RNA INTERFERENCE

Application to Drug Discovery and Challenges to Pharmaceutical Development

Edited by

PAUL H. JOHNSON
PhaseRx, Inc.
Seattle, Washington

A JOHN WILEY & SONS, INC., PUBLICATION

RNA INTERFERENCE

RNA INTERFERENCE

Application to Drug Discovery and Challenges to Pharmaceutical Development

Edited by

PAUL H. JOHNSON
PhaseRx, Inc.
Seattle, Washington

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 interference : application to drug discovery and challenges to pharmaceutical development / edited by Paul H. Johnson.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-471-77151-7 (cloth)
1. RNA. 2. Small interfering RNA. 3. Drug development. I. Johnson, Paul H., 1943-
[DNLM: 1. RNA Interference. 2. Drug Discovery. QU 475]
QP623.R573 2011
572.8′8–dc22

2010030999

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTENTS

CONTRIBUTORS

ROGER ADAMI, mdRNA, Inc., 3830 Monte Villa Parkway, Bothell, WA 98021

MARK BEHLKE, Integrated DNA Technologies, Inc., 1710 Commercial Park, Coralville, IA 52241

RODERICK L. BEIJERSBERGEN, The Netherlands Cancer Institute, Division of Molecular Carcinogenesis, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

JAMES A. BIRCHLER, Division of Biological Sciences, Tucker Hall, University of Missouri, Columbia, Missouri 65211

KUNYUAN CUI, Agave Pharma, Inc., 1124 Columbia St., Seattle, WA 98104

DIANE FRANK, AVI BioPharma, Inc., 3450 Monte Villa Parkway Suite 101, Bothell, WA 98021

YANLIN GUO, Department of Biological Sciences, University of Southern Mississippi, 118 College Dr. #5018, Hattiesburg, MS 39406

ADAM N. HARRIS, Life Technologies Corporation, 1600 Faraday Avenue, Carlsbad, CA 92008

FAQING HUANG, Department of Chemistry and Biochemistry, University of Southern Mississippi, 118 College Dr. #5043, Hattiesburg, MS 39402

MICHAEL HOUSTON, mdRNA, Inc., 3830 Monte Villa Parkway, Bothell, WA 98021

PAUL H. JOHNSON, PhaseRx, Inc., 410 West Harrison St., Suite 300, Seattle, WA 98110

ADAM JUDGE, Tekmira Pharmaceuticals Corporation, 100-8900 Glenlyon Parkway, Burnaby, BC, Canada V5J 5J8

IAN MACLACHLAN, Tekmira Pharmaceuticals Corporation, 100-8900 Glenlyon Parkway, Burnaby, BC, Canada V5J 5J8

STEVEN C. QUAY, Atossa Genetics, Inc., 4105 E Madison St., Suite 320, Seattle, WA 98112

MARJORIE ROBBINS, Tekmira Pharmaceuticals Corporation, 100-8900 Glenlyon Parkway, Burnaby, BC, Canada V5J 5J8

ALIASGER K. SALEM, Division of Pharmaceutics, College of Pharmacy, University of Iowa, Iowa City, IA 52242

OLIVER C. STEINBACH, Department Bio-Molecular Engineering, Sector Molecular Medicine, Philips Research Laboratories, High Tech Campus 11, NL-5656 AE Eindhoven, The Netherlands

STEPHEN MARK TOMPKINS, Department of Infectious Diseases, 111 Carlton St., University of Georgia, Athens, GA 30602

RALPH TRIPP, University of Georgia, Department of Infectious Diseases, 111 Carlton St., Athens, GA 30603

HARRY WANG, mdRNA, Inc., 3830 Monte Villa Parkway, Bothell, WA 98021

KRISTIN A. WIEDERHOLT, Life Technologies Corporation, 1600 Faraday Avenue, Carlsbad, CA 92008

C. J. YU, Beckman Institute, M/C 139-74, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125

PREFACE

RNA interference (RNAi) is a form of posttranscriptional gene silencing within cells involved in the control of gene expression. The RNAi pathway is found in many eukaryotes including animals and humans. It is initiated by the enzyme Dicer, which cleaves long double-stranded RNA molecules into short fragments of about 20 base pairs. The products of this reaction, small interfering RNA (siRNA) and microRNA (miRNA) molecules, bind to other homologous RNAs (transcripts) and affect their activity (protein expression) *via* degradation or translational inhibition. RNAi functions in defending cells against parasitic genes—viruses and transposons—and also in directing development as well as gene expression in general. The specific and potent effect of RNAi on gene expression makes it a valuable research tool in both cell culture and animal models to systematically shut down each gene in the cell to help identify the components necessary for a particular function/cellular process. RNAi may also be exploited in humans by introducing synthetic siRNA into cells and suppression (silencing) of disease-causing genes.

Exploitation of the RNAi pathway is a promising approach to treat a variety of diseases. Of great importance is the current view that 80% of the genome is not drugable by small molecules (10%) or biologics (10%), but is potentially drugable by gene silencing technologies. RNAi is seen as a promising way to treat cancer by silencing genes differentially upregulated in tumor cells or genes involved in cell division. Other proposed clinical applications include antiviral therapies to treat infection by herpes simplex virus type 2 and the inhibition of viral gene expression in cancerous cells, knockdown of host

receptors and coreceptors for HIV, the silencing of hepatitis A and hepatitis B genes, silencing of influenza gene expression, and inhibition of measles viral replication. Potential treatments for neurodegenerative diseases include muscular dystrophy and polyglutamine diseases such as Huntington's disease.

A key area of research in the use of RNAi for clinical applications is the development of safe and effective delivery systems that permit targeting to specific cell types and tissues with efficient cell uptake and release into the cytoplasm, the site of action. The goal of this book is to bring together a series of critical review and analysis chapters by leading scientists in the RNAi field that assess the key issues in the development of RNAi-based drugs for clinical applications.

The first section of the book covers the biology of RNAi. Chapter 1 discusses the origins and overview of RNAi. Chapter 2 describes nucleic acids as regulatory molecules including the artificial modulation of gene expression using antisense technology, triplex-forming oligonucleotides, nucleic acid decoys, aptamers, ribozymes and DNAzymes, and RNAi by siRNAs and miRNAs. Chapter 3 examines the use of siRNA oligonucleotides to study gene function including siRNA design strategies, target specificity, chemical modification, delivering siRNA in cell culture, and the experimental design and detection of gene silencing (knockdown). Chapter 4 explains genome scanning by RNAi, dissection of physiological and pathological processes with genetic screens, large-scale RNAi-based screens in mammalian cells, and the design and practical implementation of high-throughput RNAi screens.

The second section of the book deals with the development of siRNA for therapeutic applications. Chapter 5 discusses the discovery of siRNA delivery agents, focusing on issues and challenges related to cellular uptake mechanisms and tight junction dynamics, peptide-based delivery, targeting specific cell types, and high-throughput screening approaches. Chapter 6 describes the potential for use of delivery systems for synthetic siRNAs to overcome the limitations of siRNAs and enhance therapeutic efficacy including current limitations to delivery of siRNAs *in vivo*, interferon induction and the innate immune system, siRNA administration without delivery assistance, nonviral delivery vehicles for siRNAs, physical delivery, synthetic siRNA delivery systems, pegylation of siRNA delivery vehicles, cell targeting ligand conjugation to siRNA delivery vehicles, and the rational design of modular multicomponent siRNA delivery systems. Chapter 7 focuses on immunologically based *in vivo* toxicities including the mechanisms of nucleic acid-mediated immune stimulation, Toll-like receptors, TLR-independent mechanisms and protein kinase R, factors influencing immune stimulation, and the implications for pharmaceutical product development. Chapter 8 discusses synthetic siRNA drug development and therapeutic applications for respiratory syncytial virus and influenza viruses and describes the application of RNAi to viral diseases

focusing on proofs of concept for therapeutic RNAi treatment of virus infection, viral countermeasures for RNAi, translating siRNA delivery to the clinic, and RNAi versus traditional antiviral drugs.

Currently, there are more than 80 gene silencing therapeutics in clinical trials (http://clinicaltrials.gov/). They have the potential to provide a new class of broadly applicable therapies with clinically relevant efficacy and safety. Efficient delivery and cost-effective manufacturing are major challenges that must be met to achieve success. While current manufacturing processes are not yet cost-effective, they are likely to show significant improvement over time. No current delivery system appears to have the desired efficiency, cell selectivity, and safety profile necessary for an RNAi therapeutic to have a clear clinical advantage over other classes of drugs. However, on the basis of recent results and promising new systems under development, there is reason to believe that this will change in the near future.

RNAi therapeutics, a new class of drugs with unprecedented specificity, efficacy and safety, have the potential to revolutionize drug development and the future of medicine by providing broadly applicable therapies against targets that are currently undrugable.

PAUL H. JOHNSON

Seattle, WA
July 2010

FOREWORD

RNA interference (RNAi) has rapidly progressed from an intriguing scientific discovery to a powerful tool for studying gene function. In addition to the use of RNAi in the research laboratory, many investigators and commercial ventures are exploiting its potential as a sequence and gene-specific therapeutic agent. The idea that an oligonucleotide could be used to block gene function by virtue of Watson–Crick base pairing is not new, but was described over three decades ago by Paul Zamecnik and colleagues. The original antisense oligonucleotides were largely composed of DNA with various backbone modifications to stabilize the oligos or alter their Watson–Crick binding stability. For various reasons, with a few minor exceptions, the potential of antisense oligos to serve as functional genomic tools or therapeutic agents has never been fully realized. Since RNAi is in effect a form of antisense, why is it so much more potent than the conventional oligonucleotide approaches? Perhaps the biggest difference between RNAi and the antisense DNAs is that RNAi engages a specific set of cellular proteins that have evolved over millions of years for regulating gene expression. In contrast, antisense DNA approaches relied upon diffusion of the oligo to the target mRNA sequence wherein RNAseH is recruited to cleave the RNA. The two phenomena are similar in that the major mammalian Argonaute protein effecter of RNAi, Ago2, has an RNAseH-like cleavage domain at the active site. The association of Ago2 with other components of the RNAi machinery provides efficient target recognition and results in target destruction following Ago2-mediated cleavage. Thus, it is fair to state that RNAi is the most powerful sequence

specific inhibitor of gene expression available for both functional genomics and therapeutic applications.

The chapters in this volume cover a large swath of the applications and challenges of using RNAi in drug discovery and as a therapeutic modality. A couple of chapters review the use of siRNAs for therapeutic target identification and gene function analyses. There are several chapters that focus on strategies and issues of concern for using siRNAs as therapeutic agents. Of particular importance are strategies for delivery of therapeutic small RNAs. A number of clever approaches ranging from conjugation of ligands to siRNAs through encapsulation of siRNAs in nanoparticles have been published in the literature. Many of these are reviewed in the chapters of this volume. There are also applications in which siRNAs are not packaged, but backbone modified versions are applied directly, such as intraocular injection for macular degeneration treatment, or inhalation for treatment of RSV or influenza infections. Other chapters address the major concerns of systemic delivery, which are rapid clearance through the kidney and liver versus desired uptake into tumors or other tissues. Some of the delivery strategies and vehicles that are being developed can circumvent rapid kidney clearance, but the liver still remains as a major receptacle for systemic RNAs, even when the siRNAs are taken up by other tissues or tumors. That being said, there is a need to ensure that the siRNAs are not creating toxicities by triggering downregulation of targeted mRNAs in tissues other than those that are diseased. Other concerns are off-target effects caused by siRNAs functioning on nontargeted mRNAs to block translation or even direct cleavage. Thus, bio-informatics' analyses of siRNA design along with backbone modifications to block miRNA-like function or preventing passenger strand activity need to be carefully evaluated. Additional off-targeting is achieved by siRNAs engaging Toll-like receptors in immune cells, and triggering of interferon responses. A chapter is devoted to this problem with recommended strategies to abrogate interferon responses.

Overall, this volume covers the major areas of siRNA application to date. The authors of each chapter are experts in the respective topic area of the chapter, so the reader should gain valuable insight and information about the promises and pitfalls of RNAi. It is rather amazing that only 12 years ago the scientific world had no clue that such a powerful, sequence specific mechanism of gene regulation existed, yet already many clinical trials are well underway employing RNAi triggers. This volume captures the excitement and momentum of the field of siRNA applications, and should serve a useful purpose in years to come.

JOHN J. ROSSI

Duarte, CA
July 2010

PART I

BIOLOGY OF RNA INTERFERENCE

CHAPTER 1

RNA INTERFERENCE: WHAT IS IT?

JAMES A. BIRCHLER

WHAT IS RNAi?

In its most simple incarnation, it is the technique in which double-stranded RNA (dsRNA) is used to target the destruction of the homologous messenger RNA (mRNA). The first indication that such was the case involved studies of the use of antisense RNA to block the translation of specific proteins in the nematode *Caenorhabditis elegans* [1]. The concept at the time was that the introduction of antisense RNA would show base pair complementarity to the mRNA and block the progression of translation. This technique had shown some level of success in a variety of species. Interestingly, with the use of a control of the sense RNA, the same level of inhibition of protein synthesis was achieved as using the antisense. Thus, the technique of RNA interference (RNAi) was born.

Shortly thereafter, the basis of the effectiveness of both sense and antisense RNAs for RNA silencing became known [2]. This seminal contribution established that the active ingredient in RNAi was indeed a dsRNA and that the direct usage of dsRNA was very effective in eliminating the homologous mRNA. This realization ignited great interest in the use of RNAi as a reverse genetic technique throughout eukaryotic organisms.

While the technique of RNAi caused a revolution in genetics, we should drop back and review a body of work that preceded its development and that we now understand is related in mechanism. This work involves gene silencing by the introduction of transgenes into various plant species. The first example involved antibiotic resistance genes transformed into tobacco [3]. When one transgene was introduced its expression was robust. A second antibiotic resistance gene was transformed and found to be well expressed. Both

RNA Interference: Application to Drug Discovery and Challenges to Pharmaceutical Development, Edited by Paul H. Johnson

types used the same promoter. When the two transgenes were crossed together into the same plants, they both became inactive. Subsequent outcrossing to separate the two types of transgenes resulted in the recovery of the activity of both in the following generation. This type of transgene silencing was established to work at the transcriptional level.

At about the same time, experiments were underway in two laboratories to attempt to make petunia flowers darker than normal by adding an extra copy of the chalcone synthase gene, which encodes the first step of anthocyanin pigment biosynthesis, by transformation of petunias [4, 5]. However, the result found was that the extra copy of the chalcone synthase gene produced white flowers rather than darker ones. This phenomenon was referred to as cosuppression and was found to operate at the posttranscriptional level.

As the literature developed about cosuppression and additional examples were described, it was realized that a similar mechanism was operative against many plant viruses. Such viruses typically have an RNA genome and it was demonstrated that transgenes expressing a homologous RNA would serve to target the viral RNA for destruction [6].

Thus, the concept arose that this process was a defense mechanism against transposable elements and viruses [7]. Indeed, the virulence genes of plant viruses were found to inhibit the process of posttranscriptional silencing [8]. This fact strengthens the argument that host cells use this silencing mechanism against viruses and that viruses have evolved ways to circumvent such silencing. RNAi as a viral defense has also been observed in the animal kingdom [9].

After the discovery that RNAi had a basis in dsRNA, elegant experiments were performed that produced dsRNA homologous to promoters of the original transcriptionally silenced antibiotic resistance genes [10]. The formation of these RNAs was effective in silencing the target transgene. However, when separated during meiotic segregation, the transgene could regain its activity. The demonstration that dsRNA was involved with transcriptional transgene silencing drew a connection between the two types of silencing. If aberrant RNAs were produced with homology to the promoter, then transcriptional silencing would result whereas aberrant RNAs with homology to mRNAs would produce posttranscriptional silencing.

Transgene cosuppression was found in *Drosophila*, which extended this type of silencing to the animal kingdom [11, 12]. A hybrid transgene with the promoter from the *white* eye color gene was fused to the structural portion of the *Alcohol dehydrogenase* gene showed successively less expression with increased copy number. In this case, the silencing was not as strong as usually found in plants but was progressively stronger with increased number of transgenes. The transgenes that were silenced became associated

with the Polycomb repressive complex of chromatin proteins implying a transcriptional level silencing, which was later directly confirmed.

The silencing of the *Drosophila* I retroelement, which is responsible for one type of hybrid dysgenesis, possesses many characteristics of a similar mechanism to cosuppresion in that the silencing is homology dependent [13, 14]. Transgenes of a portion of the element are capable of silencing all copies in the genome. The ability to silence is transmitted only through the maternal parent to the progeny—a characteristic that is similar to the process of hybrid dysgenesis.

The posttranscriptional basis of cosuppression suggested that an RNA moiety was involved in recognizing the homologous RNAs for destruction. A search for the entity involved led to the discovery of very small RNAs that were a mere 21–23 base pairs (bp) in length [15]. This discovery together with a developing literature about RNAi inspired biochemical studies of the molecular mechanism.

In *C. elegans* and plants, the RNAi process can spread through the organism in a systemic manner [16,17]. The small RNAs likely act as primers to serve as a substrate with the homologous mRNA for RNA-dependent RNA polymerase to generate additional quantities of dsRNAs that then are acted upon by Dicer. This forms a self-perpetuating cascade that can continue via spread through the organism and at least in *C. elegans* into the next generation [18]. Such systemic spread has not been observed in *Drosophila* or mammals [19].

Much of the biochemistry of RNAi has been performed using *Drosophila* extracts. The enzyme that catalyzes the conversion of dsRNAs to single-stranded short interfering RNAs was sought. This protein was identified and was referred to as Dicer, a ribonuclease III type enzyme [20]. Further analysis of the machinery involved led to the description of the RNA interference silencing complex, which incorporates the small RNAs as guides to target the destruction of homologous mRNAs [21–27]. Parallel studies identified several genes that were required for RNAi processes, which include several RNA helicases and members of the Argonaute family of proteins. The Argonaute family of proteins supplies the "slicer function" to cleave the mRNA [28].

Further connections between the small RNA silencing machinery and transcriptional silencing were observed in *Drosophila* and fission yeast. The Polycomb-dependent transgene gene silencing in flies was established to act on the transcriptional level via run-on transcription assays [29]. A separate type of transgene silencing was also described for *Drosophila* that was post-transcriptional, namely, for a dosage series of the full-length *Alcohol dehydrogenase* gene [29]. Both types of silencing were blocked by the piwi mutation, which encodes an Argonaute family protein. The transcriptional silencing was also partially blocked by another Argonaute mutation, aubergine.

In fission yeast, the centromere repeats are silenced by the RNAi machinery [30]. In this species, there is only a single Dicer, RNA-dependent RNA polymerase and Argonaute, which facilitated the analysis. The silencing machinery generates small RNAs, which attract the histone methyltransferase that methylates lysine 9 of histone 3 (H3-K9). This modification serves to foster silenced chromatin by attracting Swi6, the yeast homologue of Heterochromatin Protein 1. In *Drosophila*, genes required for RNAi in embryos [31] suppress heterochromatic silencing and reduce the histone modifications of H3-K9 [32].

Transcriptional silencing in fission yeast is associated with a separate protein effector complex referred to as the RITS complex [33]. Furthermore, interaction with RNA polymerase II is implicated in that mutations in the largest and second largest subunit of polII disrupt the formation of small RNAs involved with centromeric silencing [34]. Indeed in plants, a separate RNA polymerase has evolved that is required for silencing functions [35–37].

While Dicer is involved with the production of small interfering RNAs (siRNAs), other classes of small RNAs, studied most thoroughly from the germline of flies and mammals, are generated in a Dicer-independent mechanism [38–40]. These RNAs are referred to as piRNAs because they are associated with the Argonaute family protein, piwi, or its mammalian homologues. The Argonaute family of proteins possess a "slicer" function capable of endonucleolytic cleavage of RNA [39, 40]. The piRNAs are slightly larger than siRNAs being about 24 bp in length. The piRNAs are heavily involved with the control of transposons in the germline but other classes of genes have also been found among the piRNA sequences [41–46]. Interestingly, there are "loci" in *Drosophila* that consist of retrotransposon fragments that are transcribed and feed into the metabolism of piRNAs that act to silence the homologous transposons [40].

While the piRNAs are presumed to operate posttranscriptionally and mainly in the germline, numerous reports of mutant effects on chromatin level phenomena have been made for the gene products involved with piRNAs. These include transcriptional transgene silencing [29], heterochromatic silencing [32], pairing sensitive silencing [47, 48], chromatin insulator function [49], nucleolar integrity [50], centromere function [51], and telomere chromatin [52]. The full understanding of the function of the Argonaute family and its involvement in Dicer-dependent and -independent small RNA biology is not yet understood.

In addition to the siRNA and piRNAs, a third major class of small RNAs are called microRNAs (miRNAs) [53–55]. These originate from endogenous loci that have a foldback structure interrupted by a spacer region. These small RNAs function to block the translation of mRNAs with which they share close but not identical homology [56]. In some cases, they serve to trigger

the destruction of the mRNA. Thus, they act as a modulation mechanism for gene expression that is posttranscriptional. In vertebrates, the miRNAs are generated by the same Dicer enzyme as siRNAs given that there is only one such enzyme encoded in the genome [57]. Because miRNAs are the prevalent small RNA in mammals, it was once thought that no endogenous siRNAs were produced, but recent deep sequencing projects have found them, thus illustrating an overlapping enzyme specificity. However, in *Drosophila*, there are two Dicer genes with a diverged preference for generating either miRNAs or siRNAs [58].

As the knowledge of small RNA silencing processes continues to grow the involvement in a variety of both posttranscriptional and transcriptional silencing (or activation) processes is increasingly evident. While "RNAi" was originally recognized as a powerful reverse genetic technique, its basis obviously rests on a biological phenomenon that the scientific community has yet to fully grasp. The interrelationship of posttranscriptional and transcriptional silencing mechanisms is far from clear and the occasional overlap of gene products involved with the generation of si, pi, and miRNAs suggests connections that are yet to be revealed [41–46]. The number of chromatin level processes affected by gene products involved with the generation and processing of small RNAs continues to expand suggesting that small RNAs might have quite prevalent roles in many mechanisms in the cell. The vast number of antisense transcripts also suggests a level of gene regulation that is yet to be fully understood [59]. Thus, the phenomenon of RNAi as a revolutionary genetic technique is likely a reflection of a much deeper and pervasive RNA biology of which we still have much to learn.

REFERENCES

1. Guo, S. and Kemphues, K. J. (1994). par-1, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**: 611–20.
2. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**: 806–11.
3. Matzke, M. A., Primig, M., Trnovsky, J., and Matzke, A. J. M. (1989). Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J* **8**: 643–9.
4. Napoli, C., Lemieux, C., and Jorgenson, R. (1990). Introduction of a chimeric chalcone synthase gene in Petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* **2**: 279–89.

5. Van Der Krol, A. R., Mur, L. A., Beld, M., Mol, J. N. M., and Stuitje, A. R. (1990). Flavanoid genes in Petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell* **2**: 291–9.
6. Lindbo, J.A., Silva-Rosales, L., Proebsting, W. M., and Dougherty, W. G. (1993). Induction of a highly specific antiviral state in transgenic plants: implications for regulation of gene expression and virus resistance. *Plant Cell* **5**: 1749–59.
7. Flavell, R. B. (1994). Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proc Natl Acad Sci USA* **91**: 3490–6.
8. Anandalakshmi, R., Pruss, G. J., Ge, X., Marathe, R., Mallory, A. C., Smith, T. H., and Vance, V. B. (1998). A viral suppressor of gene silencing in plants. *Proc Natl Acad Sci USA* **95**: 13079–84.
9. Li, H., Li, W. X., and Ding, S. W. (2002). Induction and suppression of RNA silencing by an animal virus. *Science* **296**: 1319–21.
10. Mette, M. F., Aufsatz, W., Van Der Winden, J., Matzke, M. A., and Matzke, A. J. (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J* **19**: 5194–201.
11. Pal-Bhadra, M., Bhadra, U., and Birchler, J. A. (1997). Cosuppression in Drosophila: gene silencing of *Alcohol dehydrogenase* by white-Adh transgenes is Polycomb dependent. *Cell* **90**: 479–90.
12. Pal-Bhadra, M., Bhadra, U., and Birchler, J. A. (1999). Cosuppression of nonhomologous transgenes in Drosophila involves mutually related endogenous sequences. *Cell* **99**: 35–46.
13. Chaboissier, M. C., Bucheton, A., and Finnegan, D. J. (1998). Copy number control of a transposable element, the I factor, a LINE-like element in Drosophila. *Proc Natl Acad Sci USA* **95**: 11781–85.
14. Jensen, S., Gassama, M. P., and Heidmann, T. (1999). Taming of transposable elements by homology-dependent gene silencing. *Nat Genet* **21**: 209–12.
15. Hamilton, A. J. and Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**: 950–2
16. Palauqui, J. C., Elmayan, T., Pollien, J. M., and Vaucheret, H. (1997). Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *EMBO J* **16**: 4738–45.
17. Tabara, H., Grishok, A., and Mello, C. C. (1998). RNAi in *C. elegans*: soaking in the genome sequence. *Science* **282**: 430–1.
18. Grishok, A., Tabara, H., and Mello, C. C. (2000). Genetic requirements for inheritance of RNAi in *C. elegans*. *Science* **287**: 2494–7.
19. Roignant, J.-Y., Carré, C., Mugat, B., Szymczak, D., Lepesant, J. A., and Antoniewski, C. (2003). Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in Drosophila. *RNA* **9**: 299–308.
20. Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363–6.

21. Caudy, A. A., Ketting, R. F., Hammond, S. M., Denli, A. M., Bathoorn, A. M., Tops, B. B., Silva, J. M., Myers, M. M., Hannon, G. J., and Plasterk, R. H. (2003). A micrococcal nuclease homologue in RNAi effector complexes. *Nature* **425**: 411–4.

22. Caudy, A. A., Myers, M., Hannon, G. J., and Hammond, S. M. (2002). Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev* **16**: 2491–6.

23. Elbashir, S. M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* **15**: 188–200.

24. Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001). Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J* **20**: 6877–88.

25. Ishizuka, A., Siomi, M. C., and Siomi, H. (2002). A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev* **16**: 2497–508.

26. Tomari, Y., Du, T., Haley, B., Schwarz, D. S., Bennett, R., Cook, H. A., Koppetsch, B. S., Theurkauf, W. E., and Zamore, P. D. (2004). RISC assembly defects in the Drosophila RNAi mutant armitage. *Cell* **116**: 831–41.

27. Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P. D. (2004). A protein sensor for siRNA asymmetry. *Science* **306**: 1377–80.

28. Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M. C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* **18**: 1655–66.

29. Pal Bhadra, M., Bhadra, U., and Birchler, J. A. (2002). RNAi related mechanisms affect both transcriptional and post-transcriptional transgene silencing in Drosophila. *Mol Cell* **9**: 315–27.

30. Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I., and Martienssen, R. A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**: 1833–7.

31. Kennerdell, J. R., Yamaguchi, S., and Carthew, R. W. (2002). RNAi is activated during Drosophila oocyte maturation in a manner dependent on aubergine and spindle-E. *Genes Dev* **16**: 1884–9.

32. Pal-Bhadra, M., Leibovitch, B. A., Gandhi, S. G., Rao, M., Bhadra, U., Birchler, J. A., and Elgin, S. C. (2004). Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. *Science* **303**: 669–72.

33. Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**: 672–6.

34. Schramke, V., Sheedy, D. M., Denli, A. M., Bonila, C., Ekwall, K., Hannon, G. J., and Allshire, R. C. (2005). RNA-interference-directed chromatin modification coupled to RNA polymerase II transcription. *Nature* **435**: 1275–9.

35. Onodera, Y., Haag, J. R., Ream, T., Nunes, P. C., Pontes, O., and Pikaard, C. S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**: 613–22.

36. Herr, A. J., Jensen, M. B., Dalmay, T., and Baulcombe, D. C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**: 118–20.

37. Huettel, B., Kanno, T., Daxinger, L., Aufsatz, W., Matzke, A. J., and Matzke, M. (2006). Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *EMBO J* **25**: 2828–36.

38. Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**: 320–4.

39. Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M. C. (2007). A slicer-mediated mechanisms for repeat-associated siRNA 5′ end formation in Drosophila. *Science* **315**: 1587–90.

40. Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**: 1089–103.

41. Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E. L., Zapp, M. L., Weng, Z., and Zamore, P. D. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science* **320**: 1077–81.

42. Czech, B., Malone, C. D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J. A., Sachidanandam, R., Hannon, G. J., and Brennecke, J. (2008). An endogenous small interfering RNA pathway in Drosophila. *Nature* **453**: 798–802.

43. Okamura, K., Chung, W. J., Ruby, J. G., Guo, H., Bartel, D. P., and Lai, E. C. (2008). The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* **453**: 803–6.

44. Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T. N., Siomi, M. C., and Siomi, H. (2008). Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**: 793–7.

45. Tam, O. H., Aravin, A. A., Stein, P., Girard, A., Murchison, E. P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R. M., and Hannon, G. J. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**: 534–8.

46. Chung, W.J., Okamura, K., Martin, R., and Lai, E. C. (2008). Endogenous RNA interference provides a somatic defense against Drosophila transposons. *Curr Biol* **18**: 795–802.

47. Pal-Bhadra, M., Bhadra, U., and Birchler, J. A. (2004). Interrelationship of RNA interference and transcriptional gene silencing in Drosophila. *Cold Spring Harb Symp Quant Biol* **69**: 433–8.

48. Grimaud, C., Bantignies, F., Pal-Bhadra, M., Bhadra, U., and Cavalli, G. (2006). RNAi components are required for nuclear clustering of Polycomb Group Response Elements. *Cell* **124**: 957–71.

49. Lei, E. P. and Corces, V. G. (2006). RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nat Genet* **38**: 936–41.

50. Peng, J. C. and Karpen, G. H. (2007). H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat Cell Biol* **9**: 25–35.

51. Deshpande, G., Calhoun, G., and Schedl, P. (2005). Drosophila argonaute-2 is required early in embryogenesis for the assembly of centric/centromeric heterochromatin, nuclear division, nuclear migration, and germ-cell formation. *Genes Dev* **19**: 1680–5.

52. Yin, H. and Lin, H. (2007). An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. *Nature* **450**: 3173–9.

53. Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843–54.

54. Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* **75**: 855–62.

55. Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**: 901–6.

56. Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., and Mello, C. C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**: 23–34.

57. Bernstein, E., Kim, S. Y., Carmell, M. A., Murchison, E. P., Alcorn, H., Li, M. Z., Mills, A. A., Elledge, S. J., Anderson, K. V., and Hannon, G. J. (2003). Dicer is essential for mouse development. *Nat Genet* **35**: 215–7.

58. Lee, Y. S., Nakahara, K., Pham, J. W., Kim, K., He, Z., Sontheimer, E. J., and Carthew, R. W. (2004). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**: 69–81.

59. Carlile, M., Nalbant, P., Preston-Fayers, K., McHaffie, G. S., and Werner, A. (2008). Processing of naturally occurring sense/antisense transcripts of the vertebrate Slc34a gene into short RNAs. *Physiol Genomics* **34**: 95–100.

CHAPTER 2

NUCLEIC ACIDS AS REGULATORY MOLECULES

FAQING HUANG, C. J. YU, and YANLIN GUO

2.1 GENE EXPRESSION AND ITS REGULATION

2.1.1 Genes, Chromosomes, and Genomes

The basic nature of the gene was defined by Mendel more than a century ago. Summarized in his two laws—the Law of Segregation and the Law of Independent Assortment, the gene was recognized as a "particulate factor" that passes unchanged from parent to progeny. In the 1940s, it was discovered that DNA (deoxyribonucleic acid) is the carrier of genetic information, thereby directly linking genes with DNA. Following the discovery of the double-stranded DNA (dsDNA) structure by Watson and Crick in 1953, it became clear that genetic information may be stored in DNA in the form of specific sequences of nucleotides (A, G, C, and T) and that genetic information can be transmitted by semiconservative replication through Watson–Crick base pairing. The four normal nucleotides are made of two different types—purines (A & G) and pyrimidines (C & T). Today, it is firmly established that, as the genetic material, genes on DNA molecules provide a blueprint that directs the developmental processes of cellular organisms and ultimately controls all aspects of cellular activities. At the molecular level, a gene is defined as a segment of DNA that encodes the information required to direct the synthesis of a gene product—either a particular protein or just an RNA molecule.

DNA can be a very long molecule. For instance, the total length of nuclear DNA in each human cell is about 2 m, while the diameter of the cell nucleus is only 5–8 μm. In order to ensure the equal distribution of DNA between both daughter cells during mitosis and to avoid damage of DNA

RNA Interference: Application to Drug Discovery and Challenges to Pharmaceutical Development, Edited by Paul H. Johnson

molecules, DNA molecules form complexes with proteins and are orderly packaged into chromosomes in eukaryotic cells. Every eukaryotic species has a characteristic number of chromosomes in each cell nucleus. For example, human somatic cells (all body cells except the reproductive cells) contain 23 pairs of chromosomes. Each chromosome contains one DNA molecule that carries several hundred to a few thousand genes. The complete set of genetic information carried by DNA in chromosomes of an organism is called its genome. The complete genome sequences of several model organisms including humans have been deciphered as a result of the Human Genome Project. Genome sequencing projects have revealed that in higher organisms there is a considerable amount of DNA that does not encode proteins. Therefore, the noncoding DNA is commonly referred to as nonfunctional DNA. For instance, the human genome consists of ~3.1 billion base pairs. Less than 2% of the genome actually codes for genes, which corresponds to 20,000–25,000 proteins. It is apparent that, in order to selectively express these genes from the vast majority of nonfunctional DNA in the genome and to express different genes differentially, cells must have a well-controlled gene expression system.

2.1.2 An Overview of Gene Expression

One of the central questions of molecular cell biology is how the genetic information encoded in a gene is translated into specific cellular activities. In other words, how does the information encoded by a gene contribute its particular role to the properties and functions of the cell? To exert a gene's function, the information encoded by the nucleotide sequence of a gene is used to yield a specific gene product—a functional protein or a functional noncoding RNA [1]. In this chapter, we will limit our discussion to protein as the gene product. In most cases, it is the protein that actually performs the task of that particular gene in the cell. In multicellular organisms, each cell type is programmed to express specific subsets of protein-encoding genes that determine the biochemical and phenotypic properties of that cell type. Whether it is a neuron for the transmission of neuronal signals or a lymphocyte for immune responses. Although differences between a neuron and a lymphocyte are extreme in both morphology and function, they still have the same DNA. Therefore, cells of an organism differ not because they contain different genes, but rather because they have turned on/off a specific set of genes for their specialized cellular functions.

Most cells in multicellular organisms are developmentally programmed to perform their specialized cellular functions. However, they must also be able to alter their patterns of protein synthesis to meet the needs of cells in response to environmental changes. For instance, the synthesis of several

proteins is dramatically increased in liver cells when the cells are exposed to a glucocorticoid hormone. This is because the glucocorticoid is released in the body during starvation and signals the cell the need for increased glucose production. In response to this stimulus, liver cells will turn on a set of genes that are required for the synthesis of glucose from amino acids or other small molecules. This example underscores the basic principle of gene expression in multicellular organisms. Different types of cells must selectively express the genes for their functions, and they must modify the expression patterns of certain genes in response to extracellular signals.

2.1.3 Gene Expression—from Genes to Proteins

The protein-encoding genes of an organism contain all the necessary instruction for protein synthesis. However, DNA itself does not directly participate in protein synthesis. Instead, the information in a gene is converted into the sequence of a messenger RNA (mRNA) in a process called transcription. Transcription is a general term for the DNA-templated synthesis of any RNA (including mRNA, transfer RNA—tRNA, ribosomal—rRNA, and other noncoding RNAs). An mRNA molecule with a defined sequence of A, G, C, and U is a faithful transcript of the DNA sequence of a gene. The mRNA molecule then acts as a template to direct the synthesis of a protein by the protein synthesis machinery—the ribosome (consisting of several rRNAs and tens of ribosomal proteins) through a process known as translation. During translation, tRNAs play important roles as the adapter molecules, specific amino acid carriers, and specific codon recognition. Therefore, the genetic information in the cell is first transcribed from DNA to RNA and then to protein. The entire process of genetic information flow from an information-containing gene to a functional protein product is commonly referred to as gene expression. All cells, from bacteria to human, follow the same pattern of gene expression—this principle is so fundamental that it is called the *central dogma* of molecular biology. However, the central dogma says nothing about the regulation of gene expression. Thus, obvious questions arise as to what determines the types and amounts of the proteins that characterize a particular cell type, or what factors allow the cell to modify the gene expression pattern in response to changes in its environment. These are the central questions of gene expression regulation in current molecular cell biology.

2.1.4 Gene Structure

A eukaryotic gene contains both coding and noncoding regions. The noncoding regions include the promoter, transcriptional regulatory sequences, and polyadenylation signals. The promoter is the sequence where an RNA