Automated 3D RNA Structure Prediction Using the RNAComposer Method for Riboswitches[1]

K.J. Purzycka*, M. Popenda*, M. Szachniuk*,†, M. Antczak†, P. Lukasiak*,†, J. Blazewicz*,†, R.W. Adamiak*,†,2

*Department of Structural Chemistry and Biology of Nucleic Acids, Institute of Bioorganic Chemistry Polish Academy of Sciences, Poznan, Poland
†European Center for Bioinformatics and Genomics, Institute of Computing Science, Poznan University of Technology, Poznan, Poland
2Corresponding author: e-mail address: adamiakr@ibch.poznan.pl

Contents

Abstract

Understanding the numerous functions of RNAs depends critically on the knowledge of their three-dimensional (3D) structure. In contrast to the protein field, a much smaller number of RNA 3D structures have been assessed using X-ray crystallography, NMR spectroscopy, and cryomicroscopy. This has led to a great demand to obtain the RNA 3D

[1]This work is dedicated to Professor Colin B. Reese (FRS) on the occasion of his 85th birthday anniversary.

Methods in Enzymology
ISSN 0076-6879
http://dx.doi.org/10.1016/bs.mie.2014.10.050

structures using prediction methods. The 3D structure prediction, especially of large RNAs, still remains a significant challenge and there is still a great demand for high-resolution structure prediction methods. In this chapter, we describe RNAComposer, a method and server for the automated prediction of RNA 3D structures based on the knowledge of secondary structure. Its applications are supported by other automated servers: RNA FRABASE and RNApdbee, developed to search and analyze secondary and 3D structures. Another method, RNAlyzer, offers new way to analyze and visualize quality of RNA 3D models. Scope and limitations of RNAComposer in application for an automated prediction of riboswitches' 3D structure will be presented and discussed. Analysis of the cyclic di-GMP-II riboswitch from *Clostridium acetobutylicum* (PDB ID 3Q3Z) as an example allows for 3D structure prediction of related riboswitches from *Clostridium difficile 4*, *Bacillus halodurans 1*, and *Thermus aquaticus Y5.1* of yet unknown structures.

1. INTRODUCTION

RNAs play the numerous and important roles in all major life processes (Gesteland, Cech, & Atkins, 2005). Thorough understanding of RNA functions depends critically on informative three-dimensional (3D) structure. However, it is difficult to assess 3D structures of large RNAs experimentally. In contrast to the protein field, where nearly hundred thousand structures have been deposited in the Protein Data Bank (Rose et al., 2011), a much smaller number of RNA 3D structures have been assessed experimentally (ca. 2300 entries) using X-ray crystallography, NMR spectroscopy, and, more recently, cryomicroscopy. Prediction of the RNA secondary structure using *in silico* methods is very advanced (Puton, Kozlowski, Rother, & Bujnicki, 2013; Xu, Almudevar, & Mathews, 2012) and has been recently reinforced by incorporating restraints from chemical probing methods (Mathews et al., 2004), such as SHAPE (Merino, Wilkinson, Coughlan, & Weeks, 2005). A growing number of the reasonably accurate secondary structures of large RNAs (Pang, Elazar, Pham, & Glenn, 2011; Watts et al., 2009; Wilkinson et al., 2008) created great demand in the RNA community to predict their 3D structures using computational methods (Jossinet, Ludwig, & Westhof, 2010; Leontis & Westhof, 2012; Martinez, Maizel, & Shapiro, 2008). However, 3D structure prediction still remains a significant challenge, even applying experimental restraints data (Seetin & Mathews, 2011). This situation clearly makes any RNA structure–function relationship studies difficult.

An automated prediction of the RNA 3D structure appeared of considerable and fast-growing interest. In the last years, few web-accessible tools

have been proposed for automated prediction of the RNA 3D structure. They operate on broad spectrum of input data, such as sequence, secondary structure, conformational constrains, or structural templates. Considerable differences are also visible in terms of prediction quality, which depends on the RNA strand length and topology, processor time needed, and degree of automation. Physics-based automated methods use the coarse-grained and atomic-level molecular dynamics (Cao & Chen, 2011; Jonikas, Radmer, & Altman, 2009; Jonikas, Radmer, Laederach, et al., 2009; Sharma, Ding, & Dokholyan, 2008; Xu, Zhao, & Chen, 2014), internal coordinate space dynamics (Flores & Altman, 2010; Flores, Sherman, Bruns, Eastman, & Altman, 2011), and fragment assembly (Das, Karanicolas, & Baker, 2010; Parisien & Major, 2008). Full-atomic structure predictions based on dynamics and fragment assembly are powerful tools for modeling relatively complex but small RNAs. Despite the computational cost, the coarse-grained molecular dynamics can access larger RNAs but requires demanding and not fully resolved addition of atomic details to coarse-grain models (Jonikas, Radmer, & Altman, 2009; Jonikas, Radmer, Laederach, et al., 2009). Knowledge-based comparative modeling (Rother, Rother, Puton, & Bujnicki, 2011) depends on the access to 3D structural templates and unequivocal sequence alignment. Prior to reporting on our RNAComposer method (Popenda et al., 2012), which is based on the machine translation principle and makes use of 3D fragments assembly, none of the reported methodologies has reached the stage of truly full automation, efficient access to large RNA structures, short computing time, and user-friendly performance.

To assess 3D structures of RNAs of a size too large to conduct NMR studies (see Pachulska-Wieczorek, Purzycka, & Adamiak, 2006; Purzycka, Pachulska-Wieczorek, & Adamiak, 2011), we have settled an interdisciplinary group to develop automated RNA 3D structure prediction method. Full automation of the RNA 3D structure computational prediction is a complex and very difficult task. If one considers industrial applications, robots can perform actions that man cannot or can do only with much lower precision and reproducibility. In contrast to the industrial automation, an expert scientist with extensive knowledge and experience in the field is able to perform RNA 3D modeling in much more reliable way than available automated servers. Our aim is to change this situation with a kind of the "black box" server allowing, even inexperienced users, to access RNA 3D structures very fast and with great reliability. At the beginning, based on the secondary structure, we were able to predict efficiently only 3D structures of medium-size RNA, at low-resolution level (Popenda, Bielecki, & Adamiak, 2006).

We have soon realized that an ultimate methodology must fulfill several difficult to reach criteria outlined below:

- fully automated performance,
- high prediction fidelity at atomic resolution level,
- very fast multimodel building,
- access to 3D models with assessed energy,
- application to large RNA structures,
- acceptance of experimental constraints,
- high-throughput potential,
- web accessibility, and
- user-friendly performance.

It took us nearly 6 years to accomplish most of the above criteria. Development of RNA FRABASE (Popenda, Blazewicz, Szachniuk, & Adamiak, 2008; Popenda et al., 2010) appeared to be a key step in our study and soon flourished in the RNAComposer method (Popenda et al., 2012) and servers (http://rnacomposer.ibch.poznan.pl and mirror http://rnacomposer.cs.put.poznan.pl). The RNAComposer method, supported by our allied web server tools for the RNA quality assessment (Antczak et al., 2014; Lukasiak et al., 2013), will be presented and its scopes and limitations will be discussed in this chapter. The cyclic di-GMP-II riboswitch has been selected as an application example and its 3D structure prediction will be discussed in multiple aspects.

Riboswitches are an interesting example of noncoding mRNA regions capable of binding cellular metabolites to modulate gene expression. In order to function, those RNAs must achieve high specificity. This unique metabolite sensing is possible with a diverse array of secondary and tertiary structures. However, riboswitches representing different classes adopt diverse structures, and their exceptional selectivity is encoded in conserved structural features.

Riboswitches can be divided into two general groups based on their overall special architecture: pseudoknotted and junctional (Serganov & Nudler, 2013). Pseudoknots are formed when a hairpin loop interacts with an outside region and in riboswitches usually involve a stack of two helices. Spatial structure of fluoride riboswitch is based on the small pseudoknot (Ren, Rajashankar, & Patel, 2012). Junctional riboswitches contain multihelical junction joining several helices. This group is represented by lysine (5-way junction) (Garst, Heroux, Rambo, & Batey, 2008) or TPP (3-way junction) riboswitches (Edwards & Ferre-D'Amare, 2006). For the functional architecture of several riboswitches, like c-di-GMP-II, both junction and extensive

tertiary interactions are necessary (Smith, Shanahan, Moore, Simon, & Strobel, 2011). Those riboswitches represent most complex, mixed type group (Peselis & Serganov, 2014). Differing in complexity, pseudoknots and multihelical junctions are structural elements that constitute significant challenge for the prediction methods. Extensive studies are necessary to deconvolute not only their tertiary but also secondary structures.

2. RNA FRABASE—OPENING THE ROUTE TO RNAComposer

In 2006, we have presented (Popenda et al., 2006) a computer program, called 3D-RNApredict, which implements RNA secondary structure and converts various structural data (RNA fragments coordinates, experimental data) to create the input to the CYANA software (Guntert, Mumenthaler, & Wuthrich, 1997). CYANA's torsion angle dynamics algorithm provides a very fast engine for the 3D RNA structure calculation. The method was amenable to automatization. Soon we have realized that further advance of this approach needs dedicated 3D RNA fragments search engine and database. Along this line RNA FRABASE has been developed (Popenda et al., 2008, 2010). RNA FRABASE core consists of the interfaced database, the search engine, and the web interface. Its repository collects information about PDB-deposited RNA structures and their complexes, including all models of NMR-elucidated molecules. It stores RNA sequences; secondary structures encoded in the dot-bracket notation and in graphical form; atom coordinates of the unmodified and modified nucleotide and nucleoside residues; torsion and pseudotorsion angle values; sugar pucker parameters; and complete classification of base pair types, base–base parameters for base pairs, and inter-base pair parameters for dinucleotide steps. The third component has been designed with an emphasis on friendly and ergonomic interface. RNA FRABASE is a publicly available web-interfaced system, running at http://rnafrabase.cs.put.poznan.pl. RNA FRABASE is automatically updated on a monthly basis. As of June 2014, it contained 2259 RNA structures that represent about 90% of all RNAs deposited in the Protein Data Bank.

Relevant to this chapter's topic, it should be noted that within the RNA FRABASE database, one could find fragments derived from 149 PDB-deposited structures of riboswitches. They include 1119 duplexes (constituting 0.7% of all duplexes deposited in RNA FRABASE), 2106 loops (0.7%), and 71 single-stranded fragments (0.9%).

Having this valuable tool in hand we use it as the prerequisite to generate the dedicated dictionary—a key element in our RNA 3D structure prediction method.

3. RNAComposer—FROM THE RNA SECONDARY STRUCTURE TO RNA 3D STRUCTURE

3.1. General description of the RNAComposer method

The RNAComposer method for automated prediction of RNA 3D structures (Popenda et al., 2012) was founded on the machine translation concept parallel to that used in the computational linguistics. The developed workflow system allows us to progress rapidly from the RNA secondary structure to the RNA 3D structure (Fig. 1). Pivotal part of the translation system is the dedicated dictionary. This dictionary is tailored from the RNA FRABASE database (Popenda et al., 2008, 2010) on which the translation engine operates to predict RNA 3D structure. Dictionary relates the RNA secondary structure and tertiary structure elements. The RNAComposer dictionary differs from the RNA FRABASE database. In order to enable dictionary function in the translation, all the 3D structural elements collected there were attributed to a complete set of atoms, energy calculated by the CHARMM force field (XPLOR-NIH), and are characterized by a good stereochemistry and structural properties. The dictionary does not include elements with modified residues or missing heavy-atom coordinates. Since our report (Popenda et al., 2012), the volume of the dictionary has been substantially enlarged and includes 23,092 secondary structure elements (initially 14,464) and as many as 489,599 related 3D structure elements (initially 190,928). The 3D structure elements are continuously transferred from the RNA FRABASE to the RNAComposer system and are transformed into dictionary's elements. All steps of this transfer, starting from the import of the newly deposited PDB structures, are automated. As we expected, growing dictionary volume substantially increases the quality of the predicted 3D structures.

The algorithms governing the RNAComposer engine actions allow us to build the RNA 3D models automatically in the following steps exemplified on the cyclic di-GMP-II riboswitch (PDB ID 3Q3Z) (Fig. 1):

(i) *RNA secondary structure fragmentation*. The input RNA secondary structure is divided into fragments following its tree graph representation (Gan, Pasquali, & Schlick, 2003). The fragmentation algorithm provides the secondary structure elements: stems, loops (i.e., apical, bulge,

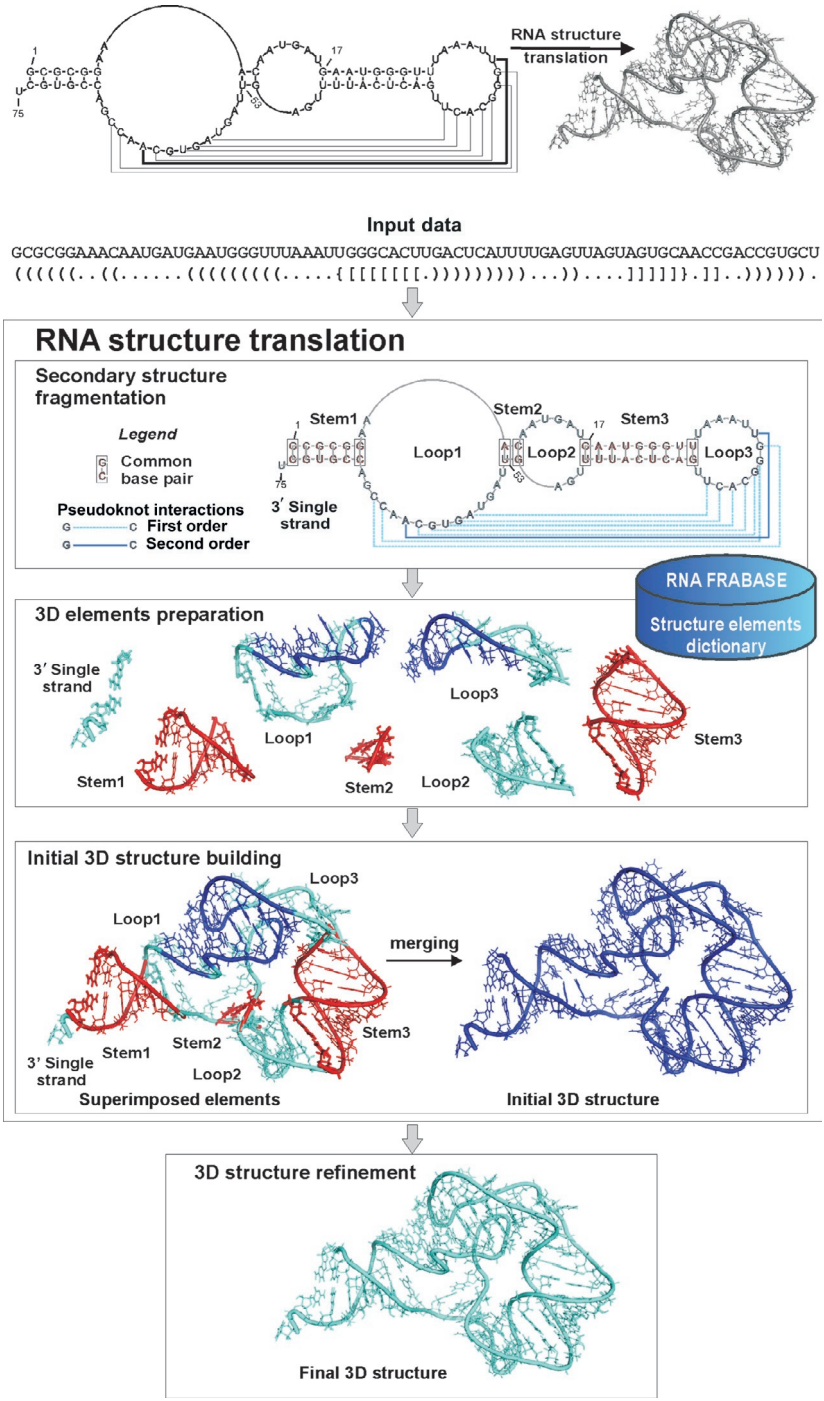

Figure 1 The RNA secondary structure to the 3D structure translation. Basic steps of RNAComposer action exemplified on the cyclic di-GMP-II riboswitch structure from *C. acetobutylicum* (PDB ID 3Q3Z).

internal, and *n*-way junctions), and single strands. Each element is closed by canonical base pair(s) according to the RNA FRABASE concept (Popenda et al., 2008). In the case of the secondary structure elements for the cyclic di-GMP-II riboswitch, three stems, three loops, and one single strand are produced by the fragmentation algorithm. At this stage pseudoknots are not considered.

(ii) *3D structure elements search*. The fragmentation constitutes input patterns (based on the RNA FRABASE notation) for an automatic search of related 3D structure elements within the dictionary. Engine makes use of only those 3D elements whose heavy-atom root mean square deviation (rmsd) is lower than 1.0 Å, when referred to the parent PDB structure.

(iii) *3D structure elements preparation*. The 3D structure elements are selected, usually from the wide spectrum of entities found in the dictionary. The selection process is crucial and is governed by an algorithm cantered on the five criteria executed in the following priority order:

- secondary structure topology,
- sequence similarity,
- pyrimidines/purines compatibility,
- source structure resolution, and
- energy.

(iv) *Initial RNA structure building*. The RNA tree graph representing secondary structure governs the building process. The 3D structure elements are superimposed with reference to the common canonical base pairs and merged to give initial, already well-shaped RNA 3D structure. Up to this step, the action of RNAComposer system is very fast. It usually takes several seconds on the single processor. For the cyclic di-GMP-II riboswitches discussed below it takes about 5 s.

(v) *3D structure refinement*. The energy minimization in the torsion angle space (Guntert et al., 1997) and subsequently in the Cartesian atom coordinate space (Schwieters, Kuszewski, Tjandra, & Clore, 2003) leads to the final, high-quality RNA 3D model. From the single secondary structure, up to 10 models can be generated in one prediction. This step takes several minutes for large RNA structures. In case of the cyclic di-GMP-II riboswitch from *Clostridium acetobutylicum*, generation of 10 models takes 2 min and 35 s.

Several important points for the users concerning RNAComposer action should be outlined. The system pipeline does not allow for changes from

the secondary structure topology set as an input (in dot-bracket notation). As far as the criterion of sequence similarity (homology) is concerned, there is often departure from the full sequence homology during 3D element selection. If the RNA sequence is not found for the fragment of the required topology, the respective bases are replaced based on the pyrimidines/purines compatibility. Fragments from the highest resolution structures are preferably selected. Crystal structures are first priority, preceded by NMR and cryo-EM structures. The translation engine can generate a family of closely related 3D models in the single prediction. Up to 10 models can be generated from one secondary structure. The first model is always built using all the selection criteria described above. Other models are generated from 3D structure elements for which the criteria of structure resolution and energy are ignored. These models often include NMR-derived 3D elements. In the batch mode, RNAComposer servers (see Section 3.4) can be loaded with up to 100 secondary structures. Moreover, when launching RNAComposer, user can enter own experimental restraints, which makes the system very powerful.

During the modeling of complex RNAs, some 3D structure elements related to the certain secondary structure topology might be missing in the dictionary. For the purpose of automation, which must be considered as the process of self-reliability, the missing 3D fragments are instantly generated with dedicated algorithm. To ensure overall speed of 3D structure generation, a very fast fd_helix routine of the NAB software was chosen to generate stems and single strands. In the case of missing RNA loops being the part of hairpins, bulges, internal loops, and *n*-way junctions, molecular mechanics in the torsion angle space using CYANA (Guntert et al., 1997) is conducted.

All the details concerning major steps of the RNAComposer action are given to the user in the log.txt file accompanying respective PDB file.

In the original report (Popenda et al., 2012), RNAComposer method was carefully evaluated to estimate its scope and the quality of the predicted 3D structure models in terms of the secondary structure topology conservation, their stereochemical properties, energy, precision, and accuracy. The predicted 3D structures were validated using representative benchmark set of RNAs with the secondary structures derived from the highest resolution X-ray structures.

The RNAComposer system is based on the concept of machine translation, new in structural bioinformatics field. In the view of existing methods, its formal classification is difficult and might raise controversies.

The RNAComposer might be classified as a knowledge-based method that employs automated fragment assembly, based on the secondary structure tree graph representation and homology of structural elements.

3.2. RNAComposer input data

RNAComposer uses RNA sequence and secondary structure topology in dot-bracket notation (Vienna notation) as an input for the 3D structure prediction. In this notation, dot represents unpaired nucleotide, while bracket refers to the nucleotide involved in canonical base pairing including GU. In order to allow prediction of more complex RNA structures, square brackets are accepted to annotate first-order pseudoknots and curly brackets for higher order structures. Accomplishing the structure translation concept, RNAComposer does not change provided secondary structure topology and entire prediction depends on the input data. Correct secondary structure is therefore critical for the accurate prediction of the 3D structure. RNA secondary structure probing methods are available to support and substantially improve *in silico* RNA structure prediction, e.g., see Deigan, Li, Mathews, and Weeks (2009). Based on our own experience (Huang et al., 2013; Lusvarghi et al., 2013), we strongly encourage the user to use experimentally adjusted secondary structures to ensure more accurate 3D structure modeling with RNAComposer.

Two modes are available for the input data into RNAComposer servers (see Section 3.4). High efficiency of RNAComposer system allowed us to make an interactive mode accessible for the unregistered user. This mode allows the user to introduce only RNA sequence and choose secondary structure prediction method from three possibilities incorporated within RNAComposer system: RNAfold (Schuster, Fontana, Stadler, & Hofacker, 1994), RNAstructure (Reuter & Mathews, 2010), and Contrafold (Do, Woods, & Batzoglou, 2006). This mode is oriented toward inexperienced users and works well only for simple RNA structures, i.e., RNA for which *in silico* secondary structure prediction is reliable enough.

Batch mode, accessible upon login, requires secondary structure topology in dot-bracket notation and is oriented toward prediction of the RNA of medium size (up to 500 nts) or higher level of complexity. This allows users to introduce their own experimentally adjusted secondary structures of RNA and gives best 3D structure predictions. Several secondary structure topologies can be introduced for a single RNA sequence. Importantly, distance restraints can be entered into the RNAComposer system in the batch

mode. This functionality allows users to incorporate additional data, e.g., from NMR or FRET experiments, into the 3D structure prediction pipeline. In addition, this option allows constraining noncanonical base pairs or pseudoknots.

3.3. Output data and quality control of the 3D models

The PDB and log files are provided as RNAComposer output. The PDB file contains coordinates of predicted 3D structure models. The log file describes details of the model generation process. This file includes several information allowing the user to inspect all steps of the RNAComposer action and to analyze obtained models: (i) a list of structure elements resulting from secondary structure fragmentation, (ii) a list of tertiary elements (sequence and topology) selected for the 3D structure assembly, (iii) origin (PDB ID code) of the 3D element, (iv) sequence similarity, and (v) the final energy given by the force field for the structure. In the log file, information are provided about 3D structure elements, which, due to the absence in the dictionary, were generated by the RNAComposer. Resulted RNA 3D structures containing such elements will be usually much less accurate and this should be taken into account. The energy of the final structure is a good indicator of the 3D structure models quality and should be inspected by the user (Popenda et al., 2012). Further quality control of the predicted 3D models could be accomplished using two complementary tools: RNApdbee (Antczak et al., 2014) and RNAlyzer (Lukasiak et al., 2013).

3.3.1 RNApdbee—Application in the RNA 3D structure validation

RNAComposer translation system does not allow for the departure from the input secondary structure topology. However, it is advisable to check whether the secondary structure pattern, especially in case of pseudoknots, was preserved after the refinement step. This can be done by reversing predicted 3D structure back to the secondary structure. Our recently reported method RNApdbee (Antczak et al., 2014) and publicly available server at http://rnapdbee.cs.put.poznan.pl allow the user to inspect the conservation of the input secondary structure topology in the predicted 3D structure models. RNApdbee web server has been designed to derive RNA secondary structure from the tertiary structure encoded in the PDB file or from the list of base pairs. It should be underlined that RNApdbee can process knotted and unknotted structures of large RNAs. RNApdbee supports an identification and classification of high-order pseudoknots, and their graphical visualization as well as dot-bracket encoding.

3.3.2 RNAlyzer—New tool for the 3D structure quality assessment

RNAComposer can generate a large ensemble of 3D structures within minutes. Having so many possible structures that can be generated for one sequence, one faces the problem of RNA 3D structure quality assessment. Currently, to assess 3D RNA models one can use several metrics such as rmsd, calculated for pairs of superimposed corresponding atoms, deformation index and interaction network fidelity, taking into account base–base interactions within the structures, and the deformation profile, taking into account both global and local differences (Parisien, Cruz, Westhof, & Major, 2009).

Recently, we have developed a new approach called RNAlyzer (Lukasiak et al., 2013) designed to solve the problem of comparison of RNA structure models to a reference structure. It is based on rmsd metrics and allows the user to identify the difference between them by calculating the dissimilarity between sets of atoms situated inside the series of spheres. Around selected atom of nucleotide, a sphere with defined radius is built. Selected atom plays a role of the center of the sphere that is constructed for every nucleotide of reference structure. The stage of sphere building gives sets of atoms corresponding to every sphere built on reference structure (Fig. 2).

In the next phase, for each set of the atoms identified in the previous phase, the corresponding set of atoms from the analyzed models are

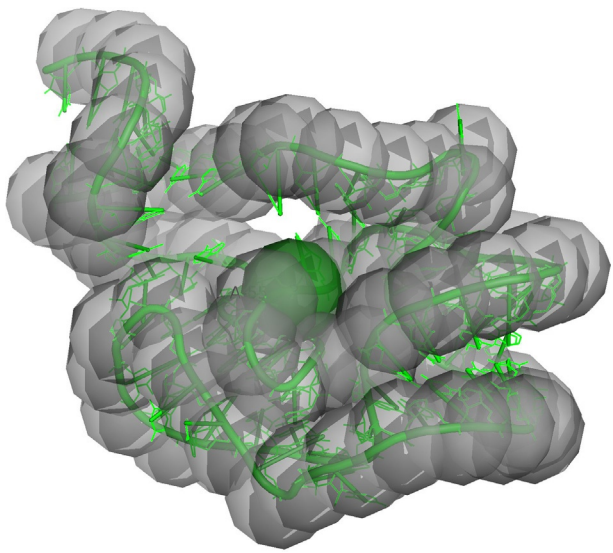

Figure 2 RNAlyzer spheres defining range of the atoms in superimposition of 3D model and reference structure, as probing structural space of every nucleotide residue.

localized. As a final stage, corresponding sets of atoms from the reference structure and the model are superimposed, and rmsd between them is calculated. The approach gives a possibility to select different radii and compare models from different levels of structural precision. Small sphere radius means the ability to analyze model from local structural point of view. With an increasing value of radius, the analysis becomes more global (lower accuracy). With sphere radius higher than radius of the analyzed molecule, the rmsd is calculated for the whole structure. The proposed quality assessment result can be analyzed and visualized on several types of 2D and 3D plots to show the model quality simultaneously for different levels of accuracy and to discriminate between correct and incorrect predictions in the entire spectrum of modeling precision. Application of RNAlyzer (http://rnalyzer.cs.put.poznan.pl) will be demonstrated on the example of cyclic di-GMP-II riboswitch from *C. acetobutylicum*.

3.4. RNAComposer web server

RNAComposer is offered to the users as a publicly available system with a web interface, running at two independent mirror sites: http://rnacomposer.ibch.poznan.pl and http://rnacomposer.cs.put.poznan.pl. Currently available version of the web server allows the user to build 3D structures of single-stranded RNAs up to 500 residues long. As a basic input, it requires a sequence and a secondary structure topology which can either be typed in directly or predicted from the sequence by one of the integrated tools (RNAstructure (Reuter & Mathews, 2010), RNAfold (Schuster et al., 1994), CONTRAfold (Do et al., 2006)) upon user choice. As already mentioned, RNAComposer works in two execution modes. An interactive mode is provided to newcomers and users processing single structure per query. Large-scale modeling is possible in a batch mode available for registered users. In this mode, one can upload a task containing even 10 RNA sequences, each associated with 1–10 secondary structures. For every secondary structure input in a batch, RNAComposer predicts up to 10 3D models. They can be generated with respect to optionally provided input atomic distance restraints. The basic output data are released in PDB-formatted files and can be immediately visualized (interactive mode), stored, and analyzed in the user workspace (batch mode) or emailed (both modes).

RNAComposer has been realized in the two-layer architecture, composed of a computational engine and a web application. The first, back-end layer hosts the translation machine (encoded in Java) and the dictionary

(stored in PostgreSQL DBMS). The computational engine supports concurrent processing and adapts dynamically to available resources such as the number of processing units and RAM capacity. The front-end layer of RNAComposer provides a user-friendly web interface implemented in SpringMVC. The system is closed in the Virtual Private Network that supports effective and safe communication between its components through the integrated message brokers (Apache Active MQ).

4. PREDICTING THE TERTIARY STRUCTURE OF RIBOSWITCHES WITH RNAComposer

4.1. RNAComposer accurately predicts 3D structure of several complex riboswitches

Taking into account the predictive power of RNAComposer, we have chosen a representative benchmark set of riboswitches of the known X-ray crystal structures. While these RNAs differ in strand length and structural complexity, all contain multihelical junctions and/or pseudoknots. Their sequence and secondary structure topologies (Table 1), derived from the highest resolution X-ray structures, were inputted into RNAComposer.

It is important to note that before prediction, all the 3D structure elements comprised by the respective crystal PDB structure were excluded from the dictionary. The results are presented in Table 2 and Fig. 3. Ten models were predicted for each 2D structure and the accuracy was analyzed by calculating network fidelity parameters and the rmsd with the crystal structure. All examples, with the exception of the c-di-GMP-II riboswitch from *C. acetobutylicum* (PDB ID 3Q3Z), were characterized by high accuracy, and canonical and noncanonical base pairing and stacking were recovered, as indicated by network fidelity parameters (Table 2). Superposition of all but one predicted and crystal structures showed very good overall agreement consistent with their rmsd values (Fig. 3).

Multihelical junction conformation and orientation of the helices were correct for those examples, including 3OWW with an rmsd of 5.7 Å. However, the c-di-GMP-II riboswitch *C. acetobutylicum* (3Q3Z) was predicted with an average rmsd of 12.3 Å, and the best model showed a value of 10.1 Å. The c-di-GMP-II from *C. acetobutylicum* represents especially difficult example due to very unique structural features (see below). Since this riboswitch is represented only by one PDB entry (3Q3Z), it was chosen as a challenging application example to discuss the scope and limitations of the RNAComposer in its current version (see below).

Table 1 Sequence and secondary structure data inputted into RNAComposer for the 3D structure prediction of selected riboswitches

RNA PDB code and chain	Sequence and secondary structure topology
2YIE X+Z	GGAUCUUCGGGGCAGGGUGAAAUUCCCGACCGGUGGUAUAGUCCACGAAUCCAUCCGGAUUGAUUUGGUGAAAUUCCAAAACCGACAGUAGAGUCUGGAUGAGAGAAGAUUCG (((((((((.....(((.......))))).[[(((((((....[[))))...(((....)))....(((((.......)))))]]))).(((....]])))).....)))))))).
3D2V A	GGGACCAGGGGUGCUUGUUCACAGGCUGAGAAAGUCCCUUUGAACCUGAACAGGGUAAUGCCUGCGCAGGGAGUGUC ..(((((((((..(((((....)))))).........)))).....(((...((((.....))))...)))..).)))
3GX5 A	GGCUUAUCAAGAGAGGUGGAGGGACUGGCCCGACGAAACCCGGCAACCAGAAAUGGUGCCAAUUCCUGCAGCGGAAACGUUGAAAGAUGAGCCG (((((((((....(.(((...(((...[[))).....)))))(((..(((....))).))))...(]]..(((((....)))))..)))))))).
4FE5 B	GGACAUAUAAUCGCGUGGAUAUGGCACGCAAGUUUCUACCGGGCACCGUAAAUGUCCGACUAUGUCC .(((((((..(.(((((.....[[))))))..)[.....)](((((]].....))))))..)))))).
3DIL A	GGCCGACGGAGGCGCGCCCGAGAUGAGUAGGCUGUCCCAUCAGGGGAGGAAUCGGGGACGGCUGAAAGGCGAGGGCGCCGAAGGGUGCAGA GUUCCUCCCGCUCUGCAUGCCUGGGGGUAUGGGGAAUACCCAUACCACUGUCACGGAGGUCUCUCCGUGGAGAGCCGUCGGUC (((((((((..(.((((((..(.......((((((((((..(..[[[[[[....))))))))))).....)..)))))))..((((((((((((.]]]]]])))))))))).)))((.(((((((((.....)))))))).)).(((((((((....)))))))))....)))))))))
3PDR X	GGGCUUCGUUAGGUGAGGCUCCUGUAUGGAGAUACGCUGCUGCCCAAAAAUGUCCAAAGACGCCAAUGGGUCAACAGAAAUCAUCGACAUAAGGUGAUUUUUAAUG CAGCUGGAUGCUUGUCCUAUGCCAUACAGUGCUAAAGCUCUACGAUUGAAGCCCA (((((((((.[(((..(((..(((((((((.....(((((((((((....((((.....))))....)))))).....((((((((((.....])))))))))..).)))))).((........))....)))))))).)))...))))..)).....))))))).
3OWW A	GGCUCUGGAGAGAACCGUUUAAUCGGUCGCCGAAGGAGCAAGCUCUGCGGAAACGCAGAGUGAAACUCUCAGGCAAAAGGACAGAGUC (((((((....(.((((.....)))))))(((...((((...(((((((((....)))))))))....))).).).))).....)))))))
3MXH R	GGUCACGCACAGGGCAAACCAUUCGAAAGAGUGGGACGCAAAGCCUCCGGCCUAAACCAUUGCACUCCGGUAGGUAGCGGGGUUACCGAUGG ...(.((.....((...(((((((....))))))..[)).(((.(((((((..((.........)))))))..])))))).))...)
3Q3Z V	GCGCGGAAACAAUGAUGAAUGGGUUUAAAUUGGGCACUUGACUCAUUUUGAGUUAGUAGUGCAACCGACCGUGCU ((((((..((.....((((((((((.....{[[[[[[[.))))))))))...)).....]]]]]}.]]..)))))).
3SD3 A	GGAGAGUAGAUGAUUCGCGUUAAGCGUGUGUGAAUGGGAUGUCGUCACACAACGAAGCGAGAGCGCGGUGAAUCAUUGCAUCCGCUCCA ((((....((((((((((((.....((((((((...[[[[....)))))))).....((....)))))).))))))))).]]]].)))))).

Table 2 Characteristics and accuracy of the 3D models predicted for selected riboswitches

RFAM				PDB				Accuracy[a]		
		Number of sequences								
RFAM ID	Family description	Full	Seed	Number of PDB entries	RNA PDB code and chain	Resolution (Å)	Strand length (nt)	rmsd (Å)	INF^{cbp}	INF^{all}
RF00050	FMN (RFN element)	4516	144	16	2YIE X, Z	2.94	113	3.2	0.98 (0.01)	0.79 (0.02)
RF00059	TPP (THI element)	11197	115	12	3D2V A	2.00	77	3.3	1.00 (0.00)	0.82 (0.02)
RF00162	SAM (S box leader)	4757	433	18	3GX5 A	2.40	94	2.7	0.99 (0.01)	0.81 (0.02)
RF00167	Purine riboswitch	2427	133	24	4FE5 B	1.32	67	2.0	1.00 (0.01)	0.85 (0.04)
RF00168	Lysine riboswitch	2422	47	14	3DIL A	1.90	174	3.0	1.00 (0.00)	0.89 (0.01)
RF00380	ykoK leader	1493	157	2	3PDR X	1.85	161	3.5	0.97 (0.01)	0.78 (0.01)
RF00504	Glycine riboswitch	6875	44	12	3OWW A	2.80	88	5.7	0.98 (0.00)	0.81 (0.03)
RF01051	Cyclic di-GMP-I	1990	155	17	3MXH R	2.30	92	3.4	0.98 (0.01)	0.79 (0.03)
RF01786	**Cyclic di-GMP-II**	**237**	**54**	**1**	**3Q3Z V**	**2.51**	**75**	**12.3**	**0.85 (0.05)**	**0.48 (0.03)**
RF01831	THF	598	98	5	3SD3 A	1.95	83	2.0	1.00 (0.01)	0.85 (0.02)

[a]Described as the average heavy-atom rmsd (in Å) between 10 individual 3D models and the crystal structure, and the average interaction network fidelity (INF) measures. INF scores range from 0.00 (worst) to 1.00 (best).

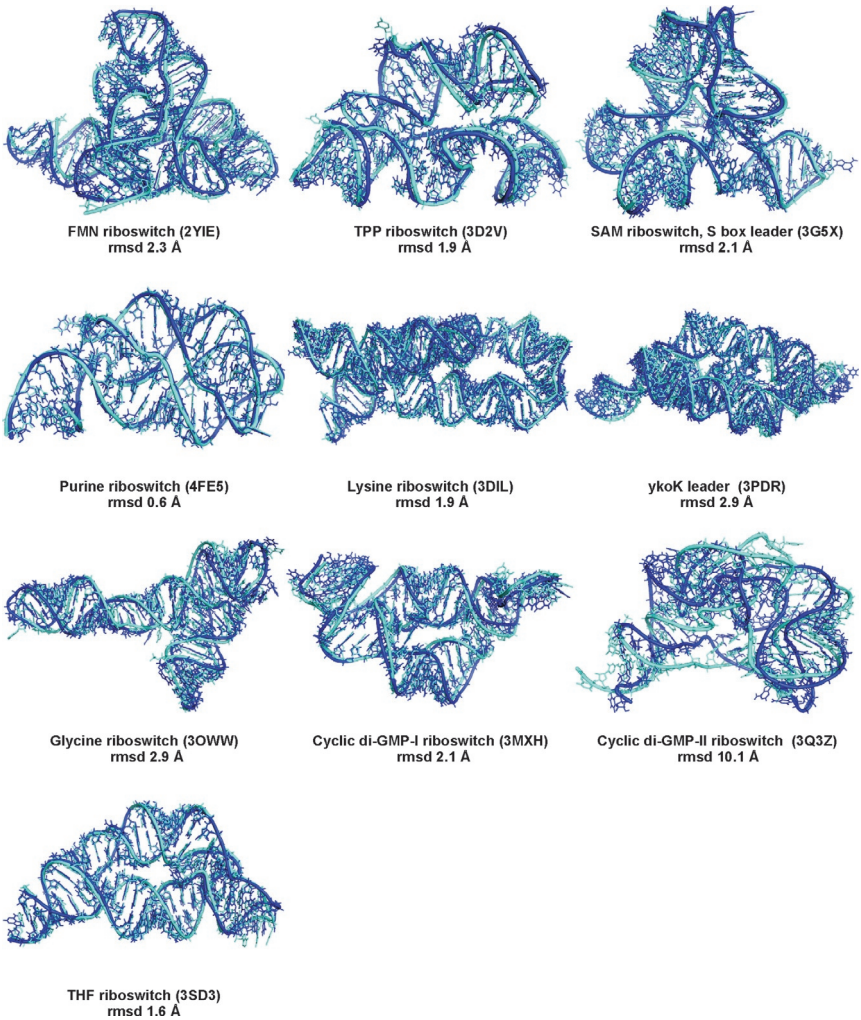

Figure 3 Superimposition of the predicted 3D models (dark blue; black in the print version) and their respective X-ray structures (cyan; light gray in the print version) for 10 selected riboswitches.

The quality of the predicted 3D models could be anticipated based on energy values calculated by RNAComposer. The energy values change linearly with the RNA strand length (Popenda et al., 2012). The models for which the calculated energy value (Table 3) was lower than the expected (Popenda et al., 2012) were in good agreement with the crystal structures.

Table 3 Energy and quality of the 3D models predicted for selected riboswitches

RNA PDB code and chain	Strand length (nt)	Energy (XPLOR) (kcal/mol)		Clash-score, all atoms		Nucleic acid geometry							
						Potentially incorrect							
						Sugar puckers		Backbone		Outlier bonds (%)		Outlier angles (%)	
		Min.	Max.	RNAComposer	X-ray	RNAComposer	X-ray	RNAComposer	X-ray	RNAComposer	X-ray	RNAComposer	X-ray
2YIE X+Z	113	−2608	−2472	7.91 (1.59)	6.26	2.70 (0.95)	1	22.00 (2.79)	12	0.00 (0.00)	0.00	0.00 (0.00)	0.00
3D2V A	77	−1941	−1852	11.87 (2.10)	5.30	1.11 (1.05)	4	16.78 (2.28)	14	0.00 (0.00)	0.00	0.00 (0.00)	0.00
3GX5 A	94	−2174	−1986	9.78 (1.43)	15.58	4.30 (1.64)	7	21.80 (2.35)	17	0.00 (0.00)	0.00	0.00 (0.01)	0.48
4FE5 B	67	−1557	−1473	12.17 (2.91)	4.66	1.00 (0.82)	1	11.70 (1.95)	3	0.00 (0.00)	0.00	0.00 (0.00)	0.00
3DIL A	174	−4628	−4462	8.87 (1.88)	0.36	1.80 (0.63)	4	26.00 (2.63)	17	0.00 (0.00)	0.00	0.00 (0.00)	0.12
3PDR X	161	−3637	−3580	10.65 (1.42)	5.07	1.50 (0.71)	2	33.30 (3.40)	18	0.00 (0.00)	0.00	0.00 (0.00)	0.32
3OWW A	88	−1985	−1859	7.28 (1.26)	1.08	1.30 (1.25)	3	15.00 (2.31)	18	0.00 (0.00)	0.00	0.00 (0.00)	0.06
3MXH R	92	−2062	−1874	11.09 (2.09)	2.72	4.50 (2.27)	1	23.50 (3.41)	10	0.00 (0.00)	0.00	0.01 (0.01)	0.09
3Q3Z V	**75**	**−1257**	**−922**	**29.26 (4.94)**	**0.43**	**3.60 (0.97)**	**5**	**26.30 (4.37)**	**10**	**0.00 (0.00)**	**0.00**	**0.08 (0.02)**	**0.29**
3SD3 A	89	−2164	−2083	8.56 (1.62)	1.76	0.20 (0.63)	0	10.80 (2.70)	1	0.00 (0.00)	0.00	0.00 (0.00)	0.00

3Q3Z riboswitch was the only example that showed higher than expected energy value, although the stereochemical quality of all models, including 3Q3Z, was high (Table 3).

4.2. Application example 1: The c-di-GMP-II riboswitch

The c-di-GMP-II riboswitch from *C. acetobutylicum* controls a gene involved in carbohydrate processing and its aptamer domain structure was solved at 2.5 Å resolution (Smith et al., 2011). This RNA assumes a compact structure comprising several structural elements that are particularly challenging for tertiary structure prediction. The structure contains not only the second-order pseudoknot, triple helix within pseudoknot major grove but also an unusual U-turn/S-turn architecture. This U-turn/S-turn motif is a unique feature not present in any other PDB-deposited structure (Smith et al., 2011). Within the predictions presented below, all structural elements derived from 3Q3Z were removed from the RNAComposer dictionary to simulate a situation in which the 3D structure of a novel RNA is predicted.

The quality of the RNAComposer-predicted 3D structure depends strongly on the input secondary structure. As the c-di-GMP-II riboswitch from *C. acetobutylicum* is a relatively small RNA (75-mer), *in silico* methods should allow prediction of the correct secondary structure. However, this is not the case. We used several web-accessible tools (Table 4) to predict secondary structure models of this riboswitch. The secondary structures obtained departed significantly from that found in the crystal structure (Table 4) as indicated by Matthews correlation coefficient (MCC) (Matthews, 1975).

The best fit was observed for the structure generated using CyloFold program (Bindewald, Kluth, & Shapiro, 2010) with MCC of 0.89 (Table 4). Secondary structures thus obtained were inputted into RNAComposer. As expected, predicted 3D structures showed little agreement with the crystal structure 3Q3Z (average rmsd of 21.17 Å). CyloFold-derived structure has rmsd of 15.85 Å despite the fact that a loop, missing in the dictionary, was generated by RNAComposer. The energy value for an RNA of this size should be lower than −1498 kcal/mol (Popenda et al., 2012) and this criterion was not fulfilled in case of the CyloFold-derived secondary structure used for 3D structure prediction (Table 4). Other structures (Table 4), with low MCC, and consequently high rmsd, showed good energy values, demonstrating that the energy criterion might not be informative if the input secondary structure topology is extensively inaccurate.

Table 4 Dependence of the 3D structure prediction of the cyclic di-GMP-II riboswitch (*C. acetobutylicum*) on the accuracy of the *in silico* predicted secondary structure

Sequence GCGCGGAAACAAUGAUGAAUGGGUUUAAAUUGGGCACUUGACUCAUUUUGAGUUAGUAGUGCAACCGACCGUGCU

Method	Secondary structure topology	MCC	Energy (XPLOR) (kcal/mol)	rmsd (Å)	Type of 3D element with lowest sequence homology
X-ray[a]	((((((..((......((((((((((..... {[[[[[[[.))))))))))...))....]]]]]}.]]..))))))).	1.00	−1257	10.1	Loop (5.26%)
CyloFold	.(((((..........((((((((((.... [[[[{{{{{.))))))))))........}}}}}..]]]])))))..	0.89	−1165	15.9	Loop (missing in the dictionary)
mfold, SFold	.((((((...((((..((((....)))).)))))). ((((((((((.....))))))..)))))......)))))..	0.41	−1398	20.9	Junction loop (62.5%)
DotKnot	.(((((...(((((........[[[[...))))). ((((((((((((...)))))))..)))))]]]]..)))))..	0.40	−1273	21.3	Loop (41.2%)
RNAfold	.(((((((((.((.....))..)))))... ((((((((((((((((...))))))).)))))..))))))..))))))))..	0.46	−1590	21.6	Stem (64.3%)
Kinefold	.(((((((((.....)))))))... ((((((((((((((.....))))))..)))))..))))))..))))))))..	0.49	−1506	22.6	Loop (64.3%)
CONTRAfold	..((((..... ((((((((((((((((...))))))).)))))..))))))..)))))))...	0.49	−1430	28.4	Loop (missing in the dictionary)
RNAstructure, CentroidFold	.(((((..... ((((((((((((((((...))))))).)))))..))))))..))))))))..	0.53	−1451	28.7	Loop (missing in the dictionary)

[a]Secondary structure topology derived from the X-ray structure (3Q3Z) using RNApdbee.

To analyze in detail how RNAComposer predicts the 3D structure of the c-di-GMP-II riboswitch from *C. acetobutylicum*, the correct secondary structure was used as an input. This was derived from the crystal structure using RNApdbee (Antczak et al., 2014) and an equivalent structure, with the exception of the U30–A62 interaction, was proposed based on the secondary structure analysis (Lee, Baker, Weinberg, Sudarsan, & Breaker, 2010). The best 3D model obtained displayed energy of −1257 kcal/mol (Table 3) and rmsd of 10.1 Å (Fig. 3). Interestingly, although one of the structural elements (loop G6-A9/U53-C69) selected by RNAComposer during assembly showed sequence homology as low as 4.8% (Table 5, first entry, exclusion of 3Q3Z from dictionary) our prediction generated a 3D model of 10.1 Å rmsd value.

This illustrates the advantage of RNAComposer over homology modeling and the importance of experimental data during input secondary structure generation. It is also important to note that the entire prediction process took about 3 min. Lack of the loop G6-A9/U53-C69 element of higher sequence homology in the RNAComposer dictionary culminated in local departure from the correct structure as visualized with RNAlyzer (Fig. 4).

Prediction of triplexes and pseudoknots as tertiary motifs annotated with the square brackets is the main limitation of RNAComposer and is currently under development. A less complex kink turn motif is well predicted; therefore, the orientation of all helices is recovered and the global structure of the c-di-GMP-II riboswitch is relatively similar to the reference structure 3Q3Z (Table 5).

As stated earlier, the quality of the predicted 3D models depends strongly on the RNAComposer dictionary content (Popenda et al., 2012). When 3D structural elements derived from 3Q3Z were returned to the dictionary (Table 5, second entry), four tertiary structure elements from the 3Q3Z were brought into play for this model building, including the G6-A9/U53-C69 loop. The 3D structure was predicted with global rmsd of 1.1 Å, indicating not only the predictive fidelity and power of the method but also the applicability of RNAComposer to model 3D structures of other members of this particular riboswitch family. Such an application case is presented below.

4.3. Application example 2: The c-di-GMP-II riboswitch relatives

As the next example we attempted to predict 3D structures of the c-di-GMP-II riboswitches, members of the family RF01786 depicted in

Table 5 3D structure prediction details for the cyclic di-GMP-II riboswitch from *C. acetobutylicum* (**3Q3Z**)

Input data

```
GCGCGGAAACAAUGAUGAAUGGGUUUAAAUUGGGCACUUGACUCAUUUUGAGUUAGUAGUGCAACCGACCGUGCU
((((((..((......((((((((((.....{[[[[[[[.))))))))))...)).....]]]]]}.]]..)))))).
```

Structural element	Target secondary structure element				Source tertiary structure element					Selection criteria			
	Localization				Localization								
	Residues		Secondary structure			Residues		Secondary structure		Secondary structure topology	Sequence similarity (%)[a]	Pyrimidines/ purines compatibility (%)[b]	Source structure resolution (Å)
	First	Last	Sequence	Topology	PDB ID	First	Last	Sequence	Topology				
First entry													
Loop 1	6	9	GAAA	(..(	1VOR	959	962	UCCC	(..(	52.4	4.8	28.6	11.5
	53	69	UUAGUAGU GCAACCGAC	)....]]]]]} .]]..)		1090	1106	GGGGCUCAAGU GAUCUA	)...]....]]]].]])				
Loop 2	10	17	CAAUGAUG	(......(	3F1H	1232	1239	CGAUGAAG	(......(	Identical	76.9	84.6	3.0
	48	52	UUGAG	)...)		1259	1263	UGGAG	)...)				
Loop 3	25	40	UUAAAUUGGG CACUUG	(..... {[[[[[[[.)	4JF2	28	43	UACUUAUUUC CUUUGA	(..... [[[[[..)	81.3	25.0	37.5	2.3
Stem 1	1	6	GCGCGG	((((((	1Q2S	22	27	GCACGG	((((((	Identical	91.7	100.0	3.2
	69	74	CCGUGC	))))))		34	39	CCGUGC	))))))				
Stem 2	9	10	AC	((	2VHN	139	140	AC	((	Identical	100.0	100.0	3.7
	52	53	GU	))		640	641	GU	))				
Stem 3	17	25	GAAUGGGUU	(((((((((	3U5H	3397	3405	GGUUGCGGC	(((((((((	Identical	44.4	66.7	3.0
	40	48	GACUCAUUU	)))))))))		3508	3516	GCUGCAAUC	)))))))))				
Single strand	74	75	CU	).	3CJZ	25	26	CU	).	Identical	100.0	100.0	1.8

Second entry													
Loop 1	6	9	GAAA	(..(	3Q3Z	81	84	GAAA	(..(	Identical	100.0	100.0	2.5
	53	69	UUAGUAGUG CAACCGAC	)....]]]]]}.]]..)		128	144	UUAGUAGUG CAACCGAC	)....]]]]]}.]]..)				
Loop 2	10	17	CAAUGAUG	(......(	3Q3Z	10	17	CAAUGAUG	(......(	Identical	100.0	100.0	2.5
	48	52	UUGAG	)...)		48	52	UUGAG	)...)				
Loop 3	25	40	UUAAAUUGG GCACUUG	(..... {[[[[[[[.)	3Q3Z	25	40	UUAAAUUG GGCACUUG	(.....{[[[[[[[.)	Identical	100.0	100.0	2.5
Stem 1	1	6	GCGCGG	((((((	1Q2S	22	27	GCACGG	((((((	Identical	91.7	100.0	3.2
	69	74	CCGUGC	))))))		34	39	CCGUGC	))))))				
Stem 2	9	10	AC	((	1JZX	328	329	AC	((	Identical	100.0	100.0	3.1
	52	53	GU	))		344	345	GU	))				
Stem 3	17	25	GAAUGGGUU	(((((((((	3Q3Z	92	100	GAAUGGGUU	(((((((((	Identical	100.0	100.0	2.5
	40	48	GACUCAUUU	)))))))))		115	123	GACUCAUUU	)))))))))				
Single strand	74	75	CU	).	3CJZ	25	26	CU	).	Identical	100.0	100.0	1.8

[a]Between the target and source RNA sequence.
[b]Matching of the purine/pyrimidine residues for the target and source RNA of given sequence.

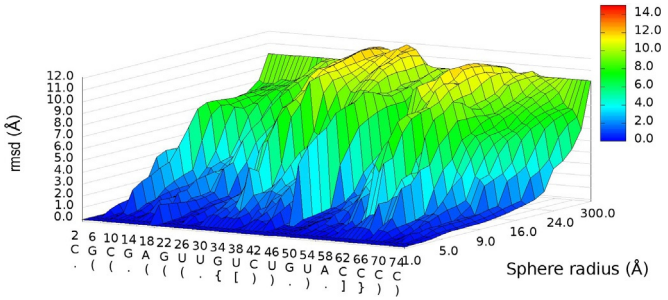

Figure 4 RNAlyzer visualization of the predicted 3D model quality for the cyclic di-GMP-II riboswitch from *C. acetobutylicum*.

the RFAM database (Burge et al., 2013). Although related to the c-di-GMP-II from *C. acetobutylicum*, these riboswitches differ in the sequence length and identity. The RF01786 family comprises 237 members, with a "seed" subgroup containing 54 entities (Table 2). Their RNA sequences are 8–15 nucleotide residues longer than c-di-GMP-II from *C. acetobutylicum*. Their consensus secondary structure, presented in the RFAM database (Stockholm notation), closely resembles that seen in the crystal structure 3Q3Z, but does not contain information about the pseudoknot. In addition, due to the length of those RNAs their graph representations differ, mostly within loops forming the pseudoknot and at the 5′- and 3′-termini. We have generated secondary structures for all RNAs present in the "seed" subgroup of the RF01786 family based on the alignment of the consensus secondary structures and respective 3Q3Z structure topology. Those structures were used for automated structure prediction using RNAComposer. One 3D model was generated for each of 54 RNA secondary structures and this ensemble was fitted into the crystal structure of the 3Q3Z using alignment of tertiary structures applying program ARTS (Dror, Nussinov, & Wolfson, 2006). Based on this, we selected three subgroup members, i.e., the riboswitches from *Clostridium difficile 4*, *Bacillus halodurans 1*, and *Thermus aquaticus Y5 1*, for which the predicted 3D structures fitted the best to the c-di-GMP-II riboswitch 3Q3Z. For each case, 10 models were generated and the best 3D structures are presented in Fig. 5.

Their alignment to 3Q3Z structure using ARTS was based on the large, consensus 3D fragment with the core size of 50–68 residues (Table 6). It appeared that within the core fragment comprising potential ligand-binding site, all three riboswitch structures very closely resemble that from *C. acetobutylicum* (rmsd less than 2 Å). Because of space limitation, 3D

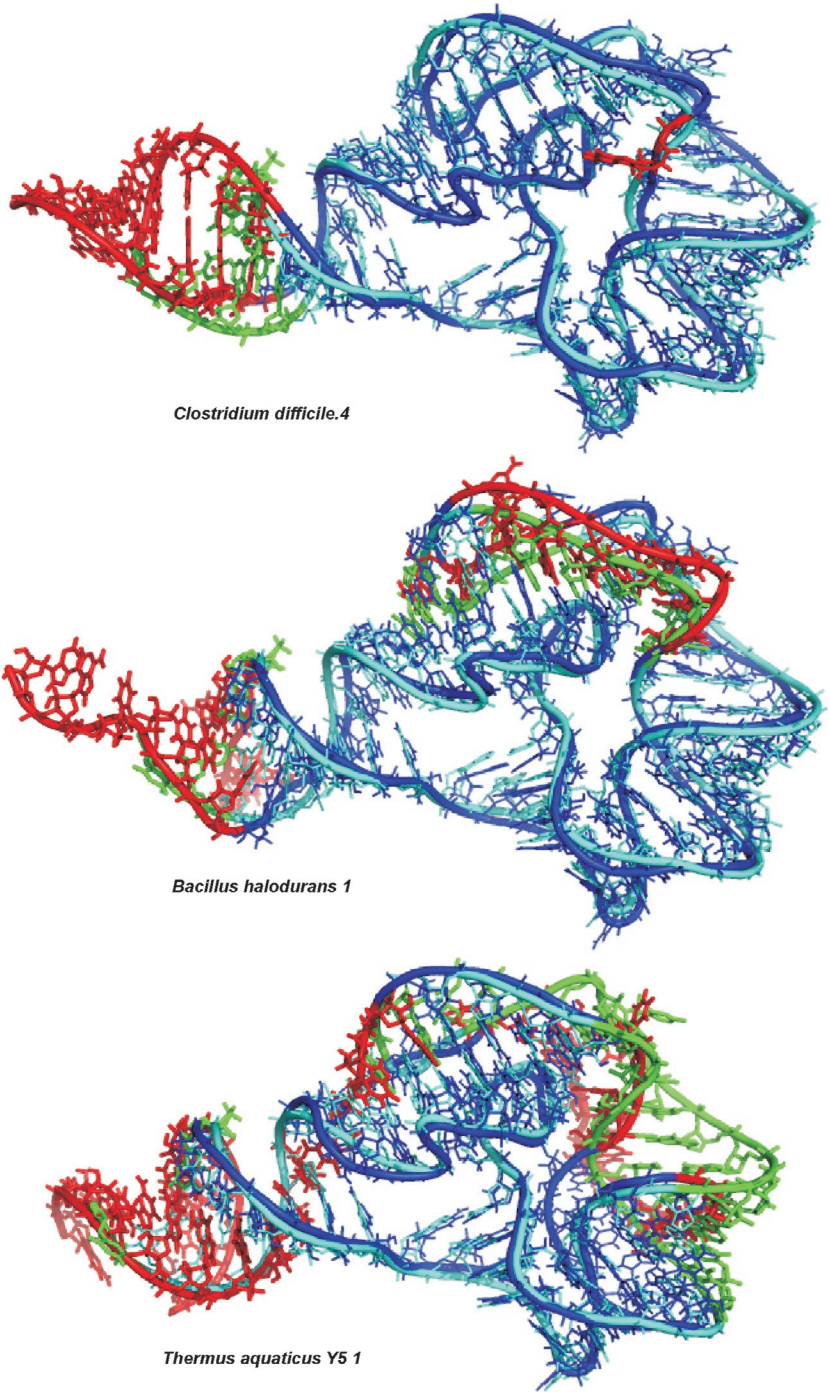

Figure 5 Pairwise superimposition of the best 3D models of the cyclic di-GMP-II riboswitch from *C. difficile 4*, *B. halodurans 1*, and *T. aquaticus Y5.1* (red and blue-core), and the X-ray structure from *C. acetobutylicum* (PDB ID 3Q3Z) (green and cyan-core).

Table 6 ARTS pairwise alignment parameters for the best 3D models of the cyclic di-GMP-II riboswitches *C. difficile 4*, *B. halodurans 1*, and *T. aquaticus Y5.1*, and the respective X-ray structure from *C. acetobutylicum* (3Q3Z)

	Cyclic di-GMP-II riboswitch		
ARTS parameters	***C. difficile 4***	***B. halodurans 1***	***T. aquaticus Y5.1***
Score	114	96	84
Core size (nt)	68	62	50
Number of matched base pairs	23	17	17
rmsd (Å)	1.4	1.8	1.3
Number of identical core residues	44	39	31
Structural identity	0.92	0.84	0.68
Core sequence identity	0.65	0.63	0.62
Sequential match score	66	59	46

structure prediction details are shown only for the c-di-GMP-II riboswitch from *C. difficile 4* (Table 7).

5. CONCLUSIONS AND PERSPECTIVES

We have demonstrated that our approach is not only fully automated but also characterized by very short computation time, efficiency, and user-friendly access to high-resolution 3D models. RNAComposer appeared to outperform existing automated methods (Popenda et al., 2012) and is especially suited for 3D RNA structure prediction of large RNAs, where experimentally adjusted secondary structure can be used as an input (see Huang et al., 2013; Lusvarghi et al., 2013). Since our report (Popenda et al., 2012), the accuracy of predicted 3D structures has increased considerably due to the 2.6-fold enlargement of the dictionary volume. Access to nearly 490,000 3D structure elements makes structure assembly more efficient. Both RNAComposer servers are very often visited (more than 75,000 entries worldwide for both sites). Apart from the 3D RNA predictions solely, RNAComposer method found new interesting application for deriving models coherent with experimental data from NMR (Krahenbuhl, Lukavsky, & Wider, 2014). We foresee the RNAComposer applications in fitting the 3D RNA patterns into the Cryo-EM data. Work

Table 7 3D structure prediction details for the cyclic di-GMP-II riboswitch from *C. difficile 4*

Input data

AAUAUUUUAGAAACUGAGAAGUAUAUCUUAUUAUUGGGCAUCUGGAGAUAUAUGGAGUUAGUGGUGCAACCGGCUAUGAAUAUA
.((((((((((..((......((((((((((.....{[[[[[[[.).)))))))))...)).....]]]]]}.]].))))).)))))).

No.	Size (nt)	Target secondary structure element				Source tertiary structure element					Selection criteria			
		Localization				Localization								
		Residues		Secondary structure			Residues		Secondary structure		Secondary structure topology	Sequence similarity (%)[a]	Pyrimidines/ purines compatibility (%)[b]	Source structure resolution (Å)
		First	Last	Sequence	Topology	PDB ID	First	Last	Sequence	Topology				
Single strands														
1	2	1	2	AA	.(	1YTU	1	2	AG	.(	Identical	50.0	100.0	2.5
Loops														
1	5	7	8	UU	((	4A1D	1351	1352	UU	((	Identical	100.0	100.0	3.5
		76	78	AUG	).)		1207	1209	AUG	).)				
2	21	10	13	GAAA	(..(	3Q3Z	6	9	GAAA	(..(	Identical	100.0	90.5	2.5
		58	74	UUAGUGGUGC AACCGGC	)....]]]]]}.]]..)		53	69	UUAGUAGUGC AACCGAC	)....]]]]]}.]]..)				
3	13	14	21	CUGAGAAG	(......(	2ZJR	1221	1228	CGGAGAAG	(......(	Identical	92.3	92.3	2.9
		53	57	UGGAG	)...)		1247	1251	UGGAG	)...)				
4	5	28	29	UU	((	2ZJR	725	716	UU	((	Identical	100.0	100.0	2.9
		44	46	GGA	).)		741	743	GGA	).)				
5	16	29	44	UAUUAUUGG GCAUCUG	(..... {[[[[[[[.)	3Q3Z	25	40	UUAAAUUGGG CACUUG	(..... {[[[[[[[.)	Identical	100.0	68.8	2.5

Continued

Table 7 3D structure prediction details for the cyclic di-GMP-II riboswitch from *C. difficile 4*—cont'd

Input data

```
AAUAUUUUAGAAACUGAGAAGUAUAUCUUAUUAUUGGGCAUCUGGAGAUAUAUGGAGUUAGUGGUGCAACCGGCUAUGAAUAUA
.((((((((((..((......((((((((((.....{[[[[[[[.).)))))))))...)).....]]]]]}.]]..))))).)))))))
```

No.	Size (nt)	Target secondary structure element: Localization: Residues: First	Last	Secondary structure: Sequence	Topology	Source tertiary structure element: Localization: PDB ID	Residues: First	Last	Secondary structure: Sequence	Topology	Selection criteria: Secondary structure topology	Sequence similarity (%)	Pyrimidines/purines compatibility (%)	Source structure resolution (Å)
Stems														
1	12	2	7	AUAUUU	((((((	3U5B	26	31	AUAUGC	((((((	Identical	66.7	88.3	3.0
		78	83	GAAUAU	))))))		595	600	GCGUAU	))))))				
2	6	8	10	UAG	(((	1VQO	2852	2854	UAG	(((	Identical	100.0	100.0	2.2
		74	76	CUA	)))		2902	2904	CUA	)))				
3	4	13	14	AC	((	1VQO	2059	2069	AC	((	Identical	100.0	100.0	2.2
		57	58	GU	))		2074	2075	GU	))				
4	16	21	28	GUAUAUCU	((((((((	3U5B	1658	1665	GAAUGGCU	((((((((	Identical	56.3	75.0	3.0
		46	53	AGAUAUAU	))))))))		1722	1729	GGUCAUUU	))))))))				
Single strands														
1	2	83	84	UA	).	3MOJ	73	74	UA	).	Identical	100.0	100.0	2.9

[a]Between the target and source RNA sequence.
[b]Matching of the purine/pyrimidine residues for the target and source RNA of given sequence.

applying RNAComposer to predict artificial RNA 3D structures is on the way (Chworos, personal communication).

Our current perspective is to enforce the RNAComposer method with new algorithms generating missing 3D structural elements of loops and single strands more effectively. This is especially important for large *n*-way junction loops and regions involved in pseudoknots formation. More efficient energy minimization protocols will be introduced in order to enlarge applicability of the method to RNAs larger than 500 nt residues. RNAComposer servers will be equipped with new functionalities allowing user to add not only the distance restraints but also the torsion angle restraints. The user, within workspace given, will be allowed to add own 3D structural elements (e.g., obtained via molecular dynamics) to the dictionary and to force their preferred usage upon fragment assembly.

ACKNOWLEDGMENTS

This work was supported by the National Science Center, Poland [MAESTRO 2012/06/A/ST6/00384 (to R. W. A.)] and the Foundation for Polish Science [HOMING PLUS/2012-6/12 (to K. J. P.)]. Poznań Supercomputing and Networking Centre is acknowledged for hosting the RNAComposer server at http://rnacomposer.ibch.poznan.pl. R.W.A. would like to acknowledge the contribution of the COST Action CM1105.

REFERENCES

Antczak, M., Zok, T., Popenda, M., Lukasiak, P., Adamiak, R. W., Blazewicz, J., et al. (2014). RNApdbee—A webserver to derive secondary structures from pdb files of knotted and unknotted RNAs. *Nucleic Acids Research*, *42*, W368–W372.

Bindewald, E., Kluth, T., & Shapiro, B. A. (2010). CyloFold: Secondary structure prediction including pseudoknots. *Nucleic Acids Research*, *38*, W368–W372.

Burge, S. W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E. P., et al. (2013). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Research*, *41*, D226–D232.

Cao, S., & Chen, S. J. (2011). Physics-based de novo prediction of RNA 3D structures. *The Journal of Physical Chemistry. B*, *115*, 4216–4226.

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

Deigan, K. E., Li, T. W., Mathews, D. H., & Weeks, K. M. (2009). Accurate SHAPE-directed RNA structure determination. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 97–102.

Do, C. B., Woods, D. A., & Batzoglou, S. (2006). CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics*, *22*, e90–e98.

Dror, O., Nussinov, R., & Wolfson, H. J. (2006). The ARTS web server for aligning RNA tertiary structures. *Nucleic Acids Research*, *34*, W412–W415.

Edwards, T. E., & Ferre-D'Amare, A. R. (2006). Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure*, *14*, 1459–1468.

Flores, S. C., & Altman, R. B. (2010). Turning limited experimental information into 3D models of RNA. *RNA*, *16*, 1769–1778.

Flores, S. C., Sherman, M. A., Bruns, C. M., Eastman, P., & Altman, R. B. (2011). Fast flexible modeling of RNA structure using internal coordinates. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *8*, 1247–1257.

Gan, H. H., Pasquali, S., & Schlick, T. (2003). Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. *Nucleic Acids Research*, *31*, 2926–2943.

Garst, A. D., Heroux, A., Rambo, R. P., & Batey, R. T. (2008). Crystal structure of the lysine riboswitch regulatory mRNA element. *The Journal of Biological Chemistry*, *283*, 22347–22351.

Gesteland, R. F., Cech, T. R., & Atkins, J. F. (2005). *The RNA world* (3rd ed.). New York: Cold Spring Harbor Press.

Guntert, P., Mumenthaler, C., & Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *Journal of Molecular Biology*, *273*, 283–298.

Huang, Q., Purzycka, K. J., Lusvarghi, S., Li, D., Legrice, S. F., & Boeke, J. D. (2013). Retrotransposon Ty1 RNA contains a 5′-terminal long-range pseudoknot required for efficient reverse transcription. *RNA*, *19*, 320–332.

Jonikas, M. A., Radmer, R. J., & Altman, R. B. (2009). Knowledge-based instantiation of full atomic detail into coarse-grain RNA 3D structural models. *Bioinformatics*, *25*, 3259–3266.

Jonikas, M. A., Radmer, R. J., Laederach, A., Das, R., Pearlman, S., Herschlag, D., et al. (2009). Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA*, *15*, 189–199.

Jossinet, F., Ludwig, T. E., & Westhof, E. (2010). Assemble: An interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics*, *26*, 2057–2059.

Krahenbuhl, B., Lukavsky, P., & Wider, G. (2014). Strategy for automated NMR resonance assignment of RNA: Application to 48-nucleotide K10. *Journal of Biomolecular NMR*, *59*, 231–240.

Lee, E. R., Baker, J. L., Weinberg, Z., Sudarsan, N., & Breaker, R. R. (2010). An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science*, *329*, 845–848.

Leontis, N., & Westhof, E. (2012). RNA 3D structure analysis and prediction. In N. Leontis, & E. Westhof (Eds.), *Nucleic acids and molecular biology series*. *Vol. 27*. Berlin and Heidelberg: Springer-Verlag.

Lukasiak, P., Antczak, M., Ratajczak, T., Bujnicki, J. M., Szachniuk, M., Adamiak, R. W., et al. (2013). RNAlyzer—Novel approach for quality analysis of RNA structural models. *Nucleic Acids Research*, *41*, 5978–5990.

Lusvarghi, S., Sztuba-Solinska, J., Purzycka, K. J., Pauly, G. T., Rausch, J. W., & Grice, S. F. (2013). The HIV-2 Rev-response element: Determining secondary structure and defining folding intermediates. *Nucleic Acids Research*, *41*, 6637–6649.

Martinez, H. M., Maizel, J. V., Jr., & Shapiro, B. A. (2008). RNA2D3D: A program for generating, viewing, and comparing 3-dimensional models of RNA. *Journal of Biomolecular Structure & Dynamics*, *25*, 669–683.

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., & Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 7287–7292.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta*, *405*, 442–451.

Merino, E. J., Wilkinson, K. A., Coughlan, J. L., & Weeks, K. M. (2005). RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *Journal of the American Chemical Society*, *127*, 4223–4231.

Pachulska-Wieczorek, K., Purzycka, K. J., & Adamiak, R. W. (2006). New, extended hairpin form of the TAR-2 RNA domain points to the structural polymorphism at the 5′ end of the HIV-2 leader RNA. *Nucleic Acids Research*, *34*, 2984–2997.

Pang, P. S., Elazar, M., Pham, E. A., & Glenn, J. S. (2011). Simplified RNA secondary structure mapping by automation of SHAPE data analysis. *Nucleic Acids Research*, *39*, e151.

Parisien, M., Cruz, J. A., Westhof, E., & Major, F. (2009). New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA*, *15*, 1875–1885.

Parisien, M., & Major, F. (2008). The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, *452*, 51–55.

Peselis, A., & Serganov, A. (2014). Themes and variations in riboswitch structure and function. *Biochimica et Biophysica Acta*, *1839*, 908–918.

Popenda, M., Bielecki, L., & Adamiak, R. W. (2006). High-throughput method for the prediction of low-resolution, three-dimensional RNA structures. *Nucleic Acids Symposium Series (Oxf)*, *50*, 67–68.

Popenda, M., Blazewicz, M., Szachniuk, M., & Adamiak, R. W. (2008). RNA FRABASE version 1.0: An engine with a database to search for the three-dimensional fragments within RNA structures. *Nucleic Acids Research*, *36*, D386–D391.

Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K. J., Lukasiak, P., Bartol, N., et al. (2012). Automated 3D structure composition for large RNAs. *Nucleic Acids Research*, *40*, e112.

Popenda, M., Szachniuk, M., Blazewicz, M., Wasik, S., Burke, E. K., Blazewicz, J., et al. (2010). RNA FRABASE 2.0: An advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *BMC Bioinformatics*, *11*, 231.

Purzycka, K. J., Pachulska-Wieczorek, K., & Adamiak, R. W. (2011). The in vitro loose dimer structure and rearrangements of the HIV-2 leader RNA. *Nucleic Acids Research*, *39*, 7234–7248.

Puton, T., Kozlowski, L. P., Rother, K. M., & Bujnicki, J. M. (2013). CompaRNA: A server for continuous benchmarking of automated methods for RNA secondary structure prediction. *Nucleic Acids Research*, *41*, 4307–4323.

Ren, A., Rajashankar, K. R., & Patel, D. J. (2012). Fluoride ion encapsulation by Mg2 + ions and phosphates in a fluoride riboswitch. *Nature*, *486*, 85–89.

Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, *11*, 129.

Rose, P. W., Beran, B., Bi, C., Bluhm, W. F., Dimitropoulos, D., Goodsell, D. S., et al. (2011). The RCSB Protein Data Bank: Redesigned web site and web services. *Nucleic Acids Research*, *39*, D392–D401.

Rother, M., Rother, K., Puton, T., & Bujnicki, J. M. (2011). ModeRNA: A tool for comparative modeling of RNA 3D structure. *Nucleic Acids Research*, *39*, 4007–4022.

Schuster, P., Fontana, W., Stadler, P. F., & Hofacker, I. L. (1994). From sequences to shapes and back: A case study in RNA secondary structures. *Proceedings of the Biological Sciences*, *255*, 279–284.

Schwieters, C. D., Kuszewski, J. J., Tjandra, N., & Clore, G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance*, *160*, 65–73.

Seetin, M. G., & Mathews, D. H. (2011). Automated RNA tertiary structure prediction from secondary structure and low-resolution restraints. *Journal of Computational Chemistry*, *32*, 2232–2244.

Serganov, A., & Nudler, E. (2013). A decade of riboswitches. *Cell*, *152*, 17–24.

Sharma, S., Ding, F., & Dokholyan, N. V. (2008). iFoldRNA: Three-dimensional RNA structure prediction and folding. *Bioinformatics*, *24*, 1951–1952.

Smith, K. D., Shanahan, C. A., Moore, E. L., Simon, A. C., & Strobel, S. A. (2011). Structural basis of differential ligand recognition by two classes of bis-(3′-5′)-cyclic dimeric guanosine monophosphate-binding riboswitches. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 7757–7762.

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

Wilkinson, K. A., Gorelick, R. J., Vasa, S. M., Guex, N., Rein, A., Mathews, D. H., et al. (2008). High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biology*, *6*, e96.

Xu, Z., Almudevar, A., & Mathews, D. H. (2012). Statistical evaluation of improvement in RNA secondary structure prediction. *Nucleic Acids Research*, *40*, e26.

Xu, X. J., Zhao, P. N., & Chen, S. J. (2014). Vfold: A web server for RNA structure and folding thermodynamics prediction. *PLoS One*, *9*, e107504. http://dx.doi.org/10.1371/journal.pone.0107504. eCollection 2014.