

# Methods of DNA and RNA Sequencing

*Edited by*

Sherman M. Weissman, M.D.

# **Methods of DNA and RNA Sequencing**

*Edited by*

**Sherman M. Weissman, M.D.**

Library of Congress Cataloging in Publication Data

Main entry under title:

Methods of DNA and RNA sequencing.

Includes bibliographical references and index.

1. Nucleotide sequence. 2. Deoxyribonucleic acid—Analysis. 3. Ribonucleic acid—Analysis.

I. Weissman, Sherman M.

QP625.N89M47 1983 574.87'328 82-22261

ISBN 0-03-059174-0

Published in 1983 by Praeger Publishers  
CBS Educational and Professional Publishing  
a Division of CBS Inc.  
521 Fifth Avenue, New York, New York 10175 U.S.A.

© 1983 by Praeger Publishers

*All rights reserved*

456789 052 98765432

Printed in the United States of America

## LIST OF CONTRIBUTORS

**Marvin H. Caruthers**

Department of Chemistry  
University of Colorado  
Boulder, CO 80309  
USA

**David J. Galas**

Département de Biologie Moléculaire  
Université de Genève  
Geneva  
Switzerland

**G. Nigel Godson**

Biochemistry Department  
New York University Medical  
Center  
New York, N.Y. 10016  
USA

**Yoshiyuki Kuchino**

Biology Division  
National Cancer Center Research  
Institute  
Tsukiji, Chuo-Ku  
Tokyo, Japan

**Allan Marshall Maxam**

Sidney Farber Cancer Institute  
Charles A. Dana Cancer Center  
Boston, MA 02115  
USA

**Stephen M. Mount**

Department of Molecular Biophysics  
and Biochemistry  
Yale University  
New Haven, CT 06510  
USA

**Susumu Nishimura**

Biology Division, National Cancer  
Center Research Institute  
Tsukiji, Chuo-Ku  
Tokyo, Japan

**Debra A. Peattie**

Department of Biochemistry  
Stanford University Medical Center  
Stanford, CA 94305  
USA

**Erika Randerath**

Department of Pharmacology  
Baylor College of Medicine  
Texas Medical Center  
Houston, TX 77030  
USA

**Kurt Randerath**

Department of Pharmacology  
Baylor College of Medicine  
Texas Medical Center  
Houston, TX 77030  
USA

**Albert Schmitz**  
Département de Biologie Moléculaire  
Université de Genève  
Geneva  
Switzerland

**Michael Smith**  
Department of Biochemistry  
Faculty of Medicine  
University of British Columbia  
Vancouver, B.C. V6T 1W5  
Canada

**Thomas Shenk**  
Department of Microbiology  
Health Sciences Center  
State University of New York  
Stony Brook, N.Y. 11794  
USA

**Joan A. Steitz**  
Department of Molecular Biophysics  
and Biochemistry  
Yale University  
New Haven, CT 06510  
USA

**Sherman M. Weissman**  
Department of Human Genetics  
Yale University School  
of Medicine  
New Haven, CT 06510  
USA

## PREFACE

The entire structure of a free-living biologic unit is, to a first approximation, encoded in its DNA. Increasingly efficient and universal approaches for cloning discrete fragments of genetic DNA defined by restriction endonuclease cleavage sites have made available in principle each part of even the most complex genomes in large amounts and of sufficient purity for detailed analysis. Development of DNA sequencing techniques based principally on restriction endonuclease-generated fragments and dependent on gel fractionation methods that are capable of resolving the oligonucleotides that differ by 0.2% or less in their chain length have made it possible for many laboratories to determine complete sequences for the genetic material of particular interest to their research. Fortunately for the field, the inventors have made available fairly detailed protocols that make it possible to routinely reproduce the methods with clean results. A particularly exemplary protocol has been provided by Drs. Maxam and Gilbert for the method of chemical sequencing of DNA. One purpose of the present volume has been to update a simple set of such protocols for the more commonly used methods of nucleic acid labeling and sequence analysis, together with a discussion of the principles underlying the methods and some presentation of discussion of common sources of technical difficulties.

The chemical synthesis of deoxynucleotides has become a fundamental tool, both in sequence analysis and in *in vitro* mutagenesis and further characterization of biologically active sequences. Dr. Caruthers has included a chapter providing an update on discussion of the current methods that are applicable for laboratories not specialized in organic chemistry, and Dr. Smith has prepared an extensive discussion of the application of synthetic oligonucleotides to problems of molecular biology. Dr. Godson has reviewed the enzymatic approaches to DNA sequencing, and Dr. Maxam has provided a historical account of the development of chemical methods for DNA sequencing. It has become an increasing challenge to interpret

sequences, once available, particularly through the use of valuable chemical methods that have been developed for studying interaction of purified proteins or protein complexes with nucleic acids, and these footprinting methods and extensive discussion of them by Drs. Schmitz and Galas are also included in the present volume.

Finally, a variety of techniques, including the classical techniques of prokaryotic genetics and *in vitro* mutagenesis of genes in functioning animal cells, followed by reintroduction of genes into cellular milieus *in vitro* enzymologic studies, have partly defined at least major subclasses of sequences responsible for transcription initiation and termination of prokaryotes and begun to provide a corresponding definition for eukaryotic sequences involved in these processes, as well as in the uniquely eukaryotic RNA splicing mechanisms. "State of the art" reviews on sequence features involved in initiation and termination of transcription (Dr. Shenk), and RNA splicing (Drs. Mount and Steitz) have been included.

The aim of the editor throughout has been to obtain experts to review their fields with the intention of describing methods in sufficient detail to facilitate ready application in laboratories without previous experience in the field and to provide discussion both of the theoretic underpinnings of the method and of the potential pitfalls. An effort has been made to minimize reduplication of material that has been authoritatively and lucidly reviewed elsewhere; for example, there is no discussion of *in vitro* mutagenesis, or of features of translation initiation sites, since several excellent reviews are available. One major feature of nucleic acid sequence analysis not covered in the present volume is the use of computers, both to expedite handling of the data and to reveal more detailed information than can be seen simply by visual inspection. This is a rapidly burgeoning subject in which many laboratories have their own favorite "data-crunching" systems, and increasing progress has been made in the theoretical aspects of problems related to the detection of potential base-pairing structures, analyses of homology among related sequences, etc. The editor can only refer the interested reader to primary papers such as the excellent collection recently published in *Nucleic Acids Research* (Jan., 1982) and to caution the novice that accessibility to computer analysis of data is almost essential for any extensive sequencing project. It is the hope of the authors that the present set of essays will expedite the work of young or novice investigators and provide a convenient reference or refresher for experienced workers in the field.

Sherman M. Weissman, M.D.

# CONTENTS

List of Contributors	v
Preface	xvii
Acknowledgments	xviii
1	
NEW METHODS FOR CHEMICALLY SYNTHESIZING DEOXYOLIGONUCLEOTIDES	1
<i>Marvin H. Caruthers</i>	
General outline of DNA chemical synthesis methodologies	2
A procedure for the rapid chemical synthesis of deoxyoligonucleotides	5
The support	6
The synthesis cycle	7
Isolation of deoxyoligonucleotides	11
Experimental section	13
General methods	13
Synthesis of the support	13
Synthesis of deoxynucleosidephosphoramidites	15
Outline of the synthesis cycle	16
Isolation of synthetic deoxyoligonucleotides	17
Acknowledgments	17
References	18



<b>SYNTHETIC OLIGODEOXYRIBONUCLEOTIDES AS PROBES FOR NUCLEIC ACIDS AND AS PRIMERS IN SEQUENCE DETERMINATION</b>	<b>23</b>
<i>Michael Smith</i>	
Introduction	23
Background	24
Oligonucleotide length and duplex stability	25
Effect of base composition on oligonucleotide duplex stability	29
Stability of oligodeoxyribonucleotide-oligoribonucleotide duplexes	31
Influence of base-pair mismatch on oligonucleotide duplex stability	34
Ionic strength, polyvalent cations, and oligodeoxyribonucleotide duplex stability	37
Oligonucleotide duplexes form very rapidly	38
Length of a genetically unique oligonucleotide	39
Synthetic oligodeoxyribonucleotide as probe for genomic DNA fragment purification	39
Oligodeoxyribonucleotides as probes for monitoring isolation of a genomic DNA clone	41
Oligodeoxyribonucleotides as probes for monitoring isolation of a cDNA clone	46
Oligodeoxyribonucleotides as primers for reverse transcriptase to produce specific cDNA probes	48
Use of synthetic oligodeoxyribonucleotides to prepare double-strand cDNAs	52
Oligodeoxyribonucleotides as primers for DNA sequence determination	53
Use of synthetic oligodeoxyribonucleotides as primers for RNA sequence determination	55
Conclusion	57
References	57

## 3

# SEQUENCING DNA BY THE SANGER CHAIN TERMINATION METHOD 69

*G. Nigel Godson*

Introduction	69
Principle of primed synthesis sequencing and the chain termination method	70
Use of the filamentous phage cloning vectors for chain termination sequencing	73
A model system for sequencing a 2-kB piece of DNA	74
Special strategies in cloning and sequencing	77
Cloning DNA fragments in a directed orientation	78
Cloning internal DNA fragments only	79
Subcloning in situ	79
Turning cloned fragments around	79
Use of internal primers	80
Checking the sequence with restriction enzymes	81
Problems of random sequencing	82
Strategies for sequencing large pieces of DNA (10–100 kB)	82
Appendixes	83
Acknowledgments	108
References	108

## 4

# NUCLEOTIDE SEQUENCE OF DNA 113

*Allan Marshall Maxam*

Introduction	113
Sequencing bases in nucleic acids	114
Distinguishing the bases	115
Ordering nucleotides	116

Phase one: Tract and block methods	123
Pyrimidine tracts	125
Copying DNA into RNA	126
Ribosubstitution	128
Phase two: Wandering-spot methods	129
Phase three: Base-specific nested-segment methods	132
The chemical method	133
Comparison and assessment of the chemical and enzymatic methods of DNA sequencing	144
Double-stranded DNA	144
Single-stranded DNA	147
Amount and specific activity of DNA	147
Sequencing labor	148
Sequence accuracy	148
Biologically significant DNA ends and modifications	149
References	150
Addendum by Sherman M. Weissman, M.D.	164
References	166

## 5

SELECTED POSTLABELING PROCEDURES FOR BASE COMPOSITION AND SEQUENCE ANALYSIS OF NUCLEIC ACIDS	169
--	-----

*Erika Randerath & Kurt Randerath*

Radioactive derivative (postlabeling) methods for structural analysis of nucleic acids	169
Fluorography and screen-intensified autoradiography for enhanced detection of radioisotopes ( $^3\text{H}$ , $^{32}\text{P}$ , $^{125}\text{I}$ ) on chromatograms or gels	173
General aspects	173
Materials	174
Fluorography for the detection of $^3\text{H}$ on thin-layer chromatograms	175

Detection of $^3\text{H}$ in polyacrylamide, mixed polyacrylamide/agarose, or agarose gels by fluorography	176
Detection of $^{32}\text{P}$ and $^{125}\text{I}$ on chromatograms or gels by screen-intensified autoradiography	178
$^3\text{H}$ derivative method for base analysis of RNA	178
General aspects	178
Outline of method	179
Materials	180
Procedure	180
Comments	186
Radioiodination method for ribonucleoside analysis	188
General aspects	188
Outline of method	188
Materials	190
Procedure	190
Comments	192
$^{32}\text{P}$ derivative method for base analysis of normal and mutagen/carcinogen-modified DNA	192
General aspects	192
Outline of method	193
Materials	194
Procedures	196
Comments	198
$^3\text{H}$ derivative method for sequence analysis of RNA	200
General aspects	200
Outline of method	200
Combined $^3\text{H}/^{32}\text{P}$ derivative method for sequence analysis of RNA	201
General aspects	201
Outline of method	201
$^{32}\text{P}$ gel readout method entailing partial enzymatic degradation of end-labeled RNA	202
General aspects	202

Outline of method	203
Materials	204
Procedures	205
Comments	210
<sup>32</sup> P thin-layer readout method based on single-hit chemical cleavages of RNA	210
General aspects	210
Outline of method	211
Materials	212
Procedure	214
Comments	220
Sequence analysis of large RNAs	223
Conclusions	224
Acknowledgments	225
References	225

## 6

### CHARACTERIZATION OF MODIFIED NUCLEOSIDES IN tRNA 235

*Susumu Nishimura & Yoshiyuki Kuchino*

Identification of modified nucleosides by two- dimensional thin-layer cellulose chromatography and paper chromatography	239
Identification of modified nucleosides by high- pressure liquid chromatography (HPLC)	243
Analysis of modified nucleosides by mass spectrometry	243
Analysis of modified nucleotides by postlabeling procedures	245
Identification of modified nucleotides during tRNA sequencing	245
Identification of modified nucleotides in total tRNA by the postlabeling procedure	251
Structural elucidation of unknown modified nucleosides	252

Materials	254
References	255

## 7

DIRECT CHEMICAL METHOD FOR SEQUENCING END-LABELED RIBONUCLEIC ACIDS	261
--	-----

*Debra A. Peattie*

Principles of the chemical RNA sequencing reactions	262
Chemical base modification	264
Aniline-induced strand scission	267
End-labeling the RNA	270
3'-end-labeling with (5'- <sup>32</sup> P)pCp and T4 RNA ligase	270
5'-end-labeling with <sup>32</sup> P	273
Practical considerations	274
Base-specific chemical cleavage reactions	275
Procedures	275
Practical considerations	284
Using the chemical reactions to probe RNA conformation	286
Guanosine reaction	288
Cytidine reaction	288
Adenosine reaction	289
Practical Considerations	289
Acknowledgments	290
References	291
Appendix	297
Outline for four base-specific reactions for chemical sequencing of end-labeled RNA molecules	297
Diagnostic table for aid in correcting aberrant RNA sequencing patterns	300

## THE STUDY OF PROTEIN-DNA BINDING SPECIFICITY: DNase FOOTPRINTING 305

*Albert Schmitz & David J. Galas*

Introduction	305
Bovine pancreatic deoxyribonuclease I (DNase I)	306
Description of the method	307
The reactions	310
Typical footprinting reaction	311
Assignment of the DNase I-produced bands	312
Other enzymes suitable for footprinting reactions	313
Interactions analyzed by footprinting	315
Footprint of <i>lac</i> repressor on operator DNA	315
Footprints of RNA polymerase on the <i>lac</i> promoter	320
Footprints of CRP on the <i>lac</i> control region DNA	327
Footprint of DNA gyrase	331
Nuclease specificities	333
Extensions of the method	334
Concluding remarks	335
Acknowledgments	336
References	336
Addendum	344

## TRANSCRIPTION INITIATION AND TERMINATION SIGNALS 349

*Thomas Shenk*

Prokaryotic promoters	349
Eukaryotic RNA polymerases	351
Cell-free transcription by eukaryotic polymerases	351
Polymerase III	352
Polymerase II	352

Polymerase I	353
5'-ends and initiation sites	353
Polymerase III	353
Polymerase II	353
Polymerase I	354
Transcriptional control regions	355
Polymerase III	355
Polymerase II	357
Polymerase I	368
Regulation of transcription by eukaryotic polymerases	368
Polymerase III	370
Polymerase II	372
Polymerase I	374
Termination of transcription in prokaryotes	374
Termination of transcription in eukaryotes	375
Acknowledgments	377
References	377

## 10

### SIGNALS FOR THE SPLICING OF EUKARYOTIC MESSENGER RNA TRANSCRIPTS 399

*Stephen M. Mount & Joan A. Steitz*

Splice junction signals	400
Are junction signals sufficient?	410
The versatility of splicing	412
What recognizes splicing signals?	417
The biochemistry of splicing	419
Why splicing?	421
Acknowledgments	423
References	424

Index	437
-------	-----



# 1

## New Methods for Chemically Synthesizing Deoxyoligonucleotides

MARVIN H. CARUTHERS

Recent developments in the recombinant DNA field have created a need for sequence-defined deoxyoligonucleotides. These compounds are being used as primers and probes for isolating natural genes (Wallace et al., 1981; Gillam et al., 1977), for experiments involving site-directed mutagenesis (Smith, 1980), and for the synthesis of genes and gene control regions (Agarwal et al., 1970; Khorana, 1979; Itakura et al., 1977; Goedell, Kleid, et al., 1979; Goedell, Heyneker, et al., 1979; Wetzel et al., 1980; Edge et al., 1981). However, until recently the synthesis and isolation of deoxyoligonucleotides has been a difficult and time-consuming task. Ideally, chemical methods should be simple, rapid, versatile, and accessible to the nonchemist. This chapter outlines a synthetic methodology that satisfies all these criteria. The approach has been used successfully by nonchemists in my own laboratory and more recently by other research groups with limited backgrounds in nucleic acid chemistry.

This chapter is divided into three sections. The first section provides a general outline of various approaches that have been used for synthesizing deoxyoligonucleotides. Such an outline is important since the field is highly specialized and not generally familiar to most biologists and biochemists. The second section outlines and discusses our methodology. The final section is a presentation of procedural details useful for those interested in duplicating our approach.