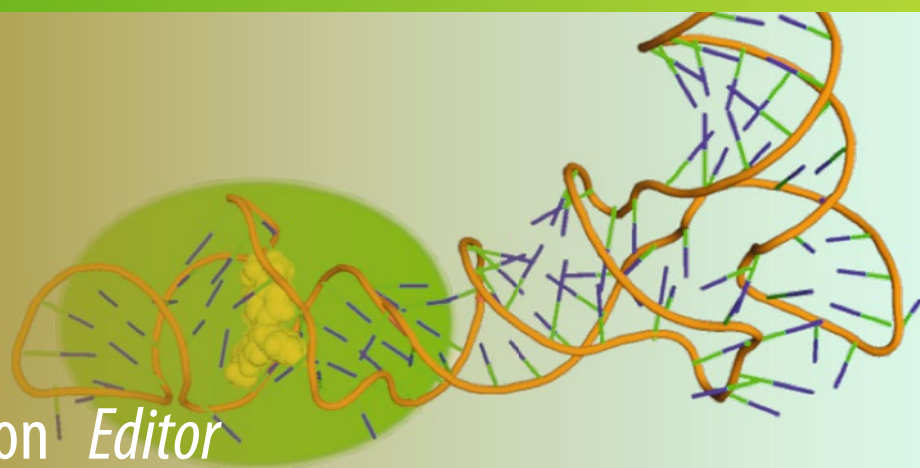


Methods in
Molecular Biology 1316

Springer Protocols



Luc Ponchon *Editor*

RNA Scaffolds

Methods and Protocols

 Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor

John M. Walker

School of Life and Medical Sciences

University of Hertfordshire

Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:

<http://www.springer.com/series/7651>

RNA Scaffolds

Methods and Protocols

Edited by

Luc Ponchon

University of Paris Descartes, Paris, France

 **Humana Press**

Editor

Luc Ponchon
University of Paris Descartes
Paris, France

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-2729-6 ISBN 978-1-4939-2730-2 (eBook)
DOI 10.1007/978-1-4939-2730-2

Library of Congress Control Number: 2015937809

Springer New York Heidelberg Dordrecht London
© Springer Science+Business Media New York 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Humana Press is a brand of Springer
Springer Science+Business Media LLC New York is part of Springer Science+Business Media (www.springer.com)

Preface

RNAs are at the center of numerous cellular phenomena and play very different roles in each. One of their roles is in particular that of organization center: the ribosome, RNA telomerase, and “Long Noncoding RNAs” are, among others, examples of RNA structures that recruit other molecules and organize biological processes. These RNAs possess structures allowing interactions with other molecules (proteins, ligands) and thus will potentialize molecular reactions.

Advances in structural biology have permitted a definition of the rules with regard to the folding of RNA allowing us today to better understand exactly how they fold and interact. As opposed to DNA, RNAs are able to adopt very variable folds and therefore are able to adopt ligand-specific structures. Contrary to proteins, we are able to create structures composed of different “RNA” modules, each of which is able to keep its activity independent from the others.

So, when they are stable, folded RNA can be used as a tool for biological, pharmacological, and/or molecular design studies. RNA presents the peculiarity, like Meccano, of being able to fold into structural domains, which can assemble and sometimes form supra-molecular objects. We can isolate, modify, or create an RNA template *de novo* to make use of its recognition or enzymatic functions.

From my point of view, an “RNA scaffold” is a synthetic or natural RNA whose structure, for example, allows one to optimize a reaction, to isolate a molecule, or to favor an interaction.

Like Biobricks, the tools based on RNA scaffolds are an example of the emergence of synthetic biology. Indeed, they participate in the creation and construction of biological objects and systems for useful purposes.

In this volume, we have tried to be as representative as possible of that which is done today. You will find detailed here processes and techniques that differ greatly from one to another.

This book reviews recently developed techniques that use “RNA scaffolds” as molecular tools. These methods cover domains as various as molecular biology, cellular biology, nanotechnology, and structural biology.

Contents

In order to design a scaffold, it is sometimes necessary to possess certain structural data. A structure can be modeled from an RNA structure data bank. In the first chapter, Chen and his colleagues describe their prediction method, Vfold, that, from a primary sequence, allows one to determine a three-dimensional model of an RNA. This method is based on a pattern-based approach. Thus, they extract different patterns (such as hairpin loops, internal loops, pseudoknot loops, and three-way junctions) from the two-dimensional structure. Based on this data, the implemented algorithm calculates a three-dimensional model.

However, if one seeks to obtain an RNA structure at an atomic level (e.g., for the rationalized design of a scaffold), crystallography is the technique of choice. RNA crystallogenesis

remains nonetheless complicated. Indeed, RNAs, as opposed to proteins, do not crystallize as easily and the X-ray diffraction is often weak. In Chapter 2, Ferré d'Amaré and his colleagues describe techniques allowing an improvement in data acquisition and notably for large RNA (over 100 nucleotides) for which only very few structures exist.

They propose a protocol allowing the increase in diffraction power of the crystals by combining two techniques. The first consisting of dehydrating the crystal, the second of substituting the divalent ions with strontium ions. Through the example of a gene-regulatory tRNA-mRNA complex, they show us the importance of these two techniques on the quality of diffraction.

A limiting factor for RNA study (notably for long RNAs) is obtaining them in large quantities. Three classic techniques exist in order to obtain the RNA: in vitro transcription, chemical synthesis, and cellular extraction.

The three following chapters describe the techniques permitting, as with proteins, the in vivo production of RNA. Chapter 3 describes a system of RNA protein co-production in the bacteria. This protocol is the logical progression of protocols already described in a previous book *Recombinant and In Vitro RNA Synthesis* in the same collection. In the aforementioned book, the authors describe the production of RNA protected from bacterial nuclease by a “tRNA” camouflage. The RNA is protected at its extremities by the tRNA chassis and accumulated in the bacteria. In this protocol, they propose to co-produce RNA-protein complexes and thus show the advantage of having a joint production of the two molecules. In Chapter 4, Wijmenga and his colleagues also use the tRNA camouflage but have added two ribozymes to the extremities of the RNA in question. Through cleaving, they can thus recuperate their RNA separated from the tRNA chassis. In their protocol, they present a labelled RNA for NMR study.

Fox and his colleagues offer a method also permitting the overproduction of RNA in bacteria in Chapter 5. They use a different chassis: the RNA ribosomal 5S. Indeed, like the tRNA, the RNA 5S possesses a fold that allows it to be accumulated in the bacteria. Their protocol thus goes into detail on the production of different RNA and how to make the cleave. So as to liberate the RNA of interest from the 5S, they propose a cleavage via the DNAzyme. These short sequences of DNA hybridize themselves with the RNA to form a structure of ribozyme-like and thus allow the RNA cleavage.

Affinity or “pull-down” techniques allow the identification of molecular complexes. In these techniques, a protein linked to a matrix serves as bait. From the cellular extracts, one can isolate the linked molecules. Macara and his colleagues propose a pull-down system in which RNA serves as bait in order to identify the RNA-binding proteins. To render the system specific and robust, their RNA bait is embedded in a tRNA scaffold and possesses an aptamer for the streptavidin in order to isolate RNA-protein complexes from extracts of mammalian cells.

Fluorescence remains a technique of choice to perform cellular localization. Green fluorescent protein (GFP) and its numerous derivatives allow one to localize the proteins in cells or tissues. However, it remains difficult to specifically locate the RNA as no naturally fluorescent RNA has been identified to date. It is therefore necessary to use indirect techniques. The two following chapters describe RNA scaffolds allowing one to make cellular localization of RNA or metabolites through fluorescence.

Broude and his colleagues describe to us in Chapter 7 a technique permitting the identification of the presence of an RNA in cells. The scaffold is composed of an antisense RNA, an RNA target, and two RNA aptamers. When the antisense hybridizes itself with the RNA target, the two aptamers can recruit two fusion proteins, which will contemplate each other

and then form the GFP and therefore fluoresce. It is therefore via the fluorescence of the GFP that they localize their RNA.

In Chapter 8, Hammond and his colleagues propose to us a technique allowing the localization of a metabolite via an RNA scaffold, which is linked to a fluorescent Chromophore. This RNA is a biosensor composed, on the one hand, of a “spinach aptamer” which links a fluorescent chromophore to DFHB1 and, on the other hand, of a specific riboswitch of the cyclic di-GMP. The idea being that the fixation of the fluorophore derives from the fixation of the di-GMP to its aptamer.

In order to observe the intracellular phenomena, it is sometimes necessary to optimize the phenomena for it to be observable. “Zhu China team” is a team of students who have participated, like many other student teams, in the IGEN 2012 competition. This team, under the guidance of their tutor, describes to us a protocol allowing one to potentialize the meeting of two proteins in a cell. Lindner, Silver, and their colleagues developed this method. It concerns an RNA composed of two aptamers, which recruit two proteins and therefore force their meeting. The Zhu China team describes to us in Chapter 9 how they have made the interaction dependent of a sensor. It is the fixation of the theophylline to the sensor that renders the RNA accessible to the proteins and allows thus their coming together and their interaction.

RNA in vitro evolution or SELEX enables the artificial evolution and selection of RNA molecules that possess a desired property, such as binding affinity for a particular ligand or an activity such as that of an enzyme or catalyst. The first such selections involved isolation of various aptamers that bind to small molecules. The first catalytic RNAs produced by in vitro evolution were RNA ligases, catalytic RNAs that join two RNA fragments to produce a single adduct.

In Chapter 10, Ikawa and his colleagues describe how they have created a DSL ribozyme (designed and selected ligase). They show us how these systems allow the reduction of the number of random sequences to screen and how these scaffolds of a well-defined structure ease the screening.

The two following chapters (11 and 12) describe two methods allowing the identification of riboswitches from a random bank of aptazymes (allosteric ribozymes). These systems are logic gates; indeed the aptazymes are composed of an aptamer RNA at a strategic position with regard to the self-cleaving ribozyme so that the structure of the ribozyme is stabilized or destabilized through the link of the ligand. These aptazymes are in the 5' position of an ARNm (coding for a reporter gene), which will or will not be degraded based on the efficiency of the riboswitch. Hartig and his colleagues have thus inserted random sequences of aptazymes in the 5'-UTR of the hRluc reporter gene on the plasmid psi-CHECK2 making the expression of the luciferase dependent on the ligand. As for Yokobayashi and his colleagues, they also work in the area of mammalian cells but use the reporter system EGFP. They describe a protocol allowing the screening, at a medium rate, around a hundred aptazymes directly into mammalian cells.

As we have just seen, certain RNAs can see their activity regulated by a ligand link like a small molecule or another RNA (in the case of aptazymes). The RNA can be used in nanotechnology like “programmable” polymers capable of assembling themselves and forming topologically complex structures. Jaramillo and his colleagues describe, in Chapter 13, a method allowing one to design scaffolds from small RNAs which stack and form nano-objects. Their protocol precisely describes how to model these objects *in silico* and the ways in which they can be characterized.

The flexibility and the versatility of the RNA allow them to adopt very varied structures. Many RNA, in particular, can assemble themselves in the form of nano-structures. We understand the determinants that govern the structure and the layout of these objects better and better. In the three following chapters, protocols allowing one to design and produce different RNA nano-structures are described. In Chapter 14, Saito and his colleagues show us how to design RNA origami. They thus manage to give their RNA a triangular structure. The angles are produced by the interaction of the L7Ae protein with recognition sequences placed in the RNA.

Guo and his colleagues have been working on the pRNA of phage phi29 for several years. Through matching bases, this RNA assembles itself naturally to form nano-structures. They have succeeded in demonstrating that this object could be functionalized and could thus have a therapeutic use.

In Chapter 16, Kumar and his colleagues offer us an original method to synthesize cadmium sulfide (cds) nanoparticles. During polymerization and formation of nanoparticles, the RNA will allow one to render the polymerization directional and optimize the interactions. More precisely, they describe to us how to form nano-objects in tube form. A whole series of techniques allowing the characterization of objects are described as a supplement.

The final chapter is devoted to an RNA scaffold that serves as a genetic tool. “Gene silencing” can be induced by small noncoding antisense RNA, which hybridizes on an RNA messenger preventing the gene translation. Silencing RNA (siRNAs) are thus small coding sequences allowing, in genetics, the switching off of a gene specifically and thereby understanding its function. In Chapter 17, Lee and his collaborators describe a method to us that allows the potentialization of the silencing activity of an siRNA. The siRNA is surrounded by two structures in a stem loop, which increase the half-life of the RNA and thus increase the silencing. This structure that they have named afsRNA (artificial small regulatory RNAs) could thus become a more efficient silencing tool.

In conclusion, I would like to thank all of the authors who have allowed this book to be published. Their work shows to what extent RNAs are fascinating and interest researchers in very different areas. Naturally, we are a long way from exploring their full potential and I hope that the reading of this book will inspire its readers, especially young researchers, to further study of this area.

Paris, France

Luc Ponchon

Contents

<i>Preface</i>	<i>v</i>
<i>Contributors</i>	<i>xi</i>
1 A Method to Predict the 3D Structure of an RNA Scaffold <i>Xiaojun Xu and Shi-Jie Chen</i>	1
2 Post-crystallization Improvement of RNA Crystal Diffraction Quality <i>Jinwei Zhang and Adrian R. Ferré-D'Amaré</i>	13
3 Expression and Purification of RNA–Protein Complexes in <i>Escherichia coli</i> <i>Margueritte El Khouri, Marjorie Catala, Bili Seijo, Johana Chabal, Carine Tisné, Frédéric Dardel, and Luc Ponchon</i>	25
4 Production of Homogeneous Recombinant RNA Using a tRNA Scaffold and Hammerhead Ribozymes <i>Frank H.T. Nelissen, Hans A. Heus, and Sybren S. Wijmenga</i>	33
5 In Vivo Production of Small Recombinant RNAs Embedded in a 5S rRNA-Derived Protective Scaffold <i>Victor G. Stepanov and George E. Fox</i>	45
6 Detection of RNA–Protein Interactions Using Tethered RNA Affinity Capture <i>Hidekazu Iioka and Ian G. Macara</i>	67
7 A Universal Method for Labeling Native RNA in Live Bacterial Cells <i>Irina Smolina and Natalia Broude</i>	75
8 Live Cell Imaging Using Riboswitch-Spinach tRNA Fusions as Metabolite-Sensing Fluorescent Biosensors <i>Colleen A. Kellenberger, Zachary F. Hallberg, and Ming C. Hammond</i>	87
9 RNA Scaffold: Designed to Co-localize Enzymes <i>ZJU_China Team (iGEM 2012) and Ming Chen</i>	105
10 Artificial Ligase Ribozymes Isolated by a “Design and Selection” Strategy <i>Shigeyoshi Matsumura and Yoshiya Ikawa</i>	113
11 Engineering Aptazyme Switches for Conditional Gene Expression in Mammalian Cells Utilizing an In Vivo Screening Approach <i>Charlotte Rehm, Benedikt Klauser, and Jörg S. Hartig</i>	127
12 Aptazyme-Based Riboswitches and Logic Gates in Mammalian Cells <i>Yoko Nomura and Yohei Yokobayashi</i>	141
13 Design and Characterization of Topological Small RNAs <i>Jack Hassall, Paul MacDonald, Teresa Cordero, William Rostain, and Alfonso Jaramillo</i>	149

14	Folding RNA–Protein Complex into Designed Nanostructures	169
	<i>Tomonori Shibata, Yuki Suzuki, Hiroshi Sugiyama, Masayuki Endo, and Hirohide Saito</i>	
15	Simple Method for Constructing RNA Triangle, Square, Pentagon by Tuning Interior RNA 3WJ Angle from 60° to 90° or 108°	181
	<i>Emil F. Khisamutdinov, My Nguyen Hoan Bui, Daniel Jasinski, Zhengyi Zhao, Zheng Cui, and Peixuan Guo</i>	
16	RNA-Mediated CdS-Based Nanostructures	195
	<i>Vinit Kumar and Anil Kumar</i>	
17	An Effective Method for Specific Gene Silencing in <i>Escherichia coli</i> Using Artificial Small RNA	211
	<i>Geunu Bak, Jee Soo Choi, Wonkyeong Kim, Shinae Suk, and Youngboon Lee</i>	
	<i>Index</i>	227

Contributors

- GEUNU BAK • *Department of Chemistry, KAIST, Daejeon, Korea*
- NATALIA BROUDE • *Department of Biomedical Engineering, Boston University, Boston, MA, USA*
- MY NGUYEN HOAN BUI • *Department of Chemistry, Ball State University, Muncie, IN, USA*
- MARJORIE CATALA • *Laboratoire de Cristallographie et RMN Biologiques, UMR CNRS 8015, Faculté de Pharmacie, Université Paris Descartes, Paris, France*
- JOHANA CHABAL • *Laboratoire de Cristallographie et RMN Biologiques, UMR CNRS 8015, Faculté de Pharmacie, Université Paris Descartes, Paris, France*
- MING CHEN • *College of Life Sciences, Zhejiang University, Hangzhou, China*
- SHI-JIE CHEN • *Department of Physics, University of Missouri, Columbia, MO, USA; Department of Biochemistry, University of Missouri, Columbia, MO, USA; Informatics Institute, University of Missouri, Columbia, MO, USA*
- JEE SOO CHOI • *Department of Chemistry, KAIST, Daejeon, Korea*
- TERESA CORDERO • *School of Life Sciences, University of Warwick, Coventry, UK*
- ZHENG CUI • *Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, Lexington, KY, USA; Nanobiotechnology Center, Markey Cancer Center, University of Kentucky, Lexington, KY, USA*
- FRÉDÉRIC DARDEL • *Laboratoire de Cristallographie et RMN Biologiques, UMR CNRS 8015, Faculté de Pharmacie, Université Paris Descartes, Paris, France*
- MASAYUKI ENDO • *Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Kyoto, Japan*
- ADRIAN R. FERRÉ-D'AMARÉ • *National Heart, Lung and Blood Institute, Bethesda, MD, USA*
- GEORGE E. FOX • *Department of Biology and Biochemistry, University of Houston, Houston, TX, USA*
- PEIXUAN GUO • *Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, Lexington, KY, USA; Nanobiotechnology Center, Markey Cancer Center, University of Kentucky, Lexington, KY, USA*
- ZACHARY F. HALLBERG • *Department of Chemistry, University of California, Berkeley, Berkeley, CA, USA*
- MING C. HAMMOND • *Department of Chemistry, University of California, Berkeley, Berkeley, CA, USA; Department of Molecular & Cell Biology, University of California, Berkeley, Berkeley, CA, USA*
- JÖRG S. HARTIG • *Department of Chemistry, Konstanz Research School Chemical Biology, University of Konstanz, Konstanz, Germany*
- JACK HASSALL • *School of Life Sciences, University of Warwick, Coventry, UK*
- HANS A. HEUS • *Biophysical Chemistry, Institute for Molecules and Materials, Radboud University, Nijmegen, The Netherlands*
- HIDEKAZU IIOKA • *Department of Cell and Developmental Biology, Vanderbilt University Medical Center, Nashville, TN, USA*
- YOSHIYA IKAWA • *Graduate School of Science and Engineering, University of Toyama Gofuku, Toyama, Japan*

- ALFONSO JARAMILLO • *School of Life Sciences, University of Warwick, Coventry, UK; Institute of Systems and Synthetic Biology (iSSB), CNRS-Université d'Evry val d'Essonne, Évry, France*
- DANIEL JASINSKI • *Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, Lexington, KY, USA; Nanobiotechnology Center, Markey Cancer Center, University of Kentucky, Lexington, KY, USA*
- COLLEEN A. KELLENBERGER • *Department of Chemistry, University of California, Berkeley, Berkeley, CA, USA*
- EMIL F. KHISAMUTDINOV • *Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, Lexington, KY, USA; Nanobiotechnology Center, Markey Cancer Center, University of Kentucky, Lexington, KY, USA*
- MARGUERITTE EL KHOURI • *Laboratoire de Cristallographie et RMN Biologiques, UMR CNRS 8015, Faculté de Pharmacie, Université Paris Descartes, Paris, France*
- WONKYEONG KIM • *Department of Chemistry, KAIST, Daejeon, Korea*
- BENEDIKT KLAUSER • *Department of Chemistry, Konstanz Research School Chemical Biology, University of Konstanz, Konstanz, Germany*
- ANIL KUMAR • *Department of Chemistry, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India*
- VINIT KUMAR • *Department of Chemistry, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India*
- YOUNGHOON LEE • *Department of Chemistry, KAIST, Daejeon, Korea*
- IAN G. MACARA • *Department of Cell and Developmental Biology, Vanderbilt University Medical Center, Nashville, TN, USA*
- PAUL MACDONALD • *School of Life Sciences, University of Warwick, Coventry, UK*
- SHIGEYOSHI MATSUMURA • *Graduate School of Science and Engineering, University of Toyama Gofuku, Toyama, Japan*
- FRANK H.T. NELISSEN • *Biophysical Chemistry, Institute for Molecules and Materials, Radboud University, Nijmegen, The Netherlands*
- YOKO NOMURA • *Department of Biomedical Engineering, University of California, Davis, CA, USA*
- LUC PONCHON • *Laboratoire de Cristallographie et RMN Biologiques, UMR CNRS 8015, Faculté de Pharmacie, Université Paris Descartes, Paris, France*
- CHARLOTTE REHM • *Department of Chemistry, Konstanz Research School Chemical Biology, University of Konstanz, Konstanz, Germany*
- WILLIAM ROSTAIN • *School of Life Sciences, University of Warwick, Coventry, UK; Institute of Systems and Synthetic Biology (iSSB) CNRS-Université d'Evry val d'Essonne, Évry, France*
- HIROHIDE SAITO • *Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, Japan*
- BILI SEIJO • *Laboratoire de Cristallographie et RMN Biologiques, UMR CNRS 8015, Faculté de Pharmacie, Université Paris Descartes, Paris, France*
- TOMONORI SHIBATA • *Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, Japan*
- IRINA SMOLINA • *Department of Biomedical Engineering, Boston University, Boston, MA, USA*
- VICTOR G. STEPANOV • *Department of Biology and Biochemistry, University of Houston, Houston, TX, USA*

- HIROSHI SUGIYAMA • *Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto, Japan*
- SHINAE SUK • *Department of Chemistry, KAIST, Daejeon, Korea*
- YUKI SUZUKI • *Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto, Japan*
- CARINE TISNÉ • *Laboratoire de Cristallographie et RMN Biologiques, UMR CNRS 8015, Faculté de Pharmacie, Université Paris Descartes, Paris, France*
- SYBREN S. WIJMEGA • *Biophysical Chemistry, Institute for Molecules and Materials, Radboud University, Nijmegen, The Netherlands*
- XIAOJUN XU • *Department of Physics, University of Missouri, Columbia, MO, USA; Department of Biochemistry, University of Missouri, Columbia, MO, USA; Informatics Institute, University of Missouri, Columbia, MO, USA*
- YOHEI YOKOBAYASHI • *Department of Biomedical Engineering, University of California, Davis, CA, USA*
- JINWEI ZHANG • *National Heart, Lung and Blood Institute, Bethesda, MD, USA*
- ZJU_CHINA TEAM (iGEM 2012) (TIANQI CHEN, XINYI GUO, XIANG LAN, XIN LI, HUACHAN LIU, XIAO LIU, JIAJUN LV, JIAHUI ZHANG) • *College of Life Sciences, Zhejiang University, Hangzhou, China*
- ZHENGYI ZHAO • *Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, Lexington, KY, USA; Nanobiotechnology Center, Markey Cancer Center, University of Kentucky, Lexington, KY, USA*

Chapter 1

A Method to Predict the 3D Structure of an RNA Scaffold

Xiaojun Xu and Shi-Jie Chen

Abstract

The ever increasing discoveries of noncoding RNA functions draw a strong demand for RNA structure determination from the sequence. In recently years, computational studies for RNA structures, at both the two-dimensional and the three-dimensional levels, led to several highly promising new developments. In this chapter, we describe a recently developed RNA structure prediction method based on the virtual bond-based coarse-grained folding model (Vfold). The main emphasis in the Vfold method is placed on the loop entropy calculations, the treatment of noncanonical (mismatch) interactions and the 3D structure assembly from motif-based template library. As case studies, we use the glycine riboswitch and the G310-U376 domain of MLV RNA to illustrate the Vfold-based prediction of RNA 3D structures from the sequences.

Key words Partition function, Loop entropy, Mismatched stacks, 2D structure motif, Structure assembly

1 Introduction

To perform crucial cellular functions, RNA molecules fold up to form compact three-dimensional (3D) structures [1–5]. The RNA structure determination by experiments alone cannot keep up the pace with the ever increasing number of RNA sequences and new functions. The gap between the number of known RNA 3D structures and the number of biologically significant RNA sequences underscores more than ever the request for accurate computational models for RNA structure prediction.

An RNA structure can be described at the two-dimensional (2D) and three-dimensional (3D) levels. A 2D structure is defined as the sum of all the base-base pairs in the structure, including long-range base pairs in tertiary folds. Computational prediction of RNA 2D structures falls into two categories [6–10]: sequence comparison (alignment) analysis and free energy-based modeling. In general, sequence comparison-based methods can give more reliable predictions than free energy-based methods, but it depends on the availability of homologous sequences and often cannot directly provide information about the alternative structures.

For the free energy-based modeling, a key problem is to determine the helix stabilities and loop free energies. The free energy parameters for a helix stem can be calculated from the Turner experimental data [11], but the loop free energy requires a model.

The recently developed Vfold model is a statistical mechanics-based RNA folding model. The model relies on a coarse-grained (virtual bond) representation of RNA structures [12–14]. Compared with other free energy-based RNA 2D structure prediction models, such as Mfold [15] and RNAstructure [16], the Vfold model computes loop entropy parameters from explicit conformational sampling. Furthermore, by enumerating all the possible (sequence-dependent) intra-loop mismatches, the Vfold model partially accounts for the sequence-dependence of the loop free energy. Through application to a broad range of experimental and biological problems, the Vfold-based predictions have shown to be able to provide novel insights for RNA mechanisms, such as pseudoknot-involved conformational switch between bistable secondary structures [17], microRNA–gene target interactions [18], and RNA–RNA kissing dimerization in viral replication [19, 20].

Knowing RNA 2D structures alone is often not sufficient to understand RNA function. We also need RNA 3D structure information in order to understand the interactions between RNA and other molecules and RNA functions [21–24]. One way to predict RNA 3D structure is to combine a coarse-grained RNA structure model with the knowledge-based force field and fold the RNA through discrete molecular dynamics (DMD) simulations [25–28]. Due to the limitation of conformational sampling, this method would be most suitable for short RNAs or large RNAs with auxiliary constraints from experimental data. Based on the assumption that 3D structure is more conserved and can be recognized by the alignment of sequences and structure motifs, (3D structure) template-based modeling has become a promising method in RNA 3D structure predictions [29–31]. The template-based methods build RNA 3D structures using known structures ranging from fragments of 1–3 nucleotides to larger structural motifs. One of the common limitations for the template (structure assembly) approaches is the completeness of the fragment library. The lack of reliable structural motifs for many loops and junctions greatly hampers the success of accurate 3D structure prediction.

For a given 2D structure, the Vfold-based 3D structure prediction method searches for the appropriate template for each loop/junction in the structure, and assembles the 3D template structures into a scaffold for further structure refinement. In comparison with other template-based (structure assembly) methods such as FARNAL/FARFAR [29] and MC-Sym [31], which sample structures from small fragments of known RNA structures, the Vfold-based method uses motif-based instead of fragment-based templates.

2 Algorithms

2.1 RNA Motif-Based Loop Entropy

Using two virtual bonds per nucleotide to represent the backbone conformation, the Vfold model samples fluctuations of loops/junction conformations in 3D space through conformational enumeration (*see* Fig. 1). By calculating the probability of loop formation, it gives the conformational entropy parameters for the formation of the different types of loops such as hairpin, bulge, internal, pseudoknot loops. The model has the advantage of accounting for chain connectivity, exclude volume and the completeness of conformational ensemble.

1. Enumerate all the possible virtual bond backbone conformations for a given chain length (*see* **Note 1**) and count the total number Ω_{coil} of the conformations.
2. From the conformational ensemble above, identify the loop conformations according to the loop closure condition. For example, for hairpin loops, the two ends of a loop conformation should be fitted to an A-form base pair. Count the total number Ω_{loop} of loop conformations.
3. Calculate the loop entropy $\Delta S_{\text{loop}} = k_B \ln(\Omega_{\text{loop}} / \Omega_{\text{coil}})$. Here, k_B is the Boltzmann constant.

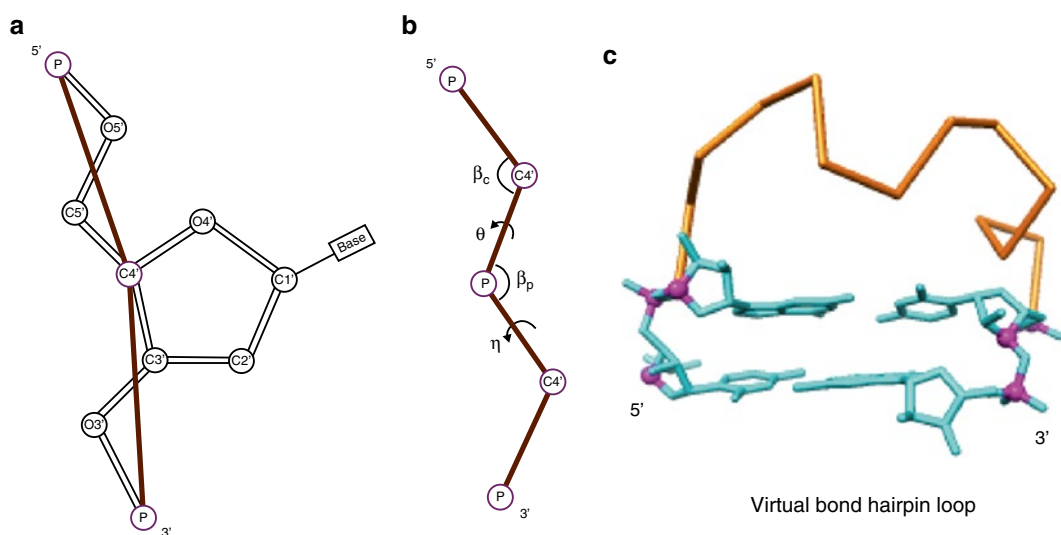


Fig. 1 Vfold computes loop entropies by sampling virtual bond conformations in 3D space. **(a)** Virtual bond representation: two bonds (P-C4' and C4'-P) per nucleotide. **(b)** The bond angles (β_c , β_p) and the torsional angles (θ , η) for the virtual bonds. Vfold enumerates RNA backbone conformations on a diamond lattice with bond length of 3.9 Å, bond angle of $\sim 109.5^\circ$ and three equiprobable torsional angles (60° , 180° , 300°). **(c)** A virtual bond backbone conformation of a hairpin loop, with two ends fitted to the base pair structure in an A-form helix

Table 1
RNA motif-based template library

Motif name	Number of templates
Hairpin loops	2,366
Internal/bulge loops	3,260
3-way junctions	820
4-way junctions	506
5-way junctions	222
6-way junctions	49
7-way junctions	61
H-type pseudoknots	56

4. Vfold computations lead to pre-tabulated entropy parameters for hairpin loops [12], internal/bulge loops [12], H-type pseudoknots with/without inter-helix junction [32, 33] and hairpin-hairpin kissing motifs [19].

2.2 RNA Motif-Based Template Library

The (3D structure) template library was built from 2,621 PDB structures (*see Note 2*), including RNA-involved complexes. It contains 3D templates for hairpin loops, internal/bulge loops, H-type pseudoknots, and multibranch junctions.

1. For a given RNA 3D structure, extract the A-form helices. From the information of helices and base pairs, the corresponding 2D structure is determined.
2. Identify all the non-helix 2D structure motifs for the given 3D structure.
3. Remove the redundant templates for those with root mean square deviation (RMSD) ≤ 1.5 Å for the same motif, same size, and identical sequence.
4. Collect all the nonredundant motif structures to construct a template library. Table 1 shows the statistics for the current template library.

3 Methods

To predict RNA 3D structures, Vfold first predicts the 2D structures from the sequence. Using the 2D structures as constraint, the model then predicts the corresponding 3D structures.

The key of the free energy-based RNA 2D structure prediction is the enthalpy and entropy parameters used to evaluate the stability of sampled structures. The enthalpy and entropy for the canonical and mismatched base stacks are calculated from Turner’s experimental data. The loop entropies are from the Vfold pre-tabulated parameters.

$$\Delta G_{\text{loop}} = -k_{\text{B}}T \ln Q_{\text{loop}}, \quad Q_{\text{loop}} = \sum_{\text{mismatches}} e^{-(\Delta G_{\text{mm}} - T\Delta S_{\text{loop}})/k_{\text{B}}T}$$

Fig. 2 Ensemble of a 9-nt hairpin loop closed by an A–U base pair, containing five different arrangements of mismatched base stacks within the loop

Here ΔG_{mm} is total free energies of the mismatched base stacks and ΔS_{loop} is the loop entropy for the given intra-loop mismatch constraints.

4. Assign the free energy for each sampled 2D structure:
 $\Delta G_s = \Delta G_{\text{helix}} + \Delta G_{\text{loop}}$.
5. Calculate the total partition function as the sum over all the possible (2D) structures:

$$Q_{\text{tot}} = \sum_{\text{structures}} e^{-\Delta G_s / k_B T}$$

6. Following the similar procedure as above for the total partition function, compute the conditional partition function Q_{ij} for all the 2D structures with nucleotides i and j base paired.
7. Calculate the probability of forming the (i, j) base pair:
 $p_{ij} = Q_{ij} / Q_{\text{tot}}$.
8. From the base pairing probability for all the possible (i, j) pairs, extract the predicted most probable (*see Note 7*) as well as alternative structures.

We use the glycine riboswitch (PDB: 3owi) as an example to show how Vfold predicts the 2D structure. Given the 84-nt RNA sequence of the glycine riboswitch (*see Note 8*), Vfold calculates the base pairing probabilities p_{ij} for all the possible base pairs. The predicted most probable 2D structure (*see Note 7*) can be predicted from p_{ij} , as shown in Fig. 3. It should be noted that depending on the sequence, the Vfold model predicts all the stable structures, including the most probable (most stable) structure as well as the alternative (metastable) structures. Therefore, it is recommended to also find out the possible alternative structures from the base pairing probabilities.

3.2 RNA 3D Structure Prediction for a Given 2D Structure

The Vfold model predicts the 3D structure from a 2D structure by assembling motif-specific structural templates. Currently, due to the limited structural template database, Vfold can only predict the 3D structures with hairpin loops, internal/bulge loops, multi-branched junctions and pseudoknots.

1. Identify the structure motifs (such as hairpin loop, internal loop, pseudoknot loop, and three-way junction) from the given 2D structure.
2. Build the virtual bond 3D structure for helices according to the A-form helix template.
3. For each non-helix motif, search for the best templates from the template library. The search criteria are based on the size (first) and sequence (second) matches (*see Note 9*).
4. From the (all-atom) templates found in the previous step (*see Notes 10 and 11*), build the virtual bond 3D structures of each motifs.