Lawrence Berkeley National Laboratory

Title
SCOR: Structural classification of RNA, Version 2.0

Permalink
https://escholarship.org/uc/item/6b604080

Authors
Tamura, Makio
Hendrix, Donna K.
Klosterman, Peter

Publication Date
2003-10-03

SCOR: Structural Classification of RNA, Version 2.0

Makio Tamura[1], Donna K. Hendrix[2], Peter S. Klosterman[1], Nancy R. B. Schimmelman[2], Steven E. Brenner[1,2] and Stephen R. Holbrook[1]

[1]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 and [2]Department of Plant and Microbial Biology, 111 Koshland Hall #3102, University of California, Berkeley, CA 94720-3102

To whom correspondence should be addressed:

      Stephen R. Holbrook

      Department of Structural Biology and Computational and Theoretical Biology

      Physical Biosciences Division

      Lawrence Berkeley National Laboratory

      1 Cyclotron Road, MS 64-123

      Berkeley, CA 94720

      Telephone: +1 510-486-4304

      FAX: +1 510-486-6059

      E-mail: srholbrook@lbl.gov

Abstract

SCOR (http://scor.lbl.gov), the Structural Classification of RNA, is a database designed to provide a comprehensive perspective and understanding of RNA motif three-dimensional structure, function, tertiary interactions, and their relationships.

SCOR 2.0 represents a major expansion and introduces a wholly new classification system. The new version represents the classification as a Directed Acyclic Graph (DAG), which allows a classification node to have multiple parents, in contrast to the strictly hierarchical classification used in SCOR 1.2. SCOR 2.0 supports three types of query terms in the updated search engine: PDB or NDB identifier, nucleotide sequence, and keyword. We also provide parseable XML files for all information.

This new release contains 511 RNA entries from the PDB as of 15 May 2003. A total of 5,880 secondary structural elements are classified; 2,104 hairpin loops and 3,776 internal loops. RNA motifs reported in the literature, such as "Kink turn" and "GNRA loops," are now incorporated into the structural classification along with definitions and descriptions.

Introduction

The SCOR database was first released in September 2001 (1). Version 1.2, released in July 2002, added the structural classification of small and large ribosomal RNA, a classification of the ribose zipper tertiary interaction (2), and enhancement of the search engine. These first versions of the SCOR database provided a hierarchical classification of structural elements; however, one RNA structural element can have several distinct features and may belong to multiple classes. For example, the subclass *GNRA tetraloops* (3, 4) should be found under both the class *U-Turn* (5, 6) and the class *Tetraloops*. That is, a classified node can have multiple parents, and thus different paths from the root can be used to reach the same subclass. Such expressivity is not supported by a hierarchical structure. We therefore reviewed the feature inheritance between classes and subclasses in SCOR 1.2 and mapped them into a Directed Acyclic Graph (DAG)

structure to allow multiple parents. Such a DAG representation has been used in the Gene

Ontology to represent terms and their relationships (7).

The number of RNA structures in the Protein Databank (8) and Nucleic Acid Database (9)

has dramatically increased in the past few years, from 259 in October 2000 to over 500 in May

2003, emphasizing the need for an organizational classification of the observed data. A unified

terminology to describe the RNA structure elements is also becoming increasingly important for

structural research. In the SCOR database, a description for each class is provided to help create a

common terminology or glossary for RNA structural motifs.

For the new database, a total of 511 PDB entries were classified. These entries contain

more than two RNA residues, are solved by NMR spectroscopy or X-ray crystallography (with

resolution better than 4.0 Angstroms), and were released before15 May 2003. Structures of

DNA-RNA hybrids and chimeras were included in the classification. If a structure were updated in

the PDB and its previous entry obsolesced, then the new structure was included and references to

the previous entry were removed from the classification. The list of structures is available on the

web site, including PDB and NDB identifiers and original literature references. Structural motifs

were identified visually with the assistance of computational tools (10-12). A comparison of the

different versions of the SCOR database is given in Table 1. A parseable XML format of the

database can be downloaded from the web site.

Database Structure

SCOR 2.0 has been fundamentally revised to use a DAG organization for the structural

motif classification. In the DAG, each class has a parent-child relationship, where the child classes,

or subclasses, contain all the features of its parent. There is no limit on the number of levels from

the root to the motif and there is no special nomenclature for each level of classification. In the

previous release of SCOR, we used special terminology according to the depth of level such as

"Class", "Subclass", and "Motif." In the current version, we follow the model from modern

phylogenetics, which has largely abandoned the Kingdom/Phylum/Class/Order nomenclature and allows arbitrary numbers of levels. Indeed, in SCOR one can be at different numbers of nodes from the root depending on the path.

Structural Classification

RNA loop elements are diverse structures that provide a scaffold for folding and recognition; thus, in SCOR, the structural classification is done at the RNA motif level for hairpin loops, internal loops, and tertiary interactions. We provide a definition and schematic diagram for each class, as well as a mappingof the parent -child relationships. Motifs are the leaves of the DAG; i.e., they have parents and no children. Structural motifs are classified manually by base pairing, base stacking, and backbone conformation. Familiar RNA motifs such as "S Turn " (13-15) and "GNRA tetraloops" (3, 4) are specifically highlighted.

A total of 5,880 secondary structural elements are classified in SCOR 2.0. This is nearly a 10-fold increase from the SCOR 1.2. While the number of available RNA structures has nearly doubled, the increase in the number of loops is due primarily to the ribosome structures 1jj2 (16) and 1j5e (17) and their associated structures with bound ligands.

Functional Classification

SCOR now classifies function at the molecular level and the motif level. In the *Molecular Function* classification, we place RNAs into two classes *Naturally occurring RNAs* and *Evolved (SELEX) RNAs*. *Naturally occurring RNAs* includes transfer RNA, ribosomal RNA and SRP RNAs, among other examples of molecules found in nature.

Motif Function is a new classification in SCOR 2.0 and represents functions specific to structural motifs. Under *Motif Function* are four classes: *Protein binding*, *Small molecule binding*, *Replicase recognition*, and *Helical packing*. These functions are related to local RNA structure rather than global RNA structure. The most specific level of this classification is linked to the structural motif classification.

Structural Models contains structures that have been created specifically to study specific aspects of RNA function and structure. These include *RNA-DNA chimeric duplexes*, *RNA-DNA four-way junctions*, *Self complementary duplexes* and *Self complementary quadruplexes*.

In the *Functional Classification*, we display the chain, or fragment of the chain, from the PDB entry as the motif, at the lowest level of the hierarchy. We provide a new feature in SCOR 2.0 that allows the user to follow a link at this motif level to the structural motif classification or the tertiary interaction classification of the element.

Tertiary Interactions

SCOR classifies the tertiary interactions responsible for the folding, stability, and maintenance of RNA three-dimensional structure into a small number of general classes including *Coaxial Helical Stacks*, *Kissing Hairpins*, *Tetraloop-Receptor*, *A-minor Motifs*, *Pseudoknots*, *tRNA D-Loop:T-Loop*, and *Ribose Zippers*. Of these, we have comprehensively examined the ribose zipper interaction, and made a detailed set of subclassifications. This type of classification is currently only partially implemented for other tertiary interactions. At the most detailed level, we display the elements of tertiary interactions as groups of structural motifs and provide links to the structural motifs.

Updated Web Interface

The web interface provides a straightforward but flexible means to explore the multi-layered classification of SCOR 2.0. Users are given a standard set of icons and buttons to navigate the classes. The interface allows the user to view multiple parallel classes, subclasses (i.e. child classes) and paths simultaneously. The user may also choose to focus on a single class or path along a branch of the hierarchy – or may choose a combination of views.

The database entries are presented in a tree view showing the path to the class from a root. By clicking + adjacent to a class, the view is expanded to display the subclasses, and by selecting – adjacent to a class, the tree is collapsed (See Figure 1.). These options expand the tree in place to

show the subclasses; the parent class and all classes at the same level remain in view. Two links are provided next to each class: the link "subclass" displays the children, and the option "list all" displays all elements of the class organized by the sequence's RNA type and species of origin. Each of these links displays a new page with all paths to the class displayed at the top of the page.

If a class has more than one parent, a designation (**o**) is displayed before +/**-**, and clicking on this designation expands allpath s to the class, with the target class emphasized in bold face. Selecting *alternate path* after the class name expands and highlights alternate paths to the class.

A textual description of each class appears when a user clicks the description icon. Literature references are displayed by selecting the reference icon. Selecting the sketch icon displays or hides two-dimensional schematic sketches for each class. RasMol (18) views are available for all examples of the motifs, and can be accessed by selecting the RasMol icon next to each description of a structural element.

Search Engine

SCOR may be searched with PDB (or NDB) identifier, PDB residue number, Sequence, or Keyword. The result of a search by PDB identifier includes a summary of the PDB entry, including links to the motif structure, function and tertiary interactions classifications. Combinations of queries, e.g. PDB identifier and Sequence, can further limit a search. Additionally, regular expressions, linked by Boolean operators, may be used with sequence and keyword searches. This is especially useful for sequence expressions with common abbreviations such as R for purine or Y for pyrimidine. The result of a sequence search returns all paths to motifs containing the sequence, and the result of a keyword search returns all paths to classes containing the keyword.

Future Directions

In the immediate future we plan to complete the classification of RNA tertiary interactions, expand the functional classification to include metal-ion binding regions, junction loops and helical regions, and include new classes as they are discovered. We will also explore additional database

search criteria, such as a secondary structure-based search. Our goal is to provide a comprehensive up-to-date classification of RNA structural motifs and tertiary interactions.

Acknowledgements

References

1   Klosterman, P. S., Tamura, M., Holbrook, S. R. and Brenner, S. E. (2002) SCOR: a
        Structural Classification of RNA database. *Nucleic Acids Res*, **30,** 392-394.

2   Tamura, M. and Holbrook, S. R. (2002) Sequence and structural conservation in RNA
        ribose zippers. *J Mol Biol*, **320,** 455-474.

3   Legault, P., Li, J., Mogridge, J., Kay, L. E. and Greenblatt, J. (1998) NMR structure of the
        bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an
        arginine-rich motif. *Cell*, **93,** 289-299.

4   Heus, H. A. and Pardi, A. (1991) Structural features that give rise to the unusual stability of
        RNA hairpins containing GNRA loops. *Science*, **253,** 191-194.

5   Holbrook, S. R., Sussman, J. L., Warrant, R. W. and Kim, S. H. (1978) Crystal structure of
        yeast phenylalanine transfer RNA. II. Structural features and functional implications. *J Mol
        Biol*, **123,** 631-660.

6   Jucker, F. M. and Pardi, A. (1995) GNRA tetraloops make a U-turn. *RNA*, **1,** 219-222.

7   Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P.,
        Dolinski, K., Dwight, S. S., Eppig, J. T.*, et al.* (2000) Gene ontology: tool for the unification
        of biology. The Gene Ontology Consortium. *Nat Genet*, **25,** 25-29.

8   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.

9   Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T.,
        Hsieh, S. H., Srinivasan, A. R. and Schneider, B. (1992) The nucleic acid database. A
        comprehensive relational database of three-dimensional structures of nucleic acids.
        Biophys J, **63,** 751-759.

10   Duarte, C. M. and Pyle, A. M. (1998) Stepping through an RNA structure: A novel
        approach to conformational analysis. *J Mol Biol*, **284,** 1465-1478.

11   Gendron, P., Lemieux, S. and Major, F. (2001) Quantitative analysis of nucleic acid
        three-dimensional structures. *J Mol Biol*, **308,** 919-936.

12   Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an
        environment for comparative protein modeling. *Electrophoresis*, **18,** 2714-2723.

13   Szewczak, A. A. and Moore, P. B. (1995) The sarcin/ricin loop, a modular RNA. *J Mol Biol*,
        247, 81-98.

14   Leontis, N. B. and Westhof, E. (1998) A common motif organizes the structure of
        multi-helix loops in 16 S and 23 S ribosomal RNAs. *J Mol Biol*, **283,** 571-583.

15   Klinck, R., Westhof, E., Walker, S., Afshar, M., Collier, A. and Aboul-Ela, F. (2000) A

potential RNA drug target in the hepatitis C virus internal ribosomal entry site. *RNA*, **6,** 1423-1431.

16   Klein, D. J., Schmeing, T. M., Moore, P. B. and Steitz, T. A. (2001) The kink-turn: a new RNA secondary structure motif. *Embo J*, **20,** 4214-4221.

17   Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T. and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407,** 327-339.

18   Sayle, R. A. and Milner-White, E. J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem Sci*, **20,** 374.

Figure Legends

Figure 1. Example of the DAG structure in SCOR. The directed acyclic graph allows multiple paths to a class. In this example, *Loops with dinucleotide platforms* are a class of internal loop with the subclass *Loops with dinucleotide platforms in a triple*. This subclass is also a subclass of *Loops with base triples*.

Table 1. Growth of the SCOR Database. The SCOR Database has grown by more than 10-fold since its inception in September 2001, reflecting the growth in the available RNA structures, particularly ribosomal RNA structures, since October 2000. SCOR 1.1 and 1.2 were based on structures available as of October 2000.

Version (Release Date)	New Features	PDB Files (Cutoff Date)	Hairpin Loops	Internal Loops
1.1 (Sep 2001)	Motif Classification	**259** (Oct 2000)	**203**	**223**
1.2 (July 2002)	Ribosomal RNA, Ribose Zippers	**261** (Oct 2000, plus 1jj2 and 1j5e)	**295**	**402**
2.0 (Sep 2003)	DAG Classification, Published Motifs	**511** (May 2003)	**2,104**	**3,776**

Figure 1.