RNA AND DNA EDITING

Molecular Mechanisms and Their Integration into Biological Systems

Edited by

HAROLD C. SMITH

Department of Biochemistry
and Biophysics University of Rochester

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

RNA AND DNA EDITING

RNA AND DNA EDITING

Molecular Mechanisms and Their Integration into Biological Systems

Edited by

HAROLD C. SMITH

Department of Biochemistry
and Biophysics University of Rochester

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

RNA and DNA editing : molecular mechanisms and their integration into biological systems / [edited by] Harold C. Smith.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-470-10991-5 (cloth)
1. RNA editing. 2. DNA. 3. Nucleotide sequence. 4. Genetic transcription. I. Smith, Harold C.
[DNLM: 1. RNA Editing. 2. Base Sequence. 3. Genomics. 4. Transcription, Genetic. QU 475 R6265 2008]
QH450.25.R57 2008
572.8'8–dc22

2007039289

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

CONTENTS

PREFACE

IN 1986, Rob Benne's research group published their finding of a posttranscriptional process in which mitochondrial messenger RNAs were altered by uridine insertions and deletions, a process he referred to as RNA editing. The finding explained a paradox that the mitochondrial genome of protozoa such as Trypanosomes encoded a scarcity of proteins and many of the genes appeared to have disrupted open reading frames or lacked a translational start codon. Benne's publication took the scientific community by surprise. The known mechanisms for nucleotide modification in RNA and alternative mRNA splicing could simply not accommodate the finding that Trypanosoma mitochondrial mRNAs contained multiple insertions of one or more non-genomically encoded uridines with no apparent consensus flanking sequence at the sites of insertion.

By the early 1990s, several forms of insertion/deletion and base modification editing had been described in amoeba, flagellates, Physarum, mammalian viruses, plants, and the kidney, intestine, liver, and neuronal tissues of mammals. However, many in the scientific community remained unaware of this emerging frontier and the sporadic nature of the identification of editing in different organisms, tissues, and organelles, and the diversity of editing mechanisms led others to question the significance that editing mechanisms would have in understanding cellular systems. For these early years the field collectively had an orphan status, finding outlets for its new discoveries largely in "catch-all" sessions at diverse scientific society meetings.

Beginning in 1994, RNA editing realized solidarity through three international conferences on RNA editing and modification organized independently by Harold Smith and Steve Hajduk (1994, Albany Conference, Rensselaerville, NY, USA), Glenn Bjork, Ted Maden, and Henri Grosjean (1994, EMBO Workshop, Aussois, France), and Paul Sloof and Rob Benne (1996, EMBO Workshop, Maastricht, The Netherlands). The first text dedicated to the topic of RNA editing was edited by Rob Benne in 1993.[*] The inaugural Gordon Research Conference dedicated to RNA editing and modification was led by Smith and co-chaired by Jonatha Gott and Maureen Hanson in 1997. By 1998, many of the RNA editing systems that are known today had been identified, and it was at this time that Grosjean and Benne co-edited a comprehensive text on RNA modification and editing.[†] The field has grown rapidly and gathered momentum as we learn how RNA and DNA editing mechanisms influence, and are influenced by, other biochemical pathways in the cell.

---

[*] *RNA Editing: The Alteration of Protein Coding Sequences of RNA*, Benne, R. (ed.), Ellis Horwood Series in Molecular Biology, Prentice Hall, Englewood Cliffs, NJ, 1993.

[†] *Modification and Editing of RNA,* Grosjean, H., and Benne, R. (eds.), *American Society of Microbiology Press, Washington, D.C., 1997.*

The topic of this book is RNA and DNA editing. The chapters were written from the perspective of the *next generation* of investigators who were formerly trainees in the field or have been newly drawn to it. The authors suggest open questions to pursue while evaluating the context of discoveries and methodologies that have led researchers to this threshold. The vitality of this text lies in its cutting-edge perspective and in its fresh introspective treatment of the progress to date. The target audience of the book are not only the aficionados of the field, but also academics and members of the private sector who are seeking to learn about the field and explore its new applications.

RNA editing is a process in which the nucleotide sequence of RNA is altered from the genomic code. Editing is accomplished through nucleotide insertion, nucleotide deletion, or base modification. It is distinguished from other forms of RNA modification in that the consequence of RNA editing is a change in the diversity and/or abundance of proteins expressed in the proteomes of organisms, their tissues, or organelles. RNA modifications that diversify RNA function or produce a gain or loss of RNA function are also considered editing. Within this rubric, numerous alterations to nucleotides have been documented affecting coding and noncoding sequences of messenger RNAs (mRNAs) as well as transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), and spliceosomal RNAs (UsnRNAs).

As might be anticipated, coordination of editing activity is essential relative to other cellular pathways involving RNA such as transcription, RNA processing, and translation. Our appreciation of this regulation has grown through the characterization of the biological occurrence of RNA editing and the macromolecules that contribute to editing mechanisms. In this regard, the factors involved in RNA substrate recognition and catalysis are diverse, ranging from lone enzymes with both substrate recognition and catalytic activity to macromolecular complexes containing both protein and small RNAs as guides for substrate recognition and multiple proteins to carry out and coordinate editing activity. In A-to-I and C-to-U base modification editing, one editing factor or editosome serves multiple sites. In other systems, such as plant organellar C-to-U editing and organellar guide RNA-dependent mRNA, UsnRNA, and rRNA editing and modification, there is more complexity and a large number of site-specific editing factors.

A recent development in the field is the identification of select members in the family of cytidine deaminase editing enzymes that use single-stranded DNA as a substrate. DNA editing is mutagenic and is responsible for diversification of the genomic coding capacity for immunoglobulins and also serves in antiviral host defense. Another exciting discovery is that A-to-I RNA editing can regulate the production of interference RNA (RNAi) and thereby may constitute an important cellular mechanism for modulating the abundance of individual sequences within the transcriptome. A-to-I RNA editing also can modulate gene silencing through RNAi-dependent regulation of the specificity and activity of the machinery involved in DNA and histone modification, leading to chromatin remodeling.

Given these considerations, RNA and DNA editing will be discussed in four thematic areas to provide a contextual map for this field. Part I, "Diversification of the Proteome through RNA and DNA Editing," highlights how editing regulates protein expression through A-to-I base modification of mRNA, dC-to-dU modification of immunoglobulin genes for somatic hypermutation and immunoglobulin class switch

recombination, guide RNA-dependent uridine insertion and deletion editing of mitochondrial RNA, and C-to-U and U-to-C mRNA editing in plant chloroplast and mitochondrial transcripts. This chapter explores the question "Why are nucleic acid sequences edited instead of encoded genomically?" through discussion of the occurrence of editing sites within transcriptomes and their distribution within individual RNAs. Depending on the biological system, editing can be seen through the lens of diversification, repair, or mutagenesis. Paramount in these discussions are mechanisms that govern RNA editing site selectivity and specificity and restrict the chromosomal domains targeted for DNA editing. Regulation depends on the temporal control of site-specific editing factor expression, subcellular localization, their interaction with nucleic acids, and the composition of individual editosomes. The reader will appreciate how diversity in *cis*- and *trans*-acting factors in different species, or in different organelles within the same species, contributes to different patterns of editing activity and thereby enables plasticity in each biological system.

Part II, "Functional Coordination of RNA Editing with Other Cellular Mechanisms," brings to the forefront why RNA and DNA editing is essential for cell survival and adaptation. This section profiles base modification of RNA and DNA and guided RNA editing as examples where cells require editing to produce functional tRNAs, process rRNA, splice pre-mRNA, regulate the stability of mRNAs, and control RNAi and viral infectivity. In some instances, editing at different sites within the same RNA is interdependent and requires coordination of the activity of different editosomes or transport of editing enzymes or their substrates within the cell and its organelles. In other examples, RNA editing site selectivity is coordinated through the interaction of A-to-I editing enzymes with the C-terminus or RNA polymerase II. In this way, editing factors have immediate access to nascent transcripts and can carry out editing before pre-mRNA splicing deletes introns that participate in RNA secondary structure necessary for editing site recognition. Transcription also makes available single-stranded DNA within chromosomes that can be targeted for mutational DNA editing leading to diversification of the genomic sequences encoding the variable regions of antibodies (as described in the prior section). Reverse transcription, coupled to RNase H activity, also regulates editing activity by exposing single-stranded viral DNA during replication for mutagenic DNA editing as a form of host defense.

The global role of RNA editing in cellular regulation is emphasized in this section of the book through several examples. Modification editing of U2 spliceosomal RNA is essential for U2-snRNP splice site binding specificity and spliceosome activity. The stability of select mRNAs is affected by binding of the factors responsible for C-to-U mRNA editing in mammalian cells to AU-rich elements in mRNA. And, modulation of RNAi production by A-to-I RNA editing is described as a mechanism for regulating gene silencing by affecting the specificity and activity of the enzymes that carry out DNA and histone modifications. The exquisite level of integration of editing with other biochemical pathways and cellular functions described in all of the chapters will lead the reader to the inescapable conclusion that RNA and DNA editing have significant roles in biology that includes, and goes well beyond, codon sequence changes and reading frame alterations.

A long-sought goal in the field has been to use our understanding of editing sites and editing factors to discover novel editing substrates and new biological roles for

editing. Part III, "Predictive Studies," underscores the power of computational approaches in identifying novel editing sites and predicting the biological consequences of editing at these sites. Historically, computational analyses have been used sporadically to validate sequences as having been edited; however, computational methods have developed to the point where comparative sequence analyses enable genome wide predictions of edited mRNA sequences. Computational approaches have also advanced comparative phylogenic analyses of edited sequences. These studies have provided unique insights into the origins of editing systems, their evolution, and an understanding of the conserved, minimally essential functional domains within editing factors.

Highly related to these discussions is Part IV, "Structural Approaches," which is the final section of the text. Structural biology is an enabling technology for basic science, biomedical research, and drug development. The structural basis for function is more conserved in many instances than is primary nucleotide or amino acid sequence. Comparative structural analyses have been vital in predicting RNA secondary structure of the substrates for A-to-I editing and guide RNAs as well as the functional folds within enzymes in both A-to-I and C-to-U families of deaminases. Comparative structural analysis suggests conserved protein folds and implicate, in some instances, ancient phylogenic origins for components of editing machinery. Importantly, computational and structural studies suggest the reaction chemistry that enzymes catalyze, and they aim to predict the physical constraints in macromolecules that determine substrate and editing site specificity.

The selection of chapters and organization of the book was conceived with multiple purposes in mind. The text serves as a reference for background information in the field. It provides an opportunity for the newest contributors to the editing field to express their vision for the future. The perspectives voiced by these authors are anticipated to be provocative and are intended to motivate discussion, lead to new experiments, and promote collaboration. Finally, this book is intended to promote new hypotheses and models to springboard the next generation of discovery in the field.

Harold C. Smith

ACKNOWLEDGMENTS

THE EDITOR is grateful to the authors whose expertise and perspectives are reflected in the sixteen chapters of this text. The editor is also grateful to colleagues and government and private funding agencies for their contributions to the vitality and growth of the editing field. Special thanks are due to Drs. Juan Alfonzo, Ryan Bennett, Andrea Bottaro, Bernie Brown, Gordon Carmichael, Nic Davidson, Donna Driscoll, Rubin Harris, Celeste MacElrevey, David Mathews, Stu Maxwell, Tom Meier, Mike Mulligan, Marie Öhman, Eric Phizicky, Laurie Read, Larry Simpson, Joseph E. Wedekind and Yi-Tao Yu for critical reading of portions of the text and helpful suggestions.

CONTRIBUTORS

Juan D. Alfonzo, Department of Microbiology, the Ohio State University, Columbus, OH 43210. Email: alfonzo.1@osu.edu

M. Wadud Bhuiya, Department of Chemistry, Wake Forest University, Winston-Salem, NC 27109. *Present address*: Department of Biology, Brookhaven National Laboratory, Upton, NY 11973. Email: bhuiya@bnl.gov

Valerie Blanc, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110. Email: vblance@IM.WUSTL.edu

Bernard A. Brown II, Department of Chemistry, Wake Forest University, Winston-Salem, NC, 27109. *Present address*: Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695.
Email: brownba2@gmail.com

Gordon G. Carmichael, Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, CT 06030.
Email: carmichael@nso2.uchc.edu

Ling-Ling Chen, Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, CT 06030. Email: LLchen@student.uchc.edu

Jorge Cruz-Reyes, Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843. Email: cruzrey@tamu.edu

Nicholas O. Davidson, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110. Email: nod@im.wustl.edu

Zachary L. Demorest, Department of Biochemistry, Molecular Biology and Biophysics, Minneapolis, MN 55455. Email: demo0034@umn.edu

Dylan E. Dupuis, Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015.

Eli Eisenberg, School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel. Email: elieis@post.tau.ac.il

Keith Gagnon, Department of Structural and Molecular Biochemistry, North Carolina State University, Raleigh, NC 27695.
Email: keith_gagnon@ncsu.edu

Willemijn M. Gommans, Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015.

Michael W. Gray, Program in Evolutionary Biology, Canadian Institute for Advanced Research, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS B3H 1X5, Canada. Email: m.w.gray@dal.ca

Reuben S. Harris, Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN 55455.
Email: rsh@umn.edu

Alfredo Hernandez, Department of Biochemistry and Biophysics, Texas A&M University, 2128 TAMU College Station, TX 77843. Email: ahernandez@tamu.edu

John Karijolich, Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642.
Email: john_karijolich@urmc.rochester.edu

Erez Y. Levanon, Department of Genetics, Harvard Medical School, Boston, MA 02115. Email: elevanon@genetics.med.harvard.edu

Senjie Lin, Department of Marine Sciences, University of Connecticut, Groton, CT 06340. Email: senjie.lin@uconn.edu

Stefan Maas, Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015. Email: swm3@lehigh.edu

Celeste MacElrevey, Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642.
Email: Celeste_MacElrevey@urmc.rochester.edu

Donna A. MacDuff, Department of Biochemistry, Molecular Biology and Biophysics, Minneapolis, MN 55455. Email: macd0110@umn.edu

Jill E. McCane, Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015.

E. Stuart Maxwell, Department of Structural and Molecular Biochemistry, North Carolina State University, Raleigh, NC 27695. Email: tu_maxwell@ncsu.edu

U. Thomas Meier, Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, Bronx, NY 10461. Email: meier@aecom.yu.edu

Jun-ichi Obokata, Center for Gene Research, Nagoya University, Nagoya, 464-8602, Japan. Email: obokata@gene.nagoya-u.ac.jp

Steven M. Offer, Department Biochemistry, Molecular Biology and Biophysics, Minneapolis, MN 55455. Email: offer014@umn.edu

Marie Öhman, Department of Molecular Biology and Functional Genomics, Stockholm University, S-106 91 Stockholm, Sweden.
Email: marie.ohman@molbio.su.se

F. Nina Papavasiliou, Laboratory of Lyphocyte Biology, the Rockefeller University, New York, NY 10021. Email: papavasilou@mail.rockefeller.edu

Evan M. Ritter, Department of Chemistry, Wake Forest University, Winston-Salem, NC 27109. *Present address*: 504 Murdocksville Road, West End, NC, 27376. Email: ritterem@gmail.com

Ann M. Sheehy, Department of Biology, College of the Holy Cross, Worcester, MA 01610. Email: asheehy@holycross.edu

Toshiharu Shikanai, Graduate School of Agriculture, Kyushu University, Fukuoka 812-8581, Japan. Email: shikanai@agr.kyushu-u.ac.jp

David Stephenson, Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642.
Email: david_stephenson@urmc.rochester.edu

Nicholas E. Tatalias, Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015.

Joseph E. Wedekind, Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642.
Email: joseph_wedekind@urmc.rochester.edu

Yi-Tao Yu, Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642.
Email: yitao_yu@urmc.rochester.edu

Huan Zhang, Department of Marine Sciences, University of Connecticut, Groton, CT 06340. Email: huan.zhang@uconn.edu

Xinxin Zhang, Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695. Email: xinxin_Zhang@ncsu.edu

Jing Zhou, Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, CT 06030. Email: zhou@student.uchc.edu

PART I

DIVERSIFICATION OF THE PROTEOME THROUGH RNA AND DNA EDITING

CHAPTER 1

DIVERSIFYING EXON CODE THROUGH A-TO-I RNA EDITING

Willemijn M. Gommans
Dylan E. Dupuis
Jill E. McCane
Nicholas E. Tatalias
Stefan Maas

AN INCREASING number of gene transcripts are found to be subject to recoding by RNA editing. RNA-targeted recoding leads to the substitution of single amino acids in the resulting proteins with subtle or sometimes drastic impact on protein function. New strategies to search for edited genes in mammals have accelerated the discovery of new targets and promise to reveal the many roles of RNA editing in gene regulation.

1.1 INTRODUCTION AND BACKGROUND

According to the central dogma, protein-coding sequences in eukaryotic genomes directly predict the primary structure of the encoded protein. However, processes such as alternative splicing of exons result in the inclusion or omission of protein domains and subdomains and thereby substantially extend the repertoire of expressible protein variants (1). Often, the occurrence and extent of alternative splicing is not predictable from analysis of genomic DNA sequences. Other posttranscriptional RNA modifications also contribute to the complexity of the proteome. One such important mechanism is RNA editing by adenosine modification (2–4), where single adenosine bases are converted into inosine. Since inosine is interpreted by the translation machinery as guanosine (5), A-to-I modification often results in nonsynonymous codon changes leading to protein variants with single amino acid substitutions. To date, it is impossible to predict with reasonable confidence a recoding event in mRNA from analyzing genomic sequence data. In this chapter we are reviewing the current knowledge regarding the prevalence and consequences of A-to-I recoding events in eukaryotic transcripts and discuss recent strategies for identifying and characterizing

RNA and DNA Editing: Molecular Mechanisms and Their Integration into Biological Systems,
Edited by Harold C. Smith

recoding editing sites in translated sequences as well as A-to-I editing events in micro RNA transcripts.

1.1.1 Initial Discovery and Context of A-to-I RNA Editing and ADARs

It came as a surprise when in 1991 an A/G discrepancy between genomic and cDNA sequences of the mammalian glutamate-gated ion channel subunit GluR-2 (6) turned out to be due to an adenosine base modification on the RNA level. Editing of this adenosine nucleotide results in the conversion of a glutamine codon into a codon specifying arginine. In fact, this single nucleotide substitution turned out to dominantly regulate ion-permeability in heteromeric receptor molecules and up to today represents maybe the most significant, intriguing, and puzzling case of adenosine modification editing in mammals (see Section 1.2.3).

The initial discovery of adenosine-modification editing quickly led to the identification of several other cases of recoding in nervous-system-specific transcripts, such as additional GluRs (7, 8) and 5HT2C-R (9). In each case a single nucleotide change resulting in an amino acid substitution could be linked to a change in protein function. Since unedited and edited protein variants often are co-expressed in the same cells RNA editing was soon recognized as a potentially important mechanism to diversify genetic information with the ability to enhance the complexity of the eukaryotic transcriptome and proteome.

At the time that the editing event in GluR-2 mRNA was discovered, neither the cellular machinery responsible for this adenosine base substitution nor the molecular mechanism at play was known. The observed A-to-G change in the cloned cDNAs was thought to be a result of either an adenosine modification process that alters this purine into another purine base functionally equivalent to a guanosine, such as hypoxanthine or due to a mechanism that involves removal of the base or of the whole nucleotide followed by introduction of the guanosine.

Interestingly, there was an enzyme known for a long time that converts adenosine mononucleotides to hypoxanthin nucleotides (also termed inosine). This evolutionary conserved adenosine deaminase (ADA) mediates an important step in eukaryotic and prokaryotic nucleotide metabolism. The ADA enzyme is well-studied and has become an important therapeutic target as ADA deficiency leads to various types of immune disorders (10). ADA modifies adenosine mononucleotides employing a hydrolytic deamination mechanism. However, the enzyme is not active on adenosines present in the context of DNA or RNA molecules. In addition to the modification of mononucleotides by ADA, the modification of genomically encoded adenosines to inosines in transfer RNAs (tRNA) has long been known (for reference see 11) and represents a critical feature for the degeneracy of the genetic code (wobble base in the anticodon of several tRNAs). The reaction mechanism and enzyme responsible for generating the wobble base was only recently revealed (12) (see below).

More importantly, a few years before the discovery of adenosine modification editing in pre-mRNAs, a novel enzymatic activity had been discovered that specifically targets adenosines embedded in dsRNA molecules (13, 14). Initially, it was

described as dsRNA unwinding activity, but the actual nature of the molecular process was soon identified as adenosine-to-inosine modifications through an analysis of reaction products (15).

With the establishment of *in vitro* systems for RNA editing based on glutamate receptor transcript minigenes and cellular extracts, the chemical mechanism of the observed A-to-G changes in mRNAs was also soon shown to be the result of A-to-I deamination, catalyzed by a zinc-dependent protein factor (16–18). Furthermore, the cis-acting features in editing targets were characterized, identifying the requirement for partial double-stranded (ds)RNA secondary structures but with no obvious primary sequence signatures (8, 19, 20). This clearly distinguished the A-to-I editing mechanism from the mammalian C-to-U deamination process that involves secondary structure elements in addition to a primary sequence motif (mooring sequence) that guides the RNA modification machinery (see Chapter 11). *In vitro* editing systems also accelerated the isolation and cloning of the first A-to-I RNA editing enzyme from mammals (21–23). It turned out that the responsible protein (initially termed dsRAD, or DRADA, which later was renamed ADAR1) had in fact been investigated by several laboratories as either an interferon-induced protein with potential antiviral functions (24) or as the dsRNA-specific A-to-I editing activity in mammalian cells (see Section 1.1.3) (25, 26). Cloning of the first mammalian ADAR (ADAR1) was followed by the identification of ADAR2 (27) and ADAR3 (28), as well as ADAR homologs in other vertebrates (29, 30), flies (31) and worms (32) (see Section 1.1.3). Also, related enzymes responsible for tRNA-specific A-to-I editing were cloned and characterized in several species (33–35).

The C-to-U editing enzyme (APOBEC1) is remotely sequence-related to the first adenosine-targeting editing enzyme ADAR1, and it is believed that APOBEC1 cytidine deaminase and the deaminase domain of ADARs may share a common ancestor gene (36, 37). Interestingly, neither ADAR1 nor APOBEC1 shows primary sequence homology to adenosine deaminase (ADA), and their predicted (ADAR1) or known (APOBEC1) three-dimensional structures also differ substantially from that of ADA even though the reaction mechanisms catalyzed by ADA, ADAR, and APOBEC1 are highly similar.

1.1.2 Important Cases of Recoding by A-to-I Modification in Pre-mRNA

The first mammalian editing events that were characterized affect several subunits of glutamate-gated ion channels (7, 8) and a prominent serotonin receptor subunit (9). These proteins were all found to be modulated in function by single or multiple site-selective adenosine modifications within their primary transcripts. Serendipity played a central role in the identification of these targets. Only recently, systematic screening methods designed to identify recoding events caused by A-to-I editing have been developed (38–41) leading to the identification of a few additional targets (see Table 1.1).

Overall, the notion that A-to-I RNA recoding editing may be particularly significant for the nervous system is supported by the preponderance of brain tissue-specific editing events. Particularly in the fly (*Drosophila melanogaster*), the

TABLE 1.1 A-to-I Editing in the Coding Regions of Mammalian, Invertebrate and Viral Genes

A. Mammalian Genes

Function	Gene (Accession Number)	aa Substitution	ADAR[a]	Functional Impact	Ref.
Serotonin receptor	5-HT_{2c}R (NM_000868)	I156V, I156M, N158S, N158D, N158G, I160V	ADAR1 & ADAR2 (B, C and E-site), ADAR1 (A-site), ADAR2 (D-site)	Reduced efficacy G-protein coupling	9
Glutamate receptor	GluR-2 (NM_000826)	R763G Q606R	ADAR1 & ADAR2 ADAR2	Decreased Ca^{2+} permeability; alteration maturation and cellular trafficking; faster recovery from desensitization	8 6
Glutamate receptor	GluR-5 (NM_175611)	Q621R	?	Variation in ion permeability	6
Glutamate receptor	GluR-6 (NM_175768)	I567V, Y571C, Q621R	? (Y571C: ADAR2)	Variation in ion permeability	6
Glutamate receptor	GluR-3 (NM_000828)	R775G	ADAR1 & ADAR2	Faster recovery from desensitization	8
Glutamate receptor	GluR-4 (NM_000829)	R765G	ADAR1 & ADAR2	Faster recovery from desensitization	8
Potassium channel	$hK_v1.1$ (NM_000217)	I400V	ADAR2	Faster recovery from desensitization	41, 59
Unknown	BC10 (NM_006698)	Y2C, Q5R, K15R	?	?	39, 40
Cross-linking actin filaments	FLNA (NM_001456)	Q2333R	?	?	40
FMR1 interacting protein	CYFIP2 (NM_001037333)	K320E	?	?	40
Chloride channel	Gabra-3 (NM_000808)	I342M	ADAR1 & ADAR2	?	103
A-to-I editing enzyme	ADAR2[b] (NM_001033049)	Intronic editing leads to frameshift	ADAR2	Alternative splicing	86

B. Invertebrate and Viral Genes

Function	Gene (Accession number)	Organism	aa Substitution	Functional Impact	Ref.
Potassium channel	SqK$_v$1.1 (U50543)	Squid	12 recoding sites	Altered channel kinetics; reduced ability to form tetramers	114
Potassium channel	sq K$_v$2 (Y14390)	Squid	12 recoding sites	Altered channel kinetics (channel closure rate & altered slowest time constant)	115
Basic fibroblast growth factor	bFGF (X16627)	*Xenopus*	Hypermutation	Unknown	116
Editing enzyme	dADAR (AF208535)	*Drosophila*	S437G	Editing activity	31, 117
Sodium channel	Para (NM_001042816)	*Drosophila*	Q473R K1455R N1587S	Unknown	118
Glutamate-gated chloride channel	GluRIIE (CG31201)	*Drosophila*	I27V, K241R, N345S	Unknown	119
Voltage gated calcium channel	cac (NM_206693)	*Drosophila*	S514G, I815M, N839S, N906S, S937G, M1016V, N1185S, N1368G, N1580D, R1602G	Unknown	120
Amine receptor	DopEcR (CG18314)	*Drosophila*	I316V Stop323W	Unknown	104
Hydrogen-transporting two-sector ATPase	CG13167	*Drosophila*	I9V	Unknown	104
Protein phosphatase type 1, regulator	CG9619	*Drosophila*	S160G	Unknown	104
Transporter activity	Spinster (CG8428)	*Drosophila*	N67G	Unknown	104

(*Coninued*)

TABLE 1.1 (*Continued*)

Function	Gene (Accession number)	Organism	aa Substitution	Functional Impact	Ref.
YT521-B	CG12076	*Drosophila*	Q636R	Unknown	104
Unknown	Tetraspanin 33B (CG14936)	*Drosophila*	Two silent sites	Unknown	104
Unknown	4f-rnp	*Drosophila*	Hypermutation	Unknown	121
Sodium channel	DSCI (CG9071)	*Drosophila*	M1174V, I1199V only *D. pseudoobscura*)	Unknown	41
Potassium channel	Sh (CG12348)	*Drosophila*	K178E, K178G, K178R, I360M, I464V, T489A, Q491R	Unknown	41
Potassium channel	Eag (CG10952)	*Drosophila*	K467R, Y548C, N567D, K699R	Unknown	41
Potassium channel	Slo (CG10693)	*Drosophila*	N264D, S977G	Unknown	41
Calcium sensor	Syt (CG3139)	*Drosophila*	I365V, K377R, I381V, I403M	Unknown	41
SNARE binding	unc-13 (CG2999)	*Drosophila*	S2371G	Unknown	41
SNARE protein	cpx (CG32490)	*Drosophila*	I124M, N129D, N129G, N129S	Unknown	41
Unknown	stnB (CG40306)	*Drosophila*	T1186A	Unknown	41
Adaptor protein	lap (CG2520)	*Drosophila*	T372A	Unknown	41
nAChRα subunit	Dα5 (CG32975)	*Drosophila*	I504V, T553A, I554V, I558M	Unknown	41

nAChR subunit	Dα6	*Drosophila*	N133S, I134V, H138R, N139G, N139S, N139D, I156M, N187S	Unknown	122
nAChR subunit	Amelα6 (ortholog of Dα6)	Apis mellifera	N164S, K176R, I181M, T184A	Unknown	123
nAChRβ subunit	ARD (CG11348)	*Drosophila*	R56G, I73M	Unknown	41
nAChRβ subunit	SBD (CG6798)	*Drosophila*	T278A	Unknown	41
GABA receptor	Rdl (CG10537)	*Drosophila*	R122G, I283V, N295G, M360V	Unknown	41
GTPase	Rab26 (GH21984)	*Drosophila*	K365R	Unknown	38
Ral GTPase activator	Rlip (GH01995)	*Drosophila*	I229V, E230G, K233E, E254G, K265R	Unknown	38
Rab3 guanyl-nucleotide exchange factor	Rab3-GEF (HL01222)	*Drosophila*	Q2022R, S2054G	Unknown	
Promotes synaptic vesicle budding	endoA (GH12907)	*Drosophila*	K129R, K137E	Unknown	38
Synapsin	Syn (CG3985)	*Drosophila*	R20G	Reduced PKA phosphorylation *in vitro*	124
AP-2 subunit	α-Adaptin (RH30202)	*Drosophila*	T207A	Unknown	38
Kinesin-dependent axonal transport	Syd (GH19969)	*Drosophila*	S983G	Unknown	38

(*Coninued*)

TABLE 1.1 (*Continued*)

Function	Gene (Accession number)	Organism	aa Substitution	Functional Impact	Ref.
Ca^{2+} binding protein	Cpn (GH08002)	*Drosophila*	S402G	Unknown	38
K^+ dependent Na^+, Ca^{2+} antiporter	Nckx30C (HL01989)	*Drosophila*	K365R	Unknown	38
K^+ dependent Na^+, Ca^{2+} antiporter	CG1090 (GH23040)	*Drosophila*	S358G	Unknown	38
Na^+, K^+ exchanging ATPase	Atpα (GH23483)	*Drosophila*	Y390C	Unknown	38
Ca^{2+} binding, acyltransferase activity	CG32699 (HL01250)	*Drosophila*	I175M	Unknown	38
Trc kinase activator	Mob1 (RH70633)	*Drosophila*	N91D	Unknown	38
G-protein coupled receptor	Boss (GH10049)	*Drosophila*	T529A, T533A	Unknown	38
Potassium channel	SK (GH16664)	*Drosophila*	Y377C	Unknown	38
Chloride channel	CG31116 (GH23529)	*Drosophila*	K232R, T259A, K268R, E269G	Unknown	38
Actin nucleation factor	Spir (GH13327)	*Drosophila*	K339R	Unknown	38
Regulator of actin filament formation	Atx2 (GH01409)	*Drosophila*	K320R, K337R	Unknown	38
Structural constituent of cytoskeleton	CG32245 (GH04632)	*Drosophila*	N297D	Unknown	38
ATPase	CG32809 (GH23439)	*Drosophila*	K179R	Unknown	38
Phosphatidylinositol transporter	Retm (GH05975)	*Drosophila*	Q245R	Unknown	38
Unknown	CG1552 (GH14443)	*Drosophila*	K121R	Unknown	38
Unknown	CG31531 (GH25780)	*Drosophila*	K679E	Unknown	38

Unknown	CG3556 (GH17087)	*Drosophila*	I572V	Unknown	38
Unknown	CG9801 (GH23026)	*Drosophila*	S345G	Unknown	38
Unknown	I(1)G0196 (GH02989)	*Drosophila*	Q1148R, S1172G, Q1176R	Unknown	38
Unknown	CG12001 (HL01040)	*Drosophila*	I325V	Unknown	38
Unknown	CG30079 (HL05615)	*Drosophila*	I127M, T303A, Q343R, Q358R, S360G	Unknown	38
nAChR subunit	α7-2 (homolog to Dα6 sites)	*H. virescens*	N133S, N139G, N139S, N139D, I156M	Unknown	122
nAChR subunit	Mdalpha6 (ortholog of Dα6)	*Musca domestica*	N129S, I130V, H134R, N135S, N135D, N135G, N137S, I152M, N183S	Unknown	125
Viral replication	HDAg-p24 (AJ307077)	Hepatitis delta virus	Stop196W	Switch from viral replication to packaging	62

[a]Preferential editing enzyme for that target site

[b]Causes indirect change in amino acid sequence

large number of neuronal editing targets and the fact that the complete elimination of the A-to-I editing machinery results in a specific neurological phenotype (42) demonstrates a critical role of editing for neural function. However, in mammals, as well as in the fly, the list of non-neuronal recoding targets is steadily growing, though knowledge regarding the physiological significance of recoding editing affecting non-neuronal transcripts is largely lacking.

1.1.2.1 Mammalian Glutamate Receptor Subunits Ionotropic glutamate receptors (iGluRs) constitute an important class of neurotransmitter receptors in the central nervous system that mediate fast excitatory neurotransmission and have been implicated in mechanisms of plasticity, such as learning and memory (43). A total of five glutamate gated ion channel subunits have been shown to be recoded at single positions within their mRNAs affecting a total of eight codons (for recent detailed reviews on the sites and regulatory roles of ion-channel receptor editing, see reference 44 and references therein). Most significantly, the GluR-2 subunit is A-to-I edited at a critical position in the ion channel molecule, which constitutes the molecular determinant for Ca-permeability (6) and in addition regulates channel trafficking (45) and receptor assembly (46). Editing at this position will therefore influence all these processes. The editing event alters a glutamine (Q) codon (CAG) to a codon (CIG) specifying arginine (R). This Q/R site of editing is further remarkable in that virtually 100% of GluR-2 pre-mRNA molecules are edited and therefore almost no GluR-2(Q) protein is present in the brain. The physiological significance of this recoding event became evident when transgenic mice with impaired RNA editing function were engineered. This resulted in mice with a severe epileptic phenotype and premature death (47, 48). It was shown that the reduction in editing at the Q/R site from 100% to ~60% results in a drastically increased conductance and Ca^{2+}-influx into principal neurons. These alterations were directly responsible for the observed phenotype, because mice that carry a genomic mutation fixing the RNA editing event in the genome showed a wild-type phenotype despite the editing deficiency (48). Why is the critical arginine codon generated by almost complete editing of the Q/R site and not genomically specified? Currently, it cannot be ruled out that nonedited versions of the GluR-2 subunit have a function during early development or within specific neuronal cell types. However, that function is dispensable for survival and normal development as judged by the lack of a discernable phenotype in transgenic mice that cannot produce nonedited GluR-2 (49). A selective deficiency in GluR-2 Q/R site editing has been implicated in a number of pathological phenotypes in humans (see reference 50 for review), such as amyotrophic lateral sclerosis (ALS). In ALS, the ensuing increase in glutamate receptor Ca^{2+}-permeability of affected neurons due to a decreased editing activity may be responsible or a contributing factor to neuronal death (51).

Another editing site changing an arginine (R) codon into a glycine (G) codon is shared between GluR subunits GluR-2, -3, and -4. Here the single amino acid alteration regulates kinetic properties of the heteromeric receptor channel (8) and also modulates receptor biogenesis (52). The extent of editing at the R/G position varies between the different GluR subunits and between neuronal cell types. It also undergoes significant regulation during embryogenesis changing from low level

editing extents during early embryonic stages to high levels in adult individuals (8). The glutamate receptor subunits GluR-5 and -6 are also edited at one or three sites (7), respectively. Here the recoding events modulate the ion-permeability and kinetic properties of the corresponding ion channels (44).

1.1.2.2 Serotonin Receptor Another prominent and well-studied example of A-to-I RNA editing is the serotonin receptor subtype 5-HT_{2C}, which is important for neuronal pathways influencing sensory and motor processes, as well as behaviors. The 5-HT_{2C} receptor is part of a G-protein-coupled transmembrane receptor that couples serotonin neurotransmitter action to intracellular signaling pathways. This mainly leads to the activation of phospholipase C, which, in turn, results in a rise in intracellular inositol phosphates and diacylglycerol. The latter elicits protein kinase activation and induces an increase in intracellular Ca^{2+} concentration. A-to-I RNA editing in 5-HT_{2C} transcripts affects five major sites, which are all located within the same second intracellular loop of the receptor protein (9). Overall, the higher the extent of modification by editing at these sites, the less sensitive the receptor becomes to serotonin activation, which is the result of a decreased G-protein coupling efficiency (53, 54). The 5-HT receptors have been implicated in the etiology of several neurological and behavioral disorders, such as depression, anxiety and schizophrenia. Intriguingly, changes in the RNA editing patterns of 5-HT_{2C} transcripts have been observed in brains of people that suffered from suicidal depression (55). Mice treated with fluoxetine (a serotonin uptake inhibitor) show the converse type of change in the RNA editing pattern of 5-HT_{2C} sequences. These data indicate that the editing extent at these modification sites may be able to change in response to external signals, such as different levels of synaptic serotonin (55). In agreement with these observations, treatment of cells with the cytokine IFN-alpha resulted in alterations in the editing pattern of 5-HT_{2C} mRNA, which may link the observed depression in patients undergoing cytokine therapy to fluctuations in editing activity (56). Recently, increasing evidence has linked changes in mood and behavior to alterations of serotonin receptor editing (for review see reference 57).

1.1.2.3 Kv1.1 Potassium Channel The mammalian Kv1 subfamily of potassium channels plays an essential role in membrane hyperpolarization during an action potential and in the propagation of action potentials along the plasma membrane (58). The tetrameric receptors form a diverse group of ion channels due to the existence of several subunits and also due to A-to-I RNA modification. Editing of the human Kv1.1 transcripts modulates the kinetics of channel inactivation (59). The editing event in the human Kv1.1 mRNA is related to the site of editing in *Drosophila melanogaster* shaker potassium channels and has independently evolved at the equivalent (analog) site in the *D. melanogaster* Shab potassium channel (59).

This more recently reported editing site in humans stands out because the pre-mRNA that undergoes modification does not contain introns, which means that the partially base-paired RNA fold-back structure is comprised entirely of exonic sequences (41). For each of the cases described above, the molecular determinants for site-selective and efficient editing involve a partially double-stranded RNA fold-back structure in the substrates that is formed between exonic sequences that surround

the to-be-edited adenosine and partially complementary sequence elements in a neighboring intron (see also Section 1.1.3 below).

1.1.2.4 Additional Recoding Targets in Vertebrates, Invertebrates, and Viruses A number of additional recoding events have been reported in mammals (see Table 1.1 and Section 1.1.2), for which there are currently no experimental data available regarding the physiological impact of editing (38–41). Interestingly, they also include non-neuronal transcripts.

In addition, a few examples exist where recoding events are predicted from A/G discrepancies that are only detected in transcripts derived from pathological tissues or cells, such as cancer (prox1, PCNP) and lupus erythematosus (60).

Intriguingly, the hepatitis delta virus (HDV) utilizes the A-to-I editing machinery to regulate viral replication. Within the antigenome of this virus, a site-specific adenosine to inosine modification converts a stop codon into a tryptophan codon (61, 62). This leads to the expression of an HDV antigen variant that suppresses replication and enhances the late stages of the viral life cycle (61). It seems that the viral genome has evolved in a way to utilize the host cell's RNA editing machinery for productive replication.

To date, a total of 77 targets for A-to-I editing have been identified in the fruit fly *D. melanogaster*. Few of these targets, of which the majority are expressed specifically in neurons, have been directly investigated with respect to the consequences of editing for protein function.

1.1.3 Cis-Acting Features for A-to-I Editing

The requirement for a partially double-stranded RNA fold-back structure for editing was first established for the GluR-2 Q/R editing site through meticulous analysis of editing extents in minigene substrates that tested the validity of computer-predicted RNA secondary structures of GluR-2 pre-mRNA transcripts (19). The partially base-paired region in the RNA is formed between sequences flanking the to-be-edited adenosine and a partially complementary sequence [termed the editing site complementary sequence (ECS)], which is often located within a downstream or upstream intron (19, 63, 64). Mutations that weaken the predicted RNA fold surrounding the editing site strongly impair or abolish editing at the Q/R site, whereas other mutations that restore the structure boost the levels of site-selective modification (19). Also, the modified adenosine in the GluR-2 Q/R site structure is in a base-paired configuration and changing the base pair into a mismatch decreases editing efficiency (19). Although this seems to be a rather simple set of parameters determining what constitutes an editing substrate, the analogous analysis of RNA fold-back structures governing editing at other sites revealed that the process is much more complex as in some cases the to-be-edited adenosine may be mismatched (8) or may be part of a loop structure embedded in base-paired regions (64). Based on the available data to date, it is not possible to define structural or sequence requirements that would allow straightforward screening for edited genes in sequence or structure databases (see Section 1.2.2).

Apart from the requirement for a partially base-paired structure, both ADAR1 and ADAR2 show certain nearest-neighbor preferences. The ADAR1 enzyme

preferentially targets adenosines that are preceded by U = A > C > G, and ADAR2 displays a 5′-preference of U = A > C = G as well as a 3′-neighbor preference of U = G > C = A (4, 65). These properties may be related to the reaction mechanism for deamination that is believed to involve a flipping of the adenosine into the enzyme's active site, similar to the mechanism of action employed by DNA methyltransferases (4). Certain sequence environments will make access for the enzyme to the to-be-edited adenosine easier. In summary, there is still an unclear picture with respect to the molecular parameters that determine substrate specificity and editing efficiency in natural recoding targets for A-to-I editing.

1.1.4 Properties of the A-to-I Editing Machinery

The family of mammalian ADAR proteins (ADAR1, ADAR2, and ADAR3) share a common general domain architecture (see Figure 1.1A) mainly comprised of two or three double-stranded RNA binding domains (dsRBDs) and a catalytic adenosine deaminase domain toward the C-terminus (for recent reviews on ADARs see references 66 and 67). Only ADAR1 and ADAR2 have been characterized functionally, whereas ADAR3, although closely sequence related to ADAR2, has not yet been assigned a function and does not display adenosine deaminase activity in established assay systems (28, 68). *In vitro* and in cellular assays, ADAR1 and ADAR2 exhibit site-selective and efficient RNA editing activity apparently without any proteinacious co-factor requirements (27, 69).

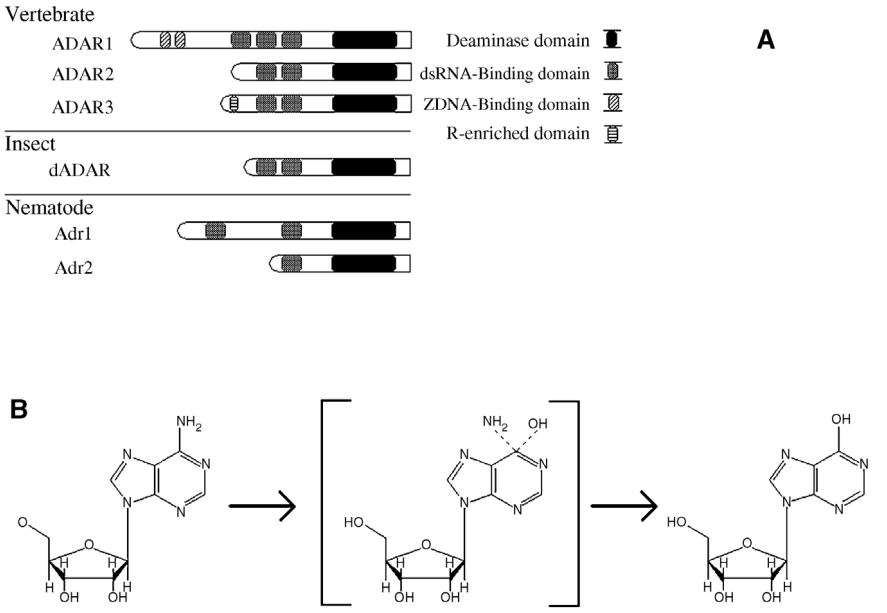

Figure 1.1 Molecular players and mechanism of A-to-I editing. (A) Schematic representation of ADAR domain structures from vertebrates, insects, and nematodes. (B) Depiction of adenosine hydrolytic deamination mechanism with transition state.

Zinc is known to be involved in the catalytic mechanism, where it activates a water molecule that initiates the nucleophilic attack at the C-6 of the targeted adenosine. However, recently it was shown that ADARs are active as dimers (70–73) and also that a small molecule inositol-6-phosphate (IP6) is an essential co-factor for function (74). A crystal structure of the catalytic domain of ADAR2 in conjunction with functional experiments demonstrated that IP6 is critical for protein folding and formation of the catalytic site (74). The mechanism of substrate recognition and site-selectivity of ADARs is not well understood, and ADAR1 and ADAR2 are known to display distinct, but overlapping specificity on known, physiological editing targets. For example, both enzymes seem to be highly active on the glutamate receptor R/G editing sites in GluR-2, -3, and -4. However, the Q/R site of GluR-B, the I/V site of the Kv1.1 potassium channel as well as the D site of the 5-HT_{2C} serotonin receptor are preferentially edited by ADAR2. The enzyme ADAR1 displays preference for the B site of 5-HT_{2C} as well as the amber/W site in the HDV antigenome (27, 56, 75, 76).

One intriguing property of ADARs is that when encountering an extended, perfectly double-stranded RNA molecule, ADARs will promiscuously deaminate up to $\sim$ 50% of all adenosines (77). Further deaminations are probably prevented due to the progressive loss of the substrate's double-strandedness such that the dsRBDs are unable to bind. This highlights that, intrinsically, the ADAR enzymes lack a particular site-selectivity but that most likely the overall three-dimensional shape and sequence environment of edited adenosines in the substrate RNAs provide the specificity seen in physiological recoding targets (4, 78). It has also been suggested that dsRBDs of ADARs may in some cases specifically interact with other structural RNA motifs such as a loop region, thereby mediating site-selective editing (79, 80).

Other organisms, such as insects (31) and nematodes (81), have A-to-I editing machineries in the form of a single ADAR enzyme (i.e., dADAR in *Drosophila melanogaster*) or a single heterodimeric adenosine deaminase targeting mRNAs (*C. elegans* adr-1 and adr-2). They harbor single dsRBDs, and their catalytic domain sequence is closely related to those of vertebrate ADARs. Plants, fungi, and yeast lack ADAR enzymes and RNA-directed adenosine deaminase activity.

Overall, the ADAR enzymes are ubiquitously expressed in most, if not all, cell types (21, 23, 27). An exception is ADAR3, which is detected only in the central nervous system (28). The diversity of ADARs is further enhanced through the expression of alternative splice variants (82–84) and the utilization of alternative promoters (85) and even by A-to-I RNA editing. Mammalian ADAR2 as well as *Drosophila* dADAR edit their own pre-mRNAs (31, 86). In the case of ADAR2, intronic editing creates a novel splice site that leads to expression of a truncated, nonfunctional enzyme (86). In dADAR, site-selective editing alters an amino acid (S-to-G substitution) that was shown to modulate enzymatic activity (31). Alternative promoter usage creates two main ADAR1 transcript variants that differ substantially with respect to subcellular localization and function (87). ADAR1-p110 is expressed from a constitutive promoter and is active primarily within the nucleus (85). Expression of ADAR1-p150 is driven by an interferon inducible promoter and generates an enzyme that harbors a unique N-terminal domain, which conveys specific binding affinity to DNA and RNA in Z-conformation (88) and is actively shuttled between the nucleus and cytoplasm (89). In contrast, both ADAR1-p110 and ADAR2 are

dynamically associated with the nucleolus and might relocalize to the nucleoplasma when substrates for editing are expressed (90, 91). Moreover, ADAR2 exists in alternative splice forms that might differ in their RNA editing efficiencies or specificities (83, 84).

1.2 MAIN QUESTIONS IN THE FIELD AND APPROACHES

A critical aspect of understanding the overall impact of A-to-I RNA editing on the regulation of gene expression and for transcriptome and proteome diversity is to delineate all A-to-I editing targets followed by characterization of the functional consequences of editing. Until recently, only a relatively small number of ADAR substrates were known (4) and the A-to-I RNA editing targets seemed to be largely brain-specific. However, the editing machinery was shown to be functionally expressed in many cell types (21, 23, 27), indicating potential additional targets for these enzymes in other tissues. In fact, increasing evidence from biochemical studies further indicated that A-to-I editing occurs in other tissues as well (92) and that overall many more edited genes should exist (see Section 1.2.1).

Previously, ADAR substrates were discovered mainly by coincidence when comparing the cDNA sequence of a cloned gene to their genomic counterparts, in which an adenosine in the genomic sequence appeared as guanosine in the cDNA molecule. In particular, cases with high editing extents, such as the glutamate receptor GluR-2 Q/R site (>99.9% editing), have a higher chance of being detected, also enhanced by the fact that the ensuing amino acid substitution involves a functionally important residue. However, in recent years, several distinct and partially complementary screening methods have been developed to detect novel ADAR substrates.

1.2.1 Biochemical Versus Computational Approaches

The first technique that was developed to systematically screen for RNA molecules that have been modified by the A-to-I editing machinery involves a biochemical procedure to isolate inosine-containing mRNA molecules (93). In this approach, the inosine-containing RNA molecules are chemically modified and then preferentially cleaved at the phosphodiester bond 3′ to the inosine nucleotide. The reaction products are subsequently cloned and sequenced followed by analysis of full lengths cDNAs spanning the identified cleavage sites. This method was successful in that it led to the identification of initially five new cases of A-to-I editing in *C. elegans* mRNAs (94). Interestingly, all detected editing sites were present in the noncoding region of the RNA molecules, and this study therefore gave a first hint that editing might have additional functions besides the alteration of specific codons. In a follow-up study, five additional substrates in *C. elegans* were discovered, as well as 19 novel editing sites in transcripts from human brain (95). Similar to the previous study, all detected editing events were within noncoding regions and the editing of 15 out of the 19 discovered human RNA substrates occurred within transposon derived repeat elements. This finding led to the speculation that repetitive sequences may generally be frequent targets of the editing machinery.

The isolation and cloning of inosine-containing mRNAs using this technique is relatively laborious, and a high background of false positives makes it impractical to use for the comprehensive identification of all editing targets (94, 95).

The recent availability of complete genome sequences and annotations, including the human genome, has made it possible to conduct specific database searches to identify edited genes (see Figure 1.2 for schematic overview of screening methods to identify editing targets). The "smoking gun" of A-to-I RNA editing is an A/G discrepancy between a gene's cDNA sequence and its genomic counterpart. However, this feature is not sufficient to distinguish genuine RNA editing sites from A/G discrepancies that are the result of a single-nucleotide polymorphism (SNP), a sequencing error in the database, or a mutation introduced during cDNA cloning. Several laboratories embarked on the genome-wide computational screening for A/G discrepancies and through application of statistical methods were able to show that human Alu-type repeat elements present in mRNA sequences are a major target for A-to-I RNA editing (65, 96–98). Alu repeats are approximately 300 bp in length and consist of two monomers linked by an A-rich region. These repeats are highly abundant within the human genome and can occur both in forward and reverse orientation (for review see reference 99). If present within the same mRNA molecule, two oppositely oriented Alu elements can form a partially double-stranded structure generating a substrate for ADAR enzymes. In fact, when validating candidate Alu elements for editing *in vivo*, it turned out that if the two interacting repeat elements are within a few kilobases of one another, the occurrence of editing could always be

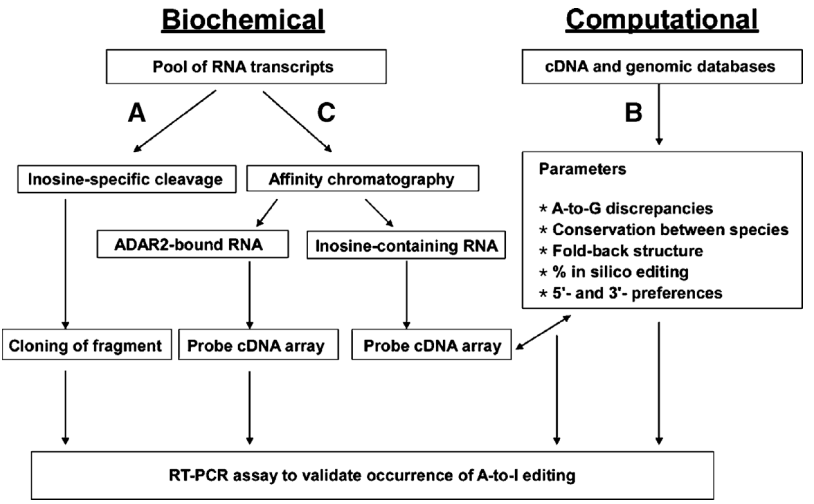

Figure 1.2 Screening strategies for identification of recoding targets. Flowchart outlining experimental strategies for delineation of editing events that lead to recoding. Approaches that use biochemical methods for the initial selection of candidate sequences (either through inosine-specific cleavage protocol or affinity chromatography using either ADAR or inosine-specific antibodies) are shown on one side. Approaches that start out by computationally filtering likely editing candidates are indicated on the right. See text for discussion of individual discussions.

confirmed experimentally (65). Between the different studies, several thousand mRNAs with a total of more than 20,000 editing sites were annotated. The main reasons for the only moderate level of overlap between the identified target sets are the different databases used for analysis (for example, either including or excluding EST sequences) and different stringencies when filtering the datasets. Since exonic Alu elements are almost always located within the noncoding regions of an mRNA molecule, it will be intriguing to see if Alu repeat editing may influence the stability, processing, or transport of mRNA molecules. In a few cases it was shown that Alu-mediated editing can destroy or create predicted splice sites (65), which represents another example of how A-to-I editing may regulate gene expression. Because Alu repeat elements are only present in primate genomes, repeat element editing in rodents was shown to occur at a much lower overall rate. This raises the intriguing question of whether editing has a specific role in primate evolution (see Chapter 13 for in-depth review on RNA editing in Alu-type repeats).

Nevertheless, rodent genomes contain other types of repetitive elements, which differ in the sequence composition (100) but which also can give rise to RNA foldback structures and subsequent A-to-I editing. In one case this has been shown to regulate the expression of a mRNA in mouse (101) (see also Chapter 13).

Alu-repeat-mediated editing targets belong to a class of editing substrates that are characterized by low site-selectivity and multiple site modification and are almost all localized in nontranslated sequences of mRNAs or introns. Figure 1.3 depicts the spectrum of currently known targets of A-to-I editing according to site-selectivity of modification, total rate of editing (inosine content), double-strandedness of the RNA fold-back structure, and the prevalence among identified substrates. Recoding targets for editing are located at the very end of the spectrum with the highest site-selectivity, the lowest relative content of inosine per transcript molecule, and the lowest double-strandedness of the substrate RNA.

The many editing events identified in noncoding mRNA sequences and within introns could explain the high amount of inosine that has been detected in mammalian

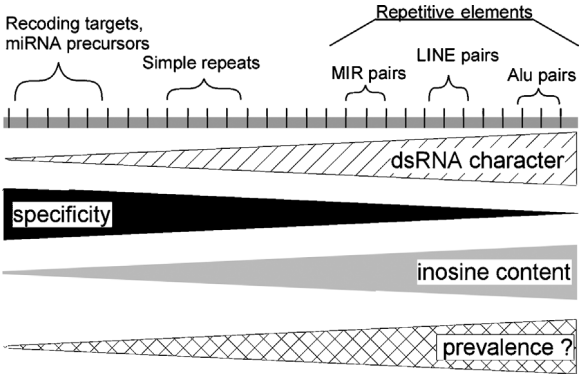

Figure 1.3 Spectrum of known A-to-I RNA editing targets. The known types of RNA editing targets are shown according to double-strandedness of the RNA fold directing editing, the site-selectivity of the editing event, total inosine content in the message, and the presumed prevalence relative to all known editing targets.

mRNA preparations (92). However, in light of the fact that rodents lack Alu elements, which constitute the main target of the RNA editing machinery in primates, a significant portion of the existing A-to-I editing events may still await discovery. If so, then site-selective targets of recoding may constitute a major fraction of these missing substrates.

Because of the predominance of inosine-containing RNAs that are edited in Alu repeats, experimental approaches to identify other sites of editing are challenging in primates. Ohlson et al. elegantly avoided this problem by using mouse brain tissue samples to selectively screen for ADAR2 specific substrates using affinity chromatography (102). For the detection of novel targets, ADAR2-specific antibodies were used to immunoprecipitate RNA molecules in complex with ADAR2. Reverse transcription and micro-array analysis of these co-precipitated RNA molecules resulted in the detection of up to 200 potential substrates (102), one of which, GABA(A) receptor subunit alpha3, has recently been verified *in vivo* (103).

In another approach to directly detect and isolate inosine-containing RNAs, an antibody was developed that selectively binds to inosine (104). Immunoprecipitated RNA molecules from wild-type and mutant ($ADAR^{-/-}$) flies were reverse transcribed and hybridized to a cDNA array. Comparison between wild-type and $ADAR^{-/-}$ cDNA array data led to the identification of 500 putative ADAR target genes (104). In addition, a database search was performed in which *Drosophila* cDNA sequences were compared to their genomic counterparts, as a genomically encoded adenosine will appear as guanosine in the expressed sequence after editing. This resulted in the detection of 800 genes that show an A/G discrepancy within the coding region. Ultimately, by comparing these two groups of putative editing targets, 62 genes were present in both groups. However, editing still has to be proven experimentally for most of these genes (104).

A prerequisite for editing is the presence of a double-stranded RNA structure. This suggests that the sequence surrounding the editing site may be conserved between species to preserve this RNA fold. Indeed, comparative genomics between 18 *Drosophila* species has demonstrated an almost complete absence of mutations in the close vicinity of the exonic edited sites (41). Using this knowledge, Hoopengardner et al. were able to discover 16 novel edited genes in *Drosophila* and one novel site in humans. Several different groups subsequently conducted database searches for finding novel edited human genes, taking into account this conservation between species as well as the established cis-sequence preferences of ADARs (39, 40). This led to the discovery of four novel human genes edited within the coding region.

Moreover, the total of known edited targets in *Drosophila* recently doubled by using *D. melanogaster* genomic and cDNA databases in a computational screen for A-to-G discrepancies, leading to the identification of 27 edited genes (38).

All of the above-described methods have proven to be effective for the discovery of novel edited genes. However, often already known substrates have been missed and there are a high number of false positives detected in each study. The methods for finding novel ADAR substrates still have to be optimized and developed further to allow for the identification of all recoding sites of editing.