

Chapter 2

Modeling RNA Molecules

Neocles Leontis and Eric Westhof

Chercher plutôt la rigueur dans l'enchaînement de la pensée plutôt que la précision dans les résultats. Le modèle le plus crédible n'est pas nécessairement le plus réaliste, car il demande l'exagération des traits caractéristiques par rapport aux traits contingents.

—Abraham Moles, *Les sciences de l'imprécis*, Paris, Seuil (1990)

Strive for rigor in the logical train of thought rather than in the precision of the results. The most enlightening scientific model is not necessarily the most realistic one, because it is necessary to exaggerate the characteristic features with respect to the contingent ones.

—Translated by the authors

2.1 Introduction

A primary activity of scientific work is the construction of models to represent the nature and workings of phenomena we observe in the world around us. Models that represent the molecular components of living system in three dimensions (3D) and at atomic resolution are highly valued in molecular and structural biology. For example, the decipherment of the 3D structures of ribosomes, the complex protein-synthesizing nanomachines of the cell, represents a tremendous achievement, recently recognized with the Nobel Prize in Chemistry (http://nobelprize.org/nobel_prizes/chemistry/laureates/2009/). Nonetheless, this phenomenal success is

N. Leontis

Department of Chemistry, Bowling Green State University, Bowling Green, OH 43403, USA

e-mail: leontis@bgsu.edu

E. Westhof (✉)

Architecture et Réactivité de l'ARN, Institut de Biologie Moléculaire et Cellulaire du Centre National de la Recherche Scientifique, Université de Strasbourg, Strasbourg, France

e-mail: e.westhof@ibmc.u-strasbg.fr

tempered by the realization that even now, over 10 years after the first ribosome structures were solved, we still do not understand fully several aspects of their functioning. For all who have grappled with the complexities of ribosome structures, Richard Feynmann's pithy statement, "What I cannot create, I do not understand," rings especially true (Hawking 2001). This physics-based realization contrasts with another point of view of modeling. To paraphrase R. W. Hamming, who said, "The purpose of computing is insight, not numbers" (Hamming 1971), we should remember that the purpose of molecular modeling is functional insight, not detailed atomic models per se. Therefore, as we seek to improve our abilities to construct 3D models for molecules for which we do not yet have experimental atomic-resolution structures, we should bear in mind that it may not be necessary to achieve some arbitrary precision in the atomic coordinates to provide insight into biological function. Rather, we should think carefully to identify those predicted features that yield important insights (Table 2.1).

Thus, for those engaged in RNA modeling, critical questions to ponder include: *What* do biologists, who are trying to unravel the roles of RNA in complex biological processes (growth and development, learning and cognition, immune and stress responses, and disease), *really* need to know about the 3D structures of the RNA molecules they study, and in *what form* do they need it? In this context, how deep do we need to go into atomic details to gain useful insights? How can knowledge of RNA 3D structure be applied to infer RNA function? It is crucial to bear in mind that, historically, some imprecise models have been richer in biological insight than other, very precise ones. The famous, original 3D model for double-stranded DNA of Watson and Crick stands out in this respect.

With these fundamental issues as background, we turn to the reasons for renewed interest in RNA 3D modeling: New high-throughput experimental approaches, developed in the postgenomic era, have revealed the pervasive role of noncoding RNA molecules in all aspects of gene expression, from chromosome remodeling and regulation of epigenetic processes to transcription, splicing, mRNA transport and targeting, and translation and its regulation. Furthermore, while the number of protein-coding genes has changed little from the genome of the tiny 1,000-cell nematode *Caenorhabditis elegans* to that of our own species, *H. sapiens*, the number of ncRNAs has exploded and appears to scale with biological complexity (Taft et al. 2007). Evidence is building that many of these ncRNAs, like those involved in splicing and translation, which have been known for many years, function at least in part by forming complex 3D structures to interact specifically with proteins, other nucleic acids, and a wide range of small molecules.

2.2 Defining the Problem

For RNA molecules that form discrete 3D structures, the folding problem can be simply stated: What is the mapping from sequence space to three-dimensional space? As many biologically active RNA molecules are very long (up to thousands

Table 2.1 The various prediction programs described in the book with their main elements and outputs

Author	Program	Strategy	Output	CG	KB	MD	MC	2D	3D
Altman	RNABuilder (1)	Method: multiresolution, KB modeling with or without templates and fragment assembly	3D atomic structure model	x	x				x
Altman	NAST	Method: coarse-grained KB modeling with MD sampling	3D coarse-grained structure model	x	x	x			x
Bujnicki	ModeRNA	Input: sequence and 3D structure of template Method: homology and template-based modeling	3D atomic structure model						x
Bujnicki	SimRNA	Input: sequence Method: de novo coarse-grained modeling with Monte Carlo sampling with KB potential	Coarse-grained 3D model from which 3D atomic structure model is built using rebuild RNA	x	x		x		x
Chen	Vfold	Input: sequence and 2D structure Method: lattice-based coarse-grained modeling. MD refinement to obtain atomistic model	Free energies of pseudoknotted structures; coarse-grained 3D structures from which 3D atomic structure model is built	x		x			x
Das	FARNA/FARFAR	Input: sequence and knowledge-based potential Method: de novo modeling using KB potentials and MC sampling	3D atomic structure model		x		x		x
Dokholyan	DMD	Input: sequence and 2D and tertiary constraints Method: coarse-grained, multiscale conformational sampling by discrete MD	Full 3D atomic coordinates deduced from coarse-grained structure model. Folding thermodynamics	x		x			x
Major Santa Lucia	MC-FOLD/MC-Sym RNA123 (homology modeling module)	KB grammar for fragment assembly Input: sequence and 3D structure template Method: homology modeling with MD energy refinement	3D atomic structure model 3D atomic structure model		x		x		x x

(continued)

Table 2.1 (continued)

Author	Program	Strategy	Output	CG	KB	MD	MC	2D	3D
Shapiro	RNA2D3D	Input: 2D structure. Atomic modeling using fragment assembly and MD energy minimization	3D structures			x			x
Shapiro	MPGAfold	Genetic algorithm to optimize 2D structure	2D structures with pseudoknots					x	
Wang	G2G (global measurements to global structure)	Inputs: sequence and 2D structure; RDC and SAXS measurements method	Global structure of relative helical orientations					x	

Abbreviations: CG coarse graining, KB knowledge-based, MD molecular dynamics, MC Monte Carlo, 2D two-dimensional, 3D three-dimensional
References: RNABuilder (Flores et al. 2011); NAST (Jonikas et al. 2009); ModERNA and SimRNA (Rother et al. 2011); Vfold (Cao and Chen 2011); Fama (Das and Baker 2007); FARFAR (Das et al. 2010); DMD (Ding et al. 2008); MC-FOLD/MC-Sym (Parisien and Major 2008); RNA2D3D (Martinez et al. 2008); MPGAfold (Shapiro et al. 2006); G2G (Wang et al. 2009)

of nucleotides), this question is relevant for those portions of RNA sequence that adopt stable architectures, required for their function during at least some period of time. In other words, given a sequence, produce a set of 3D coordinates for the nucleotides, that is biologically relevant and that satisfies the stereochemistry and physical chemistry of RNA molecules.

2.2.1 RNA Modeling Compared to Protein Modeling

In this regard, the parallels and contrasts between RNA and protein structure prediction and folding are apparent. Like proteins, RNA molecules are flexible linear polymers with astronomical conformational possibilities. Unlike proteins, RNA structures generally partition quite cleanly between secondary and tertiary hierarchical levels (Brion and Westhof 1997; Woodson 2010, 2011). Thus, as a rule, the first step in successful 3D modeling of RNA passes through a high-quality prediction of the main secondary structure elements. The state of the art in RNA secondary structure prediction is reviewed by Steger and coauthors in the third chapter of this volume. At the present state of our modeling efforts, the nature of the input data can play a decisive role at this stage of the process. Indeed, despite significant advances in 2D structure prediction, current methods still rely on theoretical approximations and an incomplete set of empirical energy parameters. Thus, working on a single RNA sequence may lead to incorrect evaluation of the importance or the role of one or more structural elements. The idiosyncrasies contained in single sequences can, however, often be ironed out by the use of multiple homologous sequences. Moreover, for RNA molecules, in contrast to proteins, one can obtain many additional experimental data containing much 3D information, using chemical or enzymatic probing and footprinting, small-angle X-ray scattering (SAXS), and cross-linking. The incorporation and computer use of such data changes the tractability of the problem. The chapters by Laederach, Wang and Fabris, and their coauthors (Chapters 15–17) address some of these issues and illustrate the challenges and power of integrating modern experimental data collection with modeling methods.

2.2.2 Defining the Inputs for RNA 3D Modeling

Inputs for the modeling of RNA 3D structure include, in addition to the sequence of the target RNA, the derived secondary structure and the sequences of available homologues, as well as all available experimental data. The database of known RNA 3D structures should also be considered an important resource for 3D modeling. This is especially the case for those approaches relying on a modular view of RNA architecture with the resulting assembly of RNA elements and modules (Jossinet et al. 2010; Westhof et al. 2011).

2.3 3D Modeling Methods and Approaches

A variety of modeling approaches are represented in the contributions to this volume. Some common themes emerge and will be summarized briefly with reference to specific chapters. As will become apparent to readers, promising approaches are rapidly adopted by multiple research groups, although specific implementations vary in ways that are usually not easy to discern. This volume focuses on methods that aim to achieve automaticity in 3D modeling, in the sense that they should require very little human intervention in the modeling process, beyond defining the inputs for the specific problem. The effort, rather, is focused “up front” on designing the algorithms and extracting and compiling relevant knowledge concerning RNA structure from structure databases for automated use by the implemented algorithms.

2.3.1 *Homology Modeling*

Automated methods generally address one or both of two distinct problems in biological structure prediction, namely, homology modeling and de novo prediction. Homology modeling concerns building atomically accurate 3D models of RNA molecules using at least one homologous 3D structure as template. RNA homology modeling draws on vast experience with protein homology modeling, and so considerable progress has been made already. The contributions of Altman, Bujnicki, and Santa Lucia focus, at least in part, on homology modeling and, between them, exhaustively address the issues involved.

2.3.2 *De Novo Modeling*

De novo prediction is necessary when no homologous 3D structure is known that can serve as a template for modeling. It is considerably more challenging than homology modeling, as it often requires generating a brand new 3D architecture from any known heretofore. As the goal is to do this without expert human input, the general approach is to generate large numbers of possible architectures and then to evaluate them, using what is already known about RNA structure. Automated, de novo 3D modeling approaches are therefore distinguished operationally by the kind of algorithm employed to generate potential 3D structural models, and also by the nature of the encapsulated knowledge concerning RNA structure that is used to score and rank models to arrive at a small set of predicted 3D structures, or in the favorable case, a single structure. The models generated by conformation-sampling algorithms are called “decoys” by practitioners. For the final output, most programs produce an all-atom predicted structure, which is generally quite “correct” in its

local, stereochemical detail, in the sense that bond lengths and angles are within allowed ranges and the model contains no unphysical nonbonded contacts. But this local precision, which most programs achieve routinely, should not mislead users of predicted 3D models into assuming the model is accurate on larger, biologically relevant length scales, ranging from structures of modular motifs to overall folds and architectures.

The contributions of Altman (Chapter 8), Bujnicki (Chapter 5), Chen (Chapter 10), Das (Chapter 4), Dokhalyan (Chapter 9), Santa Lucia (Chapter 6), and Shapiro and their coworkers (Chapter 7) address *de novo* 3D modeling and among them cover the major methods in use today. All of these methods deploy some kind of algorithm to sample conformation space and some kind of knowledge-based methods to score and rank proposed solutions to the 3D prediction problem. In addition, most approaches rely on some kind of reduced representation of the RNA structure (“coarse graining”) to speed up the calculations and allow more thorough exploration of conformational space with available computer resources. Coarse graining is an art that requires striking the right balance between speed of calculation and sufficiently detailed representation of RNA structure to capture the molecular features that stabilize the active conformations. Other ways to speed up conformational sampling involve modification of the algorithms that propagate the dynamics, as represented by the discrete molecular dynamics (DMD) method reported by Dokhalyan and coworkers.

2.3.3 Defining the Outputs of Different Modeling Approaches

The outputs of modeling studies depend on the modeling approach and the aim of the study. Indeed, output data can be full atomic coordinates for every single nucleotide or, in the case of coarse-grained methods, coordinates for only a subset of atoms or even a single pseudoatom representing each nucleotide. The different outputs are directly related to the granularity of the modeling approach. Nonetheless, nominally atomic-resolution models, when poorly refined or badly assembled, may be no better or even worse than coarse-grained models, if the characteristic base-pairing and base-stacking interactions of the structures are not represented accurately.

2.3.4 Precision of Models vs. Accuracy of Models

There is no necessary correlation between precision and accuracy, and models with comparable precision can differ substantially in the accuracy with which they predict the important interactions between nucleotides that define the RNA 3D structure. Thus, low-precision models can be very accurate (e.g., the original Watson–Crick model for DNA) and highly precise ones can be partly or totally

inaccurate and thus misleading. Clearly, less-accurate models may not be at all pertinent for structural biology, while less-precise models can be very rich and enlightening. Still, these considerations should not be taken as license for not using in model building, whenever possible, high-resolution building blocks that are precise with respect to bond lengths and angles within nucleotides, and H-bond distances, van der Waals contacts, and relative orientations within base pairs and other interactions.

2.4 Databases for Extracting Knowledge

All of the precise structural data regarding RNA comes ultimately from atomic-resolution X-ray structures of nucleotides, oligonucleotides, and various biologically relevant structures, ranging in size from individual helical elements to the full ribosome. These data comprise all our basic knowledge of bond lengths, angles, and stereochemistry, as well as interaction preferences, including all types of base pairs and most stacking and base–backbone interactions. This information is used to build force fields and to infer rules for assembly of molecular moieties. These force fields and energetic rules are then used for producing and optimizing structures, sampling the conformational space, or simulating molecular dynamics. The quality and general value of the deduced force fields will strongly depend on the number and variety of structures available. In addition, the quality of the structures is of primary importance; it is directly related to the crystallographic resolution of the X-ray data and on the refinement process since a minor fraction of X-ray structures are obtained at true atomic resolution. One key parameter for compiling reference databases for knowledge extraction is the nonredundancy of the structures that are included in order to avoid bias in the deduced parameters. The chapter by Leontis and Zirbel (Chapter 14) addresses these issues and details a nonredundant database of structures extremely valuable for extracting knowledge about RNA as well as for benchmarking modeling strategies. In this respect, it is worth noting that less than 100 nonredundant RNA structures have been solved at 2-Å resolution or better.

2.5 Evaluating Models or “The Proof of the Pudding Is in the Eating”

As discussed above, 3D models are produced either to monitor our progress in the understanding and use of the physicochemical rules governing RNA architecture or to provide insight and help to experimentalists in the interpretation and meanings of biological data and in the design of new experiments. Although objectives may differ, in every case the models produced should be evaluated to assess their relevance to biological reality. Models that make testable predictions are

especially valuable and, as emphasized above, need not be particularly precise. Additional experiments devised on the basis of a given model will provide the relevant tests for evaluating it. Depending on the outcome, the model may be retained and perhaps “tweaked,” or it may be rejected and radically revised, leading to new biological insight and further experimental tests. On the other hand, to assess the validity of force fields as well as other empirical assembly rules, precise numerical comparisons have to be performed in a systematic way. This highlights the need for discriminating and meaningful metrics to compare and evaluate predicted vs. experimental structures.

2.5.1 Metrics for Evaluating Models

The most common metric is the root mean square deviations (RMSDs) on corresponding atoms between the predicted and experimental models. RMSDs are easy to compute and yield a simple measure. However, to interpret RMSD values, some critical length scales in RNA structures should be kept in mind for comparison: First, stacking distance between bases is about 3.4 Å; second, successive P–P distance in RNA helices is about 7 Å. While RMSD values below 3.4 Å are of real value, RMSD values beyond 8 Å must be treated with caution. In addition, RMSDs, as generally calculated with rigid-body fitting, spread the errors between two sets of coordinates over the whole ensemble. Consequently, even correctly modeled regions will not superimpose properly and thus will also contribute to the overall RMSD value. Therefore, RMSD values should be supplemented with local structural comparisons, including, for example, the numbers of correct base stackings and of correct Watson–Crick base pairs and, especially for 3D architectures, the number of non-Watson–Crick pairs, correct both with respect to pairing partners and base-pair types (Leontis and Westhof 2001). For a summary of the types of non-Watson–Crick base pairs, see the Appendix of this volume. We stress the importance of predicting the correct non-WC pairings as well as the correct base stackings, both of which are key because there is no three-dimensional architecture without non-Watson–Crick pairs and additional stackings between pairs. While a simple mapping of the 2D structure into a 3D structure does lead to a three-dimensional fold, such a fold will lack the additional stackings or RNA–RNA contacts that are characteristic of the complete 3D architecture. In short, correct predictions imply correct choices of new base stackings between single-stranded nucleotides and helices as well as new long-range base-pair contacts. For these reasons, two new metrics particularly suitable to RNA were introduced: the deformation index and the deformation profile (Parisien et al. 2009). The deformation index monitors the fidelity of the interaction network and encompasses base-stacking and base-pairing interactions within the target structure. The deformation profile highlights dissimilarities between structures at the nucleotide scale for both intradomain and interdomain interactions. These tools demonstrate that there is little correlation between RMSD and interaction network fidelity. To improve force fields or modeling approaches, it is mandatory to assess the

origins of the errors. The deformation profile is a very useful tool for identifying the origins of incorrect modeling decisions.

2.5.2 Necessity for Objective Evaluation of Modeling Efforts: RNA-CASP

Structure prediction methods for proteins were boosted and consolidated by the CASP project (Critical Assessment of techniques for protein Structure Prediction), a systematic and worldwide evaluation of the predictions of new structures, prior to their publication (Kryshtafovych et al. 2005; Moult et al. 2009). CASP has proven extremely useful, productive, and constructive for benchmarking the progress made in the generation of new ideas and the objective assessment of the newly developed techniques. We believe that setting up a similar process will prove very healthy for the RNA structure-modeling field. To do so, several hurdles need to be overcome. In the case of RNA prediction, two levels would have to be distinguished, namely, the prediction of secondary structure and the modeling of 3D (tertiary) structure. The main issue, however, is how to establish efficient communication between research groups that determine RNA structures, whether at the secondary or tertiary structure levels, and research groups that predict RNA structures, so the latter can register their predictions before the structures are published. Clearly, despite the amazing advances in all aspects of the production of 3D RNA structures by X-ray crystallographic, NMR, or cryo-EM methods, the number of new structures produced per year remains rather low. The proposed process would follow these lines: (1) A structural group working on a new RNA structure (X-ray, NMR, chemical probing, cryoelectron microscopy, or mass spectroscopy) makes known their willingness to “play the game.” (2) The group sends the sequence of the RNA under investigation to the coordinator. (3) The coordinator, without disclosing the identity of the experimental laboratory or the function and origin of the RNA, distributes the sequence to the theoreticians ready to tackle the challenge. Each theoretical group must agree not to disclose the sequence or distribute it further or to disclose its own progress or results in any fashion before publication of the structure by the experimental group. (4) The deadline for submitting structure predictions to the coordinator is agreed upon at the outset and generally will coincide with the date the experimental group submits their structures for publication. (5) During a special meeting, the coordinator discloses the theoretical results, and they are compared with the published experimental structures. (6) Special guidelines and rules for the comparisons will be agreed upon before the writing and publication of the analysis. Several laboratories dedicated to RNA bioinformatics around the world have expressed their keen interest to participate in such regularly held contests. The success and real progress generated by CASP in protein structure prediction should encourage us all to pursue this endeavor in the form of an ongoing RNA-CASP process. A first test of RNA-CASP was initiated at the end of 2010 and is now in the process of being published (Cruz et al. 2012).

2.6 Complications Limiting Modeling Approaches

Biological reality is complicated, and the applicability of physicochemical approaches based fundamentally on assumptions of thermodynamic equilibrium should always be properly evaluated as part of the theoretical modeling process. First, RNA molecules begin folding almost immediately as they are transcribed (cotranscriptional folding) so the issue of kinetic vs. equilibrium control in formation of biologically relevant structures is always a real one (Cruz and Westhof 2009). When the first structure to form is not the biologically relevant one, chaperone molecules are observed to play additional important roles. RNA molecules rarely act alone; on the contrary, they almost always act by binding to other RNA molecules or to proteins, and very frequently they bind to both types of macromolecules, if not also to small molecules.

An especially complicated problem is that of “induced fit,” which occurs when the conformation adopted by an RNA molecule in isolation is not identical to that found in a complex with a small molecule ligand, antibiotic, or another RNA or a protein (Williamson 2000). Even small ligands, like hydrated magnesium ions, are difficult to treat in an appropriate fashion. Magnesium ions are especially difficult to treat when they bind, not as outer-sphere complexes (with a full share of coordinated water molecules), but instead as inner-sphere complexes, with the loss of one or more water molecules and direct coordination to the RNA, generally in a state different from the original magnesium-free ion state (see Chapter 11 by P. Auffinger). Treating induced fit, at minimum, requires that the full dynamics of an RNA fragment be known in order to be able to select the proper conformation binding a given ligand. And it is not at all proven that the range of conformations accessible by the usual methods of molecular dynamics simulations, for example, actually covers the states obtained in the presence of the ligand or protein. Thus, one can study the dynamics of the A-site of the ribosome alone or in complex with an antibiotic (because crystal structures exist for all those different states), but the docking of an antibiotic to the A-site starting from an “empty state” (which is not the same as the state of the bound complex minus the ligand) has not been achieved yet (Moitessier et al. 2006).

2.7 Challenges for the Future: Dealing with Massive Data Streams and Connecting to Biology

Several main questions of great potential for biology continue to be actively pursued, and yet we have barely scratched the surface. One is the use of modeling predictions, firstly for searching noncoding RNAs in genomes and secondly for choosing among genomic regions those that are susceptible to fold into architectural domains or fragments (e.g., as riboswitches do). Another major question is the prediction of protein-binding sites along RNA sequences. Some consensus binding

sequences are known, but in most cases, only knowledge of the RNA 3D fold allows the full understanding of the binding surface and RNA–protein contacts.

2.8 Conclusion

For modeling to be relevant to twenty-first century biological research, data pipelines need to be developed, maintained, and intelligently monitored to deal with the massive data streams produced by modern high-throughput sequencing methods. This means aiming for full automaticity at all steps of the computations. In this way, one should be able to link computational predictions with the experimental high-throughput technologies being constantly developed and refined. The establishment of such links between experimental and computational high-throughput techniques will bring us closer to the establishment of complete “RNA structuromes” for a given microbial or multicellular organism (Underwood et al. 2010; Weeks et al. 2011).

Acknowledgment NBL expresses his gratitude to Vassiliki Leontis for her support during the preparation of this book. NBL was supported by grants from the National Institutes of Health (grant numbers 1R01GM085328-01A1 to Craig Zirbel and NBL and 2R15GM055898-05 to NBL).

References

- Amaral PP, Dinger ME, Mercer TR, Mattick JS (2008) The eukaryotic genome as an RNA machine. *Science* 319:1787–1789. doi:[10.1126/science.1155472](https://doi.org/10.1126/science.1155472)
- Brion P, Westhof E (1997) Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct* 26:113–137. doi:[10.1146/annurev.biophys.26.1.113](https://doi.org/10.1146/annurev.biophys.26.1.113)
- Cruz JA, Westhof E (2009) The dynamic landscapes of RNA architecture. *Cell* 136:604–609. doi:[10.1016/j.cell.2009.02.003](https://doi.org/10.1016/j.cell.2009.02.003)
- Cruz JA, MF Blanchet, M Boniecki, JM Bujnicki, SJ Chen, S Cao, R Das, F Ding, NV Dokholyan, SC Flores, L Huang, CA Lavender, V Lisi, F Major, K Mikolajczak, DJ Patel, A Philips, T Puton, J Santalucia, F Sijeniyi, T Hermann, K Rother, M Rother, A Serganov, M Skorupski, T Soltysinski, P Sripakdeevong, I Tuszyńska, KM Weeks, C Waldsich, M Wildauer, NB Leontis and E Westhof (2012). RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* 18:610–625. doi:[10.1261/ma.031054.11](https://doi.org/10.1261/ma.031054.11)
- Cao S, Chen SJ (2011) Physics-based de novo prediction of RNA 3D structures. *J Phys Chem B* 115:4216–4226. doi:[10.1021/jp112059y](https://doi.org/10.1021/jp112059y)
- Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cao S, Das R, Ding F, Dokholyan NV, Flores SC, Huang L, Lavender CA, Lisi V, Major F, Mikolajczak K, Patel DJ, Philips A, Puton T, Santalucia J, Sijeniyi F, Hermann T, Rother K, Rother M, Serganov A, Skorupski M, Soltysinski T, Sripakdeevong P, Tuszyńska I, Weeks KM, Waldsich C, Wildauer M, Leontis NB, Westhof E (2012) RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* 18:610–625. doi:[10.1261/ma.031054.111](https://doi.org/10.1261/ma.031054.111)
- Das R, Baker D (2007) Automated de novo prediction of native-like RNA tertiary structures. *Proc Natl Acad Sci USA* 104:14664–14669. doi:[10.1073/pnas.0703836104](https://doi.org/10.1073/pnas.0703836104)

- Das R, Karanicolas J, Baker D (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat Methods* 7:291–294. doi:[10.1038/nmeth.1433](https://doi.org/10.1038/nmeth.1433)
- Flores SC, Sherman MA, Bruns CM, Eastman P, Altman RB (2011) Fast flexible modeling of RNA structure using internal coordinates. *IEEE/ACM Trans Comput Biol Bioinform* 8:1247–1257. doi:[10.1109/TCBB.2010.104](https://doi.org/10.1109/TCBB.2010.104)
- Hamming RW (1971) Introduction to applied numerical analysis. McGraw-Hill, New York
- Hawking SW (2001) The universe in a nutshell. Bantam Books, New York
- Jossinet F, Ludwig TE, Westhof E (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* 26:2057–2059. doi:[10.1093/bioinformatics/btq321](https://doi.org/10.1093/bioinformatics/btq321)
- Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB (2009) Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* 15:189–199. doi:[10.1261/rna.1270809](https://doi.org/10.1261/rna.1270809)
- Kryshtafovych A, Venclovas C, Fidelis K, Moult J (2005) Progress over the first decade of CASP experiments. *Proteins* 61(Suppl 7):225–236. doi:[10.1002/prot.20740](https://doi.org/10.1002/prot.20740)
- Leontis NB, Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* 7:499–512
- Moitessier N, Westhof E, Hanessian S (2006) Docking of aminoglycosides to hydrated and flexible RNA. *J Med Chem* 49:1023–1033. doi:[10.1021/jm0508437](https://doi.org/10.1021/jm0508437)
- Moult J, Fidelis K, Kryshtafovych A, Rost B, Tramontano A (2009) Critical assessment of methods of protein structure prediction – Round VIII. *Proteins* 77(Suppl 9):1–4. doi:[10.1002/prot.22589](https://doi.org/10.1002/prot.22589)
- Martinez HM, Maizel JV Jr, Shapiro BA (2008) RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA. *J Biomol Struct Dyn* 25:669–683
- Parisien M, Cruz JA, Westhof E, Major F (2009) New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA* 15:1875–1885. doi:[10.1261/rna.1700409](https://doi.org/10.1261/rna.1700409)
- Parisien M, Major F (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 452:51–55. doi:[10.1038/nature06684](https://doi.org/10.1038/nature06684)
- Rother M, Milanowska K, Puton T, Jeleniewicz J, Rother K, Bujnicki JM (2011) ModeRNA server: an online tool for modeling RNA 3D structures. *Bioinformatics* 27:2441–2442. doi:[10.1093/bioinformatics/btr400](https://doi.org/10.1093/bioinformatics/btr400)
- Shapiro BA, Kasprzak W, Grunewald C, Aman J (2006) Graphical exploratory data analysis of RNA secondary structure dynamics predicted by the massively parallel genetic algorithm. *J Mol Graph Model* 25:514–531. doi:[10.1016/j.jmgm.2006.04.004](https://doi.org/10.1016/j.jmgm.2006.04.004)
- Taft RJ, Pheasant M, Mattick JS (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays* 29:288–299. doi:[10.1002/bies.20544](https://doi.org/10.1002/bies.20544)
- Underwood JG, Uzilov AV, Katzman S, Onodera CS, Mainzer JE, Mathews DH, Lowe TM, Salama SR, Haussler D (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat Methods* 7:995–1001. doi:[10.1038/nmeth.1529](https://doi.org/10.1038/nmeth.1529)
- Wang J, Zuo X, Yu P, Xu H, Starich MR, Tiede DM, Shapiro BA, Schwieters CD, Wang YX (2009) A method for helical RNA global structure determination in solution using small-angle x-ray scattering and NMR measurements. *J Mol Biol* 393:717–734. doi:[10.1016/j.jmb.2009.08.001](https://doi.org/10.1016/j.jmb.2009.08.001)
- Weeks KM, Mauger DM (2011) Exploring RNA structural codes with SHAPE chemistry. *Acc Chem Res* 44:1280–1291. doi:[10.1021/ar200051h](https://doi.org/10.1021/ar200051h)
- Westhof E, Romby P (2010) The RNA structurome: high-throughput probing. *Nat Methods* 7:965–967. doi:[10.1038/nmeth1210-965](https://doi.org/10.1038/nmeth1210-965)
- Westhof E, Masquida B, Jossinet F (2011) Predicting and modeling RNA architecture. *Cold Spring Harbor Perspect Biol* 3:doi:[10.1101/cshperspect.a003632](https://doi.org/10.1101/cshperspect.a003632)
- Williamson JR (2000) Induced fit in RNA-protein recognition. *Nat Struct Biol* 7:834–837. doi:[10.1038/79575](https://doi.org/10.1038/79575)
- Woodson SA (2010) Compact intermediates in RNA folding. *Annu Rev Biophys* 39:61–77. doi:[10.1146/annurev.biophys.093008.131334](https://doi.org/10.1146/annurev.biophys.093008.131334)
- Woodson SA (2011) RNA folding pathways and the self-assembly of ribosomes. *Acc Chem Res*. doi:[10.1021/ar2000474](https://doi.org/10.1021/ar2000474)

RNA 3D Structure Analysis and Prediction

Leontis, N.; Westhof, E. (Eds.)

2012, X, 402 p., Hardcover

ISBN: 978-3-642-25739-1