# *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**

Copyright © 2008 by John Wiley & Sons, Inc. All rights reserved

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

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

***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 xv

ACKNOWLEDGMENTS xix

CONTRIBUTORS xxi

PART I DIVERSIFICATION OF THE PROTEOME THROUGH RNA AND DNA EDITING

CHAPTER 1 DIVERSIFYING EXON CODE THROUGH A-TO-I RNA EDITING 3

1.1 Introduction and Background 3
  1.1.1 Initial Discovery and Context of A-to-I RNA Editing and ADARs 4
  1.1.2 Important Cases of Recoding by A-to-I Modification in Pre-mRNA 5
  1.1.3 Cis-Acting Features for A-to-I Editing 14
  1.1.4 Properties of the A-to-I Editing Machinery 15
1.2 Main Questions in the Field and Approaches 17
  1.2.1 Biochemical Versus Computational Approaches 17
  1.2.2 Editing of miRNA Sequences 21
1.3 Future Directions: Evolution of Editing Sites and Machinery 23
  References 24

CHAPTER 2 ANTIBODY GENE DIVERSIFICATION BY AID-CATALYZED DNA EDITING 31

2.1 Introduction 31
2.2 Before AID 32
  2.2.1 Without DNA (Darkness) and with DNA (Light) 32
  2.2.2 Prominent Early Models for Antibody Diversification 32
  2.2.3 How Protein Sequencing Technology Enabled an Understanding of Antibody Diversity 34
  2.2.4 Somatic DNA Rearrangements Underpin V(D)J Joining and Create the Primary Antibody Repertoire 37
  2.2.5 Additional Antibody Diversity by Somatic Hypermutation (and Gene Conversion in Some Animals) 40
  2.2.6 Altering Antibody Function by Class Switch Recombination (Isotype Switching) 40
2.3 After AID 41
  2.3.1 A Novel Deaminase Is Required for CSR, SHM, and IGC 41

2.3.2 AID Is a DNA Cytosine Deaminase that Directly Triggers Antibody Diversification 43
2.3.3 The Importance of Uracil Bases in DNA *In Vivo* 45
2.3.4 Processing of AID-induced Lesions: The Molecular Mechanism of Somatic-Hypermutation 48
2.3.5 Processing of AID-induced Lesions: The Molecular Mechanism of Immunoglobulin Gene Conversion 49
2.3.6 Processing of AID-induced Lesions: The Molecular Mechanism of Class Switch Recombination 50
2.4 Hot Areas and Speculations 53
2.4.1 Immunodeficiency Syndromes Caused by Defects in AID-Mediated Ig Gene Diversification 53
2.4.2 Regulating the DNA Mutator Activity of AID 54
2.4.3 Misregulation of AID and Cancer 57
2.4.4 AID Is But One Member of a Much Larger Family of Polynucleotide Deaminases 58
2.5 Conclusions 60
Acknowledgments 60
References 61

CHAPTER 3 *PROTEIN–PROTEIN AND RNA–PROTEIN INTERACTIONS IN U-INSERTION/DELETION RNA EDITING COMPLEXES* 71

3.1 A Bizarre Phenomenon and its *Raison D'être* 71
3.2 The Catalytic Mechanism and Machinery 73
3.3 Extent of U-Insertion/Deletion RNA Editing in *Trypanosoma* and *Leishmania* Species 75
3.4 Functional Studies of Editing Complex Subunits 76
3.4.1 REN1, REN2, and MP67. Endonuclease Homologs 80
3.4.2 REX1 and REX2. Exonuclease Homologs 83
3.4.3 RET2. TUTase 85
3.4.4 REL1 and REL2. Ligase Homologs 85
3.4.5 MP81, MP63, MP42, MP46, MP44, MP24, MP18. Structural Components 87
3.5 RNA–Protein Interactions: Isolated Subunits and Assembled Editing Complexes 92
3.5.1 MP42 92
3.5.2 MP24 93
3.5.3 RNA–Protein Interactions in Assembled Editing Complexes 93
3.6 Concluding Remarks 96
Acknowledgments 96
References 96

CHAPTER 4 *MACHINERY OF RNA EDITING IN PLANT ORGANELLES* 99

4.1 Introduction 99
4.2 Mechanism of Target Recognition 100
4.3 PPR Protein is a *Trans*-Factor in Plastids 101
4.4 How Can the Model Be Generalized to Plant RNA Editing? 104
4.5 Can Closely Located Editing Sites Share a *Trans*-Factor? 105

4.6 Is a *Trans*-Factor Specific to a Single *Cis*-Element? 106
4.7 Mechanism Determining the Efficiency of RNA Editing 107
4.8 Co-Evolution of *Trans*-Factors and Editing Sites 109
4.9 What is an Editing Enzyme? 110
4.10 A Model of Editing Machinery in Plastids 112
4.11 Future Directions 114
Acknowledgments 114
References 114

PART II *FUNCTIONAL COORDINATION OF RNA EDITING WITH OTHER CELLULAR MECHANISMS*

CHAPTER 5 *TRANSFER RNA EDITING ENZYMES; AT THE CROSSROADS OF AFFINITY AND SPECIFICITY* 123

5.1 Introduction: Structural Versus Functional tRNA Editing 123
5.2 Transfer RNA Editing for Structure 124
5.2.1 C-to-U Editing of the tRNA Backbone 124
5.2.2 A-to-I Editing and Modification at Position 37 and 57 of tRNAs 126
5.3 Transfer RNA Editing for Function 127
5.3.1 The Lysidine Story 127
5.3.2 Nucleotide Additions at the Ends of tRNAs 130
5.3.3 C-to-U Editing in Marsupials and Trypanosomatids 133
5.3.4 A-to-I Editing of tRNAs in Yeast and Bacteria 137
5.3.5 Double Editing in Trypanosomatids 139
5.4 The Transfer RNA Editing Enzymes of Trypanosomatids: A Special Case of Catalytic Flexibility 140
5.5 Complex Formation by Transfer RNA Editing Enzymes: A Model for the Regulation of Editing Activity 141
5.6 Concluding Remarks: Evolution of Transfer RNA Editing Deaminases: Affinity Versus Specificity 142
References 143

CHAPTER 6 *A-TO-I EDITING AS A CO-TRANSCRIPTIONAL RNA PROCESSING EVENT* 146

6.1 Introduction 146
6.1.1 Overview of Co-transcriptional Pre-mRNA Processing 147
6.1.2 Localization of the ADAR Proteins 148
6.1.3 A-to-I Editing as a Pre-mRNA Processing Event 149
6.2 Main Questions in the Field and Approaches 149
6.2.1 Why Are Edited Sites Often Situated Close to Exon/Intron Border? 149
6.2.2 The Potential of A-to-I Editing in Changing the Transcriptome 151
6.2.3 RNA Editing, the Influence on Pre-mRNA Splicing and Vice Versa 152
6.3 Can Editing Influence the Fate of a Messenger RNA in Other Ways? 156
6.3.1 Editing and Its Potential Effect on RNA Export 156
6.3.2 Editing as a Modulator of RNA Stability 156
6.3.3 Editing and Its Influence on Polyadenylation 157
6.4 Prospectives for Future Research 158
References 159

CHAPTER 7 *STUDYING AND WORKING WITH RIBONUCLEOPROTEINS THAT CATALYZE H/ACA GUIDED RNA MODIFICATION* 162

7.1 Introduction 162
7.2 Discovery of Complex (RNA-guided) Pseudouridine Synthases 164
7.3 Approaches and Challenges 164
7.4 RNP Reconstitution 166
7.5 Lessons from Archaeal H/ACA RNPs 167
7.6 Biogenesis of Eukaryotic H/ACA RNPs 168
7.7 Debate on Dyskeratosis Congenita 168
7.8 Importance and Future of H/ACA RNPs 169
Acknowledgments 170
References 171

CHAPTER 8 *FUNCTIONAL ROLES OF SPLICEOSOMAL snRNA MODIFICATIONS IN PRE-mRNA SPLICING* 175

8.1 Introduction 175
8.2 Modified Nucleotides in Spliceosomal snRNAs 176
8.3 Functional Analysis of Spliceosomal snRNA Modifications 179
8.4 Modified Nucleotides of U2 snRNA are Important for Pre-mRNA Splicing 180
8.5 U2 Modifications Contribute to snRNP Biogenesis and Spliceosome Assembly 182
8.6 Genetic Analysis of U2 Modification in Yeast 183
8.7 Cytotoxicity Associated with 5FU Treatment is a Result of Inhibition on Pseudouridylation and Splicing 184
8.8 Biophysical Analysis of U2 snRNA Modification 185
8.9 Concluding Remarks 186
References 186

CHAPTER 9 *A ROLE FOR A-TO-I EDITING IN GENE SILENCING* 190

9.1 Expression of Double-Stranded RNA in Cells 190
9.2 The Activity of ADAR in the Nucleus 191
9.3 Alternative Fates of Edited RNAs in the Nucleus 192
9.4 A Possible Connection Between RNA Editing and Gene Silencing 193
9.4.1 Heterochromatin 193
9.4.2 RNAi-Directed Heterochromatin Formation 194
9.4.3 Connections Between RNAi and dsRNA Editing 196
9.4.4 Vigilin 196
9.4.5 Recognition of RNA by Vigilins 197
9.4.6 The Vigilin Complex 198
9.5 A Model for the Nuclear Function of Vigilin 199
References 200

CHAPTER 10 *BIOLOGICAL IMPLICATIONS AND BROADER-RANGE FUNCTIONS FOR APOBEC-1 AND APOBEC-1 COMPLEMENTATION FACTOR (ACF)* 203

10.1 Overview 203

10.2 Background to Our Current Understanding of C-to-U Editing of ApoB mRNA: Canonical Functions for Apobec-1 and ACF 204
10.2.1 Role of Cis-Acting Elements 204
10.2.2 Identification and Characterization of *Trans*-Acting Factors 206
10.3 Current Understanding of Apobec-1 and ACF: Structure–Function and Genetic Regulation 209
10.3.1 Apobec-1: Structure–Function Relationships 209
10.3.2 Functions of Apobec-1 Beyond apoB mRNA Editing 210
10.3.3 Apobec-1: Genetic Regulation and Gain- and Loss-of-Function 212
10.3.4 ACF: Structure–Function Relationships 214
10.3.5 Intersections of Apobec-1 and ACF Regulation in the Modulation of C-to-U RNA Editing 216
10.4 Implications and Broad-Range Function for Apobec-1 and ACF: Future Directions and Overarching Questions 218
10.4.1 Apobec-1 218
10.4.2 ACF 220
10.5 Conclusions 224
Acknowledgments 224
References 224

CHAPTER 11 *ANTIVIRAL FUNCTION OF APOBEC3 CYTIDINE DEAMINASES* 231

11.1 Explanation of Vif Phenotype Uncovers a Unique Innate Resistance to HIV-1 Infection 231
11.2 Antiviral Functionality of the APOBEC3 Family of Proteins 232
11.2.1 Mechanism of Action 232
11.2.2 APOBEC3 Proteins and the Prevention of Zoonosis 235
11.2.3 *In Vivo* Correlations Between APOBEC3 Expression and Disease Course 236
11.3 The Battle for Control: Viral Suppression of the APOBEC3 Proteins 237
11.3.1 The Hijacking of the Proteasomal Degradation Pathways 237
11.3.2 Virion Exclusion of the APOBECs via a Viral-Dependent Mechanism 238
11.4 Cellular Function and Regulation of the APOBEC3 Family 238
11.4.1 Guardians of the Genome: APOBEC3-Mediated Suppression of Cellular Retroelements 238
11.4.2 Subcellular Localization (Sequestration) 239
11.4.3 Control of the Expression of the APOBEC3 Family is Exerted Transcriptionally 240
11.5 Research Questions and the Hope of Therapeutic Manipulation of the APOBEC3 Family 243
11.5.1 The "Alternative Function" 244
11.5.2 Protein Partitioning/Subcellular Localization 245
11.5.3 Protein Cofactors and Posttranslational Modifications 245
11.5.4 Therapeutic Potential 246
References 247

PART III *PREDICTIVE STRUCTURES*

CHAPTER 12 *A-TO-I EDITING OF ALU REPEATS* 257

12.1 Background 257
12.1.1 Indirect Evidence for Abundant A-to-I Editing 257

12.1.2 Early Screens for A-to-I Editing Targets 258
12.1.3 The *Alu* Repeats 259
12.2 Computational Detection of A-to-I Editing 260
12.2.1 Sifting through db EST 260
12.2.2 Clusters of Mismatches in RNAs 262
12.2.3 Numbers of Editing Sites Detected 265
12.2.4 Characterization of the Edited Transcripts 265
12.2.5 Looking for Conserved Polymorphism Sites 267
12.2.6 Additional Potential Targets for Abundant A-to-I Editing 267
12.3 Editing in Other Organisms 268
12.3.1 Uniqueness of the *Alu* Repeat 269
12.3.2 Predicting Editing Sites from Genomic Data 270
12.4 Biological Role of *Alu* Editing 272
12.4.1 Possible Regulatory Roles 272
12.4.2 *Alu* Editing and miRNA 272
12.4.3 Alternative Splicing and *Alu* Editing 273
12.5 Concluding Remarks 275
References 276

CHAPTER 13 *RNA EDITING IN DINOFLAGELLATES AND ITS IMPLICATIONS FOR THE EVOLUTIONARY HISTORY OF THE EDITING MACHINERY* 280

13.1 Introduction 280
13.2 Inferred RNA Editing in Dinoflagellates 283
13.2.1 *cob* and *cox1* mRNA 283
13.2.2 Chloroplast Transcripts 288
13.3 Biochemical Characteristics of Editing 289
13.3.1 Unusually Diverse Types of Editing 289
13.3.2 Varying Editing Density and Discrete Distribution of Editing Events in Coding Sequences 293
13.3.3 Markedly Nonrandom Distribution in the Type of Codon Edited and in the Position of the Editing Site Within Codons 294
13.4 Consequences of Editing 297
13.5 Phylogenetic Trend 301
13.6 Implications for the Origin and Evolution of the RNA Editing Machinery 304
Acknowledgment 306
References 306

PART IV *STRUCTURAL APPROACHES*

CHAPTER 14 *THE BOX C/D RNPs: EVOLUTIONARILY ANCIENT NUCLEOTIDE MODIFICATION COMPLEXES* 313

14.1 Introduction 313
14.2 Diversity of Box C/D RNA Populations 314
14.2.1 Box C/D RNA Nomenclature 314
14.2.2 Box C/D RNA Structure 315

14.2.3 Diversity of Box C/D RNA Populations 316
14.2.4 Box C/D RNA Identification 316
14.3 Box C/D RNA Functions and Target RNAs 319
14.3.1 Folding and Cleavage of Pre-rRNA 319
14.3.2 2′-*O*-Methylation of Diverse RNA Targets 319
14.3.3 Additional Roles and Targets for Box C/D RNAs 320
14.4 Box C/D RNP Structure and Nucleotide Methylation Function 321
14.4.1 Eukaryotic Box C/D Core Proteins and snoRNP Structure 321
14.4.2 Archaeal Box C/D Core Proteins and *In Vitro* sRNP Assembly 322
14.4.3 Emerging Core Protein and RNP Crystal Structures 322
14.4.4 Investigating Methylation Function Using *In Vitro* Assembled Archaeal Box C/D sRNP 323
14.5 Box C/D RNP Biogenesis 323
14.5.1 Genomic Organization of Eukaryotic Box C/D snoRNA Genes 323
14.5.2 Independently Transcribed and Intronic Eukaryotic Box C/D snoRNA Genes 324
14.5.3 Archaeal Box C/D sRNA Genes 324
14.5.4 Transcription and Processing of Independently Transcribed Box C/D snoRNAs 325
14.5.5 Transcription and Processing of Intronic Box C/D snoRNAs 326
14.5.6 Box C/D snoRNP Transport 327
14.6 Future Directions and Experimental Challenges 327
14.6.1 Box C/D RNA Diversity, Targets, and Functions 327
14.6.2 Box C/D RNP Structure and Methylation Function 328
14.6.3 Box C/D RNP Biogenesis 330
Acknowledgments 331
References 331

CHAPTER 15 *STRUCTURAL FEATURES OF THE ADAR FAMILY OF ENZYMES AND THEIR SUBSTRATES* 340

15.1 ADAR Enzymes 340
15.2 Overview and Functions of ADARs 341
15.2.1 Double-Stranded RNA Binding Domains (dsRBDs) 344
15.2.2 *Xenopus laevis* XIrbpa 345
15.2.3 ADAR2 dsRBD1 and dsRBD2 346
15.2.4 Deaminase Domain 346
15.2.5 Zα and Zβ Domains 347
15.2.6 Zα Structure 349
15.2.7 Zβ Structure 351
15.2.8 Structural Comparison of Zα and Zβ 354
15.3 Conclusions 354
15.4 Substrates 355
15.4.1 Overview and General Features 355
15.5 Double-Stranded RNA Targets and Structural Features 355
15.5.1 Site-Selective A-to-I Editing 355
15.5.2 Structural Features of Site-Selective A-to-I Editing 358
15.5.3 Promiscuous Editing 359

15.6 Single-stranded RNA Targets 361
15.7 Z-DNA and Z-RNA Targets 361
15.8 Future Directions 362
References 362

CHAPTER 16 *CHEMISTRY, PHYLOGENY, AND THREE-DIMENSIONAL STRUCTURE OF THE APOBEC PROTEIN FAMILY* 369

16.1 Introduction to Nucleic Acid Deamination with Implications for Biological Activity 369
16.2 The Chemistry of the Zinc-Dependent Deaminase Amino Acid Signature Motif 370
16.3 The ZDD Signature Motif Implies a Specific Three-Dimensional Arrangement of Amino Acids 371
16.4 Rationale for a Combined Structural and Phylogenetic Approach to Understand APOBEC Evolution 373
16.4.1 The Starting Point for Structural and Phylogenetic Analyses of APOBEC Family Members 375
16.4.2 The CDA Superfamily: Overview of Conserved Fold Topology in the Core and Common Variations 379
16.4.3 Comparison of the Common CDA Superfamily Core Reveals Broad Peripheral Diversification 381
16.5 Modes of Oligomerization 382
16.5.1 Free Nucleotide Cytidine Deaminases (CDA): Strand β5 Antiparallel to Strand β4 382
16.5.2 Cytosine Deaminase, Guanine Deaminase, and TadA: Strand β5 Parallel to β4 382
16.5.3 Deoxcytidylate Deaminases (T4, *N. e*) and APOBEC2: Strand β5 Parallel to β4 383
16.5.4 Multidomain Enzymes RibG and ADAR2 of the CDA Superfamily: Strand β5 Parallel to β4 387
16.6 Modes of Substrate Interaction 388
16.6.1 Tetrameric fnCDAs Favor Flexible Flaps: RNA Editing and the Case of Cdd1 from Yeast 390
16.6.2 A Topological Transformation Obstructs Active Site Accessibility in Dimeric Deaminases that Bind Bases 391
16.6.3 Substrate Selection by Polynucleotide Editing Enzymes Remains Elusive 391
16.7 The APOBEC Family: Insights into a Structurally Underrepresented Family 392
16.7.1 Activation-Induced Deaminase (AID)—an Ancient Enzyme with Essential Roles in Adaptive Immunity 395
16.7.2 APOBEC2—A Divergent Ancestral Protein of Unknown Function 396
16.7.3 APOBEC1—The Historical Archetype of C-to-U Editing Enzymes 397
16.7.4 APOBEC4—Pushing the Envelope of APOBEC Boundaries 399
16.8 APOBEC3—Radiative Expansion of Proteins Involved in Viral Defense 400
16.8.1 Mechanisms of Primate-Specific Expansion of the APOBEC3 Proteins 402
16.8.2 Alternative Methods to Obtain Structure: The Molecular Envelope of APOBEC3G by Small-Angle X-Ray Scattering 405
16.8.3 Characterization of Structural Changes in APOBEC3G Morphology in the Presence of RNA 406

16.8.4 Positive Selection Exerted on the APOBEC3 Family 408
16.9 Conclusions and Future Prospects 410
References 411

*INDEX* 421

# 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
Copyright © 2008 John Wiley & Sons, Inc.

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<br>K1455R<br>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<br>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