BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 22 2012, pages 3006–3008
doi:10.1093/bioinformatics/bts554

Databases and ontologies

Advance Access publication September 12, 2012

An RNA Mapping DataBase for curating RNA structure mapping experiments

Pablo Cordero[1], Julius B. Lucks[2] and Rhiju Das[1,3,4,*]

[1]Department of Biochemistry and Biomedical Informatics Program, Stanford University, Stanford, CA 94305, [2]School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, [3]Department of Biochemistry and [4]Department of Physics, Stanford University, Stanford CA 94305, USA

Associate Editor: Ivo Hofacker

ABSTRACT

Summary: We have established an RNA mapping database (RMDB) to enable structural, thermodynamic and kinetic comparisons across single-nucleotide-resolution RNA structure mapping experiments. The volume of structure mapping data has greatly increased since the development of high-throughput sequencing techniques, accelerated software pipelines and large-scale mutagenesis. For scientists wishing to infer relationships between RNA sequence/structure and these mapping data, there is a need for a database that is curated, tagged with error estimates and interfaced with tools for sharing, visualization, search and meta-analysis. Through its on-line front-end, the RMDB allows users to explore single-nucleotide-resolution mapping data in heat-map, bar-graph and colored secondary structure graphics; to leverage these data to generate secondary structure hypotheses; and to download the data in standardized and computer-friendly files, including the RDAT and community-consensus SNRNASM formats. At the time of writing, the database houses 53 entries, describing more than 2848 experiments of 1098 RNA constructs in several solution conditions and is growing rapidly.

Availability: Freely available on the web at http://rmdb.stanford.edu

Contact: rhiju@stanford.edu

Supplementary information: Supplementary data are available at *Bioinformatics* Online.

Received on January 11, 2012; revised on August 17, 2012; accepted on September 5, 2012

1 INTRODUCTION

Understanding the secondary and tertiary structures of RNAs is critical for dissecting their diverse biological functions, ranging from catalysis in ribosomal RNAs to gene regulation in metabolite-sensing riboswitches and protein-binding elements in RNA messages (Eickbush and Eickbush, 2007; Kudla *et al*., 2009; Nudler and Mironov, 2004; Spahn *et al*., 2001; Yanofsky, 2004). RNA structure has therefore been intensely studied with a variety of biophysical and biochemical technologies (Adilakshmi *et al*., 2006; Getz *et al*., 2007; Waldsich, 2008; Varani and Tinoco, 1991; Wilkinson *et al*., 2006). Among these tools, a facile, information-rich and widely used technique is structure mapping (also called structure probing or footprinting), in which the chemical modification, enzymatic cleavage or degradation rate of an RNA nucleotide correlates with the exposure, flexibility or other structural features of the site. Modern methods often reverse transcribe probed RNA molecules into DNA fragments whose lengths can be subsequently analyzed to infer the locations of probe events. These methods permit the single-nucleotide resolution readout of structural data for RNAs as large as ribosomes (Deigan *et al*., 2009; Culver *et al*., 1999), and in recent years, investigators have developed high-throughput technologies, such as 96-well capillary electrophoresis (Mitra and Shcherbakova, 2008; Mortimer and Weeks, 2007) and deep sequencing (Lucks *et al*., 2011) to perform this step. Furthermore, several bioinformatic pipelines have been implemented to rapidly quantify, map and analyze the resulting data (Aviran *et al*., 2011; Deigan *et al*., 2009; Low and Weeks 2010; Vasa *et al*., 2008; Yoon *et al*., 2011). RNA mapping experiments are now routinely used to improve automated secondary structure modeling (Deigan *et al*., 2009), probe entire viral genomes (Watts and Dang, 2009), simultaneously map arbitrary RNA mixtures through deep sequencing (Kertesz *et al*., 2010; Lucks *et al*., 2011; Underwood *et al*., 2010; Zheng *et al*., 2010) and infer an RNA's 'contact map' by coupling to exhaustive single-nucleotide mutagenesis (Kladwang and Das, 2010; Kladwang *et al*., 2011).

These developments could enable novel methods in RNA structural biology, especially if predictive relationships between RNA sequence/structure and these data can be established. However, unlike existing structural biology fields like nuclear magnetic resonance and crystallography, there is no equivalent of the Biological Magnetic Resonance Bank (Ulrich *et al*., 2007) or the Protein Data Bank (Bernstein *et al*., 1977) that stores curated datasets. Structure mapping data are available in the supporting material of papers or self-reported in SNRNASM format (Rocca-Serra *et al*., 2011), but these formats do not typically include error estimates; are not always normalized or background-subtracted with standardized protocols; are not linked to RNA structures and are not straightforward to visualize, which would enable consistency checks during further analysis. We have therefore created an RNA mapping database (RMDB) and are populating it with curated structure mapping measurements in human and machine-readable formats amenable to inferring relationships between sequence/structure and structure mapping data. Data contained in the RMDB are freely available and can be easily integrated with future repositories such as RNAcentral (Bateman *et al*., 2011).

*To whom correspondence should be addressed.

2 DATABASE CONTENT AND STRUCTURE

For a specific RNA in defined solution conditions, each structure mapping experiment can be conceptualized as $M \times N$ matrices, where M is the number of measurements made on the RNA, e.g. normalized peak areas calculated by HiTRACE (Yoon *et al.*, 2011), CAFA (Mitra *et al.*, 2008) and ShapeFinder (Vasa *et al.*, 2008); or maximum likelihood parameters (Aviran *et al.*, 2011) and N is the number of nucleotides in the RNA. Entries in the RMDB house these data matrices and are enriched with annotations and free text to describe associated content.

The database currently includes experiments using base methylation by dimethyl sulfate, base adduct formation by 1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulfonate, selective 2′ hydroxyl acylation with primer extension (SHAPE) with either *N*-methylisatoic anhydride or 1-methyl-7-nitroisatoic anhydride and hydroxyl radical footprinting. The RMDB is further capable of storing data from enzymatic, in-line and other structure mapping experiments. The RNAs probed include riboswitches, tRNAs, ribozyme and ribosomal domains featured in several published studies as well as human-designed sequences accruing in the internet-scale RNA engineering project EteRNA (http://eterna.stanford.edu) (see Supplementary Table S2). For many RMDB entries, the complementary DNA fragment separation and analysis steps were carried out by 96-well format capillary electrophoresis and the HiTRACE pipeline (Yoon *et al.*, 2011), respectively; entries describing data from (Lucks *et al.*, 2011) were read out and processed using the SHAPE-Seq protocol (Aviran *et al.*, 2011).

New experimental data can be uploaded to the RMDB as a spreadsheet in the SNRNASM Isa-Tab format (Rocca-Serra *et al.*, 2011) or in RDAT file format (a simple text file format; detailed in the Supplement Information and at http://rmdb.stanford.edu/repository/specs/). However, like the well-curated Protein DataBank, public release then requires passing review by the RMDB team; the data must include error estimates or replicates, information on estimated or known structure (at least at the level of secondary structure), associated publications or preprints and descriptions of how the data were processed.

3 FEATURES AND EXAMPLE USE CASES

3.1 For experimentalists

The RMDB is a resource for RNA biochemists and molecular biologists interested in using existing data to guide biological hypotheses, interpret new data or to share their own experimental results. First, users interested in a particular RNA system can quickly find relevant data in the RMDB by using the full-text search field in the upper-right corner of the site. The user can then inspect each entry using the data visualization tools (Fig. 1a and b; Supplementary Material). Second, the integration with the VARNA applet allows for quickly comparing mapping data against structural models. The data can be downloaded in either RDAT or SNRNASM format or exported directly from the VARNA visualization applet for further inspection with other tools. Third, the RMDB also includes a secondary structure prediction server (located at http://rmdb.stanford.edu/structureserver); the server can use structure mapping data to generate sensible secondary structure hypotheses (see Supplementary Material). Finally, experimentalists who wish to submit their data to the RMDB can do so after registering to the site.

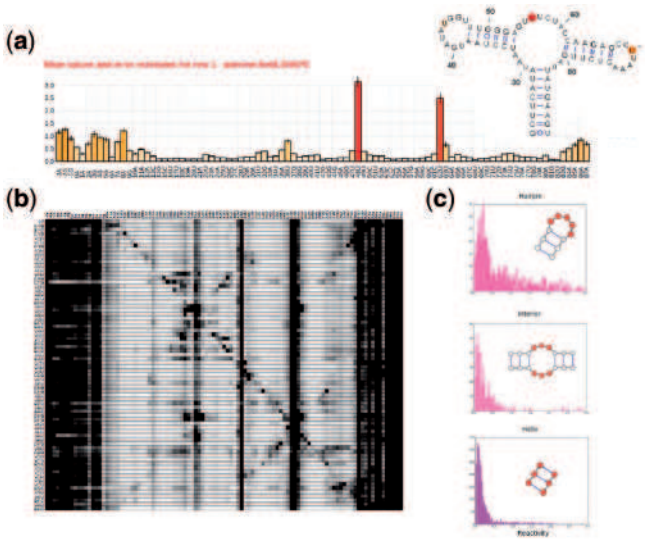


Fig. 1. Different visualization tools for entries in the RNA mapping database. (a) Classic bar plot of 2′-OH acylation (SHAPE) rates across the nucleotides of the adenine-sensing domain of the *add* riboswitch from *Vibrio vulnificus*. The data are from a 'standard state' study averaging 19 replicates across multiple experiments and estimating the resulting errors (shown as error bars); the RNA's crystallographic secondary structure, colored by the SHAPE data, is shown in the inset. (b) Through mutate-and-map data, the RMDB also allows exploring the contact map of the same riboswitch. SHAPE data are shown for constructs with single mutations at each RNA position. (c) Histograms for reactivities found in interior loops, hairpin loops and helices. Nucleotides in the motif for which reactivity data were collected are marked in red. Hairpin loops have higher average reactivities than interior loops, bulges and non-helical elements

3.2 For structural bioinformaticists

The RMDB can extract general properties of mapping data including histograms of reactivities for different secondary structure elements (Fig. 1c). Analyses of structure mapping data are facilitated by the Python/MATLAB RDATkit package (http://rdatkit.simtk.org, see Supplementary Material) for RDAT/SNRNASM-IsaTAB parsing. To demonstrate the utility of the RMDB in extracting new information from multiple datasets, we tested whether the SHAPE method can discriminate between interior and hairpin loops. Using the advanced search feature of the database, we downloaded SHAPE data for each secondary structure element (internal loops, hairpins, helices and bulges) collected in standard state experiments for all non-coding RNAs (ncRNAs) with known structure in the database. Interior and hairpin loops of ncRNA have distinct reactivity distributions, suggesting that SHAPE-directed modeling can be made more accurate by taking this effect into account (Fig. 1c).

3.3 For web-app developers

Data stored in the RMDB are exposed through a RESTful API (described in https://sites.google.com/site/rmdbwiki/web-api) in JSON format, simplifying the creation of web applications that

use the data contained in the repository. The RMDB also provides RSS feeds that are automatically updated with new entries (see https://sites.google.com/site/rmdbwiki/rss for details). These tools have allowed integration of the entries in the RMDB into the SNRNASM repository (http://snrnasm.bio.unc.edu/browse.html).

4 DISCUSSION

The throughput of structure mapping experiments has taken significant leaps with multiplexed capillary electrophoresis and next-generation sequencing that allow probing of thousands of RNAs at once. These data should, in principle, permit the development of confident structural biology tools that couple structure mapping measurements to secondary and tertiary structure modeling. However, until recently, researchers have had few resources that enable curation and sharing of high-throughput quantified RNA mapping data. It is our hope that the RMDB will make such projects possible.

ACKNOWLEDGEMENTS

We thank the authors of RNAstructure, VARNA and the Stanford Visualization Team for making their source code freely available, the Laederach laboratory for aid with SNRNASM integration, the Clote laboratory for data presentation comments and suggestions, R. Astorga for design comments and the members of the Das laboratory and C.C. VanLang for help with data curation and manuscript preparation. The server that houses the RMDB was graciously donated by K. Beauchamp.

Funding: The Burroughs-Wellcome Foundation (CASI to R.D.), a Hewlett-Packard Stanford Graduate Fellowship (to P.C.) and the National Institutes of Health (R01 GM102519 to R.D.).

Conflict of Interest: none declared.

REFERENCES

Adilakshmi,T. *et al.* (2006) Hydroxyl radical footprinting *in vivo*: mapping macromolecular structures with synchrotron radiation. *Nucleic Acids Res.*, **34**, e6.

Aviran,S. *et al.* (2011) Modeling and automation of sequencing-based characterization of RNA structure. *Proc. Natl Acad. Sci.*, **108**, 11069–11074.

Bateman,A. *et al.* (2011) RNAcentral: a vision for an international database of RNA sequences. *RNA*, **17**, 1941–1946.

Bernstein,F.C. *et al.* (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, **112**, 535–542.

Culver,G. *et al.* (1999) Identification of an RNA–protein bridge spanning the ribosomal subunit interface. *Science*, **285**, 2133–2135.

Darty,K. *et al.* (2009) VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.

Das,R. *et al.* (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods*, **7**, 291–294.

Deigan,K.E. *et al.* (2009) Accurate SHAPE directed RNA structure determination. *Proc. Natl Acad. Sci.*, **106**, 97–100.

Eickbush,T.H. and Eickbush,D. (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics*, **175**, 477–485.

Getz,M. *et al.* (2007) Review NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. *Biopolymers*, **86**, 384–402.

Kertesz,M. *et al.* (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **267**, 103–107.

Kladwang,W. *et al.* (2011) A mutate-and-map strategy accurately infers the base pairs of a 35-nucleotide model RNA. *RNA*, **17**, 522–534.

Kladwang,W. and Das,R. (2010) A mutate-and-map strategy for inferring base pairs in structured nucleic acids: proof of concept on a DNA/RNA helix. *Biochemistry*, **49**, 7414–7416.

Kladwang,W. *et al.* (2011) Two-dimensional chemical mapping of non-coding RNAs. *Nat. Chem.*, **3**, 954–962.

Kladwang,W. *et al.* (2011) Understanding the errors of SHAPE-directed RNA structure modeling. *Biochemistry*, **50**, 8049–8056.

Kudla,G. *et al.* (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*, **324**, 255–258.

Low,J.T. and Weeks,K.M. (2010) SHAPE-directed RNA secondary structure prediction. *Methods*, **52**, 150–158.

Lucks,J.B. *et al.* (2011) Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci.*, **108**, 11063–11068.

Mathews,D. and Turner,D. (2006) Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.*, **16**, 270–278.

Mitra,S. *et al.* (2008) High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Res*, **36**, e63.

Mortimer,S.A. and Weeks,K.M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE Chemistry. *J. Am. Chem. Soc.*, **129**, 4144–4145.

Nudler,E. and Mironov,A.S. (2004) The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.*, **29**, 11–17.

Rocca-Serra,P. *et al.* (2011) Sharing and archiving nucleic acid structure mapping data. *RNA*, **17**, 1204–1212.

Spahn,C.M.T. *et al.* (2001) Hepatitis C virus IRES RNA-induced changes in the conformation of the 40S ribosomal subunit. *Science*, **291**, 1959–1962.

Underwood,J.G. *et al.* (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods*, **7**, 995–1001.

Ulrich,E.L. *et al.* (2007) BioMagResBank. *Nucleic Acids Res.*, **36**, D402–D408.

Varani,G. and Tinoco,I. (1991) RNA structure and NMR spectroscopy. *Quart. Rev. Biophys.*, **24**, 479–532.

Vasa,S.M. *et al.* (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*, **14**, 1979–1990.

Waldsich,C. (2008) Dissecting RNA folding by nucleotide analog interference mapping (NAIM). *Nat. Protoc.*, **3**, 811–823.

Watts,J.M. *et al.* (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.

Wilkinson,K.A. *et al.* (2006) Selective 29-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.

Yanofsky,C. (2004) The different roles of tryptophan transfer RNA in regulating trp operon expression in E. coli versus B. subtilis. *Trends Genet.*, **20**, 367–374.

Yoon,S. *et al.* (2011) HiTRACE: high-throughput robust analysis for capillary electrophoresis. *Bioinformatics*, **27**, 1798–805.

Zheng,Q. *et al.* (2010) Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in *Arabidopsis*. *PLoS Genet.*, **6**, e1001141.

Supplemental Data and Information

For

An RNA Mapping Database for curating RNA structure mapping experiments

Pablo Cordero[1], Julius B. Lucks[2] and Rhiju Das[1,3,*]

[1]Program in Biomedical Informatics, Stanford University, Stanford CA 94305
[2] School of Chemical and Biomolecular Engineering, Cornell University, Ithaca NY 14853
[3]Departments of Biochemistry & Physics, Stanford University, Stanford CA 94305

** To whom correspondence should be addressed: rhiju@stanford.edu. Phone: (650) 723-5976. Fax: (650) 723-6783.*

The RDAT format

The RMDB makes use of the RDAT format for both its database schema (see Figure S-1) and for publishing its entries. The RNA Data (RDAT) text file format is an annotation-based specification that is both computer friendly and can be easily read by humans. RDAT files are composed of three main sections. The general section contains information about the RDAT specification version (RDAT_VERSION) used and serves as the root of the document. The construct section describes the RNA molecule that was probed (NAME) and lists information about its sequence (SEQUENCE), a user-supplied secondary structure model (STRUCTURE), solution conditions (ANNOTATION) and contains additional free text comments (COMMENTS). Optional fields can specify a sequence index offset for custom/consensus numbering (OFFSET); the sequence positions that the data describes (SEQPOS); and the mutation positions of constructs in the data (MUTPOS), if there are any. Finally, the data themselves are encapsulated in data sections, with two required lines, ANNOTATION_DATA (attributes

contained in name-value pairs, e.g., 'temperature:24C') and REACTIVITY (space-separated reactivity values). For analysis of multiple sets of data, error intervals can be reported under the REACTIVITY_ERROR data headers. The file may contain, optionally, raw data, including electropherogram traces (TRACE, such as in entry http://rmdb.stanford.edu/repository/detail/MDLOOP_SHP_0002) and the positions in the trace that were called as peaks (XSEL).

Annotations are a central part of the RDAT file format as they give information about the biological and technical context in which the experiment was performed. Annotations are composed of a name and possibly multiple values, separated by a colon (e.g. modifier: DMS) and are listed as space-separated values after the ANNOTATION or ANNOTATION_DATA headers. Annotations are hierarchical: an annotation in the ANNOTATION attribute in the construct section describes all data contained in that section unless an annotation with the same name in a particular ANNOTATION_DATA attribute overrides it. Table 1 describes the annotation syntax for the most common descriptors of an RNA structure mapping experiment. We have strived to make this annotation list sufficiently comprehensive as to encapsulate the experiments in our laboratories, but the list may not fulfill all the needs for experimentalists to accurately describe their data. Contributors are encouraged to contact the RMDB team for inclusions of more annotation types into the specification.

Complete description of the RDAT file format specification is given at http://rmdb.stanford.edu/repository/specs; see also Table S1.

Visualization Front-End

Each RMDB entry has its own details page, where its annotations, comments, associated publications, and authors are summarized. Furthermore, the entry's data matrix is displayed as a heat-map with more reactive residues given in darker shades of gray. Rows of the matrix represent individual chemical mapping experiments and columns represent nucleotide positions. This visualization mimics the familiar gel bands of a classic electrophoresis experiment and can be used to spot differences between traces, reads, or peak areas. Bar plots are also provided for each experiment, and can be browsed using the arrows of the plot visualizer. The accepted or target secondary structure of the RNA (supplied in the RMDB entry structure field) is drawn using the VARNA visualization package. Taking advantage of modern web graphics tools, each plot is rendered in real time as a support vector graphic (SVG) directly into the browser's canvas.

Implementation

The paradigm of modern web applications, which mainly revolves around the Model-View-Controller concept, allows for rapid prototyping and implementation of database visualizers using standard development frameworks. In our case, we leveraged the power of the Django python web framework for server side logic, hosted by an Apache 2.2/MySQL 14.1 setup. The client-side visualization tools make use of JavaScript and SVG, supported by libraries such as jQuery, protovis, and D3. To accelerate loading times, images commonly used in the visualizer and data browser (such as thumbnails and images for electropherogram traces) are pre-generated, using a combination of the VARNA visualization tool (Darty, Denise, et al., 2009), imagemagick, and Python's matplotlib. The VARNA applet is also used in the client side to display RNA structures interactively.

All data can be browsed freely without registration. The site has been tested on Firefox 4+, Internet Explorer 8 with the Google Chrome Frame plugin installed, Chrome 6+, and Safari 5+. When viewed in browsers that do not support native vector graphics, the RMDB site displays a warning and replaces the dynamic visualization elements with static images.

New Tools for Analysis and Meta-analysis

In addition to enabling access to chemical mapping data, the RMDB provides several tools to handle RNA mapping data, including a secondary structure prediction server. This tool is available at http://rmdb.stanford.edu/structureserver/ and is mainly used to predict RNA secondary structure guided by experimental bonuses with the RNAstructure (v5.3) package (Deigan, Li, 2009; Kladwang, VanLang, et al., 2011). These bonuses can be uploaded in RNAstructure format, in an RDAT file, or contained in an RMDB entry. Bonuses can be either one-dimensional [such as the pseudoenergy bonuses described in (Deigan, Li, 2009)] or two dimensional [for mutate-and-map experiments, such as in (Kladwang, VanLang, et al., 2011)]. Other options, such as data normalization [as performed in (Deigan, Li, 2009)] and bootstrapping (Kladwang, VanLang, et al., 2011) are also included.

The server and database applications leverage several Python modules and MATLAB scripts that serve as object oriented interfaces to popular secondary structure prediction algorithms and include parsers for the RDAT and Isa-TAB formats. We found that these software tools make it easier to handle data contained in these formats and to perform secondary structure modeling and visualization. Therefore, we bundled them into a programmer's toolkit named the RDATkit, which can be downloaded from its SVN repository (see instructions at http://rmdb.stanford.edu/repository/tools/) Both the structure server and the RDATkit are meant

to facilitate the use of the data available in the RMDB in generating and testing structural hypothesis.

Supplementary Tables

Table S-1: Some possible annotations, their respective SNRNASM terms, and RDAT syntax used in the Stanford RMDB.

Description of annotation	RDAT syntax	SNRNASM terms
Modified sample using dimethyl sulfate	modifier:DMS	DMS:OBI:001015
Modified sample using 1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-p-toluene sulfonate	modifier:CMCT	CMCT:OBI:0001006
Modified sample using N-methylisatoic anhydride	NMIA:OBI:0001026	modifier:NMIA, modifier:SHAPE
Modified sample using 1-methyl-7-nitroisatoic anhydride	1M7:in process	modifier:1M7, modifier:SHAPE
Modified sample using RNase T1	modifier:RNaseT1	T1:OBI0001030
Modified sample using an enzyme or chemical	See SNRNASM ontology for modifier identifiers	modifier:modifierName
Unmodified sample	modifier:None, modifier:Nomod	-
Chemical added to samples	chemical:chemicalName:concentration	-
Secondary RNA or DNA sequence added to samples	chemical:sequence:concentration	-
RNA sequence changed in sample	sequence:RNASequence	-
Specify temperature used	temperature:temperatureInCelsiusC	-
Specify processing method or algorithm used (such as background substraction or overmodification correction)	processing:methodName	-
Indicate that the sequence was mutated, from N to M at nucleotide i (1-based sequence indexing). M and N are IUPAC codes for	mutation:NiM	-

ribonucleoside resiudes.		
Indicate that the sequence was mutated from NNN to MMM at nucleotide range i to j (1-based indexing). M and N are IUPAC codes for ribonucleoside resiudes.	mutation:NNN(i-j)MMM	-
Specify experiment type (currently supported: standard state, mutate and map, and titration)	experimentType:StandardState[1], experimentType:MutateAndMap, or experimentType:Titration	-
Indicate that a base is modified in some manner at position i. Modifications include 2'-O-methylation, N1-methylation, and N3-methylation.	modification:modification:i	-
For chemical footprinting experiments, indicate the binding site (positions i to j) for the reverse transcription primer. If this annotation is not present and the modifier is not an enzyme, the binding site will be assumed to be the last 20 nucleotides of the RNA.	primer:(i-j)	-
Specify the source from where the sequence was extracted. Acceptable sources are pdb:pdbid or even doi:doi. If only part of the sequence was extracted from the indicated source, a position interval (i-j) that corresponds to the extracted subsequence from the original sequence can be specified.	sequenceSource:source:identifier or sequenceSource:source:identifier:(i-j)	-

[1] Standard state refers to a standard structural mapping experiment for an RNA molecule that can be used as a reference. A standard state experiment usually involves probing the RNA with and without a modifying agent under standard buffer conditions (i.e. without additional chemicals, sequence mutations, or biological agents).

Table S-2: RNAs, solvent conditions, chemical modifiers, and types of experiments that are currently featured in the RNA Mapping Database.

RNA	Length	Chemical Modifiers	Solvent Conditions	Experiment types
5S RNA, E. Coli	170	NMIA	10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0	Standard State, Mutate and Map (Kladwang, VanLang, et al., 2011; 122 mutants, 3 standard state measurements)
adenine riboswitch, add	121	NMIA	10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0, 0 and 5 mM adenine	Standard State, Mutate and Map (Kladwang, VanLang, et al., 2011; 73 mutants in three different conditions, 3 standard state measurements)
cidGMP riboswitch, V. Cholerae	135	NMIA	10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0, 0 and 10 uM ci-dGMP,	Standard State, Mutate and Map (Kladwang, VanLang, et al., 2011; 97 mutants in 2 different conditions, 3 standard state measurements)
glycine riboswitch, F. nucleatum	198	NMIA	10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0, 0 and 10 mM glycine	Standard State, Mutate and Map (Kladwang, VanLang, et al., 2011; 162 mutants in 4 different conditions, 3

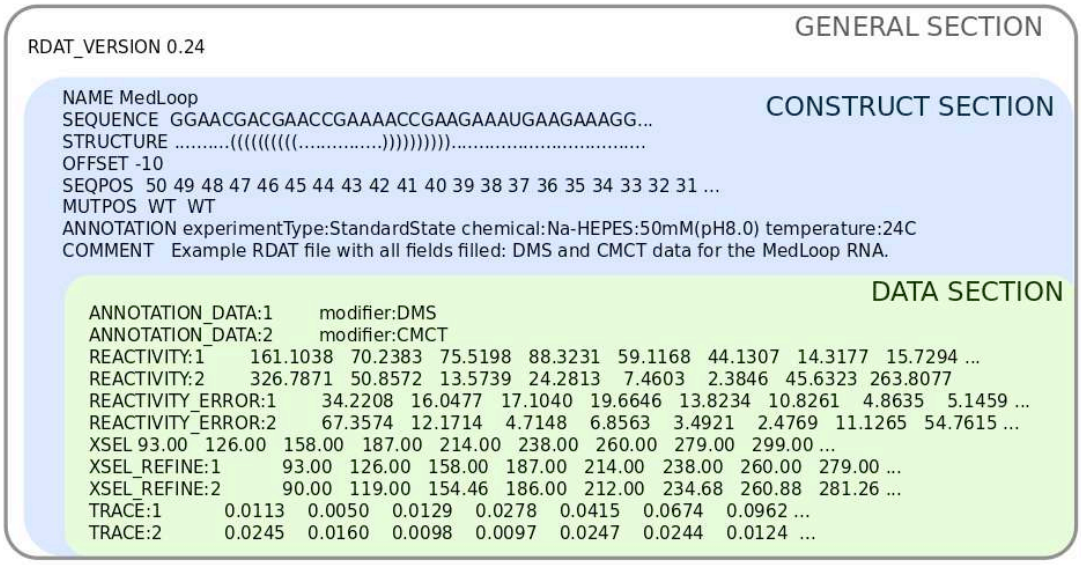
				standard state measurements)
P4-P6 domain, Tetrahymena ribozyme	202	NMIA, hydroxyl radical	10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0, 0 and 30% methylpentanediol,	Standard State, Mutate and Map (Kladwang, VanLang, et al., 2011; 161 mutants in 2 different conditions, 3 standard state measurements)
Ribonuclease P specificity domain, B. subtilis	198	1M7	10 mM $MgCl_2$, 100 mM NaCl, 100 mM Na-HEPES, pH 8.0	Standard State (Lucks, et al., 2011; 257 different barcoded constructs)
SRP Domain IV	78	DMS	0–10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0	Titration (Das, Karanicolas, et al., 2010; 3 mutants probed)
tRNA phenylalanine (yeast)	116	NMIA	10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0	Standard State, Mutate and Map (Kladwang, VanLang, et al., 2011; 81 mutants, 3 standard state measurements)
X20/H20	40	DMS	50 mM Na-HEPES, pH 8.0	Mutate and Map (Kladwang and Das., 2010; 1 construct probed)
MedLoop	80	NMIA, DMS, CMCT	50 mM Na-HEPES, pH 8.0	Standard State, Mutate and Map (Kladwang, Cordero, et al., 2011; 121 mutants in 6 different

				conditions)
One Bulge Cross	105	NMIA	10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0	Standard State (from the EteRNA community; 10 constructs)

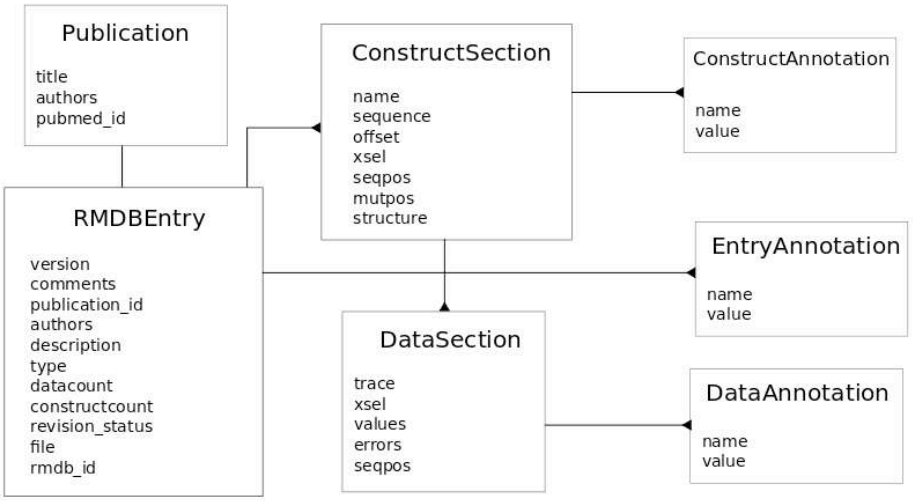
Supplementary Figures

Figure S-1 Data format and database schema.(a) The RNA Data (RDAT) file format is designed to store RNA mapping data in a standardized manner. (b) An entity-relationship representation of the RMDB database schema, based on the RDAT file format.

a)



b)



References

Adilakshmi T., Lease R.A., Woodson S.A. (2006) Hydroxyl radical footprinting in vivo: mapping macromolecular structures with synchrotron radiation. *Nucleic Acids Res*. **34**, e6

Aviran S., *et al.* (2011) Modeling and automation of sequencing-based characterization of RNA structure. *Proc. Nat. Acad. Sci.* **108**, 11069-11074

Culver G., Cate J., Yusupova G., Yusupov M., Noller H. (1999) Identification of an RNA-Protein Bridge Spanning the Ribosomal Subunit Interface. *Science*. **285**, 2133-2135

Darty K., Denise A., Ponty Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*. **25**, 1974-1975

Das R., Karanicolas J., Baker D. (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods*. **7**, 291-294

Deigan K.E., Li T.W., Mathews D.H., Weeks K.M.. (2009) Accurate SHAPE directed RNA structure determination. *Proc. Nat. Acad. Sci.* **106**, 97-10

Doi Y. and Arakawa Y. (2007)16S Ribosomal RNA Methylation: Emerging Resistance Mechanism against Aminoglycosides. *Clin. Infect. Dis*. **45**, 88-94

Eickbush T.H. and Eickbush D. (2007) Finely Orchestrated Movements: Evolution of the Ribosomal RNA Genes. *Genetics*. **175**,477-485

Getz M., Sun X., Casiano-Negroni A, Zhang Q, Al-Hashimi HM. (2007) Review NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings, *Biopolymers*. **86**,384–402

Kertesz M., Wan Y., *et al.* (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature.* **267**, 103-107

Kladwang W., Cordero P., Das R. (2011) A mutate-and-map strategy accurately infers the base pairs of a 35-nucleotide model RNA. *RNA*. **17**, 522-534

Kladwang W. and Das R. (2010) A mutate-and-map strategy for inferring base pairs in structured nucleic acids: Proof of concept on a DNA/RNA helix. *Biochemistry.* **49**, 7414-7416

Kladwang W., VanLang C., Cordero P., Das R. (2011) Two-dimensional chemical mapping of non-coding RNAs. *Nat. Chemistry.* **3**, 954–962

Kladwang W., VanLang C., Cordero P., Das R. (2011) Understanding the errors of SHAPE-directed RNA structure modeling. *Biochemistry*. **50**, 37, 8049–8056

Kudla G., Murray A., Tollervey D., Plotkin J. (2009) Coding-Sequence Determinants of Gene Expression in Escherichia coli. *Science.* **324**, 255-258

Low J.T. and Weeks K.M. (2010), SHAPE-directed RNA secondary structure prediction. *Methods.* **52**, 150-158

Lucks J.B., *et al.* (2011) Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Nat. Acad. Sci.* **108**, 11063-11068

Mandal M., *et al.* (2004) A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression. *Science.* **306**, 275-279

Mathews D. and Turner D. (2006) Prediction of RNA secondary structure by free energy minimization. *Current Opinion in Structural Biology.* **16**, 270-278

Mitra S., Shcherbakova I.V., Altman R.B., Brenowitz M., Laederach A. (2008) High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Res.* **36**, e63

Mortimer S.A. and Weeks K.M. (2007) A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry. *J. Am. Chem. Soc.* **129**, 4144-4145

Nudler E. and Mironov A.S. (2004) The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.* **29**, 11-17

Rocca-Serra P., *et al.* (2011) Sharing and archiving nucleic acid structure mapping data. *RNA,* **17**, 1204-1212

Spahn C.M.T., Kieft J.S., Grassucci R.A., Penczek P.A., Zhou K., Doudna J.A., Frank J. (2001) Hepatitis C Virus IRES RNA-Induced Changes in the Conformation of the 40S Ribosomal Subunit, *Science.* **291**, 1959-1962

Varani G. and Tinoco I. (1991) RNA structure and NMR spectroscopy. *Quart. Rev. Bioph.* **24**, 479-532

Vasa S.M., Guex N., Wilkinson K.A., Weeks K.M., Giddings M.C. (2008) ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA.* **14**, 1979-1990.

Waldsich C. (2008) Dissecting RNA folding by nucleotide analog interference mapping (NAIM). *Nat. Protoc.* **3**, 811–823

Watts J.M., Dang K.K., Gorelick R.J., Leonard C.W., Bess J.W. Jr., Swanstrom R., Burch C.L., Weeks K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature.* **460**, 711–716.

Wilkinson K.A., Merino E.J., Weeks K.M. (2006) Selective 29-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **1**, 1610–1616.

Yanofsky C. (2004) The different roles of tryptophan transfer RNA in regulating trp operon expression in E. coli versus B. subtilis. *Trends Genet.* **20**, 367-374

Yoon S., Kim J., Hum J., Kim H., Park S., Kladwang W., Das R. (2011) HiTRACE: high-throughput robust analysis for capillary electrophoresis. *Bioinformatics.* **27**, 1798-805