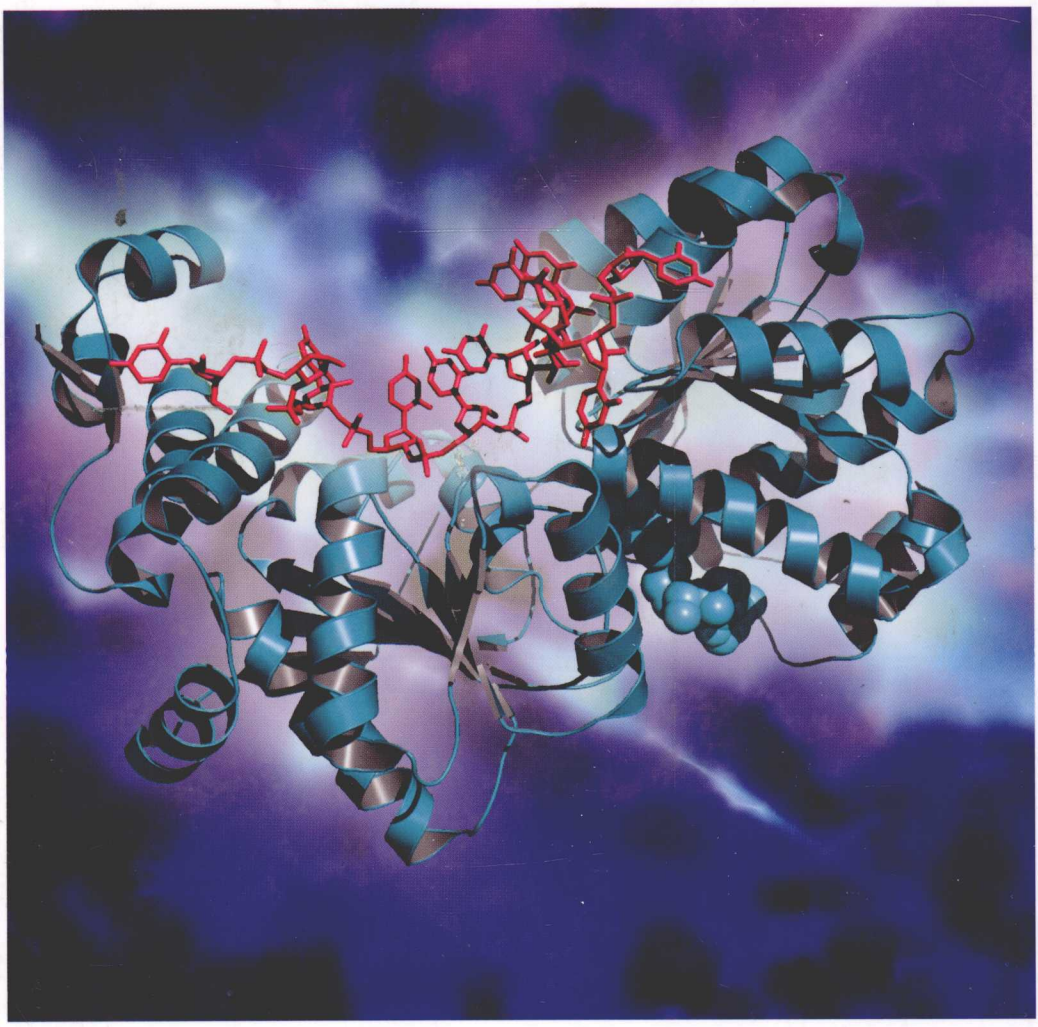
RSC Biomolecular Sciences

Edited by Eckhard Jankowsky

RNA Helicases



RSC Publishing

RNA Helicases

Edited by

Eckhard Jankowsky
Case Western Reserve University, Cleveland, OH, USA

RSC Publishing

RSC Biomolecular Sciences No. 19

ISBN: 978-1-84755-914-2
ISSN: 1757-7152

A catalogue record for this book is available from the British Library

Published by The Royal Society of Chemistry,
Thomas Graham House, Science Park, Milton Road,
Cambridge CB4 0WF, UK

Registered Charity Number 207890

For further information see our web site at www.rsc.org

RNA Helicases

RSC Biomolecular Sciences

Titles in the Series:

1: Biophysical and Structural Aspects of Bioenergetics
2: Exploiting Chemical Diversity for Drug Discovery
3: Structure-based Drug Discovery: An Overview
4: Structural Biology of Membrane Proteins
5: Protein–Carbohydrate Interactions in Infectious Disease
6: Sequence-specific DNA Binding Agents
7: Quadruplex Nucleic Acids
8: Computational and Structural Approaches to Drug Discovery: Ligand–Protein Interactions
9: Metabolomics, Metabonomics and Metabolite Profiling
10: Ribozymes and RNA Catalysis
11: Protein-Nucleic Acid Interactions: Structural Biology
12: Therapeutic Oligonucleotides
13: Protein Folding, Misfolding and Aggregation: Classical Themes and Novel Approaches
14: Nucleic Acid-Metal Ion Interactions
15: Oxidative Folding of Peptides and Proteins
16: RNA Polymerases as Molecular Motors
17: Quantum Tunnelling in Enzyme-Catalysed Reactions
18: Natural Product Chemistry for Drug Discovery
19: RNA Helicases

How to obtain future titles on publication:
A standing order plan is available for this series. A standing order will bring delivery of each new volume immediately on publication.

For further information please contact:
Book Sales Department, Royal Society of Chemistry,
Thomas Graham House, Science Park, Milton Road, Cambridge,
CB4 0WF, UK
Telephone: +44 (0)1223 420066, Fax: +44 (0)1223 420247, Email: books@rsc.org
Visit our website at http://www.rsc.org/Shop/Books/

FOREWORD

Found in Translation – The Discovery of the First RNA Helicase, eIF4A

WILLIAM C. MERRICK

Department of Biochemistry, School of Medicine, Case Western Reserve University, Cleveland, Ohio, 44106-4935, USA

More than 25 years have passed since the first RNA helicase was identified. When we were starting to look at this protein – the eukaryotic initiation factor 4A (eIF4A) – as an ATP-driven RNA unwinder – an RNA helicase – we did not imagine that RNA helicase enzymes would be among the largest enzyme classes. Nor did we anticipate how widespread and essential these proteins would be for RNA metabolism. But we are certainly pleased that the significance of RNA helicases is becoming ever more obvious, as gene regulation at the RNA level is beginning to enjoy more of the limelight usually reserved for DNA-related processes.

This book, the first volume specifically dedicated to RNA helicases, provides an impressive testimony to how far we have come from the discovery of eIF4A. Detailed mechanistic studies on several enzymes have been reported, several dozen crystal structures of RNA helicases have been solved, and a rapidly increasing body of data describes their cellular functions. Yet, despite the impressive journey from the humble beginnings of the eIF4A story to this book, many RNA helicases still cling to their secrets for specificity and mechanism of action. In this sense, these enzymes hold as much sway today as they did right after we "stumbled" across eIF4A.

RSC Biomolecular Sciences No. 19
RNA Helicases
Edited by Eckhard Jankowsky

Published by the Royal Society of Chemistry, www.rsc.org

At the end of the 1970s, we and several other laboratories were trying to reconstitute eukaryotic protein synthesis *in vitro*. Early attempts were based upon what had already been learned from the bacterial system. Although many thought the eukaryotic process might be more complicated, it had some of the same characteristics. For example, the poly(U) assay in eukaryotes showed the same requirements as in the bacterial system: ribosomes, poly(U), Phe-tRNA, GTP, eEF1A and eEF2, and Met-tRNA$_i$ was the initiator tRNA that used AUG as the start codon, just as in bacteria.[1,2] When polypeptide synthesis was measured on more complex RNAs, a number of components needed to be added: amino acids, tRNA, mRNA, ATP, GTP, ribosomes, aminoacyl-tRNA synthetases, and protein factors for initiation and elongation. However, there was a specific requirement for ATP, seen when model 80S initiation complexes were formed. The first report of a requirement for ATP to form initiation complexes with mRNA was from Giesen *et al.* using partially purified proteins from wheat germ[3] and this was soon followed by independent reports using purified initiation factors from rabbit reticulocytes.[4,5] Why ATP was needed was not clear.

Throughout the 1970s, several groups (Anderson, Gupta, Hershey, Staehelin and Voorma) went in search of all of the "factors" responsible for translation initiation. By the end of the 1970s, most groups had exchanged proteins and were able to agree on the number of individual factors and to some extent, their rough biological function. But it remained unclear which protein was responsible for mRNA binding to the ribosome. At the time, the main contenders were eIF3 and eIF4B, both incorrect, as we now know.

In the early 1980s, research began to focus on the activity of the individual translation initiation factors required for the utilisation of globin mRNA, and with it the discovery of an additional factor eIF4F.[6] The isolation of this protein was accomplished by Jamie Grifo, a graduate student in my laboratory, with the help of Drs. Stanley Tahara and Aaron Shatkin (Roche Institute for Molecular Biology, Nutley, N. J.). At the time we were attempting to separate what turned out to be eIF4B and eIF4F. Tahara and Shatkin were interested in the component (present in the high salt wash of pelleted ribosomes) that would allow a polio-virus-infected cell lysate to translate normal mRNAs. Thus, the assay for eIF4F was two-fold, one to stimulate protein synthesis with globin mRNA in a fractionated system and one to be the component that would restore the ability of polio-virus-infected cell lysates to translate normal m^{7}G-capped mRNAs. This required Jamie to begin the purification and at each step, make two aliquots from each of the column fractions. One aliquot was assayed for protein synthesis activity by Jamie. The second aliquot was sent by overnight delivery on dry ice to Stan to assay for restoring activity. The active fractions were pooled and subjected to another round of column chromatography followed by the same two assays until a purified protein was obtained that was active in both assays. As this required three separate column steps, it took about a week to complete the purification to the point that eIF4F was resolved from eIF4B (and all other initiation factors). Needless to say, for each week that this was done, Jamie and Stan got very little sleep.

Subsequent biochemical studies identified the key components for binding mRNAs as eIF4A, eIF4B and eIF4F.[7] eIF4A was a single polypeptide, eIF4B appeared to be a homodimer and eIF4F was composed of three subunits, currently named eIF4A, eIF4E and eIF4G. While all of these factors bound RNA, the interaction of eIF4A with RNA required ATP. In addition, eIF4A had another intriguing ability: it hydrolysed ATP in the presence of RNA. To convince skeptical peers of an RNA-dependent ATPase activity, it was necessary to obtain eIF4A free of other ATPases. The original eIF4A data were 8200 cpm PO_4 released from ATP in the presence of RNA and 7000 cpm release in the absence of RNA (poly(U)). As Jamie would say "But, it's reproducible!" and as I would say "The reviewers aren't going to buy it!" Due to the high background, an ATP affinity column was used in a subsequent step and then the numbers became 38 pmol of PO_4 released per mg eIF4A per 10 min in the presence of RNA and 2 pmol in the absence of RNA. This, I believed, the reviewers would accept. Although it took almost three years until our findings were independently confirmed, it was certainly satisfying to see eIF4A's RNA-dependent ATPase activity later confirmed many times over.

But why did eIF4A hydrolyse ATP in an RNA-dependent fashion? Given that the interaction of eIF4A with RNA appeared to be ATP-dependent, there were two obvious possibilities: eIF4A might be the engine for the ATP-dependent process of scanning,[8,9] or the ATP might be used as the driving force for the unwinding of duplex RNA anticipated to be in stem-loop structures in the 5′ UTR of eukaryotic mRNAs. This idea was inspired by analogy to the better studied ATP-dependent unwinding of DNA by helicases during replication. The latter option was the first to be tested, primarily by Dr. Bimal Ray in Dr. Robert Thach's laboratory (Washington University, St. Louis, MO), where it was shown that eIF4A could unwind duplexed RNA, although eIF4F was much more efficient than eIF4A on a molar basis.[10,11]

A simpler assay for RNA unwinding was then developed using self-complementary RNA transcripts.[12] Rozen *et al.* reported that the RNA-binding protein eIF4B was required to observe RNA helicase activity of eIF4A. They also made the remarkable observation that duplexes with either 5′ or 3′ singled-stranded regions were separated.[12] Thus, eIF4A (or eIF4F) was a "bidirectional" helicase, something never reported previously for any other helicase. This finding prompted my laboratory to later more carefully examine the requirements for RNA duplex unwinding using RNA duplexes of different stabilities, and, most importantly, duplexes shorter than those used in previous studies.[13,14] Our reasoning was if eIF4A catalysed RNA-dependent ATP hydrolysis, it might have been possible that previous difficulties to observe duplex unwinding with eIF4A alone reflected the stability of the substrate duplex. Using less-stable RNA duplexes, the results indicated that both eIF4A and eIF4F could effectively unwind RNA duplexes. eIF4B would stimulate the unwinding, and this was especially obvious with more-stable duplexes. We speculated at the time that these results suggested that eIF4A might affect only a single productive unwinding step (*i.e.* one ATP hydrolysis event) during which only a limited number of base-pairs would be

separated. Indeed, subsequent work has provided solid experimental proof for this scenario.

At present, the major question about eIF4A is to what extent does it function independently of the eIF4F complex (eIF4A + eIF4E + eIF4G)? It is found in ribosomal salt washes to be about 5 to 10 times the amount of either eIF4B or eIF4F (personal observation), but any eIF4A requirement can be satisfied by increased concentrations of eIF4F. Secondly, although toe printing experiments are consistent with either eIF4A or eIF4F acting as the motor for scanning, there has yet to be a definitive experiment comparing one against the other in this process. Thirdly, it is not clear whether the eIF4A/eIF4F requirement is greater to strip proteins from the 5′ end of mRNAs or to unwind secondary structure at the 5′ end of the mRNA (and this may depend on the mRNA). Thus, while we have a great deal of information on the behaviour of eIF4A in isolation, its importance in the overall process of initiation in the presence of a dozen or so other initiation factors is much less clear.

At the time when the helicase activity of eIF4A was just beginning to be explored, cloning was starting to generate protein-sequence information at a tremendous rate. By the late 1980s one could begin to search the database for a variety of similarities. In my laboratory, using the sequences of proteins known to bind GTP, we were able to map out several motifs (with spacing) that fit GTP binding proteins into two categories.[15] In a similar manner, Linder and colleagues made the observation that a number of proteins (nine used in the publication, four of which were eIF4A sequences) that displayed RNA-dependent ATPase activity and/or helicase activity with RNA duplexes shared conserved motifs in their amino acid sequence.[16] The validity of these motifs as being essential for eIF4A function was shown in a series of studies by Pause and colleagues.[17–19] Given the protein sequences available and the conserved motifs, Patrick Linder suggested the name "DEAD-box" for this family of RNA helicases due to the highly conserved sequence of aspartic acid, glutamic acid, alanine, aspartic acid. Analogous conserved motifs in other RNA helicase families reveal a series of nonwords and certainly nothing as catchy as "DEAD-box". Nonetheless, we now know that there are several different RNA helicase families with sometimes surprisingly distinct features.

This book describes these different RNA helicase families and thus shows how far the simple observation of an ATP-dependent interaction of eIF4A with RNA has gone. The group of proteins that unwind RNA duplexes has grown to a tremendous degree and along with this the pathways in RNA metabolism in which these enzymes function. Biochemical and structural work has illuminated our molecular understanding of how RNA helicases unwind RNA duplexes. Although a number of central questions regarding the function of RNA helicases remain to be answered, this unique volume is a great, and as much as this can ever be accomplished, comprehensive reference on our current knowledge of RNA helicases.

On a personal note, I wish to mention that my ability to make any headway has often been through the aid of others. Originally, I got started when Dr.

French Anderson allowed me to take the protein synthesis project he started from NIH to Case Western Reserve University in 1978. Subsequently, collaborative interactions with Drs. Aaron Shatkin, Robert Thach and Nahum Sonenberg and their colleagues have added an extraordinary dimension to the types of studies I have been associated with. And finally, I wish to thank the students and postdoctorals that have worked with me because they did all of the bench work. I'd like to say that they entered my laboratory as bums and that I turned them into successful scientists. In reality, they were all very bright young people and I did my best to not hold them back. The major players in this work were – in order of appearance – former students Dr. Jamie Grifo, Dr. Richard Abramson, Dr. Thomas Dever and Dr. George Rogers, Jr. Additionally, the other students and postdoctorals in my laboratory have contributed through discussions, alternate interpretations, proofreading of papers or galleys, and most importantly, when it came time to yet again have to purify the proteins. In addition, I have benefited from the generally supportive environment that is the field of translation where many other noted scientists have helped with their discussions, insights and on more than one occasion, useful reagents. It's been fun to be in translation and it has been great prying the secrets of RNA helicases from eIF4A.

References

1. A. Marcus, D. P. Weeks, J. P. Leis and E. B. Keller, *Proc. Natl. Acad. Sci. USA*, 1970, **67**, 1681–1687.
2. W. F. Anderson and D. A. Shafritz, *Cancer Res.*, 1971, **31**, 701–703.
3. M. Geisen, R. Roman, S. N. Seal and A. Marcus, *J. Biol. Chem.*, 1976, **251**, 6075–6081.
4. H. Trachsel, B. Erni, M. H. Schreier and T. Staehelin, *J. Mol. Biol.*, 1977, **116**, 755–767.
5. R. Benne and J. W. Hershey, *J. Biol. Chem.*, 1978, **253**, 3078–3087.
6. J. A. Grifo, S. M. Tahara, M. A. Morgan, A. J. Shatkin and W. C. Merrick, *J. Biol. Chem.*, 1983, **258**, 5804–5810.
7. W. C. Merrick, R. D. Abramson, D. D. Anthony Jr, T. E. Dever and A. M. Caliendo, in *Translational Regulation of Gene Expression*, ed. J. Ilan, Plenum Publishing Corporation, New York, N.Y., 1987, 265–286.
8. M. Kozak, *Cell*, 1978, **15**, 1109–1123.
9. M. Kozak, *Cell*, 1980, **22**, 459–467.
10. B. K. Ray, T. G. Lawson, J. C. Kramer, J. A. Cladaras, J. A. Grifo, R. D. Abramson, W. C. Merrick and R. E. Thach, *J. Biol. Chem.*, 1985, **260**, 7651–7658.
11. Y. V. Svitkin, A. Pause, A. Haghighat, S. Pyronnet, G. Witherell, G. J. Belsham and N. Sonenberg, *RNA*, 2001, **7**, 382–394.
12. F. Rozen, I. Edery, K. Meerovitch, T. E. Dever, W. C. Merrick and N. Sonenberg, *Mol. Cell. Biol.*, 1990, **10**, 1134–1144.
13. G. W. Rogers Jr, N. J. Richter and W. C. Merrick, *J. Biol. Chem.*, 1999, **274**, 12236–12244.

14. G. W. Rogers Jr, A. A. Komar and W. C. Merrick, *Prog. Nucl. Acid Res. Mol. Biol.*, 2002, **72**, 307–331.
15. T. E. Dever, M. J. Glynias and W. C. Merrick, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 1814–1818.
16. P. Linder, P. F. Lasko, M. Ashburner, P. Leroy, P. J. Nielsen, K. Nishi, J. Schneir and P. P. Slonimski, *Nature*, 1989, **337**, 121–122.
17. A. Pause and N. Sonenberg, *EMBO J.*, 1992, **11**, 2643–2654.
18. A. Pause, N. Methot and N. Sonenberg, *Mol. Cell. Biol.*, 1993, **13**, 6789–6798.
19. A. Pause, N. Methot, Y. Svitkin, W. C. Merrick and N. Sonenberg, *EMBO J.*, 1993, **13**, 1205–1215.

Preface

RNA helicases are enzymes that use ATP to bind and remodel RNA and RNA–protein complexes. In the more than 25 years that have passed since the discovery of RNA helicases, it has become clear that these proteins are found in all forms of cellular life and in many viruses. RNA helicases are involved in virtually all aspects of RNA metabolism, predominantly as parts of the RNA–protein assemblies that catalyse processes such as pre-mRNA splicing, ribosome biogenesis, and translation initiation. In eukaryotic RNA metabolism, RNA helicases are the largest class of enzymes, and many of these proteins are essential for cellular function and viability.

In this volume, we review the accumulated knowledge on RNA helicases and provide a systematic overview of these enzymes. We decided to discuss RNA helicases in the context of a sequence and structure-based, phylogenetic system that subdivides the proteins into superfamilies and families. This subdivision, which correlates with functional characteristics, is based on thousands of sequences and numerous structures, and is thus likely to remain valid for the years to come.

For this first volume specifically dedicated to RNA helicases, we have been fortunate to bring together leading experts in the vibrant field of research on these proteins. We are deeply grateful for the wonderful work of the authors and the unique set of reviews. The authors have done tremendous work integrating data from diverse fields of current and ongoing biological research. Notwithstanding, the reader will notice emphasis on structural and mechanistic themes, consistent with the scope of this series. All contributions reflect the viewpoints on the covered topics at the time of Summer 2009, when the book went into the production phase. Since then, more data have been published, some of which addresses questions posed in several chapters. New insight on RNA helicase structure and function accumulates at a vigorous pace,

RSC Biomolecular Sciences No. 19
RNA Helicases
Edited by Eckhard Jankowsky

Published by the Royal Society of Chemistry, www.rsc.org

consistent with the central role of these enzymes in RNA metabolism, but also emphasising the need for a systematic overview of the field.

It is our hope that this volume will become a valuable resource for researchers who look for a comprehensive reference on RNA helicases, but also for those who just stumbled across yet another of these enzymes during their studies. Finally, we secretly wish that this book helps to inspire experiments to solve the many questions that continue to surround the central, intriguing, and still enigmatic RNA helicases.

Eckhard Jankowsky
Cleveland

Contents

RSC Biomolecular Sciences No. 19
RNA Helicases
Edited by Eckhard Jankowsky

Published by the Royal Society of Chemistry, www.rsc.org