WILEY VCH

Edited by
John Schneekloth and Martin Pettersson

RNA as a Drug Target

The Next Frontier for Medicinal Chemistry

Volume 84
Series Editors:
R. Mannhold,
H. Buschmann, J. Holenz

Methods and Principles in Medicinal Chemistry

RNA as a Drug Target

Methods and Principles in Medicinal Chemistry

Edited by
Raimund Mannhold, Helmut Buschmann, Jörg Holenz

Previous Volumes of the Series

Daina, A., Przewosny, M., Zoete, V. (Eds.)

Open Access Databases and Datasets for Drug Discovery

2023
ISBN: 978-3-527-34839-8
Vol. 83

Bachhav, Y. (Ed.)

Targeted Drug Delivery

2023
ISBN: 978-3-527-34781-0
Vol. 82

Alza, E. (Ed.)

Flow and Microreactor Technology in Medicinal Chemistry

2022
ISBN: 978-3-527-34689-9
Vol. 81

Rübsamen-Schaeff, H., and Buschmann, H.(Eds.)

New Drug Development for Known and Emerging Viruses

2022
ISBN: 978-3-527-34337-9
Vol. 80

Gruss, M. (Ed.)

Solid State Development and Processing of Pharmaceutical Molecules Salts, Cocrystals, and Polymorphism

2021
ISBN: 978-3-527-34635-6
Vol. 79

Plowright, A.T. (Ed.)

Target Discovery and Validation Methods and Strategies for Drug Discovery

2020
ISBN: 978-3-527-34529-8
Vol. 78

Swinney, D., Pollastri, M. (Eds.)

Neglected Tropical Diseases Drug Discovery and Development

2019
ISBN: 978-3-527-34304-1
Vol. 77

Bachhav, Y. (Ed.)

Innovative Dosage Forms Design and Development at Early Stage

2019
ISBN: 978-3-527-34396-6
Vol. 76

Gervasio, F. L., Spiwok, V. (Eds.)

Biomolecular Simulations in Structure-based Drug Discovery

2018
ISBN: 978-3-527-34265-5
Vol. 75

Sippl, W., Jung, M. (Eds.)

Epigenetic Drug Discovery

2018
ISBN: 978-3-527-34314-0
Vol. 74

RNA as a Drug Target

The Next Frontier for Medicinal Chemistry

Edited by
John Schneekloth and Martin Pettersson

Series Editors
Raimund Mannhold
Helmut Buschmann
Jörg Holenz

WILEY VCH

Volume Editors
Dr. John Schneekloth, Jr.
National Cancer Institute
Building 538, Room 249
Frederick
MD 21702-1201, USA

Dr. Martin Pettersson
Promedigen
Daejeon
South Korea

Series Editors
Prof. Dr. Raimund Mannhold
Rosenweg 7 40489
Düsseldorf
Germany

Dr. Helmut Buschmann
Sperberweg 15 52076
Aachen
Germany

Dr. Jörg Holenz
BIAL – PORTELA & CA.,
S.A. São Mamede Coronado
Portugal

Cover Image: © Oleg Nebesnyy/ Shutterstock, Courtesy of John Schneekloth

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.d-nb.de>.

Print ISBN: 978-3-527-35100-8
ePDF ISBN: 978-3-527-84043-4
ePub ISBN: 978-3-527-84044-1
oBook ISBN: 978-3-527-84045-8

Typesetting Straive, Chennai, India

Contents

Series Editors' Preface *xiii*
Preface *xv*

1 Introduction *1*
John Schneekloth Jr. and Martin Pettersson
References *4*

2 RNA Structure Probing, Dynamics, and Folding *7*
Danny Incarnato
2.1 Introduction *7*
2.1.1 Relevance of RNA Structure in Disease *8*
2.1.2 Challenges in Studying RNA Structures *8*
2.2 Experimentally Guided RNA Structure Modeling *9*
2.2.1 Structural Interrogation of RNA Nucleotides via Chemical Probing *10*
2.2.1.1 Limits of RNA Chemical Probing *12*
2.2.2 Direct Mapping of RNA–RNA Interactions *14*
2.2.2.1 Limits of RNA–RNA Interaction Mapping *16*
2.2.3 Mapping Spatially Proximal Nucleotides in RNA molecules *17*
2.2.3.1 Limits of Methods for Spatial Proximity Mapping *17*
2.3 Dealing with RNA Structure Heterogeneity *19*
2.4 Querying RNA–Small Molecule Interactions with Chemical Probing *22*
2.5 Conclusions and Future Prospects *22*
References *23*

3 High-Resolution Structures of RNA *29*
Lukas Braun, Zahra Alirezaeizanjani, Roberta Tesch, and Hamed Kooshapur
3.1 Introduction *29*
3.2 X-Ray Crystallography *31*
3.3 NMR Spectroscopy *34*
3.4 Cryo-EM *37*
3.5 3D Structure Prediction and Integrative Approaches *39*
3.6 Conclusions *43*
Acknowledgments *43*
Conflicts of Interest *43*
References *43*

4 Screening and Lead Generation Techniques for RNA Binders *49*
Gary Frey, Emily Garcia Sega, and Neil Lajkiewicz
4.1 Knowledge-Based Versus Agnostic Screening *49*
4.2 Virtual Screening *50*
4.3 Screening Methods *51*
4.3.1 High-Throughput Screening (HTS) *51*
4.3.1.1 Mass Spectrometry *51*
4.3.1.2 HTS of RNA Using Direct MS Approaches *52*
4.3.1.3 HTS of RNA Using Indirect MS Approaches *54*
4.3.1.4 DNA-Encoded Libraries (DELs) *56*
4.3.1.5 Microarray Screening *57*
4.3.1.6 Fragment-Based Drug Discovery *58*
4.3.1.7 Phage Display *63*
4.3.2 Orthogonal Methods *63*
4.3.2.1 Surface Plasmon Resonance *63*
4.3.2.2 Fluorescence-Based Assays *66*
4.3.2.3 Microscale Thermophoresis (MST) *70*
4.3.2.4 Isothermal Titration Calorimetry (ITC) *70*
4.4 Binding Site Identification/Target Engagement *72*
4.4.1 Covalent Methods *72*
4.4.2 Competition with an Antisense Oligonucleotide (ASO) *74*
4.5 Defining SAR and Functional Assays *75*
4.5.1 Functional Assays *75*
4.5.2 Phenotypic Screens *76*
4.6 Identifying a Lead Series *76*
4.6.1 Hit Optimization *77*
4.6.2 Risdiplam Hit-to-Lead *78*
4.6.3 Branaplam Lead Generation *79*
4.6.4 Zotatifin Lead Generation *80*
4.7 Concluding Thoughts and Outlook *80*
Acknowledgments *81*
References *81*

5 Chemical Matter That Binds RNA *93*
Emily G. Swanson Hay, Zhengguo Cai, and Amanda E. Hargrove
5.1 Introduction *93*
5.2 Natural Ligands *94*
5.2.1 Aminoglycosides *94*
5.2.2 Tetracyclines *95*
5.2.3 Macrolides *96*
5.2.4 Native Riboswitch Ligands *96*
5.3 Commercial Ligands *97*
5.3.1 Industrial Libraries *98*
5.3.2 Academic Libraries *98*

5.4 Synthetic Ligands *99*
5.4.1 Benzimidazoles and Purines *100*
5.4.2 Naphthalenes, Quinolines, and Quinazolines *101*
5.4.3 Oxazolidinones *102*
5.4.4 Amilorides *102*
5.4.5 Diphenyl Furan *103*
5.4.6 Multivalent Ligands *103*
5.5 Computational Tools for the Exploration of Chemical Space *103*
5.5.1 Similarity Searches and Principal Component Analysis *104*
5.5.2 Additional Machine-Learning Tools *105*
5.5.3 Structure-Based Ligand Design *106*
5.6 Case Studies in Examining and Expanding RNA-Targeted Chemical Space *106*
5.6.1 Using QSAR to Probe RNA-Targeting Small-Molecule Properties *107*
5.6.2 Evaluating the Chemical Space of Natural, Synthetic, and Commercial Ligands *108*
5.7 Conclusions and Outlook *111*
Acknowledgments *111*
References *111*

6 MicroRNAs as Targets for Small-Molecule Binders *119*
Maria Duca
6.1 Introduction *119*
6.2 MicroRNAs *121*
6.3 MicroRNAs Biogenesis *122*
6.4 Targeting MicroRNAs with Small-Molecule RNA Binders *123*
6.4.1 Induction of miRNAs Expression: Tackling the Decrease of Tumor Suppressor miRNAs *124*
6.4.2 Inhibition of miRNAs Production: Pre- and Pri-miRNA Binders *125*
6.4.2.1 Discovery of miRNAs Inhibitors by Intracellular Assays *125*
6.4.2.2 Target-Based *In Vitro* Assays *127*
6.4.2.3 Design of Specific Ligands of Pre- and Pri-miRNAs *131*
6.4.2.4 Fragment-Based Drug Design *138*
6.4.2.5 DNA-Encoded Libraries (DELs) *139*
6.5 Inhibition of RNA–Protein Interactions in miRNAs Pathways *140*
6.6 Adding Cleavage Properties to miRNAs Interfering Agents *142*
6.7 Conclusions *144*
References *144*

7 Pre-mRNA Splicing Modulation *151*
Scott J. Barraza and Matthew G. Woll
7.1 Introduction *151*
7.2 Overview of Splicing Biology *152*
7.2.1 The Spliceosome *152*
7.2.2 Classes of Alternative Splicing *154*

7.3 Pharmacological Mechanisms of Splicing Modulation *155*
7.3.1 *Cis*- and *Trans*-Regulatory Elements (Splicing Factors) *155*
7.3.1.1 Stabilization of *Cis*-Regulatory Elements *156*
7.3.1.2 Destabilization of *Cis*-Regulatory Elements *158*
7.3.1.3 Inhibition of *Cis*-Regulatory RNA–Protein Interactions *158*
7.3.1.4 Inhibition of *Trans*-Regulatory Elements *160*
7.3.1.5 Degradation of *Trans*-Regulatory Elements *161*
7.3.1.6 Inhibition of *Trans*-Regulatory Element Protein–Protein Interactions (PPIs) *162*
7.3.1.7 Stabilization of *Trans*-Regulatory Element RNA–Protein Interactions (RPIs) *165*
7.3.2 Kinases and Phosphatases *165*
7.3.2.1 Challenges in Targeting Kinases *167*
7.3.2.2 Inhibition of Kinases *168*
7.3.2.3 Activation and Degradation of Kinases *168*
7.3.2.4 Inhibition and Activation of Protein Phosphatases *169*
7.3.3 Epigenetic Writers and Erasers *172*
7.3.3.1 Inhibition of Epigenetic Writers *172*
7.3.4 RNA Helicases *174*
7.3.5 Drugging the Spliceosome *175*
7.3.5.1 Inhibition of U2 snRNP Recognition of the 3′-Splice Site *176*
7.3.5.2 E7107 *176*
7.3.5.3 H3B-8800 *177*
7.3.5.4 Stabilizers of U1 snRNP Recognition of the 5′-Splice Site *180*
7.3.5.5 Introduction to Spinal Muscular Atrophy (SMA) *180*
7.3.5.6 Risdiplam (Evrysdi®) *183*
7.4 Future Outlook *186*
References *188*

8 Prospects for Riboswitches in Drug Development *203*
Michael G. Mohsen and Ronald R. Breaker
8.1 Introduction *203*
8.1.1 The Known Landscape of Riboswitches *203*
8.1.2 Riboswitches in Drug Development *203*
8.1.3 The Need for Novel Antibiotics *205*
8.2 Riboswitches as Drug Targets *207*
8.2.1 Why Target Riboswitches? *207*
8.2.2 Features of a Druggable Riboswitch *208*
8.2.3 Riboswitch-Targeted Drugs *208*
8.2.3.1 Small Molecules Targeting FMN Riboswitches *208*
8.2.3.2 Other Riboswitches Targeted in Proof-of-Principle Demonstrations *209*
8.2.4 Barriers and Future Developments *210*
8.3 Riboswitches as Tools for Antibiotic Drug Development *210*
8.3.1 Riboswitches as Biosensors *210*
8.3.2 A Riboswitch-Based Fluoride Sensor Illuminates Agonists of Fluoride Toxicity *211*

8.3.3 A Riboswitch-Based ZTP Sensor Identifies Inhibitors of Folate Biosynthesis *211*
8.3.4 A Riboswitch-Based SAH Sensor Reveals an Inhibitor of SAH Nucleosidase *212*
8.3.5 Barriers and Future Developments *213*
8.4 Application of Riboswitches in Gene Therapy *213*
8.4.1 Considerations for Designer Riboswitches *213*
8.4.2 Eukaryotic Expression Platforms *214*
8.4.3 Barriers and Future Developments *216*
8.5 Concluding Remarks *217*
Acknowledgment *218*
References *218*

9 Small Molecules That Degrade RNA *227*
Noah A. Springer, Samantha M. Meyer, Amirhossein Taghavi, Jessica L. Childs-Disney, and Matthew D. Disney
9.1 Antisense Oligonucleotide Degraders *227*
9.2 Small-Molecule Direct Degraders *228*
9.2.1 *N*-Hydroxypyridine-2(1*H*)-thione (*N*-HPT) Conjugates *229*
9.2.2 Bleomycin *229*
9.2.3 Bleomycin Conjugates *231*
9.2.3.1 Bleomycin Degraders Targeting the r(CUG) Repeat Expansion That Causes DM1 *231*
9.2.3.2 Bleomycin Degraders Targeting r(CCUG) Repeat Expansion that Causes DM2 *233*
9.2.3.3 Bleomycin Degraders Targeting Oncogenic Precursor microRNAs *233*
9.2.3.4 Conclusions and Outlook for Bleomycin-Based Direct Degraders *234*
9.3 Ribonuclease Targeting Chimeras (RiboTACs) *235*
9.3.1 RNase L is an Endogenous Endoribonuclease That Functions as Part of the Innate Immune Response *236*
9.3.2 First-Generation RiboTACs Targeting Oncogenic miRNAs *236*
9.3.3 Small-Molecule-Based RiboTACs *239*
9.3.4 Comparison of Bleomycin-Based Direct Degraders and RiboTACs *242*
9.3.5 Discovery of Additional Small-Molecule RNase L Activators *242*
9.3.6 Conclusions and Outlook for RiboTACs *243*
9.4 Summary and Outlook for Small-Molecule RNA Degraders *244*
References *246*

10 Approaches to the Identification of Molecules Altering Programmed Ribosomal Frameshifting in Viruses *253*
Elinore A. VanGraafeiland, Diego M. Arévalo, and Benjamin L. Miller
10.1 Introduction *253*
10.2 Mechanisms of Frameshifting *256*
10.3 Targeting Frameshifting in HIV *257*
10.4 Targeting Frameshifting in SARS-CoV-1 and SARS-CoV-2 *263*
10.5 Conclusions *274*
References *274*

11 RNA–Protein Interactions: A New Approach for Drugging RNA Biology *281*
Dalia M. Soueid and Amanda L. Garner
11.1 Molecular Basis of RNA–Protein Interactions *282*
11.1.1 RNA Recognition Motifs (RRMs) *282*
11.1.2 Double-Stranded RNA-Binding Domains (dsRBD) *286*
11.1.3 Zinc Finger (ZnF) Domains *287*
11.1.4 K Homology (KH) Domains *289*
11.1.5 Other RBDs *290*
11.2 Regulation and Dysregulation of RNA–Protein Interactions *290*
11.2.1 Poor Quality Control Leads to Over- and Underproduction of RBPs *292*
11.2.2 RBPs Become Out of Control, mRNA Processing Gets a Makeover (and Hates It) *294*
11.2.3 RBP Shuttling of mRNA Becomes Askew *294*
11.2.4 The RBP is Lost and Wreaks Havoc on the Cell *295*
11.2.5 RBPs Dictate Which mRNAs are Translated, Favoring their Toxic Friends *295*
11.2.6 RBPs and RNA Become Very Clique-y, Form Their Own Complex and Cause Stress to the Rest of the Cell *296*
11.3 Experimental Methods to Detect and Screen for Small Molecules that Modulate RNA–Protein Interactions *297*
11.3.1 *In vitro* Fluorescence-Based Assays *297*
11.3.2 *In vitro* Chemiluminescence-Based Assays *297*
11.3.2.1 Cell-Based RPI Detection Assays *300*
11.3.3 Cell-Based RNA–Protein Interaction Screening *301*
11.4 Closing Remarks *302*
References *303*

12 Drugging the Epitranscriptome *321*
Tanner W. Eggert and Ralph E. Kleiner
12.1 Introduction *321*
12.2 Modifications on mRNA: N^6-Methyladenosine, Pseudouridine, and Inosine *325*
12.2.1 N^6-Methyladenosine (m^6A) *325*
12.2.2 Pseudouridine (Ψ) *327*
12.2.3 Inosine (I) *328*
12.3 Modifications on tRNA and rRNA *330*
12.3.1 tRNA Modifications *330*
12.3.2 rRNA Modifications *334*
12.4 Concluding Remarks *335*
References *336*

13 Outlook *355*
Christopher R. Fullenkamp, Xiao Liang, Martin Pettersson, and John Schneekloth Jr.
13.1 Introduction *355*
13.2 Target Selection: Identification of the Most Promising RNA Intervention Points *357*
13.3 Development of Robust Biophysical Methods, Alternative Strategies for Target Engagement, and Accurate and Reliable Functional Models *358*
13.3.1 Biophysical Methods for Interrogating Small Molecule–RNA Interactions *358*
13.3.2 Cellular Target Engagement Methods *360*
13.3.3 Unique Challenges Faced in the Development of Functional Assays for Studying Small Molecule–RNA Interactions *364*
13.4 Acquisition of High-Resolution RNA and RNA–Ligand Structures is Needed to Enable the Development and Validation of Computational Tools for RNA–Small Molecule Therapeutic Discovery *367*
13.4.1 RNA Structure Prediction *367*
13.4.2 Computational Tools for Hit Optimization *369*
13.4.3 Implementation of Molecular Dynamics Simulations, Machine Learning, and AI Tools to Interrogate RNA–Small Molecule Interactions *371*
13.5 Deposition of Small Molecule–RNA Interaction Data with Rigorous Experimental Protocols and Controls is Needed *373*
13.6 Outlook: The Future of Small Molecule-Based RNA Therapeutics is Bright *375*
References *376*

Index *385*

Series Editors' Preface

As one of the leading book series in the field, the ambition and tradition of Methods and Principles in Medicinal Chemistry has been to cover the most innovative, cutting-edge areas within the field first, and in comprehensive, hands-on format enabling the reader not only to learn deeply the field but also to practise it successfully.

The interface between small molecules and ribonucleic acid (RNA) is one of those areas that has recently gained enormous attention and has lifted medicinal chemistry to the level being *"en par"* with gene therapy or antisense approaches while maintaining the convenience and safety of oral dosing and its associated transient and reversible pharmacodynamic effects.

Prompted by the success story of Risdiplam, a pre-mRNA splicing modulator small molecule drug and a break-through medication for spinal muscular atrophy (SMA), academic as well as pharmaceutical industry research has lifted off to thoroughly explore and understand the field and to deliver new, innovative medicines to patients in difficult to drug areas of disease.

John Schneekloth, Jr., and Martin Pettersson have diligently searched for the most distinguished experts in the field to assemble the team of authors sharing their insights, while giving practical advice as well. The book follows the natural logic of the drug discovery and development path, starting from understanding the structure of RNAs and its potential interaction sites for small molecules, explaining strategies for lead generation, the features of chemical matter with increased chances of binding to RNA, and in the second part tackles the fundamentally different strategies for small molecules to interact with RNA to elicit the desired pharmacological effects. Finally, an outlook is given to capture recent trends where the field will evolve to, as well as estimate and appreciate its future impact and potential within the modalities a medicinal chemist has to interact with disease pathology.

This book is a concise, up to date, must read for every medicinal chemist! It will also give important scientific and strategic insights to all scientists, as well as managers, working in this field!

The senior editors of Methods and Principles of Medicinal Chemistry would like to express their gratitude to Martin and Jay, as well as the entire author team, for delivering what we believe will be the new reference standard for *Small Molecules targeting RNA*.

Jörg Holenz
Christa Müller
Helmut Buschmann

Preface

Once thought of as a transient messenger that did little more than carry a sequence for proteins, RNA is now recognized as a key regulator of biological homeostasis, gene expression, metabolism, and disease. While the sequencing of the human genome heralded the modern era of genomics with the promise of a drug for every gene, in practice this has been much harder to realize than originally thought. Recognition that the majority of the genome is noncoding and the majority of the proteome is at best challenging to target with small molecules has led to the perspective that RNA and RNA regulation are important new frontiers for medicinal chemistry. A key motivation for considering RNA as a drug target is the potential to develop novel small-molecule therapeutics for diseases with no treatment. This could be realized in a variety of ways, for example, by targeting the RNA that encodes for proteins that are challenging to drug (some ~85% of the proteome), targeting noncoding RNA, or pharmacologically controlling RNA-related processes like pre-mRNA splicing or posttranscriptional modifications. Additionally, broad classes of diseases, including cancers, neurodegenerative diseases, viral infections, and bacterial infections, are thought to have potential relevance in this space.

While the potential upside of targeting RNA is considerable, there are fundamental ways in which RNA differs from other targets. A broad variety of biophysical and biochemical techniques are both widely available and routinely used for protein targets, far fewer have been successfully applied to RNA. Thus, new approaches are needed to accurately identify, develop, characterize, and understand RNA-binding small molecules, and more work is needed to develop better assays that are compatible with the stringent requirements and rigor needed for drug discovery programs. Additionally, the question of what makes a good RNA target for small molecules remains an area of open debate. Still, advances in other areas, including genomics technologies such as next-generation sequencing, have allowed progress to be made at a remarkable pace, enabling the characterization and annotation of increasingly large collections of RNA structures and sequences.

We are pleased to present this exciting collection of chapters from leading experts in the field. The perspectives included here include a range of topics relevant to RNA-targeted drug discovery and medicinal chemistry. We hope this superb collection is both informative and reflective of the state of the art and brings diverse insights as to how scientists in academia, government, and industry around the

world think about this important problem. Finally, we would like to extend a heartfelt thank you to all the authors and contributors for providing their important perspectives on this book. It has been an honor to work with you all, and it is a testament to the strong state of the field to have so many talented individuals willing to work on this project. Finally, we are indebted to the exceptional team at Wiley for their patience and help in guiding us through the process of editing this book. We are extremely grateful for your efforts!

John Schneekloth, Jr. (Jay)
USA

Martin Pettersson
South Korea

1

Introduction

John Schneekloth Jr.[1] *and Martin Pettersson*[2]

[1] *National Cancer Institute, Chemical Biology Laboratory, Frederick, MD 21702-1201, USA*
[2] *Promedigen, Daejeon, 34050, Republic of Korea*

The sequencing of the human genome more than 20 years ago revealed that it encodes just ~20,000 proteins, comprising only about three percent of its ~3 billion bases [1]. However, we now know that more than 80% of the genome is transcribed into RNA at some point [2]. In addition to RNAs that encode polypeptide sequences and facilitate protein biogenesis, diverse classes of non-coding RNA such as microRNAs, long non-coding RNAs (lncRNAs), small nucleolar RNAs (snoRNAs), and others continue to be discovered and characterized [3]. Numerous genome-wide association studies and other broad genomics efforts have pointed toward non-coding RNAs as relevant to diverse diseases including neurodegenerative conditions, infectious diseases, and cancer [3]. Furthermore, recent estimates indicate that only ~15% of the human proteome has been successfully targeted with small molecules, which is partly due to the lack of suitable binding pockets for small molecules on many proteins [4]. Challenges with targeting a large part of the human proteome, coupled with widespread non-coding functions of RNA in both humans and pathogenic organisms, point toward RNA and related regulatory processes as intriguing alternative drug targets for novel therapeutics.

A key promise of developing small molecules that target RNA or regulate RNA function is that, by doing so, one could potentially pharmacologically control the function of genes or signaling pathways that are challenging to modulate at the protein level [5–9]. However, understanding RNA as a small-molecule drug target comes with a number of challenges. Many RNAs are structurally dynamic, more closely resembling intrinsically disordered proteins than the ordered protein domains that are commonly associated with "traditional" drug targets [10]. Limitations in biophysical techniques to characterize RNA–ligand complexes and structural ensembles at atomic resolution complicate efforts to rationally design new potential drug molecules. In addition, the majority of RNAs are bound to and interact with RNA-binding proteins, often in a time- or stimulus-dependent fashion [11]. The RNA biopolymer consists of just four bases in contrast to the 20 canonical amino acids found in protein sequences. In addition, the chemical

RNA as a Drug Target: The Next Frontier for Medicinal Chemistry, First Edition.
Edited by John Schneekloth and Martin Pettersson.

nature of the anionic phosphodiester backbone of RNA contrasts starkly with that of the polypeptide backbone of proteins, resulting in considerably different biophysical properties of the two polymers. Still, fundamental studies aimed at understanding the targetability of RNA indicate that it is capable of folding into complex, three-dimensional structures with hydrophobic pockets that are likely suitable for small-molecule binding [7, 12]. Moreover, naturally occurring aptamers (as seen in riboswitches) and lab-evolved aptamers provide validation that RNA can interact with and sense low-molecular-weight species ranging from metal cations or halide ions to drug-like small molecules and even complex metabolites like cobalamin (vitamin B12) with exquisite selectivity [13].

Indeed, the history of drug discovery has already demonstrated the potential and significant impact that RNA-targeting medicines can have. Ribosome-targeting antibiotics, exemplified by macrolides, aminoglycosides, and tetracyclines, have been known since at least the 1940s and represent the largest class of clinically used drugs to treat infections [14]. In general, these compounds make specific contacts with ribosomal RNA (rRNA) within the large ribonucleoprotein complex of the ribosome and modulate protein synthesis. The development of these compounds has had a remarkable impact on human health. However, despite the massive clinical success of such antibiotics, efforts to develop compounds that target specific transcripts apart from the ribosome itself have proven challenging. Nevertheless, the approval of risdiplam in 2020 for the treatment of spinal muscular atrophy demonstrates that small molecules that target RNA can indeed be developed as powerful and mechanistically novel therapeutics [15].

One important factor that represents a barrier to developing RNA-targeted small molecule therapeutics is selectivity. Antisense oligonucleotides and related sequence-based probes such as peptide nucleic acids (PNAs) and locked nucleic acids (LNAs) are now routinely used as tools and provide high sequence specificity [16]. Antisense molecules have also been approved as therapeutics; however, challenges with delivery, biodistribution, and cell permeability continue to be barriers impacting the broader application and development of such molecules as therapeutics. As a complementary modality, small molecules remain highly attractive as drugs due to their ability to passively diffuse across cell membranes, achieve high metabolically stability and oral bioavailability, as well as the potential for penetration into the central nervous system (CNS). From a patient perspective, these characteristics can translate into significant benefits such as the convenience of once-a-day oral administration. The widespread recognition that RNA can directly drive disease, coupled with the potential to modulate it with small-molecule drugs, has driven significant interest in this emerging field.

This book represents a collection of perspectives from leading experts on the state of the art in understanding RNA as a target for small molecules. The chapters included represent a broad overview, covering topics ranging from how to think about and understand RNA structure, how to identify and understand druglike small molecules that bind to RNA, various approaches for controlling RNA function with small molecules, to overviews of RNA–protein interactions and

post-transcriptional modifications of RNA. Finally, we provide an outlook chapter that adds perspective to the future of RNA-targeted drug discovery, including specific challenges that need to be overcome to develop RNA-targeted therapeutics.

In Chapter 2, Incarnato provides a discussion on how to think about RNA structure at the level of individual base pairs. This chapter covers methods and applications for utilizing structure probing to characterize individual RNAs (including both well-defined structures and conformationally diverse ensembles). In Chapter 3, Braun et al. report on the history and progress in determining atomic resolution structures of RNA. This space has historically been highly challenging due to the flexible nature of the RNA biopolymer. However, advances in cryoelectron microscopy and techniques in X-Ray crystallography have resulted in exciting progress in recent years. The increasing availability of atomic resolution structures of RNA ligand complexes is bound to significantly enhance the ability to effectively design high-quality small molecules that bind RNA selectively.

Chapter 4 provides a discussion on lead generation techniques for identifying small molecules that bind to RNA. These methods, including both target-based and phenotypic screens, represent diverse approaches to discover starting points for identifying lead structures for medicinal chemistry programs. In many cases, lead generation techniques used for identifying compounds that target proteins can be adapted for RNA targets, and the authors discuss specific challenges and considerations when applying these methods to RNA. This chapter also discuss several RNA-specific approaches for establishing cellular target engagement. In Chapter 5, Hay et al. discuss the types of chemical matter that binds to RNA, how to characterize it, and how it differs from (and is similar to) protein binding small molecules. Here, it is becoming clear that diverse chemotypes can interact with RNA, both within traditional druglike chemical space and beyond.

The next series of chapters describe examples of different classes of RNA targets. In Chapter 6, Duca provides an overview of microRNAs and efforts to target them. These small RNAs are an intriguing class of non-coding RNA targets that have received considerable interest from both industrial and academic groups as they play a key role in regulating mRNA levels and as a result, control expression levels of proteins. In Chapter 7, Barraza et al. describe efforts to modulate pre-mRNA splicing with small molecules. This important process is the target of clinically validated therapies for spinal muscular atrophy and holds promise for a variety of other mechanistically novel therapies for other diseases including Huntington's diseases and various cancers. While early work has focused on rare monogenic diseases, there is considerable promise in the development of splice modulators as therapeutics for a variety of diseases. Chapter 8, by Mohsen et al. describes riboswitches, which are naturally occurring, ligand-responsive RNA aptamers. These intriguing structures have been the subject of considerable study in both structural/biochemical contexts as well as drug discovery – primarily through the lens of identifying novel antibiotics. However, they also provide promise in synthetic biology as well, particularly in the exciting area of gene therapy as a potential on/off switch for gene expression.

In Chapter 9, Meyer et al. describe examples of small molecules that degrade RNA. By a mechanism analogous to targeted protein degradation strategies such as PROTACs, these molecules recruit nucleases to target RNAs for cleavage and degradation. While some RNA targeting molecules are now FDA approved, RNA-targeting chimeras (RiboTACs) are still at an earlier stage of development. However, RiboTACs may offer specific advantages over monofunctional ligands that target RNA and present opportunities to harness induced proximity pharmacology in the RNA target space and also function in a sub-stoichiometric matter.

In Chapter 10, VanGraafeiland et al. discuss how modulation of programmed frameshifting, primarily by targeting sequences in viral genomes, is an intriguing therapeutic strategy with significant potential. With the development of several potent molecules, cellular proof of concept has now been established for this unique mechanism of controlling gene expression with small molecules. In Chapter 11, Soueid et al. broadly discuss RNA–protein interactions as targets for therapeutics. RNA-binding proteins represent nearly 10% of the human genome and regulate a wide range of cellular and disease relevant processes. Here, understanding structure and patterns of recognition is key to both unraveling biology and developing therapeutics. This chapter also describes the attributes and limitations of a several assay formats that have been developed for identifying compounds that modulate RNA–protein interactions. Chapter 12, by Eggert et al. covers epitranscriptomics. In this context, epitranscriptomics is defined as chemical modifications of RNA, and how these modifications alter biological processes. Although a wide variety of RNA modifications are known, only a few have been studied extensively. Nevertheless, this is an active area of research where small molecules have already entered clinical development.

Finally, in Chapter 13, we provide an outlook on the field and discuss some of the challenges and opportunities facing the development of RNA-targeting molecules as therapeutics. Continued efforts to develop the fundamentals of medicinal chemistry and molecular design principles within this challenging space, coupled with the importance of target validation, establishing relevant functional assays, and understanding selectivity are key topics. The latter stands to benefit tremendously as more and more atomic resolution structures of RNA–small molecules complexes become availed. While many challenges remain, the potential for RNA-targeting medicines to make a broader impact on human health stands as a compelling rationale for continued efforts. Together, we hope that these chapters provide an exciting perspective on how to think about and prosecute RNA as targets for small molecules.

References

1 Lander, E.S., Linton, L.M., Birren, B. et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.

2 Consortium, E.P (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74.

3 Iyer, M.K., Niknafs, Y.S., Malik, R. et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* 47: 199–208.

4 Spradlin, J.N., Zhang, E., and Nomura, D.K. (2021). Reimagining druggability using chemoproteomic platforms. *Acc. Chem. Res.* 54: 1801–1813.

5 Kovachka, S., Panosetti, M., Grimaldi, B. et al. (2024). Small molecule approaches to targeting RNA. *Nat. Rev. Chem.* 8: 120–135.

6 Childs-Disney, J.L., Yang, X., Gibaut, Q.M.R. et al. (2022). Targeting RNA structures with small molecules. *Nat. Rev. Drug Discovery* 21: 736–762.

7 Warner, K.D., Hajdin, C.E., and Weeks, K.M. (2018). Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discovery* 17: 547–558.

8 Connelly, C.M., Moon, M.H., and Schneekloth, J.S. Jr., (2016). The emerging role of RNA as a therapeutic target for small molecules. *Cell Chem. Biol.* 23: 1077–1090.

9 Zafferani, M. and Hargrove, A.E. (2021). Small molecule targeting of biologically relevant RNA tertiary and quaternary structures. *Cell Chem. Biol.* 28: 594–609.

10 Ganser, L.R., Kelly, M.L., Herschlag, D., and Al-Hashimi, H.M. (2019). The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* 20: 474–489.

11 Gebauer, F., Schwarzl, T., Valcarcel, J., and Hentze, M.W. (2021). RNA-binding proteins in human genetic disease. *Nat. Rev. Genet.* 22: 185–198.

12 Hewitt, W.M., Calabrese, D.R., and Schneekloth, J.S. Jr., (2019). Evidence for ligandable sites in structured RNA throughout the Protein Data Bank. *Bioorg. Med. Chem.* 27: 2253–2260.

13 Kavita, K. and Breaker, R.R. (2023). Discovering riboswitches: the past and the future. *Trends Biochem. Sci.* 48: 119–141.

14 Wilson, D.N. (2014). Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.* 12: 35–48.

15 Dhillon, S. (2020). Risdiplam: first approval. *Drugs* 80: 1853–1858.

16 Crooke, S.T., Baker, B.F., Crooke, R.M., and Liang, X.H. (2021). Antisense technology: an overview and prospectus. *Nat. Rev. Drug Discovery* 20: 427–453.

2

RNA Structure Probing, Dynamics, and Folding

Danny Incarnato

Department of Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Groningen, The Netherlands

2.1 Introduction

RNA is pivotal for the regulation of most biological processes in the cell. Although for most of the past century researchers have mainly regarded RNA as a carrier of genetic information, a simple intermediate between DNA and proteins, recent advances in the fields of genomics and epigenomics, aided by the development of high-throughput sequencing methods, have begun shedding new light on the polyhedric nature of RNA molecules.

Messenger RNAs (mRNAs), which encode for proteins, only represent a small piece in the highly complex puzzle that is the transcriptome of a cell. The advent of RNA sequencing (RNA-seq) technologies and the dawn of the ENCODE project [1] revealed that the vast majority of the nonrepetitive genome in mammalian cells is transcribed and that most of these transcripts lack the ability to encode for proteins. The human genome is now estimated to carry over 60,000 genes, generating approximately 250,000 transcripts, of which nearly two-thirds are noncoding RNAs (ncRNAs) [2].

While the cellular roles and mechanisms of action of most of these transcripts are still largely unknown, their ability to fold into stable secondary and tertiary structures is expected to be paramount to their functions, as reported for many cellular RNAs [3]. Indeed, being a single-stranded molecule, RNA has the unique ability to fold back on itself to generate very intricate secondary structures, mediated both by direct base pairing between complementary nucleobases within the RNA strand and by base stacking. Secondary structures can then be further compacted into tertiary structures thanks to the contribution of different factors such as divalent cations, non-Watson-Crick interactions, and macromolecular crowding, among others [4]. These structures have been proven essential to mediate many of the noncoding functions of RNA molecules, including macromolecular scaffolding, sensing of the environment, and even catalysis. The ribosome is a key example of these functions as proper folding of ribosomal RNAs is crucial to provide the docking platform for ribosomal proteins and to enable efficient translation of mRNAs [5].

RNA as a Drug Target: The Next Frontier for Medicinal Chemistry, First Edition.
Edited by John Schneekloth and Martin Pettersson.

Aside from ncRNAs, mRNAs themselves carry untranslated regions (UTRs) that are populated by a myriad of regulatory RNA structure elements such as G-quadruplexes (G4), internal ribosome entry sites (IRES), iron response elements (IRE), and many more, which mediate the post-transcriptional regulation of gene expression [6].

In light of these considerations, it will become immediately evident to the reader that a complete understanding of the transcriptome is deeply entangled with our ability to interrogate and understand RNA structures in the cellular context. This chapter discusses recent advances in methods for RNA structure determination in living cells, the challenges, and the future perspectives.

2.1.1 Relevance of RNA Structure in Disease

Besides controlling gene expression in a physiological context, the deregulation of RNA structure has been widely implicated in pathogenesis. This might either involve the disruption of physiologically relevant RNA structures or the genesis of pathogenic RNA structures that are not normally present under physiological conditions. Disease-associated mutations identified within transcribed noncoding regions from genome-wide association studies (GWAS), for example, have been shown to disrupt key regulatory RNA structure elements within UTRs [7]. Mutations in the 5′UTR of the ferritin light chain (FTL) gene, associated with hyperferritinemia cataract syndrome, lead to disruption of an IRE element, hence impairing binding by an iron response protein (IRP) and, consequently, deregulating FTL expression. The expansion of CTG repeats in the 3'UTR of the dystrophin myotonin protein kinase (DMPK) gene, which drives the pathogenesis of myotonic dystrophy (DM), leads to the generation of an extended stem-loop structure [8, 9] that is capable of sequestering muscleblind proteins (MBNL) [10, 11], ultimately deregulating splicing of multiple mRNAs [12].

Aside from pathogenic RNA structures within human transcripts, studies conducted on RNA viruses highlighted how their genomes are replete with regulatory RNA structure elements that are pivotal in regulating multiple aspects of the viral life cycle, including replication, translation, and packaging [13]. Translation of the two partially overlapping ORF1a and ORF1b in coronavirus genomes, for example, is regulated by the frameshifting element (FSE), which enables pausing and backtracking of the ribosome at the end of ORF1a, leading to the synthesis of the ORF1ab polyprotein, crucial for viral replication [14]. The stem-loop A (SLA) RNA structure located at the 5′ end of flavivirus genomes acts as a promoter for the replication of the viral genome by providing the docking site for the viral RNA polymerase NS5 [15]. The Rev Response Element (RRE) of HIV-1, which interacts with the Rev protein, can switch between two alternative structures to regulate the nuclear export of the viral genome [16].

2.1.2 Challenges in Studying RNA Structures

Owing to their extreme flexibility, RNA molecules can adopt a wide variety of different conformations. An *n*-nucleotides long RNA can theoretically fold into up to

1.8^n distinct secondary structures [17]. This scenario is further complicated when considering that each nucleotide in the RNA chain has eight degrees of freedom and that the biologically active structures might not be strongly favored thermodynamically over competing structures [18]. This dilemma is commonly known in the field as the *RNA folding problem* [19].

At present, for most RNAs, it is nearly impossible to accurately predict their structure solely from their primary sequence. Computational approaches based on free energy minimization have been developed to predict the minimum free energy (MFE) structure, that is the thermodynamically most stable structure, using a set of thermodynamic parameters known as nearest-neighbor model (or *Turner* rules) [20]. The accuracy of MFE structure prediction is, however, limited in practice, with only ∼66% of predicted base pairs found in experimentally validated structures from crystallographic studies or phylogenetic analyses [21].

One of the issues inherent to MFE-based approaches is that the MFE structure only represents one of the many conformations a given RNA can adopt. Every RNA can populate a multitude of structural *states*, each one having a certain probability of occurring, which are collectively referred to as *ensemble*. The probability that the MFE structure might occur within the ensemble might be very low, meaning that the MFE structure might not represent the most likely conformation for a given RNA. To address this limitation, methods to determine the partition function for secondary structure formation were developed, which allow predicting the probability that each possible base pair in the RNA strand might form within the ensemble. The accuracy of this type of approach is much higher, with ∼91% of the base pairs occurring with a probability ≥0.99 in the MFE structure found in phylogenetically determined structures [21].

More recently, a new class of predictive algorithms based on machine learning has emerged. These algorithms are trained on sets of experimentally validated RNA structures [22]. However, extensive benchmarking of these algorithms suggests that their accuracy is generally biased toward certain classes of RNAs [23] and that, typically, predicted structures tend to contain more base pairs than ground truth structures [24], a trend that grows quadratically with increasing RNA lengths.

2.2 Experimentally Guided RNA Structure Modeling

Irrespective of the chosen approach, the reliability of computational RNA structure predictions is limited, and it decreases with increasing length of the RNA strand. To address these limitations, a wide variety of complementary approaches have been developed to date to experimentally interrogate the structure of RNA molecules. Data generated by these approaches can be then incorporated as *constraints* (or *restraints*) into thermodynamics-driven structure predictions to increase their accuracy [25, 26].

This chapter will solely focus on approaches suitable for the analysis of RNA structures in living cells. Particularly, I will discuss three classes of methods: chemical

probing-based methods, methods for the direct mapping of RNA–RNA interactions, and methods to map nucleotide spatial proximity.

2.2.1 Structural Interrogation of RNA Nucleotides via Chemical Probing

Among the repertoire of available methods for RNA structural interrogation, chemical probing is perhaps the most popular. Methods based on chemical probing exploit specific reagents (hereafter referred to as *chemical probes*) that can react with specific functional groups in RNA nucleotides on the basis of their structural context.

Particularly, three broad classes of chemical probes can be identified. The first class of reagents includes nucleobase-specific chemical probes to measure the base-pairing status of RNA bases (Figure 2.1). These compounds act by chemically modifying functional groups involved in base pairing. One of the most popular probes in this class is dimethyl sulfate (DMS), an alkylating reagent that methylates the N1 of adenines and the N3 of cytosines when unpaired (Figure 2.1a) [27]. DMS has been widely employed to query the structure of both viral genomes [28–30] and entire cellular transcriptomes [31–34], owing both to its ability to readily permeate biological membranes and to its high efficiency, resulting in a very high signal-to-noise ratio. More recently, it has been suggested that, under mildly alkaline conditions, DMS might partially react also with unpaired guanines and uracils, although with much lower efficiency [35]. α-Ketoaldehydes such as glyoxal and N_3-kethoxal have been recently reported to form adducts with the N1 and N2 of unpaired guanines *in vivo* (Figure 2.1b) [36, 37]. Similarly, carbodiimides such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) are capable of forming adducts with the N1 of guanines and the N3 of uracils, both unpaired or in G:U wobbled base pairs (Figure 2.1c) [38, 39].

The second class of reagents includes selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) probes (Figure 2.1d) [40]. SHAPE reagents are electrophilic compounds capable of forming adducts with the 2′-hydroxyl (2′-OH) of the ribose moiety at structurally flexible nucleotides. These nucleotides typically lie within loops of unpaired bases, or at helix termini. A wide variety of SHAPE probes, largely differing in their hydrolysis rates in aqueous buffers, have been reported to date. Among these, acyl imidazoles such as 2-methylnicotinic acid imidazolide (NAI) and 2-aminopyridine-3-carboxylic acid imidazolide (2A3) [41, 42] and isatoic anhydrides such as 5-nitroisatoic anhydride (5NIA) [43] are the better-characterized SHAPE probes for the interrogation of RNA structures in living cells.

The third class of reagents includes compounds capable of querying the accessibility of RNA bases to the solvent. The recently characterized nicotinoyl azide (NAz), for example, forms adducts with the C8 of solvent-accessible purines upon irradiation with long-wavelength UV light (Figure 2.1e) [44]. A popular alternative to NAz involves the use of hydroxyl radicals to induce RNA strand scission at the level of solvent-accessible C3 and C4 of the ribose moiety [45]. This approach is

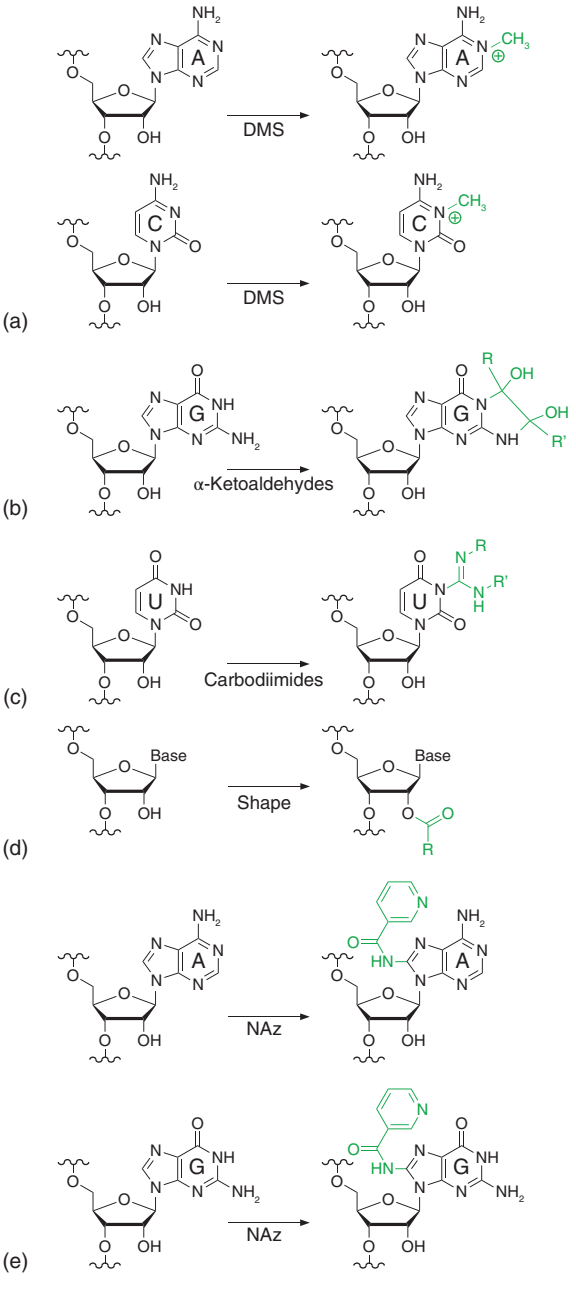

Figure 2.1 Reaction of different RNA chemical probes. (a) Reaction of dimethyl sulfate (DMS) with unpaired A and C bases. (b) Reaction of α-ketoaldehydes (such as glyoxal and N_3-kethoxal) with unpaired G bases. (c) Reaction of carbodiimides (such as EDC) with unpaired U bases. (d) Reaction of SHAPE reagents (such as 5NIA, NAI, and 2A3) with the 2′-OH of structurally flexible nucleotides. (e) Reaction of nicotinoyl azide (NAz) with the C-8 of solvent exposed A and G bases.

typically more challenging than NAz probing when it comes to interrogating RNA structures in living cells as it requires generating intracellular hydroxyl radicals, for example, using a high-flux synchrotron X-ray beam [46].

Irrespective of the probe choice, sites of modification on the RNA can be read out by reverse transcription experiments [47]. These experiments typically rely on the

inability of the reverse transcriptase (RT) to read through probe-induced modifications, resulting in RT drop-off and premature termination of cDNA synthesis, one nucleotide downstream of the modification site. Low-throughput chemical probing experiments historically involved the interrogation of a single RNA at a time and its reverse transcription using gene-specific primers, followed by resolution of cDNAs by gel electrophoresis. Thanks to the advent of high-throughput sequencing technologies, numerous protocols have been devised to extend this procedure to the study of entire cellular transcriptomes [47]. These protocols typically involve the fragmentation of cellular RNAs, followed by reverse transcription and direct ligation of a sequencing adapter to the 3′ end of the cDNAs, allowing the high-throughput detection of RT drop-off events on a transcriptome-wide scale (Figure 2.2).

More recently, an alternative approach dubbed mutational profiling (MaP) has been introduced [32, 48, 49]. This involves adjusting the reverse transcription reaction to favor the RT read-through on the site of probe-induced modification. When reading through the modified residue, the RT incorporates a random nucleotide, typically resulting in a mutation in the synthesized cDNA. The main difference between RT drop-off and MaP methods is the ability of MaP to capture multiple sites of modification within each cDNA product. The advantages of this approach will be further discussed later on in the chapter (see Section 2.3, "Dealing with RNA structure heterogeneity").

Following the sequencing of cDNAs and mapping to the reference transcriptome, RT drop-off events (or mutations) at each position of every RNA can be counted, yielding *reactivity profiles* that can be used to constrain RNA structure prediction. Highly reactive bases are interpreted as having a high likelihood of being unpaired (or structurally flexible or accessible to the solvent). Conversely, bases with low reactivity are interpreted as having a high likelihood of being base paired (or structurally rigid or inaccessible to the solvent). Reactivity profiles can then be fed into thermodynamics-driven RNA structure prediction algorithms to adjust the thermodynamic parameters by rewarding base pairing of lowly reactive positions and penalizing base pairing of highly reactive positions [25, 26].

2.2.1.1 Limits of RNA Chemical Probing

The biggest limitation of chemical probing-based methods is the fact that lack of reactivity to the chemical probe is not necessarily correlated to the structural state of the target nucleotide. For instance, protein binding, or how deeply buried a nucleotide is within the tertiary structure of the RNA, might significantly affect the accessibility of the probe to the target nucleotide. Furthermore, systematic investigation of the structural preferences of SHAPE reagents suggests that small loops are typically more reactive to SHAPE probing as compared to longer stretches of unpaired nucleotides [50]. Interestingly, a recent study introduced a number of novel acyl imidazoles, showing that by altering the reagent scaffold, it is possible to alter the structural specificity of the SHAPE probe, for example, to reduce reactivity on loop-closing base pairs [51]. Moreover, the smallest possible scaffold, acetylimidazole, is cell permeable and capable of probing RNA structures even at the level of protein-bound regions.