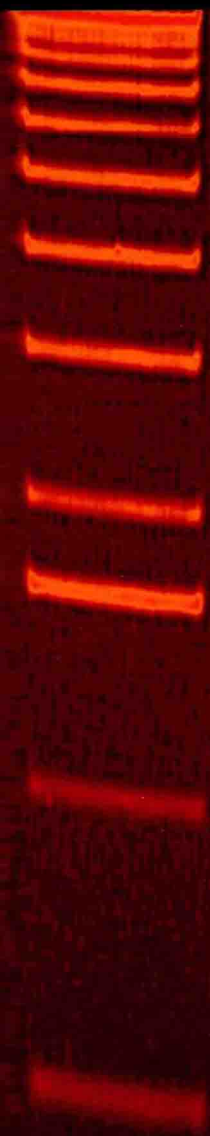
Methods in Molecular Biology™

VOLUME 221



Generation of cDNA Libraries

Methods and Protocols

Edited by

Shao-Yao Ying

HUMANA PRESS

METHODS IN MOLECULAR BIOLOGY™

Generation of cDNA Libraries

Methods and Protocols

Edited by

Shao-Yao Ying

Keck School of Medicine, University of Southern California, Los Angeles, CA

Humana Press  **Totowa, New Jersey**

999 Riverview Drive, Suite 208
Totowa, New Jersey 07512

www.humanapress.com

This publication is printed on acid-free paper. ∞
ANSI Z39.48-1984 (American Standards Institute) Permanence of Paper for Printed Library Materials.

Production Editor: Jessica Jannicelli.

Cover design by Patricia F. Cleary.

Cover Illustration: Figure 3 from Chapter 10, "Amplification of Representative cDNA Pools from Microscopic Amounts of Animal Tissue," by M. V. Matz.

For additional copies, pricing for bulk purchases, and/or information about other Humana titles, contact Humana at the above address or at any of the following numbers: Tel.: 973-256-1699; Fax: 973-256-8341; E-mail: humana@humanapr.com; or visit our Website: www.humanapress.com

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

Library of Congress Cataloging in Publication Data

Generation of cDNA libraries : methods and protocols / edited by Shao-Yao Ying.
p. cm. -- (Methods in molecular biology ; 221)
Includes bibliographical references and index.
ISBN 1-58829-066-2 (alk. paper) eISBN: 1-59259-359-3
1. Antisense DNA--Laboratory manuals. I. Series.

QP624.5.A57G46 2003
611'.0186--dc21

2002192170

Generation of cDNA Libraries

METHODS IN MOLECULAR BIOLOGY™

John M. Walker, SERIES EDITOR

236. **Plant Functional Genomics**: *Methods and Protocols*, edited by *Erich Grotewold, 2003*
235. ***E. coli* Plasmid Vectors**: *Methods and Applications*, edited by *Nicola Casali and Andrew Preston, 2003*
234. **p53 Protocols**, edited by *Sumitra Deb and Swati Palit Deb, 2003*
233. **Protein Kinase C Protocols**, edited by *Alexandra C. Newton, 2003*
232. **Protein Misfolding and Disease:** *Principles and Protocols*, edited by *Peter Bross and Niels Gregersen, 2003*
231. **Directed Evolution Library Creation**: *Methods and Protocols*, edited by *Frances H. Arnold and George Georgiou, 2003*
230. **Directed Enzyme Evolution**: *Screening and Selection Methods*, edited by *Frances H. Arnold and George Georgiou, 2003*
229. **Lentivirus Gene Engineering Protocols**, edited by *Maurizio Federico, 2003*
228. **Membrane Protein Protocols**: *Expression, Purification, and Characterization*, edited by *Barry S. Selinsky, 2003*
227. **Membrane Transporters:** *Methods and Protocols*, edited by *Qing Yan, 2003*
226. **PCR Protocols, Second Edition,** edited by *John M. S. Bartlett and David Stirling, 2003*
225. **Inflammation Protocols,** edited by *Paul G. Winyard and Derek A. Willoughby, 2003*
224. **Functional Genomics:** *Methods and Protocols*, edited by *Michael J. Brownstein and Arkady B. Khodursky, 2003*
223. **Tumor Suppressor Genes:** *Volume 2: Regulation, Function, and Medicinal Applications*, edited by *Wafik S. El-Deiry, 2003*
222. **Tumor Suppressor Genes:** *Volume 1: Pathways and Isolation Strategies*, edited by *Wafik S. El-Deiry, 2003*
221. **Generation of cDNA Libraries:** *Methods and Protocols*, edited by *Shao-Yao Ying, 2003*
220. **Cancer Cytogenetics:** *Methods and Protocols*, edited by *John Swansbury, 2003*
219. **Cardiac Cell and Gene Transfer:** *Principles, Protocols, and Applications*, edited by *Joseph M. Metzger, 2003*
218. **Cancer Cell Signaling:** *Methods and Protocols*, edited by *David M. Terrian, 2003*
217. **Neurogenetics:** *Methods and Protocols*, edited by *Nicholas T. Potter, 2003*
216. **PCR Detection of Microbial Pathogens:** *Methods and Protocols*, edited by *Konrad Sachse and Joachim Frey, 2003*
215. **Cytokines and Colony Stimulating Factors:** *Methods and Protocols*, edited by *Dieter Körholz and Wieland Kiess, 2003*
214. **Superantigen Protocols,** edited by *Teresa Krakauer, 2003*
213. **Capillary Electrophoresis of Carbohydrates,** edited by *Pierre Thibault and Susumu Honda, 2003*
212. **Single Nucleotide Polymorphisms:** *Methods and Protocols*, edited by *Pui-Yan Kwok, 2003*
211. **Protein Sequencing Protocols, Second Edition,** edited by *Bryan John Smith, 2003*
210. **MHC Protocols,** edited by *Stephen H. Powis and Robert W. Vaughan, 2003*
209. **Transgenic Mouse Methods and Protocols,** edited by *Marten Hofker and Jan van Deursen, 2003*
208. **Peptide Nucleic Acids:** *Methods and Protocols*, edited by *Peter E. Nielsen, 2002*
207. **Recombinant Antibodies for Cancer Therapy:** *Methods and Protocols*, edited by *Martin Welschof and Jürgen Krauss, 2002*
206. **Endothelin Protocols,** edited by *Janet J. Maguire and Anthony P. Davenport, 2002*
205. ***E. coli* Gene Expression Protocols,** edited by *Peter E. Vaillancourt, 2002*
204. **Molecular Cytogenetics:** *Protocols and Applications*, edited by *Yao-Shan Fan, 2002*
203. ***In Situ* Detection of DNA Damage:** *Methods and Protocols*, edited by *Vladimir V. Didenko, 2002*
202. **Thyroid Hormone Receptors:** *Methods and Protocols*, edited by *Aria Baniahmad, 2002*
201. **Combinatorial Library Methods and Protocols,** edited by *Lisa B. English, 2002*
200. **DNA Methylation Protocols,** edited by *Ken I. Mills and Bernie H, Ramsahoye, 2002*
199. **Liposome Methods and Protocols,** edited by *Subhash C. Basu and Manju Basu, 2002*
198. **Neural Stem Cells:** *Methods and Protocols*, edited by *Tanja Zigova, Juan R. Sanchez-Ramos, and Paul R. Sanberg, 2002*
197. **Mitochondrial DNA:** *Methods and Protocols*, edited by *William C. Copeland, 2002*
196. **Oxidants and Antioxidants:** *Ultrastructure and Molecular Biology Protocols*, edited by *Donald Armstrong, 2002*
195. **Quantitative Trait Loci:** *Methods and Protocols*, edited by *Nicola J. Camp and Angela Cox, 2002*
194. **Posttranslational Modifications of Proteins:** *Tools for Functional Proteomics*, edited by *Christoph Kannicht, 2002*
193. **RT-PCR Protocols,** edited by *Joe O'Connell, 2002*
192. **PCR Cloning Protocols, Second Edition,** edited by *Bing-Yuan Chen and Harry W. Janes, 2002*
191. **Telomeres and Telomerase:** *Methods and Protocols*, edited by *John A. Double and Michael J. Thompson, 2002*
190. **High Throughput Screening:** *Methods and Protocols*, edited by *William P. Janzen, 2002*
189. **GTPase Protocols:** *The RAS Superfamily*, edited by *Edward J. Manser and Thomas Leung, 2002*
188. **Epithelial Cell Culture Protocols,** edited by *Clare Wise, 2002*
187. **PCR Mutation Detection Protocols,** edited by *Bimal D. M. Theophilus and Ralph Rapley, 2002*
186. **Oxidative Stress Biomarkers and Antioxidant Protocols,** edited by *Donald Armstrong, 2002*
185. **Embryonic Stem Cells:** *Methods and Protocols*, edited by *Kursad Turksen, 2002*
184. **Biostatistical Methods,** edited by *Stephen W. Looney, 2002*
183. **Green Fluorescent Protein:** *Applications and Protocols*, edited by *Barry W. Hicks, 2002*
182. **In Vitro Mutagenesis Protocols, Second Edition,** edited by *Jeff Braman, 2002*
181. **Genomic Imprinting:** *Methods and Protocols*, edited by *Andrew Ward, 2002*
180. **Transgenesis Techniques, Second Edition:** *Principles and Protocols*, edited by *Alan R. Clarke, 2002*

Preface

Since its invention and subsequent development nearly 20 years ago, polymerase chain reaction (PCR) has been extensively utilized to identify numerous gene probes in vitro and in vivo. However, attempts to generate complete and full-length complementary cDNA libraries were, for the most part, fruitless and remained elusive until the last decade, when simple and rapid methods were developed. With current decoding and potential application of human genome information to genechips, there are urgent needs for identification of functional significance of these decoded gene sequences. Inherent in bringing these applications to fruition is the need to generate a complete and full-length cDNA library for potential functional assays of specific gene sequences.

Generation of cDNA Libraries: Methods and Protocols serves as a laboratory manual on the evolution of generation of cDNA libraries, covering both background information and step-by-step practical laboratory recipes for which protocols, reagents, operational tips, instrumentation, and other requirements are detailed. The first chapter of the book is an overview of the basics of generating cDNA libraries, which include the following: (a) the definition of a cDNA library, (b) different kinds of cDNA libraries, (c) differences between methods for cDNA library generation using conventional approaches and novel strategies, including reverse generation of RNA repertoires from cDNA libraries, and (d) the quality of cDNA libraries. In subsequent chapters, various methods are presented to provide the reader with a wide range of methodologies for enhancing the generation of complete and full-length libraries. Again, each method of cDNA library generation contains a balanced presentation of both background information and practical procedures. The remainder of this book explains how to confirm the quality of the cDNAs generated and some of the applications, including (a) electrophoresis, (b) Northern blotting, (c) microarray analysis, (d) subtractive hybridization, (e) subtractive cloning, (f) gene cloning, and (g) peptide library generation.

The final chapter of the book outlines the future use of full-length cDNA libraries in biomedical research, diagnostic utilization, drug development, and clinical therapy.

The authors contributing the various chapters are all experts in their fields, and they have either developed and/or routinely performed the methodologies

described herein. It is anticipated that the subject matter covered in *Generation of cDNA Libraries: Methods and Protocols* will be particularly useful for biologists, biochemists, molecular biologists, and clinicians, which would furnish them several ready-to-use methodologies for attacking the problems in their specific areas of interests.

Shao-Yao Ying

Contributors

NAILA O. ALIEVA • *Whitney Laboratory, University of Florida, St. Augustine, FL*

MASAHIRO ASADA • *Age Dimension Research Center, National Institute of Advanced Industrial Science and Technology, Ibaraki, Japan*

CYNTHIA CHANG • *BD Biosciences Clontech, Palo Alto, CA*

JIAN-JUN CHEN • *Section of Hematology and Oncology, University of Chicago Medical Center, Chicago, IL*

STEVEN H. M. CHEN • *Division of Hematology, Department of Pediatrics, University of Southern California, Childrens Hospital Los Angeles, Los Angeles, CA*

YE-GUANG CHEN • *Division of Biomedical Sciences, University of California, Riverside, CA*

ALEX CHENCHIK • *BD Biosciences Clontech, Palo Alto, CA*

LUDA DIATCHENKO • *BD Biosciences Clontech, Palo Alto, CA*

SCOTT C. FEARS • *Section of Hematology and Oncology, University of Chicago Medical Center, Chicago, IL*

RODNEY J. FOLZ • *Department of Medicine and Cell Biology, Division of Pulmonary and Critical Care Medicine, Duke University Medical Center, Durham, NC*

HANS-JÜRGEN FÜLLE • *Department of Cell and Neurobiology, Keck School of Medicine, University of Southern California, Los Angeles, CA*

ZE-GUANG HAN • *Chinese National Human Genome Center at Shanghai, Shanghai, China*

QIU-HUA HUANG • *Chinese National Human Genome Center at Shanghai, Shanghai, China*

SHENG-HE HUANG • *Division of Infectious Diseases, Department of Pediatrics, University of Southern California, Childrens Hospital Los Angeles, Los Angeles, CA*

TORU IMAMURA • *Age Dimension Research Center, National Institute of Advanced Industrial Science and Technology, Ibaraki, Japan*

HENRY JI • *Epiclone Inc., San Diego, CA*

MICHAEL G. K. JONES • *Western Australia State Agricultural Biotechnology Center, Murdoch University, Perth, WA, Australia*

AMBROSE Y. JONG • *Division of Hematology, Department of Pediatrics, University of Southern California, Childrens Hospital Los Angeles, Los Angeles, CA*

SANGGYU LEE • *Section of Hematology and Oncology, University of Chicago Medical Center, Chicago, IL*

SHI-LUNG LIN • *Epiclone Inc., San Diego, CA*

TSEN-YIN LIN • *Department of Cell and Neurobiology, Keck School of Medicine, University of Southern California, Los Angeles, CA*

SERGEY LUKYANOV • *Evrogen GSC, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia*

MIKHAIL V. MATZ • *Whitney Laboratory, University of Florida, St. Augustine, FL*

JOSEPH MCKINLEY • *BD Biosciences Clontech, Palo Alto, CA*

REBECCA L. MULLINAX • *Stratagene, La Jolla, CA*

IGOR NEPLUEV • *Duke University Medical Center, Durham, NC*

OSAMU OHARA • *Department of Human Gene Research, Kazusa DNA Research Institute, Chiba, Japan; Immunogenomics Research Team, Research Center for Allergy and Immunology, The Institute of Physical and Chemical Research (RIKEN), Yokohama, Japan*

JANET D. ROWLEY • *Section of Hematology and Oncology, University of Chicago Medical Center, Chicago, IL*

PAUL D. SIEBERT • *BD Biosciences Clontech, Palo Alto, CA*

JOSEPH A. SORGE • *Stratagene, La Jolla, CA*

SUMIO SUGANO • *Laboratory of Genome Structure Analysis, HGC, Institute of Medical Science, University of Tokyo, Tokyo, Japan*

YUTAKA SUZUKI • *Laboratory of Genome Structure Analysis, HGC, Institute of Medical Science, University of Tokyo, Tokyo, Japan*

SAN MING WANG • *Section of Hematology and Oncology, University of Chicago Medical Center, Chicago, IL*

ZHAOHUI WANG • *Western Australia State Agricultural Biotechnology Center, Murdoch University, Perth, WA, Australia*

BONNIE WU • *Stratagene, La Jolla, CA*

SHAO-YAO YING • *Department of Cell and Neurobiology, Keck School of Medicine, University of Southern California, Los Angeles, CA*

LIN ZHANG • *Oncology Center, Johns Hopkins University School of Medicine, Baltimore, MD*

XIN ZHANG • *Chinese National Human Genome Center at Shanghai, Shanghai, China*

YUE ZHANG • *Department of Molecular Genetics and Microbiology, Center for Infectious Diseases/CMM, State University of New York at Stony Brook, Stony Brook, NY*

GUOLIN ZHOU • *Section of Hematology and Oncology, University of Chicago Medical Center, Chicago, IL*

BAKHYT ZHUMABAYEVA • *BD Biosciences Clontech, Palo Alto, CA*

Contents

1

Complementary DNA Libraries

An Overview

Shao-Yao Ying

1. Introduction

Complementary DNA libraries reflect gene expression at certain times for specific cells, whereas genomic DNA libraries represent all genetic information in somatic cells. The complexity of cellular organization reflects a genetic program that encodes a collection of genes and the means to use them by manufacturing proteins for cellular structures, functional activities, and reproduction of cells themselves. The essential aspect of this process is protein synthesis based on the information stored in the sequence of nucleotides that make up a gene (a transcribable segment of a DNA molecule) as the blueprint. The information is transcribed as a complementary sequence of the nucleotides (mRNA or the transcript) that carries the genetic information from the nucleus to the protein-synthesizing machinery in the cytoplasm. Then, mRNA is translated into the sequence of amino acids that make up a protein. The basis of the widely used novel strategies for the generation of cDNA libraries are base pair complementarities, reverse transcription, and polymerase chain reactions. This chapter presents some general information on the principles of, biology behind, basic protocols of, and reagents used in the generation of cDNA libraries. Hopefully, this information will help researchers overcome problems encountered in actual construction of cDNA libraries.

1.1. Base Pair Complementarities

Nucleic acids exhibit base pair complementarities that faithfully convert one strand of RNA/DNA to a complementary one. Although all genetic information

From: *Methods in Molecular Biology, vol. 221: Generation of cDNA Libraries: Methods and Protocols*
Edited by: S.-Y. Ying © Humana Press Inc., Totowa, NJ

in the somatic cells of a specific organism can be expressed as a transcript, many DNA sequences are not transcribed. These segments of DNA are the coding exons and the noncoding introns. Basically, the genetic information is stored as a strand of a DNA molecule consisting of four bases: adenine, thymine, guanine, and cytosine. A second complementary strand of DNA can be formed by DNA polymerase. Polymerases, enzymes that function in DNA replication and RNA transcription, synthesize a nucleic acid from the genetic information encoded by the template strand. The polymerases are unique because they take direction from another nucleic acid template, which is either DNA or RNA. During the formation of a second strand of DNA, bases are generated according to the Watson–Crick base-pairing pattern. That is to say, every cytosine is replaced by a guanine, every guanine by a cytosine, every adenine by a thymine, and every thymine by an adenine. In this way, information in DNA is correctly transcribed into RNA.

1.2. Probe Hybridization

Another unique feature of the base pair complementarity is probe hybridization. The findings of Gillespie and Spiegelman *(1)* that viral genomic DNA and RNA in infected cells showed a base pair complementarity opened an avenue for specific hybridization between a gene and its transcript as a DNA–RNA hybrid. Subsequently, the DNA–DNA or DNA–RNA hybrids have been employed in a large number of powerful techniques for the identification and manipulation of the geneitic information stored in DNA and used by the cell via RNA. Usually, a labeled-probe nucleic acid is hybridized with a target nucleic acid. After removal of any unreacted probe, the remaining labeled probe is identified and the intensity of the labeling of the hybrid duplex is determined. As a result, the regions of complementarity between the probe and the target nucleotides are detected *(2)*. Frequently, the number of targets is quite low, perhaps only a few copies. In such cases, amplification techniques are performed to produce large numbers of copies of the target, thus increasing the amount of hybrid duplex and the observed signal. In addition, immobilization of the target on a surface, such as a nitrocellulose or nylon filter and many other solid-phase materials, is used to solve the competitive equilibrium problem. Thus, nucleic acid sequences can be quantified by molecular hybridization using complementary nucleic acids as probes, with complementarity as the essential feature for hybridization.

1.3. Polymerases Are Essential for DNA Synthesis

Polymerases that use RNA as a template to form a complementary DNA are RNA-direct DNA polymerases *(3,4)*. One of these enzymes is reverse transcriptase, usually observed as a part of the viral particle, during the life

cycle of retroviruses and other retrotransposable elements. Purified reverse transcriptase is used to generate complementary DNA from polyadenylated mRNAs; therefore, double-stranded DNA molecules can be formed from the single-stranded RNA templates. The synthesis of DNA on an RNA template mediated by the enzyme reverse transcriptase is known as reverse transcription *(5)*.

1.4. A Primer is Required for Reverse Transcription

Although polymerases copy genetic information from one nucleotide into another, including copying a mRNA to generate a complementary DNA strand in the presence of reverse transcriptase, they do need a "start signal" to tell them where to begin making the complementary copy. The short piece of DNA that is annealed to the template and serves as a signal to initiate the copying process is the primer *(6)*. The primer is annealed to the template by basepairing so that its 3′-terminus possesses a free 3′-OH group and chain growth is exclusively from 5′ end to the 3′ end for polymerization. Wherever such as primer–template pair is found, DNA polymerase will begin adding bases to the primer to create a complementary copy of the template.

1.5. Formation of cDNA

Generally, the cDNA of cells can be formed according to the following steps:

1. Isolation of the mRNA template: The source mRNAs can be enriched by increasing the abundance of specific classes of rare mRNAs via one of the following approaches: (1) antibody precipitation of the protein of interest that is synthesized in cell lines, (2) increasing the concentrations of relevant RNAs by drug-induced overexpression of genes of interest, and (3) inhibition of protein synthesis by inhibitors, resulting in extended transcription of the early genes of mammalian DNA virus.

   The integrity of the mRNA is essential for the quality of cDNA generation. The size of mRNAs isolated should range from 500 bp to 8.0 kb, and the sequence should retain the capability of synthesizing the polypeptide of interest in vitro, such as in cell-free reticulocytes. When fractionated by electrophoresis and stained with ethidium bromide, a good preparation of mRNA should appear as a smear from 500 bp to 8 kb.
2. A short oligo(dT) primer is bound to the poly(A) of each mRNA at the 3′ end.
3. The mRNA is transcribed by reverse transcriptase (the primer is needed to initiate DNA synthesis) to form the first strand of DNA, usually in the presence of a reagent to denature any regions of the secondary structure. RNAse is used to prevent RNA degradation.
4. DNA–RNA hybids are formed.
5. The RNA is nicked by treatment with RNAse H to generate the free 3′-OH groups.

6. DNA polymerase I is added to digest the RNA, using the RNA fragments as primers, and replace the RNA with DNA. In some cases, a primer–adapter method is carried out as follows: (1) terminal transferase is added to the first strand cDNA [add (dC) to provide free 3′ hydroxyl groups]; (2) the tail of hybrided cDNA with oligo(dG) serves as the primer.
7. Double-stranded cDNA is formed.

1.6. PCR

Another important development in generating DNA from mRNA is the enzymatic amplification of DNA by a technique known as polymerase chain reaction (PCR). The technique was originally reported by Saike et al. *(7)*, who employed a heat-stable DNA polymerase–*Taq* polymerase with two primers that are complementary to DNA sequences at the 3′ ends of the region of the DNA to be amplified. The oligonucleotides serve as primers to which nucleotides are added during the subsequent replication steps. Because a DNA strand can only add nucleotides at the 3′ hydroxyl terminus of an existing strand, a strand of DNA that provides the necessary 3′-OH terminus, in this case, is also called a primer. All DNA polymerases require a template and a primer.

The PCR is well established as the default method for DNA and RNA analysis. More robust formats have been introduced, improved thermal cyclers developed, and new labeling and detection methods developed. Because gene expression profiling relies on mRNA extraction from defined types and numbers of cells, in some cases the use of small number of cells or even a few cells is necessary. In this situation, the PCR technique has been used to allow synthesis of cDNAs from a small amount of mRNA *(8,9)*. Other techniques of amplifying mRNA have been developed *(10)*. For instance, the cDNA can be generated by mRNA extracted and amplified by poly(A) reverse transverse transcription and PCR.

2. Definitions

2.1. Complementary DNA

If a chromosome is defined as a supercoiled, linear DNA molecule consisting of numerous transcribable segments as genes (specific segments of DNA that code for a specific protein), the complementary DNA (cDNA) can be defined as the transcriptionally active segment of a DNA molecule that shows the base pair complementarity between the gene and its transcribed and processed mRNA molecules—the transcript. To define it differently, cDNAs are complementary DNA copies of mRNA that are generated by the enzyme—reverse transcriptase. In contrast to genomic DNA, the extra, nontranscribed DNAs in a genome are removed by this process because DNA polymerase activity depends on the