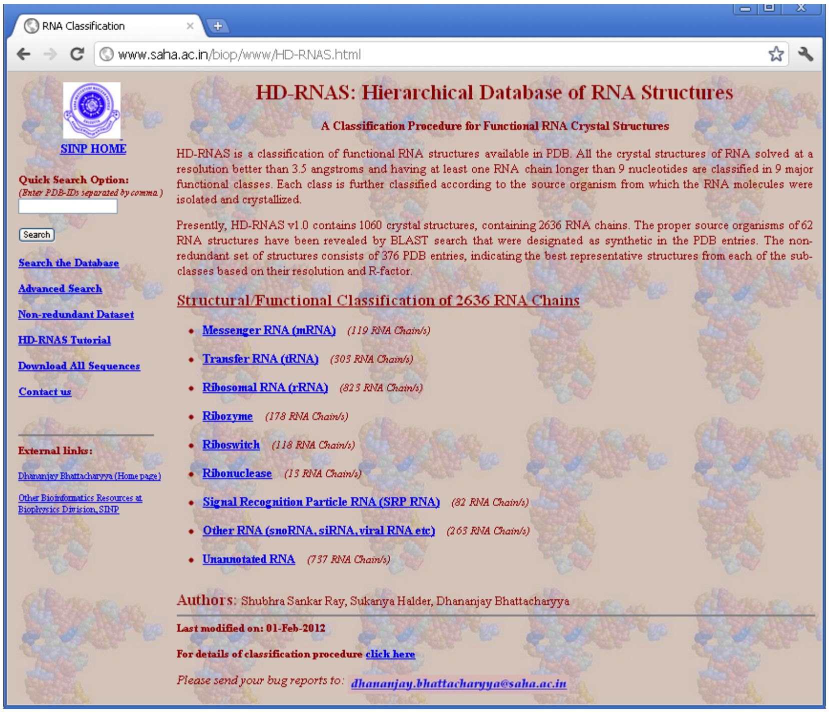
We focus on PDB files containing at least one RNA chain having length equal to or greater than 10 nucleotides, as shorter fragments than this length cannot be expected to fold back and form a secondary structure of biological relevance. RNA chains shorter than 10 nucleotides usually form double helix pairing with their complimentary strands, and do not form a secondary structure on their own.

A total of 2095 RNA structure entries were reported by the PDB search engine in July, 2011. We have developed a software in GNU-Octave, which is similar to MATLAB scripting language, that

i) Programmatically examines and reads the information of all the RNA structures from the PDB files and classifies them,
ii) Creates the necessary database files, and
iii) Creates the web-layout of HTML pages displayed in the server containing major information of each RNA chain.

These HTML files are published in the web. Text-based CGI-Perl codes have been created for back-end support of different search applications, which are available in HD-RNAS. The webserver is available at http://www.saha.ac.in/biop/www/HD-RNAS.html

At the first stage, the RNA structures are classified according to their functional classes, e.g., tRNA, rRNA, mRNA etc. Along with these most common ones, we have also included some other types like ribozymes, riboswitches, ribonucleases, and signal recognition particle (SRP) RNAs, keeping in mind their growing significance in maintaining cellular machinery and their specific structural patterns. A number of PDB files correspond to multi-molecular complexes of several RNA as well as protein chains and information about these individual RNA chains is given in the PDB file as MOL_ID 1, MOL_ID 2 etc. Hence, one PDB file can be classified as belonging to several different classes simultaneously. We have



HD-RNAS: Hierarchical Database of RNA Structures

A Classification Procedure for Functional RNA Crystal Structures

HD-RNAS is a classification of functional RNA structures available in PDB. All the crystal structures of RNA solved at a resolution better than 3.5 angstroms and having at least one RNA chain longer than 9 nucleotides are classified in 9 major functional classes. Each class is further classified according to the source organism from which the RNA molecules were isolated and crystallized.

Presently, HD-RNAS v1.0 contains 1060 crystal structures, containing 2636 RNA chains. The proper source organisms of 62 RNA structures have been revealed by BLAST search that were designated as synthetic in the PDB entries. The non-redundant set of structures consists of 376 PDB entries, indicating the best representative structures from each of the sub-classes based on their resolution and R-factor.

Structural/Functional Classification of 2636 RNA Chains

- Messenger RNA (mRNA) (119 RNA Chain/s)
- Transfer RNA (tRNA) (303 RNA Chain/s)
- Ribosomal RNA (rRNA) (823 RNA Chain/s)
- Ribozyme (178 RNA Chain/s)
- Riboswitch (118 RNA Chain/s)
- Ribonuclease (13 RNA Chain/s)
- Signal Recognition Particle RNA (SRP RNA) (82 RNA Chain/s)
- Other RNA (snoRNA, siRNA, viral RNA etc) (263 RNA Chain/s)
- Unannotated RNA (737 RNA Chain/s)

Authors: Shubhra Sankar Ray, Sukanya Halder, Dhananjay Bhattacharyya

Last modified on: 01-Feb-2012

For details of classification procedure click here

Please send your bug reports to: dhananjay.bhattacharyya@saha.ac.in

FIGURE 1 | HD-RNAS homepage.

made no attempt to classify the structures of DNA or protein chains. The main steps of our Octave code for classifying RNA can be summarized as:

(S1) We have looked for several specific keywords in the MOL_ID field to classify each chain individually into one of the seven major classes for mRNA, SRP-RNA, tRNA, rRNA, Ribonuclease, Riboswitch, and Ribozyme. **Table 1** shows the complete mapping of keywords used for the function assignment. In a few cases the functions are not clearly understood from the information in MOL_ID field alone. In those cases, we have additionally looked at HEADER and the first KEYWDS field of the PDB files also.

(S2) There are some RNA structures of various other functions. As the numbers of structures for these functional classes are very small, at present we have clustered those under the class named 'Other-RNA'. This class also contains RNA chains of significant length (at least 10 nucleotides) and obtained from a natural source, for which no appropriate function could be assigned. There are large numbers of synthetic RNA structures for which no source organism or functional type can be determined. These sequences together comprise the "Unannotated RNA" class.

(S3) The rRNA and tRNA molecules are classified into further subclasses. The rRNA structures are classified according to 5S, 16S, 23S, and 28S (for eukaryotic organisms), depending on their sedimentation coefficients and ribosomal fragments, which group only the defined partial structures. The tRNA structures are classified according to the amino acid or stop codon names.

(S4) The structures are then classified according to the source organism from which the RNA molecules were isolated and crystallized. We have placed each chain according to the organism information as supplied by the SOURCE field of PDB files. In some of the cases, where ORGANISM_SCIENTIFIC field do not produce any source information, the source organism has been extracted from the OTHER_DETAILS field. For example, chain A of PDB entry 1YSV is annotated as "SYNTHETIC" in the "SOURCE" field, whereas OTHER_DETAILS field describes that the sequence is taken from human. Thus, we have put the structure in "*Homo sapiens*" mRNA class. Quite a number of PDB files contain different types of RNA molecules obtained from different source organisms. For example, PDB entry 2J00 is consisted of 16S ribosomal RNA chains from *T. thermophilus*, A-site, P-site, and E-site tRNAs from *Escherichia coli* and a synthetic mRNA. We have, therefore, placed this PDB entry into all the RNA classes along with their corresponding chain identifiers (see **Figure 2**).

(S5) We found that many PDB files do not contain exact source of the RNA chains, and have been indicated as synthetic. However, their size and function indicate that these sequences are from some biological organism. In order to determine the actual source of these RNA chains, we have used BLAST (Altschul et al., 1997) algorithm for sequence alignment. We have used nucleotide sequence database from NCBI as available on June, 2011 and compared our synthetic RNA sequences with all of them. We have picked up the hits having E-value less than 1.0×10^{-5}, number of aligned bases greater than 99% of the complete chain length of synthetic sequence, and sequence identity 99% or greater.

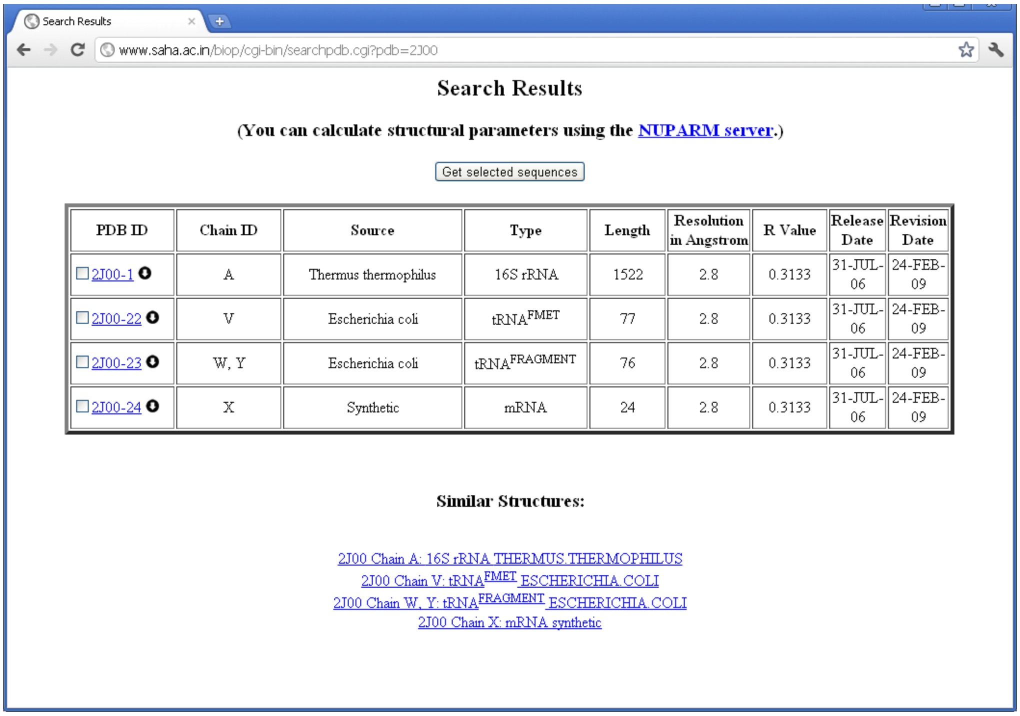
(S6) Finally, we have programmatically created a master database file and all the HTML pages that describe the classification. These HTML files contain major information of each RNA chain, such as PDB-ID, chain name, functional class, source, resolution, chain length, free R-factor, and release date of the entry. Each structure is hyper-linked to the corresponding information page on PDB-site. RNA structures solved by NMR spectroscopy have poorer resolution than X-ray crystal structures and are assigned with a large resolution value and R-factor of 99.0.

(S7) In order to determine the non-redundant set of RNA structures at a given resolution, we pick up the structures with best resolution and R-factor (free R-value) from each subclass. Sometimes smaller fragments of a functional RNA are classified as a full-length functional RNA due to improper information in PDB and often these fragments are of better resolution than the other full-length structures. In order to avoid picking up such fragmented chains as representatives, we have put a length constraint so that the representative structure from a class should be 80% or more of the average length of that structural class. For example, the best representative structure of *E. coli* 23S rRNA should be chain A from PDB 1Q9A as this structure have the best resolution in its class (1.04Å). However, this structure represents only a fragmented part (27 nucleotides) of the complete RNA chain. Thus it is replaced by chain A of PDB 3R8S (resolution: 3.0Å, length: 2903 nucleotides) as the representative of its class. The non-redundant dataset is available at the web-server for various unbiased statistical analysis purposes.

Our classified database is maintained in a flat-file format, without any database management system. This has been possible since we do not keep the large PDB files at the web-server and our complete database is quite small. The web-server also provides different

Table 1 | Special keywords and their corresponding classification for a RNA chain.

Keywords	Significance
UTR, EXON, INTRON	mRNA
CODON, ANTIOCODON, ACCEPTOR tRNA-fragment, A-site, P-site, E-site, tRNA X-MER (e.g., tRNA 30-MER)	tRNA
OPAL or AMBER or OCHRE	tRNA – OPAL or AMBER or OCHRE
RIBONUCLEASE P, RNASE P	Ribonuclease
FMET, FME, INITIATOR, INI, PRIMER	tRNA
S-TURN, CATALYTIC RNA, HAMMERHEAD	Ribozyme
APTAMER	Riboswitch
4.8S, 5S, 5.8S, 16S, 18S, 23S, 28S, 30S, 50S, 70S, 80S	rRNA
4.5S, 7S, 7SL	SRP-RNA



Search Results

(You can calculate structural parameters using the NUPARM server.)

PDB ID	Chain ID	Source	Type	Length	Resolution in Angstrom	R Value	Release Date	Revision Date
□ 2J00-1	A	Thermus thermophilus	16S rRNA	1522	2.8	0.3133	31-JUL-06	24-FEB-09
□ 2J00-22	V	Escherichia coli	tRNA^{FMET}	77	2.8	0.3133	31-JUL-06	24-FEB-09
□ 2J00-23	W, Y	Escherichia coli	tRNA^{FRAGMENT}	76	2.8	0.3133	31-JUL-06	24-FEB-09
□ 2J00-24	X	Synthetic	mRNA	24	2.8	0.3133	31-JUL-06	24-FEB-09

Similar Structures:

2J00 Chain A: 16S rRNA THERMUS THERMOPHILUS
2J00 Chain V: tRNA^{FMET} ESCHERICHIA COLI
2J00 Chain W, Y: tRNA^{FRAGMENT} ESCHERICHIA COLI
2J00 Chain X: mRNA synthetic

FIGURE 2 | Proper classification of individual RNA chains present in 2J00 in HD-RNAS.

search options with user-specified criteria like source organism or functional types. Sequence of the RNA chains in plain text formats can be obtained from the search result pages. Similarly, one can search for PDB files in the database containing a given sequence motif.

RESULTS AND DISCUSSION

PRESENT STATUS

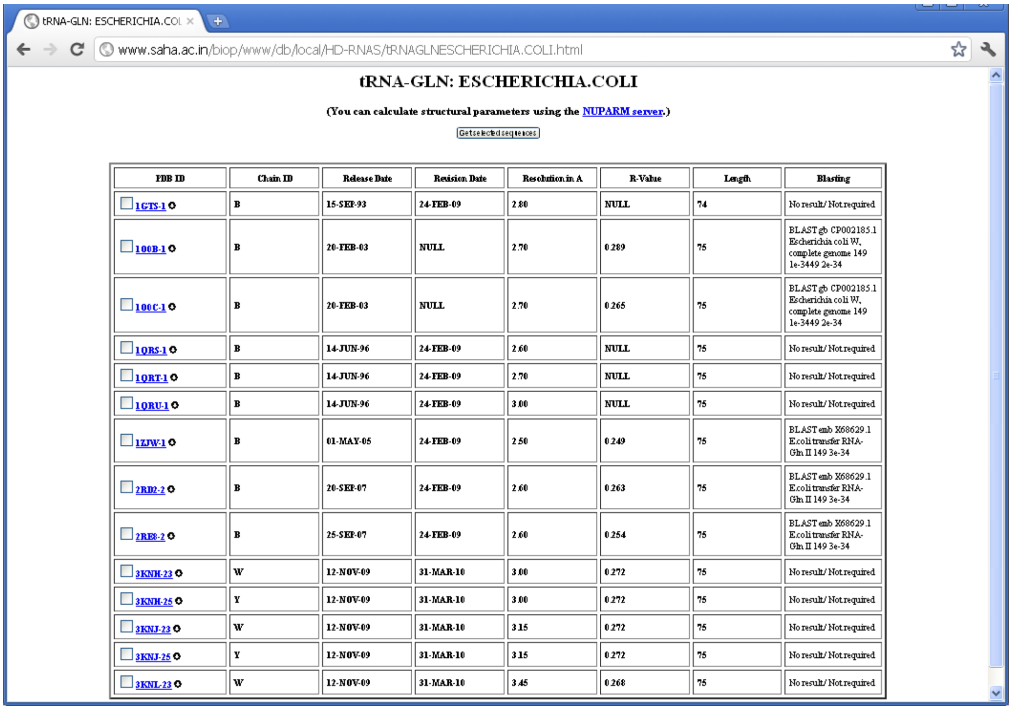
We found 2095 RNA structures were available in PDB in July, 2011. We further rejected 345 crystal structures as these do not contain RNA chains of significant length. Presently, the database contains 1750 PDB files having structures of 2636 RNA chains with significant length. We have 263 structures clustered in the “Other-RNA” group including IRES RNA, viral RNA, miRNA, snoRNA etc. For the “Unannotated RNA” class, it is seen that most of the unannotated RNA chains are shorter than 30 nucleotides (650 out of 737), whereas functional RNAs are generally larger. There are, however, a few entries in the unannotated class, such as 1KH6, 1P6V, 2B57 etc., which have significant length.

BLAST SEARCH FOR SYNTHETIC SEQUENCES

There are many structures where the source organism is mentioned as “SYNTHETIC” by the depositors, whereas one expects these sequences to be derived from some natural organisms, as they correspond to quite long RNA chains. In some of the cases, however, the depositors mentioned about the source organism in the OTHER_DETAILS field of the PDB file. But the information is not provided in machine-interpretable format. The proper source organisms of 66 RNA sequences, which were imperfectly designated as synthetic, have been revealed by BLAST search. There are also some PDB entries like 1DFU, 1EHZ etc., where natural sources of the RNA sequences are mentioned in the OTHER_DETAILS field; yet, they have been designated as synthetic ones probably because they have been synthesized by *in vitro* transcription. We have not done BLAST search for source determination of the synthetic sequences having length smaller than 30 nucleotides, as the significance of BLAST result are poor in these cases and multiple hits with identical E-values are often observed. We could determine source of five (out of 14) *E. coli* tRNA^{Gln} using BLAST (see **Figure 3**). Wherever possible, we have manually crosschecked the validity of BLAST result from the OTHER_DETAILS field of PDB entries and the results are found to be in good agreement (see **Table 2**).

NON-REDUNDANT DATASET

To obtain an unbiased set of RNA structures, we have derived a non-redundant dataset consisting of the best representative structures from each of the classes. These representative structures are the ones with best resolution or, in case there are more than one entries having resolution values identical to the best one, the structure with smallest R-factor and larger length. As the unannotated structures include huge number of structures and most of them are unrelated, we tried to pick up more than one representatives from this clan. In order to remove redundant repeats of structures, we have calculated sequence identity among the structures of synthetic RNAs in the unannotated-RNA group. In cases where two sequences are 100% identical, we have considered that of the best resolution and R-factor as the representative one. The non-redundant dataset thus obtained contains 849 RNA chains from 702 PDB entries. Sometimes, this non-redundant set may


tRNA-GLN: ESCHERICHIA.COL
www.saha.ac.in/biop/www/db/local/HD-RNAS/tRNAGLNESCHERICHIA.COLI.html

tRNA-GLN: ESCHERICHIA.COLI

(You can calculate structural parameters using the NUPARM server.)

Getselectedsequences

PDB ID	Chain ID	Release Date	Revision Date	Resolution in A	R-Value	Length	Blasting
□ 1GTS-1	B	15-SEP-93	24-FEB-09	2.80	NULL	74	No result/ Not required
□ 1O0B-1	B	20-FEB-03	NULL	2.70	0.289	75	BLAST gb CP002185.1 Escherichia coli W, complete genome 149 1e-3449 2e-34
□ 1O0C-1	B	20-FEB-03	NULL	2.70	0.265	75	BLAST gb CP002185.1 Escherichia coli W, complete genome 149 1e-3449 2e-34
□ 1QRS-1	B	14-JUN-96	24-FEB-09	2.60	NULL	75	No result/ Not required
□ 1QRT-1	B	14-JUN-96	24-FEB-09	2.70	NULL	75	No result/ Not required
□ 1QRU-1	B	14-JUN-96	24-FEB-09	3.00	NULL	75	No result/ Not required
□ 1ZJW-1	B	01-MAY-05	24-FEB-09	2.50	0.249	75	BLAST emb X68629.1 E.coli transfer RNA-Gln II 149 3e-34
□ 2RD2-2	B	20-SEP-07	24-FEB-09	2.60	0.263	75	BLAST emb X68629.1 E.coli transfer RNA-Gln II 149 3e-34
□ 2RE8-2	B	25-SEP-07	24-FEB-09	2.60	0.254	75	BLAST emb X68629.1 E.coli transfer RNA-Gln II 149 3e-34
□ 3KNH-23	W	12-NOV-09	31-MAR-10	3.00	0.272	75	No result/ Not required
□ 3KNH-25	Y	12-NOV-09	31-MAR-10	3.00	0.272	75	No result/ Not required
□ 3KNJ-23	W	12-NOV-09	31-MAR-10	3.15	0.272	75	No result/ Not required
□ 3KNJ-25	Y	12-NOV-09	31-MAR-10	3.15	0.272	75	No result/ Not required
□ 3KNL-23	W	12-NOV-09	31-MAR-10	3.45	0.268	75	No result/ Not required

FIGURE 3 | BLAST search results for E. coli tRNA^{Gln}.

contain more than one ribosomal RNA structures from the same group, as they contain RNA chains belonging to different classes, such as tRNA, mRNA, rRNA etc. We have not attempted to remove these, as inter-chain base pairings are also important in higher order organization of RNA structures. The web-server is also equipped with a search tool to determine a non-redundant set of structures with user-defined criteria of functional type, source organism, chain length, and resolution. Such non-redundant set of RNA structures was used recently in analyzing structure and energetics of different non-canonical basepairs (Panigrahi et al., 2011).

In our non-redundant dataset, we find that there are a large number of structures of RNA with length less than 30 nucleotides. These are mainly synthetic sequences for which functional annotation is unavailable. Among the 849 structures, 491 sequences do not carry any functional information, among which 427 sequences are of insignificant length (< 30 nucleotide). As the functional RNA molecules are generally of length larger than 30-residues, we also generate a suggested non-redundant set containing representative structures from each of the functional classes as well as representatives from the unannotated groups with larger length. The members are selected with resolution better than 3.5Å to make it a meaningful set of structures for real applications. This set has 159 structures, including only 22 functionally unannotated RNA structures of synthetic source. We have kept no structure solved by NMR spectroscopy in the non-redundant set, as there is no way one can determine quality of the data. The structures determined by cryo-electron microscopy are automatically removed because of their poor resolution.

APPLICATIONS

The database can be searched for a set of RNA structural entries according to functional type or source organism. Also, a combined search can be performed using advanced options where a user can specify the chain length as well as a resolution cutoff of the crystal structures. Furthermore, one can determine if there are any structures whose sequence is identical to a given nucleotide sequence. At present, only scientific names of the organisms are accepted for search criteria in "Advanced Search Options" of our web-server.

As the classification shows, there are many RNA classes where the numbers of PDB files are 10 or greater. These subclasses have been shown in **Table 3** and corresponding MOL_ID's are shown as a suffix to the PDB-ID. They carry signatures that may indicate variations introduced in the molecular structure due to ligand binding or alteration of crystallization conditions. Eventually, they can be referred to as crystallographic ensembles in analogy with statistical ensembles obtained from molecular dynamics or Monte Carlo simulations as done recently (Halder and Bhattacharyya, 2010; Samanta et al., 2010). We have performed pair-wise secondary structure comparison for these classes to compare the structural similarity between them. For secondary structure assignment, the base pairing patterns of each RNA structure in a functional class have been obtained using BPFind software tool (Das et al., 2006). BPFind gives us the secondary structure of a nucleic acid at the basepairing level. These secondary structural information of basepairing for each chain have been converted to a one-dimensional string of characters: H, N, T, and L corresponding to Watson–Crick base pairs, non-canonical base pairs, base triplets, and unpaired

Table 2 | Validation of BLAST search results.

PDB	Chain	BLAST result	PDB description
1ASY	R, S	BLAST gb M25168.1 YSCTRDCER S.cerevisiae Asp-tRNA 149 2e-34	Yeast aspartyl-tRNA synthetase complexed with tRNA Asp
1ASZ	R, S	BLAST gb M25168.1 YSCTRDCER S.cerevisiae Asp-tRNA 149 2e-34	Yeast aspartyl-tRNA synthetase complexed with tRNA Asp
1GID	A, B	BLAST emb V01416.1 Fragments of a Tetrahymena gene for 26s rRNA with... 311 8e-83	Tetrahymena group I INTRON
1HR2	A, B	BLAST emb V01416.1 Fragments of a Tetrahymena gene for 26s rRNA with... 295 4e-78	Tetrahymena group I INTRON
1IVS	C, D	BLAST dbj AB080139.1 *Thermus thermophilus* tRNA-Val 149 2e-34	Thermus thermophilus valyl-tRNA synthetase complexed with tRNA(VAL)
1U9S	A	BLAST gb AE017221.1 *Thermus thermophilus* HB27, complete genome 319 3e-85	Thermus thermophilus
1VTQ	A	BLAST gb M25168.1 YSCTRDCER S.cerevisiae Asp-tRNA 149 2e-34	Yeast t-RNA-Asp
2CSX	C, D	BLAST gb AE000657.1 Aquifex aeolicus VF5, complete genome 149 2e-34	Aquifex aeolicus methionyl-tRNA synthetase complexed with tRNA(MET)
2CT8	C, D	BLAST gb AE000657.1 Aquifex aeolicus VF5, complete genome 147 1e-33	Aquifex aeolicus methionyl-tRNA synthetase complexed with tRNA(MET)
2CV0	C, D	BLAST gb AE017221.1 *Thermus thermophilus* HB27, complete genome 149 2e-34	Glutamyl-tRNA synthetase from thermus thermophilus in complex with tRNA(GLU)
2CV1	C, D	BLAST gb AE017221.1 *Thermus thermophilus* HB27, complete genome 149 2e-34	Glutamyl-tRNA synthetase from thermus thermophilus in complex with tRNA(GLU)
2CV2	C, D	BLAST gb AE017221.1 *Thermus thermophilus* HB27, complete genome 149 2e-34	Glutamyl-tRNA synthetase from thermus thermophilus in complex with tRNA(GLU)
2DXI	C, D	BLAST gb AE017221.1 *Thermus thermophilus* HB27, complete genome 149 2e-34	Glutamyl-tRNA synthetase from thermus thermophilus complexed with tRNA(GLU)
2IHX	B	BLAST gb M21526.1 ALRSRCAC Rous sarcoma virus defective mutant PR2257... 149 2e-34	Sequence occurs naturally in rous sarcoma virus (RSV)
2NOQ	A	BLAST gb AF218039.1 AF218039 Cricket paralysis virus non-structural pol... 377 e-102	Ribosome-bound cricket paralysis virus IRES RNA
2ZNI	C, D	BLAST dbj AP008230.1 *Desulfitobacterium hafniense* Y51 DNA, complete g... 143 1e-32	Pyrrolysyl-tRNA synthetase-tRNA(PYL) complex from desulfitobacterium hafniense
3AKZ	F, H, G, E	BLAST gb AE000512.1 *Thermotoga maritima* MSB8, complete genome 147 1e-33	Thermotoga maritima non-discriminating glutamyl-tRNA synthetase in complex with tRNAGLN
3AL0	E	BLAST gb AE000512.1 *Thermotoga maritima* MSB8, complete genome 147 1e-33	Glutamine transamidosome from thermotoga maritima in the glutamylation state
3FOZ	D	BLAST gb CP002185.1 Escherichia coli W, complete genome 137 3e-3137 8e-31	*E. Coli* isopentenyl-tRNA transferase in complex with *E. coli* tRNA(PHE)
3Q50	A	BLAST gb AE008691.1 *Thermoanaerobacter tengcongensis* MB4, complete ge... 66 8e-10	Based on the sequence of class I, type 1 PREQ1 riboswitch aptamer from T. tengcongensis
3Q51	A	BLAST gb AE008691.1 *Thermoanaerobacter tengcongensis* MB4, complete ge... 66 8e-10	Based on the sequence of class I, type 1 PREQ1 riboswitch aptamer from T. tengcongensis

bases, respectively. The secondary structures thus obtained correspond to individual RNA chains separately and do not consider inter-chain base pairing information. These secondary structural sequences are then compared with that of the best representative structure of that class using Needleman–Wunsch algorithm (Needleman and Wunsch, 1970) as implemented in EMBOSS (Rice et al., 2000). An identity scoring matrix (EDNAMAT) was used for such alignment with a gap-open penalty of −10.0 and gap-extension penalty of –0.5. We find that the average similarity is 80% or more for a set of similar structures of a given class having 30 or more RNA chains. The detailed analyses of structural variation among these crystallographic ensembles are beyond the scope of this paper and would be presented elsewhere. Here a point to note is that secondary structure comparison have not been performed for crystal structures with resolution worse than 3.5Å because of the poor quality of secondary structural information obtained from such structures. RNA structures solved by NMR spectroscopy or electron microscopy are also not included in secondary structure comparison for the same reason.

Table 3 | Subclasses where the number of PDB files are 10 or greater and their structural variation.

Functional type	Organism	No. of PDB files*	Representative structure	Average similarity	SD
16S rRNA	*Escherichia coli*	26 (25)	3OFP_A	0.9053	0.0403
16S rRNA	*Thermus thermophilus*	73 (69)	2VQE_A	0.8556	0.0579
5S rRNA	*Deinococcus radiodurans*	12 (12)	2ZJR_Y	0.6792	0.1558
5S rRNA	*Escherichia coli*	27 (24)	3OFQ_B	0.8896	0.0454
5S rRNA	*Haloarcula marismortui*	61 (61)	1VQO_9	0.9077	0.0451
5S rRNA	*Thermus thermophilus*	45 (44)	2J01_B	0.8349	0.0952
23S rRNA	*Deinococcus radiodurans*	23 (23)	2ZJR_X	0.6582	0.1306
23S rRNA	*Escherichia coli*	30 (30)	3OFR_A	0.742	0.2972
23S rRNA	*Haloarcula marismortui*	65 (58)	1VQO_0	0.9714	0.0128
23S rRNA	*Thermus thermophilus*	44 (41)	2J01_A	0.9516	0.03
$tRNA^{fMet}$	*Escherichia coli*	20 (20)	2FMT_C/D	0.8499	0.0943
$tRNA^{Gln}$	*Escherichia coli*	14 (14)	1ZJW	0.7755	0.2094
$tRNA^{Phe}$	*Escherichia coli*	55 (55)	3FOZ_D	0.5269	0.2425

**Numbers in parentheses give the total no. of structures after removing outliers.*

Table 4 | Function assignment to RNAs of the other-RNA class.

Query	Length	Original function	Best match	Predicted function	Best score	Average	SD	Prediction score
1FIR-A	76	$tRNA^{Lys}$	1EVV-A	$tRNA^{Phe}$	362.00	131.60	13.62	16.92
2B57-A	65	Riboswitch	2XNZ-A	Riboswitch	307.00	99.34	11.35	18.29
2ZZM-B	88	$tRNA^{Leu}$	1WZ2-D	$tRNA^{Leu}$	326.50	118.16	14.40	14.46
2ZZN-C	75	$tRNA^{Cys}$	3KFU-L	$tRNA^{Asn}$	304.00	128.54	12.98	13.52
2ZZN-D	75	$tRNA^{Cys}$	1F7U-B	$tRNA^{Arg}$	309.00	135.00	13.32	13.06
3KIQ-a	1504	rRNA	2VQE-A	16S rRNA	5904.50	2062.09	46.95	81.85
3KIQ-v	77	tRNA	2FMT-D	$tRNA^{Fmet}$	349.00	150.39	15.36	12.93
3KIR-A	2848	rRNA	2XG0-A	rRNA fragment	12839.00	3764.09	55.97	162.13
3KIR-B	119	rRNA	2J01-B	5S rRNA	505.00	156.37	13.11	26.60
3KIS-a	1504	rRNA	2VQE-A	16S rRNA	5904.50	2068.17	46.28	82.89
3KIS-v	77	tRNA	2ZUE-B	$tRNA^{Arg}$	342.00	146.26	11.99	16.32
3KIT-A	2848	rRNA	2XG0-A	rRNA fragment	12821.00	3779.68	61.28	147.53
3KIT-B	119	rRNA	2J01-B	5S rRNA	523.00	148.25	13.61	27.54
3KNN-W	75	tRNA	1F7U-B	$tRNA^{Arg}$	279.00	129.06	12.13	12.36
3KNN-X	77	$tRNA^{Fmet}$	2FMT-D	$tRNA^{Fmet}$	331.00	147.96	14.51	12.62
1VFG-D	75	Primer tRNA	2V0G-F	$tRNA^{Leu}$	138.00	93.55	8.51	5.22
3KNN-Y	75	tRNA	2XQD-W	tRNA	208.00	146.06	11.57	5.35
3MOJ-A	74	23S rRNA	2XG0-A	rRNA fragment	196.50	158.28	8.48	4.51
3NPB-A	119	Riboswitch	2A64-A	Ribonuclease	318.00	255.50	15.32	4.08
3R9X-C	35	16S rRNA	1M5O-E	Ribozyme	142.00	83.44	13.25	4.42
3KIS-w	77	tRNA	2XG0-A	rRNA fragment	164.50	160.35	12.56	0.33
3KIQ-w	77	tRNA	2XG0-A	rRNA fragment	211.00	203.98	10.29	0.68
1P6V-B	68	tmRNA	2XG0-A	rRNA fragment	146.00	107.33	8.65	4.47
1P6V-D	68	tmRNA	3IAB-R	Ribonuclease	102.00	65.03	10.56	3.50
1S03-A	47	mRNA	2ZNI-D	$tRNA^{Lys}$	178.50	127.49	9.00	5.67
1S03-B	47	mRNA	2ZNI-D	$tRNA^{Lys}$	178.50	127.83	9.54	5.31
1VFG-C	75	Primer tRNA	2XG0-A	rRNA fragment	123.50	105.85	6.09	2.90

Comparison of secondary structure can also be useful to predict the functions of unannotated RNAs. As mentioned in the earlier section, the functional types of many RNA chains that have been placed under the other-RNA and unannotated-RNA classes could not be obtained from the PDB files unambiguously. However, we find that there are a number of structures in these two types, which are of significant length (30 nucleotide or more). We have performed structural comparison of these RNA chains with known functional RNAs available in the suggested non-redundant dataset, consisting of only X-ray crystal structures with resolution better than 3.5Å. For the secondary structure alignment, the basepairing pattern for each RNA structure in the non-redundant dataset, obtained using BPFind software, comprises a database of known structural forms of well-classified functional RNAs. In a similar way, the secondary structural data for each of the unannotated RNA chains were generated and compared against the database of known structural forms. Thus, the secondary structural information of an unannotated RNA chain has been aligned with secondary structural information of each RNA chain present in the non-redundant dataset using Needleman–Wunsch algorithm for global sequence alignment. We have assigned probable function of an unannotated RNA as identical to some known structural form with which the best similarity score is obtained. For example, chain A of 1FIR shows highest score of 362 against chain A of 1EVV (**Table 4**). To validate the results of our function prediction method and to remove any false positive, we have used the bootstrapping method. For each pair of unannotated query and its best match in the non-redundant dataset, the best match secondary structure has been shuffled to generate a set of 100 random secondary structures having the same composition. Thus, the secondary structure sequence for chain A of 1EVV has been shuffled to generate 100 random sequence files and aligned secondary structure of 1FIR (chain: A) with all these 100 random sequences. Average and SD of these alignment scores have been calculated (**Tables 4** and **5**). We find that in many cases the original best match score and the average score against random sequences are quite similar (e.g., 3NVK, 1VFG, etc.). **Figure 4** shows distributions of the random scores along with the actual predictive scores for two representative systems. Obviously, prediction

Table 5 | Function annotation to unannotated RNAs.

Unannotated RNA								
Query	**Length**	**Original function**	**Best match**	**Predicted function**	**Best score**	**Average**	**SD**	**Prediction score**
1JBR-D	31	28S rRNA	1Q96-B	28S rRNA	117.00	39.56	7.76	9.98
2HW8-B	36	mRNA	1ZHO-H	mRNA	180.00	66.01	8.86	12.86
2ZH1-B	33	tRNA$^{\mathrm{Phe}}$	3AKZ-G	tRNA$^{\mathrm{Gln}}$	154.50	87.89	9.13	7.30
3DS7-A	67	Guanine riboswitch	1U8D-A	mRNA (Riboswitch)	317.00	109.10	10.63	19.56
3DS7-B	67	Guanine riboswitch	1U8D-A	mRNA (Riboswitch)	299.00	107.86	10.67	17.92
3HJW-D	58	H/ACA snoRNA	1WZ2-D	tRNA$^{\mathrm{Leu}}$	197.00	105.06	11.43	8.04
3ICQ-D	67	tRNA	1EVV-A	tRNA$^{\mathrm{Phe}}$	212.00	102.10	10.20	10.78
3ICQ-E	67	tRNA	2Y10-V	tRNA$^{\mathrm{Trp}}$	230.00	116.67	10.18	11.13
3LWO-D	58	H/ACA RNA	1WZ2-D	tRNA$^{\mathrm{Leu}}$	197.00	105.99	13.15	6.92
3LWP-D	58	H/ACA snoRNA	1WZ2-D	tRNA$^{\mathrm{Leu}}$	197.00	107.14	12.43	7.23
3LWQ-D	58	H/ACA snoRNA	1WZ2-D	tRNA$^{\mathrm{Leu}}$	197.00	105.22	11.26	8.15
3LWR-D	58	H/ACA snoRNA	1WZ2-D	tRNA$^{\mathrm{Leu}}$	197.00	107.21	11.17	8.04
3LWV-D	58	H/ACA RNA	1WZ2-D	tRNA$^{\mathrm{Leu}}$	197.00	107.64	13.11	6.82
3OUY-C	35	tRNA$^{\mathrm{Ile}}$	3A3A-A	tRNA$^{\mathrm{Sel}}$	157.00	95.52	8.96	6.86
3OV7-C	34	tRNA$^{\mathrm{Ile}}$	3AKZ-G	tRNA$^{\mathrm{Gln}}$	134.00	78.04	9.20	6.08
3OVA-C	34	tRNA$^{\mathrm{Ile}}$	3AKZ-G	tRNA$^{\mathrm{Gln}}$	159.50	88.13	9.76	7.31
3OVB-C	35	tRNA$^{\mathrm{Ile}}$	3AKZ-G	tRNA$^{\mathrm{Gln}}$	164.50	86.64	9.72	8.01
3OVB-D	35	tRNA$^{\mathrm{Ile}}$	3A3A-A	tRNA$^{\mathrm{Sel}}$	175.00	104.03	10.02	7.09
3OVS-C	34	tRNA$^{\mathrm{Ile}}$	3A3A-A	tRNA$^{\mathrm{Sel}}$	170.00	103.36	9.39	7.10
3OVS-D	34	tRNA$^{\mathrm{Ile}}$	3AKZ-G	tRNA$^{\mathrm{Gln}}$	142.00	75.11	9.84	6.80
3OV7-D	34	tRNA$^{\mathrm{Ile}}$	3A3A-A	tRNA$^{\mathrm{Sel}}$	152.00	95.84	10.76	5.22
3OUY-D	35	tRNA$^{\mathrm{Ile}}$	3A3A-A	tRNA$^{\mathrm{Sel}}$	139.00	88.53	9.62	5.25
2HVY-E	65	H/ACA	3R8S-B	5S rRNA	196.00	109.93	12.66	6.80
3HAX-E	63	H/ACA	1M5O-E	Ribozyme	201.50	104.17	14.31	6.80
3P22-C	40	Guide RNA	3IAB-R	Ribonuclease	167.00	91.79	11.09	6.78
3P22-E	40	Guide RNA	3IAB-R	Ribonuclease	167.00	89.59	11.18	6.92
3P22-G	40	Guide RNA	3IAB-R	Ribonuclease	167.00	92.37	12.71	5.87
1DDY-A	35	Aptamer	2A64-A	Ribonuclease	94.00	68.71	9.19	2.75
1DDY-C	35	Aptamer	2A64-A	Ribonuclease	94.00	67.48	8.89	2.99
1DDY-E	35	Aptamer	2A64-A	Ribonuclease	94.00	67.40	8.49	3.14
1DDY-G	35	Aptamer	2A64-A	Ribonuclease	94.00	68.01	9.96	2.61
1KH6-A	48	IRES RNA (viral)	3AM1-B	tRNA	164.00	120.80	8.39	5.15
1XJR-A	47	s2m RNA (viral)	3ADD-D	tRNA$^{\mathrm{Sel}}$	125.50	98.22	9.22	2.96
3NVK-K	34	Box C/D snoRNA	2XG0-A	rRNA Fragment	73.50	71.68	5.87	0.31
3NVK-L	34	Box C/D snoRNA	2XQD-W	tRNA	68.00	59.05	6.88	1.30
3P22-A	40	Guide RNA	1EUY-B	tRNA$^{\mathrm{Gln}}$	155.00	86.87	9.14	7.45

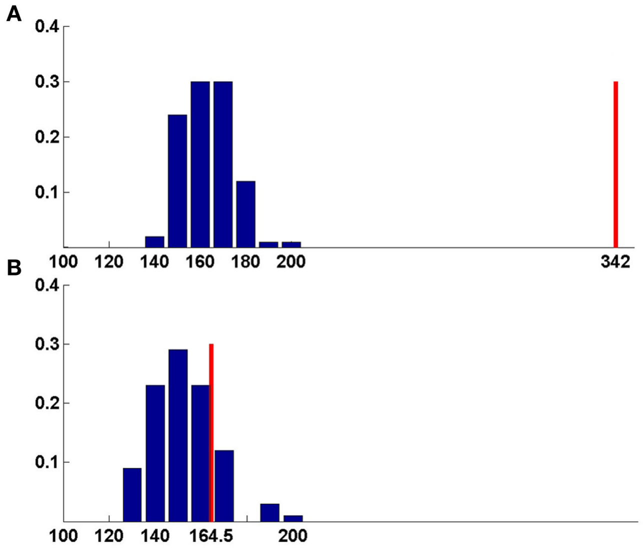


FIGURE 4 | Bootstrapping method for function prediction. Blue-random scores, red-best match score. **(A)** Chain v of 3KIS, **(B)** chain w of 3KIS. For chain v, the best score is significantly larger than random average score giving correct prediction, whereas for chain w, best match score lies within the range of random scores, leading to false prediction (**Table 4**).

of function of w-chain of 3KIS is questionable (**Figure 4b**) while that of v-chain of 3KIS is a good prediction. This quality depends on (*original score for best match – average score from random sequences*)/*SD of random scores*, as given in the last column of **Tables 4** and **5**. It is found that when these values are larger than 6 or 7, the predictions are generally correct. We have tried to manually examine our results against the original functions as obtained from literature study and found them to be in very good agreement.

Using the above procedure, we have been able to predict the functions of 17 RNA structures from the "Unannotated-RNA" class and 18 more from the "Other-RNA" class. The results of the function assignments are shown in **Tables 4** and **5**. Structures of H/ACA box snoRNAs, like 3HJW (chain D), 3LWP (chain D), 3LWQ (chain D), 3LWR (chain D), and 3P22 (chain A), have been predicted as tRNAs. In these cases, the snoRNAs are bound with tRNA-processing proteins and thus have tRNA-like structural motifs for proper recognition. Also, the sub-divisions of different tRNAs can be predicted with limited accuracy only, as the secondary structure of tRNAs is nearly universal among different sub-types.

CONCLUSION

Hierarchical Database of RNA Structures is an evolving resource that is expected to grow and incorporate more and more RNA structures as and when they are solved and made available from PDB. We have used the PDB files in plain text format, instead of XML files, as these do not contain any extra information but require significantly more storage due to their huge size. The classification of RNA structures are done by an automated tool, a code written in high-level GNU-Octave language, which takes roughly 3 h to classify 2095 PDB files in a 3.0 GHz "Pentium 4" processor with 1GB RAM. Most importantly, out of these 3 h, the classification job takes only a part and nearly 2 h are used by the BLAST search program. However, this task is carried out once in a month during creation of the database and does not affect users. There is also an inbuilt function called "pdbread" in the recent versions of Matlab which can read a PDB file into a structure and can also store the relevant information. Although, a small part of our program and the function "pdbread" are similar, the function "pdbread" demands more time to gather the complete information including the coordinate data from a big PDB file. However, our routine skips the coordinate data in the PDB file to reduce CPU time. The program code is a flexible one and certainly there would be necessity for modifications when structures of altogether new RNA with function as yet unknown or when structures of more siRNA, miRNA, virus etc., would be available. However, we can easily modify the code in future to characterize these structures into new functional classes. Similarly, when several structures of a sub–sub-class would be available, it is expected to open up new directions of research toward understanding ligand or environment-induced structural alterations. For instance, it would be interesting to understand conformational variations between ligand-bound and ligand-free states of riboswitches, between ribosome structures with or without tRNA bound to it, free tRNA and tRNA complexed with synthatase etc. Such structural comparisons can also be used to detect the functional class of unannotated RNA structures. We have used a simple mechanism for prediction of function of unknown RNA structures but it can be improved by betterment of the scoring matrix or by using graph-theoretic approach. We are planning to include the predicted functions in later versions of the database.

ACKNOWLEDGMENTS

We are thankful to Dr. Devapriya Chowdhury, JNU, New Delhi, India, and Dr. Rahul Banerjee, SINP, India, for suggestions. We are thankful to Ms. Shreya Pakrashi, Ms. Tania Sen, and Ms. Rishika Sengupta for technical support. Funding: This work has been partially supported by Dept. of Biotechnology, India and CBAUNP project of Dept. of Atomic Energy, Govt. of India.

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST a new generation of protein database search programs. *Nucl. Acids Res.* 25, 3389–3402.

Batey, R. T., Rambo, R. P., and Doudna, J. A. (1999). Tertiary motifs in RNA structure and folding. *Angew. Chem. Int. Ed. Engl.* 38, 2327–2343.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.

Das, J., Mukherjee, S., Mitra, A., and Bhattacharyya, D. (2006). Non-canonical base pairs and higher order structures in nucleic acids crystal structure database analysis. *J. Biomol. Struct. Dynam.* 24, 149–161.

Ferre-d'Amare, A. R., and Doudna, J. A. (1999). RNA FOLDS insights from recent crystal structures. *Ann. Rev. Biophys. Biomol. Struct.* 28, 57–73.

Halder, S., and Bhattacharyya, D. (2010). Structural stability of tandemly occurring noncanonical basepairs within double helical fragments: molecular dynamics studies of functional RNA. *J. Phys. Chem. B* 114, 14028–14040.

Hermann, T., and Patel, D. J. (1999). Stitching together RNA tertiary architectures. *J. Mol. Biol.,* 294, 829–849.

Holm, L., and Sander, C. (1997). Dali/FSSP: classification of three-dimensional protein folds. *Nucl. Acids Res.* 25, 231–234.

Hubbard, T. J. P., Murzin, A. G., Brenner, S. E., and Chothia, C. (1997). SCOP: A structural classification of proteins database. *Nucl. Acids Res.* 25, 236–239.

Klosterman, P. S., Tamura, M., Holbrook, S. R., and Brenner, S. E. (2002). SCOR: a structural classification of RNA database. *Nucl. Acids Res.* 30, 392–394.

Lee, S., and Blundell, T. L. (2009), BIPA: a database for protein-nucleic acid interaction in 3D structures. *Bioinformatics* 15, 1559–1560.

Leontis, N. B., Stombaugh, J., and Westhof, E. (2002). The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucl. Acids Res.* 30, 3497–3531.

Leontis, N. B., and Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *RNA* 7, 499–512.

Lu, X. J., and Olson, W. K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucl. Acids Res.* 31, 5108–5121.

Moore, P. B. (1999). Structural motifs in RNA. *Ann. Rev. Biochem.* 68, 287–300.

Murthy, V. L., and Rose, G. D. (2003). RNABase: an annotated database of RNA structures. *Nucl. Acids Res.* 31, 502–504.

Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995). SCOP – a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.,* 247, 536–540.

Needleman, S. B., and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.,* 48, 443–453.

Panigrahi, S., Pal, R., and Bhattacharyya, D. (2011). Structure and energy of non-canonical basepairs: comparison of various computational chemistry methods with crystallographic ensembles. *J. Biomol. Struct. Dynam.* 29, 541–556.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16, 276–277.

Roy, A., Panigrahi, S., Bhattacharyya, M., and Bhattacharyya, D. (2008). Structure, stability and dynamics of canonical and noncanonical base pairs: quantum chemical studies. *J. Phys. Chem. B* 112, 3786–3796.

Samanta, S., Chakrabarti, J., and Bhattacharyya, D. (2010). Changes in thermodynamic properties of DNA base pairs in protein-DNA recognition. *J. Biomol. Struct. Dynam.* 27, 429–442.

Sarver, M., Zirbel, C. L., Stombaugh, J., Mokdad, A., and Leontis, N. B. (2008). FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J. Math. Biol.,* 56, 215–252.

Scott, W. G., Finch, J. T., and Klug, A. (1995). The crystal-structure of an all-RNA hammerhead ribozyme – a proposed mechanism for RNA catalytic cleavage. *Cell* 81, 991–1002.

Stombaugh, J., Zirbel, C. L., Westhof, E., and Leontis, N. B. (2009). Frequency and isostericity of RNA base pairs. *Nucl. Acids Res.* 37, 2294–2312.

Tamura, M., Hendrix, D. K., Klosterman, P. S., Schimmelman, N. R. B., Brenner, S. E., and Holbrook, S. R. (2004). SCOR: structural classification of RNA, version 2.0. *Nucl. Acids Res.* 32, D182–D184.

Wang, G. L., and Dunbrack, R. L. (2005). PISCES: recent improvements to a PDB sequence culling server. *Nucl. Acids Res.* 33, W94–W98.

Zwieb, C., Van Nues, R. W., Rosenblad, M. A., Brown, J. D., and Samuelsson, T. (2005). A nomenclature for all signal recognition particle RNAs. *RNA* 11, 7–13.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 December 2011; accepted: 29 March 2012; published online: 18 April 2012.

Citation: Ray SS, Halder S, Kaypee S and Bhattacharyya D (2012) HD-RNAS: an automated hierarchical database of RNA structures. Front. Gene. ***3****:59. doi: 10.3389/fgene.2012.00059*

This article was submitted to Frontiers in Bioinformatics and Computational Biology, a specialty of Frontiers in Genetics.