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.

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

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.