

Updates to the RNA mapping database (RMDB), version 2

Joseph D. Yesselman^{1,†}, Siqi Tian^{1,†}, Xin Liu², Lei Shi, Jin Billy Li² and Rhiju Das^{1,3,*}

¹Department of Biochemistry, Stanford University School of Medicine, Stanford CA 94305, USA, ²Department of Genetics, Stanford University School of Medicine, Stanford CA 94305, USA and ³Department of Physics, Stanford University, Stanford, CA 94305, USA

Received August 16, 2017; Revised September 13, 2017; Editorial Decision September 16, 2017; Accepted September 19, 2017

ABSTRACT

Chemical mapping is a broadly utilized technique for probing the structure and function of RNAs. The volume of chemical mapping data continues to grow as more researchers routinely employ this information and as experimental methods increase in throughput and information content. To create a central location for these data, we established an RNA mapping database (RMDB) 5 years ago. The RMDB, which is available at <http://rmdb.stanford.edu>, now contains chemical mapping data for over 800 entries, involving 134 000 natural and engineered RNAs, *in vitro* and *in cellulo*. The entries include large data sets from multidimensional techniques that focus on RNA tertiary structure and co-transcriptional folding, resulting in over 15 million residues probed. The database interface has been redesigned and now offers interactive graphical browsing of structural, thermodynamic and kinetic data at single-nucleotide resolution. The front-end interface now uses the force-directed RNA applet for secondary structure visualization and other JavaScript-based views of bar graphs and annotations. A new interface also streamlines the process for depositing new chemical mapping data to the RMDB.

INTRODUCTION

RNA secondary and tertiary structures are critical to understanding the diverse range of cellular processes that RNA performs, including protein synthesis, gene regulation and mRNA maturation (1–3). RNA structure is being investigated using a host of biophysical and biochemical techniques (1,4,5–7). Over the past decade, chemical mapping has become a widely utilized method for determining the structural features of RNA nucleotides. Although originally devised in the 1970s (8), chemical mapping began to gain much wider use with the development of modern methods;

for example, probed RNA molecules can be reverse transcribed into DNA, which can be analyzed via sequencing to infer the nucleotides that were chemically modified (9). These new methods also include high-throughput methods that permit the generation of single-nucleotide-resolution chemical mapping data for thousands of distinct RNAs using a single-pot reaction, including entire transcriptomes probed within cells (10,11). Recently, techniques have been developed that probe the tertiary structure of RNA or monitor an RNA's secondary structure at different stages of its transcription, adding additional structural features that can be determined with chemical mapping (12,13).

The vast amount of data now being produced might enable the generation of better rules that can predict the relationships between RNA sequence and structure. Nevertheless, unlike other structural biology techniques that produce shareable data, such as nuclear magnetic resonance, crystallography and electron microscopy, there has been no equivalent of the Biological Magnetic Resonance Bank (14), Protein Data Bank (15) or Electron Microscopy Data Bank (16) for storing chemical mapping data sets with corresponding structural models. While structure mapping data are often available in the supporting material of publications, independent analyses benefit from consolidation of data, standardization of file formats, error estimates and notes on normalization. To enable such analyses, we created an RNA mapping database (RMDB) (17), a central location to deposit and search curated structure mapping measurements in both human- and machine-readable formats. In addition, the RMDB contains an interface that allows users to infer the relationships between structure and chemical mapping data. All the data contained in the RMDB are freely available and can be integrated into other repositories, such as the Protein Data Bank.

The amount of chemical mapping data deposited into the RMDB continues to rapidly increase. As of 2017, the RMDB contains data for over 134 000 RNAs. Based on the lessons learned from dealing with this significant amount of data as well as with new kinds of data and new visualization methods, we have redesigned all facets of the RMDB. It now

*To whom correspondence should be addressed. Tel: +1 650 723 5976; Email: rhiju@stanford.edu

†These authors contributed equally to the paper as first authors.

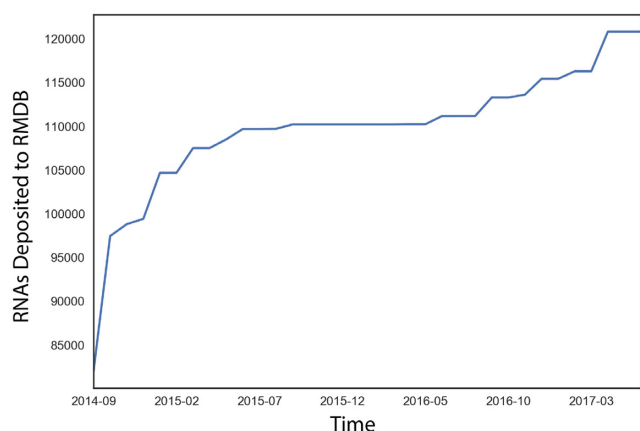


Figure 1. Recent growth in the RNA mapping databank, plotted as a cumulative distribution against deposition date.

offers improved graphical analyses of structural, thermodynamic and kinetic data, all via an interactive environment. This redesign includes a new front-end interface that provides significantly more information about the probed RNA at single-nucleotide resolution. For every entry, the RMDB provides the force-directed RNA (FORNA) applet to display the chemical mapping reactivity on secondary structures that can be interactively rearranged and an interactive heat map for reading out data values within the context of large data sets. Furthermore, the new RMDB has simplified the submission process to encourage more groups to submit their chemical mapping data.

UPDATES TO THE RMDB

New entries

There are currently 506 entries detailing 845 experimental datasets on over 134 000 RNAs and 15 million unique data points (growth of RMDB: Figure 1). To simplify browsing these data, we now organize them into distinct categories. For example, ‘RNA Puzzles’ contains the chemical mapping data generated for the community-wide RNA Puzzle structure prediction trials, starting with Puzzle 5 (18). These challenges are blind tertiary structural predictions that involve modeling groups predicting 3D models of unreleased RNA crystal structures. ‘Eterna’ archives chemical mapping data from the Eterna massive open laboratory, an internet-scale project in which participants design RNA sequences that match a given target secondary structure. These RNAs are created experimentally and probed to see how well they match the target (19). Other categories are being created such as entries with associated 3D models in the PDB. ‘General’ now includes all other chemical mapping data entries that are from published experiments, allowing users to investigate the raw data from given publications.

New interface

Since the original release of the RMDB, there have been significant improvements in web interface technology. In addition, newly developed chemical mapping techniques

now facilitate multidimensional experiments, by introducing primary sequence mutations, titrating in small molecule binders or by introducing protein road blocks to pause RNA polymerase at each sequential residue (13,20,21). To best represent these new chemical mapping data, we have redesigned the RMDB user interface using current standards in HTML5, CSS and JavaScript. This redesign ensures that all modern web browsers (Chrome, Firefox, Edge and Safari), as well as mobile browsers, have an identical experience, and all sections of the RMDB are responsive to both window and screen size. Instead of a single bar graph to represent reactivity, we now have an interactive heat map in which each row corresponds to a specific RNA sequence or condition and each column corresponds to the reactivity of each residue in a construct (Figure 2A). Furthermore, there are multiple tabs that can be accessed on the left side of the interface that give more information about experimental conditions (Figure 2B) and display a bar plot of a subset of the data for the location the user clicked on the heat map (Figure 2C).

In addition to the new interactive interface for viewing existing chemical mapping data, the RMDB has a simplified submission process for adding new chemical mapping entries (Figure 3). In addition to detailed instructions on how to generate RDAT files (the format of the RMDB) or ISA-TAB files (22), we have also developed validation tools to inspect a given RDAT file before submission; these tools can detect missing or incomplete fields.

New secondary structure viewer

The previous version of the RMDB used VARNA (23) as the RNA secondary structure viewer, but support for VARNA has been lost in most modern web browsers due to security concerns about Java plugins. We therefore upgraded RMDB to use the FORNA (24) web application to display RNA secondary structure with chemical mapping data (Figure 4). The new FORNA viewer allows RNA secondary structures to be displayed directly in a web browser without installing any software. The complete FORNA package also allows for the interactive editing of the displayed structures, the display of structures with pseudoknots, the simultaneous visualization of multiple structures and the automatic generation of secondary structure diagrams from Protein Data Bank files. These features will allow more functionality to be built into RMDB’s FORNA structure viewer in the future. The FORNA viewer in the updated RMDB displays the secondary structure model provided by the depositor and shows the experimental reactivity with a color bar scaled to the normalized data from the RDAT files deposited (Figure 4). For entries with multiple mapping datasets, the drop-down menu allows for the selection of the specific datasets to display.

CONCLUDING REMARKS

Chemical mapping has become a widely used technique for exploring both the secondary and tertiary structures of RNA. We expect its use to only increase, with more groups adopting and expanding chemical mapping techniques to supplement their other molecular and structural biology experiments. The updates presented herein bring the RMDB

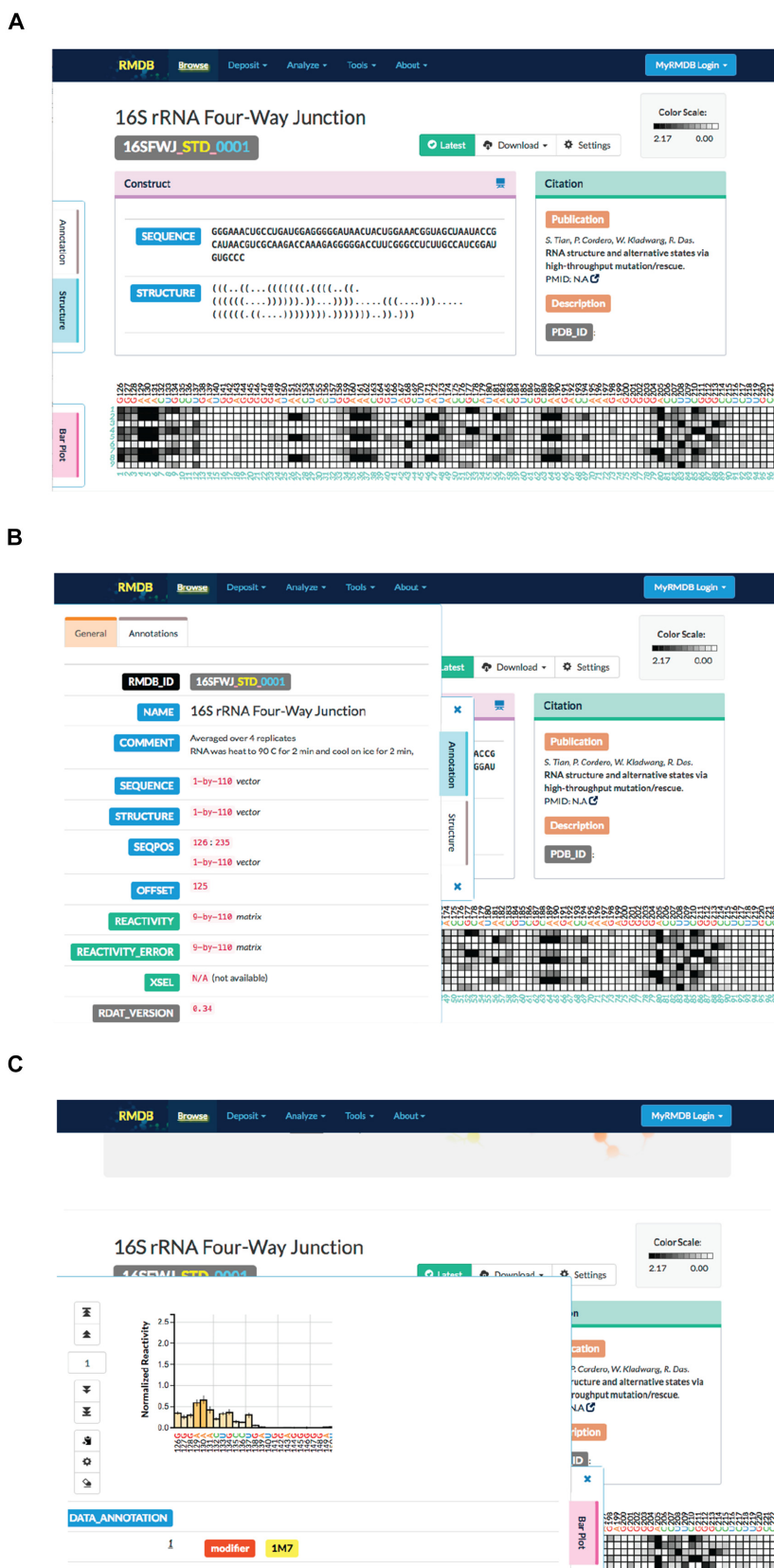


Figure 2. Screenshots of the new interactive user interface for viewing RMDB entries. (A) An example of an entry. (B) The left panel maximizes when clicked, giving additional data about the experimental conditions. (C) Clicking on a data point on the heat map opens a bar plot that gives more information about the surrounding data.

Submit RMDB Entry

New experimental data can be uploaded to the RMDB as a spreadsheet in either the SNRNASM [ISATAB](#) format or [RDAT](#) format.

Data must be reviewed and approved by RMDB team before public release. It should meet the following requirements:

- Includes error estimates;
- Contains information on estimated or known structure (at least at the level of secondary structure);
- Provides description of how the data were processed.

RDAT File

Format: Experiment Type:

RMDB_ID: [Use File Name](#)

Upload File: [Browse](#)

Description: (optional)

Associated Publication

Please specify authorship. The other publication information fields are optional. If left blank, N/A will be filled in as default.

Authors:

Pubmed ID: (optional)

Publication Title: (optional)

If this entry **supercedes** another **RMDB_ID**, please contact the admin for manual annotation.

Figure 3. The simplified submission form for adding a new entry to the RMDB.

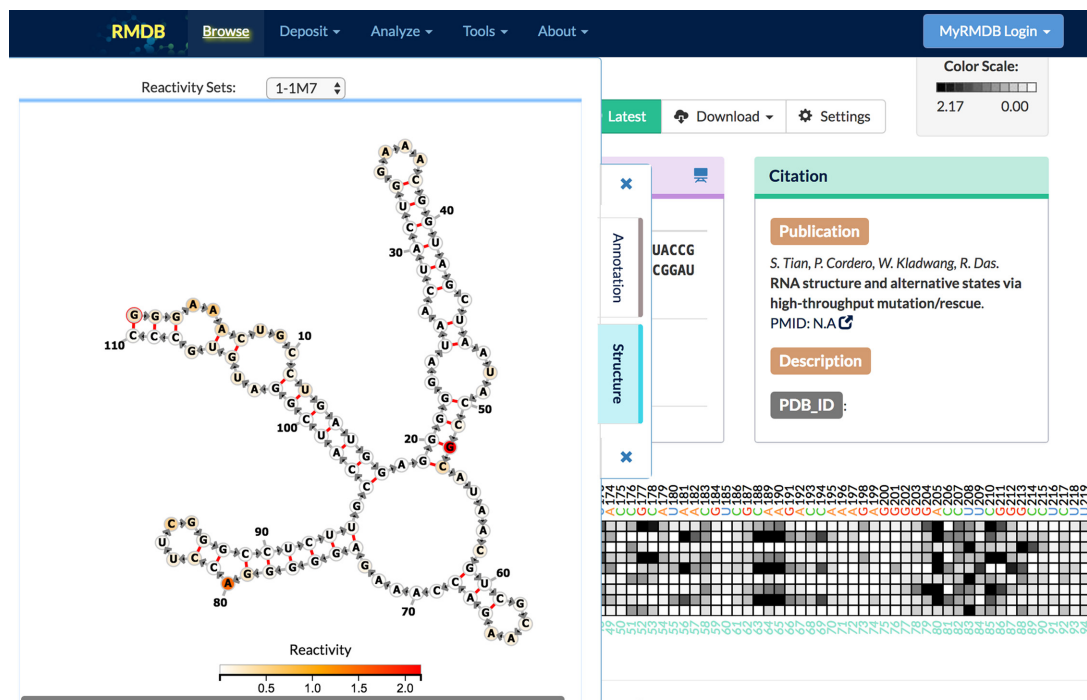


Figure 4. The FORNA JavaScript applet allows users to see each chemical mapping data set on a secondary structure supplied with the RMDB entry by the depositor.

to version 2.0. Use of current web standards and interactive interfaces for users prepare the database for more data and for further extensions to display 3D structures. In the future, we plan to link to other repositories such as RNA-central, RFAM and the PDB. The RMDB seeks to be a central location for RNA researchers to deposit their data for publication and to also serve those interested in determining the general relationship between chemical mapping data and structure using machine learning or physics-based approaches.

ACKNOWLEDGEMENTS

We would like to thank the members of the Das lab for thoughtful discussions and feedback.

FUNDING

National Institutes of Health [MIRA 1205527-100-PALHN to R.D., R01 GM100953 to R.D., R01 GM102484, GM124215 to J.B.L.]; Ruth L. Kirschstein National Research Service Award Postdoctoral Fellowships [GM112294 to J.D.Y.]; Ellison Medical Foundation [AG-NS-0959-12 to J.B.L.]. Funding for open access charge: National Institutes of Health [R01 GM100953].
Conflict of interest statement. None declared.

REFERENCES

1. Nguyen,T.H.D., Galej,W.P., Bai,X., Savva,C.G., Newman,A.J., Scheres,S.H.W. and Nagai,K. (2015) The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature*, **523**, 47–52.
2. Nissen,P., Hansen,J., Ban,N., Moore,P.B. and Steitz,T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science*, **289**, 920–930.
3. Winkler,W.C. and Breaker,R.R. (2003) Genetic control by metabolite-binding riboswitches. *ChemBiochem*, **4**, 1024–1032.
4. Adilakshmi,T., Lease,R.A. and Woodson,S.A. (2006) Hydroxyl radical footprinting in vivo: mapping macromolecular structures with synchrotron radiation. *Nucleic Acids Res.*, **34**, e64.
5. Tian,S. and Das,R. (2016) RNA structure through multidimensional chemical mapping. *Q. Rev. Biophys.*, **49**, e7.
6. Varani,G. and Tinoco,I. (1991) RNA structure and NMR spectroscopy. *Q. Rev. Biophys.*, **24**, 479–532.
7. Wilkinson,K.A., Merino,E.J. and Weeks,K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.
8. Noller,H.F. and Chaires,J.B. (1972) Functional modification of 16S ribosomal RNA by kethoxal. *Proc. Natl. Acad. Sci. U.S.A.*, **69**, 3115–3118.
9. Lucks,J.B., Mortimer,S.A., Trapnell,C., Luo,S., Aviran,S., Schroth,G.P., Pachter,L., Doudna,J.A. and Arkin,A.P. (2011) Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11063–11068.
10. Ding,Y., Tang,Y., Kwok,C.K., Zhang,Y., Bevilacqua,P.C. and Assmann,S.M. (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, **505**, 696–700.
11. Kertesz,M., Wan,Y., Mazor,E., Rinn,J.L., Nutter,R.C., Chang,H.Y. and Segal,E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **467**, 103–107.
12. Cheng,C.Y., Chou,F.-C., Kladwang,W., Tian,S., Cordero,P. and Das,R. (2015) Consistent global structures of complex RNA states through multidimensional chemical mapping. *Elife*, **4**, e07600.
13. Watters,K.E., Strobel,E.J., Yu,A.M., Lis,J.T. and Lucks,J.B. (2016) Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nat. Struct. Mol. Biol.*, **23**, 1124–1131.
14. Ulrich,E.L., Akutsu,H., Doreleijers,J.F., Harano,Y., Ioannidis,Y.E., Lin,J., Livny,M., Mading,S., Maziuk,D., Miller,Z. *et al.* (2008) BioMagResBank. *Nucleic Acids Res.*, **36**, 402–408.
15. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H. and Shindyalov,I.N. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
16. Lawson,C.L., Patwardhan,A., Baker,M.L., Hryc,C., Garcia,E.S., Hudson,B.P., Lagerstedt,I., Ludtke,S.J., Pintilie,G., Sala,R. *et al.* (2016) EMDatabank unified data resource for 3DEM. *Nucleic Acids Res.*, **44**, D396–D403.
17. Cordero,P., Lucks,J.B. and Das,R. (2012) An RNA Mapping DataBase for curating RNA structure mapping experiments. *Bioinformatics*, **28**, 3006–3008.
18. Miao,Z., Adamiak,R.W., Blanchet,M.-F., Boniecki,M., Bujnicki,J.M., Chen,S.-J., Cheng,C., Chojnowski,G., Chou,F.-C., Cordero,P. *et al.* (2015) RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA*, **21**, 1066–1084.
19. Lee,J., Kladwang,W., Lee,M., Cantu,D., Azizyan,M., Kim,H. and Limpaecher,A. (2014) RNA design rules from a massive open laboratory. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 2122–2127.
20. Kladwang,W., Chou,F.C. and Das,R. (2012) Automated RNA structure prediction uncovers a kink-turn linker in double glycine riboswitches. *J. Am. Chem. Soc.*, **134**, 1404–1407.
21. Kladwang,W., Cordero,P. and Das,R. (2011) A mutate-and-map strategy accurately infers the base pairs of a 35-nucleotide model. *RNA*, **17**, 522–534.
22. Rocca-serra,P., Bellaousov,S., Birmingham,A., Rocca-serra,P., Bellaousov,S., Birmingham,A., Chen,C., Cordero,P., Das,R., Davis-neulander,L. *et al.* (2011) Sharing and archiving nucleic acid structure mapping data. *RNA*, **17**, 1204–1212.
23. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.
24. Kerpedjiev,P., Hammer,S. and Hofacker,I.L. (2015) Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. *Bioinformatics*, **31**, 3377–3379.