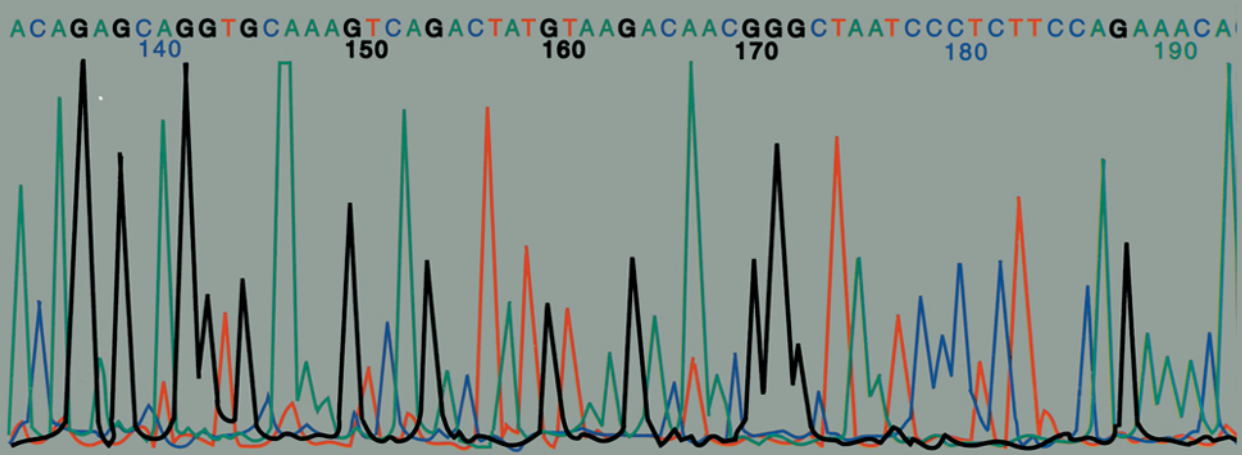
Human Molecular Genetics

Edited by

Kenneth W. Adolph



Methods in Molecular Genetics

Methods in Molecular Genetics

Volume 8

Human Molecular Genetics

This Page Intentionally Left Blank

Methods in Molecular Genetics

Edited by

Kenneth W. Adolph
Department of Biochemistry
University of Minnesota Medical School
Minneapolis, Minnesota

Volume 8

Human Molecular Genetics

ACADEMIC PRESS
San Diego New York Boston London Sydney Tokyo Toronto

Front cover photograph: ABI sequencing result using the modified cycle sequencing procedure. As a result of numerous experiments, a modified ABI cycle sequencing procedure was developed that could use nonamer primers to sequence double-stranded template DNAs. The ABI373A sequencing profile shown was generated using nonamer 197 to prime a cycle sequencing reaction from its binding site in clone pCh32-2.3-7E. This ABI sequence profile yielded about 450 bp of sequence information. Courtesy of Drs. Jerry L. Slightom and Jeffrey H. Bock, Molecular Biology Research Unit, The Upjohn Company, Kalamazoo, Michigan.

This book is printed on acid-free paper. ♾

Academic Press, Inc.
A Division of Harcourt Brace & Company
525 B Street, Suite 1900, San Diego, California 92101-4495

United Kingdom Edition published by
Academic Press Limited
24-28 Oval Road, London NW1 7DX

International Standard Serial Number: 1067-2389

International Standard Book Number: 0-12-044310-4

PRINTED IN THE UNITED STATES OF AMERICA
96 97 98 99 00 01 EB 9 8 7 6 5 4 3 2 1

Table of Contents

This Page Intentionally Left Blank

Contributors to Volume 8

Numbers in parentheses indicate the pages on which the authors' contributions begin.

ELIZABETH A. ALLEGRETTO (405), Ligand Pharmaceuticals, Inc., San Diego, California 92121

PHILIP AVNER (439), Unité de Génétique Moléculaire Murine, Institut Pasteur, 75724 Paris, France

PATRICK A. BAEUERLE (278), Institute of Biochemistry, Albert-Ludwigs-University, D-79104 Freiburg, Germany

LUTZ-PETER BERG (261), Charter Molecular Genetics Laboratory, Thrombosis Research Institute, London SW3 6LR, United Kingdom

EDOVARD BERTRAND (348), Institut Jacques Monod, 75251 Paris, France

JEFFREY H. BOCK (246), Molecular Biology Research Unit, The Upjohn Company, Kalamazoo, Michigan 49007

SUSANNA BUNGE (26), Institut für Humangenetik, Universität-Krankenhaus Eppendorf, D-22529 Hamburg, Germany

DANIELA CASTANOTTO (348), Center for Molecular Biology and Gene Therapy, Loma Linda University School of Medicine, Loma Linda, California 92350

LAWRENCE CHAN (335), Departments of Cell Biology and Medicine, Baylor College of Medicine, Houston, Texas 77030

JIANG FAN CHEN (362), Departments of Pediatrics and Molecular Pharmacology, Diabetes and Metabolic Diseases Research Program, University Medical Center, State University of New York at Stony Brook, Stony Brook, New York 11794

KENNETH R. CHIEN (470), Department of Medicine, Division of Cardiology, Center for Molecular Genetics, and the American Heart Association-Bugher Foundation, Center for Molecular Biology, University of California, San Diego, School of Medicine, La Jolla, California 92093

E. CHUNG (139), Department of Paediatrics, University College London Medical School, London WC1E 6JJ, United Kingdom

DAVID N. COOPER (261), Institute of Medical Genetics, University of Wales, College of Medicine, Cardiff CF4 4XN, United Kingdom

PIETER A. DOEVENDANS (470), Academic Hospital, Maastricht, The Netherlands

JAMES H. EBERWINE (321), Department of Pharmacology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

JONATHAN A. EPSTEIN (40), Department of Medicine, Genetics and Cardiology Divisions, Brigham and Women's Hospital, Harvard Medical School and Howard Hughes Medical Institute, Boston, Massachusetts 02115

STEPHANIE D. FLAGG (321), Department of Pharmacology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

SIGRID FUCHS (26), Institut für Humangenetik, Universität-Krankenhaus Eppendorf, D-22529 Hamburg, Germany

RICARDO FUJITA (207), Department of Ophthalmology, W. K. Kellogg Eye Center, University of Michigan, Ann Arbor, Michigan 48105

ANDREAS GAL (26), Institut für Humangenetik, Universität-Krankenhaus Eppendorf, D-22529 Hamburg, Germany

PATRICIA A. GALVIN-PARTON (362), Departments of Pediatrics and Molecular Pharmacology, Diabetes and Metabolic Diseases Research Program, University Medical Center, State University of New York at Stony Brook, Stony Brook, New York 11794

R. M. GARDINER (139), Department of Paediatrics, University College London Medical School, London WC1E 6JJ, United Kingdom

TOM GLASER (40), Howard Hughes Medical Institute and Department of Internal Medicine, Division of Molecular Medicine and Genetics, University of Michigan, Ann Arbor, Michigan 48109

MARKUS GROMPE (3), Department of Molecular and Medical Genetics, Oregon Health Sciences University, Portland, Oregon 97201

ANN HARRIS (70), Paediatric Molecular Genetics, Institute of Molecular Medicine, Oxford University, John Radcliffe Hospital, Oxford OX3 9DU, United Kingdom

EDITH HEARD (439), Unité de Génétique Moléculaire Murine, Institut Pasteur, 75724 Paris, France

JAMES A. HEJNA (3), Department of Molecular and Medical Genetics, Oregon Health Sciences University, Portland, Oregon 97201

RICHARD A. HEYMAN (405), Ligand Pharmaceuticals, Inc., San Diego, California 92121

JEREMY HULL (70), Paediatric Molecular Genetics, Institute of Molecular Medicine, Oxford University, John Radcliffe Hospital, Oxford OX3 9DU, United Kingdom

JOHN J. HUNTER (470), Department of Medicine, Division of Cardiology, Center for Molecular Genetics, and the American Heart Association-Bugher Foundation, Center for Molecular Biology, University of California, San Diego, School of Medicine, La Jolla, California 92093

TAKASHI ITO (229), Human Genome Center, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108, Japan

ETHYLIN WANG JABS (189), Center for Medical Genetics and Department of Pediatrics, The Johns Hopkins University, Baltimore, Maryland 21218

VIJAY V. KAKKAR (261), Charter Molecular Genetics Laboratory, Thrombosis Research Institute, London SW3 6LR, United Kingdom

P. MEERA KHAN (97), MGC Department of Human Genetics, Faculty of Medicine, Leiden University, 2333 AL Leiden, The Netherlands

MARGARET A. KNOWLES (113), Molecular Genetics Laboratory, Marie Curie Research Institute, The Chart, Oxted, Surrey RH8 0TL, United Kingdom

STEVEN W. KUBALAK (470), Department of Medicine, Division of Cardiology, Center for Molecular Genetics, and the American Heart Association-Bugher Foundation, Center for Molecular Biology, University of California, San Diego, School of Medicine, La Jolla, California 92093

HELIOS LEUNG (375), Bristol-Myers Squibb Pharmaceutical, Research Institute, Seattle, Washington 98121

MING-JIE LI (375), Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, Connecticut 06510

MICHAEL LOVETT (189), The McDermott Center and Department of Biochemistry, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas 75235

RICHARD MAAS (40), Department of Medicine, Genetics Division, Brigham and Women's Hospital, Harvard Medical School and Howard Hughes Medical Institute, Boston, Massachusetts 02115

NANCY MAIZELS (375), Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, Connecticut 06510

CRAIG C. MALBON (362), Department of Molecular Pharmacology, Diabetes and Metabolic Diseases Research Program, University Medical Center, State University of New York at Stony Brook, Stony Brook, New York 11794

EVIE MELANITOU (439), Unité de Génétique Moléculaire Murine, Institut Pasteur, 75724 Paris, France

ROBB E. MOSES (3), Department of Molecular and Medical Genetics, Oregon Health Sciences University, Portland, Oregon 97201

CHRISTOPHER M. MOXHAM (362), Departments of Pediatrics and Molecular Pharmacology, Diabetes and Metabolic Diseases Research Program, University Medical Center, State University of New York at Stony Brook, Stony Brook, New York 11794

HEIKE L. PAHL (278), Institute of Biochemistry, Albert-Ludwigs-University, D-79104 Freiburg, Germany

LEENA PELTONEN (14), Department of Human Molecular Genetics, National Public Health Institute, 00300 Helsinki, Finland

HOWARD A. ROCKMAN (470), Department of Medicine, Division of Cardiology, Center for Molecular Genetics, University of California, San Diego, School of Medicine, La Jolla, California 92093

JOHN ROSS, JR. (470), Department of Medicine, Division of Cardiology, University of California, San Diego, School of Medicine, La Jolla, California 92093

JOHN J. ROSSI (348), Center for Molecular Biology and Gene Therapy, Loma Linda University School of Medicine, Loma Linda, California 92350

CLAIRE ROUGEULLE (439), Unité de Génétique Moléculaire Murine, Institut Pasteur, 75724 Paris, France

YOSHIYUKI SAKAKI (229), Human Genome Center, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108, Japan

DEBORAH A. SCOPES (261), Charter Molecular Genetics Laboratory, Thrombosis Research Institute, London SW3 6LR, United Kingdom

GREGG L. SEMENZA (298), Center for Medical Genetics, Departments of Pediatrics and Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21287

SUE SHACKLETON (70), Paediatric Molecular Genetics, Institute of Molecular Medicine, Oxford University, John Radcliffe Hospital, Oxford OX3 9DU, United Kingdom

MARIE-CHRISTINE SIMMLER (439), Unité de Génétique Moléculaire Murine, Institut Pasteur, 75724 Paris, France

JERRY L. SLIGHTOM (246), Molecular Biology Research Unit, The Upjohn Company, Kalamazoo, Michigan 49007

CORINNE M. SPENCER (321), Department of Pharmacology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

TIAN-QIANG SUN (167), Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514

ANAND SWAROOP (207), Departments of Ophthalmology and Human Genetics, W. K. Kellogg Eye Center, University of Michigan, Ann Arbor, Michigan 48105

ARTHUR J. SYTKOWSKI (421), New England Deaconess Hospital, Department of Medicine, Harvard Medical School, Boston, Massachusetts 02215

ANN-CHRISTINE SYVÄNEN (14), Department of Human Molecular Genetics, National Public Health Institute, 00300 Helsinki, Finland

NOBUAKI TANAKA (470), Department of Medicine, Division of Cardiology, University of California, San Diego, School of Medicine, La Jolla, California 92093

BABIE TENG (335), Department of Medicine, Baylor College of Medicine, Houston, Texas 77030

T. D. TLSTY (388), Department of Pathology, University of California at San Francisco, San Francisco, California 94143

ROB B. VAN DER LUIJT (97), MGC Department of Human Genetics, Faculty of Medicine, Leiden University, 2333 AL Leiden, The Netherlands

JEAN-MICHEL H. VOS (167), Lineberger Comprehensive Cancer Center, Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, North Carolina 27514

GUANG L. WANG (298), Otsuka Research Laboratory, Otsuka American Pharmaceutical, Inc., St. Petersburg, Florida 33701

HSIEN-YU WANG (362), Departments of Pediatrics and Molecular Pharmacology, Diabetes and Metabolic Diseases Research Program, University Medical Center, State University of New York at Stony Brook, Stony Brook, New York 11794

CAROL A. WISE (189), The McDermott Center and Department of Biochemistry, University of Texas Southwestern Medical Center at Dallas, Texas 75235

DENISE YAN (207), Department of Ophthalmology, W. K. Kellogg Eye Center, University of Michigan, Ann Arbor, Michigan 48105

This Page Intentionally Left Blank

Preface

The series *Methods in Molecular Genetics* provides practical experimental procedures for use in the laboratory. Because the introduction of molecular genetic techniques and related methodology has revolutionized biological research, a wide range of methods are covered. The power and applicability of these techniques have led to detailed molecular answers to important biological questions and have changed the emphasis of biological research, including medical research, from the isolation and characterization of cellular material to studies of genes and their protein products.

Molecular genetics and related fields are concerned with genes: DNA sequences of genes, regulation of gene expression, and the proteins encoded by genes. The consequences of gene activity at the cellular and developmental levels are also investigated. In medical research, knowledge of the causes of human disease is reaching an increasingly sophisticated level now that disease genes and their products can be studied. The techniques of molecular genetics are also being widely applied to other biological systems, including viruses, bacteria, and plants, and the utilization of gene cloning methodology for the commercial production of proteins for medicine and industry is the foundation of biotechnology. The revolution in biology that began with the introduction of DNA sequencing and cloning techniques will continue as new procedures of increasing usefulness and convenience are developed.

In addition to the basic DNA methods, instrumentation and cell biology innovations are contributing to the advances in molecular genetics. Important examples include gel electrophoresis and DNA sequencing instrumentation, *in situ* hybridization, and transgenic animal technology. Such related methodology must be considered along with the DNA procedures.

In *Human Molecular Genetics,* methods are outlined for the investigation of gene and chromosome structure in human cells, with an emphasis on disease genes. Also emphasized are topics such as the functioning of the human genome in transcription and at the cellular level in signal transduction. In addition, methods deal with the mouse genome as a model system. The twenty-six chapters in this volume are organized into seven sections presenting related protocols: Mutation Detection in Human Genes; Gene Mapping, Cloning, Sequencing; Transcription: Promoters, Transcription Factors, mRNA; RNA Editing, Ribozymes, Antisense RNA; Genome Recominbation, Amplification; Receptors, Signal Transduction; The Mouse as a Model System for Human Molecular Genetics.

Methods in Molecular Genetics will be of value to researchers, as well as to students and technicians, in a number of biological disciplines because of the wide applicability of the procedures and the range of topics covered.

KENNETH W. ADOLPH

This Page Intentionally Left Blank

Methods in Molecular Genetics

Volume 1	Gene and Chromosome Analysis (Part A) *Edited by Kenneth W. Adolph*
Volume 2	Gene and Chromosome Analysis (Part B) *Edited by Kenneth W. Adolph*
Volume 3	Molecular Microbiology Techniques *Edited by Kenneth W. Adolph*
Volume 4	Molecular Virology Techniques *Edited by Kenneth W. Adolph*
Volume 5	Gene and Chromosome Analysis (Part C) *Edited by Kenneth W. Adolph*
Volume 6	Microbial Gene Techniques *Edited by Kenneth W. Adolph*
Volume 7	Viral Gene Techniques *Edited by Kenneth W. Adolph*
Volume 8	Human Molecular Genetics *Edited by Kenneth W. Adolph*

This Page Intentionally Left Blank

Section I

Mutation Detection in Human Genes

This Page Intentionally Left Blank

[1] Chemical Mismatch Cleavage

James A. Hejna, Markus Grompe, and Robb E. Moses

Introduction

Mutation detection and analysis have become important technical steps in correlating candidate genes with genetic diseases. Chemical mismatch cleavage is a powerful method for scanning large regions of sequence, yet it has been relatively underutilized (1, 2). The method was developed by Cotton *et al.* (3), who systematically optimized the reagents and conditions used in Maxam–Gilbert DNA sequencing for the preferential cleavage of mismatched bases in heteroduplex DNA. The reactions exploit the susceptibility of mismatched T and C residues in heteroduplex DNA to modification by osmium tetroxide and hydroxylamine, respectively, followed by piperidine cleavage at the site of the modified base. The chemical cleavage protocol has been streamlined so that all of the chemical modification and cleavage reactions can be run in a single day, and the samples run on a sequencing gel the following day.

Substrate Design

The chemical mismatch cleavage (CMC) protocol is designed to scan lengths of DNA up to 1.7 kb, with resolution limited primarily by the constraints of a sequencing gel, so primers can be chosen to scan an entire small cDNA in a single set of reactions; a large cDNA can be covered by choosing primers that preferably amplify 1.7-kb fragments that overlap by about 200 bp. For example, a 3.2-kb cDNA could be scanned by amplifying two 1.7-kb fragments overlapping by 200 bp. The method can also be used to scan genomic DNA when it is possible to amplify across small introns. The preparation of RNA for reverse transcription, and a protocol for cDNA synthesis, are not included here. Several kits are commercially available. Typically, 1 μg of total RNA is random-primed and reverse-transcribed with Moloney murine leukemia virus (MMLV) reverse transcriptase, and 1 μl from this reaction is used directly in reverse-transcribed polymerase chain reaction (RT-PCR) for the generation of either probe or target. The same primers are used to amplify both the probe and target. Because the probe is reamplified with end-labeled primers, it is usually not necessary to amplify more than 35 cycles. For the target samples, approximately 100–150 ng of target DNA is required per heteroduplex annealing. It is recommended that 1–2 μg of PCR-amplified target be stockpiled. This might entail pooling several identical PCR reactions.

Methods in Molecular Genetics, Volume 8

Probe Production

While most mismatches will be detected with either the coding or noncoding strand of the wild-type probe, it may be necessary to make an additional probe from mutant DNA. The basic procedure is the same. An initial PCR amplification is carried out, and the product is purified in a Centricon-100 (Amicon, Danvers, MA) unit: the PCR reaction is added to a Centricon-100 unit, diluted to 2 ml with 1 × Tris–EDTA (TE), then concentrated by centrifugation to approximately 40 μl. Alternatively, the PCR product can be cleaned with an ammonium acetate precipitation: the volume of the PCR reaction is first brought to 200 μl with H_2O; then 60 μl of 10 *M* ammonium acetate is added, and the tube is spun for 20 min in a microcentrifuge. The supernatant is transferred to a new tube, and the DNA is precipitated with 600 μl of cold ethanol. After 15 min at −20°C, the sample is spun for 15 min, and the pellet is washed with 70% (v/v) EtOH. The pellet is then resuspended in 200 μl of H_2O, then reprecipitated with 60 μl of 10 *M* ammonium acetate and ethanol. The final pellet is dried and dissolved in H_2O or 1 × TE to a final concentration of ~50 ng/μl. The PCR product is then quantitated by running it alongside standardized size markers on an agarose gel. It is best if the PCR reaction produces a single product; however, PCR products may be purified on agarose gels, if necessary, and PCR Wizard Preps (Promega, Madison, WI) generally give good results. If gel-purified DNA is to be used in subsequent PCR amplifications to produce labeled probes, special care must be taken to ensure that it does not become contaminated during gel purification; this could easily occur, for example, if wild-type and mutant PCR products were purified on the same gel.

Radiolabeling of Oligonucleotide Primers

The objective is to end-label primers to a high specific activity. While most commercial T4 polynucleotide kinase comes with a 10× buffer stock [e.g., 0.5 *M* Tris (pH 7.6), 0.1 *M* $MgCl_2$, 0.15 *M* dithiothreitol (DTT)], the protocol for labeling blunt ends in the first edition of *Molecular Cloning: A Laboratory Manual* (4) works well, and is recommended. The γ-^{32}P]ATP (6000 Ci/mmol, 10 mCi/μl) should be used within 1 week (2 weeks at most) of the activity date. A typical reaction uses 10 μl of [γ-^{32}P]ATP with 4 pmol of primer in a reaction volume of 50 μl. After an incubation at 37°C for 45 min, the enzyme is inactivated by heating at 70°C for 10 min. The primer is then dialyzed by diluting to 2 ml with H_2O, and concentrating in a Centricon-3 (Amicon) unit. Other methods of removing unincorporated nucleotide are also possible.

Reamplification with End-Labeled Primers

Reamplification of the wild-type DNA (or mutant DNA, if that is to be used for the probe) utilizes all of the end-labeled primer together with 4 pmol of the correspond-

ing cold primer in a single PCR reaction with ~5 ng from the initial PCR-amplified DNA. Reamplification should be limited to 25–30 cycles. Owing to the low concentration of primers in the reaction, the PCR conditions might need to be altered. Trial reamplifications with cold primers at similar concentrations can be run in order to check the conditions. It is recommended that the amplified probe be purified on an agarose gel at this stage, and that extraction from the gel be as gentle as possible, to minimize background caused by nicking. The gel is stained with ethidium bromide, and the band is visualized on a 320-nm ultraviolet (UV) light box. It is convenient to lay a sheet of UV-transparent glass or plastic wrap (e.g., Saran Wrap) on the light box beforehand, to minimize contamination of the light box with radioactivity. Local guidelines for handling and disposal of radioactivity must be followed. Electroelution of the excised DNA directly into Centricon-100 units works well, allowing subsequent dialysis versus 1 × TE and concentration in the same unit, with good recovery. A device for simultaneous electroelution of up to 12 samples is marketed by Amicon. Other methods, such as Spin-X (Costar, Cambridge, MA) or Micropure (Amicon), might also work, with slightly lower recoveries. After purification, the probes must be quantitated. The DNA concentration is estimated by running 1 μl alongside standard markers on an agarose gel, and also counting 1 μl in a scintillation counter. A minimum specific activity should be approximately 2×10^3 cpm/ng of probe, or 2×10^4 cpm/heteroduplex annealing. With fresh [γ-^{32}P]ATP, specific activities of 1×10^4 cpm/ng of probe can be achieved. The probe can be stored for up to 3 weeks before use; however, radiolysis leads to increasing background over time.

Heteroduplex Formation

Heteroduplexes are formed by combining probe with an excess of target, heat denaturing, and then reannealing. Cleaner backgrounds have been obtained by annealing under more stringent conditions than previously described (3). Approximately 100–150 ng of target DNA and 10 ng of probe ($\geq 2 \times 10^4$ cpm) are combined in 0.3 *M* NaCl–3.5 m*M* $MgCl_2$–3 m*M* Tris, pH 7.7. The final volume is typically 30 μl, although larger volumes have been used. The DNA is denatured by boiling for 5 min, then it is immediately transferred to a 65°C water bath and annealed for 1 hr. After 1 hr, the DNA is precipitated with 90 μl of cold ethanol. The samples are chilled for 15 min at −80°C, warmed briefly to 4°C, and spun for 15 min in a microcentrifuge at 4°C. The supernatant is drawn off and monitored with a Geiger counter to ensure that the pellet is not lost. The pellets are then dried [a 70% (v/v) ethanol rinse is not necessary at this step] and resuspended in 13 μl of sterile H_2O. The pellets might not go into solution immediately. Throughout the chemical cleavage protocol a Geiger counter should be used to monitor precipitation and resuspension of the samples.

A high G-C content can make full denaturation difficult in the above buffer. A simpler buffer, 100 m*M* NaCl–2 m*M* EDTA, pH 8, has been used, and the DNA precipitated after the annealing with 1/10 vol of 3 *M* sodium acetate, pH 7, and 2 vol

of ethanol. Alternatively, the DNA can be denatured by boiling in 1× TE, then adjusting the buffer to the standard conditions.

Chemical Cleavage

Siliconized 0.7-ml microfuge tubes are recommended for the chemical cleavage reactions. Colored tubes are convenient for keeping track of the different reactions; it is better to use light blue or green for the OsO_4 reactions in order to observe a yellowish precipitate that forms on addition of the OsO_4. In most instances up to 40 samples might be analyzed at once. An aluminum block drilled with an array of holes to fit 0.7-ml tubes is particularly convenient, as the tubes can be organized in the block and chilled on ice; the entire block can then be transferred to the −80°C freezer during the ethanol precipitations, and warmed rapidly by partially immersing it in a water bath. A heating apparatus often contains such removable aluminum blocks.

The heteroduplex DNA is divided into two separate, 0.7-ml siliconized tubes (6 μl/tube): one for the OsO_4 reaction, and one for the hydroxylamine reaction. The tubes are kept chilled on ice. The hydroxylamine reaction is initiated by adding 20 μl of hydroxylamine solution (see Materials, below). Mix well and incubate for 15–30 min at 37°C. The reactions are then chilled on ice and stopped by adding 100 μl of hot stop buffer, vortexing, then precipitating with 250 μl of cold ethanol. The samples are then chilled for 15 min at −80°C, warmed briefly to 4°C (to reduce the viscosity of the ethanol), and centrifuged for 15 min at 4°C. The supernatant is carefully drawn off, and then the pellet is rinsed with 400 μl of 70% (v/v) ethanol and spun for 5 min at 4°C. It is critical to rinse the pellet twice with 70% (v/v) ethanol. A single rinse does not remove traces of hydroxylamine, which leads to degradation of the DNA during the piperidine reaction. The pellets are dried briefly, then chilled on ice.

The OsO_4 reactions are initiated by combining the 6 μl of heteroduplex DNA with 10 μl of 2.5× OsO_4 buffer, mixing briefly, and then adding 8 μl of 4% (w/v) OsO_4. If droplets cling to the walls of the tube, the tube can be tapped gently to bring them down. A yellowish precipitate will form. The reaction is started by transferring the tubes to a 37°C water bath and incubating for 5 min. It is important to keep the precipitate mixed by removing tubes from the water bath and agitating them several times, then returning them to the water bath. In most instances the precipitate will gradually dissolve during the 5-min incubation, but the occasional failure of the precipitate to dissolve does not seem to affect the quality of the reaction. Because each tube requires some manipulation during the incubation, it is better to incubate fewer than 15 tubes at a time. Stopping the reaction is the same as for hydroxylamine: the tubes are chilled on ice, returned to the fume hood, and 100 μl of hot stop buffer is added, followed by 250 μl of cold ethanol. The tubes are then vortexed vigorously. Any residual yellow precipitate should go into solution at this point. The tubes are

then chilled at −80°C and spun for 15 min at 4°C. The pellets are rinsed with 400 μl of cold 70% (v/v) ethanol, spun for 5 min at 4°C, then dried briefly and chilled on ice. Unlike the hydroxylamine reactions, a single 70% (v/v) ethanol rinse is adequate at this step. All OsO_4 liquid waste, including the 70% (v/v) ethanol rinse, should be collected and disposed of according to local guidelines; note that the liquid OsO_4 waste will be slightly radioactive as well. It can usually be stored in the fume hood long enough to allow decay of the ^{32}P before it is disposed of as toxic waste.

Piperidine Cleavage

To each pellet, chilled on ice, 50 μl of 1 *M* piperidine is added, and mixed vigorously. The pellet will readily go into solution. The piperidine cleavage reaction is easily carried out in a thermocycler programmed to heat for 15 min and then to chill at 0°C. The reaction tubes should be placed in the machine, and then weighted down to ensure that the lids do not pop open during the 90°C incubation. The OsO_4 tubes normally turn a bit black during the piperidine cleavage reaction. After the 15-min incubation, the tubes are returned to ice, and the DNA is precipitated by adding 12 μl of 3 *M* ammonium acetate, pH 7, containing tRNA (1.6 mg/ml, or 2 μg/tube), followed by 150 μl of cold ethanol. The addition of carrier tRNA at this step is important; it greatly facilitates resuspending the pellets after ethanol precipitation. The tubes are then chilled at −80°C and spun as before. A single 70% (v/v) ethanol rinse of the pellet is sufficient. The pellets are dried, and then resuspended in 6–10 μl of formamide loading buffer. It may be necessary to mix vigorously or scrape the walls of the tube with a pipette tip to get the pellet back into solution. Samples can be stored at −20°C up to several days before they are run on a sequencing gel.

Electrophoresis

A 4% (w/v) denaturing polyacrylamide sequencing gel usually gives good resolution between 30 and 1500 bp. Wedge gels can be poured, and there is usually enough radioactivity in each sample for two or three loadings if resolution between 30 and 1700 bp is required. Shorter fragments can also be run on normal 6% (w/v) sequencing gels. Size markers can be prepared from commercially available φX174 *Hae*III-digested DNA: the DNA is first dephosphorylated with alkaline phosphatase, then end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. As with DNA sequencing, the samples must be denatured by heating at 85°C for 3 min, then chilling on ice just prior to loading the gel. Usually 3–4 μl is loaded at a time. The gel is run as a normal sequencing gel (e.g., 2000 V) until the bromphenol blue tracking dye has just reached the bottom of the gel. After electrophoresis, the gel is taken down and trans-

ferred directly to a piece of Whatman (Clifton, NJ) 3MM paper, and dried without fixing. A good signal generally can be detected with an overnight exposure, but exposures of several days may be needed.

Interpretation of Results

Frameshift mutations, either small deletions or insertions, are the most easily detected, as they produce strong signals with both the hydroxylamine and OsO_4 reactions; in addition, when the results of cleavage of the coding and noncoding strands are compared, the corresponding sizes of the cleavage fragments should add up to that of the full-length probe. Whether the alteration is an insertion or a deletion cannot easily be determined by this method.

Base substitutions lead to mismatches that can be detected in virtually 100% of the cases. For example, a change from a C in the wild-type coding strand to a T would yield a corresponding end-labeled C mispaired with a target A; this would be cleaved by hydroxylamine but not by OsO_4. The other heteroduplex, in which the noncoding strand is end-labeled, would yield a labeled G mispaired with a target T, and no (radioactive) signal would be observed, either with hydroxylamine or OsO_4. In this case, a signal would be generated from only one of the four reactions. Accurate graphing of the size markers can usually locate the site of the base substitution to within 5–6 bp; if the wild-type sequence is known, the precise location can often be predicted. Once a mismatch has been mapped, DNA sequencing should be employed locally to confirm the base alteration. While a combination of both wild-type and mutant probes will accurately predict the exact nature of the base alteration, that may not be the immediate objective in a preliminary scanning of a candidate gene. The use of wild-type end-labeled coding strand and noncoding strand probes is sufficient to detect and localize mismatches, which can then be sequenced by conventional methods. Mismatch signals are not always unambiguous; for example, a T·A base pair adjacent to a C·C mismatch will often be cleaved by OsO_4. A homologous run of T's within a coding region gives a ladder pattern when reacted with OsO_4, due to "breathing" of the DNA.

Because both the modification reactions and the piperidine cleavage reactions are partial reactions, slight variations in technique might alter the extent of the reactions, and a positive control carrying a known mismatch is a helpful diagnostic. Applications of the technique have included an analysis of mutations in the human dystrophin gene (5), the Wilms' tumor gene (6), and a steroid sulfatase (7).

Streamlining Procedure

Many times, mismatch scanning has been used to test a candidate gene for the presence of mutations within the coding region. At the outset, chemical mismatch cleav-

age can be used to screen samples from a large number of normal and affected individuals. If the primary objective is to detect a base alteration, the coding and noncoding strand probes can be combined; because the target DNA is present in the heteroduplex annealings in a 10-fold molar excess, heteroduplexes with probe DNA will still be favored over probe homoduplexes. This cuts the number of individual reactions in half. Only those samples producing a mismatch cleavage signal in the first round of reactions need to be analyzed further, when a second set of heteroduplexes is prepared using each probe separately. An example is presented in Fig. 1. Both the coding and noncoding probes were combined in the CMC scanning of a novel cDNA, 51C, that was mapped to human chromosome 11q23. The positive con-

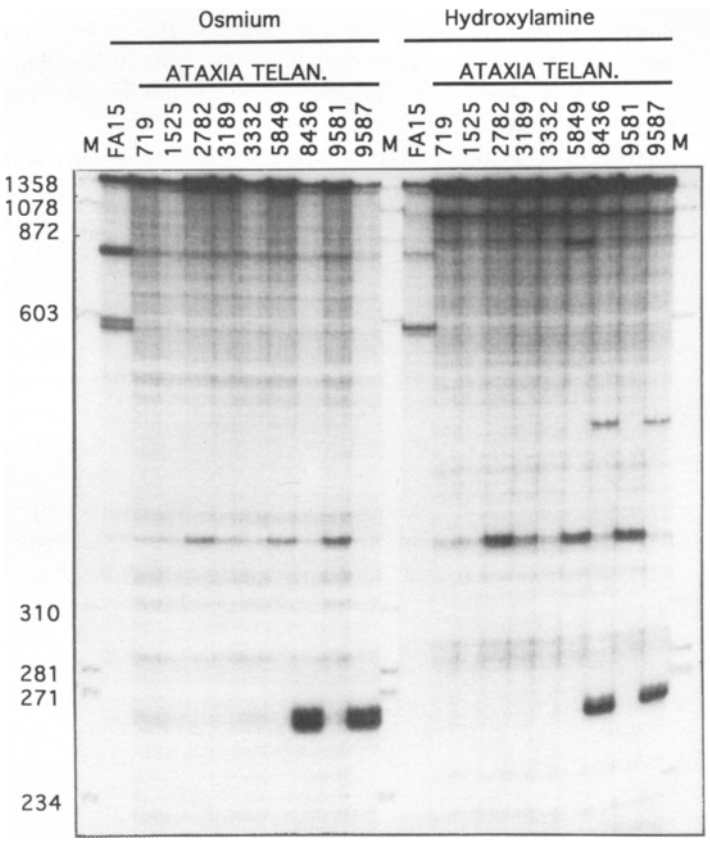


FIG. 1 Chemical mismatch cleavage of the 51C genomic DNA, combined coding and anticoding probes. Several cleavage signals can be seen. Small deletions, such as for FA15, generate strong signals. Single base substitutions, such as the signal at 450 nucleotides in AT 8436, are weaker.

trol in this case was genomic DNA from a Fanconi anemia complementation group A patient. The 3′-UTR (untranslated leader region) carries a precise deletion of a 16-bp sequence that is normally tandemly repeated. This accounts for the doublet at about 600 bp on the gel. Because both probes were combined in the reaction, an additional strong band is seen at about 800 bp. This, too, is a doublet, arising from the wild-type complementary probe looping out one or the other of the tandem repeats during heteroduplex formation. The signals add up to the size of the full-length probe. Among the ataxia telangiectasia samples scanned, there are several polymorphisms present. It is impossible to tell from this experiment whether the mismatch at about 450 bp in AT 8436 and AT 9587 is closer to the 5′-end of the heteroduplex or the 3′ end. A second cleavage experiment with individual probes would unambiguously map the signal. The doublet at about 260 bp in samples AT 8436 and AT 9587 is due to the deletion of 4 bases. In this case, the noncoding strand has only unpaired GAAG, and is not cleaved. As can be seen, simple base substitutions generate weaker signals than deletions. Sample AT 5849 carries a second polymorphism at about 870 bp. This was more pronounced when the experiment was repeated with separate probes.

With many samples a repeat pipettor would be helpful. If all of the target DNAs are adjusted to a uniform concentration (e.g., 50 ng/μl), the heteroduplex annealings can be set up more rapidly by combining buffer, probe, and H_2O in a master mix. Depending on the number of samples to be analyzed, it may be convenient to mobilize several microcentrifuges. The cleavage reactions can usually be dovetailed so that one batch of reactions can be precipitated or centrifuged while another batch is being incubated. Rinsed, dried DNA pellets can be stored at $-20°C$ overnight, so that the protocol can be interrupted after heteroduplex formation, chemical modification, or piperidine cleavage. Once the heteroduplexes have been modified, however, it is best to proceed immediately to the piperidine cleavage step. The DNA pellets do not even have to be dried thoroughly before they are resuspended in piperidine, although some care should be taken with the hydroxylamine reactions to remove as much of the first 70% (v/v) ethanol rinse as possible.

Figure 2 presents a flow chart of the steps in chemical mismatch cleavage.

Safety

Hydroxylamine, osmium tetroxide, and piperidine are all hazardous materials, and should be handled carefully. All work with concentrated stock solutions, and all work where tubes containing osmium tetroxide are opened, should be carried out in a chemical fume hood with adequate air flow. Generally, the hydroxylamine-treated samples and the OsO_4-treated samples are considered safe to open to the air after the final 70% (v/v) ethanol rinse, and the pellets can be dried in a Speedvac (Savant, Hicksville, NY). The same applies to the samples treated with piperidine: once the

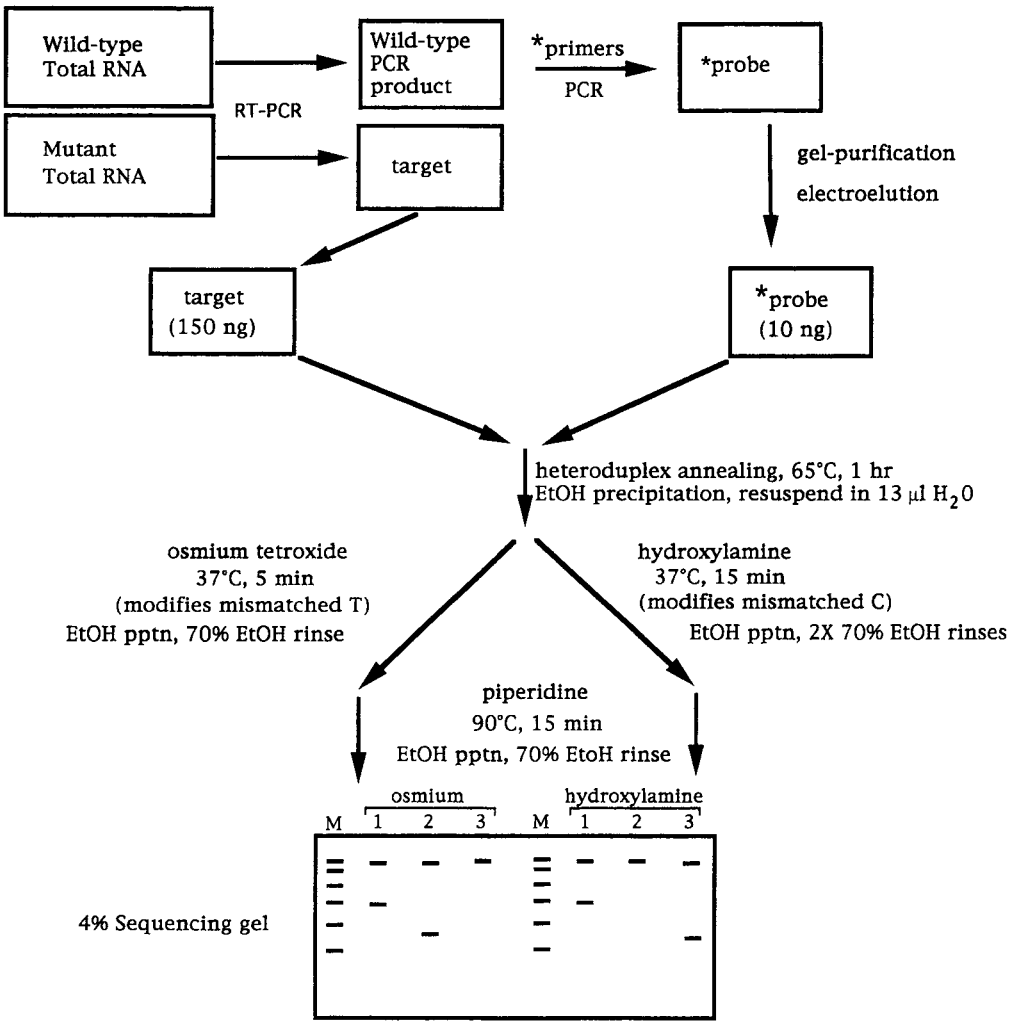


FIG. 2 Flow chart for chemical mismatch cleavage. Once the target and probe DNAs have been amplified and purified, the heteroduplex annealing and cleavage reactions can be carried out in a single day. The samples can be stored overnight at −20°C and run on a 4% sequencing gel the next day.

DNA pellets have been rinsed with 70% (v/v) ethanol, the tubes can be opened to the air, which facilitates the final resuspension in loading buffer prior to electrophoresis. All ethanol supernatants should be treated either as radioactive waste (heteroduplex

annealings) or toxic waste (hydroxylamine and osmium liquid waste should be collected in separate bottles).

Materials

Annealing buffer (10×): 3 *M* NaCl, 30 m*M* Tris (pH 7.7), 35 m*M* $MgCl_2$; autoclave; store at room temperature

Hydroxylamine solution: Dissolve 1.39 g of solid hydroxylamine chloride [Fluka (Ronkonkoma, NY), Aldrich (Milwaukee, WI), or BDH Analar] in 1.6 ml of H_2O in a sterile, polypropylene screw-cap tube, and warm gently (e.g., in a 37°C water bath) with periodic vortexing until dissolved. Add 1 ml of diethylamine (Fluka or Aldrich). Measure the pH by adding 2 drops from the stock to 2 ml of H_2O in a disposable tube, then measuring the pH of this dilution. Do not insert the pH electrode directly into the hydroxylamine stock. The pH can be adjusted to 6.0 by adding diethylamine in 50-μl increments, checking the pH each time. The stock solution can be stored at 4°C for 7–10 days. Storage at 4°C often causes some precipitation in the form of large crystals. These can be redissolved by gentle heating and vortexing

Osmium tetroxide buffer (2.5×): 25 m*M* Tris (pH 7.7), 4 m*M* EDTA, 3.75% pyridine (HPLC grade; Aldrich). Make 1 ml at a time. The buffer can be stored at −20°C for at least 1 month

Osmium tetroxide (4%): Purchase directly from Sigma (St. Louis, MO) in a sealed ampoule. On opening, the contents are transferred with a glass Pasteur pipette to a clean glass vial. Prior to storage, the vial is tightly capped, wrapped securely with Parafilm, and then once more with time tape. The vial is then wrapped in aluminum foil and carefully packed into the metal tin in which it arrived, and stored at 4°C. Local guidelines for storage might be more strict, and should be referred to before the osmium is ordered. When stored as described above, the solution remains active for more than 6 months. Eventually, the solution turns grayish green, and the intense yellow color reaction that occurs during the mixing with pyridine fades, indicating a loss of reactivity

Hot stop buffer: 0.2 *M* NaCl, 10 m*M* Tris (pH 7.7), 1 m*M* EDTA, tRNA (50 μl/ml). Autoclave the Tris–NaCl–EDTA stock, and add tRNA fresh for each experiment. A tRNA (baker's yeast; Boehringer-Mannheim, Indianapolis, IN) is prepared by dissolving 5 mg of tRNA in 1 ml of H_2O. The tRNA is then phenol extracted, phenol–chloroform extracted, chloroform extracted, precipitated with ethanol, and dissolved in 0.5 ml of sterile H_2O. The tRNA is then heat treated at 95°C for 10 min. The concentration is determined spectrophotometrically, and the final concentration adjusted accordingly with H_2O to 5 mg/ml. The tRNA stock can be stored at −20°C

Piperidine (1*M*): Combine 1 ml of piperidine (Fluka) with 9 ml of water in a clean glass vial or bottle. The solution can be stored at 4°C for up to 1 month

Formamide loading buffer: 95% Formamide, 20 m*M* EDTA (pH 8.0), 0.05% (v/v) bromphenol blue, 0.05% (w/v) xylene cyanol

Wherever possible, solutions should be autoclaved.

References

1. M. Grompe, *Nature Genetics* **5,** 111 (1993).
2. R. G. H. Cotton, *Mutation Res.* **285,** 125 (1993).
3. R. G. Cotton, N. R. Rodriques, and R. D. Campbell, *Proc. Natl. Acad. Sci. U.S.A.* **85,** 4397 (1988).
4. T. Maniatis, E. F. Fritsch, and J. Sambrook, "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1982.
5. M. W. Kilimann, A. Pizzuti, M. Grompe, and C. T. Caskey, *Hum. Genet.* **89,** 253 (1992).
6. M. H. Little, J. Prosser, A. Condie, P. J. Smith, V. V. Heyningen, and N. D. Hastie, *Proc. Natl. Acad. Sci. U.S.A.* **89,** 4791 (1992).
7. E. Basler, M. Grompe, G. Parenti, J. Yates, and A. Ballabio, *Am. J. Hum. Genet.* **50,** 483 (1992).

[2] Detection of Point Mutations by Solid-Phase Minisequencing

Ann-Christine Syvänen and Leena Peltonen

Introduction

Many mutation types, including large alterations of chromosomal structures, rearrangements, extensions, deletions, or insertions of varying size, have been identified in the human genome. However, 95% of the identified disease-causing mutations are point mutations affecting one or a few nucleotides (1). Because genetic disorders rarely are caused by a single mutation per disease gene in a population, the ability to detect several mutations simultaneously per sample is of central importance. Consequently there is a need for reliable and technically simple methods for detecting point mutations both in clinical diagnostics and in research laboratories studying human genetic disorders. Another target for monitoring single-nucleotide changes is polymorphisms that have been estimated to occur on the average at 1 nucleotide out of 500 in the human genome (1). Analysis of this allelic variation can be utilized in tissue typing, in the identification of individuals, and in population genetic studies. In the diagnosis of infectious diseases, efficient methods for detecting sequence variants of genes are also required to identify virulent forms of microbes or resistance to drugs, as well as for typing bacteria and viruses in epidemiological studies.

The methods currently used for detecting single-nucleotide variations rely in most cases on amplification of a DNA fragment spanning the variable nucleotide by the polymerase chain reaction (PCR) (2), which allows both specific and sensitive analysis of the target DNA sequence. Mutant and normal sequences can be distinguished using sequence-specific oligonucleotides either as hybridization probes or as primers in the PCR reaction, or by using nucleic acid-specific enzymes, such as restriction enzymes, DNA polymerases, or DNA ligases. Numerous modifications, combinations, and formats of the above-mentioned assay principles to distinguish between sequence variants have been developed (for reviews, see, e.g., Refs. 3 and 4).

We have devised a convenient method, solid-phase minisequencing, for the detection of single-nucleotide variations, small deletions, or insertions in DNA fragments amplified by the PCR (5), and have applied it for detection of numerous disease-causing mutations and single-nucleotide polymorphisms in human genes (for a summary see Ref. 6 and Table I). The major advantage of this method is that its high specificity allows unequivocal identification of any nucleotide variation under the same reaction conditions. Because the assay is carried out in a solid-phase format it comprises simple manipulations in a microtiter well or test tube format without gel-

Methods in Molecular Genetics, Volume 8

TABLE I Applications of Solid-Phase Minisequencing Method

Application area	Gene	Purpose	Ref.[a]
Diagnosis of monogenic disorders			
Aspartylglucosaminuria	Aspartylglucosaminidase	Carrier screening	*b*
Cystic fibrosis	*CFTR*	Carrier screening	*c*
Familial amyloidosis of Finnish type	Gelsolin	Presymptomatic diagnosis	*d*
α_1-Antitrypsin deficiency	α_1-Antitrypsin	Presymptomatic diagnosis	*e*
Marfan syndrome	Fibrillin	Quantification of mutant transcripts	*f*
Congenital FXIII deficiency	Coagulation factor XIII	Quantification of mutant transcripts	*g*
Diagnosis of mitochondrial diseases			
MERRF (myclonus with epilepsy and with ragged red fibers)	$tRNA_{Lys}$	Diagnosis, quantification of heteroplasmy	*h*
MELAS (mitochondrial encephalopathy, lactic acidosis, and strokelike episodes)	$tRNA_{Leu}$	Diagnosis, quantification of heteroplasmy	*i*
Leber's hereditary optic neuretinopathy	*ND2, ND4*	Diagnosis, quantification of heteroplasmy	*j*
NARP (neurogenic muscle weakness, ataxia, and retinis pigmeutosa)	Subunit of ATP synthase	Diagnosis, quantification of heteroplasmy	*k*
Identification of somatic mutations			
Acute myeloid leukemia, myelodysplastic syndromes	N-*ras*	Diagnosis, follow-up	*l*
Pancreatic carcinoma	K*ras*	Diagnosis	*m*
Identification of polymorphic nucleotides			
	Apolipoprotein E	Allelic identification	*n*
	Coagulation factor V	Allelic identification	*o*
	Set of 12 biallelic markers	Forensic and paternity analyses	*p*
	Set of 12 biallelic markers	Genetic mapping	*q*

[a] The primer sequences and PCR conditions are given in these references.

[b] A.-C. Syvänen, E. Ikonen, T. Manninen, M. Bengström, H. Söderlund, P. Aula, and L. Peltonen, *Genomics* **12,** 590 (1992).

[c] A. Jalanko, J. Kere, E. Savilahti, M. Schwartz, A.-C. Syvänen, M. Ranki, and H. Söderlund, *Clin. Chem.* **38,** 39 (1992).

[d] T. Paunio, S. Kiuru, V. Hongell, E. Mustonen, A.-C. Syvänen, M. Bengström, J. Palo, and L. Peltonen, *Genomics* **13,** 237 (1992).

[e] L. Harju, T. Weber, L. Alexsandrova, M. Lukin, M. Ranki, and A. Jalanko, *Clin. Chem.* **2,** 2282 (1993).

[f] K. Kainulainen, L. Y. Sakai, A. Child, F. M. Pope, L. Puhakka, L. Ryhänen, A. Palotie, I. Kaitila, and L. Peltonen, *Proc. Natl. Acad. Sci. U.S.A.* **89,** 5917 (1992).

(*continued*)

electrophoretic separation steps. Because the results of the assay are obtained as numeric values, their interpretation requires no special expertise and the numeric format facilitates computer-assisted handling of the data.

Principle of Method

Figure 1 illustrates the principle of the solid-phase minisequencing method. A DNA fragment spanning the site of the mutation is first amplified using one biotinylated and one unbiotinylated PCR primer. The amplified DNA fragment carrying a biotin residue in the 5′ end of one of its strands is captured on a solid support, taking advantage of the biotin–avidin interaction. The excess of unbiotinylated primer and nucleoside triphosphates (dNTPs) from PCR is removed by washing the solid support, and the unbiotinylated strand of the amplified fragment is removed by alkaline denaturation. The mutant and normal nucleotides are distinguished in the captured DNA strand by two separate "minisequencing" reactions. In a minisequencing reaction a DNA polymerase is used to specifically extend the 3′ end of an oligonucleotide primer that anneals immediately upstream of a variable nucleotide position with a single labeled dNTP complementary to the nucleotide at the variable position. The incorporated labeled dNTP serves as a highly specific indicator of the nucleotide present at the variable site of the template. In samples from homozygous individuals a labeled dNTP will be incorporated in only one of the reactions and in samples from heterozygous individuals a dNTP will be incorporated in both reactions. After the minisequencing reaction the amount of incorporated label is measured and the ratio between the labels incorporated in the two reactions defines the genotype of the sample.

TABLE I (*continued*)

[g] H. Mikkola, M. Syrjälä, V. Rasi, E. Vahtera, E. Hämäläinen, L. Peltonen, and A. Palotie, *Blood* **84,** 517 (1994).
[h] A. Suomalainen, P. Kollmann, J.-N. Octave, H. Söderlund, and A.-C. Syvänen, *Eur. J. Hum. Genet.* **1,** 88 (1993).
[i] A. Suomalainen, A. Majander, H. Pihko, L. Peltonen, and A.-C. Syvänen, *Hum. Mol. Genet.* **2,** 525 (1993).
[j] V. Juvonen, K. Huoponen, A.-C. Syvänen, P. Aula, E. Nikoskelainen, and M. Savontaus, *Hum. Genet.* **93,** 16 (1994).
[k] P. Mäkelä-Bengs, A. Suomalainen, A. Majander, J. Rapola, H. Kalimo, A. Nuutila, and H. Pihko, *Pediatr. Res.* **37,** 634 (1994).
[l] A.-C. Syvänen, H. Söderlund, E. Laaksonen, M. Bengtström, M. Turunen, and A. Palotie, *Int. J. Cancer* **50,** 713 (1992).
[m] J. Ihalainen, M. Taavitsainen, T. Salmivaara, and A. Palotie, *J. Clin. Pathol.* **47,** 1082 (1994).
[n] A.-C. Syvänen, K. Aalto-Setälä, L. Harju, K. Kontula, and H. Söderlund, *Genomics* **8,** 684 (1990).
[o] A. Palotie, unpublished.
[p] A.-C. Syvänen, A. Sajantila, and M. Lukka, *Am. J. Hum. Genet.* **52,** 46 (1993).
[q] J. Aaltonen, J. Komulainen, A. Wikman, A. Palotie, C. Wadelius, J. Peheentupa, and L. Peltonen, *Eur. J. Hum. Genet.* **1,** 164 (1993).

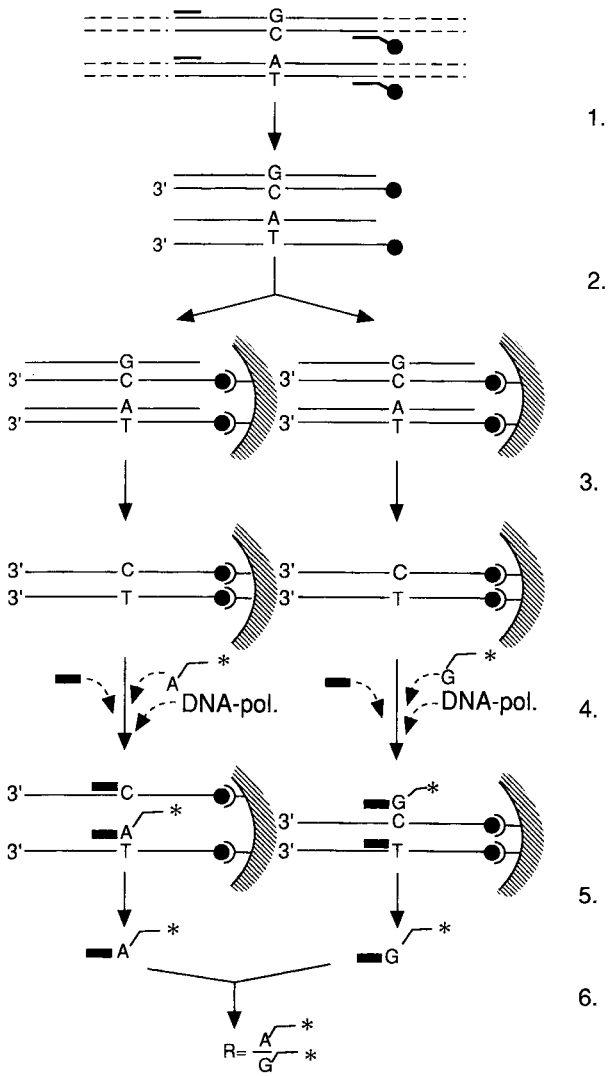


FIG. 1 Steps of the solid-phase minisequencing method. (1) PCR with one biotinylated and one unbiotinylated primer. (2) Affinity-capture of the biotinylated PCR product in streptavidin-coated microtiter wells. (3) Washing and denaturation. (4) The minisequencing primer extension reaction. (5) Measurement of the incorporated label. (6) Calculation of the result.

Performance of Assay

Equipment and Reagents

The following equipment is needed for carrying out the solid-phase minisequencing assay: access to oligonucleotide synthesis, programmable heat block for PCR, shaker at 37°C, water bath for incubation at 50°C, liquid scintillation counter, and optionally, multichannel pipette or microtiter plate washer.

Required materials and reagents include the following: thermostable DNA polymerase for PCR and for the minisequencing reaction (*Taq* DNA polymerase; Promega Biotech, Madison, WI), ^{3}H-labeled deoxynucleoside triphosphates (dATP, TRK 625; dCTP, TRK 576; dGTP, TRK 627; dTTP, TRK 633; Amersham, Arlington Heights, IL), streptavidin-coated microtiter plates (Combiplate 8; Labsystems, Helsinki, Finland), and biotinylphosphoramidite reagent for biotinylation of one of the PCR primers (RPN 2012; Amersham). Reagents and materials from sources other than those indicated here can also be used. Reagents of the highest purity grade are used for preparation of buffers and other solutions.

Design of Primers

A 5′ and a 3′ PCR primer and a detection step primer for the minisequencing reaction are required. The PCR primers should yield an amplification product spanning the variable nucleotide position preferably between 80 and 200 base pairs (bp) in size. The PCR primers should be 20–23 nucleotides long and have similar melting temperatures and noncomplementary 3′ ends (7). The 5′ end of one of the PCR primers is biotinylated in the last step of the synthesis. The minisequencing detection step primer should be 20 nucleotides long and complementary to the biotinylated strand of the PCR product immediately 3′ of the variable nucleotide position (Fig. 2). To detect a deletion or an insertion, the first nucleotide of the deletion/insertion should differ from the first nucleotide following the deletion/insertion.

Procedure

The steps are numbered as in Fig. 1.

Step 1: Polymerase Chain Reaction

Any type of DNA sample, treated as is suitable for PCR amplification (8), can be analyzed. RNA is amplified after synthesis of a first-strand cDNA by reverse transcriptase (9).

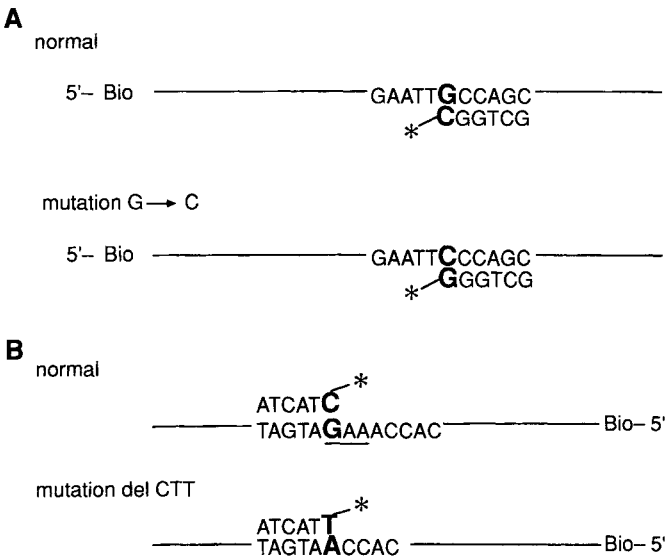


FIG. 2 Example of detection of (A) a G-to-C transversion and (B) a 3-bp (CTT) deletion by solid-phase minisequencing. Bio, Biotinylated primer; the variable nucleotide and the incorporated labeled nucleotide are indicated in boldface type; asterisk (*) symbolizes the label on the incorporated nucleotide; and the five 3′ nucleotides of the minisequencing detection step primer annealing immediately next to the variable nucleotide position are shown.

The PCR is most conveniently carried out in a 50-μl volume. Prepare a master mix containing 5 μl of 10× concentrated *Taq* DNA polymerase buffer [500 m*M* Tris-HCl (pH 8.8), 150 m*M* $(NH_4)_2SO_4$, 15 m*M* $MgCl_2$, 7% (v/v) Triton X-100, 0.7% (w/v) gelatin], 5 μl of a mixture of all four dNTS at 2 m*M* concentration, and dTTP, 2.5 μl of 20 μM unbiotinylated PCR primer, 2.5 μl of 4 μM biotinylated primer, and distilled water to 35 μl/reaction. Add 35 μl of the PCR reagent mixture to 10 μl of a solution containing 30 to 3 × 10^4 molecules of the DNA template. Initiate the PCR by a "hot start" by first incubating the samples in a programmable heat block for 3 min at 95°C, followed by addition of 1.25 units of *Taq* DNA polymerase in 5 μl at 80°C. Carry out 30 PCR cycles of 1 min at 95°C, 1 min at an annealing temperature (which is determined by the sequence of the primers, usually between 50 and 60°C), and 1 min at 72°C.

Comment

Precautions for avoiding DNA contamination in the PCR should be taken (10). Obviously other PCR conditions than those given above can be applied. However, two points of importance for the solid-phase minisequencing method should be considered.

1. The biotin-binding capacity of the streptavidin-coated microtiter wells used to capture the biotinylated PCR products (and excess of biotinylated PCR primer) sets an upper limit to the amount of biotinylated primer that can be used in the PCR. The biotin-binding capacity of the streptavidin-coated wells we use is about 2 pmol. Therefore, in the standard protocol we use 10 pmol of the biotinylated primer per PCR reaction and analyze one-fifth of the PCR product per minisequencing reaction. If "multiplex" PCR is used for amplification of more than one locus, a smaller aliquot of the PCR product should be analyzed. For multiplex PCRs of four loci we found it optimal to detect the polymorphic nucleotides in $\frac{1}{20}$ of the PCR product per well (11). Avidin- or streptavidin-coated microparticles with significantly higher biotin-binding capacity than microtiter plates can also be used as solid support (12).

2. A prerequisite for successful use of the solid-phase minisequencing method with ^{3}H (which has a low specific activity) as label is that the PCR amplification is efficient. Ten microliters of the PCR product should be clearly visible in an agarose gel by staining with ethidium bromide.

Step 2: Capturing Reaction

Transfer two 10-μl aliquots of the amplified sample to streptavidin-coated microtiter wells. Include negative controls containing 10 μl of *Taq* DNA polymerase buffer for both minisequencing reactions. Add 40 μl of 20 m*M* sodium phosphate buffer (pH 7.5), 100 m*M* NaCl, and 0.1% (v/v) Tween 20 to each well. Seal the wells with a sticker and incubate the microtiter plate for 1.5 hr at 37°C with gentle shaking. Discard the contents of the wells. Wash the wells at room temperature three times by adding 200 μl of washing solution [40 m*M* Tris-HCl (pH 8.8), 1 m*M* EDTA, 50 m*M* NaCl, 0.1% (v/v) Tween 20]. Empty the wells thoroughly between the washes.

Comment

It is important for the specificity of the minisequencing reaction that the excess of dNTPs present during the PCR are completely removed by the washing steps. The use of an automatic microtiter plate washer saves time and improves the washing efficiency.

Step 3: Denaturation

Add 100 μl of 50 m*M* NaOH to each well, and incubate at room temperature for 2–5 min. Discard the contents of the wells and wash as described in step 2.

Step 4: Minisequencing Reaction

Prepare two master mixes, one for detection of the mutant nucleotide and the other for detection of the normal nucleotide, by combining 5 μl of 10× *Taq* DNA polymerase buffer (see step 1 above), 2 μl of 5 μ*M* detection step primer, 0.1 μCi (usually 0.1 μl) of ^{3}H-labeled dNTP complementary to the nucleotide to be detected, 0.1 unit of *Taq* DNA polymerase, and distilled water to 50 μl/reaction. Prepare the master

mix during the capturing reaction and store at room temperature until use. For each sample, add 50 μl of minisequencing reaction mix for detection of the mutant nucleotide to one well and 50 μl of reaction mix for detection of the normal nucleotide to another well. Incubate at 50°C for 10 min. Discard the contents of the wells and wash as described in step 2.

Comment

1. The reaction conditions for annealing the primer to the immobilized DNA strand are nonstringent. Therefore, the same reaction conditions can be used for analysis of any DNA fragment, irrespectively of the nucleotide sequence of the detection step primer.
2. An advantage of using a thermostable polymerase for the minisequencing primer extension reaction is that annealing of the primer to the template and extension of the primer with the labeled dNTP can be carried out simultaneously at a fairly high temperature (50°C) that is favorable for both reactions.
3. Deoxynucleoside triphosphates labeled with other radioisotopes (^{32}P or ^{35}S) (5) or with haptens (13) can also be used.

Step 5: Measurement of Incorporated Label

Release the primer after the minisequencing reaction by incubating the microtiter well with 60 μl of 50 m*M* NaOH for 2–5 min at room temperature. Transfer the eluted primer to scintillation vials, add scintillation fluid, and measure the eluted ^{3}H in a liquid scintillation counter.

Comment

By using streptavidin-coated microtiter plates manufactured from scintillating polystyrene (ScintiStrips; Wallac OY, Turku, Finland) as solid support, the final washing and denaturation steps and the transfer of the eluted primer to scintillation vials can be omitted (14). This requires a scintillation counter for microtiter plates.

Step 6: Interpretation of Result

Calculate the ratio (*R* value) between the ^{3}H-labeled dNTP incorporated in the reaction for detecting the mutant nucleotide and the ^{3}H-labeled dNTP incorporated in the reaction for detecting the normal nucleotide. The *R* value will be >10 in samples from individuals homozygous for the mutant nucleotide, <0.1 in samples from individuals homozygous for the normal nucleotide, and in samples from heterozygous individuals it will usually be between 0.5 and 2.0, depending on the specific activities of the ^{3}H-labeled dNTPs used (Table II).

Comment

Calculation of the *R* value eliminates variations in the amount of incorporated ^{3}H-labeled dNTPs due to sample-to-sample variations in the efficiency of the PCR am-

TABLE II Detection of Polymorphic Nucleotides (A > G) in the *ADH3* and *METH* Genes by Solid-Phase Minisequencing

	ADH3				*METH*			
	[^{3}H]dNTP incorporated (cpm)[a]				[^{3}H]dNTP incorporated (cpm)[a]			
Sample	A allele	G allele	*R* value (A_{cpm}/G_{cpm})	Genotype	A allele	G allele	*R* value (A_{cpm}/G_{cpm})	Genotype
1	6550	61	107	AA	53	4570	0.011	GG
2	160	2800	0.057	GG	970	1530	0.63	AG
3	2610	1470	1.77	AG	1900	105	18.0	AA
H_2O	30	39	—	—	27	38	—	—

[a] At both sites, one [^{3}H]dATP will be incorporated in the A allele. At the *ADH3* site, one [^{3}H]dGTP will be incorporated, and at the *METH* site three [^{3}H]dGTPs will be incorporated in the G allele. In the analysis of the *ADH3* locus the specific activities of [^{3}H]dATP and [^{3}H]dGTP were 62 and 31 Ci/mmol, respectively. In the analysis of the *METH* locus the corresponding specific activities were 61 and 40 Ci/mmol. [Data from A.-C. Syvänen, A. Sajantila, and M. Lukka, *Am. J. Hum. Genet.* **52,** 46 (1993).]

plification. If the sequence contains one (or more) identical nucleotides immediately next to the nucleotide at the variable site, one (or more) additional [^{3}H]dNTPs will be incorporated in the minisequencing reaction, which obviously affects the *R* value. Table II shows as an example the results from analyzing polymorphic nucleotides in the alcohol dehydrogenase gene and the *METH* protooncogene, in which one or three ^{3}H-labeled dGTPs are incorporated, respectively. The small sequence-specific background misincorporation by the *Taq* DNA polymerase, which most probably is due to other dNTPs present as impurities in the ^{3}H-labeled dNTPs, has only a minor effect on the *R* value.

Discussion

In addition to unequivocally distinguishing between two sequences present either in homozygous (allele ratio 2:0) or heterozygous (allele ratio 1:1) samples, the *R* value obtained in the solid-phase minisequencing method directly reflects the ratio between two sequences when they are present in a sample as a mixture in any other ratio. Because the two sequences are essentially identical, they are amplified with equal efficiency during the PCR, which results in a linear relationship between the *R* value obtained in the minisequencing assay and the initial ratio between the two sequences (Fig. 3). Consequently, the method is a useful tool for quantitative PCR analysis. It is highly sensitive, allowing the detection of one sequence present as a small minority (<1%) in a sample (15, 16).

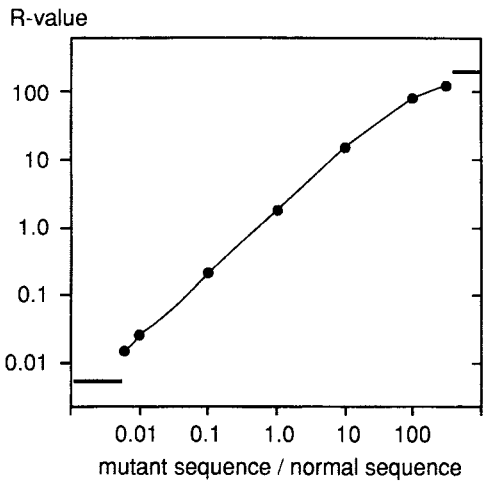


FIG. 3 Solid-phase minisequencing standard curve. Mixtures of known amounts of two 63-mer oligonucleotides differing from each other at a single nucleotide (nucleotide 3243 of the mitochondrial $tRNA_{Leu}$ gene; Ref. 16) were analyzed and the *R* values obtained in the minisequencing assay were plotted as a function of the initial ratio between the two sequences. The *R* values obtained by analyzing each oligonucleotide separately are indicated by the horizontal bars. The *R* values are the means of three parallel assays.

We have utilized the quantitative character of the solid-phase minisequencing method to analyze from pooled DNA samples two sequences present as a mixture in this DNA sample to determine population frequencies of disease-causing mutant alleles and polymorphisms (11, 16) as well as to determine the proportion of heteroplasmic mutations of the mitochondrial DNA (17–19). To determine the absolute amount of a nucleic acid sequence present in a sample, a known amount of internal standard differing from the sequence to be quantified by a single nucleotide is added to the sample before the PCR amplification (18, 20). Genomic DNA of known genotype (16), synthetic oligonucleotides (18), and RNA prepared by *in vitro* transcription (20) have been shown to be suitable for use as internal standards.

The initial ratio between the two amplified sequences can be calculated directly from the obtained *R* value, taking into account the specific activities and number of ^{3}H-labeled dNTPs incorporated in the minisequencing reactions. Alternatively, the initial ratio between the two sequences can be determined by comparing the obtained *R* value with a standard curve prepared in parallel with mixtures of known amounts of the corresponding sequences. The use of a standard curve will correct for possible misincorporation of an [^{3}H]dNTP by the DNA polymerase, which may affect the result when a sequence present as a small minority of the other sequence is to be quantified.

In conclusion, the high specificity of the single-nucleotide primer extension reaction catalyzed by a DNA polymerase is utilized in the minisequencing method, and this allows both unequivocal discrimination between sequence variants and sensitive quantitative PCR analysis. The combination of this assay principle with a solid-phase format yields a robust assay that is both suitable for large-scale use (21) and easy to set up for detecting single-nucleotide variations for different purposes both in the research laboratory and in the routine clinical laboratory.

References

1. V. A. McKusick, *in* "Mendelian Inheritance of Man," Vol. 1, 10th ed., p. xxxi. The Johns Hopkins University Press, Baltimore, Maryland, 1992.
2. K. B. Mullis and F. A. Faloona, *Methods Enzymol.* **155,** 335 (1987).
3. R. G. H. Cotton, *Mutat. Res.* **285,** 125 (1993).
4. A.-C. Syvänen and U. Landegren, *Hum. Mutat.* **3,** 172 (1994).
5. A.-C. Syvänen, K. Aalto-Setälä, L. Harju, K. Kontula, and H. Söderlund, *Genomics* **8,** 684 (1990).
6. A.-C. Syvänen, *Clin. Chim. Acta* **226,** 225, (1994).
7. M. A. Innis and D. H. Gelfand, *in* "PCR Protocols. A Guide to Methods and Applications" (M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, eds.), p. 3. Academic Press, San Diego, California, 1990.
8. R. Higuchi, *in* "PCR Technology: Principles and Applications for DNA Amplification" (H. A. Erlich, ed.), p. 35. Stockton Press, New York, 1989.
9. E. S. Kawasaki, *in* "PCR Protocols. A Guide to Methods and Applications" (M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, eds.), p. 21. Academic Press, San Diego, California, 1990.
10. C. Orrego, *in* "PCR Protocols: A Guide to Methods and Applications" (M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, eds.), p. 451. Academic Press, San Diego, California, 1990.
11. A.-C. Syvänen, A. Sajantila, and M. Lukka, *Am. J. Hum. Genet.* **52,** 46 (1993).
12. A.-C. Syvänen and H. Söderlund, *Methods Enzymol.* **218,** 474 (1993).
13. L. Harju, T. Weber, L. Alexsandrova, M. Lukin, M. Ranki, and A. Jalanko, *Clin. Chem.* **2,** 2282 (1993).
14. J. Ihalainen, H. Siitari, S. Laine, A.-C. Syvänen, and A. Palotie, *BioTechniques* **16,** 938 (1994).
15. A.-C. Syvänen, H. Söderlund, E. Laaksonen, M. Bengtström, M. Turunen, and A. Palotie, *Int. J. Cancer* **50,** 713 (1992).
16. A.-C. Syvänen, E. Ikonen, T. Manninen, M. Bengström, H. Söderlund, P. Aula, and L. Peltonen, *Genomics* **12,** 590 (1992).
17. A. Suomalainen, P. Kollmann, J.-N. Octave, H. Söderlund, and A.-C. Syvänen, *Eur. J. Hum. Genet.* **1,** 88 (1993).
18. A. Suomalainen, A. Majander, H. Pihko, L. Peltonen, and A.-C. Syvänen, *Hum. Mol. Genet.* **2,** 525 (1993).

19. V. Juvonen, K. Huoponen, A.-C. Syvänen, P. Aula, E. Nikoskelainen, and M. Savontaus, *Hum. Genet.* **93,** 16 (1994).
20. E. Ikonen, T. Manninen, L. Peltonen, and A.-C. Syvänen, *PCR Methods Appl.* **1,** 234 (1992).
21. M. Hietala, H. Grön, A.-C. Syvänen, L. Peltonen, and P. Aula, *Eur. J. Hum. Genet.* **1,** 296 (1993).

[3] Simple and Nonisotopic Methods to Detect Unknown Gene Mutations in Nucleic Acids

Susanna Bunge, Sigrid Fuchs, and Andreas Gal

Introduction

About 3000 different human diseases are thought to be caused by mutations of single genes (1). In addition, genetic factors seem to play an important role in a considerable number of other conditions such as diabetes, psychiatric disorders, and cancer. Over the last several years an increasing number of genes implicated in human inherited diseases have been discovered. In general, three different strategies have been used. Many of these genes have been isolated either by positional cloning or by functional cloning (for literature and reviews see Refs. 2 and 3). The third possibility of disease gene identification is the candidate gene approach. In this latter case, mutations of a particular gene, in many cases already cloned and extensively characterized, are predicted to be responsible for the pathology because a functionally intact gene product is essential for the physiological process in question. A typical example is rhodopsin, a protein involved in the phototransduction cascade in rod photoreceptors. It has turned out that mutations in the rhodopsin gene may lead to various forms of retinal dystrophies and dysfunctions (for a review see Ref. 4). Genes encoding other proteins involved in light perception, for example, rod cGMP-specific phosphodiesterase (PDE), arrestin, and transducin, are also potential disease genes for this group of disorders. Clearly, the combination of the above-mentioned methods may provide additional powerful tools with which to find disease genes (e.g., the approach of positional candidate; see Ref. 3).

Regardless of the method used to identify a gene, the demonstration of mutations that cosegregate with the disease phenotype is both an important step and an essential condition toward proving the involvement of the gene(product) in the pathological process. Once a reasonable candidate gene has been found, in many cases it is necessary to analyze a representative group of unrelated patients affected by the same disorder to identify predominant mutations and investigate genetic heterogeneity. The identification of disease-causing mutations in a large number of patients is the first step in examining genotype–phenotype correlation.

According to their size, gene mutations can be divided into two categories. Gross rearrangements consist of complete or partial deletions of the gene, large insertions, duplications, or inversions. Such mutations can be detected on Southern blot or by fluorescence *in situ* (FIS) hybridizations. "Short-length" mutations denote the most frequent type of gene alterations, that is, single base substitutions or rearrangements

Methods in Molecular Genetics, Volume 8

affecting only a few base pairs. For the analysis of this latter group of alterations, the polymerase chain reaction (PCR) is an important tool.

With increasing numbers of patients to be analyzed for mutations in an increasing number of genes, simple and reliable screening methods are necessary. During the last few years, many such methods have been developed (reviewed in Ref. 5). The method of chemical cleavage of mismatches (CCM) (6) is based on heteroduplex formation (see below) due to mispairing between the normal and mutant DNA strands after denaturation and reassociation. Mispaired nucleotides are modified chemically and the DNA strands are cleaved specifically at this position. While CCM as a screening method has several advantages over the other procedures discussed below, it is time consuming and requires both radiolabeled and, in part, toxic chemicals. Nevertheless, a modification of CCM circumvents many of these drawbacks (7). Another widely used method is denaturing gradient gel electrophoresis (DGGE), which relies on the physicochemical fact that the melting point of double-stranded DNA (dsDNA) depends on its base composition. Partial melting of dsDNA electrophoresed through a gradient of increasing concentration of denaturing agents results in a characteristic band pattern that may be altered by any nucleotide exchange present in the template DNA (8). Although DGGE is sensitive, it requires a special (and relatively expensive) apparatus and PCR primers designed especially for this purpose (9). The principle of the method of single-strand conformation polymorphism (SSCP) analysis is that the electrophoretic mobility of single-stranded nucleic acid fragments under nondenaturing conditions depends largely on their conformation. Owing to (partial) intramolecular base pairing, each single-stranded (ss) nucleic acid fragment assumes one (or more) conformation stable under a given electrophoretic condition (10). As the conformation is determined by the base sequence, exchange of any nucleotide may alter considerably the secondary structure of the ssDNA and result in a change of its electrophoretic mobility ("band shift"). Single-strand conformation polymorphism analysis has also been used successfully to screen RNA/cDNA for mutations (11). As SSCP analysis is straightforward and does not require complicated laboratory equipment, it has become popular and represents the most frequently used screening method. A series of papers deal with the issue of sensitivity of SSCP (12–14) and several modifications of the original protocol have already been suggested (15–19). For heteroduplex analysis, the mixture of wild-type and mutant DNA fragments is denatured and "renatured" under conditions optimal for nucleic acid association. In addition to homoduplexes (wild-type:wild-type and mutant:mutant), parts of the molecules form heteroduplexes (HDs, wild-type:mutant). Heteroduplexes contain one or more mismatches and show an altered electrophoretic behavior (for a summary of the literature see Ref. 20).

We are using a simple and nonisotopic method that can be carried out in any laboratory without special equipment and combines the SSCP and heteroduplex analyses. To date we have successfully used this method for mutation screening in 15 different genes responsible for inherited diseases.

Polymerase Chain Reaction Amplification

General Protocol

Since the invention of highly selective DNA amplification by a thermostable DNA polymerase and repeated cycles of denaturation, primer annealing, and chain elongation, the polymerase chain reaction (PCR) has become one of the most widely used techniques in molecular genetics. The PCR provides fast and highly effective synthesis of any DNA (or cDNA) fragment specified by a pair of primers from a complex mixture of nucleic acids without time-consuming cloning procedures.

A typical PCR is carried out in a volume of 10–100 μl with final concentrations of 10 m*M* Tris-HCl (pH 8.3), 1.5 m*M* $MgCl_2$, 50 m*M* KCl, 0.01% (w/v) gelatin, 0.2 m*M* dNTPs, a 0.4 μM concentration of each primer, 100–200 ng of template (genomic DNA or first-strand cDNA), and 1 unit of *Taq* polymerase (for a 50-μl reaction). In the case of performing a great number of PCRs at the same time, it is convenient to prepare a mixture containing all components but template DNA. The inclusion of *Taq* polymerase in the reaction mixture is not generally recommended. It has been suggested that unspecific priming and amplification may occur below the annealing temperature. Therefore, in many cases, a "hot start" of the PCR is suggested, that is, the enzyme is added only after/during the initial denaturation step. A standard PCR protocol starts with an initial 5 to 8 min denaturation at 94°C. If a hot start is preferred, *Taq* polymerase should be added at this time. In general, 35 cycles are performed, each cycle consisting of a denaturation step for 1 min at 94°C, annealing for 1 min at a temperature selected according to the length and base composition of the primers used, and an extension for 2 min at 72°C. The latter step can be prolonged by a few seconds in every cycle to compensate for the gradual loss of enzyme activity. After PCR has been completed, a final extension of 10 min at 72°C can be performed to obtain as many "complete" products as possible. The length of each step can be considerably shortened by using PCR machines of the latest generation equipped with cover heating, temperature control in the reference tube, and blocks for thin-walled tubes or microtiter plates.

Optimizing Polymerase Chain Reaction

Although standard conditions may work well in many cases, adjustment of PCR conditions is often necessary to obtain optimal results. The most common problem is coamplification of unspecific products and/or low yield of specific product. In the case of poor-quality DNA (contamination by proteins), supplementation with 2.5 m*M* spermidine may improve amplification efficiency. As *Taq* polymerase requires magnesium, the concentration of $MgCl_2$ may also be critical (EDTA, phosphate, or free nucleotides may chelate magnesium ions). Therefore, in some cases,

higher $MgCl_2$ concentrations (2.25 m*M* instead of 1.5 m*M*) may also be helpful. The amount of oligonucleotide primers and free nucleotides is also crucial. The two primers should be present in equimolar amounts. "Difficult" PCRs sometimes work much better at elevated concentrations of nucleotides. Denaturing agents such as formamide, dimethyl sulfoxide (DMSO), or various detergents [e.g. Nonidet P-40 (NP-40), Tween 20] may be used to reduce unspecific primer annealing. In our hands, the addition of 5–10% (v/v) DMSO was the most effective. A widely used additive is "Perfect Match" (Stratagene, La Jolla, CA), probably a single-strand binding protein.

Screening of Genomic DNA and/or cDNA

Instead of genomic DNA, mRNA/cDNA can also be screened for mutations. Total RNA (5 μg) is reverse-transcribed in 100 μl of PCR buffer supplemented with 400 μM dNTPs, 5 μg random hexamers, 80 U of human placental ribonuclease inhibitor, and 200 U of reverse transcriptase at 37°C for 60 min. About one-tenth of the reverse transcription (RT) reaction mixture is used for a 100-μl PCR. Instead of random hexamers, oligo(dT) or target-specific oligonucleotide primers can also be used in the RT reaction. However, aliquots of first-strand cDNAs prepared by unspecific primers can be further amplified by PCRs of different specificity without eliminating the primers. For transcripts present only in low levels under physiological conditions in blood cells or fibroblasts, the two tissues frequently used and readily available for molecular diagnostics, a second "nested" PCR may be necessary to amplify effectively messages of low abundance synthesized by "illegitimate" transcription.

An obvious advantage of the analysis of cDNA is that (the coding sequence of) large genes can be examined by a number of overlapping fragments. Nevertheless, caution is necessary if cDNA is analyzed exclusively. In cases of autosomal recessive traits, different gene alterations may be present on the two alleles (compound heterozygosity). The relative proportion of transcripts originating from the two alleles may vary considerably owing to the different (relative) stability of the corresponding messages. mRNA molecules containing a premature stop codon have been shown to have reduced stability. Figure 1 gives an example of this phenomenon. cDNA analysis was performed on individuals heterozygous for the common *amber* mutation W402X of the gene encoding α-L-iduronidase (IDUA), an enzyme involved in the lysosomal degradation of mucopolysaccharides. Owing to the dramatically reduced stability of the message with the stop mutation, the abundance of the W402X allele is low and it can hardly be detected by PCR if wild-type sequence or a mutation not interfering with mRNA stability is present on the other allele. Certainly, the situation is similar in the case of carrier diagnostics for X-linked conditions.

In case of splice site mutations, analysis of both the mRNA and genomic DNA is necessary. Mutations affecting the intronic part of the splice site consensus sequence

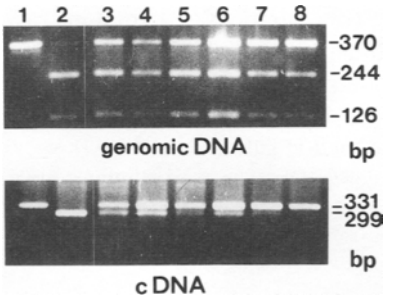


FIG. 1 Detection of the W402X mutation of α-L-iduronidase (IDUA) in cDNA and genomic DNA by restriction analysis. Exon IX of the gene was PCR amplified from genomic DNA (*top*) or cDNA (*bottom*) of the same person, restricted by *Mae*I, and separated on an agarose gel. The nonsense mutation W402X (TGG→TAG) creates a new *Mae*I restriction site leading to the cleavage of the 370-bp genomic fragment into two fragments of 244 and 126 bp. The 331-bp cDNA fragment is cleaved into fragments of 299 and 32 bp (the latter is not visible on the gel). Lane 1, normal control; lane 2, patient homozygous for W402X; lanes 3–6 and 8, patients compound heterozygous for W402X and a second mutation on the other allele (lane 3, 225insC; lane 4, Q70X; lane 5, L218P; lane 6, R621X; and lane 8, R489P); lane 7: heterozygote for W402X/wild type.

can be identified only by examining genomic DNA; the effect a given alteration will have on the splicing (exon skipping, use of cryptic splice sites, etc.) can be elucidated only by the analysis of the transcript. An unusual mutation was identified in the gene for the X-chromosomal form of mucopolysaccharidosis (type II, Hunter syndrome) encoding the enzyme iduronate-2-sulfatase (IDS). One patient showed, in addition to the normal-sized message, a predominant transcript with a 78-bp insertion between exons VII and VIII. Further analysis revealed that a point mutation in intron 7 created a new donor splice site, which together with a sequence resembling a splice acceptor site present 78 bp upstream in the intron resulted in an alternatively spliced IDS mRNA (M. Rathmann, A. Gal, and S. Bunge, unpublished observation, 1995). As the intronic sequence containing the point mutation is outside the gene region routinely amplified/screened, this particular gene alteration would escape detection performed by exon screening only. For many genes, sequences of high homology exist in the human genome that may represent pseudogenes. Although pseudo-genes are not transcribed, their presence complicates the analysis of the functional gene at the genomic DNA level, owing to coamplification. Therefore, on the one hand, the analysis of cDNA should be preferred if pseudogenes exist. On the other hand, if a sufficient amount of sequence information is available on both the functional gene and pseudogene, allele-specific primers may be designed that discriminate between the two.

Simple Nonisotopic Single-Strand Conformation Polymorphism Method Combined with Heteroduplex Analysis

While many laboratories perform SSCP with radiolabeled PCR products, we have had success using nonisotopic SSCP analysis that, in addition, allows the simultaneous detection of heteroduplex formation.

Experimental Procedure

Prior to SSCP, PCR products should be viewed on an agarose gel to ensure that there is a sufficient amount of specific product. It is important that an equal amount of DNA be loaded in each lane. Fragments larger than 450 bp should be digested with an appropriate restriction enzyme prior to SSCP analysis, as the method is more sensitive for smaller fragments.

Five to 10 μl of PCR product is mixed with 10 μl of formamide solution [95% (v/v) formamide, 20 m*M* EDTA, 0.05% (v/v) bromphenol blue, 0.05% (v/v) xylene cyanole FF]. Denaturation is performed for 10 min at 95°C. Thereafter samples are cooled quickly on ice, and loaded immediately on the gel.

Although optimal gel conditions must be tested empirically for each DNA fragment, the following protocol may be used initially: 6 or 8% (w/v) acrylamide (2.6% cross-linking), 1× TBE as running buffer [90 m*M* Tris–borate (pH 8.3), 2.5 m*M* EDTA], and 10% (v/v) glycerol. Depending on the size of the PCR products to be analyzed, the gel is run overnight at 20–25 W at room temperature. If these conditions do not yield a clear pattern of distinct and sharp bands, a gel without glycerol and run at 30 W for 5–7 hr at room temperature can be tried. Alternatively, gels with or without glycerol can be run at 4°C. The use of 10% (w/v) sucrose instead of glycerol may improve separation.

We use a sequencing gel apparatus with a gel size of about 35 × 38 × 0.08 cm and a 32-well comb. A considerable portion of the PCR product reassociates, probably during the first few minutes of electrophoresis, and the gel size given above allows us to detect both double-stranded and single-stranded DNA fragments on the same gel at the same time. In the upper part of the gel, there are mobility shifts in the SSCP analysis, while in the bottom part heteroduplex formation can be visualized. Figure 2 shows a typical example of a mutation in exon III of the rhodopsin gene that causes both a mobility shift of single strands and heteroduplex formation. While the aberrant SSCP pattern is obvious for every sample carrying the mutation, heteroduplexes are resolved less clearly in the slightly overloaded lanes.

Many parameters influence the single-stranded pattern produced by any DNA fragment, which may also vary slightly from experiment to experiment. Similarly, the mobility shift resulting from a given mutation may look different under various

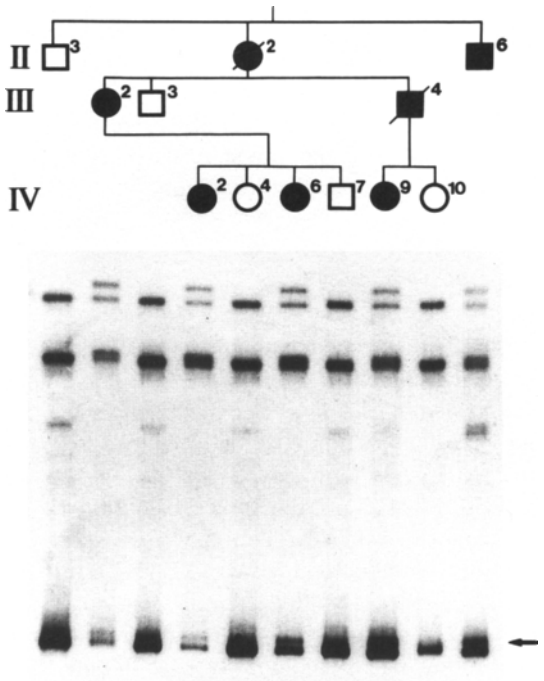


FIG. 2 Combined SSCP and heteroduplex analyses of a family with autosomal dominant retinitis pigmentosa. Exon III of the rhodopsin gene was PCR amplified and examined as described in text. All affected individuals show a mobility shift of single-stranded (*top*) bands and a heteroduplex (arrow) of the double-stranded (*bottom*) band. Unaffected individuals show patterns identical to those of controls. The mutation was shown to be a T-to-C transition (TGC→CGC) predicting the replacement of Cys-222 by arginine (Cys222Arg) in the protein. Cosegregation of the mutation with the disease phenotype and its absence in 100 unaffected control persons strongly suggest that this mutation is causative for the disease.

gel conditions. Therefore, the combination of the different parameters may considerably increase detection efficiency. Figure 3 shows a mutation in the gene encoding the β subunit of phosphodiesterase (PDE). The band shift is hard to detect under standard SSCP conditions [2.6% cross-linking, 10% (v/v) glycerol] whereas it is clearly visible when 1.3% cross-linking and 10% (w/v) sucrose are used. The parameters most frequently altered in the SSCP analysis are ion strength of the gel buffer, proportion of bisacrylamide and acrylamide (percent cross-linking), temperature of the gel, presence or absence of glycerol in the gel, and amount of DNA loaded.

We estimate that heteroduplexes can be demonstrated in about two-thirds of the samples presenting single-stranded mobility shifts due to a heterozygous mutation,

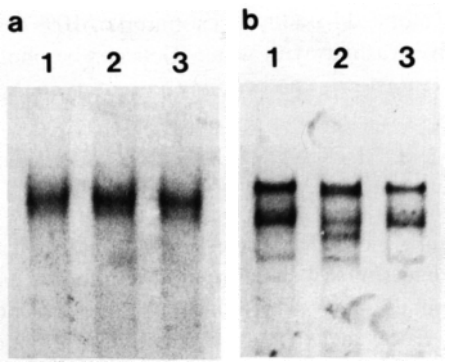


FIG. 3 Effect of different electrophoretic conditions on the mobility shift produced by the C→G transversion in codon 854 of the human *βPDE* gene in SSCP analysis. Lanes 1 and 3, unaffected control; lane 2, patient heterozygous for the C→G transversion (Leu854Val). (a) 2.6% cross-linking with 10% (v/v) glycerol; (b) 1.3% cross-linking with 10% (w/v) sucrose. Reproduced with permission of Plenum Publishing.

especially if small deletions or insertions are present. In the case of a mutation in homo- or hemizygous form, the PCR product can be mixed with wild-type DNA to allow heteroduplex formation. Although heteroduplex analysis is certainly less sensitive than SSCP, heteroduplexes may help to pick up mutations, especially when the single-stranded shift is not prominent (under the condition used).

At the end of electrophoresis, the position of double strands is estimated in relation to the xylene cyanole dye; single strands run much slower, especially when glycerol is added to the gel. The appropriate part of the gel is transferred with a plastic foil in a glass dish containing 10% (v/v) ethanol as fixative. If the gel cannot be stained immediately by silver, it can be kept in 40% (v/v) methanol and transferred to ethanol just before staining. Although commercial silver staining kits are available, staining with self-made solutions is less expensive and gives good results. The silver-staining protocol used is essentially the same as that described by Budowle *et al.* (21). All steps take place with gentle agitation of the gel. Ethanol fixation is performed for 5 min followed by oxidization with 1% (v/v) nitric acid for 3 min. It is important to remove the ethanol completely and to ensure that the gel is covered with the oxidizer. The acid solution is removed and the gel is rinsed twice with distilled water. The water used for all staining solutions should be deionized to minimize background staining. Silver nitrate (12 m*M*) is applied for 20 min followed by two rinsing steps. Development is performed with a solution of 0.28 *M* sodium carbonate and 0.019% (v/v) formaldehyde. It is essential that the gel be permanently agitated and the developer solution be changed after about 30 sec to remove the precipitate that is formed. As soon as the staining intensity is sufficient, the developer can be removed and the

reaction is stopped by adding a solution of 10% (v/v) acetic acid for 5 min followed by extensive washing with water. Gels can be photodocumented on a transilluminator. Sealed moist gels can be kept (preferably in the dark) for years.

Identification of Mutations by DNA Sequencing

The screening method described above provides a rapid and sensitive technique with which to detect most "short-length" mutations, including single base pair substitutions and small deletions. However, for the identification of a yet unknown sequence alteration, DNA sequencing is essential. The major steps of the dideoxy-mediated chain termination sequencing method (22) are the following: annealing of specific oligonucleotides that serve as a starter for extension by the DNA polymerase, labeling reaction (either radioactive or fluorescent), and base-specific chain termination by incorporation of dideoxynucleotides (ddNTPs). Finally, sequencing products are separated on a denaturing polyacrylamide gel. Different commercial sequencing kits are also available. For special applications, for example, if the amount of template is limited, cycle-sequencing using *Taq* DNA polymerase (23) may be favorable. However, as the relative intensity of the bands in the four nucleotide lanes may vary greatly, cycle sequencing is not optimal for detection of heterozygotes. Automated sequencing using fluorescent dyes combined with computer-mediated sequence analysis is becoming more common.

Direct Sequencing of Polymerase Chain Reaction

The sequence of DNA fragments generated by PCR can be determined either after subcloning in appropriate sequencing vectors (M13 or Bluescript) or by direct sequencing of double-stranded or single-stranded PCR products.

Sequencing after the initial subcloning is time consuming. In addition, several clones must be sequenced to exclude sequence artifacts introduced by DNA polymerase used for PCR amplification. A major advantage of direct sequencing is that both alleles can be analyzed simultaneously. However, occasionally this can lead to difficulties in the interpretation owing to overlaps of two sequences if, for example, a deletion is present on one allele. Similarly, if templates are generated from cDNA, and one of the alleles is less abundant in the sample, subcloning of PCR products would be the easier way to sequence the rare allele.

Preparation of Templates

Different protocols are in use for the preparation of adequate templates for direct sequencing of PCR products by the chain termination method. In all cases, the pri-

mary aim is to obtain single strands suitable for the subsequent primer annealing. A purification of templates is highly recommended in each of the methods described below, to remove primers, free nucleotides, and unspecific PCR products that may disturb the reaction.

Alkaline Denaturation of Double-Stranded DNA: Polymerase Chain Reaction Product

Denaturation of templates with alkali is commonly used if plasmid DNA is sequenced. Treatment of purified dsPCR products with alkali may also generate single-stranded templates. The DNA is denatured by adding 0.25 vol of 2 *M* NaOH for 10 min at room temperature. The reaction is neutralized with 0.3 vol of 3 *M* sodium acetate (pH 4.8) and 0.7 vol of distilled water followed by precipitation of the DNA with 2–3 vol of 96% (v/v) ethanol (−70°C, 15 min). The pelleted DNA is washed with 70% (v/v) ethanol, redissolved in 7–10 μl of distilled water (the volume depends on the sequencing kit/system used), and the template is ready for primer annealing. Although alkaline denaturation of PCR products may be appropriate, it is not widely used in direct sequencing of PCR products because the two strands may rapidly reassociate during the sequencing reaction, inhibiting the primer from annealing. This problem may be overcome by generating single-stranded PCR products.

Generation of Single-Stranded DNA by Use of Asymmetric Polymerase Chain Reaction

This procedure represents a modified PCR reaction in which one of the primers is diluted (1:50), leading to an excess synthesis of ssDNA corresponding to the strand primed by the undiluted primer (24). To increase the purity of the dsPCR products used as template in the asymmetric PCR, they should be separated on a 12% (w/v) preparative agarose gel (Seakem agarose; FMC, Philadelphia, PA). The piece of agarose containing the band of interest is cut out and overlaid with 200–300 μl of Tris–EDTA (TE), pH 8.0. After incubation at room temperature for 4–5 hr or at 4°C overnight, 5–10 μl of the supernatant is used for a second round of PCR, which is carried out as an asymmetric PCR in a 100-μl reaction volume. To sequence both the sense and antisense strands, usually two asymmetric PCRs are performed in which either the "sense" or the "antisense" primer is diluted.

Asymmetric PCR is carried out under the same conditions as the PCR to synthesize double-stranded products, except that the primer used subsequently as sequencing primer is diluted to 1:50 (final concentration, 8 n*M*) whereas the second primer has the standard concentration (0.4 μ*M*). In most cases, it is useful to increase the annealing temperature by 2°C (to achieve higher specificity) and the number of cycles to 40 (to obtain a better yield). An aliquot of the asymmetric PCR product is analyzed on an agarose gel stained with ethidium bromide. As ssPCR products are

poorly stained with ethidium bromide, it is difficult to calculate the amount obtained in the asymmetric PCR.

Generation of Single-Stranded DNA by Use of Magnetic Beads

The separation of dsDNA strands by using the avidin–biotin system and magnetic beads (Dynabeads M-280 streptavidin; Dynall, Oslo, Norway), is a quick, elegant method to obtain pure single-stranded PCR products (25, 26). In principle, biotinylated double-stranded PCR products are captured on avidin-coated magnetic beads and the nonbiotinylated strand is melted away using NaOH. The sequencing reaction is carried out on the immobilized single-stranded template. The nonbiotinylated strand can also be recovered and sequenced following neutralization with HCl. By using this technique, one of the PCR primers must be labeled by biotin at its 5′ end during synthesis. Alternatively, oligonucleotide PCR primers coupled with an amino modifier at the final step of synthesis can be labeled subsequently by commercially available biotin–NHS ester.

Twenty-microliter Dynabeads are used for a 50-μl PCR. Dynabeads are washed twice with phosphate-buffered saline (PBS), pH 7.0, containing 0.1% bovine serum albumin (BSA) to remove NaN_3 and once with 1 × B&W buffer [5 m*M* Tris (pH 7.5), 0.5 m*M* EDTA, and 1 *M* NaCl]. All reaction steps including washing are performed using a Dynal MPC-E (magnetic particle concentrator for microcentrifuge tubes) to immobilize Dynabeads so that the supernatant can be removed easily.

Prewashed Dynabeads are resuspended in 40 μl of 2 × B&W buffer (per reaction). Forty microliters of biotinylated and purified PCR product is added and the mixture is incubated at room temperature for 15 min with occasional shaking to achieve adsorption of dsDNA to Dynabeads. The supernatant is removed using a Dynal MPC-E, and after an additional wash with 1 × B&W buffer Dynabeads are resuspended in 8 μl of freshly prepared 0.1 *M* NaOH and incubated for 10 min at room temperature to melt away the nonbiotinylated complementary strand. The supernatant, which now contains the nonbiotinylated DNA strand, is removed and transferred to a new reaction tube. After neutralization with 4 μl of 0.2 *M* HCl and 1 μl of 1 *M* Tris-HCl, pH 7.5, this complementary strand can also serve as single-stranded template for sequencing.

Dynabeads, coated with the biotinylated ssDNA, are washed with 50 μl of 0.1 *M* NaOH, 40 μl of 1 × B&W buffer, and 50 μl of TE. Supernatants are removed and discarded and Dynabeads are resuspended in 20 μl of distilled H_2O. An aliquot of 7–8 μl of the suspension is used directly in the sequencing reaction.

Purification of Polymerase Chain Reaction Products

Ultrafiltration through a Centricon-100 centrifugal concentrator membrane (Amicon, Danvers, MA) equilibrated with 1 × TE, pH 8.0, can be used to purify PCR products (single-stranded or double-stranded). After chloroform extraction (to remove paraffin oil), the volume is adjusted to 1 ml with 1 × TE and applied to the concentrator.

Centrifugation is carried out at room temperature in a fixed angle-head rotor for 5 min at 1000 *g*. Subsequently, three washing steps are performed, each with 1 ml of TE, pH 8.0, and a 5-min centrifugation. Only the last centrifugation step is prolonged to 20 min to achieve a suitable concentration of the purified PCR product (usually about 40 μl).

Sequencing Reaction

Sequencing reactions are carried out using commercially available kits and according to manufacturer instructions. The use of the 7-deaza analog of dGTP (c^7dGTP) may overcome problems of "band compression" due to GC-rich sequences.

Primer Annealing

Seven microliters of single-stranded template is mixed with 2 μl of reaction buffer (included in the kit) and 1 μl of sequencing primer (2 μM) (we use either the complementary PCR primer or an internal sequencing primer). The mixture is heated for 2 min at 65°C and cooled down slowly to room temperature. The reaction mixture is spinned down briefly and the annealing reaction product is chilled on ice until use.

During primer annealing, four tubes (1.5 ml) are prepared for each template with 2.5 μl of the termination mixture (each containing one of the dideoxynucleotides ddATP, ddCTP, ddGTP, and ddTTP) and prewarmed at 37°C. The labeling mix is diluted fivefold to a working concentration.

Labeling Reaction

To the 10 μl of ice-cold annealed DNA mixture, 1 μl of dithiothreitol (DTT, 0.1 *M*), 2 μl of diluted labeling mix, 0.5 μl of [^{35}S]dATP, and 2 μl of diluted Sequenase [diluted to 8:1 (v/v) in ice-cold enzyme dilution buffer] are added. The reaction is mixed and incubated at room temperature for 2–5 min.

Termination Reaction

A 3.5-μl portion of the labeling reaction is transferred to each termination tube (A, C, G, and T), mixed, and incubated at 37°C for 5 min. The reaction is stopped by adding 4 μl of stop solution [95% (v/v) formamide, 20 m*M* EDTA, 0.05% (v/v) bromphenol blue, 0.05% (v/v) xylene cyanol FF]. Immediately before loading them onto a sequencing gel, samples are denatured at 94°C for 2 min and chilled on ice.

Sequencing Gel

In general, 3 μl of each probe is loaded on a sequencing gel consisting of 6% (w/v) polyacrylamide and 7 *M* urea, 1 × TBE. Electrophoresis buffer is 1 × TBE. Running time varies between 1.5 and 3.5 hr with 50 W, and depends on the separation desired. After electrophoresis the gel is fixed in 5% (v/v) methanol, 10% (v/v) acetic acid, and dried before exposure for autoradiography.

After a previously unknown mutation has been identified by the methods discussed above, it is important to determine whether the sequence alteration cosegregates with the disease phenotype, that is, if it has pathogenic relevance, or, alternatively, if it represents a polymorphism also present in unaffected controls. In addition, the typing of relatives of the index case is frequently requested to permit a reliable genetic counseling. Restriction analysis is a simple and highly specific way of detecting defined mutations if the sequence variation in question alters the recognition site of a restriction enzyme (Fig. 1). Hybridization with allele-specific oligonucleotides (ASOs) is another commonly used method to detect known mutations in a large collective of individuals (27). A further assay is "restriction map modification by site-directed mutagenesis" (28), in which specially designed primers introduce mismatches during PCR and thereby create a new restriction site in the DNA fragment to be analyzed either in combination with the mutation or together with the wild-type sequence.

Acknowledgment

The experimental work described in this article was financially supported by the Deutsche Forschungsgemeinschaft.

References

1. V. A. McKusick, "Mendelian Inheritance in Man. Catalogs of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes," 10th ed. Johns Hopkins University Press, Baltimore, Maryland, 1992.
2. F. S. Collins, *Nature Genet.* **1,** 3 (1992).
3. A. Ballabio, *Nature Genet.* **3,** 277 (1993).
4. G. J. Farrar, S. A. Jordan, R. Kumar-Singh, C. F. Inglehearn, A. Gal, C. Greggory, M. AlMaghtheh, P. Kenna, M. M. Humphries, E. M. Sharp, D. M. Sheils, S. Bunge, P. A. Hargrave, M. J. Denton, E. Schwinger, S. S. Bhattacharya, and P. Humphries, *in* "Retinal Degeneration" (J. G. Hollyfield, R. E. Anderson, and M. M. LaVail, eds.), p. 63. Plenum Press, New York, 1993.
5. M. Grompe, *Nature Genet.* **5,** 111 (1993).
6. R. G. Cotton, N. R. Rodrigueus, and R. D. Campbell, *Proc. Natl. Acad. Sci. U.S.A.* **85,** 4397 (1988).
7. R. Youil, B. W. Kemper, and R. G. H. Cotton, *Proc. Natl. Acad. Sci. U.S.A.* **92,** 87 (1995).
8. R. M. Myers, T. Maniatis, and L. S. Lerman, *Methods Enzymol.* **155,** 501 (1987).
9. V. C. Sheffield, D. R. Cox, L. S. Lerman, and R. M. Myers, *Proc. Natl. Acad. Sci. U.S.A.* **86,** 232 (1989).
10. M. Orita, Y. Suzuki, T. Sekiya, and K. Hayashi, *Genomics* **5,** 874 (1989).
11. G. Sarkar, H.-S. Yoon, and S. Sommer, *Nucleic Acids Res.* **20,** 871 (1992).
12. J. Michaud, L. C. Brody, G. Steel, G. Fontaine, L. S. Martin, D. Valle, and G. Mitchell, *Genomics* **13,** 389 (1992).

13. K. Hayashi and D. W. Yandell, *Hum. Mutat.* **2,** 338 (1993).
14. M. Ravnik-Glavac, D. Glavac, and M. Dean, *Hum. Mol. Genet.* **3,** 801 (1994).
15. L. Spinardi, R. Mazars, and C. Theillet, *Nucleic Acids Res.* **19,** 4009 (1991).
16. P. J. Ainsworth, L. C. Surh, and M. B. Coulter-Mackie, *Nucleic Acids Res.* **19,** 405 (1991).
17. D. Glavac and M. Dean, *Hum. Mutat.* **2,** 404 (1993).
18. J. Weidner, A. Eigel, J. Horst, and W. Köhnlein, *Hum. Mutat.* **4,** 55 (1994).
19. A. Winterpacht, K. Hilbert, U. Schwarze, and B. Zabel, *Hum. Genet.* **95,** 437 (1995).
20. M. B. White, M. Carvalho, D. Derse, S. J. O'Brien, and M. Dean, *Genomics* **12,** 301 (1992).
21. B. Budowle, R. Chakraborty, A. Ginsti, A. Eisenberg, and R. Allen, *Am. J. Hum. Genet.* **48,** 137 (1991).
22. F. Sanger, S. Nicklen, and A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74,** 5463 (1977).
23. M. A. Innis, K. B. Myambo, D. H. Gelfand, and M. A. D. Brow, *Proc. Natl. Acad. Sci. U.S.A.* **85,** 9436 (1988).
24. U. B. Gyllensten and H. A. Erlich, *Proc. Natl. Acad. Sci. U.S.A.* **85,** 7652 (1988).
25. T. Hultman, S. Stahl, E. Hornes, and M. Uhlen, *Nucleic Acids Res.* **17,** 4937 (1989).
26. S. L. Thein and J. Hinton, *Br. J. Haematol.* **79,** 113 (1991).
27. R. K. Saiki, T. L. Bugawan, G. T. Horn, K. B. Mullis, and H. A. Erlich, *Nature (London)* **324,** 163 (1986).
28. A. Haliassos, J. C. Chomel, L. Tesson, M. Baudis, J. Kruh, J. C. Kaplan, and A. Kitzis, *Nucleic Acids Res.* **17,** 3606 (1989).

[4] Identification and Functional Analysis of Mutations in the Human *PAX6* Gene

Richard Maas, Jonathan A. Epstein, and Tom Glaser

Introduction

The mutational analysis of genes important in human disease has received great attention. Such studies have resulted in the identification of mutations in the genes responsible for cystic fibrosis, Duchenne's muscular dystrophy, Huntington's disease, and fragile X syndrome, among others (reviewed in Ref. 1). One class of human disorders only recently subjected to mutation analysis at the molecular level includes those responsible for congenital malformation syndromes. In several cases, these syndromes are due to mutations in transcription factors (2). In this chapter, we review the methodologies we have employed to detect and to analyze the functional consequences of mutations in one such transcription factor, that encoded by the *PAX6* developmental control gene. Mutations in *PAX6* are responsible for a dominantly inherited defect of human ocular and central nervous system (CNS) development called aniridia (reviewed in Refs. 3 and 4).

PAX6 belongs to a family of evolutionarily conserved *Pax* genes, unified by the presence of a paired box, which encodes a 128-amino acid bipartite DNA-binding domain (5). Because loss of function mutations in *Pax* genes give rise to semidominant phenotypes, *PAX* mutations have been identified in a number of human and mouse developmental syndromes (reviewed in Refs. 6 and 7), including in humans, aniridia (*PAX6*), Waardenburg's syndrome (*PAX3*), and renal agenesis with coloboma (*PAX2*). Thus, the identification of *PAX* gene mutations has direct clinical relevance. In addition, human *PAX* gene mutations provide valuable insight into the molecular function of *PAX* gene products. Here, we review a PCR (polymerase chain reaction)-based single-strand conformational polymorphism (SSCP) assay for detecting mutations in the human *PAX6* gene, and methods pertinent to the functional analysis of select *PAX6* mutations.

Detection of *PAX6* Mutations

Concise reviews of mutation detection methods have been published (8, 9). A central starting point for all these methods is some knowledge of either the genomic or cDNA structure of the gene being analyzed. In cases where such knowledge is limited, the corresponding analysis will necessarily be incomplete. Single-stranded conformational polymorphism analysis is based on the principle that single-stranded

Methods in Molecular Genetics, Volume 8

DNA normally adopts a highly ordered conformational state that is remarkably sensitive to DNA sequence composition (10). A single nucleotide substitution can disturb the secondary structure of single-stranded DNA, resulting in an altered mobility on a nondenaturing polyacrylamide gel. If the DNA fragment is radiolabeled, the novel conformer(s) can be readily detected when compared to a wild-type or unsubstituted DNA fragment run on the same gel. Ideally, the SSCP approach requires thorough knowledge of the intron–exon structure of the gene, including the DNA sequence flanking the exons, so that splice acceptors and donors, frequent sites of mutation, can be included in the PCR products being analyzed.

The sensitivity of SSCP analyses for detecting individual point mutations is obviously of great interest. Estimates of the detection sensitivity for SSCP vary, ranging from ~80 to 95% (11, 12). We have not systematically examined this, but mutations have been detected in 80% of aniridia patients whose genomic DNAs have been analyzed by SSCP (3). Because it is possible that some aniridia patients contain mutations in parts of the *PAX6* gene, such as regulatory elements, that were not analyzed, it is likely that the real detection sensitivity is even higher. A number of factors influence the sensitivity of SSCP for detecting point mutations, including fragment size and sequence composition, the nucleotide sequence surrounding the mutation, and electrophoresis and gel matrix conditions (12). Single-stranded conformational polymorphism analysis appears to be most sensitive when the fragments being analyzed are in the 100- to 250-nucleotide size range. Longer exons must be analyzed as separate overlapping fragments. A possible compensating benefit is that SSCP assay results may be confirmed and further localized if present on overlapping fragments. In the case of the vertebrate *PAX6* gene, of the 16 known exons (13–15), all except for 2 [that encoding the terminal portion of the coding region and the 3′ untranslated region (UTR), designated exon 13, and an alternative exon designated the α exon (14)] are encoded by small 42- to 216-bp exons, ideally suited for SSCP analysis. Thus, each primer set is specifically designed to amplify one exon. The sequences of these oligonucleotide primers have been previously published (13), and are therefore not reproduced here. A diagram illustrating the various steps in the SSCP assay we have employed for identifying *PAX6* mutations is provided in Fig. 1. Because exon 13 of *PAX6*, encoding both the 3′ UTR and the terminal portion of the *PAX6* coding region, is ~820 bp, we have subdivided it into six smaller segments, each amenable to PCR amplification and SSCP analysis. The locations of these additional primers and their sequences are provided in Figure 2.

An important point to emphasize is that although aniridia is a dominant disorder, like other *PAX* gene mutations, it results from loss of function mutations, that is, from haploinsufficiency. Thus, in aniridia cases where one *PAX6* allele is deleted, such as the WAGR 11p13 contiguous gene deletion syndrome (3), the SSCP assay will result only in amplification of the wild-type allele, confounding the analysis. Even a deletion involving only DNA sequences corresponding to a portion of one PCR primer will escape detection by SSCP, as the mutant allele will not be amplified.

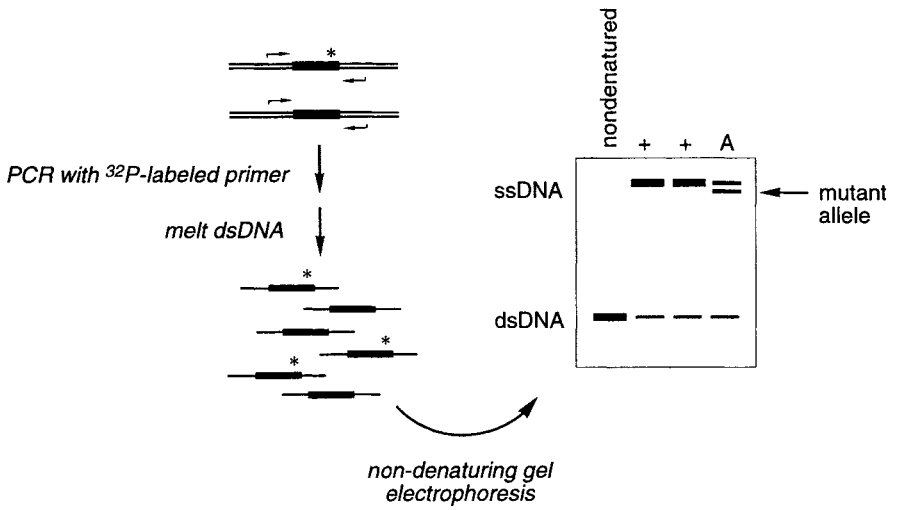


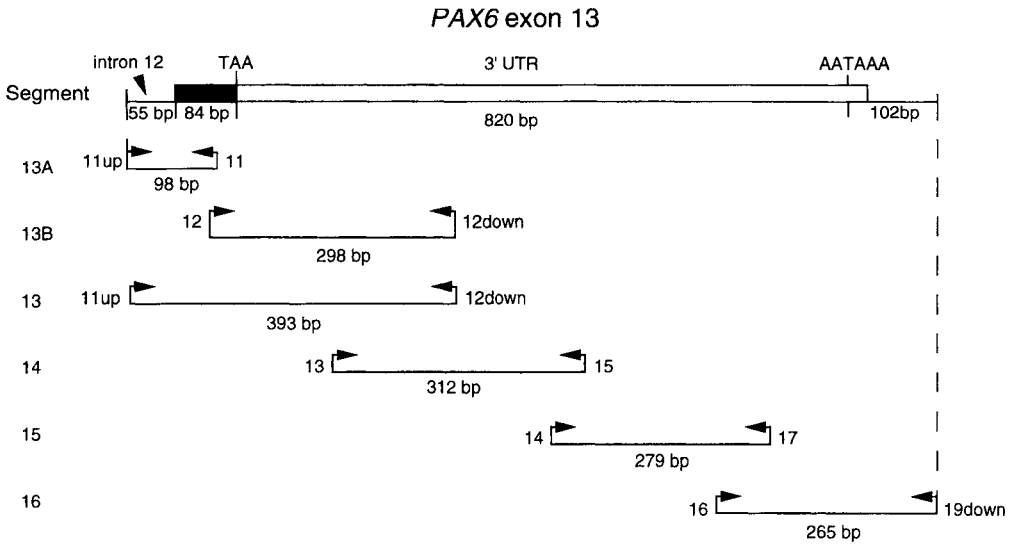
FIG. 1 Scheme illustrating the various steps in the SSCP assay for mutations in the human *PAX6* gene. Following PCR amplification with a labeled primer, the products are analyzed on a nondenaturing polyacrylamide gel. Mutant alleles can be detected as single-stranded bands migrating with novel mobilities.

In such cases, reverse transcription-PCR (RT-PCR) approaches (discussed later), or Southern blot analyses employing multiple enzyme digests with the human *PAX6* cDNA as a probe, may prove successful alternatives. When suspected, gross deletions or rearrangements can be detected by cytogenetic methods, such as a banded karyotype or fluorescence *in situ* hybridization (FISH) of metaphase chromosomes.

Methods for Sample Preparation

Extraction of Genomic DNA from Peripheral Blood

Blood drawn from peripheral blood to extract genomic DNA (16) can be stored in yellow-top ACD (acid–citrate–dextrose) tubes for up to 3–5 days at room temperature, while that drawn into green-top tubes containing lithium heparin does not store as well, and less successful transformation rates with Epstein–Barr virus (EBV) may be encountered. For preparation of genomic DNA within 1–2 days of the time of blood drawing, however, the more readily obtainable green-top or purple-top (sodium EDTA) tubes are perfectly suitable. It has also proven possible, using a protocol published elsewhere (17), to identify *PAX6* mutations in genomic DNA extracted from paraffin-embedded postmortem specimens (18).



HP6-11up	5'-CATGTCTGTTTCTCAAAGGGA-3'	HP6-11	5'-CCGGGAACTTGAACTGGAAC-3'
HP6-12	5'-AAGTGAACCTGATATGTCTC-3'	HP6-12down	5'-CACCAAAATGAATAAAAGTTTG-3'
HP6-13	5'-AATGGCTGTTAGAGCCGCTT-3'	HP6-15	5'-ATAAAACTCTTAAATTCGTGGC-3'
HP6-14	5'-CCATGATCAACAAGCTTTGCC-3'	HP6-17	5'-ATATCGAAGACACACTCTACC-3'
HP6-16	5'-TTAACAGATGAAGTTTATGTGC-3'	HP6-19down	5'-GAACAATTAACTTTTGCTGGCC-3'

FIG. 2 Diagram showing the structure of *PAX6* exon 13, encoding the terminal portion of the *PAX6* coding region (shown in black) and the entire 3′ untranslated region. Note that there are two additional polyadenylation sites, 5′ and 3′ to the one shown. The positions and sizes of the individual PCR products (denoted segments) are shown, and the sequences and names of the individual PCR primers (all shown 5′ to 3′) are listed below. For additional details regarding the intron 12-exon 13 junction, consult ref. (13).

Reagents

Acid–citrate–dextrose (ACD) solution A, yellow-top Vacutainer blood tubes (Cat. No. B3003-51; Scientific Products, Bridgeport, NJ) or lithium heparin green-top Vacutainer blood tubes (Cat. No. B2988-54; Scientific Products)

Sodium EDTA purple-top Vacutainer blood tubes (Cat. No. B2992-52; Scientific Products), DNA extraction purposes only

Proteinase K (Cat. No. 745-723; Boehringer Mannheim, Indianapolis, IN)

Proteinase K digestion buffer:10 m*M* Tris (pH 7.5), 100 m*M* NaCl, 1 m*M* EDTA

Sodium dodecyl sulfate (SDS) solution, 20% (w/v)

Lysis solution: 50 m*M* *N*-2-hydroxyethylpiperazine-N′-2-ethanesulfonic acid (HEPES, pH 8.0), 50 m*M* NaCl, 5 m*M* $MgCl_2$, 10% (w/v) sucrose, 0.5% (v/v) Triton-X; filter sterilize and store at 4°C

Method

1. Draw 4–10 ml of peripheral blood by venipuncture into an appropriate Vacutainer tube (see above), using universal precautions. Invert and mix well.
2. Add 4–10 ml of anticoagulated blood to 3 vol of 1× phosphate-buffered saline (PBS, without Ca^{2+} or Mg^{2+}) in a 50-ml centrifuge tube and spin at 1500 *g* for 15 min at room temperature.
3. Aspirate the supernatant carefully, leaving the white "buffy coat" of mononuclear cells intact. Resuspend the cell pellet in 20 ml of ice-cold lysis buffer. Pipette the cells up and down to lyse and spin at 3000 *g* for 15 min at 4°C.
4. Carefully aspirate the supernatant and resuspend the nuclear pellet in 5 ml of proteinase K digestion buffer containing 100 μg of proteinase K per milliliter. Transfer the solution into a polypropylene (organic solvent resistant) 15-ml Falcon tube and add 0.125 ml of 20% (w/v) SDS (0.5% final concentration) and incubate overnight at 40–50°C, mixing the tube gently on a small rotator to ensure efficient digestion.
5. Add an equal volume of Tris-equilibrated phenol (pH 7–8) and gently rotate the tube at room temperature for 2–16 hr. Discard the organic phase.
6. Repeat the extraction once with phenol and twice with chloroform. Precipitate the DNA with 2 vol of cold ethanol. Isolate the DNA by spooling on a glass rod. Drain but do not dry.
7. Resuspend the genomic DNA in ~100 μl of TE [10 m*M* Tris (pH 7.6), 1 m*M* EDTA]. Check the yield and purity of the genomic DNA by measuring the absorbance of an aliquot at 260 and 280 nm in an ultraviolet (UV) spectrophotometer, taking care to ensure that the DNA is entirely in solution. The yield of genomic DNA should be ~200 μg per 5 ml of blood. Store at 4°C.

Immortalization of Lymphoblasts with Epstein–Barr Virus

Lymphocytes can be immortalized with EBV (19).

Reagents

B95-8 Marmoset leukocyte Epstein–Barr virus producer cell line, available from the Coriell Cell Repository (Camden, NJ)

Cyclosporin A (Sandoz, East Hanover, NJ)

LeucoPrep (sodium diatrizoate) mononuclear cell separation tubes (Cat. No. 2750 or 2751; Becton Dickinson, Mountain View, CA)

Roswell Park Memorial Institute medium (RPMI 1640) with 10% (v/v) fetal calf serum (FCS; GIBCO-BRL, Gaithersburg, MD)

Methods

1. Draw 4–10 ml of peripheral blood by venipuncture into a yellow- or green-top Vacutainer tube (see above), using universal precautions. Invert and mix well.
2. Add 4–8 ml of anticoagulated blood to the Leucoprep tube and spin at 1500 *g* for 30 min at room temperature.
3. Remove the mononuclear cells, which form a distinct band, with a cotton plugged Pasteur pipette, keeping the volume aspirated along with the cells as small as possible.
4. Place the mononuclear cells into a 15-ml tissue Falcon tube and wash once at room temperature by centrifugation at 300 *g* in 1× PBS (without Ca^{2+}, Mg^{2+}).
5. Place mononuclear cells isolated from 10 ml of blood in 3 ml of RPMI 1640 plus 10% (v/v) heat-inactivated FCS culture medium containing cyclosporin A (20 μg/ml) (Sandoz). Add 0.5 ml of virus stock (prepared by filter sterilizing the supernatant from a 1-week saturated culture of producer cells). Transfer to an upright 25-cm^2 tissue culture flask and incubate at 37°C in 5–10% (v/v) CO_2.
6. After 7 days, replace with fresh medium lacking cyclosporin A. After 2 weeks, it may be possible to detect the presence of clumps of cells, indicating successful transformation.
7. Continue to culture for about 4 weeks, increasing the volume of medium slowly and replacing it every 3–5 days or when the pH indicator turns yellow. Increasingly frequent pH changes of the medium and clumping of cells usually signifies successful transformation.
8. Aliquots of cells should be frozen in medium containing 10% (v/v) dimethyl sulfoxide (DMSO) and 50% (v/v) FCS in 1.2-ml plastic freezing vials, archived in liquid nitrogen, and tested by thawing and subculturing.
9. Prepare genomic DNA by the method outlined above, or RNA by the phenol–guanidinium thiocyanate method (20).

Detection of PAX6 Mutations by SSCP Assay

Polymerase Chain Reaction Amplification

The sequences of the oligonucleotide primers used for the PCR step of the SSCP assay of *PAX6* mutations are given in Ref. 13 and Fig. 2. These were designed to permit amplification of all *PAX6* exons under similar PCR conditions (i.e., using only two different annealing temperatures). In general, the primers are 20 nucleotides in length with ~50% GC content. A significant danger is the accidental introduction of contaminating sequences into the primers themselves. We recommend storing the PCR oligonucleotide primers in H_2O in multiple aliquots at −80°C, separate from the working stocks. In this case, introduction of contaminants into the primers does not necessitate primer resynthesis.

Reagents

Genomic DNA (template), prepared as described above
Taq polymerase, prepared as described (21)
Taq polymerase buffer (10×): 100 m*M* Tris-HCl, 500 m*M* KCl, 15 m*M* $MgCl_2$, 0.1% (w/v) gelatin, pH 8.3
dNTP stock solution (10×): dATP, dCTP, dGTP, dTTP (200 μM each)
Primer stocks (10–20 μM)

Methods

1. Assemble the first PCR reaction as follows:

Taq polymerase	1 μl
Taq buffer, 10×	10 μl
Forward primer	20 pmol
Reverse primer	20 pmol
dNTPs (200 μM each), 10×	10 μl
Add genomic DNA template, prepared as described above	100–200 ng
Water	To 100 μl

2. Control reactions lacking template should always be included and should be devoid of product. Polymerase chain reaction contamination may result in reisolation of a previously amplified mutation or wild-type fragments.
3. Perform the PCR as follows (13): for exons 1–12, 5 min at 95°C; followed by 40 cycles of 1 min at 95°C, 2 min at 60°C for annealing, and 2 min at 72°C for extension. For exons 5a and 13, the annealing temperature is 55°C. In a case where DNA was recovered from paraffin-embedded specimens, it was necessary to perform 10 additional rounds of amplification (18).
4. Monitor the PCR products on a 1.5–2.0% (w/v) agarose gel containing ethidium bromide to make sure the PCR reaction worked as expected, before proceeding to the next step.

SSCP Assay

For SSCP analysis, 1 μl from the first PCR reaction is used to direct a second PCR of seven cycles, which is chosen empirically to render a signal sufficiently strong for autoradiographic detection on an overnight exposure. By including only one radiolabeled primer in the second PCR reaction, each strand is analyzed separately, making the gel easier to interpret. Performing a first round of PCR prior to the SSCP analysis permits the quality of the PCR result to be evaluated before running the SSCP gel and facilitates the cloning and subsequent analysis of mutant PCR products. In some cases involving small deletions or insertions, heterozygosity may be apparent when the products of the initial PCR are analyzed by agarose gel electrophoresis. In such cases, the mutant allele may appear as a fragment of different mo-

bility than the wild-type allele, and the SSCP assay simply becomes confirmatory. Experiments analyzing *PAX6* mutations under the conditions described below and also on an MDE gel matrix (butylamide-doped polyacrylamide; AT Biochem, Malvern, PA) have shown a modest advantage to the latter, for a small number of mutations detected by only one of these techniques.

Reagents

[γ-^{32}P]ATP, ~6000 Ci/mmol (Cat. No. NEG002Z; NEN, Boston, MA)
Oligonucleotide primers, as described above
Polynucleotide kinase (Cat. No. 174-645; Boehringer Mannheim)
Kinase buffer (10×): 500 m*M* Tris-HCl, 100 m*M* $MgCl_2$, 1 m*M* EDTA, 50 m*M* dithiothreitol (DTT), 1 m*M* spermidine (pH 8.2)

Methods

1. End label both PCR primers of each set, combining 50 pmol of primer, 1 μl of polynucleotide kinase, 5 μl of 10× kinase buffer, and 100 μCi of [γ-^{32}P]ATP in a 50-μl total volume at 37°C for 30 min. Check the reaction by determining the amount of incorporation on Whatman DE-81 filters (Cat. No. 21426–168; VWR, Bridgeport, NJ) (22), or after step 2.
2. Remove the unincorporated label by passage through a 1 ml Sephadex G-25 spin column (22).
3. Set up a second PCR reaction as follows:

First PCR reaction (from prior 100-μl reaction)	1 μl
Taq polymerase	1 μl
Taq polymerase buffer, 10×	10 μl
Forward- or reverse-radiolabeled oligonucleotide primer	2 pmol (~10 μCi)
Unlabeled forward primer	20 pmol
Unlabeled reverse primer	20 pmol
dNTPs, 200 μ*M*	16 μl
Water	To 100 μl

4. Conduct the PCR reaction under the same conditions as the first PCR reaction, except that only seven cycles are performed (see above). The reaction volume can be scaled down for convenience.
5. Mix a 2-μl aliquot of the PCR reaction with 2 μl of loading dye mix [95% (v/v) formamide, 0.1% (w/v) xylene cyanol, 0.1% (w/v) bromphenol blue] in a microcentrifuge tube, and heat to 90°C for 5 min. Centrifuge the samples and chill on ice.
6. Load the samples on a 0.4 mm × 20 cm × 40 cm nondenaturing 6% (w/v) polyacrylamide gel containing 10% (v/v) glycerol and 0.5× Tris–borate EDTA (TBE) buffer and run at room temperature at 4 W overnight for 16 hr.

7. Dry the gel on Whatman (Clifton, NJ) 3MM filter paper (paper chromatography grade) and expose overnight to X-ray film (XAR; Kodak, Rochester, NY) with an intensifying screen at −80°C.

It is often helpful to run samples from all potentially informative family members at the same time, because it can immediately be determined whether or not there is appropriate segregation of any novel conformers that are detected. In the case of a heterozygous disease like aniridia, samples containing a candidate mutation should also contain a band corresponding to the wild-type allele. We have found that standard shark's tooth sequencing combs are easily loaded, afford good resolution, and allow analysis of multiple sets of samples at a time. Because most diseases are analyzed in batch fashion when a sufficient number of samples are available, it is most efficient to run PCR products derived from several exons simultaneously on the same gel. If fragments of different sizes are to be run together on the same gel, they should be grouped so that they are of nominally equivalent sizes, so that the maximum run time and hence maximal resolution can be utilized for all samples. In each case, we typically include both denatured (heated prior to loading, step 5 above) and nondenatured wild-type control samples; this permits determination of which bands are single stranded. As mutations are identified, and with the obvious caveats about PCR contamination in mind, it may prove useful to include select mutant fragments in future assays to control for the quality of the gel and electrophoresis conditions. This could prove reassuring, because in any random batch of candidate mutant samples for any specific PCR fragment, only a minority of the PCR fragments, if any, are expected to contain a novel conformer representing a mutation.

A premise of the SSCP assay is that the labeled DNA fragments are rendered single stranded by thermal denaturation prior to electrophoresis. However, a variable fraction of the single-stranded DNA normally reanneals prior to electrophoresis. If a heterozygous mutation is present, half of these renatured molecules are expected to form heteroduplexes, and these may be resolved from the reannealed homoduplexes (23), as shown in Fig. 3A. The identity of these latter forms can be assigned, as noted above, by comparison to a wild-type sample that is specifically not heat denatured prior to loading the gel. Thus, the nondenatured double-stranded bands running at the bottom of an SSCP gel may display altered mobilities not only in cases of small deletions or insertions, but also even when a simple base substitution mutation exists. The original PCR products corresponding to samples displaying novel conformational bands are then cloned into the *Sma*I site of pGEM3Z by A-T annealing (24) for further analysis as described below.

Confirmation of Mutations Detected by SSCP Assay

DNA sequence analysis is still the definitive method for establishing the identity of mutations. In the event that protocols for the direct sequencing of PCR products are

available (25), then speed considerations make this the method of choice. If such a method is not available, then subcloning the PCR product is required. In the case of heterozygosity, the resulting subclones should contain equal numbers of wild-type and mutant sequences. Because the inclusion of restriction sites on the ends of primers increases cost and may introduce uncertainty about the effectiveness of the primers for PCR, we have used T-vector (T:A vector) cloning followed by hybridization with the radiolabeled oligonucleotide primers to identify insert-containing colonies, and plasmid miniprep sequencing for final analysis. We have arbitrarily chosen to sequence five independent subclones for each PCR product that shows a novel SSCP band because from statistical considerations the mutant sequence should be represented one to two times in most sets of five subclones. While the failure to find a mutation in five independent clones does not preclude its existence, it would seem advisable to repeat the SSCP analysis before sequencing additional clones. It is important to keep in mind that it is not uncommon to find isolated sequence changes in any PCR-derived fragment as an artifact of *Taq* polymerase misincorporation. Moreover, such changes are rarely responsible for an abnormal SSCP shift observed for the same fragment. While finding a presumptive mutation in two or more subclones makes it likely to be authentic, some care must be employed to make sure that these clones are truly independent, that is, not simply sibs. Criteria for establishing the independence of subclones include: (a) clones in opposite orientations (an advantage of T vector and symmetric cloning strategies, as opposed to asymmetric cloning strategies), and (b) finding nucleotide sequence differences at the vector–insert junction, reflecting the variable insertion of a T:A nucleotide pair at one or both vector–insert junctions.

As mentioned above, because SSCP is a PCR-based approach, it is possible that the $\sim\frac{1}{1000}$ error rate of *Taq* polymerase will introduce artifactual changes into the nucleotide sequence, especially if template molecules are damaged, often resulting in the spurious substitution of an adenosine residue (26). Newer thermophile polymerases have been introduced that possess higher fidelities, making this concern less likely when these are employed. However, caution should still be employed when any novel DNA sequence change is detected using a PCR-based approach. Simple mathematical considerations indicate that in order to be quantitatively significant, a PCR error must be introduced in the first several rounds of PCR amplification. Approaches to reducing the risk of a *Taq*-induced error include minimizing the amount of template, so that low dNTP concentrations are not encountered during early rounds of PCR amplification, as this tends to favor misincorporation events, and using multiple independent PCR reactions, which are then combined. In the case of *PAX6* aniridia mutations, because aniridia is a heterozygous disease, SSCP bands representing authentic mutations are expected to occur in an $\sim 1:1$ ratio with wild-type products. Although conditions may exist in which wild-type and mutant alleles are not amplified equally (27), we have not observed this to be a significant problem,

and suspicion should be aroused when novel products are observed on SSCP gels or at the DNA sequence level in unexpected, especially submolar, proportions.

A simple method of further verifying a mutation in a PCR product that consists of a mixture of wild-type and mutant sequences exists if a restriction enzyme site is either created or destroyed by the mutation. Because in general these PCR products are small, the digestion products will be even smaller, and these can therefore be most effectively displayed by nondenaturing polyacrylamide gel electrophoresis, staining with ethidium bromide, and visualization by UV illumination. Alternatively, small fragments may be visualized by silver staining or, if radiolabeled primers are used for the PCR, the gel may be fixed and dried and the product detected by autoradiography. These approaches are illustrated elsewhere (13, 18, 28).

Reverse Transcription-Polymerase Chain Reaction Analysis

A second approach to the analysis of mutations in a gene such as *PAX6* with many discrete exons is to assay transcripts directly using RT-PCR. Ordinarily, this approach is limited to tissues from which RNA is readily obtained, such as skin fibroblasts or peripheral blood leukocytes. The *PAX6* gene is primarily expressed in ocular and CNS tissues, however, and it is impractical to isolate RNA from these tissues when seeking to analyze *PAX6* in patients with aniridia. Because of the extraordinary sensitivity afforded by PCR, it has proven possible to use EBV-transformed lymphocytes from aniridia patients as a source of RNA from which to assay the low-abundance illegitimate transcripts that are ordinarily synthesized from almost all genes (reviewed in Ref. 29). In some cases, depending on the levels of transcripts that are expressed, this may require a nested PCR primer approach. Such an approach has proven successful for the analysis of *PAX6* transcripts (30, 31), and offers the considerable advantage of not requiring knowledge of the complete genomic structure for the gene of interest. In addition, splicing errors may be immediately apparent, appearing as novel-sized band(s) on agarose gel electrophoresis after RT-PCR. A caveat is that it may not be immediately possible to determine the precise molecular nature of mutations residing within the intronic portion of a splice acceptor or splice donor. Although we have primarily utilized the SSCP method presented above, readers should be aware of the potential utility of the RT-PCR approach, and note that the two approaches are to some extent complementary. A specific application of the RT-PCR assay is described below, in connection with the assay of a particular splicing mutation in *PAX6* (28).

Analysis of PAX6 Mutations Affecting Splicing

Splicing mutations are in general more readily identified by RT-PCR approaches than by SSCP, because the accuracy of multiple splicing events can be assayed simultaneously. Analysis of PCR products by agarose gel electrophoresis will readily reveal

significant alterations in transcript sizes, resulting from use of cryptic splice donors and acceptors. The finding that *PAX6* transcripts in EBV-transformed lymphocytes, normally below the threshold of detection in conventional Northern blot and RNase protection analyses, can be detected by RT-PCR offers an efficient approach to this problem. In the example described below, a candidate T→C transition was found by SSCP assay in the −3 position of the splice acceptor flanking the alternatively spliced *PAX6* exon 5a (Fig. 3A). In this case, the presumed T→C transition mutation at the −3 position of the exon 5a splice acceptor is actually predicted to favor inclusion of the alternatively spliced exon 5a, and disfavor expression of the non-exon 5a-containing form of *PAX6*. Although RT-PCR was not used to identify the mutation originally, the *in vivo* splicing assay was essential to prove that the observed change was functionally significant, because both pyrimidines are allowed at this position.

To determine the effect of the presumptive splice site mutation on splicing efficiency of *PAX6* exon 5a, a genomic fragment containing the three exons of interest from the wild-type *PAX6* gene (exons 5, 5a, and 6) was cloned in the sense orientation into the *Bam*HI site of the exon amplification vector pSPL1 (32). The approach used is shown schematically in Figure 4. This vector contains a cloning site located within a 2.3-kb intron derived from the human immunodeficiency virus (HIV-1) *tat* gene. The intron forms part of a transcription unit that begins with the simian virus 40 (SV40) *ori* promoter, and is flanked by *tat* splice donor and acceptor sites, and is designed to permit splicing from these sites to endogenous splice sites that reside within the cloned insert. The resulting plasmid was then subjected to site-directed mutagenesis (33) to introduce the desired point mutation into the wild-type *PAX6* gene fragment for comparison. Site-directed mutagenesis was chosen to introduce this mutation into the wild-type sequence, rather than subcloning a fragment from the mutant into the larger wild-type fragment, for two reasons. The intronic sequence flanking exon 5a was limited as were available restriction sites, and PCR amplification of a large region of the mutant allele carries a risk of introducing undesired sequence changes, a possibility difficult to exclude without fully sequencing the amplified fragment. To assess specifically the ratio of the alternative products transcribed from the introduced plasmid DNA, the PCR primers were modified from the original exon amplification scheme (32). The forward PCR primer sequences are located in human *PAX6* exon 5 and the reverse primer encompasses the *PAX6–tat* junction. This strategy prevents cDNAs derived from the endogenous COS-7 (simian) *PAX6* gene or from the pSPL1 vector alone from being amplified. The plasmids are then introduced into COS-7 cells in separate transfections. After 3 days, RNA is isolated and analyzed by RT-PCR, using random hexamer priming for the reverse transcriptase step, and PCR amplification with the primers 5′ GTCACAGCGGAGTGAATCAGCTCGG 3′ and 5′ TTGGGAGGTGGGTGCTTGGTATGTTATCG 3′. As described above, these primers correspond, respectively, to the upstream primer within exon 5 and the downstream primer that crosses the *tat*–exon 6 splice junction (*tat* sequence underlined). The results from this assay are shown in Figure 3B.

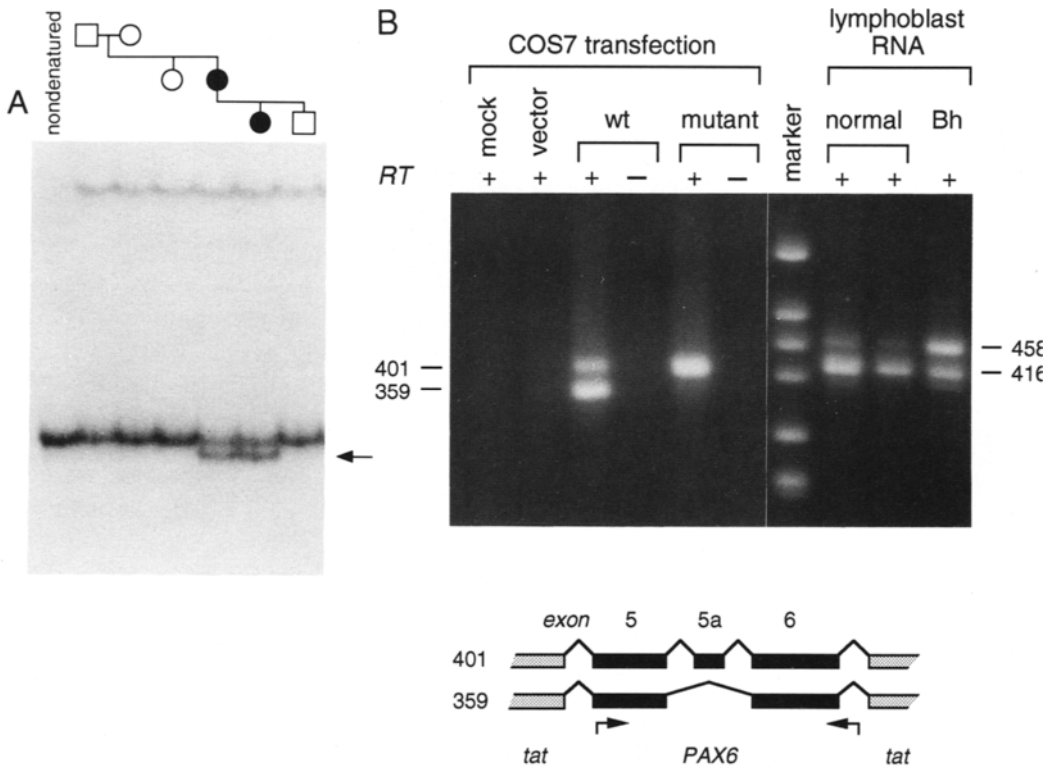


FIG. 3 Analysis of the *PAX6* mutation Bh by *in vivo* splicing assay. (A) Heteroduplex analysis on an SSCP gel demonstrating the presence of novel *PAX6* heteroduplexes in the affected mother and daughter (arrow) in the Bh pedigree shown. The fragment amplified spans *PAX6* exon 5a, and the primers used are described in Ref. 13. In this pedigree, the mutant ocular phenotype appeared first in the mother, consistent with the molecular findings. Note that the single-stranded band at the top of the gel fails to reveal this mutation. (B) *In vivo* splicing assay of the *PAX6* Bh mutation in COS7 cells (left), and assay of illegitimate *PAX6* transcripts in lymphoblastoid lines by RT-PCR (right). An ethidium bromide-stained agarose gel illustrating that the Bh mutation increases the proportion of 401-bp transcripts containing exon 5a relative to the 359-bp transcripts lacking exon 5a. COS7 cells were transfected with pSPL1 expression plasmids containing either wild-type or mutant *PAX6* genomic sequences, no DNA (mock) or pSPL1 DNA alone (vector). The addition (+) or omission (−) of reverse transcriptase (RT) is indicated above the lanes. The upstream primer used for the RT-PCR analysis is located in exon 5 of *PAX6,* while the downstream primer spans the exon 6–HIV–*tat* gene junction (28). In the lymphoblastoid cell RNA from normals or patient Bh, the same relative effect is observed, with an increased proportion of 458-bp exon 5a-containing transcripts relative to 416-bp nonexon 5a transcripts. These products were amplified with primers located in exons 3 and 6, thus accounting for the slightly larger products than those observed in the *in vivo* splicing assay. [Reproduced from Ref. 28, with permission of Cold Spring Harbor Laboratory Press.]

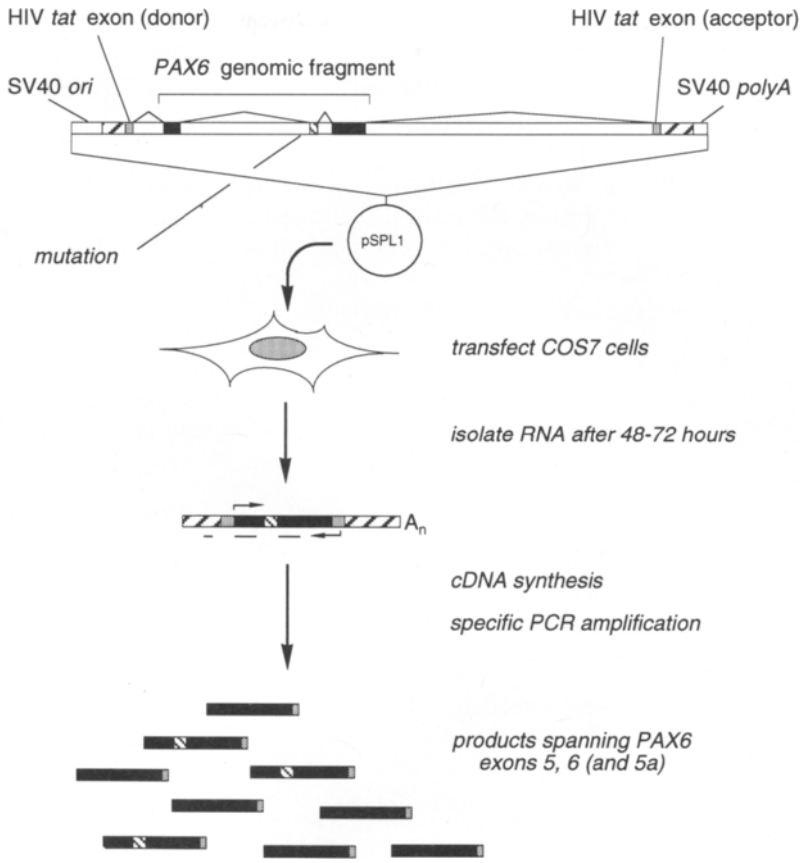


FIG. 4 Scheme illustrating steps in the *in vivo* splicing assay. Shown is a portion of the exon trapping vector pSPL1 containing a portion of the *PAX6* genomic fragment cloned between the HIV-*tat* gene splice donor and splice acceptor. The SV40 origin permits amplification in COS cells after transient transfection. Following RT-PCR analysis, the amplified product is a mixture of fragments corresponding to the two alternatively spliced transcripts containing or lacking exon 5a. The relative ratio of the fragments approximates the relative abundance of the respective transcripts.

Reagents

Plasmid vector pSPL1 (Cat. No. 18447-011; GIBCO)
Taq polymerase (19)
Taq buffer (10×): 100 m*M* Tris-HCl, 500 m*M* KCl, 15 m*M* $MgCl_2$, pH 8.3
RNazol B (guanidinium thiocyanate–phenol–chloroform) (Biotecx, Houston, TX)

RNase H⁻Moloney murine leukemia virus (MMLV) reverse transcriptase (Cat. No. 18053-017; BRL Superscript, GIBCO-BRI, Gaithersburg, MD)
Reverse transcriptase (RT) buffer (5×): 250 m*M* Tris-HCl (pH 8.3), 375 m*M* KCl, 15 m*M* $MgCl_2$
dNTPs, 10 m*M*
RNasin, 40 units/ml (Promega, Madison, WI)
$pd(N)_6$ primer, 100 μg/ml (Boehringer Mannheim)
PCR primers encompassing HIV *tat* and human *PAX6* gene sequences (see above)
COS-7 cells (ATCC CRL1651)

Methods

1. Plate COS cells at 10^6/60-mm culture dish in DMEM with 10% (v/v) FCS.
2. Transfect 1.5 μg of the above plasmids (wild-type, mutant, or vector alone) coated with cationic lipopolyamine (Transfectam; Promega), and replace with fresh medium after 6 hr.
3. Harvest the cells after 72 hr and prepare total cellular RNA using the phenol–guanidinium thiocyanate (RNazolB) method (20).
4. Perform a reverse transcription reaction on ~2 μg of total RNA. Heat the RNA to 90°C for 5 min, then quench on ice before adding the remaining reagents. Assemble the following and incubate at 37°C for 1 hr:

Total RNA (isolated as described above)	2 μg
RNasin	1 μl (40 units)
Reverse transcriptase	1 μl (200 units)
Reverse transcriptase buffer, 5×	4 μl
$pd(N)_6$ primer	0.5 μl (~25 pmol)
dNTPs (10 m*M* each)	2 μl (final concentration, 1 m*M* each)
Water	to 20 μl

   Prepare control reactions lacking reverse transcriptase enzyme and incubate in parallel.
5. Subject the entire cDNA template derived from the above RT reaction to PCR, using the following conditions (for the primers described above): 1 min at 95°C, 2 min at 63°C, and 2 min at 72°C, for 45 cycles.
6. Analyze the resulting products by agarose gel electrophoresis. Examine carefully for the presence of aberrant splicing products or for an altered ratio of products. For an example, see Fig. 3B.

Functional Analysis of *PAX6* Mutations Affecting DNA Binding

The *PAX6* gene product contains two independent DNA-binding domains, an N-terminal paired domain (PD), and a centrally located homeodomain (HD). Mutations

in either domain are expected to affect the DNA-binding ability of the entire PAX6 protein. An assumption that is valid for many DNA-binding proteins is that the individual DNA-binding domains are modular. In general, the DNA-binding properties of the entire protein are accurately reflected by the binding properties of the individual DNA-binding domains. This observation greatly simplifies the task of analyzing the functional consequences of mutations in DNA-binding domains, because only these domains themselves need to be expressed as proteins. Problems commonly associated with attempts to express full-length proteins in *Escherichia coli,* such as insolubility and low expression due to unfavorable codon usage, are thus avoided. Several excellent prokaryotic expression vectors are available for this purpose, including bacteriophage T7-based glutathione *S*-transferase (GST) pGEX vectors (34), enabling affinity purification of the GST fusion protein on glutathione (GSH) agarose beads, and T7-based vectors containing a polyhistidine epitope tag at either the N or C terminus, permitting affinity purification on an Ni-chelate resin (35). If the DNA recognition sequence is known for the DNA-binding domain in question, the electrophoretic mobility shift assay (EMSA) provides an easy and powerful assay to test for alterations in affinity or sequence recognition. If, on the other hand, the DNA recognition sequence is unknown, as was initially the case with PAX6, then it must first be determined. In the following sections we describe methods for expression and purification of polyhistidine PAX6 paired domain fusion proteins, and a binding site selection method to determine an optimal consensus DNA recognition sequence for the PAX6 paired and homeodomains. Mutant proteins can then be tested for their ability to bind to DNA recognition sequences, or subjected to a binding site selection to determine their altered sequence specificity.

Expression of PAX6 Paired and Homeodomains in Escherichia coli as Polyhistidine Fusion Proteins

Several relatively simple methods now exist for expression of proteins in bacteria that facilitate purification. In most cases, this involves expression of a fusion protein with a moiety that has high affinity for a coated matrix, thus allowing column or batch purification. The expression of proteins as fusions with glutathione *S*-transferase has been well described [34] and has been successful in our hands. We have also expressed the PAX6 paired domain with an amino-terminal polyhistidine tag that allows for rapid purification by affinity to a nickel resin.

Reagents

pET expression vector (Novagen, Madison, WI) (pET16b or pET14b)
Ni^{2+}–NTA–agarose (Cat. No. 30230; Qiagen, Chatsworth, CA)
French press buffer (FPB): 10% (w/v) sucrose, 25 m*M* $NaPO_4$ (pH 7.5), 150 m*M* NaCl, 4 m*M* 2-mercaptoethanol
Wash buffer: 60 m*M* imidazole, 0.5 *M* NaCl, 20 m*M* Tris (pH 7.9)
Binding buffer: 5 m*M* imidazole, 0.5 *M* NaCl, 20 m*M* Tris (pH 7.9)

Elution buffer: 0.5 *M* imidazole, 0.5 *M* NaCl, 20 m*M* Tris (pH 7.9)
Isopropyl-β-D-thiogalactopyranoside (IPTG) stock solution: 100 m*M* in water
Phosphate-buffered saline (PBS)

Methods

1. Clone the appropriate coding region of the gene of interest in frame into a pET expression vector and transform BL21(DE3) bacteria (protease deficient).
2. Pick a single colony into 10 ml of LB-ampicillin medium and grow to saturation.
3. Dilute 10 ml of culture into 1 liter of LB-ampicillin medium and incubate at 37°C at 300 rpm for 3 hr.
4. Add 10 ml of IPTG stock solution and continue incubation for 6–12 hr.
5. Collect the bacterial pellet by centrifugation at 5000 *g* for 10 min at 4°C.
6. Wash the pellet in 200 ml of ice-cold PBS and centrifuge as above.
7. Resuspend in 35 ml of ice-cold FPB with 2 m*M* EDTA.
8. Lyse the bacteria by sonication or (preferably) by passage through a French press at 1100 psi twice.
9. Clear the lysate of debris by centrifugation at 16,000 *g* for 10 min at 4°C.
10. Wash the pellet with cold FPB (without EDTA) and recentrifuge.
11. Elute DNA-binding proteins from the pellet by vortexing in 50 ml of cold FPB with 0.5 *M* NaCl (without EDTA) for 2 min. Centrifuge at 16,000 *g* for 10 min at 4°C.
12. Add 5–10 ml of nickel resin slurry (prewashed with binding buffer) and rock at 4°C for 15 min. Collect the resin by centrifugation at 1000 *g* for 5 min at 4°C.
13. Wash the resin by repeated resuspension and centrifugation in wash buffer until no further protein can be detected in the supernatant. We determine this rapidly by using the Bio-Rad (Richmond, CA) protein assay reagent according to the supplied protocol.
14. Elute the protein in 10 ml of elution buffer, and repeat until no further protein is eluted.
15. Perform buffer exchange as needed for subsequent experiments using a Sephadex G-25 column (e.g., NAP-10 column; Pharmacia, Piscataway, NJ) and standard procedures. We elute the PAX6 paired domain from the G-25 column in 10 m*M* NaH_2PO_4 (pH 7.0)–150 m*M* KCl. This step is critical to prevent precipitation of the PAX6 paired domain on freezing and thawing.
16. Freeze aliquots of final protein elutions in liquid nitrogen and store at −80°C.

Identifying the Optimal DNA-Binding Site for a DNA-Binding Protein

We have utilized an *in vitro* DNA-binding site selection assay to identify the optimal DNA sequence bound by several Pax DNA-binding domains including the paired domains of Pax2, Pax3, and Pax6 and the paired-type homeodomains of Pax3 and

Pax6 (36 and unpublished data). This procedure, modified from existing procedures (37) and illustrated in Figure 5, allows the selection of high-affinity DNA-binding sequences and provides obligatory reagents for studies on the mechanism of DNA binding for this class of transcription factors. The assay involves expressing the isolated DNA-binding domain in *E. coli* as a fusion protein with glutathione *S*-transferase (GST), and mixing the purified protein with a pool of oligonucleotides containing a core of random DNA sequence. Oligonucleotides bound under stringent binding conditions are amplified by PCR, and the selection process repeated until the vast majority of the selected oligonucleotides contain high-affinity binding sites.

Prior to performing this assay several parameters must be defined. These include defining the stringency of the binding reaction in terms of salt concentration and nonspecific competitor DNA concentration, and determining the length of the core of random nucleotides sufficient to accommodate the particular DNA-binding domain being used.

Binding Site Selection Conditions

Preliminary tests should be performed in order to ensure that binding conditions are stringent enough to prevent nonspecific binding of the protein in question to random sequence DNA. This can be performed by setting up binding reactions at a range of salt (KCl) and competitor DNA [poly(dI-dC)] concentrations with a radioactively labeled oligonucleotide probe consisting of a pool of "randomers," each containing a core of random DNA sequence. Initially, we utilized a pool of 66-mers that were synthesized with 27 bp of fixed sequence flanking a core of 12 positions, each of which contained G, A, T, or C at equal frequency. The randomer pool is conveniently synthesized *en masse* simply by introducing an equimolar mixture of A, C, G, or T phosphoramidite at each of the degenerate positions during the oligo synthesis. A 22-bp reverse primer complementary to the 3′ fixed sequence was annealed and the complementary strand filled in using the Klenow fragment of DNA polymerase I and dNTPs, including [α-^{32}P]dCTP. An example of the oligonucleotides we have used is as follows (the restriction sites are underlined):

5′-GCGAATTCGCATGGTCTCAGTATCGAT-$(N)_{12}$-
GCGGCCGCAGATCCTAGATCAGGATCCTC-3′ Randomer pool
3′-GTCTAGGATCTAGTCCTAGGAG 5′ Reverse primer

Reagents

Oligonucleotides: 66- to 89-mer with 12–35 random core positions and 22-mer complementary to the 3′ end of the random oligonucleotide; each at ~400 ng/μl

Glutathione *S*-transferase (GST) fusion protein, expressed as referenced above (34)

Glutathione (GSH)-conjugated agarose beads (Cat. No. G2879; Molecular Probes, Eugene, OR)

Binding buffer (2×): 40 m*M* Tris (pH 8), 1 m*M* EDTA, 5% (w/v) bovine serum

albumin fraction V (Cat. No. 100030; Boehringer Mannheim), 2 m*M* dithiothreitol (added fresh), 20% (v/v) glycerol

Poly(dI-dC): 5 μg/μl stock solution in water (Cat. No. 1219847; Boehringer Mannheim)

[α-^{32}P]dCTP, 3000 Ci/mmol (Cat. No. NEG-013H; NEN)

dNTPs (each 1 m*M*)

Klenow buffer (10×): 100 m*M* Tris (pH 7.5), 100 m*M* $MgCl_2$, 0.5 *M* NaCl, 10 m*M* dithioerythritol (DTE)

Elution buffer: 0.5 *M* ammonium acetate, 1 m*M* EDTA

Methods

1. Anneal equal molar amounts of reverse primer and randomer in dH_2O by heating at 90°C for 10 min and allow to cool to room temperature. Synthesize the second strand at 30°C for 30 min as follows:

Klenow buffer, 10×	2 μl
dNTPs, 1 m*M* each	4 μl
[α-^{32}P]dCTP	5 μl
Klenow fragment (2 units/μl)	1 μl
Annealing reaction, described above	5 μl
Water	To 20 μl

2. Purify the resulting double-stranded pool of oligonucleotides on a nondenaturing 10% (w/v) polyacrylamide gel, wrap the gel in Saran Wrap, expose it briefly to film, and cut out the desired band. Elute in 400 μl of elution buffer, check the absorbance at 260 nm to determine the recovery, and count an aliquot to determine the specific activity, which should be 10^6–10^7 cpm/μg. A typical recovery should be 40–80%. Ethanol precipitate the DNA and resuspend in 20 μl of TE.

3. Mix an equal molar ratio of DNA and GST–fusion protein (typically 250–500 ng each, or 0.5–1.0 μM) attached to glutathione–agarose beads in a volume of 50 μl in 1× binding buffer supplemented with KCl ranging from 0 to 500 m*M*, and poly(dI-dC) ranging from 0 to 500 μg/ml.

4. Perform binding reactions at room temperature for 45 min with agitation to prevent the agarose beads from settling to the bottom of the tube. The tubes can be taped to a Vortex mixer.

5. Pellet the agarose beads with associated protein and DNA by brief centrifugation at 14,000 rpm in a microcentrifuge, wash twice with 1× binding buffer, and count the radioactivity that remains associated with the protein–bead complexes. Select the conditions at which less than 0.1% of the initial radioactivity remains bound for further binding site selection experiments. For the Pax6 homeodomain, these conditions are 50 m*M* KCl, 50 μg of poly(dI-dC) per milliliter, while for the Pax6 paired domain higher stringency [250 m*M* KCl and 250 μg of poly(dI-dC) per milliliter] is required.

Determination of Minimal Length of Random Nucleotide Positions Necessary to Accommodate DNA-Binding Domain

Choosing the appropriate number of random positions to include within the core of the starting oligonucleotide pool used for the binding site selection assay is of critical importance. Too short a stretch of randomers may not allow optimal binding to occur, while too long a stretch increases the complexity of the mixture and may reduce the representation of the optimal site within the pool. Prior information regarding the length of DNA contacted or footprinted by similar DNA-binding domains is helpful in making this decision but, as in the case of the PAX6 paired domain, this may not be available. Our initial studies utilized a pool of oligonucleotides with a core of 12 random positions, a length that proved sufficient for high-affinity binding site selection by the PAX6 homeodomain. However, when sequences from this pool bound by the PAX6 paired domain were selected, we found that the DNase I footprint of the paired domain on these sequences overlapped the flanking fixed sequence, indicating that our initial choice of 12 random positions was not long enough to allow a selection over the total binding site. To recognize this problem, we recommend including restriction sites in the fixed sequence immediately flanking the random core, as shown in the oligonucleotide above. After selection for high-affinity binding sites (see below), individual tightly bound oligonucleotides are cleaved first 5′ and then 3′ to the core. The ability of the paired domain to bind the full-length oligonucleotide can then be compared to the deleted versions in electrophoretic mobility shift assays. If deletion from either end results in a diminished binding affinity, it can be concluded that the initial core of randomers was too short, and the process repeated with an initial pool containing a longer central random core. In the case of the PAX6 paired domain, these experiments indicate that 35 random positions are sufficient to encompass the entire footprint.

It is conceivable that if the central core of randomers is too short and the arbitrarily chosen, fixed flanking sequences are incompatible with binding, then no high-affinity sites will be selected. In this case it would be advisable to try empirically a longer core of randomers. Likewise, the flanking sequences may preclude binding of sufficient affinity to demonstrate an overlapping footprint, or to be recognized by restriction enzyme truncation. It is therefore recommended that the binding site selection assay be repeated with a random core slightly longer than that which has previously proven to be most successful with other DNA-binding domains to ensure that the footprint does not extend beyond the selected core positions.

DNA-Binding Site Selection

The protocol that we have used for selecting binding sites recognized by several Pax paired domain and homeodomains is outlined below and is summarized in Fig. 5. We use pGEX-2T (Pharmacia), which encodes a thrombin cleavage site between the GST moiety and the protein of interest. In the first round of selection, the amount of DNA used depends on the length of the random core. Enough DNA should be used to

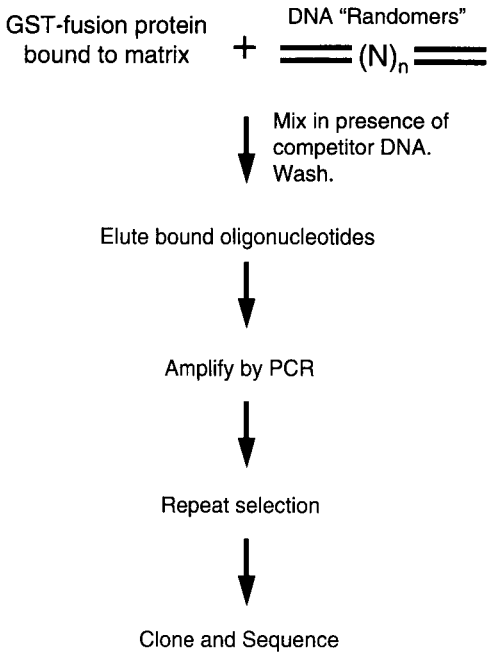


FIG. 5 DNA-binding site selection assay protocol. A DNA-binding domain is expressed in bacteria as a fusion protein with glutathione *S*-transferase and attached to glutathione-coated agarose beads. Oligonucleotides are synthesized with fixed sequences flanking a core of random nucleotides (N). The number of random positions (n) is chosen empirically (see text). The DNA and protein are mixed under stringent binding conditions (determined as described) and the protein–DNA complexes are washed to remove unbound DNA. Specifically bound oligonucleotides are eluted, amplified by polymerase chain reaction, and subjected to further rounds of selection. Finally, after sufficient rounds of selection, the bound oligonucleotides are cloned and sequenced, and a consensus DNA-binding sequence is deduced.

include as complete a representation of sequences as possible, assuming that the relative abundance of any single sequence is Poisson distributed. This can be easily calculated because the number of possible sequences in the original mixture of oligonucleotides equals 4^n, where n is the number of random positions. After each round of selection, we release the bound oligonucleotides by thrombin cleavage of the fusion protein. Releasing the bound DNA with thrombin avoids contamination of the supernatant with DNA bound nonspecifically to the agarose beads as occurs if the sample is boiled to denature the protein at this step, as has been recommended (37). Specifically bound DNA is amplified by PCR and the selection process repeated. We incorporate radioactively labeled nucleotides into the PCR products at each stage to

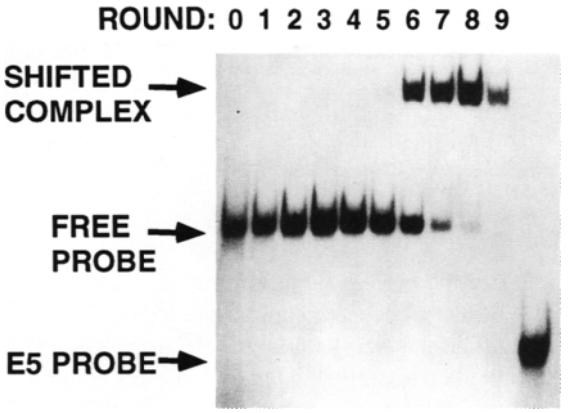


FIG. 6 Selection of specifically bound sequences with the *PAX6* paired domain, expressed as a glutathione S-transferase (GST) fusion protein. Shown is an electrophoretic mobility shift assay (EMSA) demonstrating progressive enrichment of specifically bound sequences after successive rounds of selection. An 83-bp oligonucleotide containing a central region of 25 random nucleotides was used as a probe in lane 1. After each round of selection, the enriched pool of sequences was used as a probe in successive lanes. For comparison, a 45-bp probe containing a previously identified paired domain-binding sequence e5 (38) was included in the last lane with 1 μM PAX6 paired domain, demonstrating that none of the e5 probe is shifted under these conditions. [Reproduced from Ref. 36 with permission of The American Society for Biochemistry and Molecular Biology.]

facilitate monitoring progress of the selection process. Also, the amplified oligonucleotides can be used directly as probes in electrophoretic mobility shift assay (EMSA) experiments (as in Fig. 6). Also, we include two different restriction sites at the extremes of the original oligonucleotides to facilitate cloning and subsequent sequencing after selection.

Reagents

Appropriate DNA-binding domain expressed as a pGEX-2T GST fusion according to standard protocols (34)

Oligonucleotide pool with random core (as described above) and oligonucleotide primers homologous to each flanking fixed sequence

Binding buffer (2×): 40 m*M* Tris (pH 8.0), 1 m*M* EDTA, 0.5% (w/v) bovine serum albumin fraction V (Cat. No. 100030; Boehringer Mannheim), 2 m*M* dithiothreitol (added fresh), 20% (v/v) glycerol

KC1, 1 *M*

Thrombin (Cat. No. T6634; Sigma, St. Louis, MO)

Thrombin cleavage buffer: 150 m*M* NaCl, 50 m*M* Tris (pH 7.5), 2.5 m*M* $CaCl_2$
Tris–borate–EDTA buffer (TBE): 50 m*M* Tris base, 50 m*M* boric acid, 1 m*M* EDTA
Poly(dI-dC): 5-μg/μl stock solution in water (Cat. No. 1219847; Boehringer Mannheim)
Other reagents as described above

Method

1. Purify the protein by affinity to glutathione–agarose beads but do not elute from the beads. Check for appropriate expression and for degradation by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). [The protein attached to the beads can be added directly to Laemmli sample buffer, boiled briefly, quenched on ice, and analyzed by 10% (w/v) SDS-PAGE; the agarose beads remain in the well.]

2. Dilute the protein–bead suspension to 100 ng/μl by adding fresh glutathione–agarose beads in a 50% slurry with appropriate buffer and rocking at 4°C for several hours.

3. Prepare double-stranded oligonucleotides containing a random sequence core as described above (see *Binding Site Selection Conditions:* step 1). Gel purify on a 10% (w/v) nondenaturing polyacrylamide gel.

4. Extract the oligonucleotides from the gel slice by soaking in elution buffer at 37°C for several hours or overnight, or by electroelution in a dialysis bag in 0.2× TBE. Precipitate the DNA and resuspend in 20 μl of water.

5. Prepare the binding reaction with empirically determined KCl and poly(dI-dC) concentrations (see *Binding Site Selection Conditions:* above):

Binding buffer, 2×	10 μl
KCl	x μl
Poly(dI-dC)	y μl
GST–protein/beads	3 μl
Oligonucleotides	2 μl
Water	To 20 μl

Shake for 45 min at room temperature to maintain suspension of the agarose beads (tape to a Vortex mixer at medium speed).

6. Wash by adding 1 ml of ice-cold 1× binding buffer with the appropriate KCl concentration [omit poly(dI-dC), because the cost would be prohibitive], briefly vortex, and centrifuge for 1 min at 14,000 rpm in a microcentrifuge. Occasionally the agarose beads will stick to the sides of the 1.6-ml Eppendorf tube, in which case siliconized tubes can be used. Remove most of the supernatant but leave about 50 μl. Centrifuge briefly. The agarose beads can just be seen at the bottom of the tube.

Carefully remove the remaining supernatant with a fine sequencing pipette tip. Wash the beads two more times in binding buffer.

7. Release the specifically bound oligonucleotides by adding 50 μl of thrombin solution containing 2 units of thrombin in thrombin cleavage buffer. Incubate with agitation for 15 min at room temperature. Extract the supernatant sequentially with phenol–chloroform (1 : 1, v : v) and chloroform.

8. Amplify the selected oligonucleotides by PCR. Add 5 μl of the extracted oligonucleotide solution to a standard PCR reaction containing 100 ng of each primer homologous to the fixed flanking region of the original oligonucleotide pool, 5 μl of [α-^{32}P]dCTP (3000 Ci/mmol), 2 μM dCTP, 20 μM dTTP, 20 μM dGTP, and 20 μM dATP in a volume of 50 μl. Amplify for 20 cycles at 94, 65, and 72°C for 30 sec each step, for denaturation, annealing and extension, respectively.

9. Ensure homoduplex formation by a final round of denaturing, annealing, and extending after adding an additional 100 ng of each primer and enough dNTPs to yield a final concentration of 200 μM. Extend for 10 min.

10. Purify the PCR product to remove primer dimers and other contaminants by 10% (w/v) nondenaturing polyacrylamide gel electrophoresis. Elute from the gel as described above, precipitate, and resuspend in 20 μl of water.

11. Repeat the selection process from step 3 above. Progress during subsequent rounds of selection can be qualitatively assessed by determining the amount of radioactivity associated with the washed agarose bead–protein complex. If the selection process is successful, more and more radiolabel will be retained by the beads after each round.

12. Determine the progress of the selection process more quantitatively by performing an EMSA with the DNA-binding protein of interest and the gel-purified PCR products after each round of selection as probe (see Fig. 5 and EMSA protocol below). If the selection is successful, an increasing fraction of the DNA should be shifted after each round.

13. Once much of the PCR product is bound at high affinity, or if the selection process is not progressing further, clone and sequence the final selected oligonucleotides. Align the selected sequences using available computer software such as the Wisconsin Genetics Computer Group program Pileup. (GCG, Madison, WI)

Use of Consensus DNA-Binding Sites to Assess Binding Properties of DNA-Binding Proteins

We have utilized the consensus binding sites determined for the Pax paired and homeodomains to study various aspects of protein–DNA interactions including binding affinity, protein dimerization properties, and the effect of alternative splicing on DNA-binding characteristics. Some general procedures used in these analyses are described below.

Electrophoretic Mobility Shift Assay

Protein–DNA interactions can be conveniently assessed by performing a binding reaction followed by electrophoretic separation of bound complex from unbound protein and DNA. Usually, the DNA is radioactively labeled and protein binding results in retardation of the migration of DNA through the gel.

Reagents

Purified DNA-binding protein or nuclear extract (see above)
Oligonucleotides
[α-^{32}P]dCTP, ~3000 Ci/mmol (Cat. No. NEG-013H; NEN)
Klenow fragment (Cat. No. 1008412, 2 U/μl; Boehringer Mannheim)
TBE (10×): 0.5 *M* Tris base, 0.5 *M* boric acid, 10 m*M* EDTA
Gel shift buffer (2×): 40 m*M* HEPES (pH 7.4), 100 m*M* KCl, 40% (v/v) glycerol, 2 m*M* 2-mercaptoethanol
Poly(dI-dC): 5 μg/ul (Cat. No. 1219847; Boehringer Mannheim)
Bovine serum albumin fraction V [1%, w/v] (Cat. No. 100030; Boehringer Mannheim)
Acrylamide stock (acrylamide: bisacrylamide, 19:1, v/v)
N,N,N′,N′-Tetramethylethylenediamine (TEMED) (Cat. No. 100139; Boehringer Mannheim)
Ammonium persulfate (10%, w/v) (Cat. No. 100190; Boehringer Mannheim)

Methods

1. To prepare radioactively labeled DNA probe, anneal equimolar quantities of complementary oligonucleotides (full-length forward strand and incomplete reverse strand leaving at least one cytosine to be filled in) in dH_2O by heating to 90°C for 10 min and cooling slowly to room temperature. Fill in by mixing:

Annealed oligonucleotides (200 pmol/μl)	5 μl
Klenow polymerase buffer, 10×	2 μl
dGTP, dTTP, dATP (1.25 m*M* each)	6 μl
[α-^{32}P]dCTP	5 μl
Klenow fragment	2 μl

Incubate at 30°C for 1 hr. Then add:

dNTPs (1.25 m*M* each)	8 μl
Klenow fragment	1 μl

Incubate at 30°C for 15 min.

2. Extract with phenol–chloroform and purify on a 10% (w/v) polyacrylamide gel. Recover the radioactive DNA after exposure of the gel to film and excision of the appropriate position of the gel by electroelution in a dialysis bag or by soaking in elution buffer (0.5 *M* ammonium acetate, 1 m*M* EDTA) for 3 hr or overnight. Ethanol precipitate and resuspend in water at 10,000 cpm/μl.

3. Prepare purified DNA-binding protein or nuclear extract by standard techniques. We routinely use GST-fusion proteins purified by affinity to glutathione–agarose beads (34).

4. Prepare 6% (w/v) polyacrylamide gel in 0.5× TBE the night before and store at 4°C. For a 14 × 14 × 0.15 cm gel:

Acrylamide stock, 40% (w/v)	6 ml
TBE, 10×	2 ml
Water	32 ml
Ammonium persulfate, 10% (w/v)	0.4 ml
TEMED	15 μl

Pour immediately, insert a 12- or 15-well comb, cover with plastic wrap, allow to polymerize at room temperature, and store at 4°C.

5. Prepare the binding reaction by mixing at room temperature in a total volume of 20 μl:

Gel shift buffer, 2×	10 μl
BSA	5 μl
Poly(dI-dC)	0.2–1 μl
DNA probe (10,000 cpm/μl)	1 μl
Purified protein	0–4 μl

6. Incubate on ice for 30 min.

7. Load the samples without running dye (which may interfere with binding) on a precooled gel immersed in cold 0.5× TBE. Use an empty lane to load a small sample of tracking dye such as bromphenol blue in 20% (v/v) glycerol. Electrophorese at 240 V for 90 min at 4°C.

8. Dry on a standard gel dryer and expose to autoradiography film overnight.

Direct Measurement of Protein–DNA Complex Stoichiometry Using Double-Label EMSA Protocol

Many transcription factors bind to DNA as dimers, or form higher order complexes. Careful analysis of dimer formation was required for the study of the PAX6 paired domain. Two forms of the PAX6 paired domain exist owing to alternative mRNA splicing, which can result in a 14-amino acid insertion within the paired domain. We

have demonstrated that this 14-amino acid insertion alters DNA binding and results in an altered consensus binding site as determined by the binding site selection assay described above. Electrophoretic mobility shift assay experiments performed with the consensus binding site selected from the alternatively spliced form of PAX6 demonstrated several shifted complexes, and we sought to determine as directly as possible if the more slowly migrating complexes represented dimers of protein binding to DNA.

It should be noted that a more simple procedure has been described to determine if a protein is capable of dimerization when binding to a specific DNA recognition sequence (39). This procedure involves expressing the protein of interest in two forms of different molecular weights. On binding DNA, the two protein forms will migrate differently during EMSA. If the two protein forms are mixed, and dimers are able to form between the two forms, a heterodimeric complex with an intermediate mobility on will be seen. This technique will indicate whether protein–protein complexes can form, but does not yield a measure of protein–DNA-binding stoichiometry.

Our method relies on the ability to express the fusion protein of interest engineered to contain a cAMP-dependent protein kinase site that can be specifically phosphorylated with radioactive phosphate. The DNA is labeled with a second isotope, and the ratio of the two isotopes in the complex is directly measured after excising it from an EMSA gel (28).

Reagents

pGEX-2TK (Cat. No. 27-4587-01; Pharmacia) bacterial expression vector
[γ-^{32}P]ATP (Cat. No. NEG002Z; NEN)
[γ-^{33}P]ATP (Cat. No. NEG-302H; NEN)
Polynucleotide kinase: 10 U/μl (Cat. No. 174645; Boehringer Mannheim)
Bovine heart kinase (Cat. No. P-2645; Sigma)
Solvable (DuPont-NEN)
Formula 989 (DuPont-NEN) or other aqueous fluor
Elution buffer: 10 m*M* reduced glutathione in 50 m*M* Tris (pH 8.0)
Other materials as per EMSA (above)

Methods

1. Express the protein of interest as a GST fusion protein (using pGEX-2TK) with a cAMP-dependent protein kinase site located between the GST moiety and the DNA-binding protein according to manufacturer directions. Do not elute from glutathione–agarose beads.

2. Radiolabel with [γ-^{32}P]ATP, using 5 U of bovine heart kinase according to the protocol provided with the pGEX-2TK vector (Pharmacia). Wash extensively to remove unincorporated radioactivity, and elute the labeled protein from the beads with elution buffer. The specific activity should be $\sim 2 \times 10^6$ cpm/μg protein.

3. Radiolabel the forward- and reverse-strand oligonucleotides with ^{33}P using [γ-^{33}P]ATP and polynucleotide kinase. Prepare ^{32}P-labeled DNA probe by using [γ-^{32}P]ATP in a separate reaction to be used as a marker of DNA migration on electrophoresis. Anneal and fill in to make double-stranded oligonucleotide probes as described above under EMSA methods. The specific activities should be $\sim 4 \times 10^6$ cpm/μg.

4. Perform the EMSA binding reaction as described above, using ^{32}P-labeled protein and ^{33}P-labeled DNA. Include a sample with no protein, and another with no DNA.

5. Run the samples on an EMSA gel, and include a lane with the ^{32}P-labeled DNA probe alone as a marker of oligonucleotide migration. We found a 4% (w/v) polyacrylamide gel to provide maximal separation of free protein from protein–DNA complex after electrophoresis.

6. Dry the gel and expose it to autoradiography film.

7. The developed film will indicate the location of ^{32}P-labeled protein alone, and as part of a DNA–protein complex. The location of free ^{33}P-labeled DNA probe can be identified by the location of adjacent ^{32}P-labeled DNA marker. Excise the bands corresponding to free ^{33}P-labeled DNA probe, protein–DNA complexes, and free protein from all sample lanes.

8. Rehydrate the excised gel slices in 0.5 ml of water.

9. Elute the protein and DNA from the gel by adding 0.5 ml of Solvable and incubating for 3 hr at 50°C.

10. Neutralize by addition of 10 ml of Formula 989 fluor and perform dual-channel liquid scintillation counting.

11. The final counts provide an internal control for quenching and decay because the specific activity of the protein and DNA can be calculated from the amount of protein and DNA in the reactions and compared to the recovered counts from the protein-alone and DNA-alone lanes. Also, comigrating free protein or DNA counts can be subtracted from complex counts. These values can then be used to determine the absolute stoichiometry of DNA–protein binding for a given complex.

12. Within a single sample lane, the relative ratio of protein to DNA within two distinct shifted complexes can be directly compared. This does not require the calculation of the specific activity of protein or DNA. Thus, we were able to demonstrate in the case of the alternatively spliced form of Pax6 that a more slowly migrating complex contained twice as much protein per unit DNA than did the more rapidly migrating complex.

Other Functions

In the case of Pax proteins, two additional assayable properties that should not be overlooked are the ability of the protein to function as a transcriptional activator

(18, 28), and to localize to the cell nucleus (3, 14, 31). In the case of the *PAX6* gene product, mutations affecting both of these functions have been identified. The former can be assayed by assaying the transcription activation domain as part of a GAL4 fusion (18). Several excellent compilations of nuclear localization signals (NLS) exist (40), and in general, if a missense mutation falls within one of these regions, it is worth testing this property. It should be cautioned that interference with nuclear localization per se does not rigorously prove that the sequence mutated contains an NLS, because the protein may be denatured or rendered unstable. Ordinarily, this proof requires showing that the candidate NLS confers nuclear transport on an unrelated protein that is not ordinarily localized to the nucleus, such as the *E. coli* enzyme β-galactosidase.

Conclusions

In summary, we have outlined an SSCP-based technique for detecting mutations in the human *PAX6* gene. Although the number of *PAX6* exons is relatively large, the technique is highly effective for detecting such mutations, with a sensitivity of approximately 80%. Second, we have summarized some analytical methodologies useful for assaying the molecular consequences of these mutations, including an *in vivo* splicing assay and DNA-binding assays. Such functional experiments provide the opportunity not only to confirm the mutation, but also to use human genetics to gain insight into basic questions of molecular function.

References

1. F. S. Collins, *Nature Genet.* **9,** 347 (1995).
2. C. J. Epstein, *Proc. Nat. Acad. Sci. USA* **92,** 8566 (1995).
3. T. Glaser, D. S. Walton, J. Cai, J. A. Epstein, L. Jepeal, and R. L. Maas, *in* "Molecular Genetics of Ocular Disease" (J. L. Wiggs, ed.), p. 51. Wiley-Liss, New York, 1995.
4. I. Hanson and V. van Heyningen, *Trends Genet.* **11,** 268, 1995.
5. C. Walther, J. L. Guenet, D. Simon, U. Deutsch, B. Jostes, M. D. Goulding, D. Plachov, R. Balling, and P. Gruss, *Genomics* **11,** 424 (1991).
6. A. P. Read, *Nature Genet.* **9,** 333 (1995).
7. T. Strachan and A. P. Read, *Curr. Opin. Genet. Dev.* **4,** 427 (1994).
8. M. Dean, *Nature Genet.* **9,** 103 (1995).
9. E. P. Lessa and G. Applebaum, *Mol. Ecol.* **2,** 119 (1993).
10. M. Orita, Y. Suzuki, T. Sekiya, and K. Hayashi, *Genomics* **5,** 874 (1989).
11. J. Michaud, L. C. Brody, G. Steel, G. Fontaine, L. S. Martin, D. Valle, and G. Mitchell. *Genomics* **13,** 389 (1992).
12. K. Hayashi and D. W. Yandell, *Hum. Mutat.* **2,** 338 (1993).
13. T. Glaser, D. S. Walton, and R. L. Maas, *Nature Genet.* **2,** 232 (1992).

14. C. Carrière, S. Plaza, P. Martin, B. Quatannens, M. Bailly, D. Stehelin, and S. Saule, *Mol. Cell. Biol.* **13,** 7257 (1993).
15. S. Plaza, C. Dozier, N. Turque, and S. Saule, *Mol. Cell. Biol.* **15,** 3344 (1995).
16. R. A. Clark and W. M. Nauseef, *in* "Current Protocols in Immunology" (J. E. Coligan, ed.), p. 723. Greene Publishing Associates, New York, 1991.
17. D. S. Wright and M. M. Manos, *in* "PCR Protocols: A Guide to Methods and Applications" (M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, eds.), p. 153. Academic Press, New York, 1990.
18. T. Glaser, L. Jepeal, J. G. Edwards, S. R. Young, J. Favor, and R. L. Maas, *Nature Genet.* **7,** 463 (1994).
19. M. Anderson and J. S. Gusella, *In Vitro* **20,** 856 (1984).
20. P. Chomczynski and N. Sacchi, *Anal. Biochem.* **162,** 156 (1987).
21. D. R. Engelke, A. Krikos, M. E. Bruck, and D. Ginsburg, *Anal. Biochem.* **191,** 396 (1990).
22. J. Sambrook, E. F. Fritsch, and T. Maniatis, "Molecular Cloning: A Laboratory Manual," Vol. 3, 2nd Ed., p. E.19. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
23. M. B. White, D. Carvalho, D. Derse, S. J. O'Brien, and M. Dean, *Genomics* **12,** 301 (1992).
24. D. Marchuk, M. Drumm, A. Sauklino, and F. S. Collins, *Nucleic Acids Res.* **19,** 1154 (1991).
25. L. G. Mitchell and C. R. Merril, *Anal. Biochem.* **178,** 239 (1989).
26. S. Paabo, D. M. Irwin, and A. C. Wilson, *J. Biol. Chem.* **265,** 4718 (1990).
27. P. S. Walsh, H. A. Erlich, and R. Higuchi, *PCR Methods Appl.* **1,** 241 (1992).
28. J. A. Epstein, T. Glaser, J. Cai, L. Jepeal, D. S. Walton, and R. Maas, *Genes Dev.* **8,** 2022 (1994).
29. D. N. Cooper, L. P. Berg, V. V. Kakkar, and J. Reiss, *Ann. Med.* **26,** 9 (1994).
30. T. Jordan, I. Hanson, D. Zaletayev, S. Hodgson, J. Prosser, A. Seawright, N. Hastie, and V. van Heyningen, *Nature Genet.* **1,** 328 (1992).
31. I. M. Hanson, A. Seawright, K. Hardman, S. Hodgson, D. Zaletayev, G. Fekete, and V. van Heyningen, *Hum. Mol. Genet.* **2,** 915 (1993).
32. A. J. Buckler, D. D. Chang, S. L. Graw, D. Brook, D. A. Haber, P. A. Sharp, and D. E. Housman, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 4005 (1991).
33. W. P. Deng and J. A. Nickoloff, *Anal. Biochem.* **200,** 81 (1992).
34. D. B. Smith and K. S. Johnson, *Gene* **67,** 31 (1988).
35. F. W. Studier, A. H. Rosenberg, I. J. Dunn, and J. W. Dubendorf, *Methods Enzymol.* **185,** 60 (1990).
36. J. A. Epstein, J. Cai, T. Glaser, L. Jepeal, and R. Maas, *J. Biol. Chem.* **269,** 8355 (1994).
37. D. Wilson, G. Sheng, T. Lecuit, N. Dostatni, and C. Desplan, *Genes Dev.* **7,** 2120 (1993).
38. J. Treisman, E. Harris, and C. Desplan, *Genes Dev.* **5,** 594 (1991).
39. J. A. Hope and K. Struhl, *EMBO J.* **6,** 2781 (1987).
40. B. Roberts, *Biochim. Biophys, Acta* **1008,** 263 (1989).

[5] Mutation Detection in the Cystic Fibrosis Transmembrane Conductance Regulator Gene

Ann Harris, Sue Shackleton, and Jeremy Hull

Introduction

The isolation of the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene in 1989 by the groups of L.-C. Tsui in Toronto, Canada and F. Collins in Michigan was the culmination of more than 5 years of work by various research groups to move from linked DNA markers to the locus itself (1–3). Mutations in the *CFTR* gene result in the common severe autosomal recessive disease, cystic fibrosis.

Gene Function

The *CFTR* gene product was named the cystic fibrosis transmembrane conductance regulator (CFTR), on the basis of its predicted structure and function. The 1480-amino acid protein predicted from the cDNA sequence functions as a small cAMP-activated chloride ion channel with a conductance of 6–9 pS (4–8). CFTR would appear to be a member of the ABC-transporter/traffic ATPase superfamily, on the basis of amino acid homology (9). However, to date no active transport of a substrate by CFTR has been established unequivocally. CFTR moves chloride ions down a concentration gradient across specialized epithelia *in vivo*.

Structure of the Gene

The *CFTR* gene is large, spanning about 230 kb of genomic DNA at chromosome 7q31. The locus has been extensively analyzed by chromosome walking and jumping from flanking, linked polymorphic markers on genomic DNA (1), by pulsed-field gel mapping (10–12), and also by isolation of yeast artificial chromosomes (YACs) (13, 14).

Coding Region

The coding region of the *CFTR* gene is made up of 27 exons, numbered 1 to 24, but including 6a and 6b, 14a and 14b, and 17a and 17b (Fig. 1). Most of these are between 50 and 250 bp (see Table I) with the exception of exon 13, which spans 723 bp of genomic DNA.

Methods in Molecular Genetics, Volume 8

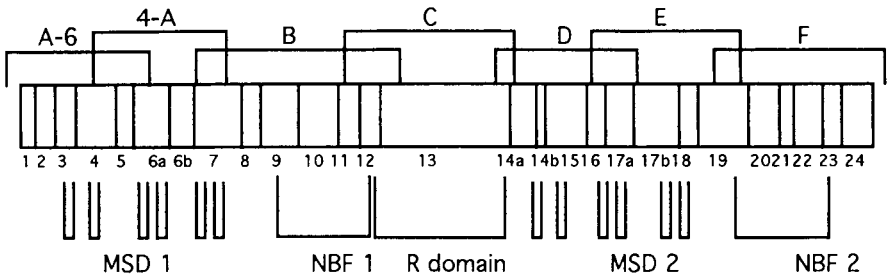


FIG. 1 Linear representation of the *CFTR* cDNA, showing exons, functional domains, and overlapping fragments amplified by reverse transcriptase PCR (RT-PCR). The size of each of the fragments are as follows: A-6, 814 bp; 4-A, 708 bp; B, 898 bp; C, 769 bp; D. 729 bp; E, 699 bp; F, 845 bp.

The CFTR protein consists of two similar halves that each contain a nucleotide-binding fold (NBF) and a membrane-spanning domain, the two halves being separated by the "regulatory" (R) domain. The R domain, encoded largely by a single exon, (exon 13), contains multiple predicted substrate-binding sites for protein kinases A and C, and is a unique feature of CFTR that is not found in the other ABC transporters.

Flanking Regions

The 5′ and 3′ untranslated regions are 134 and 1556 bp, respectively. Sequence analysis of the 1 kb of DNA immediately 5′ to the *CFTR* gene shows that it includes

TABLE I *CFTR* Exon Sizes

Exon	Start	Size (bp)	Exon	Start	Size (bp)
1	121	65	14a	2622	130
2	186	111	14b	2752	37
3	297	109	15	2789	52
4	406	116	16	3041	80
5	622	90	17a	3121	150
6a	712	164	17b	3271	229
6b	876	126	18	3500	99
7	1002	247	19	3599	151
8	1249	93	20	3850	156
9	1342	182	21	4006	90
10	1524	193	22	4096	173
11	1717	95	23	4269	106
12	1812	87	24	4375	198
13	1899	723			

a GC-rich region but does not contain a TATA element. Furthermore, there are several putative regulatory signals, including two consensus GC boxes (GGGCGG) and three potential AP1 sites, that may be sites of regulation by phorbol esters (15). Hence, in certain respects the *CFTR* gene shows certain features of housekeeping genes, that is, genes that are uniformly expressed at similar levels in all cell types. However, the *CFTR* gene is clearly not a housekeeping gene by any functional criteria because its expression is highly tissue specific (see below).

There appear to be at least three transcription initiation sites and three regulatory elements (two positive and one negative) have been described 5′ to the gene (16, 17). However, the major genetic elements controlling temporal and spatial patterns of gene expression have yet to be identified.

Localization of Expression

The *CFTR* gene exhibits a complex mechanism of regulation. The gene shows tight tissue-specific regulation, being expressed mainly in specialized epithelial cells (18–20). In humans, these sites include certain cells of the respiratory system (within the nasal and tracheal epithelium and the serous portion of submucosal glands), the intestine (specifically within crypt cells in the ileum, duodenum, and colon), the pancreas and gallbladder (throughout the ductal epithelium), urogenital tract, and sweat gland.

Mutations

The Cystic Fibrosis Genetic Analysis Consortium collates mutation information on the *CFTR* gene. Members of the consortium are individuals/laboratories worldwide who are making a contribution either to the definition of novel mutations in the gene or to the investigation of the population genetics of different CF mutations. To date more than 500 mutations have been described (CF Genetic Analysis Consortium, unpublished data, May 1995). The vast majority of these have been defined only in one or a few patients. Most types of mutations have been detected within the gene: missense mutations, single base substitutions resulting in alterations of a single amino acid; nonsense mutations leading to "stop" signals; frameshift mutations; splice mutations; and small insertions or deletions. Very few major deletions of large parts of the gene have been described (21), although this may partly reflect the mutation screening methods employed by most research groups, which, being polymerase chain reaction (PCR)-based, might fail to detect an individual who was hemizygous for a particular exon under analysis.

Many of the predicted mutations have been detected only at the DNA level. It is

possible that some of them may be polymorphisms rather than disease-causing mutations. However, an increasing number have been expressed in heterologous systems to establish their deleterious effect on CFTR function [reviewed by Welsh and Smith (22)].

Methods of Detecting Mutations in Cystic Fibrosis Gene

As with mutation detection in all disease-associated genes, a range of techniques is being employed by different laboratories, either analyzing genomic DNA or mRNA. In this chapter we concentrate on those that have been used in our laboratory.

Genomic DNA Analysis

Because the *CFTR* gene encompasses at least 230 kb of genomic DNA, the task of mutation definition is enormous. Clearly it is not feasible to analyze 230 kb of genomic DNA for each mutant gene. Nor is it apparently warranted (except for a few mutant alleles) in a locus that does not frequently exhibit major deletions or structural rearrangements, as are seen in many X-linked genes.

Mutations in the *CFTR* gene have been analyzed primarily by the PCR amplification (23) of exons and 100–200 bp of flanking DNA, followed by use of a method of detecting any mutations within the amplified segment. The most laborious means of detecting mutations is to sequence the whole segment. This is feasible, particularly with an adequate automatic sequencing facility. However, many techniques have been developed to attempt to find, by more rapid means, errors at the level of single base changes within the PCR product. The most frequently used methods of rapid mutation scanning are described in the following sections, although there are many other methods that are not mentioned here.

Single-Stranded Conformational Polymorphism Analysis

Single-stranded conformational polymorphism (SSCP)(24) relies on the slightly different electrophoretic mobilities of two fragments of DNA that are the same size but differ in base composition at one or more sites. The PCR amplification of the fragment of interest is carried out using one or more radioactive nucleotides. This yields a PCR product that can be visualized by autoradiography following denaturation and electrophoresis on an acrylamide gel. Alternatively, nonradioactive DNA fragments of different mobilities can be visualized by silver staining of acrylamide gels. The upper size limit of efficacy of this technique is not exactly certain, although the technique clearly detects more mutations on a fragment of 250 bp than in one of 400 bp, which is probably close to the maximum size that can be analyzed.

Denaturing Gradient Gel Electrophoresis

Denaturing gradient gel electrophoresis (DGGE)(25) exploits the different melting properties of fragments of DNA that have a variant base composition. As for SSCP, this method is able to detect single base changes in PCR products from CF patient DNA. Usually, one of the two PCR primers carries a GC clamp, that is, a GC-rich domain, at its 5′ end to create a high-temperature melting domain adjacent to the region being analyzed. This ensures that the test region is the first melting domain.

Chemical Cleavage of Mismatch

In the chemical cleavage of mismatch (CCM) (26, 27) technique, detection of mutations relies on the formation of a hybrid between control DNA and patient DNA, followed by chemical modification of any single-stranded bases arising from mutations in patient DNA, causing a mismatch in the hybrid. Chemically modified bases (C residues being modified by hydroxylamine hydrochloride and T residues by osmium tetroxide) are then cleaved by piperidine and fragments separated by acrylamide gel electrophoresis. The technique involves end labeling a PCR product from control and/or patient DNAs in order to be able to visualize the products of cleavage. The upper size limit on this technique, under standard conditions, is probably in the region of 1.5–2 kb.

In all cases in which single base changes or other mutations have been detected in *CFTR* by one of the three methods described above, they are subsequently confirmed by sequence analysis. In this respect the chemical mismatch method has a major advantage over SSCP and DGGE because it provides accurate information on the location of the mutation, rather than necessitating sequence analysis of the whole PCR product under test.

cDNA or mRNA Analysis

Analysis of CFTR mRNA to Detect Mutations

The *CFTR* mRNA is 6.2 kb and hence its use in mutation screening is advantageous in comparison to scanning the 27 individual exons of the genomic DNA. Analysis of cDNA by reverse transcription-based PCR (RT-PCR) can rapidly reveal the existence of splicing mutations and some stop mutations that result in exon skipping. In addition, sections of the *CFTR* cDNA generated by RT-PCR can be analyzed by the techniques discussed above for rapid mutation detection. We have generated data on the analysis of *CFTR* segments by CCM (see below).

One of the problems in using *CFTR* mRNA as a resource for mutation detection is the lack of availability of cells that express significant amounts of CFTR, as discussed above. However, the gene is transcribed at a readily detectable level in nasal epithelial cells and because these are frequently obtainable by nasal brushing, several

groups have defined mutations in the *CFTR* mRNA in these cells. Clearly the success of this approach is dependent on the relative abundance and stability of mutant *CFTR* mRNA, particularly in the case of compound heterozygotes, in which the presence of a more abundant or more stable transcript from one chromosome might mask the mutant transcript.

Materials and Methods

Isolation of Cells

Nasal Epithelial Cells

Method of Collection

Brushed human nasal epithelial (BHNE) cells are collected using a bronchoscopy cytology brush (Olympus Optical Co. Ltd., Tokyo, Japan), modified so that it is shortened to 6 cm and mounted on a plastic handle. The inferior turbinate is visualized using an auroscope with the magnifying lens removed. Three to five minutes after application of local anesthetic the cytology brush is passed into the nose and rubbed against the inferior turbinate for a few seconds under direct vision. The brush is then removed and attached cells are transferred to a universal container containing Dulbecco's modified Eagle's medium (DMEM). The cells are transported on ice in DMEM and are processed within 4 hr of collection.

Peripheral Blood Lymphocytes

Another potential resource for analyzing *CFTR* mRNA is peripheral blood lymphocytes. Although CFTR is not expressed in these cells at physiological levels, by exploiting the phenomenon of "ectopic" transcription of genes in lymphocytes, *CFTR* mRNA can be detected. This is achieved by nested PCR, using two sets of PCR primers, the second set being located within the first, to amplify *CFTR* cDNA that has been reverse transcribed from total RNA. This phenomenon has been exploited successfully for several X-linked genes (28–30) and *CFTR* (31). However, in the case of an autosomal locus such as the *CFTR* gene, the potential problems of mutant mRNA abundance and stability in compound heterozygotes are even more severe than in the case of nasal epithelial cells. This is due in part to the numerous cycles (30 cycles with each nested set) of PCR required to obtain adequate amounts of DNA for analysis.

Lymphocyte Preparation

Lymphocytes are prepared from whole blood for subsequent RNA extraction. Fifteen milliliters of fresh whole blood is diluted with an equal volume of phosphate-buffered saline (PBS). The diluted blood is then layered onto 15 ml of Histopaque (relative

density 1.077; Sigma, St. Louis, MO) and then centrifuged at 1700 rpm for 30 min at 4°C. The buffy layer of lymphocytes is aspirated into a clean tube and diluted with 2 vol of PBS. This cell suspension is then centrifuged at 1500 rpm for 15 min. The supernatant is removed and the cell pellet washed two further times in 10 ml of cold PBS. The cells are either used immediately or stored in liquid nitrogen as a cell pellet.

Analysis of Mutations

Genomic DNA

Isolation of Genomic DNA from Lymphocytes

Blood samples (5–10 ml) are mailed to the laboratory in EDTA and frozen at −20°C until processed. Genomic DNA is prepared by standard methods (32).

RNA

RNA is extracted either directly from BHNE cells or lymphocytes.

Extraction with Guanidinium Thiocyanate and Cesium Chloride Centrifugation

The cells are washed twice in ice-cold PBS and pelleted by centrifugation at 1500 rpm for 5 min. RNA is prepared according to the method of Chirgwin *et al.* (33). Briefly, the cells are lysed by adding 1–2 ml of 4 *M* guanidinium thiocyanate solution [4 *M* guanidinium thiocyanate, 0.05 *M* Tris-HCl (pH 7.6), 0.01 *M* EDTA, 2% (w/v) Sarkosyl, and 0.14 *M* 2-mercaptoethanol]. The homogenate is then layered onto a 1.6-ml cushion of 6.7 *M* CsCl in 0.1 *M* EDTA, pH 8.0 (relative density 1.7) in a Beckman (Fullerton, CA) SW50.1 polyallomer tube that has been previously treated with 30% (w/v) hydrogen peroxide for 15 min and then rinsed in diethyl pyrocarbonate (DEPC)-treated water. The tubes are then spun at 30,000 rpm for 18 hr at 20°C. After centrifugation the supernatant is removed and the RNA is resuspended in 400 μl of T0.1E (10 m*M* Tris, 0·1 m*M* EDTA.

Preparation of Poly (A)$^+$ RNA by Use of QuickPrep Micro mRNA Purification Kit (Pharmacia)

The QuickPrep Micro mRNA purification kit (Pharmacia, Piscataway, NJ) allows extraction of polyadenylated RNA direct from the cells. The cells are lysed in a solution containing guanidinium thiocyanate. Oligo(dT)-cellulose is then added. In the high salt concentration of the solution, poly(A)$^+$ RNA binds to the cellulose. The bound poly(A)$^+$ RNA is then washed several times to remove contaminants, using spin columns. The poly(A)$^+$ RNA is finally eluted with DEPC-treated water or T0.1E. The protocol provided by Pharmacia is followed without modification.

Amplification of DNA and RNA

Oligonucleotide Synthesis Oligonucleotides are synthesized on a 392 DNA/RNA synthesizer (Applied Biosystems, Foster City, CA). After synthesis the oligonucleotide is deprotected by incubating it in a solution of 35% (v/v) ammonium hydroxide at 55°C overnight. The oligonucleotide is then dried in a heated vacuum centrifuge and resuspended in 200 ml of T0.1E. Yield is estimated by spectrophotometry. The approximate concentration of nucleic acid is calculated by determining the optical density (OD) at 260 nm.

Polymerase Chain Reaction The PCR (23) is used to amplify DNA using human genomic DNA or cDNA as starting material. The general protocol used is as follows. One microliter containing 50–200 ng of the DNA to be amplified is added to a PCR premix containing 5 μl of each 5′ and 3′ oligonucleotide primer (100 ng/μl), 5μl of 10× PCR buffer [670 m*M* Tris-HCl (pH 8.8), 166 m*M* $(NH_4)_2SO_4$, 67 m*M* $MgCl_2$], 5 μl of a dNTP solution (5m*M* each), 1.7 μl of 0.5% (w/v) bovine serum albumin, 0.7 μl of 5% (w/v) 2-mercaptoethanol, 0.6 μl of *Taq* DNA polymerase, and made up to 50 μl with T0.1E. The premix is exposed to ultraviolet (UV) light for 5 min before adding the *Taq* DNA polymerase. The reaction mix is covered with one drop of mineral oil to minimize evaporation. After the reactions are complete, 5 μl of the sample is electrophoresed on an agarose gel to check the size and specificity of the product. A sample in which no DNA is present is used as a control for contamination.

Reverse Transcription-Based Polymerase Chain Reaction cDNA synthesis: For RT-PCR (29,31), 30–50 ng of poly(a)$^+$ RNA or 1 μg of total RNA, resuspended in 5 μl of DEPC-treated water, is used for each reaction. "Premix 1" (4.5 μl), containing 0.5 μl of a 3′ oligonucleotide primer (100 ng/μl) and 4 μl of T0.1E, is added to the RNA, covered with mineral oil, and incubated at 65°C for 10 min. The sample is transferred to ice and 10.5 μl of "premix 2" containing 4 μl of 5× RT buffer [250 m*M* Tris-HCl (pH 8.5), 375 m*M* KCl, 50 m*M* dithiothreitol (DTT), 15 m*M* $MgCl_2$], 5 μl of a solution of 5 m*M* dNTPs, 0.5 μl of RNase inhibitor (human placental; Boehringer Mannheim, Indianapolis, IN), 0.5 μl of distilled H_2O, and 1 μl of reverse transcriptase [Maloney murine leukemia virus (MMLV), 200 U/μl; GIBCO, Grande Island, NY] is added below the mineral oil and the mixture is incubated at 42°C for 1 hr. Controls in which either no RNA or no reverse transcriptase is added are included to indicate the presence of contamination.

Amplification (First-round PCR): The cDNA is amplified by PCR as described above, except that the entire cDNA sample (a volume of 20 μl) and only 4.5 μl of 3′ primer are used. The total volume of the reaction mix is 50 μl, as for the standard PCR. Thirty cycles of PCR are performed.

Amplification [Second-round (nested) PCR]: To increase the sensitivity of the RT-PCR technique, a second round of PCR is performed. One microliter of the first-

TABLE II Oligonucleotides and Conditions Used for Polymerase Chain Reaction Amplification and Methods Used for Mutation Detection in Genomic DNA Fragments[a]

Exon	5′ Primer	3′ Primer	Annealing temperature (°C)	Extension time (min)	Product size (bp)	Method of analysis
1	α1R	10D	63	0.8	220	SSCP
2	2bPF	2bPR	60	2.5	330	SSCP
3	3i-5	3i-3	60	1.5	309	SSCP
4	4i-5	4i-3	60	3	438	SSCP
5	5i-5	5i-3	62	2	395	SSCP
6a	6ai-5	6ai-3	62	3	385	SSCP
6b	6bi-5	6bi-3	62	3	417	SSCP
7	7i-5	7i-3	60	3	410	SSCP
8	I8D1	I8R1	62	2	328	SSCP
9	9i-5s	9ii-3	62	1	199	SSCP
10	10i-5	10i-3	60	3	491	CCM
11	11i-5	11i-3	60	3	425	SSCP
12	12i-5	12i-3	62	3	426	SSCP
13	Ex13 5′	Ex13 3′	60	5	714	CCM
14a	14aii-5	14aii-3	62	1	276	SSCP
14b	14bi-5	14bii-3	62	2.5	386	SSCP
15	15i-5	15i-3	62	3	485	CCM
16	16ii-5	16i-3	64	3	409	CCM
17a	17aii-5	17aii-3	57	1	283	SSCP
17b	17bi-5	17bi-3	62	3	463	CCM
18	18ii-5	18i-3	60	1	304	SSCP
19	19i-5	E2L	62	1	309	SSCP
	F1R	19i-3	62	1	204	SSCP
20	20i-5	20i-3	62	3	473	CCM
21	21i-5	2li-3	62	3	477	CCM
22/23	524F	23i-3	60	7	1000	CCM
24	24i-5	G1L	62	7	997	CCM

[a] See Refs. 33a and 46 for descriptions of primers.

round PCR reaction is added to 49 μl of a fresh PCR mix containing a nested set of 5′ and 3′ primers located inside the first-round primers. The composition of the PCR mix is as described above and a further 30 cycles of PCR are carried out [see Table II (33a)].

Cloning of PCR products: Generally, if an abnormality is detected by SSCP or CCM, mutation analysis is carried out by direct sequence analysis of PCR products of genomic DNA amplification. In cases where it is necessary to separate the PCR products derived from the two *CFTR* alleles of a compound heterozygote patient or CF carrier, this is achieved by cloning the PCR products into plasmid vectors. Similarly, most RT-PCR products are cloned prior to sequence analysis. DNA fragments

are separated by agarose gel electrophoresis. The desired bands are then cut out from the gel and isolated using a Geneclean II kit (Bio 101, La Jolla, CA) according to the manufacturer instructions. The PCR products are either cloned into pBluescript (Stratagene, La Jolla, CA) by standard methods or more commonly into the plasmid vector (pCR) by TA cloning (Invitrogen, San Diego, CA), employing the recommended methods.

Detection of Mutations by Direct Analysis of cDNA

Splice Site Mutations

In an effort to detect mutations by direct analysis of cDNA (34–36), we investigated three putative splicing mutations, 621+1G→T (37) and 1898+1G→A (38) (which disrupt the 5′ splice donor sites of introns 4 and 12, respectively) and 1717−1G→A (39) (which disrupts the 3′ splice acceptor site of intron 10) by analyzing mRNA from nasal epithelial cells harvested from patients with cystic fibrosis. All three mutations do, as predicted, result in the production of aberrantly spliced *CFTR* mRNA.

In the cDNA amplified from patients with the 621+1G→T mutation (Figs. 2–4),

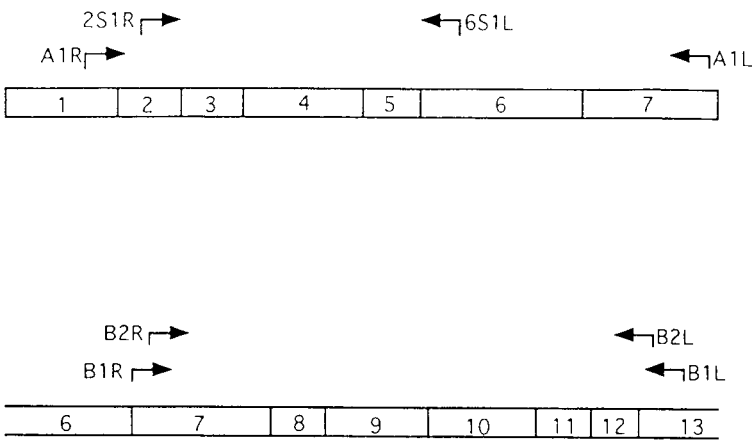


FIG. 2 Positions of primers used in cDNA synthesis and PCR reactions for the analysis of splicing mutations in exons 4, 11, and 12. The numbered boxes represent exons in the *CFTR* gene. For analysis of the 621+1G→T mutation, cDNA was synthesized from the A1L primer, the first-round PCR was carried out with A1L and A1R, and the second-round PCR was performed using the 2S1R and 6S1L primers (34,35). For analysis of the 1717−1G→A and 1898+1G→A mutations, cDNA was synthesized from the B1L primer, the first-round PCR was carried out with the B1R and B1L primers, and the second-round PCR was performed using the B2R and B2L primers. No RNA and no reverse transcriptase controls were used in all reactions.

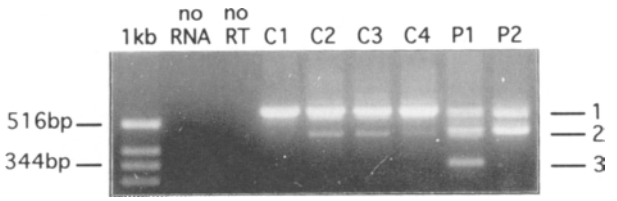


FIG. 3 A 2% (w/v) agarose gel showing PCR products from the 2S1R–6S1L second-round reaction. C1 to C4 are control subjects. P1 and P2 are the two patients with the 621+1G→T mutation. The faint bands below bands 1 and 2 are likely to be heteroduplexes.

a proportion of the PCR product lacked the last 93 bp of exon 4, the result of activation of an alternative splice site at this position in exon 4. One patient with the 621+1G→T mutation also had a PCR product that lacked all of exon 4. The reason for the same mutation resulting in different effects on mRNA splicing in these two patients is not clear, although sequence variation within intron 4 may be relevant (40). When cDNA was amplified from the patients with the 1717−1G→A mutation, a proportion of the PCR product lacked exon 11 (Figs. 2 and 5). Transcripts lacking exon 11 will have an altered reading frame, resulting in a stop at codon 566. For the

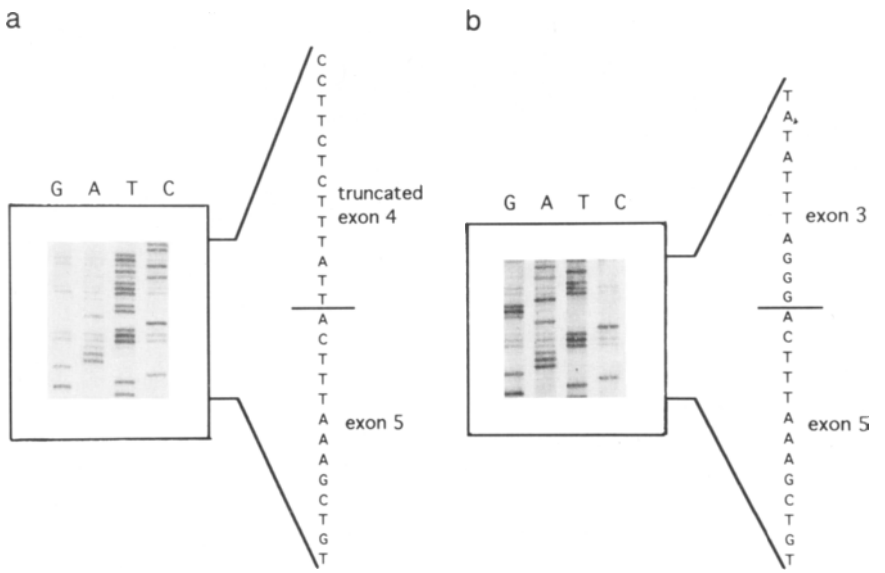


FIG. 4 (a) Sequence analysis of band 2 in one of the patients with the 621+1G→T mutation, using the 6S1L primer. The sequence reads from exon 5 to a site 93 bp within exon 4. (b) Sequence analysis of band 3 in patient P1 with the 621+1G→T mutation, using the 6S1L primer. The sequence reads from exon 5 to exon 3.

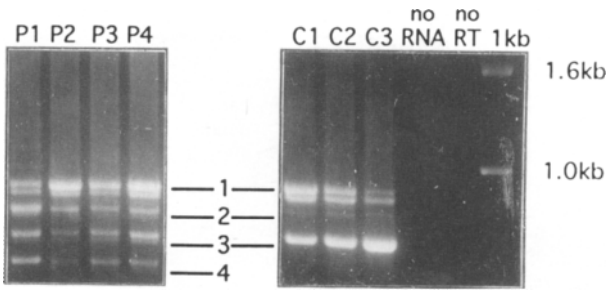


FIG. 5 A 2% (w/v) agarose gel showing PCR products from the B2R–B2L second-round reaction. P1 to P3 are compound heterozygotes for the 1717−1G→A mutation. P4 has the 1898+1G→A mutation. C1 to C3 are control subjects. Band 1 is the full-length fragment B (see Fig. 1) containing exons 7–12, and in P1–P4 is largely derived from the *CFTR* gene that does not carry the splicing mutation. Band 2 lacks exon 11 in P1–P3 and lacks exon 12 in P4. Band 3 in patients and control lacks exon 9 [a common *CFTR* splicing variant (52)]. Band 4 lacks exons 9 and 12.

patient with the 1898+1G→A mutation a proportion of the PCR product was found to lack exon 12 (Figs. 2 and 5). Loss of exon 12 resulting from the 1898+1G→A mutation has also been demonstrated by Strong *et al.* (41).

Stop Mutations

The stop mutation R553X, caused by a C→T substitution at base 1789 in exon 11 of the *CFTR* gene (42), is the fourth most common mutation worldwide, with a frequency of just over 1% in the CF population (43). Stop mutations in other genes have been shown to result in either decreased or undetectable levels of mRNA, skipping of the exon containing the stop mutation, or the production of truncated protein. We have investigated the effects of the R553X mutation on mRNA processing by studying mRNA extracted from nasal epithelial cells harvested from four CF patients who are compound heterozygotes for this mutation.

When cDNA from the four patients carrying the R553X mutation was amplified using the 9S1R and B2L primers (Fig. 2), three PCR products were identified (bands 1, 2, and 3; see Fig. 6). Restriction enzyme digests and direct sequence analysis of the largest band (band 1) showed that this was the full-length 504-bp fragment (B9) predicted from the position of the primers. Band 2a contained part of exon 9, exons 10 and 12, and part of exon 13, as expected, but lacked exon 11 (Fig. 7). Band 3 contained part of exon 9, exon 10, and part of exon 13, as expected, but lacked exons 11 and 12. The same primers were used to amplify cDNA from the 10 control subjects. The full-length fragment B9 was identified in all subjects. In eight subjects, variable amounts of an additional PCR product, band 2b, were also seen. This product was slightly larger than band 2a seen in the R553X patients and subsequent analysis showed that it contained part of exon 9, exons 10 and 11, and part of exon 13, as

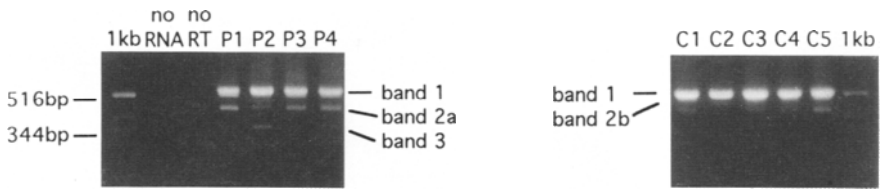


FIG. 6 A 2% (w/v) agarose gel showing PCR products from the 9S1R–B2L second-round reaction. The 9S1R primer is located within exon 9 (34), giving a normal PCR product of 522 bp. *Left:* Patients with the R553X mutation (P1 to P4). Band 1 is the full-length fragment B9. Band 2a lacks exon 11, and band 3 lacks exons 11 and 12. *Right:* Five control subjects (C1 to C5). Band 1 is the full-length fragment B9 and band 2b lacks exon 12.

expected, but lacked exon 12. Skipping of exon 12 has been shown previously in normal subjects (44).

Detection of Mutations by Analysis of Genomic DNA

Single-Stranded Conformational Polymorphism Analysis

For PCR of exons to be analyzed by SSCP (24), 0.1 μl of [α-^{32}P]dCTP (3000 Ci/mmol) is added to the reaction mixture and the concentration of unlabeled dCTP

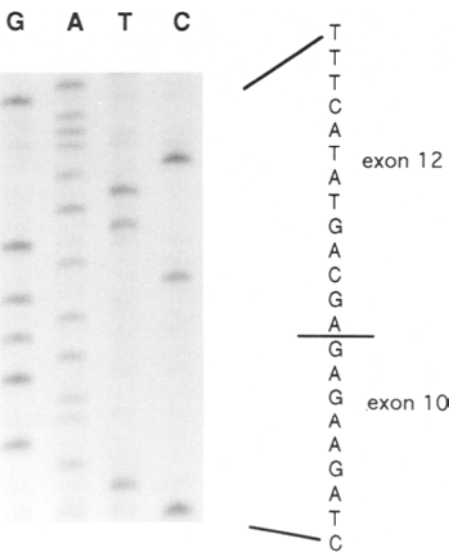


FIG. 7 Direct double-stranded sequence analysis of PCR products. Sequencing primers used were O41E and O43E. Shown here is the sequence analysis, using the O43E primer, of band 2a in one of the R553X patients. The sequence reads from exon 10 to exon 12.

reduced to 125 μM, in a total volume of 25 μl. The labeled PCR product is mixed with an equal volume of stop buffer [95% (v/v) formamide, 10 mM NaOH, 0.05% (w/v) xylene cyanol, and 0.05% (w/v) bromphenol blue]. Samples are then boiled for 5 min and placed on ice prior to loading on a 5% (w/v) acrylamide (99:1, acrylamide–bisacrylamide)–5% (v/v) glycerol gel. Gels are run in Tris–borate–EDTA (TBE) buffer at 8 W for 20–24hr at 4°C. The gels are then fixed in 10% (v/v) methanol–10% (v/v) acetic acid, dried, and exposed to Kodak (Rochester, NY) XAR film for 24–48 hr at room temperature. We have detected a number of novel mutations by SSCP, including G27X (45) and 182delT (46) shown in Fig. 8 (see also Table III).

Chemical Cleavage of Mismatch

In CCM (26, 27), the fragment of DNA of interest is amplified by PCR and subsequently isolated using the Geneclean II kit (Stratech Scientific Ltd.) according to manufacturer instructions.

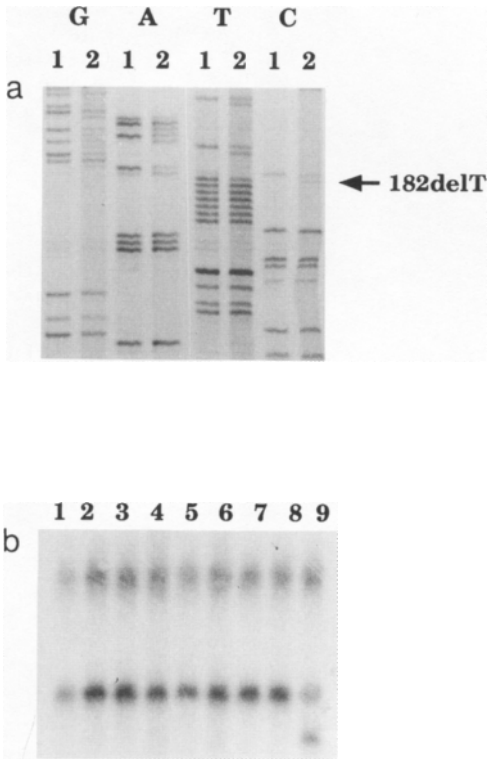


FIG. 8 A novel mutation in exon 1 was detected by SSCP in a patient [lane 9 in (B)] of UK origin. The change is the deletion of a T residue at base 182 (182delT) (46) in (A).

TABLE III Mutations Detected on Genomic DNA by Single-Stranded Conformational Polymorphism or Chemical Cleavage of Mismatch[a]

Exon	SSCP/CCM analysis	Mutations/ polymorphisms detected	Frequency
1	SSCP	182delT[b]	1
2	SSCP	G27X[b]	1
3	SSCP/CCM	G85E	4
		394delTT	1
4	SSCP/CCM	E92K	1
		R117C	2
		Q151X[b]	1
5	SSCP	G178R	2
		711+1G→T	1
6a	SSCP	Q220X[b]	1
6b	SSCP	0	0
7	SSCP/CCM	1078delT	1
		Ins1154TC	2
		R347H	1
		R352Q	1
8	SSCP	0	0
9	SSCP	0	0
10	CCM	Q525X[b]	1
11	SSCP	1717−1G→A	3
		R560T	2
12	SSCP	1898+1G→A	3
13	CCM	2184delA	2
14a	SSCP	E831X	1
		W846X1	1
14b	SSCP	2789+5G→A	2
15	CCM	0	0
16	CCM	3041delG[b]	1
17a	SSCP	0	0
17b	CCM	W1098R	1
18	SSCP	0	0
19	SSCP	R1158X	1
		R1162X	2
		3695delC	5
20	CCM	0	0
21	CCM	N1303K	4
22/23	CCM	4271delC[b]	1
24	CCM	0	0
			50

[a] From Ref. 46.
[b] A novel mutation.

Labeling Control DNA

Control DNA (100 ng) is added to a premix containing 1 μl of kinase buffer [0.5 *M* Tris-HCl (pH 7.6), 0.1 *M* $MgCl_2$], 0.5 μl of [γ-32P]dATP (185 TBq/mmol), and 0.5 μl of T4 polynucleotide kinase. The total volume is made up to 10 μl with T0.1E and the sample is then incubated at 37°C for 30 min. In some cases it is necessary to label the patient DNA instead of or in addition to the control DNA in order to detect all mismatches (47).

Heteroduplex Formation

One microliter of solution containing the labeled control DNA is added to 200 ng of test DNA in a premix containing 1 μl of 10× hybridization buffer [1 *M* Tris-HCl (pH 8.0), 3 *M* NaCl] and made up to a total volume of 10 μl. This gives a final ratio of control DNA to test DNA of 1 : 20. The sample is covered with one drop of mineral oil, and after boiling for 5 min is incubated overnight at 65°C.

Modification Reactions

The hybridized sample is removed from under the mineral oil and transfered to a 1.5-ml siliconized Eppendorf tube. The DNA is precipitated on dry ice by the addition of 4 μl of glycogen and 800 μl of a solution containing 1 part 0.3 *M* sodium acetate–0.1 *M* EDTA to 4 parts 100% ethanol. After 10 min the DNA is pelleted by centrifugation and washed with 70% (v/v) ethanol. The sample is then resuspended in 14 μl of T0.1E and then divided into two 7-μl volumes.

Hydroxylamine Modification: Modification of Mismatched Cytosine A fresh 4 *M* solution of hydroxylamine hydrochloride (Aldrich, Milwaukee, WI) is made in water and titrated to pH 6 by the addition of 0.3 vol of diethylamine (Aldrich). Twenty microliters of this solution is incubated with one of the 7-μl volumes containing the heteroduplex DNA at 37°C for 2 hr.

Osmium Tetroxide Modification: Modification of Mismatched Thymidine The other 7-μl volume of heteroduplex DNA is incubated with a premix containing 5 m*M* Tris-HCl (pH 8), 0.5 m*M* EDTA, 3% (v/v) pyridine, and 0.025% (w/v) osmium tetroxide (Aldrich) in a total volume of 25 μl, at 37°C for 2 hr.

Piperidine Cleavage

After modification, the DNA is ethanol precipitated on dry ice as before and washed with 70% (v/v) ethanol. The pellet is resuspended in 50 μl of a fresh 1 *M* solution of piperidine (Aldrich) by vortexing and incubated at 90°C for 30 min. The samples are then ethanol precipitated on dry ice and washed with 70% (v/v) ethanol as before.

Electrophoreis and Autoradiography

The samples are resuspended in 3 μl of T0.1E, 2 μl of formamide dyes is added, and the samples are then loaded onto 6% (w/v) denaturing acrylamide gels. Size markers,

generated by end labeling the products of an *Msp*I digest of the pATX plasmid, are also loaded onto the gel. After electrophoresis (1.5–2 hr), the gel is fixed in a solution of 10% (v/v) methanol–10% (v/v) acetic acid. The toxic nature of many of the chemicals used in CCM makes drying the gel impractical, so that after fixing, the gel and plate are covered in Saran Wrap and exposed overnight to preflashed XAR film (Kodak).

Chemical Mismatch on Genomic DNA Fragments

A number of novel mutations in the *CFTR* gene have been detected by CCM, as is shown in Figs. 9–12. These include the Q525X, 3041delG, and 4271delC mutations (46).

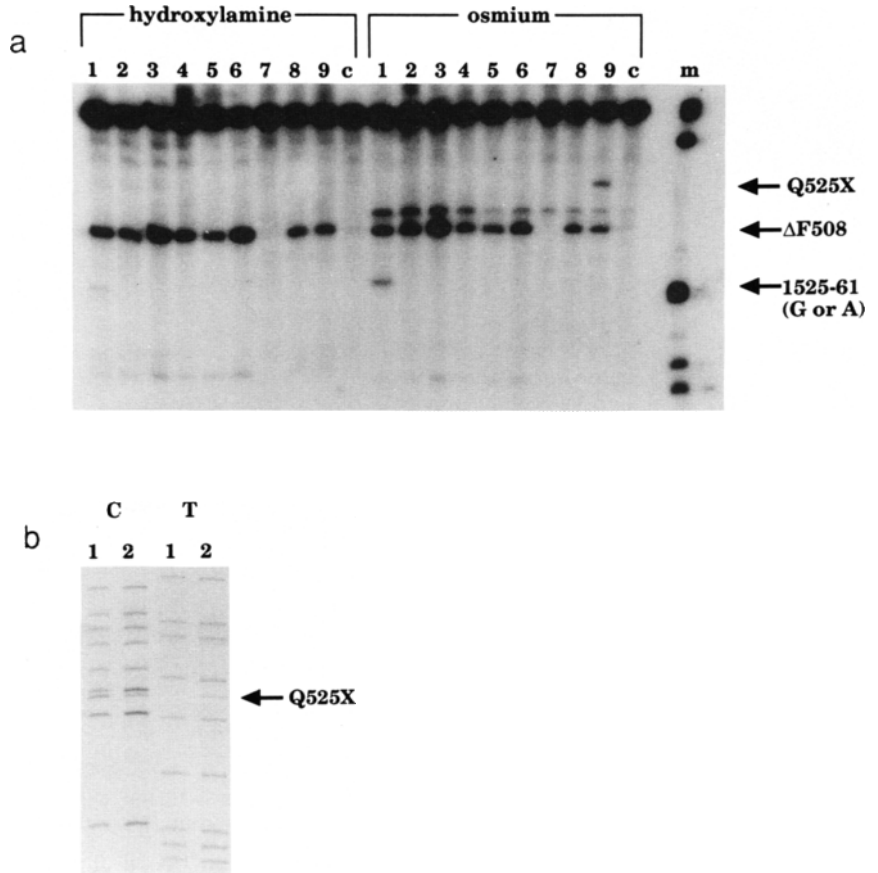


FIG. 9 (a) In exon 10 amplified from genomic DNA a novel mutation was detected by CCM in one patient with classic CF who also carries the F508 deletion. Sequence analysis (b) revealed this mutation to be a C-T substitution at base 1705, resulting in a "stop" mutation at Glu-525 (Q525X) (46).

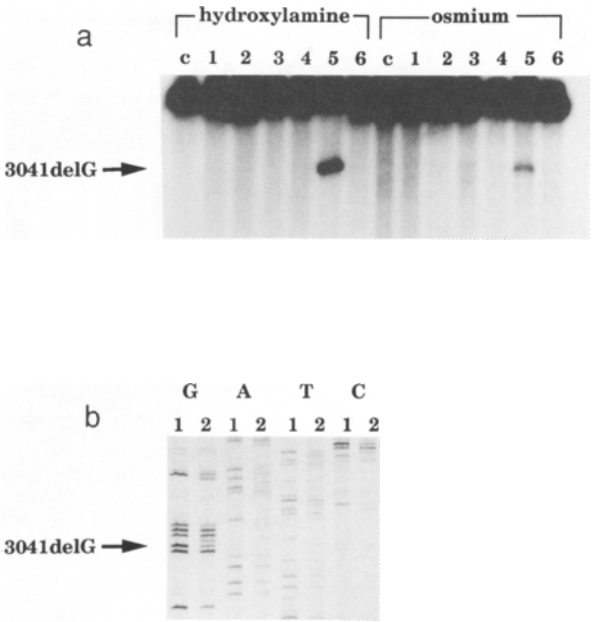


FIG. 10 (a) Within exon 16 of *CFTR* a novel frameshift mutation was detected by chemical mismatch analysis in the maternal *CFTR* gene of a British patient carrying the G551D mutation in his paternal *CFTR* gene. (b) This mutation is the deletion of base 3041 (3041delG), the first base in exon 16 (46).

Chemical Mismatch on cDNA Fragments

We have employed the chemical cleavage method to detect sequence changes in *CFTR* cDNA amplified from nasal epithelial cells of CF patients. We have detected 14 sequence changes in the *CFTR* gene using this technique [see Fig. 12 and Table IV (3, 37, 42, 45–47a–c)].

Sequencing and Mutation Detection

Dideoxy Sequencing of Plasmid DNA and PCR Products The dideoxy sequencing method of Sanger *et al.* (48) [modified by Green *et al.* (49)] with Sequenase version 2.0 is used for direct sequencing of PCR products and plasmids.

Sequencing reactions: The PCR product (100–200 ng) or 500 ng of plasmid DNA, resuspended in 3 μl of distilled H_2O, is used for each sequencing reaction. Two solutions are made, each sufficient for eight sequencing reactions. The primer premix solution contains 5 μl of sequencing primer (100 ng/μl), 18 μl of sequencing buffer [0.2 *M* Tris-HCl (pH 7.5), 0.1 *M* $MgCl_2$, 0.25 *M* NaCl], 9 μl of T0.1E, and 6 μl of dimethyl sulfoxide (DMSO). The enzyme/label premix contains 8 μl of 0.1 *M* DTT, 3 μl of DMSO, 19 μl of T0.1E, 4 μl of [α-^{35}S]dATP (185 TBq/mmol), and 4 μl of Sequenase.

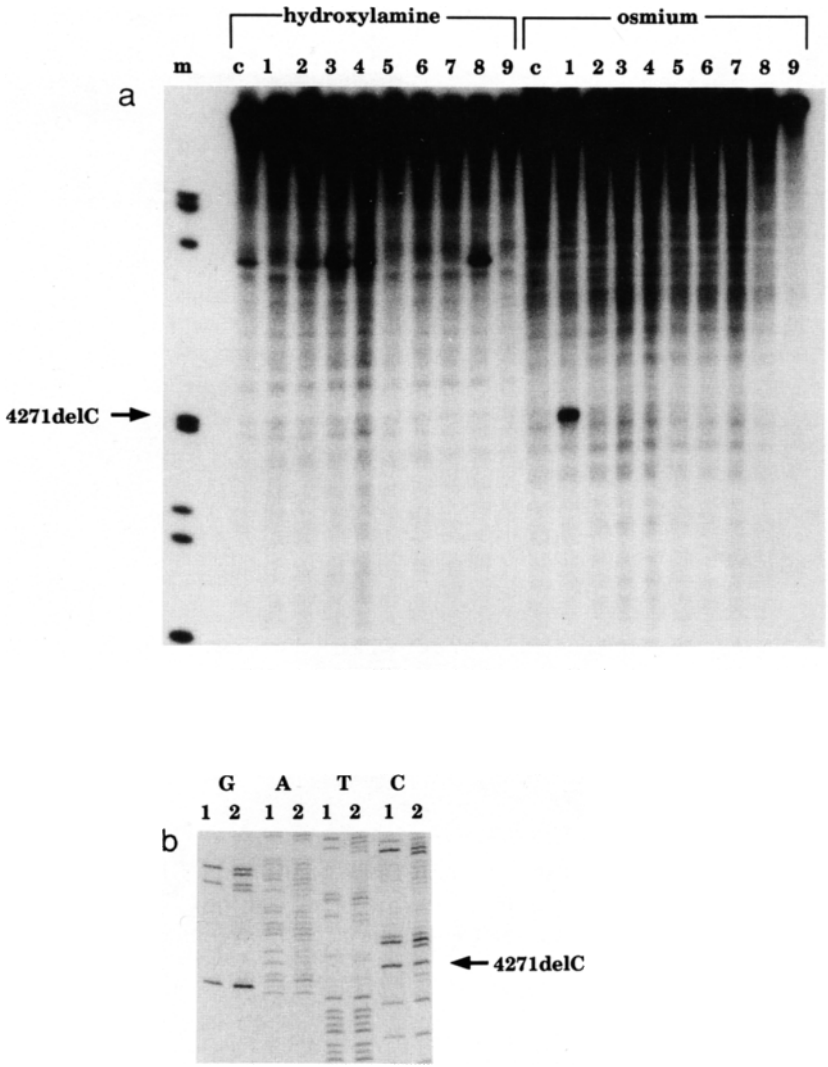


FIG. 11 (a) Chemical mismatch analysis of exons 22 and 23 in one genomic DNA fragment revealed another novel frameshift mutation. (b) 4271delC lies within exon 23 and was found in one CF patient (46).

The DNA is boiled for 3 min with 4 μl of the primer premix solution and then immediatedly chilled on ice. Four microliters of the enzyme/label premix is then added. The solution is mixed, 2 μl is added to 2 μl of each of the four termination mixes containing the relevant terminating [dideoxy (dd)NTP] nucleotide. Each termination mix contains 180 μl of T0.1E, 80 μl each of 0.5 m*M* solutions of dCTP,

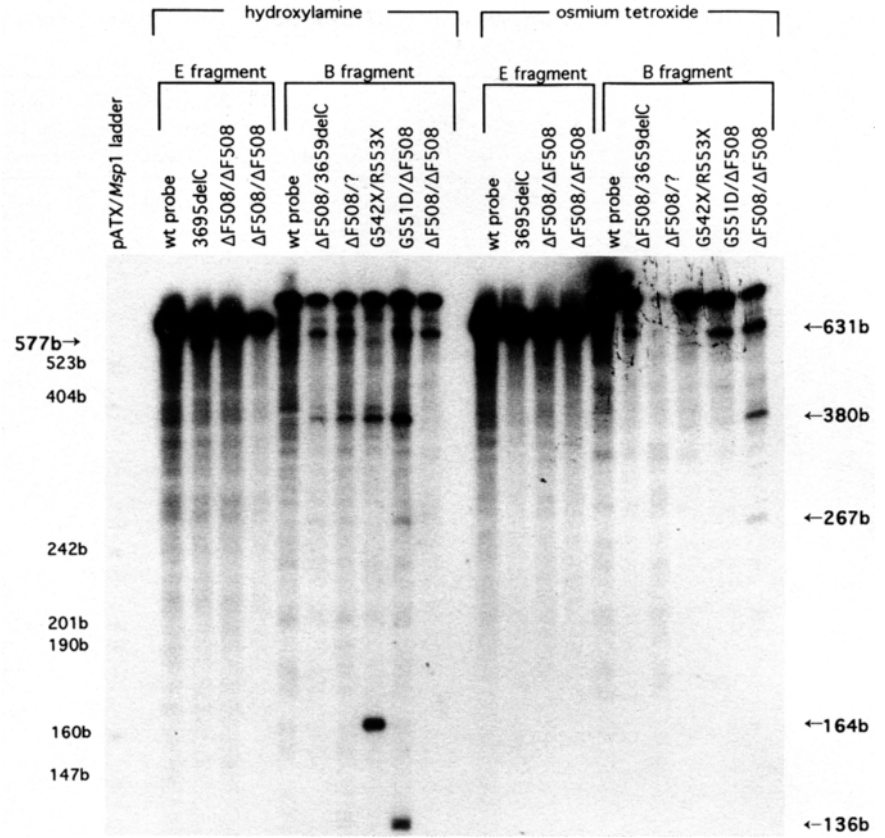


FIG. 12 Detection of mutations by CCM analysis of *CFTR* cDNA fragments (see Fig. 1). Autoradiograph showing cleavage products for the 3659delC mutation (577 bases) in the E fragment, F508 (634+267 bases), G542X (164 bases), and G551D (136 bases) mutations. The band at 380 bases is the cleavage product resulting from the 1540A→G polymorphism present in the control probe. The absence of this band in the control and the F507/F508 lanes suggests that these two subjects were homozygous for this common polymorphism. The pATX ladder was generated by end labeling the products of an *Msp*I digest of the pATX plasmid. Full-length E fragment, 699 bases; full-length B fragment, 898 bases. b, Bases; wt, wild type.

dGTP, and dTTP, and 80 μl of a 0.05 m*M* solution of the appropriate ddNTP, apart from the adenosine termination mix, which has only 0.8 μl of ddATP and a correspondingly larger amount of T0.1E. After incubation of the sequencing reactions at 37°C for 5 min, 2 μl of a chase solution (composed of 0.75 μl of each 100 m*M* dNTP, 30 μl of DMSO, and 267 μl of distilled H_2O) is added to each reaction. After a further 5 min at 37°C, the reactions are stopped by the addition of 2 μl of formam-

TABLE IV Position and Nature of Sequence Changes Used as Positive Controls for the Chemical Cleavage Experiments[a]

Mutation	Nucleotide change	Exon and fragment	Modification chemical	Predicted size of labeled product (bases)	Ref.
182delT	Deletion T at 182	Exon 1, fragment A-6	Hydroxylamine (N) Osmium (Y)	26	46
G27X	G→T at 211	Exon 2, fragment A-6	Hydroxylamine (Y) Osmium (N)	759	45
G85E[b]	G→A at 386	Exon 3, fragment A-6	Hydroxylamine (Y) Osmium (N)	584	37
E92K[b]	G→A at 406	Exon 4, fragment A-6	Hydroxylamine (Y) Osmium (N)	564	47a
Q220X[b]	C→T at 790	Exon 6a, fragment 4-A	Hydroxylamine (Y) Osmium (N)	304	46
ΔF508	Deletion of CTT, 1652–1655	Exon 10, fragment B	Hydroxylamine (Y) Osmium (Y)	631+267 631+267	3
G542X[b]	G→T at 1756	Exon 11, fragment B	Hydroxylamine (Y) Osmium (N)	164	47b
G551D[b]	G→A at 1784	Exon 11, fragment B	Hydroxylamine (Y) Osmium (N)	136	42
R553X	C→T at 1789	Exon 11, fragment B	Hydroxylamine (Y) Osmium (N)	767	42
W1098R[b]	T→C at 3424	Exon 17b, fragment E	Hydroxylamine (N) Osmium (Y)	342	54
3659delC[b]	Deletion of C at 3659	Exon 19, fragment E	Hydroxylamine (Y) Osmium (N)	577	47b
W1282X[b]	G→A at 3978	Exon 20, fragment F	Hydroxylamine (Y) Osmium (N)	621	47c
4271delC[b]	Deletion of C at 4271	Exon 23, fragment F	Hydroxylamine (Y) Osmium (N)	517	46

[a] The chemical predicted to modify the labeled strand of the probe is indicated by Y (yes) or N (no), and the size of the resulting cleavage product is given.
[b] Sequence changes that were successfully detected.

ide dyes [deionized formamide, 0.1% (w/v) xylene cyanol, and 0.1% (w/v) bromphenol blue].

Standard denaturing 6% (w/v) acrylamide gels are run. The DNA samples are denatured by boiling for 3 min and then stored on ice prior to loading. Gels are run in 0.5× TBE at 55 W for an appropriate length of time to resolve specific parts of the DNA sequence of interest. Gels are fixed in 10% (v/v) methanol–10% (v/v) glacial acetic acid for 30 min and dried onto Whatman (Clifton, NJ) 3MM paper and exposed to XAR film at room temperature. The films are usually developed after an exposure of 12–24 hr.

Alternative splicing of *CFTR* Gene

The *CFTR* cDNA has been successfully amplified in seven overlapping segments, which, starting at the 5′ end, have been called fragments A-6, 4-A, B, C, D, E, and F (see Fig. 1) [A6 and 4A are overlapping fragments of the segment A described by Chalkley and Harris (31)]. We have identified eight alternatively spliced forms of *CFTR* mRNA through analysis of these fragments, derived from normal *CFTR* genes, four of which have not been described previously (36). Other alternatively spliced products that are routinely seen in RT-PCR analysis of the *CFTR* cDNA are shown in Table V. In addition to the full-length PCR product predicted from the position of the primers, additional smaller products were seen in all fragments apart from the 4-A and F fragments, as shown in Fig. 13. The PCR products labeled in Fig. 13 were cut out of the agarose gels, the DNA was extracted, and then cloned and sequenced. The alternatively spliced forms of *CFTR* identified included transcripts lacking (a) exon 3, (b) exons 2 and 3, (c) exons 9 and 12, and (d) the final 357 bp of

TABLE V Sequence Analysis of Additional RT-PCR Products Produced by Amplification of CFTR cDNA[a]

CFTR cDNA fragments[b]	Number of control subjects (10) in which band was identified	Alternatively spliced form of CFTR mRNA[c]
A-6-1	7	Transcript lacks exon 3
A-6-2	3	Transcript lacks exon 2 and exon 3
A-6-2	3	Proportion of this band also represents transcript lacking exon 4
B-1	4	Transcript lacks exon 12
B-2	10	Transcript lacks exon 9
B-3	4	Transcript lacks exon 9 and exon 12
B-4	2	PCR artifact
C-1	5	PCR artifact
C-2	4	Transcript lacks first part of exon 13 as a result of use of a cryptic splice acceptor site, TCATTAG/A_{2147}A
D-1	4	Transcript lacks exon 14a
D-2	4	Transcript lacks final part of exon 15 as a result of use of a cryptic splice donor site, CA_{2863}GTTTCGT
E-1	1	PCR artifact

[a] Products are identified by the fragment name assigned in Fig. 1.
[b] As labeled on agarose gels shown on Fig. 13.
[c] As shown by sequence analysis of cloned PCR product.

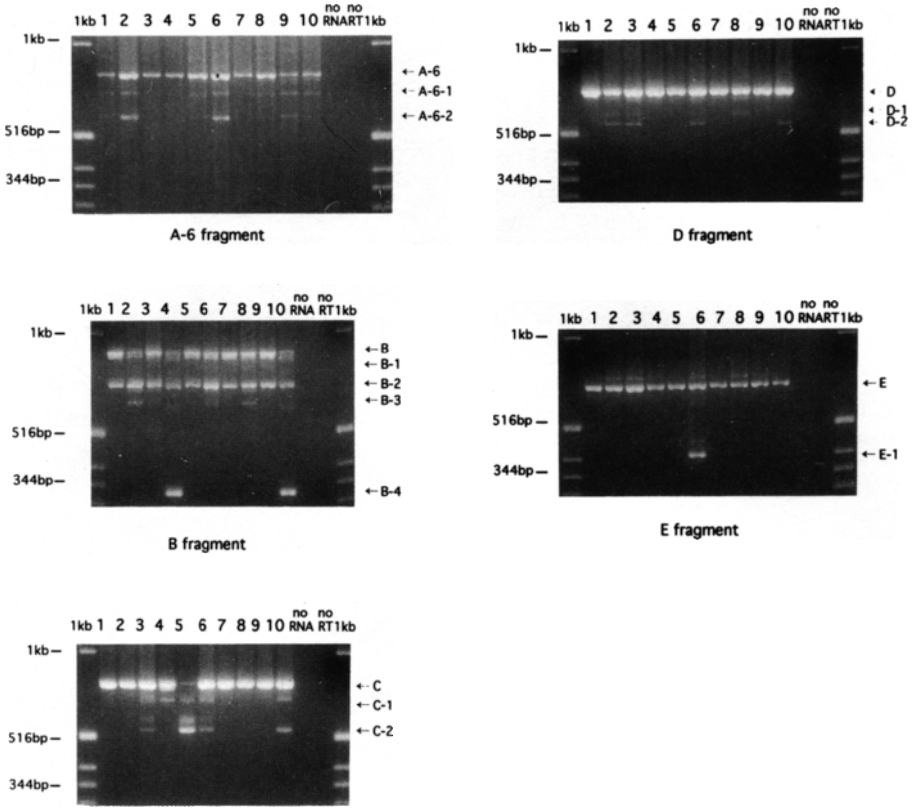


FIG. 13 Two percent (w/v) agarose gels showing the full-length and smaller PCR products for the A-6, B, C, D, and E fragments (see Fig. 1). The largest PCR product for each fragment is the full-length product expected from the position of the PCR primers. The smaller products were cloned and sequenced; the results of this analysis are shown in Table V. The bands immediately below B and above C-2 are heteroduplexes. The faint band in two subjects between B1 and B3 was not analyzed. Only the full-length products expected from the position of the PCR primers were seen for fragments 4-A and F (data not shown).

exon 15 as a result of use of the cryptic splice donor site CA_{2863}/GTTCGT. The PCR products B4, C1, and E1 were PCR artifacts formed by aberrant annealing between two distant but homologous regions of the *CFTR* cDNA.

Other studies investigating the alternatively spliced forms of *CFTR* have identified transcripts that we have not detected, including transcripts lacking exons 11 and 12, and 12 plus the 5′ portion of exon 13 as well as alternatively spliced forms of exon 1 (50). In addition, a novel exon, 119 bp in length, has been identified in nasal epithelial cell mRNA in 4 of 20 patients, as the result of utilization of cryptic splice donor and splice acceptor sites within intron 10 (51).

Some of the alternatively spliced forms of *CFTR* mRNA occur at relatively high proportions of total *CFTR* mRNA; studies using quantitative PCR and Southern blot analysis have estimated that *CFTR* mRNA from airway epithelial cells of normal individuals lacking exon 9 accounted for 10–90% of total *CFTR* mRNA (52) and transcripts lacking exon 12 accounted for 5–30% of total CFTR mRNA (53). The physiological importance of these alternatively spliced transcripts is uncertain.

Final Comments

We have used a combination of SSCP and CCM to analyze individual exons of the *CFTR* gene in order to maximize the efficiency of mutation detection. Single-stranded conformational polymorphism analysis, although less labor intensive than CCM, is restricted to the upper size limit of DNA fragments for which it is effective. The precise maximum size of DNA fragment that can be analyzed productively is difficult to estimate. Under the experimental conditions that we have employed, single base changes have been detected by SSCP analysis of DNA fragments of 400–450 bp. However, subsequent analysis of smaller fragments of 200–250 bp lying within some of the fragments analyzed previously has revealed a number of mutations that were not detected by analysis of the larger product (T. McDowell and S. Shackleton, unpublished data).

Chemical cleavage of mismatch analysis was carried out when (a) exons were too large to analyze by SSCP, (b) the DNA sequence of the exon itself gave complex or inconsistent patterns on SSCP, (c) the PCR product from amplification of the exon required gel purification to eliminate minor nonspecific amplification products.

The specific and consistent patterns produced by many mutations in the *CFTR* gene on SSCP analysis could provide a useful method for rapid screening for these mutations. Specifically, the band patterns produced by a variety of mutations in exons 3, 4, 7 (46), 11, and 19, which together encompass a large number of reported mutations, are particularly amenable to this type of analysis.

Any base change that is detected by the methods we have discussed should be looked for in at least 100 normal *CFTR* genes to exclude the possibility that a rare polymorphism exists within the population under analysis. Clearly, this is particularly important for missense mutations that merely alter an amino acid in the CFTR protein. The ultimate proof that a proposed novel mutation in the *CFTR* gene is actually disease causing is to express the mutant cDNA in a heterologous system and show a phenotypic abnormality in the function of the resultant CFTR protein.

Acknowledgments

We thank the many CF patients and their parents who provided the DNA samples that we have analyzed, and Dr. Lap-Chee Tsui for running the CF Consortium. In

addition we are grateful to the many referring physicians and DNA laboratories that have sent us samples. This work was supported by the Cystic Fibrosis Research Trust UK (S.S.), the Medical Research Council, the Medical Research Fund, Oxford, The Wellcome Trust, and the HC Roscoe award from the British Medical Association (J.H.).

References

1. J. M. Rommens, M. C. Iannuzzi, B.-S. Kerem, M. J. Drumm, G. Melmer, M. Dean, R. Rozmahel, J. Cole, D. Kennedy, N. Hidaka, M. Zsiga, M. Buchwald, J. R. Riordan, L.-C. Tsui, and F. S. Collins, *Science* **245,** 1059 (1989).
2. J. R. Riordan, J. M. Rommens, B.-S. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, S. Lok, N. Plavsic, J.-L. Chou, M. L. Drumm, M. C. Iannuzzi, F. S. Collins, and L.-C. Tsui, *Science* **245,** 1066 (1989).
3. B. Kerem, J. M. Rommens, J. A. Buchanan, D. Markiewicz, T. K. Cox, A. Chakravarti, M. Buchwald, and L.-C. Tsui, *Science* 245, 1073 (1989).
4. M. A. Gray, A. Harris, L. Coleman, J. R. Greenwell, and B. E. Argent, *Am. J. Physiol.* **257,** C240 (1989).
5. M. P. Anderson, D. P. Rich, R. J. Gregory, A. E. Smith, and M. J. Welsh, *Science* **251,** 679 (1991).
6. C. E. Bear, C. Li, N. Kartner, R. J. Bridges, T. J. Jensen, M. Ramjeesingh, and J. R. Riordan, *Cell* **68,** 809 (1992).
7. N. Kartner, J. W. Hanrahan, T. I. Jensen, A. T. Naismith, S. Sun, C. A. Ackerley, E. F. Reyes, L.-C. Tsui, J. M. Rommens, C. E. Bear, and J. R. Riordan, *Cell* **64,** 681 (1991).
8. B. C. Tilly, M. C. Winter, L. S. Ostergaard, C. O'Riordan, A. E. Smith, and M. J. Welsh, *J. Biol. Chem.* **267,** 9470 (1992).
9. S. C. Hyde, P. Emsley, M. J. Hartshorn, M. M. Mimmack, U. Gileadi, S. R. Pearce, M. P. Gallagher, D. R. Gill, R. E. Hubbard, and C. F. Higgins, *Nature (London)* **346,** 362 (1990).
10. M. L. Drumm, C. L. Smith, M. Dean, J. L. Cole, M. C. Iannuzzi, and F. S. Collins, *Genomics* **2,** (1988).
11. T. R. Fulton, A. M. Bowcock, D. R. Smith, L. Daneshvar, P. Green, L. L. Cavalli-Sforza, and H. Donis-Keller, *Nucleic Acids Res.* **17,** 271 (1989).
12. A. M. Poustka, H. Lehrach, R. Williamson, and G. Bates, *Genomics* **2,** 337 (1988).
13. R. Anand, D. J. Ogilvie, R. Butler, J. H. Riley, R. S. Finniear, J. S. Powell, J. C. Smith, and A. F. Markham, *Genomics* **9,** 124 (1991).
14. E. D. Green and M. V. Olson, *Science* **250,** 94 (1990).
15. K. Yoshimura, H. Nakamura, B. C. Trapnell, W. Dalemans, A. Pavirani, J.-P. Lecocq, and R. G. Crystal, *J. Biol. Chem.* **266,** 9140 (1991).
16. J.-L. Chou, R. Rozmahel, and L.-C. Tsui, *J. Biol. Chem.* **266,** 24471 (1991).
17. J. Koh, J. Sferra, and F. S. Collins, *J. Biol. Chem.* **21,** 15912 (1993).
18. I. C. Crawford, P. C. Maloney, P. L. Zeitlin, W. B. Guggino, S. C. Hyde, H. Turley, K. C. Gatter, A. Harris, and C. F. Higgins, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 9262 (1991).
19. G. M. Denning, L. S. Ostedgaard, S. H. Cheng, A. E. Smith, and M. J. Welsh, *J. Clin. Invest.* **89,** 339 (1992).

20. J. F. Engelhardt, J. R. Yankaskas, S. A. Ernst, Y. Yang, C. R. Marino, R. C. Boucher, J. A. Cohn, and J. M. Wilson, *Nature Genet.* **2,** 240 (1993).
21. N. Morral, V. Nunes, T. Casals, N. Cobos, O. Asensio, J. Dapena, and X. Estivill, *Hum. Mol. Genet.* **2,** 677 (1993).
22. M. J. Welsh and A. E. Smith, *Cell* **73,** 1251 (1993).
23. R. K. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuichi, G. T. Horn, K. B. Mullis, and H. A. Erlich, **Science 239,** 487 (1988).
24. M. Orita, Y. Suzuki, T. Sekiya, and K. Hayashi, *Genomics* **5,** 874 (1989).
25. R. M. Myers, T. Maniatis, and L. Lerman, *Methods Enzymol* **155,** 501 (1987).
26. R. G. H. Cotton, N. R. Rodrigues, and R. D. Campbell, *Proc. Natl. Acad. Sci. U.S.A.* **85,** 4397 (1988).
27. A. J. Montandon, P. M. Green, F. Giannelli, and D. R. Bentley, *Nucleic Acids Res.* **17,** 3347 (1989).
28. J. Chelley, J. C. Kaplan, P. Maire, S. Gautron, and A. Kahn, *Nature (London)* **333,** 858 (1988).
29. R. Roberts, D. R. Bentley, T. F. M. Barby, E. Manners, and M. Bobrow, *Lancet* **336,** 1523 (1990).
30. J. A. Naylor, P. M. Green, A. J. Montandon, C. R. Rizza, and F. Giannelli, *Lancet* **337,** 635 (1991).
31. G. Chalkley and A. Harris, *J. Med. Genet.* **28,** 777 (1991).
32. G. I. Bell, K. H. Karam, and W. J. Rutter, *Proc. Natl. Acad. Sci. U.S.A.* **78,** 5759 (1981).
33. J. M. Chirgwin, A. E. Przybyla, R. J. Macdonald, and W. J. Rutter, *Biochemistry* **18,** 5294 (1979).
33a. J. Zielenski, R. Rozmahel, D. Bozon, B. Kerem, Z. Grzelczak, J. R. Riordan, J. M. Rommens, and L.-C. Tsui, *Genomics* **10,** 214 (1991).
34. J. Hull, S. Shackleton, and A. Harris, *Genomics* **19,** 362 (1994).
35. J. Hull, S. Shackleton, and A. Harris, *Hum. Mol. Genet.* **2,** 689 (1993).
36. J. Hull, S. Shackleton, and A. Harris, *Hum. Mol. Genet.* **3,** 1141 (1994).
37. J. Zielenski, D. Bozon, B. Kerem, D. Markiewicz, J. M. Rommens, and L.-C. Tsui, *Genomics* **10,** 229 (1991).
38. L. Smit, T. Strong, J. Cole, M. Lnowles, S. Turpin, M. Iannuzzi, T. Petty, L.-C. Tsui, and F. S. Collins, *Paediatr. Pulm.* Suppl. **6,** 244 (1991).
39. B.-S. Kerem, J. Zielenski, D. Markiewicz, D. Bozon, E. Gazit, J. Yahaf, D. Kennedy, J. R. Riordan, F. S. Collins, J. R. Rommens, and L.-C. Tsui, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 8447 (1990).
40. C. S. Chu, B. C. Trapnell, S. Curristin, G. R. Cutting, and R. G. Crystal, *J. Clin. Invest.* **90,** 785 (1992).
41. T. V. Strong, L. S. Smit, N. Nasr, D. L. Wood, J. L. Cole, M. C. Iannuzzi, R. C. Stern, and F. S. Collins, *Hum. Mutat.* **1,** 380 (1992).
42. G. R. Cutting, L. M. Kasch, B. J. Rosenstein, J. Zielenski, L.-C. Tsui, S. E. Antonorakis, and H. H. Kazazian, *Nature (London)* **346,** 366 (1990).
43. L.-C. Tsui, *Trends Genet.* **8,** 392 (1992).
44. R. Slomski, M. Schloesser, L.-P. Berg, M. Wagner, V. V. Kakkar, D. N. Cooper, and J. Reiss, *Hum. Genet.* **89,** 615 (1992).
45. S. Shackleton and A. Harris, *Hum. Mol. Genet.* **1,** 445 (1992).
46. S. Shackleton, J. Hull, S. Dear, A. Seller, A. Thomson, and A. Harris, *Hum. Mutat.* **3,** 141 (1994).

47. S. M. Forrest, H. H. Dahl, D. W. Howells, I. Dianzani, and R. G. H. Cotton, *Am. J. Hum. Genet.* **49,** 175 (1991).
47a. V. Nunes, M. Chillon, T. Dork, B. Tummler, T. Casals, and X. Estivill, *Hum. Mol. Genet.* **2,** 79 (1993).
47b. B. Kerem, J. Zielenski, D. Markiewicz, D. Bozon, E. Gazit, J. Yahav, D. Kennedy, J. R. Riordan, F. S. Collins, J. M. Rommens, and L.-C. Tsui, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 8447 (1990).
47c. M. Vidaud, P. Fanen, J. Martin, N. Ghanem, S. Nicolas, and M. Gossens, *Hum. Genet.* **85,** (1990).
48. F. Sanger, S. Nicklen, and A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74,** 5463 (1977).
49. P. M. Green, D. R. Bentley, R. S. Mibashan, I. M. Milsson, and F. Gianelli, *EMBO J.* **8,** 1067 (1989).
50. T. V. Strong and F. S. Collins, *in* "Cystic Fibrosis—Current Topics" (J. A. Dodge, D. J. H. Brock, and J. H. Widdicombe, eds.), p. 3. John Wiley & Sons, Chichester, 1993.
51. K. Will, M. Stuhrmann, M. Dean, and J. Schmidtke, *Hum. Mol. Genet.* **2,** 231 (1993).
52. C. S. Chu, B. C. Trapnell, J. J. Murtagh, Jr., J. Moss, W. Dalemans, S. Jallat, A. Mercenier, A. Pavirani, J. P. Lecocq, G. R. Cutting, W. B. Guggino, and R. G. Crystal, *EMBO J.* **10,** 1355 (1991).
53. S. Bremer, T. Hoof, M. Wilke, R. Busche, B. Scholte, J. R. Riordan, and G. Maass, *et al., Eur. J. Biochem.* **799,** 1 (1992).
54. J. Zielenski, D. Markiewicz, H. S. Chen, K. Schappert, A. Seller, P. Durie, M. Corey, and L.-C. Tsui, *Hum. Mutat.* **5,** 43 (1995).

[6] Protein Truncation Test for Presymptomatic Diagnosis of Familial Adenomatous Polyposis

Rob B. van der Luijt and P. Meera Khan

Introduction

The protein truncation test (PTT), also known as the *in vitro*-synthesized protein (IVSP) assay, is a method of direct mutation detection based on the *in vitro* transcription and translation of polymerase chain reaction (PCR)-amplified sequences. The PTT allows the selective detection of chain-terminating mutations in large amplified segments and has been shown to be uniquely useful for the molecular diagnosis of familial adenomatous polyposis (FAP) (1, 2).

Familial adenomatous polyposis is an autosomal dominant predisposition to colorectal cancer, characterized by the development of numerous adenomatous polyps located in the colorectum. Without surgical treatment, progression of one or more of these benign tumors will inevitably result in colorectal adenocarcinoma. Identification of predisposed individuals before the onset of symptoms is extremely important in order to prevent morbidity and mortality in FAP. For this reason, the first-degree relatives of FAP patients (who are at a 50% *a priori* risk) are subjected to periodic endoscopic examination of the colorectum.

The localization of the gene responsible for FAP on chromosome 5q21-q22 led to the identification of tightly linked polymorphic markers, which were found to be valuable tools for the presymptomatic DNA diagnosis of FAP by genetic linkage analysis. A more accurate and direct approach to the presymptomatic DNA diagnosis of FAP became available when a positional cloning strategy resulted in the isolation and characterization of the gene involved in FAP. Germ line mutations of this gene, designated as *APC* (adenomatous polyposis coli), were identified in individuals affected with FAP, while somatic *APC* mutations were found in sporadic colorectal cancers. The *APC* gene spans about 200 kb of genomic DNA and has a coding region of 8532 base pairs (bp). The *APC* coding region is divided into 15 exons, of which the first 14 range in size between 78 and 379 bp, and the most 3′ exon, numbered 15, is 6.5 kb long.

The identification of the *APC* gene has facilitated the detection of disease-causing mutations in individuals with FAP. Mutation studies of *APC* are of importance for the presymptomatic diagnosis of FAP discussed above, but also for the establishment of possible genotype–phenotype correlations. For these reasons, many investigators have performed mutation analysis of the *APC* gene, using a variety of methods of direct mutation detection including single-strand conformational polymorphism (SSCP) analysis (3), the RNase protection assay (RPA) (4), and denaturing gradient

Methods in Molecular Genetics, Volume 8

gel electrophoresis (DGGE) (5). Mutation studies of the *APC* gene in FAP, reviewed by Nagase and Nakamura (6), revealed that the mutation spectrum of *APC* is extremely heterogeneous, with the mutations being scattered over almost the entire coding region. Using SSCP, RPA, or DGGE, which permit the analysis of sequences ranging in size between 100 and 600 bp, the molecular analysis of the entire *APC* coding region is a laborious and technically difficult procedure. To screen the entire *APC* coding sequence by the above-mentioned methods, the coding region must be divided into up to 40 segments. Each of these segments is amplified by PCR and analyzed to detect genetic alterations separately. Once a variant has been found in one of these fragments, its sequencing is necessary to determine whether the alteration represents a disease-causing mutation or an inconsequential polymorphism.

A striking feature of the observed mutation spectrum of the *APC* gene is that the majority (> 95%) of the germ line mutations identified in FAP patients would lead to the premature termination of translation. Thus, chain-terminating mutations, introduced by small deletions, insertions, as well as single base substitutions, are predicted to result in the synthesis of truncated gene products.

Because of the distribution and nature of the mutations in the large *APC* gene, an ideal method for mutation analysis would be to allow the selective detection of chain-terminating mutations in large stretches of the coding sequence. Such a method, called the protein truncation test (PTT), was developed for the detection of point mutations in the Duchenne muscular dystrophy (*DMD*) gene (7). The PTT allowed the rapid detection of translation-terminating mutations in the *DMD* gene, which has a coding region of about 12 kb dispersed among 79 exons.

Principle of Protein Truncation Test

The PTT is based on the *in vitro* coupled transcription and translation of coding sequences amplified by the PCR. A flowchart of the PTT procedure, in which the different steps and the principle of the method are outlined, is presented in Fig. 1. For PTT, stretches of coding regions are amplified by PCR with a modified primer containing a promoter and translation initiation sequence. This modifiction allows *in vitro* transcription and translation of the PCR products. During translation, radiolabeled amino acids are incorporated into the translation products, which are resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and detected by fluorography. If a normal template is subjected to the PTT, a full-length protein is produced by runoff translation from the amplified segment. The appearance of a shortened polypeptide indicates the presence of a translation–termination mutation in the corresponding segment.

Generation of Templates

The purpose of the first step in the PTT protocol is to obtain large, uninterrupted stretches of coding sequence that can be subjected to transcription and translation *in*

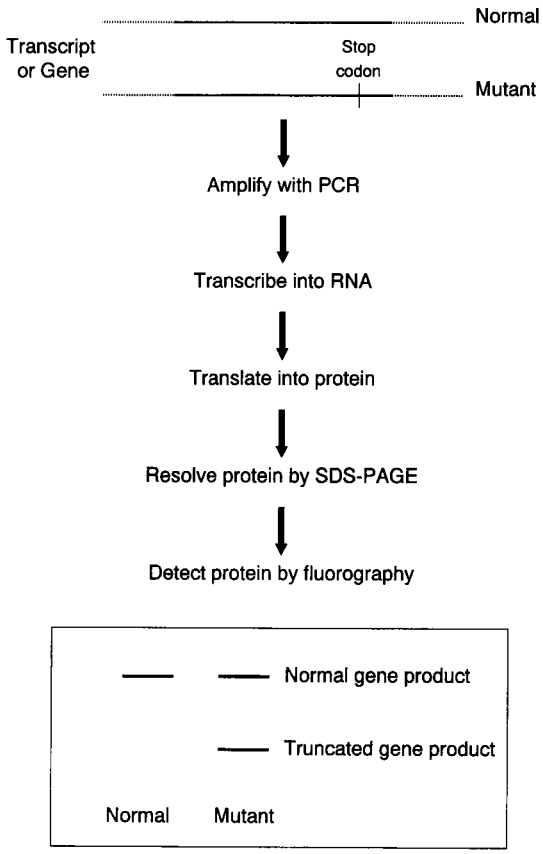


FIG. 1 Flowchart and principle of the protein truncation test (PTT). *Top:* The two horizontal lines schematically represent a normal and mutated gene or transcript. *Middle:* The consecutive steps of the PTT are shown. The *in vitro* transcription and translation, shown here as two separate steps, are usually performed as a one-step, coupled reaction. *Bottom:* The result of PTT analysis performed on a gene or transcript in a normal individual as well as in an individual heterozygous for a chain-terminating mutation.

vitro. Depending on the intron–exon structure of the gene of interest, templates suitable for the PTT are generated either by reverse transcription-based PCR (RT-PCR) of RNA or by genomic DNA amplification. For PCR, the coding region is divided into overlapping segments, which may be several kilobases in length. Conveniently sized templates for PTT usually are between 1 and 2 kb long. To allow *in vitro* transcription and translation of the amplified segment, PCR is performed with a T7-modified, gene-specific sense oligonucleotide that places transcriptional and translational sequences at the 5′ end.

If the gene of interest has a favorable exonic structure, templates for the PTT can be obtained by PCR performed directly on genomic DNA samples. A large part of the coding region of *APC* can be analyzed by PTT using genomic DNA, because it contains a 6.5-kb-long exon.

In the majority of the human disease genes similarly large exons are not present. Therefore, templates for PTT cannot be generated by PCR on genomic DNA. In such cases, the templates are prepared by RT-PCR using samples of RNA (2, 7, 8). In the RT-PCR-based procedure, multiple coding exons form parts of a cDNA template. Reverse transcription of RNA, using either random hexamers or gene-specific primers, is employed to synthesize a cDNA, which is used as a template for the primary PCR. The resulting RT-PCR product is then subjected to a secondary, nested PCR with the T7-modified primer pair.

Guidelines for Designing T7-Modified Polymerase Chain Reaction Primers

A primer pair employed to generate templates for PTT consists of a T7-modified sense (or forward) primer and a conventional antisense (or reverse) primer. The sense oligonucleotide contains three important elements: a promoter sequence necessary for transcription, an initiation signal for translation, and a sequence complementary to the gene of interest. Primers for the PTT are based on the 37-base sequence originally described by Sarkar and Sommer (9), which contains the 18-bp bacteriophage T7 promoter and the 8-bp eukaryotic translation initiation signal required for *in vitro* transcription and translation, respectively (Fig. 2). On the basis of this sequence, two

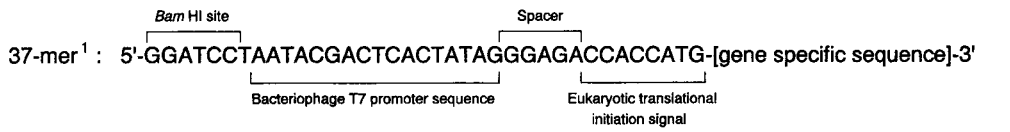


FIG. 2 Sequences of three different T7-modified oligonucleotide primers, employed for *in vitro* transcription and translation of PCR products. In the upper sequence, the individual elements (*Bam*HI restriction site, T7 promoter, spacer, initiation signal, and gene-specific sequence) of the primer are depicted. Sequence differences between the three primers are shown in boldface. *Key to references*: (1) G. Sarkar and S. S. Sommer, *Science* **244,** 331 (1989); (2) S. M. Powell, G. M. Petersen, A. J. Krush, S. Booker, J. Jen, F. M. Giardiello, S. R. Hamilton, B. Vogelstein, and K. W. Kinzler, *N. Engl. J. Med.* **329,** 1982 (1993); and (3) P. A. M. Roest, R. G. Roberts, S. Sugino, G.-J. B. van Ommen, and J. T. den Dunnen, *Hum. Mol. Genet.* **2,** 1719 (1993).

T7-modified primers were developed that are currently employed for PTT. These primers carry modifications at the spacer region (located between the T7 promoter and initiation sequences) and in the translation initiation signal. The promoter sequence and initiation signal usually are separated by a spacer of 5–7 bp. The initiation signal located in the primer ends with the methionine start codon ATG or, alternatively, with the sequence ATGG, which corresponds better with the translation initiation consensus or optimal Kozak sequence (10, 11). If cloning of the amplified segment is desired, a sequence may be added to the 5′ end of the primer, which introduces a recognition site for a restriction endonuclease in the PCR product. By using the primer sequences shown in Fig. 2, a *Bam*HI restriction site is created in the amplified segment.

Polymerase chain reaction primers suitable for the PTT are obtained by extending the 3′ end of either of the three transcription/translation sequences shown in Fig. 2 with a sequence complementary to the gene of interest. To allow translation of the amplified segment, the gene-specific primer sequence should be selected in such a way that the open reading frame is maintained within the T7-modified PCR product. If a primer is designed that is based on the 37- or 39-base sequence (Fig. 2), the first three bases of the gene-specific sequence should correspond to a codon in the open reading frame of the gene of interest. However, if the 38-base sequence is used to develop a PTT primer, the first two bases of the gene-specific oligonucleotide should represent the second and the third base of a codon in the open reading frame. This difference with the former two primer sequences is due to the extra G that follows the ATG initiation codon in the 38-base sequence. Oligonucleotide sequences complementary to the gene to be analyzed are typically between 20 and 30 bases long, while the total length of T7-modified primers usually ranges between 60 and 70 bases. Antisense oligonucleotides do not have to meet special requirements for allowing transcription and translation of the corresponding PCR product. Their selection should follow the rules of normal primer set design, making the PCR amplification as optimal as possible. If cloning of the PCR product is required, the antisense primer may contain a restriction site at the 5′ end. It is not necessary for antisense primers to harbor a translation termination signal, because protein products are generated by run off translation from the amplified segment.

In Vitro Transcription and Translation of Polymerase Chain Reaction Products

Amplified segments generated by PCR with T7-modified primers can be efficiently transcribed and translated *in vitro* using a T7-coupled reticulocyte lysate system. The TnT T7 coupled rabbit reticulocyte lysate system (Promega, Madison, WI) is commonly used for the PTT. The advantage of using this system is that the *in vitro* transcription and translation reactions are coupled and can be performed as a single-step, one-tube reaction. The *in vitro* transcription/translation reaction requires the following components:

DNA template
Reticulocyte lysate
Amino acid mixture omitting the amino acid used for labeling
Reaction buffer
T7 RNA polymerase
Radiolabeled amino acid

The PTT is usually performed in a reaction volume of 25 μl, using 200–500 ng of the T7 PCR product. Purification of the T7 PCR products prior to the *in vitro* transcription/translation reaction is not necessary. The radiolabeled amino acids most commonly used for the PTT are [^{3}H]leucine and [^{35}S]methionine. The radiolabeled amino acid employed to label the *in vitro*-synthesized proteins should not be present (in its nonlabeled form) in the amino acid mixture used for the translation reaction. To prevent degradation of *in vitro*-synthesized RNA, the addition of an inhibitor of ribonuclease activity (RNasin ribonuclease inhibitor; Promega) is recommended.

SDS-PAGE Analysis and Detection of Translation Products

After translation, the products are subjected to polyacrylamide slab gel electrophoresis in the presence of 0.1% (w/v) sodium dodecyl sulfate. Electrophoresis is performed using a discontinuous buffer system (12). The labeled protein bands in the gel are visualized by fluorography, which increases the sensitivity of detection of the ^{3}H (and ^{35}S)-labeled translation products. Fluorography is performed by infusing an organic scintillant into the gel, which converts the energy emitted by the isotope to visible light. In this chapter, the method based on the organic scintillant 2,5-diphenyloxazole (PPO) in dimethyl sulfoxide (DMSO) (13) was employed. To allow infusion of PPO into the gel, the gel is first dehydrated with DMSO. After the PPO treatment, the gel is washed with water and dried using a vacuum slab gel dryer. Instead of using PPO and DMSO, fluorographic enhancement of the signal can also be achieved using commercially available scintillant solutions such as Amplify (Amersham, Arlington Heights, IL) or En^{3}Hance (New England Nuclear, Boston, MA).

Methods

Polymerase Chain Reaction Amplification of APC exon 15 for Protein Trucation Test

The generation of templates for the PTT analysis of *APC* exon 15 is based on the use of genomic DNA samples. For the PTT analysis, the 6.5k-kb exon 15 is divided into four overlapping segments: 15a-f (codons 654–1264), 15d-j (codons 989–1700),

TABLE I Primer Sequences Employed for Protein Truncation Test Analysis Covering *APC* Exon 15

Segment	Length (kb)	Codons	Primer sequence[a]
Exon 15a–f	1.8	654–1264	5′ [T7-trans]-CAAATCCTAAGAGAGAACAACTGTC 3′ 5′ CACAATAAGTCTGTATTGTTTCTT 3′
Exon 15d–j	2.1	989–1700	5′ [T7-trans]-GATGATGAAAGTAAGTTTTGCAGTT 3′ 5′ GAGCCTCATCTGTACTTCTGC 3′
Exon 15j–q	2.2	1595–2337	5′ [T7-trans]-GCCCAGACTGCTTCAAAATTAC 3′ 5′ CTTATTCCATTTCTACCAGGGGAA 3′
Exon 15p–3′utr	2.4	2101–2844	5′ [T7-trans]-TGGAAAGCTATTCAGGAAGGTG 3′ 5′ CCAGAACAAAAACCCTCTAACAAG 3′

[a] In the primer sequences listed, [T7-trans] means: 5′ GGATCCTAATACGACTCACTATAGGAACAGACCACCATG 3′. (See also Fig. 2.)

15j-q (codons 1595–2337), and 15p-3′utr (codons 2101–2844). The primer sequences and the sizes of the corresponding PCR products are listed in Table I.

1. Dispense between 100 and 500 ng of genomic DNA and 12.5 pmol of each primer in 50 μl of PCR reaction buffer.

PCR reaction buffer
dNTPs (200 μM)
Tris-HCl (pH 9.0), 10 m*M*
KCl, 50 m*M*
$MgCl_2$, 1.0 or 1.5 m*M*
Gelatin (0.1%, w/v)
Triton X-100 (0.1%, v/v)
SuperTaq polymerase (HT Biotechnology), 0.5 unit

Note: For segments 15a-f, 15j-q, and 15p-3′utr, the final concentration of $MgCl_2$ is 1.5 m*M*, while for segment 15d-j the $MgCl_2$ concentration is 1.0m*M*.

2. Perform PCR by subjecting the reaction mixture to 35 cycles (1 cycle is 60 sec at 95°C, 90 sec at 55°C, and 120 sec at 70°C).

3. After the PCR, analyze the reaction products by standard agarose gel electrophoresis. *Note*: For the PCR products obtained using the primers for *APC* exon 15, a 1% (w/v) agarose gel is appropriate.

In Vitro Transcription and Translation of Polymerase Chain Reaction Products

The following protocol is based on the TnT T7 coupled Reticulocyte Lysate System (Promega), using [^{3}H]leucine (Amersham) as the radiolabeled amino acid.

1. Remove the reagents from storage at −70°C and immediately place the RNA polymerase on ice. Rapidly thaw the TnT reticulocyte lysate by hand warming and place on ice. The other components can be thawed at room temperature and then stored on ice.

2. The reaction components are assembled in a microcentrifuge tube (0.5 or 1.5 ml), following the example below. After all the components are added, the lysate is gently mixed by flicking the tube. A brief centrifuge spin may be necessary to return the reaction to the bottom of the tube. An example of the TnT T7 lysate reaction using [^{3}H]leucine would require the following:

TnT rabbit reticulocyte lysate	12.5 μl
TnT reaction buffer	1 μl
TnT T7 RNA polymerase	0.5 μl
Amino acid mixture minus leucine, 1 m*M*	0.5 μl
RNasin ribonuclease inhibitor, 40 U/μl	0.5 μl
[^{3}H]leucine (100–200 Ci/mmol) at 0.5 mCi/ml	2 μl
DNA template	0.2–0.5 μg
Nuclease-free water	To final volume 25 μl

Note: As a control for the *in vitro* transcription/translation reaction, the TnT lysate system contains a luciferase-encoding control plasmid. When this template DNA is subjected to the *in vitro* transcription/translation, a 62-kDa protein is synthesized.

3. Incubate the reaction at 30°C for 60 min.

4. Stop the reaction by adding 25 μl of SDS sample buffer (2×).

5. Cap the tube and incubate the sample at 100°C for 2–5 min to denature the proteins.

6. If the translation products are not analyzed by SDS-PAGE directly, they can be stored at −20°C.

Analysis of Translation Products by SDS-PAGE

The following example describes the analysis of translation products by electrophoresis on a 14% (w/v) SDS-polyacrylamide gel, 1 mm thick and 200 × 200 mm. Formulations for preparing separating and stacking gels are provided in Tables II and III, respectively.

1. Pour the separating gel mix into the assembled gel plates, leaving sufficient space above the gel for the stacking gel to be added later. Gently overlay the gel mix with distilled water.

2. After polymerization (20–30 min), remove the overlay and rinse the surface of

TABLE II Formulation of Sodium Dodecyl Sulfate-Polyacrylamide Separating Gel[a]

Component	Volume
Acrylamide–bisacrylamide [30% (w/v)/0.8% (w/v)]	12.1 ml
Distilled water	6.8 ml
Separating gel buffer, 4×	6.5 ml
SDS, 10% (w/v)	260 μl
Ammonium persulfate, 10% (w/v)	200 μl
TEMED	20 μl
	26 ml

[a] Acrylamide (14%, w/v), 1 mm thick, 200 × 200 mm.

TABLE III Formulation of Sodium Dodecyl Sulfate-Polyacrylamide Stacking Gel[a]

Component	Volume
Acrylamide–bisacrylamide [30% (w/v)/0.8% (w/v)]	1.5 ml
Distilled water	5.7 ml
Stacking gel, 4× buffer	2.5 ml
SDS, 10% (w/v)	100 μl
Ammonium persulfate, 10% (w/v)	150 μl
TEMED	10 μl
	10 ml

[a] Acrylamide, 4.5% (w/v).

the separating gel with water to remove any unpolymerized acrylamide and then with a small volume of electrophoresis buffer.

3. Pour the stacking gel mix on top of the separating gel to fill the remaining space. Immediately insert a comb of the appropriate thickness.

4. After the stacking gel has polymerized (10–15 min), remove the comb and rinse the wells with electrophoresis buffer to remove unpolymerized acrylamide.

5. Place the gel sandwich in the electrophoresis apparatus. Do not prerun the gel.

6. Load a prestained or ^{14}C-labeled SDS-PAGE molecular weight marker on the gel.

7. Briefly spin down the samples. Load half (25 μl) of each *in vitro* transcription/translation reaction on the gel. Also load 5 μl of the luciferase control reaction on the gel.

8. Electrophoresis is performed at a constant current of 30 mA in the stacking gel

and 40 mA in the separating gel. The run is stopped when the bromphenol blue dye reaches the bottom end of the gel. Under these conditions, electrophoretic separation of translation products on a 14% (w/v) polyacrylamide gel of 1 mm thickness and 200 × 200 mm usually takes about 3 hr.

Composition of Buffers and Solutions for SDS-PAGE

Acrylamide–bisacrylamide solution [30% (w/v) acrylamide, 0.8% (w/v) bisacrylamide]: Combine the following:

Acrylamide	146.0 g
N′,N′-Bismethyleneacrylamide	4.0 g

Adjust the volume to 500 ml with distilled water. Filter and store at 4°C in the dark. *Note:* Acrylamide is a neurotoxin. Always wear gloves when handling acrylamide, preparing monomer solutions, or pouring gels (polyacrylamide is not toxic)

Sodium dodecyl sulfate (SDS), 10% (w/v): Dissolve 10 g of SDS in distilled water with gentle stirring. Add distilled water to 100 ml

SDS sample buffer, 2× concentrated [100 m*M* Tris-HCl, (pH 6.8), 20% (v/v) glycerol, 4% (w/v) SDS, 0.005% (w/v) bromphenol blue, 5% (v/v) 2-mercaptoethanol]: Combine the following:

Tris-HCl (pH 6.8), 0.5 *M*	2 ml
Glycerol	2 ml
SDS, 10% (w/v)	4 ml
Bromphenol blue	5 mg
2-Mercaptoethanol	0.5 ml

Add distilled water to a final volume of 10 ml. Store at room temperature. 2-Mercaptoethanol is added to the sample buffer immediately before use

Electrophoresis buffer, 5× concentrated [125 m*M* Tris, 960 m*M* glycine, 0.5% (w/v) SDS]: Combine the following:

Tris base	30 g
Glycine	144 g
SDS, 10% (w/v)	100 ml

Dissolve in distilled water, and adjust the final volume to 2 liters

Separating gel buffer, 4× concentrated (1.5 *M* Tris-HCl, pH 8.8): Dissolve 36.34 g of Tris base in distilled water. Adjust the pH to 8.8 with 12 *N* HCl and add distilled water to a 200-ml final volume

Stacking gel buffer, 4× concentrated (0.5 *M* Tris-HCl, pH 6.8): Dissolve 12.12 g of Tris base in distilled water. Adjust the pH to 6.8 with 12 *N* HCl and add distilled water to a 200-ml final volume

Detection of Translation Products by Fluorography

1. Remove the gel sandwich from the electrophoresis apparatus. Cut off the stacking gel and discard. Place the separating gel in a plastic tray.
2. Submerge the separating gel in once-used DMSO and incubate for 30 min while gently shaking. *Note:* DMSO is a hazardous chemical, therefore steps 2–5 of the protocol should be performed in a fume hood.
3. Remove the DMSO and submerge the gel in fresh DMSO. Incubate for 30 min while gently shaking.
4. Remove the DMSO. Submerge the gel in a solution of PPO in DMSO and incubate for at least 2 hr.

   PPO/DMSO solution: Dissolve 226 g of PPO in 1000 ml of DMSO. Stir until the solution becomes clear.

5. Remove the PPO/DMSO solution and wash the gel in distilled water for 1 hr. Replace the water every 15 min.
6. Remove the gel from the tray and place it on two sheets of thin filter or chromatography paper. Cover the gel with Saran Wrap.
7. Dry the gel in a vacuum slab gel dryer at 60°C for about 1 hr.
8. Autoradiograph the gel at −70°C using Kodak (Rochester, NY) X-Omat AR film. Usually, an overnight exposure gives sufficiently strong signals.

Mutation Analysis of *APC* Gene by Protein Truncation Test

The PTT has been employed successfully for the identification of germ line and somatic mutations of the *APC* gene in FAP patients and colorectal tumors, respectively (1, 2). For the PTT of the *APC* gene, the total coding region is divided into five overlapping segments. As described in detail by Powell *et al.* (2), the first 14 exons of *APC* can be analyzed as a single segment of about 2.4 kb, using a two-stage RT-PCR-based procedure performed on samples of RNA. For the PTT analysis of the 6.5-kb exon 15 of *APC*, genomic DNA samples are used. For PCR, exon 15 is divided into four overlapping segments, the length of which ranges from 1.8 to 2.4 kb. Although the germ line mutations of *APC* are scattered over almost the entire coding sequence, the majority of them are located within the 5′ half of the gene, which in PTT analysis is covered by the first three segments (codons 1–1700). In contrast to the germ line mutations, the somatic *APC* mutations identified in colorectal adenomas and carcinomas have been shown to be less dispersed over the coding region. The segment of exon 15 covering codons 989 to 1700 of *APC* encompasses a region of the gene in which the somatic mutations are clustered, the so-called mutation clus-

ter region (MCR). The PTT analysis of this single segment in colorectal tumors will result in the efficient identification of the majority of the somatic *APC* mutations.

To demonstrate the usefulness of this approach for the detection of germ line and somatic *APC* mutations, we have subjected a set of samples containing chain-terminating *APC* mutations to the PTT analysis. The PTT was performed on two overlapping segments of the *APC* gene in a set of samples containing truncating *APC* mutations previously identified by DGGE (5). Templates for PTT were generated by PCR amplification of genomic DNA samples obtained from FAP patients as well as from colorectal tumor-derived cell lines. For amplification of the 1.8-kb segment of *APC* exon 15 encompassing codons 653–1264, the T7-modified sense primer apc15aT7 was used in combination with the antisense primer apc15f3b. The 2.1-kb segment of exon 15 covering codons 989–1700 was amplified using the sense and antisense primer apc15dT7 and apc15j3, respectively. Sequences of the oligonucleotides employed for the amplification of the *APC* exon 15 are listed in Table I. The PCR amplifications were performed as described in Methods. After PCR, amplification products were first analyzed by conventional electrophoresis on 1% (w/v) agarose gels, to check whether their sizes were correct and to estimate the amount of PCR product required for the *in vitro* transcription/translation reaction. Subsequently, unpurified PCR products were subjected to PTT analysis. Three DNA samples were analyzed by PTT of segment 15a-f (codons 653–1264), while segment 15d-j (codons 989–1700) was investigated in five DNA samples (Fig. 3).

The PTT analysis of the 1.8-kb exon 15 segment a-f (encompassing codons 653–1264) in samples from two FAP patients with different chain-terminating mutations revealed shortened translation products in both of them. The mutation previously identified in patient 23 resulted in a premature termination at codon 820; the size of the truncated peptide is about 19 kDa (Fig. 3B, lane 2). In patient 46, the mutation resulted in a stop codon at 1061. The PTT showed a truncated peptide of 45 kDa (Fig. 3B, lane 3). In both the patients, the normal product of 68 kDa corresponding to the wild-type *APC* allele was present in addition to the truncated gene product. As a control, a DNA sample from a normal individual was also subjected to the PTT, which showed only the full-length translation product of 68 kDa (Fig. 3B, lane 1).

Analysis by PTT of a 2.1-kb mutation-dense region from exon 15 (segment 15d–j; codons 989–1700) performed in DNA samples from two FAP patients and two colorectal tumor-derived cell lines confirmed the presence of chain-terminating *APC* mutations in all four of them. In patient 68, a 5-bp deletion was found at codon 1309, which generated a stop codon immediately downstream. The PTT showed a truncated protein of 35 kDa (Fig. 3B, lane 5). In the colorectal tumor-derived cell line SW403, both *APC* alleles harbor a truncating mutation, located at positions 1198 and 1278. As can be seen in Fig. 3B (lane 6), two truncated proteins (23 and 32 kDa, respectively) and no wild-type APC product were detected by PTT. SW480 is also a colorectal tumor-derived cell line, which carries a truncating mutation at position 1338 and has lost the remaining *APC* allele. The PTT analysis of exon 15 segment

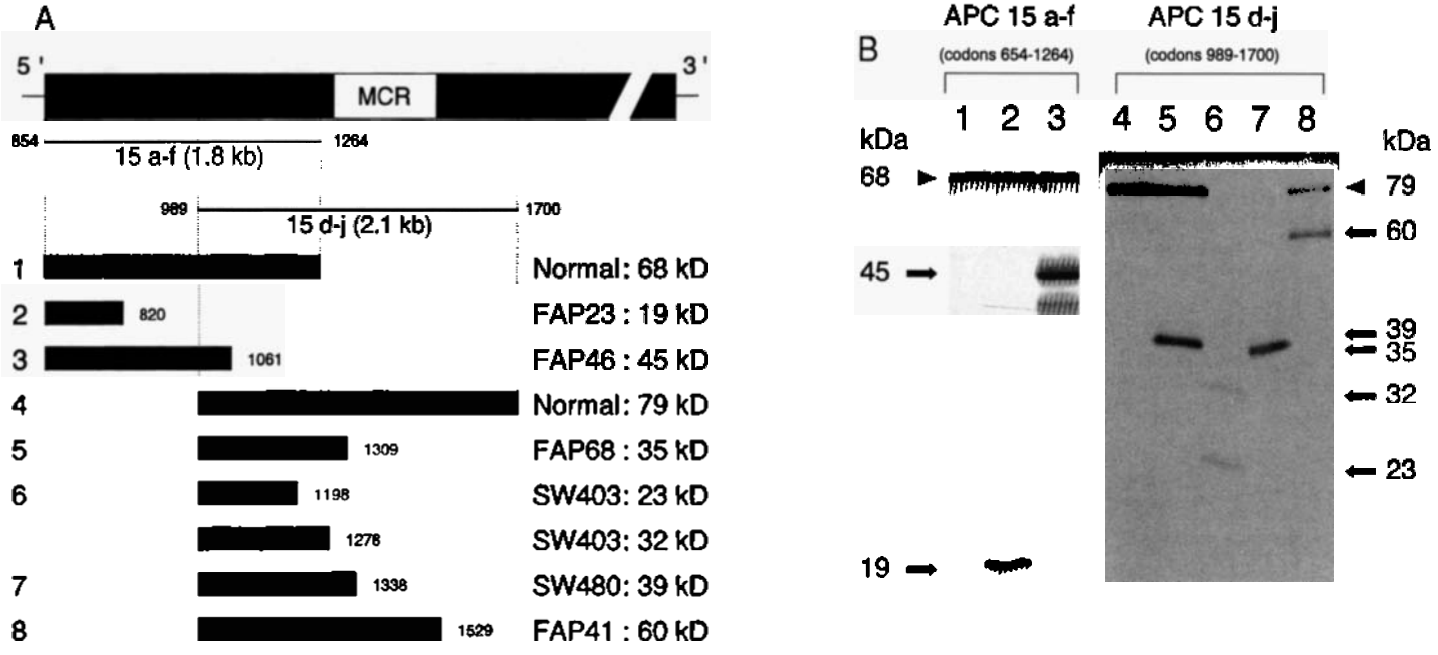


FIG. 3 Mutation detection in the *APC* gene by the protein truncation test (PTT). (**A**) Schematic representation of *APC* exon 15 (black bar, top), in which the location of the two amplified segments 15a-f and 15d-j and the mutation cluster region (MCR) are indicated. Below the exon, translation products are drawn as black bars, at the right of which the position of the stop codon, name of the sample, and the size of the expected protein product are shown. The samples are as follows: Normal controls (1 and 4), Dutch patients with FAP (2, 3, 5, and 8), and well-known colorectal cancer-derived cell lines (6 and 7). The numbers on the left refer to the gel lanes in (B). (**B**) SDS-PAGE analysis of translation products from the 1.8-kb *APC* exon 15 segment a-f (lanes 1–3) and the 2.1-kb segment 15d-j (lanes 4–8). The sizes of full-length (arrowheads) and truncated translation products (arrows) are indicated. Lane 1 contains the translation product obtained by PTT of exon 15a-f in a normal control, while lanes 2 and 3 show the products from samples with truncating mutations. In addition to the normal and truncated proteins, a translation product due to internal initiation is seen in lane 3. In lane 4, the product of PTT analysis of exon 15d-j in a normal control is shown, while *in vitro*-synthesized truncated proteins of various sizes can be seen in lanes 5–8. Note that in two of the samples no full-length translation product is present: the cell line SW403 (lane 6) has two truncated proteins and SW480 (lane 7) has a mutated allele, while the other allele is lost.

d-j in SW480 (Fig. 3B, lane 7) revealed a shortened polypeptide of 39 kDa and, as expected, no wild-type APC product. In patient 41, PTT analysis showed a truncated protein of 60 kDa in addition to the wild-type product of 79 kDa (Fig. 3B, lane 8). The size of the truncated protein corresponded with the presence of a nonsense mutation at codon 1529 in this patient. In a normal control, PTT showed only the *in vitro*-synthesized protein of 79 kDa produced by runoff translation from the wild-type exon 15d–j fragment (Fig. 3B, lane 4).

Advantages and Limitations of Technique

The PTT method has a number of advantages over other methods of direct mutation detection. Large coding segments (of 2– kb) can be screened for the presence of mutations in a single assay. In the case of *APC,* the entire coding region (8.5 kb) can be analyzed in five protein truncation tests using overlapping primer pairs. Mutations that result in the termination of translation are selectively detected by the PTT, while missene mutations and polymorphisms due to silent mutations are not detected. In contrast to the other methods such as DGGE, SSCP, and RPA, sequence analysis is not strictly necessary to confirm the mutation. The mutation analysis of disease genes such as *APC,* which are characterized by a predominance of truncating mutations, is greatly simplified by the PTT method. If further characterization of the mutation by sequencing is desired, this analysis will be facilitated by the size of the truncated protein, which indicates the approximate position of the stop codon.

A limitation of the PTT method for its general use in human genetic diseases is the fact that in its present form, in which translation products are resolved by conventional SDS-PAGE analysis, the PTT does not detect missense mutations. However, other methods of protein separation (such as isoelectric focusing) may be used to extend the detection range to this type of mutations. For the PTT analysis of most disease genes, RNA is required. In those laboratories where DNA-based screening methods have so far been employed to perform the molecular diagnostics, the introduction of the PTT as a screening method may require resampling of individuals. Therefore, it is evident that the feasibility of setting up the PTT as a screening procedure for a particular genetic disorder will depend on the availability of RNA samples. Finally, the use of radioactivity may represent a disadvantage of the PTT. The possible development of nonradioactive protocols for the PTT, in which proteins are detected by fluorescence or chemiluminescence, may alleviate this problem.

General Applicability of Protein Truncation Test

Apart from its use for the mutation analysis of the *APC* and *DMD* genes, the PTT has proven to be applicable to a variety of human genetic diseases. A summary of

TABLE IV Applications of Protein Truncation Test or *in Vitro*-Synthesized Protein Assay in Human Molecular Genetics

Disease	Gene	Length ORF (kb)	Number of exons	Template	Number of segments	Refs.[a]
Duchenne muscular dystrophy	*DMD*	12	79	RNA	5–10	1
Familial adenomatous polyposis	*APC*	8.5	15	RNA, DNA	5	2, 3
Hereditary nonpolyposis colorectal cancer	*hMSH2*	2.8	16	RNA	2	4
	hMLH1	2.3	19	RNA	2	5
	hPMS1	2.8	?	RNA	3	6
	hPMS2	2.6	?	RNA	2	6
Hunter syndrome	*IDS*	1.7	9	RNA	1	7
Neurofibromatosis-1	*NF1*	8.4	>56	RNA	5	8, 9
Neurofibromatosis-2	*NF2*	1.8	17	RNA	1	10
Polycystic kidney disease	*PKD1*	4.8	?	RNA	?	11
Breast–ovarian cancer	*BRCA1*	5.6	22	RNA, DNA	5	12

[a] *Key to references:* (1) P. A. M. Roest, R. G. Roberts, S. Sugino, G.-J. B. van Ommen, and J. T. den Dunnen, *Hum. Mol. Genet.* **2,** 1719 (1993); (2) S. M. Powell, G. M. Petersen, A. J. Krush, S. Booker, J. Jen, F. M. Giardiello, S. R. Hamilton, B. Vogelstein, and K. W. Kinzler, *N. Engl. J. Med.* **329,** 1982 (1993); (3) R. van der Luijt, P. Meera Khan, H. Vasen, C. van Leeuwen, C. Tops, P. Roest, J. den Dunnen, and R. Fodde, *Genomics* **20,** 1 (1994); (4) B. Liu, R. E. Parsons, S. R. Hamilton, G. M. Petersen, H. T. Lynch, P. Watson, S. Markowitz, J. K. V. Willson, J. Green, A. de la Chapelle, K. W. Kinzler, and B. Vogelstein, *Cancer Res.* **54,** 4590 (1994); (5) N. Papadopoulos, N. Nicolaides, Y.-F. Wei, S. M. Ruben, K. C. Carter, C. A. Rosen, W. A. Haseltine, R. D. Fleischmann, C. M. Fraser, M. D. Adams, J. C. Venter, S. R. Hamilton, G. M. Petersen, P. Watson, H. T. Lynch, P. Peltomäki, J.-P. Mecklin, A. de la Chapelle, K. W. Kinzler, and B. Vogelstein, *Science* **263,** 1625 (1994); (6) N. C. Nicolaides, N. Papadopoulos, B. Liu, Y.-F. Wei, K. C. Carter, S. M. Ruben, C. A. Rosen, W. A. Haseltine, R. D. Fleischmann, C. M. Fraser, M. D. Adams, J. C. Venter, M. G. Dunlop, S. R. Hamilton, G. M. Petersen, A. de la Chapelle, B. Vogelstein, and K. W. Kinzler, *Nature (London)* **371,** 75 (1994); (7) F. B. L. Hogervorst, A. C. van der Tuijn, B. Poorthuis, W. Kleyer, E. Bakker, G. J. B. van Ommen, and J. T. den Dunnen, *Am. J. Hum. Genet.* **55** (Suppl.), 1303 (1994); (8) R. Cawthon and H. D. Breidenbach, *Am. J. Hum. Genet.* **55** (Suppl.), 1256 (1994); (9) R. A. Heim, L. M. Silverman, R. A. Farber, L. N. W. Kam-Morgan, and M. C. Luce, *Nature Genet.* **8,** 218 (1994); (10) K. Pulaski, W. Pettingell, M. MacCollin, and J. F. Gusella, *Am. J. Hum. Genet.* **55** (Suppl.), 1383 (1994); (11) J. H. Roelfsema, L. Spruit, C. J. Ward, G.-J. B. van Ommen, D. J. M. Peters, P. C. Harris, and M. H. Breuning, *Am. J. Hum. Genet.* **55** (Suppl.), 1400 (1994); (12) F. B. L. Hogervorst, R. S. Cornelis, M. Bout, M. van Vliet, J. C. Oosterwijk, R. Olmer, E. Bakker, J. G. M. Klijn, H. F. A. Vasen, E. J. Meijers-Heijboer, F. H. Menko, C. J. Cornelisse, J. T. den Dunnen, P. Devilee, and G. J. B. van Ommen, *Nature Genet.*, in press (1995).

applications of the protein truncation methods in human molecular genetics is presented in Table IV. The PTT is used in several inherited forms of cancer, neuromuscular disorders, inborn errors of metabolism, and other genetic defects. For most disease genes, templates for the PTT are generated by the RT-PCR-based procedure, which uses RNA as the starting material. The *APC* gene as well as the *BRCA1* gene

contain a large exon, encompassing 76 and 61% of the corresponding proteins, respectively. Therefore, genomic DNA samples can be employed to screen for mutations in a large part of these two genes, which greatly simplifies the analysis and allows the use of stored DNA samples. To extend the analysis to the entire coding region, the smaller exons of the *APC* and *BRCA1* genes are screened by the RT-PCR-based PTT procedure.

References

1. R. van der Luijt, P. Meera Khan, H. Vasen, C. van Leeuwen, C. Tops, P. Roest, J. den Dunnen, and R. Fodde, *Genomics* 20, 1 (1994).
2. S. M. Powell, G. M. Petersen, A. J. Krush, S. Booker, J. Jen, F. M. Giardiello, S. R. Hamilton, B. Vogelstein, and K. W. Kinzler, *N. Engl. J. Med.* 329, 1982 (1993).
3. J. Groden, A. Thliveris, W. Samowitz, M. Carlson, L. Gelbert, H. Albertsen, G. Joslyn, J. Stevens, L. Spirio, M. Robertson, L. Sargeant, K. Krapcho, E. Wolff, R. Burt, J. P. Hughes, J. Warrington, J. McPherson, J. Wasmuth, D. Le Paslier, H. Abderrahim, D. Cohen, M. Leppert, and R. White, *Cell* 66, 589 (1991).
4. Y. Miyoshi, H. Ando, H. Nagase, I. Nishisho, A. Horii, Y. Miki, T. Mori, J. Utsunomiya, S. Baba, G. Petersen, S. R. Hamilton, K. W. Kinzler, B. Vogelstein, and Y. Nakamura, *Proc. Natl. Acad. Sci.* USA 89, 4452 (1992).
5. R. Fodde, R. van der Luijt, J. Wijnen, C. Tops, H. van der Klift, I. van Leeuwen-Cornelisse, G. Griffioen, H. Vasen, and P. Meera Khan, *Genomics* 13, 1162 (1992).
6. H. Nagase and Y. Nakamura, *Hum. Mutat.* 2, 425 (1993).
7. P. A. M. Roest, R. G. Roberts, S. Sugino, G.-J. B. van Ommen, and J. T. den Dunnen, *Hum. Mol. Genet.* 2, 1719 (1993).
8. F. B. L. Hogervorst, R. S. Cornelis, M. Bout, M. van Vliet, J. C. Oosterwijk, R. Olmer, E. Bakker, J. G. M. Klijn, H. F. A. Vasen, E. J. Meijers-Heijboer, F. H. Menko, C. J. Cornelisse, J. T. den Dunnen, P. Devilee, and G. J. B. van Ommen, *Nature Genet., 10,* 208 (1995).
9. G. Sarkar and S. S. Sommer, *Science* 244, 331 (1989).
10. M. Kozak, *Nucleic Acids Res.* 12, 857 (1984).
11. M. Kozak, *Mol. Cell. Biol.* 8, 2737, (1988).
12. U. K. Laemmli, *Nature (London)* 277, 680 (1970).
13. R. A. Laskey and A. D. Mills, *Eur. J. Biochem.* 56, 335 (1975).

[7] Localization of Tumor Suppressor Genes by Loss of Heterozygosity and Homozygous Deletion Analysis

Margaret A. Knowles

Introduction

Tumor suppressor genes (TSGs) are defined as genes whose functional inactivation contributes to tumor development. A series of TSGs (*RB, WT1, TP53, NF1, NF2, DCC,* and *APC*) with widely differing functions have been isolated. In the examples studied to date, inactivation of both copies of a TSG in a somatic cell are required for tumorigenesis. This may occur by epigenetic or genetic mechanisms, but in most cases mutations have been identified in both alleles of a TSG in relevant tumors (1).

The initial observation in cases of familial retinoblastoma, that some patients showed a constitutional deletion at 13q14, led to the suggestion that the inherited predisposition to the disease resulted from mutational inactivation of one allele of a gene at 13q14 and that subsequent somatic inactivation of the second allele at this locus in any retinal cell led to tumor development. This so-called "two-hit" hypothesis for tumor formation proposed by Knudson (2) has since been shown to be applicable to *RB* and to other TSGs. In familial cancers, a mutation in one allele is inherited via the germ line and the second allele sustains a mutation later in life as the result of spontaneous or carcinogen-induced somatic damage. In sporadic tumors, mutations to both alleles occur in a somatic cell.

One of the early observations made on retinoblastomas was that one allele, typically that carrying the inherited mutation, showed a small genetic alteration, often a point mutation, but that the second mutation usually involved a more gross genetic change, commonly a large deletion that could be detected by standard molecular and cytogenetic techniques. Studies of sporadic tumors have revealed gross genetic deletions at the *RB* locus and at other locations in the genome, some specific to particular tumor types. It was postulated that common regions of deletion identified the likely location of TSGs and this was confirmed when novel genes showing inactivation of both alleles by mutation were isolated from such regions in familial cancers, for example, *NF1* and *APC* from chromosomes 17q and 5q in neurofibromatosis type I and familial adenomatous polyposis, respectively (3, 4). The likely explanation for one small and one gross genetic inactivating event is that large deletions on both homologs would inactivate many genes and this would usually be lethal. Very occasionally, deletions can be detected on both alleles. Such homozygous deletions may occur within a TSG or are restricted to the immediate vicinity of such a gene, and in

Methods in Molecular Genetics, Volume 8

the only case described to date where such deletions have been identified at high frequency (in the region of the *CDKN2* locus at 9p21) may identify a region of low gene density.

Following mutation to one allele of a TSG, the cell is heterozygous for the mutation and subsequent deletion of the second allele generates homozygosity for this mutation. The identification of human genetic polymorphisms has allowed both parental copies of specific genetic sequences to be identified. Thus, in an individual in whom two alleles can be identified at a particular locus (an informative individual), gross deletion of one allele can be detected as loss of heterozygosity (LOH). Figure 1 illustrates a general model for LOH in somatic cells of a normal individual.

In the past ten years, LOH has been detected in all tumor types studied. Because tumor development is a multistep process believed to result from accumulation of mutations over a prolonged period, it comes as no surprise that common regions of LOH are present on several chromosomes in the majority of tumors studied. Our ability to detect LOH, a hallmark of TSG mutation, has improved dramatically with the characterization of many thousands of human polymorphisms and their precise mapping as the direct result of the Human Genome Project.

Today the search for novel TSGs involved in the development of the majority of sporadic cancers in adults is based largely on the identification of common regions of chromosomal deletion and subsequent precise mapping of the minimum regions of deletion by LOH analysis. Such mapping provides a starting point for isolation of candidate TSGs from within the region by positional cloning strategies. This chapter describes molecular techniques commonly used to identify LOH and homozygous deletions in human tumors.

Human Genetic Polymorphisms

Until 1985, DNA polymorphisms identified were essentially limited to restriction fragment length polymorphisms (RFLPs), most of which were two-allele systems with low information content. The identification of minisatellite and variable number of tandem repeat (VNTR) polymorphisms in 1985 and 1987, respectively (5, 6), provided more highly polymorphic markers with multiple alleles generated by the presence of tandemly repeated core sequences of 10–35 bases. Although these markers commonly show frequencies of heterozygosity of >70%, the large size of some alleles precludes polymerase chain reaction (PCR) analysis and these markers are unevenly distributed, many VNTRs being clustered in the proterminal regions of human chromosomes. The description in 1989 by several laboratories (7–10) of highly polymorphic simple sequence repeats or microsatellites has revolutionized human genetics. Microsatellites consist of ~10–50 copies of simple sequence motifs of 1–6 bp that can occur in perfect tandem repetition, as imperfect (interrupted) repeats, or as more complex combinations with another type of repeat (11). It has been estimated that dinucleotide repeats of the type $(dC\text{-}dA)_n\cdot(dG\text{-}dT)_n$ occur on average

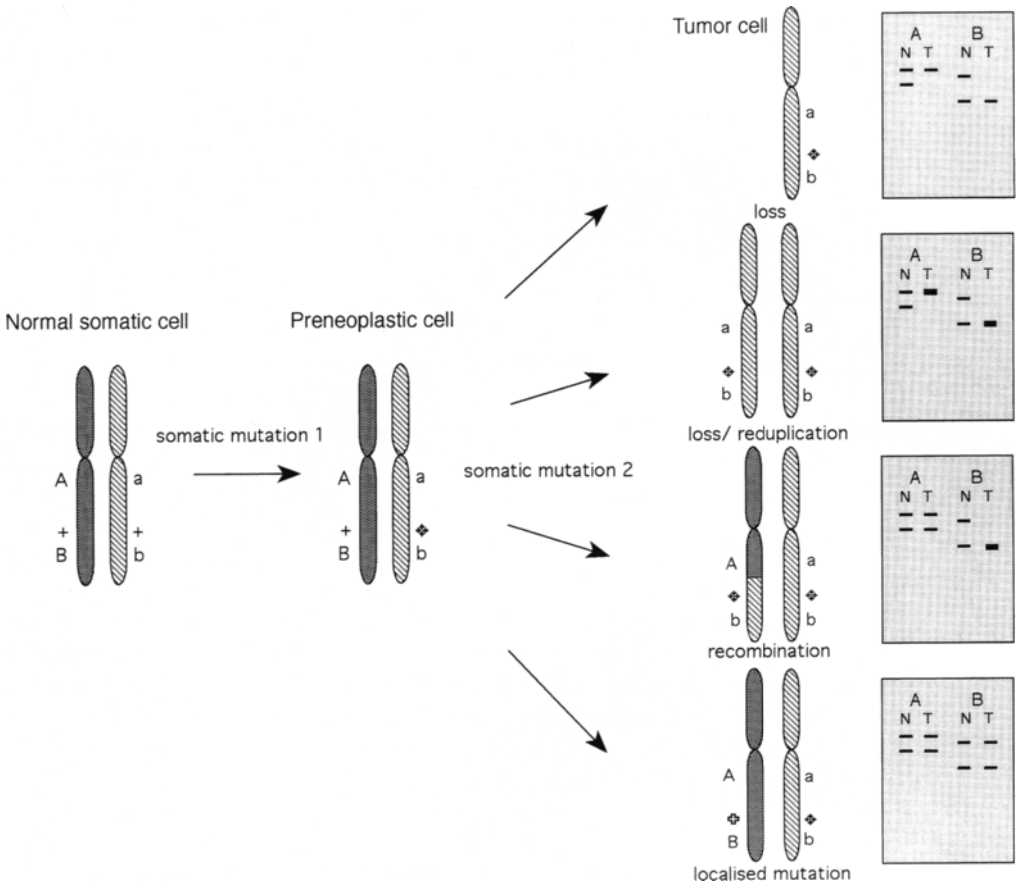


FIG. 1 Mechanisms for loss of constitutional heterozygosity in somatic cells. *Left:* Two parental alleles are shown at three loci in a somatic cell (A, B, and a tumor suppressor locus designated + at which both alleles are wild type). The individual is heterozygous at two loci (*Aa, Bb*). A point mutation in one allele at the tumor suppressor locus (❖) gives rise to a preneoplastic cell in which no LOH can be detected. Loss or mutation (✣) of the wild-type allele can be achieved in several ways (right), most of which result in LOH at one or both of the informative loci nearby. *Far right:* Autoradiograms of Southern blots or sequencing gels of microsatellite markers A and B. In addition to detection of LOH, differences in the representation of the two alleles are detected in the case of loss of one allele and reduplication of the other and in the case of some recombination events.

once every 30 kb in human genomic DNA, distributed evenly throughout 3′ and 5′ untranslated regions of genes and introns. Tri- and tetranucleotide repeats are less frequent (~1/100 kb) but are easier to type than the dinucleotide repeats. Many microsatellite markers have polymorphic information content (PIC) >0.80. It ap-

pears that the mutation rates at these loci are low (12), permitting their use for many types of genetic study including LOH analysis. In 1994, the linkage map of the human genome consisted of 5840 polymorphic loci covering the map at an average density of 1 per 0.7 centimorgan (cM). Of these markers, 3617 are PCR-formated microsatellite polymorphisms, 427 are genes, and the remainder are mostly RFLPs of anonymous DNA segments (13). All types of polymorphism have been used for LOH analysis and many studies still use a combination of RFLPs, VNTRs, and microsatellites, all of which have some advantages and some problems, as described below.

Extraction of DNA from Tumor Tissue and Blood Leukocytes

All LOH studies require paired samples of constitutional and tumor DNA from a set of patients. For Southern analysis this must be of high quality. For PCR analysis, partially degraded samples will give results but it is desirable to collect and store pure DNA samples of high molecular weight in readiness for later studies of candidate suppressor genes.

Most laboratories handle large quantities of cloned DNA (e.g., plasmids) and it must be emphasized that contamination of genomic DNA with such material is easily achieved. Because PCR is extremely sensitive, a very few contaminating copies of a cloned sequence can give misleading results. Similarly, the presence of contaminating plasmids on a Southern blot will be detected by most probes used. Once in a sample, such contaminants cannot be removed, and for this reason all manipulations involving genomic DNA should be carried out in a designated area away from those used for handling plasmids and using equipment (e.g., micropipettes and pipetters) dedicated to genomic DNA manipulation only.

Tumor tissue should be extracted immediately or frozen as soon as possible after removal from the patient. Ideally this should be done in liquid nitrogen or on dry ice so that samples will be suitable for analyses of DNA, RNA, and protein. However, if only DNA is to be extracted, we have found that placement of small fragments in closed, dry containers in a standard domestic −20°C freezer is sufficient and DNA integrity is preserved. Ten to 20 ml of venous blood or a sample of normal tissue should be collected as a source of constitutional DNA. Blood should be collected into EDTA tubes (not heparin, as this interferes with DNA digestion) and extracted or frozen immediately. This can also be done at −20°C if no other means is available.

Thaw frozen blood samples on a rotary mixer at room temperature for about 1 hr, pour into 50-ml tubes (Falcon 2070), add 30 ml of solution A [0.32 *M* sucrose, 10 m*M* Tris base (pH 7.5), 1% (v/v) Triton X-100; this should be autoclaved before use and stored at 4°C], and leave on ice for 5 min. Spin for 10 min at 4000 rpm at room temperature. Pour the supernatant from the white cell pellet, add 2.8 ml of solution B [0.075 *M* NaCl, 0.024 *M* EDTA (pH 8.0)], sodium dodecyl sulfate (SDS)

to a final concentration of 1% (w/v), and proteinase K to a final concentration of 100 μg/ml. Incubate at 37°C overnight or at 50–55°C for 3–4 hr. Addition of more proteinase K and incubation at the higher temperature should be used if tissue is not solubilized after overnight incubation. Extract with an equal volume of phenol:water:chloroform (68:18:14) for 10–15 min on a rotator, spin for 10 min at 4000 rpm at room temperature, and remove the aqueous layer containing the DNA. At this stage the interface will be dirty and this should be removed with the aqueous phase. Extract with phenol–chloroform one to three more times until the interface is clear, leaving any material at the interface after these extractions. Finally, extract with an equal volume of chloroform containing 5% (v/v) isoamyl alcohol. Precipitate DNA by adding 1/10 vol of 3 *M* sodium acetate, pH 5.5, and 2 vol of absolute ethanol (at −20°C). High molecular weight DNA can be spooled from this mixture using a heat-sealed glass Pasteur pipette and air dried on the glass for 10–15 min. The tip of the Pasteur pipette with the adherent DNA can be snapped off and placed in a 1.5-ml tube and the DNA dissolved in 1× TE (10 m*M* Tris-Cl, 1 m*M* EDTA) at 4°C for at least 24 hr before quantitation. To extract DNA from tumor tissue, place tissue in a sterile petri dish, observe under a dissecting microscope, and dissect viable tumor tissue for extraction. The ease with which this can be done depends on the tissue and may be better carried out at the time of surgery. Chop the tissue, using scalpels or curved scissors, to give fragments not exceeding 1 mm^3. Place the tissue (~500 mg) into a Falcon 2059 tube (15 ml), add 2.8 ml of solution B, SDS, and proteinase K and proceed as for blood. Long-term stocks of extracted DNA are stored at −20 or −70°C. DNA for immediate use should be stored at 4°C. Repeated freeze–thawing should be avoided.

Analysis of Loss of Heterozygosity by Southern Blotting

Before the advent of PCR, Southern blot analysis provided the mainstay for most studies of human polymorphisms. Southern blotting is the transfer of DNA fragments from an agarose gel to a membrane support. After blotting, the membrane carries a representative sample of the DNA fragments that were present in the gel. Many alternative methods exist for Southern blotting using different types of membrane and different methods of DNA transfer and immobilization on the membrane. The procedures described below represent a standard method for capillary transfer onto uncharged nylon membranes. This generates blots suitable for multiple reprobings and requires no specialized apparatus.

Digestion of Genomic DNA and Agarose Gel Electrophoresis

Because a large amount of DNA (5–10 μg) is used per track, it is important to select restriction enzymes that will generate RFLPs at as many useful loci as possible and

then to rehybridize the blots with many (>10) probes in turn to obtain the maximum amount of information. Some enzymes recognize more polymorphisms than others. Among these are enzymes such as *Taq*I and *Msp*I, which cut at sequences containing CpG dinucleotides where spontaneous deamination of a methylated cytosine to thymine can occur, making this a common site of point mutation. For a more detailed discussion, see Ref. 14.

In a total volume of 50 μl combine 5–10 μg of DNA, 10× restriction buffer (supplied by enzyme manufacturer), and 25–50 units of enzyme. Incubate at the appropriate temperature (overnight at 37°C or for 4 hr at 65°C). A small aliquot of the digest (e.g., 0.5 μg) can be electrophoresed in agarose to check that digestion is complete. Add 5 μl of 10× loading buffer [25% (v/v) Ficoll, 0.4% (w/v) bromphenol blue, 0.4% (v/v) xylene cyanol in 1× TBE or 1× TAE] to each sample, mix, and load onto 0.8–2% (w/v) agarose gels in 1× TAE or 1× TBE (1× TAE is 40 m*M* Tris–acetate–1 m*M* EDTA; 1× TBE is 90 m*M* Tris–borate–2 m*M* EDTA) containing ethidium bromide (1 μg/ml). Large-format gels, for example, the Life Technologies (Paisley, Scotland) Horizon 20.25 system, which uses 300 ml of agarose and provides 20 wells, are ideal. Samples should be loaded in pairs (blood, tumor) from the same patient and quantitation should be as accurate as possible for ease of interpretation of results. Depending on the size of fragment to be separated and the concentration of the gel, run in 1× TAE or 1× TBE at 25–30 V for 24 to 48 hr. Shorter running times at higher voltage will decrease the resolution of the gel. Include radioactively labeled size markers (e.g., *Hin*dIII fragments of λ 3′ end-labeled by in-filling with Klenow fragment of DNA polymerase I in the presence of [α-^{32}P]dATP) at 5000 cpm/track to provide a useful size marker on the blot that will be visible for several hybridizations. End-labeled size markers have the advantage that, unlike internally labeled fragments, they are stable for several months at −20°C. Another lane containing 2 μg of unlabeled DNA size markers provides a means to determine when the gel has run the desired distance. Following electrophoresis, photograph the gel under ultraviolet (UV) illumination (254 nm) to give a record of gel loading, completeness of digestion, and possible DNA degradation.

Transfer of DNA to Membrane Support

Rinse the gel with distilled water. If large fragments of DNA (>10 kb) are to be transferred, place in 0.25 *M* HCl and agitate gently on a platform shaker for 20–30 min. This partially depurinates the DNA, which causes strand breakage and facilitates transfer. Rinse in distilled water, cover with denaturing solution (1.5 *M* NaCl, 0.5 *M* NaOH), and agitate for 15 min. Repeat twice, leaving the final solution for 30 min. Replace the solution with neutralizing solution [1.5 *M* NaCl, 0.5 *M* Tris-Cl (pH 7.2), 1 m*M* EDTA] and incubate for 30 min. Repeat twice. Drain the gel and set up the transfer as shown in Fig. 2. Cut membrane and three sheets of What-

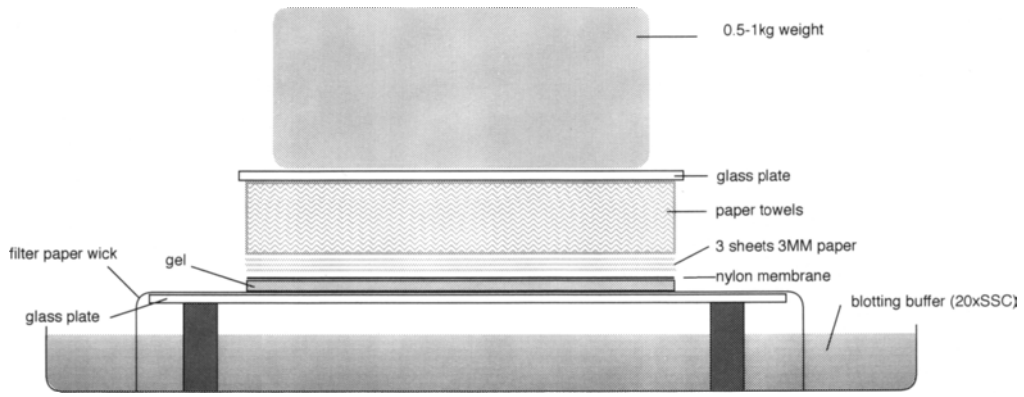


FIG. 2 Apparatus for capillary Southern transfer.

man (Clifton, NJ) 3MM paper to the same size as the gel. Cut filter paper to act as a wick between the buffer reservoir and the gel, wet the paper, and fill the reservoir with 20× SSC (1× SSC is 150 m*M* NaCl, 15 m*M* trisodium citrate). Place the gel on this and the membrane (e.g., Hybond N; Amersham, Arlington Heights, IL) on top. One corner of the membrane should be marked, for example, with indelible fine marker pen or pencil, for later orientation. Care should be taken to ensure that there are no air bubbles between the wick and the gel or between the gel and the membrane. These can be removed by rolling a pipette over the surface. On top of the membrane place a sheet of Whatman 3MM paper prewetted in 2× SSC, followed by three dry sheets and paper towels or other absorbent pad (4–5 cm high). Place a sheet of glass on top and compress with a weight of approximately 0.5–1 kg. Excess weight will crush the gel and distort the DNA bands. The edges of the gel should be covered with Saran Wrap or strips of used X-ray film to ensure that the buffer flows through and not around the gel. Leave overnight or up to 48 hr if the agarose concentration is >1% and large fragments are desired. After transfer, remove the towels and filter papers and briefly rinse the membrane in 2× SSC to remove any adherent agarose. At this stage the DNA is adherent to the membrane but requires permanent fixation. This is done by wrapping the membrane in Saran Wrap and placing DNA side down on a UV transilluminator (254-nm wavelength). The optimal length of irradiation required on each individual transilluminator should be determined and rechecked periodically, as too little exposure will not cross-link all DNA and excessive exposure appears to cause DNA degradation leading to reduced hybridization signals. This can be done with several strips of membrane onto which dots of a test DNA (e.g., a plasmid) are made at a series of concentrations in the nanogram range. These should initially be dried and then dampened in 2× SSC and wrapped in Saran Wrap. Strips should be irradiated for different periods of time and all hybridized together to a suitable probe.

Hybridization

Place a membrane in a heat-sealable plastic bag or glass hybridization tube and add prehybridization mix prewarmed to 65°C. Prehybridization mix, which should be made fresh, contains 6× SSC, 5× Denhardt's solution [100× Denhardt's is 2% (w/v) bovine serum albumin (BSA), 2% (w/v) Ficoll, 2% (w/v) polyvinylpyrrolidone (PVP)], and 0.5% (w/v) SDS. The 100× Denhardt's can be stored at −20°C. Denature sonicated nonhomologous DNA (e.g., calf thymus or salmon sperm) by heating to 100°C for 5 min, chill on ice, and add to prehybridization mix at a final concentration of 25 μg/ml. Prehybridize with shaking or rotation for at least 1 hr at 65°C. Hybridization mix is as for prehybridization but with the optional addition of 10% (w/v) dextran sulfate, which is added from a 50% stock stored at 4°C. This decreases the hybridization time required. Probes should be labeled with [α-^{32}P]dNTP to high specific activity (10^7–10^8 cpm/μg) using random priming (15) and separated from unincorporated dNTPs by precipitation or preferably by passage through a small Sephadex G-50 column. Denature the probe by boiling for 5 min, chill on ice, and add to hybridization mix at $\simeq 10^6$ counts per minute (cpm)/ml. Heterologous DNA is added as in prehybridization and can be denatured with the probe. Probes that contain repetitive human DNA sequences require blocking with total human DNA. In this case, add 1–2 mg of total human DNA to the probe before boiling, chill on ice, and incubate at 65°C for 10–30 min before adding to the hybridization mixture. Hybridize with shaking or rotation at 65°C for >12 hr. Carefully discard the hybridization mix. Carry out the following washes at 65°C using preheated solutions: two washes in 2× SSC for 15 min, one wash in 2× SSC containing 0.1% (w/v) SDS for 30 min, and one wash in 0.1× SSC for 10 min. Reduction of the salt concentration and/or increase in washing temperature increases the stringency of the wash. If hybridization to nonidentical sequences is present after these washes, increase the final wash temperature or incubate in the final wash for longer. Do not allow the membrane to dry out at any stage. After the final wash, blot the membrane on Whatman 3MM paper (do not dry completely) and seal in a plastic bag or wrap in Saran Wrap. Expose to X-ray film (typically for 24–48 hr) at −70°C with intensifying screens or generate a phosphorimage. Following exposure, the membrane can be stripped by placing in a solution of 0.1% (w/v) SDS, bringing to the boil and allowing to cool slowly at room temperature. Ascertain that all probe has been removed by autoradiography before rehybridization. Multiple preprobings (10 to 20) are usually possible. Store the membranes damp for short periods at 4°C or for long-term storage at −20°C.

Interpretation of Results

Results for the majority of RFLPs and VNTRs are usually straightforward to interpret. The Genome Database (GDB) lists allele sizes and frequencies. Details of access to GDB worldwide are given by the Cooperative Human Linkage Center

(CHLC) (13). Unless contamination of the tumor tissue with large amounts of normal stromal DNA is a problem, LOH is easy to score as a gross reduction in signal from one allele in the tumor sample. In the case of solid tumors, a small amount of signal, usually given the misnomer "residual signal," is seen from the lost allele and this comes from normal cells present in the sample. If several loci are studied in the same tumor this residual signal should be constant at all loci. Any deviation from this at loci where the signal from two alleles is clearly imbalanced must be scrutinized more carefully. One of the advantages of Southern blots compared with PCR-based methods is that the amount of signal observed on the film is, broadly speaking, proportional to the amount of template present on the gel. This is most closely approximated when preflashed film is used and exposures of films are such that bands are gray and not "burned in." An example of apparent differences in "residual signal" at loci along chromosome 8 is shown in Fig. 3 (15a). This tumor shows complete loss of

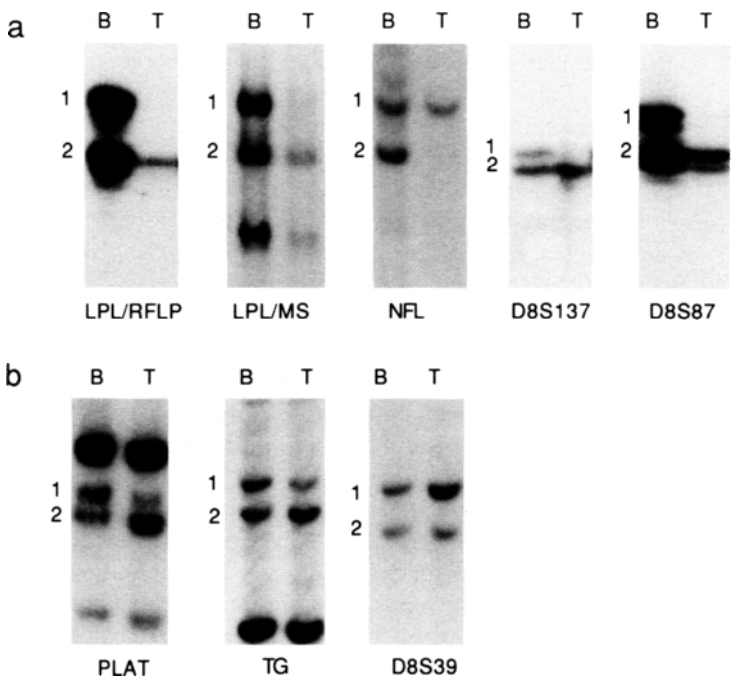


FIG. 3 Loss of heterozygosity at loci on chromosome 8 in a bladder carcinoma. *LPL*/RFLP, *NFL, PLAT, TG,* and *D9S39* were analyzed on Southern blots and *LPL*/MS, *D8S137,* and *D8S87* by microsatellite analysis. B, Blood leukocyte DNA; T, tumor DNA. (a) Clear LOH is seen at *LPL, NFL, D8S137,* and *D8S87.* (b) At *PLAT, TG,* and *D8S39,* a clear signal is seen from both alleles with amplification of signal from one allele, indicating retention of heterozygosity and the presence of several copies of one allele (for discussion, see text). (From Ref. 15a with permission.)

signal from one allele at *LPL, NFL, D8S137,* and *D8S87* (all on 8p). However, at *PLAT* (8p), *TG,* and *D8S39* (both on 8q), allele intensities are imbalanced but there is significant signal from both alleles. It is clear, given the low level of stromal contamination apparent from the results for 8p markers, that the weaker signal is not signal generated solely by normal stromal contamination. The result can be interpreted as retention of heterozygosity at these latter loci on proximal 8p and 8q but with a difference in the number of copies of these alleles. This type of discrepancy is easy to identify if multiple loci are assessed in the same tumor. However, if only one marker was used, this might be scored as LOH on 8q unless the amount of stromal contamination was known, for example, if DNA had been extracted from dissected frozen sections or careful loading controls were carried out on the blot. In the case shown in Fig. 3, the total amount of DNA present in the tumor tracks for *PLAT, TG,* and *D8S39* was lower than in the corresponding constitutional DNA track. The correct interpretation of this result is therefore an overrepresentation of one allele at *PLAT, TG,* and *D8S39.* A single hybridization of each blot to a nonpolymorphic probe detecting a single-copy fragment of 3–6 kb (which will be less affected by DNA degradation or incomplete digestion than larger or smaller fragments, respectively) will give an adequate loading control. Such a sequence should be selected from a region of the genome where no alterations have been identified in the tumor type in question. Quantitation of signal from each allele and from a loading control probe can be most accurately achieved by computer-assisted quantitation of a phosphorimage or a scanned image of an autoradiograph. We use the ImageQuant software supplied by Molecular Dynamics (Sevenoaks, Kent, UK) for use with their Phosphorimager to analyze both phosphorimages and scanned images of autoradiograms for this purpose.

Many publications refer to "allelic imbalance" rather than LOH where loading controls have not been carried out on all samples or when PCR has been used. Clearly, overrepresentation of an allele may suggest amplification of a particular chromosomal region and this would more likely indicate the presence of a dominantly acting target such as a protooncogene. It is important therefore to obtain a good idea of the amount of stromal contamination in the tumor samples, to keep a good photographic record of the appearance of the ethidium bromide-stained gels, and to carry out control hybridizations if possible. If this is not done, a so-called "deletion mapping" exercise could result in the identification of a novel oncogene!

Analysis of Loss of Heterozygosity Following Polymerase Chain Reaction Amplification of Genomic Fragments

One of the major drawbacks of LOH analysis by Southern blotting, apart from the low level of heterozygosity of the majority of markers, is the requirement for a large amounts of DNA from each tumor sample. In contrast, PCR may be used to amplify

products from extremely small amounts of template DNA. Many of the polymorphisms detectable by Southern blotting are theoretically detectable by PCR. For example, fragments containing RFLPs may be amplified, cut with appropriate restriction enzymes, and the resulting allele-specific products analyzed by agarose or polyacrylamide gel electrophoresis. The major restriction is that genomic sequence information is required to allow synthesis of appropriate PCR primers. Detailed information on numerous RFLP, VNTR, and minisatellite polymorphisms that can be analyzed by PCR can be found in the GDB. Several methods can be used to detect the alleles generated. Methods for direct visualization of PCR-generated RFLP and VNTR products and for detection by autoradiography are detailed below.

Polymerase Chain Reaction Amplification, Generation of Restriction Fragment Length Polymorphisms and Analysis by Agarose or Polyacrylamide Gel Electrophoresis

For RFLPs and VNTRs, agarose or polyacrylamide gel electrophoresis with ethidium bromide staining is commonly used. This is simple and rapid but it is less easy to quantitate the products than when using methods involving ^{32}P incorporation followed by autoradiography or phosphorimaging. However, if stromal contamination of the sample is known to be minimal, this method will yield good results.

Polymerase Chain Reaction

Because contamination with only a few copies of an unwanted template or primers can generate spurious results by PCR, stringent precautions should be taken to avoid this. A clean working area, careful aseptic technique using dedicated pipettes, preferably with plugged tips, the use of gloves, and control reactions on all runs are appropriate. Most markers can be amplified using the conditions described here. However, for new primers it is advisable to test a range of Mg^{2+} concentrations (e.g., 1–4 m*M*) and several annealing temperatures (e.g., from 5°C below to 5°C above the theoretical T_m of the primers).

Carry out reactions in a total volume of 50 μl, containing 50–100 ng of DNA template, 50 pmol of each primer, dNTPs at a final concentration of 200 μM each, and $MgCl_2$ at 1.5 m*M*. We have used amplification buffers (10×) supplied by Life Technologies, Amersham, and Advanced Biotechnologies (Leatherhead, UK) with similar results. To avoid pipetting inaccuracies, make a master mix containing 10× buffer, dNTPs, and primers and add this to the DNA template in 96-well plates or 0.5-ml Eppendorf tubes. Always include "no DNA" and "no primer" controls. Overlay with a drop of mineral oil. Use a "hot start" procedure to inactivate proteases which might degrade the polymerase and to avoid mispriming. Denature the samples for 10 min at 95°C, pause the PCR machine, add 1 unit of *Taq* DNA polymerase through the oil, and mix by pipetting several times. Amplify for 30 cycles (each cycle

consisting of 95°C for 1 min, 55°C for 1 min, and 72°C for 1.5 min) at the maximum ramp rate. After the final cycle, elongate for a further 10 min.

Digestion with Restriction Endonucleases

Check the efficiency of amplification by electrophoresis of 10 μl of the reactions through agarose. In general, 10–20 μl of a reaction is sufficient for restriction. If 10 μl is used, add 1.5 μl of the appropriate 10× reaction buffer and 5 units of restriction endonuclease, adjust the volume to 15 μl, and incubate at 37°C for 1–2 hr.

Electrophoresis

The products may be resolved by electrophoresis in agarose or polyacrylamide gels. The latter give superior resolution. For agarose gel electrophoresis add 1.5 μl of 10× loading buffer and load the entire reaction onto an agarose gel of appropriate concentration, containing ethidium bromide (1 μg/ml.) For optimum resolution for fragments in the range of 50–500 bp, we use 4% (w/v) NuSieve gels (FMC BioProducts, Rockland, ME) in 1× TAE.

Rapid results by polyacrylamide gel electrophoresis can be obtained using one of the small-format gel apparatuses [e.g., Bio-Rad (Richmond, CA) Mini-PROTEAN gel apparatus] and running conditions recommended by the supplier. Gel concentrations in the range of 8–15% will resolve most PCR products. Make nondenaturing polyacrylamide gels from 40% (w/v) acrylamide stock (19:1, acrylamide–bisacrylamide) in 1× TBE and polymerize by the addition of 70 μl of 10% (w/v) ammonium persulfate solution (made fresh daily) and 3.5 μl of *N,N,N′,N′*-tetramethylethylenediamine (TEMED) per 10 ml of final gel mix. Premixed acrylamide solutions sold by several manufacturers are suitable for these gels. Run 1 μg of cold size markers and stain the gels in ethidium bromide (1 μg/ml) in 1× TBE for 10–20 min or use a silver-staining kit to visualize the bands. Ethidium bromide-stained gels can be photographed and the negative used for densitometry if required. An example of analysis of an *Acc*II RFLP in exon 4 of *TP53* (16) is shown in Fig. 4a.

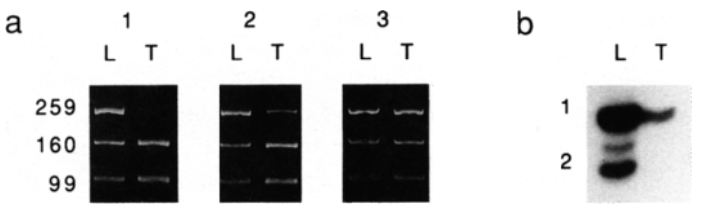


FIG. 4 (a) Loss of heterozygosity within the *TP53* gene detected by RFLP-PCR. *Acc*II detects a polymorphism in exon 4 of *TP53* (16). A1, 259 bp; A2, 160 bp + 99 bp. Tumor 1 shows clear LOH, tumor 2 shows allele imbalance, and tumor 3 shows retention of heterozygosity. L, Blood leukocyte DNA; T, tumor DNA template. Nondenaturing polyacrylamide gel, stained with ethidium bromide. (b) Loss of heterozygosity within the *Rb* gene detected by PCR of a VNTR in intron 20 (19). L, Leukocyte; T, tumor DNA template.

Polymerase Chain Reaction Amplification of Polymorphic Fragments Using ^{32}P Label and Analysis by Polyacrylamide Gel Electrophoresis

An alternative method by which to visualize PCR products is to incorporate ^{32}P during the PCR reaction to allow later detection by autoradiography or phosphorimaging. This can be done either by incorporation of an [α-^{32}P]dNTP or by end labeling of one of the PCR primers. The former method results in a product in which both DNA strands are labeled. The products may be resolved in a nondenaturing polyacrylamide gel and will give rise to a single double-stranded DNA (dsDNA) band. On a denaturing gel, the two complementary strands from a single duplex migrate at different speeds and each allele will typically be resolved as two bands. Generally the resolution of denaturing gels is superior but for most allele loss applications this does not present a problem and nondenaturing gels can be used. An advantage of incorporation of hot dNTPs is that if required, a product of very high specific activity can be generated and film exposure times are consequently short.

The use of one end-labeled primer generates a simple high-resolution image on a denaturing gel. This is highly desirable for LOH studies using microsatellite markers, where allele size differences may be 1 or 2 nucleotides. The only situation in which the lower specific activity of these products becomes limiting is when restricted numbers of PCR cycles are used, for example, to detect and/or confirm homozygous deletions (see below), and exposure times required become much longer (days rather than minutes).

Polymerase Chain Reaction Using [α-^{32}P]dCTP Incorporation

Because the detection of ^{32}P-labeled PCR products is highly sensitive, it is not essential to generate a large amount of product. Limiting amounts of template can be used (5–10 ng) and smaller numbers of PCR cycles can be used than when unlabeled product is to be visualized directly. As discussed above, for LOH analysis, it is important that the end result be directly proportional to the amount of template present so that direct comparisons can be made between constitutional and tumor DNAs. The PCR reaction is therefore adjusted to give a quantitative result. A detailed discussion of factors affecting the linear range of amplification is beyond the scope of this chapter (for more detailed information, see Refs. 17 and 18). Rigorous standardization and control of reactions such as can be achieved when a single target is the subject of study are not practicable when many primer pairs are to be used. However, we have found that because the linear range is greater for smaller amounts of template, restriction of template to 5–10 ng is helpful. Restriction of the number of cycles used to 25–27 ensures that product is sampled in the linear part of the curve. For less efficient primer sets 27 cycles can be used and for others 25 may be sufficient to generate a good signal after relatively short exposures. This becomes more critical for homozygous deletion analysis in tumor tissue with stromal contamination (see below).

Set up PCR reactions as described above, adding a master mix to templates in 96-well plates and using a hot-start procedure. Reactions of 10 μl are used. These contain 5 ng of template, 1× reaction buffer, 1.5 m*M* $MgCl_2$, dATP, dGTP, and dTTP (200 μM each), 20 μM dCTP, 2 pmol of each primer, 1 μCi of [α-^{32}P]dCTP, and 1 U *Taq* DNA polymerase. Carry out 25–27 cycles consisting of 95°C for 1 min, 55°C (or appropriate annealing temperature) for 1 min, and 72°C for 1.5 min, with a final elongation for 10 min at 72°C. This protocol generates a product of very high specific activity. A smaller amount of labeled dCTP may be used (e.g., 0.1 μCi). A lower incorporation can also be achieved by increasing the amount of cold dCTP in the reaction. However, this is not desirable because large amounts of unincorporated [α-^{32}P]dCTP are then present in the reaction and run into the lower buffer tank.

Electrophoresis in Nondenaturing Polyacrylamide Gels

Add 7.5 μl of formamide loading buffer [98% (v/v) formamide containing 10 m*M* EDTA, 0.05% (w/v) bromphenol blue, and 0.05% (w/v) xylene cyanol] to the PCR products and mix. Run 2–4 μl of the nondenatured products in a 5–10% (w/v) nondenaturing polyacrylamide (19:1) gel containing 1× TBE. To polymerize 100 ml of gel (suitable for a single 30 cm × 40 cm × 0.4 mm gel) use 700 μl of 10% (w/v) ammonium persulfate and 35 μl of TEMED. Allow the gel to polymerize at room temperature for at least 2 hr. Run the gel at 30 W at 4°C for 4–6 hr, according to product size and polyacrylamide concentration. Gels may also be run more slowly (e.g., 8–10 W) at room temperature overnight. It is important that the gel does not heat up, as this will cause "smiling" of the DNA bands and may melt the DNA duplexes. At the end of the run, remove the top plate, place a sheet of Whatman 3MM paper onto the gel, and lift the gel from the lower plate. Cover with Saran Wrap, trim to the size of the gel, and dry on a heated gel dryer (80°C for 1–1.5 hr). Place the dried gel in a film cassette and expose to X-ray film with an intensifying screen. Exposure times range from 15 min to several hours. Alternatively, make a phosphorimage of the gel. As for Southern blot hybridizations, the images obtained can be analyzed by densitometry.

Polymerase Chain Reaction of Microsatellite Markers Using One ^{32}P End-Labeled Primer

The method of choice for visualizing small polymorphic fragments, particularly microsatellites, is to label only one strand of the PCR product and to denature the product and run in a sequencing gel. This gives a high-resolution image in which differences in strand length of one nucleotide can be resolved.

Primers are 5′ end-labeled using T4 polynucleotide kinase and [γ-^{32}P]dATP. One of each primer pair should be labeled. This can be done the day before and primers stored overnight at −20°C. To label primer sufficient for 25 PCR reactions, combine (in a total volume of 12.5 μl) 50 pmol of primer, 1.25 μl of 10× kinase buffer

(supplied by the manufacturer), 3 U of T4 polynucleotide kinase, and 25 μCi of [γ-^{32}P]dATP. Incubate at 37°C for 45 min. Heat to 68°C for 10 min to inactivate the enzyme, chill on ice, and freeze if not required immediately.

The PCR reactions are similar to those used for [α-^{32}P]dNTP incorporation except that all dNTPs are used at 200 μM and 0.5 μl of labeled primer = 2 pmol and 2 pmol of the other, cold primer are used per reaction. Samples may be stored at −20°C for several days before the gels are run.

Electrophoresis in Denaturing (Sequencing) Gels

Add 7.5 μl of formamide loading buffer to each reaction, denature the PCR products by heating to 95°C for 5 min, chill on ice, and load 2–4 μl immediately onto 6% (w/v) denaturing gels. We have found that Seqagel XR premixed gel solution supplied by National Diagnostics (Atlanta, GA) gives excellent resolution. Load a labeled sequencing ladder to allow accurate sizing of alleles. For LOH analysis, sizing is not as critical as for linkage analysis, but it is a good idea to check the size and allele frequencies for each marker when used for the first time. This method can also be applied to larger DNA products generated at VNTR loci. Figure 4b is an auto

radiogram showing LOH at the *RB* locus detected by PCR of the VNTR region in intron 20 (19), using one end-labeled primer.

Interpretation of Results

Microsatellite analysis is rapidly becoming the method of choice for LOH studies. Because solid tumor tissues inevitably contain some normal cells that will contribute to the final PCR product, it is important, as discussed above, to terminate PCR reactions while the relationship between template and product is still linear. Figure 5b (15a) shows an example of a titration of two templates carried out to check that at levels of stromal contamination commonly found (e.g., up to 20%), LOH can be scored easily. It will be noted that a shadow or "stutter" band is present below the major product band in each case. Clearly, the stutter from the larger allele makes a contribution to the intensity of the smaller allele in this case. Nevertheless, using the conditions described above, we estimate that microsatellite typing is at least as sensitive as Southern blotting for the detection of LOH.

Some primer pairs, particularly those for some dinucleotide repeats, can show two or three significant "stutter" bands and these should be avoided if possible because interpretation of results is more difficult. It is thought that these bands are generated by slipped-strand mispairing (20). This mechanism results in deletions or insertions within a dinucleotide repeat region that always result in length changes of two nucleotides or multiples of two nucleotides. Possibly the greatest difficulty is experienced in distinguishing a homozygote (or tumor with LOH) from a heterozygote with alleles differing by a single dinucleotide repeat. If no stutter bands were present, this would be easy but their presence usually results in the appearance of three bands,

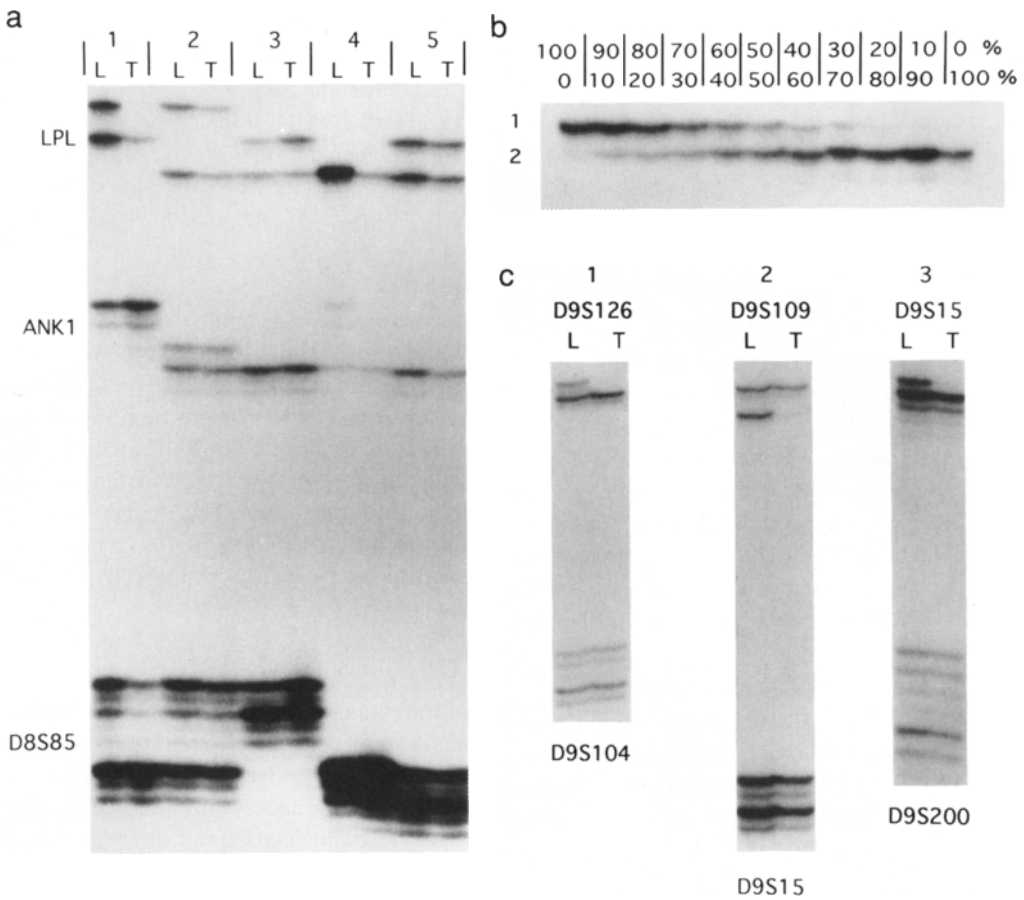


FIG. 5 (a) Multiplex PCR analysis of three chromosome 8 loci in human bladder tumors. Loss of heterozygosity at *LPL* and *D8S85* is illustrated by patient 1, and LOH at *ANK1* and retention of heterozygosity at *D9S85* by patient 4. (From Ref. 15a with permission.) (b) Analysis of different template ratios of a tetranucleotide repeat at the *LPL* locus. DNA samples from two individuals homozygous for different alleles were mixed to give a series of ratios with the same total amount of template (10 ng). Percentages of each allele (1 and 2) are shown above. Twenty-seven PCR cycles were used. (From Ref. 15a with permission.) (c) Duplex PCR analysis of microsatellite loci on chromosome 9 to illustrate breakpoints on one allele. In each tumor, the locus with larger alleles shows LOH, while that with smaller alleles shows retention of heterozygosity. (From Ref. 20a with permission.)

differing in their relative intensities. Examples of LOH at such loci are shown in Fig. 5c (tumor 3 at *D9S15*) (20a), Fig. 6b (right-hand tumor at *D9S161*) (33), and Fig. 6c (at *D9S736*). Slipped-strand mispairing is sequence-dependent and is expected to be more frequent at dinucleotide than at tetranucleotide repeats. This is

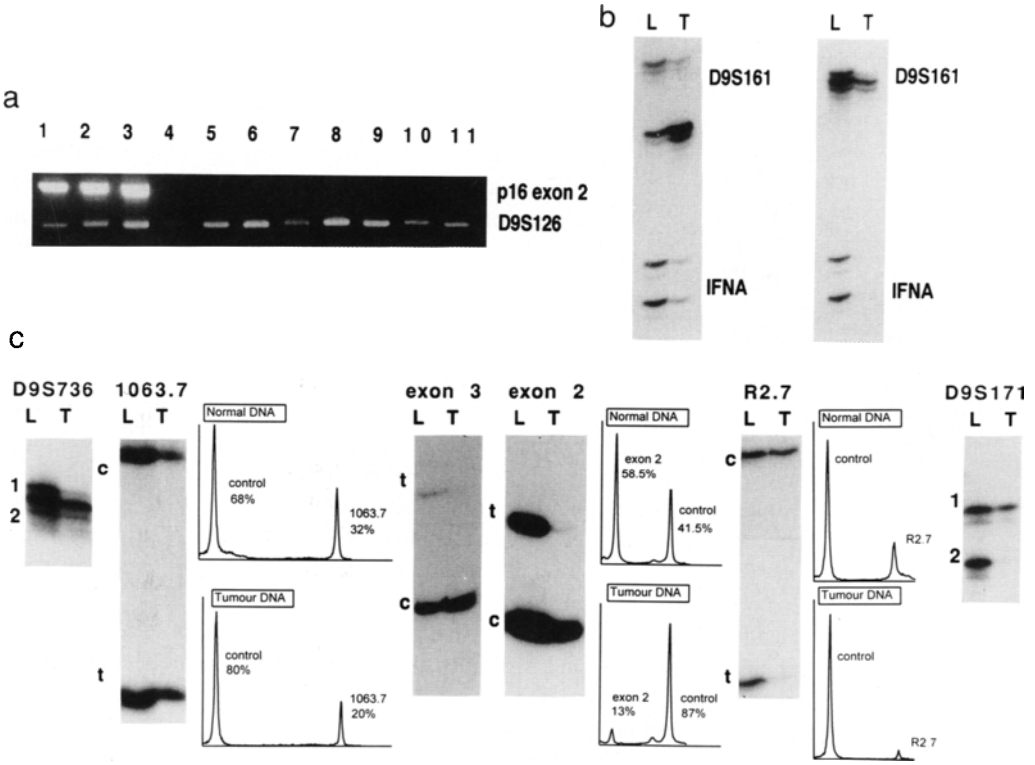


FIG. 6 Detection of homozygous deletions. (a) Homozygous deletion of exon 2 of p16 (*CDKN2*) in bladder cancer cell lines detected by cold duplex PCR with a retained marker from the same chromosomal region. (b) Homozygous deletion of *IFNA* (9p21) in bladder tumors detected using duplex quantitative PCR with end-labeled primers. In each case *D9S161* shows LOH and *IFNA* has a weak signal showing apparent retention. By reference to the control, this is interpreted as homozygous deletion of *IFNA*. The weak signal is generated from two alleles present in the normal stromal cells in the tumor tissue sample. L, leukocyte; T, tumor DNA template. (From Ref. 33 with permission.) (c) Homozygous deletion mapping in the region of p16 (*CDKN2*) in a bladder tumor. Markers are shown in linkage order from left (tel) to right (cen). *D9S736* and *D9S171* are polymorphic microsatellites (two alleles designated 1 and 2); 1063.7, p16 exons 2 and 3, and R2.7 are STSs analyzed by duplex quantitative PCR with primers for control sequences on chromosome 12 or 14. c, Control; t, test product. This tumor shows LOH (hemizygous deletion) at *D9S736*, hemizygous deletion at the STS 1063.7, homozygous deletion of p16 and of R2.7, and LOH at *D9S171*. The "residual signal" from the lost allele at *D9S171* and *D9S736* gives a signal strength similar to that seen for p16 and R2.7 in the tumor, indicating that this signal comes from stromal contamination of the tumor sample. Phosphorimager profiles for 1063.7, p16 and R2.7 are shown. Reproduced from Ref. 24 with permission.

what is observed. Most tetranucleotide repeats give extremely clean products and these should be used where possible. An example is the *LPL* tetranucleotide polymorphism shown in Figs. 3 and 5a (*LPL/MS*).

As shown in Fig. 5a, several primer pairs can be used in the same PCR reaction if the alleles do not overlap. Such multiplex reactions have the advantage that fewer gels are needed and results at different loci in the same tumor are easily compared. Before carrying out multiplex reactions, test each primer pair individually and in combination on a small number of samples. It cannot be assumed that all primers with the same theoretical melting temperature and no obvious homology will work well together. Some primer pairs will not work together at all. To set up duplex or multiplex reactions, we use 2 pmol of each primer and leave all other components as for single primer pairs. We have found duplex reactions a useful way to confirm critical breakpoints in specific tumors. Some examples of breakpoints on chromosome 9 in bladder tumors are shown in Fig. 5c. The large-scale linkage mapping project conducted at Genethon (Evry, France) has successfully used a method in which multiple cold PCR products are combined, run in acrylamide gels, blotted onto membranes, and then hybridized successively to combinations of PCR primers, using a nonradioactive detection method. This allows many products with overlapping alleles to be run on the same gel and detected sequentially. This method, which has been described in detail by Vignal *et al.* in a previous volume in this series (21), has not yet been applied to LOH analysis but could be useful if large numbers of tumors and/or markers are to be analyzed.

In general, microsatellite typings can be read by eye. Densitometry and phosphorimager analysis are not straightforward if stutter bands are present. However, tetranucleotide repeats usually give a clear band pattern that is suitable for these analyses.

Detection of Homozygous Deletions

As already indicated, the most common alterations to the two copies of a suppressor gene consist of one small sequence variation, for example, a point mutation, and one large alteration, commonly a deletion that may involve an entire chromosome arm. Rarely, homozygous deletion is observed and when it is, this generally indicates close proximity of a suppressor gene. To date, only one region of the genome has been shown to have frequent homozygous deletion in human tumors, and this is at 9p21. A candidate tumor suppressor gene, *CDKN2* (p16), an inhibitor of cyclin-dependent kinases, has been identified within this region (22, 23). This gene is homozygously deleted in cell lines derived from a wide range of tumors, and mutations have been found in some *in vivo* tumors. This has prompted many attempts to look for point mutations and deletions of the gene to verify that it is indeed the target for deletion in the many types of tumor that show LOH at 9p21.

The identification of homozygous deletions in cell line DNA is an easy task. The

population is clonal and no signal from the test sequence will be seen whether Southern blotting or PCR-based methods are used. However, in a tumor tissue sample, contaminating normal cells will generate some signal that at an informative polymorphic locus will resemble retention of heterozygosity. The first hint that both alleles of a region at 9p21 might be deleted in some bladder tumors came from our repeated observation of apparent retention of heterozygosity at the *IFNA* locus in tumors with LOH at many other loci on 9p. The methods outlined below have since been used to confirm these deletions and to map the extent of homozygous and hemizygous deletion in this region in detail, using both polymorphic markers and nonpolymorphic sequence-tagged sites (STSs).

Southern Blotting

Southern blots give good quantitation of gene copy number as discussed above. However, few laboratories who have not already made blots of tumor DNAs will wish to use Southern blotting as a routine method to look for homozygous deletions, because it is so wasteful of valuable DNA. If blots are available, hybridizations to the test sequence and a control probe should be carried out and signal intensities compared. In addition, because the test sequence is unlikely to be polymorphic, it will be necessary to set some criteria by which to decide whether one or both alleles are deleted. This requires a measure of the amount of normal tissue present in the tumor, which can be estimated by reference to a polymorphic locus where there is LOH as assessed by Southern blotting in the same tumor.

If blots are not already available, it is helpful to confirm by Southern blotting some homozygous deletions identified by quantitative PCR (see below).

Polymerase Chain Reaction

To detect homozygous deletions, PCR reactions must be carefully controlled. An internal control should be used to confirm that each template can be amplified efficiently and for quantitative comparison with the test sequence.

Detection of Homozygous Deletion by Duplex Polymerase Chain Reaction and Agarose or Polyacrylamide Gel Electrophoresis

If clonal populations of tumor cells are to be analyzed, for example, cell lines or flow-sorted tumor cells, the PCR reactions need not be sampled in the linear range of amplification, because in the case of a homozygous deletion no product will be produced no matter how many cycles are performed. Thus 30 cycles of PCR can be carried out, using 50–100 ng of template and primers for the test sequence and a control sequence. The control product in this case will be used only to verify that the

PCR reaction has worked. Run products on agarose or small, nondenaturing polyacrylamide gels as described for PCR analysis of RFLPs and VNTRs (above). An example of homozygous deletion of exon 2 of the *CDKN2* (p16) gene in a series of bladder tumor cell lines is shown in Fig. 6a. This type of analysis is not suitable for tumor DNA templates where normal cell contamination is present.

Detection of Homozygous Deletion by Quantitative Polymerase Chain Reaction Using ^{32}P End-Labeled Primers

Conditions for quantitative PCR of a possible homozygous deletion in tumor tissues are similar to those used for microsatellite analysis of LOH. However, in this case, an internal control is used. It is unlikely that most test sequences will be polymorphic. So that subsequent quantitation is easy, control primers should be designed to amplify a control product of similar size to the test sequence. For analysis of p16 deletions in bladder tumors, we made primers that amplified each of the three exons of the gene as fragments of just over 300 bp. Control primers were 300-bp fragments of genes from two chromosomes that to date have shown no alterations in bladder tumors. For polymorphic loci, markers close to the test sequence can be used because it will be easy to tell whether there is LOH or retention of heterozygosity.

Carry out duplex PCR reactions using end-labeled primers. For this application, it is vital that product be sampled on the linear part of the curve. Initial tests using various numbers of cycles should be carried out to determine the minimum number of cycles that will give a product that can be detected within a reasonable time (e.g., 48–72 hr). This is likely to be 23–26 cycles. Carry out the PCRs, run the products on denaturing polyacrylamide gels as described for microsatellite analysis, and generate phosphorimages for quantitation.

If polymorphic markers are assessed, homozygous deletions appear as weak signals resembling retention of heterozygosity. The intensity of such a signal will be similar to the "residual signal" observed at loci showing LOH in the same tumor. An example of this is shown in Fig. 6b, where two tumors with homozygous deletion at *IFNA* are illustrated. In each case, LOH at the nearby locus *D9S162* is seen but the product from the retained allele is much more intense than the *IFNA* signal, confirming that there is not a reduced amount of template in this track. Results such as these can be scored by eye.

For nonpolymorphic test sequences, some deletions can be scored by eye if there is little normal tissue contamination. This is much easier if the control product runs close to the test product. Others require careful quantitation of phosphorimages to discriminate between hemizygous and homozygous deletion. The relative representation of the test sequence can be calculated as follows:

$$\frac{\text{Intensity of test sequence in tumor}}{\text{Intensity of control in tumor}} \times \frac{\text{Intensity of control in constitutional DNA}}{\text{Intensity of test in constitutional DNA}}$$

We consider bladder tumors with values <0.5 to have homozygous deletions and those with values ≥ 0.5 and ≤ 0.8 to have hemizygous deletions. Clearly, the criteria adopted in any particular study must be decided with reference to the known amount of stromal contamination of the samples. Figure 6c illustrates results obtained by duplex PCR for markers in the region of p16 in a bladder tumor (24). The difference between hemi- and homozygous deletion is clearly defined and the critical region of deletion can be pinpointed between 1063.7 and *D9S171*.

Construction of Deletion Maps

The objective of any study of allele loss is to pinpoint the likely location of a tumor suppressor gene. An initial survey may use only two or three markers per chromosome arm to identify the regions of most frequent deletion in a specific type of tumor. A study of LOH on all chromosome arms has been termed an allelotype (by analogy with karyotype) (25). Allelotypes of many types of tumor have now been published (e.g., Refs. 25–30) and it is clear that tumors arising in different tissues have distinctive allelotypes. Some regions of deletion are shared by many tumors, while others are unique to a specific type of tumor or subset of tumors. In the allelotypes presented to date, marker density has been low and it is likely that some important tumor suppressor loci have remained undetected, particularly if for some reason (e.g., high gene density) all deletions are small. Nevertheless, many regions have been identified and the hunt is now on for the target genes within these.

Having established where deletions occur at highest frequency, high-density deletion mapping can be carried out. Here, large numbers of markers spread as evenly as possible across the minimum region of deletion [sometimes termed the smallest region of overlap (SRO)] are used. There are few regions of the genome where inadequate numbers of markers are available for this type of study. The number of tumors that should be studied to narrow the region significantly varies according to tumor type, frequency of deletion, and the chromosomal region in question. For example, in bladder tumors, deletions of the short arm of chromosome 4, although relatively infrequent ($\sim$20%), are usually small. Thus, using a panel of only 23 tumors with known deletions in the region, and only 6 microsatellite markers, a minimum region of deletion of only 750 kb could be defined (31). In contrast, deletions of chromosome 9 in the same series of tumors are extremely frequent ($>$60%) but often involve loss of an entire homolog. It is easy, therefore, to measure the frequency of the event but precise mapping of a common region is extremely difficult. In this particular example, it is now clear that there are several relevant deletion targets on 9q in addition to p16 on 9p (32). Studies by several laboratories involving literally hundreds of tumors have identified only a handful of deletions that are helpful in pinpointing these targets. Most regions of deletion studied by LOH to date fall somewhere between these two extremes.

Our approach to more detailed deletion mapping is to increase gradually the number of markers used. For example, an initial screen might use 6–8 markers spread along the chromosome arm and might screen 100–200 tumors. This should identify tumors with LOH at all loci, those with smaller deletions and those with retention of heterozygosity at all loci studied. A second screen then uses ~10 new markers within the minimum region defined by the first screen and analyzes only tumors that showed partial losses or retentions in the first screen. This should narrow the region further and the process can be repeated until no more smaller deletions can be identified. With luck, after three or four rounds of screening the minimum region of deletion defined may be only a few centimorgans in size. It is helpful to construct diagrams of the chromosome in question so that breakpoints can be visualized. An example of a typical representation of LOH data is shown in Fig. 7 (33). This depicts deletions of 9p in bladder tumors and includes information about homozygous deletions. The information pinpoints a critical region between the loci *IFNA* (LOH in tumors 14 and 15) and *D9S171* (retention of both alleles in tumors 6 and 14). It will usually be clear as the study progresses whether more than one region of deletion may be present. Large deletions in the majority of tumors may indicate that more than one target is present. However, it has been demonstrated in some cases that two regions of deletion can be clearly defined by relatively small interstitial deletions identified in only a small series of tumors. At present the basis for the differences in the sizes of

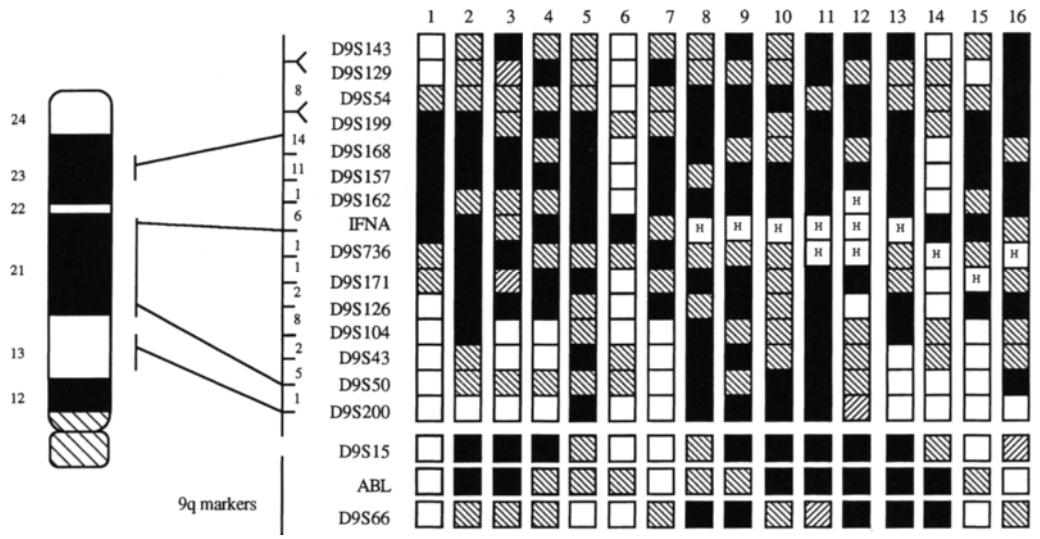


FIG. 7 Examples of deletion maps of the short arm of chromosome 9 in bladder tumors. (■) Loss of heterozygosity; (H) Homozygous deletion; (□) retention of heterozygosity; (▨) not informative; (▨) not done. (From Ref. 33 with permission.)

deletions observed is poorly understood. It seems likely that the relative importance to cell viability of genes in the immediate vicinity of a tumor suppressor gene will impose constraints on the size of deletion that can be tolerated. Undoubtedly, we shall learn more as the genome map improves and new tumor suppressor genes are cloned.

Acknowledgments

I am grateful to many colleagues who have contributed to work referred to in this chapter and in particular to Jayne Devlin, Paddy Elder, and Alison Keen, who contributed illustrations. I also thank Jennie Wood and Lynda Rasmusen for secretarial help in preparation of the manuscript. Work in the author's laboratory is funded by The Marie Curie Research Institute and grants from the Medical Research Council and The Association for International Cancer Research.

References

1. A. G. Knudson, *Nature Genet.* **5,** 103 (1993).
2. A. G. Knudson, *Proc. Natl. Acad. Sci. U.S.A.* **68,** 820 (1971).
3. R. M. Cawthon, R. Weiss, G. Xu, D. Viskochil, M. Culver, J. Stevens, M. Robertson, D. Dunn, R. Gesteland, P. O'Connell, and R. White, *Cell* **62,** 193 (1990).
4. J. Groden, A. Thliveris, W. Samowitz, M. Carlson, L. Gelbert, H. Albertsen, G. Joslyn, J. Stevens, L. Spirio, M. Robertson, L. Sargeant, K. Krapcho, E. Wolff, R. Burt, J. P. Hughes, J. Warrington, J. McPherson, J. Wasmuth, D. Le Paslier, H. Abderrahim, D. Cohen, M. Leppert, and R. White, *Cell* **66,** 589 (1991).
5. A. J. Jeffreys, V. Wilson, and S. L. Thein, *Nature (London)* **314,** 67 (1985).
6. Y. Nakamura, M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin, and R. White, *Science* **235,** 1616 (1987).
7. J. L. Weber and P. E. May, *Am. J. Hum. Genet.* **44,** 388 (1989).
8. M. Litt and J. A. Luty, *Am. J. Hum. Genet.* **44,** 397 (1989).
9. H. J. M. Smeets, H. G. Brunner, H. H. Ropers, and B. Wieringa, *Hum. Genet.* **83,** 245 (1989).
10. D. Tautz, *Nucleic Acids Res.* **17,** 6463 (1989).
11. J. L. Weber, *Genome Anal.* **1,** 159 (1990).
12. J. L. Weber and C. Wong, *Hum. Mol. Genet.* **2,** 1123 (1993).
13. J. C. Murray, K. H. Buetow, J. L. Weber, S. Ludwigsen, T. Scherpbier-Heddema, F. Manion, J. Quillen, V. C. Sheffield, S. Sunden, G. M. Duyk, J. Weissenbach, G. Gyapay, C. Dib, J. Morrissette, G. M. Lathrop, A. Vignal, R. White, N. Matsunami, S. Gerken, R. Melis, H. Albertsen, R. Plaetke, S. Odelberg, D. Ward, J. Dausset, D. Cohen, and H. Cann, *Science* **265,** 2049 (1994).
14. E. M. Wijsman, *Nucleic Acids Res.* **12,** 9209 (1984).
15. A. P. Feinberg and B. Vogelstein, *Anal. Biochem.* **132,** 6 (1983).

15a. M. A. Knowles, M. E. Shaw, and A. J. Proctor, *Oncogene* **8,** 1357 (1993).

16. O. de la Calle-Martin, V. Fabregat, M. Romero, J. Soler, J. Vives, and J. Yagüe, *Nucleic Acids Res.* **18,** 4963 (1990).
17. F. Ferre, *PCR Methods Appl.* **2,** 1 (1992).
18. P. L. Crotty, R. A. Staggs, P. T. Porter, A. A. Killeen, and R. C. McGlennen, *Hum. Pathol.* **25,** 572 (1994).
19. D. W. Yandell and T. Dryja, *Am. J. Hum. Genet.* **45,** 547 (1989).
20. X. Y. Hauge and M. Litt, *Hum. Mol. Genet.* **2,** 411 (1993).
20a. A. J. Keen and M. A. Knowles, *Oncogene* **9,** 2083 (1994).
21. A. Vignal, G. Gyapay, J. Hazan, S. Nguyen, C. Dupraz, N. Cheron, N. Becuwe, M. Tranchant, and J. Weissenbach, *Methods Mol. Genet.* **1,** 211 (1993).
22. T. Nobori, K. Miura, D. J. Wu, A. Lois, K. Takabayashi, and D. A. Carson, *Nature (London)* **368,** 753 (1994).
23. A. Kamb, N. A. Gruis, J. Weaver-Feldhaus, Q. Liu, K. Harshman, S. V. Tavtigian, E. Stockert, R. S. Day III, B. E. Johnson, and M. H. Skolnick, *Science* **264,** 436 (1994).
24. M. P. Williamson, P. A. Elder, M. E. Shaw, J. Devlin, and M. A. Knowles, *Hum. Mol. Genet.* **4,** 1569–1577 (1995).
25. B. Vogelstein, E. R. Fearon, S. E. Kern, S. R. Hamilton, A. C. Preisinger, Y. Nakamura, and R. White, *Science* **244,** 207 (1989).
26. T. Sato, A. Tanigami, K. Yamakawa, F. Akiyama, F. Kasumi, G. Sakamoto, and Y. Nakamura, *Cancer Res.* **50,** 7184 (1990).
27. D. Fults, C. A. Pedone, G. A. Thomas, and R. White, *Cancer Res.* **50,** 5784 (1990).
28. E. Tsuchiya, Y. Nakamura, S.-Y. Weng, K. Nakagawa, S. Tsuchiya, H. Sugano, and T. Kitagawa, *Cancer Res.* **52,** 2478 (1992).
29. W. Cliby, S. Ritland, L. Hartmann, M. Dodson, K. C. Halling, G. Keeney, K. C. Podratz, and R. B. Jenkins, *Cancer Res.* **53,** 2393 (1993).
30. M. A. Knowles, P. A. Elder, M. Williamson, J. P. Cairns, M. E. Shaw, and M. G. Law, *Cancer Res.* **54,** 531 (1994).
31. P. A. Elder, S. M. Bell, and M. A. Knowles, *Oncogene* **9,** 3433 (1994).
32. T. Habuchi, J. Devlin, P. A. Elder, and M. A. Knowles, *Oncogene.* **11,** 1671 (1995).
33. J. Devlin, A. J. Keen, and M. A. Knowles, *Oncogene* **9,** 2757 (1994).

Section II

Gene Mapping, Cloning, Sequencing

This Page Intentionally Left Blank

[8] Mapping Human Disease Genes by Linkage Analysis

E. Chung and R. M. Gardiner

Introduction

The identification of human disease genes can be accomplished by two strategies: functional cloning and positional cloning. In functional cloning the gene is isolated based on the information regarding the protein product or its function. In positional cloning the isolation of a disease gene starts from the knowledge of its genetic or physical location in the genome and usually with no or little information regarding its function. Gene mapping, the assignment of human genes to specific chromosomes and parts of chromosomes, is the first step toward positional cloning. This can be done using a variety of methods, the most powerful of which is linkage analysis, especially if the disease is known only by its phenotype. Genetic mapping and positional cloning offer an opportunity for understanding the molecular basis of inherited diseases for which the biochemical basis is unknown. In the last decade or so, the development of methods that allow access to polymorphism at the level of DNA sequence variation has led to the construction of high-resolution genetic maps covering most of the human genome. These maps should in theory permit the gene for any Mendelian trait to be localized by linkage analysis.

The human genome is large. The haploid human genome is estimated to have a genetic length of 3000 centimorgans (cM), corresponding to a physical size of 3000 megabases (Mb). The most recent estimate of the number of genes is about 60,000 to 70,000, which corresponds to an average of 1 gene every 40 to 50 kilobases (kb) throughout the human genome and gene-poor regions represent roughly $\frac{1}{10}$ the gene density of gene-rich regions (1). By the end of 1993, more than 3000 genes had been mapped and there are more than 5000 polymorphic markers identified throughout the human genome with an average distance of 3 cM between loci.

In this chapter, approaches to mapping human disease genes by linkage analysis are discussed with a minimum emphasis on the mathematical principles involved. We hope the chapter will serve as an introduction to any investigator who plans to study an inherited disease in humans by linkage analysis.

Genetic Disease in Humans

Genes are units of heredity, and in modern times, a specific coding sequence of DNA. Each individual carries two copies of each autosomal gene, of which one was re-

Methods in Molecular Genetics, Volume 8

ceived from the mother and the other from the father. As a result of mutation, a gene may occur in different forms or states called alleles. Each allele potentially has a different physical expression. The relative frequencies in the population of the different alleles of a gene are called gene frequencies. Each individual contributes two alleles (for an autosomal gene) to the population gene pool. The pair of alleles in an individual constitutes the genotype of that individual. The two alleles in an individual can be the same, in which case the individual is a homozygote, or they are different, in which case the individual is a heterozygote. The phase refers to whether alleles at two (or more) linked loci inherited by a child from his/her parent are from the same chromosome (coupling or cis) or homologous chromosomes (repulsion or trans). Phenotype is the expression of a particular genotype but phenotypes do not always fully exhibit the underlying genotypes. The concept of penetrance establishes the connection between genotypes and phenotypes. Penetrance can be defined as the conditional probability that an individual with a given genotype expresses the phenotype or simply the probability of being affected (with a disease) given a certain genotype. Genetically determined diseases are often classified into three major categories: chromosomal, single-gene disorders, and complex diseases (2).

Chromosomal Disorders

Chromosomal disorders are the result of the addition or deletion of entire chromosomes or parts of chromosomes. Major chromosomal anomalies are often incompatible with survival, and such abnormalities are found in almost one-half of spontaneous abortuses and account for almost 3% of childhood deaths. In gene mapping, chromosomal abnormalities such as small deletions or translocations associated with a disease may provide important clues and allow localization of the disease gene. The inheritance of some diseases, for example, Prader-Willi syndrome, can depend on which parent contributes the mutant gene (3). This phenomenon of differential genetic contributions of parents to their offsprings is called imprinting.

Single-Gene Disorders

Single-gene disorders are caused by the presence of disease alleles at one genetic locus. Single-gene disorders are inherited in a simple Mendelian fashion and are also referred to as Mendelian diseases. Some 3000 distinct disorders are now known or suspected to be single-gene diseases inherited in autosomal dominant, autosomal recessive, or X-linked fashion. It is important to realize that it is a given allele that is dominant or recessive. Some diseases due to mutations at the same locus display dominant or recessive inheritance depending on the mutant allele segregating.

Autosomal Dominant Inheritance

A dominant allele is one that manifests its phenotypic (recognizable) effect in heterozygotes. About 1500 autosomal dominant conditions are known. People affected by rare dominant diseases are almost always heterozygotes. Mutations at loci encoding structural proteins often show dominant inheritance.

Autosomal Recessive Inheritance

A recessive allele causes a phenotypic effect only when present in the homozygous state. More than 1000 autosomal recessive conditions are known. The parents of affected children are phenotypically normal carriers and each child has a one-in-four chance of being affected. A typical pedigree would show a single affected child, with unaffected nonconsanguineous parents and no previous family history of the disease. Genes encoding enzymes often display recessive mutant alleles: one normal allele allows synthesis of sufficient enzymes to allow normal function.

Autosomal Codominant Inheritance

If both alleles of a pair are fully expressed in a heterozygote, the mode of inheritance is termed codominant and is said to demonstrate a gene dosage effect. The majority of genetic marker loci display codominant inheritance.

X-Linked Inheritance

If the responsible gene is on the X chromosome, the disease is termed X-linked.

Complex Diseases

Complex disease is a broad category that covers any genetic disease that has a mode of inheritance that cannot be accurately defined or explained. A disease can also be considered complex when we do not really know who is affected or at least who is a potential carrier of a specific genetic defect. Genetic heterogeneity (discussed below), and the possibility of large rates of sporadic nongenetic causes of the same or similar disease phenotypes (phenocopies), also make a disease complex. Various models have been used in the analysis of complex diseases. These include the polygenic model (additive effects of alleles at several or many genetic loci), the epistatic model (nonadditive effects of interaction of alleles at multiple loci), the multifactorial model (combined effects of genetic and environmental factors), the threshold of liability model (as in multifactorial inheritance, but the trait is expressed only when the liability of an individual exceeds a critical threshold value), the single major locus model, and the mixed model. This group of diseases is both the most common and the least understood of human genetic disease. They constitute the greatest component of disease burden in the whole population. Important diseases in this category

include asthma, diabetes mellitus, the epilepsies, major psychoses, cardiovascular disorders, and cancer.

As the effect of other genetic and environmental factors increases, pedigrees with single-gene disorders become less regular, until it becomes more convenient to describe the condition as multifactorial and non-Mendelian. The distinction between single-gene and complex genetic disorders may not always be very clear-cut.

Mitochondrial DNA

The mitochondrion has its own DNA, which encodes about 12 polypeptides, while nuclear DNA controls the synthesis of the majority of mitochondrial (mt) proteins. Most known mitochondrially encoded proteins are subunits of respiratory chain complexes located in the inner mitochondrial membrane. The complete nucleotide sequence of the mitochondrial genome has been determined, all 16,569 base pairs (bp) (4). The mitochondria are derived exclusively from the mother. Diseases arising from mutations in mtDNA therefore display maternal inheritance, that is, an affected mother could pass the disease on to her children, both sons and daughters could be affected, but only her daughters could pass the trait on to subsequent generations. Examples of mitochondrial DNA disorders include Kearns–Sayre syndrome and myoclonic epilepsy with ragged red fibers (MERRF) (5, 6).

Linkage Analysis

Several methods exist for the mapping of genes (7). These include interspecific somatic cell hybridization with DNA and phenotypic markers, gene dosage, *in situ* DNA hybridization to banded chromosomes, chromosome sorting, amino acid sequence of protein resulting from gene deletion and fusion, linkage homology between species, comparative mapping, centromere mapping, and linkage analysis. We discuss only linkage analysis in this chapter.

Genetic linkage analysis is the only method that allows mapping of disease genes that are detectable only as phenotypic traits. The majority of genes that underlie genetic diseases fall into this category. Linkage was first demonstrated by T. H. Morgan in 1910. He observed that in fruit flies the gene for white eye or its alternative, that for red, traveled with the X chromosome (8). This was the first gene in any creature to be assigned to a specific chromosome. Three years later, A. H. Sturtevant, a student of Morgan's, analyzed mating results for fruit flies with six different mutant factors, each known to be recessive and X-linked. He published the results in a paper in 1913 that contained a diagram of the first gene map (9). Establishment of human linkage did not occur until in 1911, with the realization that color blindness was expressed

only in males but was, like the X chromosome, inherited in the female line, and therefore X-linked (10). The first identification of linkage with a particular human autosome was as late as 1968 by R. Donahue, a doctoral student who observed the cosegregation of blood group antigens and a distinctive abnormal chromosome 1 through his own pedigree (11).

Linkage analysis is based on the observed segregation of homologous chromosomes (and thus the alleles they carry) in meiosis in family studies. Genetic linkage can be defined as the tendency for alleles at loci close together on the same chromosome to be transmitted together through meiosis, the process of cell division which leads to the formation of gametes (i.e., egg or sperm cells). In meiosis the two homologous copies of each chromosome (one from the mother and the other from the father) pair up before each goes into one or the other daughter cell. During this process, a portion of the maternal homolog recombines with the paternal homolog to form a hybrid chromosome in the place of the original ones, that is, the chromosome that a child has inherited from the mother is a patchwork consisting of alternating portions of the maternal grandfather's chromosome and the maternal grandmother's chromosome. The entire basis of genetic linkage analysis is that recombination events between two genetic loci on the same chromosome occur at a rate related to the distance between them. Alleles at loci on different chromosomes are transmitted independently of each other and are separated at 50% of meioses. Loci, extremely close together on the same chromosome, are transmitted together essentially all of the time and they are said to be tightly linked. Alleles at two loci that are located on the same chromosome but some distance apart tend to be transmitted together unless a single recombination event in meiosis separates the two. The extent of genetic linkage between loci is measured by the recombination fraction (θ) between them, which is the probability that a gamete produced by a parent is a recombinant. Two loci are said to be genetically linked when the recombination fraction between them is less than 0.5. The recombination fraction is a measure of the genetic distance between loci. Because of the occurrence of multiple cross-overs, the recombination fraction is not additive and must be transformed by a map function into the map distance measured in centimorgans. The two most commonly used map functions are the Haldane map function (12)

$$X \text{ (cM)} = -\frac{1}{2} \ln(1 - 2\theta)$$

and the Kosambi map function (13)

$$W \text{ (cM)} = \frac{1}{4} \ln[(1 + 2\theta)/(1 - 2\theta)]$$

At $\theta = 0.22$, $X = 29$ cM and $W = 23.6$ cM. In practice, the recombination fraction (in percent) and genetic map distance (in centimorgans) are equal for small values of θ less than 10%. Genetic studies of extended pedigrees to determine how frequently two loci are inherited together can be used for the production of genetic maps: the loci on a chromosome arranged in the correct linear order with the genetic distance between them estimated. The relationship between genetic and physical distances varies in different locations depending on the frequency of recombination in the region. In recombination "hot spots," a given physical distance corresponds to an extended genetic distance. On average, 1 cM corresponds to 1 Mb. It should also be recalled that the recombination rate is not equal in male and female meioses, being generally higher in the human female. The female genetic map is therefore usually "longer" than the corresponding male map, although of course the physical size of the section of chromosome is the same.

The estimate of the recombination fraction, in simple cases, can be obtained by calculating the proportion of recombinants out of all opportunities for recombination. It is often not possible to count recombinants and nonrecombinants directly, and therefore likelihood-based methods are usually used in linkage analysis (14). The probability is expressed as a lod score (logarithm of the odds), which is the logarithm of the ratio between the probability that the two alleles are linked and the probability that they are not linked. Thus the lod score Z at a given value of θ is

$$Z(\theta) = \log_{10}[L(\theta)/L\left(\frac{1}{2}\right)]$$

By convention, an lod score of +3 or greater (equivalent to greater than 1000:1 odds in favor of linkage) is considered evidence that two loci are linked. This critical value of 3 has taken into account the fact that any two loci have a certain probability (1 in 20) of being on the same chromosome. For X-linked disorders, a lod score of 2 is sufficient evidence for linkage because the prior possibility of linkage is much higher than that for an autosomal disease. The lod score test is usually performed by calculating the lod score (Z) for all values of θ between 0 and 0.5, and determining the recombination fraction at which the lod score is maximal. One advantage of the lod score method is that values for different pedigrees can be added together. This of course requires the assumption that there is genetic (locus) homogeneity for the phenotype under investigation.

Negative lod scores suggest absence of linkage and lod scores below -2 are accepted as evidence that linkage has been excluded at the corresponding values of the recombination fraction. Lod scores between -2 and 3 are inconclusive and indicate more data are needed. This can be obtained by studying more families using more informative or additional markers close by.

The whole subject of human genetic linkage is dealt with in detail by Ott and Terwilliger in two books (15, 16).

Design of Linkage Project

The success of any genetic mapping project depends mainly on the nature of the disease studied. This in turn determines the availability, structure, and power of pedigrees. One of course cannot carry out linkage analysis without a set of highly informative genetic markers covering the entire human genome. These factors are discussed in turn below.

The Disease

There are a few important questions one needs to ask about the disease before embarking on a linkage study. These include the following: (a) Is the disease genetic? (b) What is the mode of inheritance? (c) Is it fully penetrant? (d) Are phenocopies likely to be present? and (e) Is there evidence of heterogeneity?

The first and the most important question about the disease is the following: Is the disease truly genetic and not a result of common environmental factors? The genetic contribution to etiology may be established with family, twin, and adoption studies, as well as studies involving different ethnic groups. Once genetic factors have been established, the goal of further analysis, such as segregation analysis, is to investigate the number of gene(s) that influence the disease and to determine, if possible, the mode of inheritance of the gene. However, it is often not easy to define the precise genetic mechanism of a disease because of complicating factors such as heterogeneity, variable expression, phenocopies, and nonpenetrance.

Variable expression means the condition can be severe in one person and mild in another. Nonpenetrance is the extreme of variable expression: a person with no signs of the condition (asymptomatic) but who carries the disease allele. Penetrance, as discussed earlier, is the probability of being affected (with a disease) given a certain genotype or the probability that a susceptible individual is affected. It is estimated as the proportion of affecteds among susceptible individuals. However, it is often not easy to find a suitable collection of susceptible individuals and to estimate in an unbiased manner the proportion of affecteds. When "high-density" pedigrees with multiple affected individuals in multiple generations are collected, penetrance values estimated from such pedigree data are likely to be falsely high. This constitutes ascertainment bias. In the absence of good estimates, one may try a number of reasonable-appearing values, or simply work with a reduced penetrance of, say, 0.7 for susceptible genotypes. Many factors, including environmental factors, interaction with other genes, age, or sex, can all have a significant effect on the penetrance. This can sometimes be taken into account by forming different liability classes and for each such class, defining an appropriate set of penetrance. Incomplete or reduced penetrance obscures our view of the genotypes, that is, phenotypes no longer unambiguously exhibit underlying genotypes, which results in a loss of information.

In many diseases, phenocopies (sporadic cases or false positives) occur. These are individuals whose affection status is not the result of genetic predisposition but is due to some unspecified environmental factors such as epilepsy secondary to head injury. Genetic and nongenetic forms of disease may not always be distinguishable. The presence of phenocopies is modeled by a nonzero penetrance for nonsusceptible genotypes (homozygous normal for the disease locus).

There are two types of heterogeneity: allelic and nonallelic (locus). In locus heterogeneity, a given phenotype is caused by mutations at more than one locus. This variety of heterogeneity is particularly important and often a major obstacle in linkage analysis. In allelic heterogeneity, different phenotypes (and sometimes modes of inheritance) are caused by a variety of mutant alleles at the same locus. Allelic heterogeneity can be detected through linkage analysis. The possibility of locus heterogeneity may be minimized by studying a single large pedigree and a uniform phenotype. Because a single pedigree large enough to detect linkage is rarely available, one often relies on studying a number of families with a uniform phenotype assuming locus homogeneity. A uniform phenotype depends on an accurate diagnosis of the affected individuals. This is achieved by acquiring as much detail as possible of the diagnostic features of the affected individuals. Clearly, misdiagnosis is an important source of error and can adversely affect the outcome of a linkage analysis, as a recombinant may be mistaken as a nonrecombinant. If one is uncertain about the diagnosis of an individual, it is advisable to leave the individual as unknown but they could still be included in the analysis to provide phase information (affecteds-only analysis). This is often necessary in studying diseases with incomplete penetrance.

Family Resource

Sampling strategies vary, influenced most by the observed mode of inheritance of the disease studied. Ideally linkage analysis should be carried out on a single large extended pedigree, as in the case of Huntington's chorea or the autosomal recessive Hirschsprung's disease (17, 18). This will minimize the risk of genetic locus heterogeneity. Unfortunately such a family is rarely available.

Because genetic markers are not perfect, not all matings within a given family will be informative. In general the larger the number of affected individuals in the family from whom samples can be obtained, the better the estimate of recombination fraction between marker and disease, and the greater the support for linkage. A single larger family of a given size produces more potential linkage information than does a combination of several smaller families with the same number of individuals. Pedigrees with larger sibship sizes are more efficient than those with smaller sibships because the proportion of meioses scored per genotype is higher when many offspring are available for inclusion in the analysis. For the same reason, three-generation families are more useful than two-generation families.

For an autosomal dominant disease, it is often possible to ascertain so-called multiplex families with multiple affected individuals in multiple generations. For an autosomal recessive disease, one is usually only able to collect small nuclear families in which there is at least one affected child with a number of unaffected siblings in two generations. For rare recessive traits, the parents of an affected individual tend to be related more often than in the general population because they have an increased probability of sharing two copies of the same ancestral disease allele—identity by descent (IBD). The two alleles of a marker locus tightly linked to the disease gene tend to be IBD as well. An affected child of a first-cousin marriage is estimated to inherit the same alleles (i.e., to be homozygous) at marker loci for about 15 cM around the disease locus from the parents. This concept of homozygosity mapping, introduced by Lander and Botstein (19), greatly enhances the power of linkage analysis in rare recessive diseases. Under ideal conditions a single affected child born of first cousins is roughly as informative as a nuclear family with three affected children. For a complex disease, multiplex families would be suitable, perhaps with an emphasis on collecting families with at least two affected siblings so that sib-pair and other nonparametric methods of analysis can be carried out.

Before embarking on a linkage project, it is useful to estimate the number of families required to find linkage to a hypothetical marker. In general, 20 phase-known meioses are required to produce a lod score of 3. However, many factors such as unknown phase, heterogeneity, misclassification (including errors in marker typing), and incomplete penetrance can all reduce the informativeness of the family and must be allowed for in the analysis. In addition some of the family members may be unwilling to cooperate or some cell lines may not grow. Computer simulation methods are available (e.g., SLINK), and they are now the standard techniques by which to approximate the expected lod score (ELOD) before marker typing.

Genetic Markers and Linkage Map

Genetic markers are any Mendelian characters that can be used to follow the inheritance of a small section of a chromosome through a pedigree. For a marker to be useful in genetic mapping, it must be polymorphic and its chromosomal location must be known. It is also desirable for a marker to show codominant inheritance (so that the genotype is always apparent from the phenotype) and to be easily scored in a readily available tissue such as blood. In the past, protein polymorphisms, such as red blood cell (ABO, Rh, etc.) and HLA antigens, were the main markers used in linkage studies. Unfortunately, highly polymorphic protein markers are rare and the number is limited, covering only a small percentage of the human genome: approximately 10–15%.

Genetic marker loci may be more or less polymorphic depending on the number of alleles and their population frequencies. The degree of polymorphism of a marker

may be assessed by the proportion of individuals in the population who are heterozygous for that marker, that is, the probability that a random individual is heterozygous is used as a measure of the degree of polymorphism. This probability may be estimated by the observed heterozygosity, which is simply the proportion of heterozygous individuals observed in a random sample of unrelated individuals, or the expected heterozygosity (H), which is a more precise estimate assuming that the genotypes are in Hardy–Weinberg equilibrium. The expected heterozygosity is defined as

$$H = 1 - p_i^2$$

where the sum is taken over all alleles, with p_i denoting the frequency of the ith allele. An older measure of heterozygosity is the polymorphism information content (PIC), which takes into account the probability that half of the offspring would be uninformative if both parents are identically heterozygous. The PIC may be somewhat more appropriate for family data.

More recently, with the appearance of recombinant DNA technology, DNA markers that detect restriction fragment length polymorphisms (RFLPs) have added enormously to the number of polymorphic loci (20). Restriction enzymes are prokaryotic proteins that cleave at, or very near to, specific recognition sequences within DNA. These restriction sites are present in some people and absent in others, causing the DNA to be cut into different-sized fragments in different people by the respective restriction enzyme. RFLPs have only two alleles, giving rise to only three possible combinations and therefore are often not very informative (PIC < 0.5). The polymorphism is detected by Southern blotting.

In 1988 the introduction of the polymerase chain reaction (PCR), using thermalstable DNA polymerase, provided an entirely new means of analyzing polymorphism and made practical the analysis of length variation in simple-sequence tandemly repeated DNA (21). Variable number tandem repeats (VNTRs) (22) were the first type of what are now collectively known as simple-sequence length polymorphisms (SSLPs). These are short, tandemly repeated DNA sequences, where the number of repeat units at a particular chromosomal location varies from person to person. The size of different repeated units varies from a single nucleotide up to tens of nucleotides. Repeats of more than four nucleotides are called minisatellites and those of one to four nucleotides are called microsatellites. Microsatellites are best detected using the polymerase chain reaction. These markers are simple to use and, compared to Southern blotting, they reduce the time to score a polymorphism from 1 week to less than 2 days. They have many possible alleles and are therefore usually very informative. They appear to be distributed randomly throughout the genome, unlike VNTR minisatellites, which tend to be clustered near the ends of chromosomes.

These SSLPs form the majority of DNA polymorphisms in use today for linkage analysis. The Genome Database (GDB) as of September 1994 contained information on a total of 6691 SSLPs, and 3752 of those (56%) show heterozygosity in the population studied of greater than 60%. The publication of a series of genetic maps, con-

structed from genotypic data derived from a large number of reference linkage families, has made the search for disease genes easier. The linkage maps of the human chromosomes, mainly based on genotypes from the 40 Centre de l'Etude du Polymorphisme Humain (CEPH) reference families, are becoming increasingly dense, with the average distance between adjacent markers being not more than 10 cM (23–25).

The sequence and working conditions for most of the microsatellite markers are easily obtainable from the published linkage maps or accessed from the Human Genome Mapping Project database. The polymorphism is commonly detected by radioactive labeling, 6% (w/v) acrylamide gel electrophoresis, and autoradiography. The radioactive labeling is commonly done by either incorporating ^{32}P-labeled deoxynucleotide into the elongated strands during PCR or by end labeling one of the primers using [^{32}P]ATP and polynucleotide kinase. This process has limitations in large-scale projects requiring hundreds of thousands of genotypes. More recently, using fluorescent dyes and semiautomated laser-based technology, the speed of marker typing has been improved. The PCR primers can now be labeled with fluorescent dyes (four possible), each with a different emission wavelength that can be detected by a laser-based automated DNA fragment analyzer such as the ABI373A sequencer (Perkin Elmer, Applied Biosystems Division, UK) (26). This technique allows simultaneous analysis within a single lane of a polyacrylamide gel of 24 PCR products of microsatellite markers (3 sets of 8 marker loci with allele size ranges in each set arranged so that they do not overlap; a fourth dye is used to incorporate a size standard in each lane) and 864 (36 lanes $\times$ 24 markers) genotypes per gel.

Linkage Strategy

Having established a family resource with sufficient power to detect linkage, a strategy is required to maximize the speed and efficiency with which success is achieved. Clearly the strategy will be determined by the disease in question and the mode of inheritance involved, but in each case the following questions need to be answered.

How Many and Which Individuals Should Be Typed?

The minimum number of individuals necessary to detect linkage should be analyzed. A pragmatic approach is required, as there is a relationship between the number of meioses examined and the genetic distance at which a linkage may be detected. In general one can only hope to find linkage at a distance not more than 20 cM from the disease locus because one usually cannot find sufficient family material to establish significant evidence for linkage at longer distances. The more limited the family resource (or number of individuals used), the higher the density of marker loci required, as a significant lod score will emerge only if the locus is close enough for there to be

few recombinants among the meioses examined. The power of individual families and the total resource can be estimated using SLINK, which will also determine the minimum number of individuals required to detect linkage by finding a "positive" lod score. There is initially no need to aim for an expected lod score greatly in excess of +3 for detection of linkage: the speed of detection will be increased by screening marker loci more quickly in fewer individuals. When a significantly positive lod score is found or linkage is detected, additional individuals or families can then be analyzed to confirm the linkage and refine the localization.

In autosomal recessive disorders, in which data from several small nuclear pedigrees must be pooled, the possibility of locus heterogeneity is a particular problem. Unfortunately, if several different single loci are responsible for an apparently homogeneous phenotype, the subset of families linked to a particular locus will be hard to detect unless the total number of families is large. If a number of consanguineous families can be ascertained as in some rare recessive disorders, initial typing may be confined to affected individuals only, utilising the concept of homozygosity mapping.

In autosomal dominant disorders with high penetrance, multigeneration families with multiple affected individuals are likely to be available and one of these may be used to detect linkage. Individuals of uncertain affection status may still be required to be typed to provide information useful for establishing the phase at a particular locus.

In diseases with complex inheritance, a nonparametric approach may be considered, such as affected sib-pair (ASP) analysis or affected pedigree member (APM) analysis. Using these approaches, marker typing may be restricted initially to affected individuals.

Which Marker Loci Should Be Examined and in What Order?

As a general rule, highly polymorphic loci that are detected by PCR-based automated methodology and part of a well-supported genetic map are preferable. The availability of approaching 6000 SSLPs renders microsatellites the markers of choice now under most circumstances. The large number of alleles at these loci and therefore the high level of heterozygosity maximizes the chance of rendering a meiosis informative.

The selection of which microsatellites to analyze and the order in which they are done depends initially on the disease, and the existence or nonexistence of candidate genes or regions for the disease. If there are no plausible candidate genes or regions for the disease, a systematic search of the genome is required. These options are considered in turn.

Candidate Genes or Regions

A candidate gene may be identifiable on the basis of existing knowledge of the biochemical or physiological basis of the disease or on the existence of a murine homo-

log of the disease. The role of such a gene may be explored by linkage analysis or direct sequence analysis: the approach adopted depends on the family resource available and the existing knowledge of the genomic structure of the gene. If an SSLP has been identified in the gene (an intragenic microsatellite), this should be used first. Identification of obligate recombinants with an intragenic marker can exclude the gene as the disease locus in a particular family. Alternatively, highly polymorphic loci flanking the candidate gene may be utilized.

A candidate region may be identified on the basis of cytogenetic abnormalities associated with the disease. These are more likely to underlie dominant disorders, and are potentially extremely valuable not only as an indicator of the possible location of the disease gene but also for high-resolution mapping. An example was the 10q deletion identified in autosomal dominant Hirschsprung's disease (27). Cosegregation of diseases within a family may represent an observation of potential significance. An example was the pedigree segregating both tuberose sclerosis complex and autosomal dominant adult polycystic kidney disease (28).

Systematic Search

If there are no *a priori* clues as to the disease gene location, a so-called systematic search is undertaken. Such an approach depends on two factors: the family resource and gene density. The question of family resource has been discussed. Gene density is known to vary significantly among chromosomes, for example, chromosomes 17 and 19 have the highest density and chromosomes 13 and 18 the lowest. If one uses markers strategically located in the middle of the GC-rich or gene-rich regions, they can scan up to 10 times as many genes as are present in a gene-poor region. Localization of the elusive genes in theory should be achieved with much less effort and expense (29). This approach is particularly attractive in the search for susceptibility genes for common human disorders with strong genetic components. It has been estimated that about 70 markers in the autosomes would be sufficient for this gene-rich region scanning of the genome, and this initial effort will probably localize more than 70% of the genes. The task of selecting the most valuable markers is by no means an easy one, because the precise localization of a DNA segment to a chromosomal band is far from perfect.

The Genome Database (GDB) contains 672 genes mapped by linkage analysis out of a total 3485 cloned genes. In most of the successful linkage studies so far, it usually took between 100 and 200 polymorphic markers before the localization of the disease gene was achieved. An initial search at high density increases the likelihood of generating an increasingly large amount of exclusion data before linkage is found. An initial search should preferably be carried out at low density, with an effort made to select and choose the appropriate markers (conveniently located in the gene-rich regions, within a candidate gene or region, easy to read and interpret, and highly polymorphic), in the hope of finding a genuine "hint" of linkage. Subsequent searches can be undertaken at increasing density down to an average of 10-cM inter-

vals between loci, corresponding to about 300 in total to cover the genome. At this density, multipoint analysis may be used to scan the regions between loci. SLINK provides a guide to the threshold at which a positive lod score should be pursued by typing additional individuals and marker loci in the vicinity.

Which Method of Linkage Analysis Should Be Applied?

The method of linkage analysis depends, once again, mainly on the disease studied. For a simple Mendelian trait, the most common method is two-point and multipoint lod score calculation using MLINK and LINKMAP, respectively.

For a complex disease, the mode of inheritance is often uncertain. An investigator may try various models, involving, for example, combinations of different penetrance values and diagnostic schemes and then quote the highest lod score. Under such an approach, the traditional critical limit of a lod score of 3 in the test for linkage may no longer be appropriate because a portion of the lod score is not due to linkage but merely to maximization over models. This procedure amounts to maximizing the lod score, which has been shown to lead to an increased false-positive rate, although using an incorrect model for the disease does not (30). Complex diseases are sometimes thought to be produced by a combination of different genes working together to produce a phenotype. It is possible that some unaffected individuals may possess the disease-predisposing genotype at one of the loci but lack the required second disease locus genotype necessary for development of the disease. It is therefore advisable to consider all unaffected individuals to have "unknown" phenotype when doing linkage analysis of a complex disease. This is known as an affecteds-only study.

Alternatively, one can apply nonparametric methods of analysis. They have a common principle in assessing the chromosomal regions that tend to be inherited in common among affected relatives without making any assumption about the mode of inheritance. They are mostly based on the concepts of identity by descent (IBD: copies of the same allele that are inherited from a common ancestral source) and identity by state (IBS: copies of the same allele at a given locus). If two alleles are IBD, they are definitely also IBS, but the converse is not necessarily true. Two nonparametric methods are commonly used, namely affected sib-pair analysis and affected pedigree member analysis, which are discussed below.

Programs for Linkage Analysis

Most linkage programs in current use compute lod scores by calculating pedigree likelihoods numerically for each set of parameter values. A list of addresses from which these programs can be obtained is available in the book by Terwilliger (16).

LINKAGE LINKAGE is a package consisting of three analysis programs, each of which calculates two-point or multipoint likelihoods in human pedigrees (31, 32). MLINK calculates likelihood by stepwise variation of the recombination fraction in one of the interlocus intervals. It also computes genetic risks. LINKMAP assumes a fixed map of markers and calculates likelihoods for a new locus at various points in each interval along the known map. ILINK estimates recombination fractions and finds the maximum of the log likelihood.

LIPED LIPED (for Likelihood in PEDigrees) is the first generally available program to compute likelihoods numerically in large pedigrees (33, 34). It is similar to MLINK and calculates two-point lod scores. It has been generalized to allow for allelic association, age-dependent penetrance, and sex-specific recombination fractions.

SLINK SLINK is a simulation program designed to generate an expected lod score of a marker locus at various recombination fractions from the disease locus (35). It randomly generates replicates of marker data and phenotypes of a family and calculates an expected lod score. It requires one to predetermine the phenotypes of certain individuals for the disease locus (usually the affected child or children) and genotypes of a marker locus for certain individuals (usually the parents of the affecteds) in order to generate the replicates.

HOMOG HOMOG is a program for detecting the presence of heterogeneity (36, 37). It carries out a homogeneity test (A-test) under the following hypothesis: two family types, one with linkage between a trait (or any gene locus) to a marker or map of markers, the other without linkage. It is commonly carried out in two situations. First, in the presence of a positive linkage result but when not all families appear to be linked, one asks if there is significant evidence for a proportion of the families segregating the linked gene and another proportion segregating a putative unlinked gene for the same disease. Second, in the absence of linkage, one asks if a proportion of families is assumed to be segregating an unlinked gene; is there significant evidence for linkage to a disease gene in a proportion of the families? Various extensions of the homogeneity test are available to allow for more than two family types and sex difference in the recombination fractions.

LINKSYS LINKSYS is a utility program written by J. Attwood to make the LIPED and LINKAGE packages more easily accessible (38). Genetic and pedigree data are first entered into a database from which the information can be extracted in a form that is suitable for entry into the LIPED or LINKAGE program. Data need only be entered once and can be extracted in almost any desired combination in a form suitable for analysis.

CYRILLIC The objectives of CYRILLIC, a Window-based utility package, are similar to those of the LINKSYS program. It is written by C. Chapman (39). The program is essentially a database for pedigree and genetic data that can be extracted

in a variety of forms suitable for analysis by the LIPED and LINKAGE packages. It has additional abilities to carry out haplotype analysis, to draw pedigrees, and to perform risk calculations.

Affected Sib Pair Analysis Affected sib pair (ASP) analysis was first used by Penrose in 1935 (40). The tests compare the observed and expected IBD distributions in pairs of affected siblings. It assumes that if a given marker is cosegregating with a disease-predisposing allele, then affected siblings of affected persons are more likely to receive the same allele IBD at a closely linked marker locus than if the marker locus were segregating independently (i.e., unlinked) of the disease-predisposing allele. This type of analysis requires no assumption about the mode of inheritance. The problem of trying multiple models and correcting for inflation of the lod score is avoided. The statistic used here is the χ^2 test, which can be performed by hand or with a simple calculator.

Affected Pedigree Member Method The affected pedigree member (APM) method has been adapted from ASP to examine more distantly related individuals in extended pedigrees by IBS rather than IBD (41). Again these sibs tend to share alleles IBS at loci linked to disease-predisposing loci, because of an increased probability that they inherit the marker locus IBD with the disease locus or because the marker locus itself is contributing to the disease phenotype in some manner. Because some individuals are more distantly related, it is not typically possible to determine whether alleles are inherited IBD or not, so a comparison of their IBS status is made instead. This method is critically sensitive to the allele frequency.

Examples

Two projects of gene mapping by linkage analysis are discussed briefly. The aim is to clarify and illustrate some of the points discussed earlier. The first project involves a rare disease called pseudohypoaldosteronism type 1. Families with this disorder have been reported to exhibit different modes of inheritance and the clinical spectrum (severity and number of organs affected) varies greatly. The problems anticipated here would include a limited family resource and possible heterogeneity (allelic or locus or both). The second project is to try to map the gene of a common disorder affecting young infants called infantile pyloric stenosis. The phenotype appears uniform and there are numerous affected adults in multiplex families available for the study, as the disease is nonlethal. However, the inheritance is complex and problems anticipated here would include uncertain mode of inheritance and incomplete penetrance. In each case a plausible candidate gene or region has been excluded and a genome search is being undertaken. At the time of writing, neither of these genes has been mapped.

Pseudohypoaldosteronism type 1

Disease

Pseudohypoaldosteronism type 1 (PHA1) is an example of a hormone-resistance or end-organ insensitivity syndrome. This condition was first described by Cheek and Perry in 1958. Since then more than 100 cases have been reported in the literature, although the prevalence is unclear. Various aspects of this condition have been reviewed extensively (42).

Clinical Features

Affected infants often present in the first few months of life with evidence of salt loss in their urine, sweat, stool, or saliva. Other features commonly include vomiting and failure to thrive. Diagnosis is made from the typical biochemical features, which include hyponatremia, hyperkalemia, metabolic acidosis, and an increased renin activity with a markedly elevated aldosterone level in the presence of normal renal and adrenal functions. Patients do not respond to exogenous mineralocorticoids and treatment is by salt supplementation, which can often be discontinued in early childhood. The raised aldosterone level almost invariably persists into adulthood. The clinical spectrum varies greatly with mild or asymptomatic to severe fatal cases. Some patients have salt loss restricted to the kidneys (renal form) while some have widespread abnormalities involving the skin, salivary glands, and colon (generalized form). Armanini *et al.* demonstrated that intracellular aldosterone-binding sites with characteristics similar to those reported in kidney and colon could be detected in human peripheral blood mononuclear leukocytes (43). Such binding studies have been shown to be reduced in some patients with PHA1 and some of their unaffected relatives (44). It has been suggested that results of such binding studies could represent a marker for the disease and possibly the subclinical state in identifying a mildly affected person or a carrier in a recessive family. On the other hand, reduced aldosterone binding in PHA1 may merely be a result of downregulation secondary to the elevated serum aldosterone level.

Genetics

A familial incidence of PHA1 has been well documented although formal segregation analysis has not been carried out. Both autosomal dominant and autosomal recessive models of inheritance have been reported. It has been suggested that the autosomal dominant form is milder and the end-organ defect is often restricted to the level of the renal tubule. In contrast, families with the generalized form usually exhibit a recessive inheritance pattern. There is an increased incidence of consanguinity among the families with an autosomal recessive inheritance. The different models of

inheritance and the variable (severity and number of organs affected) clinical spectrum suggest the possibility of heterogeneity (locus, allelic, or both).

Pathophysiology

The pathophysiology of PHA1 remains unclear. Since the first description of hormone resistance in the context of pseudohypoparathyroidism was made by Albright in 1948, many other such syndromes have been described. These include resistance to insulin, androgen, vitamin D, cortisol, thyroid hormone, and aldosterone. In many cases the underlying abnormality has been traced to a defect in the gene encoding the receptor protein with a variety of mutations reported. However, a number of hormone-resistance conditions have been shown to be the result of nonreceptor defects. There is increasing evidence that aldosterone exerts an immediate effect on its target cells via a membrane-bound receptor (now known as nongenomic action). This is followed by the classic genomic action of aldosterone, which is thought to act by binding onto the cytoplasmic mineralocorticoid receptor, forming an active complex. This complex initiates protein synthesis (aldosterone-induced proteins, AIPs) by binding onto a number of hormone response elements (HREs) of the nuclear DNA. The relationship between the nongenomic and genomic actions is unclear and the nature of the AIPs has remained uncertain. The proposed AIPs include Na^+/K^+ channels and Na^+,K^+-ATPase.

Families

Ascertainment of families has been achieved through collaboration with the physicians who had reported cases of PHA1 in the literature and British pediatric endocrinologists or nephrologists who would normally look after children with PHA1. A total of 30 families has been identified. In 14 of these families, the parents are consanguineous (see Fig. 1, top) and an autosomal recessive inheritance can be assumed, while the other families have an uncertain mode of inheritance (autosomal dominant, recessive, or sporadic).

Linkage Strategy

Because of the possibility of heterogeneity as suggested by the different modes of inheritance and wide spectrum of clinical features, we have initially concentrated on the 14 families in which the parents are consanguineous with an assumed autosomal recessive inheritance and the risk of heterogeneity may be minimized. DNA is available on 10 of these families (families 001–010). Computer simulation confirms the power of these 10 families with an average expected lod score of $+2$ at $\theta = 0.15$ from the disease gene. Linkage analysis is carried out using the lod score method, assuming a fully penetrant autosomal recessive inheritance with a disease allele frequency of 0.0001 and a zero phenocopy rate.

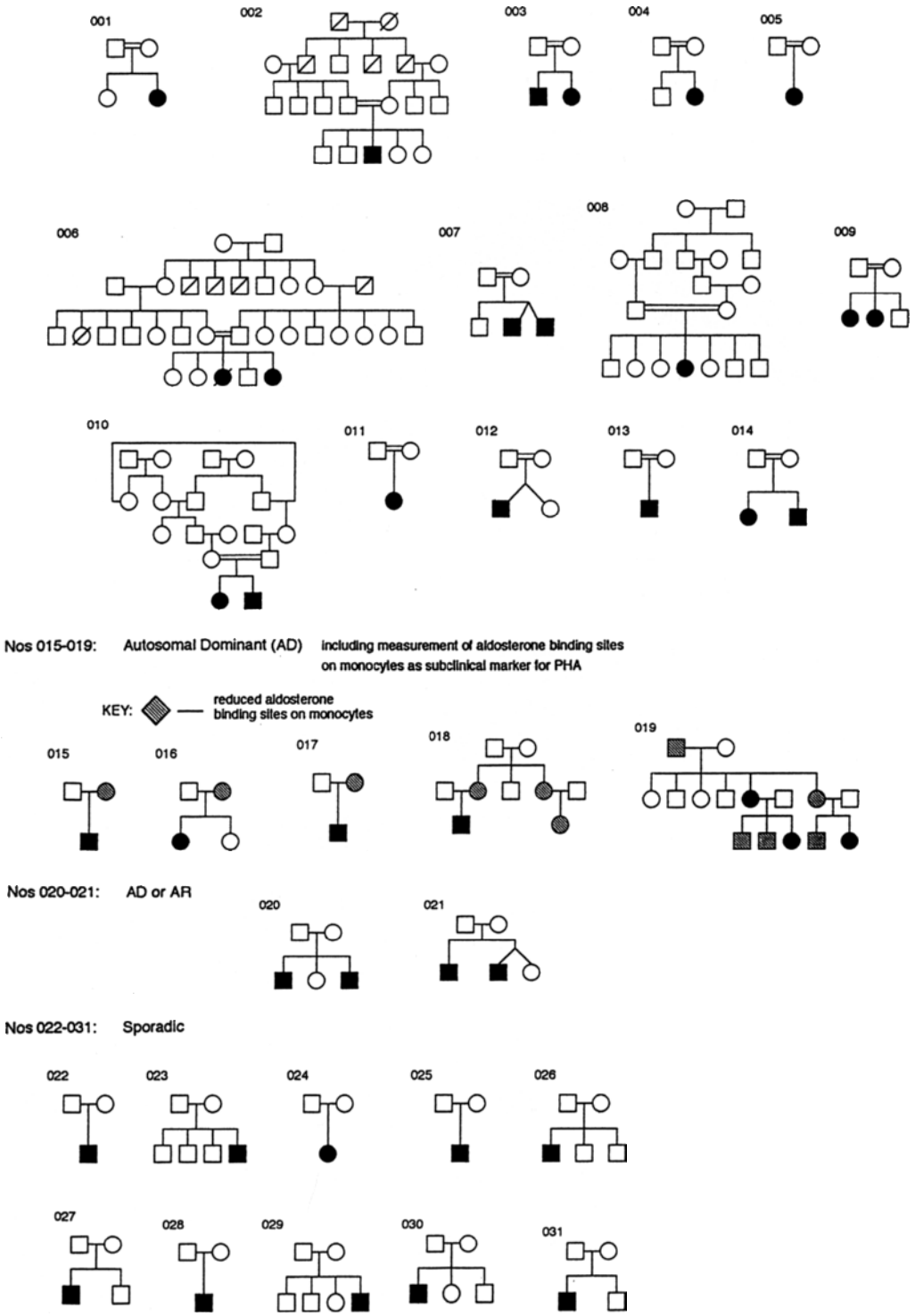


FIG. 1 Pseudohypoaldosteronism type 1.

Candidate Genes and Regions

There is little evidence to suggest a problem at the prereceptor level. The defect is most likely to be at either the receptor or at the postreceptor level. The mineralocorticoid receptor (MLR) is a prime candidate, as are the so-called aldosterone-induced proteins, which include amiloride-sensitive epithelial sodium channels and the Na^+,K^+-ATPase. The MLR locus is a strong candidate gene for PHA1 because the gene encoding the receptor protein has been shown to be the disease gene in many other hormone-resistance syndromes. In addition, reduced intracellular aldosterone binding has been demonstrated in patients with PHA1. The MLR locus has been mapped to chromosome 4q31.1-31.2 by *in situ* hybridization. Ideally one should use an intragenic marker, as exclusion is achieved simply by observing obligate recombinants. Unfortunately an informative MLR intragenic marker is not available. We have therefore investigated the etiological role of the MLR gene in PHA1 by linkage analysis using microsatellite markers (*D4S192, D4S1548, D4S413*) covering the region where the MLR has been located (see Fig. 2) (45).

Results of Linkage Analysis between PHA1 and MLR Locus Region

The haplotypes of the affected individuals for the three markers are shown in Table I. There is an obvious lack of homozygotes among the affected individuals (only 2 of 14 for *D4S192*, 4 of 13 for *D4S1548*, and 3 of 14 for *D4S413* were homozygous). The pairwise lod score is less than -2 for at least 10 cM around *D4S192* and *D4S413*, and 6 cM around *D4S1548* as shown in Table II. There is therefore no evidence of linkage of PHA1 and the gene for the MLR locus. These data exclude MLR as the gene responsible for PHA1, at least in our families of PHA1 with an

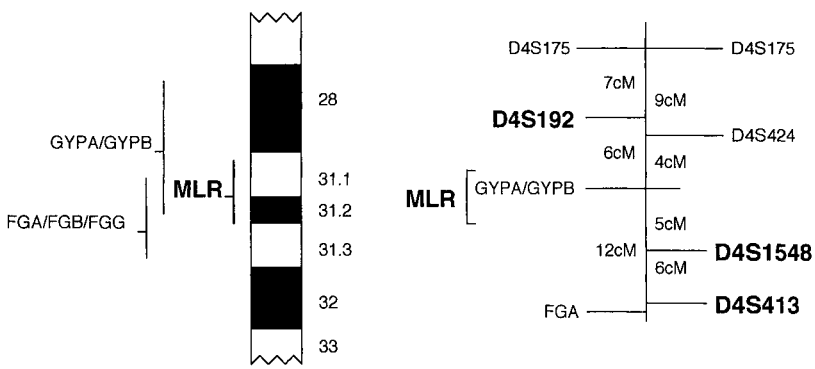


FIG. 2 Physical and genetic maps of human chromosome 4q, indicating relative position of the MLR and marker loci *D4S192, D4S1548,* and *D4S413.*

TABLE I Haplotype Data of PHA1-Affected Individuals at Marker Loci *D4S192, D4S1548,* and *D4S413*[a]

Family:	001	002	003		004	005	006	007		008	009		010	
Person:	II-2	IV-3	II-1	II-2	II-2	II-1	IV-5	II-2	II-3	II-4	II-1	II-2	V-1	V-2
D4S192	21	54	**44**	**44**	14	23	34	24	24	35	23	23	12	12
D4S1548	13	12	**22**	**22**	14	**44**	**22**	34	34	[b]	34	34	13	13
D4S413	**55**	11	13	23	53	15	13	42	42	24	**33**	**33**	23	23

[a] Numbers in boldface indicate homozygotes.
[b] Not available because of failure of amplification.

autosomal recessive mode of inheritance. This is in keeping with the published results of sequence analysis of the MLR gene on patients with PHA1 in whom no significant sequence variation has been identified (46, 47).

Systematic Search

A systematic search forms part of our strategy with special attention and priority being paid to candidate genes or regions and gene-rich regions. For maximal efficiency, a mapping panel is formed with 8 of the affected individuals, aiming initially to detect a hint of linkage (i.e., when 4 of the 8 individuals are homozygous for a marker), considering only a 1-in-16 chance that an affected child from a first cousin's marriage will be homozygous at a fully informative locus. Marker typing will be carried out using fluorescence-based semiautomated technology (ABI373A sequencer Perkin Elmer, Applied Biosystems Division, UK). A set of about 300 fluorescently labeled primers has been developed and is available (48).

TABLE II Two-Point lod Score between PHA1 and Marker Loci around MLR Locus Region

	lod score at θ =						θ at lod
PHA–Marker	0.00	0.05	0.10	0.20	0.30	0.40	score = −2
PHA–*D4S192*	−17.91	−2.01	−0.37	0.42	0.33	0.10	0.05
PHA–*D4S1548*	−11.59	−0.49	0.51	0.81	0.54	0.21	0.03
PHA–*D4S413*	$-\infty$	−2.35	−0.80	0.12	0.22	0.11	0.05

Infantile Pyloric Stenosis

Disease

Infantile pyloric stenosis (PS) is the commonest condition requiring surgical intervention in the first year of life. It was first described by Hirschsprung in 1887. Its prevalence is estimated to be 1 to 5 per 1000 live births in Britain. Mortality was high until successful treatment by pyloromyotomy was developed by Rammstedt in 1911.

Clinical Features

Affected infants present between the ages of 3 to 12 weeks with forceful vomiting typically described as projectile. On examination there are visible peristalsis and a palpable pyloric tumor. Diagnosis is usually made clinically, supported on occasion by ultrasound or barium studies, and confirmed at laparotomy.

Genetics

Pyloric stenosis is a complex disease. Its familial incidence was first reported by Cockayne and Penrose and has been confirmed by many subsequent family studies. The twin concordance rate is estimated to be about 30–40%. Male infants are two to five times more at risk than female infants. The inheritance does not follow a simple Mendelian pattern. Segregation analysis has failed to identify the exact inheritance mechanism, although a multifactorial sex-modified threshold model of inheritance is thought to be the most compatible. However, a single gene with reduced penetrance has not been excluded. An asymptomatic twin (with diagnostic features of PS on barium studies) of an affected infant has been reported, which indicates an incomplete penetrance.

Pathophysiology

The pathophysiological basis of PS is unknown. It is characterized by hypertrophy of the pyloric circular muscle and gastric outlet obstruction. In biopsy specimens of infants with PS, Vanderwidden *et al.* found NADPH-diaphorase staining to be absent in the neurons that innervate the circular muscle of the pylorus (49). NADPH-diaphorase histochemical enzymatic activity is due to neuronal nitric-oxide synthase (nNOS), the enzyme that catalyzes the formation of nitric oxide, a mediator of smooth muscle relaxation in the pylorus. A defect in the relaxation of the pylorus has therefore been suggested as a possible underlying mechanism of PS and the gene for nNOS (NOS1) is a strong candidate gene. Further evidence in support of this hypothesis was provided by the demonstration by Huang *et al.* that the NOS1 knockout mice showed phenotypic features of enlarged stomach with hypertrophy of the pyloric sphincter and the circular muscle layer, which are diagnostic of PS (50).

Families

Thirty-five multiplex families have been ascertained from a variety of sources, including families of probands identified in previous family studies (Fig. 3). This includes a total of 305 individuals, 114 of whom are affected (82 males, 32 females). Fifteen families have at least 4 affected members and there are 3 affected members in 11 families. Diagnosis was made according to standard clinical criteria and confirmed at laparotomy in all but five of the affected individuals who were treated medically.

Linkage Strategy

In the absence of a satisfactory estimate of the mode of inheritance, a major single gene is assumed and linkage analysis is carried out using the lod score method under both autosomal dominant (AD) and autosomal recessive (AR) models of inheritance with a reduced sex-specific penetrance (0.60 for males and 0.15 for females). The gene frequency of the disease allele is set at 0.002 and a phenocopy rate of 0.002 is used. All unaffected individuals are classified as unknown (affecteds-only study).

Candidate Genes or Regions

As already mentioned, the gene for nNOS has become a strong candidate for PS, which is at present under investigation. Two siblings in one family have been reported to have PS as well as a translocation involving the telomeres of 17q and 8q (S. Hodgson, personal communication, 1995), which makes these regions of some increased interest. Pyloric stenosis has also been described in association with various chromosomal disorders, including Edward's syndrome (trisomy 18), Down's syndrome (trisomy 21), Turner's syndrome (XO), and duplication of chromosome 9q11-q33. Pyloric stenosis was reported to occur in four of seven (57%) of the cases with duplication of chromosome 9q11-q33 (51). We have therefore carried out linkage analysis of 20 of our PS families and 7 microsatellite markers from this region of chromosome 9.

Results of Linkage Analysis between Infantile Pyloric Stenosis and Markers on Chromosome 9q

Linkage analysis was carried out between 20 PS families (families 001–020; DNA was available in these 20 of the 35 families ascertained at the time of analysis) and marker loci (*D9S55, D9S111, D9S15, D9S12, D9S56, D9S59,* and *ASS*) covering the region of chromosome 9q11-q33 the duplication of which has been associated with PS (52). Two-point lod scores for all six markers are shown in Table III. The two-point lod score was -2 at greater or equal to 0.04 for all marker loci. By convention an lod score of -2 is taken as proof of exclusion and the region of exclusion with each marker overlaps into each other. These data therefore provide evidence against

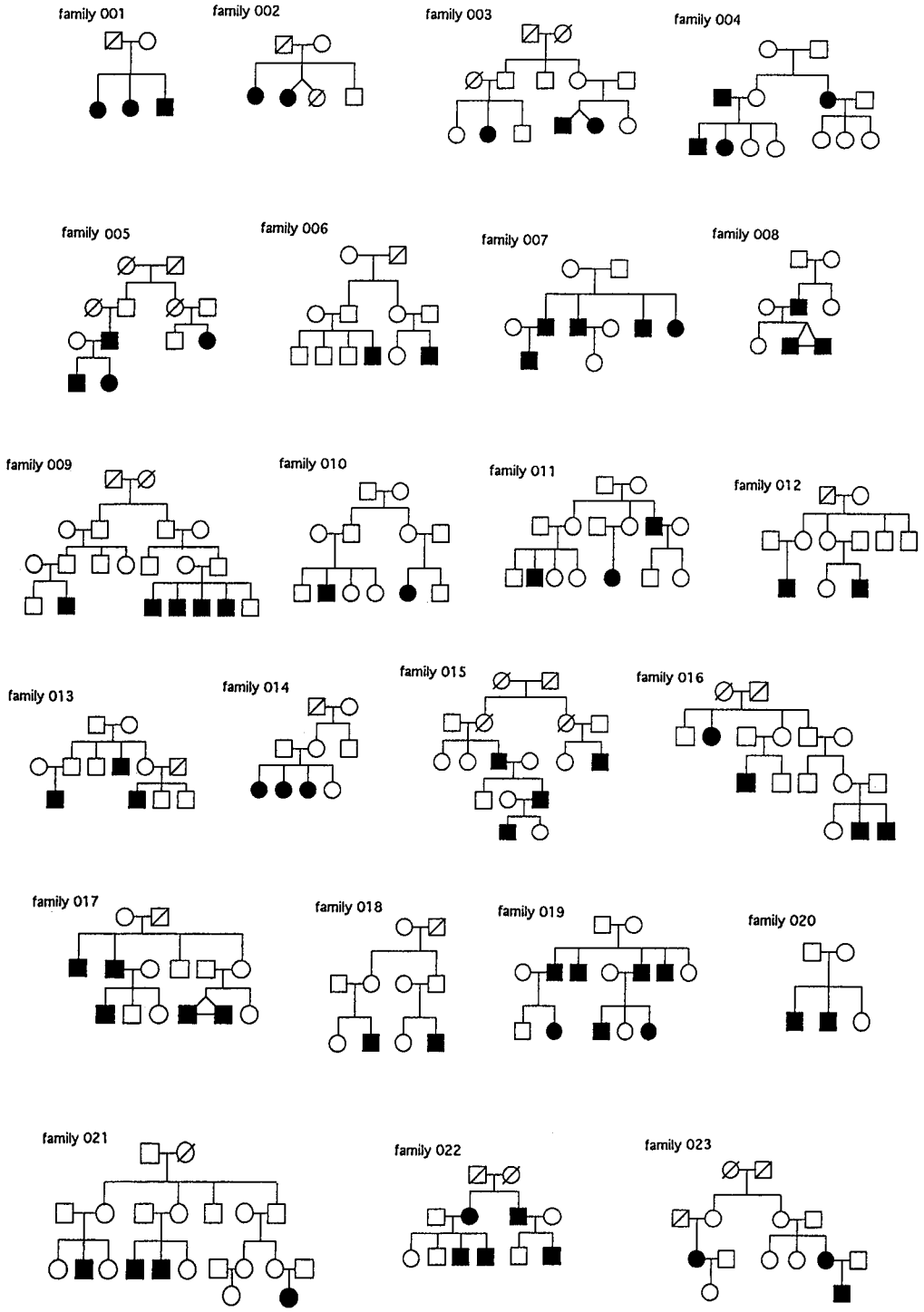


FIG. 3 Pyloric stenosis pedigrees.

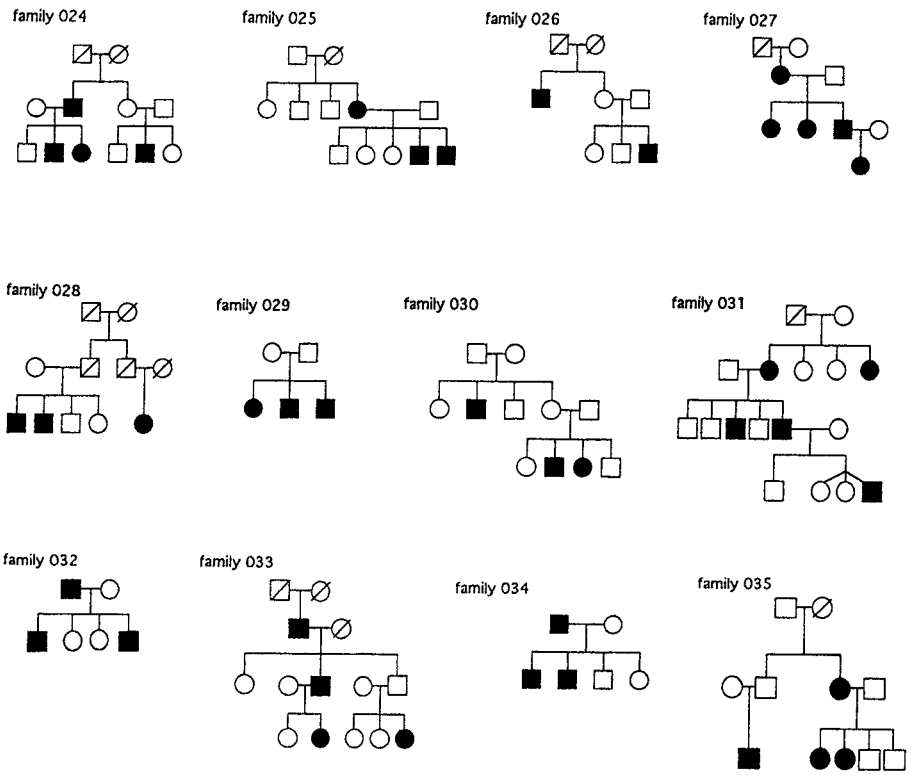


FIG. 3 *Continued.*

TABLE III Two-Point lod Scores between Pyloric Stenosis Trait and Chromosome 9 Marker Loci

	Lod score at recombination fraction						
PS–marker	0.00	0.05	0.10	0.20	0.30	0.40	θ at lod score = −2
Autosomal dominant							
PS–*D9S55*	−12.05	−4.30	−2.00	−0.29	0.14	0.18	0.10
PS–*D9S111*	−12.41	−4.83	−2.50	−0.66	−0.09	0.05	0.11
PS–*D9S15*	−9.25	−3.23	−1.31	0.09	0.38	0.27	0.07
PS–*D9S12*	−9.89	−3.37	−1.34	0.15	0.41	0.27	0.07
PS–*D9S56*	−9.32	−3.61	−1.50	0.08	0.38	0.26	0.08
PS–*D9S59*	−10.31	−4.69	−2.63	−0.75	−0.05	0.13	0.11
PS–*ASS*	−11.80	−4.39	−1.84	0.03	0.38	0.24	0.09
Autosomal recessive							
PS–*D9S55*	−4.27	−1.78	−0.56	0.33	0.36	0.15	0.04
PS–*D9S111*	−3.70	−1.77	−0.73	0.19	0.31	0.14	0.04
PS–*D9S15*	−6.00	−3.63	−2.25	−0.81	−0.23	−0.04	0.11
PS–*D9S12*	−5.51	−2.82	−1.46	−0.26	0.07	0.06	0.07
PS–*D9S56*	−5.35	−2.87	−1.56	−0.35	0.02	0.05	0.07
PS–*D9S59*	−5.50	−3.13	−1.91	−0.65	−0.14	0.03	0.09
PS–*ASS*	−4.05	−2.05	−1.01	−0.05	0.19	0.14	0.05

the existence of a major locus predisposing to PS over the whole region from *D9S55* (9p12) to *ASS* (9q34.1).

Systematic Search

As with PHA1, a systematic search is performed using the set of 300 highly informative, fluorescently labeled microsatellite markers covering the human genome. Six of the most informative families form the initial mapping which consists of 70 individuals. Assuming homogeneity, these 70 individuals should give an lod score of 1.5 at $\theta = 0.15$ from a major locus under our assumed models of inheritance.

Update on results

PHA1

Following exclusion of the gene for the human mineralocorticoid receptor as the disease locus for PHA1 (53), a genome search was undertaken using homozygosity mapping in 11 consanguineous families. Evidence of locus heterogeneity was obtained with six families linked to chromosome 16p12.2-13.11 and five families to 12p13.1-pter. A two-locus admixture lod score of 9.9 was obtained (54).

PS

The etiological role of NOS1 in PS was investigated by analysis of two intragenic polymorphisms (NOS1a, NOS1b) in 27 families. Using the multi-allele transmission disequilibrium test (55), significant overall transmission disequilibrium was observed between PS and NOS1a (p-0.006) (56). Consideration of each allele independently revealed a highly significant tendency for allele 7 (210bp) of NOS1a to be preferentially transmitted to the affected offspring ($p = 0.0006$). These observations suggest that NOS1 is a susceptibility locus for PS.

References

1. C. Fields, M. D. Adams, O. White, and J. C. Venter, *Nat. Genet.* **7,** 345 (1994).
2. V. A. McKusick, "Mendelian Inheritance of Man," 10th ed. The Johns Hopkins University Press, Baltimore, Maryland, 1991.
3. S. B. Cassidy, L. W. Lai, and R. P. Erickson, *Am. J. Hum. Genet.* **51,** 701 (1992).
4. P. E. M. Fine, *Lancet* **1,** 659 (1978).
5. P. Lestienne and G. Ponsot, *Lancet* **1,** 885 (1988). [Letter]
6. J. M. Shaffaer, M. T. Lott, and A. M. S. Lezza, *Cell* **61,** 931 (1990).
7. R. S. Sparkes, *in* "The Genetic Basis of Common Diseases" (R. A. King, J. I. Rotter, and A. G. Motulsky, eds.), Chap. 3. Oxford University Press, Oxford, 1992.
8. T. H. Morgan, *Science* **32,** 120 (1910).
9. A. H. Sturtevant, *J. Exp. Zool.* **14,** 43 (1913).
10. E. B. Wilson, *Arch. Mikrosk. Anat. Entwicklungsmech.* **77,** 249 (1911).
11. R. P. Donahue, W. B. Bias, J. H. Renwick, and V. A. McKusick, *Proc. Natl. Acad. Sci. U.S.A.* **61,** 949 (1968).
12. J. B. S. Haldane, *J. Genet.* **8,** 299 (1919).

13. D. D. Kosambi, *Ann. Eugen.* **12,** 172 (1944).
14. N. E. Morton, *Am. J. Hum. Genet.* **7,** 277 (1955).
15. J. Ott, "Analysis of Human Genetic Linkage" (revised edition). The Johns Hopkins University Press, Baltimore, Maryland, 1991.
16. J. D. Terwilliger and J. Ott, "Handbook of Human Genetic Linkage." The Johns Hopkins University Press, Baltimore, Maryland, 1994.
17. J. Gusella, R. E. Tanzi, P. I. Bader, M. C. Phelan, R. Stevenson, M. R. Hayden, K. J. Hofman, A. G. Faryniarz, and K. Gibbons, *Nature (London)* **318,** 75 (1985).
18. E. G. Puffenberger, E. R. Kauffman, S. Bolk, T. C. Matise, S. S. Washington, M. Angrist, J. Weissenbach, K. L. Garver, M. Mascari, R. Ladda, S. A. Slaugenhaupt, and A. Chakravarti, *Hum. Mol. Genet.* **3,** 1217 (1994).
19. E. S. Lander and D. Botstein, *Science* **236,** 1567 (1987).
20. K. F. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Ehrlich, *Cold Spring Harbor Symp. Quant. Biol.* **51,** 263 (1986).
21. D. Botstein, R. L. White, M. Skolnick, and R. W. Davis, *Am. J. Hum. Genet.* **32,** 314 (1980).
22. Y. Nakamura, M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin, and R. White, *Science* **235,** 1616 (1987).
23. NIH/CEPH Collaborative Mapping Group, *Science* **258,** 67 (1992).
24. K. H. Buetow, J. L. Weber, S. Ludwigsen, T. Scherpbier-Heddema, G. M. Duyk, V. C. Sheffield, Z. Wang, and J. C. Murray, *Nat. Genet.* **6,** 391 (1994).
25. G. Gyapay, J. Morissette, A. Vignal, C. Dib, C. Fizames, P. Millasseay, S. Marc, G. Bernardi, M. Lathrop, and J. Weissenbach, *Nat. Genet.* **7,** 246 (1994).
26. S. R. Diehl, J. Ziegle, G. A. Buck, T. R. Reynolds, and J. L. Weber, *Am. J. Hum. Genet.* **47,** A177 (1990).
27. M. Angrist, E. Kauffman, S. A. Slaugenhaupt, T. C. Matise, E. G. Puffenberger, S. S. Washington, A. Lipson, D. T. Cass, T. Reyna, D. E. Weeks, W. Sieber, and A. Chakravarti, *Nat. Genet.* **4,** 351 (1993).
28. European Polycystic Kidney Disease Consortium, *Cell* **77,** 881 (1994).
29. S. E. Antonarakis, *Nat. Genet.* **8,** 211 (1994).
30. D. E. Weeks, T. Lehner, and E. Squires-Wheeler, *Genet. Epidemiol.* **7,** 237 (1990).
31. G. M. Lathrop, J. M. Lalouel, C. Julier, and J. Ott, *Proc. Natl. Acad. Sci. U.S.A.* **81,** 3443 (1984).
32. G. M. Lathrop, J. M. Lalouel, C. Julier, and J. Ott, *Am. J. Hum. Genet.* **37,** 482 (1985).
33. J. Ott, *Am. J. Hum. Genet.* **26,** 588 (1975).
34. J. Ott, *Am. J. Hum. Genet.* **28,** 528 (1976).
35. D. E. Weeks, J. Ott, and G. M. Lathrop, *Am. J. Hum. Genet.* **47,** A204 (1990).
36. C. A. B. Smith, *J. R. Statist. Soc.* **15B,** 153 (1961).
37. J. Ott, *Ann. Hum. Genet.* **47,** 311 (1983).
38. J. Attwood and S. A. Bryant, *Ann. Hum. Genet.* **52,** 259 (1988).
39. C. J. Chapman, *Am. J. Med. Genet.* **36,** 155 (1990).
40. L. S. Penrose, *Ann. Eugen.* **6,** 133 (1935).
41. D. E. Weeks and K. Lange, *Am. J. Hum. Genet.* **42,** 315 (1988).
42. P. A. Komesaroff, J. W. Funder, and P. J. Fuller, *in* "Bailliere's Clinical Endocrinology and Metabolism" (M. C. Sheppard and P. M. Stewart, eds.), Chap. 5, Vol. 8. 1994. Bailliere Tindall, London.

43. D. Armanini, T. Strasser, and P. C. Weber, *Am. J. Physiol.* **248,** E388 (1985).
44. D. Armanini, U. Kuhnle, T. Strasser, H. Dorr, J. Butenandt, P. C. Weber, J. R. Stockigt, P. Pearce, and J. W. Funder, *N. Engl. J. Med.* **313,** 1178 (1985).
45. K. A. Mills, K. H. Buetow, Y. Xu, J. L. Weber, M. R. Altherr, J. J. Wasmuth, and J. C. Murray, *Genomics* **14,** 209 (1992).
46. M.-C. Zennaro, P. Borensztein, X. Jeunemaitre, D. Armanini, and F. Soubrier, *J. Clin. Endocrinol. Metab.* **79,** 32 (1994).
47. P. Komesaroff, U. Kuhnle, K. Verity, and P. J. Fuller, *J. Clin. Endocrinol. Metab.* **79,** 1580 (1994).
48. P. W. Reed, J. L. Davies, J. B. Coperman, S. T. Bennett, S. M. Palmer, L. E. Pritchardet, S. C. L. Gough, Y. Kawaguchi, H. J. Cordell, K. M. Balfour, S. C. Jenkins, E. E. Powell, A. Vignal, and J. A. Todd, *Nat. Genet.* **7,** 390 (1994).
49. J.-M. Vanderwindden, P. Mailleux, S. N. Schiffmann, J.-J. Vanderhaeghen, and M.-H. L. De Laet, *N. Engl. J. Med.* **327,** 511 (1992).
50. P. L. Huang, T. M. Dawson, D. S. Bredt, S. H. Snyder, and M. C. Fishman, *Cell* **75,** 1273 (1993).
51. Y. Yamamoto, N. Oguro, T. Nara, H. Horita, N. Niitsu, and S. Imaizumi, *Am. J. Med. Genet.* **31,** 378 (1988).
52. D. J. Kwiatkowski, E. P. Henske, K. Weimer, L. Ozelius, J. F. Gusella, and J. Haines, *Genomics* **12,** 229 (1992).
53. E. Chung, A. Hanukoglu, M. Rees, R. Thompson, M. Dillon, I. Hanukoglu, T. Bistritzer, U. Kuhnle, J. Seckl, and R. M. Gardiner, *J. Clin. Endocrinol. Metab.* **80,** 3341 (1995).
54. S. S. Strautnieks, R. J. Thompson, A. Hanukoglu, M. J. Dillon, I. Hanukoglu, U. Kuhnle, J. Seckle, R. M. Gardiner, and E. Chung. *Hum. Mol. Genet.* in press (1996).
55. P. C. Sham and D. Curtis. *Ann. Hum. Genet.* **59,** 323 (1995).
56. E. Chung, D. Curtis, G. Chen, P. A. Marsden, R. Twells, W. Xu, and R. M. Gardiner. *Am. J. Hum. Genet.* in press (1996).

[9] Engineering 100- to 300-kb DNA as Persisting Extrachromosomal Elements in Human Cells Using Human Artificial Episomal Chromosome System

Tian-Qiang Sun and Jean-Michel H. Vos

Introduction

The human genome consists of numerous genes and other functional genetic units, many of them spanning hundreds of kilobase pairs of DNA. The ability to clone large DNA fragments expands our capacity to study the complexity of the human genome. Several microbial-based systems, including yeast artificial chromosomes (YACs) (1, 2), bacterial chromosomes (BACs) (3, 4, 5), and P1 artificial chromosomes (PACs) (6, 7), have been developed for cloning 100–1000 kb of DNA. Advances in genomic cloning and construction of detailed whole genome maps have been greatly facilitated by the development of these large-capacity DNA cloning systems. But functional studies of mammalian genetic units have been impeded by the lack of a suitable system that permits the stable propagation and maintenance of large DNA segments in mammalian cells. To conduct such functional studies, one must be able to transfer entire genetic units into mammalian cells, and such cloned DNA should be stably maintained in the recipient cells. Because YACs, BACs, or PACs do not carry the genetic elements necessary for their episomal maintenance in mammalian cells, they could not be stably maintained unless integrated into host chromosomes. However, the integration processes often caused rearrangement, deletion, and disruption of inserts.

To address this issue, we have developed a human artificial episomal chromosome (HAEC) system for cloning and functional analysis of large DNA fragments in human cells (8). The HAEC system is based on the latent replication origin (*oriP*) (9, 10) of the Epstein–Barr virus (EBV), a large human herpes virus. This system permits the propagation and stable maintenance of large DNA inserts, ranging from 60 to 330 kb, in human cells as artificial episomal chromosomes. A pilot HAEC library containing 1500 independent colonies with insert sizes ranging from 100 to 200 kb has been constructed, which covers approximately 10% of the human genome. The HAEC DNA established at 50–100 copies per cell could be recovered as large intact supercoiled DNA. Most cell transformants carried one or two distinct HAEC clones, and individual HAECs carried single human genomic inserts. The large genomic DNA inserts were genetically distinct and derived from different regions of the human genome, and appeared genetically stable. In summary, the multicopy DNA, clonality, and stability of the HAEC system makes it suitable for cloning

Methods in Molecular Genetics, Volume 8

various large human sequences in human cells. It provides an experimental approach for the functional study of large DNA regions in human cells and potentially for episome-based gene therapy.

Strategy for Construction of Human Artificial Episomal Chromosomes

Figure 1 shows the scheme for the construction of HAECs. The HAEC vector pH211 carries the EBV latent replication origin, oriP, which permits the propagation and stable episomal maintenance of 100–300 kb of HAECs in human cells. The EBV transactivator EBNA-1, which is required for *oriP* function, is provided *in trans* from the resident EBV in host cells. The hygromycin resistance gene allows the selection of transformant cells carrying HAECs. The expressible *lacZ* gene acts as a reporter gene. In summary, the HAEC vector is ligated to partially digested human genomic DNA of 100–400 kb. The unligated vector DNA and genomic DNA with sizes smaller than 100 kb are removed by a sucrose gradient. Ligated DNA (100–300 kb) is introduced into human cells by lipofection, which permits the transfer of large DNA into human cells. The cell transformants are selected by growing cells in the presence of hygromycin. The following sections detail the procedures for the construction of a HAEC library and isolation of individual clones.

Preparation of High Molecular Weight Genomic DNA

Cultured cells, such as the EBV-negative human lymphoblastoid cell line Ramos, are used for source DNA. High molecular weight (HMW) genomic DNA is encapsulated in agarose microbeads; therefore it is much easier to manipulate and it can be pipetted without DNA shearing. Because microbeads have the advantage of a large surface area and short surface-to-center distance, restriction enzymes can easily access DNA in beads and result in good partial digestion. High molecular weight DNA in agarose beads is isolated as described (11–13) with modification. This procedure produces HMW genomic DNA with very large sizes (>5 megabase pairs), free of nuclease contamination and of good yield.

1. Grow 100 ml of Ramos cells in RPMI 1640 with 10% fetal bovine serum (FBS), pen–strep (penicillin–streptomycin), and glutamine to a density of 2.0×10^6 cells/ml with total cell number of 2×10^8.

2. Collect the cells and pipette the cell suspension several times with a 25-ml pipette to break any cell clumps. Transfer to two 50-ml tubes and pellet at 1500 rpm for 10 min at 4°C. Pour off the supernatant, wash the cells once with an equal volume of phosphate-buffered saline (PBS) without calcium and magnesium, spin down the

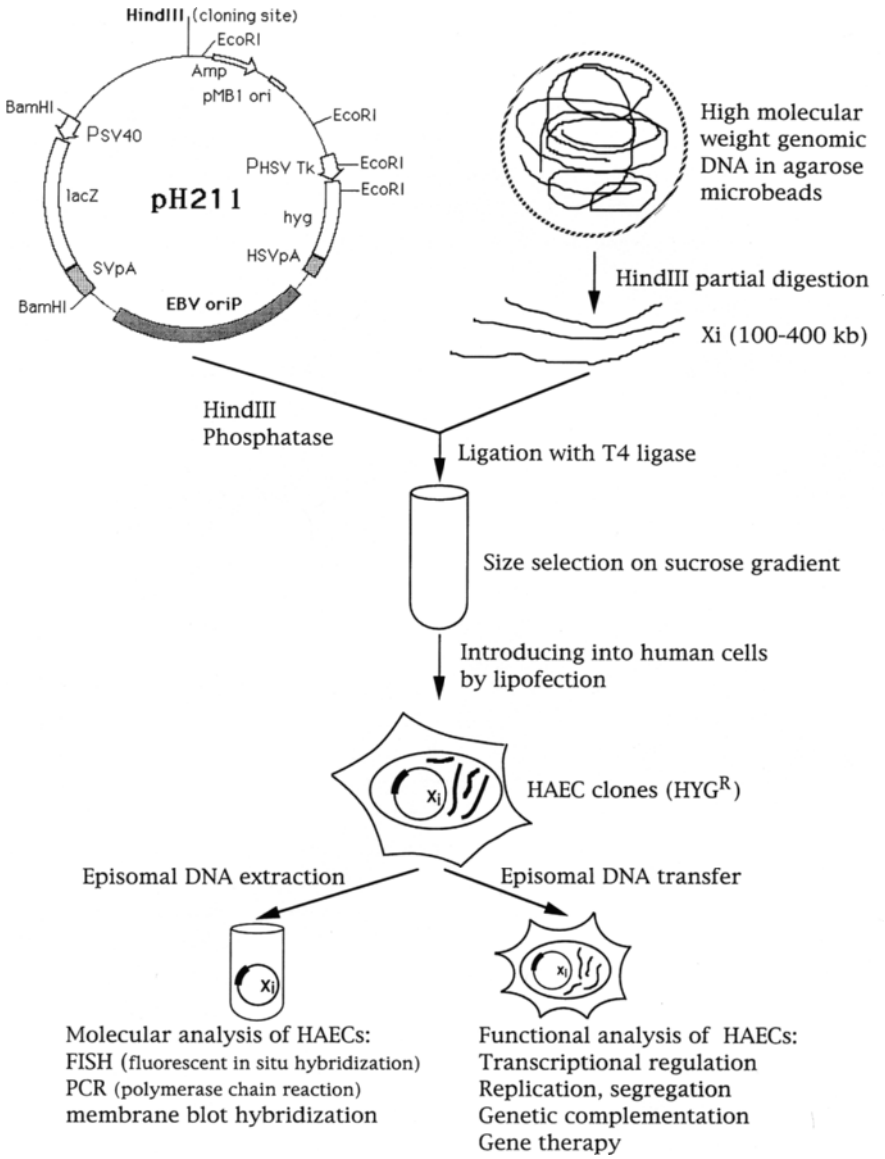


FIG. 1 Schematic illustration of the strategy for construction of human artificial episomal chromosomes. The HAEC vector, pH211, carries the EBV latent replication origin (*oriP*), hygromycin resistance gene (*hyg*) driven by SV40 promoter (P_{SV40}) and SV40 poly(A) site, and β-galactosidase gene (*lacZ*) driven by HSV TK promoter ($P_{HSV\ TK}$) and HSV TK poly(A) site. A unique *Hin*dIII site is the cloning site. B, *Bam*HI site; E, *Eco*RI site.

cells, and discard the supernatant. Resuspend the cell pellet in about 4.7 ml of PBS to achieve 4×10^7 cells/ml. The total volume is 5 ml.

3. Place a 50-ml conical tube containing 5 ml of 1.6% (w/v) molten agarose (low melting temperature Sea Plaque agarose, GTG grade; FMC, Philadelphia, PA) in PBS and a bottle containing 20 ml of mineral oil (light white oil, molecular biology reagent, DNase free; Sigma, St. Louis, MO) in a 65°C water bath. Place a 250-ml Erlenmeyer flask containing 100 ml of PBS in an ice bucket with ice. Place the ice bucket on a stirplate set to moderate speed and add a stir bar to the flask.

4. Transfer 5 ml of cells (2×10^8 cells) to a 125-ml Erlenmeyer flask and place it in a 65°C water bath for 1.5 min.

5. Add the 5 ml of 1.6% (w/v) molten agarose to the cells, mix well by swirling, and pour the 20 ml of light mineral oil into the cells while the cells are still in the 65°C water bath.

6. Take the flask containing the cells out of the water bath and swirl by hand as fast as possible at room temperature for 45 sec. The mixture should form a uniform emulsion by the vigorous swirling. Quickly pour the emulsion into the 100 ml of ice-cold PBS while stirring. Large globules should form immediately. Stir for 10 min on ice.

7. Pour them into three 50-ml conical tubes and spin at 1600 rpm (approximately $500 \times g$) for 10 min at room temperature in a centrifuge with a swinging-bucket rotor.

8. Aspirate the mineral oil at the top of each tube and the PBS in the middle of each tube. Usually there will be a layer of beads directly beneath the mineral oil and at the bottom of the tube. The beads in both levels are dispersed by repeated pipetting with a disposable 25-ml plastic pipette that has a wide orifice.

9. Add 20–30 ml of fresh PBS to each tube and invert (gently) several times. Spin down the beads at 1600 rpm for 10 min at room temperature. After this centrifugation, most beads should be in the pellets. Remove the supernatant and any beads that are not pelleted, and combine the remaining pellets in a single tube and spin as before. Remove the supernatant and wipe the inside of the tube with a tissue to remove any excess mineral oil. About 10 ml of agarose bead pellet should be obtained.

10. Resuspend the bead in 40 ml (4 vol) of lysis solution [1% (w/v) sodium dodecyl sulfate (SDS), 25 m*M* EDTA (pH 8.0), 10 m*M* Tris (pH 8.0)] by pipetting five times with a 25-ml pipette (wide mouth) to break the agarose clumps. Leave at room temperature for 30 min with occasional inversion to maintain the beads in suspension.

11. Spin down as before and resuspend the beads in 40 ml (4 vol) of proteinase K solution [1% sodium (v/v) *N*-lauroylsarcosinate, 25 m*M* EDTA (pH 8.0), proteinase K (50 μg/ml)] by pipetting five times with a wide-mouth 25-ml plastic pipette. Incubate at 50°C overnight.

12. The next morning, pellet the agarose beads as before and aspirate the supernatant. The supernatant becomes very viscous due to HWM DNA releasing from

cells embedded on the surface of the beads. Resuspend the beads in 40 ml of TE 7.5 [10 m*M* Tris (pH 7.5), 1 m*M* EDTA (pH 8.0)] and add 40 μl of 100 m*M* phenylmethylsulfonyl fluoride (PMSF; final concentration 100 μM). Invert the tube several times and pipette five times with a 25-ml pipette with a large orifice to break large clumps. Spin down and aspirate the supernatant as before. Repeat the wash with TE 7.5 containing 100 μM PMSF three more times.

13. After the last spin, aspirate the supernatant. About 7–10 ml of agarose beads should be recovered. Add TE to 20 ml and disperse the beads as before. Store at 4°C. The HMW DNA should be stable for several months.

14. The HMW DNA is now in 0.8% (w/v) LMT agarose beads (0.2–0.4 mm diameter). The DNA concentration is estimated from the number of cells (2×10^7 cells/ml) to be about 100 μg/ml solid agarose beads. About 1000 μg of HMW DNA can be isolated from 2×10^8 cells.

15. The prepared HMW DNA should be examined for nuclease contamination and digestibility by restriction enzyme. Incubate the DNA in digestion buffer containing Mg^{2+} with and without restriction enzyme and analyze on a pulse-field gel along with untreated DNA. The undigested DNA, which is very large, should remain in the wells and should stay the same after incubation in restriction buffer without a restriction enzyme, but the restriction-digested DNA should all have migrated out of the wells (Fig. 2). If the DNA becomes smaller after incubation in restriction buffer without restriction enzyme, discard this HMW DNA and repeat the isolation procedure, because starting from clean HMW DNA is important for the later manipulations.

Partial Digestion of High Molecular Weight DNA

A small amount of HMW DNA is partially digested with a serial dilution of *Hin*dIII to determine the optimal conditions needed to generate approximately 100- to 400-kb DNA (Fig. 2). The partial digestion of several hundred micrograms of HMW DNA is then performed under the determined optimal conditions.

1. Just before restriction digestion, wash 500 μl of agarose beads in 1 ml of 10 m*M* Tris-HCl (pH 7.4). Resuspend the beads by pipetting several times with a 200-μl large-orifice tip. Leave at room temperature for 5 min. Spin down at 4000 rpm for 5 min at 4°C in an Eppendorf centrifuge. Aspirate the supernatant and repeat the wash two more times.

2. Combine 40 μl of agarose beads (2 μg of DNA) with 12.5 μl of magnesium-free 5× restriction buffer [50 m*M* Tris-HCl (pH 7.5), 250 m*M* NaCl, 5 m*M* dithiothreitol, bovine serum albumin (BSA, 0.5 mg/ml)] and 5 μl of enzyme containing 0, 0.01, 0.03, 0.1, 0.3, 1, 3, and 10 units of *Hin*dIII, respectively. Mix with a 200-μl large-orifice tip by pipetting up and down several times.

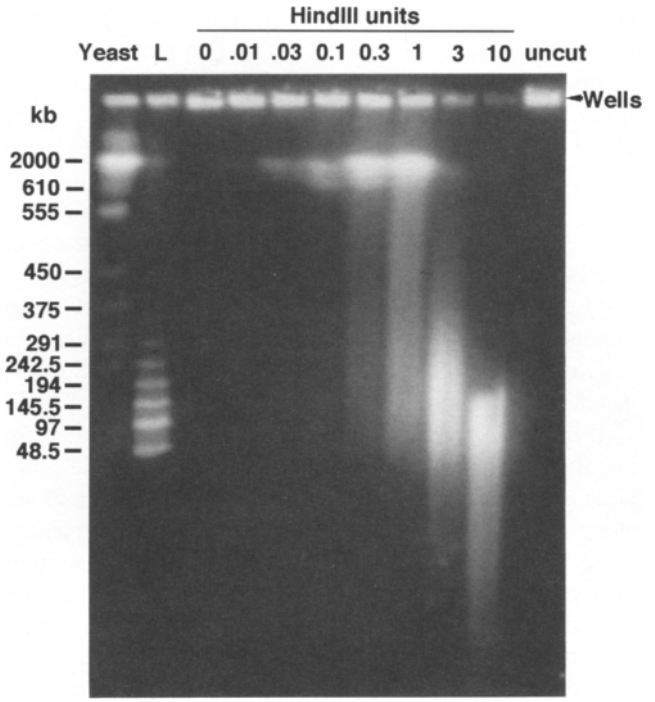


FIG. 2 Partial digestion of high molecular weight genomic DNA with a restriction enzyme. Genomic DNA in agarose beads is incubated in restriction buffer containing Mg^{2+} with 0–10 units of *Hin*dIII. The digested DNA is analyzed on a pulse-field gel with yeast chromosomes (yeast) and λ ladder (L) as molecular weight markers. An untreated DNA (uncut) is also included to examine whether the size of DNA changes after incubation in restriction buffer containing Mg^{2+} but no enzyme.

3. Incubate at 37°C for 30–45 min to allow the enzyme to diffuse into the agarose beads and then put on ice for 5 min. Add 5 μl of 120 m*M* $MgCl_2$, mix with a large-orifice tip, and leave on ice for 20–30 min. Incubate at 37°C for 15 min and then immediately put on ice. Add 12 μl of 0.5 *M* EDTA, pH 8.0, to stop the reaction.

4. Analyze 40–50 μl of digested DNA on a pulse-field gel with a short run.

5. Use the condition that produces most 100- to 400-kb partially digested DNA, in this case, 3 units of *Hin*dIII per reaction (Fig. 2), as a starting point to finely titrate the optimal condition. Set up a second round of reactions, the same as above, but each containing 2, 2.5, 3, 3.5, and 4 units of *Hin*dIII, respectively. Analyze the digested DNA on a pulse-field gel.

6. Choose the condition that gives the most 100- to 400-kb DNA and scale up the reaction to digest 400 μg of HMW DNA in two reactions, each containing 200 μg

of DNA. Stop the reactions with EDTA as above. Verify the partial digestion on a pulse-field gel.

7. Release the DNA from the agarose beads by digesting with β-agarase I. To each 200 μg of digested DNA, add 300 μl of 10× β-agarase I buffer and water to a final volume of 3 ml. Incubate at 65°C for 8 min to melt the agarose. Agitate gently with a pipette tip and incubate at 65°C for another 8 min.

8. Immediately transfer to a 40°C water bath, add 20 units of β-agarase I (1 U/μl; New England Biolabs, Beverly, MA) and incubate for 2 hr. Then incubate on ice for 20 min to solidify any undigested agarose. Spin at 15,000 rpm for 20 min at room temperature.

9. Gently transfer the supernatant to a dialysis bag (molecular weight cutoff: 12,000–14,000) with a large-orifice pipette tip. Dialyze against 1× TE.

10. If necessary, concentrate the DNA by ultrafiltration. Place the DNA in an ultrafiltration bag (collodion concentration thimble from Spectrum, Inc., Houston, TX) and dialyze against TE under house vacuum. Following this procedure, DNA can be concentrated several-fold without significant loss. After the ultrafiltration, recover the DNA from the dialysis bag. Rinse the dialysis bag with fresh TE using a large-orifice pipette tip to recover DNA that sticks on the membrane during ultrafiltration.

Ligation to HAEC Vector

The linearized HAEC vector is dephosphorylated to prevent formation of concatemers of vector that would later be difficult to separate from the desired ligation products by size fractionation. The vector DNA is ligated to *Hin*dIII partially digested DNA with sizes of 100–400 kb. The ligation products are then size fractionated on a sucrose gradient to remove unligated vector DNA.

Preparation of Vector DNA

1. HAEC vector DNA (500 μg) is digested by 1500 units of *Hin*dIII [Promega (Madison, WI) high concentration *Hin*dIII] in a 1000-μl reaction at 37°C for 2 hr. Run a minigel to check the completion of digestion. Inactivate the enzyme at 75°C for 15 min.

2. To test the optimal condition for dephosphorylation, 4 μg of linearized vector DNA is incubated with 0.001, 0.005, 0.01, and 0.05 U of calf intestinal alkaline phosphatase (CIP), respectively, in a 10-μl reaction at 37°C for 45 min. Then extract the DNA once with phenol and once with chloroform, precipitate with ethanol, and resuspend in 5 μl of TE.

3. In a 10-μl ligation reaction, 0.2 μg of CIP-treated vector DNA is incubated with 1.5 U of ligase at 16°C for 12 hr to examine the inability to ligate between ends following CIP treatment (Fig. 3). In parallel, 0.2 μg of CIP-treated vector and 0.5 μg of linearized vector DNA that has not been treated with CIP are incubated in a 10-μl

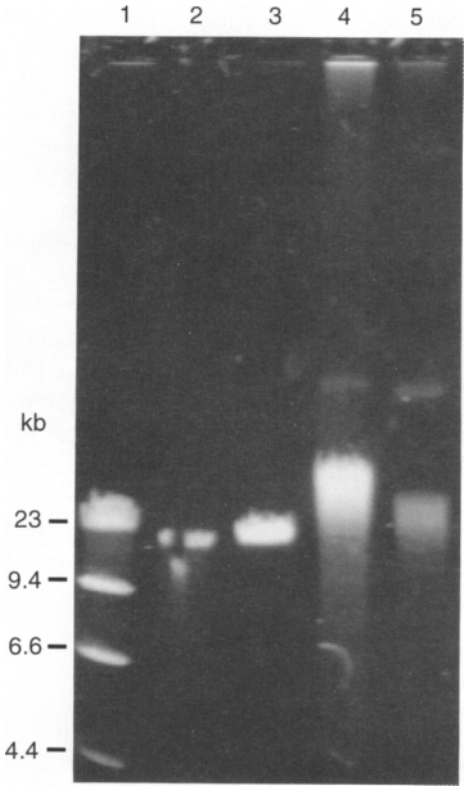


FIG. 3 Testing the ligation of CIP-treated HAEC vector ends. Ligation reaction is set up to check the effect of CIP treatment on (a) ligation between two treated *Hin*dIII ends and (b) ligation between treated and nontreated ends. Lane 1, λ *Hin*dIII markers; lane 2, *Hin*dIII linearized HAEC vector; lane 3, ligation between two CIP-treated ends; lane 4, ligation between CIP-treated and nontreated ends; lane 5, ligation between two nontreated ends.

reaction as above to check the ability to ligate between CIP-treated ends and nontreated ligatable ends (Fig. 3).

4. Choose the condition that completely dephosphorylates ends with a minimal amount of CIP and does not yield damaged ends. Scale up the CIP reaction to treat 450 μg of *Hin*dIII-linearized HAEC vector. Extract with phenol and chloroform and precipitate with ethanol. Resuspend the DNA in 100 μl of 0.5× TE.

Ligation between Inserts and Vector

Ligate the prepared vector DNA to partially digested human DNA at a molar ratio of approximately 10 to 1 in excess of vector DNA. Set up two ligation reactions, each

containing 160 μg of *Hin*dIII partially digested human DNA (100–400 kb), 200 μg of *Hin*dIII-linearized and CIP-treated HAEC vector DNA, and 50 units of T4 ligase in 1000 μl of ligation buffer [50 m*M* Tris (pH 7.5), 10 m*M* $MgCl_2$, 1 m*M* ATP, and 10 m*M* dithiothreitol]. Incubate the reactions at 16°C for 24 hr. Stop the reactions at 75°C for 15 min.

Size Fractionation of Ligated DNA by Sucrose Gradient

The ligation products are size fractionated on an exponential 5–25% (w/v) sucrose gradient as previously described (14) to remove unligated vector DNA and smaller genomic DNA. A two-chamber gradient maker with a closed mixing chamber with a volume only slightly larger than the volume of the gradient is used to generate gradients.

1. Place the gradient maker on a stir plate. Add 34 ml of 5% (w/v) sucrose gradient solution [0.8 *M* NaCl, 0.02 *M* Tris-HCl (pH 8.0), 0.01 *M* EDTA (pH 8.0)] in the closed mixing chamber (chamber A) with a stir bar. Add 34 ml of 25% (w/v) sucrose gradient solution (same as above) in another chamber (chamber B). Pump air into chamber B at a flow rate of 2–3 ml/min and let 25% (w/v) sucrose solution enter the closed chamber to mix with the 5% (w/v) sucrose while the stir bar is spinning. Displace 34 ml of sucrose solution into a 36-ml ultracentrifuge tube (polyallomer, 1 × 3.5 inches; Beckman, Fullerton, CA) through outlet tubing that is placed down the bottom of the centrifuge tube. A total of four 34-ml sucrose gradients are required.
2. Overlay 500 μl of ligation product on top of each of these gradients, using a large-orifice pipette tip.
3. Spin the gradients at 10,000 rpm in a swinging bucket rotor (Beckman SW28) for 24 hr at 15°C with slow acceleration and slow deceleration.
4. Place a rubber stopper with an inlet for mineral oil in the top of the centrifuge tube.
5. Puncture the bottom of the tube with a 16-gauge hypodermic needle. Caution is necessary to avoid introduction of air bubbles.
6. Collect 750-μl fractions into the well of 24-well plates by slowly pumping oil in the top of the centrifuge tube.
7. A 40-μl aliquot is taken from every other fraction as a sample. A total of 19 samples are taken and analyzed on a pulse-field gel to examine the size fractionation (Fig. 4).
8. Fractions of 100 to 300 kb are pooled and dialyzed extensively against TE at 4°C.

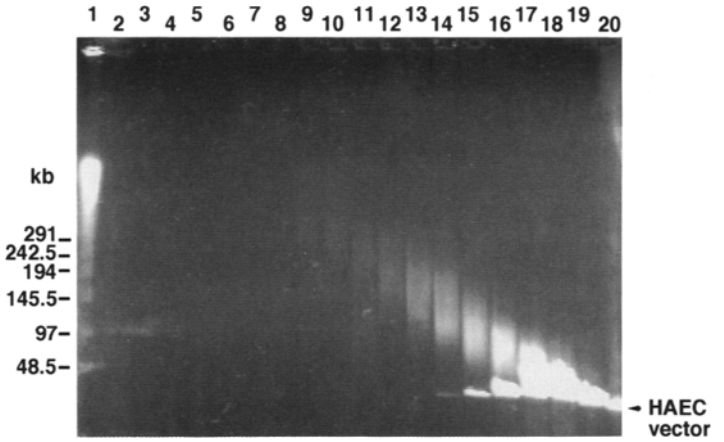


FIG. 4 Size fractionation of *in vitro*-constructed large DNA on a sucrose gradient. The ligated large genomic DNA is separated on a 5–25% (w/v) sucrose gradient and about 45 fractions are collected from the bottom of the sucrose gradient. Samples, taken every other fraction, are analyzed on a pulse-field gel with λ concatemers as molecular weight markers (lane 1). Lanes 2 to 20: fractions 6, 8, 10, 12, 14, and so on to fraction 42. The arrow indicates the unligated HAEC vector DNA (20.3 kb).

Introducing Size-Selected DNA Ligation Products into Cells by Lipofection

One hundred- to 300-kb size-selected DNA is introduced into the human cell line D98/Raji by lipofection. D98/Raji is a somatic cell hybrid between human HeLa cell variant D98 and the EBV-positive lymphoid cell line Raji (15). The resident EBV in D98/Raji provides the viral transactivator EBNA-1 *in trans* for the EBV latent replication origin, *oriP,* which is carried by the HAECs and required for the propagation of HAEC.

1. Grow D98/Raji cells in Eagle's minimum essential medium (MEM) supplemented with 10% (v/v) fetal bovine serum (FBS).
2. The day before transfection, seed cells into a six-well plate (35-mm-diameter well) at 5×10^5 cells in 3 ml of medium per well.
3. In one well of a separate six-well plate, gently mix 5 μg of size-selected, ligated DNA with 300 μl of 5× EMEM medium (prepared from powder medium) and add water to a final volume of 1.5 ml. In another well, add 15 μl of lipofectin reagent (GIBCO/BRL , Gaithersburg, MD) into 1.5 ml Opti-MEM medium (GIBCO/BRL). Gently mix DNA in 1.5 ml of 1× EMEM and lipofectin in 1.5 ml of Opti-MEM to form the DNA–liposome complex. Leave at room temperature for 15 min with gentle rocking every 3 min.

4. Meanwhile, wash D98/Raji cells, grown on a six-well plate, twice with PBS and once with Opti-MEM.

5. Add 3 ml of DNA–liposome complex to D98/Raji cells per well and incubate in 5% CO_2 at 37°C for 20 hr. Remove the DNA–liposome mix and add 3 ml of fresh EMEM with 10% (v/v) FBS. Let the cells grow overnight.

6. Trypsinize the cells and seed to one 100-mm dish in 10 ml of EMEM with 10% (v/v) FBS per well of a six-well plate. Add hygromycin to 200 μg/ml the next day to select transformants. *Note:* Hygromycin-resistant colonies will appear 2 weeks after transfection. Approximately 200–300 transformant clones can be obtained from 5 μg of size-selected ligated DNA.

7. About 3 weeks after transfection, pick up 5–10 individual hygromycin resistant clones from one 100-mm dish. Using a pipetman with a 200-μl large-orifice tip, overlay 10 μl of 1× trypsin-EDTA on a colony, pipette up and down several times while scratching the colony with the pipette tip, and recover the dispersed cells with the pipetman. Seed cells from one colony to 1 well of a 24-well plate and add 0.5 ml of EMEM with 10% (v/v) FBS. Incubate the cells at 37°C in a CO_2 incubator.

8. Trypsinize the rest of the colonies together, seed to a new 100-mm dish and grow as a subpool of the HAEC library.

A total of 25 μg of size-selected ligated DNA was transfected into 5×10^6 D98/Raji cells grown in five dishes. Three days after transfection, hygromycin was added to a final concentration of 200 μg/ml.

Analysis of Human Artificial Episomal Chromosomes

Isolation of Large Episomal DNA from Cells

Large supercoiled episomal DNA is isolated by the alkaline lysis method modified from a previous procedure (16). Up to 350 kb of intact supercoiled DNA can be isolated by this procedure.

1. Grow HAEC-transformed cells in the presence of hygromycin (200 μg/ml) in 150-mm plates until confluency. Change the medium and let grow for another day to achieve higher cell density. Up to $3–4 \times 10^7$ cells can be obtained from one 150-mm plate.

2. The next day, trypsinize the cells, break cell clumps by pipetting, resuspend in medium with 10% (v/v) FBS, and count the number of cells with a hemocytometer.

3. Pellet the cells at 1200 rpm for 5 min at room temperature; discard the supernatant. Wash the cells in an equal volume of PBS without calcium and magnesium, then pellet the cells at 1200 rpm for 5 min at room temperature; discard the supernatant.

4. Resuspend the pellet in an appropriate volume of PBS to bring the cells to 1×10^7 cells/ml. Leave the cells on ice.

5. Transfer 5 ml of cells (5×10^7 cells) to a 50-ml screw-cap conical tube (a maximum of 15 ml of cells can be processed in one tube), pellet at 1200 rpm for 5 min at room temperature, and carefully aspirate the supernatant.

6. Resuspend the cells by vortexing and make sure to break all cell clumps. Add 300 μl of H_2O and vortex again. Add 5 ml of freshly made lysis solution using a Pipet Aid (Drummond Scientific Co., Broomall, PA) at maximum speed. Vortex at highest speed for 2 min. At this point, the white cell suspension should become clear and not viscous. If still viscous, vortex longer.

> Lysis solution: 50 m*M* NaCl, 8 m*M* EDTA (pH 8.0), 1% (w/v) SDS; adjust the pH to 12.45 by adding 10 *N* NaOH. Calibrate the pH meter with a pH 12.40 standard (50 m*M* KCl, 32.4 m*M* NaOH)

7. Incubate at 30°C for 30 min.

8. Add 1 ml of 1 *M* Tris (pH 7.0) to neutralize the DNA (should yield a final pH of ~8.5) and slowly invert the tube three to five times to mix. Add 660 μl of 5 *M* NaCl (it may become cloudy after addition of NaCl) and slowly invert the tube three to five times (it should become clear). Spool out any large clump (which comes from the unlysed cell clump) with a 200-μl tip. Add 60 μl of proteinase K (10 mg/ml) by inserting the tip into the DNA solution and slowly swirling, then mix gently by inverting three to five times.

9. Incubate at 37°C for 30 min, then at room temperature for 5 min to cool down.

10. Add 4 ml of phenol that has been saturated twice with NaCl–Tris (see below), invert very gently 30 times, and leave on its side for 8 min. Repeat this process three more times. The aqueous phase should become very cloudy at this stage. Balance the tubes with phenol and spin at 3500 rpm for 15 min at room temperature in a tabletop centrifuge with a swinging-bucket rotor. A thick white interface, which contains protein and denatured DNA, should be generated. To make NaCl–Tris saturated phenol: add TE-saturated phenol to an equal volume of NaCl–Tris solution [0.2 *M* NaCl, 0.2 *M* Tris-HCl (pH 8.0)], shake vigorously for 1 min (be sure the mixing is complete), and leave for 5–10 min; repeat this mixing once. Aspirate the aqueous phase, add an equal volume of NaCl–Tris solution, and repeat this saturation process once.

11. Carefully transfer the aqueous phase to a fresh tube, using a 5-ml disposable plastic pipette with the tip snapped off. Avoid the transfer of any white interface. The aqueous phase may turn cloudy during transfer but this does not affect the solution.

12. Repeat steps 10 and 11, two more times. After the third phenol extraction, the aqueous phase becomes translucent (less cloudy) before the spin and remains clear for a long time after the spin. Little or no white precipitate will be present at the interface. If the entire interface is still covered by the white precipitate at this point, do one more phenol extraction.

13. Gently extract once with 5 ml of chloroform–isoamyl alcohol (24:1, v/v) as in phenol extraction. At this point, the chloroform phase should be clear; a turbid

chloroform phase may indicate that the phenol extraction is not complete. Spin at 3500 rpm for 5 min at room temperature.

14. Carefully transfer the aqueous phase to a 34-ml plastic tube (for a Sorvall SS-34 rotor) as before. *Note:* A successful preparation relies on the minimum breakage of large supercoiled DNA and complete removal of denatured DNA. Therefore, handle gently when mixing and always use a disposable 5-ml plastic pipette with the tip snapped off or a large-orifice pipette tip to transfer the aqueous phase. Care should be taken to ensure complete mixing of aqueous and phenol phases. Avoid carryover of any white interface because it contains denatured DNA, which blocks large supercoiled DNA from entering the gel.

15. Add 11 ml of cold 100% ethanol (store in a freezer), mix by gently inverting 20–30 times, and leave at −20°C overnight.

16. Spin down the DNA at 12,000 rpm for 30 min in a Sorvall SS-34 rotor at 4°C. Wash the pellet with 70% (v/v) ethanol and spin at 12,000 rpm for 5 min. Carefully and slowly aspirate all supernatant and dry the pellet in air [using a Speedvac (Savant, Hicksville, NY) might nick the supercoiled DNA] until the white pellet becomes clear (it may take 10–20 min).

17. Slowly add 150 μl of TE, pH 8.0 (3 μl of DNA/10^6 cells), to the pellet and leave at room temperature for 2 hr to dissolve DNA; slowly pipette five times with a 200-μl large-orifice tip to break up any clumps and store at 4°C. It is stable at 4°C for 2 months.

18. Run 3 μl of DNA (3 μl of DNA + 6 μl of TE + 1 μl of 10× loading dye) on a 0.8% (w/v) minigel to check the quality of supercoiled DNA. *Note:* For a successfully prepared episomal DNA, an undistorted closed-circular mitochondria DNA band migrates slower than the linear DNA band; it should have little DNA stuck in the well. If there are large amounts of denatured DNA stuck in the wells, clean up the DNA by either of the following: (a) One more round of alkaline lysis. Add 10 times the volume of lysis solution to 1 vol of DNA, gently mix by inverting, and leave at room temperature for 10 min. Neutralize with Tris, pH 7.0, and add 5 *M* NaCl as above. Extract once with phenol saturated with NaCl–Tris. Extract once with chloroform–isoamyl alcohol. Precipitate the DNA with ethanol; (b) LiCl precipitation of denatured DNA. Add an equal volume of 10 *M* LiCl to the DNA, and leave on ice for 1 hr. Spin at 13,500 rpm for 20 min in an Eppendorf centrifuge. Transfer the supernatant to a fresh tube with a 200-μl large-orifice tip. Precipitate the DNA with 2.5 vol (2.5 times the volume of the supernatant) of cold 100% ethanol.

19. The DNA is ready for enzyme digestion or agarose gel analysis.

Agarose Gel Analysis of Large Supercoiled Episomal DNA

The large supercoiled episomal DNA of different sizes can be separated on an agarose gel at high voltage (Fig. 5B). For best results, run at 4°C with buffer constantly circulated.

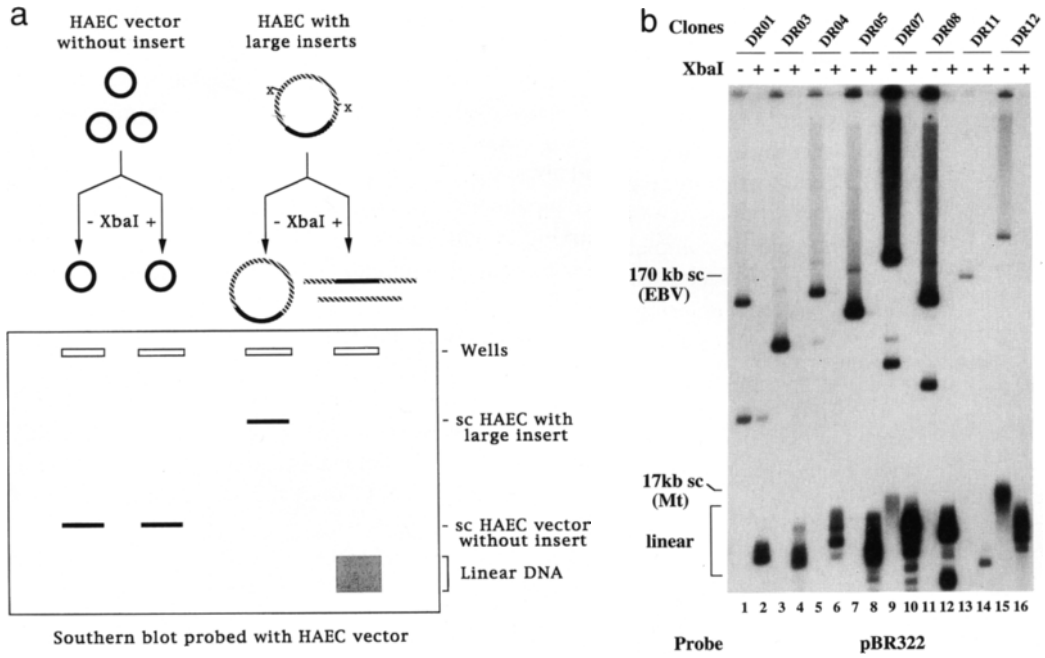


FIG. 5 Analysis of HAECs by agarose gel electrophoresis. (A) Schematic illustration of insert-cleavage assay to examine the presence of inserts in HAECs. *Xba*I, a restriction enzyme that does not cleave HAEC vector. X, *Xba*I site in HAEC inserts; sc, supercoiled. (B) Analysis of the HAECs by agarose gel electrophoresis and DNA hybridization. Episomal DNA is extracted from human cells, separated on an agarose gel before and after *Xba*I digestion. Southern analysis is performed with the ^{32}P-labeled HAEC vector-specific probe pBR322. DR01, DR03, and so on are individual HAEC clones. The positions and molecular weight of supercoiled (sc) markers are indicated to the left. When separated on an agarose gel, the supercoiled episomal DNA migrates proportionally to its size, the nicked or open circle DNAs of large size remain in the wells, while linear DNA runs ahead of supercoiled vector DNA.

1. Cast a 0.8% (w/v) agarose gel in 1× TBE in a large gel tray (e.g., 35 × 15 cm). Place the gel in the gel box and fill with 1× TBE without ethidium bromide. Put the gel box in a 4°C cold room for 1 hr before running. Load and run the gel in the cold room.

2. Load episomal DNA in 1× loading dye with a 200-μl large-orifice tip (DNA from 1–2 × 10^6 cells/well).

3. Run the gel at 5.4 V/cm (189 V for the 35-cm-long large gel box). Begin buffer circulation 1 hr after the gel is started. Let the gel run a total of 10–11 hr.

4. After electrophoresis, stain the gel in ethidium bromide (0.5 μg/ml) for 1 hr.

5. After photography, leave the gel on the ultraviolet (UV) box (UV transilluminator, 302-nm UV) for 12 min with the UV light on to nick the DNA.

6. Treat the gel in 300 ml of alkaline solution (0.5 *M* NaOH, 1.5 *M* NaCl) for 20 min; repeat once. Neutralize the gel in 300 ml of neutralization solution [0.5 *M* Tris-HCl (pH 7.8), 1.5 *M* NaCl] for 20 min; repeat once.

7. Transfer the DNA to a nylon membrane (such as Magna Graph nylon transfer membrane from MSI, Westborough, MA) in 10× SSC, using a vacuum transfer device (vacuum blotter from Hoefer Scientific, San Francisco, CA). Transfer at 5 cmHG for 1.5 hr and at 7.5 cmHG for 0.5 hr.

8. Immobilize the DNA on the membrane by cross-linking in a UV cross-linker.

9. Wash the membrane in 5× SSC for 5 min. The membrane is ready for prehybridization.

10. Hybridize the DNA with appropriate probes.

Episomal DNA Size Measurement

The sizes of large supercoiled episomal DNA can be determined based on their migration rate in an agarose gel. When separated in an agarose gel, the supercoiled episomal DNA migrates proportionally to its size, while linear DNA runs ahead of the large supercoiled DNA and nicked open-circular DNA is retained in wells (Fig. 5B) (8, 17, 18). Using proper supercoiled DNA size standards, such as 17-kb mitochondrial DNA, 170-kb EBV DNA, and an HAEC DNA of known size, a standard curve can be generated. The sizes of HAEC DNA can be directly determined from their migration rate in the agarose gel.

Insert–Cleavage Assay to Examine Presence of Inserts in HAECs

The presence of inserts in HAECs is determined by the sensitivity to cleavage by *Xba*I, a restriction enzyme with a 6-base pair (bp) recognition sequence which is expected to cut random DNA every 4 kb on average. The HAEC vector, which does not have an *Xba*I site, will remain supercoiled (sc) and migrate at the same position, irrespective of *Xba*I digestion. The HAECs with large inserts are expected to be cleaved into smaller fragments and become fast-migrating linear DNA (Fig. 5A and B).

Gel Purification of HAEC DNA

When required to perform Alu polymerase chain reaction (PCR) analysis on HAECs or to use HAEC DNA as probes, purify supercoiled HAEC DNA on a 0.8% (w/v) low melting temperature agarose (LMT) minigel in 1× TBE (without ethidium bromide) at 4°C (Fig. 6).

1. Prepare a 0.8% (w/v) LMT minigel in 1× TBE (i.e., 8 × 7 cm). When the gel solidifies, place the gel and gel box filled with 1× TBE in the cold room.

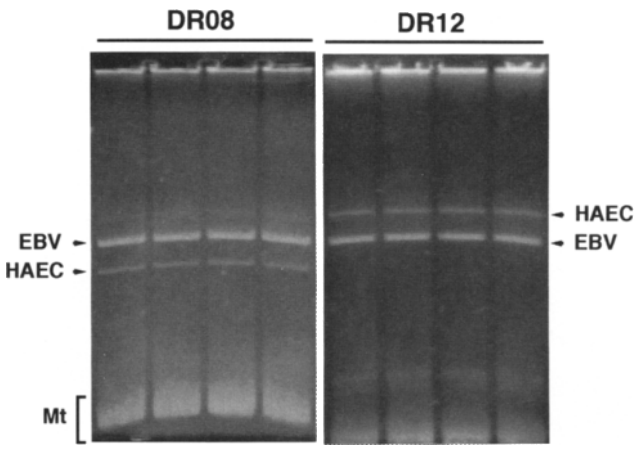


FIG. 6 Purification of HAEC DNA by agarose gel electrophoresis. Episomal DNA is extracted from individual HAEC clones and separated on a 0.8% (w/v) low melting temperature agarose gel (Sea Plaque; FMC) at 5.4 V/cm, 4°C for 7 hr. Each well (9 × 1.5 mm) is loaded with episomal DNA from 1–1.5 × 10^7 cells. After electrophoresis, the gel is stained with ethidium bromide. HAEC DNA bands are cut out from gel and recovered with β-agarase I digestion. DR08 and DR12 are HAEC clones carrying 120- and 350-kb HAECs, respectively. EBV DNA (170 kb) is present in the host cell line D98/Raji.

2. Load the episomal DNA in 1× loading dye with a 200-μl large-orifice tip (DNA from 1–1.5 × 10^7 cells in 30–40 μl/well, 9 × 1.5 mm).
3. Run the gel at 5.4 V/cm for 7–9 hr at 4°C. Mix the electrophoresis buffer in the two chambers every 1.5 hr. Change the buffer after 3- to 4-hr run.
4. Stain the gel in ethidium bromide (0.5 μg/ml) for 30 min and destain in water for 20 min.
5. Visualize the DNA on a UV box (302nm). Cut out HAEC DNA bands.
6. Recover the DNA from gel slices using β-agarase I (New England Biolabs) as recommended by the manufacturer.

Alu-PCR Analysis of HAEC DNA

Alu-PCR (19) can be performed to analyze the identity and diversity of human inserts in HAECs. A pair of *Alu* primers, which complement the 5′ and 3′ ends of the *Alu* sequence, is used to amplify sequences between two *Alu* repeats irrespective of the orientation of the two *Alu* repeats. Gel-purified HAEC DNA from individual HAEC clones is amplified with Alu-1 and Alu-2 primers (20).

1. Perform PCR reactions in 50 m*M* KCl, 10 m*M* Tris (pH 9.0), 1.5 m*M* $MgCl_2$, 0.1% (v/v) Triton X-100 Alu-1 and Alu-2 (1 μM each), 0.2 μM dNTPs and 0.125 U *Taq* polymerase in a 50-μl reaction. Each reaction includes gel-purified HAEC DNA from 1×10^5 cells as template DNA.
2. PCR conditions: 94°C for 5 min for one cycle; 94°C for 1 min, 37°C for 30 sec, and 72°C for 6 min for 35 cycles; and 72°C for 10 min for one cycle.
3. Analyze the PCR products on a 4% (w/v) polyacrylamide gel in 1× TBE. After running, stain the gel in ethidium bromide (0.5 μg/ml).

Hybridization for Studying Insert Organization

1. Digest HAEC DNA with appropriate restriction enzymes. Separate on an agarose gel and transfer to a membrane by Southern blotting.
2. Prehybridize the membrane blot in 4 ml of 7% (w/v) SDS, 0.25 *M* sodium phosphate (pH 7.2), 1 m*M* EDTA, human Cot-1 DNA (20 μg/ml; GIBCO-BRL, Gaithersburg, MD) and 5 μg of sonicated HAEC vector DNA per milliliter at 65°C for 6 hr.
3. Label 5–10 ng of gel-purified HAEC DNA with ^{32}P by random priming. Remove unincorporated nucleotides by spun column.
4. Mix the labeled probe with 80 μg of human Cot-1 (GIBCO-BRL) and 22 μg of sonicated HAEC vector DNA in a total volume of 200 μl of 5× SSC. Boil the probe for 5 min and incubate at 65°C for 20 min.
5. Add the probe to the prehybridized membrane and hybridize at 65°C for 20 to 24 hr.
6. After hybridization, wash the membrane once with 5× SSC for 5 min at room temperature, twice with 2× SSC–0.1% (w/v) SDS for 15 min at room temperature, and twice with 0.1× SSC–1% (w/v) SDS for 15 min at 65°C.
7. Expose to X-ray film with intensifying screen at −70°C.

General Discussion and Future Directions

In this chapter, we have described the experimental procedure for the construction of human artificial episomal chromosomes (HAECs). In summary, the capability to transfer, maintain, and recover large DNA segments in human cells will provide a number of opportunities for the study of the human genome, and potentially of genomes from other species. It will allow the analysis of large human genes or gene clusters in their natural configuration and genetic background, as well as the study of the influence of normally remote control elements on gene expression. The analysis of long-range effects of regulatory elements such as those involved in chromosome inactivation and imprinting may also be feasible using the HAEC system. Other po-

tential applications of the HAEC system are the identification and analysis of cis-acting DNA sequences necessary for human chromosome function, such as DNA replication and mitotic segregation.

HAEC-Based Genomic Libraries

The HAEC cloning system allows the establishment of genomic libraries with large inserts in human cells. Providing that a HAEC library covering the entire human genome is available, the introduction of such human genomic libraries into disease cells could allow for complementation of mutant phenotypes and isolation of genes. To generate a single-hit coverage of the human genome, one needs approximately 20,000 HAEC clones with an average insert size of 150 kb. A pilot human genomic library of 1500 clones with 150- to 200-kb average insert size was constructed, which covers approximately 10% of the human genome (8). Following the current HAEC cloning procedure, one needs to transfect approximately 25 μg of vector-ligated human DNA to obtain 1500 independent clones of 150-kb average insert size. Thus, the current efficiency of the HAEC system appears adequate to prepare a comprehensive library of the human genome. Future improvement with the HAEC cloning system may focus on (a) insert size, (b) transfection efficiency, and (c) region-specific HAEC libraries.

a. If clones with larger HAEC inserts can be made and manipulated, fewer numbers of clones will be required to cover the human genome. It may indeed be possible to engineer very large HAECs, that is, up to the megabase range. The current HAEC average insert size, 150 kb, has been size selected in the range of the latent EBV genome, 170 kb. The identification of HAECs up to 350 kb (8) indicated, however, that the HAEC-based system can replicate human DNA inserts at least twice the size of the natural EBV replicon. After all, episomal double-minute chromosomes in human cells can reach the megabase range (21).

b. Although the transfection efficiency obtained with the HAEC system is in the range typically observed in human cells transfected with episomal vectors using standard methods (i.e., lipofection, electroporation, or calcium phosphate coprecipitation), it is one to two orders of magnitude lower than bacterial or yeast systems. However, efficient DNA transfection procedures have become available, such as adenoviral-assisted transfection, which rely on receptor-mediated gene delivery (22, 23). The adaptation of such methods for the transfer of large HAEC DNA could decrease the amount of episomal DNA required for the construction of a complete human genome library.

c. Finally, if chromosomal or subchromosomal HAEC libraries can be made, one could isolate genes already localized by genetic and physical means to specific chromosomal regions with fewer HAEC clones.

HAEC-Based Phenotypic Rescue

The potential to transfer large HAEC inserts into appropriate mutant or disease cells could be used to isolate specific genes from HAEC libraries via genetic complementation. However, any utilization of HAEC libraries to isolate specific large human genomic inserts by phenotypic rescue/screening will require the development of methods to shuttle such libraries between cells. Episomal DNA extracted from pooled HAEC libraries could be transferred directly into specific target cells using the various transfection techniques discussed above. Alternatively, the large genome size of the Epstein–Barr virus, 172 kb, may allow packaging of the large human inserts into virions. Efficient packaging of 200-kb engineered DNA as infectious EBV using a mini-EBV vector and EBV helper cell line has been observed (24). This approach could be particularly adapted to human lymphoblastoid cells, which are readily available for most human syndromes and are efficiently infected by engineered EBV (24, 25). Indeed, *in vitro* correction of disease phenotypes has been accomplished via EBV infection of B lymphoblastoid cells from several human syndromes (26, 27). Although limited by the development of appropriate selection schemes, this complementation strategy in combination with the extensive physical and genetic maps would be of considerable utility for the identification of disease genes and their future correction via gene therapy.

Physical Screening of HAEC Libraries

The system may also provide an additional source to yeast and bacteria for large human genomic fragments. Specifically, the HAEC system may be an alternative to these single-cell systems for cloning of human DNA regions that are refractory to faithful cloning in such microbial and fungal hosts. Unlike the heterologous BAC, PAC, and YAC cloning systems, the HAEC is a complete autologous system with large human DNA segments solely cloned, propagated, and maintained in human cells. Some human DNA regions appear to be unstable, rearranged, or deleted when propagated in distinct genetic backgrounds such as yeast and bacteria. Because human DNA cloned as HAEC remains in its own genetic background, it is expected to be replicated faithfully. In addition, human genetic imprinting based on covalent modification of DNA, such as DNA methylation, should be maintained more faithfully with the human-based HAEC system than with the bacteria or yeast large cloning systems. However, any utilization of HAEC libraries to isolate specific large human genomic inserts will require the development of methods to store and screen such libraries. Arraying HAEC libraries in medium-size pools (i.e., 100–1000 clones per pool) appears the most practical. As suggested by the *Alu*-PCR fingerprinting experiment (8), positive pools could be detected via PCR-based screening of extracted HAEC DNA. Providing minimal copurification of genomic DNA, HAEC

screening by PCR on pools of this size should be feasible. Because of the high copy number of HAECs per cell and the capacity for further EBV-mediated HAEC amplification (T.-Q. Sun and J.-M. H. Vos, unpublished data, 1995), positive clones could then be screened by membrane replica plating and *in situ* hybridization (28). One additional parameter to be considered is the coexistence of more than one HAEC clone per cell. Although the HAEC system is multicopy in nature, there are only on average two unrelated HAECs per cell clone. However, such low cocloning frequency could still complicate any future screening procedure. Two factors probably contribute to this cocloning phenomenon. First, the transfection procedure via lipofection most likely transfers many DNA molecules per cell. Second, the cell line used in this study can accommodate up to 50–100 copies of EBV-based episomes (T.-Q. Sun and J.-M. H. Vos, unpublished observations, 1995). Because EBV *oriP* can function in many human cell lines with copy numbers ranging from a few to several hundreds per cell, a cell line that supports only a low copy number of EBV plasmid could be used instead. Thus, with improved transfection procedure and low copy cell lines, the cocloning problem should be resolved. If so, candidates for HAEC screening could be human regions refractory to faithful cloning in yeast or bacteria, or contiguous DNA gaps that cannot be bridged using the microbial-biased libraries.

HAEC-Based Shuttling of Microbial-Based Large Clones into Human Cells

The potential for shuttling large clones between yeast or bacteria and human cells using the HAEC system may offer the possibility of functional studies of large human inserts, and the implementation of genetic complementation strategies, using the available YAC, BAC, and PAC human genomic libraries. If so, the power of genetic manipulation and screening in these microbial systems would synergize the potential of the HAEC system for assaying large genes and other functional units of human chromosomes. It may be possible to conduct functional analysis of large DNA fragments cloned as YACs or BACs/PACs through transfer and long-term episomal maintenance of such large clones in human cells using the HAEC system. By combining HAECs with YACs, BACs or PACs, it will be possible to shuttle large human DNA fragments between the microbial hosts and human cells, and take advantage of the simple genetic manipulation and physical screening strategies available in these unicellular organisms together with the power of functional analysis in human cells.

Engineering HAECs for Gene Therapy

As a long-term approach, one could use the large cloning potential of the HAEC system to isolate functional human centromeres. Despite extensive progress on the isolation of centromeric sequences and mapping of functional human centromeres

(29), the DNA required for an active human centromere has not yet been determined. Randomly chosen HAEC clones from the available human genomic library cover various regions of the human genome. The identification in the current HAEC library of HAEC clones spanning human centromeres (8) suggests that such potential exists. Thus, in addition to offering an experimental approach to study large functional units, such as genes and, potentially, replication units in human cells, the HAEC cloning system may allow the development of human artificial chromosomes. By analogy to the YAC system, the addition of a functional human centromere and telomeric ends on the HAEC vector would permit the establishment of single-copy, stable, and linear minichromosomes in human cells. Engineering large human artificial chromosomes would set the stage for the development of gene therapy strategies for long-term disease correction based on rationalized extrachromosomal elements.

Acknowledgments

We thank the members of our laboratory, especially Linda Briley for laboratory assistance and Dr. Michael Grosz for critical reading of the manuscript. This work was supported by the DOE Human Genome Initiative Grant DEFG0591ER611350, T.-Q.S. was recipient of a Lineberger Cancer Center Graduate Student Award from the School of Medicine at the University of North Carolina at Chapel Hill. J.-M.H.V. was recipient of an American Cancer Society Junior Faculty Award and a Basil O'Connor Starter Scholar Research Award from the March of Dimes Birth Defects Foundation.

References

1. A. W. Murray and J. W. Szostak, *Nature (London)* **305,** 189 (1983).
2. D. T. Burke, G. F. Carle, and M. V. Olson, *Science* **236,** 806 (1987).
3. M. O'Connor, M. Peifer, and W. Bender, *Science* **244,** 1307 (1989).
4. F. Hosoda, S. Nishimura, H. Uchida, and M. Ohki, *Nucleic Acids Res.* **18,** 3863 (1990).
5. H. Shizuya, B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiri, and M. Simon, *Proc. Natl. Acad. Sci. U.S.A.* **89,** 8794 (1992).
6. N. Sternberg, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 103 (1990).
7. P. A. Ioannou, C. T. Amemiya, J. Garnes, P. M. Kroisel, H. Shizuya, C. Chen, M. A. Batzer, and P. J. de Jong, *Nature Genet.* **6,** 84 (1994).
8. T.-Q. Sun, D. A. Fenstermacher, and J.-M. H. Vos, *Nature Genet.* **8,** 33 (1994).
9. J. L. Yates, N. Warren, D. Reisman, and B. Sugden, *Proc. Natl. Acad. Sci. U.S.A.* **81,** 3806 (1984).
10. S. Lupton and A. J. Levine, *Mol. Cell. Biol.* **5,** 2533 (1985).
11. D. A. Jackson and P. R. Cook, *EMBO J.* **4,** 913 (1985).
12. J. Overhauser and M. Z. Radic, *Focus* **9,** 8 (1987).
13. T. Imai and M. V. Olson, *Genomics* **8,** 297 (1990).
14. D. T. Burke and M. V. Olson, *Methods Enzymol.* **194,** 251 (1991).

15. R. Glaser and M. Nonoyama, *J. Virol.* **14,** 174 (1974).
16. S. M. Carroll, P. Gaudray, M. L. De Rose, J. F. Emery, J. L. Meinkoth, E. Nakkim, M. Subler, D. D. Von Hoff, and G. M. Wahl, *Mol. Cell. Biol.* **7,** 1740 (1987).
17. S. Mickel, V. Arena, and W. Bauer, *Nucleic Acids Res.* **4,** 1465 (1977).
18. T. Eckhardt, *Plasmid* **1,** 584 (1978).
19. D. L. Nelson, S. A. Ledbetter, L. Corbo, M. F. Victoria, R. Ramirez-Solis, T. D. Webster, D. H. Ledbetter, and C. T. Caskey, *Proc. Natl. Acad. Sci. U.S.A.* **86,** 6686.
20. P. Liu, J. Siciliano, D. Seong, J. Craig, Y. Zhao, P. J. de Jong, and M. J. Siciliano, *Cancer Genet. Cytogenet.* **65,** 93 (1993).
21. J. C. Ruiz, K. H. Choi, D. D. von Hoff, I. B. Robinson, and G. M. Wahl, *Mol. Cell. Biol.* **9,** 109 (1989).
22. Z. T. Kelleher and J.-M. H. Vos, *BioTechniques* **17,** 1110 (1994).
23. D. Curiel, *in* "Viruses in Human Gene Therapy" (J.-M. H. Vos, ed.), p. 164. Academic Press, Durham, North Carolina, 1994.
24. T.-Q. Sun and J.-M. H. Vos, *Int. J. Genome Res.* **1,** 45 (1992).
25. J.-M. H. Vos, *in* "Viruses in Human Gene Therapy" (J.-M. H. Vos, ed.), p. 109. Academic Press, Durham, North Carolina, 1995.
26. S. Banerjee, E. Livanos, and J.-M. K. Vos, *Nature Medicine* **1,** 1303 (1995).
27. T.-Q. Sun, E. Livanos, and J.-M. K. Vos, submitted, 1996.
28. H. Avraham, S. Avraham, and S. C. Bernstein, *Anal. Biochem.* **179,** 217 (1989).
29. C. Tyler-Smith and H. F. Willard, *Curr. Opin. Gen. Dev.* **3,** 390 (1993).

[10] Isolating and Mapping Coding Regions from Complex Genomes: Direct cDNA Selection

Carol A. Wise, Ethylin Wang Jabs, and Michael Lovett

Introduction

As detailed physical maps of mammalian genomes come closer to completion, the next logical step in the human genome project is to use these maps to isolate and map coding sequences. Annotated maps of this type not only add to the biological information content of the physical map, but are of enormous utility for investigators involved in the positional cloning of genetic disease loci. Finding coding regions in complex genomes, which contain a small percentage of coding sequence amid a sea of noncoding sequence, has been a formidable task. In positional cloning projects this usually consumes the majority of the time and effort in projects that can last several years between genetic linkage and discovery of a mutated gene. Several methods have been developed to address this problem specifically (1–5). Here we describe direct cDNA selection, a polymerase chain reaction (PCR)-based method developed in our laboratory to enrich for coding sequences from large regions of the human genome. The introduction to this chapter gives an overview of the method, including the rationale for various steps. The Methods section outlines a detailed protocol, and the Discussion addresses special issues in the analysis of selected cDNAs.

General Principles

One major limitation in isolating particular human genes as cDNAs is the abundance problem—some genes are expressed at levels as low as 1 in 10^6, making it difficult to identify these transcripts by conventional screening of a cloned cDNA library. This problem is exacerbated when the target transcript is expressed at a low level only in a small subpopulation of cells within a complex tissue. In these cases screening a conventional cDNA library is unlikely to be productive. Another critical limitation is the need for a cDNA source; if one does not have some *a priori* idea of which tissue or developmental stage to search within, then the search may be fruitless. This is particularly relevant when one wishes to build a "complete" transcription map of a genomic region. If the region contains genes that are spatially and/or temporally regulated then many of these may be missed if the search is confined to a few tissues. For this reason it is unlikely that any cDNA-based approach by itself will identify

Methods in Molecular Genetics, Volume 8

"all" of the genes in the genome. Exon trapping and genomic DNA sequencing are not subject to these limitations and can together (in theory) identify all human genes. However, because these techniques are still less than 100% accurate, a "Catch 22" situation exists in which the validation of a putative exon is dependent on identifying a corresponding mRNA or cDNA clone. Direct cDNA selection has several important limitations (see Ref. 6 for a more detailed discussion): it is PCR based and relies on a cDNA source, but it can circumvent the aforementioned abundance problems and lead to the efficient isolation of very low-abundance cDNAs. Briefly, a genomic target, typically a yeast artificial chromosome (YAC) or cosmid (see below), is biotinylated and hybridized in solution to a cDNA pool that is uncloned, but is ligated to oligonucleotide adapters that can be used as priming sites in the PCR. These uncloned pools of cDNAs are important because they preserve the sequence complexity of the starting source; a cDNA present at 1 in 10^7 would probably not be present in a conventional library of 1 million cDNA clones, but would still be present in a cDNA pool. Selections are initially conducted under conditions such that the genomic target is in molar excess over its homologous low-abundance cDNA. After hybridization to an intermediate C_0t value, the hybrids are captured on streptavidin-coated magnetic beads and nonhybridizing cDNAs are washed away. The hybridized cDNAs are eluted off and amplified in the PCR with primers complementary to the end linkers (Fig. 1). Typical enrichments of low-abundance cDNAs in this first step are 1000- to 10,000-fold (5, 7). The amplified cDNAs are usually selected a second time for a further 10- to 20-fold enrichment. Secondary rounds of selection are generally conducted under conditions in which genomic DNA is limiting. In this way high-abundance cDNAs have few targets with which to hybridize, and conversely low-abundance cDNAs have many targets, thus enriching low-abundance cDNAs and reducing highly abundant ones. This rather simplistic theory ignores the problems inherent in having multiple exons, gene families, or pseudogenes present in some contiguous sequences of DNA (contigs), but nevertheless to a first approximation it appears to work. By playing this hybridization-based game of "musical chairs," it is possible to reduce the relative abundance of a cDNA that comprises a large proportion of the primary selected cDNAs and to increase the relative abundance of a cDNA that was less efficiently selected in the primary selection. Thus, some cDNAs show a dramatic primary enrichment and then fall slightly, whereas others show more modest primary enrichments, but continue to increase in the secondary selections. Some examples of this are shown in Fig. 2. Three rounds of selection, in our experience, do not lead to any noticeable improvements (Fig. 2C).

Genomic Targets

We have used a number of genomic targets in selection. Our original description of the technique (5) utilized a 550-kb purified YAC containing the EPO gene encoded by chromosome 7 and a 20- to 40-kb cosmid containing the interleukin 3 (IL-3)

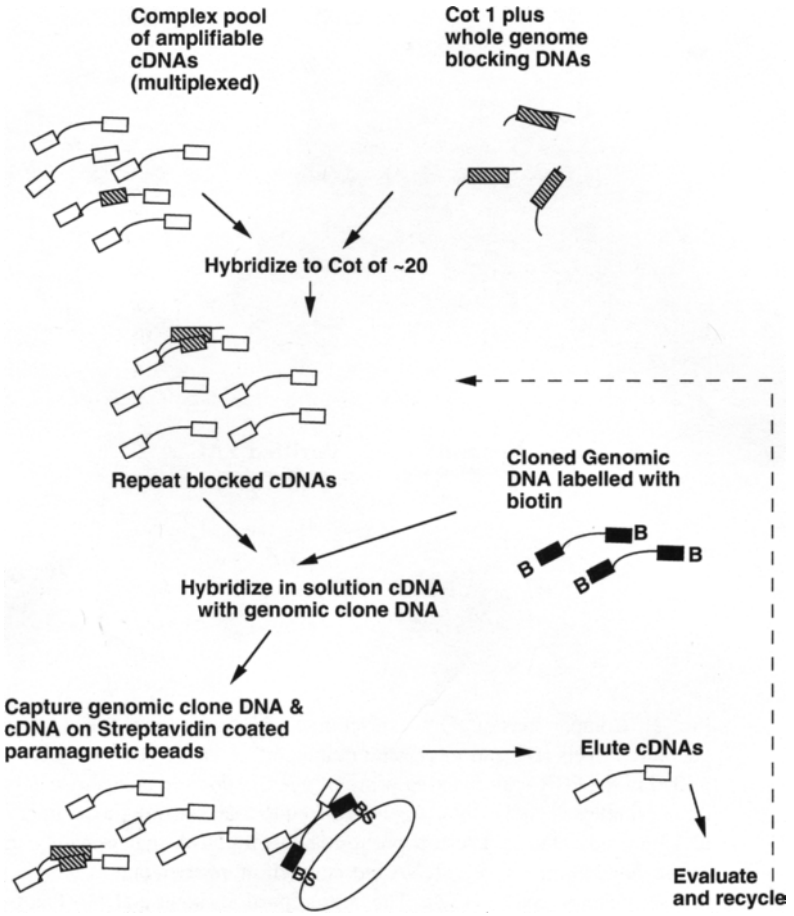


FIG. 1 Direct cDNA selection. Complex pools of multiplexed, amplifiable cDNAs are preblocked in solution to Cot-1 (GIBCO-BRL) as well as genomic blocking DNAs (both shown as shaded boxes) to a C_0t of approximately 20 mol nucleotide liter^{-1} sec. Cloned, biotinylated genomic DNA (shown as black boxes) is then hybridized in solution to the repeat blocked cDNAs to an intermediate $C_0t_{1/2}$ (approximately 100 mol nucleotide liter^{-1} sec). Genomic DNA–cDNA hybrids are then captured on streptavidin-coated magnetic beads; nonhybridizing cDNAs (including repeat blocked cDNAs shown) are washed away. The captured cDNA is eluted, amplified in the PCR, evaluated for enrichment of reporter genes, and recycled through the selection procedure.

and granulocyte/macrophage colony-stimulating factor (GM-CSF) genes encoded by chromosome 5. These genomic targets were bound to filters in the selections, which yielded approximately 1000-fold enrichments of the monitored genes after 1 round of selection. Hybridizations on filters were subsequently modified to solution hybrid-

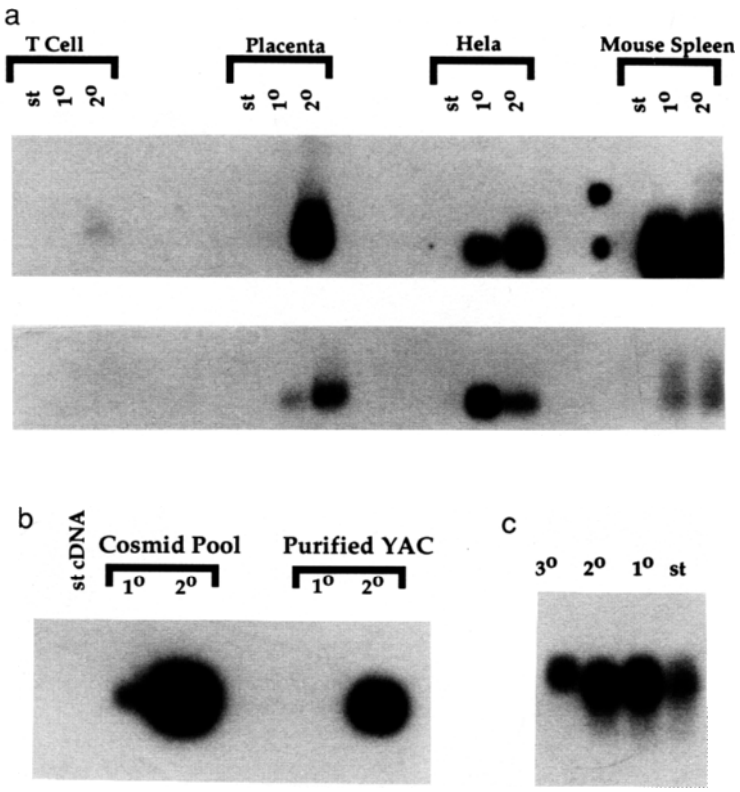


FIG. 2 Comparisons of cDNA enrichments from various tissues (A), using YAC or cosmid genomic targets (B), and increasing rounds of selections (C). (A) Selected material was amplified in the PCR with flanking primers, electrophoresed on agarose gels, and Southern blotted. A single selected cDNA was hybridized to total starting (st), primary (1°), and secondary (2°) selected cDNAs. The top portion shows a cDNA that is barely detectable in secondary selected human T cell cDNA but is much more abundant in human placenta, HeLa, and mouse spleen selected cDNAs. The bottom portion shows a cDNA that is not detectable at all in T cells but is present in both primary and secondary selected placenta, HeLa, and mouse spleen cDNAs and in fact decreases in the secondary selected HeLa material. This decrease in the secondary selected material is a phenomenon seen after selection of more abundant cDNAs and appears to be indicative of some level of normalization. (B) cDNA enrichments from different genomic targets. Selected material was amplified and Southern blotted as in (A). A single selected cDNA was hybridized to total material selected using either a cosmid pool or a purified YAC as the genomic target. Although this assessment is not quantitative it is clear that both genomic targets selected this particular clone quite well. In general, the efficiencies of selections using YAC or cosmid targets are comparable. (C) Tertiary selection. Total selected material was amplified and Southern blotted as described in (A). A single selected cDNA was hybridized to starting, primary, secondary, and tertiary selected material. The average size and abundance of this cDNA decrease in the tertiary selected material. We have not found that more than two rounds of selection leads to significant enrichments of the genes we have analyzed.

izations using biotinylated targets in order to better control the kinetics of hybridization. Using this technique a purified 450-kb YAC was used to select the interleukin 4 and interleukin 5 genes encoded by chromosome 5 and known to be contained in the YAC. The known and novel genes found in this contig were enriched by between 1000-fold to several 100,000-fold in 2 rounds of selection. Thus, purified YACs can be used successfully in direct selections and we now routinely use YACs of >2 Mb to achieve enrichments of >10,000-fold. In the course of our initial studies it was discovered that a major contaminant in YAC selections is ribosomal RNA. All cDNA libraries have some clones derived from contaminating ribosomal RNA, and purified YACs are usually contaminated with some residual yeast genomic DNA containing multiple ribosomal genes. This problem has been circumvented by blocking the cDNAs before selection with ribosomal cDNA (see below) or by "counterscreening" selected cDNAs with a ribosomal probe. One limitation to the use of purified YACs in selections is the time and labor required to purify sufficient quantities of DNA by separation on pulsed-field gels. We (L. Wang and M. Lovett, unpublished observation, 1995) and others (8) have used total yeast DNA containing both the YAC of interest and yeast genomic DNA as a selection substrate, achieving a few hundred-fold enrichment. Alternatively, we have used inter-*Alu* PCR product from YACs as selection targets with variable results. The success of this approach depends, of course, on having inter-*Alu* PCR products that span a coding region; however, high levels of enrichment can be achieved (L. Wang and M. Lovett, unpublished observation, 1995), comparable to those selections using purified YACs.

Genomic cosmid clones work well in direct selections and provide high levels of enrichment with low background, even when several hundred are pooled together. However, when the ribosomal contaminants are removed from YAC selections, cosmid and YAC selections yield almost identical levels of enrichment (Fig. 2). When isolating coding sequences from large genomic regions (on the order of one to several megabases), a cosmid contig is typically not readily derived or available. We have used two methods to address this. In one case a YAC contig covering approximately 1 Mb was converted to cosmids by using inter-*Alu* PCR products derived from the YAC to screen a chromosome-specific cosmid library. Although these "binned" cosmids may not represent a complete contig, they compensate for chimerism and some microdeletions common in YACs. In the second method a chromosome band was converted to cosmids. The q35 band of chromosome 5 was microdissected, cloned, and used as a probe to screen the Los Alamos National Laboratory (LANL) chromosome 5 flow-sorted cosmid library (9). Approximately 450 cosmids were identified and 410 of these were pooled and used in selections (the remaining 40 cosmids contained a chromosome-specific repeat element and were removed from the set). In both these cases, enrichments of ~10,000-fold were achieved. Finally, the entirety of chromosome 5 (approximately 174 Mb), represented in 25,000 cosmid clones in the LANL library, has been used to enrich successfully for chromosome 5-specific cDNAs (9a). Two important points to note in contemplating selections on very large genomic targets are (a) the genomic DNA source must be of high

quality because "minor" contaminants can actually be present at levels comparable to a single gene in the DNA, and (b) the amount of genomic target must be increased in the hybridization reaction to preserve the relative ratios of target to cDNA species (see Discussion).

One contaminant commonly selected by purified YAC, total yeast, or cosmid selections is mitochondrial transcripts. Several mitochondrial transcripts are polyadenylated and are thus present in cDNA sources. Total yeast DNA contains yeast mitochondrial sequences, some of which are highly conserved between yeast and humans. Purified YACs often have comigrating yeast mitochondrial DNA, and cosmids are often contaminated with *Escherichia coli* chromosomal DNA, which has some homology to mitochondrial genes. Thus mitochondrial genes can be a significant contaminant in selected material. This problem can be reduced by preelectrophoresing YAC blocks in standard agarose gels to remove the faster-migrating circular mitochondrial genome. This method also removes the yeast circular 2μ plasmid, which is a contaminant in some commercial cDNA libraries (see below). Cosmid preparations can be banded on CsCl gradients to reduce the chromosomal background. Preblocking cDNAs before selection (see below) can reduce these contaminants further, but it will not completely eliminate them. A final counterscreen of the cloned material after selection does eliminate these remaining contaminants.

cDNA Sources

Using cloned commercial libraries in early experiments in this laboratory revealed a number of unforeseen problems and contaminants. As mentioned, the 2μ plasmid can be a contaminant of commercial libraries because they are often precipitated with yeast tRNA as carrier, which contains some 2μ plasmid. Conventional amplified libraries may also have a skewed representation, and it is unlikely that very rare transcripts will be represented. For these reasons we now use primary cDNAs made in our laboratory to better represent the complexity of the starting RNA source. cDNA is made by both oligo(dT) priming and random priming; the method gives relatively short cDNA that is more reproducibly amplified in the PCR. Because approximately 10% of genes contain an intermediate repeat within an untranslated region, a significant number of transcripts might be lost in the repeat suppression step or counterscreened in subsequent steps; however, use of random-primed cDNA potentially circumvents this problem by providing other single parts of the same coding region. The cDNA is linkered to amplification cassettes and amplified in the PCR to provide enough material for selection. In our laboratory cDNA pools are routinely multiplexed. Different cDNA sources are linkered to different amplification cassettes, pooled, and used in a single selection; the cDNA source can be distinguished afterward by amplification with cassette-specific oligonucleotides or by sequencing. We have used up to eight different sources successfully in one selection, and have not yet explored the upper limits of this application.

The choice of methods to prepare mRNA is an important consideration when embarking on selection experiments. Preparation of total cellular RNA by methods such as the guanidinium isothiocyanate method usually yields intact RNA of high quality. However, even after oligo(dT) selections for polyadenylated RNAs, these preparations tend to contain low levels of two contaminants that can complicate the analytical process later. The first of these is heterogeneous nuclear RNA (in this category we also include partially spliced intermediates), and the second is contaminating genomic DNA. In most conventional applications these are trivial contaminants and can be discounted. However, because direct selections efficiently enrich any sequence that is homologous to the genomic target, these can become a substantial problem. The problem is further exacerbated by the fact that both of these contaminants are colinear with the genomic DNA and thus hybridize well. Because the goal of direct selection is to identify cDNAs that are homologous to *bona fide* cytoplasmic transcripts, it is advisable to make cytoplasmic RNA preparations whenever possible, to avoid a great deal of work later on confirming the origin of all the selected clones.

As discussed previously, proper blocking of the cDNA before selection is critical to avoid hybridization to any cDNA with a repeat. For all selections we routinely preblock the cDNA with Cot-1 DNA (GIBCO-BRL, Gaithersburg, MD) to suppress high copy repeats present in the genomic target, and with total human genomic DNA to suppress mitochondrial and lower-copy repeats. pUC-based plasmid DNA such as Bluescript (Stratagene, La Jolla, CA) is also used to preblock cDNA because it is a common contaminant of both commercial and home-made cDNAs, probably because of trace amounts of plasmid that contaminate laboratory ware. In the case of cosmid selections, cosmid vector DNA such as pWE is used for preblocking because we have found portions of this vector not found in pUC-based vectors as a contaminant in cDNA preparations (L. Wang, T. Gallardo, C. Hilliard, M. Lovett, unpublished, 1995). For YAC selections, yeast DNA from the host strain AB1380 (10) and/or yeast ribosomal DNA is used to preblock. Alternatively, perhaps the best reagent for blocking or counterscreening ribosomal contamination has been cDNA prepared from the poly(A)$^-$ fraction of the primary cDNA used for selection.

It is important to remember that direct cDNA selection is entirely dependent on the cDNA sources used; in other words, a gene cannot be selected if it is not expressed in the tissue source. Therefore choice of tissue is critical, particularly if one is embarking on a positional cloning project. If the choice of cDNA source is not obvious from the pathogenesis of the disorder then it is advisable to use two or more gene hunting techniques in combination.

Cloning and Analysis

After each round of selection the PCR products of the eluted cDNA are evaluated for enrichment. This is accomplished either by monitoring the enrichment of a known gene in the contig or by seeding a reporter gene into the experiment to act as a sur-

rogate marker for enrichment. This is an important positive control and should not be dispensed with; a selection experiment usually takes at least 2 weeks and requires fairly extensive molecular biological manipulations that are sometimes liable to failure. Thus a positive control is an absolute necessity to avoid characterizing clones from experiments that failed. After two rounds of selection and PCR amplification the selected material is first cloned into a vector such as λgt10. This has the advantage that any plasmid contaminants can be easily evaluated by plaque lifts, as can the presence of repetitive elements or any other known contaminant. After this initial validation stage, the secondary selected material is cloned into the high-efficiency UdG plasmid-based cloning system (CloneAmp; BRL), so that clones can be arrayed at high density and rapidly screened. Common contaminants mentioned previously, if present, are then counterscreened from selected clones. The remaining clones are validated by hybridization to the starting source, to whole genome Southern blots, and by DNA sequencing.

Methods

Preparation of Genomic DNA

This section describes purification of YACs and biotinylation of YACs or cosmid DNAs to be used in direct selection.

Materials

YAC clones in agarose blocks
Restriction enzyme
GeneClean II kit (Cat. No. 3113; Bio 101, La Jolla, CA)
Biotin-16-dUTP, 0.4 m*M* (Cat. No. 1093 070; Boehringer Mannheim, Indianapolis, IN)
Nick translation kit (Cat. No. 976 776; Boehringer Mannheim)
[α-^{32}P]dCTP
Sephadex G-50 spin columns (11)
Control reporter DNA only if there is no control gene within the contig (see below)
Additional reagents for standard plasmid minipreps [for cosmid DNA purification (11)], agarose gel electrophoresis, and pulsed-field gel electrophoresis (for YAC purification)

1. Purify cosmid DNAs by standard alkaline lysis minipreparation procedure (11). As described above, cosmid DNAs can be prepared by banding on CsCl gradients to reduce mitochondrial background, but this can be cumbersome for large numbers of cosmids.
2. Preelectrophorese yeast chromosomes in agarose blocks by running them for

1 hr at 50 V (3 V/cm) on a conventional 1% (w/v) agarose gel. This eliminates the mitochondrial and 2μ DNAs described above. Remove the agarose blocks from the gel wells and transfer them to a contour-clamped homogeneous electric field (CHEF) gel. Run the CHEF gel and excise the YAC. Typically every lane on a CHEF gel is required to yield enough of a single YAC for selection.

3. Digest the excised gel slice *in situ* with a restriction enzyme, preferably one with a four-base recognition, to facilitate extraction of the DNA from the gel. Extract the DNA with GeneClean by adding 100 μl of glass milk to approximately 5 ml of liquefied agarose gel slice and incubate at room temperature for 5 min. Wash as directed, then elute in 400 μl of Tris–EDTA (TE). Measure the concentration of DNA. The yield should be a few hundred nanograms of DNA.

4. If there is no known gene within the contig that is also present in the cDNAs to be selected, then a reporter DNA must be diluted into the YAC (or cosmid) DNA. For example, for 200 ng of YAC DNA spanning 1 Mb, add approximately 6–8 ng of control cosmid.

5. Label the YAC or cosmid DNA (plus reporter) with biotin-16-dUTP, using a nick translation kit. Maintain a ratio of TTP to biotin-16-dUTP of 20:1 and add 1 μl of [α-^{32}P]dCTP. Incubate according to the manufacturer instructions and remove unincorporated dNTPs by passage over a Sephadex G-50 spin column.

6. Check incorporation of [α-^{32}P]dCTP by scintillation counting.

7. Check that biotin was incorporated by incubating 1 μl of the reaction with 20 μl of washed streptavidin-coated magnetic beads (see washing protocol under *Hybridization Selection*) at room temperature for 15 min with occasional shaking. Separate the beads from the solution with a particle separator. Count the radioactivity in the bead fraction and the solution fraction. The majority of radioactivity (>5:1) should be in the bead fraction, indicating that biotin was incorporated along with the radiolabeled dCTP.

8. Two alternative biotinylation methods make use of commercially available photobiotin systems or amplification of the genomic DNA in the PCR with biotinylated primers. The latter requires that the DNA be linked first with oligonucleotides complementary to the biotinylated primers, but has the advantage that large amounts of the DNA can be generated via the PCR.

9. Ethanol precipitate the genomic DNA, resuspend in water at a concentration of 20 μg/ml, and store at $-20°$C.

Preparation of Primary cDNA

Materials

Random-primed, blunt-ended, double-stranded cDNA in 22 μl (average length, 500 bp–1 kb), 3 μg

Phosphorylated cDNA linkers containing a restriction site such that the cDNA can be cloned subsequently, 1 μg/μl

T4 DNA ligase buffer (10×) containing fresh 5 m*M* ATP
T4 DNA ligase (Cat. No. 481 220; Boehringer Mannheim), 1 unit/μl
Sephadex G-50 spin column

1. Ligate the linkers to the cDNA by adding 3 μl of 10× T4 DNA ligase buffer containing 5 m*M* ATP and 2 μl of linkers to the 22 μl of blunt-ended cDNA. Add 3 μl (3 units) of T4 DNA ligase, mix gently, and incubate at 4° C overnight.
2. Microcentrifuge briefly and inactivate the ligase by incubation at 65° C for 10 min.
3. We use linkers containing a two base 3′ overhang, which will form linker dimers but will not form larger oligomers. Typically, unligated linkers are not removed at this stage, but a portion of the ligation is amplified in the PCR along with positive and negative controls to confirm that an appropriate smear is generated. If a tissue can be obtained in only limited quantities it may not be realistic to expect to use primary cDNA in the selection; in this case the PCR-amplified material is used.
4. Extract the cDNA with phenol–chloroform (50:50, v/v) and desalt on a Sephadex G-50 spin column. Store 1 μg ($\frac{1}{3}$ of total) at −20° C for future use.
5. Ethanol precipitate the remaining cDNA (2 μg) and resuspend in 5 μl of water.

Repeat Suppression

Materials

Human Cot-1 DNA (Cat. No. 5279SA; GIBCO-BRL), 2 μg
Ribosomal blocking DNA (for YAC selections). This can be either 1 μg of total DNA from the YAC host strain AB1380, 100 ng of cloned ribosomal genes including 5.8S RNA, or preferably 100 ng of cDNA from the nonpolyadenylated fraction of a cDNA synthesis (dT-cellulose flowthrough)
pYAC4 vector blocking DNA ((Cat. No. V7504; Sigma, St. Louis, MO), linearized, for YAC selections (10 ng); otherwise use the appropriate linearized vector DNA for other selections
pUC-based plasmid DNA, linearized (10 ng)
Hybridization solution, 2× (see below)

Mix 2 μg of cDNA with 2 μg of Cot-1 DNA, ribosomal blocker, and vector and plasmid DNAs in a total volume of 10 μl. Overlay with mineral oil and heat to 95° C for 5 min. Add 10 μl of 2× hybridization solution under the oil and mix gently. Incubate at 65° C for 4 hr. At a concentration of 200 μg of cDNA per milliliter this corresponds to a $C_0t_{1/2}$ of about 10 mol nucleotide liter^{-1} sec. As a rule of thumb, the $C_0t_{1/2}$ can be calculated based on a hybridization of 1 hr at a concentration of 80 μg/ml, resulting in a $C_0t_{1/2}$ of approximately 1 mol nucleotide liter^{-1} sec. This protocol can be modified to attain the same ending $C_0t_{1/2}$.

Hybridization Selection

Materials

Dynabeads M-280 streptavidin-coated magnetic beads (Cat. No. 112.05; Dynal, Inc., Lake Success, NY)
Magnetic particle separator (Cat. No. MPC-E-1; Dynal Inc., Lake Success, NY)
Wash solution 1: 1× SSC, 0.1% (w/v) sodium dodecyl sulfate (SDS), 10 ml
Wash solution 2: 0.1× SSC, 0.1% (w/v) SDS, 10 ml
Elution solution 1: 100 m*M* NaOH (make fresh)
Elution solution 2: 1 *M* Tris-HCl (pH 7.5)
Hybridization solution (2×): 1.5 *M* NaCl–40 m*M* sodium phosphate (pH 7.2)–10 m*M* EDTA (pH 8.0)–10× Denhardt's solution–0.2% (w/v) SDS
Binding buffer: 10 m*M* Tris (pH 7.5)–1 m*M* EDTA (pH 8.0)–1 *M* NaCl

1. Denature the genomic DNA contig by overlaying 5 μl of a 20-μg/ml solution of the biotinylated DNA (100 ng) with mineral oil and heating to 95° C for 5 min.
2. Immediately add the prehybridized cDNA (20 μl) plus 5 μl of 2× hybridization solution to the denatured genomic DNA. Hybridize the cDNA and genomic contig for 54 hr at 65° C with gentle shaking. This corresponds to a total $C_0t_{1/2}$ of 100 nucleotides liter^{-1} sec. For very large contigs (tens of megabases) the ratio of genomic DNA to cDNA should be altered (see Discussion).
3. Wash 200 μl (2 mg) of Dynabeads three times in binding buffer. At each washing cycle remove the beads from solution with the magnetic separator for 1 min. Resuspend the washed beads at 10 mg/ml in binding buffer.
4. Capture genomic DNA by incubating the hybridized DNAs with the washed beads at room temperature for 15 min with occasional shaking. Remove the beads with the magnetic separator for 1 min.
5. Wash the beads and captured DNA at room temperature with 1 ml of wash solution 1 for 15 min followed by three washes at 65° C with wash solution 2.
6. Elute the cDNAs from the genomic DNA by incubating the beads in 50 μl of 100 m*M* NaOH for 10 min at room temperature followed by neutralization with 50 μl of 1 *M* Tris-HCl (pH 7.5). Desalt by running this material through a Sephadex G-50 column.

Analysis

Materials

Reagents for standard amplification by PCR
Starting cDNA, 1-ng/μl concentration
Southern blotting reagents
Probe for reporter gene

1. Amplify the eluted cDNAs in 50-μl PCR reactions, using the appropriate primers. Use 1, 2, and 5 μl of eluted material in the reactions and include no DNA and no primer controls. Also include a control using 1 ng of the starting cDNA.

2. Electrophorese the PCR products on a 1% (w/v) agarose gel. At least one reaction of the primary selection should show a smear of products that may or may not include background bands; use this for further analysis. Other reactions may show no product, predominant bands, or a smear extending all the way to the well, indicating template switching due to too much input DNA. Negative controls should be negative.

3. Electrophorese the chosen PCR product on a 1% (w/v) agarose gel alongside reamplified starting cDNA, starting cDNA that has not been amplified, negative controls, and appropriate size markers. Typically the average size of the reamplified material will be smaller than the starting cDNA because of preferential amplification of shorter products. Transfer the gel to nylon membrane by Southern blotting.

4. Hybridize the blot with a probe for the control reporter gene. The autoradiograph should show enrichment of the positive control in the primary selected material relative to the starting material. This blot also can be hybridized to a probe for a negative control gene (one that is not present in the contig but is expressed in the starting cDNA) and with radiolabeled Cot-1 DNA to monitor repeat sequences. Both of these should decrease in the selected material, although repetitive sequences may have remained constant. The critical point is that repeats have not been enriched. Likewise, ribosomal and plasmid background can be measured; neither should be enriched.

5. If the positive control reporter has been enriched then recycle the primary selected cDNAs through a second round of selection. It may be necessary to scale up the appropriate PCR of the eluted material from the primary selection to generate enough cDNA for the next selection. This cDNA should be blocked as described for the first round of selection. Hybrid select as described using the other half (100 ng) of the biotinylated genomic template.

6. Capture the genomic DNA, wash, elute, amplify, and assess enrichment as described above, except for adding a lane of secondary selected cDNA. Also include a control of primary selected cDNA that has been reamplified. At this point the amplified, secondary selected material may show discrete bands on an ethidium bromide-stained agarose gel. We typically see some enrichment after the second round of selection that is not as dramatic as the first round. Some reporter genes have been observed to decrease slightly after the second round of selection, probably due to a higher abundance in the starting material.

7. Clone the secondary selected material; primary selected material may also be cloned if it is deemed appropriate. To clone the PCR products, digest with a restriction enzyme that will remove most of the linker, extract the digests with phenol–chloroform (50:50, v/v), desalt over Sephadex G-50 spin columns, and ethanol pre-

cipitate. Quantitate the DNA, then ligate into the vector of choice. As mentioned previously λgt10 has particular advantages due to its high cloning efficiency and lack of homology to plasmid sequences; however, we now routinely clone into a plasmid vector using the UdG cloning system so that large numbers of selected clones can be arrayed for high-throughput analysis.

8. Pick five clones (colonies or plaques) from each selected library and resuspend in 0.5 ml of LB or SM, respectively (see Ref. 11 for recipes). Analyze the inserts by amplifying 1 μl from each in a PCR reaction using the appropriate primers. Electrophorese the products on an agarose gel; there should be a range of different sizes.

9. If the cloning appeared to be successful in step 8, then pick individual clones (colonies or plaques) into two or three 96-well Titertube boxes (Bio-Rad, Richmond, CA), 500 ml of medium per tube (see below for a discussion of numbers of clones to analyze). Allow the clones to grow or elute overnight. Make two duplicate plates by transferring 100-μl aliquots into two microtiter dishes. To one plate add dimethyl sulfoxide (DMSO) to 100% (v/v) in each well; store this at −80° C as a master plate. Use the other microtiter plate for making "stamps" or arrays (below). Use the original box as a PCR source.

10. Clones in a bacterial host can be stamped onto appropriate agar plates; if phage, they can be stamped onto bacterial lawns. Grow plates at 37° C overnight, then make duplicate lifts. Hybridize these lifts with probes to screen out the background. These include plasmid, repeats (screened by labeling total human genomic DNA), and ribosomal DNA in the case of a YAC selection. We have found that clones that hybridize weakly to total human DNA often contain mitochondrial inserts; therefore even weakly hybridizing clones should be eliminated.

11. If the cDNAs are cloned into a bacterial host then an alternative to the stamps described above is to pick clones into microtiter plates, then array them at medium to high density on filters with an automated system such as a Biomek robot (Beckman, Fullerton, CA). We routinely array clones at 12-fold density in a 96-well format. In this way potentially many more clones can be analyzed in a shorter time. Obviously the same backup sources should be made.

12. Amplify via the PCR those cDNAs that were not screened out and electrophorese through low melting point agarose gels. Cut out gel slices and store at 4° C.

13. Label individual inserts contained in the low melt agarose gel slices and hybridize to Southern blots or colony lifts containing the original cosmid or YAC template DNAs to validate the cDNAs. Also hybridize to cDNA stamps or arrays described above to eliminate redundant clones and to Southern blots containing digested total human DNA to detect repetitive cDNAs (see below). Figure 3 shows analysis of a single-selected cDNA clone by simultaneous hybridization to gridded arrayed cosmids that were used as the genomic target in the selection (Fig. 3A), to arrayed cDNA plaques (Fig. 3B), and to lifts of selected cDNAs containing 10^3 plaques. Hybridization to the cosmids validates the clone and places it within the

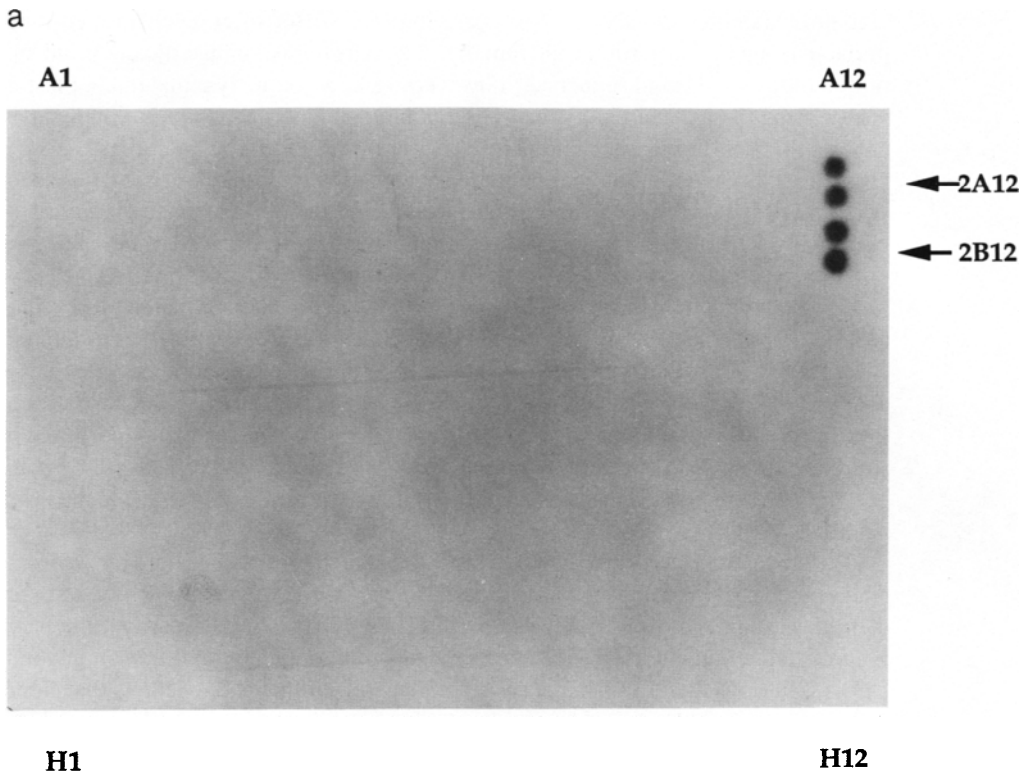


FIG. 3 Analyses of a single-selected cDNA. A selected cDNA clone was hybridized simultaneously to a cosmid grid containing the cosmids used as the genomic target (A), to gridded selected cDNAs (B), and to approximately 10^3 nongridded selected cDNAs (C). These hybridizations effectively place the cDNA within a cosmid contig, eliminate redundancy in the picked and arrayed set, and allow its representation in the total selected set to be calculated, respectively. (A) This clone hybridized in duplicate to cosmids 1A12 and 2B12 (the top and bottom spots are duplicates). (B) This clone hybridized to itself plus four other clones in an arrayed set of 96; the redundant clones were then eliminated from further analysis. (C) The same clone hybridized to 15 clones out of approximately 10^3. Hybridization to greater numbers of selected clones as in this example allows more accurate calculation of redundancy, which in turn allows one to determine the number of clones that should be analyzed to maximize coverage. Results shown in (B) and (C) are each half of a duplicate set.

contig. Hybridization to picked arrayed clones eliminates future redundant analyses, and redundancy measurements in higher density lifts gives some measure of the number of clones to analyze in order to saturate the selected set (see Discussion).

14. An alternative to the described hybridizations, and for higher throughput

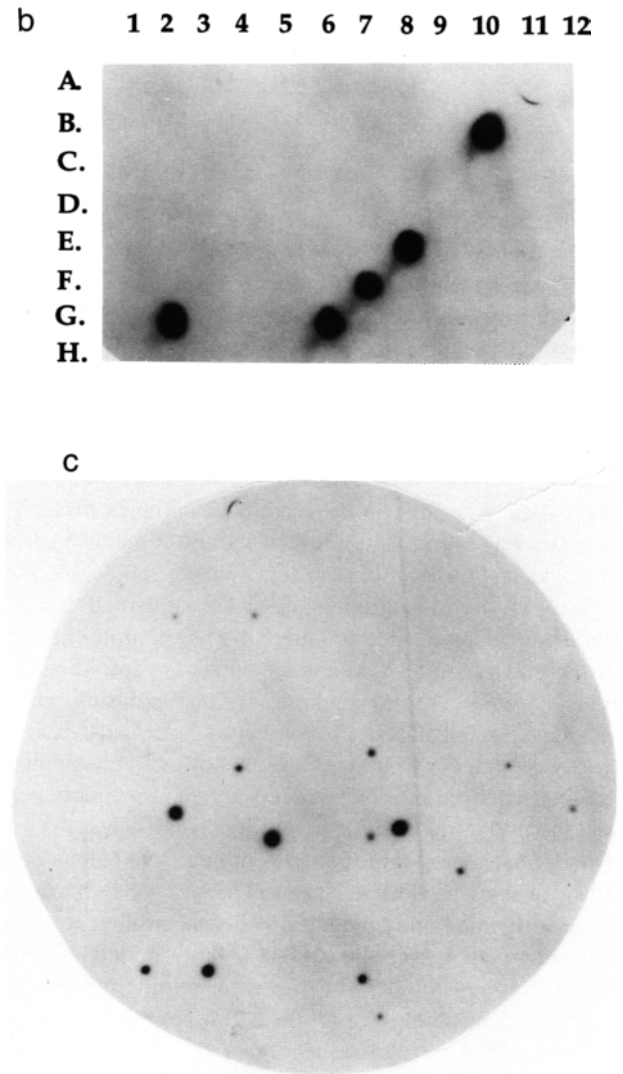


FIG. 3 *Continued.*

analysis, is validation of the selected cDNAs via the PCR. After elimination of background and redundant clones, remaining clones are sequenced using PCR-based cycle sequencing. Single-pass sequence is generated by automated sequencing (Applied Biosystems, Foster City, CA). Primers developed from this sequence are then

used in the PCR to determine if they amplify the starting genomic targets used in the original selection.

Discussion

Direct cDNA selection can be used to efficiently isolate cDNAs that are encoded by large regions of human genomic DNA. Examples in the literature of the success of this method in the isolation of disease genes include the identification of the diastrophic dysplasia (12), the breast and ovarian cancer susceptibility BCRA1 (13), and the ataxia telangiectasia (14) genes. Two critical parameters, discussed previously, that must be considered are the starting cDNA source (choice of tissue) and abundance of the targeted cDNAs. In some searches for disease genes, neither of these is clearly predicted by the pathogenesis of the disorder. The most expedient position, should this occur, is to search within diverse and complex cDNA sources (e.g., placenta, neonatal brain, or testes) that have complex patterns of gene expression. Our experience has been that adult/fetal kidney contains a surprisingly high complexity of mRNAs and this would represent another good choice for a "shotgun" search. In designing the hybridization conditions it is also advisable to assume that the transcript is low in abundance and therefore use complex cDNA pools, rather than resort to using commercially available or otherwise amplified cDNA libraries. Repetitive elements within complex genomes are one enduring problem with confirming and validating selected cDNAs. The protocol described here is successful in blocking the majority of high copy number repeats; however, this unmasks a number of low-level repeats or diverged intermediate repeats that inevitably persist in a selection. These reveal themselves as a smear or multiple bands when the individual cDNA is hybridized to a Southern blot of human genomic DNA. Gene families also complicate the analysis of selected clones. Thus, a cDNA may be selected by the genomic locus that encodes a related family member. The same argument holds for pseudogenes, which can select a cDNA from the structural gene and lead to erroneous mapping data. All of these examples point to the need for parallel approaches to gene identification. Gene identification algorithms based on genomic DNA sequence cannot currently lead to the unequivocal identification of all genes. However, isolating and sequencing cDNAs can serve to validate the genomic sequence and vice versa; the genomic DNA sequence will immediately reveal disparities between a cDNA and the genome that would point to the presence of a pseudogene or gene family. From a practical standpoint it is usually more efficient to focus initially on selected cDNAs that are single copy; if there is enough redundancy in the selected set and they are derived by random priming, rather than exclusively by oligo(dT) priming, then a portion of the gene containing single-copy sequence can frequently be found, even if a repeat resides in an untranslated region.

The redundancy of individual selected cDNA clones can be measured and used to calculate the number of clones that must be analyzed to sample theoretically a complete cross-section of a selected set. For example, Fig. 3C shows a selected cDNA hybridized to plaque lifts containing approximately 10^3 selected clones. The cDNA hybridized to 15 clones, in Fig. 3B another cDNA detected 18 clones, and in Fig. 3C a third cDNA detected 3 clones. Interestingly, this same clone hybridized to 5 others in the first set of 96 analyzed; however, hybridization to more arrayed clones was consistent with the 1.5% measured in 1000. This demonstrates the utility of eliminating redundancies in arrayed sets. Analysis of 10 clones from this selected set indicated that the cDNAs were pseudonormalized, ranging from 1 in 50 to 1 in 300 in abundance. In this particular case 300 clones were picked and sequenced. Analyzing 300 clones is a fairly rigorous requirement because the cDNA fragments typically do not represent full-length genes and overlaps can be expected in a significant proportion of clones; however, in the absence of gene density and transcript abundance information for a particular contig, it is currently the best course of action.

Direct selection is a PCR-based method and sequences that do not efficiently amplify in the PCR will be lost in the selection. Thus, particularly GC-rich stretches or long cDNA fragments, for example, may not be amplified. This argues that shorter, random-primed cDNA should be used as a starting source so that some portion of each gene will be represented. However, clones with an average molecular length of >250 bp (and sometimes up to 2 kb in length) should be routinely selected by the method as described here.

Finally, the derivation of full-length cDNAs is worth mentioning. This process of deconvolution is a time-consuming and important step in any gene hunt. Selected cDNAs can be used as probes and hybridized to cloned, full-length sources. This has met with some success in our laboratory (R. Del Mastro, unpublished, 1995). However, this is laborious as one must perform primary, secondary, and tertiary screens on hundred of thousands of clones with a number of different probes. One quick screen that can be employed to decide on which cDNA source to target is to produce Southern blots of cDNAs from a wide variety of cDNA sources and probe these with the clones of interest. Alternatively, expression can be tested by PCR amplification of phage supernatants of the various cDNAs with primers to individual selected cDNAs. This is more rapid and sensitive than hybridization and is another reason to consider generating PCR primers to selected cDNAs. In this way a specific full-length cDNA library can be targeted for clone extension. We are currently investigating using selected cDNAs as single-copy, colinear targets to select full-length cDNAs *en masse* from both cloned sources and primary, linked full-length sources. Long-range PCR methods may prove effective in this application. In the future, as sequencing technology improves it may be possible to sequence through tens of kilobases of a contig to which particular selected cDNAs hybridize, then "stitch" genes together using the selected cDNA sequences as templates and exon prediction com-

puter algorithms as needle and thread. At the present, however, this is not a reality and the rapid derivation of any part of a gene, far less a full-length gene, still remains a problem in the analysis of human and other complex genomes.

Acknowledgment

Supported in part by NIH Grant P50DE11131.

References

1. A. J. Buckler, D. D. Chang, S. L. Graw, J. D. Brook, *et al., Proc. Natl. Acad. Sci. U.S.A.* **88,** 4005 (1991).
2. D. M. Kurnit and B. Seed, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 3166 (1990).
3. P. Liu, R. Legerski, and M. J. Siciliano, *Science* **246,** 813 (1989).
4. L. Corbo, J. A. Maley, D. L. Nelson, and C. T. Caskey, *Science* **249,** 652 (1990).
5. M. Lovett, J. Kere, and L. Hinton, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 9628 (1991).
6. M. Lovett, *Trends Genet.* **10,** 352 (1994).
7. J. G. Morgan, G. M. Dolganov, S. E. Robbins, L. M. Hinton, and M. Lovett, *Nucleic Acids Res.* **20,** 5173 (1992).
8. S. Parimoo, R. Kolluri, and S. M. Weissman, *Nucleic Acids Res.* **21,** 4422 (1993).
9. D. L. Saltman, G. M. Dolganov, B. S. Pearce, S. S. Kuo, P. J. Callahan, M. L. Cleary, and M. Lovett, *Nucleic Acids Res.* **20,** 1401 (1992).

9a. R. Del Mastro, L. Wang, A. D. Simmons, T. D. Gallardo, G. A. Clines, J. A. Ashley, C. J. Hilliard, J. J. Wasmuth, J. D. McPherson, and M. Lovett, *Genome Research* **5,** 185 (1995).

10. D. T. Burke, G. F. Carle, and M. V. Olson, *Science* **236,** 806 (1987).
11. J. Sambrook, E. F. Fritsch, and T. Maniatis, "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
12. J. Hastbacka, A. de la Chapelle, M. M. Mahtani, G. Clines, *et al., Cell* **78,** 1073 (1994).
13. Y. Miki, J. Swensen, D. Shattuckeidens, P. A. Futreal, *et al., Science* **266,** 66 (1994).
14. K. Savitsky, A. Bar-shira, S. Gilad, G. Rotman, Y. Zui, L. Vanagaite, D. A. Tagle, S. Smith, T. Uziel, S. Sfez, M. Ashkenazi, I. Pecker, M. Frydman, R. Harnik, S. R. Patanjali, A. Simmons, G. Clines, A. Sartiel, R. A. Gatti, L. Chessa, O. Sanal, M. F. Lavin, N. G. J. Jaspers, A. Malcom, R. Taylor, C. F. Arlett, T. Miki, S. M. Weissman, M. Lovett, F. S. Collins, and Y. Shiloh, *Science,* **268.** 1749 (1995).

[11] Isolation of Candidate Genes for Inherited Diseases: Application to X-Linked Retinal Degenerations

Denise Yan, Ricardo Fujita, and Anand Swaroop

Introduction

Identification of genes responsible for inherited diseases is a daunting task because of the enormous size and complexity of the human genome. "Positional cloning" for disease gene isolation is a laborious and time-consuming strategy (1–3). Now that the cloning of large genomic regions is no longer a major obstacle, the success with this strategy depends on efficient isolation of transcribed sequences from the cloned genomic DNA. Sequencing and cDNA library screening methods are cumbersome and less efficient for gene hunting in large genomic regions. Exon amplification (4) and cDNA selection using filter hybridization (5, 6) or magnetic capture (7, 8) are preferred for identifying genes in the genomic DNA that has been cloned in cosmid and/or yeast artificial chromosome (YAC) vectors. The search for mutations in functionally relevant "candidate genes," although successful at times, is a random approach that requires educated guesswork (1–3). It is believed that a combination of "positional cloning" and "candidate gene" approaches, referred to as the "positional candidate" strategy (3), will greatly facilitate the cloning of disease genes. Availability of a large number of specific candidate genes that are precisely mapped to human chromosomal regions will be required for the success of this strategy. The candidate gene repertoire is now augmented by a large number of expressed sequence tags (9, 10); however, their chromosomal localization is proceeding slowly (11, 12). New and improved methods for cDNA selection and/or candidate gene isolation are needed to expedite the disease gene identification.

In *Drosophila,* more than 2500 genes (out of ~17,000 encoded by its genome) are predicted to be involved in eye morphogenesis (13). Because the human genome contains ~100,000 genes (14), we estimate that as many as 10,000 of these may participate in the development of considerably more complex ocular tissues in humans. It is interesting to note that almost one-third of more than 4000 Mendelian disorders affect the visual function (15). To facilitate disease gene identification by the "positional candidate" strategy, we are developing methods for efficient isolation of retinal genes that can be used as candidates for inherited retinal degenerations. However, these methods can be extended to other systems and diseases as well. Here, we describe a general subtraction–selection strategy that has been employed to iden-

Methods in Molecular Genetics, Volume 8

tify putative candidate gene regions for X-linked retinal diseases. This strategy involves subtraction cloning to reduce the complexity of specific cDNA libraries and utilizes the subtracted probe to isolate genomic clones from specific regions or chromosomes. We also delineate an efficient hybridization-based sandwich-selection method, which does not employ the polymerase chain reaction (PCR) during any step and allows the isolation of larger-size cDNA clones. In addition, we have modified the magnetic capture and filter hybridization methods and used large-insert libraries for cDNA selection from cloned genomic regions.

Subtraction–Selection Strategy

Subtraction–selection strategy is designed for isolation of candidate genes from specific regions of human chromosomes, and has been employed for identifying X-linked retinal genes. The strategy (represented schematically in Fig. 1) involves the following steps: (a) subtraction cDNA cloning, (b) isolation of chromosome- or region-specific genomic clones using the subtracted cDNA probe, (c) regional localization and microsatellite content analysis of isolated genomic clones, and (d) identification of transcribed sequences in the genomic clones of interest.

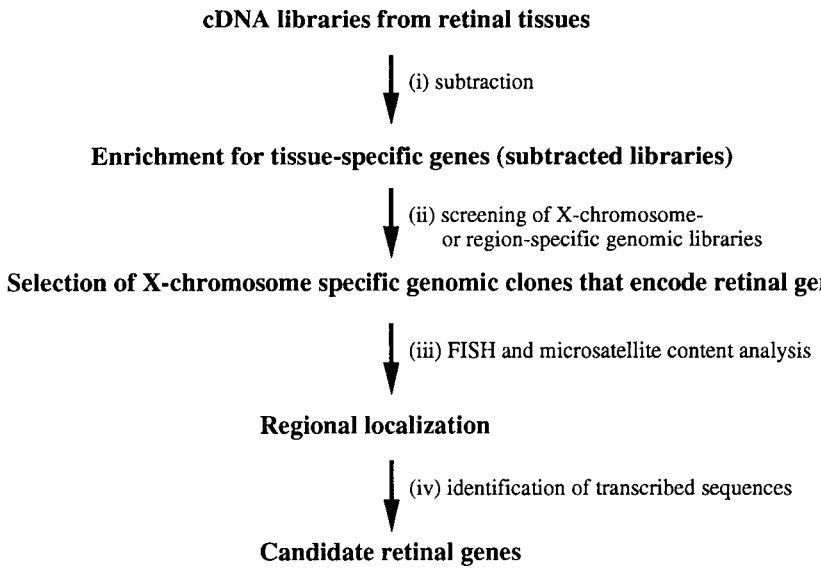


FIG. 1 Subtraction–selection strategy for isolation of X-linked retinal genes.

Subtraction cDNA Cloning

It is expected that genes responsible for retinal degenerations are expressed in the retina. We use subtraction cDNA cloning to reduce the complexity of retinal cDNA libraries. This also results in enrichment of tissue-specific genes. Several subtraction methods utilizing cDNA libraries or uncloned nucleic acids have been described (16–19). We earlier constructed high-complexity, directional cDNA libraries from a large number of human tissues and cell lines (20), and enriched adult retina and retinal pigment epithelial cell line libraries using a novel biotin-based subtraction procedure (19). The details of this method and further modifications/improvements have been published (21–23). Because the "disease gene" may be expressed in other tissues as well, only one round of subtraction is performed against a distant cell type. This results in the removal and/or reduction of abundant constitutively expressed genes in the subtracted library (19; A. Swaroop, unpublished data, 1995).

Isolation of Chromosome- or Region-Specific Genomic Clones

cDNA inserts from the subtracted library are used as a probe to screen a chromosome- or region-specific genomic library. We prefer the libraries constructed in a lambda (λ) phage vector to minimize the background hybridization due to contaminating plasmid sequences that may be present in the subtracted cDNA probe. The human chromosome-specific genomic libraries are available from American Type Culture Collection (ATCC; Rockville, MD). For our studies, we screened a human X-chromosome-specific genomic library (LA0XNL01), which was prepared by cloning partial *Sau* 3A-digested flow-sorted X-chromosomal DNA in the λ Charon 35 phage vector. The method for generating the ^{32}P-labeled subtracted cDNA probe is dependent on the choice of subtraction procedure that is employed for enrichment. The biotin-based subtraction method that we used (19, 21) yields cDNA clone in the Bluescript plasmid vector (Stratagene Cloning Systems, La Jolla, CA). This allows the preparation of a ^{32}P-labeled high specific activity probe by one of the following methods: (a) PCR amplification using T3 and T7 primers (or M13 forward and reverse primers), (b) use of T3 or T7 RNA polymerase to prepare the riboprobe, or (c) random-primer labeling of the purified cDNA inserts (24).

Materials, Media, and Reagents

Host *Escherichia coli* bacteria

NZCYM medium: 10 g of NZ-amine, 5 g of NaCl, 2 g of $MgSO_4 \cdot H_2O$, 5 g of Bacto Yeast extract (Difco, Detroit, MI), 1 g of Casamino Acids per liter. Autoclave to sterilize

SM buffer: 5.8 g of NaCl, 2 g of $MgSO_4 \cdot 7\ H_2O$, 50 ml of 1 *M* Tris-HCl (pH 7.5), 5 ml of 2% (w/v) gelatin per liter. Autoclave to sterilize
Hybond N membrane (Amersham Life Science, Arlington Heights, IL)
Denaturation solution: 1.5 *M* NaCl, 0.5 *N* NaOH
Neutralization solution: 1.5 *M* NaCl, 0.5 *M* Tris-HCl (pH 8.0)
SSC, 20×: 1× SSC is 0.15 *M* NaCl, 0.015 *M* sodium citrate
Denhardt's solution, 5×: 1× Denhardt's solution is 0.02% (w/v) Ficoll, 0.2% (w/v) polyvinylpyrrolidone, 0.02% (w/v) bovine serum albumin (BSA) fraction V
TE: 10 m*M* Tris-HCl (pH 8.0), 1 m*M* EDTA
Sodium dodecyl sulfate (SDS)
PEG 8000 (Polyethylene glycol 8000; Sigma Chemical Co., St. Louis, MO)
Salmon sperm DNA (Sigma)
Cot-1 DNA (GIBCO-BRL, Life Technologies, Inc., Gaithersburg, MD)
dNTP solutions (Boehringer Mannheim, Indianapolis, IN)
AmpliTaq DNA polymerase (Perkin-Elmer Corp., Foster City, CA)
Prehybridization solution [5× Denhardt's solution, 6× SSC, 0.5% (w/v) SDS, 100 μg of sonicated and denatured salmon sperm DNA per milliliter]
Restriction enzymes: Purchase from New England Biolabs (Beverly, MA)

Plating of Phage Genomic Library

The number of plaques that need to be screened depends on the complexity of the genomic library. For example, we screen ~100,000 independent plaques from the human X-chromosome-specific genomic library LA0XNL01, where an average insert size is 15–20 kb ($\simeq$ 7 genome equivalent).

1. Prepare large (150-mm diameter) NZ–agar petri plates [NZCYM medium with 1.5% (w/v) agar]. Keep at 37° C for a few hours before use.
2. To prepare *E. coli* cells for plating the phage library, inoculate 0.5 ml of an overnight culture of the appropriate host bacteria (e.g., LE392) to 4.5 ml of NZCYM medium containing 0.2% (w/v) maltose. Grow until the OD_{600} is approximately 1.0. Centrifuge the cells at 4000 rpm for 10 min at 4° C. Decant the supernatant and resuspend the pellet in 2.5 ml of SM. The plating cells can be stored for a few days at 4° C. A small decrease in plating efficiency may be observed.
3. In a 15-ml tube, mix 0.2 ml of plating bacteria and 0.2 ml of phage (5×10^3 to 2×10^4 particles). (Dilute the phage library in SM, if needed. Calculate the number of phage particles from a recently titered library stock, because the phage titer can drop over time.) Mix gently.
4. Allow adsorption of phage to bacteria for 15–20 min at 37° C.
5. To each tube, add 6–7 ml of NZ–top agarose [NZCYM medium containing 0.7% (w/v) agarose] kept at 48° C. Immediately pour the mixture onto an NZ–agar

plate. Gently rotate the slightly tilted plate. This allows even spreading of bacteria across the surface of the plate before the top agarose hardens.
6. Allow the top agarose to harden for 15 min at room temperature. Invert the plates and incubate for 6–12 hr at 37° C, until plaques appear as clear spots against the opaque background of bacterial lawn.
7. Store the plates inverted at 4° C until screening.

Transfer of Phage Library to Membrane Filters

1. Label each library plate and the corresponding Hybond N filter membrane. Always wear gloves when handling the membrane filters.
2. Put the membrane filter on the plate. Be careful not to trap air bubbles. Once the filter is wet, let it set for 1 min to allow transfer of phage plaques.
3. Mark the filter in three asymmetric locations by stabbing through it and into the agar beneath with an 18-gauge needle attached to a syringe containing waterproof black India ink.
4. Remove the filter, using a blunt-ended forcep, without peeling off the top agarose overlay and immerse it sequentially, plaque side up, in the denaturation, neutralization and 5× SSC solutions for 3–4 min each.
5. A second set of filter lifts can be made from each plate. Hybridization to duplicate replica filters is helpful in eliminating the false positives.
6. Cross-link DNA to the filters by exposure to ultraviolet (UV) (using Stratalinker UV cross-linker from Stratagene Cloning Systems).

Generation of Subtracted cDNA Probe

1. Generate the probe from subtracted cDNA library by PCR amplification. The PCR reaction consists of the following.

DNA from the subtracted cDNA library	~50 ng
T3 primer (100 ng/μl)	1 μl
T7 primer (100 ng/μl)	1 μl
$MgCl_2$ (25 m*M*)	5 μl
Buffer (10×): 0.5 *M* KCl, 0.1 *M* Tris-HCl (pH 8.3)	5 μl
dATP (10 m*M*)	1 μl
dGTP (10 m*M*)	1 μl
dTTP (10 m*M*)	1 μl
dCTP (0.1 m*M*)	10 μl
[α-^{32}P]dCTP (3000 Ci/mmol)	50 μCi
AmpliTaq (5 U/μl)	0.5 μl

Make up the volume to 50 μl with water.
2. Cover the reaction mixture with mineral oil and place the tubes in a thermocycler.

3. Carry out amplification for 40 cycles with denaturation at 94° C for 1 min, annealing at 55° C for 1 min, and extension at 72° C for 2 min. A "hot-start" procedure can also be used.
4. Pass the ^{32}P-labeled amplified cDNA probe through a Sephadex G-50 column (Boehringer Mannheim).
5. Preanneal the probe to an excess amount of human Cot-1 DNA to remove repetitive sequences (according to GIBCO-BRL supplied protocol). For this, mix ^{32}P-labeled cDNA probe with 50 μl of the human Cot-1 DNA (1 mg/ml) and 50 μl of 20× SSC. Make up to 200 μl with TE. Incubate the mixture in boiling water for 10 min and cool in ice for 3–5 min. Incubate at 65° C for 90 min before adding to the prewarmed hybridization solution.

Hybridization of Filters

1. After UV cross-linking, library filters are incubated at 65° C in the prehybridization solution.
2. After 4 hr, discard the prehybridization solution. Hybridize the filters for 16–20 hr at 65° C in fresh prehybridization solution containing the preannealed ^{32}P-labeled subtracted cDNA probe.
3. Wash the filters sequentially in solution A [2× SSC, 0.1% (w/v) sodium dodecyl sulfate (SDS)] and solution B [0.2× SSC, 0.1% (w/v) SDS] with several changes. Washings are performed for 15 min each at room temperature. If needed, wash the filters at 60° C.
4. Expose the filters to Kodak (Rochester, NY) XAR-5 film with intensifying screen at −70° C.

Plaque Purification

1. Develop the X-ray film, align it to the filters, and mark appropriately.
2. Align the plates with the positive hybridizing signals on the X-ray film, using a light-box. Hybridization to replica filters helps in identifying real positive clones.
3. Pick up positive plaques by stabbing into the agar plate with a cut pipette tip. Place each agar plug into 1 ml of SM containing a drop of chloroform.
4. Allow phage particles to diffuse out of the agar at room temperature for 4 hr to overnight. The first screening does not allow the isolation of a single hybridizing plaque. The positive phage clones are, therefore, plaque purified by an additional round of screening.
5. Replate an aliquot of phage clones obtained after the first screening. Dilute the phage particles in SM to obtain 200–500 plaques for the second screening. Methods for plating, preparation, and hybridization of filters are essentially as described above.
6. Pick up a single, well-isolated, positive plaque from each independent plate for preparing DNA.

Phage DNA Preparation

1. Amplify the purified phage clone by plating ~100,000 plaques on a 150-mm NZ–agarose plate [NZCYM medium containing 1.5% (w/v) agarose]. This allows complete lysis of the plating bacteria.
2. Add 10 ml of SM to each plate. Rotate the plates gently on a shaker for 4 hr to overnight at room temperature.
3. Collect SM containing the phage lysate off the plates and centrifuge at 4000 *g* for 10 min at 4° C. Discard the pellet of bacterial debris. Store 1 ml of the phage lysate at 4° C with 50 μl of chloroform. (This serves as phage stock.)
4. To the remaining phage lysate, add an equal volume of PEG solution [20% (w/v) PEG 8000, 2 *M* NaCl in SM]. Mix gently.
5. Keep the mixture on ice for 1–2 hr to allow the precipitation of phage particles.
6. Centrifuge at 10,000 rpm for 20 min at 4° C. Discard the supernatant. Invert the tubes on a paper towel and wipe the rim because it is important to remove any remaining PEG solution.
7. Resuspend the phage pellet into 0.6 ml of SM.
8. Transfer the suspension into a microcentrifuge tube. Extract once by mixing with an equal volume of chloroform–isoamyl alcohol (25:1, v/v), and centrifuge at 14,000 rpm for 5 min at room temperature.
9. Transfer the aqueous (upper) phase into a new microcentrifuge tube and add DNase I (10 μg) and RNase A (100 μg). Incubate at 37° C for 30 min.
10. Add 30 μl of 0.5 *M* EDTA (pH 8.0), 30 μl of 10% (w/v) SDS, and 5 μl of proteinase K (25 μg/μl). Mix well by inverting the tubes. Centrifuge briefly and incubate at 65° C for 60 min.
11. Extract twice with an equal volume of phenol–chloroform mixture (1:1, v/v) and once with chloroform–isoamyl alcohol (24:1, v/v). Centrifuge at 14,000 rpm for 5 min at room temperature to separate the phases.
12. To the aqueous phase, add $\frac{1}{10}$ volume of 3 *M* sodium acetate (pH 5.2) and 2 vol of ethanol. Mix gently. The DNA fiber may be visible.
13. Pellet or spool the phage DNA. Wash with 70% (v/v) ethanol, dry in a vacuum (*caution:* do not overdry), and dissolve in 100–200 μl of TE containing RNase (100 μg/ml).

Southern Analysis of Purified Genomic Clones with Subtracted cDNA Probe

1. Digest the phage DNA from purified clones with a restriction enzyme (e.g., *Hin*dIII), and separate the fragments by electrophoresis using a 1% (w/v) agarose gel.
2. Transfer DNA to Hybond N membrane (24).
3. Hybridize Southern blots of phage DNAs to the original subtracted cDNA probe that was used for library screening. Select positive phage clones for further characterization.

4. Also, hybridize the blots to ^{32}P-labeled Cot-1 DNA probe to identify fragments containing highly repetitive sequences.

Regional Localization and Microsatellite Content Analysis

The genomic clones that hybridize to the subtracted cDNA probe on Southern analysis are then regionally localized by fluorescence *in situ* hybridization (FISH) (25). Because the average size of inserts in genomic clones from the sorted X-chromosome library is over 15 kb, it is also possible to identify polymorphic microsatellite repeats in several of the clones (26). Fluorescence *in situ* hybridization and linkage analysis using polymorphic repeats (these methods are beyond the scope of this chapter) allow physical and genetic localization of the genomic clones. One can, therefore, focus on the isolation of transcribed sequences in the genomic clones that are in the region of interest.

Isolation of Encoded cDNAs

1. Identify DNA fragments that hybridize to the subtracted cDNA probe on Southern analysis. Purify these fragments by low melting point gel electrophoresis.
2. Label with ^{32}P, using random primers (Megaprime labeling kit; Amersham Life Sciences).
3. Use the probe after preannealing with the human Cot-1 DNA to screen corresponding cDNA libraries.
4. Purify positive clones and characterize the cDNAs.

Standard recombinant cDNA methods are described in the *Current Protocols* manual (24). The exon-amplification method (4) can also be used to identify transcribed sequences present within the genomic clones isolated with the subtracted cDNA probe. For this, we recommend using the exon-amplification kit from GIBCO-BRL.

Comments

We have isolated at least 14 independent clones by screening the human X-chromosome-specific genomic library (LA0XNL01) with ^{32}P-labeled subtracted retinal cDNA probe. Of these, 11 clones hybridized back to the subtracted probe on Southern analysis. Most of the fragments that hybridize to the subtracted probe did not contain highly repetitive sequences as revealed by hybridization to the Cot-1 DNA probe. Four of the genomic clones were regionally localized by FISH to

Xp11.23–Xp21.1, where several genetic loci for retinal diseases [including *RP2* and *RP3* responsible for X-linked retinitis pigmentosa (XLRP)] have previously been mapped (27). We have characterized polymorphic dinucleotide repeats from several of the X-clones, and identified transcribed sequences in at least one of these clones (27a).

The subtracted cDNA probe can also be used for screening genomic libraries from specific microdissected chromosomal regions (28) or for hybridization to YAC and cosmid clones spanning particular disease gene regions.

Sandwich–Selection Method for Isolating cDNA Clones from Large Genomic Regions

Sandwich–selection (schematically represented in Fig. 2) is a hybridization-based method for isolating longer cDNAs from large genomic regions (29). It consists of (a) hybridization of single-stranded (ss) circular cDNAs to the simian virus 40 (SV40)-tagged *in vitro*-synthesized RNA (called "genomic RNA"), (b) retention of "genomic RNA"–cDNA hybrids on an avidin matrix through a biotin-labeled RNA (called "capture RNA") via an SV40-tagged sequence, (c) elution of specific cDNAs, which hybridize to the "genomic RNA," from the avidin matrix with RNase A, and (d) electroporation of eluted ss cDNAs into electrocompetent *E. coli* cells either directly or after conversion to double-stranded DNA.

Materials

Genomic DNA: DNA from YAC clones spanning *OTC, DXS140,* and the X-linked retinitis pigmentosa locus *RP3* and/or their derivative cosmids (30) is used for selection experiments

cDNA libraries: Because this method requires single-stranded cDNAs for hybridization, cDNA libraries should be in vectors containing *ori* from single-stranded phage (e.g., f1). A number of cDNA libraries from human tissues and cell lines have been constructed in the Charon BS phage vector (20) and can be transferred to the Bluescript plasmid (31). These libraries are available from ATCC and/or from the authors. cDNA libraries in the λZAP vector can be obtained from commercial sources

Vectors: Details for constructing the pSV9Zf9 vector and the pSV7Zf3 plasmid, used for cloning the genomic DNA and for generating the biotinylated "capture RNA," respectively, have been described (29). Both of these plasmids are available from the authors

YT (2×): 16 g of Bacto Tryptone, 10 g of yeast extract, 10 g of NaCl, per liter of H_2O; adjust pH to 7.4 with 5 *N* NaOH

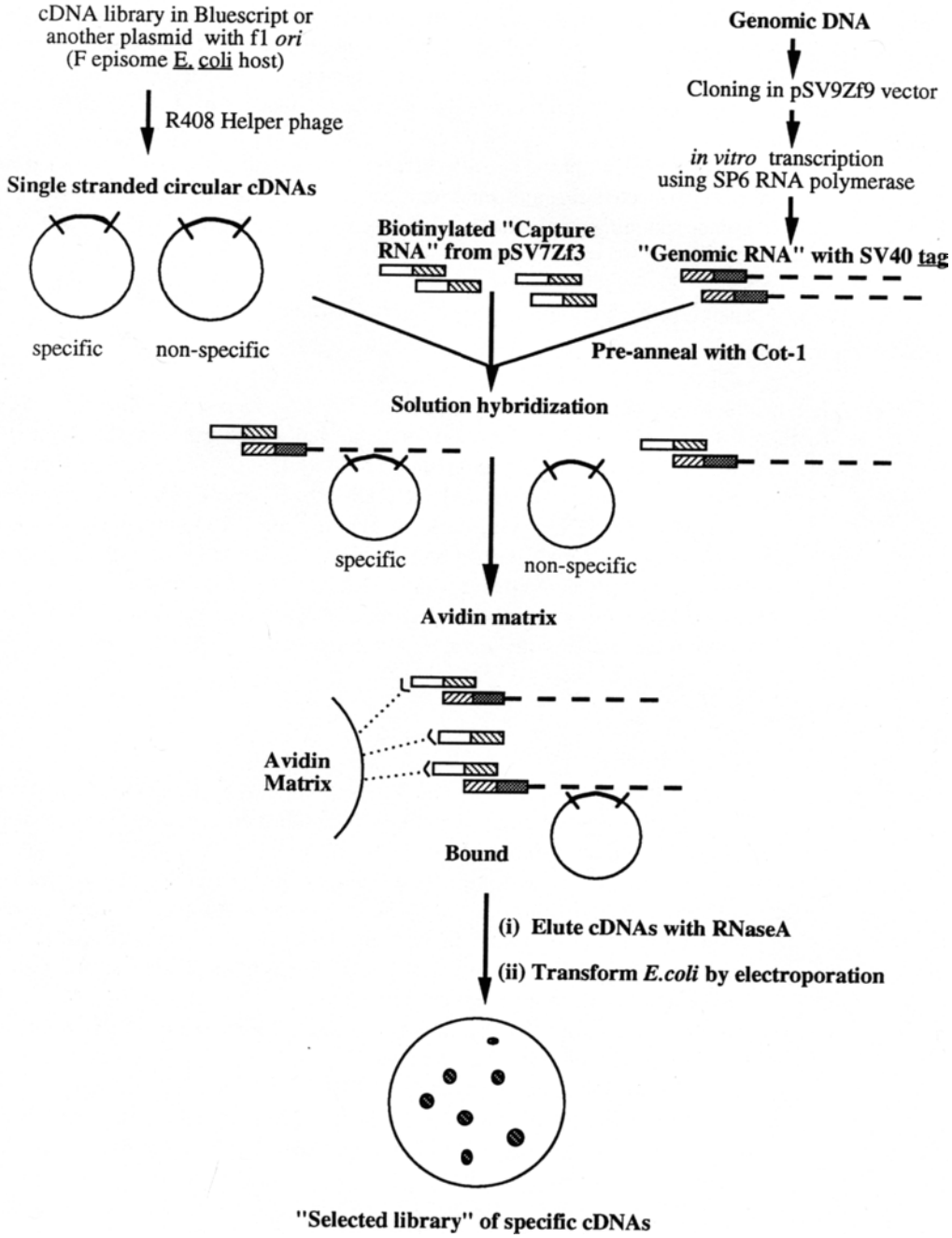


FIG. 2 The sandwich–selection strategy.

R408 helper phage (Stratagene Cloning Systems)
Calf intestinal alkaline phosphatase (CIP): From Boehringer Mannheim
Ethylene glycol bis(β-aminoethyl ether)-*N,N,N',N'*-tetraacetic acid (EGTA): From Sigma Chemical Co.
Reagents for *in vitro* transcription: Purchase from Promega Corp. (Madison, WI)
RNasin ribonuclease inhibitor (Promega)
Transcription buffer (5×): 200 m*M* Tris-HCl (pH 7.5), 30 m*M* $MgCl_2$, 10 m*M* spermidine, 50 m*M* NaCl
Bio-11-UTP (Enzo Biochem, Farmingdale, NY)
RNase-free DNase I (Boehringer Mannheim)
Sephadex G-50 column (Boehringer Mannheim)
Vectrex–avidin (Vector Laboratories, Burlingame, CA)
Streptavidin-conjugated magnetic beads (Dynal, Inc., Lake Success, NY)
Vectrex–avidin binding buffer: (100 m*M* Tris-HCl (pH 7.5), 150 m*M* NaCl, 0.5 (w/v) SDS
Wash buffer A: 10 m*M* Tris-HCl (pH 7.5), 0.5% (w/v) SDS, 2 m*M* EDTA, 100 m*M* NaCl
Wash buffer B: Same as buffer A except 50 m*M* NaCl
Preelution buffer: 10 m*M* Tris-HCl (pH 7.5), 80 m*M* NaCl, 2 m*M* EDTA
Elution buffer: Preelution buffer containing RNase A (125 μg/ml)
Restriction enzymes, Klenow polymerase, and T4 DNA ligase: Purchase from New England Biolabs or GIBCO-BRL

Single-Stranded DNA from cDNA Libraries

Single-stranded DNA is generated from cDNA libraries in the Bluescript vector using R408 helper phage (protocol from Stratagene).

1. Grow an overnight culture of Bluescript plasmid library in XL1-Blue in the presence of tetracycline (12 μg/ml) to allow selection of the F′ episome. Inoculate 3 ml of 2× YT in a 15-ml tube with 300 μl of the overnight culture. Incubate at 37° C in a shaker for 2–3 hr.
2. Add helper phage to the culture during log phase (OD_{600} −0.3 or 2.5 × 10^8 cells/ml) at a multiplicity of infection of 20. Continue shaking for 8 hr with vigorous agitation.
3. Pellet the cells by centrifugation at 12,000 *g* for 15 min at room temperature. Transfer the supernatant to a fresh tube and centrifuge again for 15 min.
4. Precipitate the ss phage by adding 0.25 vol of solution containing 3.5 *M* ammonium acetate, pH 7.5, and 20% (w/v) PEG. Invert to mix, leave at room temperature for 15 min, and then centrifuge for 20 min at 12,000 *g* at room temperature.

Remove the supernatant. Centrifuge again for 1 min and discard any remaining supernatant.

5. Resuspend the pellet in 200 μl of TE. Add 200 μl of phenol–chloroform (1:1, v/v). Mix by vortexing for 1 min. Centrifuge at 12,000 *g* for 5 min at room temperature.
6. Transfer the upper, aqueous phase to a fresh tube. Repeat this extraction until there is no visible material at the interface. Extract with 200 μl of chloroform–isoamyl alcohol (24:1, v/v). Mix by vortexing for 1 min. Centrifuge at 12,000 *g* for 3 min at room temperature and collect the aqueous phase.
7. Precipite the ssDNA by adding 0.5 vol of 7.5 *M* ammonium acetate (pH 7.5) and 2 vol of cold 100% ethanol. Mix and leave at −20° C for 30 min.
8. Centrifuge at 12,000 *g* for 20 min at 4° C. Carefully wash the pellet with cold 70% (v/v) ethanol, and centrifuge again at 12,000 *g* for 2 min at 4° C. Dry the pellet in vacuum.
9. Resuspend the DNA in 20 μl of H_2O. The quantity of DNA can be estimated after agarose gel electrophoresis by comparison of band intensities between the experimental and control samples.

Cloning of Genomic DNA into pSV9Zf9 Vector

1. Digest the pSV9Zf9 vector DNA to completion with *Eco*RI and treat with Klenow polymerase to create blunt ends.
2. Dephosphorylate the vector DNA by incubation with CIP (24; per protocol supplied by the manufacturer). Add $\frac{1}{10}$ volume of 0.5 *M* EGTA and incubate at 65° C for 45 min. Extract the DNA twice with phenol–chloroform (1:1, v/v) and once with chloroform–isoamyl alcohol (24:1, v/v). After ethanol precipitation, dissolve the pelleted DNA in TE.
3. Digest the YAC/cosmid genomic DNA (being used for cDNA selection) separately with two or three frequent-cutting restriction enzymes (e.g., *Rsa*I, *Mbo*I, and *Bst*UI). Treat the DNA with Klenow polymerase to create blunt ends for cloning, if required.
4. Ligate the genomic DNA inserts to the pSV9Zf9 vector (molar ratio 3:1).
5. Electroporate 1 μl of the ligation mixture into electrocompetent *E. coli* cells to generate a sublibrary of genomic clones.
6. Prepare plasmid DNA from the sublibrary. This is used as a template for *in vitro* transcription to generate the driver "genomic RNA" with SP6 RNA polymerase.

Synthesis of Genomic RNA

1. Digest the DNA from the genomic sublibrary in the pSV9Zf9 vector with *Sac*I or *Sal*I.

2. Incubate with proteinase K (50 μg/ml) in the restriction buffer at 37° C for 30 min.
3. Extract twice with phenol–chloroform (1 : 1, v/v) and precipitate with ethanol.
4. For *in vitro* synthesis, mix the following components:

Transcription buffer (5×)	10 μl
Dithiothreitol (DTT), 100 m*M*	5 μl
RNasin	20 U
ATP, GTP, CTP, and UTP (25 m*M* each)	3 μl
*Sac*I- or *Sal*I-digested template DNA (2–5 μg)	■ *x* μl
SP6 RNA polymerase (15–20 U/μl)	2 μl

Make up the volume to 50 μl with RNase-free H_2O. Incubate for 60–120 min at 37–40° C.

5. Add RNase-free DNase I (1 U/μg template DNA). Incubate for 15 min at 37° C.
6. Extract once with phenol–chloroform (1 : 1, v/v) and once with chloroform–isoamyl alcohol (24 : 1, v/v). Transfer the aqueous phase to a fresh tube, and precipitate the RNA with 2.5 vol of 100% ethanol in the presence of 2 *M* ammonium acetate. Place on dry ice for 10 min, and centrifuge at 14,000 rpm for 15 min at 4° C. Wash the pellet with 70% (v/v) ethanol. Dry it and resuspend in 20 μl of RNase-free H_2O.

All genomic RNA molecules that are synthesized using SP6 RNA polymerase have the SV40-tag sequence at their 5′ end.

Synthesis of Biotinylated Capture RNA

Biotinylated "capture RNA" is derived from the pSV7Zf3 construct by *in vitro* transcription with T7 RNA polymerase in the presence of Bio-11-UTP. It contains a region complementary to a 225-nucleotide sequence in the SV40-tag of the genomic RNA.

1. Digest the pSV7Zf3 plasmid DNA with *Sst*I and *Pvu*II. Extract with phenol–chloroform and precipitate with ethanol. Dissolve the pelleted DNA in RNase-free H_2O.
2. The *in vitro* transcription reaction contains the following:

Transcription buffer (5×)	10 μl
DTT (100 m*M*)	5 μl
RNasin	20 U
ATP, GTP, and CTP (25 m*M* each)	3 μl
Bio-11-UTP (20 m*M*)	5 μl
*Sst*I- and *Pvu*II-digested template DNA (2–5 μg)	■ *x* μl
T7 RNA polymerase (15–20 U/μl)	2 μl

Make up the volume to 50 μl with RNase-free H_2O. Incubate for 60–90 min at 37° C.

3. Add RNase-free DNase I (1 U/μg of template DNA). Incubate for 15 min at 37° C.
4. Add 2 μl of 0.5 *M* EDTA to stop the reaction.
5. Purify the *in vitro*-synthesized biotinylated RNA by passing it through a Sephadex G-50 column.
6. Use an aliquot to estimate the amount by OD_{260} measurement and by gel electrophoresis. Store the RNA at −70° C until use.

cDNA Selection Method

1. Mix the genomic RNA (10–50 μg) with Cot-1 DNA (~20 μg) in 0.5 *M* sodium phosphate (pH 7.2)–10 m*M* EDTA. Denature by incubation in boiling water for 10 min. Preanneal at 60° C for 90 min (per manufacturer-supplied protocol).
2. Hybridize the preannealed genomic RNA with ss cDNAs and the biotinylated capture RNA in 0.5 *M* sodium phosphate (pH 7.2)–10 m*M* EDTA–0.5% (w/v) SDS in a final volume of 50 μl. Carry out the hybridization in a sealed capillary (sterilized, 100-μl size; sealed on both ends by rotating on a flame) at 60° C for 16 hr. Hybridization can also be performed in a 0.2-ml Eppendorf tube with a drop of mineral oil on top of the solution.

Genomic RNA is generally in 10- to 100-fold molar excess compared to the ss cDNAs. The amount of biotinylated capture RNA should be in 2- to 10-fold molar excess compared to the genomic RNA. In general, we use 0.5–2 μg of ss cDNA with 10–40 μg of genomic RNA and 10–50 μg of the biotinylated capture RNA.

After hybridization, the cDNA–genomic RNA hybrids are retained on an avidin matrix via the biotinylated capture RNA. We used Vectrex–avidin for capturing these hybrid molecules. However, this matrix can be replaced with streptavidin-conjugated magnetic beads, which may provide even greater enrichment.

3. Before use, preequilibrate Vectrex–avidin with binding buffer containing sonicated, denatured herring sperm DNA (50–100 μg/ml) at room temperature.
4. Incubate the hybridization mixture with Vectrex–avidin (100–250 mg) in the binding buffer for 60 min at room temperature with intermittent mixing.
5. Wash the Vectrex–avidin beads twice each with the binding buffer, buffer A, and buffer B.
6. Equilibrate Vectrex–avidin in the preelution buffer.
7. Elute the specifically hybridizing ss cDNAs from the matrix by incubation at 37° C in 400–600 μl of elution buffer for two successive periods of 30 min each.
8. Extract the eluates once with phenol–chloroform (1 : 1, v/v) and once with chloroform–isoamyl alcohol (24 : 1, v/v).

9. Precipitate the ss cDNA with ethanol in the presence of 20 μg of glycogen (Boehringer Mannheim).
10. Use the ss cDNA for electrotransformation of XL1-Blue cells, either directly or after converting these into the double-stranded DNA (18, 21).

The cDNA clones are then characterized further to identify those mapping back to the original cosmid and YAC clones used for selection.

Comments

Sandwich–selection is complementary to the existing hybridization-based selection methods (5–8). The use of a sandwich for capturing cDNAs on an avidin matrix and their elution with RNase A provide higher efficiency and specificity in this method. A major advantage is that larger-size cDNA clones are obtained after selection. In addition, PCR is not used during any step in this procedure, eliminating size bias and potential artifacts. However, the use of ssDNAs can introduce some bias and lead to partially chimeric clones.

cDNA Selection by Magnetic Capture and Filter Hybridization (Direct Selection)

Two hybridization-based selection methods—magnetic capture (7, 8) and filter hybridization (5, 6)—have been widely used for isolating cDNA clones from large genomic regions. For magnetic capture, biotinylated genomic DNA is hybridized to PCR-amplified cDNAs in solution and specific genomic DNA–cDNA hybrids are captured with streptavidin-conjugated magnetic beads. In the filter hybridization method, genomic DNA is cross-linked to a nylon membrane that is then hybridized to PCR-amplified cDNA inserts. These methods allow simultaneous selection from a number of cDNA sources. Although a high degree of enrichment is reported by different groups, the selection of small-size cDNA clones remains a major problem. We have modified the two methods to obtain larger clones from high-complexity cDNA libraries. The modified magnetic capture method is schematically illustrated in Fig. 3.

Materials and Reagents

cDNA libraries: We have used cDNA libraries constructed in several different phage or plasmid vectors (e.g., Charon BS, λZAPII, λgt11, Bluescript) for cDNA selection experiments (see Ref. 20 for a list of cDNA libraries). cDNAs used in our experiments are primarily from oligo(dT)-primed li-

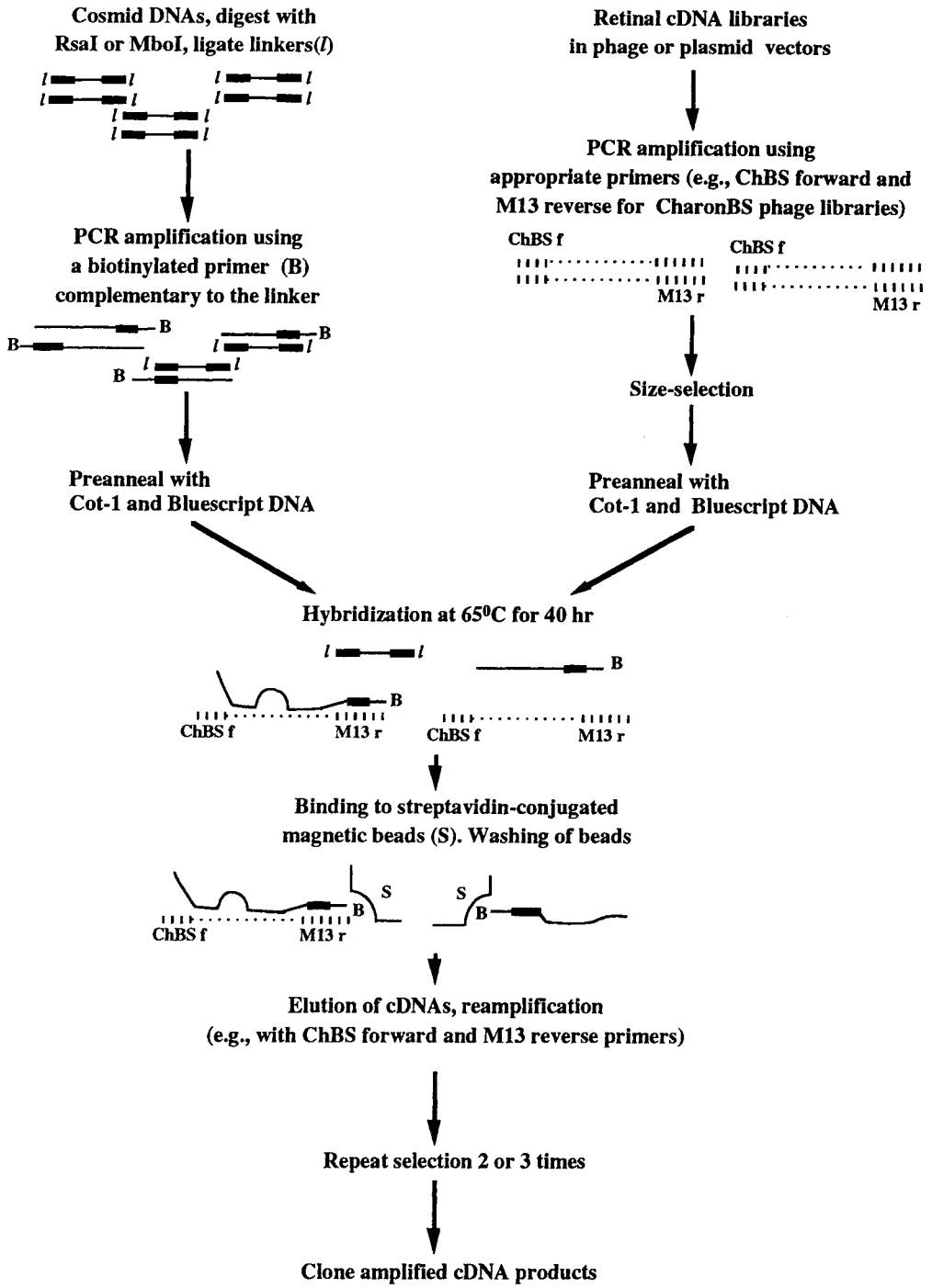


FIG. 3 A modified magnetic capture method for selecting cDNAs from retinal libraries.

braries that contain inserts ranging from 0.4 to 3 kb. If mRNA from appropriate tissue(s) is available, primary uncloned cDNAs can be used for selection

Genomic DNA: We have used eight partially overlapping cosmid clones (derived from a YAC clone) spanning ~250 kb from the RP3 critical region (30) as the source of genomic DNA for cDNA selection experiments described here

SeaKem GTG and SeaPlaque agarose (FMC Bioproducts, Rockland, ME)

Agarase enzyme (New England Biolabs)

Streptavidin-conjugated magnetic beads, Dynabeads (Dynal, Inc.)

Dynabeads binding buffer: 10 m*M* Tris-HCl (pH 7.5), 1 m*M* EDTA, 2 *M* NaCl

Hybridization solution for magnetic capture: 0.75 *M* NaCl, 20 m*M* $NaPO_4$ (pH 7.2), 10 m*M* EDTA, 5× Denhardt's solution, 0.6% (w/v) SDS

Hybridization solution for filter selection: 5× SSC, 5× Denhardt's solution, 2% (w/v) SDS

Microcon 100 (Amicon, Beverly, MA)

TA cloning system (Invitrogen Corp., San Diego, CA)

Preparation of cDNA Inserts

Phage lysate or DNA from cDNA libraries are used to generate inserts by PCR amplification.

1. Amplify different quantities of the template (1, 2, and 5 μl) using appropriate primer sets for cDNA insert amplification from different libraries (as shown in Table I). A typical 100-μl PCR reaction includes a 2 μM concentration of each primer, 10 m*M* Tris-HCl (pH 8.2), 50 m*M* KCl, 1.5 m*M* $MgCl_2$, a 200 μM concentration of each dNTP, and 2.5 U of AmpliTaq. The recommended thermal cycling profile is 30 cycles of 94° C for 45 sec, 55° C (Charon BS or M13 primers) or 60° C (λgt11 primers) for 45 sec, at 72° C for 2 min. Include a reaction without primers and another without template as controls.

TABLE I Primers for Polymerase Chain Reaction Amplification of cDNA Inserts from Libraries

Primer	Sequence	Vector
ChBS forward	5′ GCA ACG CCA AAG GCG GTT 3′	Charon BS
M13 reverse	5′ GGA AAC AGC TAT GAC CAT G 3′	Charon BS, Bluescript, and λZAP
M13 forward	5′ GTA AAA CGA CGG CCA GT 3′	Bluescript and λZAP
gt-forward	5′ GGT GGC GAC GAC TCC TGG AGC CCG 3′	λgt11
gt-reverse	5′ TTG ACA CCA GAC CAA CTG GTA ATG 3′	λgt11

2. Analyze an aliquot of the PCR reaction by gel electrophoresis [2% (w/v) agarose, SeaKem GTG] to determine template quantity that yields high complexity of cDNAs.
3. Obtain >3 μg of amplified products from various cDNA libraries using PCR conditions optimized earlier.
4. Size-select amplified cDNAs on a low melting point agarose gel. Isolate cDNA products ranging from 0.6 to 5 kb in size for libraries in the Charon BS vector, and from 0.2 to 3 kb for libraries in the Bluescript plasmid, or λZAP and λgt11 phage vectors. Gel electrophoresis also helps in removing excess primers and certain vector-specific products.
5. Melt the agarose slices (containing the DNA) at 65° C. Cool at 37° C for 2 min, and add agarase (1 U/500 μl of gel). Incubate overnight at 37° C.
6. Centrifuge at 13,000 rpm for 2 min to pellet any remaining agarose particles.
7. Transfer the supernatant to a fresh tube and precipitate the DNA with ethanol.

Magnetic Capture (Solution Hybridization)

Preparation of Biotinylated Genomic DNA

1. Digest the cosmid (or YAC insert) DNA (~1 μg, or 100–150 ng each of several different cosmids) to completion with a frequent-cutting restriction enzyme [e.g., *Rsa*I (blunt ends) or *Mbo*I (5′ overhang, cohesive ends)] in a final volume of 40 μl. Analyze an aliquot by agarose gel electrophoresis to verify complete digestion.
2. Ligate an aliquot (10 μl) of the digested DNA to 400 ng of appropriate linkers (see Table II) in a total volume of 15 μl for 16 hr at 22° C (blunt ends) or 14° C (cohesive ends).
3. Amplify the genomic DNA by PCR, using biotinylated primers. A typical 100-μl PCR reaction includes 5 μl of ligation mixture, 10 m*M* Tris-HCl (pH 8.2), 50 m*M* KCl, 1.5 m*M* $MgCl_2$, a 200 m*M* concentration of each dNTP, 2.5 U of AmpliTaq, and 200 pmol of the biotinylated primer. The recommended thermal cycling profile is 30 cycles of 94° C for 45 sec, 60° C for 45 sec, and 72° C for 2 min.

TABLE II Linkers and Primers

Type	Sequence
Linker 5′ overhang	5′ GATC TC GAC GAATTC GTGAGACCA 3′[a]
	3′ AG CTG CTTAAG CACTCTGGT 5′[b]
Linker blunt	5′ GAT CTC GAC GAA TTC GTG AGA CCA 3′[a]
	3′ CTA GAG CTG CTT AAG CAC TC 5′
Biotinylated primer	5′ biotin-TGG TCT CAC GAA TTC GTC GA 3′[b]

[a] Same primer.
[b] Same sequence.

4. Verify the complexity of amplified products by analyzing an aliquot of the PCR reaction on a 2% (w/v) agarose gel. (It should be similar to that observed for the digested DNA.)

Hybridization of cDNAs to Biotinylated Genomic DNA

1. Preanneal the amplified genomic DNA and cDNA products separately to human Cot-1 and plasmid (Bluescript or pBR) DNA. Denature a mixture of cosmid DNA (100 ng; containing an equal quantity for each cosmid), Cot-1 (4 μg), and Bluescript DNA (50 ng; digested with *Rsa*I) in 20 μl of hybridization solution in boiling water for 10 min, and anneal for 4 hr at 65° C. Similarly, denature and preanneal the cDNA products (2 μg; containing an equal quantity for each library) with Cot-1 (2 μg) and Bluescript DNA (50 ng; digested with *Rsa*I) in 20 μl of hybridization solution.
2. Combine the genomic mixture with the cDNA solution and hybridize for 40 hr at 65° C.
3. Before use, wash Dynabeads (10 mg) twice in 1 ml of binding buffer. Recover the beads with a magnetic concentrator (Dynal) for 1 min. Resuspended in 50 μl of 2× binding buffer.
4. Mix the Dynabeads with the hybridization solution (containing genomic DNA–cDNA hybrids). Shake gently for 30 min at room temperature.
5. Wash the Dynabeads with genomic DNA–cDNA hybrids twice with 2× SSC–0.1% (w/v) SDS for 15 min at room temperature, and three times with 0.1× SSC–0.1% (w/v) SDS at 65° C.
6. Resuspend the Dynabeads in 100 μl of TE and elute the selected cDNAs by incubation in boiling water for 5 min.
7. Collect the eluted cDNAs, and amplify different aliquots (1, 2, and 5 μl) in a 100-μl PCR reaction under the conditions used earlier for preparing cDNA inserts.
8. Use an amount of template that yields a smear with some strong bands to amplify larger quantities of cDNAs (3–4 μg).
9. Purify the amplified products by passing them through a Microcon 100 column. This step removes the remaining primers from the reaction and concentrates the selected cDNAs.
10. Repeat the magnetic capture procedure to enrich for specific cDNAs by hybridizing again to the biotinylated genomic DNA.

Filter Hybridization

Preparation of Membrane Filters

1. Digest the cosmid (or YAC) DNA (~500 ng) with a restriction enzyme that recognizes a 6-bp sequence (e.g., *Hin*dIII) in a total volume of 20 μl. Analyze an aliquot of the digested DNA (5 μl) on a 1% (w/v) agarose gel to verify complete digestion.

2. Denature the remaining 15 μl of DNA by incubation with 1 μl of 2 *N* NaOH at 37° C for 10 min. Pool the DNA from several cosmid clones.
3. Apply a 1-μl aliquot (containing 3 ng of each cosmid DNA) onto a Hybond N membrane. Label the wet spot with pencil for identification.
4. Air dry the membrane for 2 min, cross-link with the UV light, and keep in a plastic bag until use.

Hybridization of Filters

1. Prehybridize membrane filters containing the genomic DNA with denaturated Cot-1 (10 μg) and pBR322 (2 μg) DNA in 60 μl of filter hybridization solution, covered with mineral oil, for 4 hr at 65° C.
2. Prehybridize the amplified cDNA (2–3 μg) with Cot-1 (10 μg) and pBR322 (2 μg) DNA in 60 μl of filter hybridization solution for 4 hr at 65° C.
3. Transfer the membrane filters to the cDNA tube and incubate for 40 hr at 65° C.
4. Wash the filters in Eppendorf tubes twice with 2× SSC for 15 min each at room temperature, twice with 0.2× SSC for 15 min each at 65° C, and once with 0.1× SSC for 15 min at 65° C.
5. Elute the hybridized cDNAs by incubating the filter in 40 μl of TE for 5 min at 100° C.
6. Use aliquots from the eluted sample (1, 2, and 5 μl) for PCR to amplify cDNAs (conditions will be similar to those used for preparation of cDNA inserts).
7. Choose an amount of sample and PCR conditions yielding a smear (with some strong bands) to obtain additional product (about 3–4 μg).
8. Purify the amplified cDNA products by passing them through a Microcon 100 column.
9. Repeat selection (two or three cycles) by hybridization of "selected cDNAs" to fresh genomic DNA on filter membranes.

Cloning of Selected cDNAs

Amplified cDNAs obtained by either solution or filter hybridization method are cloned using the TA cloning system and following the manufacturer recommendations.

1. Ligate ~4 ng of purified selected cDNAs to 50 ng of TA cloning vector, pCRII.
2. Electroporate 1 μl of the ligation mixture into electrocompetent XL1-Blue cells, and spread on LB plates containing ampicillin or kanamycin and X-Gal/IPTG to identify the recombinant bacterial colonies. Most of the colonies should be white and will contain insert.
3. Lift the replica filters from the plates, and hybridize these to the ^{32}P-labeled cosmid DNA (used initially for cDNA selection) (24).
4. Pick up positive colonies (of selected cDNA clones) for further characterization.

Comments

The direct selection approach of magnetic capture or filter hybridization (5–8) is widely used for isolation of cDNAs from large genomic regions because of its relative simplicity in comparison to the screening of cDNA libraries with cosmid or YAC DNA probes. The direct cDNA selection methods also permit the simultaneous search for rare transcripts in several cDNA libraries with varying complexity and from different tissues or developmental stages. A major problem, however, encountered when YAC DNA is used for cDNA selection is caused by simultaneous enrichment of rRNA clones present in random-primed cDNA libraries. The use of a mixture of blocking reagents including rRNA, tRNA, and yeast DNA can alleviate this problem considerably. Genomic DNA cloned in cosmid vectors is a preferred substrate for cDNA selection. Another problem with direct selection is the recovery of short cDNA fragments, which must be used for screening cDNA libraries to isolate larger clones. We have used a group of cosmids spanning ~250 kb in the region of XLRP locus *RP3* for selecting cDNAs from a number of large-insert retinal libraries. Using the methods described here, we have been successful in isolating large cDNAs (0.5–2 kb) by both magnetic capture and filter hybridization.

Acknowledgments

We thank Dr. Julia E. Richards for helpful comments and Ms. Dorothy Giebel for secretarial assistance. This research is supported by grants from the National Institutes of Health (RO1 EY07961 and R55 ODO1322) and from the Foundation Fighting Blindness (Baltimore, MD).

References

1. F. S. Collins, *Nature Genet.* **1,** 3 (1992).
2. A. Ballabio, *Nature Genet.* **3,** 277 (1993).
3. F. S. Collins, *Nature Genet.* **9,** 347 (1995).
4. A. J. Buckler, D. D. Chang, S. L. Graw, J. D. Brook, D. A. Haber, P. A. Sharp, and D. E. Housman, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 4005 (1991).
5. S. Parimoo, S. R. Patanjali, H. Shukla, D. D. Chaplin, and S. M. Weissman, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 9623 (1991).
6. M. Lovett, J. Kere, and L. M. Hinton, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 9628 (1991).
7. D. A. Tagle, M. Swaroop, M. Lovett, and F. S. Collins, *Nature (London)* **361,** 751 (1993).
8. M. Lovett, *in* "Current Protocols in Human Genetics" (N. C. Dracopoli, J. L. Haines, B. R. Korf, D. T. Moir, C. C. Morton, C. E. Seidman, J. G. Seidman, and D. R. Smith, eds.), Vol. 1, p. 6.3.1. John Wiley & Sons, New York, 1994.
9. M. D. Adams, J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, H. Xiao, C. R. Merril, A. Wu, B. Olde, R. F. Moreno, A. R. Kerlavage, W. R. McCombie, and J. C. Venter, *Science* **252,** 1651 (1991).

10. J. M. Sikela and C. Auffray, *Nature Genet.* **3,** 189 (1993).
11. D. R. Maglott, A. S. Durkin, and W. C. Nierman, *in* "Identification of Transcribed Sequences" (U. Hochgeschwender and K. Gardiner, eds.), p. 273. Plenum Press, New York, 1994.
12. D. R. Moir, R. Lundstorm, P. Richterich, X. Wang, M. Atkinson, K. Falls, J. Mao, D. R. Smith, and G. F. Vovis, *in* "Identification of Transcribed Sequences" (U. Hochgeschwender and K. Gardiner, eds.), p. 289. Plenum Press, New York, 1994.
13. G. Halder, P. Callaerts, and W. J. Gehring, *Science* **267,** 1788 (1995).
14. C. Fields, M. D. Adams, O. White, and J. C. Venter, *Nature Genet.* **7,** 345 (1994).
15. V. A. McKusick, "Online Mendelian Inheritence in Man." Johns Hopkins University School of Medicine, Baltimore, Maryland, 1995.
16. J. R. Duguid, R. G. Rohwer, and B. Seed, *Proc. Natl. Acad. Sci. U.S.A.* **85,** 5738 (1988).
17. H. L. Sive and T. St. John., *Nucleic Acids Res.* **16,** 10937 (1988).
18. J. L. R. Rubenstein, A. E. J. Brice, R. D. Ciaranello, D. Denney, M. H. Porteus, and T. B. Usdin, *Nucleic Acids Res.* **18,** 4833 (1990).
19. A. Swaroop, J. Xu, N. Agarwal, and S. M. Weissman, *Nucleic Acids Res.* **19,** 1954 (1991).
20. A. Swaroop and J. Xu, *Cytogenet. Cell Genet.* **64,** 292 (1993).
21. A. Swaroop, *in* "Methods in Neurosciences" (P. A. Hargrave, ed.), Vol. 15, p. 285. Academic Press, San Diego, California, 1993.
22. W.-B. Li, C. E. Gruber, J.-J. Lin, R. Lim, J. M. D'Alessio, and J. A. Jessee, *BioTechniques* **16,** 722 (1994).
23. H.-C. Aasheim, A. Deggerdal, E. B. Smeland, and E. Hornes, *BioTechniques* **16,** 716 (1994).
24. F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl, eds., "Current Protocols in Molecular Biology." John Wiley & Sons, New York, 1994.
25. J. H. M. Knoll and P. Lichter, *in* "Current Protocols in Human Genetics" (N. C. Dracopoli, J. L. Haines, B. R. Korf, D. T. Moir, C. C. Morton, C. E. Seidman, J. G. Seidman, and D. R. Smith, eds.), Vol. 1, p. 4.3.1. John Wiley & Sons, New York, 1994.
26. J. L. Weber, *in* "Genome Analysis 1: Genetic and Physical Mapping" (K. E. Davies and S. M. Tilghman, eds.), p. 159. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1990.
27. H. F. Willard, F. Cremers, J. L. Mandel, A. P. Monaco, D. L. Nelson, and D. Schlessinger, *Cytogenet. Cell Genet.* **67,** 296 (1994).
27a. D. Yan, H. Pawar, S. A. Statum-Brown, K. Zheng, H. Ehrenberg, D. Wong, L. Gibson, M. Coffey, J. E. Richards, T. L. Yang-Feng, and A. Swaroop, manuscript submitted (1995).
28. X.-Y. Guan, P. S. Meltzer, J. Cao, and J. M. Trent, *Genomics* **14,** 680 (1992).
29. D. Yan and A. Swaroop, *in* "Identification of Transcribed Sequences" (U. Hochgeschwender and K. Gardiner, eds.), p. 91. Plenum Press, New York, 1994.
30. R. Fujita, D. Ayon, L. Gieser, C. Skolnick, T. Yang-Feng, P. Sieving, and A. Swaroop, *Am. J. Hum. Genet.* **53**(Suppl.), A1290 (1993).
31. A. Swaroop and S. M. Weissman, *Nucleic Acids Res.* **16,** 8739 (1988).

[12] Fluorescent Differential Display Method for High-Speed Scanning of Tissue- or Cell-Specific Transcripts

Takashi Ito and Yoshiyuki Sakaki

Introduction

The identification of messages transcribed in a spatially and/or temporally regulated manner and the elucidation of the molecular mechanisms controlling differential gene expression are of interest to molecular biologists because these concepts form the basis for various biological phenomena such as development, growth, differentiation, and cell death. The first step in these studies is the identification of differentially expressed transcripts using various subtractive and differential hybridization-based techniques.

An alternative approach to this goal has been independently developed by two groups, and termed differential display (DD) and RNA fingerprinting by arbitrarily primed polymerase chain reaction (RAP–PCR), respectively (1, 2). Both are based on the arbitrary primer PCR technique that had been developed as a method for genomic fingerprinting (3, 4). The basic concept of DD and RAP-PCR is to amplify selectively a defined subset of transcripts by means of reverse transcription (RT)-PCR using primers of arbitrarily chosen sequences and to display them as fingerprints on the gel. Using this technique, different subsets of cDNAs can be sampled from the original complex RNA population by simply changing the primers used in each analysis.

In DD of mRNA (1), the first strand of cDNA is synthesized by reverse transcription using a unique 3′-anchored oligo(dT) primer to reduce the complexity of the cDNA and to fix the priming site to the beginning of the poly(A) tail. The second-strand synthesis and subsequent amplification of 3′-end portions of selected cDNAs are conducted using PCR with a low annealing temperature (~40° C), an arbitrary primer (usually, 10-mers), and the anchored primer used in the reverse transcription step (Fig. 1). In the original RAP-PCR (2), a single arbitrarily chosen primer(s) ~20 nucleotides long is used throughout the reverse transcription and subsequent PCR steps. The first cycle of RAP-PCR for the second-strand synthesis employs a low annealing temperature (~40° C) so that low stringency annealing allows even the longer primers to initiate priming on multiple cDNAs, whereas the following cycles are run at a higher annealing temperature to specifically amplify the cDNA fragments that had been tagged by the primer in the preceding steps.

Methods in Molecular Genetics, Volume 8

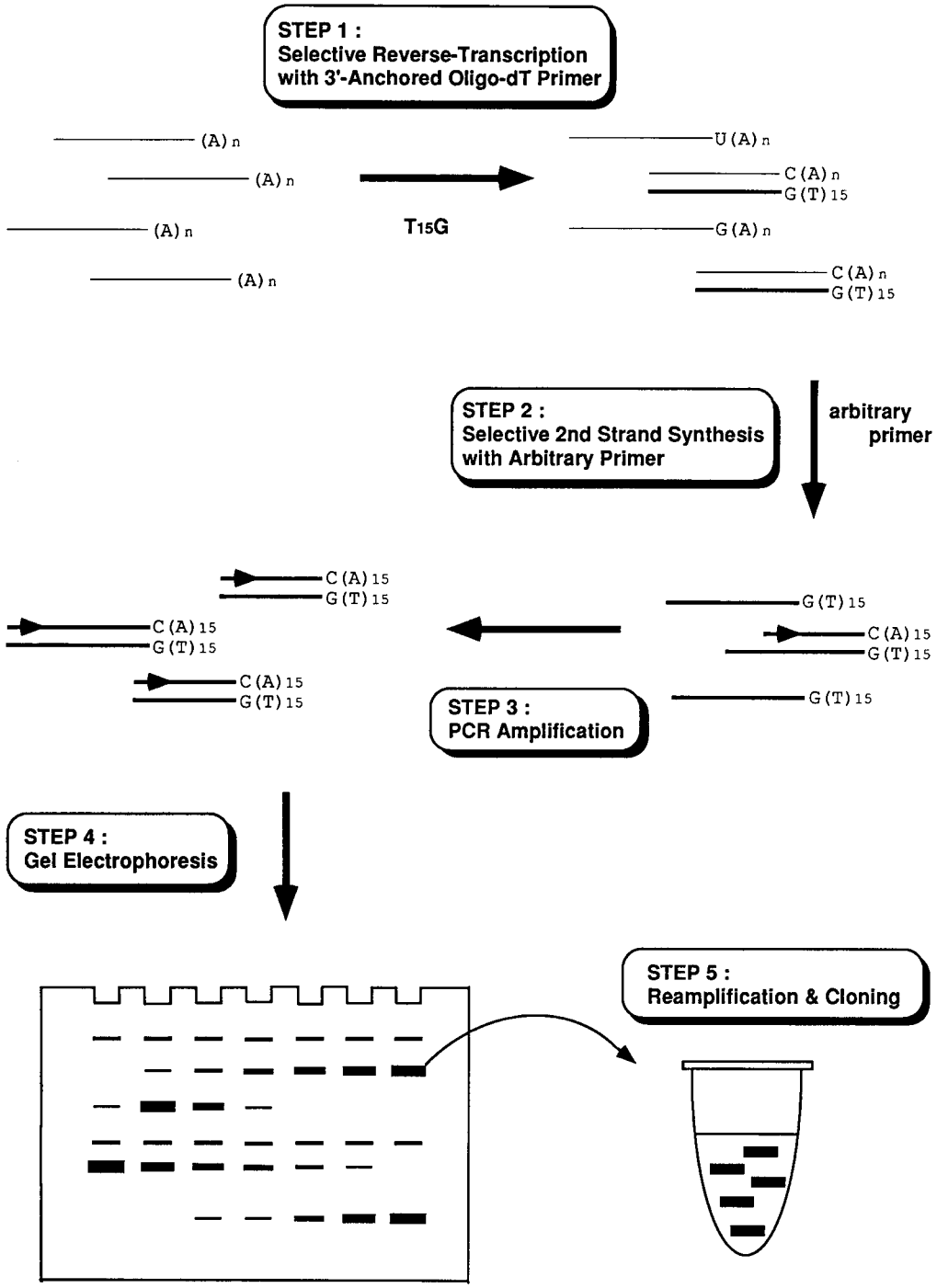


FIG. 1 Principle of differential display. See text for details.

Both methods generate fingerprints on polyacrylamide gels, each of which is composed of a subset of cDNAs sampled from the original complex RNA source in a primer-dependent manner. Parallel comparison of such fingerprints (i.e., lane-to-lane comparison), therefore, allows one to identify differentially expressed species with ease. Such bands can be recovered from the gel and reamplified with the PCR using the same primers to obtain the band in sufficient amounts for use in further characterization. It should be noted that more than two samples can be readily compared by these methods, unlike subtraction methods that are essentially for the isolation of the transcripts present in one sample but not in the other. Furthermore, these methods can perform much more complex comparison among the samples and detect transcripts of various behaviors, including ones with subtle changes that are often overlooked by subtractive hybridization. The sensitivity of the PCR can give access to low-abundance transcripts that are too rare to detect by conventional differential hybridization-based approaches. The amount of RNAs required for DD and RAP–PCR is quite small, so that one can apply these techniques to biological sources where massive RNA preparation is difficult or practically impossible.

These features make DD and RAP-PCR attractive to researchers from various fields where the identification of differentially expressed transcripts is of particular interest and importance. However, to scan most transcripts generated from complex genomes, a considerable number of reactions must be carried out. The number of reactions required (N) is

$$N = \log(1 - P)/\log(1 - B/E)$$

where P is the probability desired, E is the number of expressed transcript species, and B is the number of bands detected per each reaction. Therefore, if each DD reaction amplifies 100 cDNAs ($B = 100$), one must perform about 450 reactions ($N = 448$) to cover all the transcripts in typical mammalian cells ($E = 15{,}000$) with 95% probability ($P = 0.95$).

To run this scale of analysis, one must increase the speed of each analysis, in particular that of the time-consuming and labor-intensive steps of electrophoresis, gel processing after the run, and signal detection. Furthermore, the safety problem inherent in radioactive DD (5) would not be negligible on that scale. Bearing this in mind, we have developed a high-throughput DD termed fluorescent DD (FDD), based on an improved, highly reliable DD protocol and a modified fluorescent DNA sequencer (6).

Here, we describe the protocols for FDD using an automated DNA sequencer and a fluorescent image analyzer. We developed two FDD protocols: protocol S uses shorter anchor primers and arbitrary 10-mers like the original DD (1) and protocol L employs primers having the usual lengths (~20-mer) as arbitrary primers for DD, that is, primers generally found in the laboratory for DD analysis and eliminating the need for preparing arbitrary 10-mers. A fluorescent RAP-PCR procedure is also de-

scribed. In addition, a simple, albeit less sensitive, protocol using highly sensitive dye and ultraviolet illumination is described as an alternative that requires no special equipment.

Fluorescent Differential Display

Materials

The reverse transcriptase of choice is SuperScriptII obtained from Bethesda Research Laboratories (BRL; Gaithersburg, MD). The preamplification kit (BRL) containing the enzyme, 10× buffer, 0.1 *M* dithiothreitol (DTT), 10 m*M* dNTP, and diethyl pyrocarbonate (DEPC)-treated water is used. We have also used the first-strand synthesis kit from Stratagene (La Jolla, CA) to obtain comparable results.

The anchor primers are synthesized on a model 392 DNA synthesizer (Applied Biosystems, Foster City, CA) with FluorPrime (Pharmacia, Uppsala, Sweden). Deprotected oligonucleotides are dried, dissolved in 1 ml of DEPC-treated water, and separated on NAP-10 columns (Pharmacia). Following spectrophotometric measurement (aliquots of oligonucleotides), calculate the concentration of oligonucleotides using the appropriate extinction coefficient (7). [Note the biased base composition of oligo(dT) primers.] Adjust the primer concentrations to 50 μM, and keep them frozen at −20° C until use. If the coupling of fluorescein isothiocyanate (FITC) seemed to be inefficient, purify the oligonucleotides by preparative polyacrylamide gel electrophoresis as described (7). Throughout the preparation steps, avoid contamination with RNase, because the anchor primers are to be incubated with RNAs.

We use AmpliTaq DNA polymerase (Perkin-Elmer, Norwalk, CT), but enzymes from other suppliers give comparable results. We recommend using the 10× buffer supplied by each manufacturer. However, we recently found that a truncated Taq DNA polymerase called "GeneTaq DNA polymerase" supplied from Nippon Gene (Toyama, Japan) gives much stronger signals, in particular, in shorter size range (~500 bp), than any other enzymes we have tested so far. Highly efficient amplification with this enzyme makes the signal detection by Vistra 725 DNA sequencer and FluorImager SI (Molecular Dynamics, Sunnyvale CA.) much easier.

As arbitrary 10-mers, we use 10-mer kits obtained from Operon Technologies (CA). Each kit contains 20 different 10-mers. Dissolve the dried 10-mers in Tris–EDTA (TE) buffer (7) to give a 100 μM stock solution. The working solutions are made by further diluting the stock solution 10-fold with distilled water (10 μM). Both solutions are kept frozen at −20° C. For longer arbitrary primers, use oligonucleotides having ~20 nucleotides.

The acrylamide monomer and *N,N'*-methylenebis(acrylamide) are obtained from Pharmacia. The HydroLink LongRanger (AT Biochem, Malvern, PA) can be used as an alternative gel matrix. Electrophoresis-grade urea for denaturing gels is obtained

from Bio-Rad (Richmond, CA). Tris–base, boric acid, and EDTA for 10× TBE buffer (7) are from Pharmacia. The gel solution, 5 or 6% (w/v) polyacrylamide (acrylamide monomer–bisacrylamide, 19:1) or LongRanger–7 *M* urea–1× TBE, should be filtered and degassed before pouring.

The gel loading solution is composed of 98% (v/v) formamide–10 m*M* EDTA containing 0.01% (w/v) methyl violet. (*Note:* Methyl violet that is too dense may considerably disturb the banding patterns.)

Step 1: Reverse Transcription

Procedure

1. Mix the following ingredients in a 0.5-ml tube:

DEPC-treated water	8.0 μl
Anchor primer (50 μM)	1.0 μl (50 pmol)
Total RNA (2.5 μg/μl)	1.0 μl (2.5 μg)

2. Incubate the tube at 70° C for 5 to 10 min.
3. During the incubation, prepare the following 2× RT solution using solutions supplied with the BRL kit. Prepare enough, taking the loss at each pipetting into account. (For instance, let N be 1.1 × n for n reactions.) Following thorough mixing, place the tube on ice.

DEPC-treated water	(2.0 × N) μl
PCR Buffer (10×)	(2.0 × N) μl
25 mM $MgCl_2$	(2.0 × N) μl
DTT (100 mM)	(2.0 × N) μl
dNTP (10 mM)	(1.0 × N) μl
SuperScriptII (200 U/μl)	(1.0 × N) μl

4. Put the heated tubes containing RNA and primer into an ice–water bath for a few minutes. Spin the tube briefly.
5. Add 10 μl of 2× RT solution into the tubes and mix with the RNA–primer solution thoroughly.
6. Place the tubes in a thermal cycler programmed as follows:

10 min at 25° C
50 min at 42° C
15 min at 70° C
Soak at 4° C

7. Spin the tubes briefly. Then, add 76 μl of TE and 4 μl of the same anchor primer solution as the one used in reverse transcription. Following thorough mixing, store the tube containing the cDNA solution at $-20°$ C.

Comments

Handle the RNAs (7) carefully until the RNA is converted to cDNA.

Although we usually prepare RNAs using a modified acid–guanidine thiocyanate–phenol–chloroform method using TRIzol (BRL) reagent (8), RNAs isolated with other methods can be used for the procedure. However, some samples require further purification. For instance, RNAs prepared from *Xenopus* egg with the TRIzol procedure required further LiCl precipitation for successful FDD (9). Purification as poly(A)$^+$ RNA by oligo(dT)-cellulose chromatography was reported to be rather hazardous for DD (10).

Prior to starting extensive analysis, test whether the RNA samples are contaminated with genomic DNAs. If mock reverse-transcribed cDNAs gave a considerable number of bands, treat the RNA sample with RNase-free DNase (Promega, Madison, WI) to degrade the residual genomic DNAs. Note that some primers may amplify several faint bands even after extensive DNase treatment. However, such bands will not disturb the differential display analysis.

The anchor primers we are presently using are listed below (6) and are no longer the ones (T_{11-12} M*N*) originally reported.

(1) gT_{15}(A/C/G)*N*	*N* = A, C, G, or T
(2) gT_{15}*N*	*N* = A, C, or G
(3) cccggatccT_{15}*N* or cgtacgcgT_{15}*N*	*N* = A, C, or G

We found that the priming efficiency of the original anchor primers was poor and might lead to instability of the reaction, causing run-to-run variation (6). The use of the modified primers shown above drastically increased the amplification efficiency (6). [A similar observation was reported by Liang *et al.*, who also recommended the use of novel anchor primers similar to ours (11).] Primers (1) and (2) contain additional dG at their 5′ ends that are necessary for the generation of signals intense enough for fluorescent detection. Addition of further nucleotides did not improve the signals significantly. The dG would form a terminal G-A mismatch at the reverse transcription step and is thus expected to have the least effect on sequence discrimination. Primer (3) has restriction enzyme sites (*Bam*HI or *Mlu*I) at the 5′ end. These are attached to increase the T_m of the primers to allow conjunctive use with longer arbitrary primers (protocol L) as well as to allow the unidirectional cloning of the DD products. These longer anchor primers can also be used with arbitrary 10-mers, as in protocol S.

Although the original DD protocol was intended to subdivide the transcripts into 12 fractions by reverse transcription with two-base anchored primers (1), later studies revealed that the secondmost 3′ nucleotides barely affect the base discrimination

(10). We also confirmed this and designed our primers to subdivide the transcripts into three or four subsets.

The concentration of anchor primer at the reverse transcription step can be lowered 5- to 10-fold in some samples without affecting the overall yield, whereas significant reduction was observed in others. The concentrations given are only a starting point.

Step 2: Polymerase Chain Reaction

Procedure

1. Combine the following solutions to make the PCR mix I. Prepare sufficient amounts, taking into account the loss during repeated pipetting.

Distilled water	$(13.7 \times N)$ μl
PCR buffer II (10×)	$(2.0 \times N)$ μl (Perkin-Elmer)
$MgCl_2$ (25 m*M*)	$(1.2 \times N)$ μl
dNTP (10 m*M*)	$(0.1 \times N)$ μl
cDNA solution	$(2.0 \times N)$ μl

2. Put 1.0 μl of arbitrary primers (10 pmol) in each well of a Techne HI-TEMP 96 plate (Techne, Cambridge, UK).
3. Mix the PCR mix I and AmpliTaq DNA polymerase gently but thoroughly (PCR mix II). Avoid forming bubbles.

PCR mix I	$(18.8 \times N)$ μl
AmpliTaq DNA polymerase	$(0.2 \times N)$ μl (1 unit)

4. Dispense 19.0 μl of PCR mix II into each well containing 1.0 μl of arbitrary primer solution.
5. Overlay each well with a drop of mineral oil.
6. Place the plate in a Techne PHC-3 thermal cycler, and run the following program.

Protocol S (short anchor primer and 10-mer)
3 min at 94° C, 5 min at 40° C, and 5 min at 72° C (1 cycle)
(for the second-strand synthesis)
15 sec at 95° C, 2 min at 40° C, and 1 min at 72° C (20–24 cycles)
(for the amplification)
5 min at 72° C
(for the final extension)

Protocol L (long anchor and arbitrary primers of ~20-mer)
3 min at 94° C, 5 min at 37° C, and 5 min at 72° C (1 cycle)
(for the second-strand synthesis)

15 sec at 95° C, 1 min at 55° C, and 2 min at 72° C (20–24 cycles)
(for the amplification)
5 min at 72° C
(for the final extension)

Comments

We routinely use cDNA equivalent to 50 ng of total RNA, but one can further reduce the amount of cDNA (for instance, ~10 ng) in case the RNA source is limited.

The Mg^{2+} concentration can be varied. In our experience, any single condition cannot amplify every band at its optimum. We must compromise and usually use the most standard Mg^{2+} concentration (1.5 m*M*) for PCR. The concentration of dNTP can be changed in our protocol, unlike the original DD where the labeling is performed by the incorporation of labeled dNTP. Even without the addition of dNTPs at the PCR step, those carried over from the reverse transcription reaction can support the cDNA band generation, albeit less efficiently. We usually add additional dNTP (~50 μM each). In some samples, addition of other dNTPs (200 μM each) significantly improved the signal intensities, especially of longer products. The number of thermal cycles should be empirically determined in pilot studies. In general, 20–24 cycles gave satisfactory results.

Variation due to thermal cycler makeup does exist, so that for comparison studies the same machine and model should be used. The run-to-run variation can be minimized by faithfully repeating the experiments in our procedure. We confirmed that 95% (or more) of the bands were amplified reproducibly when the experiments are carefully repeated. However, in the case of pursuing subtle changes, we strongly recommend testing the reproducibility using different batches of RNA samples or, at least, different batches of cDNAs generated from the same RNA samples.

In principle, the so-called hot start method is desirable in DD. However, manual hot start of 96 samples is impractical. Our choice is the use of antibody to *Taq* DNA polymerase (Clontech, Palo Alto, CA). However, in our experience, the fingerprinting pattern was barely affected by the use of the hot start technique.

Step 3: Analytical Run on Hitachi SQ-3000 DNA Sequencer

Procedure

1. Degas 15 ml of gel solution *in vacuo* during the gel plate assembly.
2. Assemble the gel plates (17 × 20 cm) with 0.30 mm thick spacers.
3. Following the addition of 100 μl of 10% (w/v) ammonium persulfate and 10 μl of *N,N,N′,N′*-tetramethylethylenediamine (TEMED), pour the degassed gel solution between the glass plates. Allow the gel to polymerize for ~2 hr at room temperature. Make sample wells with a shark's tooth comb (42–54 wells) and attach the assembled gel to the DNA sequencer. Run at 500 V for ~30 min (prerun).

4. Mix 2 μl of PCR products with 2 μl of loading solution in a 0.5-ml tube. Heat the tube at 90° C for 2 min, quench on ice, and apply the sample (~2 μl) onto each well of the shark's tooth comb.
5. Run the gel at 800 V for 2–3 hr.
6. Search for the band of interest using the track file mode that displays the fluorogram.

Comments

We routinely use the Hitachi SQ-3000 DNA sequencer equipped with a half-length gel plate for the FDD screening. This machine can display a gel image as a fluorogram on the NEWS workstation (Sony, Tokyo, Japan) controlling the sequencer (see Fig. 2, for example). Visual inspection of such fluorograms was found to be much

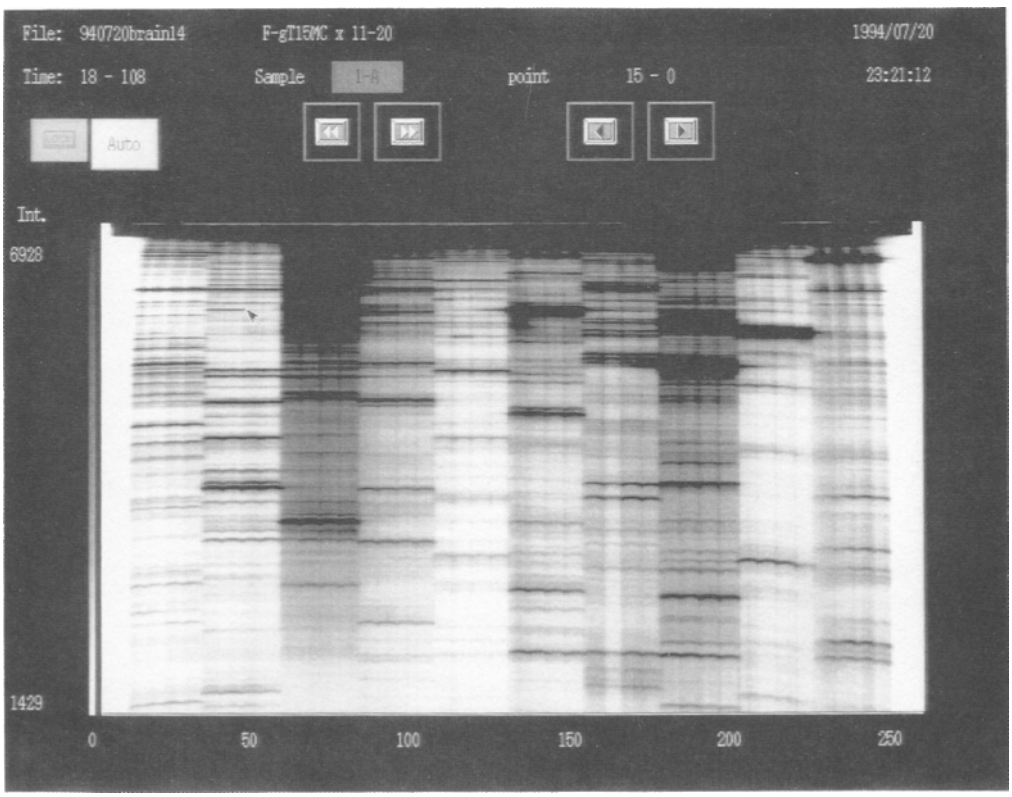


FIG. 2 Typical result of FDD. Brain RNAs from male and female mice were compared with FDD protocol S (from left to right, two males and two females). This picture is a screen dump from the Sony NEWS workstation controlling the Hitachi SQ-3000 DNA sequencer. The arrow indicates a male-specific band.

faster and more sensitive than the examination of chromatogram traces. Also, the detector on this machine uses any comb, thereby theoretically delimiting the number of lanes. However, loading of more than 48 samples seems to be impractical.

The Hitachi SQ-5500 sequencer [available as the Vistra 725 DNA sequencer from Molecular Dynamics (CA) and Amersham Life Science (UK)] can be used essentially in the same manner as the SQ-3000, except that Texas Red-labeled anchor primers must be prepared.

The use of the short gels drastically saves running time, while keeping the resolution required for DD analysis. Because the short running time minimizes damage to gel matrices, a second set of samples can be loaded onto the same gel immediately after the first run. Because the same gel is used twice and 48 samples are loaded per run, samples from two 96-well plates can be analyzed per day with 2 glass plates. Each lane contains 50–100 cDNAs, so that data on more than 10,000 cDNA bands can be obtained with minimum operator interaction.

The other advantage of using the DNA sequencer is its ability to scan a wider size range than conventional gels. With conventional gels, the run must be stopped at some point for signal detection: some short bands have already run out of the gel, whereas the longer ones are not well resolved. In contrast, the DNA sequencer can detect all of the bands migrating in front of the detector, ranging from the one running just behind the primers to the longest amplified products, so that bands of interest are not overlooked as would occur with conventional gels.

Step 4: Preparative Run and Band Excision with Vistra FluorImager SI

Procedure

1. Prepare a gel between 20 × 33 cm glass plates with an 0.35-mm-thick spacer.
2. Load the heat-denatured samples (5–6 μl) as described above.
3. After the run, remove the upper glass plate carefully and put the gel attached to the lower plate into the scanner.
4. Scan the gel using the "high sensitivity" mode.
5. Print the image by "actual size" mode. Overlay the gel onto the printout, and excise the band of interest with a razor blade. Following the excision, scan the gel again to confirm the precise excision of the band of interest.

Comments

When the analytical run has been done on the sequencer, the approximate size of the band of interest is already known, so that one can choose the gel concentration most suitable for the separation of the target band. In contrast, when the FluorImager 575 is used for the initial screening, as many bands as possible should be separated. The use of a buffer gradient system would be helpful for the scanning of a wider size

range than the conventional one. We use a 6% (w/v) gel in 1× TBE buffer run with 0.5× TBE buffer and TBE–sodium acetate mix (1× TBE: 3 *M* sodium acetate, 2:1 (v/v)) as upper and lower buffers, respectively.

When the signals are too weak for the FluorImager, overlay SYBR Green I (Molecular Probes, Eugene, OR) solution (1:10,000 dilution in TE) onto the gel (~5 ml) for signal intensification. Note that overlaying too much fluid onto thin gels for a prolonged time will cause the diffusion of the bands to disturb the imaging. Keep the incubation time and volume of the fluid at a minimum. A few minutes of incubation is usually enough for a 0.35-mm-thick gel. If using the SYBR Green I overlay technique, do not prepare FITC-labeled primers. This technique is well suited to the use of a nondenaturing gel, which is recommended by some authors to reduce the artificial complexity of DD patterns on denaturing gels (12).

Another commercially available fluorescent gel imager is the FMBIO-100 (TaKaRa, Japan) and this machine can also be used for FDD analysis with rhodamine X (ROX)-labeled primers. Using the FMBIO-100 the gel image is obtained without removing the upper glass plate, so that the gel can be scanned at different gel running times to cover wider size ranges per single run.

For preparative gel electrophoresis, a native polyacrylamide gel can be used, followed by SYBR Green I staining and visualization by ultraviolet (UV) illumination. According to the manufacturer, less than 20 pg of double-stranded DNA can be detected by SYBR Green I staining with UV (254-nm) epiilumination. Load ~10 μl of DD reaction. Excise the band of interest or the region that should contain the faint band of interest. Note that a yellow gelatin filter (Eastman Kodak, Rochester, NY) is necessary for photographing the stained gels. This simple staining method provides a simple, albeit less sensitive, alternative detection method that requires no special equipment.

Step 5: Reamplification and Cloning of Excised Band

1. Rinse the excised gel piece with distilled water.
2. Put half of the gel piece into a PCR tube. Cut the piece into several smaller pieces by using a flat loading tip as a blade.
3. Add 50–100 μl of the PCR reaction mix described above except for the lower concentration (0.25 μM each) of arbitrary and anchor primers. The thermal cycling profiles are as follows:

Protocol S	
3 min at 94°C	1 cycle
30 sec at 94°C, 2 min at 40°C, and 1 min at 72°C	20 cycles
5 min at 72°C	1 cycle

Protocol L

3 min at 94°C	1 cycle
30 sec at 94°C, 1 min at 55°C, and ~2 min at 72°C	20 cycles
5 min at 72°C	1 cycle

4. A portion ($\frac{1}{10}$) is run on a polyacrylamide gel to confirm the amount and purity of the reamplification products.
5. Following the ethanol precipitation, clone the band into T-vector according to the conventional method (7).

Comments

We found that elution steps, such as boiling or electroelution, from the gel pieces were unnecessary. Direct addition of the gel piece into the PCR reaction usually gave satisfactory results. If the reamplified fragment is sufficiently pure, the sequence information can be obtained by direct cycle sequencing using the anchor primer as the sequencing primer.

Step 6: Selection of Correct Clones

Procedure

1. To perform "colony PCR," select the colonies of subclones using toothpicks. The conditions are same as for the reamplification step from the gel piece, except that the number of cycles can be reduced to 20 cycles.
2. Load 0.1 – 1 μl of colony PCR product onto the DNA sequencer and run in parallel with the original FDD reaction (comigration test). Select the clones (candidate clones) bearing the inserts that comigrated precisely with the band of interest.
3. Determine the nucleotide sequences of the candidate clones.
4. Search for appropriate restriction enzyme sites in the sequences of candidate clones, and digest the original FDD reaction and the amplified inserts of the candidate clones with the enzyme. Run the digested products on the DNA sequencer to confirm that the digestion pattern of the band of interest is similar to that of candidate clones (restriction test).
5. Confirm the expression pattern of the corresponding transcripts by Northern blot hybridization, RNase protection assay, or RT-PCR assay.

Comments

Although the excised bands are occasionally pure, they are often contaminated with neighboring bands and sometimes contain cDNAs that comigrate with but are distinct from the target species. Therefore, we usually avoid using the reamplified products directly for further analysis. We recommend purification by molecular cloning.

For correct clone selection various approaches can be used, including differential techniques or Southern hybridization (13). We prefer the approach based on DNA

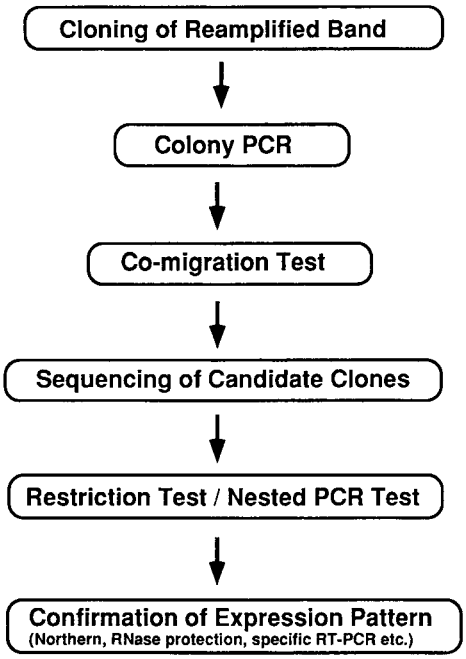


FIG. 3 Steps for the cloning of the band of interest. See text for details.

sequencing as described above and summarized in Fig. 3, because it is straightforward and convincing. The restriction test is the simplest assay to perform. If suitable enzyme sites are not found in the sequence of candidate clones, synthesize primers to perform a simple PCR-based test. In the nested PCR test, diluted FDD products (or FDD with much reduced thermal cycling) are used as template for the PCR to confirm that the cDNA band cloned has a differential expression pattern. The same primers can be used for the confirmation of the expression pattern of the corresponding transcript by specific RT-PCR, where cDNA generated with oligo(dT) and/or random primers is used as the template. These primers can also be used in the PCR-based isolation of 5′ portions or occasionally 3′ portions of the transcripts. (Note that the band obtained by DD is not always derived from the 3′ end of the corresponding transcript. We have so far experienced several such cases, and assume that priming from internal A-rich regions is not so rare.)

In general, the clone most frequently recovered is the correct target. However, some cDNAs are refractory to cloning and are considerably underrepresented in the subclone populations. In such cases, direct cycle sequencing of the reamplified fragments should be tried in order to obtain the sequence of the major species in the reamplified products. This often helps in identifying the correct clone, even when it is considerably underrepresented.

Owing to the principle of DD, cDNA fragments obtained are largely derived from the 3′-untranslated regions. As this portion is generally AT rich, some short fragments work poorly as hybridization probes. Isolation of the upstream portion often solves such problems.

Fluorescent RNA Fingerprinting by Arbitrarily Primed Polymerase Chain Reaction

Step 1: Reverse Transcription

Procedure

1. Mix:

DEPC-treated water	8.2 μl
Arbitrary primer (50 μM)	0.8 μl
Total RNA	1.0 μl (50–100 ng)

2. Incubate the tube at 70° C for 5 to 10 min.
3. During the incubation, prepare the following solution (2× RT solution):

DEPC-treated water	(4.8 × N) μl
RT buffer (10×)	(2.0 × N) μl
DTT (100 mM)	(2.0 × N) μl
dNTP (10 mM)	(0.2 × N) μl
SuperScriptII (200 U/μl)	(1.0 × N) μl

4. Put the heated tube containing RNA and primer into an ice–water bath. Spin the tube briefly.
5. Add the 10 μl of 2× RT solution to the tube and mix well.
6. Put the tubes in the thermal cycler programmed for 25° C for 10 min, 42° C for 50 min, and 90° C for 10 min.

Step 2: Polymerase Chain Reaction

Procedure

1. Add the following PCR mix (20 μl) to each tube containing the cDNA:

Distilled water	(15.5 × N) μl
PCR buffer II (10×)	(2.0 × N) μl (Perkin-Elmer)
$MgCl_2$ (25 mM)	(2.0 × N) μl
AmpliTaq DNA polymerase	(0.5 × N) μl

2. Overlay with mineral oil.
3. Run the PCR according to the following program.

3 min at 94° C, 5 min at 37° C, and 5 min at 72° C (1 cycle)
(for the second-strand synthesis)
30 sec at 95° C, 1 min at 55° C, and 2 min at 72° C (~25 cycles)
(for the amplification)
5 min at 72° C
(for the final extension)

Step 3: Electrophoresis, Band Excision, and Cloning

These steps are performed as described above for FDD.

Comments

RAP-PCR can be performed without any significant modification of the original protocol (2). (However, some primers work poorly in RAP-PCR.) The concentration of primer, Mg^{2+}, and dNTP can be varied, and the number of thermal cycles should be empirically determined in each case. It is possible to use two arbitrary primers sequentially to enlarge the subsets of cDNA to be surveyed (14) or "motif" primers to bias the sampled population to members of particular gene families (15). If arbitrary primers are suspected to be contaminated with RNase, avoid incubation of primers with RNAs. Instead, include the primers in the 2× RT solution with RNase inhibitor.

The fluorescent labeling by incorporation of fluorescently labeled dNTPs during PCR did not provide us with satisfactory results. We recommend using fluorescently labeled primers for the fluorescent RAP-PCR on the DNA sequencer. An alternative approach is the use of nonfluorescent primers and detection by SYBR Green I staining as described above.

Conclusion

The FDD protocols described here have been used successfully in our laboratory to clone ~40 bands of differential expression patterns from various biological sources. Subsequent analyses (Northern blot hybridization or specific RT-PCR assay) using the clones proved that the expression pattern of the corresponding transcripts had been faithfully, albeit semiquantitatively, reflected on FDD fingerprints in all the cases. Although some literature pointed out the irreproducibility and the high incidence of false-positive signals with the original DD procedure (16–18), our protocol has so far brought satisfactory results as noted above. We attribute this to the reliable protocol developed using the novel anchor primers under more standard PCR condi-

tions, as well as the careful clone examination steps. The procedure described here will thus provide useful techniques for those performing DD analysis using both radioactive and nonradioactive detection systems.

The FDD with the use of a modified DNA sequencer was developed to establish a high throughput transcript scanning system. In our current system, we can obtain data on more than ~10,000 cDNAs per day (see above) with minimum operator interaction. Although another procedure for DD on a DNA sequencer has been published, it requires not only two-step PCR but also the preparation of fluorescently labeled 10-mers (12). The unsurpassed speed of our FDD system will make "saturation transcript scanning" feasible, and accelerate the identification of differentially expressed transcripts to address various biological questions.

Acknowledgments

We are grateful to K. Kito for contributions to the establishment and extensive testing of protocol L, as well as various useful discussions. We also thank H. Hagiwara and T. Muramatsu for the modification of the Hitachi DNA sequencers. This work is partly supported by grants from the Ministry of Education, Science and Culture (MESC) and the Science and Technology Agency (STA), Japan.

References

1. P. Liang and A. B. Pardee, *Science* **257,** 967 (1992).
2. J. Welsh, K. Chada, S. S. Dalal, R. Cheng, D. Ralph, and M. McClelland, *Nucleic Acids Res.* **20,** 4965 (1992).
3. J. Welsh and M. McClelland, *Nucleic Acids Res.* **18,** 7213 (1990).
4. J. G. K. Williams, A. R. Kubelik, K. J. Livak, J. A. Rafalaski, and S. V. Tingey, *Nucleic Acids Res.* **18,** 6531 (1990).
5. S. M. Trentmann, E. van der Knaap, and H. Kende, *Science* **267,** 1186 (1995).
6. T. Ito, K. Kito, N. Adati, Y. Mitsui, H. Hagiwara, and Y. Sakaki, *FEBS Lett.* **351,** 231 (1994).
7. J. Sambrook, E. F. Fritsch, and T. Maniatis, "Molecular Cloning: A Laboratory Manual," 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
8. P. Chomczynski, *BioTechniques* **15,** 532 (1993).
9. N. Adati, T. Ito, C. Koga, K. Kito, Y. Sakaki, and K. Shiokawa, *Biochim. Biophys. Acta* **1262,** 43 (1995).
10. P. Liang, L. Averboukh, and A. B. Pardee, *Nucleic Acids Res.* **21,** 3269 (1993).
11. P. Liang, W. Zhu, X. Zhang, Z. Guo, R. P. O'Connell, L. Averboukh, F. Wang, and A. B. Pardee, *Nucleic Acids Res.* **22,** 5763 (1994).
12. D. Bauer, H. Müller, J. Reich, H. Reidel, V. Ahrenkiel, P. Warthoe, and M. Strauss, *Nucleic Acids Res.* **21,** 4272 (1992).

13. K. K. Wong and M. McClelland, *Proc. Natl. Acad. Sci. U.S.A.* **91,** 639 (1994).
14. M. McClelland, D. Ralph, R. Chang, and J. Welsh, *Nucleic Acids Res.* **22,** 4361 (1994).
15. B. Stone and W. Wharton, *Nucleic Acids Res.* **22,** 2612 (1994).
16. Y. Nishio, L. P. Aiello, and G. L. King, *FASEB J.* **8,** 103 (1994).
17. L. Mou, H. Miller, J. Li, E. Wang, and L. Chalifour, *Biochem. Biophys. Res. Commun.* **199,** 564 (1994).
18. Y. Sun, G. Hegemyer, and N. H. Colburn, *Cancer Res.* **54,** 1139 (1994).

[13] ^{32}P- and Fluorescence-Labeled DNA Sequencing Using Primers Selected from Nonamer Library

Jeffrey H. Bock and Jerry L. Slightom

Introduction

Efforts are underway to obtain the complete genome sequence of several model organisms, such as *Escherichia coli,* yeast, *Caenorhabditis elegans, Drosophila melanogaster,* and also from far more complex organisms, such as mouse and human (1–6). These sequencing goals are extremely ambitious considering that the smallest of these genomes contains about 5×10^6 base pairs (bp) and that a mammalian haploid genome contains about 3×10^9 bp. Present nucleotide sequencing technologies may not be adequate to achieve these goals, thus there is a great need and interest in the development of improved DNA sequencing technologies and strategies.

Current sequencing strategies include the straightforward primer walking (7, 8), random M13 phage "shotgun" (9), and bacterial transposon tagging (10, 11). The major strength of the M13 phage shotgun and transposon strategies is that they facilitate high-throughput sequencing using automated instrumentation. However, major weaknesses of these strategies are the need to generate a large number of subclones and to sequence many of them to ensure five- to eightfold sequence coverage (6, 12, 13). The primer-walking strategy is potentially more efficient because it sequentially extends known sequences (obtained from flanking vector sequences or from previous walking steps) to complete the sequence of a cloned insert, thus greatly reducing the number of subclones and amount of redundant sequencing. Theoretically, complete double-stranded coverage could be achieved with as little as a 2.5-fold redundancy. Inserts as large as 40 kilobases (kb) have been completely sequenced using the primer-walking strategy (7, 8, 14). The major disadvantage of the primer-walking strategy is its inherent low throughput due to time delays needed to obtain new sequence information (film exposures and sequence readings) for the selection of new walking primers and for primer synthesis. In addition, primers represent a considerable expense because only a small fraction (>0.001%) of the synthesized primer is used.

Despite these disadvantages, there is still considerable interest in developing a high-throughput primer-based DNA sequencing strategy. Such a primer-based sequencing strategy would be based on the use of a presynthesized library of primers, so as to eliminate the delay and waste problems. Presently, the synthesis of a library of standard size (20-mer) primers is impractical because of the enormous number of primers required for the library (4^{20}). A potential solution to this problem would be to use short primers for library construction (15), in the hexamer-to-decamer size

Methods in Molecular Genetics, Volume 8

range, because these complete short primer libraries would require fewer members (4^6 to 4^{10} members, respectively). In such a primer library-based sequencing strategy, the high-throughput sequencing capability would be derived from the primer library (15) instead of from an M13 subclone library. Redundant sequencing could be minimized with the primer library-based strategy because it still involves the sequential sequencing from known into unknown DNA regions, which allows for operator interaction.

The short primer library strategy described by Studier (15) has subsequently been refined to include the use of ligated or nonligated pentamer/hexamer strings (16–18) and a reduced library of selected nonamers (19). Both pentamer/hexamer string and reduced library strategies have their potential strengths and weaknesses. The pentamer/hexamer library strategy has the advantage of being complete and should be able to supply all the primers needed for any project, while their most notable weakness (ligated or nonligated) is the additional steps required for selection and stable alignment of multiple short primers into a continuous string (17, 18). The nonamer library strategy has the advantage of requiring the selection of only one primer; however, the major disadvantage is that the library may not contain all the primers needed for a sequencing project. Both strategies, however, suffer from the major trade-off of using short primers: a reduction in the number of successful primed reactions. Thus far, even after the development of modified DNA sequencing procedures, only marginal success rates (about 50%) have been achieved (17, 18, 20, 21). Our efforts have been focused on the development of procedures for using the nonamer library described by Siemieniak and Slightom (19). We present here both the ^{32}P label- and fluorescence-based methodologies that we have developed for nonamer primer sequencing of double-stranded plasmid DNAs, and we demonstrate that a reduced nonamer library consisting of only 2391 members can be of considerable value in aiding the completion of many sequencing projects.

Methods

General Reagents and Equipment

^{32}P Label

T7 polymerase (Sequenase version 2) and Sequenase sequencing kits are purchased from U.S. Biochemical (Cleveland, OH). Deoxynucleotides (dNTPs) and dideoxynucleotides (ddNTPs) are obtained from Pharmacia (Piscataway, NJ) and [α-^{32}P]dATP and [α-^{32}P]dCTP (specific activities > 3000 Ci/mmol, 10 mCi/ml) are purchased from Amersham (Arlington Heights, IL). X-Ray roll film (20 cm × 25 m, XAR-351) is purchased from Kodak (Rochester, NY), and the DNA sequencing gel stands and safety cabinets used have previously been described (22).

Fluorescence Label

The ABI PRISM Ready Dye-Deoxy Terminator kits for *Taq* DNA polymerase are purchased from Perkin-Elmer/Applied Biosystems Division (PE/ABI, Foster City,

CA) and used in conjunction with the ABI373A fluorescence-based sequencer (software version 1.2). The sequencing gel used consists of 6% (w/v) polyacrylamide and 7.5 *M* urea (as described by ABI), which is poured between special optically pure glass plates to obtain a gel measuring 0.4 mm thick, 254 mm wide, and 406 mm long. Samples are loaded using the 24-well shark's tooth comb format.

Oligonucleotide Library

The 2391-member nonamer library of deoxyribonucleotide primers is synthesized at Genosys Biotechnologies, Inc. (The Woodlands, TX) with specially designed multichannel synthesis instruments using phosphoramidite chemistry (23). The synthesized primers are deprotected in NH_4OH at 55°C and the NH_4OH is removed by evaporation. Dried oligomer primers are resuspended in sterile double-distilled H_2O at a final concentration of 10 μg/ml. Each member of the library is dispensed into 1 well of a 96-well microtiter plate and its position recorded. The 2391-member nonamer library is cataloged in a total of 25 microtiter plates and the location of each nonamer primer is noted on a data sheet (Microsoft EXCEL database format). The locations of selected primers are identified using the EXCEL database search routine. After locating a selected nonamer, it is removed from the microtiter plate and used without any further purification or analysis.

Plasmid Subclone and DNA Preparation

Plasmid pCh32-2.3-7E contains a 7.0-kb *Eco*RI fragment from the human $\psi\eta$-globin gene region (24) cloned into pBR322. Plasmid DNAs are isolated and purified using Qiagen (Chatsworth, CA) protocols (Plasmid Midi kit). Qiagen-purified DNA is of sufficient purity to be used for both the T7 polymerase-based manual and *Taq* polymerase-based ABI cycle sequencing methods.

Computer-Aided Nonamer Mapping and DNA Sequence Analyses

The location of a nonamer primer within a known nucleotide sequence region is determined using the MAP program available in GCG (Genetics Computer Group, Madison, WI) software package. This is accomplished by creating a local data map program file that contains each sequence of the 2391 nonamers. The output of the MAP program lists the number of each nonamer at the exact match position, similar to that done for restriction enzyme site matches.

^{32}P-Labeled Sequencing with Primers Selected from Nonamer Library

Method

The major problem associated with using short oligomer primers is directly related to their size: they have relatively low annealing temperatures, and their annealing T_m values vary widely with their G+C composition (T_m values can range between 18 to 27°C for nonamers). However, even the highest expected nonamer T_m for a nonamer

is considerably lower than that calculated for an average 20-mer (50°C). The nonamer T_m range is reduced to 22 to 24°C in the nonamer library described by Siemieniak and Slightom (19) by restricting the G+C compositions to between 45 to 60%. DNA sequencing reactions are done using the dideoxy chain termination (25) method, using a modified T7 polymerase (Sequenase) on double-stranded plasmid DNAs. About 4 μg of plasmid DNA is subjected to dideoxy sequencing reactions that included both [α-^{32}P]dATP and [α-^{32}P]dCTP for radioactive labeling, as described by Siemieniak *et al.* (7). The ratio of ddNTP:dNTP is reduced to 1:30, which allows many of these sequence reactions to be read beyond 600 bp from the primer. A slightly modified Sequenase reaction procedure is used (Table I), in that

TABLE I Manual ^{32}P-Labeled Sequencing Method for Nonamer Primers[a]

I. Denature template and anneal nonamer primer to template.
   A. Add *X* μl of double-distilled sterile H_2O to *X* μl of template DNA (3–4 μg of DNA) to obtain a final volume of 32 μl. Then add 8 μl of freshly prepared 1 *M* NaOH.
   B. Incubate at 22°C for 5 min.
   C. Add 4 μl of 5 *M* NH_4OAc, mix, and add 100 μl of 100% ethanol.
   D. Incubate at −70°C for 10 min. Pellet the DNA by spinning in a microcentrifuge at 10,000 rpm for 15 min.
   E. Wash the DNA pellet once with 70% (v/v) ethanol. Centrifuge for 2 min.
   F. Dry the DNA in a Savant Speed-Vac.
   G. Resuspend the DNA in *X* μl of double-distilled H_2O, 2 μl of 5× reaction buffer, and *X* μl of the nonamer primer (10 to 20 ng) to obtain a final volume of 10 μl.
   H. Anneal the nonamer primer and denatured template DNA for 10 min at 37°C. Place the tubes on ice (~5 min) while preparing labeling tubes (see step III).

II. Aliquot the appropriate termination mixes into tubes labeled G, A, T, and C. The following termination mixes should be aliquoted:

    G mix: 2.5 μl of G termination mix + 2.5 μl of extension mix (180 μl of each dNTP)
    A mix: 3.0 μl of A termination mix + 3.0 μl of extension mix (180 μl of each dNTP)
    T mix: 3.0 μl of T termination mix + 3.0 μl of extension mix (180 μl of each dNTP)
    C mix: 2.5 μl of C termination mix + 2.5 μl of extension mix (180 μl of each dNTP)

III. Aliquot the following reagents to the annealing reaction tubes (from step II). Volumes represent one reaction tube:

    1 μl of 0.1 *M* DTT, 2 μl of 1× labeling mix, 1 μl of [α-^{32}P]dATP, 1 μl of [α-^{32}P]dCTP, 2 μl of dilute Sequenase (dilute enzyme 1:8 into cold enzyme dilution buffer—3.25 U final).

IV. Incubate the annealing tubes at 22°C for 5 min. During this incubation step, place the termination tubes (from step II) in a 37°C heat block for approximately 1 min.

V. Transfer 3.5 μl of the reaction mixture (step IV) to each termination tube (step II), mix, and incubate at 37°C for 10 min.

VI. Stop the reaction by drying the reactions (step V) in a Savant Speed-Vac. Resuspend in 10 μl of loading buffer [80% (v/v) formamide, 10 m*M* NaOH, 1.0 m*M* EDTA, and 0.1% (v/w) xylene cyanol].

VII. Heat samples to 90°C for 3 min, cool on ice, and load on gel.

[a] Termination mixes, reaction buffer, and enzyme dilution buffer are from the Sequenase kit.

the primer-template is first annealed at 37°C for 10 min followed by annealing at 0°C for an additional 5 min. Obviously, most of the nonamer-template annealing occurs during the 0°C incubation step. The higher 37°C template-primer annealing step (Table I, step I,H) can be skipped without adversely affecting the quality of the sequencing ladder. After the reaction termination step (Table I, step VI) the sequenced samples are dried in a Savant Speed-Vac (Farmingdale, NY), followed by resuspension in 10 μl of sequence gel loading buffer (Table I, step VI), heating to 90°C, and then loading onto a sequencing gel. We routinely use 1-m-long gels (22) to maximize sequence readings.

Example of ^{32}P-Labeled Nonamer-Primer Sequencing Method

A total of 59 one-hit nonamers were found to match the sequence of the 7.0-kb human $\psi\eta$-globin gene region (Fig. 1A). From these nonamers, 27 were selected and 16 (59%) produced readable sequence ladders. These successful reactions yielded 90% single-stranded and 67% double-stranded sequence coverage (Fig. 1A). A total of 10,558 bp of sequence information was accumulated within the human $\psi\eta$-globin gene region and the average reading length was 754 bp for the 14 reactions that were read to their full length (see Table II). The human $\psi\eta$-globin gene region contains only one *Alu* element (between positions 1060 and 1359) and no primer matches were found in this region (Fig. 1). *Alu* element nonamer matches were removed from the library prior to synthesis (19). This *Alu* element was not a problem, except that the nonamer primers positioned to sequence across it failed, which resulted in the only substantial single-stranded coverage gap (Fig. 1A). We estimate that the sequence of this insert could be completed with the use of only eight additional custom oligomers. This represents a potential 70% savings in the number of custom oligomers needed to sequence this 7.0-kb insert.

Fluorescence-Based DNA Sequencing with Primers Selected from Nonamer Library

Background

In the ^{32}P-labeled DNA sequencing procedure described above, T7 polymerase was used for DNA synthesis. Because of this success, we expected a T7 polymerase-based procedure to work best for reactions run on the fluorescence-based ABI373A sequencer. The alternative would be to use cycle sequencing, which uses *Taq* polymerase and a DNA extension temperature (60°C). We expected that this high extension temperature would be incompatible for use with nonamer primers. Thus, nonamers were first used with the ABI PRISM Ready Reaction Dye-Deoxy Terminator kit for T7 DNA polymerase. Although good results were generated for double-stranded sequence reactions using standard oligomers (21 bases in length), no sequence information was generated when nonamers were used (21). This lack of suc-

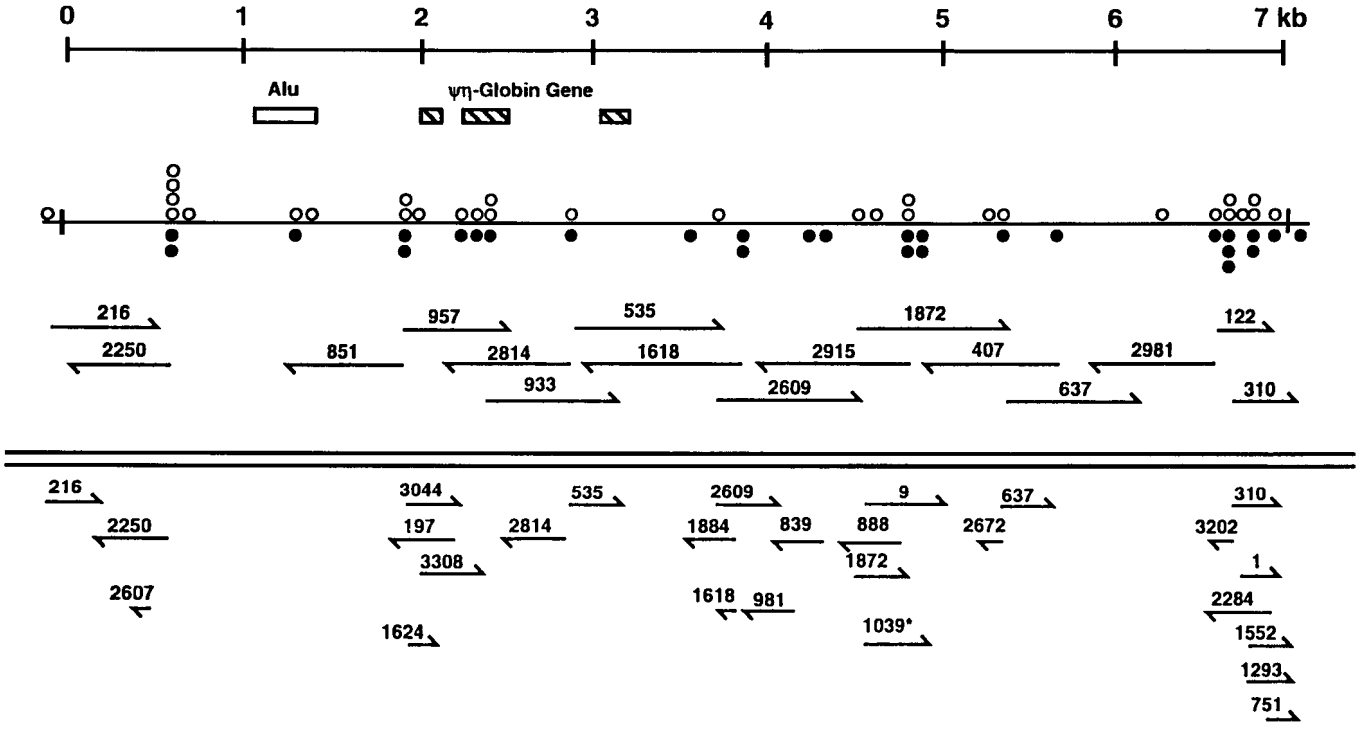


FIG. 1 Physical map of the human clone pCh32-2.3-7E and results from manual and ABI373A nonamer-primed sequencing experiments. The physical map of clone pCh32-2.3-7E shows the location of the human $\psi\eta$-globin gene within the cloned insert. A total of 59 nonamers was mapped and these are represented as open circles (○) and closed circles (●) for nonamer-binding sites on the top and bottom DNA strands, respectively. The arrows (→) and (←) represent the locations of nonamers that primed sequence reactions (top and bottom DNA strands, respectively) and the lengths of the arrows represent the sequence reading lengths. The number above each arrow corresponds to the nonamer library location. (A) Placement and coverage for nonamer-primed sequence reactions using the manual procedure. (B) Placement and coverage of successful nonamer-primed sequencing reactions using the modified ABI cycle sequencing procedure. Note that the primer number followed by the asterisk (*) primed at a location different than expected.

TABLE II Nonamer Test Results Obtained for Individual Nonamers Using Manual or ABI Sequencing Methods with Template DNA from Clone pCh32-2.3-7E[a]

Oligo-nucleotide number	Sequenase (bp)	ABI (bp)	Oligo-nucleotide number	Sequenase (bp)	ABI (bp)
1039		370**	1872	890	300
1487		—	9		410
776		—	12		—
2250	567	400	342		—
1681		—	3309		—
2607	Weak	30	888		370
1049	Weak	—	2915	874	—
1059	—	—	1626		—
2324	Multiprime	—	637	768	280
1910	Weak	—	2672		100
957	621	—	407	802	—
3044		300	786	Multiprime	—
2723		—	122	316	—
851	670	—	2981	646	—
1624		160	310	254	270
3308		390	2698	**	—
1906		—	3124		—
460	Multiprime	—	1536		—
197		370	3202		50
933	791	—	1		230
2960		—	1552		170
2992		—	1293		150
535	867	310	2622		50
2814	717	370	1119		—
1978	—	—	751		80
2609	842	350	2284		390
1884		300			
1618	936	30			
981	—	250	216	567 (from vec)	280
839		290	724	**	—

[a]Nonamer reference numbers are listed in the oligonucleotide column and the numbers in the Sequenase and ABI columns represent the reading lengths obtained from either the film or generated by the ABI373A sequencer, respectively. A dash (—) indicates that the nonamer tested did not yield any usable sequence information; a double asterisk (**) indicates that a nonamer primed at a location different than that expected.

cess appeared to be because little, if any, DNA synthesis was being initiated from the nonamer primers. For manual sequencing this low amount of new DNA strand synthesis was compensated for by using two ^{32}P-labeled nucleotides (Table I) to enhance detection sensitivity. It appears that in the case of nonamer-primed reactions, success on the ABI373A sequencer will require a much more sensitive detection system.

Because of the lack of success with the ABI PRISM T7 polymerase kit, we switched to the *Taq* polymerase cycle sequencing procedure.

Cycle Sequencing with Nonamer Primers

The standard ABI cycling sequencing conditions for the PRISM Ready Reaction Dye-Deoxy Terminator Cycle Sequencing kit using *Taq* polymerase could not be used with nonamer-size primers without some modifications. The initial cycle sequence reaction did yield some signals, and obviously, the reason for these poor results was due to the fact that the standard ABI cycle sequencing procedure uses annealing (50°C) and extension (60°C) temperatures that are well above the nonamer T_m values. Resolution of this problem meant that lower temperatures would have to be used for at least the annealing step and possibly for the extension step. However, *Taq* polymerase activity is low at temperatures less than 60°C; its maximum activity is at 72°C (26). Thus, we anticipated that if an annealing temperature near the nonamer's T_m value of 22 to 24°C is used, DNA synthesis by *Taq* polymerase should not be expected. If no, or very little, DNA strand synthesis occurs at this low annealing temperature we would expect the nonamer to rapidly melt away from the template during the switch to the 60°C extension temperature. This problem could potentially be circumvented by using a low annealing temperature followed by a slow ramp to the 60°C extension temperature. We reasoned that during the time required for the slow ramp to 60°C, *Taq* polymerase would have a chance to initiate DNA synthesis (at some increased temperature reached during this ramp) from a nonamer before it melted away from the template DNA. Limited DNA synthesis by *Taq* polymerase of as little as 11 nucleotides would provide a 20-mer, which should have an average T_m of about 50°C. Thus, with the use of these (and other) modified cycle sequencing steps we were able to develop a nonamer-primed cycle sequencing reaction procedure for the ABI373A sequencer. An example of an ABI nonamer-generated sequencing profile is presented in Fig. 2. However, it should be noted that a considerable amount of experimentation was needed (21) before the modified cycle sequencing method described below was developed.

Method

Linear plasmid template for 20-cycle sequencing reactions is prepared (in this case) by digesting 40 μg of plasmid DNA (pCh32-12.3-7E) with *Hin*dIII (New England Biolabs, Beverly, MA), followed by incubation at 65°C for 10 min, phenol and chloroform extraction, and ethanol precipitation. The linear plasmid DNA is then resolubilized in 100 μl of sterile double-distilled H_2O to obtain a final DNA concentration of about 0.4 μg/μl. Before adding to the cycle sequencing reaction mix, the linear DNA is alkaline denatured by adding 25 μl of freshly prepared 1 *M* NaOH (0.2 *M* final NaOH), incubated at 22°C for 5 min, neutralized by adding 13 μl of 5 *M* ammonium acetate (0.47 *M* final ammonium acetate), and then ethanol precipitated. The DNA pellet is washed twice with 70% (v/v) ethanol and allowed to air dry. The linear

and alkaline-treated DNA is then resolubilized in 80 μl of sterile double-distilled H_2O (0.5 μg/μl) and 4 μl (2 μg) is added to each 2× (40 μl) cycle sequence reaction that contained a final concentration of 2% (v/v) dimethyl sulfoxide (DMSO).

Each cycle sequencing reaction is primed with between 8 and 16 pmol (20 to 40 ng) of a nonamer primer. The temperature cycles are controlled using an automated PCR thermal cycler 9600 (Perkin-Elmer, Norwalk, CT). The best modified cycle sequencing procedure developed, thus far, for nonamers uses the following steps: an initial denaturation at 98°C for 1 min, followed by 50 cycles in which the samples are denatured at 96°C for 30 sec, annealed at 20°C for 5 min followed by a 5-min ramp to the extension temperature, and extended at 60°C for 4 min. After completion of the modified cycle sequencing steps, which required 16 hr, the samples are held at 4°C. Extension products are purified using 1-ml spin columns (Nu-Clean D50; IBI, New Haven, CT). Each reaction product is loaded by pipette onto the column, which is then centrifuged in a swinging bucket centrifuge (Sorvall model RT6000B tabletop centrifuge) at 750 *g* for 2 min at room temperature. Column-purified samples are dried under vacuum for about 40 min and then dissolved in 5 μl of a DNA loading solution [83% (v/v) deionized formamide, 8.3 m*M* EDTA, and Blue dextran (1.6 mg/ml)]. The samples are then heated to 90°C for 3 min and loaded into the gel sample wells for sequence analysis by the ABI373A sequencer.

Example

As mentioned above, 59 nonamers map to the 7.0-kb insert of clone pCh32-2.3-7E (Fig. 1), and all of these nonamers were used to prime cycle sequencing reactions for the ABI373A. Twenty-eight of these nonamers primed a successful sequence reaction, which are summarized in Fig. 1, frame B, and Table II. Fig. 1, frame B, offers a more visual comparison of the sequence coverage obtained by the manual and ABI373A sequencing methods. From this comparison, it is obvious that coverage by the ABI373A was much less than that obtained by the manual method, despite the fact that almost twice as many nonamer primers were used with the ABI373A sequencer. Two reasons for this lack of coverage are apparent from the results shown in Fig. 1: first, many of the primers are from the same location on the cloned insert, and, second, much shorter sequence readings were obtained on the ABI373A sequencer versus that obtained manually (average reading lengths were about 257 bp on the ABI373A versus 754 bp obtained manually). Single-stranded sequence coverage was about 60% for the ABI373A results, while nearly 90% coverage was obtained by the manual method.

Conclusions

The results presented here clearly demonstrate that nonamers can be used to prime nucleotide sequencing reactions using double-stranded plasmids as large as 11 kb

(vector plus insert). The fact that the ^{32}P-labeled manual method achieved a 60% success rate while using only a slightly modified T7 polymerase reaction procedure is an encouraging result. Improvement of this success rate may be possible, as we need to test more reaction conditions and try to determine why some reactions fail. The most plausible reason for primer failure is intramolecular base pairing resulting in the formation of strong stem–loop structures. Secondary structure formation has been shown by Kieleczawa *et al.* (17) to interfere with sequence reactions that use strings of contiguous hexamers. We have also developed a cycle sequencing procedures for using nonamer with the ABI PRISM Ready Dye-Deoxy Terminator Cycle Sequencing kit. These cycle sequencing procedures differ considerably from the original ABI373A cycle sequencing procedure (see above). The use of linear and predenatured plasmid DNAs with a 2× reaction mix must be considered unusual steps for a cycle sequencing procedure. However, we found their combined use to be most effective in increasing the success rate and quality of nonamer-primed cycle sequence reactions. Apparently, the use of linear and alkaline-denatured template decreases the amount of secondary structure, providing nonamers better access to their template binding site. The 2× reaction mix seems to increase the signal-to-noise ratio; that is, increasing the DNA synthesis fluorescence signal derived from the correctly annealed nonamer–template duplex above that derived from random nonspecific DNA synthesis (background noise). Other peculiarities associated with the method include the use of low annealing temperature (20°C for 5 min) followed by a 5-min ramp to the 60°C extension temperature and the use of 50 cycles. The lower annealing temperature is needed to increase the concentration of the annealed nonamer–template duplex. Use of the slow ramp to the much higher extension temperature appears to be critical in keeping the nonamer bound to its template site and providing *Taq* polymerase time to initiate some limited amount of DNA synthesis at a lower temperature. We also found the use of DMSO [2% (v/v) final concentration] to be helpful, as it appears to increase nonamer specificity for its intended 100% matching complement within the template DNA.

It appears that a nonamer library can be effectively used to aid a manual primer walk across an unknown insert (20, 21); however, such a primer-walking strategy would be difficult using our present ABI373A-based method due to the high rate of primer failures and short reading lengths. The results shown in Fig. 1 suggest that the nonamer library may best be used with the ABI373A sequencer to obtain initial sequence contigs, followed by a switch to a primer-walking sequencing strategy using either custom primers or hexamer string primers (unligated or ligated; 16–18) selected from a complete hexamer library. However, this initial nonamer-based sequencing phase could only be accomplished provided that a method is developed to rapidly identify all of the nonamer primers that share a 100% match within an unknown insert. This type of information could be obtained by hybridizing the labeled (^{32}P or fluorescent) cloned DNA against all the members of the nonamer library. Such a process would be similar, but not as extensive, as the method known as se-

quence by hybridization (SHB) (27). Having the ability to identify 50 to 60 one-hit nonamers rapidly would essentially allow the nonamer library to be used in an initial shotgun mode to obtain initial internal contigs, followed by a switch to the primer-walking strategy to extend and eventually close the contigs. The integrated and sequential use of both shotgun and primer-walking strategies might allow high-throughput sequencing to be done without the need to generate a large number of M13-based subclones.

As stated above, the overall goal of using short primers is to develop a procedure that is not dependent on the use of custom primers and numerous subclones, while retaining a high-throughput sequencing capability. For this to be achieved for any primer library-based strategy, techniques must be developed for using short primers with fluorescence-based instruments. Thus far, it is clear that we and others have only partially satisfied these goals. We need to find techniques that will increase our success rate to greater than 80% before the use of any short primer library would be considered practical for large scale genome-type sequencing. Competing library strategies include the use of hexamer strings, which have also shown limited success with fluorescence-based sequencing instruments (28–30). Possibly the most exciting and potentially most effective method of using short primers may be derived from the ligation of these hexamer strings, as shown by Kaczorowski and Szybalski (16), who have effectively ligated hexamers to form 18-mers that were subsequently used to prime manual sequencing reactions. However, it remains to be determined whether ligated hexamers can be used effectively to prime sequencing reactions for fluorescence-based sequencing instruments.

References

1. J. D. Watson, *Science* **248,** 44 (1990).
2. D. L. Daniels, G. Plunkett III, V. Burland, and F. R. Blattner, *Science* **257,** 771 (1992).
3. S. G. Oliver, Q. J. M. van der Aart, M. L. Agonstoni-Carbone, *et al., Nature (London)* **357,** 38 (1992).
4. J. Sulston, Z. Du, K. Thomas, R. Wilson, L. Hillier, R. Staden, N. Halloran, P. Green, J. Thierry-Mieg, L. Qui, S. Dear, A. Coulson, M. Craxton, R. Durbin, M. Berks, M. Metzstein, T. Hawkins, R. Ainscough, and R. Waterston, *Nature (London)* **356,** 37 (1992).
5. B. F. Koop and L. Hood, *Nature Genet.* **7,** 48 (1994).
6. R. K. Wilson, B. F. Koop, C. Chen, N. Halloran, R. Sciammis, and L. Hood, *Genomics* **13,** 1198 (1992).
7. D. R. Siemieniak, L. C. Sieu, and J. L. Slightom, *Anal. Biochem.* **192,** 441 (1991).
8. J. L. Slightom, D. R. Siemieniak, L. C. Sieu, B. F. Koop, and L. Hood, *Genomics* **20,** 149 (1994).
9. P. L. Deininger, *Anal. Biochem.* **129,** 216 (1983).
10. B. R. Krishnan, D. Kersulyte, I. Brikun, C. M. Berg, and D. E. Berg, *Nucleic Acids Res.* **19,** 6177 (1991).

11. M. Strathmann, B. A. Hamilton, C. A. Mayeda, M. L. Simon, E. M. Meyerowitz, and M. J. Palazzolo, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 1247 (1991).
12. A. Edward, H. Voss, P. Rice, A. Civitello, J. Stegemann, C. Schwager, J. Zimmermann, H. Erfle, C. T. Caskey, and W. Ansorge, *Genomics* **6,** 593 (1990).
13. B. F. Koop, L. Rowan, W.-Q. Chen, P. Deshpande, H. Lee, and L. Hood, *BioTechniques* **14,** 442 (1993).
14. H. Voss, S. Wiemann, D. Grothues, C. Sensen, J. Zimmermann, C. Schwager, J. Stegemann, H. Erfle, T. Rupp, and W. Ansorge, *BioTechniques* **15,** 714 (1993).
15. F. W. Studier, *Proc. Natl. Acad. Sci. U.S.A.* **86,** 6917 (1989).
16. T. Kaczorowski and W. Szybalski, *Anal. Biochem.* **221,** 127 (1994).
17. J. Kieleczawa, J. J. Dunn, and F. W. Studier, *Science* **258,** 1787 (1992).
18. L. E. Kotler, D. Zevin-Sonkin, I. A. Sobolev, A. D. Beskin, and L. E. Ulanovsky, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 4241 (1993).
19. D. R. Siemieniak and J. L. Slightom, *Gene* **96,** 121 (1990).
20. J. L. Slightom, J. H. Bock, D. R. Siemieniak, G. D. Hurst, and K. L. Beattie, *BioTechniques* **17,** 536 (1994).
21. J. H. Bock and J. L. Slightom, *BioTechniques* **19,** 60 (1995).
22. J. L. Slightom, D. R. Siemieniak, and L. C. Sieu, *in* "Phylogenetic Analysis of DNA Sequences" (M. M. Miyamoto and J. Cracraft, eds.), p. 18. Oxford University Press, New York, 1991.
23. K. L. Beattie, N. J. Logsdon, R. S. Anderson, J. M. Espinosa-Lara, R. Maldonado-Rodriguez, and J. D. Frost III, *Biotech. Appl. Biochem.* **10,** 510 (1988).
24. L.-Y. E. Chang and J. L. Slightom, *J. Mol. Biol.* **180,** 767 (1984).
25. F. S. Sanger, S. Nicklen, and A. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74,** 5463 (1977).
26. R. K. Saiki, D. H. Gelfand, S. Stoffel, S. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich, *Science* **239,** 487 (1988).
27. R. Drmanac, I. Labat, I. Brukner, and R. Crkvenjakov, *Genomics* **4,** 114 (1989).
28. W. Hon and L. M. Smith, *Anal. Biochem.* **221,** 136 (1994).
29. L. Kotler, I. Sobolev, and L. Ulanovsky, *BioTechniques* **17,** 554 (1994).
30. W. R. McCombie and J. Kieleczawa, *BioTechniques* **17,** 574 (1994).

This Page Intentionally Left Blank

Section III

Transcription: Promoters, Transcription Factors, mRNA

This Page Intentionally Left Blank

[14] Analysis of Promoter Mutations Causing Human Genetic Disease

Lutz-Peter Berg, Deborah A. Scopes, Vijay V. Kakkar, and David N. Cooper

Introduction

Disease-causing mutations can exert their pathological effect at any level of the gene expression pathway. Mutations within gene coding regions may cause a disease phenotype via amino acid substitutions, defective splicing, alterations of the translational reading frame, or truncation of the encoded protein. Such mutations are routinely detected because most screening strategies employ the polymerase chain reaction (PCR) amplification of all exons and splice junctions of the disease gene under study. A different class of lesion, however, may often be overlooked by conventional screening strategies: mutations within sequences that control gene expression. These mutations are generally found outside the gene coding sequence; they serve either to increase or decrease the level of mRNA/protein synthesized rather than altering its nature.

The DNA sequence elements that direct the expression of a gene are typically found in its 5′-flanking region. Promoter elements, required to drive a basal level of transcription, are usually contained within the first 200 bp immediately upstream of the first exon; additional control sequences involved in enhancing or regulating transcription may be located at a considerable distance from the promoter, or even within the gene. The binding of transcription factors (DNA-binding proteins) both to DNA sequence motifs and to other transcription factors forms the molecular basis of transcriptional regulation. These interactions serve to promote the assembly of the basal transcription initiation complex and to regulate the rate of transcription. Only a small proportion of transcription factors present in a particular cell type are involved in the regulation of any given gene expressed in that cell type. The *cis*-acting DNA motifs contained within the regulatory region of a gene determine which transcription factors are able to bind and with what relative affinity.

Comparatively few regulatory mutations have been studied to date. The majority of reported mutations have been found in the regulatory regions of the human globin genes, where they are responsible for β- and δ-thalassemia as well as the clinically benign hereditary persistence of fetal hemoglobin (HPFH)(reviewed in Ref. 1). The study of these mutations has led to the identification of functional regulatory motifs in the promoter regions of the human globin genes and has provided evidence for the involvement of several known transcription factors, both ubiquitous and erythroid-specific, in their transcriptional regulation. Further examples of regulatory mutations

Methods in Molecular Genetics, Volume 8

that cause human genetic disease by altering transcription factor-binding sites include two point mutations in the *RB* gene that disrupt recognition sequences for ATF and Sp1, resulting in retinoblastoma (2), and a mutation in the *AFP* gene that causes hereditary persistence of α-fetoprotein by increasing the affinity of a binding site for the liver-enriched transcription factor HNF-1 (3). A fascinating example of the functional analysis of a gene promoter based on the study of regulatory mutations is hemophilia B Leyden, an initially severe bleeding condition that is ameliorated at puberty. Different single-base pair substitutions, clustered around the transcription initiation site of the *F9* gene, have been identified in different patients. Functional studies on these mutations have identified several liver transcription factors involved in the regulation of the *F9* gene and have provided tentative explanations for the observed increase of factor IX levels at puberty (4–6).

The above studies illustrate the two major goals of promoter mutation analysis: the identification of both DNA motifs and DNA-binding proteins, which together govern gene regulation *in vitro* (and by inference *in vivo*), and the use of this information to explain the clinical phenotypes observed in human genetic disease.

The protocols provided in this chapter are basic, established methods for the detection of disease-causing mutations and for the functional analysis of sequence elements involved in transcriptional regulation. Because promoter mutations have so far been reported only for a handful of genes, it would seem appropriate to combine these two methodologies in order to perform a systematic search for regulatory mutations and determine their effect on promoter function.

An example from our own laboratory is presented alongside the protocols in order to illustrate this strategy and the expected results. In an attempt to identify regulatory mechanisms responsible for the expression of the liver-expressed protein C(*PROC*) gene, we have screened the putative *PROC* gene promoter region from patients with protein C deficiency and a history of recurrent venous thrombosis. Protein C is the zymogen of a serine proteinase that plays an important role in the regulation of hemostasis; familial deficiency of protein C is an important risk factor for venous thrombophilia (7). Numerous mutations underlying hereditary protein C deficiency have been identified in the coding sequence and splice sites of the *PROC* gene (8). Although some 600 bp of DNA sequence upstream of exon 1 have been determined, the structure and function of the *PROC* gene promoter have not yet been studied. The detection of a novel mutation within this promoter and its functional analysis represents a first step toward this goal and provides the first example of a promoter mutation causing hereditary thrombosis.

Experimental Strategy

The search for promoter mutations involves the PCR amplification of DNA fragments containing (putative) transcriptional control elements from individual patients. By comparison with mutations in the protein-coding region of a gene, promoter mu-

tations are relatively rare. Therefore, a search for such mutations is likely to be successful only if a large collection of patient samples is available. Because the direct sequencing of a large number of patient-derived DNA fragments is a daunting prospect, one of several screening methods can be used to detect sequence variation in a particular DNA fragment. Direct sequencing of that fragment will then reveal the nature and location of the sequence variation.

Mutation Screening

Of the various screening methods currently in use, single-strand conformational polymorphism (SSCP) analysis is probably the easiest to establish and a basic protocol is therefore provided here. Single-strand conformational polymorphism was first described by Orita *et al.* (9) and relies on the detection of a difference in electrophoretic mobility of single-stranded DNA fragments differing by one nucleotide or more. Commonly, PCR products derived from a patient and a control individual are denatured to separate the two strands and then electrophoresed under nondenaturing conditions through an acrylamide gel. A difference in the band patterns of patient and control, observed either after silver staining of the gel, or on a resulting autoradiograph if radioactively labeled fragments have been used, is indicative of sequence variation. Of course, the location of the mutation within the DNA fragment remains to be established. One of the important parameters determining the efficiency of this method appears to be fragment size; Hayashi (10) identified 97% of mutations in fragments of 100–300 bp but only 67% in nucleotide strands between 300 and 450 bp in size. The fact that not all mutations are routinely detected suggests that the local sequence environment also plays a role in determining if a given nucleotide substitution will result in a mobility shift. Single-strand conformational polymorphism has a slightly lower detection efficiency than other screening methods such as denaturing gradient gel electrophoresis (DGGE) (11) and chemical mismatch cleavage (CMC) (12), but its simplicity made it the method of choice for the experiments reported in this study.

Functional Analysis

Once a mutation has been found, several types of experiment can be employed to elucidate both its effect on gene expression and the precise mechanism by which this effect is achieved. The effect of the mutation of the transcriptional activity of the promoter is studied by comparing the ability of the wild-type and mutant promoter to drive expression of a reporter gene. This type of assay relies on the linkage of the regulatory sequences to a reporter gene whose transcription can easily be quantified on transient transfection of the chimeric constructs into cultured human cells. Several genes have successfully been used as reporters; the most widely used systems are vectors containing the bacterial genes encoding chloramphenicol acetyltransferase (CAT) and β-galactosidase and the firefly-derived luciferase gene.

The recipient cell lines should be derived from a tissue in which the gene under study is normally expressed, for example, hepatoma cell lines for the analysis of liver gene promoters. Because cultured cell lines do not necessarily possess all the characteristics of the tissue they are derived from, a suitable cell line should be selected by confirming the expression of the gene of interest by Northern blotting.

The binding of transcription factors to the mutated region can be studied using a gel retardation assay (also known as gel shift assay or electrophoretic mobility shift assay, EMSA). In this type of assay, synthetic, double-stranded oligonucleotides, corresponding both to the wild-type and the mutant promoter sequence, are incubated with nuclear protein extracts prepared from cells known to express the gene under study. If protein contained in these extracts can bind to the radiolabeled oligonucleotide probe, then the resulting DNA–protein complex(es) can easily be distinguished from the unbound probe by virtue of their retarded migration during subsequent gel electrophoresis. The effect of a mutation on transcription factor binding can be manifested in different ways: it may alter the specificity of the wild-type binding site (the mutated site binds a different protein or no protein at all) or it may alter its affinity for the bound protein. Both possibilities can be investigated using gel retardation assays.

To characterize the molecular mechanism by which a promoter mutation exerts its pathological effect, it is necessary to identify the transcription factor(s) involved. Comparison of the DNA sequence around the mutation with published consensus sequences of known transcription factor-binding sites will often reveal candidate transcription factors. The inclusion of oligonucleotide probes corresponding to known consensus sequences and of transcription factor-specific antibodies in gel retardation assays then usually allows the identification of already known DNA-binding proteins affected by the mutation.

The two experimental approaches outlined above provide information about the effect of the mutation on promoter strength and on the binding of transcription factors. If expression plasmids containing cDNA encoding the identified transcription factors are available, cotransfection of such expression plasmids with reporter constructs containing the wild-type target promoter can be performed in order to demonstrate that the characterized transcription factor is indeed involved in the regulation of this gene under study. Furthermore, cotransfection experiments employing reporter gene constructs containing the mutant promoter have successfully been used to provide direct evidence for the molecular mechanism of pathogenesis.

Methods

Mutation Detection by Single-Strand Conformational Polymorphism Analysis

1. Use the PCR to amplify fragments of suitable size (<500 bp) and containing the putative promoter of the gene under study from the genomic DNA of patients.

Note: It is advisable to include two samples from control individuals in each SSCP experiment: one should be treated identically to the patient samples to reveal the normal band pattern, the second sample should not be denatured prior to gel electrophoresis and can be used to identify those bands that are derived from undenatured or reannealed DNA.

2. Precipitate the PCR products for 20 min at −20°C after addition of a 1/10 vol of 3 *M* sodium acetate and 2 vol of ethanol.
3. Wash the pellets with 300 μl of 70% (v/v) ethanol.
4. Resuspend the pellets in 10 μl of water.
5. Incubate 1 μl of precipitated PCR products with 0.5 μl (5 μCi) of [γ-^{32}P]ATP (3000 Ci/mmol), 1 μl (1-unit/μl dilution) of T4 polynucleotide kinase, 0.5 μl of 10× kinase buffer [500 m*M* Tris-HCl(pH 7.4), 100 m*M* $MgCl_2$, 50 m*M* dithiothreitol (DTT), and 100 m*M* spermidine] and 2 μl of water at 37°C for 30 min.
6. Add 5 μl of loading dye [95% (v/v) formamide, 0.05% (w/v) bromophenol blue, 0.05% (v/v) xylene cyanol, 20 m*M* EDTA] to the labeled PCR products.
7. Boil the samples for 3 min (except for the undenatured control) and snap-chill on ice.
8. Using standard sequencing gel apparatus, subject 5 μl of each sample to electrophoresis through a 6% (w/v) polyacrylamide gel [acrylamide–bisacrylamide 29:1; 5% (v/v) glycerol] at room temperature and 40 W; the duration of electrophoresis required for optimal separation should be determined empirically.

Note: It has been reported that some mutations are preferentially detected when the electrophoresis temperature is below room temperature. The remaining 5 μl of the denatured sample could therefore be used to run a second gel at a lower temperature, for example in the cold room.

9. After completion of the electrophoresis, fix the gel in 10% (v/v) methanol, 5% (v/v) acetic acid for 10 min.
10. Rinse the gel in water for 10 min and dry it in an 80°C oven.
11. Expose the dried gel for 12–72 hr at −80°C.

Note: The resulting autoradiograph should exhibit a pattern of two or more bands in each lane (the two single DNA strands in one or more secondary conformation); the undenatured sample should indicate whether one of the bands is derived from reannealed DNA. A variation in the pattern characteristic for a given PCR product is indicative of sequence variation and should be investigated by direct sequencing.

Example: Identification of a Novel Mutation in the PROC Gene Promoter

A 240-bp fragment containing the putative *PROC* gene promoter was PCR amplified from 151 unrelated individuals with recurrent thrombotic disease and a familial deficiency of protein C. These DNA fragments contained 203 bp 5′ and 36 bp 3′ to the reported transcription initiation site. Screening of these fragments by SSCP analysis revealed an aberrant band pattern in one patient as compared to the remaining patients and healthy controls (Fig. 1A). Direct sequencing of this fragment identified a heterozygous T→C transition 14 bp 5′ to the transcription initiation site (Fig. 1B).

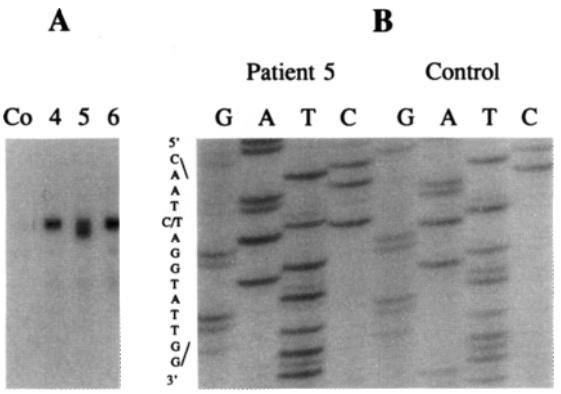


FIG. 1 Detection and characterization of a novel point mutation in the *PROC* gene promoter. (A) SSCP analysis of 240-bp fragments PCR amplified from the 5′ flanking region of a control individual (Co) and three patients (4, 5, and 6) with inherited protein C deficiency. (B) Sequence analysis of the 240-bp fragments of a control individual and patient 5. The nucleotide sequence of the sense strands from patient 5, showing the heterozygous T→C transition at position −14 relative to the transcription initiation site, is presented.

Further sequencing of all exons, splice sites, and 630-bp DNA sequence 5′ to exon 1 of the patient's *PROC* gene failed to detect any other deviation from the wild-type sequence. This lesion was not found in any of the other 150 protein C-deficient patients screened (by SSCP analysis) or in 20 control individuals (by restriction enzyme cleavage).

The lesion occurred within a sequence motif with strong homology to the consensus binding site for the liver-enriched transcription factor, hepatocyte nuclear factor 1 (HNF-1), also known as LF-B1, HP1 and APF (13) (Fig. 2). The palindromic

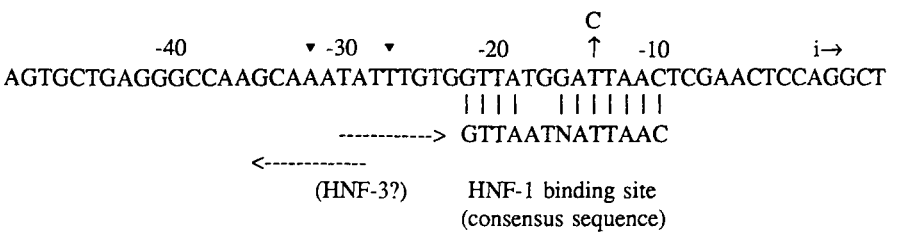


FIG. 2 Location of the HNF-1 binding site in the *PROC* gene promoter. i, Major site of transcription initiation, defined as +1; →, direction of transcription. The consensus sequence for the HNF-1-binding site is shown below the *PROC* gene promoter sequence; matching nucleotides are denoted by vertical lines; N, any nucleotide. The position of the T→C transition is indicated (↑). ▼, Previously reported but uncharacterized point mutations. The location of the putative HNF-3-binding site (TGT_3-type) is shown.

HNF-1 recognition sequence can bind two distinct homeodomain-containing proteins, HNF-1α and HNF-1β (also referred to as HNF-1 and vHNF-1), either as homo- or heterodimers. HNF-1α appears to be the predominant HNF-1 protein in the liver, where it plays a prominent role in determining the strength and cell specificity of promoters of a large number of liver-expressed genes.

Reporter Gene Analysis

Reporter Gene Constructs

The CAT (14) and luciferase (15, 16) reporter genes were used in the experiments described below to test the transcriptional activity of wild-type and mutant promoters; a construct containing the β-galactosidase gene linked to a strong promoter (simian virus 40, SV40) was cotransfected to correct for differences in transfection efficiency (17). Both reporter gene systems are suitable for comparing the strength of wild-type and mutant promoters. CAT is a bacterial drug resistance gene that inactivates chloramphenicol by acetylating the drug at one or both of its hydroxyl groups (18). In the CAT assay, extracts of transfected cells are incubated with ^{14}C-labeled chloramphenicol and the cofactor n-butyryl coenzyme-A. The substrate chloramphenicol is converted to its butyrylated forms by CAT. The converted forms of the radioactively labeled substrate are then separated either by thin-layer chromatography (TLC)(14) and quantitated by exposure to X-ray film or, alternatively, by phase extraction (19) and liquid scintillation counting. The activity of the reporter enzyme is taken as a measure of the strength of the promoter sequence linked to the reporter gene. Because the assay is based on the measurement of radioactivity, no specialized equipment is required; however, the assays are comparatively laborious and time consuming.

Firefly luciferase produces light by ATP-dependent oxidation of the substrate luciferin. A large excess of substrate is used in the luciferase assay; under these conditions, the light emission is proportional to the level of enzyme. The luciferase system offers several distinct advantages over the CAT system: the luciferase level can be determined in a rapid and simple bioluminescence assay that exhibits a linear relationship between luminescence and luciferase level over many orders of magnitude. Furthermore, the assay is not based on the use of radioactivity, yet its sensitivity is significantly higher than that of the CAT assay (20). However, these advantages are balanced by the need for a luminometer.

Vectors containing these or other reporter genes are available commercially [e.g., Promega (Madison, WI) pCAT and pGL (GeneLight, contains luciferase gene) vectors], together with constructs that can be employed as a positive control (the reporter gene driven by SV40 promoter sequences) or used to correct for transfection efficiency (pSV40βGal). The putative promoter region to be assayed was PCR amplified both from a control individual and from the patient carrying the identified mutation.

Restriction sites present in the polylinker region of the chosen reporter gene vector were incorporated into the PCR primers to facilitate cloning of the PCR products. The use of a high-fidelity polymerase is essential in these experiments because, in contrast to SSCP analysis and direct sequencing, where misincorporations usually affect only a fraction of the template molecules, individual clones are used here.

The ligation of the PCR fragments into the pCATBasic of pGLBasic vectors yielded the wild-type and mutant constructs to be tested, while the Basic vectors themselves served as negative controls. It is important to confirm that the transcripts derived from these chimeric constructs are indeed initiated at the correct site(s) within the test promoter rather than at cryptic sites in the vector. Primer extension analysis or, better still, S1 nuclease mapping (21) can be used to this end.

Calcium Phosphate Transfection

Several methods are currently in use for the introduction of DNA into cultured cells: calcium phosphate transfection, DEAE-dextran transfection, electroporation, and liposome-mediated transfection. The highest transfection efficiencies are normally achieved with electroporation and liposome-mediated transfection; however, the need for specialized equipment for electroporation and the relatively high cost of commercially available liposomes are distinct disadvantages of these methods. The calcium phosphate transfection employs an *N*-2-hydroxyethylpiperazine-*N'*-2-anesulfonic acid (HEPES)-buffered solution to form a DNA precipitate that adheres to the cell surface, resulting in subsequent incorporation of the DNA (22). A subsequent "glycerol shock" step (23) increases the transfection efficiency for many, but not all, cell lines. It should be established for each cell type whether or not this step is required.

1. Plate cells at about 3×10^6 cells/60-mm dish in Dulbecco's modified Eagle's medium (DMEM)–10% (v/v) fetal calf serum (FCS).

Note: The optimal density varies between cell types; in the experiments presented below, trypsinized cells derived from one confluent 150-mm^2 flask were usually plated onto nine 60-mm dishes.

2. Replace the medium on the morning of the following day (the cells should be about 50% confluent).
3. Mix plasmid DNA (5 μg of test plasmid and 5 μg of pSV β-galactosidase control plasmid) with 37 μl of 2 *M* $CaCl_2$ in a final volume of 300 μl (made up with water).
4. Gently vortex 300 μl of 2$\times$ HEPES-buffered saline (HBS), pH 7.10 (50 m*M* HEPES, 280 m*M* NaCl, 1.5 m*M* Na_2HPO_4, pH adjusted with NaOH), while adding the DNA–calcium chloride solution (300 μl) dropwise.
5. Incubate the precipitate at room temperature for 30 min before adding it to the growth medium [DMEM–10% (v/v) FCS] in the dishes.

6. Incubate the cells for 12–14 hr.

Note: The pH of the 2× HBS appears to be a critical determinant of transfection efficiency. It is recommended to test different 2× HBS solutions with varying pH values (within the range of 7.08 to 7.12) for optimal performance.

Glycerol Shock

1. Remove the medium and wash the cells with 2 ml of phosphate-buffered saline (PBS).
2. Expose the cells to 2 ml of glycerol solution [15% (v/v) glycerol in PBS] at room temperature for 2 min.
3. Wash the cells with 2 ml of PBS.
4. Add 5 ml of growth medium and incubate the cells for 12–24 hr at 37°C.

Preparation of Cell Extracts for Luciferase and β-Galactosidase Assays

1. After removal of the growth medium, wash the cells twice with 2 ml of PBS (Ca^{2+} and Mg^{2+} free).
2. Spread 400 μl of 1× reporter lysis buffer (Promega) evenly on the plates and leave for 15 min.
3. Scrape the cells off the plate and transfer to a microcentrifuge tube (on ice).
4. Subject the tubes to one freeze (liquid nitrogen)/thaw (37°C water bath) cycle.
5. Centrifuge the lysed cells for 2 min at 4°C (13,000 rpm) and transfer the supernatant to fresh microcentrifuge tubes.

Preparation of Cell Extracts for CAT Assays

1. After removal of the growth medium, wash the cells twice with 2 ml of PBS.
2. Spread 400 μl of TEN buffer [40 m*M* Tris-HCl (pH 7.5), 1 m*M* EDTA, 15 m*M* NaCl] evenly on the plates and leave for 5 min.
3. Scrape the cells off the plate and transfer to a microcentrifuge tube.
4. Centrifuge at 4°C for 10 min and resuspend the pellet in 200 μl of 0.25 *M* Tris-HCl, pH 8.0.
5. Subject the extract to three consecutive freeze/thaw cycles.
6. Centrifuge at 4°C for 10 min and transfer the supernatant to a fresh tube.

Note: Both types of cell extract are either used directly for CAT, luciferase, or β-galactosidase assays or stored at −70°C.

Reporter Gene Assays

For any reporter gene system used, it is imperative to establish that the obtained activity values are within the linear range of that particular assay. Therefore, serial dilutions of (commercially available) β-galactosidase, CAT, or luciferase should be used to generate standard curves for each type of assay.

CAT Assay

The separation of the modified chloramphenicol by phase extraction has several advantages over TLC: it is substantially easier, less expensive, and quantitation by liquid scintillation counting gives accurate results over a wider range of concentrations. For these reasons, only the phase extraction protocol is provided here.

1. Incubate 200 μl of lysate from each culture dish at 70°C for 10 min to inactivate proteinases.
2. Mix 50 μl of lysate with 65 μl of 250 m*M* Tris (pH 8.0), 5 μl of *n*-butyryl coenzyme-A (5 mg/ml) and 4 μl of ^{14}C-labeled chloramphenicol (0.025 mCi/ml; DuPont/NEN, Boston, MA) and incubate at 37°C for 3–12 hr.
3. Stop the incubations by addition of 350 μl of mixed xylenes (Aldrich, Milwaukee, WI).
4. After 30 sec of vortexing and a 2-min centrifugation at 10,000 *g*, transfer the xylene-containing (top) phase to a fresh microcentrifuge tube.
5. Add 100 μl of 250 m*M* Tris, pH 8.0.
6. Vortex the samples for 30 sec.
7. Centrifuge for 2 min (13,000 rpm) and transfer the xylene-containing (top) phase to a fresh tube.
8. Repeat this back-extraction once more.

Note: The back-extractions serve to remove any unmodified chloramphenicol, thereby increasing the sensitivity of the assay.

9. Add 200-μl aliquots of organic solution to appropriate scintillation fluid and determine the CAT activity of the cell extracts by liquid scintillation counting.

β-Galactosidase Assay

β-Galactosidase activity is quantitated using a simple colorimetric assay. The values can then be used to correct the results obtained with the test promoter/reporter gene system for differences in transfection efficiency. The reactions should be terminated when a faint yellow color has developed. If left for too long, the samples will be outside the linear range of the colorimetric assay and cannot therefore be used to correct for differences between dishes.

1. Mix 150 μl of 2× assay buffer [Promega; 120 m*M* Na_2HPO^4, 80 m*M* NaH_2PO^4, 2 m*M* $MgCl_2$, 100 m*M* 2-mercaptoethanol, 1.33 mg of *o*-nitrophyle-β-D-galactopyranoside (ONPG) per milliliter] with an equal volume of cell extract (either undiluted or diluted with 1× reporter lysis buffer) and incubate at 37°C for 1–12 hr.
2. Stop the reactions by adding 500 μl of 1 *M* sodium carbonate.
3. Quantify color development spectrophotometrically at a 420-nm wavelength.

Luciferase Assay

1. Equilibrate the cell extract and luciferase assay reagent (Promega) to room temperature.
2. Add 20–40 μl of cell extract to 100 μl of luciferase reagent and immediately insert samples into the luminometer (Turner); measure the luminescence over 20 sec.

Example: Effect of Mutation on Promoter Strength

The effect of the detected *PROC* mutation on promoter strength was studied by comparing the ability of the wild-type and mutant *PROC* gene promoters to drive *CAT* reporter gene expression in Hep G2 (human hepatocellular carcinoma) cells. Two *CAT* reporter gene constructs were made, containing 626 bp of putative *PROC* gene promoter sequence, PCR amplified from a control individual (pPCwtCAT) and the patient (pPCMuCAT), respectively. DNA sequence analysis confirmed that the two constructs differed only in the T→C transition at position −14 (data not shown). On transient transfection, CAT expression from pPCwtCAT was observed in hepatoma cells (Hep G2), but not in fibroblasts (MRC-5, human diploid lung fibroblasts). This indicated that the 626-bp fragment of *PROC* gene 5′-flanking sequence contains regulatory elements involved in the liver-specific transcriptional activation of the *PROC* gene. The promoter strength of pPCMuCAT, as assessed by measured CAT activity in Hep G2 cells, was drastically reduced [6.5% of that of the normal promoter (pPCwtCAT) in the typical experiment shown in the Fig. 3]. This suggests that the lesion disrupts the binding site(s) of a *trans*-acting factor(s) required for *PROC* gene expression.

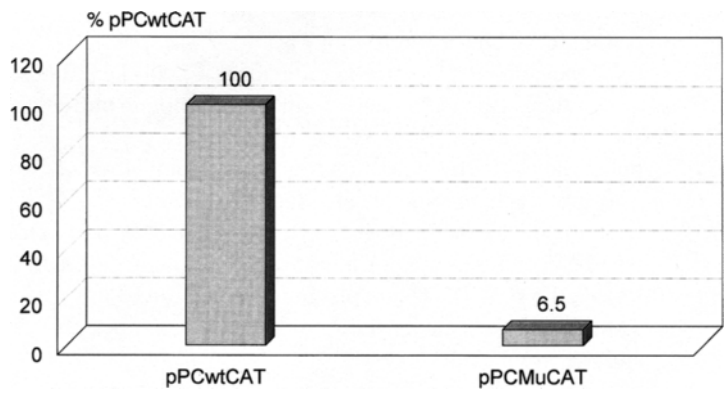


FIG. 3 Relative CAT activities from pPCwtCAT and pPCMuCAT in Hep G2 hepatoma cells. The average of three independent transfection experiments is shown; background activity (activity of pCAT-Basic) was subtracted.

Study of DNA–Protein Interactions

The gel retardation assay (24, 25) represents a simple, rapid, and sensitive tool to detect and analyze the binding of nuclear proteins to putative binding sites. Crude nuclear extracts from cells known to express the gene under study can be used together with synthetic double-stranded oligonucleotide corresponding to both the wild-type and mutant promoter region. If oligonucleotides between 18 and 30 bp in length are employed as radiolabeled probes, only one or two DNA complexes are normally observed. This should facilitate the comparison between binding affinities of the wild-type and mutant binding sites.

Preparation of Nuclear Protein Extracts

The procedure for the preparation of nuclear extracts from cultured cells follows that of Dignam *et al.* (26), but with simplifying modifications (27). The salt concentration in the extraction buffer given in this protocol worked well for several ubiquitous and liver-enriched DNA-binding proteins studied in our laboratory. However, the optimal KCl concentration for the preparation of active extracts may vary and should be determined empirically for each oligonucleotide probe.

1. Wash the cells contained in one 75-cm^2 culture flask with 20 ml of PBS.
2. Scrape the cells into 10 ml of PBS.
3. Pellet the cells by centrifugation at 500 *g* for 5 min.
4. Redissolve the pellet in 5 ml of ice-cold hypotonic buffer [10 m*M* HEPES (pH 7.9 at 4°C), 1.5 m*M* $MgCl_2$, 10 m*M* KCl, 0.2 m*M* phenylmethylsulfonyl fluoride (PMSF), 0.5 m*M* DTT] to which 0.1% Nonidet P-40 (NP-40) has been added; leave on ice for 10 min.
5. Lyse the cells in a Dounce homogenizer (20 strokes, B pestle).
6. Layer the suspension onto 5 ml of hypotonic buffer containing 10% (w/v) sucrose.
7. Pellet the nuclei by centrifugation at 400 *g* and 4°C for 5 min.

Note: At this stage, the nuclei can either be stored at −80°C or processed directly.

8. Lyse the nuclei by resuspending the pellet in 300 μl of extraction buffer [10 m*M* HEPES (pH 7.9 at 4°C), 1.5 m*M* $MgCl_2$, 400 m*M* KCl, 0.2 m*M* EDTA, 0.2 m*M* PMSF, 0.5 m*M* DTT] and placing it on ice for 10 min.
9. Transfer the suspension into microcentrifuge tubes and centrifuge at 10,000 *g* and 4°C for 15 min.
10. Dilute the supernatant containing the nuclear proteins with 3 vol of dilution buffer [20 m*M* HEPES (pH 7.9 at 4°C), 20% (v/v) glycerol, 0.2 m*M* EDTA, 0.2 m*M* PMSF, 0.5 m*M* DTT].
11. Aliquot the extract, freeze in liquid nitrogen, and store at −80°C.

Note: If fresh tissue samples are available to isolate nuclear extracts, the samples are finely chopped (on ice), suspended in hypotonic buffer, homogenized, and then treated identically to cultured cells.

Generation of Radiolabeled, Double-Stranded Oligonucleotide Probes

1. Pool 50-ng/μl solutions of two complementary oligonucleotides, incubate at 95°C for 5 min, and then allow to cool to room temperature.
2. Incubate 2 μl (100 ng) of these double-stranded oligonucleotides with 5 μl (50 μCi) of [γ-^{32}P]ATP (3000 Ci/mmol), 1 μl (8 units) of T4 polynucleotide kinase, and 2 μl of 10× kinase buffer [500 m*M* Tris-HCl (pH 7.4), 100 m*M* $MgCl_2$, 50 m*M* DTT, and 100 m*M* spermidine] at 37°C for 30 min.
3. Equilibrate Sephadex G-50 columns (Nick columns; Pharmacia, Piscataway, NJ) with 1 ml of water.
4. Add the sample in a volume of 500 μl (made up with water) and discard the eluate (containing mainly unincorporated ATP).
5. Elute the labeled oligonucleotide with 400 μl of water and store at −70°C.

Note: Aliquots of one batch of labeled oligonucleotide probe can be routinely used for several weeks after labeling. The resulting concentration of oligonucleotide probe (0.25 ng/μl) is convenient for the experiments described below. However, the probe concentration can easily be altered by using different amounts of oligonucleotide in the labeling reaction and eluting with a constant volume of 400 μl.

Gel Retardation Assay

1. Incubate 0.5–1 ng of labeled double-stranded oligonucleotides at room temperature with 5–10 μg of nuclear extract in the presence of 1 μl (2 μg) of nonspecific competitor [poly(dI-dC)·poly(dI-dC) (Pharmacia), 4 μl of 5× buffer [50 m*M* Tris(pH 7.5), 250 m*M* NaCl, 2.5 m*M* EDTA, 25% (v/v) glycerol], and, if required, unlabeled competitor oligonucleotides in a total volume of 20 μl for 15–60 min. If antiserum is used to produce a "supershift," 1 μl of the respective antiserum is added to the reaction after probe and protein extract.

Note: The binding conditions for individual DNA–protein complexes can be optimized by varying the composition of the binding buffer (different NaCl concentration, inclusion of divalent ions, e.g., $MgCl_2$, etc.). In addition, the duration and temperature of incubation can be varied. If the mutated sequence already reveals a candidate binding protein, appropriate binding conditions can be taken from the literature.

2. Load the reactions onto a 4% (w/v) acrylamide minigel (acrylamide–bisacrylamide, 19:1, 0.75-mm spacers) and electrophoresed at 100 V for 30–45 min.
3. After removal of the "front" plate, fix the gel in 10% (v/v) methanol–5% (v/v) acetic acid for 5 min.
4. Dry the gel (45 min at 80°C) and autoradiograph for 2–24 hr at −80°C.

Example: Promoter Mutation Abolishes Binding of HNF-1

The potential binding site for HNF-1 within which the mutation occurs differs from the consensus sequence in two positions (Fig. 2). To determine whether this site is capable of binding HNF-1 *in vitro* and whether the mutation affects its affinity for HNF-1, gel retardation assays were performed. These assays employed double-stranded synthetic oligonucleotides corresponding to the potential HNF-1-binding site from the normal (N) and mutant (Mu) alleles of the *PROC* gene as well as the consensus HNF-1-binding site (CS). On incubation with nuclear extracts from Hep 3B cells, the normal binding site (N) formed a DNA–protein complex (Fig. 4A, lane 2). In addition to the major complex, a faint band representing a low level of a complex of slightly higher electrophoretic mobility was sometimes visible (Fig. 4A–C). The binding was specific because it could be competed out by a 50-fold excess of the unlabeled normal binding site (N)(Fig. 4A, lane 3). The ability of the HNF-1 consensus binding sites (CS) to compete out the *PROC* gene HNF-1-binding site (N) (from which it differs at two positions) (Fig. 4A, lane 5) suggested that the bound protein was indeed HNF-1. The effect of the mutation on the affinity of the HNF-1-binding site was then demonstrated: the oligonucleotide containing the

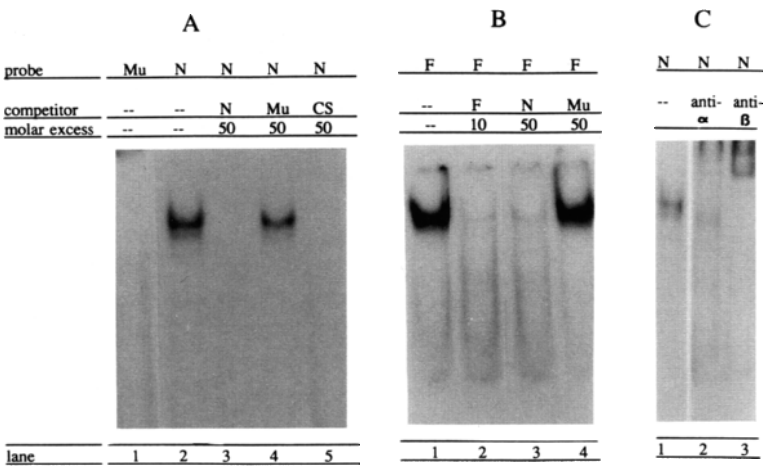


FIG. 4 Analysis of HNF-1 binding to the wild-type and mutant *PROC* gene promoter by gel retardation assay. The radioactively labeled probes and unlabeled competitors used were double-stranded oligonucleotides corresponding to the normal (N) and mutant (Mu) HNF-1-binding site between −8 and −20 in the *PROC* gene promoter, the consensus sequence for the binding site of HNF-1 (CS), and an HNF-1-binding site from the β-fibrinogen gene promoter (β28, designated F). The molar excess of unlabeled competitor used is indicated. "anti α" and "anti β" denote the use of specific antisera, raised against HNF-1α and HNF-1β, respectively, in the assays.

mutation (Mu) was unable to bind the nuclear protein either as labeled probe (Fig. 4A, lane 1) or as unlabeled competitor in 50-fold excess (Fig. 4A, lane 4).

The identity of the nuclear protein from Hep 3B cells, already presumed to be HNF-1, was confirmed in two ways: (a) the protein that bound to the normal HNF-1-binding site in the *PROC* gene promoter also bound specifically to the well-characterized HNF-1-binding site in the β-fibrinogen (*FGB*) gene (β28, designated F), albeit with lower affinity (Fig. 4B, lanes1–3). Again, the mutant binding site failed to bind HNF-1 (Fig. 4B, lane 4); (b) the inclusion in the assay of specific antisera against either one of the two distinct HNF-1 proteins (α and β) led to further retardation of the protein–DNA complexes (Fig. 4C). The resulting "supershifts" were characteristic for HNF-1α (lane 2) and HNF-1β (Fig. 4C, lane 3), respectively; this identifies the bound protein(s) as HNF-1 and indicates that both HNF-1α and HNF-1β are present in Hep 3B cells.

Discussion

The methodology outlined in this chapter provides the basic tools for the detection and functional characterization of regulatory mutations causing human genetic disease. This is illustrated by the analysis of a *PROC* promoter mutation in a patient with levels of circulating plasma protein and protein C activity concomitantly reduced to about half of normal. The observed laboratory phenotype had pointed to the virtually complete loss of transcription from the mutant *PROC* allele in the patient's liver. Consistent with this postulate, *CAT* reporter gene experiments demonstrated a drastic reduction in transcriptional strength of the mutant promoter in cultured human hepatoma cells. Using gel retardation assays, the *in vitro* binding of the liver transcription factor HNF-1 to the *PROC* gene promoter was shown to be abolished by the mutation.

To test whether the loss of HNF-1 binding was the cause of the observed reduction in promoter strength, cotransfection experiments using an expression plasmid for HNF-1α were performed. Overexpression of HNF-1α in Hep G2 cells (cotransfection of 3 μg of HNF-1α expression plasmid) resulted in *trans*-activation of the wild-type *PROC* gene promoter (average: 1.5-fold using a *CAT* reporter gene and 3-fold using a luciferase reporter gene). No *trans*-activation was observed with the mutant *PROC* gene promoter, suggesting that the mutation exerts its pathological effect by interfering with the HNF-1-mediated activation of the *PROC* gene promoter. Because both *in vitro* and *in cellula* promoter studies were consistent with the observed protein C deficiency state, it was concluded that the T→C transition 14 bp upstream of the transcription initiation site caused the loss of transcription from the disease allele by abolishing the binding of the transcriptional activator HNF-1 to the *PROC* gene promoter.

This analysis is indicative of a vital role for HNF-1 in the expression of the *PROC* gene *in vivo*. The HNF-1 proteins, together with the HNF-3, HNF-4, C/EBP, and DBP families of transcription factors, appear to represent the major regulators of liver-specific gene expression. Sequence motifs capable of binding HNF-1 have been found in the regulatory regions of more than 20 liver-expressed genes. However, evidence from *in vivo* footprinting experiments suggests that not all transcription factor-binding sites that are functional *in vitro* are necessarily occupied *in vivo*. This study therefore implicates HNF-1 in the transcriptional activation of the *PROC* gene *in vivo* and provides the first example of the abolition of HNF-1 binding as a cause of human genetic disease.

HNF-1-binding sites are usually found in the vicinity of other binding sites for both ubiquitous transcription factors and liver-enriched DNA-binding proteins. A search of the *PROC* gene promoter for recognition sequences of transcription factors that could in principle act in concert with HNF-1 revealed two TGT_3-type binding sites for hepatocyte nuclear factor 3 (HNF-3), partially overlapping and in opposite orientations, between -35 and -24 relative to the transcription initiation site (Fig. 2). Three putative *PROC* gene promoter mutations have been reported previously (8), although they have not yet been characterized at a functional level. Interestingly, two of these occur at the same relative position in either one or the other HNF-3-binding site (Fig. 2). HNF-1 and HNF-3 may therefore act in concert so as to drive *PROC* gene expression.

This study illustrates that the detection and analysis of naturally occurring regulatory mutations is an invaluable tool in understanding the molecular mechanisms underlying gene expression in health and disease. Unfortunately, this category of mutation has so far been comparatively neglected in the study of disease-causing lesions, presumably because the pathological effect of any sequence variation in 5′-flanking regions needs to be confirmed by functional tests. However, the important information generated whenever promoter mutations have been analyzed clearly argues for the routine inclusion of putative regulatory regions into mutation screening strategies. In addition to pathological lesions, such screening efforts may reveal polymorphisms in regulatory regions that may have an effect on gene expression and whose study may aid in understanding the often complex relationship between mutant genotype and clinical phenotype.

References

1. D. N. Cooper, *Ann. Med.* **24,** 427 (1992).
2. T. Sakai, N. Ohtani, T. L. McGee, P. D. Robbins, and P. D. Drya, *Nature (London)* **353,** 83 (1991).
3. J. H. McVey, K. Michaelides, L. P. Hansen, M. Ferguson-Smith, S. Tilghman, R. Krumlauf, and E. G. D. Tuddenham, *Hum. Mol. Genet.* **2,** 379 (1993).

4. M. Crossley, M. Ludwig, K. M. Stowell, P. De Vos, K. Olek, and G. G. Brownlee, *Science* **257,** 377 (1992).
5. M. J. Reijnen, F. M. Sladek, R. M. Bertina, and P. H. Reitsma, *Proc. Natl. Acad. Sci. U.S.A.* **89,** 6300 (1992).
6. D. J. Picketts, D. J. Lillicrap, and C. R. Mueller, *Nature Genet.* **3,** 175 (1993).
7. C. F. Allaart, S. R. Poort, F. R. Rosendaal, P. H. Reitsma, R. M. Bertina, and E. Briet, *Lancet* **341,** 134 (1993).
8. P. H. Reitsma, S. R. Poort, F. Bernardi, S. Gandrille, G. L. Long, N. Sala, and D. N. Cooper, *Thromb. Haemost.* **69,** 77 (1993).
9. M. Orita, H. Iwahana, H. Kanazawa, and T. Sekia, *Proc. Natl. Acad. Sci. U.S.A.* **86,** 2766 (1989).
10. K. Hayashi, *PCR Methods Appl.* **1,** 34 (1991).
11. R. M. Myers, S. G. Fisher, T. Maniatis, and L. S. Lerman, *Nucleic Acids Res.* **13,** 3111 (1985).
12. S. M. Forrest, H. H. Dahl, D. W. Howells, I. Dianzani, and R. G. Cotton, *Am. J. Hum. Genet.* **49,** 175 (1991).
13. F. Tronche and M. Yaniv, *BioEssays* **14,** 579 (1992).
14. C. M. Gorman, L. F. Moffat, and B. H. Howard, *Mol. Cell. Biol.* **2,** 1044 (1982).
15. J. R. De Wet, K. V. Wood, M. De Luca, D. R. Helinski, and S. Subramani, *Mol. Cell. Biol.* **7,** 725 (1987).
16. A. R. Brazier, J. E. Tate, and J. F. Habener, *BioTechniques* **7,** 1116 (1989).
17. P. Herbomel, B. Bourachot, and M. Yaniv, *Cell* **39,** 1044 (1984).
18. W. V. Shaw, *Methods Enzymol.* **43,** 737 (1975).
19. B. Seed and J.-Y. Sheen, *Gene* **67,** 271 (1988).
20. T. M. Williams, J. E. Buerlein, S. Ogden, L. F. Kricka, and J. A. Kant, *Anal. Biochem.* **176,** 28 (1989).
21. J. M. Greene and K. Struhl, *in* "Current Protocols in Molecular Biology" (F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidmann, J. A. Smith, and K. Struhl, eds.), Chap. 4.6. John Wiley & Sons, New York, 1995.
22. F. L. Graham and A. J. van der Eb, *Virology* **52,** 456 (1973).
23. E. Frost and J. Williams, *Virology* **91,** 39 (1978).
24. M. Fried and D. M. Crothers, *Nucleic Acids Res.* **9,** 6505 (1981).
25. M. M. Garner and A. Revzin, *Nucleic Acids Res.* **9,** 3047 (1981).
26. J. D. Dignam, R. M. Lebovitz and R. G. Roeder, *Nucleic Acids Res.* **11,** 1475 (1983).
27. L.-P. Berg, D. A. Scopes, A. Alhaq, V. V. Kakkar, and D. N. Cooper, *Hum. Mol. Genet.* **3,** 2147 (1994).

[15] Identification of Transcription Factors and Their Target Genes

Heike L. Pahl and Patrick A. Baeuerle

Transcription Factors Modulating Biological Processes

Various biological processes, such as differentiation or the response of a cell to environmental stimuli, are controlled by transcription factors. These proteins bind specific DNA sequences in promoters and enhancers, thereby inducing or repressing expression of their target genes. Transcription factors play a critical role, for example, during the differentiation of muscle fibers and red blood cells. In these tissues, the presence of a single transactivating protein, myoD or GATA-1, respectively, can decide whether an immature cell differentiates into a mature myelocyte or erythrocyte. Moreover, the aberrant expression of myoD in a fibroblast causes this cell to become a myoblast (1). Conversely, erythropoiesis does not occur in the absence of GATA-1, as was shown in "knock-out" mice (2). myoD and GATA-1 represent so-called "master transcription factors." They control the expression of many if not all genes required for the progression of a certain differentiation program. A novel master transcription factor, PU.1, has been described (3). PU.1 regulates the transcription of many myeloid (monocyte and macrophage)-specific genes. Interestingly, mice deficient in PU.1 lack hematopoietic progenitor cells and die during embryogenesis (4), illustrating the importance of this transcriptional regulator.

Steroid hormones such as estrogen, progesterone, glucocorticoids, thyroid hormone, retinoic acid, and vitamin D exert their effect on cell growth and development by activating transcription factors. These hormones enter cells by diffusion through the lipid bilayer. In the cytoplasm or nucleus, they bind to steroid hormone receptors, which become active transcription factors on ligand binding (5). Activated receptors bind to steroid response elements [e.g., glucocorticoid response elements (GREs) and thyroid hormone response elements (TREs)], thereby inducing or repressing transcription of their target genes (6). One steroid hormone receptor, the retinoic acid receptor-α (RAR-α), has been implicated in the etiology of some acute myelogenous leukemias (AML type M3) (7). Leukemic cells of this type carry a chromosomal translocation, which fuses the retinoic acid receptor gene on chromosome 17 to *PML*, a gene of yet unknown function on chromosome 9 (8). The function of the PML/RAR fusion protein remains unclear, but it must exert a dominant effect, as one translocation suffices to confer the leukemic phenotype.

Transcription factors also mediate the activation of cells by pathogenic stimuli such as bacterial or viral infection, ultraviolet (UV) irradiation, or oxidative stress. One central player in this response is the transcription factor NF-κB, which is present

Methods in Molecular Genetics, Volume 8

in an inactive, inducible cytoplasmic complex in almost all cell types. On stimulation of cells by a large variety of pathological agents (for a complete list see Ref. 9), the inhibitory subunit of NF-κB, IκB, is degraded, releasing the active NF-κB heterodimer, which subsequently translocates to the nucleus (10). There the transcription factor binds its cognate DNA sequence, inducing transcription of a wide variety of target genes including cytokines, adhesion molecules, cell surface receptors, and hematopoietic growth factors (9). Other stress-inducible transcription factors include HIF-1, a transcription factor that is induced by hypoxia (11), and AP-1, which is activated by UV and phorbol esters and induces transcription of a variety of genes, including those of its own subunits (12).

These examples illustrate the wide variety of physiological processes that are mediated by transcription factors. Thus, many biological mechanisms can be investigated at the molecular level by studying transcription factors and their target genes. In the following sections we describe the techniques used to study transcription factors and their interaction with DNA sequences of target genes. This chapter provides an overview of the experimental procedures, their usefulness, advantages, and disadvantages. For more detailed information, the reader is referred to publications that have successfully used the technique.

Methods Used to Study Transcription Factors

Two completely different categories of experiments are used to study transcription factors. One type of experiment demonstrates the ability of a transcription factor to bind a certain DNA sequence with high affinity. In these assays, which include the electrophoretic mobility shift assay (EMSA), footprinting techniques, and Southwestern blotting, a transcription factor will bind a given DNA sequence, taken, for example, from the promoter region of a gene of interest, as long as it agrees closely enough with the ideal binding site for this transcription factor. Whether the transcription factor functions to regulate transcription of this gene remains to be determined by a second category of experiments, the functional assays. These include deletion and linker scanning analysis of promoters and enhancers coupled to transient or stable transfections and reporter gene assays. If a transcription factor is identified and its cDNA available, additional experiments are possible to investigate its importance in regulating a particular gene or a biological process.

Defining the Target

In the following we describe a logical and efficient set of experiments that an investigator might use to identify the transcription factors that interact with the promoter and enhancer of his or her favorite gene (called *HFG* in the following) (for flowchart see Fig. 1). We will assume that a sizable portion of the *HFG* promoter has been

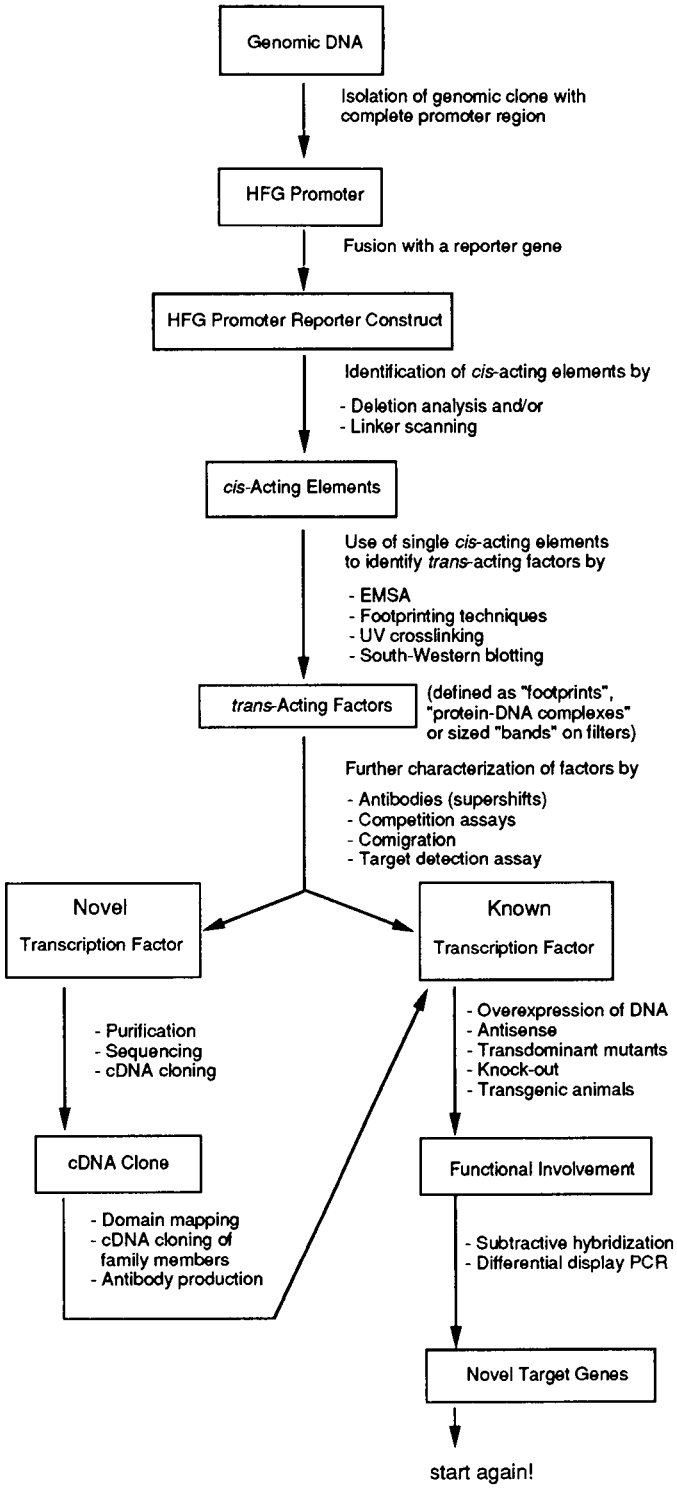


FIG. 1 Flowchart of techniques used to identify and characterize a transcription factor. For references see text.

cloned: The exact size of DNA required to obtain a functional promoter and enhancer varies from gene to gene, but it is often sufficient to begin with 1 to 5 kb of 5′-flanking DNA. Later, it may become necessary to look for control elements in the 3′-flanking regions or in introns as well, as it has become evident that these regions often also influence gene expression.

The 5′-flanking DNA should be cloned in front of a reporter gene to facilitate the detection of promoter function. If necessary, transcript levels can also be determined by S1 RNA analysis. Frequently used reporter genes include those encoding chloramphenicol acetyltransferase (CAT), luciferase (Luc), growth hormone (GH) and β-galactosidase (β-Gal). Of these, luciferase allows for the most sensitive detection of reporter gene activity, the luciferase assay being 10- to 100-fold more sensitive than the CAT assay. However, the measurement of luciferase activity requires a luminometer, so that not every investigator may be able to use this system. Several companies have marketed CAT enzyme-linked immunosorbent assay (ELISA) kits, abrogating the use of ^{14}C-labeled chloramphenicol and making CAT assays safer, faster, as well as slightly more sensitive. Growth hormone levels in the culture supernatant can be measured either by radioimmunoassay (RIA) or by ELISA, but both assays are not as sensitive as luciferase or CAT. β-Galactosidase is not recommended as a reporter gene *in vitro* because, besides lacking sensitivity, the spectrophotometric assay used for quantification of β-Gal activity varies up to 100% in duplicate samples, making small differences in activity impossible to detect. In transgenic mice, however, β-Gal is often the reporter gene of choice because tissue sections can be stained and reporter gene activity detected at the single-cell level. Several companies offer reporter gene vectors with convenient multiple cloning sites for the insertion of the *HFG* promoter upstream of either *Luc* or *CAT*.

The green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* was introduced as a reporter gene (13). As with the β-Gal reporter, gene expression can be measured in single cells; however, the GFP has the additional advantage that reporter gene activity can be examined in living cells using fluorescence microscopy.

Before any transcription factor–DNA interactions can be characterized, a functional promoter must be defined. This is done by transiently transfecting the plasmid containing the *HFG* 5′-flanking DNA upstream of the reporter gene into a cell line that expresses HFG. If HFG is expressed only in stimulated cells, cells should be stimulated for 12–24 hr after transfection and at least 6 hr prior to harvest for the reporter gene assay. A positive control for the transfection should be included and in most cell lines a cytomegalovirus (CMV) enhancer coupled to a reporter gene will give extremely high levels of activity. Depending on the level of expression of the endogenous gene, reporter gene activities between 20- and 1000-fold above background (untransfected cells or cells transfected with the promoterless reporter gene vector) should be obtained. This represents a functional *HFG* promoter, which, as mentioned above, may not contain all the information required for correct expression of the gene, but represents a good starting point in the search for required transcrip-

tion factors. Alternatively, if the 5′-flanking DNA fails to direct reporter gene activity, a larger fragment must be tested or introns and 3′-flanking DNA added to the construct until activity is detected. The DNase hypersensitivity assay, described below, can be used to find additional DNA segments required for gene transcription.

Finding Distant Enhancers: DNase Hypersensitivity

Often, DNA sequences required for promoter function are located at large distances from the transcriptional start site and are thus difficult to locate. To find these regions, the DNase hypersensitivity assay can be used (14). This assay is based on the observation that DNA segments that are bound by proteins, such as transcription factors, are "unraveled," that is, they are not condensed as tightly in the chromatin structure as other DNA sequences. Thus, these protein-binding sites are more easily digested by DNase I than the tightly coiled DNA. A limited DNase I digest will thus preferentially cut at these sites. In this assay, cells are exposed either to various concentrations of DNase I or to a fixed concentration for different periods of time. The goal is to find a DNA concentration or time point where the DNase-hypersensitive sites have been cut, but the rest of the DNA remains fairly intact. This often requires several trials. The genomic DNA is then extracted from the cells and digested to completion with a restriction enzyme. This enzyme is chosen so that at least one site has been mapped and can be used as a reference point. The digested DNA is run on an agarose gel and transferred to a nitrocellulose membrane in a Southern blot. This blot is probed with a DNA fragment from the *HFG* gene. DNA not treated with DNase but also digested with the restriction enzyme serves as a control. The probe will hybridize to a fragment in the control DNA. If a DNase-hypersensitive site is present within this fragment, hybridization will become increasingly weaker in the DNase-treated samples as the DNase I concentration or the time of digestion is increased. In return, a novel, smaller band will appear, which represents the DNA digested by DNase. Ideally, two novel bands should appear, which add up in size to the original fragment. In practice, however, this is rarely seen and the presence of one novel band is used to map the hypersensitive site. The location of the restriction enzyme site and the size of the novel band can be used to locate the DNase I hypersensitive site. This represents a possible enhancer sequence and should be cloned upstream of the promoter in the reporter gene constructs and tested for activity. The DNase I hypersensitivity assay may require several repetitions before results are obtained. For one, the optimal DNase I concentration and treatment time must be found, but in addition, several DNA fragments must be used as probes before all hypersensitive sites are found. Experiments that fail to show a hypersensitive site may result from technical errors, such that the DNA was over- or underdigested. In contrast, they may simply reflect the fact that no hypersensitive site is present in this region. Rehybridizing the blot with a DNA probe known to detect a hypersensitive site, such as a fragment from the rat tryptophan oxygenase gene (15), can serve as a control.

Landmarks: Deletion Mutants

The first set of experiments aims to delineate critical segments within the promoter, whose deletion strongly reduces its activity. For this purpose, a set of large deletion mutants is produced. Conveniently, the end points are defined by unique restriction sites within the promoter region. These constructs are again tested by transient transfection into HFG-expressing cells. For purposes of comparison, an internal transfection control should be included in each sample, as transfection efficiencies vary up to fivefold between samples, depending on the transfection method used. A detailed discussion of the different transfection techniques exceeds the scope of this chapter; however, they should be mentioned briefly. Four major transfection methods have been established for eukaryotic cells: calcium phosphate precipitation, DEAE-dextran transfection, electroporation, and lipofection. The optimal transfection technique differs for each cell line and the parameters of a given technique must be optimized for each cell line. In general, electroporation is the most efficient method of transfection and cells that previously seemed "untransfectable" have been successfully electroporated (16). Deletion analysis represents a first, rough assessment of the functional regions of a promoter. It is inherent in the procedure that large portions of promoter sequence are removed, which may contain clusters of several synergizing transcription factor-binding sites. This also explains the gradual increases or decreases in the overall activity of constructs.

Detailed Analysis: Linker Scanning

The results of the deletion analysis show where important regulatory regions of the *HFG* promoter are located. These areas should be investigated in more detail. One elaborate but extremely informative method of achieving a more detailed analysis is the use of linker scanning mutants (17). This form of mutation substitutes the same oligonucleotide (i.e., the linker) at consecutive positions in the promoter, for example, every 10 bp. Care should be taken not to choose an oligonucleotide that itself encodes a transcription factor-binding site. A great advantage of linker scanning over deletion analysis is that the context of the promoter is much better conserved. The easiest method of constructing linker scanning mutants is by using the polymerase chain reaction (PCR) (18). Two oligonucleotide primers (P-1 and P-2) are needed for each linker scanning mutant to be constructed. In addition, primers encoding the 5′ and 3′ boundaries of the promoter region under investigation are needed. Finally, a double-stranded oligonucleotide encoding the constant linker must be prepared. The oligonucleotide linker is chosen so that on annealing of the two strands, overhangs for two restriction enzyme sites are generated. The two primers that create the mutation also carry these sites in their sequence. The two constant primers encoding the

5′ and 3′ boundaries of the promoter region (P-5′ and P-3′) carry different restriction endonuclease sites. Care should be taken to choose those restriction enzymes that cut close to DNA ends. A helpful table describing enzyme activity on DNA ends is included in the reference section of the New England Biolabs (Beverly, MA) catalog. Two PCR reactions are run: PCR 1 with primers P-5′ and P-1, and PCR 2 with primers P-3′ and P-2. The two PCR products are cut with the appropriate restriction enzymes, purified, and used in a four-way ligation reaction. The ligation includes the reporter gene vector, cut with two different restriction enzymes, the two PCR products, and the oligonucleotide linker. Following ligation, the plasmids are transformed into *Escherichia coli* and the correct linker scanning mutants selected by minipreparation analysis. An extremely fast PCR method, which substitutes for conventional minipreparations, has been described by Henkel and Baeuerle (19). Instead of growing overnight cultures of individual colonies and isolating the plasmid DNA, this protocol uses the entire bacterial colony as the template in a PCR reaction. A PCR product of the correct size indicates the presence of the desired insert. In the case of our linker scanning mutants either the two outside primers, P-5′ and P-3′, or a combination of primers used to generate the mutants, such as P-5′ and P-1 or P-3′ and P-2, could be used to detect the correct clones in PCR minipreparations. Although linker scanning mutants seem difficult to generate at first, by using the PCR method 10 linker scanning mutants can be generated in 1 week, making this powerful approach feasible. The linker scanning mutants are again tested for their activity by transient transfection and reporter gene assay (20).

Visualizing Protein–DNA Interactions: Electrophoretic Mobility Shift Assay

If the linker scanning mutants show that promoter activity is drastically reduced when a sequence is substituted by the oligonucleotide linker, this strongly suggests that a regulatory protein binds this region, thereby activating transcription of the *HFG* promoter. An electrophoretic mobility shift assay (EMSA) can be used to verify that proteins bind this region. In an EMSA, a radioactive DNA probe, usually a synthetic double-stranded oligonucleotide, containing the region of interest is exposed to proteins in a cellular extract (total cell extracts or fractionated, nuclear extracts may be used). The proteins and the DNA probe are incubated in a binding reaction for 10 min to 1 hr, during which protein–DNA interactions occur. The exact composition of the binding reaction varies and must be optimized for each transcription factor. Below, we discuss the different components of the binding reaction in detail. In general, it contains the cellular proteins, the ^{32}P-labeled oligonucleotide, NaCl or KCl, $MgCl_2$, a nonspecific competitor DNA, bovine serum albumin (BSA; to coat absorbing surfaces), and either glycerol or Ficoll. The binding reaction is applied to a native polyacrylamide gel and electrophoresed at low amperage (15–25 mA) for 1 to 4 hr. The gel is dried without fixation and subjected to autoradiography. DNA that is not bound

by proteins will migrate quickly through the gel, whereas DNA bound in a protein complex will be slower, that is, retarded on the gel (hence the alternative name for an EMSA, gel retardation assay). Because various proteins may interact with a given DNA probe, several retarded complexes may appear in the gel. These may differ in their size or conformation. In general, the larger a protein–DNA complex, the slower it will migrate through the gel. DNA bending, which can be induced by the bound protein, will also slow down the mobility of the protein–DNA complex. Likewise, covalent modification, such as phosphorylation, has been shown to alter the mobility of ternary complex factor (TCF) proteins (21).

The protein–DNA complexes observed with a given DNA probe depend on many different parameters. When using a novel DNA probe, such as the fragment of the *HFG* promoter mentioned above, care should be taken to vary several parameters and investigate the different complexes obtained. In this way the chances of finding the correct, functionally important complex are maximized. The critical parameters in an EMSA include the following:

1. The buffer system: Three buffers are commonly used in an EMSA (22): a Tris–borate–EDTA buffer (TBE), a Tris–acetate–EDTA buffer (TAE), or a Tris–glycine buffer (TG). Because, in order to support continued protein–DNA interaction within the gel, all these buffers are of very low ionic strength, they often need to be recirculated during the electrophoresis. Depending on the buffer system used, different protein complexes may be observed on a given DNA probe. Likewise, the resolution of bands may vary.
2. Both the salt (Na^+/K^+) concentration and the magnesium (Mg^{2+}) concentration are critical parameters in an EMSA. When investigating a novel DNA probe, salt concentrations between 20 and 100 m*M* and Mg^{2+} concentrations between 0 and 5 m*M* as well as all combinations of the two should be tried. Again, several different patterns of protein–DNA complexes may be observed.
3. Other parameters that can be varied include the concentration and composition of the nonspecific DNA competitor. Poly(dI-dC) is most commonly used, but poly(dA-dT) or poly(dG-dC) should also be tried. Final concentrations between 0.02 and 0.2 mg/ml can be used in the binding reaction.
4. The pH of the binding reaction may also be varied, but it should be remembered that relevant protein–DNA interactions must take place under physiological conditions between pH 7.0 and 7.8.
5. Occasionally, the addition of polyamines, nucleotides, or metals, such as Zn^{2+}, can strongly increase DNA binding.
6. The inclusion of glycerol or Ficoll in the binding reaction has been shown to alter the migration of complexes in the EMSA gel, but not the overall composition of the complexes observed. Ficoll appears beneficial for a better resolution of protein–DNA complexes in TBE gels.

The binding conditions for the transcription factor NF-κB were optimized by Zabel *et al.* (23). This publication illustrates the points discussed above.

Specificity: Competition Assay

Because both specific and nonspecific protein–DNA interactions are observed in an EMSA, the next step must distinguish between the two, as only specific interactions with the *HFG* promoter are of interest. A competition assay accomplishes this as well as providing additional information about the proteins binding. In a competition assay a 10- to 1000-fold excess of nonradioactive DNA, with respect to the radioactive probe, is added to the binding reaction. Several different competitor DNAs are used to define the specificity of a protein–DNA interaction. First, the same DNA that is used as a radioactive probe ("self") is added as a competitor. By definition, this should abolish all binding activity in the autoradiograph. Because proteins have the same affinity for the radioactive and the nonradioactive probe, but binding to the nonradioactive DNA will occur 10- to 1000-fold more frequently, no protein binding is detected after addition of excess nonradioactive "self" DNA. Any binding activity not abolished by the addition of 100-fold excess self nonradioactive DNA surely represents an artifact. Second, a completely unrelated DNA is added as a competitor. In the case of promoter analysis, a sequence adjacent to the segment under investigation is often convenient. However, care should be taken to ensure that adjacent DNAs do not carry two copies of the same sequence, which may represent two binding sites for the same transcription factor and hence would compete with each other. It is advisable to mix the labeled and unlabeled oligonucleotides before addition to the proteins. In case of a high-affinity binding, if the protein is incubated first with the hot probe, the off-rate may be too slow to allow an efficient competition by the unlabeled probe. A protein interacting specifically with the DNA probe in an EMSA should, by definition, be competed by the addition of excess self, but not by the addition of excess, even up to 1000-fold, of nonself. Under defined conditions, competition assays can be used to determine the half-life of a protein–DNA complex (24).

Once specific binding to a small fragment of the *HFG* promoter has been established, other sequences may be tried in competition assays. Because many transcription factors have already been described, it is possible that the sequence in the *HFG* promoter resembles or is even identical to the binding site of a known transcription factor. In this case, characterized binding sites for this transcription factor may be used as competitor DNAs. If the known transcription factor indeed binds the *HFG* promoter, these DNAs should compete the protein–DNA complex in an EMSA. This is a strong indication but not yet a proof of the identity of the transcription factor under investigation. If the site resembles no known transcription factor-binding site,

it is possible that a novel transcription factor has been discovered. However, transcription factors may bind various sites with different affinities and care should be taken to investigate all possibilities. The National Library of Medicine has compiled a database of all identified transcription factor-binding sites and investigators may subject their sequences to a search on this database via computer. In this chapter, we mainly discuss methods of establishing the function of a known transcription factor on the *HFG* promoter. If, however, the transcription factor has not been previously described, techniques described toward the end of this chapter can be used to characterize the factor and isolate its cDNA.

Final Proof: Supershift Assay

Final identification of the transcription factor can be achieved by so-called "supershift assays." These are EMSAs to which an antibody has been added in the binding reaction. If the antibody recognizes the protein under investigation, a new, larger complex, containing the DNA probe, the protein, and the antibody, will be formed. This complex will migrate more slowly in the gel than the protein–DNA complex and thus will generate a novel band, in which the previous complex is supershifted. Some antibodies will not supershift but interfere with DNA binding. These antibodies simply abolish complex formation. If the correct controls, that is, unrelated antibodies that cause no change in the protein–DNA complex, are included and the specificity of the antibody has been established, an EMSA supershift can conclusively identify a protein–DNA complex as containing a known transcription factor. If the complex consists of several distinct subunits, antibodies can also be used to determine the subunit composition.

Alternative Methods: *In Vitro* Transcription/Translation

Often, a transcription factor has been previously described and characterized, but antibodies are not available to the investigator. If the cDNA clone for the transcription factor can be obtained, small quantities of the protein can be produced by *in vitro* transcription and translation. This material can be used in an EMSA with a probe from the HFG promoter and run adjacent to a binding reaction performed with a cell extract. If the *in vitro*-translated protein binds the probe and forms a complex similar in size to the complex formed by the protein in the cell extract, this suggests that the two proteins might be identical. In addition, the cDNA can be inserted into a eukaryotic expression vector and transfected. Using cells that transfect extremely well, such as the human embryonal kidney cell line 293, the expression vector can

direct levels of protein production that far exceed the physiological levels of the transcription factor. Extracts from transfected and untransfected cells can be compared in an EMSA. Untransfected cells should contain little or no DNA-binding activity; cells transfected with the expression vector should contain a DNA-binding protein that comigrates with the complex observed using cell extracts from HFG-expressing cells. Comigration does not prove the identity of the transcription factor under investigation, but represents suggestive evidence that, taken together for example with competition assays and point mutants, makes a strong argument.

Alternatively to the cDNA clone, a small quantity of purified protein may be obtained and compared to proteins from a cell extract in an EMSA. Again, complexes with similar migration patterns indicate a possible identity of the proteins. Several companies market purified transcription factors for this purpose.

Footprinting Techniques: Steps on DNA Ladder

DNase Footprinting

Besides the EMSA, a second group of techniques can be used to visualize DNA–protein interactions. These "footprinting techniques" include DNase I footprinting (25), methylation interference (26), and hydroxyl-radical footprinting (27). Like the EMSA, they reveal DNA–protein interaction based on affinity but give no information on the functionality of the binding. They are slightly more difficult to perform than the EMSA, but should be used in parallel to find proteins binding a novel promoter, because some transcription factors (such as SP1) do not bind effectively under EMSA conditions, but give nice footprints.

For DNase I footprinting, a DNA fragment is labeled at both ends using Klenow polymerase and radioactive nucleotides. The fragment is then cut with a restriction enzyme to generate a fragment between 50 and 250 bp in length. This cut also removes the radioactive label at one end of the DNA, leaving the resulting fragment labeled only on one strand. The labeled DNA fragment is purified by polyacrylamide gel electrophoresis (PAGE) and recovered by elution from the gel. The DNA is exposed to cell extracts in binding reactions similar to those used in the EMSA. The protein–DNA mixture is exposed to DNase I, and, as a control, DNA not bound to protein is also digested. As in the DNase I hypersensitivity assay described above, the amount of DNase used and the time of digestion need to be optimized. Under the correct conditions, the unbound, free DNA will yield a characteristic pattern of bands on electrophoresis on a standard sequencing gel. If the DNA exposed to cell extracts is bound by proteins, these will protect it from DNase digestion and bands resulting from digestion at these sites will be missing in the bound sample. Hence the name "footprint": the bound protein produces an interruption in the ladder of DNA fragments. DNase footprinting is subject to the same restrictions as protein–DNA inter-

actions in the EMSA discussed above. Nonspecific interactions occur, and in order to discern them competitor DNAs must be added in separate binding reactions.

Methylation Interference

As a DNase footprint is between 20 and 50 bp in length, this limits the precision with which the binding site can be mapped. Precise characterization of protein–DNA interactions can be performed using a related assay, methylation interference (26). In this assay, DNA fragments are again radioactively labeled at only one end and subjected to methylation by exposure to dimethylsulfate (DMS). On average, one methylation event per DNA strand should be achieved. Again, the concentration and time of exposure to DMS must be optimized. Methylated DNA is exposed to cell extracts in a standard EMSA binding reaction and protein-bound DNA is separated from free DNA by nondenaturing PAGE, as in an EMSA. The gel, however, is not dried, but exposed to film as is. Both the free and the bound DNA are cut from the gel and recovered by precipitation. The DNA is then cleaved at methylated guanidine residues, using piperidine, and analyzed on a sequencing gel. Because on average each DNA strand will be cleaved only at one site, a pattern of bands is obtained that represents all fragments with a terminal G residue in the free DNA. Because methylation at a G residue interferes with protein binding, the "bound" fraction will contain only those DNA fragments whose methylated residues did not interfere with protein binding. By comparison to the free DNA, G residues contacted by a protein can be identified as missing bands. Moreover, because each transcription factor shows a characteristic pattern of G residue interactions, methylation interference can suggest its identity.

Hydroxyl Radical Footprinting

A third footprinting technique, hydroxyl radical footprinting (27), precisely maps the borders of the footprint. In this technique, the radioactively labeled DNA fragment is cleaved evenly at the phosphate backbone by hydroxyl radicals, produced from H_2O_2 by the iron-dependent Fenton reaction. Again, titration is required to achieve only one cleavage per DNA molecule. Cleavage of the DNA at a certain position will prevent DNA binding. Protein–DNA complexes will contain only DNA fragments cleaved in positions that do not interfere with DNA binding. Comparison to the unbound DNA precisely defines sequence positions crucial for protein–DNA interactions.

Both DMS and hydroxyl radicals can also be used to treat preformed protein–DNA complexes. Because both are small molecules, the resolution of the footprint is much higher than with the bulky enzyme DNase I. Moreover, DMS and H_2O_2 can be used on intact cells. This is called *in vivo* footprinting—a powerful method to verify that a protein–DNA interaction occurs in living cells (19, 28, 29).

Point Mutants

Assuming that (a) the identified promoter sequence resembles the binding site of a known transcription factor, that (b) previously characterized binding sites for this factor compete the binding, and that (c) the methylation interference pattern matches the one previously described for this factor, several additional experiments can be performed to confirm the identity of the transcription factor. For many characterized transcription factors, the precise bases of their recognition sequence that mediates the protein–DNA interaction have been described, mainly by methylation interference analysis. This allows the creation of point mutations in which only 1–3 base pairs (bp) in the recognition sequence have been changed, completely abrogating protein binding. A similar mutation in the *HFG* promoter sequence under investigation should also abolish protein binding in EMSA and footprinting, if this sequence binds the same transcription factor. Furthermore, the same 2- or 3-bp mutation, when introduced in the context of the entire promoter and tested in transient transfections with reporter gene assays, should have the same effect as the linker scanning mutant in this region. This result would strongly support the idea that the protein observed in EMSA is crucially involved in transactivation of the *HFG* promoter in this region.

Functional Analysis: Transactivation in Transient Transfections

Besides providing protein through *in vitro* transcription and translation, the cDNA clone of a transcription factor suspected to direct transcription of the *HFG* gene can be used in functional assays. The *HFG* promoter/reporter gene construct can be transfected into cells that do not express HFG and in which the promoter is thus inactive. When the promoter construct is cotransfected into these cells together with an expression vector containing the transcription factor cDNA, the previously inactive promoter should be transactivated by the transcription factor and thus direct reporter gene expression. Likewise, expression vectors encoding transdominant negative mutants of a transcription factor can be cotransfected along with the reporter. The mutant protein can, for instance, inactivate the endogenous factor by forming a heterodimer that has a deletion in the DNA-binding domain of one subunit. Alternatively, a protein lacking a transactivating domain can be overexpressed. The mutant protein would occupy the DNA-binding sites, preventing the endogenous factor from binding and activating gene expression. Such experiments give convincing evidence that the transcription factor modulates HFG expression. Coupled to a control experiment that uses the promoter point mutant, which no longer binds the transcription factor in an EMSA or footprint and fails to be transactivated by cotransfection, this assay comes close to "proving" an involvement of the transcription factor in *HFG* regulation. However, this assay is not as simple as it seems. Many transcription fac-

tors exert their activity only in combination with other proteins or as heterodimers. Therefore, a mutation of a protein-binding site, for example during linker scanning analysis or by point mutations, causes a loss in promoter activity. However, the factor is not sufficient to transactivate the promoter on its own, requiring other accessory proteins that may not be present in all cells.

A variation of this assay does not use the entire promoter, but only an oligonucleotide encoding the protein-binding site, cloned in front of a minimal promoter [often the thymidine kinase (*tk*) promoter] and a reporter gene. The oligonucleotide is often multimerized, providing several adjacent binding sites for the protein and potentiating reporter gene activity. This construct is used in the cotransfection assay described above. However, such an experiment is rather unphysiological, as the protein-binding site is taken out of the context of the promoter. Moreover, effects that are only detectable once the binding site has been multimerized may be so weak as to be irrelevant for gene regulation; in the worst case they simply represent artifacts.

Size Determination: Southwestern Blotting

Often, an investigator will possess neither antibodies nor the cDNA clone or purified protein of the transcription factor suspected to control HFG expression. In this case a different method can be used, one that characterizes the DNA-binding protein but cannot identify it unambiguously. Southwestern blotting (30) determines the size of the DNA-binding protein. As the name suggests, this technique combines elements of the Southern and the Western blot. As in a Western blot, the substances under investigation are proteins, which have been separated according to size in a denaturing polyacrylamide gel and transferred onto a nylon membrane. Unlike the Western blot, however, the Southwestern uses a radioactive DNA fragment instead of an antibody as a probe. This probe contains the sequence recognized by the transcription factor. After incubation under conditions optimal for this particular protein–DNA interaction, a band corresponding to the size of the DNA-binding protein may be detected by autoradiography. If this size corresponds to that of the previously described transcription factor, this suggests that the two proteins are identical. Sometimes antibodies to the transcription factor exist, which are unsuitable for the supershift assay described above. Either these antibodies recognize only the denatured and not the native protein, or they do not bind under the conditions of an EMSA. In these cases the Southwestern blot may be probed sequentially, first with the DNA probe and then with the antibody. If both probes recognize bands of similar size this is suggestive evidence that this transcription factor binds the sequence under investigation. The use of two-dimensional gels in Southwestern blots could unequivocally identify transcription factors; this, to our knowledge, has not yet been done. Of course, this technique can also be used to establish the size of a novel transcription factor.

Size Determination: Ultraviolet Cross-Linking

A second method to investigate the size of a transcription factor is UV cross-linking (31). An oligonucleotide with the factor-binding site of interest is annealed to a shorter, complementary oligonucleotide. The missing bases are inserted using Klenow polymerase and a mix of ^{32}P-labeled dNTPs, in which dTTP is substituted by the photoreactive nucleotide bromodeoxyuridine (BrdU). It must be assured that BrdU is present within the sequence contacted by this protein. On UV irradiation, BrdU will covalently bind to a nearby amino acid of the transcription factor. The radioactive, covalent complex between the DNA probe and the protein is then subjected to SDS-PAGE and autoradiography. Usually, the addition of DNA will decrease the relative mobility of the protein by several kilodaltons. Trimming of the DNA in the complex using micrococcal nuclease and DNase I will minimize the increase in size. Should the DNA sequence interact with a heterodimer, the BrdU can be placed in the sequence such that individual polypeptides may be photoaffinity labeled. This was successfully performed with p50–p65 NF-κB complexes (32).

Cloning a Transcription Factor: Conventional or Southwestern Style

So far we have discussed methods of identifying and characterizing a transcription factor. In this section we discuss one possible method of cloning its gene should it represent a novel protein. Of course, classic methods, by which a total cellular extract is serially fractionated over a variety of columns until finally one single, pure protein has been isolated, can be used. In fact, because the EMSA (in addition to being fast, simple, and sensitive) requires only a small amount of protein, this represents a convenient method of assaying various column fractions for the presence of the transcription factor. The purified protein is then sequenced and degenerate oligonucleotides from this sequence are used to probe a cDNA library to clone the transcription factor. Nonetheless, fractionating cellular extracts is laborious and requires large amounts of cells to prepare extracts, which is costly. An alternative method of cloning transcription factors was described by Singh *et al.* (33). A cDNA library is generated in a λ phage (λgt11), which allows for the expression of the protein encoded by the cDNA. The phage are allowed to infect bacteria and are plated on petri dishes. The proteins produced by the phage are immobilized on nylon membranes by performing plaque lifts. As in the Southwestern blot, the proteins are incubated with a radioactive DNA probe, which contains the binding site of the transcription factor to be cloned. As in any cDNA cloning strategy, primary screens are performed in duplicate and followed by rounds of secondary and tertiary screening. Mutant DNA probes are used to test isolated cDNA clones for specific DNA binding at an early stage. Several transcription factors have been successfully isolated using this technique (34, 35).

Transcription Factors and Target Sequences

So far we have described assays that aim to find transcription factors acting on a known DNA sequence. What, however, if we have a suspected transcription factor and want to know which sequence it binds? Such, for example, was the case for a long time with the protooncogene c-*myc*. The example of c-*myc* immediately illustrates a disadvantage of the method we describe here. It is successful only if the protein under investigation binds DNA as a monomer or as a homodimer. Proteins that bind as heterodimers or require cofactors for binding will not yield results in this assay. The assay, called a target detection assay, relies on two techniques already mentioned above: the EMSA and PCR (36, 37). The purified protein of interest is incubated in an EMSA binding reaction, but in this case a DNA probe is used that consists of highly degenerate oligonucleotides, 30 bp in length, which have been radioactively marked. Because the central part of the oligonucleotide carries any of the four bases A, C, G, T at each position, every possible DNA sequence will be represented in the pool. As with any unknown protein–DNA interaction, the binding reaction must be carried out under various conditions. If the protein binds DNA, it will form a slower migrating complex in the EMSA. This protein–DNA complex will contain only oligonucleotides with an appropriate recognition sequence; the faster migrating free DNA will contain the sequences not recognized by the protein. The bound protein–DNA complex is eluted from the dried gel. To obtain only those oligonucleotides whose sequence is bound specifically with high affinity, this process needs to be repeated several times. However, the amount of DNA recovered from the dried gel is not sufficient for several rounds of EMSA and elution. Therefore, degenerate oligonucleotides were designed to contain short, constant sequences on either side of the probe (36, 37). Using primers that hybridize to these constant regions and PCR, the degenerate oligonucleotides can be amplified. They are then subjected to two more rounds of EMSA, elution, and PCR. The oligonucleotides isolated after the last EMSA are sequenced and from their similarities a consensus binding site for the transcription factor can be deduced. This sequence can subsequently be tested with specific, unique oligonucleotides in EMSAs. Once the binding site for a novel transcription factor has been determined in this way, comparison with characterized promoter sequences may reveal possible target genes. The target detection assay was used to determine the binding specificities of the NF-κB subunits p50 and p65 (38).

Transcription Factors and Novel Target Genes

Although it is possible to find transcription factor target genes by scanning characterized promoter sequences, this depends on the availability of cloned and sequenced promoters and by definition will never lead to the discovery of a novel gene. Several

strategies that require the inducible expression of a transcription factor have been developed to clone novel target genes. These methods use different techniques to achieve the same goal: mRNAs produced in the absence of the transcription factor are compared to those produced after the factor has been expressed or activated. Novel mRNAs, synthesized only in the presence of the transcription factor, are potential target genes. Induction of the transcription factor can be achieved (a) by cloning the gene into an inducible eukaryotic expression system, such as the tet system (39), (b) by treating with stimulators if the factor is naturally inducible, and (c) by generating a chimeric protein, for example by fusing the hormone-binding domain of the estrogen receptor to the transcription factor. The latter method renders the transcription factor inducible, as the fusion protein is inactive in untreated cells. Exposure to estrogen converts the chimeric protein to the active form. mRNA from unstimulated and stimulated cells is harvested and used in one of several assays. Historically, comparison between different mRNAs was done by generating subtractive libraries, in which only those mRNAs present in the induced but not in the uninduced sample were represented. Alternatively, the mRNAs were used to generate a "subtracted probe," a set of radioactively labeled mRNAs, again representing only those present in the induced sample but not in the uninduced mRNA. These mRNAs were used to probe a cDNA library generated from induced mRNAs. Both techniques have been used successfully to clone many genes (40, 41); however, they are difficult and time consuming, requiring skill and dexterity to achieve results.

More recently the differential display technique has been developed to compare different mRNA samples (42, 43). This technique uses reverse transcription coupled to radioactive PCR to display the mRNA pool as a characteristic pattern of bands in a sequencing gel. By comparing different mRNA samples, those bands present in the induced but not in the uninduced sample can be identified. The PCR bands are cut from the dried sequencing gel and the DNA eluted, precipitated, and reamplified in a second PCR reaction. The DNA obtained can be used directly as a probe on a Northern blot to verify that the gene is expressed only in the induced sample. The PCR product is easily subcloned and sequenced to investigate whether a known or a novel gene was cloned. In a final step, the PCR product is used as a probe to screen a cDNA library to clone the complete cDNA of the novel target gene. Although this technique is extremely fast it has disadvantages as well. The PCR technique is subject to artifacts and variability. Care should be taken to use at least four samples, two uninduced and two induced, and to demand that the band of interest be absent in both uninduced and present in both induced samples. Alternatively, if this is not practical, the reverse transcription and the PCR reactions should be performed twice independently. Only bands that appear to be induced in both experiments should be analyzed further. Several laboratories have successfully used differential display to clone novel genes (44, 45). The technique requires far less mRNA than the conventional subtractive hybridization methods and may thus represent the only feasible method for some investigations.

Function of Transcription Factor

Having found a novel transcription factor and perhaps some target genes, the final question remains to be answered: what physiological role does this factor play? The function of a protein can be deduced, for example, by observing the effect of eliminating it. To do so, either the gene encoding the factor, its mRNA, or the protein itself must be removed or rendered nonfunctional. To remove the gene, a transgenic mouse (a so-called "knock-out" mouse) that carries a deletion or substitution in the gene for the transcription factor can be created. By mating two mice that carry such mutations, homozygous offspring entirely lacking the transcription factor can be generated. Such animals were generated for the two hematopoietic transcription factors mentioned at the beginning of this chapter: GATA-1 and PU.1 (2, 4).

In the case of GATA-1, the task was made easier by the chromosomal location of the transcription factor gene: *GATA-1* is located on the X chromosome. In females, one X chromosome in each cell is inactivated by Barr body formation, leaving only one functional X chromosome. Mice heterozygous for a mutation in the *GATA-1* locus produced only red blood cells carrying the normal X chromosome. Cells in which the normal X chromosome had been inactivated did not differentiate past the erythroblastic stage, demonstrating the importance of *GATA-1* for erythropoietic differentiation (2). Mice carrying mutations in the gene for PU.1, a factor essential for the transcription of many myeloid (monocyte, granulocyte, and macrophage) genes, die at a late gestational stage (4). They show a multilineage defect in the generation of T cell, B cell, monocyte, and granulocyte progenitor cells. Thus, this transcription factor is essential for the generation of both lymphoid and myeloid cells in mice.

The transcription factor mRNA can be removed by transfecting a gene encoding antisense mRNA or by the exogenous application of antisense oligonucleotides. The antisense probes bind the mRNA, thereby prohibiting its translation into protein. In the case of vav, a protein of previously unknown function, it was shown that ES cells stably transfected with a *vav* antisense gene were no longer able to differentiate into hematopoietic cells (46). Consistent with the function of PU.1 observed in transgenic mice, antisense oligonucleotides against this transcription factor applied to hematopoietic stem cells prohibited their differentiation into hematopoietic colonies (47).

Transcription factor function can likewise be elucidated by the transfection of transdominant negative mutants. These proteins are altered so that they either heterodimerize with the wild-type factor and subsequently prevent its activity, or they bind and occupy DNA elements but no longer modulate transcription. This, in essence, has the same effect as the absence of the factor.

Another powerful approach to study function is the use of drugs that block activity of the transcription factor. Basically, all derivatives of steroid hormone receptors used now to control ovulation, sexual phenotype, or the growth of hormone receptive tumors target transcription factors. A second example is the immunosuppressive drug cyclosporin. It binds to an enzyme called cyclophiline (48). This complex inhibits

the protein phosphatase calcineurin. As a consequence, calcium can no longer activate the T cell transcription factor NF-AT by dephosphorylation. This perturbs T cell activation and, macroscopically, prevents the rejection of grafted organs (49).

The techniques outlined in this chapter can be used to identify, characterize, and clone transcription factors (for a summary see Fig. 1). The description of a biological role, as was achieved for NF-AT, GATA-1 and PU.1, represents the culmination of these investigations.

References

1. R. L. Davis, H. Weintraub, and A. B. Lassar, *Cell* **51,** 987 (1987).
2. L. Pevny, M. C. Simon, E. Robertson, W. H. Klein, S. I. Tsai, V. D'Agati, S. F. Orkin, and F. Costantini, *Nature (London)* **349,** 257 (1991).
3. F. Moreau-Gachelin, A. Tavitian, and P. Tambourin, *Nature (London)* **331,** 277 (1988).
4. E. W. Scott, M. C. Simon, J. Anastasi, and H. Singh, *Science* **265,** 1573 (1994).
5. U. Gehring and G. M. Tomkins, *Cell* **3,** 301 (1974).
6. R. Renkawitz, G. Schutz, D. von der Ahe, and M. Beato, *Cell* **37,** 503 (1984).
7. H. de Thé, C. Chomienne, M. Lanotte, L. Degos, and A. Dejean, *Nature (London)* **347,** 558 (1990).
8. K. Weis, S. Rambaud, C. Lavau, J. Jansen, T. Carvahlo, M. Carmo-Fonseca, A. Lamond, and A. Dejean, *Cell* **76,** 345 (1994).
9. P. A. Baeuerle and T. Henkel, *Annu. Rev. Immunol.* **12,** 141 (1994).
10. T. Henkel, T. Machleidt, I. Alkalay, M. Krönke Y. Ben-Neriah, and P. A. Baeuerle, *Nature (London)* **365,** 182 (1993).
11. G. L. Wang and G. L. Semenzy, *J. Biol. Chem.* **268,** 21513 (1993).
12. P. Angel and M. Karin, *Biochim. Biophys. Acta* **1072,** 129 (1991).
13. M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher, *Science* **263,** 802 (1994).
14. J. Stalder, A. Larsen, J. D. Engel, M. Dolan, M. Groudine, and H. Weintraub, *Cell* **20,** 451 (1980).
15. C. Kaltschmidt, M. Muller, G. Brem, and R. Renkawitz, *Mech. Dev.* **45,** 203 (1993).
16. H. L. Pahl, T. C. Burns, and D. G. Tenen, *Exp. Hematol.* **19,** 1078 (1991).
17. S. L. McKnight and R. Kingsbury, *Science* **217,** 316 (1982).
18. K. S. Zaret, J.-K. Lui, and C. M. Di Persio, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 5469 (1990).
19. T. Henkel and P. A. Baeuerle, *Anal. Biochem.* **214,** 351 (1993).
20. H. M. Chen, H. L. Pahl, R. J. Scheibe, D. E. Zhang, and D. G. Tenen, *J. Biol. Chem.* **268,** 8230 (1993).
21. J. M. Müller, M. Cahill, P. A. Baeuerle, and A. Nordheim, submitted (1995).
22. L. A. Chodosh, *in* "Current Protocols in Molecular Biology" (F. M. Ausubel *et al.*, eds), Vol. 2. Greene Publishing Associates and John Wiley & Sons, Brooklyn, New York, 1994.
23. U. Zabel, R. Schreck, and P. A. Baeuerle, *J. Biol. Chem.* **266,** 252 (1991).
24. U. Zabel and P. A. Baeuerle, *Cell* **61,** 255 (1990).
25. D. Galas and A. Schmitz, *Nucleic Acids Res.* **5,** 3157 (1978).
26. U. Siebenlist and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **77,** 122 (1980).

27. T. D. Tullius, B. A. Dombroski, M. E. A. Churchill, and L. Kam, *Methods Enzymol.* **155,** 537 (1987).
28. P. R. Mueller and B. Wold, *Science* **246,** 780 (1989).
29. R. A. Hipskind, V. N. Rao, C. G. F. Mueller, E. S. P. Reddy, and A. Nordheim, *Nature* **354,** 531 (1991).
30. W. K. Miskimins, M. P. Roberts, A. Mc Clelland, and F. H. Ruddle, *Proc. Natl. Acad. Sci. U.S.A.* **82,** 6741 (1985).
31. S.-Y. Lin and A. D. Riggs, *Proc. Natl. Acad. Sci. U.S.A.* **71,** 947 (1974).
32. M. B. Urban, R. Schreck, and P. A. Baeuerle, *EMBO J.* **10,** 1817 (1991).
33. H. Singh, J. H. LeBowitz, A. S. Baldwin, and P. A. Sharp, *Cell* **52,** 415 (1988).
34. C. R. Vinson, K. L. LaMarco, P. F. Johnson, W. H. Landschulz, and S. L. McKnight, *Genes Dev.* **2,** 801 (1988).
35. L. M. Staudt, R. G. Clerc, H. Singh, J. H. LeBowitz, P. A. Sharp, and D. Baltimore, *Science* **241,** 577 (1988).
36. G. Mavrothalassitis, G. Beal, and T. Papas, *DNA Cell Biol.* **9,** 783 (1990).
37. H.-J. Thiesen and C. Bach, *Nucleic Acids Res.* **18,** 3203 (1990).
38. C. Kunsch, S. M. Ruben, and C. A. Rosen, *Mol. Cell. Biol.* **12,** 4412 (1992).
39. M. Gossen and H. Bujard, *Proc. Natl. Acad. Sci. U.S.A.* **89,** 5547 (1992).
40. D. Barila, C. Murgia, F. Nobili, S. Gaetani, and G. Perozzi, *Eur. J. Biochem.* **223,** 701 (1994).
41. M. Eilers, S. Schirm, and J. M. Bishop, *EMBO J.* **10,** 133 (1991).
42. P. Liang and A. B. Pardee, *Science* **257,** 967 (1992).
43. D. Bauer, H. Müller, J. Reich, H. Riedel, V. Ahrenkiel, P. Warthoe, and M. Strauss, *Nucleic Acids Res.* **21,** 4272 (1993).
44. R. Sager, A. Anisowicz, M. Neveu, P. Liang, and G. Sotiropoulou, *FASEB J.* **7,** 964 (1993).
45. T. C. Burns and D. G. Tenen, *Blood,* in press (1995).
46. G. M. Wulf, C. N. Adra, and B. Lim, *EMBO J.* **12,** 5065 (1993).
47. M. T. Voso, T. C. Burn, G. Wulf, B. Lim, G. Leone, and D. G. Tenen, *Proc. Natl. Acad. Sci. U.S.A.* **91,** 7932 (1994).
48. R. E. Handschumacher, M. W. Harding, J. Rice, R. J. Drugge, and D. W. Speicher, *Science* **226,** 68 (1984).
49. S. L. Schreiber, *Cell* **70,** 365 (1992).

[16] Identification and Characterization of Transcription Factors from Mammalian Cells

Guang L. Wang and Gregg L. Semenza

Introduction

Rapid progress in molecular biology techniques during the past decade has made analysis of gene regulatory networks possible in mammalian cells, where little can be done using classic genetic approaches. Cells constantly express a set of genes to maintain essential cellular activities and modulate the expression of others in response to environmental and developmental signals. Gene transcription is regulated by transcription factors that recognize specific DNA sequences and interact with the basic transcriptional machinery to promote the assembly of active transcription-initiation complexes. Some transcription factors are constitutively active in cells whereas others are present in cells in an inactive state or are newly synthesized in response to cellular signaling. Some latent factors are sequestered in the cytoplasm and translocate into the nucleus to bind DNA after posttranslational modification or ligand binding. The activation of immediate response genes may result in *de novo* synthesis of other transcription factors and a cascade of transcriptional activation events, which ultimately modulate cellular metabolism, growth, or differentiation.

To identify transcription factors involved in regulating transcription of a specific gene it is necessary first to define the *cis*-acting DNA sequences that confer constitutive or inducible gene transcription. This can be carried out by systematically introducing an overlapping series of DNA fragments into a reporter plasmid vector and transfecting the plasmid constructs into suitable cells to test their ability to support transcription. After the promoter or enhancer sequences are localized, scanning or site-directed mutagenesis can be utilized to identify specific nucleotides that are critical for function. Eukaryotic genes are usually regulated by a complex array of regulatory sequences that extend over a large region surrounding the gene. Each enhancer may contain multiple binding sites for general or tissue-specific factors. It is the interactions among DNA-binding proteins that give rise to the immensely diverse expression patterns of different genes. Because of the modular structure of enhancers, a particular protein-binding sequence can be analyzed in a heterologous context to assess its individual conribution to enhancer activity. To analyze a complex regulatory sequence it may be helpful to perform initial protein-binding studies, such as DNase I footprinting analysis and gel shift assays, to identify putative protein-

Methods in Molecular Genetics, Volume 8

binding sites in the region. Caution should be taken, however, in evaluating the *in vitro* data because they may not accurately reflect protein–DNA interactions *in vivo*.

After definition of essential enhancer elements through functional analysis, *in vitro* protein–binding analysis can then be used to identify *trans*-acting protein factors that specifically recognize the *cis*-acting DNA sequences. Physiological relevance of a detected DNA-binding protein should be evaluated by multiple criteria: (a) whether mutations that abolish the enhancer activity also ablate DNA binding of the protein factor, (b) whether cell type-specific distribution of the protein factor coincides with the enhancer activity in different tissues, (c) whether treatments that activate or block the expression of the gene also activate or block the DNA-binding activity of the protein factor, and (d) whether induction kinetics of the DNA-binding activity correlate with induction kinetics for transcriptional activation of the gene. The next step is to study the biochemical nature of the DNA-binding protein, because this will determine the method taken to isolate cDNA sequences encoding the protein. Different cell types should be examined for abundance of the factor and the feasibility of using the cells for large-scale protein purification.

In this chapter we discuss approaches that are generally employed to (a) identify DNA-binding proteins that recognize specific *cis*-acting regulatory elements and (b) characterize the DNA-binding activity biochemically. We present these techniques as we have applied them to our analysis of hypoxia-inducible factor 1 (HIF-1). HIF-1 is a heterodimeric protein factor that is synthesized *de novo* when mammalian cells are exposed to hypoxia. HIF-1 binds a sequence in the hypoxia-inducible enhancer of the erythropoietin (*EPO*) gene that is required for transcriptional activation in hypoxic cells (1–5).

Methods

Analysis of Protein–DNA Interactions

Nuclear Extract Preparation

To identify protein factors that recognize a specific DNA sequence, protein extracts can be prepared from appropriate cell types and used for *in vitro* DNA-binding analyses. Because active transcription factors are nuclear proteins, isolated nuclei are an enriched source of DNA-binding proteins (6). Cytoplasmic extracts may also be prepared from the same cells to investigate the latent form of transcription factors. In some cases, latent proteins can be released from inhibitory proteins *in vitro* by treating the cytoplasmic extract with detergents or can be activated with specific ligands. The cytoplasmic fraction of the cells can also be used to prepare mRNA (7) to study the correlation between gene expression and a DNA-binding activity under different developmental or physiological conditions.

For initial studies of DNA-binding proteins, small-scale nuclear extract preparation may be desirable in order to quickly evaluate DNA-binding activity in different cell types or in cells subjected to various treatments. Several short protocols are available to prepare nuclear extracts from small numbers of cells (8, 9). The standard method for nuclear extract preparation developed by Dignam *et al.* (6) can also be adapted to prepare nuclear extracts from as few as 10^7 cells. This protocol allows preparation of both nuclear and cytoplasmic extracts. Nuclear proteins are extracted from isolated nuclei in 0.3 *M* KCl, which is usually sufficient to release soluble nuclear proteins. Higher salt concentrations will result in lysis of nuclei and extraction of inhibitory proteins. Isolated nuclei can be stored at −80°C for many months without loss of protein activity. A microdialysis step at the end of nuclear extract preparation serves to exchange the buffer to one suitable for subsequent *in vitro* DNA-binding analysis and to precipitate some structural nuclear proteins that can be removed by centrifugation. The protocol provided below has been used to detect HIF-1 DNA-binding activity in a variety of different cell types (1, 2).

Several important precautions should be taken in preparing nuclear protein extracts. Most proteins are sensitive to heat, oxidation, and mechanical shearing. Protein should be stored at −80°C and kept on ice or at 4°C when in use. Proteins are usually stored in a buffer containing 20% (v/v) glycerol to minimize denaturation during freezing and thawing. Protein samples should not be vortexed but mixed by gentle tapping or swirling instead. A cocktail of proteinase and phosphatase inhibitors is routinely added to the protein extract to minimize protein degradation and dephosphorylation, respectively. Choice of the final dialysis buffer depends on the optimal binding conditions for a particular factor. In the case of HIF-1, divalent cation is not required for the binding reaction and Mg^{2+} is therefore left out of the storage buffer, which inactivates most proteinases and nucleases. In addition, reducing reagents such as dithiothreitol (DTT) are generally included in all buffers to prevent protein oxidation.

Protocol 1: Nuclear Extract Preparation

This protocol is adapted from Dignam *et al.* (6).

1. Collect cells by centrifugation at 4°C for 10 min at 2000 *g*. Cells grown in a monolayer are washed and scraped into cold phosphate-buffered saline (PBS) before centrifugation.
2. Wash the cells with 5 PCV (packed cell volume) of buffer A* [10 m*M* Tris-HCl (pH 7.5), 1.5 m*M* $MgCl_2$, 10 m*M* KCl], resuspend in 4 PCV of buffer A,* and incubate on ice for 10 min.

*Freshly supplemented with 1:500 dilutions of separate stock solutions of 1 *M* DTT, 0.2 *M* PMSF, leupeptin (1 mg/ml), aprotinin (1 mg/ml), pepstatin (1 mg/ml), and 0.5 *M* Na_3VO_4.

3. Homogenize the cell suspension with 15 strokes in a glass Dounce homogenizer with type B pestle and transfer to a clean tube. Check for cell lysis by microscopic examination.
4. Pellet the nuclei from the cell lysate by centrifugation at 10,000 *g* for 10 min at 4°C. Mix the supernatant with 0.5 vol of buffer B* [75 m*M* Tris-HCl (pH 7.5), 60% (v/v) glycerol, 0.3 *M* KCl, 0.6 m*M* EDTA], clarify by centrifugation at 100,000 *g* for 30 min at 4°C, dialyze as in step 6, and designate as cytoplasmic extract. Alternatively, the supernatant can be used to prepare cytoplasmic RNA.
5. Resuspend the nuclei in 3 PNV (packed nuclear volume) of buffer C* [0.42 *M* KCl, 20 m*M* Tris-HCl (pH 7.5), 20% (v/v) glycerol, 1.5 m*M* $MgCl_2$] and rotate for 30 min at 4°C.
6. Centrifuge the suspension for 30 min at 15,000 *g* at 4°C and dialyze the supernatant against two changes of buffer D-0.1 [25 m*M* Tris-HCl (pH 7.5), 20% (v/v) glycerol, and 0.2 m*M* EDTA containing 0.1 *M* KCl], freshly supplemented with a 1 : 1000 dilution of 1 *M* DTT, 0.2 *M* phenylmethylsulfonyl fluoride (PMSF), and 0.5 *M* Na_3VO_4, for at least 4 hr at 4°C.
7. Clarify the dialysate by centrifugation for 10 min at 15,000 *g* at 4°C, divide into aliquots, freeze in liquid N_2, and store at −80°C.
8. Determine the protein concentration by the method of Bradford [e.g., Bio-Rad (Richmond, CA) protein assay kit].

Gel Shift Assay

In the gel shift assay, nuclear protein extracts are incubated with probe DNA in the presence of carrier DNA to allow the formation of specific protein–DNA complexes, and the reaction mixture is resolved on a native polyacrylamide gel. Protein–DNA complexes migrate slower than the unbound (free) probe DNA during gel electrophoresis (10, 11) and appear as retarded bands on autoradiography. Carrier DNA is used to adsorb nonspecific DNA-binding proteins that, because of their abundance, might otherwise fully occupy the probe molecules. The type of carrier DNA and ratio of protein extract to carrier DNA are of critical importance because the method is based on the relative affinity of sequence-specific and nonspecific DNA-binding proteins to probe and carrier DNA. Some transcription factors may bind with high affinity to poly(dI-dC) or poly(dA-dT). Genomic DNA such calf thymus DNA, salmon sperm DNA, and *Escherichia coli* DNA should therefore also be used in the initial optimization of binding conditions. The ratio of nuclear protein extract to carrier DNA should be tested over a wide range varying from 1 : 1 to 1 : 100 (w/w) to maximize the likelihood of detecting a new DNA-binding protein. Insufficient carrier DNA will result in binding of probe DNA by nonspecific DNA-binding proteins. Too much carrier DNA, on the other hand, will compete with probe DNA for specific DNA-binding proteins. In some cases, a compromise must made in which, in order to detect specific protein–DNA interactions, some nonspecific DNA-binding activi-

ties still remain. Because many DNA-binding proteins are active only in a reduced state, DTT should be included in the binding reaction buffer. Some protein factors also require the presence of specific ions or cofactors for DNA-binding activity.

The specificity of DNA binding can be determined by competition experiments with specific DNA, usually unlabeled probe DNA, or nonspecific DNA, usually a mutant version of the probe or an unrelated DNA sequence. In addition, mutant DNA when used as probe should not detect the same sequence-specific DNA-binding activity as the wild-type probe under the same assay conditions.

Figure 1 is an example of a gel shift assay. Nuclear extracts, prepared from Hep 3B human hepatoblastoma cells grown at 20% O_2 or 1% O_2, were incubated with either wild-type or mutant oligonucleotide probe DNA in buffer D-0.1 freshly supplemented with 5 m*M* DTT for 15 min on ice and resolved on a 5% (w/v) polyacrylamide gel. The wild-type probe contained an 18-bp sequence (Fig. 3B) from the *EPO* gene enhancer. Compared to the wild-type probe, the mutant probe contained a 3-bp

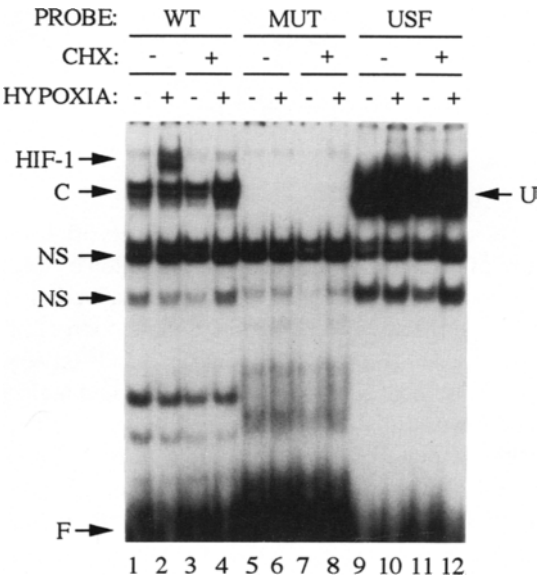


FIG. 1 Gel shift assay to detect HIF-1 DNA-binding activity in Hep 3B cells. Nuclear extracts were prepared from Hep 3B cells pretreated for 2 hr in the absence (−) or presence (+) of 100 m*M* cycloheximide (CHX) and placed in 20% O_2 (−) or 1% O_2 (+) for 4 hr. Extracts (5 μg) were incubated with wild-type (WT, lanes 1–4), mutant (MUT, lanes 5–8), and unrelated (USF) (lanes 9–12) oligonucleotide probes in the presence of 0.1 μg of calf thymus DNA for 15 min on ice, followed by electrophoresis on a 5% (w/v) native polyacrylamide gel in 0.3× TBE buffer, and autoradiography. Positions of HIF-1, constitutive (C), nonspecific (NS), and USF (U) DNA-binding activities and free probe (F) are indicated. [Reproduced with permission from G. L. Semenza and G. L. Wang, *Mol. Cell. Biol.* **12,** 5446 (1992).]

substitution that eliminated enhancer activity in transient expression assays (1). A probe for another transcription factor (USF) was used as a control for the overall DNA-binding activity of different nuclear extract preparations. By comparing the gel shift pattern of the wild-type and mutant probes, a hypoxia-induced (HIF-1) and constitutive activity (C) can be identified as binding specifically to the wild-type probe. Both HIF-1 and constitutive DNA-binding activities are actually doublet bands. Pretreatment of cells with the protein synthesis inhibitor cycloheximide (CHX) blocked induction of HIF-1, suggesting that HIF-1 is synthesized *de novo* in hypoxic cells (1).

In addition to detecting DNA-binding proteins, the gel shift assay can also be used to study the kinetics of protein binding (3) and to monitor the activity of a DNA-binding protein during analytical or preparative procedures, as described below. DNA–protein complexes are stabilized once they enter the gel and otherwise transient protein–DNA interactions can therefore be captured by this assay.

As more transcription factors are identified, it becomes increasingly likely that a DNA-binding activity represents a previously characterized factor. This possibility can be investigated by performing "supershift" experiments, in which antibodies against a suspected factor are incubated with nuclear extract and probe DNA. If the factor is part of a protein–DNA complex, binding of the antibodies to the complex will further retard migration of the protein–DNA complex on the gel. Alternatively, formation of the protein–DNA complex may be disrupted by the antibodies.

Protocol 2: Gel Shift Assay

1. Label one strand of the synthetic oligonucleotide at the 5′ end with T4 polynucleotide kinase and [γ-^{32}P]ATP. Anneal the labeled oligonucleotide to the second strand by heating to 85°C for 5 min and allowing the tube to cool down to room temperature over a period of 2 hr.
2. Run the annealed double-stranded oligonucleotide on a polyacrylamide gel in TBE buffer [89 m*M* Tris–borate (pH 8.3), 2 m*M* EDTA]. Isolate a gel slice containing the probe DNA after brief exposure of the wet gel to X-ray film.
3. Crush the gel with a pipette tip in a microfuge tube and elute the DNA by mixing with 800 μl of 10 m*M* Tris-HCl (pH 7.6), 50 m*M* NaCl, 1 m*M* EDTA at 37°C for 2 hr. After polyacrylamide gel fragments are pelleted by centrifugation, filter the supernatant through a 0.22-μm pore size filter and recover the DNA by ethanol precipitation.
4. Carry out the binding reactions in 20 μl of buffer D-0.1 freshly supplied with 5 m*M* DTT. Add probe DNA (10^4 cpm) after the nuclear protein extract (5 μg) is mixed with carrier DNA and incubated on ice for 5 min.
5. After a 15-min incubation on ice, load the reaction mixture on a 5% (w/v) polyacrylamide gel using 0.3× TBE as electrophoresis buffer (the gel is first prerun for 30 min at 185 V at 4°C). Load a separate lane with buffer D-0.1 containing 0.01% (w/v) bromphenol blue as a marker for electrophoresis.

6. Resolve by electrophoresis at 185 V for 2.5 hr at 4°C or until the dye is near the bottom of the gel.
7. Vacuum dry the gel on Whatman (Clifton, NJ) 3MM paper and expose to X-ray film at −80°C with intensifying screens.

DNase I Footprint Analysis

The DNase I footprinting method first described by Galas and Schmitz (12) is suitable for examining protein–DNA interactions over a region of several hundred base pairs. Nuclear extracts are incubated with probe DNA labeled at one end in a binding reaction buffer that has previously been optimized by gel shift assay. DNase I is added to the reaction to introduce limited digestion, at an average of one nick per DNA molecule. Protein-bound DNA sequences will be protected from digestion and appear as "missing" bands on the sequencing gel compared to free DNA probe subjected to the same treatment. Not all nucleotides, however, are equally accessible to DNase I digestion. DNase I-hypersensitive sites may also appear at the edge of a protected region. Compared to the gel shift assay, DNase I footprinting is less sensitive because it requires full saturation of DNA probe by a protein factor at a specific site. This technique also cannot distinguish different protein complexes that bind to the same region of DNA. In addition, proteins with rapid dissociation kinetics may not remain bound long enough to protect the probe from digestion.

DNA probes in the range of 50 to 250 bp with the anticipated position of protein binding near the middle of the fragment give best resolution. The probe should be labeled to high specific activity because decreasing the amount of probe increases the fraction of sites that are occupied. The amount of nuclear extract and nonspecific carrier DNA should also be varied to achieve complete protection. An optimal concentration of DNase I is the amount that will cleave approximately half of the probe DNA used in the reaction. Maxam–Gilbert chemical sequencing reactions (13) using probe DNA as substrate are usually performed and run in parallel with the footprinting reaction to determine precisely the sequence of the protected region. This can be substituted with the more convenient dideoxy chain termination method of DNA sequencing if the probe is a polymerase chain reaction (PCR) product, because the PCR primer can be used as the sequencing primer.

The results of a footprinting experiment are shown in Fig. 2. In this experiment, the probe was a 256-bp DNA fragment amplified by PCR with one primer labeled at its 5′ end. Nuclear extract (45 μg) from Hep 3B cells was incubated with 1 μg of poly(dI-dC) and probe DNA (10^4 cpm) in a buffer consisting of 10 m*M* Tris-HCl (pH 7.9), 50 m*M* KCl, 0.1 m*M* EDTA, 0.5 m*M* DTT, 5 m*M* $MgCl_2$ and 10% (v/v) glycerol for 15 min on ice. DNase I was added to final concentration of 0, 1, and 10 μg/ml and incubated for 1 min at 25°C. After stopping the reaction the DNA probe was purified and analyzed by denaturing polyacrylamide gel electrophoresis. The footprinted region in the enhancer sequence is a consensus binding site for members of the steroid/thyroid hormone receptor superfamily (1).

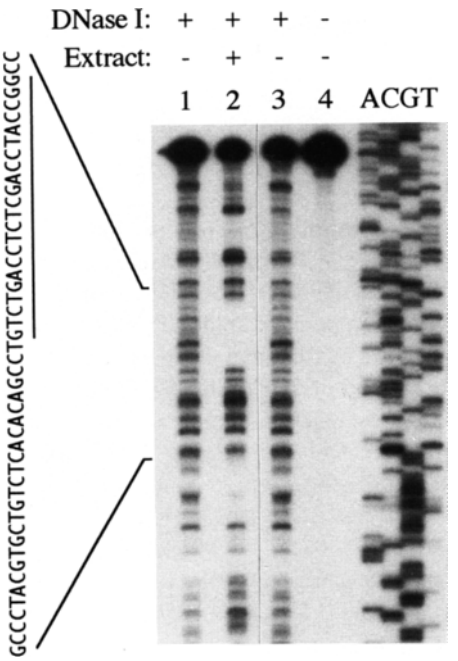


FIG. 2 DNase I footprinting analysis of the *EPO* gene enhancer. The PCR probe was incubated with 0 (lanes 1, 3, and 4) or 45 μg (lane 2) of nuclear extract from Hep 3B cells in the presence of 1 μg of poly(dI-dC), digested with 0 (lane 4), 0.05 (lanes 1 and 3), or 0.5 μg (lane 2) of DNase I for 1 min at 25°C, and analyzed by 8 *M* urea–8% (w/v) polyacrylamide gel electrophoresis (only the relevant region of the gel is shown). The nucleotide sequence of the enhancer, with footprinted region underlined, is shown at left. Dideoxy chain termination sequencing was also performed (ACGT) with the same 5′ end-labeled primer used for PCR. [Reproduced with permission from G. L. Semenza and G. L. Wang, *Mol. Cell. Biol.* **12,** 5447 (1992).]

Specificity of footprints can be examined by competition experiments. The footprint should be competed away by including in the reaction mixture an excess amount of an oligonucleotide spanning the footprinted region but not by an oligonucleotide of unrelated sequence.

An alternative method is to cleave protein-bound DNA with hydroxyl radical generated by the reduction of hydrogen peroxide with iron(II) (14). Hydroxyl radical attacks the DNA backbone primarily through the minor groove of the DNA helix and shows no sequence preference in cleaving DNA. Because hydroxyl radicals are very small, this method allows the detection of precise nucleotides that are protected by the bound protein.

Protocol 3: DNase I Footprinting

1. To prepare probe, label one PCR primer at the 5′ end with [γ-^{32}P]ATP and T4 polynucleotide kinase and perform PCR with the second primer and template DNA.
2. Purify the PCR product by polyacrylamide gel electrophoresis as described in Protocol 2, steps 2 and 3.
3. Set up the footprinting reaction in binding buffer optimized for a gel shift assay by incubating 50–100 μg of nuclear extract with 10^4 cpm of probe DNA in the presence of carrier DNA in a total volume of 50 μl for 15 min on ice.
4. Add a room temperature solution (50 μl) containing 5 m*M* $CaCl_2$, 10 m*M* $MgCl_2$ to the reaction mixture and incubate in a 25°C water bath for 2 min.
5. Add 1 μl of appropriate DNase I (Worthington Biochemical Corp., Freehold, NJ) dilutions to the reaction mixture to give final concentrations that range from 1 to 10 μg/ml and incubate in a 25°C water bath for exactly 1 min.
6. Add 100 μl of stop solution containing 0.6 *M* sodium acetate (pH 7.0), 50 m*M* EDTA, and yeast tRNA (0.1 mg/ml).
7. Extract the DNA once with phenol–chloroform and precipitate with ethanol.
8. Resuspend the pellet in 50% (v/v) formamide, 1 m*M* EDTA, 0.1% (w/v) xylene cyanol, and 0.1% (w/v) bromphenol blue, heat to 90°C for 2 min, and chill on ice.
9. Resolve the sample on an 8 *M* urea–8% (w/v) polyacrylamide sequencing gel (13) and visualize by autoradiography at −80°C with intensifying screens. In parallel, resolve a dideoxy chain termination sequencing reaction performed using the same radioactively labeled oligonucleotide as primer.

Methylation Interference Analysis

Methylation interference analysis, initially developed by Siebenlist and Gilbert (15), can be used to identify guanine residues that are closely contacted by DNA-binding proteins. The DNA probe is 5′-end labeled on either the coding or noncoding strand and subjected to limited methylation with dimethyl sulfate (DMS), methylating an average of one guanine residue per DNA molecule. The partially methylated probe is used in a preparative gel shift assay. After exposing the wet gel to X-ray film, the regions of polyacrylamide gel containing DNA–protein complexes and free probe are excised and the DNA probe is recovered from the gel slices. The purified probe DNA is then digested with piperidine, which cleaves the DNA after methylated guanine residues, and resolved on a polyacrylamide sequencing gel. If the methylation of a guanine residue interferes with protein binding, the DNA molecules recovered from the protein–DNA complex will not be cleaved at that guanine residue. If methylation decreases but does not eliminate protein–DNA interaction, a partial interference pattern will be detected.

To recover sufficient amounts of radioactively labeled DNA probe from the protein–DNA complex it is usually necessary to scale up the gel shift assay several-fold. Because a large sample volume results in distorted migration of retarded bands, conditions should be optimized so that the binding reaction can be performed with a higher concentration of nuclear extract. It is also important that the recovered DNA be purified by binding to a DEAE membrane followed by phenol–chloroform extraction. For optimal results, approximately half of the probe should remain uncleaved after piperidine treatment, indicating an average of one methylated guanine residue per DNA molecule. Overmethylation will result in a nonhomogeneous ladder, with bands that are more intense near the bottom of the gel.

Because DMS methylates guanine residues mainly at the N-7 position that protrudes into the major groove of DNA, interference with protein binding by guanine methylation indicates that the DNA–protein interaction occurs in the major groove. Dimethyl sulfate also methylates the N-3 position of adenines that protrude into the minor groove. Therefore, close inspection of the weak cleavage pattern at adenine residues will provide information regarding minor groove interactions as well. Alternatively, the probe can be partially alkylated by diethyl pyrocarbonate (DEPC). Because modification occurs at both adenine and guanine, this method is useful for analyzing binding sites that are adenine rich (16).

To identify the specific nucleotides contacted by HIF-1, the gel shift probe, which contains an 18-bp *EPO* enhancer sequence, was used for methylation interference analysis (Fig. 3). A preparative gel shift assay was performed with 100 μg of nuclear extract, 2 μg of calf thymus DNA, and 2×10^5 cpm of probe DNA in a reaction volume of 60 μl (as in Fig. 4, bottom, except that probe was partially methylated by DMS). Bands containing HIF-1, constitutive, and nonspecific DNA-binding activities, as well as free probe, were isolated, and the purified DNA was cleaved with piperidine. Both HIF-1 bands gave similar patterns of methylation interference. On the coding strand, methylation of G8 and G10 eliminated HIF-1 binding. Methylation of G10 only partially interfered with the binding of the constitutive factor. On the noncoding strand, methylation of G7 and G11 blocked HIF-1 binding to the probe. Only the methylation of G7 interfered with binding of the constitutive factor. The nonspecific binding factors were unaffected by DNA methylation on either strand. Methylation of adenine residues on either strand had no effect on HIF-1 binding. The results indicated that HIF-1 closely contacts G8 and G10 on the coding strand and G7 and G11 on the noncoding strand through the major groove of the DNA helix, and that HIF-1 and the constitutive DNA-binding factors can be distinguished by the nature of their DNA-binding site contacts (3).

Protocol 4: Methylation Interference Analysis

1. Prepare the DNA probe 5′-end labeled on one strand as described in Protocol 2, steps 1 to 3.

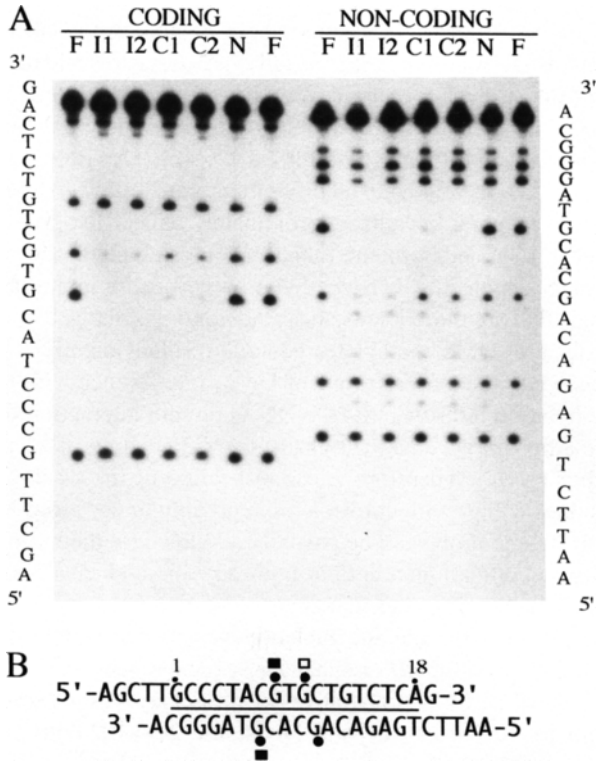


FIG. 3 Methylation interference analysis of HIF-1 binding to the *EPO* enhancer. (A) Polyacrylamide gel analysis. Probe was labeled on either the coding or noncoding strand, partially methylated, and incubated with nuclear extracts from hypoxic Hep 3B cells. DNA–protein complexes representing HIF-1 (I1, I2), constitutive (C1 and C2), and nonspecific (N) DNA-binding activities, as well as free probe (F), were isolated from a preparative gel shift assay (as in Fig. 4, bottom), and the DNA was cleaved with piperidine and analyzed by electrophoresis on an 8 *M* urea–15% (w/v) polyacrylamide gel in 1× TBE. (B) Summary of results. The oligonucleotide probe is shown with the *EPO* gene coding strand. Nucleotides 1 to 18 from the *EPO* enhancer are indicated by the line. Guanine residues whose methylation completely interfered with HIF-1 binding are shown as filled circles. Residues whose methylation completely and partially interfered with the binding of constitutive activities are indicated by the filled and open squares, respectively. [Reproduced with permission from G. L. Wang and G. L. Semenza, *J. Biol. Chem.* **268,** 21513 (1993).]

2. Resuspend 10^6 cpm of probe in 10 μl of TE buffer [10 m*M* Tris-HCl (pH 8.0), 1 m*M* EDTA]. Add 200 μl of DMS reaction buffer [50 m*M* sodium cacodylate (pH 8.0), 1 m*M* EDTA, 1 μl of DMS], mix well, and incubate at 22°C for 5 min.
3. Add 40 μl of DMS stop buffer [1.5 *M* sodium acetate (pH 7.0), 1 *M* 2-mercaptoethanol], 1 μl of tRNA (10 mg/ml), and 600 μl of ethanol to the re-

action mixture. Pellet the DNA after incubation in a dry ice–ethanol bath for 10 min.

4. Resuspend the pellet in 250 μl of 0.3 *M* sodium acetate–1 m*M* EDTA. Add 750 μl of ethanol and precipitate as in step 3.
5. Repeat the ethanol precipitation once as in step 4. Wash the pellet once with 70% (v/v) ethanol and resuspend thoroughly in TE buffer after air drying at 22°C for 10 min.
6. Perform the preparative gel shift assay and excise the gel slices corresponding to bound and free probes after autoradiography of the wet gel at 4°C overnight.
7. Recover the DNA probe as in Protocol 2, step 3. Spot the DNA onto a DEAE membrane (Schleicher & Schuell, Keene, NH) and wash the membrane extensively with TE buffer.
8. Recover the DNA from the DEAE membrane by incubating the membrane in DEAE elution buffer containing 10 m*M* Tris-HCl (pH 8.0), 1 m*M* EDTA, 1 *M* NaCl at 65°C for 30 min.
9. Extract the eluted DNA twice with phenol–chloroform and precipitate by adding 10 μg of tRNA and 2 vol of ethanol.
10. Resuspend the DNA pellet in 100 μl of 1 *M* piperidine and incubate at 95°C for 30 min.
11. Lyophilize the DNA and resuspend in 100 μl of distilled water. Repeat the lyophilization three times. Dissolve the DNA in 5–10 μl of formamide loading buffer and determine the radioactivity of the samples.
12. Load 2500 cpm of each sample on a denaturing polyacrylamide gel, resolve by electrophoresis, and expose to X-ray film with intensifying screens at −80°C overnight.

Biochemical Characterization of Transcription Factors

Ultraviolet Cross-Linking Analysis

Once a sequence-specific DNA-binding activity has been detected in crude nuclear extract, ultraviolet (UV) cross-linking analysis (17) and Southwestern blot hybridization can be performed to determine the mass of the DNA-binding polypeptide(s). In the Southwestern hybridization procedure, nuclear protein extracts are separated by denaturing sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). After blotting onto a nitrocellulose membrane, the proteins are renatured *in situ* and probed with the radioactively labeled DNA fragment bearing the binding site for the factor (e.g., probe used for gel shift assays). This method has several limitations. Many protein factors cannot be renatured into proper conformation to restore DNA-binding activity. In addition, the protein must be able to bind DNA as a monomer in order to be detected. Ultraviolet cross-linking analysis, on the other hand, is carried out with the protein in a native intact state, and DNA–protein complexes with relatively short half-lives can also be detected.

Ultraviolet irradiation causes the formation of covalent linkages between DNA and proteins that are in close contact with the DNA. The time and wavelength of UV irradiation need to be optimized for different DNA–protein complexes. The DNA-binding component of the protein can then be resolved on an SDS-polyacrylamide gel together with molecular weight standards and detected by autoradiography. When crude nuclear extracts are used in the binding reaction, multiple cross-linked proteins may be detected if the probe is bound by more than one factor. To identify a cross-linked protein as arising from a specific DNA-binding activity, the protein–DNA complex can be excised from a preparative gel shift assay after UV irradiation and analyzed by SDS-PAGE. If the binding reaction mixture is loaded directly on an SDS-polyacrylamide gel after UV irradiation, NaCl should be substituted for KCl in the binding buffer to avoid formation of precipitates when it is mixed with Laemmli buffer (18). The molecular weight of the DNA-binding protein can be estimated by subtracting the molecular weight of the probe from the apparent molecular weight of the whole protein–DNA complex. Alternatively, the probe can be uniformly labeled by incorporating ^{32}P-labeled deoxynucleotide with Klenow fragment of DNA polymerase. After UV irradiation, DNase I is added to remove most of the DNA sequence from the protein, leaving behind only a small radioactive tag representing nucleotides protected from digestion by protein binding. To increase the efficiency of the cross-linking reaction, thymidine residues can be substituted with bromodeoxyuridine (BrdU) by a Klenow extension reaction in the presence of 5′-bromo-2′-deoxyuridine triphosphate with a primer, which is synthesized with BrdU instead of thymidine.

Specificity of the DNA-binding protein can be examined by competition experiments with wild-type and mutant DNA sequences. Formation of a specific DNA–protein complex should be inhibited by an excess amount of unlabeled wild-type, but not by mutant, probe DNA in the binding reactions. In addition, a control UV cross-linking experiment can be performed with the mutant probe to localize nonspecific DNA-binding proteins.

For detecting HIF-1 DNA-binding subunits, binding reactions were performed with either nonhypoxic or hypoxic Hep 3B nuclear extracts (100 μg) and 10^6 cpm of probe DNA in the presence of 2 μg of calf thymus DNA carrier (Fig. 4). After UV irradiation for 30 or 60 min, cross-linked proteins were identified by SDS-polyacrylamide gel electrophoresis and autoradiography. A protein–DNA complex corresponding to a molecular mass of ~140 kDa was detected only in the hypoxic nuclear extract. After subtraction of the molecular weight contributed by the probe (~16 kDa), the apparent molecular weight of the DNA-binding protein is ~124 kDa. In a parallel experiment the reaction mixture was resolved on a native polyacrylamide gel, UV-irradiated, and exposed to X-ray film to identify different DNA-binding activities. Gel slices containing HIF-1, constitutive, and nonspecific activities were isolated, and eluted proteins were analyzed by SDS-PAGE. The same DNA-binding protein was identified. Longer exposure of the gel revealed two additional cross-linked complexes in the isolated HIF-1 bands, of apparent molecular weight ~94 and

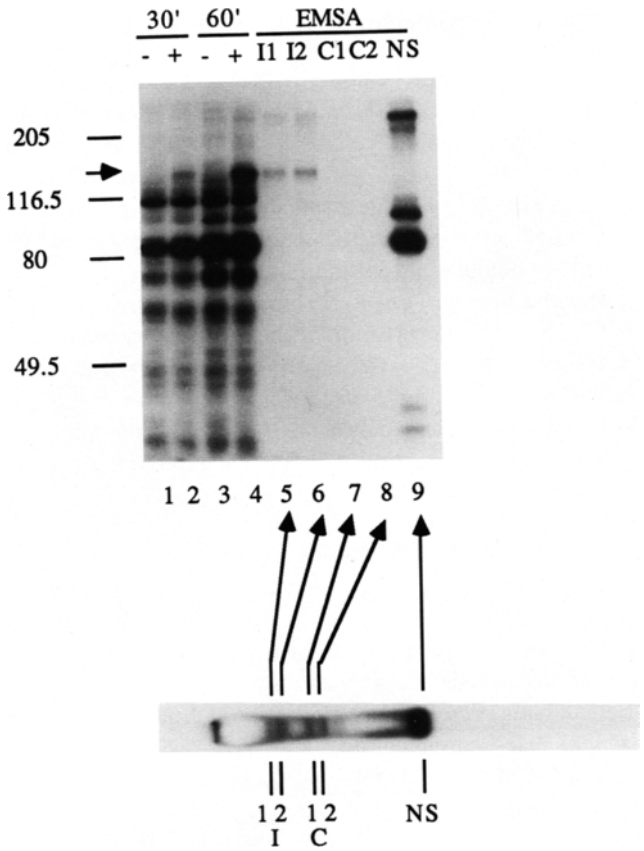


FIG. 4 Detection of HIF-1 DNA-binding subunits by UV cross-linking analysis. Top: In lanes 1–4, nuclear extracts from nonhypoxic (−) (lanes 1 and 3) and hypoxic (+) (lanes 2 and 4) Hep 3B cells were incubated with probe DNA, UV irradiated for 30 (lanes 1 and 2) or 60 min (lanes 3 and 4), and resolved on a 7% (w/v) SDS–polyacrylamide gel in 1× TBE. The arrow indicates a DNA-binding protein that was cross-linked to the probe only in nuclear extract from hypoxic cells. In lanes 5–9, individual retarded bands representing HIF-1 (I1 and I2), constitutive (C1 and C2), and nonspecific (NS) activities, were isolated from a preparative gel shift assay performed with nuclear extract from hypoxic Hep 3B cells (bottom) and analyzed on the same gel. Migration of prestained molecular weight markers (kDa; Bio-Rad) are indicated. [Reproduced with permission from G. L. Wang and G. L. Semenza, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 4304 (1993).]

>190 kDa, which represent the second DNA-binding component of the HIF-1 heterodimer and a cross-linked complex containing both members of the heterodimer, respectively (3, 4).

Protocol 5: Ultraviolet Cross-Linking Analysis

1. Perform a large-scale binding reaction as for gel shift assay, using 50–100 μg of nuclear extract and 10^6 cpm of probe DNA in a 50-μl reaction volume.
2. After incubation on ice for 15 min, resolve the reaction mixture on a 5% (w/v) native polyacrylamide gel. Irradiate the gel on a UV box (320 nm) for 30–60 min.
3. Excise the gel slices corresponding to DNA–protein complexes after exposing the wet polyacrylamide to X-ray film. Elute the proteins by crushing the gel and boiling in Laemmli buffer for 5 min.
4. Perform another large-scale binding reaction as in step 1. After incubation on ice for 15 min, cut off the top of the tubes, cover the top with plastic wrap, and irradiate with the UV source 5 cm from the top of the tubes for 30–60 min at 4°C. Add an equal volume of 2× Laemmli buffer to the reaction and boil for 5 min.
5. Fractionate the samples on an SDS-polyacrylamide gel with prestained molecular weight markers run in a parallel lane. Expose the gel to X-ray film at −80°C with intensifying screens.

Glycerol Gradient Sedimentation Analysis

Glycerol gradient sedimentation analysis can be used to estimate the molecular weight of a protein under native conditions, which may aid in determining if the protein is a monomer, dimer, or higher order complex. In glycerol gradient sedimentation analysis, the formation of a glycerol gradient and sedimentation of an applied sample by ultracentrifugation permits the fractionation of proteins on the basis of their sedimentation coefficients, which in general approximates their molecular weights (19). DNA-binding proteins can be detected by gel shift assay to identify peak fractions containing the DNA-binding activity. The glycerol gradient can be calibrated by sedimentation of protein standards of known molecular weight. The position of molecular weight standards in the gradient can be determined by electrophoresis of aliquots from different fractions on an SDS-polyacrylamide gel and staining with silver nitrate. The molecular weight of the DNA-binding protein can then be estimated by calculating the sedimentation value of the protein from a standard curve generated with molecular weight markers.

In glycerol sedimentation analysis the buffer should have a salt concentration of at least 100 m*M* to prevent nonspecific association among proteins. This problem can also be minimized by using highly purified protein samples. For proteins that are not able to exist as a stable complex in solution, the glycerol gradient sedimentation can be performed in the presence of radioactively labeled probe DNA. Gradient fractions are then analyzed by native polyacrylamide gel electrophoresis to detect the DNA-binding protein (20). The gradient can be fractionated either from the top by pipetting aliquots or from the bottom by punching a hole and collecting drops.

Glycerol gradient sedimentation analysis for HIF-1 was performed in buffer D-0.1 freshly supplemented with DTT and PMSF. The molecular weight of the faster migrating predominant HIF-1 band was estimated to be 200–220 kDa (Fig. 5), con-

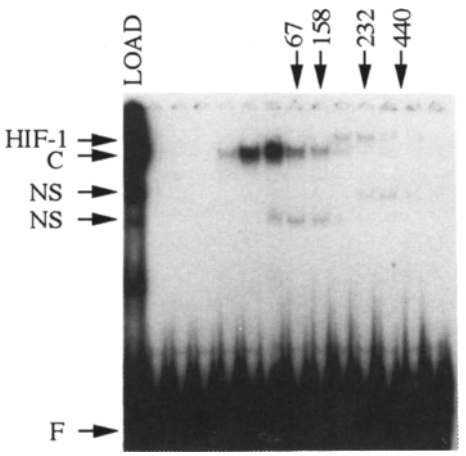


FIG. 5 Glycerol gradient sedimentation analysis of HIF-1. Nuclear extract prepared from Hep 3B cells exposed to 1% O_2 for 4 hr (LOAD) was sedimented through a 10 to 30% (v/v) linear glycerol gradient at 4°C. Aliquots (10 μl) from each fraction were analyzed by gel shift assay. The arrows at the top indicate the peak migration for ferritin (440 kDa), catalase (232 kDa), aldolase (158 kDa), and BSA (67 kDa). [Reproduced with permission from G. L. Wang and G. L. Semenza, *J. Biol. Chem.* **270,** 1230 (1995).]

sistent with the hypothesis that HIF-1 is a heterodimer consisting of the 124- and 94-kDa DNA-binding polypeptides identified by UV cross-linking analysis. Longer exposure of the autoradiograph also revealed that the slower migrating HIF-1 band comigrated with ferritin, which has a molecular mass of 440 kDa, suggesting a tetrameric complex (4).

An alternative method that can also complement the results of glycerol gradient sedimentation is gel-filtration chromatography, which separates proteins on the basis of their physical dimensions or Stokes radius, rather than the molecular weight. For globular proteins, however, the Stokes radius is proportional to molecular weight. In gel-filtration chromatography, the protein sample is layered on top of a gel-filtration column. For initial determination of molecular weight, broad fractionation-range gel matrices should be used. More accurate determination can be achieved with a matrix in which the factor will elute midway through the fractionation range. The molecular weight of the unknown protein can be determined by comparison with the elution profile of protein standards of known molecular weight on the same column.

Protocol 6: Glycerol Gradient Sedimentation Analysis

1. Using a gradient mixer, prepare two identical 12-ml gradients [10–30% (v/v) glycerol] in buffer D-0.1 freshly supplemented with 5 m*M* DTT and 0.4 m*M* PMSF in 12.5-ml ultracentrifuge tubes.

2. Overlay the nuclear protein samples (100 μg) or molecular weight standards (50 μg each) in 0.2 ml of buffer D-0.1 containing 10% (v/v) glycerol onto the gradients.
3. Centrifuge the two glycerol gradients in parallel for 42 hr at 40,000 rpm at 4°C in an SW40 rotor (Beckman, Fullerton, CA).
4. Fractionate the gradients by taking 500-μl aliquots from the top surface of the gradient.
5. Incubate a 20-μl aliquot from each fraction with the probe DNA and analyze by gel shift assay.
6. Determine the positions of molecular weight standards in the fractions by SDS-PAGE and silver staining.
7. Generate a standard curve by plotting the known sedimentation coefficients ($s_{20,w}$) of standards against fraction number. From the slope of the standard curve and the observed sedimentation position of the protein, the s value of the DNA-binding protein is determined and used to estimate the molecular weight.

Purification of DNA-Binding Proteins

One approach to clone the genes encoding a transcription factor is to purify a sufficient amount of protein for microsequence analysis. In addition, the purification itself will reveal important biochemical characteristics of the protein factor, including its molecular composition. The purification procedure usually consists of an initial fractionation step followed by multiple rounds of oligonucleotide–DNA affinity chromatography. DNA-binding activity in the chromatography fractions is detected by gel shift assay. The initial step is used to separate the DNA-binding protein of interest from nucleases and other DNA-binding proteins that also recognize the same DNA sequence. Nucleases, if not removed, will damage the oligonucleotide DNA in the DNA resin. Other DNA-binding proteins that specifically recognize the same DNA sequence cannot be separated from the factor of interest by DNA affinity chromatography. The choice of chromatography method suitable for the purpose of initial purification should be determined empirically for each particular factor. Some protein factors are modified by glycosylation and for these factors wheat germ agglutinin resin is an excellent choice. Many ion-exchange resins, especially fast protein liquid chromatography (FPLC) resins such as Mono-Q and Mono-S, have high binding capacity and can serve to concentrate the sample in addition to purification. Gel-filtration chromatography usually provides insufficient resolution for large-scale purification. For some factors, such as HIF-1, Mg^{2+} is not required for DNA-binding activity. Removal of nucleases in this case is not a consideration because they are mostly inactive in the absence of Mg^{2+}.

The partially purified protein sample is then applied to an oligonucleotide–DNA affinity column. Oligonucleotide–DNA affinity chromatography, developed by Kadonaga and Tjian (21), takes advantage of the sequence-specific DNA-binding properties of transcription factors and the high protein-binding capacity of concatenated

oligonucleotide–DNA resins. To prepare the affinity resin, oligonucleotides containing the binding sequence for a protein factor are concatenated through ligation with T4 DNA ligase and coupled to CNBr-activated Sepharose 4B matrix (Pharmacia, Piscataway, NJ). The two complementary DNA strands should be annealed in equimolar amounts so that no single-stranded oligonucleotide will be attached to the Sepharose. A small amount of [γ-^{32}P]ATP is usually included in the kinase reaction to monitor the concatenation and coupling efficiency. The concatenated oligonucleotide should be at least a 10-mer. Coupling efficiency should be at least 2 nmol of DNA/ml. The concatenation serves both to avoid steric hindrance of proteins binding to the oligonucleotide by gel matrix and to increase the binding affinity by virtue of cooperative protein–protein interactions.

The most critical factor for affinity chromatography is the use of carrier DNA. The type of carrier DNA should be the same as that used in the gel shift assay. The amount of carrier DNA depends on the purity of the protein sample, namely the presence/absence of nonspecific DNA-binding proteins. Insufficient carrier DNA results in nonspecific binding to DNA. Too much carrier DNA, on the other hand, will decrease or prevent specific binding to the column. It is advisable first to determine the optimal ratio of protein to carrier DNA by gel shift assay. However, the optimal protein-to-carrier DNA ratio for affinity columns can range from 1/5 to equal to the optimal ratio for gel shift assays. After incubation with carrier DNA in binding buffer the sample should be centrifuged to remove precipitates before loading on to the affinity column. Either one large column with 2–6 ml of DNA-affinity resin or a series of small (1-ml) columns may be used. Care should be taken, however, to avoid allowing the column to run dry when the larger column is used. In subsequent rounds of affinity chromatography, smaller columns of resin are required due to both the reduction of amount in total protein and the increase in purity of the protein sample. Correspondingly less carrier DNA should be used. The DNA-binding proteins are eluted from the column by increasing the ionic strength of the buffer. The use of a wash step with an intermediate salt concentration may help to remove loosely associated nonspecific DNA-binding proteins from the column. Eluted protein must be dialyzed to decrease the salt concentration before application to the next affinity column.

A chromatography step with a column containing mutant oligonucleotide may be helpful in removing (a) nonspecific DNA-binding proteins that fail to be competed away by carrier DNA and (b) proteins that recognize DNA sequences surrounding the specific protein-binding site in the oligonucleotide. The mutant DNA resin is prepared exactly as for the wild-type DNA affinity resin. A gel shift assay with the mutant oligonucleotide as probe should be performed to optimize binding of nonspecific DNA-binding proteins in the presence of carrier DNA. The mutant oligonucleotide–DNA column may be used after the first round of affinity chromatography. The flowthrough fraction from the mutant column can be loaded directly onto the next wild-type DNA affinity column.

All protein purification procedures should be performed at 4°C. The nonionic de-

tergent Nonidet P-40 (NP-40) should be included at a concentration 0.05–0.1% in all buffers used for DNA affinity chromatography to reduce loss of protein sample by nonspecific adsorption to plastic and glassware. The purity of protein samples from each step can be examined by SDS-PAGE and silver staining.

Protein factors with low binding affinity or rapid dissociation kinetics may be difficult to purify by affinity chromatography. A compromise may have to be made between purity and yield. For the purpose of protein microsequencing it is not necessary to purify a protein to homogeneity as long as the polypeptide subunits of the DNA-binding factor can be clearly identified on an SDS-polyacrylamide gel and excised from a polyvinylidene difluoride (PVDF) membrane after electrophoretic transfer. One method to identify protein components of a factor is to excise the protein–DNA complex from a preparative gel shift assay using highly purified protein preparations (4, 20). The excised gel slice is placed directly onto the stacking gel of an SDS-polyacrylamide gel, incubated with Laemmli buffer *in situ* for 15 min, and fractionated in parallel with an aliquot of the highly purified sample. After visualization by silver staining, polypeptides of the protein–DNA complex can be used to identify the corresponding polypeptide species in the sample of partially purified protein.

The procedure given below was used for the purification of HIF-1 from 120 liters of HeLa S3 cells. The initial fractionation step was DEAE ion-exchange chromatography, which reduced total protein from 3 g to 550 mg with a yield of HIF-1 DNA-binding activity of 72%. Most importantly, this step separated HIF-1 from the constitutive DNA-binding activities (Fig. 6A, lane 1), with the constitutive activities found in the flowthrough in buffer D-01, whereas HIF-1 was retained on the column and eluted with buffer D-0.25. The fractions containing HIF-1 activity were then applied to a wild-type oligonucleotide–DNA affinity column, followed by a mutant DNA affinity column and the second wild-type DNA affinity column. Figure 6 shows the gel shift assay and SDS-PAGE analysis of selected fractions.

Protocol 7: Oligonucleotide–DNA Affinity Chromatography

1. Phosphorylate 20 nmol of each of the two complementary oligonucleotides with 100 U of T4 polynucleotide kinase, 1 m*M* ATP, and 1 μCi of [γ-^{32}P]ATP in a buffer (100 μl) containing 40 m*M* Tris-HCl (pH 8.0), 10 m*M* $MgCl_2$, 0.1 m*M* EDTA by incubating at room temperature for 2 hr.
2. After ethanol precipitation, dissolve the DNA in 86 μl of ligation buffer [50 m*M* Tris-HCl (pH 7.8), 10 m*M* $MgCl_2$], anneal by heating to 95°C for 5 min, and allow to cool to room temperature over a period of 2 hr.
3. Ligate the annealed DNA by adding 2 μl of 10 m*M* ATP, 2 μl of 1 *M* DTT, and 10 U of T4 DNA ligase, and incubate at 4°C for 18 hr.
4. Purify the DNA by phenol–chloroform extraction and three rounds of ethanol

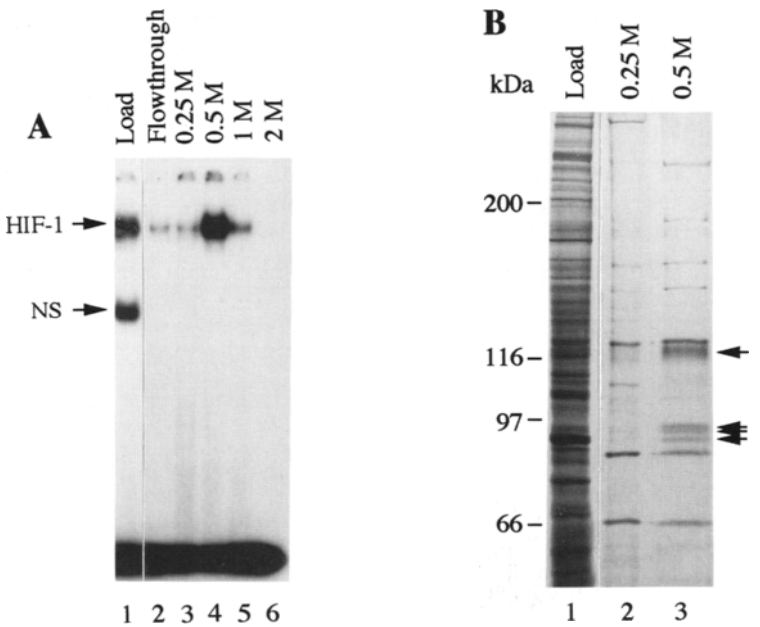


FIG. 6 Purification of HIF-1 by oligonucleotide–DNA affinity column. (A) Gel shift assay of column fractions for HIF-1 DNA-binding activity. Fractions from a DEAE-Sepharose column containing HIF-1 activity were applied to a wild-type DNA-affinity column (Load, lane 1). Aliquots (1 μl) from each fraction were incubated with 5 ng of calf thymus DNA and analyzed by gel shift assay. The position of HIF-1 and nonspecific activity (NS) are indicated. 0.25 M, 0.5 M, 1 M, and 2 M: Fractions eluted with the indicated concentration of KCl in buffer D (lanes 3–6). (B) Analysis of HIF-1 purification by SDS-PAGE. Flowthrough fraction from a mutant oligonucleotide–DNA column (Load, lane 1) and 0.25 *M* KCl and 0.5 *M* KCl fractions from subsequent wild-type DNA-affinity column (lanes 2 and 3) were analyzed. An aliquot of each fraction (5 μg of load or 1 μg of affinity column fractions) was resolved by 6% (w/v) SDS-PAGE and silver stained. The arrows indicate HIF-1 polypeptides. Molecular mass markers are myosin (200 kDa), β-galactosidase (116 kDa), phosphorylase (97 kDa), and bovine serum albumin (BSA) (66 kDa). [Reproduced with permission from G. L. Wang and G. L. Semenza, *J. Biol. Chem.* **270,** 1230 (1995).]

precipitation to remove proteins and ATP. Check the extent of ligation by polyacrylamide gel electrophoresis and autoradiography.

5. Couple the concatenated oligonucleotides to 10 ml of CNBr-activated Sepharose 4B (Pharmacia) according to the instructions of the manufacturer. DNA resin is stored in column storage buffer containing 0.3 *M* NaCl, 10 m*M* Tris-HCl (pH 7.5), 1 m*M* EDTA, and 0.02% (w/v) NaN_3.

6. For affinity chromatography, pack a 6-ml column with wild-type oligonucleotide–DNA affinity resin. Equilibrate the column with 6 column volumes of buffer D-0.1.
7. Mix the protein fractions with carrier DNA and incubate for 15 min on ice. Remove the precipitates by centrifugation at 15,000 *g* for 10 min.
8. Apply the supernatant to the wild-type oligonucleotide–DNA affinity column. Wash the column with 6 column volumes of buffer D-0.1 and 3 column volumes of buffer D-0.25.
9. Elute the bound proteins from the column with 3 column volumes of buffer D-0.5. Dialyze the eluted fraction against 20 vol of buffer D-0.1 with two changes for a total of 4 hr.
10. Mix the elute fraction with carrier DNA and incubate for 15 min on ice. After centrifugation apply the sample to a 2-ml mutant oligonucleotide–DNA column. Collect the flowthrough.
11. Apply the flowthrough fraction directly to the second 2-ml wild-type DNA affinity column.
12. Wash and elute the bound proteins as in steps 5 and 6. Dialyze a small aliquot of the active fraction for gel shift analysis. Freeze the remainder in liquid N_2 and store at −80°C.
13. Regenerate the oligonucleotide–DNA resin by washing twice with column regeneration buffer [2.5 *M* NaCl, 10 m*M* Tris-HCl (pH 7.5), 1 m*M* EDTA] and store in column storage buffer at 4°C.

Concluding Remarks

To further characterize the biochemical nature and analyze the regulatory mechanisms of a transcription factor it is necessary to isolate the cDNA clones encoding the factor. The approach many researchers try first is to screen a cDNA expression library with a DNA probe containing the binding site of the factor to be cloned (22). DNA-binding proteins are often modular in design, and DNA-binding domains are contained in a relatively small region that may be expressed in a functional form as fusion proteins in bacteria. This method is simple and straightforward. In many cases, however, it has proven unsuccessful due to intrinsic limitations. First, efficient DNA-binding activity of many eukaryotic transcription factors requires posttranslational modifications such as phosphorylation and glycosylation whereas proteins expressed in bacteria lack such modifications. Second, many factors bind to DNA as a heteromeric complex, but only one fusion protein can be expressed in any given clone.

The next choice, in many cases, is to purify the DNA-binding protein from tissues or cultured cells. Development of oligonucleotide–DNA affinity purification has greatly simplified the purification procedure and made this approach feasible even for nonbiochemistry laboratories. This method, however, requires large amounts of

starting materials, which is especially difficult for transcription factors that are expressed only in unique cell types. The purified proteins are then subjected to protein microsequence analysis. Oligonucleotides can be designed according to codon preference rules (23) and used as primers for PCR with cDNA to amplify a fragment of cDNA, either within a single tryptic peptide or between two different tryptic peptides. Alternatively, a nested anchoring PCR can be performed using one primer against a lambda (λ) phage arm and another two nested primers whose sequence is based on a tryptic peptide (24). The PCR fragment can then be used as a probe to screen cDNA libraries to isolate a full-length cDNA (25). An advantage of molecular cloning through biochemical purification is that it allows isolation of genes encoding all components of a transcription factor including both DNA-binding and non-DNA-binding subunits.

When cDNA sequences encoding a DNA-binding protein are available, recombinant proteins can be expressed either in bacterial cells or by *in vitro* transcription/translation in reticulocyte or wheat germ lysates. Recombinant proteins can be used to generate antibodies as well as to analyze DNA–protein interactions and regulatory mechanisms of the protein factor. Identity of the cloned factor can be established by (a) "supershift" assay with the antibodies, (b) reconstituting DNA-binding activity with the recombinant proteins, and (c) *in vitro* transcription analysis to demonstrate transcriptional activation mediated by the recombinant protein.

References

1. G. L. Semenza and G. L. Wang, *Mol. Cell. Biol.* **12,** 5447 (1992).
2. G. L. Wang and G. L. Semenza, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 4304 (1993).
3. G. L. Wang and G. L. Semenza, *J. Biol. Chem.* **268,** 21513 (1993).
4. G. L. Wang and G. L. Semenza, *J. Biol. Chem.* **270,** 1230 (1995).
5. G. L. Semenza, *Hematol. Oncol. Clin. N. Am.* **8,** 863 (1994).
6. J. D. Dignam, R. M. Lebovitz, and R. G. Roeder, *Nucleic Acids Res.* **11,** 1475 (1983).
7. N. M. Gough, *Anal. Biochem.* **173,** 93 (1988).
8. E. Schreiber, P. Matthias, M. M. Muller, and W. Schaffer, *Nucleic Acids Res.* **17,** 6419 (1989).
9. N. C. Andrews and D. V. Faller, *Nucleic Acids Res.* **19,** 2499 (1991).
10. M. M. Garner and A. Rezvin, *Nucleic Acids Res.* **9,** 3047 (1981).
11. M. Fried and D. M. Crothers, *Nucleic Acids Res.* **9,** 6505 (1981).
12. D. Galas and A. Schmitz, *Nucleic Acids Res.* **5,** 3157 (1978).
13. A. Maxam and W. Gilbert, *Methods Enzymol.* **65,** 499 (1980).
14. T. D. Tullius and B. A. Dombroski, *Proc. Natl. Acad. Sci. U.S.A.* **83,** 5469 (1986).
15. U. Siebenlist and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **77,** 122 (1980).
16. R. Sturm, T. Baumruker, R. Franza, and W. Herr, *Genes Dev.* **1,** 1147 (1987).
17. L. A. Chodosh, R. W. Carthew, and P. A. Sharp, *Mol. Cell. Biol.* **6,** 4723 (1986).
18. U. K. Laemmli, *Nature (London)* **227,** 680 (1970).
19. R. Martin and B. Ames, *J. Biol. Chem.* **236,** 1372 (1961).

20. D. S. Kessler, S. A. Veals, X.-Y. Fu, and D. E. Levy, *Genes Dev.* **4,** 1753 (1990).
21. J. T. Kadonaga and R. Tjian, *Proc. Natl. Acad. Sci. U.S.A.* **83,** 5889 (1986).
22. H. Singh, J. J. LeBowitz, A. S. Baldwin, and P. A. Sharp, *Cell* **52,** 415 (1988).
23. R. Lathe, *J. Mol. Biol.* **183,** 1 (1985).
24. G. M. Preston, *Methods Mol. Biol.* **15,** 317 (1993).
25. G. L. Wang, B.-H. Jiang, E. A. Rue, and G. L. Semenza, *Proc. Natl. Acad. Sci. U.S.A.* **92,** 5510 (1995).

[17] *In Vitro* Assays for Characterization of RNA–Protein Complexes Involved in Pro-opiomelanocortin mRNA Translation

Stephanie D. Flagg, Corinne M. Spencer, and James H. Eberwine

Introduction

Proper regulation of pro-opiomelanocortin (POMC) gene expression is critical to the mammalian stress response. In response to physical or psychological stress, the hypothalamus secretes vasopressin and corticotropin-releasing hormone (CRH), which in turn stimulate POMC gene expression in the anterior and intermediate lobes of the pituitary. This stimulation results in the increased synthesis of POMC-derived peptides including adrenocorticotropin (ACTH) and β-endorphin. Adrenocorticotropin then stimulates cortisol synthesis and secretion in the adrenal cortex. Glucocorticoids negatively regulate POMC gene expression in the anterior pituitary, while dopamine assumes this role in the intermediate lobe. Several groups have noted a discordance in the time courses over which POMC peptide and mRNA levels are altered in response to alterations in stress hormones (1–3). In pulse–chase experiments, Shiomi *et al.* (4) have noted a 50% increase in the efficiency of POMC peptide synthesis in rats exposed to acute stress at a time when no change in POMC mRNA levels is detected. Additionally, we have observed a change in the polysome distribution of POMC mRNA in response to treatment of AtT-20 cells (a mouse pituitary tumor) with dexamethasone or forskolin, a compound that mimicks the action of CRH (5). These findings suggest that POMC gene expression is regulated at the level of translation.

Translational control refers to the regulation of any of several factors that determine the efficiency of protein synthesis. These factors include the rate at which ribosomal binding and initiation of translation occur, the rate at which each ribosome translocates along a mRNA molecule, and the fraction of mRNA that is actively being translated in polysomes. In addition, several features of a mRNA transcript are thought to influence its translation, including the exposure of the 7-methylguanosine cap, the context of the initiator methionine, and the degree of secondary and tertiary structure present within the mRNA (6).

Control of gene expression at the level of translation allows for protein levels to be rapidly altered without necessitating a change in steady state mRNA transcript levels, thereby bypassing the time delay that is associated with RNA transcription, processing, and transport. The added fact that translation is an amplification process in which one mRNA molecule can be translated into thousands of molecules of pro-

Methods in Molecular Genetics, Volume 8

tein product allows this rapid control mechanism to alter protein levels dramatically. Translational control is thus a mechanism by which cells may rapidly respond to changes in the environment.

One commonly utilized mechanism by which translational control is effected is through the interaction of an RNA-binding protein with a specific RNA sequence or structure. The binding of such proteins can alter the stability of RNA structures formed by intramolecular base pairing, which in turn can influence the efficiency by which a mRNA is translated. Protein–RNA interactions have been shown to impede the translation of many prokaryotic genes (7, 8), and have been demonstrated to play an important role in the expression of several eukaryotic genes as well (9, 10).

Such protein–RNA interactions may play a role in the translational regulation of POMC gene expression. Computer modeling of POMC mRNA predicts the presence of a stem–loop structure of -45 kcal/mol stability within the 5′ coding region (Fig. 1). Data from *in situ* transcription experiments suggest that this predicted stem–loop sequence exists *in situ* in rat pituitary tissue (5). On the basis of these data, we hypothesized that cytoplasmic proteins in rat pituitary and AtT-20 cells might recognize and bind to the stem–loop structure in POMC mRNA to effect its translational control. With the use of *in vitro* assays we have demonstrated the plausibility of this hypothesis. Cytoplasmic extracts from rat pituitary and AtT-20, but not COS cells (a monkey kidney fibroblast cell line), impede the translation of POMC mRNA in an *in vitro* translation assay (11). We have characterized RNA–protein interactions that play a role in POMC translation with the use of RNA gel mobility shift and ultraviolet (UV) cross-linking assays. These techniques are the focus of this chapter and are useful tools in the study of RNA–protein interactions that assume roles in RNA translation as well as those involved in RNA processing or transport.

In Vitro Transcription

DNA sequences encoding the RNA sequence of interest should be cloned into a vector downstream of a strong bacteriophage RNA polymerase promoter such as T7, T3, or SP6. To obtain capped, full-length POMC mRNA that can be readily translated in rabbit reticulocyte lysates, we have used linearized DNA template and T7 RNA polymerase with the protocol for the mMessage mMachine *in vitro* transcription kit (Ambion, Inc.). For the production of large quantities of nonradioactive competitor RNAs, we have achieved optimal success with the use of the appropriate (T7, T3, or SP6) MEGAscript kit using the unmodified procedure as outlined by the manufacturer (Ambion, Inc.) or by using a high concentration of T7 RNA polymerase (1000 U/μl; Epicentre Technologies) as described below:

1. DNA template must be free of RNases and thus should be purified by cesium chloride gradient centrifugation rather than by commercially available column pro-

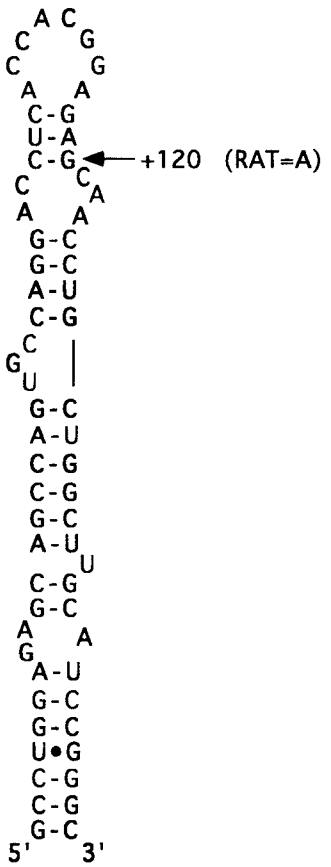


FIG. 1 Computer-predicted secondary structure of mouse POMC RNA stem–loop. The rat POMC stem–loop is identical to that of the mouse, with the exception of a single base pair substitution (A for G at position 120).

cedures, which rely on RNase treatment to remove contaminating RNA. DNA template should be linearized with the restriction enzyme that cuts 3′ to the insert so as to minimize the amount of vector sequences transcribed and to ensure that all RNA transcripts synthesized will be of the same length. Restriction enzyme-cut plasmid DNA is extracted with equal volumes of phenol–chloroform and precipitated with a $\frac{1}{10}$ vol of 3 *M* sodium acetate, pH 5.2, and 2.5 vol of ethanol. The pellet should be resuspended in Tris–EDTA (TE) buffer that has been filter sterilized. Because they stick to the nitrocellulose membrane, RNases can be removed from solutions during the filtration process.

2. Set up *in vitro* transcription reactions in RNase-free microfuge tubes. All buff-

ers are made RNase free by sterile filtration or by treatment of buffers that do not contain primary amines with diethyl pyrocarbonate (DEPC).

H_2O	43 μl
Transcription buffer (5×): 200 m*M* Tris (pH 7.5), 35 m*M* $MgCl_2$, 50 m*M* NaCl, 10 m*M* spermidine	20 μl
Dithiothreitol (DTT, 100 m*M*)	10 μl
RNasin (40 U; Promega, Madison, WI)	1 μl
ATP (10 m*M*)	5 μl
CTP (10 m*M*)	5 μl
GTP (10 m*M*)	5 μl
UTP (10 m*M*)	5 μl
Linearized template DNA	5 μl = 1 μg
T7 RNA polymerase (1000 U; Epicentre Technologies)	1 μl

3. Incubate the reactions for 90 min at 37°C.
4. Remove the DNA template by addition of 2 μl of RNase-free DNase I (10 U/μl; BMB) and continue the incubation at 37°C for 30 min.
5. Extract the RNA with phenol–chloroform, and ethanol precipitate.
6. Resuspend the RNA in 10 μl of sterile-filtered H_2O and drop dialyze the sample on nitrocellulose filters (Cat. No. VSWP01300; Millipore, Bedford, MA) for 1–4 hr to remove free triphosphates. Repeat the phenol–chloroform extraction and ethanol precipitation.
7. Obtain a spectrophotometric reading of sample at 260 nm to quantitate the RNA.

Radiolabeled riboprobe can be synthesized by the above protocol if 30 μCi of [α-^{32}P]CTP is added to the reaction tubes and the final concentration of unlabeled CTP is reduced to 0.5 μ*M*. T7 RNA polymerase of a lower specific activity (GIBCO-BRL, Gaithersburg, MD) may be employed in riboprobe synthesis if a smaller quantity of RNA needs to be produced. Additionally, free triphosphates can be removed with two sequential ethanol precipitations. Counts per minute are determined by scintillation counting.

Cytoplasmic Extract Preparation

S100 Cytoplasmic extracts are prepared as described (12). Briefly:

1. Grow cells in culture to 5×10^5 and harvest by centrifugation for 10 min at 2000 rpm in a Sorvall HG4L rotor. All subsequent steps should be performed at 4°C.
2. Resuspend the pellet in 5 vol of phosphate-buffered saline.

3. Centrifuge as in step 1.
4. Resuspend in five packed cell pellet volumes of buffer A [10 m*M* *N*-2-hydroxyethylpiperazine-*N*-2-ethanesulfonic acid (HEPES, pH 7.9 at 4°C), 1.5 m*M* $MgCl_2$, 10 m*M* KCl, 0.5 m*M* DTT] and incubate for 10 min on ice.
5. Centrifuge as in step 1.
6. Resuspend the pellet in 2 vol of buffer A, and lyse cells with at least 10 strokes of an all-glass Kontes Dunce homogenizer, type B pestle. The homogenate should be checked at this point for the efficiency of cell lysis. For effective lysis of AtT-20 cells, an additional one or two rounds of homogenization are required.
7. Centrifuge as in step 1 to pellet nuclei.
8. Mix the resulting supernatant with 0.11 vol of buffer B [300 m*M* HEPES (pH 7.9 at 4°C), 1.4 *M* KCl, 30 m*M* $MgCl_2$] and centrifuge for 60 min at 100,000 g_{av}.
9. Dialyze the supernatant for 5 to 8 hr against 20 vol of buffer D [20 m*M* HEPES (pH 7.9 at 4°C), 100 m*M* KCl, 0.2 m*M* EDTA, 20% (v/v) glycerol, 0.1 *M* KCl, 0.5 m*M* phenylmethylsulfonyl fluoride (PMSF), and 0.5 m*M* DTT]. Protein determinations can be made by the use of the Bradford assay (Bio-Rad, Richmond, CA). Aliquot cytoplasmic extracts and store at −80°C.

Note: Dithiothreitol and PMSF should be added to buffers at the time of use.

In Vitro Translations

We have used a rabbit reticulocyte lysate system (Amersham, Arlington Heights, IL) to demonstrate the ability of protein–RNA interactions to regulate the *in vitro* translation of POMC mRNA (Fig. 2). Reaction conditions are outlined by the manufacturer. Briefly, a 10-μl reaction is set up consisting of 200 ng of the full-length mRNA of interest (or control RNA such as brome mosaic virus; Amersham), 10 μCi of L-[^{35}S]methionine (translation grade; Amersham), 4 μl of cytoplasmic extract or Dignam buffer D (see cytoplasmic extract preparation), and 4 μl of nuclease-treated, message-dependent rabbit reticulocyte lysate. Translation reactions are incubated for 1 hr (time course should be empirically determined) at 30°C and then mixed with 2× sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) sample buffer, boiled for 5 min, and electrophoresed on the percentage denaturing polyacrylamide gels that will achieve effective separation of the translated protein product. After treatment with Entensify universal autoradiography enhancer (NEN-Dupont, Boston, MA), gels are dried and exposed to film for 3–7 days at −80°C. Competitor RNAs can be included in the translation reactions to evaluate the ability of an RNA sequence or structure to compete with the full-length mRNA for interaction with cytoplasmic protein and attenuate the ability of cytoplasmic extract to alter the translation of the full-length mRNA. Certain issues must be kept in mind in performing these experiments and in interpreting the resulting data. High concentrations of RNA inhibit protein synthesis in rabbit reticulocyte lysates, perhaps by sequestering initia-

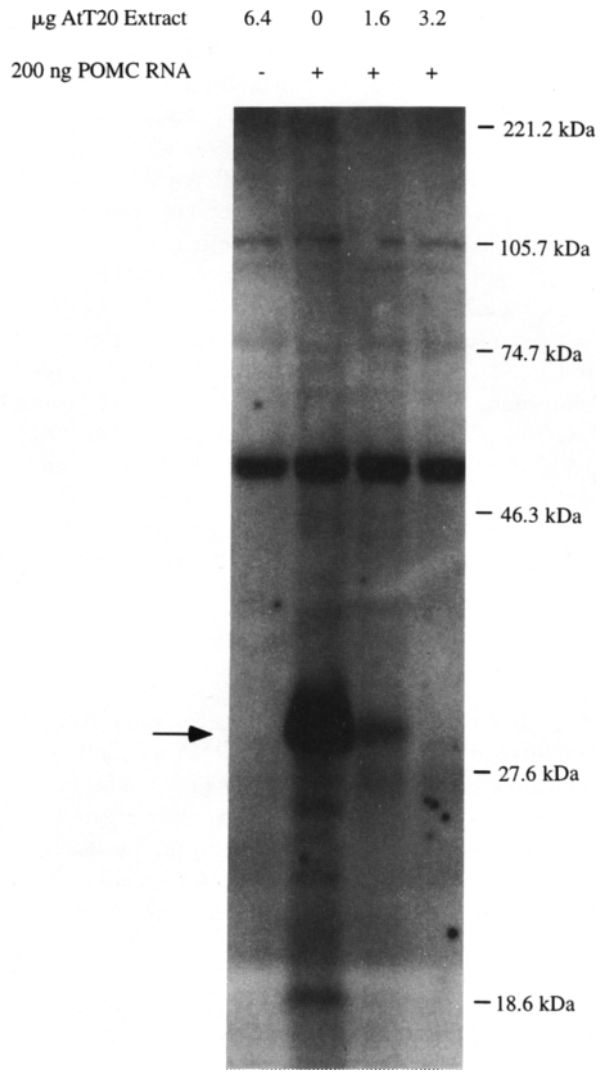


FIG. 2 The translation of POMC RNA *in vitro* is inhibited by AtT-20 cytoplasmic extract. Two hundred nanograms of capped full-length POMC mRNA was translated in a rabbit reticulocyte lysate system in the presence of increasing amounts of AtT-20 cytoplasmic extract (lanes 2–4). The first lane is a control reaction in which no RNA was added to the reticulocyte lysate. The arrow indicates the position of the POMC peptide as determined by immune precipitation (data not shown).

tion and elongation factors (13). Furthermore, the addition of high concentrations (>1 μg/ml) of double-stranded RNA to reticulocyte lysate can lead to a generalized translational arrest (14, 15) by blocking the activation of a double stranded RNA-dependent protein kinase that phosphorylates initiation factor eIF-2. Thus, proper controls are required to evaluate data from these experiments.

RNA Gel Mobility Shift Assay

RNA gel mobility shift assays are performed essentially as described (16).

1. Set up RNA–protein binding reactions in RNase-free microfuge tubes with 5–25 μg of cytoplasmic extracts (or lesser amounts of purified protein), 40 U of RNasin (Promega), and 50,000 cpm of labeled riboprobe in 10 m*M* HEPES (pH 7.6), 3 m*M* $MgCl_2$, 40 m*M* KCl, 5% (v/v) glycerol, and 1 m*M* DTT in a 20-μl reaction volume. One reaction should be set up with riboprobe in the absence of protein extract to demonstrate the mobility of unretarded RNA transcript.
2. Incubate binding reactions at room temperature for 20 min to allow proteins to interact with the riboprobe. Salt concentrations, incubation time, and temperature may need to be modified to optimize binding.
3. Add 1 μl of heparin (100 mg/ml; Sigma, St. Louis, MO) and continue the incubation at room temperature for 10 min. The addition of heparin should decrease the binding of proteins that interact nonspecifically with the phosphate backbone of the riboprobe.
4. Electrophoresis is performed as described (Ref. 17, with modifications). Load binding reactions directly onto a 5% (w/v) nondenaturing polyacrylamide (60:1, acrylamide–bisacrylamide) gel that has been preelectrophoresed for 30–60 min at 8 V/cm in 45 m*M* Tris–borate, 1 m*M* EDTA. Electrophorese the gel for 2 to 2.5 hr at 8 V/cm.
5. Dry the gel and autoradiograph at room temperature for 12 to 24 hr.

The RNA gel mobility shift assay is a rapid method that can be used to demonstrate whether a purified protein or proteins within an extract preparation can bind specifically to an RNA fragment. Control reactions performed in the absence of protein demonstrate the unretarded migration of riboprobe within the gel. The addition of cytoplasmic extract containing proteins that recognize the riboprobe results in the formation of riboprobe–protein complexes that are distinguishable by their slower migration through the gel (Fig. 3).

Competition experiments can be performed by the inclusion of a 10- to 1000-fold molar excess of unlabeled competitor RNAs to the binding reactions (step 1). Unlabeled competitor RNAs that specifically compete with the riboprobe for protein binding will disrupt the formation of riboprobe–protein complexes (less riboprobe

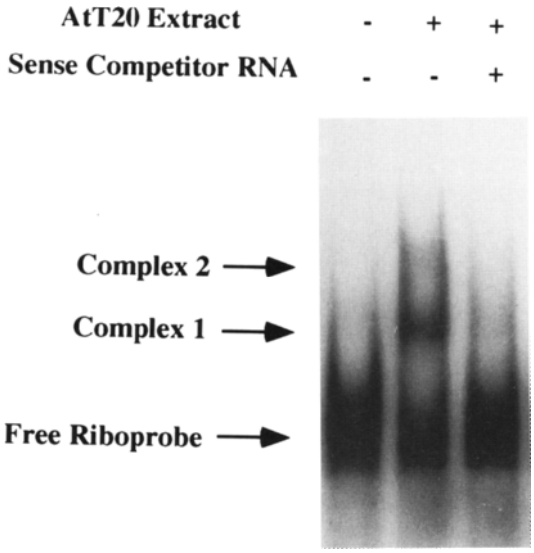


FIG. 3 RNA gel mobility shift assay with a ^{32}P-labeled POMC stem–loop RNA in the absence (lane 1) or presence (lane 2) of 5 μg of cytoplasmic AtT-20 extract. Two RNA protein complexes are observed on the addition of AtT-20 cytoplasmic extract. These complexes are competed away by the addition of 100 ng of unlabeled POMC stem–loop RNA (lane 3).

will be shifted) to a greater degree than will cold nonspecific RNAs. Nonspecific RNAs that are commonly used in this assay include RNAs of the same length as the riboprobe that are synthesized from linearized vector DNA templates, tRNA, and poly(rI-rC). Proteinase K should be included in a control reaction to demonstrate unequivocally that the retarded migration of riboprobe is due to the interaction of protein with the RNA.

The RNA gel mobility shift assay does not provide information on the molecular weight of the proteins that interact with the riboprobe, and does not indicate the number of proteins that contribute to the riboprobe–protein complexes formed. For this type of information, a UV cross-linking assay may be performed.

Ultraviolet Cross-Linking Assay

Binding reactions are performed as described in steps 1 to 3 for the gel mobility shift assay, except that reactions are set up in the caps of microfuge tubes. Competition experiments can be performed by adding unlabeled competitor RNAs in the binding reactions as was described for the gel mobility shift assay. After the addition of heparin:

1. Binding reactions are exposed to UV light (Sylvania G15T8; GTE Sylvania, Inc.) for 10 min at a distance of 4.5 cm (18). A time course should be performed with the UV source at various distances to optimize conditions.
2. Incubate the reactions at room temperature with 5 μg of RNase A to digest RNA not protected by covalent linkage to protein. Some protocols include a variety of RNases (T1, T2, and V1) in this incubation. RNase A cleaves after pyrimidine residues; RNase T1 cuts after guanosine residues; T2 cleaves after any residue; and RNase VI cleaves double-stranded RNA. The addition of these enzymes to the reactions may be important for the complete digestion of highly structured RNAs.
3. Add an equal volume of 2× SDS sample buffer to the reaction tubes, and boil the samples for 5 min before separating the proteins by electrophoresis on 8–10% (w/v) SDS-polyacrylamide gels according to a standard protocol (19).
4. Dry the gel and visualize complexes by autoradiography.

Ultraviolet treatment of the binding reactions covalently cross-links to RNA proteins that are in direct association with it. RNase treatment of the RNA–protein complexes degrades all riboprobe that is not protected by direct cross-links to protein. When these samples are run on denaturing polyacrylamide gels, the proteins that are cross-linked to small fragments of riboprobe are separated by virtue of their molecular weight and can be visualized by autoradiography. This assay allows for the estimation of molecular weight of the RNA-binding proteins and gives information as to the number of proteins that can bind directly to the riboprobe (Fig. 4). This technique, however, has several limitations. A single band on a UV cross-linking experiment does not correspond necessarily to one unique RNA-binding protein because one band may represent either a complex of proteins (if two or more proteins become cross-linked by the UV treatment) or the partial degradation product of an RNA-binding protein. In addition, a protein that binds indirectly to the RNA via a protein–protein interaction may not become covalently cross-linked, and therefore would not be visualized on autoradiography.

In Vitro Mutagenesis of RNA Recognition Site

The RNA-binding site recognized by extract proteins can be further delineated by site-directed mutagenesis of the RNA fragment. Mutant RNAs can be synthesized from mutagenized DNA templates and used as riboprobes or competitors in gel mobility shift and UV cross-linking assays. Additionally, mutant RNAs can be studied in *in vitro* translation assays and *in vivo*, after transfection of cells with reporter gene constructs. From these types of experiments, one can ascertain the degree to which sequence and structure of an RNA contribute to the RNA–protein interactions being studied.

We have created mutants in the stem–loop structure of POMC mRNA with the use

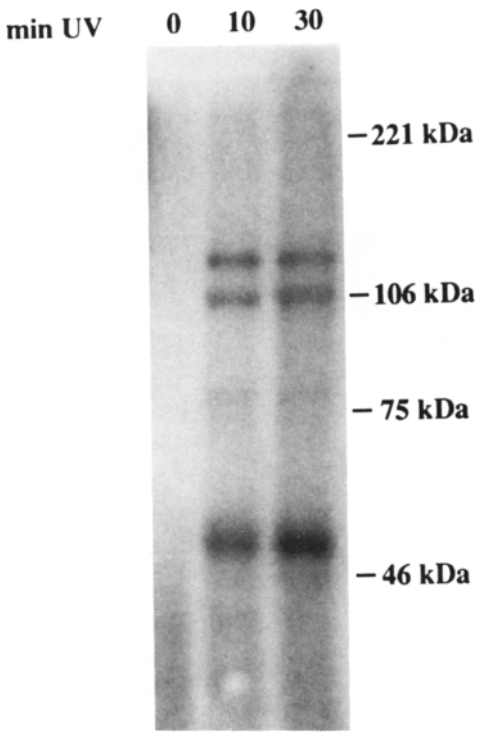


FIG. 4 Ultraviolet cross-linking experiment with ^{32}P-labeled POMC RNA stem–loop and 18.7 μg of AtT-20 cytoplasmic extracts. Samples were exposed to UV light for 0, 10, or 30 minutes as indicated (lanes 1–3). Four RNA–protein complexes are observed.

of the Altered Sites *in vitro* mutagenesis system (Promega). This system requires subcloning the DNA encoding the RNA of interest into the multiple cloning site of the pSELECT-1 phagemid-plasmid, which contains SP6 and T7 promoters. Single-stranded DNA (ssDNA) is produced by infection of host bacterial cells with helper phage. This ssDNA is annealed to a 5′-phosphorylated mutant oligonucleotide and the synthesis of the mutant strand is completed with the addition of T4 DNA polymerase and T4 DNA ligase. The percentage of mutants that is created is optimized by performing a first round of transformation in BMH 71-18 *MutS*, a repair-defective *Escherichia coli* strain. A second round of transformation is performed in JM109 cells in order to ensure segregation of wild-type and mutant clones. Clones are sequenced to confirm the identity of mutants. The pSELECT-1 plasmid contains a gene for tetracycline resistance and a nonfunctional ampicillin resistance gene. This feature, and the provision of an oligonucleotide that when included in the annealing reaction results in the repair of the ampicillin resistance gene, allows for

ampicillin selection for mutant strand synthesis and increases the percentages of mutants obtained.

A newer version of this system (Altered Sites II *in vitro* mutagenesis system; Promega) is currently available that boasts as advantages (a) the inclusion of a protocol for mutagenesis of double-stranded DNA and (b) the ability to perform multiple rounds of mutagenesis without resubcloning by alternating the knockout and repair of the tetracycline and ampicillin genes. We have found the ES1301 *MutS* repair-defective *E. coli* strain that currently must be used for multiple rounds of mutagenesis to be difficult to transform. Competent ES1301 *MutS* cells soon may be available from Promega.

Summary

Although we have chosen to illustrate the utility of these techniques in the characterization of RNA-binding proteins that are involved in translation, these techniques are also useful in characterizing RNA-binding proteins that have other functional roles within the cell, such as the processing and transport of RNA. In particular, these techniques are being utilized to characterize RNA-binding proteins involved in the subcellular distribution of RNAs as well as those thought to be involved in regulating the course of various neurological diseases. Once the specificity of an RNA–protein interaction has been characterized by these methods, one can isolate the protein(s) involved by the use of a direct cloning approach via a Northwestern hybridization screen (20, 21) or by biochemical purification techniques that take advantage of RNA affinity chromatography (22, 23).

Acknowledgments

J.H.E. is an established investigator of the American Heart Association. S.D.F. is supported by NIGMS and is a Dupont-Merck Scholar.

References

1. C. L. C. Chen, F. T. Dionne, and J. L. Roberts, *Proc. Natl. Acad. Sci. U.S.A.* **80,** 2211 (1983).
2. N. C. Birnberg, J.-C. Lissitzky, M. Hinman, and E. Herbert, *Proc. Natl. Acad. Sci. U.S.A.* **80,** 6982 (1983).
3. G. S. Wand, V. May, and B. A. Eipper, *Endocrinology* **123,** 1153 (1988).
4. H. Shiomi, S. J. Watson, J. E. Kelsey, and H. Akil, *Endocrinology* **119,** 1793 (1986).
5. J. H. Eberwine, C. Spencer, D. Newell, and A. R. Hoffman, *Microsc. Res. Tech.* **25,** 19 (1993).

6. M. Kozak, *J. Biol. Chem.* **266,** 19867 (1991).
7. J. E. G. McCarthy and C. Gualerzi, *Trends Genet.* **6,** 78 (1990).
8. L. Gold, *Annu. Rev. Biochem.* **57,** 199 (1988).
9. M. Kozak, *Annu. Rev. Biochem.* **8,** 197 (1992).
10. R. D. Klausner, T. A. Rouault, and J. B. Harford, *Cell* **72,** 19 (1993).
11. C. M. Spencer and J. H. Eberwine, submitted (1994).
12. J. D. Dignam, R. M. Lebovitz, and R. G. Roeder, *Nucleic Acids Res.* **11,** 1475 (1983).
13. L. P. Ovchinnikov, A. S. Spirin, B. Erni, and T. Staehelin, *FEBS Lett.* **88,** 21 (1978).
14. T. Hunter, T. Hunt, R. J. Jackson, and H. D. Robertson, *J. Biol. Chem.* **250,** 409 (1975).
15. P. J. Farrell, G. C. Sen, M. F. Dubois, L. Ratner, E. Slattery, and P. Lengyel, *Proc. Natl. Acad. Sci. U.S.A.* **75,** 5893 (1978).
16. E. A. Leibold and H. N. Munro, *Proc. Natl. Acad. Sci. U.S.A.* **85,** 2171 (1988).
17. M. M. Konarska and P. A. Sharp, *Cell* **46,** 845 (1986).
18. S. Piñol-Roma, S. A. Adam, Y. D. Choi, and G. Dreyfuss, *in* "Methods in Enzymology," Vol. 180, p. 410. Academic Press, San Diego, California, 1989.
19. U. K. Laemmli, *Nature (London)* **227,** 680 (1970).
20. A. Gatignol, A. Buckler-White, B. Berkhout, and K.- T. Jeang, *Science* **251,** 1597 (1991).
21. A. Gatignol and K.-T. Jeang, this series, Vol. 4, p. 18.
22. T. A. Rouault, M. W. Hentze, D. J. Haile, J. B. Harford, and R. D. Klausner, *Proc. Natl. Acad. Sci. U.S.A.* **86,** 5768 (1989).
23. B. Neupert, N. A. Thompson, C. Meyer, and L. C. Kuhn, *Nucleic Acids Res.* **18,** 51 (1990).

Section IV

RNA Editing, Ribozymes, Antisense RNA

This Page Intentionally Left Blank

[18] Mammalian RNA Editing

Lawrence Chan and BaBie Teng

Apolipoprotein (apo) B mRNA is a novel genetic phenomenon first described in 1987 (1, 2). It consists of the conversion of a specific cytidine residue (number 6666) in apoB mRNA to a uridine residue. This posttranscriptional modification (3) changes codon 2153 from CAA, encoding a glutamine residue, to UAA, a termination codon. The normal (genomic) termination codon for apoB mRNA transcript is located some 7 kb downstream. Therefore, the unedited apoB mRNA encodes a protein, known as apoB-100, that contains 4536 amino acid residues, whereas the edited mRNA encodes a much shorter protein, called apoB-48, that contains only 2152 residues. Not only do apoB-100 and apoB-48 differ in size, they also differ drastically in their physicochemical properties. ApoB-100 is an essential component of the plasma lipoproteins called very low-density lipoproteins, intermediate-density lipoproteins, and low-density lipoproteins, and a special lipoprotein called lipoprotein (a). ApoB-48, in contrast, is an obligatory component of the intestinally derived lipoproteins known as chylomicrons and their metabolic product, chylomicron remnants. The physiological implications of apoB mRNA editing and the pathophysiology of the various apoB-containing lipoproteins have been reviewed (4–8).

From a genetic standpoint, apoB mRNA editing is important because it represents an exception to the central dogma of molecular biology. Here is an instance of one gene, two polypeptides. It was the first example of RNA editing described in mammals. The other well-characterized example is the editing of the mRNA of the subunits for high-affinity kainate receptors in brain (6, 9, 10).

ApoB mRNA editing occurs in the small intestine in all mammals examined. In the liver of most mammals, including humans, apoB mRNA remains unedited. However, in a few species, notably rats and mice, editing also occurs in the liver. A sequence-specific hydrolytic deamination of cytidine-6666 appears to mediate the editing reaction. Editing requires a multiprotein component complex called an editosome (11). A catalytic component of the editing machinery has been cloned (12–15). This component has been named apobec-1. Human apobec-1 is a 28-kDa protein that exists as a homodimer (15). The *APOBEC1* locus has been mapped to chromosome band 12p13.1-p13.2 (15). Apobec-1 is a cytidine deaminase that requires zinc for activity (16, 17). By itself, however, it does not have the competence to edit apoB mRNA *in vitro* (12). Editing requires the participation of complementation factor(s) that have not been fully characterized.

ApoB mRNA editing and kainate receptor subunit editing were first discovered when discrepancies between cDNA and genomic DNA sequences were noted and subsequently fully characterized. It is likely that other instances of RNA editing have

Methods in Molecular Genetics, Volume 8

been missed because discrepancies in DNA sequences were assumed to be sequencing errors. Other likely examples of RNA editing in mammals are now beginning to be described (18, 19). We believe that the methods described in the chapter on the characterization of apoB mRNA editing can be readily adapted for studying RNA editing in other systems.

Methods

In characterizing apoB mRNA editing, one needs to address two major questions: (a) within a sample of RNA (with apoB mRNA as one of its components), what proportion of the apoB mRNA is in the edited form, and (b) what is the editing activity in a particular sample (of tissue extract or cloned editing components expressed *in vitro*). The experimental approaches used to answer these questions are described below (Assays for Degree of Editing in RNA Sample and *In Vitro* Assays for Apolipoprotein B mRNA Editing Activity).

Assays for Degree of Editing in RNA Sample

ApoB mRNA editing consists of a C→U conversion involving C-6666, the first base of the codon CAA encoding Glu-2153, to UAA, a stop codon (Fig. 1). A number of

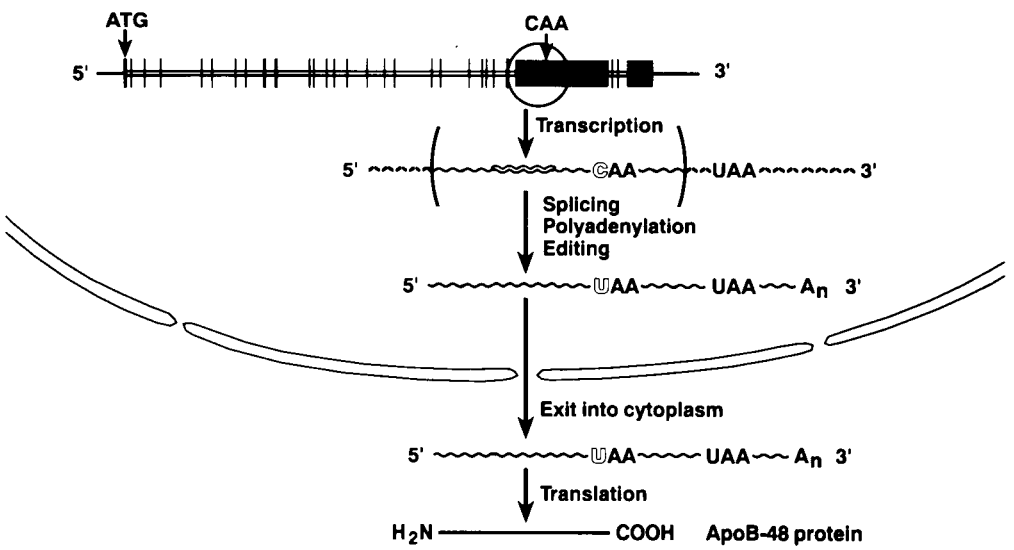


FIG. 1 Schematic model of apoB mRNA editing. [Reprinted from Chan (4) with permission.]

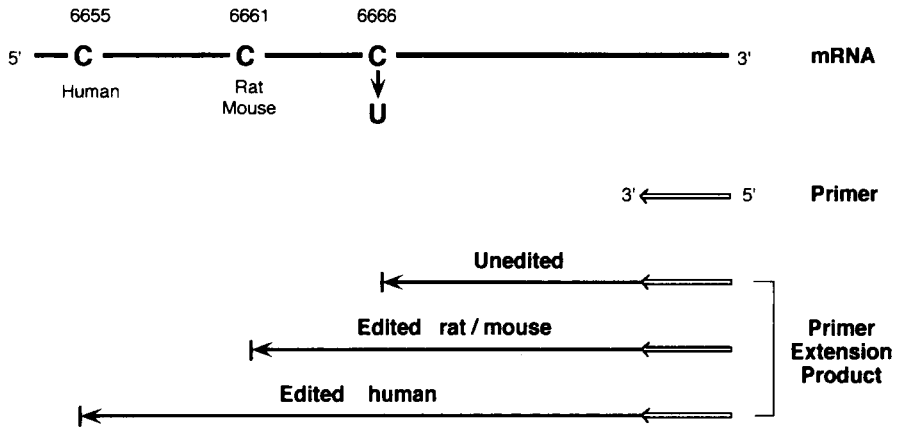


FIG. 2 Principle of the primer extension assay.

assays have been developed to quantify the proportion of C-6666 that has been converted to U-6666. They are based on one of two techniques for detecting single nucleotide substitutions in an RNA sample: (a) a primer extension method and (b) an oligonucleotide colony hybridization detection method.

The primer extension assay was first developed by Driscoll *et al.* (20). It has been widely used and modified by different laboratories (21–26). The principle of the method is illustrated in Fig. 2. In the edited region of apoB mRNA, there are no C's in the immediate vicinity of C-6666. The next upstream C is C-6655 for human and C-6661 for rat and mouse apoB mRNA. An oligonucleotide annealed to sequences downstream to C-6666 is used as a primer for synthesizing a cDNA complementary to the edited region by reverse transcription. This primer extension reaction is performed using reverse transcriptase in the presence of high concentrations of dideoxy-GTP. For an unedited apoB mRNA, the primer extension product stops at C-6666, the first C encountered in the reaction. For an edited mRNA where a U replaces C-6666, the primer extension product goes past U-6666 and stops at the next upstream C, at position 6661 for rat and mouse, and at position 6655 for human apoB mRNA. Use of an end-labeled primer will allow the primer extension products to be analyzed on a sequencing gel; the different-sized bands represent the products from unedited or edited apoB mRNA. One can estimate the proportion of edited apoB mRNA from the ratio of the intensity of these bands.

Primer extension can be performed directly on a sample of RNA if the amount of apoB mRNA within the sample is moderately high. If its concentration is low and cannot be used effectively as a template for direct primer extension, the apoB mRNA-specific sequence can be amplified by reverse transcription (RT) and the polymerase chain reaction (PCR) before the primer extension assay is performed on the PCR product.

In the last few years, primer extension assays (directly or following PCR) for quantitation of apoB mRNA editing have been optimized. They are the method of choice for most situations. If the highest degree of accuracy is required, however, a method involving oligonucleotide colony hybridization (27) should be used instead. In this method, the apoB mRNA-specific sequences are amplified by reverse transcriptase-PCR, subcloned into plasmids; plasmids containing unedited or edited sequences are then identified by oligonucleotide colony hybridization. The method is laborious and takes several days to complete and is not recommended for routine use. However, in addition to its being an extremely sensitive and accurate assay, this method is also suitable for the detection of multiple nucleotide changes in an RNA sample. For example, it has been used for detecting the editing of multiple, closely clustered C's in mutated apoB mRNA substances (28).

Direct RNA Primer Extension Assay

The following protocols are for measuring the degree of RNA editing in rat and human apoB mRNAs, the species that have been studied most extensively. We will also provide the primer sequences for mouse, pig, and rabbit. However, as long as the sequences immediately flanking the edited C's are known, the primer sequence can be devised for any species. The only requirement is that there be no other C's between the edited C and the last base (i.e., 3′ end) of the oligonucleotide primer.

Materials

Oligonucleotide primers: Primers of 35 nucleotides (nt 6674–6708) complementary to apoB mRNA of different species (26, 29) are listed as follows:

Human:	5′ AATCATGTAAATCATAACTATCTTTAATATACTGA
Pig:	5′ AATCATGTAAATCATAATTATCTTTAATATACTGA
Rabbit:	5′ AATCATGTAGATCAAAATTATCTTTAATATACTGA
Rat:	5′ AGTCCTGTGCATCATAATTATCTCTAATATACTGA
Mouse:	5′ AGTCATGTGGATCATAATTATCTTTAATATACTGA
Chicken:*	5′ TGTCAAACTGATCGTAATTCTCTTTGATGTACTGC

Annealing buffer: 50 m*M* piperazine-*N,N′*-bis(2-ethanesulfonic acid) (PIPES, pH 6.4), 0.2 *M* NaCl

Extension buffer: 50 m*M* Tris-HCl (pH 8.2), 6 m*M* $MgCl_2$, 10 m*M* dithiothreitol (DTT)

1. In a 500-μl microcentrifuge tube, total RNA (5 μg) is denatured at 75° C for 10 min and annealed to ~20 pg of ^{32}P end-labeled oligonucleotide primer at 45° C

* Note that chicken apoB mRNA is not edited (26).

overnight in annealing buffer with a final volume of 10 μl. The reaction mixture is overlaid with mineral oil (Sigma, St. Louis, MO) to prevent evaporation (30).

2. The following day, the annealed products are transferred to a new tube and precipitated in ethanol. Recovery of annealed products should be quantitative and can be monitored by a Geiger counter.

3. After ethanol precipitation, the primers annealed to the respective mRNAs are extended in extension buffer at 45° C for 60 min in the presence of 500 μM each of dATP, dCTP, dTTP, and dideoxy-GTP (Pharmacia LKB Technology, Inc., Alameda, CA) and 10 units of Moloney murine leukemia virus reverse transcriptase (MMLV-RT) (Bethesda Research Laboratories, Gaithersburg, MD).

4. The primer extension products are fractionated by electrophoresis on 8% (w/v) polyacrylamide–urea gels followed by autoradiography at −70° C. Under these conditions, the primer extension product for the unedited apoB mRNA (nucleotide 6666 = C) contains 43 bases; for the edited mRNAs, the product varies in length depending on the species under study (taking nucleotide 6666 as the canonical C that is edited to U, the next upstream C for human, pig, and rabbit is at nucleotide 6655; for rat and mouse it is at nucleotide 6661, and for chicken, at nucleotide 6660). The proportion of edited to unedited apoB mRNA is determined by the ratio of intensity of the two primer extension bands determined by the Phosphoimager SF (Molecular Dynamics, Sunnyvale, CA).

Reverse Transcriptase Polymerase Chain Reaction Primer Extension

1. Total cellular RNA (1 to 5 μg) is treated with 10 units of DNase RQ1 (Promega, Madison, WI) to remove traces of genomic DNA contamination. The sample is then extracted with phenol–chloroform and chloroform, and precipitated in ethanol. The DNA-free RNA is resuspended in H_2O at a concentration of 0.5 μg/μl (30) for further analysis.

2. Reverse transcriptase polymerase chain reaction (RT-PCR) is carried out with 500 ng of total RNA by using the rTth reverse transcriptase RNA PCR kit (Perkin-Elmer Branchburg, NJ) or by synthesizing first-strand cDNA with MMLV-RTase followed by the PCR amplification protocol above. The following is the protocol for the rTth RNA PCR method that we routinely use. Total cellular RNA (500 ng) is annealed to 100 pmol of apoB-specific downstream primer flanking the edited base in rTth buffer [10 mM Tris-HCl (pH 8.3), 90 mM KCl]. Reverse transcription is carried out at 70° C for 10 min in a 20-μl final volume in 1 mM $MnCl_2$, 200 μM dNTPs, and 5 U of rTth DNA polymerase. This is followed by PCR amplification, which is started by the addition of 8 μl of chelating buffer [50% (v/v) glycerol, 100 mM Tris-HCl (pH 8.3), 1 M KCl, 7.5 mM ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA), 0.5% (v/v) Tween 20], 6 μl of 25 mM $MgCl_2$, 100 pmol of upstream primer, and water to a final volume of 100 μl. The PCR is carried out for 40 cycles at 95° C for 30 sec, 55° C for 30 sec, and extension at 72° C for 90 sec, with a final 10-min extension at 72° C. Control PCR amplification

should be carried out on every sample using rTth but leaving the reaction on ice during reverse transcription. The RT-PCR products are then purified over a Qiagen column (Qiagen, Chatworth, CA) and approximately 200 pg of amplified DNA is used for primer extension analysis as follows.

3. In a 500-μl microcentrifuge tube, approximately 200 pg of amplified DNA is denatured at 95° C for 5 min and annealed at 37° C for 10 min to a ^{32}P end-labeled primer in 7 μl of 40 m*M* Tris-HCl (pH 8.3), 20 m*M* $MgCl_2$, and 50 m*M* NaCl. Subsequently, the reaction mixture is extended in the presence of 10 m*M* DTT, 0.1 m*M* dATP, dCTP, and dTTP, 0.25 m*M* dideoxy-GTP, and 2 U of T7 DNA polymerase (U.S. Biochemical, Cleveland, OH) in a final volume of 10 μl. The reaction mixture is incubated at 42° C for 10 min and the reaction terminated by ethanol precipitation. The products are analyzed by 8% (w/v) polyacrylamide-urea gel electrophoresis as described above.

Polymerase Chain Reaction cDNA Cloning and Oligonucleotide Colony Hybridization

Reverse Transcriptase-Polymerase Chain Reaction Amplification

A fragment of apoB mRNA can be amplified from an RNA sample by RT-PCR. For human apoB mRNA, the primers are 5′ GGAATTCGAAACTGACTGCTCTCACAA 3′ for the 5′ primer and 5′ CGGATCCGCTATTTTCAAATCATGTAA 3′ for the 3′ primer, producing a 171-base pair (bp) product. For rat apoB mRNA, the corresponding primers are 5′ GGAATTCTGGGAGAGACAAGTAGCTGG 3′ and 5′ CGGATCCGTTCTTTTTAAGTCCTGTGC 3′, producing a 203-bp product. Amplifications can be performed with a Perkin-Elmer DNA amplification reagent kit.

A 100-μl reaction mixture (10 μl of 10× reaction buffer, 16 μl of 1.25 m*M* deoxynucleoside triphosphate mixture, 1 μg each of primer and RNA) is first denatured at 70° C for 7 min; 4 units of reverse transcriptase is then added. The reaction is incubated at 50° C for 10 min and then at 91° C for 1 min before addition of 5 units of *Taq* polymerase. Thirty cycles (each cycle consisting of extention at 58° C for 6 min and denaturation at 92° C for 1 min) are carried out on a Perkin-Elmer Cetus thermocycler.

Cloning of Polymerase Chain Reaction Fragments and Colony Hybridization to Sequence-Specific Oligonucleotides

The PCR-amplified fragments are phenol–chloroform extracted and ethanol precipitated before being digested with *Eco*RI and *Bam*HI. The *Eco*RI/*Bam*HI-linearized vector (pGEM3Z) and PCR fragments are purified on 1% (w/v) agarose gels, ligated together, and used to transform competent JM109 bacterial cells by the method of Hanahan (31). Colonies are lifted onto nictrocellulose filters and amplified on chloramphenicol (170 μg/ml) plates overnight. Filters are denatured with 0.5 *N* NaOH–1.5 *M* NaCl; neutralized with 0.5 *M* Tris-HCl (pH 7.4)–1.5 *M* NaCl, and baked at

80° C for 2 hr. The sequence-specific oligonucleotide probes used for hybridization are B-Gln: 5′ AGACATATATGATACAATTT 3′ (for human) and 5′ AGACATACGCGATACAATTT 3′ (for rat), respectively, and B-Stop: 5′ AGACATATATGATATAATTT 3′ (for human) and 5′ AGACATACGCGATATAATTT 3′ (for rat), respectively. Hybridization is performed in 6× SSC (1× SSC is 0.15 *M* NaCl plus 0.015 *M* sodium citrate) 5× Denhardt's solution, 0.1% (w/v) sodium dodecyl sulfate (SDS), salmon sperm DNA (3.5 μg/ml) at 50° C (for B-Stop) or 52.5° C (for B-Gln) for 12–24 hr using 5′ end-labeled B-Stop or B-Gln probes. Washings are performed in 6× SSC, 0.1% (w/v) SDS at room temperature for 20 min, followed by 2× SSC, 0.1% (w/v) SDS at 50° C for 1.5 min (for B-Stop) or 52.5° C for 2 min (for B-Gln). Blots are exposed to Kodak (Rochester, NY) X-Omat AR film at −70° C for 16 hr. All blots are first hybridized to the B-Stop probe, stripped, and rehybridized to the B-Gln probe. These conditions should be verified in individual laboratories using standard mixtures of synthetic apoB-100 and apoB-48 mRNAs to ensure 100% sensitivity and 100% specificity. Most (>99%) of the colonies should generate a signal and there should be 100% congruency between the sequence of randomly picked clones and the B-Gln- or B-Stop-specific signal (27).

In Vitro Assays for Apolipoprotein B mRNA Editing Activity

Apolipoprotein B mRNA editing can be demonstrated *in vitro* using synthetic apoB-100 mRNA substrate and cytosolic S-100 extracts (20) or nuclear extracts (28) from intestine or liver of different mammalian species. Such *in vitro* systems allow the analysis of the sequence requirements for efficient editing of apoB mRNA (32). Furthermore, the use of editing assays *in vitro* enabled Teng and Davidson (26) to show that chicken (a species that does not edit) intestinal extracts stimulate the editing activity of mammalian tissue extracts. The development of an *in vitro* enhancement (complementation) assay also enabled them to clone the cDNA for rat apobec-1, the catalytic component of the editing machinery (12).

The following protocols provide the methods for an S-100 cytosolic editing system (which is more active than any nuclear-derived editing system), an enhancement assay, and an editing system reconstituted from cloned rat apobec-1.

S-100 in Vitro Editing System

There are two steps in this procedure: the preparation of S-100 extracts and the *in vitro* editing reaction.

Preparation of Cytosolic S-100 Extracts from Rat and Chicken Small Intestines

Isolation of enterocytes from small intestine using citrate–EDTA chelation is carried out as described by Weiser (33) with the exception that a single 30-min incubation in solution B is used (26).

Materials

Weiser citrate solution A: 5.6 g of NaCl, 0.11 g of KCl, 7.95 g of trisodium citrate, 1.1 g of KH_2PO_4, and 0.8 g of Na_2HPO_4 per liter

Weiser EDTA solution B: 1.18 m*M* EDTA and 0.5 m*M* DTT in phosphate-buffered saline (PBS), pH 7.2

1. Segments of small intestines are isolated from adult Sprague-Dawley rats and 10- to 15-day-old White Leghorn chickens. They are cleaned with ice-cold saline containing 1 m*M* DTT to remove intestinal contents.
2. Each segment of cleaned small intestine is clamped at one end, filled with warm (37° C) Weiser citrate solution A, and clamped shut at the other end. The segment of small intestine is incubated at 37° C in PBS solution for 15 min with gentle shaking on a water bath.
3. The citrate solution is removed carefully, which should leave the intestine free of the mucus that normally lines the lumen of the small intestine. The segment of intestine is then filled with warm (37° C) Weiser EDTA solution B and incubated in PBS solution at 37° C for 30 min with gentle shaking. Intact enterocytes are dislodged easily under these conditions.
4. The EDTA solution containing the isolated enterocytes is transferred to a polypropylene tube and the cells are pelleted at 4° C by centrifugation at 2000 rpm for 10 min. The isolated enterocytes are washed twice with ice-cold PBS and pelleted by centrifugation to remove any residual EDTA.

Preparation of cytosolic S-100 extracts from enterocytes is as described by Dignam *et al.* (34).

Materials

Dignam buffer A: 10 m*M* *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES, pH 7.9), 1.5 m*M* $MgCl_2$, 10 m*M* KCl, 0.5 m*M* DTT

Dignam buffer B: 300 m*M* HEPES (pH 7.9), 1.4 *M* KCl, 30 m*M* $MgCl_2$

Dignam buffer D: 20 m*M* HEPES (pH 7.9), 25% (v/v) glycerol, 100 m*M* KCl, 0.5 m*M* EDTA

1. Cells are suspended in 5 vol of Dignam buffer A containing the following protease inhibitors: leupeptin (10 μg/ml), antipain (10 μg/ml), 1 m*M* benzamidine, and 0.5 m*M* phenylmethylsulfonyl fluoride (PMSF). They are allowed to sit for 10 min on ice and are then collected by spinning at 2000 rpm at 4° C for 10 min. They are resuspended in 2 vol of Dignam buffer A containing the same protease inhibitors.
2. Cells are homogenized by ~100 strokes of a type B pestle on a Dounce homogenizer. The extent of cell lysis should be checked by microscopy.
3. The homogenate is centrifuged at 4°C for 10 min at 2000 rpm on a Sorvall RT6000 centrifuge. The pellet containing crude cell nuclei is discarded. To the su-

pernatant is added 0.11 vol of Dignam buffer B and the homogenate is centrifuged for 1 hr at 100,000 *g* at 4°C.

4. After centrifugation, the S-100 supernatant fraction is removed and dialyzed for 8 hr against Dignam buffer D containing 0.5 m*M* DTT and 0.5 m*M* PMSF. The protein concentration of the S-100 extract is determined by the Bradford method (Bio-Rad, Richmond, CA) and stored in aliquots at −80°C. It should remain active for as long as 5 years under these conditions.

In Vitro Editing Assay

1. Each assay contains 0.8–2.0 ng of synthetic apoB-100 RNA prepared by *in vitro* transcription, 1 μg of cytosolic S-100 extract, 200 ng of yeast tRNA, 10 m*M* HEPEs (pH 7.9), 100 m*M* KCl, 0.25 m*M* EDTA, 0.25 m*M* DTT, and 20 units of RNasin (Promega) in a final volume of 20 μl (26). After incubation at 30°C for 2–3 hr, the reaction is stopped by the addition of an equal volume of 2× stop solution [200 m*M* Tris-HCl (pH 8.0), 20 m*M* EDTA, 0.8% (w/v) SDS, 0.4 *M* NaCl, proteinase K (400 μg/ml), and yeast tRNA (200 μg/ml)]. The reaction is further incubated at 30°C for 20 min and RNA is reisolated by phenol–chloroform/isoamyl alcohol (1:1, v/v), extraction followed by ethanol precipitation at −20°C overnight.

2. The following day, RNA is collected by centrifugation at 13,000 *g* for 20 min at 4°C in a microcentrifuge. The RNA pellet is washed with 75% (v/v) alcohol and analyzed by primer extension. The RNA is denatured at 75°C for 10 min and annealed to approximately 20 pg of ^{32}P-labeled oligonucleotide primer at 37°C for 1 hr in a final volume of 5 μl. The annealed products are extended in extension buffer containing dATP, dCTP, dTTP, and dideoxy-GtP (500 μ*M* each; Pharmacia LKB Biotechnology, Alameda, Inc.) and 10 units of MMLV-RT (BRL) at 45°C for 60 min in a final volume of 20 μl. Products are analyzed by 8% (w/v) polyacrylamide–urea electrophoresis and autoradiography at −70°C.

In Vitro Editing Enhancement Assay

This assay is performed in exactly the same manner as the *in vitro* editing assay except that 1–10 μg of chicken enterocyte cytosolic S-100 extract is added to mammalian cytosolic S-100 extracts or recombinant apobec-1 produced *in vitro*. Results of the representative experiments shown in Fig. 3 indicate that the addition of chicken intestinal extracts stimulated the editing efficiency of rat enterocyte S-100 extracts more than fourfold; with human intestinal S-100 extracts, the editing efficiency increased from 0.9 to 7%, and with pig intestinal S-100 extracts the editing efficiency increased from 16 to 33%.

Expression of Recombinant Rat Small Intestine Apolipoprotein B mRNA Editing Enzyme

We have been able to obtain biologically active recombinant apobec-1 from protein expression systems using bacculovirus expressed in insect cells (Invitrogen, San Diego, CA), adenovirus-transduced mammalian cells, and GST vector expressed

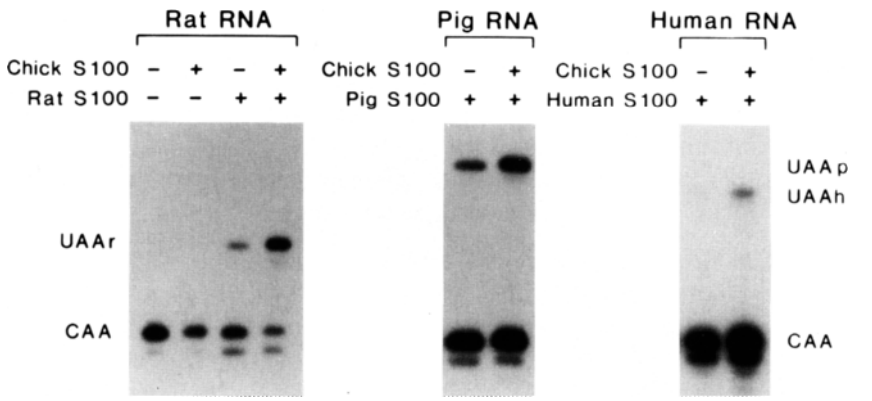


FIG. 3 ApoB mRNA editing enhancement assay. This is an example of primer extension assays performed on S-100 enterocyte extracts prepared from different vertebrate species in the absence and presence (i.e., enhancement) of chicken enterocyte S-100 extract. Note that chicken enterocyte S-100 extract by itself has no editing activity. Synthetic rat, pig, or human apoB RNA was incubated for 2 hr at 30°C (see method) with buffer alone; with 5 μg of chicken enterocyte S-100 extract alone; with 5 μg of rat, pig, or human enterocyte S-100 extract alone; or in combination with 5 μg of chicken enterocyte S-100 extract. ApoB RNA was subsequently analyzed by primer extension. The unedited (CAA) and edited products (human UAAh, pig UAAp, and rat UAAr) on the autoradiographs are as indicated.

in *Escherichia coli* (Pharmacia, Piscataway, NJ) as well as a TNT-coupled reticulocyte lysate system (Promega), which is a coupled transcription–translation system. Here we describe the TNT-coupled reticulocyte lysate system, because it is the most convenient method to test the activity of wild-type apobec-1 and large numbers of mutant apobec-1 and the competence of various tissue extracts to complement the editing activity of apobec-1 *in vitro*.

1. TNT-coupled reticulocyte lysate system is a simple, rapid, *in vitro*-coupled transcription–translation system. Apobec-1 is cloned into a plasmid vector pSPORT 1 containing the SP6 and T7 RNA polymerase promoter (BRL). From the TNT-coupled reticulocyte lysate kit, reaction components are assembled in a 1.5-ml microcentrifuge tube as follows:

TNT rabbit reticulocyte lysate	13 μl
TNT reaction buffer	1 μl
RNA polymerase (T7 or SP6)	1 μl
Amino acid mixture (1 m*M*)	1 μl
RNasin ribonuclease inhibitor (40 μ/μl)	1 μl
Supercoil DNA template	1 μg
Nuclease-free H_2O	To a final volume of 25 μl

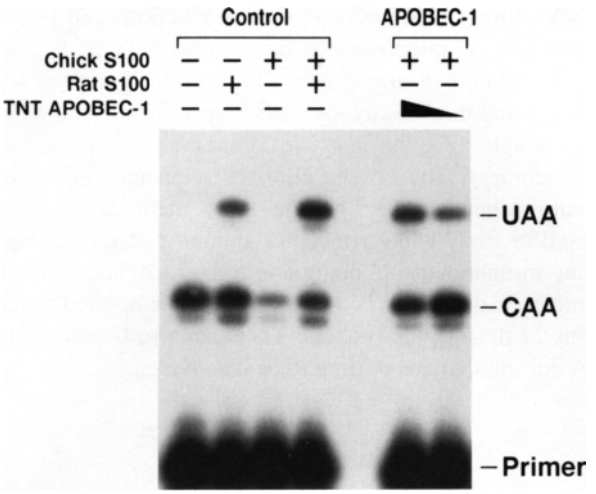


FIG. 4 ApoB mRNA editing activity of recombinant apobec-1 expressed in TNT-coupled reticulocyte lysate system. *Left* (four lanes): Synthetic rat apoB RNA was incubated at 30° C for 2 hr in buffer alone, with 5 μg of rat enterocyte S-100 extract alone, with 5 μg of chicken enterocyte S-100 extract alone, or with 5 μg each of rat and chicken enterocyte S-100 extract. *Right* (two lanes): 3 ng and 0.3 ng of TNT-apobec-1 plus 5 μg of chicken enterocyte S-100 extract. ApoB RNA products were analyzed by primer extension. APOBEC-1, Recombinant rat apobec-1 produced by the TNT-coupled reticulocyte lysate system.

After addition of all the components, the lysate is mixed gently and allowed to incubate at 30° C for 60–120 min. [^{35}S]methionine can be used to quantitate the amount of protein translated by separating the products on a 12% (w/v) SDS-polyacrylamide gel and detecting them by fluorography.

2. Following *in vitro* transcription–translation, aliquots (1 μl $\cong$ 3 ng of apobec-1) of the lysate are used to carry out *in vitro*-editing enhancement assay using synthetic rat apoB RNA as template in the presence of 5 μg of chicken enterocyte S-100 extract. A representative result of *in vitro* editing enhancement assay using TNT-coupled reticulocyte lysate translated recombinant apobec-1 protein is shown in Fig. 4. No editing activity was detected in TNT rabbit reticulocyte lysate alone. There was greater than 60% editing activity from lysate-expressed apobec-1 when the assay was performed in the presence of chicken enterocyte S-100 extracts.

Concluding Remarks

In this chapter we have outlined the protocols involved in the characterization of apoB mRNA editing and the components required for this activity. There are interesting parallels between apoB mRNA editing and kainate receptor subunit

mRNA editing: both are deamination reactions and both appear to require the participation of complementation factors (see Refs. 35 and 36 for details on the kainate receptor subunit editing system). Furthermore, there is sequence similarity between apobec-1 and double-stranded RNA-specific adenosine deaminase (37, 38), the enzyme thought to be the catalytic component of kainate receptor subunit mRNA editing machinery. The conservation of sequences in these two enzymes suggests a similar mechanism of action. Therefore, methods similar to those described here can be used to study kainate receptor subunit mRNA editing. Other examples of RNA editing in mammals are beginning to be recognized (18, 19). The principles behind the methods described in this chapter can be applied to the characterization of RNA editing of these other systems. They can also be used to identify novel instances of RNA editing that are waiting to be discovered.

References

1. L. M. Powell, S. G. Wallis, R. J. Pease, Y. H. Edwards, T. J. Knott, and J. Scott, *Cell* **50,** 831 (1987).
2. S. H. Chen, G. Habib, C.-Y. Yang, Z.-W. Gu, G. R. Lee, S.-a. Weng, S. R. Silberman, S.-J. Cai, J. P. Deslypere, M. Rosseneu, A. M. Gotto, Jr., W.-H. Li, and L. Chan, *Science* **238,** 363 (1987).
3. P. P. Lau, W. Xiong, H.-J. Zhu, S.-H. Chen, and L. Chan, *J. Biol. Chem.* **266,** 20550 (1991).
4. L. Chan, *BioEssays* **15,** 33 (1993).
5. J. Scott, N. Navaratnam, S. Bhattacharya, and J. R. Morrison, *Curr. Opin. Lipidol.* **5,** 87 (1994).
6. L. Chan and P. H. Seeburg, *Sci. Am. Sci. Med.* **2,** 68 (1995).
7. N. O. Davidson, S. Anant, and A. J. MacGinnitie, *Curr. Opin. Lipidol.* **6,** 70 (1995).
8. J. P. Kane and R. J. Havel, *in* "The Metabolic and Molecular Bases of Inherited Disease" (C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, eds.), pp. 1853–1885. McGraw-Hill, New York, 1995.
9. B. Sommer, M. Köhler, R. Sprenzel, and P. H. Seeburg, *Cell* **67,** 11 (1991).
10. M. Higuchi, F. N. Single, M. Köhler, B. Sommer, R. Sprengel, and P. H. Seeburg, *Cell* **75,** 1361 (1993).
11. S. G. Harris, I. Sabio, E. Mayer, M. F. Steinberg, J. W. Backus, J. D. Sparks, C. E. Sparks, and H. C. Smith, *J. Biol. Chem.* **268,** 7382 (1993).
12. B. Teng, C. F. Burant, and N. O. Davidson, *Science* **260,** 1816 (1993).
13. S. Yamanaka, K. S. Poksay, M. E. Balestra, G.-Q. Zeng, and T. L. Innerarity, *J. Biol. Chem.* **269,** 21725 (1994).
14. C. Hadjiagapiou, F. Giannoni, T. Funahashi, S. F. Skarosi, and N. O. Davidson, *Nucleic Acids Res.* **22,** 1874 (1994).
15. P. P. Lau, H.-J. Zhu, A. Baldini, C. Charnsangavej, and L. Chan, *Proc. Natl. Acad. Sci. U.S.A.* **91,** 8522 (1994).
16. N. Navaratnam, J. R. Morrison, S. Bhattacharya, D. Patel, T. Funahashi, F. Giannoni, B. Teng, N. O. Davidson, and J. Scott, *J. Biol. Chem.* **268,** 20709 (1993).

17. C. Barnes and H. C. Smith, *Biochem. Biophys. Res. Commun.* **197,** 1410 (1993).
18. P. M. Sharma, M. Bowman, S. L. Madden, F. J. Rauscher III, and S. Sukumar, *Genes Dev.* **8,** 720 (1994).
19. S. R. Nagalla, B. J. Barry, and E. R. Spindel, *Mol. Endocrinol.* **8,** 943 (1994).
20. D. M. Driscoll, J. K. Wynne, S. C. Wallis, and J. Scott, *Cell* **58,** 519 (1989).
21. P. P. Lau, S.-H. Chen, J. C. Wang, and L. Chan, *Nucleic Acids Res.* **18,** 5817 (1990).
22. H. C. Smith, S.-R. Kuo, J. W. Backus, S. G. Harris, C. E. Sparks, and J. D. Sparks, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 1489 (1991).
23. P. E. Hodges, N. Navaratnam, J. C. Greeve, and J. Scott, *Nucleic Acids Res.* **19,** 1197 (1991).
24. K. Bostrom, Z. Garcia, K. S. Poksay, D. F. Johnson, A. J. Lusis, and T. L. Innerarity, *J. Biol. Chem.* **265,** 22446 (1990).
25. D. M. Driscoll and E. Casanova, *J. Biol. Chem.* **265,** 21401 (1990).
26. B. Teng and N. O. Davidson, *J. Biol. Chem.* **267,** 21265 (1992).
27. J. H. Wu, C. F. Semenkovich, S.-H. Chen, W.-H. Li, and L. Chan, *J. Biol. Chem.* **265,** 12312 (1990).
28. S.-H. Chen, X. Li, W. S. L. Liao, J. H. Wu, and L. Chan, *J. Biol. Chem.* **265,** 6811 (1990).
29. B. Teng, D. D. Black, and N. O. Davidson, *Biochem. Biophys. Res. Commun.* **173,** 74 (1990).
30. B. Teng, M. Verp, J. Solomon, and N. O. Davidson, *J. Biol. Chem.* **265,** 20616 (1990).
31. D. Hanahan, *J. Mol. Biol.* **166,** 557 (1983).
32. H. C. Smith, *Semin. Cell Biol.* **4,** 267 (1993).
33. M. M. Weiser, *J. Biol. Chem.* **248,** 2536 (1973).
34. T. D. Dignam, R. M. Lebovitz, and R. G. Roeder, *Nucleic Acids Res.* **11,** 1475 (1983).
35. J.-H. Yang, P. Sklar, R. Axel, and T. Maniatis, *Nature (London)* **374,** 77 (1995).
36. S. M. Reuter, C. M. Burns, S. A. Coode, P. Mookherjee, and R. B. Emeson, *Science* **267,** 1491 (1995).
37. U. Kim, Y. Wang, T. Sanford, Y. Zeng, and K. Nishikura, *Proc. Natl. Acad. Sci. U.S.A.* **91,** 11457 (1994).
38. M. Nakamuta, K. Oka, J. Krushkal, K. Kobayashi, M. Yamamoto, W.-H. Li, and L. Chan, *J. Biol. Chem.* **270,** 13042 (1995).

[19] Ribozymes as Therapeutic Agents and Tools for Gene Analysis

John J. Rossi, Edovard Bertrand, and Daniela Castanotto

Introduction

Ribozymes are RNA molecules with enzymatic activity. These RNAs can be engineered to base pair with and cleave targeted RNAs, be they of cellular or viral origins. As such, they represent a new approach for specific inactivation of gene expression, and for the therapeutic intervention of disease, both of viral and genetic origin. Ribozymes have been effectively used in cell cultures against viral targets, endogenous transcripts, and transfected genes. They have also been demonstrated to function effectively *in vivo* in transgenic animals (mouse and *Drosophila*), allowing the generation of phenotypes that are difficult to achieve by conventional genetic approaches. Ribozymes are receiving considerable attention as therapeutic agents for the treatment of virally initiated as well as endogenous hereditary diseases. Ribozyme engineering involves the design of RNA molecules in which the antisense and enzymatic functions are combined into a single transcript, which can specifically pair and cleave any target RNA. Because of their high target specificity, coupled with the possibility of being delivered either exogenously or endogenously, ribozymes have potential for use in both basic studies as well as therapeutic applications.

Catalytic Motifs

There are five catalytic motifs that have been successfully adapted for use in ribozyme applications. These are the group I introns, RNase P, the hammerhead and hairpin motifs, and the self-cleaving domain of the hepatitis delta virus. The catalytic and base pairing domains are depicted in a schematic fashion in Fig. 1. Each of the ribozymes requires a divalent metal cation for activity (usually Mg^{2+}), which may participate in the chemistry of the cleavage reaction and/or may be important for maintaining the structure of the ribozyme (1). The group I intron of *Tetrahymena* was the first ribozyme for which the *cis*-cleaving (on a portion of the same RNA strand) reaction was converted into a *trans* (on an exogenous RNA molecule) reaction (2). RNase P, which is present in all prokaryotic and eukaryotic cells, has as its primary role in nature the maturation of the 5′ ends of precursor tRNA transcripts. *In vitro* cleavage of non-tRNA molecules has been demonstrated via the use of RNAs designated as external guide sequences (EGS, Fig. 1). These sequences pair with the

Methods in Molecular Genetics, Volume 8

target RNAs and provide the structural features required for recognition and cleavage by RNase P (3, 4).

Small ribozymes have been derived from a motif found in single-stranded plant viroid and virusoid RNAs that replicate via a rolling circle mechanism (5). Based on a shared secondary structure and a conserved set of nucleotides, the term *hammerhead* has been given to one group of these self-cleavage domains (5). A self-cleaving motif resembling the hammerhead domain has also been found in a newt satellite RNA (6). The simplicity of the hammerhead catalytic domain (Fig. 1) has made it a popular choice in the design of *trans*-acting ribozymes (7, 8). The requirements at the cleavage site are relatively simple, and virtually any UX (where X is U, C, or A) can be targeted, although the efficiency of the cleavage reaction can vary more than 100-fold between the different cleavage site combinations (9). A second plant-derived self-cleavage motif thus far identified only in the negative strand of the tobacco ringspot satellite RNA has been termed the *hairpin* or *paperclip* (Fig. 1) (10, 11). An engineered version of this catalytic motif has also been shown to be capable of cleaving and turning over multiple copies of a variety of targets *in trans* (11). The cleavage site requirements have been discerned and include an obligatory G at the nucleotide 3′ to the site of cleavage (12). There is otherwise a great deal of flexibility in the substrate–guide sequence combinations, making this a potentially useful catalyst against a wide array of RNA targets.

The other RNA catalytic motif that has been analyzed intensely and has now been engineered for cleavage *in trans* is the hepatitis delta (δ) ribozyme (13–15). The delta virus has two different catalytic domains, one deriving from the plus strand and the other from the minus strand of the virus. Two different versions of the delta ribozyme, both capable of cleavage *in trans,* have been described (14, 15). One of these utilizes a 73-nucleotide catalyst that has been demonstrated to turn over multiple target RNAs, and has a guide sequence that can be modified to bind and cleave a variety of RNA targets (15).

Design of Ribozymes and Antisense Oligonucleotides

In this chapter we focus our attention on the hammerhead ribozyme motif because it has received the most attention from those interested in developing ribozymes as surrogate genetic tools. The choice of whether to use a ribozyme or nonenzymatic antisense is the first decision to be made. The advantages of the ribozyme approach are still somewhat theoretical because it is difficult to prove formally that a ribozyme is cleaving the target RNA or turning over multiple substrates in an intracellular environment. However, numerous experiments in which a standard antisense has been used as a control suggest that there is an added effectiveness if there is a catalytic center included within the antisense domain.

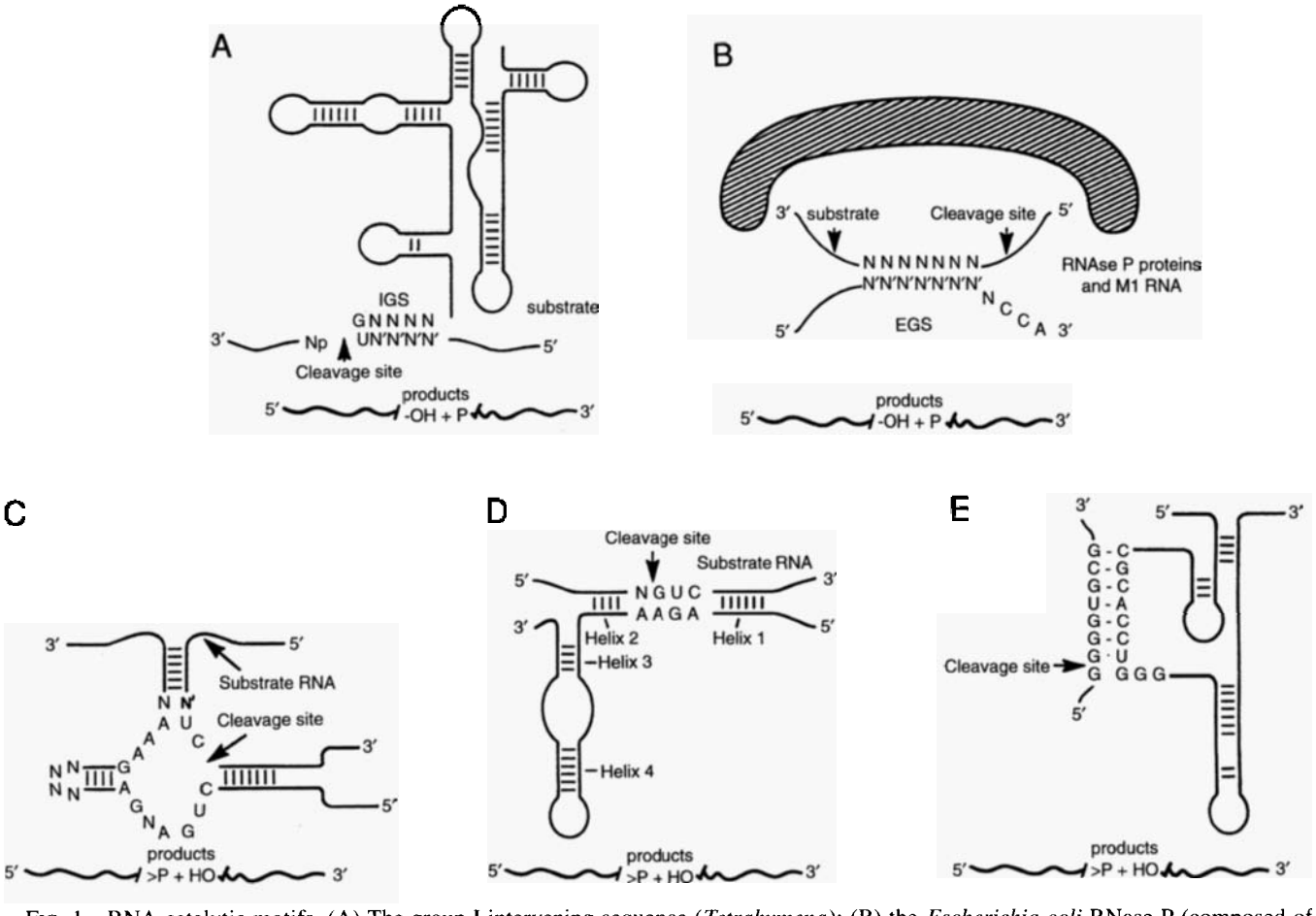


FIG. 1 RNA catalytic motifs. (A) The group I intervening sequence (*Tetrahymena*); (B) the *Escherichia coli* RNase P (composed of RNase P proteins and M1 RNA to make up the holoenzyme); (C) the hammerhead ribozyme; (D) the hairpin ribozyme; (E) the hepatitis delta ribozyme. IGS (internal guide sequence) is the guide sequence for the group 1 introns, and EGS is the external guide sequence for the RNase P. The IGS is part of the ribozyme core and is the primary binding site for the substrate. Specific complementary base-pairing interactions between this site and the substrate confer specificity to the ribozyme reaction. The EGS is an "antisense" RNA with an appended CCA at the 3′ end, which is provided *in trans*. Complementary base pairing of the EGS with the target RNA results in a duplex substrate for RNase P, by mimicking the CCA stem of the natural precursor tRNA substrate molecules. N, No specificity for nucleotides at these positions. All five ribozymes are shown as they would interact with substrates *in trans*. The substrates, cleavage sites, and cleavage products are indicated.

Choice of Target

As a first step in the design of a ribozyme, the target RNA needs to be selected: theoretical considerations suggest that stable rather than unstable RNAs may be the best targets (16). RNAs encoding regulatory proteins are often preferred over the RNAs encoding structural proteins, because small variations in the expression of the former can result in a significant downregulation of the genes with which these proteins interact.

In choosing the region(s) to target within the selected RNA molecule, several aspects of mRNA biosynthesis in eukaryotic cells need to be considered: RNAs are always associated with proteins inside the cell; in the nucleus, pre-mRNAs are complexed with heterogeneous nuclear ribonucleoproteins (hnRNPs) (17, 18) and small nuclear ribonucleoproteins (snRNPs) (19, 20), whereas in the cytoplasm mRNAs are associated with the translational machinery. These factors are not distributed equally along the target RNA, so that some regions of the target may be highly accessible to the antisense or the ribozyme, and other regions may be totally inaccessible. The structure of the RNA can affect the availability of a targeted site. If the target sequence is sequestered by a secondary or tertiary structure, the accessibility of this sequence to degradation by a ribozyme or binding by an antisense molecule will probably be impaired. The prediction of RNA secondary structure using computer-derived energy minimization programs is the most common method for target site selection (21), although the reliability of such computer-generated structures is not always good. If variations in the sequences are known [such as for variants of human immunodeficiency virus type 1 (HIV-1), a phylogenetic analysis can be a useful adjunct to the computer analysis for the prediction of the RNA structures. It is also useful to probe the structure of RNA *in vitro* using enzyme or chemical reagents (22). Although it is possible to determine directly the accessibility of a particular sequence *in vivo* (23) or in cell extracts, it requires intensive labor. Thus, accessible regions of the target should be found empirically by designing separate ribozymes to target several potential sites and testing their relative effectiveness.

Ribozyme Design

In the simplest design of the hammerhead motif, the cleavage site sequences are defined by XUY (where X is U, G, A, or C, and Y is U, C, or A) with cleavage occurring 3′ to Y. Different XUY sequences are cleaved with different efficiencies (9). Because the most commonly used site in nature is GUC (5), it is often the site of choice when present in the target RNA. When long and/or highly structured RNAs are targeted, the limiting step in the cleavage reaction is likely to be the binding/hybridization step (24, 25); thus the choice of the XUY trinucleotide may not be so critical. A potentially useful approach for target site selections is to incubate the RNA

target (end labeled) with a high molar excess of ribozymes with degenerate flanking sequences (26). This ribozyme population can select sites most accessible to binding and cleavage on the RNA substrate. The analysis of the timed appearance of cleavage products on a sequencing gel should result in the generation of a ladder of cleaved RNAs. Sites that are more accessible to ribozyme-mediate cleavage will preferentially show up as the strongest bands of the ladder.

The number of base pairs required for optimal ribozyme cleavage at different sites is variable and must be determined empirically for each specific site, because the stability of the ribozyme–substrate duplex will be affected by several factors including G-C content, temperature, and RNA structure (27, 28). (a) Twelve to 14 base pairs (bp) is a good starting point, because it has been shown that fewer than 12 bp can result in poor binding and more than 14 bp can reduce turnover by slowing dissociation of ribozyme and cleavage products (25). (b) The inhibition that longer flanking sequences exert on product release may be overcome with the design of asymmetric ribozymes, in which the ribozyme–target pairing is extended on one side of the hybrid and shortened on the other, such that one of the two cleavage products is bound to the ribozyme by only a few bases (four or five; Fig. 2). Ribozymes with long flanking arms might be modified and inactivated by the activity of the dsRNA unwinding/modifying enzyme, which requires only 15–20 bp of double-stranded RNA (29). Increasing the length of the hybrid will decrease ribozyme specificity by stabilizing binding and, thus, cleavage of mismatched targets (30).

Because the complexity of human RNA is about 100-fold less than that for human DNA (31), target specificity can be achieved with as few as 12 bp.

Chemical Synthesis of Ribozymes

Antisense oligonucleotides and ribozymes can be efficiently synthesized on any DNA/RNA synthesizer. The major advantage that chemical synthesis offers over an endogenous expression system (from a transcriptional unit) is that base or backbone chemical modifications can be readily included in the molecules. Several chemical modifications introduced into ribozymes have been shown to increase stability without impairing catalytic capability.

The 2′-ribose hydroxyl group renders RNA more sensitive to nucleases than DNA, and modification of this group can increase ribozyme stability. Only the 2′-OH required for catalysis must be preserved (32). For example, a ribozyme with DNA flanking sequences is catalytically active and two to three times more stable in the cell (33).

Ribozymes containing 2′-fluorocytidine and 2′-fluorouridine, or 2′-aminouridines, are considerably more stable in serum and maintain catalytic activity (34). Modification of the 2′-OH in all but the six nucleotides of the hammerhead's conserved catalytic core to 2′-*O*-allylribonucleotides gave similar results (35).

Phosphorothioate backbone substitutions can also render ribozymes more resistant

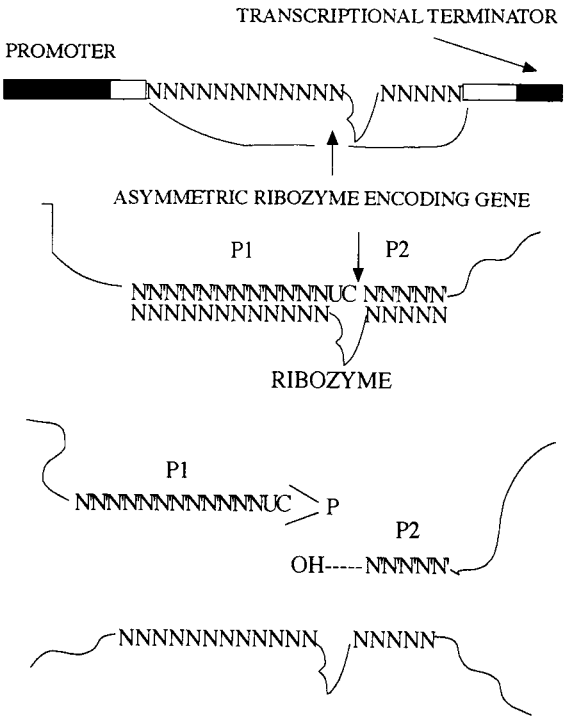


FIG. 2 Schematic representation of a ribozyme expression vector and cleavage reaction. An asymmetric ribozyme gene is cloned between a promoter and an appropriate transcriptional terminator. The N's indicate complementary nucleotides on both sides of the cleavage site. The adjacent open boxes represent the non-base-pairing flanking sequences. Below, a diagrammatic representation of the *trans*-acting ribozyme cleavage reaction. The substrate, cleavage site, and ribozyme generated cleavage products (P1 and P2) are shown. After product dissociation the ribozyme is recycled and can catalyze a new reaction. The cleavage reaction occurs only in the presence of magnesium ions.

to nucleases (25). In addition to increased stability, these types of modifications can be particularly informative for mechanistic, functional, and structural studies (32).

Suppliers

The following is a list of suppliers we have used for obtaining research materials:

HIV strains, HIV molecular clones, and HIV-specific antibodies for quantitating HIV infection: National Institutes of Health AIDS Research and Reference Reagent Program (Rockville, MD)

Tissue culture supplies: Costar (Cambridge, MA)
Tissue culture media and lipofectin reagents: GIBCO-BRL Life Technologies, Inc. (Grand Island, NY)
Other sources of cationic liposomes, cloning vectors, and *in vitro* transcription reagents: Stratagene (La Jolla, CA) and Boehringer-Mannheim (Indianapolis, IN)

Protocol 1: *In Vitro* Transcription and Testing of Ribozymes

1. RNA molecules may be enzymatically synthesized using RNA polymerases (T7, T3, and SP6 are commonly employed). The DNA bearing the RNA molecule (ribozyme or the target for the *in vitro* assays) should be cloned in a transcriptional unit. It should be inserted as close to the promoter as possible, and the plasmic should be linearized immediately downstream of the ribozyme gene in order to minimize vector-derived flanking sequences.

2. *In vitro* transcription can yield RNA up to 50 times the amount of DNA template utilized in the reaction. To produce labeled transcripts use only 0.01–0.5 m*M* cold UTP and 10 μCi of [γ-^{32}P]UTP (3000 Ci/mmol). More than 90% of the radioactivity can be incorporated into the transcripts. The addition of spermidine increases the efficiency of the reaction. This step should be performed at room temperature to avoid template DNA precipitation.

3. Standard cleavage reactions are performed by first heating, to 90° C for 2 min, two separate tubes containing either the ribozyme or substrate (^{32}P-labeled target) in a solution of 20 m*M* Tris-Cl (pH 7.5) and 0–140 m*M* KCl (or NaCl). A noncleaving, mutant version of the ribozyme should be used as a control. This is best accomplished by altering one or more of the conserved nucleotides in the catalytic core (9).

4. Renature the samples for 5 min at the temperature chosen for the cleavage reaction (37–55° C) in the presence of 10 m*M* $MgCl_2$.

5. Mix different amounts of ribozyme and target (depending on the purpose of the analysis, at either equimolar, excess of ribozyme, or excess of target) at the desired temperature. Different time points should be taken, from 0 to 3 hr, the reactions can be stopped by adding an equal volume of formamide gel loading dye containing 20 m*M* EDTA.

6. Denature the samples by heating to 90° C for 2 min and analyze the cleavage products on a 7 *M* urea polyacrylamide gel. Values for k_{cat}/K_m can be determined for ribozymes in which product release is not rate limiting by incubating a constant concentration of substrate (around 1 n*M*) with an increasing excess of ribozyme for a constant time. The k_{cat}/K_m value is derived using the equation $-\ln(\text{Frac S})/t = k_{cat}/K_m[\text{ribozyme}]$, where Frac S is the fraction of remaining substrate and t is time (25).

In Vivo Analysis of Ribozyme Activity

Exogenous Delivery

To examine the effects of a ribozyme *in vivo,* the first obstacle to overcome is the delivery into the cells of choice. A major drawback of exogenous delivery is that the inhibitory effects of the nucleic acids so delivered are transient and require repeated administrations. Despite this, exogenously delivered molecules can incorporate chemical modifications that increase stability; although these modified bases and sugar–phosphate backbones can contribute to toxicity of the molecules. Further investigation is needed to establish whether the advantages of chemical modifications will overcome the general disadvantages of an exogenous delivery.

RNA synthesized *in vitro* by T7 RNA polymerase can also be exogenously delivered. An advantage of this approach is that the ribozyme can be inserted within transcripts containing structural features (stem–loops) that confer stability. A disadvantage is that these extra sequences can negatively affect the catalytic activity of the ribozyme (36).

Many techniques have been developed to introduce functional DNA into cells, and some of these may also be applicable to the delivery of synthetic RNA. DNA is complexed with various compounds (e.g., polylysine) or to lipophilic groups (37–39), which increase cellular uptake or can be complexed with receptor ligands for specific targeting and localization to defined types of cells (40). Conjugates of DNA and lipophilic derivatives can have higher antiviral activity as shown in the case of HIV-1 (39, 41).

Antisense oligonucleotides can also be introduced into cells by direct addition to culture medium. RNA molecules, unless extensively modified, are extremely sensitive to degradation in the medium, and require protection from serum ribonucleases. To date, the most commonly utilized technique for delivering presynthesized RNA molecules into cells in culture is via liposome encapsulation.

Liposomes are composed of one or more concentric phospholipid bilayers (which can incorporate lipid-soluble substances) surrounding an aqueous compartment that can incorporate water-soluble substances. Size and lipid composition can vary, and different liposomes exhibit different characteristics as *in vivo* delivery systems. Negatively charged lipids can increase the efficiency of cellular uptake; saturated lipids and the presence of cholesterol can increase liposome stability.

Liposomes can be covalently attached to antibody molecules, resulting in specific binding to cellular antigens (42) allowing specific targeting to different types of cells. pH-sensitive liposomes on exposure to the low-pH environment of the endosomes fuse with the endosome membranes. Immunoliposomes (pH-sensitive liposomes conjugated to monoclonal antibodies) were successfully targeted to cell surface receptors *in vitro* and *in vivo* (43).

The primary mechanism for cellular uptake of liposomes seems to be endocytosis

(44). Once in the cytoplasm, the liposomes are degraded and the nucleic acids contained inside are released.

Protocol 2: Encapsulation of Ribozymes into Cationic Liposomes

Equipment and Reagents

Vortex mixer
CO_2 incubator
Laminar flow hood
Lipofectin (GIBCO-BRL) or other cationic lipid analog reagents: The reagent consists of DOTMA (0.5 mg/ml) and dioleoylphosphatidylethanolamine (0.5 mg/ml) in sterile water. Reagents functionally similar to the lipofectin agent, such as Lipofect ACE (GIBCO-BRL), Transfectase (GIBCO-BRL), Lipofect-AMINE (GIBCO-BRL), Transfectam RM (Promega, Madison, WI), and DOTAP (Boehringer Mannheim Corp.) can all be used in this procedure
Opti-MEM I (GIBCO) medium
Serum (dependent on cell lines)
DMEM high glucose (Irvine, Santa Ana, CA)
PBS (1×) (Irvine)
Fungi Bact (Irvine)
Pen-Strep (Irvine)
Amphotericin B (Fungizone; Irvine)
2-Mercaptoethanol
Sodium pyruvate (Irvine)
Trypsin (1×) (Irvine)
Sodium bicarbonate (7.5%) solution
L-Glutamine (200 m*M*)
Plasmid DNA containing ribozyme gene transcriptional unit
RNA produced from *in vitro* transcription or chemically synthesized
Synthetic antisense DNA oligonucleotides
Polystyrene tubes, sterile, 17 × 100 mm
Culture dish (60 mm; Costar)

Method

1. Plate exponentially growing cells in tissue culture dishes at 5×10^5 cells/well and grow overnight in a CO_2 incubator at 37° C to 80% confluency.

2. Dilute the presynthesized RNA molecules and the lipofectin reagent (BRL) with Opti-MEM I (GIBCO) medium. The amounts of nucleic acids and liposome suspension need to be optimized for each cell type.
3. Mix the diluted reagent from the previous step, vortex gently for several seconds at room temperature, and incubate for 5 to 10 min at room temperature (prepare this complex in a polystyrene tube, because it can stick to polypropylene).
4. Wash the cells three times with serum-free medium.
5. Add the liposome complex and incubate the cells at 37° C in a CO_2 incubator (5–10% CO_2) for 3–6 hr. In general, transfection efficiency increases with time, although after 8 hr toxic conditions may develop.
6. Add 3 ml of medium with 20% (v/v) serum (the serum is dependent on the cell type; fetal calf serum may be used).
7. Incubate the cells for 24–48 hr at 37° C in a CO_2 incubator.
8. Harvest the cells and assay for gene activity.

Endogenous Delivery of Ribozymes

Endogenous delivery involves the expression of a ribozyme from a DNA template permanently maintained within the cell. For an example of an expression vector, see Fig. 2. Expression of these molecules can be directed by RNA polymerase II and III (Pol II or Pol III) promoters. Pol II promoters include those promoters of viral origin, the long terminal repeat (LTR) promoter sequences of retroviruses, and strong cellular promoters (e.g., the actin promoter). Tandem repeats of the ribozyme genes, under the control of the same promoter, may help to increase the effective concentration from each transcript. Inducible, repressible, or tissue-specific promoters can be used to confer temporal or cell type specific expression and may temper other problems, such as cellular toxicity generated by high levels of expression within the cells.

Endogenous expression from a Pol II promoter, with the exception of a few specialized cases such as the human U1 snRNA Pol II promoter, necessitates a polyadenylation signal, which allows addition of a poly(A) tail. This, along with the 5′-m^7 GpppG cap, common to Pol II transcripts, may prolong the intracellular half-life of the RNA molecules.

Expressing these molecules under the control of Pol III promoters affords additional advantages: Pol III-driven gene expression seems to occur at high levels in all tissues. The size of Pol III-transcribed genes is smaller, presenting a more defined transcript. Other expression strategies are possible. For instance, an snRNA transcription unit (45), which incorporates portions of the snRNA structural sequence and protein-binding sites, could be used. This can facilitate targeting to the nucleus. Another option is to insert a ribozyme into the acceptor arm or the anticodon loop of a tRNA gene, which has resulted in higher levels of expression and stability of the

ribozyme (46). However, this tRNA expression system has been shown to alter post-transcriptional processing and cellular transport (46).

Viral Vectors

Although the ribozyme or antisense expression vector can be delivered with the previously described techniques, integration of the foreign DNA into the host genome does not occur with a high frequency. More promising and efficient technologies employ viral vectors.

Different viral vectors have the capacity to infect a variety of cell types with high efficiency. Several classes of viral vectors are being exploited for delivery of genes *in vivo,* including DNA viruses (adenoviruses, herpes virus, and adeno-associated virus) and RNA retroviruses. General concerns persist with the use of viral pathogens such as residual infectivity, toxicity, and rescue of infectivity by recombination. Additionally, each viral vector has its own set of advantages and disadvantages that ultimately dictate its use in a specific application.

To date the most extensively utilized viral vectors have been retroviruses. This class of viruses can infect a wide variety of cell types resulting in long-term persistence as a consequence of integration into the host chromosome. However, the integration process requires cell replication, thereby restricting retroviral use to actively dividing cells.

Another promising vector system is adeno-associated virus (AAV). The adeno-associated viruses are nonpathogenic, integrating viruses that require helper viruses for replication of their genome. The AAV vector can exist autonomously at high copy number within a cell, and can integrate into a specific site within chromosome 19 (47).

The intracellular transfer of nucleic acids is continuously improving, but many problems remain to be solved. The efficiency of cellular transformation, targeting to specific tissue and organs, subcellular localization of the RNAs, and maintenance of ribozyme activity within the intracellular environment are some of the major issues to be addressed. Other obstacles are the timing of expression and regulation of the level of endogenously synthesized molecules inside the target cells.

In Vivo Assays

For *in vivo* analyses, standard techniques are performed at the levels of RNA [Northern gels, primer extension, polymerase chain reaction (PCR)] and proteins. RNA analysis is not by itself conclusive because it is possible that the cleavage can occur

during the RNA extraction. Use a mutant, noncleaving ribozyme as control to establish that any effect seen *in vivo* is a result of a specific ribozyme activity.

The following steps are used for *in vivo* assays of ribozyme genes.

1. Clone the ribozyme gene in a mammalian expression vector, downstream of a strong promoter (e.g., human β-actin promoter). The vector must contain a marker for the isolation of transfected clones.
2. Transfect the resulting vector into the appropriate cells and isolate stable clones. It is also possible to work with pools of stable clones, so that clonal variability can be eliminated.
3. Examine these clones for the presence of the ribozyme RNA by Northern gel, RNase protection, or RT-PCR (reverse transcription–polymerase chain reaction) analysis. Assay for ribozyme-mediated inhibition of the targeted function using enzymatic, immunological, or physiological assays for the protein product.

Conclusions

The use of ribozymes as surrogate genetic tools and as therapeutic reagents for the treatment of disease is developing rapidly. Ribozymes have the inherent advantage over conventional antisense molecules of not only binding to the target RNA, but also cleaving it, thereby ensuring its permanent inactivation. Because ribozymes base pair with their substrates, antisense effects may contribute to a decrease in steady state levels of the targeted RNA-encoded product. Although this is advantageous from a practical point of view, the optimal design of ribozymes for *in vivo* use requires the development of assays in which the antisense effects of ribozymes can be quantitated separately from the cleavage effects. To date, despite the potential for catalytic activity, excesses of ribozyme over substrate are often required to achieve an inhibition of expression. We have a great deal to learn about maximizing intracellular targeting and stability of ribozymes, as well as learning more about the mechanisms governing intracellular localization before the maximal effectiveness of ribozymes can be achieved. Despite this, ribozymes have been successfully utilized in cell culture as well as animal models to study the effects of inhibiting specific gene expression, and can therefore be considered as a new surrogate genetic tool.

Acknowledgments

This work was supported by American Foundation for AIDS Research (AmFAR) Grant 01917-14-RG, the American Foundation for Pediatric AIDS Research (PAF/AmFAR) Grant 500331-14-PG, and the National Institutes of Health Grants AI25959 and AI29329.

References

1. S. C. Dahm, W. B. Derrick, and O. C. Uhlenbeck, *Biochemistry* **32,** 13040 (1993).
2. A. J. Zaug and T. R. Cech, *Science* **231,** 470 (1986).
3. A. C. Forster and S. Altman, *Science* **249,** 783 (1990).
4. Y. Yuan, E. Hwang, and S. Altman, *Proc. Natl. Acad. Sci. U.S.A.* **89,** 8006 (1992).
5. A. C. Forster and R. H. Symons, *Cell* **49,** 211 (1987).
6. L. M. Epstein and J. G. Gall, *Cell* **48,** 535 (1987).
7. O. C. Uhlenbeck, *Nature (London)* **328,** 596 (1987).
8. J. Haseloff and W. L. Gerlach, *Nature (London)* **334,** 585 (1988).
9. D. E. Ruffner, G. E. Stormo, and O. C. Uhlenbeck, *Biochemistry* **29,** 10695 (1990).
10. J. M. Buzayan, A. Hampel, and G. Bruening, *Nucleic Acids Res.* **14,** 9729 (1986).
11. A. Hampel and R. Tritz, *Biochemistry* **28,** 4929 (1989).
12. S. Joseph *et al., Genes Dev.* **7,** 130 (1993).
13. M. G. Belinsky, G. Dinter-Gottlieb, and J. Taylor, *Nucleic Acids Res.* **19,** 559 (1991).
14. A. D. Branch and H. D. Robertson, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 10163 (1991).
15. A. T. Perotta and M. D. Been, *Biochemistry* **31,** 16 (1992).
16. E. Bertrand, R. Grange, and R. Pictet, *in* "Gene Regulation Biology of Antisense RNA and DNA" (R. P. Erickson and J. G. Izant, eds.), p. 71. Raven Press, New York, 1992.
17. G. Conway, J. Wooley, T. Bibring, and W. M. Lestourgeon, *Mol. Cell. Biol.* **8,** 2884 (1988).
18. G. Dreyfuss, M. S. Swanson, and S. Pinol-Roma, *Trends Biochem. Sci.* **13,** 86 (1988).
19. J. Rinke, A. Bernd, H. Blocker, R. Frank, and R. Lurhmann, *Nucleic Acids Res.* **12,** 4111 (1984).
20. D. L. Black, B. Chabot, and J. A. Steitz, *Cell* **42,** 737 (1985).
21. J. A. Jaeger, D. A. Turner, and M. Zucker, *Proc. Natl. Acad. Sci. U.S.A.* **86,** 7706 (1989).
22. C. Ehresmann, F. Baudin, M. Mougel, P. Romby, J. P. Ebel, and B. Ehresmann, *Nucleic Acids Res.* **15,** 9101 (1987).
23. E. Bertrand, M. Fromont-Racine, R. Pictet, and T. Grange, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 3496 (1993).
24. A. Xing and J. L. Whitton, *Biochemistry* **66,** 1361 (1992).
25. O. Heidenreich and F. Eckstein, *J. Biol. Chem.* **267,** 1904 (1992).
26. C. Rittner, C. Burmester, and G. Sczakiel, *Nucleic Acids Res.* **21,** 1381 (1993).
27. S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner, *Proc. Natl. Acad. Sci. U.S.A.* **83,** 9373 (1986).
28. M. J. Fedor and O. Uhlenheck, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 1668 (1990).
29. K. Nishikura, C. Yoo, U. Kim, J. M. Murray, P. A. Estes, F. E. Cash, and S. A. Liebhaber, *EMBO J.* **10,** 3523 (1991).
30. D. Hershlag, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 6921 (1991).
31. B. Lewin, "Genes," p. 295. John Wiley & Sons, New York, 1983.
32. O. Heidenreich, W. Pieken, and F. Eckstein, *FASEB J.* **7,** 90 (1993).
33. N. R. Taylor, B. E. Kaplan, O. P. Seiderski, H. Li, and J. J. Rossi, *Nucleic Acids Res.* **20,** 4559 (1992).
34. W. A. Pieken, D. B. Olsen, F. Benseler, H. Aurup, and F. Eckstein, *Science* **253,** 314 (1991).

35. G. Paolella, B. S. Sproat, and A. I. Lamond, *EMBO J.* **11,** 1913 (1992).
36. N. Taylor and J. J. Rossi, *Antisense Res. Dev.* **1,** 173 (1991).
37. J. P. Leonetti, G. Degols, and B. Lebleu, *Bioconjugate Chem.* **1,** 149 (1990).
38. A. S. Boutorin, L. V. Guskova, E. M. Inanova, N. D. Kobetz, V. F. Zarytova, A. S. Ryte, L. V. Yurchenko, and V. V. Vlassov, *FEBS Lett.* **254,** 129 (1989).
39. R. L. Letsinger, G. Zhang, D. K. Sun, T. Ikeuchi, and P. S. Sarin, *Proc. Natl. Acad. Sci. U.S.A.* **86,** 6553 (1989).
40. G. Y. Wu, J. M. Wilson, F. Shalaby, M. Grossman, D. A. Shafritz, and C. H. Wu, *J. Biol. Chem.* **266,** 14338 (1991).
41. T. V. Abromova, V. M. Blinov, V. V. Vlassov, V. V. Gorn, V. F. Zarytova, E. M. Ivanova, D. A. Konevets, O. A. Plyasunova, A. G. Pokrovsky, L. S. Sandahchiev, F. P. Svinarchuk, V. P. Starostin, and S. R. Chaplygina, *Nucleotides Nucleosides* **10,** 419 (1991).
42. S. Wright and L. Huang, *Adv. Drug Delivery Rev.* **3,** 343 (1989).
43. R. J. Y. Ho, B. T. Rouse, and L. Huang, *J. Biol. Chem.* **262,** 13973 (1987).
44. C. R. Alving, *Adv. Drug Delivery Rev.* **2,** 107 (1988).
45. C. Guthrie and B. Patterson, *Annu. Rev. Genet.* **22,** 387 (1988).
46. M. Cotten and M. L. Birnstiel, *EMBO J.* **12,** 3861 (1989).
47. R. M. Kotin, M. Siniscalo, J. Samulski, X. Zhu, L. Hunter, C. Laughlin, S. McLaughlin, N. Muzycka, M. Rocchi, and K. I. Berns, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 2211 (1990).

[20] *In Vivo* Analysis of Signaling Elements via Targeted, Inducible Antisense RNA

Patricia A. Galvin-Parton, Hsien-yu Wang, Jiang Fan Chen, Craig C. Malbon, and Christopher M. Moxham

Introduction

One of the most exciting advances in human genetics is the translation of new technologies in molecular biology into therapeutics. Although only a subset, human pathologies derivative of genetic alterations provide an important target for molecular therapeutics, including gene therapy. Aside from introduction of genes to compensate for genetic deficiencies, few approaches seem to offer the potential of antisense RNA technology to target messenger RNA (mRNA) products of a specific gene or perhaps family of genes. Once a captive of the imagination of molecular biologists, antisense DNA and RNA technologies have advanced to the point of near-universal acceptance with regard to versatility and feasibility.

Application of antisense RNA to suppress mRNA targets has been demonstrated in countless instances *in vitro* and more recently *in vivo*. This chapter seeks to outline the approaches available for applying antisense RNA technology toward scoring a "technical knockout" (TKO) *in vivo* with a specific gene product and tissue site in mind. Practical experimental procedures are provided to enable the reader to address the applicability, technical feasibility, and translation of this technology to the laboratory as the first step. Although successful to date in several applications *in vivo*, antisense RNA technology and its application to human diseases will require expanded and more detailed evaluation. Neither technical nor ethical considerations of application of antisense RNA technology to human diseases are offered, by intent. The aim of the work is to catalyze a thoughtful and critical analysis of its promise and current limitations. G proteins are critical elements in cell signaling and biology. This chapter focuses on our applications of antisense RNA *in vivo* to probe the role of specific G proteins in neonatal growth and development.

G-Proteins and Human Diseases

Based on the prominent role of G proteins in regulation, one would predict the occurrence of human disease in which G protein function has been altered (1). With respect to infectious diseases, cholera and whooping cough highlight G-protein targets (2). Genetic mutations in $G_{s\alpha}$ are associated with pseudohypoparathyroid Albright hereditary osteodystrophy and McCune–Albright syndrome (1). Mutations of

Methods in Molecular Genetics, Volume 8

$G_{i\alpha 2}$ have been identified in several endocrine tumors, including adrenal cortex and ovary (3, 4). In addition, mutations in cell surface receptors that are coupled through G proteins are responsible for a variety of diseases (5, 6).

Antisense DNA/RNA Strategies

Dominant-negative control of gene expression in cells has proved to be an effective means of evaluating the function of a specific gene product (7). Although first reported in 1977 (8), use of complementary or "antisense" RNA and DNA sequences to block expression of a gene product in order to study its function is only in its infancy. Through base pairing, a DNA–RNA or RNA duplex is formed and is either rapidly degraded or becomes an inhibitor of nuclear processing and/or translation of the target protein (9). A variety of examples exist demonstrating the utility of the antisense strategy (7–15). Application of antisense RNA techniques to the study of the functions of G-protein subunits provides a powerful new tool for establishing G-protein coupling partners by blocking the expression of a specific subunit. Our laboratory has succeeded in using antisense oligodeoxynucleotides (16), stable expression of antisense RNA in cells following stable, retroviral infection (10), and inducible antisense RNA expression in transgenic mice (11, 12) to block production of G-protein subunits (see below).

Vector-Driven Expression of Antisense RNA *in Vitro*

The use of antisense DNA oligodeoxynucleotides to suppress gene expression is effective and straightforward for most experiments conducted on a relatively small scale (15). For studies demanding biochemical analyses, introducing vectors containing the antisense RNA sequence of interest into cells by stable transfection or retroviral infection is a more practical approach (10). Our laboratory has developed several antisense RNA approaches utilizing vectors capable of producing high-level expression of antisense RNA either constitutively or inducibly. It is the latter strategy on which we shall focus.

Inducible Suppression of $G_{i\alpha 2}$ in FTO-2B Cells

We have created a novel, inducible antisense RNA expression system with applications for both *in vitro* and *in vivo* studies (11, 12). Our system takes advantage of several key features of phosphoenolpyruvate carboxykinase (*PEPCK*) gene expression that proved ideal for the studies of G-protein function *in vivo* (see Table I). Leading considerations were as follows: (a) undetectable gene expression *in utero* (17, 18), thus bypassing any deleterious effects of suppressed G-protein expression *in utero;* (b) tissue-specific expression, in this case targeting liver, kidney, and white

TABLE I Strategy for Suppression of $G_{i\alpha 2}$ *in Vivo* Using Antisense RNA

Consideration	*PEPCK* gene expression
Developmental restriction	Undetectable expression *in utero*, initial appearances after birth
Tissue-specific expression	Predominant expression in liver, kidney and adipose tissue
Regulated expression	
Inducers	Glucagon, catecholamines (via cAMP)
	Glucocorticoids (liver and kidney)
	Thyroid hormones (synergistic with cAMP)
	Retinoic acid (liver)
	Metabolic acidosis (kidney)
	High-protein diet (via glucagon)
Repressors	Insulin
	Glucocorticoids (adipose)
	Metabolic alkalosis (kidney)
	High-carbohydrate diet (via insulin)

adipose tissue (19) that possess well-characterized G-protein-linked signaling systems; and (c) ability to induce or repress the *PEPCK* promoter *in vivo* by hormonal and dietary manipulation, allowing for regulated antisense RNA expression (20). Figure 1 illustrates the approach: expression of the antisense RNA sequence harbored within a larger hybrid 2.8-kb mRNA of the *PEPCK* gene.

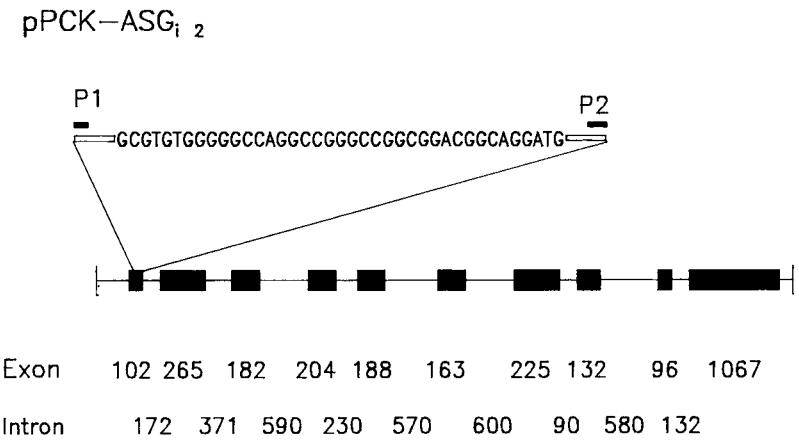


FIG. 1 The pPCK-AS construct designed for inducible and tissue-specific expression *in vitro* and *in vivo*. [Reprinted from Moxham *et al.* (11) with permission (AAAS, 1993).]

The employment of the construct can be initially tested in FTO-2B rat hepatoma cells. These cells display cAMP-inducible *PEPCK* gene expression (20, 21) and express several different G-protein α-subunits including $G_{i\alpha 2}$ and $G_{q\alpha}$. The inducible nature of the pPCK-AS construct is observed following a time course of treatment with 8-(4-chlorophenylthio)-cAMP (CPT-cAMP), a nonhydrolyzable analog of cAMP. As shown in Fig. 2, FTO-2B clones transfected with the pPCK-$ASG_{i\alpha 2}$ construct displayed normal amounts of $G_{i\alpha 2}$ in the absence of cAMP. After a 12-day induction of the antisense gene with CPT-cAMP, $G_{i\alpha 2}$ expression declined >85% in these same cells. The steady state levels of $G_{s\alpha}$ and $G_{i\alpha 3}$, in contrast, were not changed during the CPT-cAMP treatment of cells expressing the antisense RNA to $G_{i\alpha 2}$. Similarly, FTO-2B clones transfected with the vector lacking the antisense sequence to $G_{i\alpha 2}$ displayed no change in $G_{i\alpha 2}$ expression. Thus the FTO-2B hepatoma cell line is

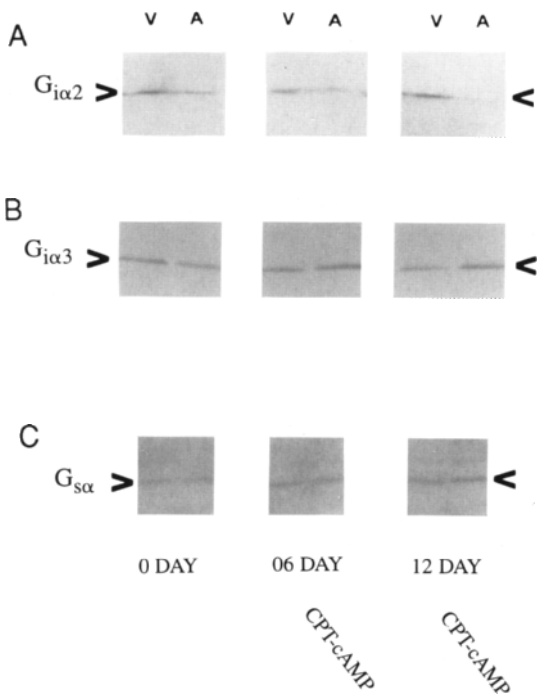


FIG. 2 Cyclic-AMP induces the expression of RNA antisense to $G_{i\alpha 2}$ in FTO-2B rat hepatoma cells transfected with pPCK-$ASG_{i\alpha 2}$. FTO-2B cells are transfected with either the $G_{i\alpha 2}$ antisense RNA construct (A) or vector alone (V) as a control. Neomycin-resistant colonies are selected (G-418, 400 μg/ml) and cultured in the absence or presence of the cAMP analog, CPT-cAMP (25 μM) for either 6 or 12 days. Immunoblot analyses are performed on at least three separate preparations as described and probed with antibodies specific for $G_{i\alpha 2}$, $G_{i\alpha 3}$, and $G_{s\alpha}$ [(A)–(C), respectively]. [Reprinted from Moxham *et al.* (12) with permission (copyright Wiley-Liss, 1993).]

an efficient screening system for the inducibility and specificity of antisense RNA constructs prior to the introduction of the gene into transgenic mice.

Selection of Target Gene Sequences

The two most important criteria for selection of the target gene sequence are range and specificity. Range of suppression can be varied, depending on the desired outcome. If multiple protein family members exist, the choice of sequence will depend on whether the goal is to target a single member (i.e., choosing a sequence unique to one member) or to target all family members (i.e., choosing a gene sequence common to all members). Specificity of the selected gene sequence can be evaluated by comparison against the GenBank database. Sequence evaluation is crucial to avoid nonspecific interactions with other gene sequences. In the case of the G-protein α-subunit family, the 39 base pairs (bp) immediately upstream and including the translation initiation ATG codon were sufficiently unique to ensure both selectivity among the known α subunits and specificity with respect to other gene sequences present in the GenBank database. Once the target sequence is selected, insertion of the antisense DNA sequence into the expression vector is accomplished using standard recombinant DNA technology (22). For our studies utilizing the *PEPCK* gene, the antisense DNA sequence was inserted into the first exon of the *PEPCK* gene 70 bp downstream of the transcription start site (for a detailed description see Ref. 12)

Cell Culture

FTO-2B cells are cultured in Ham's F12–Dulbecco's modified Eagle's medium (DMEM) (1:1) supplemented with 10% (v/v) fetal bovine serum (FBS) at 37° C in an atmosphere of 5% CO_2/95% O_2. Stable transfection with either the vector lacking the antisense sequence as a control or the antisense RNA construct is performed using the Lipofectin (Gibco BRL, Gaithersburg, MD) reagent according to the manufacturer protocol. *PEPCK* promoter activity is induced with 25 μM 8-(4-chlorophenylthio)-cAMP (CPT-cAMP). The CPT-cAMP is added for the period of induction indicated and replenished every 24 hr. CPT-cAMP is prepared in water and sterile filtered prior to use. On the last day of the induction period, the cells are washed once with ice-cold phosphate-buffered saline (PBS), harvested with 5 ml of ice-cold PBS–EDTA, and then centrifuged at 1000 *g* for 10 min at 4° C.

Preparation of Crude Membranes

Transfectant cells are harvested with PBS–EDTA and pelleted by centrifugation at 1000 *g* for 10 min at 4° C. The cell pellet is resuspended in ice-cold HME buffer supplemented with protease inhibitors [20 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES, pH 7.2), 2 mM $MgCl_2$, 1 mM EDTA, aprotinin (5 μg/ml), and leupeptin (5 μg/ml). Cells are disrupted with 20 strokes of a hand-held Dounce tissue homogenizer, prechilled on ice. The cell homogenate is centrifuged at

750 *g* for 10 min at 4° C. The resultant supernatant containing the plasma membranes is centrifuged at 48,000 *g* for 20 min at 4° C. The crude membrane pellet is resuspended in 0.5–1.0 ml of ice-cold HME and protein concentration determined by the method of Lowry *et al.* (23).

SDS-PAGE and Western Blotting of Crude Membranes

Equal amounts of membrane protein (50–100 μg) are resuspended in 30 μl of TDS buffer [40 m*M*Tris (pH 6.8), 2 m*M* dithiothreitol (DTT), 2% (w/v) SDS] and placed in a boiling water bath. After 5 min, 10 μl of *N*-ethylmaleimide (25 mg/ml) is added and the samples incubated at room temperature for 15 min. Prior to electrophoresis, 60 μl of 2× Laemmli buffer is added and the samples boiled for another 5 min. The membrane proteins are separated on a 10% (w/v) polyacrylamide gel. Following electrophoresis the separated proteins are transferred to nitrocellulose with a constant current of 0.05 A for 12 hr at 4° C. The blot is incubated with $G_{i\alpha 2}$-specific antisera for 2 hr at room temperature. Immunoreactive bands are made visible using an alkaline phosphatase-conjugated goat anti-rabbit second antibody (12).

Targeted Expression of $G_{i\alpha 2}$-Specific Antisense RNA *in Vivo*

As shown in Fig. 3, the target tissues for expression of the $G_{i\alpha 2}$ antisense RNA are liver, kidney, and white adipose tissue. We chose to assess the consequences of $G_{i\alpha 2}$ deficiency in adipocytes first, as these cells are easily isolated and possess a well-characterized inhibitory adenylylcyclase (adenylate cyclase) response (24, 25). Expression of the pPCK-$ASG_{i\alpha 2}$ transgene in white adipocytes yielded a >90% reduction in $G_{i\alpha 2}$ expression (11, 12). We probed the inhibitory regulation of adenylylcyclase in acutely prepared adipocytes from the pPCK-$ASG_{i\alpha 2}$ transgenic mice, using the A1-adenosine receptor agonist (−)-*R*-N^6-phenylisopropyladenosine (R-PIA). Intracellular levels of cAMP were elevated 3.1-fold in the untreated adipocytes isolated from the transgenic mice when compared to control mice (Fig. 4A). These data suggest that $G_{i\alpha 2}$ exerts a tonic suppression of the basal adenylylcyclase activity *in vivo*. A comparison of the inhibitory response to several different concentrations of R-PIA in adipocytes from control and transgenic mice demonstrated that the inhibitory control of adenylylcyclase is severely attenuated as result of $G_{i\alpha 2}$ deficiency (Fig. 4B). The residual inhibitory response observed in the adipocytes of the transgenic mice may reflect the residual $G_{i\alpha 2}$ or the presence of some other inhibitory G-protein such as $G_{i\alpha 1}$ or $G_{i\alpha 3}$. Cross-over to A2-adenosine receptors at higher concentrations of R-PIA, represented by a stimulatory adenylylcyclase response, was observed in adipocytes from both transgenic and control mice (Fig. 4B). These results demonstrate that the loss of $G_{i\alpha 2}$ is manifest at the level of the inhibitory adenylylcylase response, suggesting that *in vivo* $G_{i\alpha 2}$ is the major transducer of the inhibitory adenylylcyclase response in adipose tissue.

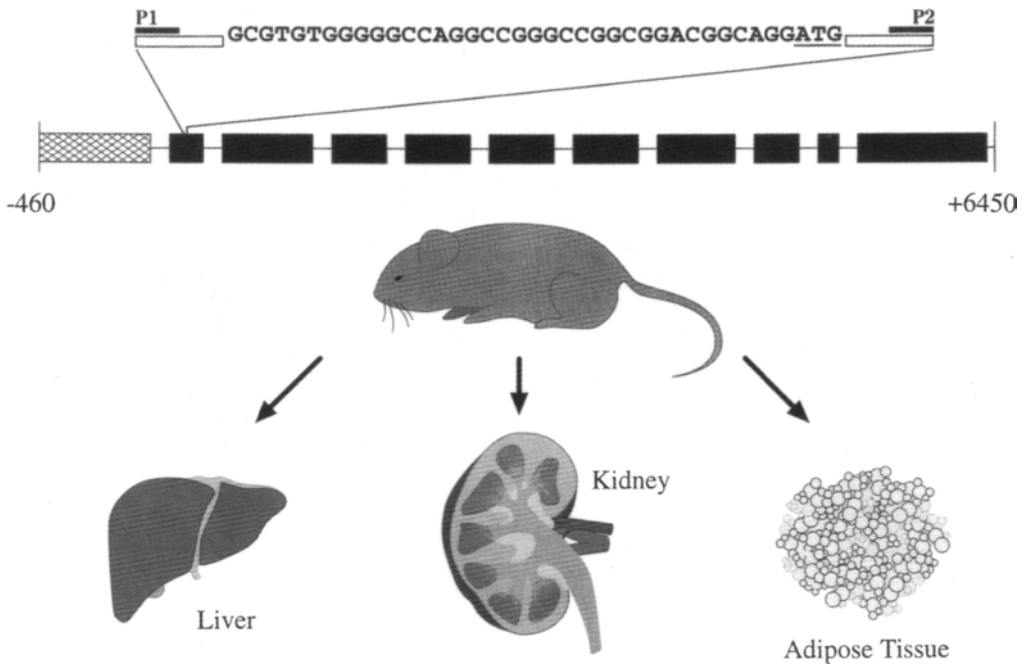


FIG. 3 Target tissues for antisense RNA expression in transgenic mice utilizing the pPCK-AS construct.

Isolation of White Adipocytes

White adipocytes are isolated from epididymal and parametrial fat pads of control and transgenic mice by collagenase digestion, as described previously (26). Briefly, 0.5–1.0 g of white adipose tissue is added to an equal volume of Krebs–Ringer phosphate (KRP) buffer (120 m*M* NaCl, 4.75 m*M* KCl, 1.2 m*M* Mg_2SO_4, 10 m*M* Na_2HPO_4, 1.2 m*M* $CaCl_2$) containing 3% (w/v) bovine serum albumin (BSA) prewarmed to 37°C. Collagenase (type 1, 1 mg/ml; Worthington, Freehold, N.J.) is added and the tissue digested for 30 min at 37°C in an orbital, shaking water bath. After two washes in KRP–BSA the cells are resuspended at a final concentration of 62.5 mg/ml in the same buffer. The KRP–BSA buffer is supplemented with adenosine deaminase (Sigma, St. Louis, MO) at a concentration of 0.5 U/ml.

Measurement of cAMP Accumulation

Cyclic AMP accumulation in adipocytes is measured as described previously (27). Briefly, 80 μl of fat cells (~5 mg/tube) are treated with various agents for 6 min at

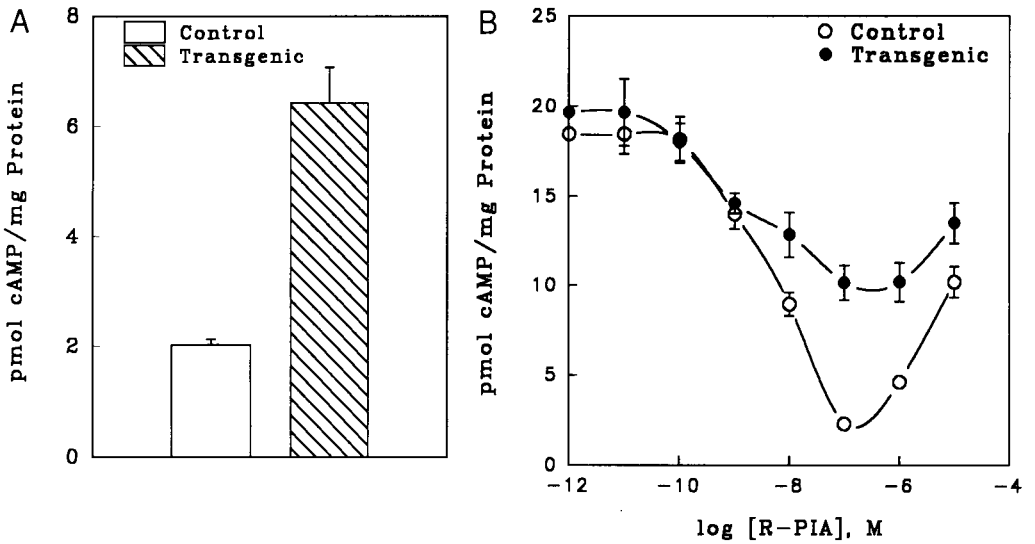


FIG. 4 Inhibitory adenylylcyclase response is attenuated by suppression of $G_{i\alpha 2}$ *in vivo*. The inhibitory adenylylcyclase response is measured using inhibition of cAMP accumulation in adipocytes isolated from mice carrying the pPCK-$ASG_{i\alpha 2}$ transgene and from their littermates. Cyclic-AMP accumulation is measured in unstimulated cells (basal) (A) as well as cells stimulated with 10 μM epinephrine in the absence or presence of increasing concentrations of the inhibitory agonist N^6-R-phenylisopropyladenosine (R-PIA) (B). The data are expressed as the mean values in picomoles of cAMP ($\pm$ SEM) per milligram cellular protein from three independent trials, each performed in triplicate. [Reprinted from Moxham *et al.* (12) with permission (copyright Wiley-Liss, 1993).]

37° C. The reaction is terminated with HCl (0.1 *N* final concentration) and boiling for 1 min. The samples are neutralized with NaOH and assayed for cAMP content using a competitive binding assay (28).

Hormonal inhibition of cAMP accumulation is measured in adipocytes stimulated with epinephrine (10 μM) and the indicated concentrations of the inhibitory A1-adenosine receptor agonist, R-PIA. The results are normalized for total cellular protein added per tube using the protein determination assay of Lowry *et al.* (23).

Application of Targeted Expression of G_α Subunit-Specific Antisense RNA as Therapy

McCune–Albright syndrome is a sporadic disease characterized by polyostotic fibrous dysplasia, *cafe au lait* spots, sexual precocity, and hyperfunction in a variety of endocrine tissues, including thryoid, adrenal, pituitary, and the gonads. It has been

postulated that this syndrome is the result of a dominant somatic mutation occurring in early development that produces an individual who is a "mosaic," possessing both normal cells as well as cells with the mutation. Molecular analysis suggests that the syndrome results from an activating mutation of the G-protein that controls the stimulatory adenylylcyclase pathway, elevating intracellular cAMP levels (29). Similar activating mutations have been identified for other G-protein α-subunits, including G_{i2}. Patients with McCune–Albright syndrome that die in childhood have profound endocrine and nonendocrine disease. Antisense RNA technology may provide a useful strategy to targeted suppression of a constitutively activated G protein, if the somatic mutation is expressed in a known tissue. For patients at risk for hepatomegaly, for example, introduction of an antisense RNA within a modified pPCK-AS construct would permit suppression in the liver. Although only speculation, the ability to create molecular vehicles for expression of antisense RNA that could be induced or suppressed *in vivo* in a tissue-specific manner offers new strategies for approaching a variety of human diseases, including certain specific types of cancer.

Targeted Expression of G_α Subunits *in Vivo*

At the end of the therapeutic spectrum from tissue-specific expression of RNA antisense to a mutant G-protein α subunit is tissue-specific expression of wild-type or constitutively activated forms of the subunit *in vivo* to ameliorate the absence of a functional element in an important signaling cascade. Prominent among the examples is pseudohypoparathyroidism, in which the lack of functional $G_{s\alpha}$ subunit in the proximal tubule impairs signaling through adenylylcyclase and hormonal stimulation of intracellular cyclic AMP accumulation (1). Targeting the proximal tubule with tissue-specific elements from the *PEPCK* gene may provide the vehicle for expression of wild-type $G_{s\alpha}$. We have succeeded in the targeted expression of a constitutively activated mutant of $G_{i\alpha 2}$ in liver and adipose tissue of transgenic mice (30). With these developments, targeted expression of G-protein α subunits may become a therapeutic goal, well within technical reach.

Summary

Antisense RNA technology provides for selective suppression of proteins of interest and thus a new strategy with which to probe the emerging complexity of the various regulatory networks of signal transduction pathways. Considering sheer economics, the use of antisense DNA oligodeoxynucleotides is practical for studies requiring small-scale culture of cells, pilot studies seeking to test the antisense DNA strategy, and in systems in which single-cell assays can be performed (e.g., patch-clamping or histochemical analysis). Vector-driven antisense RNA expression, both constitutive

and inducible, in cell culture allows for large-scale cell growth capacities enabling biochemical analyses. Expanding the antisense RNA approach to transgenic mice provides the means to generate unique animal models with which to explore the role of transmembrane signaling elements in complex biological processes *in vivo.*

In our studies, suppression of $G_{s\alpha}$ with antisense DNA oligodeoxynucleotides provided exciting information concerning the role(s) of this G protein in adipogenesis (16). Similarly, the role of $G_{i\alpha 2}$ in early mouse development has been addressed in F9 embryonic stem cells stably expressing antisense RNA (10). Finally, investigation of the role of $G_{i\alpha 2}$ in adipose tissue and liver function as well as its role in whole-body metabolism, growth, and development have been made possible only through the hybrid *PEPCK* gene construct employed in our laboratory (11). Using a variety of antisense DNA/RNA-based approaches, investigators are now able to explore the roles of signaling elements at several distinctly different levels, selectively targeting the expression of a protein of interest *in vitro* or in tissues *in vivo.*

Our knowledge of the role of transmembrane signaling elements in disease is growing rapidly. Our success with antisense DNA/RNA-based approaches *in vitro* and *in vivo* highlights the potential applications of this technology for use in gene therapy to treat pathological disorders. Expression of antisense RNA in a global set of tissues by retroviral infection or expression in a tissue-specific manner using selective promoters has implications not only for our basic understanding of how signal transduction pathways impinge on these complex events, but also for the development of new therapeutic agents with which to treat human disease.

References

1. G. Milligan and M. Wakelam (eds.), "G-Proteins, Signal Transduction and Disease." Academic Press, New York, 1992.
2. A. G. Gilman, *Annu. Rev. Biochem.* **56,** 615 (1987).
3. J. Lyons, C. A. Landis, G. Harsh, L. Vallar, K. Grunewald, H. Feichtinger, Q.-Y. Suh, O. H. Clark, E. Kawasaki, H. R. Bourne, and F. McCormick, *Science* **249,** 655 (1990).
4. A. M. L. Chan, T. P. Fleming, E. McGovern, M. Chedid, T. Miki, and S. A. Aaronson, *Mol. Cell. Biol.* **13,** 762 (1993).
5. R. J. Lefkowitz, *Nature (London)* **365,** 603 (1994).
6. D. E. Clapham, *Cell* **75,** 1237 (1994).
7. J. Haseloff and W. L. Gerlach, *Nature (London)* **334,** 585 (1988).
8. P. S. Miller, *Biochemistry* **16,** 1988 (1977).
9. J. Goodchild, *in* "Oligodeoxynucleotides—Antisense Inhibitors of Gene Expression" (C. H. Cohen, ed.). CRC Press, Boca Raton, Florida, 1989.
10. D. C. Watkins, G. L. Johnson, and C. C. Malbon, *Science* **258,** 1373 (1992).
11. C. M. Moxham, Y. Hod, and C. C. Malbon, *Science* **260,** 991 (1993).
12. C. M. Moxham, Y. Hod, and C. C. Malbon, *Dev. Genet.* **14,** 266 (1993).
13. M. Shih and C. C. Malbon, *Proc. Natl. Acad. Sci. U.S.A.* **91,** 12193 (1994).

14. N. M. Dean and R. McKay, *Proc. Natl. Acad. Sci. U.S.A.* **91,** 11762 (1994).
15. R. W. Wagner, *Nature (London)* **372,** 333 (1995).
16. H. Y. Wang, D. C. Watkins, and C. C. Malbon, *Nature (London)* **358,** 334 (1992).
17. J. P. Garcia Ruiz, R. Ingram, and R. W. Hanson, *Proc. Natl. Acad. Sci. U.S.A.* **75,** 4189 (1978).
18. F. J. Ballard and R. W. Hanson, *Biochem J.* **104,** 866 (1967).
19. R. W. Hanson and M. A. Mehlman, *in* "Gluconeogenesis: Its Regulation in Mammalian Species." John Wiley & Sons, New York, 1976.
20. Y. Hod and R. W. Hanson, *J. Biol. Chem.* **263,** 7747 (1988).
21. J. Liu, E. A. Park, A. L. Gurney, W. J. Roesler, and R. W. Hanson, *J. Biol. Chem.* **266,** 19095 (1991).
22. J. Sambrook, E. F. Fritsch, and T. Maniatis, *in* "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
23. O. H. Lowry, N. J. Rosebrough, A. L. Farr, and R. J. Randall, *J. Biol. Chem.* **193,** 265 (1951).
24. J. N. Fain and C. C. Malbon, *Mol. Cell. Biochem.* **25,** 143 (1979).
25. C. C. Malbon, P. J. Rapiejko, and T. J. Mangano, *J. Biol. Chem.* **260,** 2558 (1985).
26. C. C. Malbon, R. C. Hert, and J. N. Fain, *J. Biol. Chem.* **253,** 3114 (1978).
27. M. P. Czech, C. C. Malbon, K. Kerman, W. Gitomer, and P. F. Pilch, *J. Clin. Invest.* **66,** 574 (1980).
28. B. L. Brown, D. M. Albano, R. P. Ekins, and A. M. Sgherzi, *Biochem. J.* **121,** 561 (1971).
29. A. Shenker, *J. Pediatrics* **123,** 509 (1993).
30. J. F. Chen, H. Y. Wang, C. M. Moxham, and C. C. Malbon, submitted (1996).

Section V

Genome Recombination, Amplification

This Page Intentionally Left Blank

[21] Extrachromosomal Substrates to Study Immunoglobulin Heavy Chain Class Switch Recombination in Mammalian B Cells

Ming-Jie Li, Helios Leung, and Nancy Maizels

Introduction

Immunoglobulin heavy chain class switch recombination is a regulated recombination event that joins an expressed immunoglobulin heavy chain variable (VDJ) region to a new downstream constant (C) region of a different class or isotype [reviewed in Refs. 1–4; see Fig. 1 (4a–c)]. The result of switch recombination is to modify the antigen clearance properties of an immunoglobulin molecule without affecting its specificity for antigen. Switch recombination involves G-rich regions of repeated sequence that occur upstream of each C region that undergoes switch recombination; Cδ is the only C region that lacks upstream G-rich regions of sequence, and expression of δ is regulated by alternative RNA processing, rather than recombination.

Switch recombination is distinct from other targeted, regulated processes of recombination because it is region specific, but not sequence specific. Chromosomal switch junctions show no evidence of sequence specificity or of homologous pairing during switch recombination, nor do chromosomal junctions cluster within the S regions (5). This apparent imprecision of switch recombination makes its mechanism particularly intriguing.

We have established an assay for genetic elements and factors that regulate switch recombination (6, 7). This assay measures recombination of extrachromosomal substrates that carry sequences from the Sμ and Sγ3 switch regions flanking a conditionally lethal marker, the leftward (P_L) promoter of bacteriophage λ. Constructs carrying λP_L can be propagated in lysogenic strains of *Escherichia coli*, where the endogenous *c*I repressor shuts off transcription from λP_L, but they cannot give rise to ampicillin-resistant colonies in nonlysogens. Recombination that occurs during transfection of activated primary B cells and certain B cell lines will delete λP_L to generate a molecule that will give rise to an ampicillin-resistant colony after transformation of a nonlysogenic *E. coli* host. When the extrachromosomal substrates carry appropriate genetic elements, the frequency of recombination is quite high: as many as 25% of replicated DNA molecules will have recombined during transfection.

The switch substrate recombination assay is rapid, and the recombination substrates can be readily manipulated to test the effect of different sequences and regulatory elements on recombination. The assay also allows easy recovery of recombi-

Methods in Molecular Genetics, Volume 8

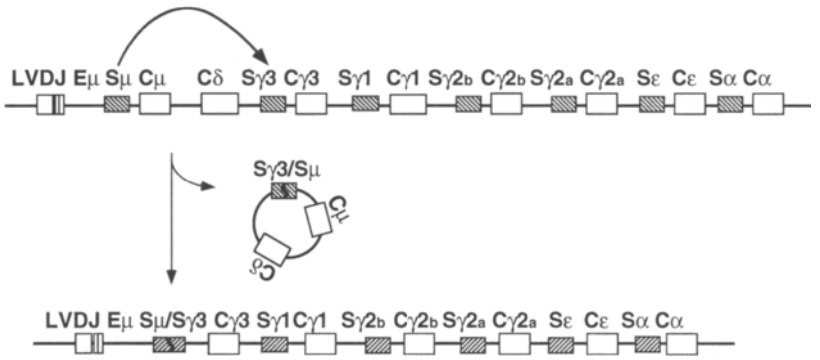


FIG. 1 Switch recombination. Switch recombination juxtaposes a rearranged variable region (VDJ) with a downstream C region. In the example shown, recombination is from μ to $\gamma 3$. A rearranged VDJ region is shown just upstream of the heavy chain intron enhancer, Eμ; switch regions (S) are shown by striped boxes. Circles carrying deleted DNA sequences are found in cells that have undergone switch recombination (4a–c).

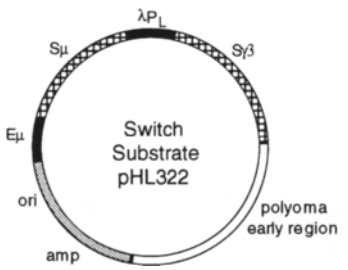
nation products for analysis. For these reasons, it has provided a valuable tool in the study of *cis*-acting elements that regulate switch recombination.

Extrachromosomal Recombination Assay

There are two steps to the extrachromosomal recombination assay, as outlined in Fig. 2.

Step 1. Recombination takes place in the first step, during transfection of mammalian B cells: Switch substrates are transfected into primary murine spleen cells that have been cultured with bacterial lipopolysaccharide (LPS) to induce switching at the chromosomal heavy chain loci, or into B cell lines. If recombination between S region sequences in the extrachromosomal constructs results in deletion of λP_L, it will generate a DNA molecule that can transform a nonlysogen to ampicillin resistance.

Step 2. Recombinants are assayed in the second step, by transformation of two *E. coli* strains, one a λ lysogen and the other a nonlysogen: Low molecular weight DNA is isolated from the transfected B cells. A compatible plasmid, pACYC184 (cm^R), is added to the recovered DNA as an internal control for DNA recovery and transformation. The recovered DNA is treated with *Dpn*I to enrich for molecules that have entered the nucleus and replicated in the eukaryotic cells. Deletion of λP_L as well as plasmid recovery are assayed by transformation of two *E. coli* strains, the



Two step assay for extrachromosomal switch substrate recombination

Step 1: Substrates Recombine in Mammalian B cells

Culture primary splenocytes with LPS
Transfect with recombination substrate on day 3
Isolate low molecular weight DNA on day 5

Step II: Assay for Recombination by Transformation of E. coli

Destroy unreplicated molecules by digestion with Dpn I
Transform E. coli strains DH10B and DH10B(λ)
Score ampicillin-resistant colonies on each strain

FIG. 2 Extrachromosomal switch substrate recombination assay. All substrates carry Sμ and Sγ3 switch region sequences flanking a conditionally lethal marker, the leftward promoter of phage λ (λP_L). Substrates are propagated on a shuttle vector carrying both a bacterial origin (ori) and ampicillin-resistance marker (amp), and polyoma sequences (Py) to support replication in murine B cells. The substrate pHL322 carries the immunoglobulin heavy chain intron enhancer, Eμ, just upstream of the Sμ region.

nonlysogen DH10B and the lysogen DH10B(λ). Recombination frequency is calculated as the ratio of ampicillin-resistant transformants in the nonlysogen and lysogen, normalized for the relative transformation efficiencies of the two strains.

Components of Extrachromosomal Switch Substrates

Extrachromosomal recombination substrates have several components, as shown in Fig. 2: the shuttle vector backbone, the switch regions, the conditionally lethal λP_L promoter, and an enhancer to stimulate switch substrate recombination.

Shuttle vector: The shuttle vector contains both the polyomavirus early region, to drive replication in murine B cells, and the pBR322 plasmid origin and ampicillin-resistance marker, to drive replication and facilitate selection in *E. coli.* An analogous shuttle vector has been used in constructs designed to assay V(D)J joining in lymphocytes (8, 9).

λP_L promoter: The λP_L promoter is a conditionally lethal marker because the high level of transcription from this promoter interferes with plasmid replication. Plasmids carrying λP_L will not give rise to ampicillin-resistant colonies in nonlysogens. In λ lysogens, the *c*I repressor protein produced by the endogenous prophage shuts off transcription from λP_L. Plasmids carrying λP_L can therefore be propagated in strains of *E. coli* that are lysogenic for phage λ.

S region sequences: The switch region sequences are 2- to 3-kb sequences subcloned from the murine Sμ and Sγ3 switch regions. These switch regions were used because chromosomal μ- to γ3 switching can be efficiently induced in primary B cells cultured with LPS. These sequences are stable during propagation in *E. coli* (see Controls, below). Efficient recombination depends on the presence of S region sequences in the substrates.

Enhancer: High frequency switch substrate recombination depends on the presence of an element in the construct that can function as a recombinational enhancer (6, 7). Either the cytomegalovirus (CMV) immediate early region promoter or the immunoglobulin heavy chain intron enhancer (Eμ) function comparably well as recombinational enhancers; the function of Eμ does not depend on the presence of a promoter. The substrate shown in Fig. 2, pHL322, carries the Eμ enhancer just upstream of the Sμ switch region. Following transfection of LPS-stimulated primary B cells with pHL322, about 24% of replicated substrates have undergone recombination, while without an enhancer, only about 2% of replicated molecules recombine (7). Gene targeting experiments have demonstrated the importance of the Eμ enhancer to chromosomal switch recombination *in vivo* (10).

Preparation of DNA for Transfection

1. Prepare supercoiled plasmid DNA for transfection by equilibrium gradient centrifugation or using Qiagen P-500 (Qiagen, Chatsworth, CA) or a comparable column.
2. To ensure that a plasmid preparation does not contain any contaminating plasmid or a jackpot of molecules that have lost the λP_L marker either by mutation or recombination, assay it by transformation of DH10B and DH10B(λ) (see below). The ratio of transformants of the nonlysogen compared to the nonlysogen should be about 5×10^{-5}. This control should be carried out on each plasmid DNA preparation before it is used for transfection.

Transfection of Murine B Cells

Transfection conditions are optimized by comparing expression of a reporter construct in cells transfected by electroporation or DEAE-dextran, using a range

of transfection conditions. Conditions are given below for transfection of LPS-stimulated primary murine B cells and cell lines representing several stages of B cell development.

Activation of Primary Murine B Cells with Lipopolysaccharide

Culturing primary murine B cells with LPS induces efficient switch recombination at the chromosomal heavy chain loci.

1. Using sterile instruments, remove spleens from three or more 8- to 12-week-old BALB/c mice that have been sacrificed by cervical dislocation. Place spleens into a petri dish containing about 3–5 ml of sterile, room temperature Hanks' balanced salt solution (HBSS) (GIBCO, Grand Island, NY) per spleen.
   Note: All subsequent manipulations should be carried out in a tissue culture hood under aseptic conditions.
2. Disaggregate each spleen by grinding between the frosted ends of two sterilized microscope slides. When all spleens have been disaggregated, use a pipette to transfer the cell suspension to a sterile disposable conical tube, leaving behind the spleen capsules and any fatty tissue. Rinse the petri dish with 3–5 ml of sterile, room temperature HBSS per spleen, again leaving behind the spleen capsules and any fatty tissue, and transfer the rinse liquid to the tube with the cell suspension. The final volume should be 8–10 ml/spleen.
3. Allow the cell suspension to settle for 3–5 min at room temperature, so that any clumps aggregate at the bottom of the tube.
4. Transfer the cell suspension to a clean tube, leaving behind the clumps.
5. Pellet the cells at 400 *g* for 10 min at room temperature. Remove the supernatant by decanting or by aspiration, taking care not to dislodge the pellet. Tap the bottom of the tube gently but firmly five or six times to disaggregate the pellet. Add about 8 ml of HBSS per spleen; invert gently to mix.
6. Remove 100 μl of cells for counting and for a cell viability check (step 7). Pellet the remaining cells at 400 *g* for 10 min at room temperature.
7. To determine cell number and viability, add 10 μl of cells to 90 μl of 0.02% (w/v) trypan blue, mix gently, then add 10 μl of these trypan blue-stained cells to a hemacytometer chamber for cell counting. Viable lymphocytes will appear large, round, and white; red cells will be smaller, elongate, and golden; and dead cells will be blue, as they do not exclude the dye. Typically, at least 90% of the lymphocytes are viable. After the 1:10 dilution with trypan blue, cell concentration can be calculated as

$$\text{Cells/ml} = 10^5 \times \text{number of cells in central grid of hemacytometer}$$

Each spleen contains from 0.5×10^8 to 1×10^8 lymphocytes, depending on the size of the mouse, so there will be about 50–100 cells in the central grid.

8. Resuspend the pelleted cells (step 6) in culture medium at $1-2 \times 10^6$ cells/ml, and transfer to a sterile plastic tissue culture flask. Culture medium is RPMI 1640 supplemented with 10% (v/v) fetal bovine serum, 20 μM 2-mercaptoethanol, 2 mM L-glutamine, LPS (40 μg/ml) (from *E. coli;* Sigma, St. Louis, MO), penicillin (50 U/ml), streptomycin (50 μg/ml), gentamicin (50 μg/ml). Lipopolysaccharide can be made up at 10 mg/ml in water, filter sterilized, and stored at −20° C. Culture medium should be prewarmed to about 37° C and preequilibrated in 5% CO_2. *Warning:* LPS is potentially toxic. It should be handled with care.
9. Culture the cells for 3 days at 37° C, 5% CO_2. By this time, a majority of lymphocytes should have enlarged and appear to be forming blasts. Chromosomal switching to immunoglobulin γ isotopes (IgG) can be assayed by staining with fluorescein isothiocyanate (FITC)-conjugated rabbit anti-mouse IgG, followed by analytical sorting on a fluorescence-activated cell sorter (FACS).

Transfection of Lipopolysaccharide-Activated Primary Murine B Cells

When LPS-activated primary murine B cells are transfected with switch substrates, efficient recombination can occur.

1. Wash the cells once with one-half volume of STBS [25 mM Tris-HCl (pH 7.4), 137 mM NaCl, 5 mM KCl, 0.6 mM Na_2HPO_4, 0.7 mM $CaCl_2$, 0.5 mM $MgCl_2$].
2. Resuspend at 10^7 cells/ml in STBS containing DEAE-dextran (100 μg/ml) [Pharmacia (Piscataway, NJ), 5×10^5 daltons; this can be made up as a 10-mg/ml stock in STBS and filter sterilized].
3. Incubate 5×10^7 cells (5 ml of the final suspension) with 10–20 μg of purified plasmid DNA in a 50-ml conical tube for 15 min at 37° C.
4. Add 30 ml of room temperature STBS and pellet the cells immediately.
5. Wash twice with 30 ml of STBS.
6. Resuspend in 40 ml of prewarmed culture medium containing LPS (see above).
7. Culture the cells an additional 40–48 hr prior to DNA isolation. Cells should maintain excellent viability throughout the period of incubation, as assayed by trypan blue exclusion.

Transfection of Cultured B Cell Lines

Two different methods of transfection have been used with cultured B cell lines: DEAE-dextran and electroporation. DEAE-dextran transfection has been found to be optimal with the pre-B cell line PD31, the B cell line A20, and the plasmacytoma

cell line J558L. Electroporation has been found optimal for the myeloma Ag8 and hybridomas made by fusion of primary cells with Ag8.

1. Culture cells in RPMI 1640 supplemented with 10% (v/v) fetal bovine serum, 20 μM 2-mercaptoethanol, 2 mM L-glutamine, penicillin (50 U/ml), streptomycin (50 μg/ml), at 37° C, 5% CO_2.
2. Harvest the cells in log phase (5×10^5–1×10^6 cells/ml) and pellet the cells at 400 g for 10 min at room temperature.
3. Wash once with one-half volume RPMI 1640 medium (medium alone, without supplements). While the cells are washing, count the viable cells with trypan blue as described above.

For DEAE-Dextran Transfection

4. Resuspend the cells at 5×10^7 cells/ml in RPMI 1640 medium containing DEAE-dextran (500 μg/ml) (prepared as a 10-mg/ml stock in STBS).
5. Transfer 1 ml of cells (5×10^7 cells) to a sterile 15-ml conical centrifuge tube. Add 10 μg of purified plasmid DNA, mix gently, and incubate at 37° C for 30 min.
6. Add 8 ml of RPMI 1640 medium (no supplement), pellet the cells at 400 g for 10 min at room temperature, and aspirate the supernatant.
7. Wash the cells once with 10 ml of RPMI 1640 medium.
8. Resuspend the cells in 30 ml of prewarmed culture medium.
9. Culture the cells for 40–48 hr prior to DNA isolation.

For Transfection by Electroporation

4. Resuspend the cells at 1.25×10^8 cells in 0.8 ml of PBS at room temperature (PBS: 120 mM sodium chloride, 20 mM potassium phosphate, pH 7.2–7.4). Addition of DEAE-dextran (100 μg/ml) to the electroporation mixture (11) enhances transfection of certain hybridomas generated by fusion with Ag8, although not of Ag8.
5. Aliquot 0.8 ml of cells (1×10^8 cells) to a 0.4-cm gap electroporation cuvette (Bio-Rad, Richmond, CA). Add 10 μg of DNA.
6. Electroporate at 350 V, 960 μF on a Bio-Rad gene pulser. The time constant should be about 12–15 msec.
7. Immediately resuspend the cell suspension in 30 ml of culture medium. The dead cell cluster caused by electroporation should be removed prior to culture.

Isolation of Low Molecular DNA from Transfected Cells

Low molecular weight DNA is isolated 42 hr after transfection by alkaline lysis, following a procedure similar to that described by Hesse *et al.* (8).

1. Pellet the transfected cells in a 50-ml conical tube by centrifugation at 1200 rpm for 10 min at room temperature.
2. Wash once with 25 ml of PBS (120 m*M* sodium chloride, 20 m*M* potassium phosphate, pH 7.2). After centrifugation, aspirate the supernatant and resuspend the cell pellet in 1 ml of PBS.
3. Aliquot 0.5 ml/tube to two 1.5-ml Eppendorf tubes. Spin for 10 sec at room temperature to pellet cells. Aspirate the supernatant. (*Note:* Increasing the numbers of tubes in this step and in step 6 results in a cleaner preparation of DNA, which is essential to efficient transformation).
4. Add to each cell pellet 200 μl of GET (GET is 50 m*M* glucose, 10 m*M* EDTA, 25 m*M* Tris, pH 8); vortex to resuspend. Add 400 μl of freshly made 0.2 *N* NaOH–1% (w/v) sodium dodecyl sulfate (SDS), and mix gently but well. Incubate at room temperature for 5 min.
5. Add 300 μl of 3 *M* sodium acetate, pH 4.8; mix gently but well. Incubate on ice for 20–30 min. Spin for 10 min at top speed in a microcentrifuge to pellet cell debris.
6. Transfer the supernatant from each tube to two 1.5-ml Eppendorf tubes. There will now be four Eppendorf tubes to work with, each containing about 400 μl. Add to each tube 1 ml of 100% ethanol. Incubate for 10 min on dry ice, then spin for 10 min at 12,000 rpm in a microcentrifuge to pellet DNA. Aspirate the supernatants.
7. Pool the four pellets into two tubes, by combining two pellets per 375 μl of sterile water. To each tube add 25 μl of 5 *M* ammonium acetate, vortex to mix, then add 800 μl of 100% ethanol. Incubate for 10 min on dry ice and then spin for 10 min at 12,000 rpm in a microcentrifuge to pellet DNA.
8. Rinse each pellet twice with 100 μl of 75% (v/v) ethanol (made by diluting 100% ethanol with water). Carefully remove the ethanol by aspiration and let the pellets air dry briefly at room temperature.
9. Resuspend and pool the two pellets in a final volume of 30 μl of sterile water.

Digestion with *Dpn*I to Destroy Molecules That Have Not Replicated during Transfection

To enrich for molecules that have entered the mammalian nucleus during transfection, the low molecular weight DNA recovered from transfection is digested with the restriction nuclease *Dpn*I. *Dpn*I recognizes the sequence GATC but cuts only at sites at which the A in both strands is methylated. *Escherichia coli* efficiently methylates GATC sequences, but after replication in the eukaryotic cell this prokaryotic modification will be lost. Molecules that are *Dpn*I resistant must therefore have undergone at least one round of replication in the mammalian cell. A single cleavage in a supercoiled DNA molecule will render it incapable of transformation, and as each sub-

strate contains numerous *Dpn*I sites (e.g., pHL322, shown in Fig. 2, has 34 *Dpn*I sites) resistance to *Dpn*I is a rigorous enrichment for replicated DNAs.

1. To 30 μl of low molecular weight DNA preparation, add 1 ng of pACYC184 (New England BioLabs, Beverly, MA) as a transformation standard for subsequent steps in the experiment. This is a pBR-compatible plasmid that carries a chloramphenicol resistance marker. It should be grown on an *E. coli dam*$^-$ strain, such as GM2727, so that the DNA is not methylated and therefore not digested by *Dpn*I in the next step.
2. Digest the DNA preparation with 5–10 units of *Dpn*I, under conditions suggested by the manufacturer. Incubate for 2 hr at 37° C.
3. Add 158 μl of water, 12 μl of 5 *M* ammonium acetate, vortex to mix, then add 400 μl of 100% ethanol and incubate for 10 min on dry ice. Spin for 10 min at top speed in a microcentrifuge to pellet the DNA; aspirate the supernatant.
4. Reprecipitate by adding 180 μl of water to the pellet, vortexing well to resuspend, then adding 12 μl of 5 *M* ammonium acetate, 400 μl of 100% ethanol, and chilling and centrifuging as above.
5. Aspirate the supernatant; rinse the pellet twice with 100 μl of 75% (v/v) ethanol. After removing the ethanol allow the pellet to air dry.
6. Resuspend in 20 μl of water and store at −20° C until transformation.

Preparation of Competent Bacterial Cells for Electroporation

Recovered plasmid DNA is assayed by transformation of bacterial cells. To obtain reliable data, transformation efficiency must be on the order of $0.5-1 \times 10^9$ transformants per microgram of 3-kb plasmid (pUC or pBluescript are appropriate controls). This requires that transformation be carried out by electroporation, and that the electrocompetent cells be carefully prepared.

1. Streak DH10B and DH10B(λ) on rich agar plates; grow overnight at 37° C.
2. Starting with cells from a single colony, grow 5 ml overnight culture in SOB [2.0% (w/v) Bacto Tryptone, 0.5% (w/v) yeast extract, 10 m*M* NaCl, 2.5 m*M* KCl].
3. Dilute 0.5 ml from the overnight culture into 500 ml of SOB (2% w/v Bacto-Tryptone, 0.5% w/v yeast extract, 0.05% w/v NaCl) in a 2-liter baffle flask (Bellco Glass, Vineland, NJ) and grow with excellent aeration at 37° C, until the OD_{600} is 0.8. This takes about 4 hr. (See note below regarding flasks.)
   Note: The importance of keeping everything cold during preparation of electrocompetent cells cannot be overstated. In the following steps, all solutions are prechilled in an ice–water bath to ensure that they are at 0° C; and all tubes, pipettes, and pipette tips are prechilled for several hours prior to use.
4. Chill the flask containing cells for 15 min in an ice–water bath in a cold room. Decant the cells to two sterile, chilled 250-ml centrifuge bottles, spin at 5000 rpm

for 10 min at 4° C in a Sorvall GSA rotor, Beckman JA14 rotor, or the equivalent. (See note below regarding centrifuge bottles.)
5. Pour off the supernatants; briefly drain the cell pellets. Resuspend each pellet in 250 ml of ice-cold 10% (w/v) glycerol and spin at 5000 rpm for 15 min at 4° C.
6. Pour off the supernatants, briefly drain the cell pellets, pool the pellets in 250 ml of ice-cold 10% (w/v) glycerol in a single centrifuge bucket, and spin at 5000 rpm for 15 min at 4° C.
7. Pour off the supernatant, briefly drain the cell pellet, and resuspend it in the small amount of liquid remaining, with no added liquid. The final volume should be about 2 ml.
8. Aliquot 40-μl aliquots into Eppendorf tubes, freeze immediately on dry ice, and store at $-70°$ C. Cells can be stored at least 1 year at $-70°$ C without apparent loss in transformation efficiency. A single 40-μl aliquot from a good preparation of competent cells can be used for 10–20 separate transformations.
9. Test transformation efficiency by transforming cells with 100 pg of pUC or pBluescript (see below). The transformation efficiency should be $0.4-1.0 \times 10^9$ transformants per microgram.

Note: The sensitivity of transformation by electroporation is sufficiently high that it will score even a minute amount of contaminating plasmid, and this in turn will jeopardize the transformation assay. To be sure that the baffle flasks and centrifuge bottles contain no trace contaminating plasmid, they should be washed with Clorox and thoroughly rinsed before autoclaving. Alternatively, new glass/plasticware can be dedicated exclusively to preparation of competent cells. This is especially important in laboratories that work routinely with plasmids.

Electroporation of Competent Bacterial Cells

1. Thaw competent cells on ice. Dilute with an appropriate volume of sterile 10% (w/v) glycerol at 4° C.
2. Aliquot 40 μl of cells to a prechilled 0.1 cm gap electroporation cuvette (Bio-Rad). Add 2–4 μl of DNA.
3. Electroporate at 1.8 kV, 200 Ω, 25 μFD on a Bio-Rad GenePulser. (The Teflon cuvette holder from the GenePulser may be prechilled at $-20°$ C before the electroporations are carried out.) The time constant should be about 4.7 msec.
4. Immediately add 0.5–1 ml of SOC and transfer to a sterile culture tube with a loose-fitting top. (SOC: SOB containing 10 m*M* glucose, diluted from a 1 *M* glucose stock solution that has been filter sterilized, not autoclaved; the 1 *M* glucose stock solution may be stored at $-20°$ C.)
5. Grow transformed cells at 37° C for 1 hr with good aeration.
6. Following electroporation, a culture may be somewhat viscous after growth due to cell lysis. If so, to ensure that volume measurements are accurate, dilute 1:5

into 10 m*M* $MgSO_4$ and mix well prior to plating under selective conditions. To assay for ampicillin-resistant transformants, plate on LB-agar [LB-agar is 0.8% (w/v) Bacto-Tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 1.5% (w/v) agar] containing carbenicillin (50 μg/ml). Carbenicillin (made up as a 50-mg/ml stock in water, stored at $-20°$ C) is used instead of ampicillin because this drug is much more stable and only marginally more expensive. To assay for chloramphenicol-resistant transformants, plate on LB-agar containing chloramphenicol (20 μg/ml) (made up as a 20-mg/ml stock in 100% ethanol, stored at $-20°$ C). Grow overnight at 37° C.

Agar plates should be sufficiently dry that no drips form on the surface during overnight incubation. If plates have been poured recently, they can be dried by incubation for 30–60 min at 37° C with lids ajar.

Controls for Transformation

In each set of transformation assays, two controls are included: a negative control, in which no exogenous DNA is added; and a positive control, in which 100 pg of control plasmid, such as pUC, is added to assay transformation frequency. Control plasmid is diluted from a concentrated stock to a final concentration of 0.1 μg/ml (100 pg/μl), and the dilute solution aliquoted and the aliquots stored frozen and used only once. (If stored at 4° C, the transforming capacity of the DNA diminishes with time, probably because the dilute DNA in solution adheres to the walls of the tube.)

Analysis of Recombination Frequency

The recombination frequency, *R*, is calculated as the ratio of ampicillin-resistant (amp^R) transformants of the nonlysogen compared to the lysogen; this value is normalized for transformation efficiency as measured by chloramphenicol-resistant transformants.

$$R = \left[\frac{\text{Amp}^R \text{ transformants of DH10B}}{\text{Amp}^R \text{ transformants of DH10B}(\lambda)}\right] \times \left[\frac{\text{cm}^R \text{ transformants of DH10B}(\lambda)}{\text{cm}^R \text{ transformants of DH10B}}\right]$$

Each transfection is carried out at least twice, with at least two different DNA preparations; and DNA from each transfection is analyzed in at least two separate transformations.

Analysis of Recombinants

To map recombinants, plasmid DNA is prepared from DH10B transformants using a standard minipreparation protocol and analyzed by restriction digestion. Recombination in LPS-activated primary B cells and certain B cell lines generates deletions that remove λP_L, and these deletions are quite distinct from the point mutations or small deletions that occur at a low level during propagation in *E. coli* (see below). Deletion end points are heterogeneous, as expected from independent recombination events mediated by the switch recombination apparatus.

Controls to Verify That Assay Quantitates Recombination Events That Occur during Transfection of Mammalian Cells

The rate of spontaneous mutation of λP_L during propagation in *E. coli* is about 2×10^{-5}, and the mutations that occur during bacterial propagation are all point mutations or deletions too small to be apparent on restriction mapping. Prior to use for transfection, each preparation of DNA is assayed by transformation to verify that fewer than 10^{-4} molecules can transform the nonlysogen. This ensures that molecules identified as recombinants after transfection do not derive from contamination or from a jackpot of mutants in the DNA preparation.

Constructs carrying $S\mu$–λP_L–$S\gamma 3$ are stable during propagation in λ lysogens; in particular, the S region sequences do not stimulate recombination. Constructs based on a pUC backbone (high copy number) appear to undergo more recombination during bacterial propagation than do constructs based on the pBR322 backbone (lower copy number). For this reason the pBR322 backbone was used in construction of the extrachromosomal substrates.

Restriction mapping of recombinants isolated from DH10B, described above, confirms that molecules carry heterogeneous deletions of λP_L which are characteristic of products of switch recombination.

Acknowledgments

This research was supported by NIH Grants R01 GM39799 and P01 CA16038.

References

1. T. K. Blackwell and F. Alt, *Annu. Rev. Genet.* **23,** 605 (1989).
2. C. Esser and A. Radbruch, *Annu. Rev. Immunol.* **8,** 717 (1990).
3. R. L. Coffman, B. W. P. Seymour, D. A. Lebman, D. D. Hiraki, J. A. Christiansen,

B. Shrader, H. M. Cherwinski, H. F. J. Savelkoul, F. D. Finkelman, M. W. Bond, and T. R. Mosmann, *Immunol. Rev.* **102,** 5 (1988).
4. M. Hengstschläger, H. Leung, and N. Maizels, *Prog. Nucleic Acids Res. Mol. Biol.,* in press (1996).
4a. U. von Schwedler, H.-M. Jack, and M. Wabl, *Nature (London)* **345,** 452 (1990).
4b. M. Matsuoka, K. Yoshida, T. Maeda, S. Usuda, and H. Sakano, *Cell* **62,** 135 (1990).
4c. T. Iwasato, A. Shimizu, T. Honjo, and H. Yamagishi, *Cell* **62,** 143 (1990).
5. W. Dunnick, G. Z. Hertz, L. Scappino, and C. Gritzmacher, *Nucleic Acids Res.* **21,** 365 (1993).
6. H. Leung and N. Maizels, *Proc. Natl. Acad. Sci. U.S.A.* **89,** 4154 (1992).
7. H. Leung and N. Maizels, *Mol. Cell. Biol.* **14,** 1450 (1994).
8. J. E. Hesse, M. R. Lieber, M. Gellert, and K. Mizuuchi, *Cell* **49,** 775 (1987).
9. M. R. Lieber, J. E. Hesse, K. Mizuuchi, and M. Gellert, *Genes Dev.* **1,** 751 (1988).
10. H. Gu, R.-R. Zou, and K. Rajewsky, *Cell* **73,** 1155 (1993).
11. G. H. Gauss and M. R. Lieber, *Nucleic Acids Res.* **20,** 6749 (1992).

[22] Gene Amplification as Marker for Studying Genomic Instability

T. D. Tlsty

Introduction

It has long been recognized that genomic instability is widespread in tumor cells. Evidence for instability is manifested in the continually changing karyotype that is evident in the chromosome analysis of tumor biopsy materials. Many types of changes are evident on both gross and molecular levels. Three categories of abnormalities are detected in tumor cells. Aneuploidy, which involves the gain or loss of a whole chromosome, probably results from defects in the machinery that is responsible for the proper segregation of chromosomes. A second category, gene rearrangements includes such changes as deletions, amplifications, inversions, translocations, and chromosome breaks. All these aberrations share the common defect of improper chromosome breakage and religation. Finally, on a molecular level we know that point mutations occur. The generation of these base changes involves the action of repair enzymes such as exonucleases and ligases.

Several years ago it was postulated that one of the first steps in the generation of transformed cells could be the acquisition of increased genomic instability (1). The karyotypic changes that result could then fuel the phenotypic changes that culminate in transformation. More recently, data that support this hypothesis have become available. To study genomic instability (or conversely, the maintenance of genomic integrity) it would be helpful to focus the efforts on one type of abnormality. As the molecular basis of this change is determined, its involvement in other types of abnormalities could be examined. Our laboratory has adopted this approach in our study of genomic integrity of human cells.

Some years back we posited that the in-depth genetic and molecular analysis of one type of genome rearrangement might give insights into the cellular processes that maintain genomic integrity or promote genomic lability. To this purpose we initiated studies that examined the regulation of gene amplification. We have used amplification ability as a marker of genomic instability and as a prototype of how the acquisition of genetic instability may be involved in neoplastic initiation and progression.

Gene amplification is an excellent marker for genomic instability for a number of reasons. For several decades, the manifestations of gene amplification, homogeneously staining regions and double minute chromosomes, have been observed in neoplastic tissue, although at the time the molecular basis of these chromosomal abnormalities was not known. In the 1970s, studies by Biedler and Spengler (2) associated these chromosomal abnormalities with drug resistance; Alt and co-workers (3) found

Methods in Molecular Genetics, Volume 8

that the basis of the resistance in this instance was an increase in gene copy number (gene amplification). These observations prompted investigations into the identity of the sequences that are carried on homogeneously staining regions and double minute chromosomes in human tumor biopsy materials and led to the discovery that oncogenes are often amplified, especially in certain tumor types. Additionally, gene amplification is an example of a rearrangement that is easy to measure on a molecular level. Unlike random chromosomal breaks of translocations, amplification of a targeted sequence can be measured using colony formation and easily verified by determination of gene copy number. Finally, some studies have suggested a relationship between the frequency and extent of oncogene amplification and the progression of malignancy.

A powerful method for studying the frequency of gene amplification is to assay the generation of drug-resistant colonies. In this type of clonogenic assay only the cells that have acquired an increase in gene copy number will survive the selection protocol and produce colonies. Analysis of the colonies then indicates the status of the genome.

One could choose to study genetic instability using *N*-(phosphonoacetyl)-L-aspartate (PALA) resistance and the ensuing amplification of the *CAD* gene because, unlike methotrexate resistance which may occur through multiple mechanisms, the major mechanism of resistance to PALA is through amplification of the *CAD* gene (4). The *CAD* gene encodes the multifunctional CAD enzyme, and PALA inhibits the aspartate transcarbamylase activity of this enzyme. Thus, in asking whether tumorigenic cells amplify DNA sequences more often than nontumorigenic cells, one can use the incidence of resistance to PALA as an indirect indicator of the ability of a cell to amplify DNA. This chapter describes the measurement of *CAD* gene amplification frequency in monolayered mammalian cells.

Measurement of incidence of amplification is a more rapid determination than measurement of rate. The incidence measurement entails characterization of the cell lines for plating efficiency, population doubling time, initial sensitivity to the drug, and quantification of the initial gene copy number. Once the cell line is characterized it is challenged at a stringency of selection equivalent to $9 \times LD_{50}$ (50% lethal dose), and the number of resistant colonies is tabulated. Gene copy number, karyotype, and cytometric analysis of DNA content are determined to verify that gene amplification (not polyploidy) is the mechanism of resistance. One can use this method with most cell lines to estimate their propensity to amplify.

Measuring Incidence of Gene Amplification

The advantages of using this method are several. First, this assay is rapid. It consists of determination of the plating efficiency and LD_{50} followed by subsequent determination of the incidence of drug resistance (percent survival) at $9 \times LD_{50}$. Although

the time required for these measurements is dependent on the doubling times of the cells, determination for the lines generally required approximately 2 months. Second, this assay is direct. Because amplification is the major reported mechanism for resistance to PALA, complications in interpreting the underlying basis of resistance appear to be avoided by using this drug, thus allowing direct assessment of amplification ability from the incidence of drug resistance. Third, this assay is quantitative; it allows the comparison of cell lines with differing growth properties or sensitivities to the drug. Finally, this assay determines the incidence of gene amplification, in contrast to other studies that measure rate (5). Both values are determined by (a) the rate at which the cells are generated, (b) their intrinsic stability, and (c) their growth rates relative to the rest of the cell population. This assay, which measures incidence, accounts for these variables and provides a sensitive and direct measurement of gene amplification. Rate measurements, however, have an additional component of time and the preceding three variables should be taken into account at each step in the selection (rate determination).

Measurement of the incidence of gene amplification has an added advantage in that it can measure *CAD* gene amplification in cell populations that have a limited capacity for proliferating. This is particularly important in analyzing mortal cell populations. A single cell must double 6 times, at most, to be visible as a drug-resistant colony (>50 cells). This is in contrast to the second method of analyzing amplification potential: the use of the Luria–Delbrück fluctuation analysis. In the Luria–Delbrück fluctuation assay one uses the conditions previously determined for the incidence measurements and applies them to the estimation of the amplification rate. Plating efficiency and LD_{50} determination are particularly relevant here. The fluctuation analysis requires propagation of the population before placing the cells in selection. To accurately measure an amplification rate on the order of 10^{-6}, a population of 100 cells must be expanded to 10^6–10^7 cells. A total of 18–25 population doublings is required. At this point, the cells are placed in selection and an additional six cell doublings are required to produce a visible colony. Both methods have given comparable results with the cell lines examined (6).

In designing the incidence assay to measure amplification ability in these lines, one wants to develop as rapid an assay as possible. To this end, one would choose to use a single-step selection protocol and determine the minimal concentration of selective drug that allowed the correlation of drug resistance with amplification ability. Cells placed in a low stringency of selection ($3 \times LD_{50}$) produced colonies within 1 week. Rechallenging these colonies with the same drug concentration during clonal propagation revealed that a large proportion of the cells were killed, with the remainder (those used for molecular analyses) growing very slowly. This result indicates that these cells were not completely resistant to PALA but were merely capable of tolerating the low drug concentrations. In contrast, cells placed in a higher stringency of selection ($9 \times LD_{50}$) take 2–4 weeks to emerge as colonies. Cells from virtually all of these colonies were resistant to killing by the drug and grew in the

drug at a rate similar to that of unselected cells during clonal propagation. These observations are consistent with our results showing that subclones resistant to concentrations of PALA are equivalent to $3 \times LD_{50}$ were not amplified, while those grown at $9 \times LD_{50}$ were amplified in each case (7). It is important to note that at drug concentrations greater than $9 \times LD_{50}$, greater amplification can result, and the relationship between the incidence of drug resistance and tumorigenicity can remain. Placing cells at drug doses that are very high, however, often eliminates the appearance of colonies. It is probable that a single amplification event can provide only a limited degree of resistance.

Two variables are especially important in the determination of incidence of amplification ability: the plating efficiency and LD_{50} of each line. The plating efficiency, the baseline number of colonies obtained without selection (defined as 100% survival), is the value to which the number of colonies obtained after drug selection is compared. Overestimating or underestimating this number would lead to an incorrect determination of the incidence of drug resistance. Likewise, accurate determination of the LD_{50}, the pivotal value for the final stringency of selection, is required.

Measuring Rate of Gene Amplification

The Luria–Delbrück fluctuation analysis is a statistical method that allows one to distinguish between variant cells arising by rare spontaneous mutations and variant cells arising through adaptation to an environmental selection (8). Figure 1 shows a schematic diagram of a fluctuation experiment and the hypothetical results one could obtain when analyzing a spontaneous mutation. A large parent population of cells is procured and the experimental samples are divided into two categories. The left arrow indicates the first category, which contains replicate samples in which an appropriate aliquot of cells from the parent population is plated directly into selection medium and analyzed for the number of resistant colonies that emerge. The colonies on these plates will represent rare resistant mutants. The number of colonies on each plate should follow a Poisson distribution and the mean number of colonies per plate will reflect the prevalence of resistant mutants in the parental population. Replicate platings from the same parent population should show variation due to random sampling only; the variance from these replicate samples should equal the mean. This category of samples is used to demonstrate that the method of sampling and plating has not introduced any fluctuations into the results besides random sampling error.

The right arrow in Fig. 1 indicates replicate samples of the second category. A small aliquot of cells (small enough to assure no preexisting mutants are present) is plated and allowed to propagate under nonselective conditions for a given amount of time. When the individual populations have reached the same cell density plated in category 1, they are transferred to fresh plates for an even distribution of cells and

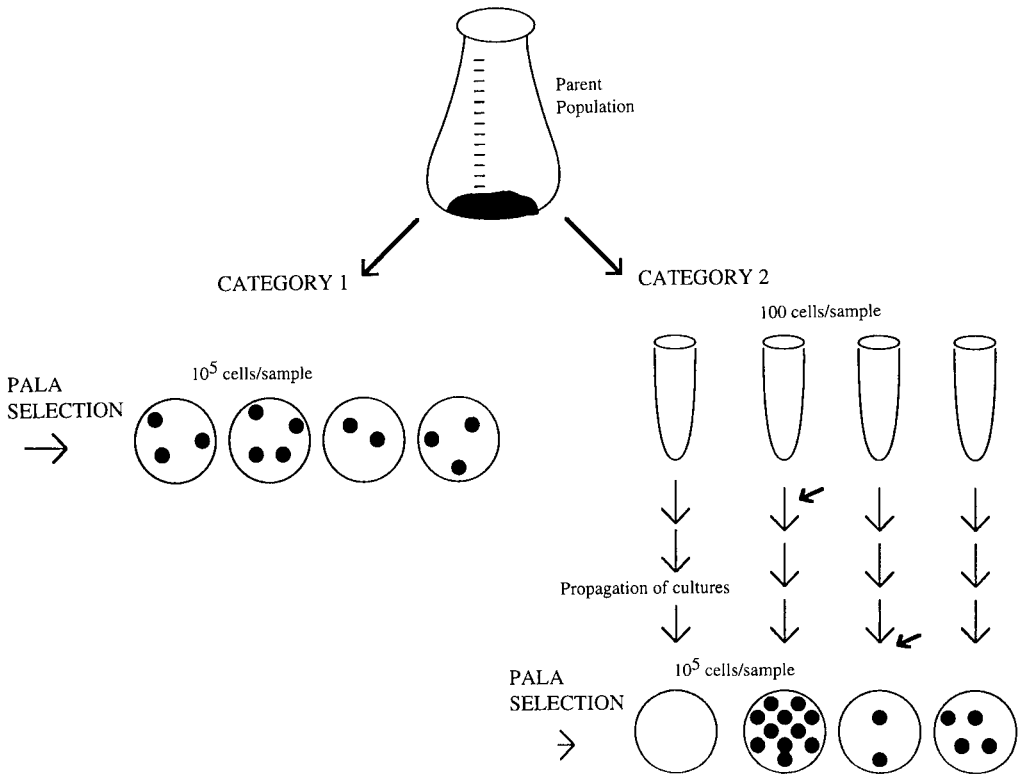


FIG. 1 Luria–Delbrück fluctuation analysis.

placed under selection (in this case PALA). If the drug-resistant cells are the result of exposure to the PALA (i.e., adaptive), each cell should have a probability for survival in selective media and the appearance of resistant mutants should be similar in all plates. The variation from plate to plate will be consistent with the Poisson model. If the events are spontaneous, as each parallel culture expands it will have a given probability for generating resistant mutants with each cell division. In some cultures the event (amplification) will occur early (as indicated by the small horizontal arrow in the second sample) and many of the progeny of the resistant cell will be present to form drug-resistant colonies. In others, the event will occur during one of the last cell divisions (as indicated by the small horizontal arrow in the third sample) and few drug-resistant progeny will result. The appearance of mutants will be random and the contribution to the surviving colonies when the cells are placed in selection will vary greatly depending on when in the propagation of the population the mutation occurred. Statistical analysis of this variation allows one to calculate the rate of appearance of spontaneous mutants.

Variables

In analyzing cells for their incidence of *CAD* gene amplification, particular attention should be paid to the following considerations.

Analysis of Multiple Subclones

Clonal heterogeneity exists in most populations for any marker that is being analyzed. The analysis of *CAD* gene amplification is no exception. Previous work has shown that populations that have been passaged for an extended period of time exhibit heterogeneity in their ability to amplify their dihydrofolate reductase gene (9). Newly subcloned populations, on the other hand, are fairly homogeneous. The determination of amplification frequency should be conducted with several isolates of each subclone.

Characterization of Subclones

As noted earlier, several characteristics of a cell line must be determined before a successful selection may be carried out. The growth rate of the cellular populations as well as the plating efficiencies of the subclones must be determined as described in Methods.

Determination of LD_{50}

Of paramount importance in designing a selection study is the initial subclonal sensitivity to the drug. A common parameter that is used for characterization is the mean lethal dose at which 50% of the population is eliminated by the drug (LD_{50}). To accomplish this, predetermined numbers of cells should be placed in various concentrations of PALA and incubated. The concentration of PALA that achieves a 50% kill of the population will be designated as the LD_{50}.

Initial CAD Gene Copy Number

To study *CAD* gene amplification, we must know the initial *CAD* gene copy number in the unchallenged population. DNA should be isolated from each cell line as described in Methods. *CAD* gene copy number and arrangement in the genome should be analyzed by Southern hybridization and slot blots. Selective hybridizations with

additional markers should allow one to determine if the chromosome carrying the *CAD* gene is present in single or multiple copies, thereby allowing discrimination between amplification and aneuploidy. Karyotypic analysis of the cell lines should be performed. Alternatively, one can use the sensitivity of fluorescent *in situ* hybridization to visualize the endogenous gene number. Both procedures are described under Methods.

Density Effects

For many cell lines, the density effects of plating increasing numbers of cells are negligible. If cells generate 1 to 2 colonies when 100 cells are plated, they generate between 10 and 20 colonies when 1000 cells are plated and between 100 and 200 colonies when 10^4 cells are plated. However, there are two considerations in this regard. The first is that if the cells are too close together (i.e., too close to confluence) a proper selection pressure cannot be achieved and a density effect is evident. Cells should always be plated at a density that is well below subconfluence. If the frequency of amplification is so low that millions of cells need to be analyzed, fewer cells in a greater number of plates will give a better determination than a greater number of cells in fewer plates. The second consideration has to do with cell volume. Cells vary in size and the area they occupy on the plating surface. In order for a proper selection to occur, the cells should be well spaced. This means that cells that occupy a greater area on the plate will need to be plated at a lower density to obtain accurate colony measurements.

Characterization of PALA-Resistant Mutants for CAD Gene Amplification

The PALA-resistant subclones that emerge from the selections should be characterized for their *CAD* gene copy number as described in Methods. The most direct route for determining if *CAD* gene amplification is the basis of PALA resistance in a cell population is to quantitate its *CAD* gene copy number through hybridization with labeled probes. Normalization to the copy number of an endogenous gene is an important part of this analysis because variations in DNA concentration in individual samples could lead to false positives for amplification of the gene in question. This conceptual control becomes even more critical in tumorigenic cells because of the reported increase in frequency of polyploid cells and cells that endoreduplicate; two (or more) complete genomes would be present. Two additional classes of chromosomal abnormalities can also lead to an increase in gene copy number and must be ruled out before amplification as a mechanism of resistance is claimed: (a) formation of a polyploid cell line could be resistant to PALA without selective amplification of its *CAD* gene; (b) formation of a cell that is aneuploid for the chromosome carrying the *CAD* gene.

We can determine the formation of polyploid cells through karyotypic analysis of the PALA-resistant cell lines. Cells are collected during mitosis from parental cell lines and from the PALA-resistant sublines. Chromosome counts are performed on fixed preparations. Mean chromosome counts for each population will reveal whether polyploid cells have given rise to the resistant cell line. One can anticipate that the majority of cells will not be polyploid. One can detect the presence of a trisomic cell through hybridization with a panel of gene probes. Hybridization with selected gene markers will allow determination of the chromosome dosage in each cell line.

Frequency of Gene Amplification

The frequency (or incidence) of gene amplification can be determined by graphing the relative plating efficiency (RPE) versus the concentration of PALA expressed in relative LD_{50}. The RPE is the number of PALA-resistant colonies normalized to the number of colonies that have emerged in the nonselected population. As indicated previously, the determination of gene amplification frequency is made with the number of colonies observed at $9 \times LD_{50}$ to ensure that one is dealing with true drug-resistant colonies. If the protocol of the experiment has not provided for cells to be plated directly at $9 \times LD_{50}$, this number can be extrapolated from data that bracket this dose. Figure 2 depicts representative data from a typical selection experiment.

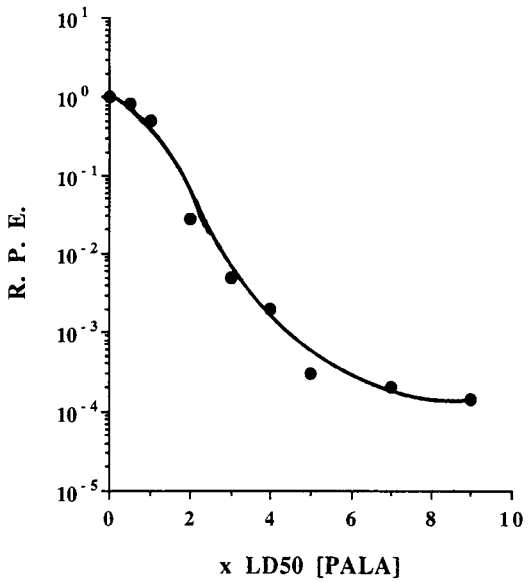


FIG. 2 Relative plating efficiency (RPE) versus concentration of PALA to determine frequency of gene amplification.

Methods

All cell lines should be grown in α-modified Eagle's medium (α-MEM) without deoxynucleotides, supplemented with 10% (v/v) dialyzed fetal calf serum, penicillin, and streptomycin and kept at 5% CO_2. It is critical that the serum be adequately dialyzed so as to remove low molecular weight ribo- and deoxyribonucleotides. If these molecules are present, the metabolic selection does not proceed because nucleotides are provided exogenously, short-circuiting the selection. Exponentially growing cells should be used for all experiments. Stock cell lines should be used for a maximum period of 3 months, at which time frozen aliquots of cells should be thawed for use in subsequent experiments to minimize changes resulting from extended propagation. As an additional control, dialyzed serum from a single lot should be used in all experiments.

Plating Efficiencies and Cell Cycle Times

To determine the plating efficiencies of unselected lines, 100 cells of each will be seeded into 100-mm-diameter dishes containing complete medium in a uniform monolayer. After 6–10 hr this medium will be replaced with fresh medium to remove any unattached cells. During this period of time, cells have an adequate amount of time to attach but not enough time to divide (forming two cells). If cells are too close to each other, adequate selection cannot proceed and calculations are confounded. Colonies should appear within 5–7 days; they will be fixed with 3:1 (v/v) methanol–acetic acid, stained with 2% (w/v) Giemsa (Gurr's), and only those colonies with more than 50 cells will be scored. The population doubling time for each line can be determined by plating cells into 100-mm-diameter dishes at a density of 5×10^4 cells/dish. Each day, plates should be trypsinized and the total number of cells determined. The population doubling time is the time required for cells to double during exponential growth in complete medium.

Drug Selections

For all selection experiments, cells should be seeded into complete medium at the appropriate density, allowed to attach for 6–10 hr, and then exposed to PALA by replacing this medium with medium supplemented with the drug; PALA can be obtained from the Drug Evaluation Branch of the National Cancer Institute (Bethesda, MD). After 3 days the medium should be changed to remove dead cells and their toxic products. Except for additional replacements with fresh selection medium each week, cells should grow undisturbed until colonies are visible: colonies selected in low PALA concentrations ($3 \times LD_{50}$) should become visible within 6–9 days; colo-

nies resistant to higher concentrations should require 2–4 weeks of growth. Colonies should then be fixed, stained, and counted as described above. The incidence of PALA resistance (percent survival) is the proportion of attached cells that give rise to resistant colonies, and thus is relative to the plating efficiencies of the cells in medium without drug (100% survival). To minimize variation between experiments, these plating efficiencies will be determined independently and in triplicate for each drug selection experiment. The LD_{50} values represent the concentration of PALA that allows 50% survival and are estimated by interpolation from values of percent survival in increasing concentrations of PALA; these values should be determined in triplicate at both the beginning and end of experiments.

Subcloning PALA-Resistant Lines

Cells from individual PALA-resistant colonies should be scraped from the bottom of 100-mm-diameter culture dishes using a sterile micropipette tip and collected in a 25-μl drop of PALA medium. These cells can be transferred to a 24-well cloning dish (Cat. No. 3047; Falcon, Oxnard, CA) and expanded in the same concentration of PALA-containing medium (3 or 9 $\times$ LD_{50}) until a total of 4×10^7 cells has been obtained. A small portion of these cells (10%) should be frozen in 90% (v/v) dialyzed serum–10% (v/v) dimethyl sulfoxide; genomic DNA can be isolated from the remainder (see below).

Molecular Analyses

Genomic DNA can be isolated as described in Brown *et al.* (9). Briefly, cells are lysed by the addition of 0.01 *M* Tris-HCl (pH 8), 0.01 *M* EDTA, 0.01 *M* sodium chloride, 0.02% sodium dodecyl sulfate (SDS), treated with proteinase K, extracted with phenol–chloroform (1 : 1, v/v), and treated with RNase A. DNA is quantitated spectrophotometrically and the indicated amounts electrophoresed on 0.7% (w/v) agarose gels after digestion with *Eco*RI. DNA is transferred to nitrocellulose (10) and hybridized to probes at 65°C under standard conditions: 10–20 ng of probe DNA per milliliter, 0.45 *M* sodium chloride, 0.045 *M* sodium citrate, 0.5% (w/v) SDS, 2× Denhardt's solution, and salmon sperm DNA (50 μg/ml). Radioactive probes can be prepared by either nick translation or randomly primed DNA synthesis (multiprime system; Amersham, Arlington Heights, IL) using [^{32}P]dCTP.

DNA Cytofluorometric Analysis

Cells can be prepared for analysis by the procedure of Gray and Coffino (11) with the following modifications: fixed cells are treated with RNase A (5 mg/ml) for

30 min at 37° C, and then stained with propidium iodide (0.5 mg/ml) for 30 min at room temperature. Cell suspensions ($2-5 \times 10^6$ cells/ml) are kept on ice and analyzed at a flow rate of approximately 500 cells/sec using a Becton-Dickinson (Mountain View, CA) FACS IV with an argon laser (488 nm). The average cellular DNA contents are relative to a karyotypically determined diploid ($2n$) cell standard.

Chromosome Analysis

Chromosome spreads can be prepared and stained by the method of Nelson-Rees *et al.* (12). Mid-log-phase populations of cells are exposed to Colcemid (5–10 μg/ml) for 1 hr or more and removed from the plate. Cells are collected by centrifugation and resuspended in hypotonic KCl plus sodium citrate and incubated at room temperature for 30 min. Cells are fixed in 3:1 (v/v) methanol–acetic acid and dropped onto a wet glass slide. For banding, slides are dried for 2 days and briefly incubated with 0.01% (w/v) highly purified trypsin on ice. The trypsin is neutralized with 10% (v/v) calf serum and the slides are stained with 2% (w/v) Giemsa. Approximately 20–50 spreads per cell line should be examined. Currently the fastest and most definitive method for verifying amplification of a specific locus is FISH (fluorescent *in situ* hybridization). It is routinely used on both rodent and human preparations. We use the *in situ* hybridization protocols of Trask and Hamlin (13) to localize the amplified *CAD* genes in these cells as follows.

Fluorescent in Situ Hybridization

Pretreatment

Initially, slides must be prepared and baked overnight at 60° C to anchor chromatin to the slides. See slide making protocol for details. Throughout the FISH protocol the slides will be immersed in liquid. Handle the slides gently to prevent sample loss. Do not shake vigorously. Slides aged 2 days to one or 2 months give the best results.

RNase A Treatment

RNase (100 μg/ml) in 2× SSC (1× SSC is 0.15 *M* NaCl plus 0.015 *M* sodium citrate) for 1 hr at 37° C removes transcripts and prevents sequestering of probe and high background. The treatment is followed by a dehydration series in ice-cold 70, 85, and 100% fresh ethanol for 5 min each (3 min each is sufficient if pressed for time). Air dry the slides on the benchtop for at least 2 hr and use immediately or store in a slide box until needed. Slides are good for approximately 1 or 2 months.

Denaturation of chromatin is in 70% (v/v) formamide (FA) (nucleic acid grade;

BRL, Gaithersburg, MD)–2× SSC for 2 min at 72° C. The H_2O bath should be set so that the temperature is 1° C above 72° C for each slide because each slide added decreases the temperature of the solution by 1° C. Prewarm the denaturing solution in a 15-ml conical tube for at least 15 min before use. Add 5 ml of denaturing solution to a hybridization chamber and immediately drop the slide in and begin timing. Denaturation time is critical. Less than 2 min may not denature the chromatin, and the probe will not bind. More than 2 min may overdenature the DNA, breaking the backbone. When overdenatured, the site complementary to the probe may be degraded, the chromosomes look hairy, and propidium iodide (PI) does not intercalate as well, so PI intensity is significantly decreased. After precisely 2 min, quickly remove the slides and dip into ice-cold 70% (v/v) ethanol, followed by 85 and 100% as before. Slides are then air dried for at least 10 min before the next step. (Stock denaturing solution can be stored at 4° C.)

Proteinase K (PK) treatment degrades both cytoplasmic and nuclear proteins, decreasing background and facilitating more efficient probe binding. For hamster cells PK (60 ng/ml) in 20 n*M* Tris–2 m*M* $CaCl_2$ (pH 7.4) for 4 min at 37° C worked best. For human cells, PK (600 ng/ml) for 3–4 min worked best. The time and concentration of PK treatment may vary depending on the cell line. Proteinase K incubation is followed by a quick room temperature rinse in the Tris-$CaCl_2$ solution to remove PK. Dehydration in the ethanol series follows. The 70 and 85% (v/v) solutions from the denaturing step can be reused, but the 100% ethanol must be fresh. Air dry the slides on the benchtop for 5 min.

Hybridization Mixture

I. Formamide (50%, v/v)	II. Formamide (55%, v/v)
SSC (2×)	SSC (1×)
Dextran sulfate (10%, w/v)	Dextran sulfate (10%, w/v)
Salmon sperm DNA (5 μg/μl)	Salmon sperm DNA (5 μg/μl)
Competitor DNA (5 μg/μl)	Competitor DNA (5 μg/μl)
Biotinylated probe (2 μg/μl)	Biotinylated probe (2 μg/μl)

Mixture II is recommended for chromosome-specific repetitive probes. It is more stringent (T_m of DNA is lower) than mixture I. Each added percent formamide decreases the T_m of DNA by 0.7° C. Salts and SSC, increase the T_m of DNA. Dextran sulfate is for volume exclusion to increase effective concentration of the mixture components. Salmon sperm DNA is for nonspecific blocking. Competitor DNA should be from the same species as the probe to block repetitive sequences in the probe. For most effective blocking, the blocking DNAs should be fragmented to 200–600 bp. Probe concentration can be increased if necessary, but 2 μg/μl is the minimum concentration for adequate signal. *Note:* Total DNA in the reaction should not exceed 10 μg.

Hybridization Procedure

Repetitive Sequence or Small Unique Sequence Probes

1. Denature the probe mix (70° C for 5 min, simultaneously with 5-min air drying of slides).
2. Ice quench (<2 min).
3. Apply 100 μl of probe to each $\frac{1}{2}$ slide and coverslip.
4. Seal the edges with rubber cement and place in a moist chamber at 37° C.

Unique Sequence Probes with Repetitive Sequences That Must Be Suppressed

1. Denature.
2. Incubate at 37° C for 5 min to 1 hr.
3. Follow steps 3 and 4 from above.

Posthybridization Washes/Fluorescein Isothiocyanate Staining

The purpose of the posthybridization washes is to remove excess probe and nonspecifically bound probe. The fluorescein isothiocyanate (FITC) staining is a sandwiching technique. A layer of avidin-DN (Vector Laboratories, Burlingame, CA) is added to the biotinylated probe followed by biotinylated anti-avidin antibody (Vector Laboratories) and another layer of avidin-DN (see Table I).

The washes need to be at the temperature specified before slides are immersed. The 50% (v/v) formamide–2× SSC can be reused two or three times. Store at 4° C between uses. The blocking solution must be made fresh each time. Five milliliters of solution in the hybridization chamber covers the slides well. The avidin-DN and anti-avidin solutions can be prepared beforehand (5 ml), but add the avidin or anti-avidin just before needed. The FITC conjugated to the avidin is very light sensitive. Cover the incubation chambers with foil. The solutions can be used for four to six slides. Store in foil-covered conical tube at 4° C between uses.

TABLE I Posthybridization Washes

Step	Washes	Temperature (°C)	Time (min)	Reagents
1	3	42	5 each	Formamide (50%, v/v), 2× SSC, pH 7
2	3	42	5 each	SSC (2×)
3	1	Room temperature	Quick rinse	SSC (4×)
4	Blocking	37	30	SSC (4×), 3% (w/v) BSA
5	Avidin	37	30	Avidin-DN (5 μg/ml) 4× SSC, 1% (w/v) BSA, 0.1% (v/v) Tween 20
6	3	37	5 each	SSC (4×), 0.1% (v/v) Tween 20
7	3	37	5 each	BN
8	Anti-avidin antibody	37	30	Biotinylated anti-avidin antibody (5 μg/ml) in PBS
9	Repeat step 7			
10	Repeat step 5			
10	Repeat step 7			

After the first two steps, the procedure can be stopped at any following step. Store the slides in a Coplin jar of BN (a blocking solution composed of 8.41 g sodium bicarbonate, 5 ml Nonidet P-40 added to 995 ml of distilled water) at 4°C until needed. If stopped after the avidin-DN step, cover the jar with foil. If stopped after the initial blocking step, another blocking step is necessary before resuming the detection and washes.

Mounting

The mounting medium is a glycerol-based antifade solution with 0.4 μg of PI/ml. Propidium iodide is a DNA intercalator for counterstaining the DNA. Apply 50 μl of antifade/PI to each slide and coverslip. Let the slides incubate in the dark for at least 15 min before viewing, or store in a dark box at 4°C until time permits viewing. Do not expose the slides to direct light; the signal will fade.

Viewing

For scanning, the 10× objective and green light, filter cube G, to see PI-stained chromatin. Once a good spread is located, center the spread between the cross-hatches, add oil, and switch to the 100× objective. Focus the spread and try to identify by shape the chromosome where the signal may be. Switch to the blue light, filter cube B-2, to view the FITC and PI simultaneously. Refocus if necessary and photograph. Spend only the minimum amount of time necessary under the blue light because the signal bleaches very quickly, and the PI bleaches somewhat. Also, surrounding spreads are being affected as one views a particular spread. A spread is much dimmer after it has been viewed.

References

1. P. Nowell, *Science* **194,** 23 (1976).
2. J. L. Biedler and B. A. Spengler, *Science* **191,** 185 (1976).
3. F. W. Alt, R. E. Kellems, J. R. Bertino, and R. T. Schimke, *J. Biol. Chem.* **253,** 1357 (1978).
4. T. Kempe, E. A. Swyryd, M. Bruist, and G. A. Stark, *Cell* **9,** 541 (1976).
5. R. Sager, I. Gadi, L. Stephens, and C. T. Grabowy, *Proc. Natl. Acad. Sci. U.S.A.* **82,** 7015 (1985).
6. T. D. Tlsty, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 3132 (1990).
7. E. Otto, S. McCord, and T. D. Tlsty, *J. Biol. Chem.* **264,** 3390 (1989).
8. S. E. Luria and Delbrück, *Genetics* **28,** 491 (1943).
9. P. C. Brown, T. D. Tlsty, and R. T. Schimke, *Mol. Cell. Biol.* **3,** 1097 (1983).
10. E. Southern, *J. Mol. Biol.* **98,** 503 (1975).
11. J. W. Gray and P. Coffino, *Methods Enzymol.* **58,** 233 (1979).
12. W. A. Nelson-Rees, R. R. Flandermeyer, and P. K. Hawthorne, *Science* **184,** 1093 (1974).
13. B. Trask and J. Hamlin, *Genes Dev.* **3,** 913 (1989).

This Page Intentionally Left Blank

Section VI

Receptors, Signal Transduction

This Page Intentionally Left Blank

[23] Intracellular Receptor Characterization and Ligand Screening by Transactivation and Hormone-Binding Assays

Elizabeth A. Allegretto and Richard A. Heyman

Introduction

The intracellular receptor superfamily includes the receptors for steroid hormones, thyroid hormone, and the active hormonal metabolites of vitamins A and D. Most of these receptor proteins were first identified in the late 1960s and early 1970s by virtue of the fact that tissue extracts containing the receptors had the ability to bind to their hormonal ligands with high affinity and specificity (1–5). Receptor purification and characterization led to the cloning of a number of these receptors including those for glucocorticoids (GR), estrogen (ER), progesterone (PR), and vitamin D (VDR) (6–10). Certain receptors, such as those for the retinoids, were identified through low-stringency screening utilizing homologous regions of previously cloned receptor cDNA sequences (11, 12). A large number of these newly identified cloned sequences shared high homology with the known intracellular receptors, but did not have an identified ligand associated with them. These receptors have been termed orphan receptors (12).

Assessment of receptor hormone-binding activity remains today an essential tool for the characterization of both the receptors with known ligands and the putative orphan receptors. In addition, the technique has been instrumental for evaluation of synthetic compounds that are potential drug candidates. One of the first hormone-binding "assays" utilized activated charcoal to remove unbound tritiated steroidal ligand from that bound to protein (13). Additional assays that utilized either DEAE filters or hydroxyapatite (HAP) to separate bound ligand from free were subsequently developed (14, 15). Saturation-binding curves in combination with Scatchard analysis (16) allow calculation of the dissociation constants (K_d) of the receptors for their various hormonal ligands. Because circulating levels of the hormones *in vivo* were known to be low, it was expected that the affinity of an intracellular receptor for its endogenous hormone might be in the low nanomolar or subnanomolar range. This has proved to be true of all of the receptors in the intracellular receptor superfamily.

In addition to binding to their cognate hormonal ligands, intracellular receptors have also been shown to bind to nuclear components, chromatin, and naked nonspecific DNA, leading to the hypothesis that the cellular response to hormone was due to modulation of gene transcription by the steroid hormone–receptor complex (17,

Methods in Molecular Genetics, Volume 8

18). On identification of the specific DNA sequences to which the receptors bound in the process of activating transcription [hormone response elements (HREs) (19)], it became possible to analyze receptor transcriptional function by the use of transactivation assays in cells. The first such assays consisted of transfecting cells that expressed a certain receptor with a reporter plasmid containing a promoter linked to an HRE for that receptor, which directed transcription of a cDNA whose product was readily assayable, such as β-galactosidase (β-Gal), luciferase, or chloramphenicol acetyltransferase (CAT) (20). The induction of the reporter enzyme thus depends on the hormonal stimulus given to the cell. This type of functional assay utilizing a steroid receptor-driven heterologous promoter was first performed by Chandler *et al.* in 1983 with rat XC cells containing endogenous GR that transactivated a mammary tumor virus promoter–*CAT* construct in response to glucocorticoids (21). With the advent of the cloning of the receptors and their subsequent recombinant expression, it became possible to cotransfect the receptor cDNA sequence as an effector plasmid into a receptor-negative cell line along with a reporter plasmid to demonstrate hormone-dependent transactivation directly (22). This type of assay has been referred to as a cotransfection or "cis/trans" assay.

This hormone- and receptor-dependent transactivation function, along with saturable, high-affinity, specific ligand binding, satisfy the criteria by which proteins have been defined as intracellular hormone receptors. Both hormone-binding and transactivation assays are indispensable tools for characterization of receptors, for screening potential drug candidate ligands for a known receptor, and for screening compounds that are potential orphan receptor ligands. Although this chapter focuses on the first two of these applications, a brief description of the latter follows. Because the receptors exhibit modular functional domain structure, the hormone-binding domain (HBD) of a known receptor can be replaced with the HBD of an orphan receptor (within the context of the intact known receptor) to generate a chimeric receptor that transcribes a reporter enzyme plasmid containing the HRE of the known receptor in response to a ligand that binds to the HBD of the orphan receptor (for a review see Ref. 23). Various compounds can then be tested for their ability to activate the reporter plasmid driven by the chimeric orphan receptor. The hormone-binding assay is essential to determine if the activating compound directly binds to the receptor. This approach led to the cloning of the retinoic acid receptor (RAR) (11).

This chapter concentrates primarily on utilization of transactivation and hormone-binding assays for characterization of known receptors and briefly discusses their application in drug screening of these known receptors. Specifically, the six retinoid receptors, including the three known RARs and the three retinoid X receptors (RXRs), and their known retinoid isomer ligands, all-*trans*-retinoic acid (tRA) and 9-*cis*-retinoic acid (9cRA) (23), are used as an example system. These receptors have been characterized by both hormone-binding assays and transactivation assays, the latter performed in both mammalian cells and yeast.

Mammalian Cell Transactivation Assays

Since the first use of the transcriptional cis/trans cotransfection assay with a cloned, recombinant receptor, the GR, by Giguere *et al.* in 1986 (22), many groups have utilized the assay or variations thereof to analyze the transcriptional activation function of other cloned intracellular receptors. CV-1 (monkey kidney) cells were used by Evans and co-workers at that time because they were easily transfected and they exhibited low expression levels of endogenous GR. CV-1 cells remain today as one cell line of choice for intracellular receptor transactivation assays. We utilize CV-1 cells for many of our cotransfection studies and prefer luciferase as the reporter enzyme as it has a wide window of sensitivity and is easily detected. Cells are transfected with a receptor expression vector (effector plasmid) under the control of a strong constitutive promoter. The HRE-containing promoter of the reporter plasmid should exhibit low basal activity that is dependent on hormone-activated receptor for increased transcription levels. In addition to transfecting cells with effector and reporter plasmid DNA, the cells are also cotransfected with a β-Gal-expressing plasmid under the control of a constitutive promoter, which serves as an internal standard for monitoring transfection efficiency, thereby allowing normalization of the luciferase values (24).

Reagents and Procedures

Transfection

$CaCl_2$ (2.5 *M*)
HEPES buffer (2×, pH 6.9): 50 m*M* HEPES, 0.28 *M* NaCl
Phosphate-buffered saline (PBS): 150 m*M*
Dulbecco's modified Eagle's medium (DMEM) with 10% (v/v) charcoal-absorbed fetal bovine serum
Test ligands, preferably dissolved in ethanol, to be dispensed at 5–10% (v/v) ethanol/medium
Pure plasmid DNA
Plated CV-1 cells (60–75% confluent)

Lysis

Cells can be lysed by a freeze/thaw method in a suitable buffer such as Tris–phosphate or, if 96-well plates are used, a lysis buffer made especially for luciferase assays such as that from Promega (Madison, WI) can be used.

Luciferase Assay

ATP (5 m*M*) in 0.1 *M* potassium phosphate buffer, pH 8.0
Sodium luciferin (0.5 m*M*)

β-Galactosidase Assay

o-Nitrophenyl-β-galactoside (ONPG), 1.15 m*M*
2-Mercaptoethanol (2-ME), 46 m*M*
Sodium carbonate, 1 *M* (100 μl to stop)

The calcium phosphate transfection procedure (25) is essentially used with few modifications. Generally 1–10 μg of each plasmid DNA (effector, reporter, and internal standard) is used per 10-cm plate of cells. DNA precipitate is allowed to form slowly by gradual mixing of the DNA–$CaCl_2$ with the HEPES–PBS mixture and by letting it stand for 20 min prior to addition to cells. The precipitate is gently added to the cell monolayer and allowed to incubate at 37° C for 6 hr. The medium containing DNA is then washed off the cells, cells are washed gently with PBS, and the ligands are added in fresh medium at the desired concentrations and incubated for 24–48 hr. Lysis buffer is then added, and after the cells are lysed the substrates for luciferase and β-Gal are added to separate aliquots of cell lysate. Luciferase activity is quantitated using a luminometer and β-Gal activity is measured spectrophotometrically at 415 nm. The luciferase activity is divided by the rate of β-Gal activity to yield the transactivation response normalized for transfection efficiency (normalized response). The EC_{50} is the effective concentration at 50% of the maximum response. The procedure is amenable to automation using 96-well tissue culture plates and a Beckman (Fullerton, CA) Biomek automated workstation (24).

Retinoid Receptor Transactivation Profiles in Mammalian Cells

The retinoid receptors are classified into two subfamilies, the RARs (α, β, and γ) and the RXRs (α, β, and γ), on the basis of their sequence homology, hormone-binding specificity, and target gene selectivity (23). all-*trans*-Retinoic acid has been shown to induce RAR-dependent transactivation (11, 26) as well as RXR-dependent transactivation, albeit with less potency (12). While tRA bound directly to RAR (11), it did not bind to RXR (12). This observation led to the hypothesis that a tRA metabolite was responsible for RXR activation and subsequent transcriptional stimulation of target promoter sequences (12). This theory was borne out by the identification of 9cRA, an isomer of tRA, which was shown to be a potent activator of RXRα and directly bound *in vitro* to RXRα (27, 28). Figure 1 illustrates typical dose–response curves for tRA and 9cRA activation of hRARα and hRXRα in CV-1 cells via their respective HRE-driven luciferase constructs (28). The potency of the retinoid is described by the effective concentration that produces a 50% response (EC_{50}) with the maximum response assumed to be at 10^{-5} *M*. The potency of tRA via RARα is ~300 n*M* (Fig. 1A) while it is less potent through RXRα (~1 μ*M*, Fig. 1B). 9-*cis*-Retinoic acid, however, activates transcription via both RXRα and RARα with approximately equal EC_{50} values of ~150 n*M* (Fig. 1A and B).

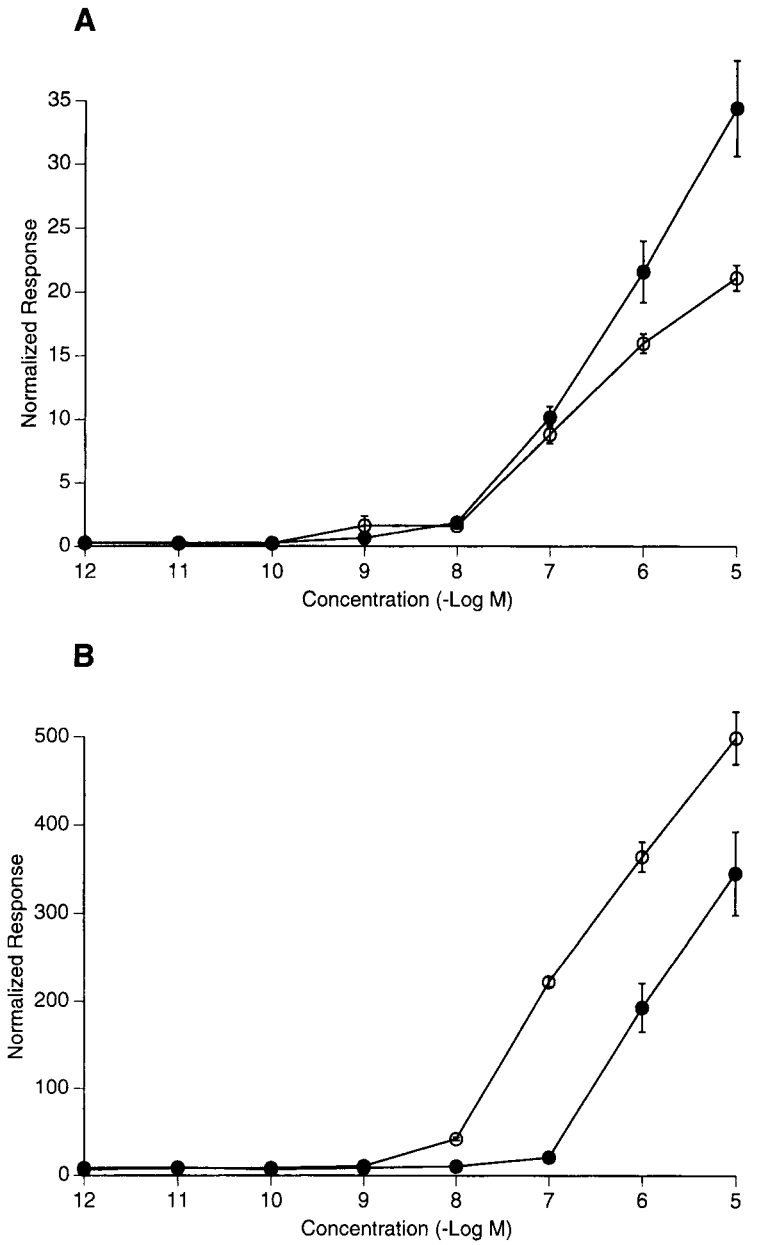


FIG. 1 CV-1 cell RXRα and RARα cotransactivation assay. CRBPII- or TREpal-driven luciferase reporter plasmids were cotransfected with RARα (A) or RXRα (B) expression vectors into CV-1 cells. The indicated concentrations of 9cRA (○) or tRA (●) were added and assay development was as described in text. Luciferase values were corrected for transfection efficiency by the use of an internal cotransfected β-Gal-expressing plasmid.

TABLE I Potencies of all-*trans*-Retinoic Acid and 9-*cis*-Retinoic Acid via Retinoic Acid Receptors and Retinoid X Receptors in CV-1 Cells

	EC_{50}(n*M*)[a]					
Ligand	RARα	RARβ	RARγ	RXRα	RXRβ	RXRγ
tRA	352 ± 31	82 ± 9	10 ± 2	916 ± 72	1492 ± 134	1130 ± 85
9cRA	191 ± 20	51 ± 17	45 ± 5	253 ± 41	221 ± 29	147 ± 13

[a] Values are mean ± standard error of the mean for each value.

The transactivation profiles of both tRA and 9cRA through each of the other RAR and RXR subtypes transfected into CV-1 cells were also examined. Table I is a compilation of the EC_{50} values derived from at least 25 such dose–response experiments performed in triplicate via each of the six retinoid receptors in CV-1 cells (29). 9-*cis*-Retinoic acid activates the RARs and the RXRs with similar potencies (45–253 n*M*). In contrast, tRA displays a much wider range of potencies, through the various receptor subtypes ranging from 10 n*M* via RARγ to 1492 n*M* for RXRβ. Thus, in CV-1 cells, tRA is a much less potent activator of the RXR subfamily (916–1492 n*M*) than it is of the RAR subfamily (10–352 n*M*).

Receptor Hormone-Binding Assays

Having established the transcriptional properties of tRA and 9cRA in CV-1 cells transfected with the retinoid receptors, it was then determined if these measured retinoid potencies correlated with the ability of the retinoids to bind directly to the receptor proteins. Receptor hormone-binding assays are relatively quick ways to measure affinities (dissociation constants) of the receptors for their ligands. These assays rely on a technique to separate ligand bound to protein from "free" ligand such as the use of DEAE filters or hydroxyapatite (HAP) resin, which retain receptors but not free ligand (14, 15), or dextran-coated charcoal, to which only the free ligand binds (13). All three of these techniques have been used successfully to determine intracellular receptor affinities for various ligands.

In the past, the retinoids tRA and retinol have been used extensively for ligand-binding studies of the cellular retinoic acid and retinol-binding proteins (CRABPs and CRBPs). The high levels of nonspecific retinoid binding to CRABP- and CRBP-containing extracts dictated that more rigorous methods of separation be employed than those of the typical hormone-binding assays. Therefore, in order to determine the specific binding component, techniques such as sucrose density gradient centrifugation (30) or gel-filtration high-performance liquid chromatography (HPLC) (31) have been necessary. High nonspecific binding of retinoids to extract components

has continued to be a problem with the nuclear retinoid receptors (32–34). To alleviate this problem, we tested a variety of detergents in retinoid receptor-binding assays and found that the inclusion of the zwitterionic detergent 3-[(3-cholamidopropyl)-dimethyl-ammonio]-1-propanesulfonate (CHAPS) in the binding and wash buffers resulted in decreased levels of nonspecific binding (less than 10% of the total binding) and increased total and specific binding. This modification allowed the determination of K_d values of the retinoid receptors for tRA and 9cRA that are in accord with the low nanomolar values obtained for other intracellular receptors (29). We generally use the HAP method (15) with various modifications depending on the receptor being studied. This technique is easily automated using Beckman minitubes and the Biomek workstation.

Reagents and Procedures

[^{3}H]tRA, ~50 Ci/mmol (NEN, Boston, MA)
[^{3}H]9cRA, ~29 Ci/mmol [Ligand Pharmaceuticals (35) San Diego, CA]
Unlabeled tRA (Sigma, St. Louis, MO)
Unlabeled 9cRA (Ligand Pharmaceuticals, San Diego, CA)

Whole-cell, high-salt protein extracts containing receptor (and extracts without receptor as a negative control, if using recombinant receptor)

Binding and wash buffer
KCl (0.15 *M*)
Tris-HCl, pH 7.5 (10 m*M*)
Dithiothreitol (DTT), 1–5 m*M*
CHAPS detergent (Boehringer Mannheim, Indianapolis, IN), 0.5% (w/v)

Hydroxyapatite resin (fined, 1 : 1 slurry in binding buffer; Bio-Rad)

For saturation binding analyses, tritiated retinoid is added to final concentrations of 0.3–10 n*M* in 12 × 75 mm borosilicate glass tubes under dim light. Solvents other than ethanol are evaporated by the application of a gentle stream of nitrogen gas and the ligand is resuspended in ethanol prior to distribution to tubes at a final concentration of 5–10% (v/v) ethanol. Each ligand concentration point is generally done in duplicate or triplicate. High-salt protein extracts (5–50 μg of total soluble protein) from yeast or insect Sf21 (*Spodoptera frugiperda*) cells expressing individual recombinant retinoid receptors are added to the tubes in binding buffer. Nonspecific binding is measured at each point in the presence of a 200-fold molar excess of unlabeled ligand. The retinoid–protein mixture is incubated for 4–16 hr at 4° C, shielded from light. Receptor–ligand binding is determined by the addition of 50 μl of a 1 : 1 slurry

of HAP for 30 min at 4° C, mixing by gentle vortexing every 10 min. The HAP pellets are washed three times in binding/wash buffer and spun at ~1500 *g* for 2 min. The HAP pellets are then transferred to scintillation vials in 1 ml of H_2O, 15 ml of liquid scintillation fluid is added, and tritium disintegrations per minute (dpm) are determined with a Beckman liquid scintillation counter. Tritium disintegrations per minute representing nonspecific binding is subtracted from total binding to yield specific binding. From the specific saturation binding curve data, Scatchard analysis (19) is employed to determine K_d values of the receptor–ligand interactions.

Retinoid Receptor Hormone-Binding Characteristics

To characterize the hormone-binding properties of the retinoid receptors, we utilized recombinantly expressed retinoid receptors from insect Sf21 cells and yeast (29). The receptor cDNAs [hRARα (11), hRARβ (36), hRARγ (37), hRXRα (12), mRXRβ (38), and mRXRγ (38)] were constructed into expression vectors as described in detail elsewhere (29). Protein extracts were tested in immunoblotting experiments with anti-peptide, subtype-selective, retinoid receptor antibodies to confirm that full-length receptor proteins were produced (29). These extracts were utilized in saturation binding analyses with 9cRA and tRA.

Figure 2 shows three representative saturation curves and Scatchard analyses using Sf21-expressed receptors: RXRα binding to 9cRA (Fig. 2A) and RARα binding to 9cRA (Fig. 2B) and tRA (Fig. 2C). The retinoic acid isomers exhibit high-affinity binding to the receptors in a concentration-dependent, saturable manner, with the nonspecific binding component representing less than 10% of the total binding. The K_d average value of 1.38 n*M* for RXRα binding to 9cRA in the presence of CHAPS is 8- to 10-fold lower than our (28) and other (27, 34) previously reported values determined without the detergent. Table II is a compilation of the resultant K_d values obtained for each of the recombinant receptors with 9cRA and tRA. The RARs bind to both tRA and 9cRA with high affinity and display approximately equal K_d values of 0.2–0.5 n*M* for each receptor subtype–ligand combination, with slightly weaker binding demonstrated for RARγ binding to 9cRA (K_d = 0.8 n*M*). These values are essentially identical to values determined for the RARs expressed in COS cells (34). Therefore, the widely variable potencies that are observed with tRA and 9cRA through the RARs in CV-1 cell transactivation assays are not simply explained by correspondingly varying K_d values of the RARs for the ligands. Other factors, such as the presence of various transcriptional accessory proteins, which display varying affinities for the different retinoid receptors, may cause the resultant distinct transcriptional profiles of each receptor subtype.

Although tRA stimulated transactivation of RXR-dependent pathways in CV-1 cells (Fig. 1, Table I), it did not bind to RXRα (12, 28). However, 9cRA, an isomer of tRA, activated the RXRs in CV-1 cells in a more potent manner than did tRA

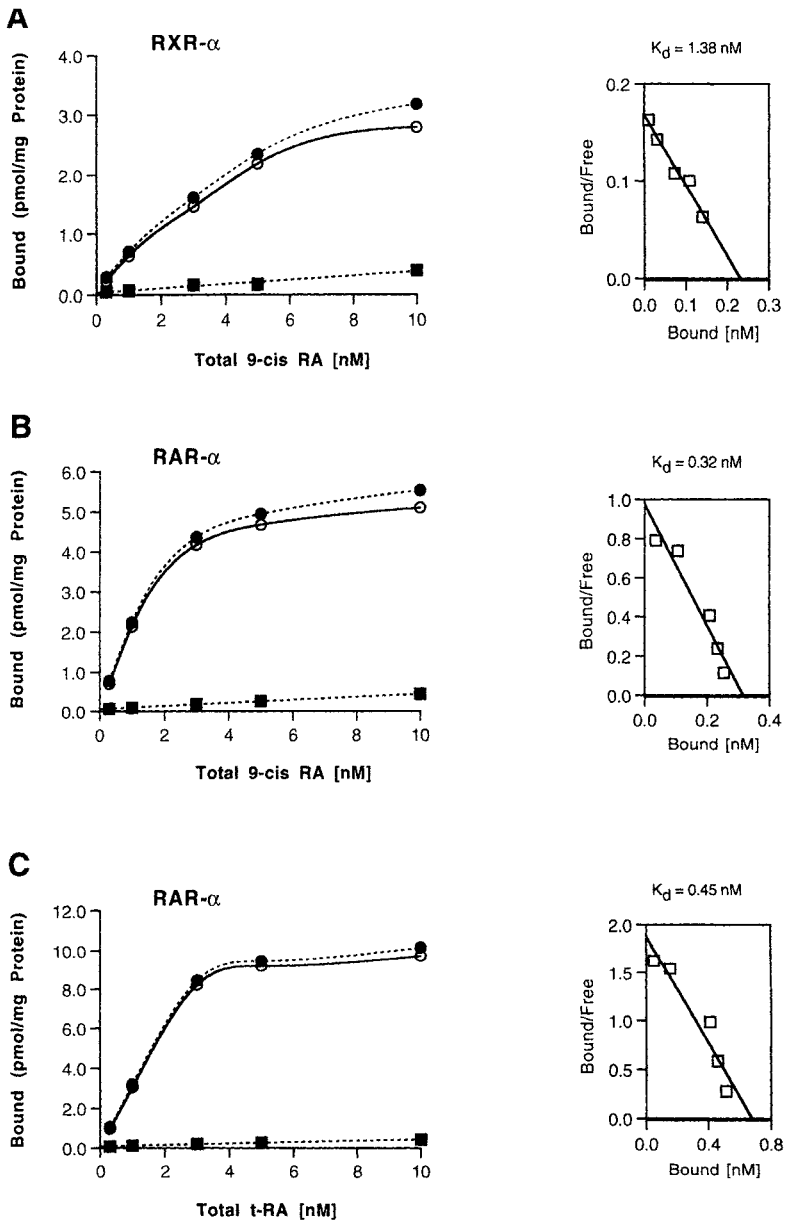


FIG. 2 Saturation binding and Scatchard analysis. Extracts (50 μg) from Sf21 cells containing RXRα (A) or containing RARα (B and C) were incubated with the indicated concentrations of [^{3}H]9cRA (A and B) or [^{3}H]tRA (C) in the absence or presence of a 200-fold molar excess of unlabeled 9cRA (A and B) or tRA (C). Nonspecific binding (■) was subtracted from total binding (●) to generate specific binding activity (○). Scatchard analysis (□) was performed on specific binding data to yield the indicated K_d values for each receptor. A representative experiment for each receptor–ligand combination is shown.

TABLE II Binding Dissociation Constants of Retinoid Receptors for all-*trans*-Retinoic Acid and 9-*cis*-Retinoic Acid

	K_d(nM)[a]					
Ligand	RARα	RARβ	RARγ	RXRα	RXRβ	RXRγ
tRA	0.37 ± 0.13	0.37 ± 0.13	0.22 ± 0.11	nb[b]	nb	nb
9cRA	0.31 ± 0.07	0.20 ± 0.09	0.78 ± 0.14	1.38 ± 0.37	2.11 ± 0.75	1.94 ± 0.76

[a] Dissociation constants (K_d) were determined by Scatchard analysis using recombinantly expressed receptors from Sf21 cells and yeast. Numbers represent the mean ± standard deviation of at least three individual saturation curves done at each point in duplicate.
[b] nb, No measurable binding.

(Table I) and also displayed high-affinity, saturable, and specific binding to RXRα (27, 28). We then tested the ligand-binding characteristics of the other RXRs with 9cRA. Table II shows that average K_d values of ~1.4–2.1 nM were obtained for the three recombinantly expressed RXRs, either from yeast or insect Sf21 cells. Interestingly, the interaction of 9cRA with the RARs is of higher affinity (K_d values of 0.2–0.8 nM) than it is with the RXRs. This corresponds to the observation that 9cRA activates transcription through the RARs at lower concentrations than it does through the RXRs in CV-1 cell transactivation assays (Table I).

Yeast Cotransactivation Assays

It is known that mammalian cells harbor factors that will enhance intracellular receptor binding to their specific HREs (32, 39, 40) and that at least a subset of these factors consists of members of the RXR subfamily. Each of the RXRs is known to heterodimerize with VDR, RARs, and thyroid receptors (TRs) (41–44), thereby modulating their transcriptional activity. Retinoid X receptors seem to be ubiquitously expressed (38) and CV-1 cells are known to contain RXR(s) because cells transfected with an RXR-dependent reporter alone display weak transcriptional activation in response to 9cRA (data not shown) and by direct immunoprecipitation techniques (data not shown). In an effort to avoid these complications in mammalian cell assays, we tested the transcriptional activation properties of the retinoid receptors in yeast, a cell devoid of known intracellular receptors (45, 46), and compared these results with the mammalian cell transactivation profiles and hormone-binding properties of the receptors. The transcriptional activity of ER (47), GR (48), and VDR (45) have been reconstituted in yeast, indicating that the fundamental factors necessary for intracellular receptor transactivation are conserved in yeast.

Reagents and Procedures

Yeast strain cotransformed with effector and reporter plasmids
$CuSO_4$
Appropriate yeast selectable growth media
Test ligands (added to the yeast cultures in H_2O)

Lysis/development buffer
Na_2HPO_4 (60 m*M*), 40 m*M* NaH_2PO_4
KCl (10 m*M*)
$MgSO_4$ (1 m*M*)
β-Gal substrate (ONPG), 2 mg/ml
SDS (0.2%, w/v)
2-Mercaptoethanol (34 m*M*)
Oxalyticase (Enzogenetics, Portland, Oregon), 2000 units/ml

Stop solution
Na_2CO_3 (2 *M*)
Antifoam (0.05%, w/v)

The retinoid receptor cDNAs were constructed into yeast expression vectors under the control of the copper-inducible yeast metallothionein gene promoter as described in detail elsewhere (29). The RXR response element (RXRE) from the CRBPII gene promoter (49) and the RAR response element from the RARβ gene promoter (βRE) (50–52) were constructed into a yeast β-Gal reporter vector as previously described (29). The transformed yeast are plated in 96-well tissue culture dishes (100 μl/well) at an OD_{600} of 0.05–0.2 and ligands are dispensed in concentrations between 10^{-12} and 10^{-5} *M* along with 0–50 μ*M* $CuSO_4$ for 8–16 hr at 30° C. After incubation, the OD_{600} is determined as an indication of cell number. Yeast are then lysed (100 μl of lysis buffer/well) and substrate is added (50 μl/well) for 10–30 min prior to OD_{415} determination. Normalized β-Gal values are determined as follows: $(OD_{415}/OD_{600}) \times 1000$/min development time.

Retinoid Receptor Transactivation Profiles in Yeast Cells

To examine the transcriptional properties of retinoid receptors in yeast, the RXRs were cotransformed with either a CRBPII-containing or a βRE-containing β-Gal reporter plasmid into the yeast strain BJ5409 (29). 9-*cis*-Retinoic acid stimulated transcription from the CRBPII RXRE in yeast, as evidenced by the increase of β-Gal in a dose-dependent manner (Fig. 3A). The induction in response to 9cRA ranged from

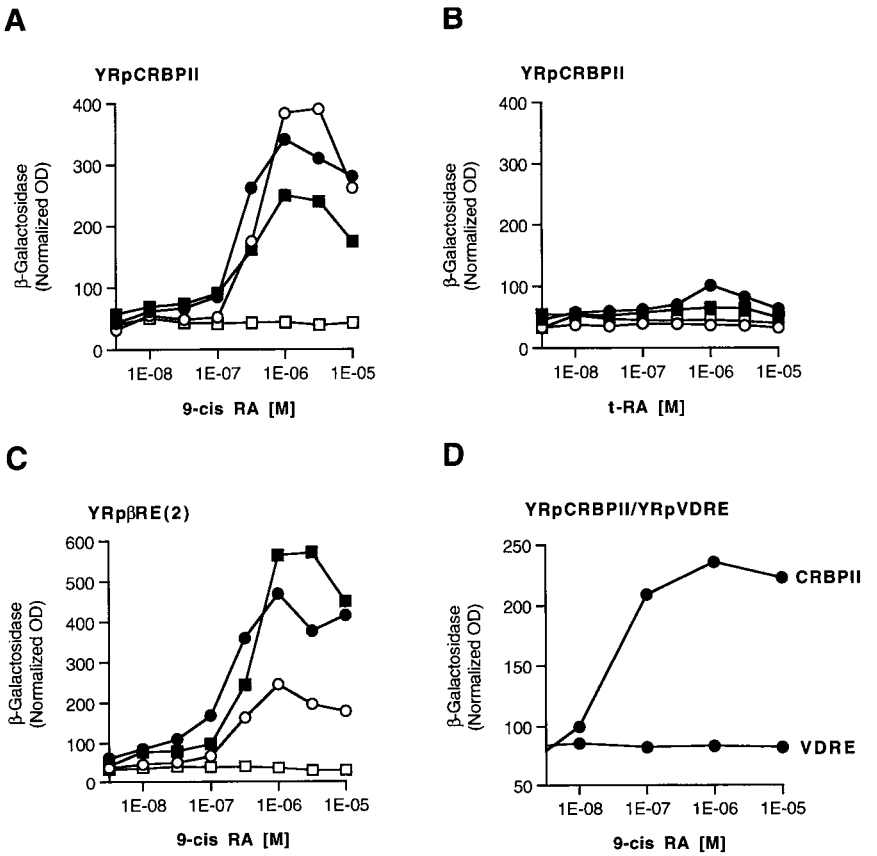


FIG. 3 Yeast RXR cotransactivation assay. CRBPII–β-Gal (A, B, and D), βRE(2)–β-Gal (C), or VDRE–β-Gal (D) reporter plasmids were transformed without (□) or with RXRα (●), RXRβ (○), or RXRγ (■) expression vectors into yeast strain BJ5409. Cultures were grown and induced with 0–10 μM $CuSO_4$ and with varying concentrations of 9cRA (A, C, and D) or tRA (B). β-Gal readings were determined and corrected for cell density and for time of development as described in text.

five- to eight-fold for the three RXRs. The potencies were similar for each of the RXRs and ranged from 200 to 300 nM 9cRA. These values were similar to those obtained in CV-1 cells (Table I). Yeast transformed with reporter plasmid alone did not respond to 9cRA (Fig. 3A), while in CV-1 cells a transcriptional response to 9cRA in the absence of transfected receptor-expressing plasmid was evident, indicating that CV-1 cells express endogenous RXR(s) (data not shown) and yeast do not. Additionally, in CV-1 cells, tRA activated the RXRs to stimulate reporter transcription of an EC_{50} of ~1 μM (28; Table 1), even though tRA did not bind to the

RXRs. In yeast, however, tRA did not stimulate the RXRs to activate transcription (Fig. 3B). Therefore, the yeast *in vivo* transcriptional assay data directly correlate with the *in vitro* hormone-binding assay information, whereas the CV-1 assay data do not. This may be an indication of more active retinoid isomerization and/or metabolism in CV-1 cells than in yeast (29). Interestingly, 9cRA also stimulated transcription from the βRE-driven reporter plasmid in an RXR-dependent manner (Fig. 3C) with a 5- to 10-fold induction and 100–300 n*M* EC_{50} values. While 9cRA activated the RXRs to induce transcription from the CRBPII- and βRE-containing promoters, it did not activate transcription through a vitamin D-responsive element (40; Fig. 3D), indicating specificity of the receptor–DNA interaction.

Discussion

In vitro hormone-binding assays and cell-based transcriptional assays are two essential methods for the characterization of intracellular receptor function. These assays are also the techniques of choice for the evaluation of the activity of ligand drug candidates, as well as that of natural ligands. This chapter has focused on the use of these assays to characterize the six known retinoid receptors. The hormone-binding affinities of the RARs and RXRs with their two known ligands, tRA and 9cRA, have been examined and these values have been compared with the ligand potencies in CV-1 cell transactivation assays. 9-*cis*-Retinoic acid is a bifunctional ligand as it binds to both the RARs and RXRs with high affinities in a saturable manner. all-*trans*-Retinoic acid, on the other hand, binds to the RARs with high affinities, but does not display binding to the RXRs. Transcriptional assays in CV-1 cells transfected with the individual retinoid receptors yield some data that correlate with the *in vitro* binding affinity information and other data that do not. For instance, there is agreement between the more potent 9cRA activation of the RARs versus the RXRs in CV-1 cells and the higher 9cRA-binding affinities displayed by the RARs versus the RXRs. However, the wide range of potencies that is exhibited by tRA via the different RAR subtypes in CV-1 cell assays is not explained by variable binding affinities of tRA for the RARs—they are about equal to each other. Additionally, while tRA stimulates transcription via the RXRs in CV-1 cell assays, it does not bind to the RXRs in hormone-binding assays.

In an attempt to avoid some of the complications of transactivation assays performed in mammalian cells, such as their endogenous intracellular receptor content, transcriptional assays were also performed in yeast, an organism devoid of known intracellular receptors, such as VDR and ER (45, 46). It was found that RAR and RXR transcriptional activity could be reconstituted in the yeast system (Fig. 3) (29, 53, 54). Yeast do not express RXRs or RARs, as retinoid-induced transcriptional activation did not occur in the absence of receptor as was observed in CV-1 cells, and no binding of retinoid to wild-type yeast extracts was observed. In some aspects, the

yeast cis/trans assay data displayed better agreement with the hormone-binding data than did the CV-1 cell assays. For example, there was no activation of RXRs by tRA in yeast as has been seen in CV-1 cells. This finding directly correlates with the hormone-binding assay data, which demonstrate that tRA does not bind to the RXRs.

Yeast cotransactivation systems have some advantages over mammalian cell assays; they are stably transformed, they provide a null receptor background into which the desired receptors can be introduced, and they seem to be inactive in metabolism of hormonal compounds such as the retinoids (29). Mammalian cell lines, on the other hand, may provide valuable information predictive of the *in vivo* behavior of receptor ligands in various cell types and may more accurately predict compound metabolism, stability, and so on, that may occur in the human organism. Therefore, both types of transcriptional assays are useful, as they elicit different, complementary information.

Receptor transactivation and hormone-binding assays, in addition to serving as fundamental tools for receptor characterization, also are essential for evaluation of candidate synthetic ligands or other compounds to be tested for receptor-dependent activity. The transactivation assays are performed as described within, except that the test ligands or extracts are tested in parallel with known receptor ligands and their activity profiles are compared. Test compounds can also be assayed in hormone-binding experiments to determine if they directly interact with the receptor of interest. Hormone-binding assays are modified to allow affinity determinations of unlabeled compounds. Competition hormone-binding assays are performed utilizing an amount of isotopically labelled standard (known control) ligand that is just sufficient to saturate the receptor in the sample. Test ligand is added in a range of concentrations 200-fold higher than the labeled standard ligand at each point and specific binding is calculated and compared with a corresponding competition curve derived from the unlabeled standard ligand. On identification of a hormone-binding active compound, a more accurate assessment of its affinity for the receptor can be made by producing isotopically labeled compound and performing saturation binding and Scatchard analyses. These types of assays will be useful for the further characterization of receptor functional mechanisms. This is especially true with regard to the dissection and understanding of receptor–protein interactions, such as that of the RXRs with the RARs and other intracellular receptors, and for the development of new classes of ligand drugs such as retinoid receptor subtype-selective drugs.

Acknowledgments

We thank Dr. Carla Suto for assistance with the CV-1 cell cotransfection assay section. We thank Michael McClurg, Bryan Macy, and the New Leads Discovery Department at Ligand Pharmaceuticals for experimental contributions.

References

1. D. Toft and J. Gorski, *Proc. Natl. Acad. Sci. U.S.A.* **55,** 1574 (1966).
2. E. V. Jensen, D. J. Hurst, E. R. DeSombre, and P. W. Jungblut, *Science* **158,** 385 (1967).
3. R. S. Gardner and G. M. Tomkins, *J. Biol. Chem.* **244,** 4761 (1969).
4. T. C. Spelsberg, A. W. Steggles, and B. W. O'Malley, *J. Biol. Chem.* **246,** 4188 (1971).
5. P. F. Brumbaugh and M. R. Haussler, *J. Biol. Chem.* **249,** 1258 (1974).
6. S. M. Hollenberg, C. Weinberger, E. S. Ong, G. Cerelli, A. Oro, R. Lebo, E. B. Thompson, M. G. Rosenfeld, and R. M. Evans, *Nature (London)* **318,** 635 (1985).
7. P. Walter, S. Green, G. Greene, A. Krust, J. Bornert, J. Jeltsch, A. Staub, E. Jensen, G. Scrace, M. Waterfield, and P. Chambon, *Proc. Natl. Acad. Sci. U.S.A.* **82,** 7889 (1985).
8. O. M. Conneely, W. P. Sullivan, D. O. Toft, M. Birnbaumer, R. G. Cook, B. L. Maxwell, T. Zarucki-Schulz, G. L. Greene, W. T. Schrader, and B. W. O'Malley, *Science* **233,** 767 (1986).
9. J. M. Jeltsch, Z. Krozowski, C. Quirin-Stricker, H. Gronemeyer, R. J. Simpson, J. M. Garnier, A. Krust, F. Jacob, and P. Chambon, *Proc. Natl. Acad. Sci. U.S.A.* **83,** 5424 (1986).
10. D. P. McDonnell, D. J. Mangelsdorf, J. W. Pike, M. R. Haussler, and B. W. O'Malley, *Science* **235,** 1214 (1987).
11. V. Giguere, E. S. Ong, P. Segui, and R. M. Evans, *Nature (London)* **330,** 624 (1987).
12. D. J. Mangelsdorf, E. S. Ong, J. A. Dyck, and R. M. Evans, *Nature (London)* **345,** 224 (1990).
13. E. Milgrom and E. E. Bavlieu, *Biochim. Biophys. Acta* **194,** 602 (1969).
14. D. V. Santi, C. H. Sibley, E. R. Perriard, G. M. Tomkins, and J. D. Baxter, *Biochemistry* **12,** 2412 (1973).
15. D. Williams and J. Gorski, *Biochemistry* **13,** 5537 (1974).
16. G. Scatchard, *Ann. N.Y. Acad. Sci.* **51,** 660 (1949).
17. R. M. Evans, *Science* **240,** 889 (1988).
18. M. Trusse and M. Beato, *Endocr. Rev.* **14,** 459 (1993).
19. K. Umesono and R. M. Evans, *Cell* **57,** 1139 (1989).
20. C. M. Gorman, L. F. Moffat, and B. H. Howard, *Mol. Cell. Biol.* **2,** 1044 (1982).
21. V. L. Chandler, B. A. Maler, and K. R. Yamamoto, *Cell* **33,** 489 (1983).
22. V. Giguere, S. M. Hollenberg, M. G. Rosenfeld, and R. M. Evans, *Cell* **46,** 645 (1986).
23. D. J. Mangelsdorf, K. Umesono, and R. M. Evans, *in* "The Retinoids: Biology, Chemistry, and Medicine" (M. Sporn, A. B. Roberts, and D. S. Goodman, eds.), p. 319. Raven Press, New York, 1994.
24. T. S. Berger, Z. Parandoosh, B. W. Perry, and R. B. Stein, *J. Steroid Biochem.* **41,** 733 (1992).
25. M. Wigler, R. Sweet, G. K. Sim, B. Wold, A. Pellicer, E. Lacy, T. Maniatis, S. Silverstein, and R. Axel, *Cell* **16,** 777 (1979).
26. M. Petkovitch, N. J. Brand, A. Krust, and P. Chambon, *Nature (London)* **330,** 444 (1987).
27. A. A. Levin, L. J. Sturzenbecker, S. Kazmer, T. Bosakowski, C. Huselton, G. Allenby, J. Speck, C. I. Kratzeisen, M. Rosenberger, A. Lovey, and J. F. Grippo, *Nature (London)* **355,** 359 (1992).
28. R. A. Heyman, D. J. Mangelsdorf, J. A. Dyck, R. B. Stein, G. Eichele, R. M. Evans, and C. Thaller, *Cell* **68,** 397 (1992).

29. E. A. Allegretto, M. R. McClurg, S. B. Lazarchik, D. L. Clemm, S. A. Kerner, M. G. Elgort, M. F. Boehm, S. K. White, J. W. Pike, and R. A. Heyman, *J. Biol. Chem.* **268,** 26625 (1993).
30. R. Lotan, *Biochim. Biophys. Acta* **605,** 33 (1980).
31. E. A. Allegretto, M. A. Kelley, C. A. Donaldson, N. Levine, J. W. Pike, and M. R. Haussler, *Biochem. Biophys. Res. Commun.* **116,** 75 (1983).
32. N. Yang, R. Schule, D. J. Mangelsdorf, and R. M. Evans, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 3559 (1991).
33. S. Keidel, E. Rupp, and M. Szardenings, *Eur. J. Biochem.* **204,** 1141 (1992).
34. G. Allenby, M. T. Bocquel, M. Saunders, S. Kazmer, T. Speck, M. Rosenberger, A. Lovey, P. Kastner, J. F. Grippo, P. Chambon, and A. A. Levin, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 30 (1993).
35. M. F. Boehm, M. R. McClurg, C. Pathirana, D. J. Mangelsdorf, S. K. White, J. Hebert, D. Winn, M. E. Goldman, and R. A. Heyman, *J. Med. Chem.* **37,** 408 (1994).
36. D. Benbrook, E. Lernhardt, and M. Pfahl, *Nature (London)* **333,** 669 (1988).
37. T. Ishikawa, K. Umesono, D. J. Mangelsdorf, H. Aburtani, V. Z. Stanger, Y. Shibasaki, M. Imaware, R. M. Evans, and F. Takaku, *Mol. Endocrinol.* **4,** 837 (1990).
38. D. J. Mangelsdorf, U. Borgmeyer, R. A. Heyman, J. Y. Zhou, E. S. Ong, A. E. Oro, A. Kakizuka, and R. M. Evans, *Genes Dev.* **6,** 329 (1992).
39. J. Burnside, D. S. Darling, and W. W. Chin, *J. Biol. Chem.* **265,** 2500 (1990).
40. T. Sone, K. Ozona, and J. W. Pike, *Mol. Endocrinol.* **5,** 1578 (1991).
41. V. C. Yu, C. Delsert, B. Andersen, J. M. Holloway, O. Devary, A. M. Naar, S. Y. Kim, J. Boutin, C. K. Glass, and M. G. Rosenfeld, *Cell* **67,** 1251 (1991).
42. X. Zhang, B. Hoffmann, P. Tran, G. Graupner, and M. Pfahl, *Nature (London)* **355,** 441 (1992).
43. M. Leid, P. Kastner, R. Lyons, H. Nakshatri, M. Saunders, T. Zacharewski, J. Chen, A. Staub, J. Garniert, S. Mader, and P. Chambon, *Cell* **68,** 377 (1992).
44. S. A. Kliewer, K. Umesono, D. J. Mangelsdorf, and R. M. Evans, *Nature (London)* **355,** 446 (1992).
45. D. P. McDonnell, J. W. Pike, D. J. Drutz, T. R. Butt, and B. W. O'Malley, *Mol. Cell. Biol.* **9,** 3517 (1989).
46. D. P. McDonnell, Z. Nawaz, C. Densmore, N. L. Weigel, T. A. Pham, J. H. Clark, and B. W. O'Malley, *J. Steroid Biochem. Mol. Biol.* **39,** 291 (1991).
47. D. Metzger, J. H. White, and P. Chambon, *Nature (London)* **334,** 31 (1988).
48. M. Schena and K. R. Yamamoto, *Science* **241,** 965 (1988).
49. D. J. Mangelsdorf, K. Umesono, S. A. Kliewer, U. Borgmeyer, E. S. Ong, and R. M. Evans, *Cell* **66,** 555 (1991).
50. H. M. Sucov, K. K. Murakami, and R. M. Evans, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 5392 (1990).
51. B. Hoffmann, J. M. Lehmann, X. Zhang, T. Hermann, M. Husmann, G. Graupner, and M. Pfahl, *Mol. Endocrinol.* **4,** 1727 (1990).
52. H. de Thé, M. M. Vivanco-Ruiz, P. Tiollais, H. Stunnenberg, and A. Dejean, *Nature (London)* **343,** 177 (1990).
53. D. M. Heery, T. Zacharewski, B. Pierrat, H. Gronemeyer, P. Chambon, and R. Losson, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 4281 (1993).
54. B. L. Hall, Z. Smit-McBride, and M. Privalsky, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 6929 (1993).

[24] Control of Blood Cell Differentiation and Mechanisms of Neoplasia

Arthur J. Sytkowski

Introduction

The experimental approaches discussed below relate not only to the specific problems of the control of blood cell differentiation and neoplastic transformation but also to the more general problem in developmental biology: differential gene expression. This chapter touches on such elementary issues as what makes a neuron a neuron or a liver cell a liver cell. Both of these cell types contain identical genetic information that is differentially expressed in a tissue-specific fashion.

How the process of differential gene expression is regulated is a complex question. However, a question of even higher complexity is "How does a cell differentiate?" That is, how is the coordinated change in the expression of multiple genes regulated during the execution of a so-called "differentiation program"?

The differentiation process is initiated and, in some cases, maintained by changes in a variety of stimuli. Documented or proposed stimuli include nutritional changes, metabolic products, light, temperature, gravity, cell–cell contact, and intercellular messenger molecules including small molecules such as steroids and prostaglandins as well as macromolecules, especially polypeptide hormones and growth factors. This last group has a special relevance to gene expression in differentiating hematopoietic cells. In this chapter we use the action of erythropoietin in inducing growth and differentiation of erythroid progenitor cells as a model for investigating the regulation of change in coordinated gene expression. It will be seen that knowledge of early changes in gene expression seen in experimentally induced erythroid differentiation can be used to move back out of the nucleus in order to investigate those signal pathways responsible for these changes. Moreover, further insight into gene regulation by such signaling pathways is gained from artificially downregulating the signal transduction elements themselves. This combination of experimental approaches offers a powerful series of tools with which to elucidate the regulation of human gene expression.

Finding a Suitable Model System

The profound heterogeneity of human hematopoietic cells found in the bone marrow makes the study of molecular events regulating their differentiation extremely diffi-

Methods in Molecular Genetics, Volume 8

cult if not impossible. The difficulty brought about by this heterogeneity extends even to the use of primary malignant cells such as those from patients with polycythemia vera. A similar problem exists using normal bone marrow cells from experimental animals. Therefore, a suitable model system representing a single lineage of hematopoietic cells must be sought.

The study of erythropoiesis was advanced greatly with the discovery of Friend erythroleukemia cells (1). The isolation of cells from mice infected with the Friend leukemia virus and their subsequent growth in culture resulted in continuous lines of cells resembling malignant proerythroblasts. The discovery that these cells could be induced to differentiate with the chemical agent dimethyl sulfoxide was of profound importance and led to numerous critical studies of the molecular biology of globin gene expression. Unfortunately, because these cells do not hemoglobinize in response to the natural inducer erythropoietin, they were deemed unresponsive to the hormone and thus not used at all for signal transduction studies until very recently. In fact, when the expression of certain primary response genes is studied, these cells apparently respond to the hormone equally as well as do other cell types that can go on to produce hemoglobin (2).

The first erythropoietin-responsive continuous cell line was reported in 1978 and was designated Rauscher murine erythroleukemia (3, 4). These cells are derived from animals infected with the Rauscher erythroleukemia virus and grow *in vitro* in the absence of the hormone. Nevertheless, they can be induced to differentiate in response to erythropoietin as well as to other small-molecule chemical inducers. It is this model erythroid progenitor system that has been used extensively in the author's laboratory in studies of signal transduction pathways of erythropoietin and the regulation of primary response genes.

Other model systems have since been developed. The investigator is advised to make a systematic study of those cell types that may be available. The investigator should always have in mind the questions to be asked, because a particular model system may be quite suited to one avenue of experimentation and not at all suited to another.

Identifying Endogenous Reporter Genes

In studying signal transduction pathways that induce cell growth and differentiation, one must identify a response or group of responses that will be used as an "assay" for experimentation. For example, studies of the erythropoietin receptor have relied on the absence or presence of a mitogenic response in cells transfected with mutant receptors and stimulated with erythropoietin (5, 6). This approach using cell growth provides a convenient assay for structural studies of effector molecules that are far upstream, such as the receptor itself. However, the growth response requires the con-

certed expression of many genes, and thus may be less useful in analyzing individual intermediate or downstream signal transducers. Moreover, it does not provide any information on the regulation of individual response genes on arrival of the signal in the nucleus.

The ideal reporter gene for arrival of a signal in the nucleus is one whose change in expression, be it an increase or decrease, is easily monitored by Northern blot analysis. Moreover, because one wishes to study molecular events occurring immediately prior to this change in reporter gene expression, the gene selected should be preferentially a primary (immediate/early) response gene, that is, one whose change in expression is not dependent on *de novo* protein synthesis.

The investigator may begin this effort by choosing from among several well-studied genes that have proven to be primary response genes in one or more other cell systems. For example, the members of the c-*fos* and c-*jun* families of proto-oncogenes may be suitable. In the case of erythropoiesis, previous studies had shown that the protooncogenes c-*myc* and c-*myb* were both downregulated by the chemical inducer dimethyl sulfoxide (7, 8). Using this information as a clue, it was relatively straightforward to determine whether c-*myc* and c-*myb* might serve as endogenous primary response reporter genes for the erythropoietin signal (2). A general method for identifying ideal endogenous reporter genes is as follows.

1. Select one or more genes from the groups suggested above.
2. Obtain suitable cDNA probes to be used in Northern blot analyses. These may be obtained from sources such as the American Type Culture Collection (Rockville, MD), other commercial sources, or colleagues in the field. An alternative approach is to use synthetic oligodeoxynucleotides as probes for Northern analyses (9). In this case, better results are obtained when the oligomer is relatively long, for example, 30–40 nucleotides.
3. Incubate $1-10 \times 10^6$ cells in the absence or presence of 10–100 U of erythropoietin per milliliter for 0–4 hr. Harvest the cells and prepare total cellular RNA using guanidinium isothiocyanate (10). Fractionate the RNA electrophoretically in 1.2% (w/v) agarose containing 5.5% (v/v) formaldehyde and transfer to GeneScreen Plus filters. Hybridize the filters with ^{32}P-labeled cDNA probes (or oligodeoxynucleotides) for the gene under study. Radiolabeled cDNA probes can be conveniently prepared by nick translation or random priming using standard techniques. Because change in the expression of the target gene must be quantified, it is important to include a control "housekeeping" gene such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH) or β-actin (11). This allows comparison of the change in steady state mRNA levels of the target gene with those of the (presumably) unchanging levels of the housekeeping genes. The synthetic oligomer should be end labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP.

In initial studies, it is preferable to use single (individual) radiolabeled probes in

the hybridization bag. For example, filters are first probed for the target gene mRNA. After autoradiography, the filters can be stripped using boiling water and then reprobed for GAPDH or β-actin and autoradiographed. The X-ray film should confirm a lack of change in steady state mRNA levels for the control gene while the target "reporter gene" should exhibit a substantial increase or decrease. If the sizes of the two transcripts are sufficiently different, then later studies can be done by performing double hybridizations, including both probes in a single hybridization bag. This significantly increases the rapidity with which repeat experiments can be performed.

An alternative to using a separate probe for a control gene is to employ ethidium bromide staining of the agarose gel followed by photography of the ethidium bromide-stained gel prior to Northern transfer. Some investigators advocate this method of ensuring equal loading of the lanes as equivalent or even superior to the use of a control gene. However, others suggest that ethidium bromide staining of RNA gels reduces the efficiency of transfer and/or detection by radiolabeled probes. In the experience of the author, it is relatively difficult to quantify the fluorescence images obtained by photographing ultraviolet (UV)-exposed ethidium bromide-stained RNA gels. Therefore, the author's laboratory routinely uses a control gene (GAPDH) for all Northern blot quantification studies.

4. The stability of the control gene and the changes in expression of the target "reporter gene" should be evident on simple inspection of the autoradiographs. Further quantification can be obtained using scanning laser densitometry.

5. Once a clear, repeatable change in the reporter gene has been documented to occur within 0–4 hr after exposure to erythropoietin, the experiment should be repeated using cells treated with the protein synthesis inhibitor cycloheximide. Preliminary experiments should be carried out using different concentrations of the compound in order to demonstrate its efficacy in inhibiting protein synthesis in the particular cell line under study. This is carried out by first incubating cells for a few hours in the presence of [^{3}H]leucine and in the presence of 1–10 μg of cycloheximide per milliliter. The cells are then lysed with hypotonic buffer or distilled water, and protein is precipitated at 0° C in the presence of 10% (v/v) trichloroacetic acid. Precipitated protein is harvested on Millipore (Bedford, MA) filters (HAWP) and radioactivity incorporated into newly synthesized protein is quantified by liquid scintillation spectrometry.

The results should demonstrate that cycloheximide inhibits protein synthesis in a concentration-dependent manner. In the case of Rauscher erythroleukemia cells, protein synthesis is 95% inhibited by 10 μg of cycloheximide per milliliter within 1 hr. Using this information, the effect of cycloheximide on the ability of erythropoietin to change the level of expression of the reporter gene is determined. In the case of both c-*myc* and c-*myb,* cycloheximide has no effect on the ability of erythropoietin to regulate their expression, thus identifying these protooncogenes as primary (immediate/early) response genes for erythropoietin.

Pharmacological Analysis of Potential Signaling Pathways to Primary Response Genes

Pharmacological agents can serve as powerful tools in the identification and analysis of signaling pathways activated by growth factors such as erythropoietin. In studies of this type, one must constantly bear in mind the potential difficulties that the use of such agents can cause. Almost no agent, be it agonist or antagonist (inhibitor), is entirely specific for just a single target. Therefore, the experimental analysis is greatly strengthened when more than one agent is employed. Ideally, these several inhibitors should be chosen from among different molecular classes of compounds with different mechanisms of inhibition toward the target. In addition, concentration effects are of great importance. Although inhibition constants derived from *in vitro* studies on purified molecules are of great assistance, the use of molecular blocking agents on intact cells is likely to result not only in the desired effect on the particular molecular target, but also in numerous other effects that are both undesirable and nonanalyzable. Nevertheless, if these potential difficulties are kept in mind, significant and important results can be obtained using such pharmacological agonists and antagonists of potential signaling pathway intermediates.

At this stage, it is assumed that the investigator has in hand a quantifiable response, that is, a reporter gene that is ideally a primary response gene whose change in expression is readily quantifiable in a simple and reproducible fashion. One must then select a potential target signal transduction intermediate or pathway. Because the current literature is rich with studies on a variety of signaling pathways, it is up to the investigator first to develop a thorough understanding of current knowledge in this area as well as an appreciation for the numerous pharmacological agents that are available.

The demonstration of a protein kinase C (PKC)-dependent pathway activated by erythropoietin (11, 12) will serve here as an example of this type of study.

1. Rauscher murine erythroleukemia cells are grown *in vitro* and kept at a density of less than 1×10^6/ml. Cells are pelleted by centrifugation and resuspended at 5×10^6/ml in fresh medium with 10% (v/v) serum 18 hr prior to the start of each experiment.

2. Aliquots of cells (2×10^6) are then incubated in the presence of 5–50 U of erythropoietin per milliliter in a specified concentration of PKC inhibitor. Some PKC inhibitors used and the ranges of concentrations to be employed are as follows: H7, 0–40 μM; H8, 0–120 μM; staurosporine, 40–160 nM; sphingosine, 0–40 μM; sangivamycin, 0–100 μM.

3. After 1 hr, RNA is extracted. Because of the multiple samples employed, the author's laboratory has used the procedure of Hatch and Bonner (13), which is readily applicable to multiple samples. RNA is then subjected to electrophoresis in agarose

gels containing formaldehyde and transferred to GeneScreen Plus filters. Northern blots are probed with ^{32}P-labeled cDNA for the reporter gene (c-*myc*) and the control gene (*GAPDH*). Changes in expression are quantified by autoradiography and scanning densitometry.

4. From the above studies, inhibition curves are derived comparing the erythropoietin-induced upregulation of c-*myc* expression in the absence of inhibitor to that seen in the presence of the specified concentration of inhibitor. In this way, the rank order of potency of these five inhibitors of PKC toward erythropoietin is derived. This order of potency is found experimentally to be staurosporine > sphingosine > H7 > sangivamycin > H8. This rank order is identical to that derived from the scientific literature and is considered to be diagnostic of a PKC-mediated signal.

In the course of such experiments, it is critical to eliminate, insofar as it is possible, toxic effects as a cause for blocking a signal. This can be done in several ways, including (a) the use of a control gene such as *GAPDH* in the analysis, (b) standard tests of cell viability such as trypan blue exclusion, and (c) demonstration of reversibility of the inhibition on the removal of the inhibitor. In this regard, the use of a primary response gene as a reporter is especially useful because the relatively short time of exposure needed to assess blocking of a signal allows the investigator to minimize the time of exposure of cells to the inhibitor and, thus, to avoid complex effects.

The pharmacological analysis of a PKC-dependent signal described here is quite different from the "PKC-depletion" approach that has been widely used. In PKC-depletion analysis, the PKC agonist tetradecanoyl phorbol acetate (TPA) is used to treat cells for longer periods of time, usually 6–24 hr. This treatment results in prolonged stimulation of PKC and translocation of the enzyme to the plasma membrane, ultimately depleting PKC from the cytosol. These PKC-depleted cells can then be studied further and the potential role of PKC in any signaling is deduced from the results obtained in this depleted state. The inherent problem with such PKC depletion studies is the obvious question regarding the effects of prolonged depletion on multiple cellular processes. Therefore, such PKC depletion experiments are somewhat less useful in a more definitive molecular analysis of signal transduction pathways, especially regarding primary response genes.

Once a signaling intermediate such as PKC has been identified as required for the pathway to c-*myc*, one should then ask whether this intermediate is necessary and sufficient for c-*myc* upregulation or necessary but not sufficient. When the studies described above were carried out, investigators in the author's laboratory showed that treatment of cells with the PKC agonist TPA failed to upregulate c-*myc* and, therefore, concluded that PKC was necessary but not sufficient for c-*myc* upregulation in erythroid cells. In contrast to this result, TPA readily upregulates c-*myc* in lymphocytes. The investigators concluded that c-*myc* upregulation by erythropoietin required a second signal in addition to a PKC-dependent one. However, other interpretations of the data should be considered. For example, knowledge of PKC isoforms

has increased markedly in the last few years since these original experiments were carried out. At this writing, 10 isoforms have been described that have been divided into classes based on functional domains found in the protein structure. Class I, also designated cPKC (common, classic), contains a calcium-binding domain and a phorbol ester (TPA) binding domain. Class II isoforms (nPKC; nonclassic, new) lack the calcium-binding domain but have a phorbol ester-binding domain. Class III (aPKC; atypical) have neither a calcium- nor a phorbol-binding domain but yet retain significant sequence homology to the cPKC isoforms. It is now known that the various inhibitors of PKC used to analyze erythropoietin signal to c-*myc* exhibit markedly different potencies toward these various isoforms. Additionally, one must consider the fact that these experiments are carried out *in vivo,* that is, in the living cell. Therefore, uptake of pharmacological agents and potential compartmentalization of these agents and/or their targets also complicate interpretation. For example, PKC-epsilon (ϵ) is a class II isoform and, therefore, retains a phorbol ester (TPA)-binding domain. Yet, depending on the cell type studied, TPA treatment variously results in membrane translocation of PKC-ϵ, indicating activation of the isoform, or lack of such translocation, consistent with resistance of the isoform to the agonist. Taken together, such complex data illustrate the great care that must be taken in interpreting such inhibitor studies.

Selection of Second Target to Identify Second, Discrete Signaling Pathway

The availability of more than one reporter gene makes possible an investigation into additional potential signaling pathways. Because erythropoietin upregulates c-*myc* and downregulates c-*myb* as primary response genes, it has been possible to demonstrate two discrete signaling pathways that differentially regulate gene expression in erythroid cells (14). To carry out such a potential analysis, one identifies a second potential signal transduction intermediate as a target for pharmacological agents. In this instance, the potential role of protein phosphatases, specifically class I,2a protein phosphatases, had been suggested by earlier studies. Such target intermediate enzymes are amenable to probing with the pharmacological agent okadaic acid. Here again, the use of a potentially toxic agent is made possible by employing a primary response gene whose regulation by erythropoietin requires only 1–2 hr. Okadaic acid is used as follows.

1. Rauscher cells are grown as described above.
2. Cells are stimulated at 5×10^6/ml for 2 hr with 10–50 U of erythropoietin per milliliter and specified concentrations of okadaic acid (0–400 n*M*).
3. Total celluar RNA is isolated and subjected to Northern analysis as described above. Filters are probed with ^{32}P-labeled c-*myc* cDNA, c-*myb* cDNA, or *GAPDH* cDNA as internal control. Levels of c-*myc* and c-*myb* mRNA are normalized to the level of *GAPDH* expression within the same lane.

4. Simultaneously, experiments are carried out using a PKC inhibitor, such as H7, to distinguish a PKC-dependent signal from a protein phosphatase-dependent signal.

5. The results of these experiments demonstrate that PKC inhibitors such as H7 completely block the erythropoietin signal to c-*myc,* but have no effect whatsoever on the ability of erythropoietin to downregulate c-*myb.* Thus, PKC is not required for erythropoietin signal to c-*myb.* In contrast, the protein phosphatase inhibitor okadaic acid, while having no effect at all on erythropoietin signal to c-*myc,* completely blocks the erythropoietin signal to c-*myb* in a concentration-dependent fashion. Thus, the use of pharmacological inhibitors of two potential signal transduction intermediates, PKC and protein phosphatase, results in the demonstration of two independent signaling pathways that regulate two different primary response genes.

Use of Pharmacological Agents to Study Mechanism of Regulation of Primary Response Gene Expression

The mechanism by which the PKC-dependent or protein phosphatase-dependent signals regulate expression of c-*myc* and c-*myb,* respectively, can also be studied through the use of pharmacological agents. The c-*myc* gene contains two transcription initiation sites as well as a transcriptional attenuation (arrest) site at the exon I/intron I border. The c-*myb* gene contains one transcription initiation site and a transcription attenuation site within its first intron. The mechanism by which the erythropoietin signal pathway regulates these genes, be it by transcriptional initiation, attenuation, and/or mRNA stability, is readily investigated (14).

mRNA stability assays are carried out as follows.

1. Rauscher cells are grown *in vitro* under standard conditions. They are then incubated for 2 hr in the presence of 50 U of erythropoietin per milliliter.

2. Actinomycin D (20 μg/ml) is added to inhibit RNA synthesis. Aliquots of cells are removed at 30-min intervals. RNA is isolated and analyzed by Northern blot analysis measuring steady state mRNA levels of c-*myc* and c-*myb.* Results are expressed as percent c-*myc* or c-*myb* mRNA versus time.

The results of these experiments, when carried out in the absence or presence of H7 or okadaic acid, demonstrate that neither erythropoietin alone nor erythropoietin in the presence of either inhibitor changes the stability of the mRNA of either reporter gene.

The effect of erythropoietin on regulating c-*myc* and c-*myb* transcription as well as the mechanism by which PKC blockers and protein phosphatase blockers inhibit this regulation are approached through transcription run-on assays. These are carried out as follows.

1. Cells are grown under standard conditions and stimulated with 50 μ of erythropoietin per milliliter for 2 hr in the absence or presence of 50 μM H7 or 400 nM okadaic acid.

2. Nuclei are isolated by the method of Natalis and Godbolt (15) as follows. Cells (1×10^7) are collected by centrifugation. The cell pellet is resuspended in 400 μl of lysis buffer [10 nM Tris-HCl, 10 mM NaCl, 3 nM $MgCl_2$, 1 mM phenylmethylsulfonyl fluoride, 0.05% (w/v) Nonidet P-40, pH 7.5] and recentrifuged at 1000 g for 5 min at 4° C. The supernatant is removed by aspiration and the pellet is resuspended in lysis buffer (2 ml) without Nonidet P-40. Nuclei are centrifuged for 5 min at 1000 g at 4° C. Nuclei may be stored in 50 mM Tris-HCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 40% (v/v) glycerol, pH 8.3, at −70° C for more than 30 days.

3. Transcription reactions are carried out as follows. Each reaction consists of 210 μl of nuclei, 60 μl of 5× runoff buffer (5× runoff buffer consists of 25 mM Tris-HCl, 12.5 mM $MgCl_2$, 750 mM KCl, and 1.25 mM triphosphates of A, G, and C), pH 8.0. α-^{32}P-labeled uridine triphosphate (30 μl; ~450 μCi; 3000 Ci/mM) is added, and the suspension is incubated at 30° C for 30 min; 15 μl of DNase I (5 μg/ml) is added. After 5 min at 30° C, the reaction is made 1× SET [1% (w/v) sodium dodecyl sulfate (SDS), 5 mM EDTA, 10 mM Tris-HCl (pH 7.4)] and proteinase K is added to a concentration of 200 μg/ml. After incubation at 37° C for 45 min, the solution is extracted with an equal volume of phenol–chloroform, and the interphase is again extracted with 100 μl of 1× SET. Ammonium acetate (10 mM) is added to the combined aqueous phases (original plus reextraction) to a final concentration of 2.3 M. An equal volume of 2-propanol is added, and nucleic acid is precipitated (−70° C, 15 min). The precipitate is centrifuged in a microcentrifuge for 10 min, and the pellet is resuspended in 100 μl of TE (10 mM Tris-HCl, 1 mM EDTA) and centrifuged through a Sephadex G-50 (medium) spin column. The eluate is made 0.2 M in NaOH, and after 10 min on ice N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES) is added to a concentration of 0.24 M. Two and one-half volumes of ethanol is then added, and the solution containing the precipitate is kept overnight at −20° C. After centrifugation in a microcentrifuge for 5 min at 4° C, the pellet is resuspended in hybridization buffer [10 mM N-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid (TES), 0.2% (w/v) SDS, 10 mM EDTA, 0.3 M NaCl, 1× Denhardt's solution, and *Escherichia coli* RNA (250 μg/ml), pH 7.4].

4. The labeled transcripts (10^7 cpm/ml) are hybridized to the following gene fragments immobilized on GeneScreen Plus filters using a dot-blot apparatus: c-*myc* exon I, c-*myc* exon II, c-*myb* exon I, c-*myb* exon II, *neo,* and genomic DNA. Hybridization and washing are performed as in Ref. 16. Filters are autoradiographed using standard methods.

The results show that there is virtually no difference in the expression of *myc* exon I in the absence or presence of erythropoietin with or without H7 or okadaic

acid. In contrast, in the absence of erythropoietin, little *myc* exon II expression is detected, whereas the addition of erythropoietin results in a marked increase in *myc* exon II. Therefore, erythropoietin upregulates c-*myc* expression by release of transcriptional attenuation at the exon I–intron border and not through transcriptional initiation. Addition of H7 results in the disappearance of exon II transcription, indicating that the erythropoietin PKC signal to c-*myc* upregulates the expression of the gene by release of transcriptional attenuation.

A comparable result is seen when c-*myb* is analyzed. In the absence of erythropoietin both exon I and exon II of c-*myb* are expressed, consistent with the high steady state levels of full-length c-*myb* transcript seen in cells in the absence of erythropoietin. Addition of erythropoietin has no effect on first intron expression but markedly inhibits expression of *myb* exon II, consistent with the imposition of transcriptional attenuation by the erythropoietin signal. Addition of okadaic acid blocks protein phosphatase and restores expression of exon II even in the presence of erythropoietin, indicating that the erythropoietin protein phosphatase-dependent signal to c-*myb* downregulates the gene by imposing transcriptional attenuation within the first intron.

In summary, although the use of pharmacological inhibitors of signal transduction intermediates is at times subject to complex interpretation, this approach, when judiciously applied, can result in significant detailed information about growth factor signal transduction pathways and gene regulation.

Gain of Function and Loss of Function Experiments to Probe Signal Transduction Intermediates and Their Primary Response Genes

Among the most powerful approaches to the study of growth factor signal transduction experiments are the types of experiments that have been designated "gain of function" and "loss of function." Generally speaking, such experiments employ the introduction of sense or antisense gene constructs or antisense oligodeoxynucleotides into cells as a means of artificially upregulating or downregulating a particular gene that encodes either a signal transduction effector molecule or a gene that is a target of a signaling pathway. This gain of function/loss of function approach has a major advantage over the pharmacological analysis described above insofar as it can be designed to be absolutely specific for a particular target protein. This advantage is also a disadvantage of the method because the gene or at least part of it must be cloned and sequenced in order for such approaches to be employed successfully. Nonetheless, the power of these approaches is easily seen in present-day scientific literature.

In general, gain of function experiments are carried out as follows. The appropriate cell line under study is contransfected with a construct containing a full-length cDNA encoding the gene whose expression is to be increased along with a means of bio-

chemical selection, such as *neo*. The transfected cells are then treated with the appropriate biochemical, for example G418, and resistant cells are grown and assessed for their stable expression of the exogenous gene. The increase in expression of this gene can be either stable or inducible depending on the construct transfected. Numerous examples of the use of gain of function experiments in analyzing the importance of PKC isoforms have been published.

Loss of function experiments can be of two general types. The first involves the stable transfection, similar to that described above, of an antisense construct whose expression leads to the synthesis of an RNA strand complementary to the mRNA of the exogenous target gene. Translation is prevented by a variety of mechanisms, including rapid degradation of the resulting hybrid by RNase H.

The singular disadvantage of these powerful approaches lies in the requirement of integration and stable expression of the construct and biochemical selection of the transfected cells. The investigator is at all times required to ask whether or not the transfected cells resulting from this procedure in all other ways are representative of the cells prior to introduction of the gene construct.

Another type of loss of function study is one that employs antisense oligodeoxynucleotides that are introduced into the cell directly simply by addition to the growth medium of the cell. The use of antisense oligonucleotides does not require transfection, integration of a construct, stable expression or biochemical selection. Therefore, it can be carried out much more quickly than the other described studies. Once the antisense oligonucleotide has gained entry into the cell, the nucleotide forms an RNA–DNA hybrid with the target mRNA. This leads to rapid degradation of the target mRNA by RNase H. A diagram generally illustrating the mechanism by which antisense oligonucleotides block synthesis of the c-*myb* gene product in Rauscher erythroleukemia cells is shown in Fig. 1.

There are several chemical forms of antisense oligodeoxynucleotides. "O-oligo-

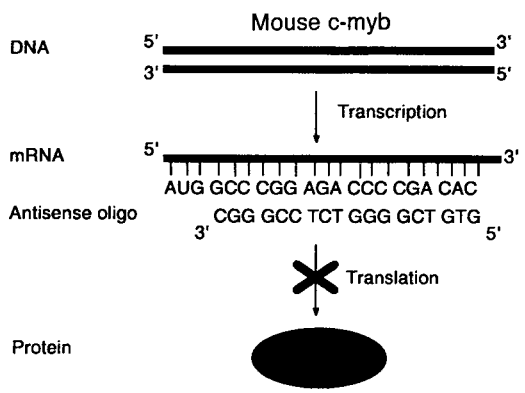


FIG. 1 Downregulation of murine c-*myb* expression with an antisense oligodeoxynucleotide.

nucleotides" are of standard deoxyribonucleotide structure. "S-oligonucleotides" or "thiolated (sulfurated) oligonucleotides" incorporate sulfur into the carbohydrate backbone of the oligonucleotide, rendering it more resistant to nuclease digestion. However, in many instances, the sulfurated form is less permeant than the conventional O-oligonucleotide. In the experience of the author's laboratory, relatively equal effects have been seen when cells have been incubated with equivalent concentrations of either O- or S-oligonucleotides. Additionally, methylated forms have been employed that reportedly exhibit increased permeation into the cell. However, they are much less used than either O or S forms.

In general, investigators have employed antisense oligonucleotides complementary to codons 2 through 6 or 7 of their respective target genes. These 15- to 18-mer compounds are usually of sufficient length to ensure good hybridization in the cell provided that the sequence is exactly complementary to that of the endogenous gene. The first codon is ordinarily not included because it provides no specificity to the antisense oligonucleotide.

Dose–response experiments must be carried out for each oligonucleotide used. This is because cellular uptake of these compounds is unpredictable. Indeed, the same cell line may require as much as a 10- to 50-fold difference in exogenous oligonucleotide concentration depending on the target gene to be studied.

It is critically important that an appropriate assay be available to evaluate efficacy of antisense oligonucleotide treatment. In the case of targeting potential signal transduction intermediates, appropriate readouts include (a) the change in expression of the primary response gene, (b) change in the expression of the subsequent secondary response exhibited by numerous genes, or (c) changes in phenotypic properties such as hemoglobin formation. In studies carried out in the author's laboratory on the role of c-*myb* in erythropoietin action, downregulation of the primary response gene itself (c-*myb*) with antisense oligonucleotides led to the induction of hemoglobin formation readily assessed by histochemical staining of cells in suspension (17). In the case of the erythropoietin PKC-dependent signal to c-*myc,* the availability of numerous PKC isoform sequences has allowed the investigators to downregulate isoforms individually and measure the effect of such downregulation on the ability of erythropoietin to upregulate c-*myc* (18).

Realizing that, in each case, experiments must be optimized to the particular cell line and target gene employed, the following procedure utilized to elucidate the erythropoietin PKC-dependent pathway to c-*myc* can serve as an example protocol.

1. Cells are grown under standard conditions and are then incubated for 16–18 hr (overnight) in a specified concentration of antisense oligonucleotide to PKC-ϵ. The antisense oligonucleotide is complementary to codons 2–7 of rat PKC-ϵ. It is assumed that the murine sequence is identical or almost identical to that of the published rat sequence. Concentrations found to be effective are 50–200 μg/ml.
2. The following morning a second addition of antisense oligonucleotide equaling

25% that of the initial addition is made. The cells are then treated with 10–50 U of erythropoietin per milliliter for 1–2 hr, sufficient to achieve upregulation of c-*myc* expression.

3. The cells are harvested. RNA is prepared and subjected to Northern blot analysis for c-*myc* and *GAPDH* expression as described.

4. The levels of steady state c-*myc* mRNA seen in the absence and presence of erythropoietin without oligonucleotides are compared to those found in the presence of erythropoietin and the specified concentration of oligonucleotide. The data are expressed as percent inhibition of erythropoietin-induced c-*myc* expression.

The results indicate that of the six antisense oligonucleotides to PKC isoforms employed (α, β, γ, δ, ϵ, and ζ) only antisense PKC-ϵ oligonucleotides block the erythropoietin signal to c-*myc*. Importantly, this inhibition is antisense oligonucleotide concentration dependent. Moreover, antisense ϵ oligonucleotide blocks the erythropoietin induction of c-*myc* expression only down to that level of expression seen in the absence of erythropoietin and not below. This is consistent with the blocking of erythropoietin signal to the transcriptional attenuation site in the c-*myc* gene without affecting transcriptional initiation of c-*myc*.

Simultaneous with carrying out the above experimental protocol, appropriate controls should be employed. Of special importance are the use of sense oligonucleotides, that is, oligonucleotides whose sequence is identical to the coding sequence of the target codons, as well as so-called "missense" oligonucleotides, which are oligonucleotides of the same base composition as the antisense oligonucleotides employed but of scrambled sequence.

Verification of the specificity of antisense oligonucleotide action is critical. In this regard, two means of verification can be used and, if possible, both should be attempted. The first is the demonstration that the antisense oligonucleotide actually downregulates the target transcript. This is most conveniently done by reverse transcription-polymerase chain reaction (RT-PCR). Primers should be selected so that a single readily identifiable RT-PCR product is found in RNA obtained from control cells and is not detected using RNA obtained from cells treated with antisense oligodeoxynucleotide. Simultaneous amplification of a control gene, such as *GAPDH,* should be included in all PCR reactions to verify specificity of the downregulation. Obviously the RT-PCR analysis of missense and sense oligonucleotide-treated cells should reveal maintenance of expression of the target gene.

In addition to RT-PCR demonstration of mRNA downregulation, downregulation of the protein should also be demonstrated. This is most conveniently done using various antibody methods amenable to unequivocal analysis. Experiments in the author's laboratory have utilized both immunohistochemical staining of cells fixed onto glass slides and Western blot analysis using specific antibodies. Obviously, it is not essential that the target protein be completely absent from the cell in order to verify that the antisense oligonucleotide has resulted in a specific downregulation. However,

reason dictates that unless something near to 100% disappearance (e.g., 60–80% inhibition) is achieved the specificity of the antisense oligonucleotide effect and the reliability of conclusions based on it are open to question.

Conclusion

The methods described above serve to illustrate some of the approaches available in the study of signal transduction pathways and the genes involved in normal and neoplastic blood cell differentiation. It is anticipated that the investigator will individualize these methods for the target signal/effector molecules under study, the genes whose expression is being regulated, and the cells in which these studies are being carried out. Undoubtedly, as information on growth factor signaling and gene regulation in hematopoietic cell growth and differentiation accumulates, even more powerful techniques will be developed.

Acknowledgments

This work was supported in part by a contract from the U.S. Navy (#N000149310776) and a grant from the NIH (#5RO1 DK38841-08).

References

1. C. Friend, W. Scher, J. G. Holland, and T. Sato, *Proc. Natl. Acad. Sci. U.S.A.* **68,** 378 (1971).
2. Y. E. Chern, R. Spangler, H.-S. Choi, and A. J. Sytkowski, *J. Biol. Chem.* **266,** 2009 (1991).
3. N. J. DeBoth, M. Vermey, E. van't Hull, E. Klootwijk-van Dijke, L. J. L. D. van Griensven, L. N. M. Mol, and R. J. Stoof, *Nature (London)* **272,** 626 (1978).
4. A. J. Sytkowski, J. J. Salvado, G. M. Smith, C. J. McIntyre, and N. J. DeBoth, *Science* **210,** 74 (1980).
5. A. Yoshimura, G. Longmore, and H. F. Lodish, *Nature (London)* **348,** 647 (1990).
6. O. Miura, A. D'Andrea, D. Kabat, and J. N. Ihle, *Mol. Cell. Biol.* **11,** 4895 (1991).
7. H. Lachman and A. L. Skoultchi, *Nature (London)* **310,** 592 (1984).
8. R. Ramsay, K. Ikeda, R. A. Rifkind, and P. A. Marks, *Proc. Natl. Acad. Sci. U.S.A.* **83,** 6849 (1986).
9. S. Yonekura, Y. Chern, K. A. Donahue, L. Feldman, G. J. Vanasse, and A. J. Sytkowski, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 2535 (1991).
10. J. M. Chirgwin, A. E. Przybyla, R. J. MacDonald, and W. J. Rutter, *Biochemistry* **18,** 5294 (1979).
11. R. Spangler, S. C. Bailey, and A. J. Sytkowski, *J. Biol. Chem.* **266,** 681 (1991).

12. R. Spangler and A. J. Sytkowski, *Blood* **79,** 52 (1992).
13. C. L. Hatch and W. M. Bonner, *Anal. Biochem.* **85,** 8900 (1987).
14. H. R. Patel, H.-S. Choi, and A. J. Sytkowski, *J. Biol. Chem.* **267,** 21300 (1992).
15. T. M. Natalis and D. Godbolt, *Nucleic Acids Res.* **19,** 4301 (1991).
16. M. Linial, N. Gunderson, and M. Groudine, *Science* **230,** 1126 (1986).
17. Y. Chern, C. O'Hara, and A. J. Sytkowski, *Eur. J. Biochem.* **202,** 78 (1991).
18. Y. Li and A. J. Sytkowski, Presented at the 35th Annual Meeting of the American Society of Hematology Meeting, St. Louis, Missouri, December 3–7, 1993.

This Page Intentionally Left Blank

Section VII

The Mouse as a Model System for Human Molecular Genetics

This Page Intentionally Left Blank

[25] Selected Methods Related to the Mouse as a Model System

Evie Melanitou, Marie-Christine Simmler, Edith Heard, Claire Rougeulle, and Philip Avner

Introduction

The power of model systems for the study of genome structure and function and for the dissection of complex traits is, by now, well established. Advances in molecular genetics have made it possible to combine classic genetics with molecular biology and to throw light on the causes of monogenic diseases as well as on complex multigenic traits.

For a variety of reasons the mouse remains the experimental mammal of choice for genetic research. The possibility of establishing crosses that exploit the many strains and mutations accumulated over the years constitutes just one reason for its extensive use. Its relatively small size, short generation time, and ease of handling provide additional advantages. Unlike in humans, for example, heterogeneity is not a problem in the mouse due to the ability to engineer breeding with defined parents: usually all offspring will be informative for particular markers. By generating backcrosses of several hundreds of animals, single-gene mutations can be positioned to within a few hundred kilobases.

Mouse models for diseases or traits under complex genetic control are also becoming increasingly amenable to detailed study. Indeed, using backcrosses or intercrosses between polymorphic strains followed by a comprehensive scan of the genome to characterize the inheritance patterns, coupled with careful quantitative phenotypic analysis, it is possible to establish the contribution of chromosomal regions to a disease phenotype. Moreover, the possibility of constructing congenic lines by introgressing a genetic region of interest from one parent into the genetic background of the other should facilitate the genetic dissection of a particular trait.

The combination in the mouse of powerful genetics, well-characterized biology, and highly developed experimental embryology should prove of particular value to such studies.

Additional demands on mouse genetics are likely to arise both from the increasing need to validate and extend findings concerning the human genome by experimental manipulation and the requirement to define new experimental models for human disease. Continuing technical innovations, including the use of techniques initially applied to the human genome, will underpin what may turn out to be a new golden age of mouse genetics and embryology.

Some of the techniques and experimental approaches underlying advances in human genetics, such as yeast artificial chromosome (YAC) cloning or microsatellite identification, can be directly applied to the mouse. Others are of less direct interest

Methods in Molecular Genetics, Volume 8

for mouse genome analysis. The high level of informativeness of the genetic markers used in the inter- and intraspecies crosses that construct the mouse genetic map has, for instance, meant that until recently radiation hybrids have been less widely exploited in the mouse than in humans. Still other techniques used in human genome studies, while of interest in the mouse, have had to be adapted to allow their potential to be realized.

This chapter does not aim to provide a compendium of all molecular techniques that are relevant to the mouse. Rather, it seeks to provide insights into some of the strengths and originality of mouse genetics that may prove of help to the nonmouse geneticist and to highlight selected techniques that have contributed to progress in mouse genetics.

Mouse Crosses

The initial steps in positional cloning involve the definition of the genetic interval containing a disease gene or mutation. The size of the genetic interval depends on both the number of informative meioses and the number of polymorphic markers available for the region under study. The latter can be maximized by using crosses in which one of the parents is derived from a strain that will introduce a high degree of genetic diversity. Because standard inbred laboratory mice are closely related in genetic terms due to their derivation from a fairly limited number of progenitor strains, the use of inbred lines derived from feral mice belonging to either a distinct *Mus* species such as *Mus spretus* or to one of the *Mus* subspecies can prove profitable (1, 2). Although feral mice, depending on their species or subspecies relationship, often do not breed together in the wild, most can be bred either with each other or with standard laboratory mouse strains under animal house conditions.

Genetic Diversity within Genus Mus

The genus *Mus* has been divided into a complex species containing the four major biochemical groupings *Mus musculus domesticus, Mus m. musculus, Mus castaneus,* and *Mus bactrianus,* all of which can be treated as subspecies, in addition to several distinct species such as the western Mediterranean mouse *M. spretus* (3). The degree of fertility between different members of the genus ranges from complete sterility to reduced fertility and is correlated with the degree of taxon divergence.

Interspecies and Intersubspecies Crosses

A major thrust in mouse genetics over the last 10 years has revolved around the exploitation of interspecies crosses involving *M. spretus*-derived strains and, more

recently, intersubspecies crosses. The degree of genetic diversity introduced is sufficient to allow the easy mapping of the majority of transcribed sequences by restriction fragment length polymorphism (RFLP) analysis using a small number of restriction enzymes (4, 5). Although genetic diversity is maximized through the use of crosses involving *M. spretus,* as for example in the establishment of the European Backcross Collaborative Group (EUCIB) mapping panel (6), its use imposes constraints on the type of cross-configurations that can be exploited.

Interspecies crosses involving *M. spretus* and laboratory mice are, for reasons of efficiency, normally set up using the laboratory mouse as female, although progeny can be obtained in both cross-configurations. The resulting F_1 females are fertile and are used to produce the backcross generation, while F_1 males, which are usually sterile, are not used for further breeding. Hybrid sterility is one of the major disadvantages inherent in interspecies crosses, making it almost impossible to obtain progeny from intercrosses. Not only are F_2 progeny rarely recovered but the development of both recombinant inbred and recombinant congenic strains is severely compromised. Subspecies crosses, on the other hand, do not suffer from this disavantage in general and their use normally confers a sufficient level of exploitable polymorphism. Male infertility is not observed, for example, when the inbred strain PWK derived from a feral mouse belonging to the *M. musculus musculus* subspecies is used as a mating partner with laboratory mouse strains. This strain can therefore be used for both intercross, recombinant inbred strain, and recombinant congenic strain construction (for reference to these breeding systems see Ref. 7).

Although the initial use of interspecies and intersubspecies crosses predates the availability of highly informative microsatellite markers, such crosses not only permit the integration of both microsatellite markers and transcribed sequences but also increase the number of informative microsatellite markers that can be exploited (8).

The use of interspecies and intersubspecies crosses for mapping single-gene defects is well established. In contrast, their use for defining and cloning genes involved in multigenic defects, such as type 1 diabetes, and quantitative trait loci (QTL) is less well developed and it is important to appreciate the advantages and disavantages introduced by their use.

The basic strategy involved in establishing the contribution of particular chromosomal regions to the overall genetic variance and phenotype of multigenic diseases does not differ in an interspecies/intersubspecies cross from that applied to intraspecies crosses. In each case the pattern of inheritance revealed by a detailed and comprehensive genome-wide scan must be correlated with a detailed analysis of the phenotypes recovered. Among the advantages conferred by interspecies and subspecies crosses is the greater genetic variation introduced, which may allow additional controlling loci to be identified (9). A second advantage is conferred on interspecies crosses by the higher density of informative markers available throughout the genome. This facilitates the initial detection of linkage to the disease and renders the subsequent refinement of the genetic region responsible for the phenotype much

more efficient. Disadvantages of such crosses are, first, that the increased genetic complexity can sometimes reduce the power of the genetic analysis, in part because the distribution of the most highly informative extreme phenotypes is strongly reduced (9). Second, the higher levels of "irrelevant" allelic sequence variation between the parental strains used in such interspecies crosses may complicate the use of comparative sequence analysis in the ultimate identification of the gene(s) responsible for a particular trait.

Genome Scanning in Mouse

Irrespective of the type of cross used, efficient and rapid scanning of the entire mouse genome is necessary for linkage analysis. Although microsatellite typing by PCR (polymerase chain reaction) is the most widely used technique, the use of defined inbred parents and the accompanying absence of heterogeneity has facilitated the application of techniques such as random amplified polymorphic DNA analysis (RAPD) (10) and multilocus DNA fingerprint analysis (12).

Multilocus DNA Fingerprint Analysis

Human VNTR (variable number of tandem repeat) loci correspond to stretches of minisatellite DNA, 30–50 nucleotides in length, that are repeated in tandem from 10 up to many of hundreds of times at a particular genomic location. Variation in the length or number of the basic repeat unit is responsible for the high degree of polymorphism and the detection of multiple alleles. When a single VNTR probe is hybridized at low stringency to DNAs restricted with a frequent cutter enzyme such as *Hin*fI, the resulting extensive cross-hybridization leads to characteristic complex "fingerprint" hybridization patterns (11). Such patterns are linked to the presence in common in VNTRs of a limited number of "core sequences" of similar nucleotide composition.

Many human VNTRs cross-hybridize under conditions of reduced stringency to reveal complex minisatellite polymorphisms not only in humans but also in other species such as the mouse, demonstrating the widespread nature of such sequences (12, 13).

Fingerprint patterns provide an abundance of information for linkage studies and can be easily used for genome scanning in appropriately designed crosses between inbred mouse strains (14). The detection of several unlinked loci by a single probe (12) ensures rapid screening of the genome.

Sample Preparation, Electrophoresis, and Hybridization

Because the use of the *Taq*I, *Hae*III, and *Hin*fI restriction enzymes has been shown to allow the majority of minisatellite loci to be identified, mouse DNAs restricted

with these three enzymes under standard conditions are subjected to agarose gel electrophoresis and blotted prior to hybridization with human VNTR probes. *Taq*I blots often detect cross-hybridizing fragments that are not revealed by the other restriction enzymes used and, in addition, allow standard RFLP detection of other loci. *Hae*III- and *Hin*fI-restricted Southern blots must be run under electrophoresis conditions that allow retention of fragments >1.5 kb (14). Electrophoresis conditions for *Taq*I-restricted DNAs should be set to allow retention of fragments >0.5 kb.

After electrophoresis the gels are denatured in 0.4 *N* NaOH for 15 min, then transferred overnight onto charged nylon membranes, such as Gelman or Pal, in 0.4 *N* NaOH. After neutralization in 50 m*M* sodium phosphate (pH 6.5) for 30 min at room temperature, blots are prehybridized for 4 hr in 6× SSC (1× SSC: 0.15 *M* NaCl–15 m*M* sodium citrate), 2× Denhardt's solution [0.02% (w/v) bovine serum albumin (BSA)–0.02% (v/v) Ficoll–0.02% (w/v) polyvinylpyrrolidone], 2 m*M* sodium phosphate, 0–100 μg of salmon sperm DNA per milliliter, 1% (w/v) sodium dodecyl sulfate (SDS), and 40% (v/v) formamide at 37 or 42° C. Ten million cpm/ml of probe is then added to the hybridization buffer and hybridization carried out overnight under identical temperature conditions. Probe is labeled with [^{32}P]dCTP (3000 Ci/mmol; Amersham, Arlington Heights, IL) using the Klenow fragment of DNA polymerase I under standard conditions in a total volume of 50 μl (15).

The blots are initially washed twice at room temperature in 2× SSC–0.1% (w/v) SDS buffer for 15 min. Subsequent washes are at 42 or 65° C, depending on the intensity of the hybridization signals.

The human VNTR probes that have been shown to reveal complex fingerprint patterns with mouse DNA and used successfully for fingerprint analysis include *D1S57, D2S44, D3S45, D14S1,* and *D17S24.* These probes are described in Ref. 14.

Microsatellite Typing

Most microsatellite mouse markers have been identified by screening genomic libraries with repetitive oligonucleotide probes (16). A minority, mainly associated with genes, have been characterized by database searches (17). The 6000 microsatellite loci that have been identified in the mouse appear to be relatively uniformly distributed throughout the genome (18). Although 30–60% of such microsatellite markers are informative in crosses involving inbred laboratory strains, over 80% are informative when tested in interspecies and intersubspecies crosses. Primers for these microsatellites are commercially available from Research Genetics (18). "Multiplexing" of PCR reactions by coamplifying two or more pairs of PCR primers for distinct microsatellite fragments can be achieved by judicious choice of primer pairs and is routinely used in our laboratory. The choice of primer pair combinations is mainly based on the size of the fragments amplified, but the efficiency of amplification is also determinant. The use of radioactive [^{32}P]dCTP in the protocol outlined below allows direct detection of the corresponding amplification products without prior

transfer of the acrylamide gel or hybridization. The technique is relatively inexpensive and highly sensitive.

Polymerase Chain Reactions

All polymerase chain reactions are carried out in 96-well U-bottom flexible plates (Cat. No. 3911; Falcon, Oxnard, CA) in a total volume of 10 μl.

Genomic DNA (12.5 ng/μl)	4 μl
Distilled H_2O	3.19 μl
Taq PCR buffer (10×)	1 μl
dNTPs (10 μM each)	0.2 μl
Primer A (6.6 μM)	0.24 μl
Primer B (6.6 μM)	0.24 μl
Taq polymerase (5 U/μl)	0.2 μl
[^{32}P]dCTP (3000 Ci/mmol)	0.033 μl
$MgCl_2$ (25 mM)	0.6 μl

The 10× PCR buffer used contains 500 mM KCl–100 mM Tris-HCl (pH 8.3) and is provided with the *Thermus aquaticus* (*Taq*) polymerase (Perkin-Elmer/Cetus, Norwalk, CT) used in our laboratory.

The procedure used is as follows: DNA is first loaded into the 96-well plates, then 6 μl of the radioactive PCR mix prepared in bulk is distributed to each well, using a multichannel pipette. Samples are amplified in a Techne PH-3 DNA thermal cycler, using an initial denaturation step for 3 min at 94° C, followed by 30 cycles of denaturation at 94° C for 15 sec, annealing at 55° C for 2 min, and elongation at 72° C for 2 min. This is followed by a final elongation step of 7 min.

Sequencing Gel Loading and Migration

Samples are analyzed by polyacrylamide gel electrophoresis on sequencing gels.

Acrylamide (6%, w/v) sequencing gel solution (per liter)

Urea	500 g
Tris–borate–EDTA (TBE, 10×)	100 ml
Acrylamide (40%, w/v)	150 ml
Acrylamide–bisacrylamide, 19:1	
Distilled H_2O	To volume

Incubate at 55° C to dissolve. Filter the solution through 0.45-μm pore size disposable sterile filters (Corning, Corning, NY)

TBE buffer (10×) (per liter)

Tris base	121.14 g
Boric acid	51.32 g
EDTA	3.72 g

Sequencing solution (75 ml), to which 400 μl of 10% (v/v) ammonium persulfate solution and 10 μl of *N,N,N',N'*-tetramethylethylenediamine (TEMED) are added, is used per gel.

After polymerization the gels are prerun for 10–15 min at a setting of 45 W, 25 mA, 1400 V. Usually gels are run for 2–3 hr, although the precise gel running time is dependent on the size of the amplification fragments and the size difference between the detected alleles.

The gels are directly exposed to X-ray autoradiographic films [Kodak (Rochester, NY)] (XAR). Exposure time is usually 12 hr at −80° C.

Somatic Cell Hybrids and Irradiation/Fusion Gene Transfer Hybrids for Genome Analysis in Mouse

Somatic cell hybrids, particularly those carrying well-defined subchromosomal regions, have proved to be a valuable resource for the physical mapping of the human genome. Such hybrids have not only been used directly for mapping purposes but also for the isolation of new probes, thereby increasing the density of markers and the resolution of the genetic and physical maps of the region in question. Relatively few somatic cell hybrids are available in the mouse compared to humans. This is partly because the power of meiotic mapping has led to less emphasis being placed on the isolation of somatic cell hybrids in the mouse. This tendency has, however, been exacerbated by the somewhat poor segregation of mouse chromosomes in mouse × Chinese hamster hybrids, which results in the retention of relatively large numbers of mouse chromosomes in the majority of such hybrids. Efforts to isolate monochromosomal hybrids for the mouse by microcell fusion represent one laborious way of improving the availability of reduced content of mouse chromosome hybrids (P. Avner and P. Goodfellow, in preparation, 1995). Somatic cell hybrids carrying specific chromosomal subregions of the mouse genome can, as in humans, be isolated by exploiting well-characterized chromosomal variants, in Ref. 19. The relatively small number of such variants for any particular chromosome in the mouse has, however, limited the scope of such strategies. An alternative approach to isolating small subchromosomal fragments on a hybrid background involves the use of the irradiation and fusion gene transfer technique (IFGT) (20). This procedure makes use of a lethal dose of irradiation to break chromosomes into small but intact fragments. Such fragments are rescued by fusion of the irradiated donor cell line with a nonirradiated recipient Chinese hamster cell line. The method, initially used to generate human IFGT hybrids (21, 22), has been shown to be applicable, with only small modifications, to the generation of mouse IFGT hybrids (23). Such hybrids represent a potentially valuable cloning resource for generating probes from particular chromosomal regions.

Genome Amplification Using Primers Directed to Interspersed Repetitive Sequences

Isolation of chromosome-specific DNA sequences from monochromosomal or IFGT hybrids is complicated by the presence of the heterologous rodent background. One method for rapidly isolating sequences from the chromosome of interest present in somatic cell hybrids is the interspersed repetitive sequence polymerase chain reaction (IRS-PCR), which combines the specificity conferred by the divergence of interspersed repetitive sequences among different species to the application of the PCR. This approach, as originally described in the human by Nelson and co-workers (24, 25), enables the amplification of unique DNA sequence situated between repeated *Alu* elements that are in the appropriate orientation. The underlying methodology can be applied to all the major classes of interspersed repetitive elements whether of the SINE (short interspersed nuclear element) or the LINE (long interspersed nuclear element) families in humans and mouse.

The most highly repeated member of the mouse SINE family is the B1 sequence, which is partly homologous to the left-hand monomer of the *Alu* sequence, the major human short interspersed element. Although B1 elements in primates and rodents differ in their overall structure, they all share extensive homology with each other and with a small nuclear RNA. The other major murine repeat element belonging to the SINE family, B2, is found only in rodent genomes and has no human equivalent (26–28). The mouse L1 repeat family (L1Md), belonging to the long interspersed element family (LINE), has, like B1, a human homolog. In both species, most LINEs are variably truncated at their 5′ end. The R repeat element located within the 3′-most conserved end of such murine L1Md elements is, however, rodent specific (26–28).

A variety of primers have been designed from such repeat elements and PCR conditions developed that allow amplification of murine sequences from heterospecific somatic cell hybrids. The efficiency of the reactions and the complexity of the murine PCR products obtained are generally greater from mouse × human hybrid cell lines than from mouse × Chinese hamster hybrid cell lines. One consequence of this is that primers and PCR conditions used to specifically amplify mouse sequences from mouse × Chinese hamster hybrids can be used successfully with mouse × human hybrid DNAs, but the converse is not true.

IRS-PCR Conditions and Primers

Inter-L1 and B2-L1 PCR: Mouse versus Chinese Hamster

Simmler and co-workers (28) have described mouse-specific amplification of mouse × hamster hybrids using L1 and B2 primers.

Reagents

Tris-HCl (pH 8.8), 67 m*M*
$(NH_4)_2SO_4$, 16 m*M*

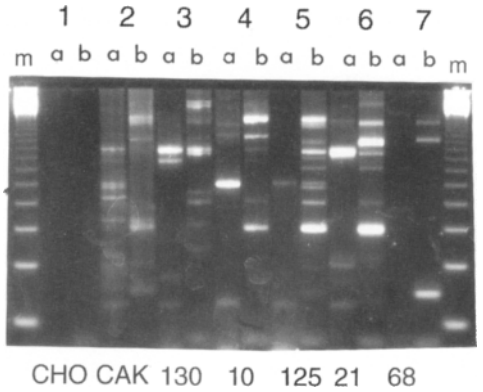


FIG. 1 IRS-PCR of Chinese hamster ovary (CHO) (lane 1), HYBX (CAK) (lane 2), and IFGT hybrid DNAs: hybrids 130 (lane 3), 10 (lane 4), 125 (lane 5), 21 (lane 6), and 68 (lane 7), using (a) the B2A.1(*Not*I)/R2(*Eco*RI) or (b) the R2(*Not*I)/R2(*Eco*RI) primer combination. m, 123-bp BRL ladder size marker (28).

$MgCl_2$, 6.7 m*M*
2-Mercaptoethanol, 10 m*M*
dNTPs, 200 μM
Thermus aquaticus polymerase, 4 U (Perkin-Elmer/Cetus)
Primer pairs R2(*Eco*RI)/R2(*Not*I) or R2(*Eco*RI)/B2A.1(*Not*I) (final concentration of each primer, 2 μM)

Both combinations of primers specifically amplify unique bands from mouse DNA and not from Chinese hamster DNA (see Fig. 1). The PCR reactions are carried out in a total volume of 50 μl, using 1 μg of genomic DNA.

PCR Cycle Conditions

The PCR cycle conditions consist of an initial denaturation step of 4 min at 94° C. This is followed by 35 cycles of 94° C denaturation for 1 min, 60° C annealing for 48 sec, and 68° C extension for 5 min. A final extension step consists of 7 min at 68° C. The reaction conditions have been optimized for use with a Techne PH-3 DNA thermal cycler.

Oligonucleotide Primers

R2(*Eco*RI) primer: CCGAATTCAGCATTTGAAATGTAAATGA
R2(*Not*I) primer: AAGTCGCGGCCGCAGCATTTGAAATGTAAATGA
B2A.1(*Not*I) primer: AAGTCGCGGCCGCTCTTCTGGAGTGTCTGAAGA

All the primers have a 5′ end extension containing either a *Not*I or an *Eco*RI cloning site (underlined in the sequence) and are always used in association to facilitate further cloning procedures.

Inter-L1 PCR: Mouse versus Chinese Hamster

Irving and Brown (29) have described conditions from mouse-specific amplification of mouse × Chinese hamster hybrids using inter-L1 sequence from the R element of the mouse L1Md repeat.

Reagents

Tris-HCl (pH 8.5), 10 m*M*
KCl, 50 m*M*
$MgCl_2$, 1.5 m*M*
Gelatin, 0.01% (w/v)
dNTPs, 200 μM
Thermus aquaticus polymerase, 2.5 U (Perkin-Elmer/Cetus)
L1 repeat primer (1 μM final)

The PCR is carried out in a total volume of 100 μl, using 500 ng of genomic DNA.

PCR Cycle Conditions

The PCR cycle conditions consist of an initial denaturation step of 6 min at 94° C. This is followed by 35 cycles of 94° C denaturation for 1 min, 60° C annealing for 45 sec, and 68° C extension for 5 min. The conditions are optimized for the Perkin-Elmer/Cetus thermal cycler.

Oligonucleotide Primer

L1 primer: GGTATGGGGGACTTTTGGGAT

Inter-B1 PCR: Mouse versus Human

Herman and co-workers (30) have described mouse-specific amplification of mouse × human hybrids using inter-B1 PCR.

Reagents

Tris-HCl (pH 8.0), 10 m*M*
KCl, 50 m*M*
$MgCl_2$, 1.5 m*M*
Gelatin, 0.01% (w/v)
dNTPs, 200 μM
Thermus aquaticus polymerase, 2.5 U (Perkin-Elmer/Cetus)
B1 primer (1 μM final)

The PCR is carried out in a final volume of 100 μl, using 300 ng of genomic DNA.

PCR Cycle Conditions

An initial denaturation step of 8.5 min at 94° C is followed by 30 cycles of 92° C denaturation for 1 min and 72° C annealing/extension for 6 min. A final extension step consists of 7 min at 72° C. Conditions are optimized for use with a Perkin-Elmer/Cetus thermal cycler.

Oligonucleotide Primer

B1 primer: CAGAGTGAGTTCCAGGACAGCCAGGG

Characterization and Isolation of Polymerase Chain Reaction Products as Hybridization Probes

To visualize the IRS-PCR products, 15-μl aliquots are electrophoresed on 4% (3% NuSieve: 1% SeaKem; FMC, Philadelphia, PA) agarose gels in 1× TAE buffer and stained with ethidium bromide.

For band isolation, 25 μl of the PCR products to be processed is electrophoresed on 3% NuSieve gels in 1× TAE on a 20-cm gel tray at 50 V overnight using a 132-bp BRL (Gaithersburg, MD) ladder as size marker. Individual fragments are cut out of the gel, melted at 68° C, and diluted with 100 μl of water. To further purify the fragment, 5-μl samples are then reamplified for another 35 cycles. After further electrophoresis and band isolation, aliquots containing 20 ng of melted fragments are labeled by random priming (31). Such probes are then characterized for their specificity by hybridization to Southern blots containing hybrid DNAs. In the example presented in Fig. 1, probes were tested against *Eco*RI-digested DNAs of hybrid cell lines containing all or part of the mouse X chromosome. These DNAs allowed both the specificity and the subchromosomal localization of the derived probe to be assessed. Chinese hamster and mouse DNAs are systematically included as controls.

The PCR product hybridizations are performed in Church's buffer [0.5 *M* sodium phosphate (pH 7.2), 1 m*M* EDTA, and 7% (w/v) SDS], in the presence of an excess of unlabeled competitor mouse DNA to eliminate any signal produced by the repeated primer sequences: 2–4 ng of radioactively labeled probe adjusted to 100 μl with water is boiled for 10 min in the presence of 76 μl of 1 m*M* EDTA–10 m*M* Tris (pH 8.0) containing 300 μg of total mouse spleen DNA sonicated to obtain 200- to 500-bp fragments. Immediately after boiling, 24 μl of a 1.0 *M* sodium phosphate solution adjusted to pH 6.8 is added (0.12 *M* final concentration) and the DNA mix incubated for 2–4 hr at 65° C, and then used directly for hybridization. Washing conditions depend on the probe: in 30–40 min, two washes with 2× SSC–0.1% (w/v) SDS followed by one wash with 1× SSC–0.1% (w/v) SDS or one wash with 0.5× SSC (or 0.1× SSC)–0.1% (w/v) SDS at 65° C.

After characterization for their specificity, PCR fragments are isolated using a direct cloning procedure. Aliquots containing 100 ng of melted DNA fragments are

ligated into 10 ng of *Eco*RI/*Not*I-cut pBluescript vector (Stratagene, La Jolla, CA) using standard procedures and ligation products transformed into the DH5α *Escherichia coli* strain (GIBCO-BRL, Gaithersburg, MD).

Alternative protocols have been developed in which the products of the PCRs are directly cloned without prior characterization. The cloned PCR products are then characterized for their specificity by hybridization as outlined above.

Applications of Yeast Artificial Chromosomes to Mouse

Regions that have been defined by meiotic mapping, radiation hybrid mapping, or through the availability of chromosomal deletions and translocations may typically span several megabases of DNA. Yeast artificial chromosomes (YACs) with their large cloning capacity (50–1000 kb) allow analysis of such extensive regions. The availability of physical maps based on the assembly of overlapping YACs within contiguous sequences of DNA (contigs) that span a chromosomal region containing loci of interest has greatly facilitated the isolation of such genes in humans and is proving to be an equally powerful tool in the mouse.

The YAC cloning system also offers the advantage of high-efficiency homologous recombination in the yeast host, allowing mammalian DNA cloned in a YAC to be readily manipulated *in vitro* to generate a range of alterations including deletions, insertions, and even base pair substitutions (32; for review see Ref. 33). One example of such homologous recombination-based modifications is YAC fragmentation, which allows the generation of a series of nested deletion derivatives from a YAC, and the application of this technique to YACs carrying mouse inserts is described below. Another example is the introduction of a marker gene into a YAC, either for selection of its uptake into mammalian cells or as a reporter gene to investigate the influence of regulatory sequences present in the YAC. An example of the latter is also described below.

Yeast Artificial Chromosome Fragmentation

Yeast artificial chromosome fragmentation involves the generation of deleted or shortened derivatives of a parental YAC by introduction into the yeast of a linearized fragmentation vector, one end of which is homologous to sequences within the YAC insert and the other end of which carries a telomere (34). Homologous recombination between the incoming plasmid and the YAC results in the introduction of a telomere, terminating the YAC at the integration site and resulting in the deletion of all of the distal sequences. Clones containing shortened YACs are selected based on the acquisition and loss of appropriate markers. The target sequence may be unique, resulting in the production of a specific fragment, or repetitive, resulting in the generation of a nested series of deletion derivatives. The former can be used to facilitate the identi-

fication and mapping of exons within a gene by using a cDNA as the targeting sequence (35). The latter is particularly useful for the construction of restriction maps of very large YACs (34) or, for example, in determining the position of regulatory elements or genes in functional assays involving the transfer of YACs into mammalian cells.

To create shortened derivatives from YACs containing mouse inserts, fragmentation vectors have been created that specifically target mouse genomic DNA using the mouse B1 repeat element (36). The distribution of B1 repeats in the mouse genome, if it were random, is estimated to be approximately once every 20 kb and the divergence of individual sequences from the B1 consensus is thought to be minimal. The YAC fragmentation strategy and vectors are based on those originally developed by Pavan *et al.* (34) for human YACs and are shown in Fig. 2 (36a).

The YAC to be fragmented must be present in a strain mutated for *HIS3*, as this marker is carried by the fragmentation vectors. The YACs present in the AB1380 strain (*his5 HIS3*), which has been used for the majority of YAC libraries, must therefore be transferred into a *his3* background. Ideally a strain carrying a *his3* deletion mutation rather than a point mutation should be used as this avoids the generation of spontaneous His^+ revertants. The fragmentation vector is then introduced into the YAC-containing yeast strain by spheroplast transformation, as described below.

Alternative mouse B1 fragmentation vectors that incorporate a *LYS2* gene rather than the *HIS3* gene and hence do not necessitate transfer of a YAC from AB1380 (which is a *lys2* mutant) have been developed (by J. Edmondson and R. Rothstein, personal communication, 1995). However, the presence of a *lys2* point mutation rather than a deletion mutation in AB1380 means that a certain proportion of *LYS2* revertants will be obtained among the Lys^+ transformants.

Yeast Media

See Ref. 37 for standard recipes.

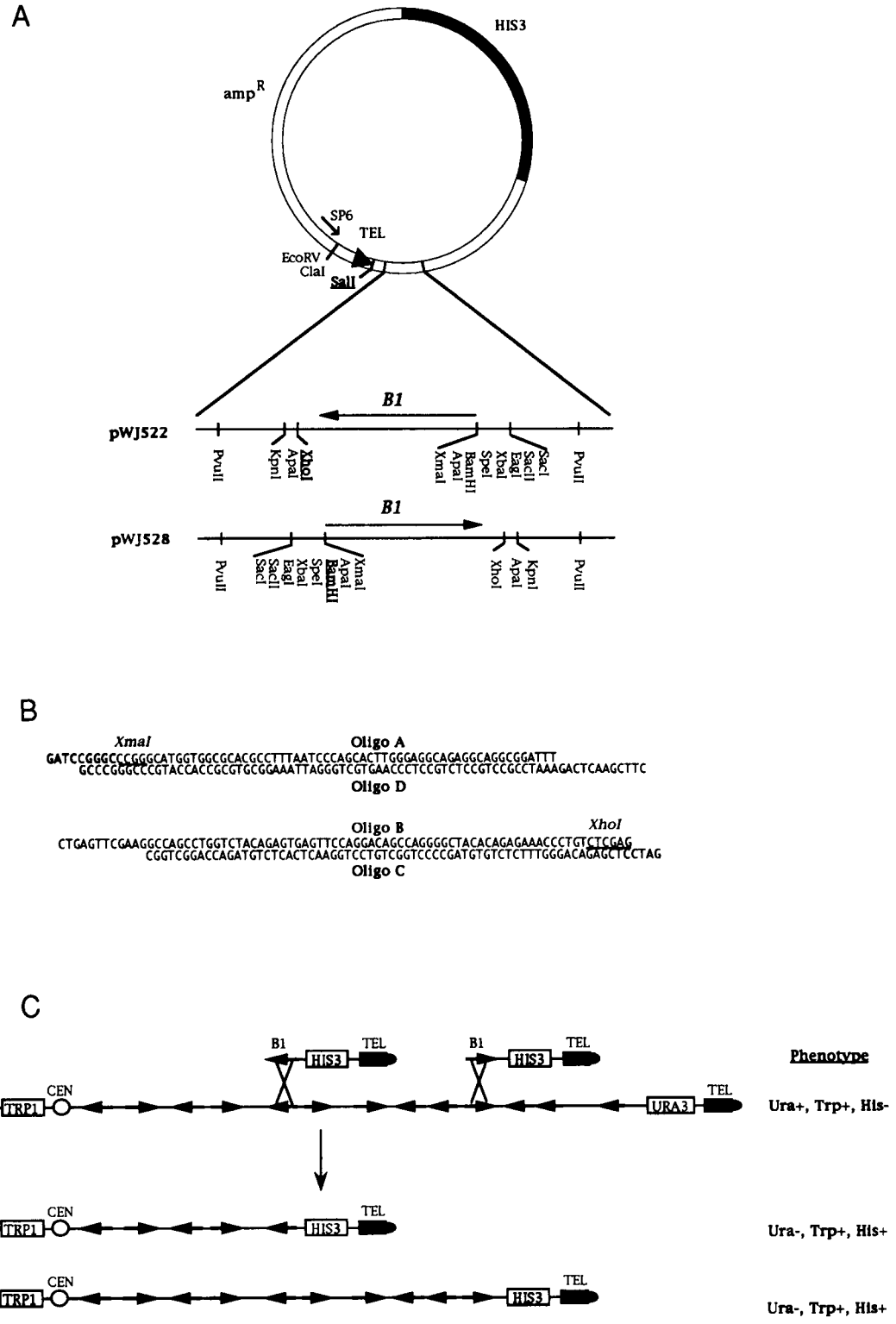
AHC (a rich-Ura-Trp medium for routine growing of YACs): per liter:

Glucose	20	g
Yeast nitrogen base with and without amino acids (Difco, Detroit, MI)	6.7	g
Casein acid hydrolysate (Sigma, St. Louis, MO)	10	g
Adenine hemisulfate (warm to dissolve), 1% (w/v)	1	ml

Add water, and adjust the pH to 5.8 with a few drops of 4 *M* NaOH; for solid medium add 17 g of agar (Difco). Autoclave for 20 min/1 liter, 30 min/2 liter, etc.

SORB plates

Synthetic medium (SD) plates [0.17% (w/v) yeast nitrogen base without amino acids, 2% (w/v) glucose (pH 5.8)] containing 0.9 *M* sorbitol and


A
HIS3
amp^R
SP6
TEL
EcoRV
ClaI
SalI
B1
pWJ522
PvuII
KpnI
ApaI
XhoI
XmaI
ApaI
BamHI
SpeI
XbaI
EagI
SacII
SacI
PvuII
B1
pWJ528
PvuII
SacI
SacII
EagI
XbaI
SpeI
BamHI
ApaI
XmaI
XhoI
ApaI
KpnI
PvuII
B
XmaI
Oligo A
GATCCGGGCCCGGGCATGGTGGCGCACGCCTTTAATCCCAGCACTTGGGAGGCAGAGGCAGGCGGATTT
GCCCGGGCCCGTACCACCGCGTGCGGAAATTAGGGTCGTGAACCCTCCGTCTCCGTCCGCCTAAAGACTCAAGCTTC
Oligo D
Oligo B
XhoI
CTGAGTTCGAAGGCCAGCCTGGTCTACAGAGTGAGTTCCAGGACAGCCAGGGGCTACACAGAGAAACCCTGTCTCGAG
CGGTCGGACCAGATGTCTCACTCAAGGTCCTGTCGGTCCCCGATGTGTCTCTTTGGGACAGAGCTCCTAG
Oligo C
C
B1
HIS3
TEL
B1
HIS3
TEL
Phenotype
TEL
CEN
TRP1
URA3
TEL
Ura+, Trp+, His-
TEL
CEN
TRP1
HIS3
TEL
Ura-, Trp+, His+
TEL
CEN
TRP1
HIS3
TEL
Ura-, Trp+, His+

supplemented with appropriate amino acids (20 μg/ml each of adenine sulfate, L-arginine hydrochloride, L-histidine hydrochloride, L-methionine, L-tryptophan, and uracil; 30 μg/ml each of L-isoleucine, L-lysine hydrochloride, and L-tyrosine; 50 μg/ml of L-phenylalanine; 60 μg/ml of L-leucine; 150 μg/ml of L-valine). For example, −His SORB plates contain all of the above amino acids except histidine

TOP

Sorbitol (1 *M*), 2.5% (w/v) agar in SD medium supplemented with appropriate amino acids (see above)

Yeast Spheroplast Transformation Reagents

Sorbitol (M_r 182.2), 1 *M*

SCEM

Sorbitol, 1 *M*
Sodium citrate (pH 5.8) (M_r 294.1), 0.1 *M*
EDTA, 10 m*M*
2-Mercaptoethanol, 30 m*M* (14.7 *M* stock solution; add just prior to use)

Lyticase (Sigma) stock solution of 10,000 U/ml in 50 m*M* KPO_4 (pH 7.5)

STC

Sorbitol, 1 *M*
Tris-Cl (pH 7.5), 10 m*M*
$CaCl_2$, 10 m*M*

FIG. 2 Mouse YAC fragmentation vectors. (A) The pBP103 backbone plasmid of Pavan *et al.* (36a) is shown with the two possible orientations of the B1 sequence inserted into the polylinker, generating pWJ522 for one orientation and pWJ528 for the other orientation. The underlined restriction sites are used to linearize the vector: *Sal*I and *Xho*I for pWJ522, *Sal*I and *Bam*HI for pWJ528. The *Eco*RV and *Cla*I sites indicated can be used to isolate the sequences adjacent to the fragmentation site by plasmid rescue in *E. coli.* (B) The sequences of the four oligomers (A, B, C, and D) used to construct the B1 repeat element are shown: the 4-bp (GATC, *Bam*HI site) overlap between the extremities of oligos A and C and should be noted, as should the 12-bp overlap between oligo D and oligo B, which forms the central portion of the B1 sequence. The orientation of this B1 element is as shown for the pWJ528 vector above. (C) Fragmentation mechanism: the linearized fragmentation vector undergoes homologous recombination at one of the B1 repeats in the parental YAC insert, introducing a telomere and deleting all the distal sequences. A vector carrying each orientation of the B1 element is shown pairing with the target YAC. In practice each vector is transformed separately. Deletion derivatives are selected based on acquisition of *HIS3* and screened for loss of *URA3.*

PEG (filter sterilize)
Tris-Cl (pH 7.5)
$CaCl_2$
Polyethylene glycol (PEG) 8000, 20% (w/v)

SOS (filter sterilize)
Sorbitol, 1 *M*
$CaCl_2$, 6.5 m*M*
yeast extract, 0.25% (w/v)
Bacto-Peptone, 0.5% (w/v)

1. Transfer the YAC, if necessary, into a *his3* genetic background by crossing to a strain such as YPH857 (*Mat***a** *leu2-D1 trp1-D63 ura3-52 ade2-101 his3-D200 lys2-801 cyh*^{R}) (37). Crossing, sporulation, and tetrad analysis should be carried out using standard yeast genetics (38). Alternatively, use direct transformation of purified YAC DNA in the presence of polyamines into a *his3* strain (39).
2. Streak the YAC-containing yeast strain onto a selective AHC plate and grow at 30° C for 2–3 days. Inoculate a single colony into 10 ml of liquid AHC medium and grow overnight at 30° C.
3. Use this fresh overnight culture to inoculate 50 ml of AHC. Grow with vigorous aeration at 30° C to a density of about 3×10^7 cells/ml (approximately 3–4 hr).
4. Harvest cells by centrifugation at 3500 rpm for 5 min at room temperature.
5. Wash once with 20 ml of sterile water and once with 20 ml of 1 *M* sorbitol, centrifuging as in step 4 each time.
6. Resuspend in 20 ml of SCEM and add 1000 U of lyticase. Incubate at 30° C, gently inverting the tube from time to time. Monitor the degree of spheroplasting by observing the decrease in OD_{800}, which should reach 10% of the original value after 15–30 min. If this is not the case, add further lyticase and continue the incubation.
7. Centrifuge the spheroplasts at 1000 rpm for 5 min at room temperature. Discard the supernatant carefully and gently resuspend the pellet in 20 ml of 1 *M* sorbitol. Centrifuge again. Resuspend the pellet gently in 2 ml of STC.
8. Aliquot 200 μl of spheroplasts into 15-ml tubes. Add 1 μg of plasmid DNA, appropriately linearized between the B1 repeat and the telomere sequence (see Fig. 2) and 50 μg of sonicated carrier DNA (herring sperm). The total volume of DNA should not exceed $\frac{1}{10}$ the volume of the spheroplasts.
9. Leave for 10 min at room temperature, then gently resuspend the cells and add 1 ml of PEG solution, mixing thoroughly but gently. Leave for 10 min and then centrifuge at 1000 rpm for 5 min.
10. Gently resuspend the spheroplasts in 150 μl of SOS and incubate at 30° C for 30–40 min.

11. Plate out onto −His SORB plates after adding 5 ml of −His TOP agar (kept at 46° C) and inverting once to mix the cells.

12. Incubate at 30° C for 3–4 days. Colonies embedded in the agar should be visible. Colonies should be replica plated onto medium lacking tryptophan or uracil to identify those that have lost the ability to grow in the absence of uracil but still grow in the absence of tryptophan. $His^+Trp^+Ura^-$ clones are maintained on −Trp−His medium.

We have observed transformation frequencies of 50–100 colonies per microgram of plasmid with these vectors. The frequency of transformants with the appropriate $His^+Trp^+Ura^-$ phenotype ranges from 75 to 90%. This high efficiency is probably partly because of the high conservation of B1 repeats and because of the absence of *HIS3* revertants due to the use of a *his3* deletion mutant strain. The sizes of the YAC fragmentation products should be estimated and their marker content assessed as thoroughly as possible, in order to try and detect any aberrant YACs generated during the course of the fragmentation. For example, our experience has shown that 13% of the fragmentation products derived from a 460-kb YAC are aberrant, based on the order of their internal markers, although only one of these aberrant YACs could have been detected by its size alone (36).

Analysis of Yeast Artificial Chromosome Fragmentation Products

1. Using standard methods, prepare high molecular weight DNA in plugs and analyze by pulsed-field gel electrophoresis and Southern blot hybridization with all the available probes present in the YAC.

2. Any sequence-tagged sites (STSs) present should be analyzed by colony PCR (40).

3. In the absence of numerous markers, a simple fingerprint analysis using mouse repeat sequences as probes can be conducted. Alternatively, restriction mapping of the YAC fragments for comparison with the original YAC could be undertaken (34).

4. These fragmentation vectors provide the possibility of rescuing the sequences adjacent to a given fragmentation site following digestion with *Eco*RV or *Cla*I and recircularization followed by transformation into bacteria as previously described (34). Appropriate end-rescue plasmids can be selected for by transformation into BA1, a *hisB trpC* bacterial strain (41), where the yeast *HIS3* gene, which should be present on the plasmid, can complement the bacterial *hisB* mutation. The isolation of end probes from a selection of YAC fragments is a useful way of generating new markers in a region and can also help in further assessing the integrity of the panel of YAC fragments generated.

Transfer of Yeast Artificial Chromosomes into Mouse Germ Line

In addition to facilitating the chromosome walking and mapping involved in positional cloning, YACs are increasingly being used in a functional approach to locus identification and investigation. In certain cases of diseases with cellular phenotypes, the gene responsible can be identified within a large DNA fragment by transfer of YACs into mutant mammalian cells for the observation of complementation *in vitro.* Yeast artificial chromosomes have been successfully introduced into the mouse genome *in vivo,* thus opening up a wide range of new possibilities involving the mouse as a model system (for review see Ref. 33). The probability of including all of the control elements of a gene, even those located several hundred kilobases from the expressed sequence, is greatly increased when YACs are used as transgenes. The likelihood of rescuing mouse mutant phenotypes is thus increased, as was shown, for example, in the rescue of the albino phenotype in transgenic mice carrying a 250-kb YAC that included the 80-kb tyrosinase coding region and 155 kb of upstream sequences (42). Furthermore, the nature of the long-range elements involved in producing the appropriate developmental and tissue-specific expression of a gene can be identified by using appropriate deletion derivatives of the original YAC as transgenes. The coordinated function of large clusters of genes can also be investigated using YAC transgenes that span the intact cluster and surrounding sequences, an example of this being the correct developmental control of the genes at the human β-globin locus observed in transgenic mice when a 248-kb YAC, spanning the 82-kb locus and 148 kb of flanking sequence, was used as a transgene (43). Finally, the search for long-range *cis*-acting elements such as those controlling X chromosome inactivation (the X-inactivation center) can now be approached in a systematic manner through the use of YAC transgenes.

Methods for Transferring Yeast Artificial Chromosomes into Mouse Germ Line

Two types of approach have been used to create transgenic mice using YACs: direct microinjection of purified YAC DNA into fertilized oocytes; or initial transfer into embryonic stem (ES) cells by spheroplast fusion or lipofection, followed by incorporation of transfected ES cells into blastocysts and production of chimeric mice (for review see Ref. 33). The former approach, used with success, for example, by Peterson *et al.* (43) and Schedl *et al.* (42) and described below, is the most direct. It does, however, necessitate purification of the YAC DNA using conditions that avoid fragmentation of such large DNA molecules and there seems to be an upper limit to the size of YACs (500–600 kb) that can pass through an injection needle without shearing (45).

Although the ES cell-mediated approach is more time consuming, requiring trans-

mission of the YAC through the germ line to produce nonchimeric transgenic mice, it has the advantage that the integrity, copy number, and other features of the integrated YAC can be characterized in the ES clones *in vitro* prior to their reintroduction into mice. Lipofection into ES cells requires large quantities of purified YAC DNA and this technique seems to have a tendency to generate complex rearrangements of the YAC during transfer into cells (45). Spheroplast fusion to ES cells does not require purification of the YAC DNA, thus making it much more straightforward than lipofection, and it has the important advantage that there is no apparent upper size limit for the YAC transferred. Although large amounts of yeast DNA, in addition to the YAC, are integrated into the mouse genome using this technique, this does not appear to affect the pluripotency of the ES cells adversely, nor the functional properties of genes within the YAC, either in ES cells or transgenic mice (46).

Yeast Artificial Chromosome Retrofitting

Although microinjection is efficient enough to avoid the necessity for selection, selectable markers are required for the transfer of a YAC into ES cells. Several markers have been used with success, including the gene for neomycin resistance, the hypoxanthine phosphoribosyltransferase (HPRT) gene, and the herpes virus TK gene (33). Reporter genes can also be used as visual markers to detect the presence of a YAC following injection into cells or spheroplast fusion. The function of particular regulatory sequences within a YAC in transgenic mice can also be studied using visual markers, such as β-galactosidase, in particular tissues or at particular developmental stages.

To retrofit a YAC with a suitable marker gene, homologous recombination can be used to target the marker either into repetitive elements or into a defined sequence within the YAC insert, or into one of the vector arms. The incorporation of a plasmid carrying the bacterial β-galactosidase gene into a YAC by homologous recombination with one of the YAC vector arms is described.

The pLβ-galA plasmid, shown in Fig. 3, was constructed by incorporation of the yeast *LYS2* gene into the pCMVβ plasmid (Clontech). The β-galactosidase gene is under the control of a cytomegalovirus early promoter and enhancer, and contains a simian virus 40 (SV40) poly(A) site. Integration of the plasmid, linearized at a unique *Sca*I site, into the TRP CEN YAC vector arm involves a simple insertion via the ampicillin resistance gene as shown in Fig. 3 and is selected for by acquisition of the *LYS2* gene.

1. Grow the YAC (on a *lys2* mutant background, e.g., AB1380) in AHC medium and prepare spheroplasts as described above (steps 2–7).
2. Transformation (as in step 8 above) involves 1 μg of *Sca*I-digested pLβ-galA plasmid and 50 μg of carrier DNA.

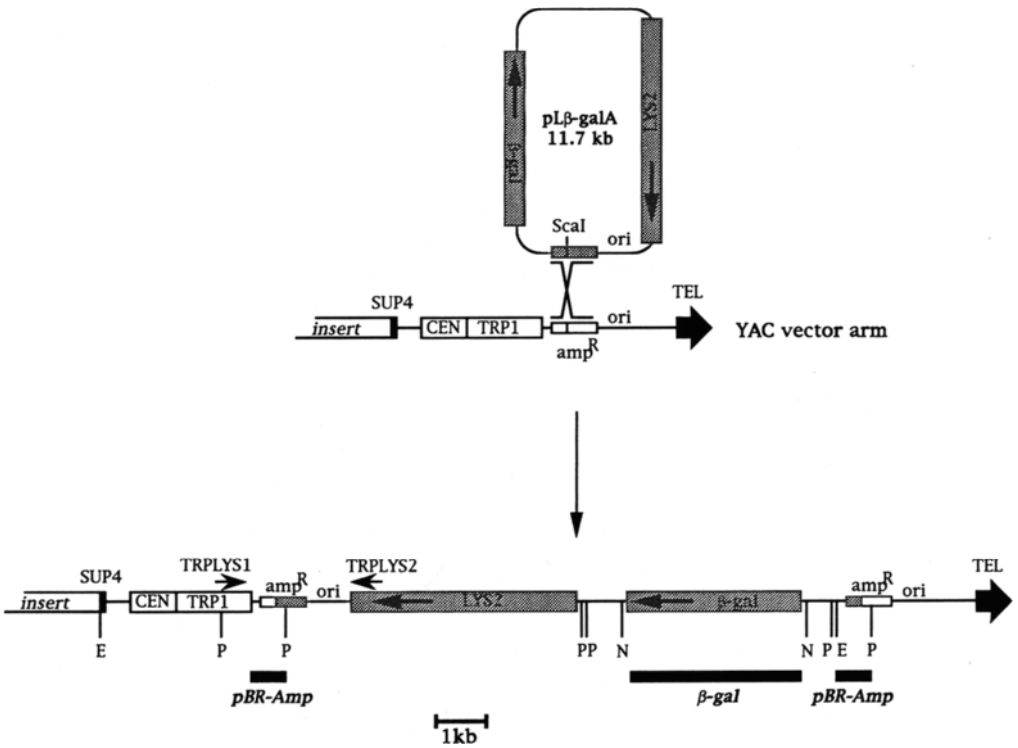


FIG. 3 Modification of the TRP CEN YAC vector arm using pLβ-galA. The YAC modification vector pLβ-galA is digested with *Sca*I, which cuts at a unique site within the bacterial ampicillin resistance gene (Amp), to create recombinogenic ends. Homologous recombination between the linearized ends of pLβ-galA and the Amp gene in the TRP CEN YAC vector arm results in the insertion of the complete pLβ-galA plasmid into the Amp gene and thus in the duplication of the region of homology mediated by the cross-over points. The *LYS2* gene is used to select for successful uptake of the pLβ-galA plasmid. Arrows indicate the transcriptional orientation of *LYS2* and β-Gal. The positions of the primers TRPLYS1 and TRPLYS2 used to detect correctly retrofitted Lys$^+$ transformants are also indicated with arrows. Selected restriction sites that can be used to verify correct insertion of pLβ-galA into the YAC arm are indicated. Probes indicated as solid bars are as follows: β-Gal (4-kb *Not*I fragment derived from pCMVβ; Clontech) and pBR-Amp (0.7-kb *Eco*RI–*Pst*I fragment derived from pBR322). Note that the pBR-Amp probe detects two regions in the correctly retrofitted YAC, owing to the duplication of the region of homology. E, *Eco*RI; N, *Not*I; P, *Pst*I.

3. Plate out onto −Lys SORB plates and incubate at 30°C for 3–4 days. Screen colonies for correct insertion of pLβ-galA by colony PCR (39), using the following primers (see Fig. 3):

TRPLYS1 5′ CTC CCT CTT GGC CCT CTC CTT TTC 3′
TRPLYS2 5′ CGT GGA AGC TCC GCA GCA GCT TAA 3′

This should generate an approximately 2.5-kb product.

4. Analyze a number of retrofitted YAC clones by standard and pulsed-field gel electrophoresis followed by Southern blot hybridization using the probes indicated in Fig. 3 to verify the integrity of the YAC following the insertion.

Preparation of Yeast Artificial Chromosome DNA for Pronuclear Injection of Fertilized Mouse Oocytes

The following technique is based on that developed by Gnirke *et al.* (45) for microinjection into cultured cells, with certain modifications adapting it for use with fertilized oocytes [C. Huxley, personal communication (1995) and Ref. 43]. The upper size limit for microinjection of YAC DNA prepared in the following way appears to be about 500–600 kb. If the preparation is not efficient, or larger YACs are used, shearing occurs and the transgenic mice generated will contain fragments rather than the intact YAC integrated into their genome. The quality of the YAC preparation should therefore be checked by pulsed-field gel electrophoresis prior to its use.

1. Prepare agarose plugs at a density of about 4×10^9 cells/ml, using standard methods.

2. Resolve the YAC on a pulsed field gel using 1% (w/v) low melting agarose (SeaPlaque; FMC). Low melting point agarose gels should be run using the same switching interval as for a normal agarose gel, but the run time should be 1.5 times longer to achieve the same resolution (e.g., 72 hr instead of 48 hr for a 470-kb YAC).

3. Cut out the central portion of the gel and keep in 0.5× TBE. Take the edge strips of the gel and stain in ethidium bromide solution. Visualize the DNA on an ultraviolet (UV) light box and cut a nick in the gel at the position of the YAC. Reassemble the gel and cut out the gel slice (about 3 mm wide) containing the YAC. Cut the gel slice into 1-cm pieces and transfer to a 50-ml tube. Stain the remains of the gel to check that the YAC was cut out.

4. Equilibrate twice with 10 ml of agarase buffer [10 m*M* Bis–Tris (pH 6.5), 0.5 m*M* EDTA, 100 m*M* NaCl] for 2 hr.

5. Place two agarose pieces into a microcentrifuge tube, estimate their weight

(should be approximately 500 mg), and spin the agarose to the bottom of the tube in a microcentrifuge for about 6 sec.

6. Place in a 68° C water bath for 10 min.

7. Place in a 40° C water bath for 5 min.

8. Add 1 U of β-agarase I [New England Biolabs (Beverly, MA) at 1000 U/ml] per 100 mg of agarose. Mix by pipetting once very slowly with a cut-off tip.

9. Incubate at 40° C for 2 hr.

10. Centrifuge at 13,000 rpm in a microcentrifuge at room temperature for 30 min and carefully transfer the supernatant to a new tube using a cut-off tip. Store the DNA at 4° C.

11. The quality of the DNA should be checked by pulsed-field gel electrophoresis and the concentration estimated on a 0.8% (w/v) agarose minigel using lambda (λ) phage DNA as a standard.

12. Just prior to injection into fertilized oocytes, the purified YAC DNA is diluted twofold with water to reduce the concentrations of EDTA and NaCl.

13. Load at the back of a microinjection needle that has an inner capillary, so that the DNA solution runs to the tip of the needle by capillary action.

14. Microinjection into the pronuclei of fertilized mouse oocytes should be performed according to standard procedures for the generation of transgenic mice (44).

Using this technology with a 470-kb YAC, we have prepared intact YAC DNA at concentrations of 2–4 ng/μl, which was then injected at concentrations of 1–2 ng/μl to give an efficiency of transgenesis of 10% (i.e., 9 animals carrying some or all of the YAC, of 90 pups born). Although one animal appears to carry the entire YAC intact and to transmit it, the majority of the transgenic animals contained only part of the YAC, suggesting that substantial fragmentation must have occurred during the injection process. This is probably not surprising for a 470-kb YAC, if the cutoff size for successful injection of intact DNA is 500–600 kb as suggested by Gnirke *et al.* (45).

The assessment of the presence of YAC-derived sequences in potentially transgenic mice or in ES cell clones may be approached in a number of ways as described by Huxley (33). The transfer of YACs containing a mouse insert into mouse cells poses a particular problem as the YAC DNA cannot easily be distinguished from the host DNA by hybridization and the only probes unique to the YAC are the vector arms or other foreign sequences such as marker genes. Such probes can be used to provide an idea of the copy number of the YAC present by comparing hybridization intensities and by counting the number of end fragments. The most efficient way to assess rapidly for the presence of different parts of a mouse-derived YAC in mouse cells is to ensure that the YAC insert and the recipient cells/animals are derived from different strains (e.g., see Refs. 42 and 46) such that restriction fragment length differences and, more likely, microsatellite polymorphisms will exist between the two strains in the region covered by the YAC.

Nevertheless, the presence of all YAC-derived markers in a particular cell line or transgenic animal is no guarantee that the YAC is intact and not fragmented or rearranged. The RecA-assisted restriction endonuclease (RARE) cleavage technique (48) provides one way of demonstrating that the YAC insert is intact in the mouse genome (e.g., see Ref. 45). A more straightforward solution to this problem would be to retrofit the YAC vector arms, prior to introduction of the YAC into mammalian cells, with restriction sites for enzymes that cut very rarely in the mammalian genome, such as the meganuclease I–*Sce*I (see below). Vectors of this kind, which also incorporate a selectable marker gene, have recently been created (49).

Subcloning of Yeast Artificial Chromosomes and Contig Assembly

Although YACs have clearly made a major contribution to the characterization of complex genomes in many situations, including the preparation of DNA for sequencing, contigs constructed from libraries cloned in smaller capacity cloning vectors such as cosmids and λ will be necessary. The YAC inserts can easily be subcloned into such phage or cosmid vectors, but the assembly of these into contigs of overlapping clones can be time consuming, especially when DNA extraction of individual clones is required.

The following approach combines an IRS-PCR-based technique for establishing clone overlap and a meganuclease I–*Sce*I-based YAC fragmentation technique for the positioning and orientation of the primary contigs, thus allowing rapid establishment of contig arrays corresponding to entire YACs.

Random Overlap Strategy for Organization of λ Subclones into Primary Contigs

This technique is based on the amplification of random fragments within cloned inserts by direct IRS-PCR (interspersed repeat sequence-PCR) of λ lysis supernatants. The PCR-amplified fragments are then used to probe the clone matrices and identify overlapping clones by hybridization. First applied to human DNA by Pieretti *et al.* (50), this technique has been modified for use in the mouse (51). The technique can be easily modified for use with cosmid clones.

In the initial phase of contig establishment, phage clones are chosen randomly from the YAC-derived library. They are subsequently PCR amplified using a mix of three primer pairs corresponding to the 3′ and 5′ ends of the B2 sequence (primers B2A1 and B2A3, respectively), the 3′ end of the L1 sequence (primers R1 and R2), and the T3 and T7 promoter sequences flanking the cloning site of the vector. The former are known to allow efficient IRS-PCR amplification of mouse DNA in IFGT hybrids (see above and Ref. 28), whereas the latter allow amplification of IRS–vector products. Simultaneous use of more than a single primer pair significantly increases the number of clones yielding useful amplification products.

PCR Reactions and Labeling of IRS-PCR Products

Phage plaques are touched with a toothpick, which is then soaked in 50 μl of Tris–EDTA (TE), 2 μl of which is then used for PCR amplification.

All PCRs are carried out in a total volume of 50 μl under the following conditions:

KCl, 50 m*M*
Tris-HCl (pH 8.0), 10 m*M*
$MgCl_2$, 1.5 m*M*
Gelatin, 0.01% (w/v)
dNTPs (125 μM each)
Primer (1 μM each)
Taq polymerase (Perkin-Elmer/Cetus), 2.5 units

Oligonucleotide primer sequences
B2A1: TCTTCTGGAGTGTCTGAAGA
B2A3: CAGGTGCTCTTAACCACTG
R1: GCTGATTATGGGATGGATCC
R2: AGCATTTGAAATGTAAATGA
T3: ATTAACCCTCACTAAAGGGA
T7: TAATACGACTCACTATAGGG

Cycle Conditions

An initial denaturation step of 4 min at 94° C is followed by 30 cycles consisting of 94° C denaturation for 1 min, 51° C annealing for 1 min, and 72° C extension for 4 min. A final extension step consists of 8 min at 72° C. The PCR reactions are optimized for a Techne PH-3 DNA thermal cycler.

Following amplification, the reaction products are loaded onto 1% (w/v) low melting point agarose gels. Bands corresponding to the PCR products are excised, and the equivalent of 50 ng of DNA labeled by random priming. The probe is prehybridized with total mouse DNA in order to block mouse repetitive sequence, then used to probe a filter matrix representative of the entire cloned YAC sublibrary.

Prehybridization and Hybridization

Labeled DNA	100 μl (20–50 ng)
Total mouse sonicated DNA	20 μl (100–300 μg)
TE (pH 8.0)	56 μl

1. Heat at 100° C for 10 min.
2. Add immediately 24 μl of 1 *M* sodium phosphate (pH 6.8).
3. Incubate for 2–3 hr at 65° C.

Hybridizations are carried out with 0.5 *M* sodium phosphate (pH 7.2)–1 m*M* EDTA–7% (w/v) SDS at 65°C, modified from Church and Gilbert (52). Filters are washed in 40 m*M* sodium phosphate (pH 7.2)–1% (w/v) SDS at 65°C.

This random approach allows the establishment of primary contigs whose position and orientation relative to each other and to the YAC remain, however, unknown.

Walking Strategy

Once the majority of the YAC insert is covered by primary contigs, closure is achieved by the isolation of end-specific probes generated by IRS–vector PCR from clones located at the extremities of the existing primary contigs. The efficiency of this approach depends on both the representation and orientation of repeats in proximity to the vector arms, which serve as the second primer. In our hands the B2A3 primer has been shown to give the highest percentage of amplification products in such IRS–vector PCR reactions. For each amplification reaction, a control PCR using the B2A3 primer alone is performed in order to identify and exclude fragments due to internal IRS amplification between neighboring B2 sequences. In case of difficulties in obtaining such IRS–vector products, the PCR reactions should be repeated using purified DNA rather than lysis products, because higher end-specific amplification efficiency is then observed.

Localization and Orientation of Primary Contigs Using Endonuclease I–SceI-Based Fragmentation

Fragmentation of endogenous yeast chromosomes by the introduction of artificial restriction sites for the endonuclease I–*Sce*I has been described previously (53). We took advantage of the presence in the mouse genome of widely distributed repeat sequences and of the efficiency of yeast homologous recombination to randomly introduce endonuclease I–*Sce*I sites within YACs carrying inserts of murine origin (54). Among the different classes of repeats of the mouse genome that can be used, the SINEs (short interspersed nuclear elements) and LINEs (long interspersed nuclear elements) are known to be of high copy number and well dispersed throughout the genome (approximately once every 30 kb).

To achieve random integration of endonuclease I–*Sce*I sites into the YAC insert, an integrative vector is used that contains the endonuclease I–*Sce*I site, the mouse B2 sequence (which belongs to the SINE family), and the *LYS2* selectable marker (Fig. 4). To increase the efficiency of homologous recombination within the B2 sequences, DNA from the integrative vector is first linearized *in vitro* at a restriction

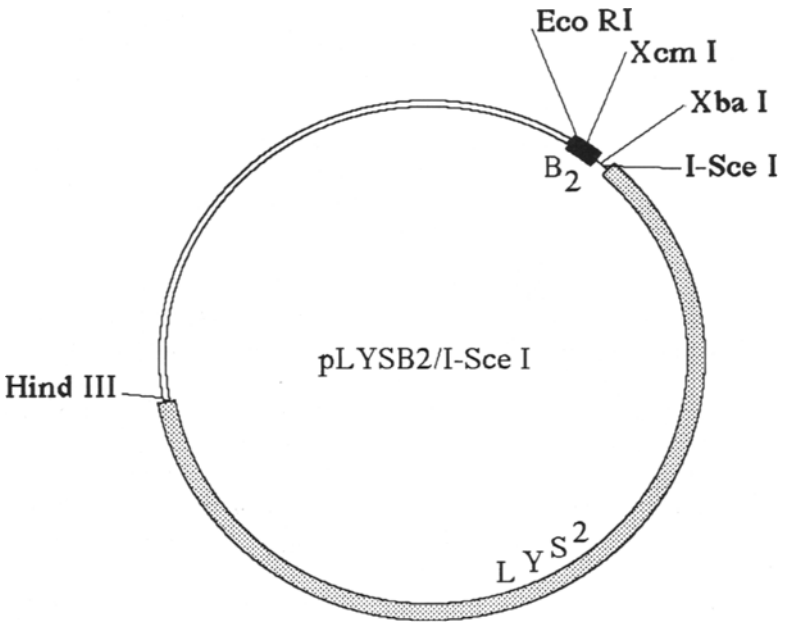


FIG. 4 Random integration vector. The pLYSB2/I-*Sce*I plasmid is a pBluescript SK$^+$ derivative that contains a B2 repeat element (black box) as well as the *LYS2* marker. The plasmid can be digested with *Xcm*I to create recombinogenic ends.

site lying within the B2 sequence (*Xcm*I) prior to transformation by electroporation of the yeast strain carrying the mouse YAC.

Endonuclease I–SceI Site Integration

1. Yeast cells are grown in 15 ml of YPD (36) at 30° C with aeration until the culture reaches ca. 10^7 cells/ml or an OD of 0.8–1.0.
2. The cells are then centrifuged for 5 min at 3000 *g*, and resuspended in 0.025 *M* dithiothreitol and incubated for 10 min at 30° C.
3. The cells are centrifuged for 5 min at 3000 *g* and cells resuspended at a concentration of 10^9 cells/ml in a solution of 0.27 *M* sucrose–0.01 *M* Tris-HCl (pH 7.5)–0.001 *M* $MgCl_2$.
4. Cell suspension (100 μl) is pipetted into 0.2-cm electroporation cuvettes (Bio-Rad, Richmond, CA).
5. Two micrograms of *in vitro*-linearized vector is added and the solution electroporated (2250 V/cm, 250 μF, 200 Ω, 30–35 msec).
6. Immediately after electroporation, the cells are plated out on SC-LYS medium (38).

Transformants with a Lys^+ phenotype are isolated by streaking on selective medium for both uracil and tryptophan to eliminate cells that may have lost the YAC during transformation.

Embedded DNA of $Ura^+Trp^+Lys^+$ transformants is prepared and digested *in vitro* with endonuclease I–*Sce*I using the following protocol.

1. Plugs are equilibrated overnight in an excess volume of 0.1 *M* diethanolamine hydrochloride (pH 9.5) at 4° C.
2. Rinse in 1× incubation buffer for 1 hr at 4° C.
3. One plug is placed in a microfuge tube containing 150 μl of 1× incubation buffer at 4° C. Twenty units of endonuclease I–*Sce*I is added together with 20 ng of its "enhancer" and incubated for 1.5 hr at 4° C to allow for adequate diffusion of the endonuclease I–*Sce*I enzyme.
4. The reaction is initiated by adding 1.6 μl of 1 *M* $MgCl_2$ solution (0.008 *M* final concentration), then incubated for 30 min to 1 hr at 37° C (do not exceed 1 hr, as nonspecific degradation may occur).
5. The plugs are cooled to 0° C and the reaction stopped by adding 500 μl of 0.5 *M* EDTA (pH 9).
6. Combine the following ingredients to make 1× incubation buffer:

   Diethanolamine hydrochloride (pH 9.5), 0.1 *M*
   Dithiothreitol (0.001 *M*)
   Bovine serum albumin (0.2 mg/ml)

Endonuclease I–*Sce*I-digested transformants are analyzed by pulsed-field gel electrophoresis (PFGE). Agarose plugs containing 1 μg of yeast DNA are introduced into wells and immobilized with molten agarose. Agarose gels (1%, w/v) (SeaKem GTG; FMC) made up in 0.25× TBE buffer are run in the same buffer at 13 V and 12° C on a Rotaphor apparatus (Biometra) for 70 hr, using 30- to 100-sec increasing pulse times.

It has been shown that YACs having undergone such homologous recombination at a B2 site within the YAC normally carry a unique endonuclease I–*Sce*I site, suggesting that multiple recombination events occur only rarely. Digestion of such recombinant YACs with endonuclease I–*Sce*I generates two fragments whose size depends on the localization of the plasmid insertion and can be determined by successive hybridization with YAC arm-specific probes (*URA3* and *TRP1*). The localization of each integration event allows the construction of an endonuclease I–*Sce*I insertion map. The position and left/right orientation of the primary contigs with respect to the YAC arms can be deduced from hybridization experiments with PCR probes corresponding to the most terminal clones in each contig. In all cases, only one arm of each digested YAC gives a positive signal, allowing easy positioning of the probe and the corresponding contig. The labeling and hybridization conditions are the same as those described in the previous section (Fig. 4).

To map phage clones efficiently using this approach, the distance between each endonuclease I–*Sce*I site must be around twice the average size of the phage inserts. If such a resolution cannot be obtained using a single type of repeat, targeting by recombination to other repetitive elements can be undertaken. A second integrative vector with 3′ sequence of the LINE repetitive elements can, for example, be used to increase the number of insertions.

If analysis of the B2 sequence targeted is required, plasmid rescue can be performed. Total yeast DNA of the transformant is extracted, digested with *Hin*dIII (which cleaves within the integrative vector and at unknown positions in the mouse DNA), ligated, and the ligation mixture used to transform *E. coli.* DNA is prepared from *E. coli* transformants and sequenced. Only integration into B2 repeat sequences is observed. Because no sequence identical to the cloned B2 element present in the integrative vector was recovered, this suggests that homologous recombination between divergent repeat sequences is occurring (the divergence rates are from 2 to 6%).

Concluding Remarks

Rapid progress toward completing and verifying physical maps of the human and the mouse genomes is currently being made. Commensurate progress toward the establishment of transcription maps should lead to archived collections of YACs, and eventually of smaller capacity vectors such as P1 and cosmids, that are of known gene content and even of known arrangement. The availability of this material should lead, through a candidate gene approach, to the rapid identification of many of the thousand or more mouse mutations, both spontaneous and induced, that have been isolated over about the last 70 years of mouse genetics. The emphasis being put onto model systems for analysis of human disease and development is likely to be increasingly oriented toward those such as the mouse, which permit extensive experimental manipulation of the mammalian genome. The development of systems able to generate targeted deletions of regions extending over several hundred kilobases are likely to go hand in hand with the rapid progress in systems for addition of large pieces of genetic material to the genome. Both targeted YAC/P1 transgenesis as well as the development of mammalian artificial chromosomes should turn out to be important tools in this type of analysis. The methods outlined in this chapter represent part of the increasing repertoire of techniques available for mouse genetics and embryology that are aimed at efficiently characterizing regions of the mouse genome. The application of such approaches should allow improved understanding not only of traits under monogenic or simple genetic control but equally those under polygenic control or those showing multifactorial inheritance, where the influence of allelic variation at multiple interacting loci can as yet only vaguely be perceived.

Acknowledgments

We thank Clare Huxley for encouragement and advice concerning YAC transfer into mice, Rodney Rothstein and colleagues at the Institut Pasteur (Bernard Dujon, Laurence Colleaux, Charles Babinet, and Chantal Kress), as well as members of the Unité de Génétique Moléculaire Murine who have contributed in a variety of ways to the techniques and experimental approaches outlined in this chapter. Our work is supported by grants from le Ministère de la Recherche et de l'Espace (MRE), le Groupement de Recherches et d'Etudes sur les Génomes (GREG), l'Association Française pour la Recherche contre les Myopathies (AFM), l'Association pour la Recherche contre le Cancer (ARC), and le Centre National de la Recherche Scientifique (CNRS). C.R. is supported by grants from the MRE and the AFM.

References

1. F. Bonhomme and J.-L. Guénet, *in* "Genetic Variants and Strains of the Laboratory Mouse" (M. F. Lyon and A. G. Searle, eds.), p. 649. Oxford University Press, London, 1989.
2. J.-L. Guénet, D. Simon-Chazottes, and P. R. Avner, *Curr. Topics Microbiol. Immunol.* **137,** 13 (1988).
3. F. Bonhomme, J. Catalan, J. Britton-Davidian, V. M. Chapman, K. Moriwaki, E. Nevo, and L. Thaler, *Biochem. Genet.* **22,** 275 (1984).
4. P. Avner, L. Amar, L. Dandolo, and J.-L. Guénet, *Trends Genet.* **4,** 18 (1988).
5. N. G. Copeland and N. A. Jenkins, *Trends Genet.* **7,** 113 (1991).
6. The European Backcross Collaborative Group, *Hum. Mol. Genet.* **3,** 621 (1994).
7. M. F. W. Festing, *in* "Inbred Strains in Biomedical Research," p. 104. Macmillan Press, London, 1979.
8. C. M. Hearne, M. A. McAleer, J. M. Love, T. J. Aitman, R. J. Cornall, S. Ghosh, A. M. Knight, J.-B. Prins, and J. Todd, *Mammalian Genome* **1,** 273 (1991).
9. B. de Gouyon, E. Melanitou, M. F. Richard, M. Requarth, I. H. Hahn, J. L. Guénet, F. Demesnais, C. Julier, G. M. Lathrop, C. Boitard, and P. Avner, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 1977 (1993).
10. J. Welsh, C. Peterson, and M. McClelland, *Nucleic Acids Res.* **19,** 303 (1991).
11. A. Jeffreys, V. Wilson, R. Kelly, B. Taylor, and G. Bulfield, *Nucleic Acids Res.* **15,** 2832 (1987).
12. B. de Gouyon, C. Julier, P. Avner, M. Georges, and M. Lathrop, *in* "DNA Fingerprinting: Approaches and Applications" (T. Burke, G. Dolf, A. J. Jeffreys, and R. Wolf, eds.), p. 85. Birkhauser Verlag, Basel, Switzerland, 1991.
13. M. Georges, A. S. Lequarré, M. Castelli, R. Hanset, and G. Vassart, *Cytogenet. Cell Genet.* **47,** 127 (1988).
14. C. Julier, B. de Gouyon, M. Georges, J. L. Guénet, Y. Nakamura, P. Avner, and M. G. Lathrop, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 4585 (1990).
15. J. Sambrook, E. F. Fritsch, and T. Maniatis, *in* "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.

16. D. Tautz, *Nucleic Acids Res.* **17,** 6463 (1989).
17. J. M. Love, M. Knight, A. McAleer, and J. A. Todd, *Nucleic Acids Res.* **18,** 4123 (1990).
18. W. F. Dietrich, J. C. Miller, R. G. Sreen, M. Merchant, D. Damron, R. Nahf, A. Gross, D. C. Joyce, M. Wessel, R. D. Dredge, A. Marquis, L. D. Stein, N. Goodman, D. C. Page, and E. S. Lander, *Nature Genet.* **7,** 220 (1994).
19. A. G. Searle and C. V. Beechey, *in* "Genetics Variants and Strains of the Laboratory Mouse" (M. F. Lyon and A. G. Searle, eds.), p. 620. Oxford University Press, London, (1989).
20. S. J. Goss and H. Harris, *Nature (London)* **255,** 680 (1975).
21. D. R. Cox, C. A. Pritchard, E. Uglum, D. Casher, J. Kobori, and R. M. Myers, *Genomics* **4,** 397 (1989).
22. F. Benham, K. Hart, J. Crolla, M. Bobrow, M. Francavilla, and P. N. Goodfellow, *Genomics* **4,** 509 (1989).
23. L. Sefton, D. Arnaud, P. N. Goodfellow, M. C. Simmler, and P. Avner, *Mammalian Genome* **2,** 21 (1992).
24. D. L. Nelson, S. A. Ledbetter, L. Corbo, M. F. Victoria, R. Ramirez-Solis, T. D. Webster, D. H. Ledbetter, and C. T. Caskey, *Proc. Natl. Acad. Sci. U.S.A.* **86,** 6686 (1989).
25. S. A. Ledbetter, D. L. Nelson, S. T. Warren, and D. H. Ledbetter, *Genomics* **6,** 475 (1990).
26. P. L. Deininger and G. R. Daniels, *Trends Genet.* **2,** 76 (1986).
27. N. D. Hastie, *in* "Genetics Variants and Strains of the Laboratory Mouse" (M. F. Lyon and A. G. Searle, eds.), p. 559. Oxford University Press, London, (1989).
28. M. C. Simmler, R. D. Cox, and P. Avner, *Genomics* **10,** 770 (1991).
29. N. G. Irving and S. D. M. Brown, *Genomics* **11,** 679 (1991).
30. G. E. Herman, M. Berry, E. Munro, I. W. Graig, and E. R. Levy, *Genomics* **10,** 961 (1991).
31. A. P. Feinberg and B. Vogelstein, *Anal. Biochem.* **137,** 266 (1984).
32. R. Rothstein, *Methods Enzymol.* **194,** 281 (1991).
33. C. Huxley, *in* "Genetic Engineering" (J. K. Setlow, ed.), Vol. 16, p. 65. Plenum Press, New York, (1994).
34. W. J. Pavan, P. Hieter, and R. H. Reeves, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 1300 (1990).
35. R. Das Gupta, B. Morrow, I. Marondel, S. Parimoo, V. L. Goei, J. Gruen, S. Weissman, A. Skoultchi, and R. Kucherlapati, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 4364 (1993).
36. E. Heard, P. Avner, and R. Rothstein, *Nucleic Acids Res.* **22,** 1830 (1994).
36a. W. J. Pavan, P. Hieter, D. Sears, Al Burkhoff, and R. H. Reeves, *Gene* **106,** 125 (1991).
37. M. Rose, F. Winston, and P. Hieter, *in* "Methods in Yeast Genetics: A Course Manual." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, (1990).
38. R. H. Reeves, W. J. Pavan, and P. Hieter, *Methods Enzymol.* **216,** 584 (1992).
39. C. Connelly, M. McCormick, J. Shero, and P. Hieter, *Genomics* **10,** 10 (1991).
40. E. Heard, B. Davies, S. Feo, and M. Fried, *Nucleic Acids Res.* **17,** 5861.
41. A. W. Murray, N. P. Schultes, and J. W. Szostak, *Cell* **45,** 529 (1986).
42. A. Schedl, L. Montoliu, G. Kelsey, and G. Schutz, *Nature (London)* **362,** 258 (1993).
43. K. R. Peterson, C. H. Clegg, C. Huxley, B. M. Josephsen, H. S. Haugen, T. Furukawa, and G. Stamatoyannopoulos, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 7593 (1993).
44. B. Hogan, F. Constantini, and E. Lacy, "Manipulating the Mouse Embryo." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1986.
45. A. Gnirke, C. Huxley, K. Peterson, and M. V. Olson, *Genomics* **15,** 659 (1993).

46. A. Jakobovits, A. L. Moore, L. L. Green, G. J. Vergara, C. E. Maynard-Currie, H. A. Austin, and S. Klapholz, *Nature (London)* **362,** 255 (1993).
47. W. M. Strauss, J. Dausman, C. Beard, C. Johnson, J. B. Lawrence, and R. Jaenisch, *Science* **259,** 1904 (1993).
48. L. J. Ferrin and R. D. Camerini-Otero, *Science* **254,** 1494.
49. C. Fairhead, E. Heard, D. Arnaud, P. Avner, and B. Dujon, *Nucleic Acids Res.,* in press.
50. M. Pieretti, R. Tonlorenzi, and A. Ballabio, *Nucleic Acids Res.* **19,** 2795 (1991).
51. C. Rougeulle, L. Colleaux, B. Dujon, and P. Avner, *Mammalian Genome* **5,** 416 (1994).
52. G. M. Church and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **81,** 1991 (1984).
53. A. Thierry and B. Dujon, *Nucleic Acids Res.* **21,** 5625 (1992).
54. L. Colleaux, C. Rougeulle, P. Avner, and B. Dujon, *Hum. Mol. Genet.* **2,** 265 (1993).

[26] Molecular Analysis of Cardiac Muscle Diseases Based on Mouse Genetics

Steven W. Kubalak, Pieter A. Doevendans, Howard A. Rockman, John J. Hunter, Nobuaki Tanaka, John Ross, Jr., and Kenneth R. Chien

Introduction

Significant advances have been made in understanding important problems in cardiovascular biology and medicine through the use of transgenic and gene-targeting techniques (1–3). The establishment of genetically based mouse models of atherogenesis, hypertension, cardiac hypertrophy/failure, and cardiovascular developmental defects has demonstrated the possibility of dissecting complex cardiovascular phenotypes through the utilization of mouse genetics (4). In the past few years, there has been an exponential growth in the application of transgenic and gene-targeted approaches to study the cardiovascular system, as well as in coupling traditional cardiovascular physiological and biochemical analyses to the mouse. Concomitantly, miniaturized technology has been developed to assess physiological cardiovascular phenotypes, which has given birth to a new field: molecular physiology of the cardiovascular system (4). In addition to monitoring the electrocardiogram (ECG) and vascular and left ventricular pressures, newly developed techniques for contrast microangiography allow the quantitative assessment of ventricular volumes and ejection fractions in adult and embryogenic mouse heart. The culture of cardiomyocytes from individual hearts of mouse embryos and the use of embryonic stem cell differentiation into embryoid bodies as an *in vitro* model of cardiogenesis provide the opportunity to address gene regulation during cardiac growth and development. Accordingly, this chapter outlines several of the experimental approaches that can be utilized in the analysis of cardiac muscle diseases in transgenic mice.

Methods in Molecular Physiology

Mouse Preparation

For microsurgery, the adult mouse ($\geq$8 weeks) weighing $\pm$20 g is anesthetized using a mixture of ketamine (100 mg/kg), xylazine (5 μg/kg), and morphine (2.5 μg/kg) intraperitoneally. Morphine may be eliminated for minor surgery or measurement of hemodynamic parameters. The animal is placed supine and a blunt 20-gauge needle

Methods in Molecular Genetics, Volume 8

is placed in the trachea to serve as a tracheal cannula. The cannula is connected to a volume-cycled rodent respirator (model 683; Harvard Apparatus, South Natick, MA) on room air with a tidal volume of 0.2 ml and a respiratory rate of 110 breaths/min. Adequacy of ventilation can be determined from inspection of chest expansion. The skin is cleaned with chlorhexidine solution and the surgery is performed under clean but not sterile conditions.

Mouse Models for Cardiac Muscle Disease

Left Ventricular Pressure Overload Induced by Transverse Aortic Constriction

The chest cavity is approached through the second intercostal space at the left sternal border through a small incision, allowing deflection of the thymus, to expose the aortic arch. The transverse aorta is isolated between the carotid arteries and surrounded by a 7-0 nylon suture (Fig. 1A). The vessel is constricted against a 27-gauge needle placed within the suture. After prompt removal of the needle, a constriction with a diameter of 0.4 mm can be achieved, which is 65–70% reduction (normal diameter 1.2 mm, for a 20-g mouse). Postsurgery, the chest is closed, the pneumothorax evacuated by a PE 50 catheter attached to a 1-ml syringe (5–7).

Cannulation of Carotid Arteries

A midline incision in the cervical region is made to expose both carotid arteries. The right and left carotid artery can be isolated and cannulated with flame-stretched PE 50 tubing. The catheters are connected to modified P50 Statham transducers for measurement of simultaneous right and left carotid arterial pressures (Fig. 1A). This technique can be used to assess acute effects of aortic constriction in short studies (2 hr); alternatively this procedure can be applied to assess the pressure gradient achieved by a previous transverse aortic constriction (TAC) procedure days earlier before the experiment is terminated. Following TAC, a predictable systolic gradient across the aortic stenosis results with a range of 20 to 180 mmHg (7).

Molecular Analysis

Cardiac hypertrophy is characterized by an increase in cell size and a selective up- and downregulation of the expression of different genes. A hallmark of the hypertrophic response appears to be the reexpression of fetal protein isoforms. *In vitro* studies have identified the reexpression of protooncogenes, immediate early genes, and atrial natriuretic factor (ANF) in ventricular myocytes following hypertrophic stimuli.

With the development of a reproducible murine model of cardiac hypertrophy, molecular studies of the *in vivo* hypertrophy response have become feasible. mRNA levels were studied by isolation of total RNA from flash-frozen ventricular cardiac

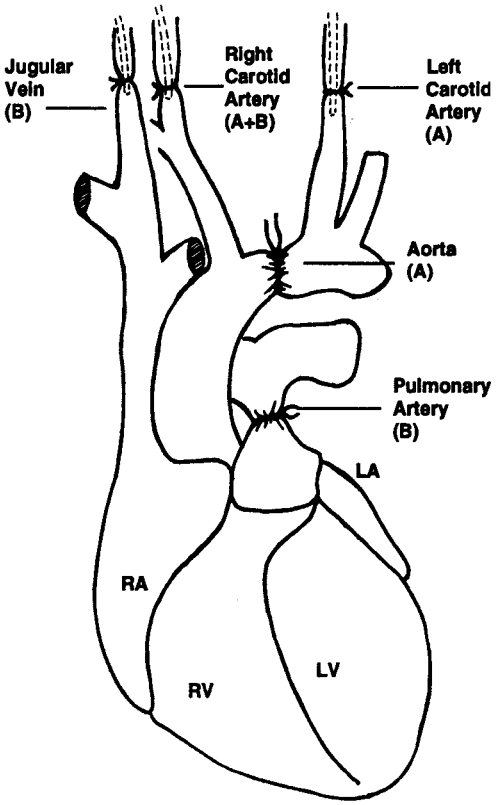


FIG. 1 Schematic of surgical procedures. (A) Transverse aortic constriction with cannulation of the carotid arteries. (B) Pulmonary artery banding with cannulation of the right carotid artery and jugular vein.

tissue, transfer to a nylon filter, and hybridization with probes specific for the immediate early genes (c-*fos,* c-*jun, junB,* and *Egr-1*), ANF, isoforms of sarcomeric proteins (myosin heavy chain, actin, troponin, and tropomyosin), and calcium-handling proteins [sarco(endo)plasmic reticulum Ca^{2+}-ATPase and phospholamban]. A rapid and transient induction of the early gene program has been found acutely and a marked increase in the message level for ANF occurs over days (6). This model of murine cardiac hypertrophy was then applied to investigate the important role of angiotensin II in the hypertrophic process. The normal increase of the heart-to-body weight ratio in banded mice was attenuated by treatment with a specific antagonist (losartan) for the angiotensin II subtype 1 receptor and by inhibition of the angiotensin-converting enzyme (captopril) (6). Application of this technique in

transgenic mice can potentially be used to unravel the myriad of signaling pathways involved in cardiac hypertrophy.

Right Ventricular Hypertrophy and Volume Overload Induced by Pulmonary Artery Banding

For this procedure the chest is opened through the right third intercostal space. The pericardium is opened and the pulmonary artery is identified and isolated. A 7-0 nylon suture ligature is placed around the artery and tied against a 25-gauge needle for a moderate stenosis or a 26-gauge needle for a severe stenosis (Fig. 1B). The needle is removed promptly after the ligature has been made (8).

Imaging and Data Processing

The right jugular vein is identified and cannulated in an anesthetized mouse. Angiographic images are acquired on a $\frac{3}{4}$-inch videotape under constant fluoroscopy after injection of 120 μl of nonionic contrast over 1–2 sec in 30° right anterior oblique and the 60° left anterior oblique projections. X-Ray images are digitized using a Gould De Anza video processing system interfaced to a VAX 11/750 computer system. First-pass right ventricle (RV) and levophase left ventricle (LV) video-density curves are generated by using field-by-field subtraction of RV and LV regions of interest from the corresponding background region of interest: images are interlaced to provide a temporal resolution equivalent to 60 frames/sec. Ejection fraction values can be calculated from the maximal (end-diastole) and minimal (end-systole) value of each beat. The left ventricular and biventricular end-diastolic volumes are obtained from biplane images using the area length method. The right ventricular volume is obtained by subtraction of the left ventricular end diastolic volume and the calculated septal volume from the biventricular volume. In mice with a moderate stenosis, significant hypertrophy was seen of the RV without a fall in RV function; however, in mice with severe pulmonary artery constriction, there was RV enlargement and depressed function (8).

Molecular Analysis

The cardiac tissue is frozen in liquid nitrogen after the weight of the cardiac chambers is measured. RNA and protein can be isolated from these tissues and used for Northern and Western blot analysis. To date the gene program involved in the transition from a hypertrophic response to muscle dysfunction is unknown. It has been shown that impaired Ca^{2+} handling coincides with end-stage pump failure in the human heart (9). In our study we focused on proteins involved in calcium handling by the sarcoplasmatic reticulum: phospholamban and sarcoplasmatic reticulum Ca^{2+}-ATPase. The most prominent finding was the downregulation of phospholamban both at the RNA and protein level, with a more pronounced downregulation in the failing state. Atrial natriuretic factor was clearly upregulated in the hypertrophic

and failing state. This mouse model, allowing cardiac function assessment *in vivo,* combined with mouse genetics may provide a method to determine the sequence and characteristics of the molecular events that are associated with the transition from compensated to decompensated cardiac pump function.

Physiological Measurement in Adult Mice

Cardiac Output Measurement in Conscious Mice

The right carotid artery and femoral artery are exposed and cannulated with flame-stretched and tapered PE 50 tubing. The outer diameter at the tip is ± 367 μm. The PE 50 is connected to PE 10, filled with heparinized saline (10 IU/ml), and exteriorized at the back of the neck. After the animals are recovered from surgery they are trained to stay in a restraining tube during several days. With this surgical approach, radioactive microspheres can be used to assess cardiac output and regional blood flow (10).

Data analysis in the mice studied revealed a cardiac output in the mouse of 16 ± 1.4 ml/min. The calculated cardiac index was 591 ml/min/kg (human values: 4–5 liter/min, 55–70 ml/min/kg). The blood volume measured 2.3 ± 0.009 ml, which is 80 ml/kg. The changes in hemodynamic parameters in the anesthetized animal were a drop in heart rate (653 to 459/min), reduction of the mean arterial pressure (92 to 77 mmHg), and increased blood flow to the kidney and splanchnic region (10).

Systolic and Diastolic Left Ventricular Function by Micromanometers

Following anesthesia, jugular vein and a carotid artery are cannulated and the chest and pericardium are opened. Through a puncture in the left atrium a 2F high-fidelity micromanometer catheter (Millar Instruments, Houston, Texas) can be positioned across the mitral value into the LV (11, 12). Aortic pressure, left ventricular systolic and diastolic pressure, and the derivative of the left ventricular pressure (LV dP/dt) can be recorded continuously during basal conditions and following infusion of various vasoactive agents. The time constant of left ventricular relaxation (Tau) can also be accessed.

Myocardial function was measured in transgenic mice constitutively overexpressing the human β_2-adrenergic receptor driven by α-myosin heavy chain promoter. These studies revealed that there was an increased heart rate and maximum values of dP/dt in the basal state that was not responsive to adrenergic stimulation. In nontransgenics the maximal dP/dt was reached with β-adrenergic stimulation with isoproterenol. The downregulation of the β-adrenergic system has been shown to occur in human pump failure. Induction of β-adrenergic receptor formation in the failing heart through gene therapy may be a nonpharmacologic means to improve cardiac function (11).

Passive Ventricular Mechanics

After opening of the chest the ascending aorta is clamped. The heart is arrested by slowly injecting up to 0.3 ml of a hypothermic, hyperkalemic, cardioplegic solution through the apex into the left ventricle. The arresting solution contains (in grams per liter) 4.0 NaCl, 1.0 $NaHCO_3$, 2.0 glucose, 3.0 2,3-butanedione monoxime, 3.8 ethylene glycol bis(β-aminoethyl ether)-*N,N,N',N'*-tetraacetic acid, 0.0002 nifedipine, and heparin (10,000 U/liter). The arrested heart is excised and rinsed in saline. A left ventricular balloon is made by stretching a polypropylene membrane into the shape of the left ventricular cavity. The balloon is inserted in the left ventricular cavity and secured with a purse string suture around the mitral valve. The unloaded ventricular volume is 13 μl (6 μl for balloon and cannula, 7 μl of water) to achieve a pressure reading of 0 mmHg in a 20-g control mouse. The cannula is connected to a Statham P23 Db pressure transducer and a Harvard infusion pump with a 100-μl syringe. To obtain pressure volume data the balloon is inflated at a constant rate (25–35 μl/min) until the pressure is at least 20 mmHg. For a typical control mouse the volume needed to reach 20 mmHg is 30 μl (13).

The tight skin (TSk) mouse is known to develop cardiomegaly and have increased collagen content in skin, lung, and heart (14). The increased collagen content of the cardiac muscle might impair compliance of the left ventricle. By measuring the passive volume–pressure curves, the compliance in TSk and wild-type hearts can be compared. Although the variability within the group of TSk mice was large, there was no significant difference comparing the curve representing the whole group, compared to control groups. Transgenic mice that harbor a constitutively active Harvey ras (H-*ras*) gene driven by the ventricular myosin light chain-2 (MLC-2v) promoter display a hypertrophic phenotype and concomitant cardiac chamber dysfunction. These mice are presently being studied with this technique (15). In mice with marked hypertrophy the passive left ventricular pressure–volume curve is steeper and shifted to the left compared to "wild-type" littermates, indicating reduced left ventricular compliance.

Echocardiography

With the use of a high-frequency ultrasonic transducer system, transcutaneous two-dimensional, M-mode, and Doppler echocardiographic studies can now be performed in intact anesthetized mice. Adult mice are sedated with ketamine (100 mg/kg) and xylazine (5 μg/kg) administered intraperitoneally. Using an Interspec Apogee CX echocardiograph (Interspec, Ambler, PA), with the transducer in the 9-mHz mode, the left ventricle can generally be visualized in the transverse (short axis) or parasternal (long axis) views by two-dimensional imaging, with recordings of interventricular septum and posterior wall motion using M-mode display. Aortic and mitral valve velocity profiles can be assessed by pulsed Doppler echocardiography. With this new ultrasonic approach in mice it is possible to study the LV chamber diameter at end-

diastole (EDD) and end-systole (ESD), as well as the fractional shortening as a measure of LV function (EDD − ESD/EDD). It is also feasible to measure LV posterior wall thickness in most animals for the evaluation of chamber hypertrophy.

Digitized Microangiography to Study Embryonic Cardiac Malformations

Embryo Dissection

Timed pregnant females are prepared as described above. After making a longitudinal abdominal incision the uterus can be identified and mobilized by dissecting the connection to the oviducts, allowing externalization of the two uterine horns. Using a dissecting microscope (model ZDX-80; Scope Instruments, San Diego, CA) the uterine horns are opened by splitting the uterine wall along the antimesometrial side. This can be done by cleaving the muscular tissue along the long axis of the uterine horn with two pairs of forceps. By careful dissection the intact decidua of the individual embryos can be exposed one by one. After identification of the placenta the extraembryonic membranes can be opened, leaving the placental vessels intact. A V-shaped, thoracic incision anterior exposes the fetal heart. The exposed embryos are maintained by a continuous superfusion with Krebs–Henseleit solution (in millimolar units: NaCl 118.0, $NaHCO_3$ 14.5, KCl 2.6, KH_2PO_4 1.2, $MgSO_4$ 1.2, and glucose 11.1; pH 7.4) at 37° C and saturated with 15% O_2, 5% CO_2, and 80% N_2.

Imaging and Data Processing

The fetus is placed on a water-heated glass block within a molded clay ring to maintain the perfusate around the embryo. As the embryos are relatively translucent, real-time images of cardiac contraction can be obtained using transmitted light, an intravital Leitz microscope (Deerfield, IL), and an Optronics VI-470 color video camera (Goleta, CA). Glass micropipettes are drawn to a 2- to 3-μm diameter. Via the micropipette, 5 to 15 μl of 1.5% (v/v) fluorescein-conjugated bovine serum albumin is injected into the left atrium. If the left atrium is inaccessible, alternate cardiac chambers can be used. Fluorescent images of the blood flow can be recorded using a fluorescent attachment (Ploempak; Leitz) and a mercury light source. The left ventricular ejection fraction can be calculated by digitizing end-systolic and -diastolic images by using NIH 1.52 software (Bethesda, MD) and measuring areas by edge detection. The imaging system is calibrated in 50-μm increments, using a 2-mm scale. The left ventricular volumes can be estimated using the formula $\frac{4}{3}\pi a^2 b$, where a is the minor axis and b the major axis of the ellipse measured on the single plane digital image. After the imaging study the embryos can be processed for genotyping when analyzing transgenic embryos, histology, or electron microscopy.

This method was developed to study cardiac malformations in homozygous mutant embryos from the RXRα knock-out mice (16). RXRα is one of six retinoic acid receptors [three retinoic acid receptors (RAR) and three retinoid X receptors (RXR)],

which function as ligand-dependent transcription factors. Studies of the cardiac abnormalities found in RXRα knockout mice have documented an important role of retinoids in ventricular chamber growth and morphogenesis (17). RXRα homozygous mutant embryos appear normal until day 11.5 postcoitum but the embryos die at day 14.5–15.5 due to cardiac defects and generalized edema, which is a manifestation of embryonic congestive heart failure. The heart at this stage displayed a ventricular septal defect and ventricular hypoplasia with the absence of proliferation of the compact zone of the myocardium. Left ventricular ejection fractions in the wild-type and heterozygous 13.5- to 14.5-day-old embryos were 50%, whereas in homozygous mutant embryos this was reduced to approximately 15%.

Analysis of Mouse Cardiomyocytes

Culture of Dispersed Myocytes from Single Embryonic Hearts

To better understand the individual cellular characteristics of cardiac myocytes, we have developed a method to culture cells derived from ventricular or atrial chambers isolated from single hearts of mouse embryos from between day 9.5 and 14.5 of embryonic development. The advantage of this method is that hearts from single mouse embryos (of initially unknown genotypes) can be studied on an individual basis. All cells studied are derived from a single heart. Therefore, this method results in the culture of representative cells from a separate heart with no further purification of cardiomyocytes (i.e., smooth muscle cells, fibroblasts, etc., are not removed). Consequently, protocols are combined with a final step utilizing an immunohistochemical analysis to determine which cells are indeed myocytes. These cardiocytes have been used to study the expression of myofilaments such as the ventricular and atrial isoforms of the contractile protein myosin light chain 2 and myomesin (18) and have been demonstrated to contain ionic currents indicative of channels typically associated with the heart (19).

An excellent method of dissecting embryos from the uterus is described in detail by Sturm and Tam (20). The following are modifications applicable to the isolation and culture of individual ventricular- and atrial-derived cells. Pregnant mice are sacrificed by cervical dislocation and the uterus is removed into phosphate-buffered saline (PBS, in millimolar units: NaCl 137, KCl 2.7, Na_2HPO_4 10, and KH_2PO_4 1.8, pH 7.4). The outer surface of the uterus is rinsed free of blood and other debris and placed into fresh PBS for dissection. With watchmakers forceps, the uterus is opened longitudinally to expose the embryos still contained within the decidua. Care should be taken not to impart excessive compression of the decidua as this may compromise the integrity of the embryo. After each embryo has been dissected free from the uterine wall, they are placed in ADS buffer for removal of the heart [in millimolar

units: NaCl 116, *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES) 20, NaH_2PO_4 1.0, KCl 5.0, $MgSO_4$ 0.8, and glucose 5.5, pH 7.35, and filter sterilized with a 0.22-μm pore size filter]. The decidual mass, Reichert's membrane, the parietal and visceral yolk sacs, and the amnion are all removed, exposing the embryo proper. For day 9.5 to 12.5 the hearts are removed from the embryo by using a pair of forceps to separate the connection between the outflow tracts and the ventricular segment (day 9.5) or ventricular chamber (day 10.5–12.5), and the inflow tracts and the common atrial chamber. The heart is then placed in fresh ADS buffer for separation of the atria from ventricles. (No further dissection is attempted between right versus left chambers.) Removal of the common atria is accomplished by grasping the atria at the point of connection to the ventricles with one forceps and, with the other forceps, pulling free the atria. Each of these segments or chambers is then placed into an individual 1.5-ml Eppendorf tube containing 0.5 ml of ADS buffer on ice. Alternatively, more than one embryonic atrium or ventricle could be combined to give larger numbers of cells for culture. If this is desired, several chambers can be placed together into the same Eppendorf tube for subsequent digestion.

Following dissection of all hearts, the tubes are quick-spun in a tabletop microcentrifuge for 1 sec. The supernatant is removed and replaced with fresh ADS buffer containing collagenase type II (0.5 mg/ml; Worthington Biochemical Corp., Freehold, NJ) and pancreatin (1.0 mg/ml; GIBCO, Grand Island, NY). All tubes are then placed on a rotating wheel and allowed to digest for 15 min at 37° C. The Eppendorf tubes are quick-spun for 1 sec to remove buffer from the top of the tube. The tissue is further digested by triturating up and down in the pipette tip about five times. After mixing in this manner, the tissue from single chambers should be largely digested and barely visible. Tubes are then quick-spun again and the supernatant added to 2 ml of Dulbecco's modified Eagle's medium (DMEM)/199 supplemented with 10% (v/v) horse serum, 5% (v/v) fetal calf serum, penicillin (100 U/ml), and streptomycin (0.1 mg/ml). The process of digestion, incubation, and trituration is repeated one more time (for single hearts) or until complete dissociation of the embryonic heart chambers is achieved (for more than one heart). After trituration, the solution is combined with the first sample to give a final volume of 3 ml. This solution can either be plated directly or centrifuged at 500 *g* for 5 min at room temperature, resuspended in fresh medium, and then plated. Generally, from one embryonic day 9.5 to day 14.5 heart, atrial cells can be split into two wells of a four-chamber slide (Nunc, Roskilde, Denmark) that has been pretreated with 1% (w/v) gelatin for 10 min at room temperature followed by a 4-hr treatment with laminin (20 μg/ml; Sigma, St. Louis, MO) in PBS. Ventricular cells from day 9.5 to day 10.5 embryonic hearts are also split into two wells, while heart cells from embryos between day 11.5 and day 14.5 can be split into four chambers. Myocytes isolated by this method can be observed to be vigorously beating by the following day and, if left in culture for more than 3–4 days, begin to divide and proliferate. A disadvantage of extended culture, however,

is the opportunity for the noncardiac cells (i.e., fibroblasts) to overgrow the myocytes. Therefore, an ideal time frame for the usefulness of these cells is less than 3 days after isolation.

Discrimination between Atrial and Ventricular Cardiomyocytes

As indicated by the proportion of cells spontaneously contracting in culture, the majority (>90%) are either atrial or ventricular myocytes (21). A more definitive discrimination between cardiac myocytes and other cells such as smooth muscle cells or fibroblasts (which are also cocultured by this method) is accomplished through the use of antibodies that specifically recognize either the atrial (22) or ventricular (23) isoform of the contractile protein myosin light chain 2. If myocytes are plated on gridded coverslips, then experiments such as patch-clamp studies can be performed on the cells followed by *in situ* fixation and antibody staining to determine the specific cell type. Consequently, this method allows the direct comparison of atrial and ventricular myocytes from the same heart at a specific day of embryonic development. The following protocol describes the immunohistochemical staining of cells isolated from single hearts from mouse embryos.

All solutions are made in 1× PBS. Isolated cells on either chamber slides or coverslips are first briefly rinsed in PBS to remove any culture medium, then fixed in 4% (w/v) paraformaldehyde for 15 min. Cells are then neutralized in 50 m*M* NH_4Cl followed by two 5-min washes in PBS. If desired, slides may be stored at this point for up to 2 weeks at 4° C under PBS without compromising the integrity of the cells for immunostaining. After washing, myocytes are permeabilized and blocked simultaneously by incubation for 15 min in 0.3% (v/v) Triton X-100 and 1% (w/v) bovine serum albumin. Cardiomyocytes are then incubated for 60 min with primary antibody in 1% (w/v) bovine serum albumin, 2% (v/v) normal goat serum, and 0.1% (v/v) Tween 20 followed by four 5-min washes in PBS. Fluorescein isothiocyanate (FITC)-conjugated or Texas Red-conjugated secondary antibodies are used to visualize the phenotype of the cultured cells. Incubation with secondary antibodies is also for 60 min and is performed in the same buffer as for the primary antibodies. Because fluorochromes are sensitive to the light, this incubation is allowed to proceed in the dark. Following another PBS washing (four times, 5 min each), cells are briefly rinsed in distilled water to remove residual PBS. Using a Kimwipe tissue, as much water as possible should be removed without letting the cells dry out prior to mounting with Gelvatol plus 2.5% (w/v) triethylenediamine (DABCO) [see Harlow and Lane for instructions on how to make this solution (24)]. Slides are then stored overnight in the dark at room temperature to allow the Gelvatol to set. They can then be examined the following day using a microscope equipped with epifluorescence optics. Figure 2A shows atrial cells detected by using the MLC-2a primary anti-

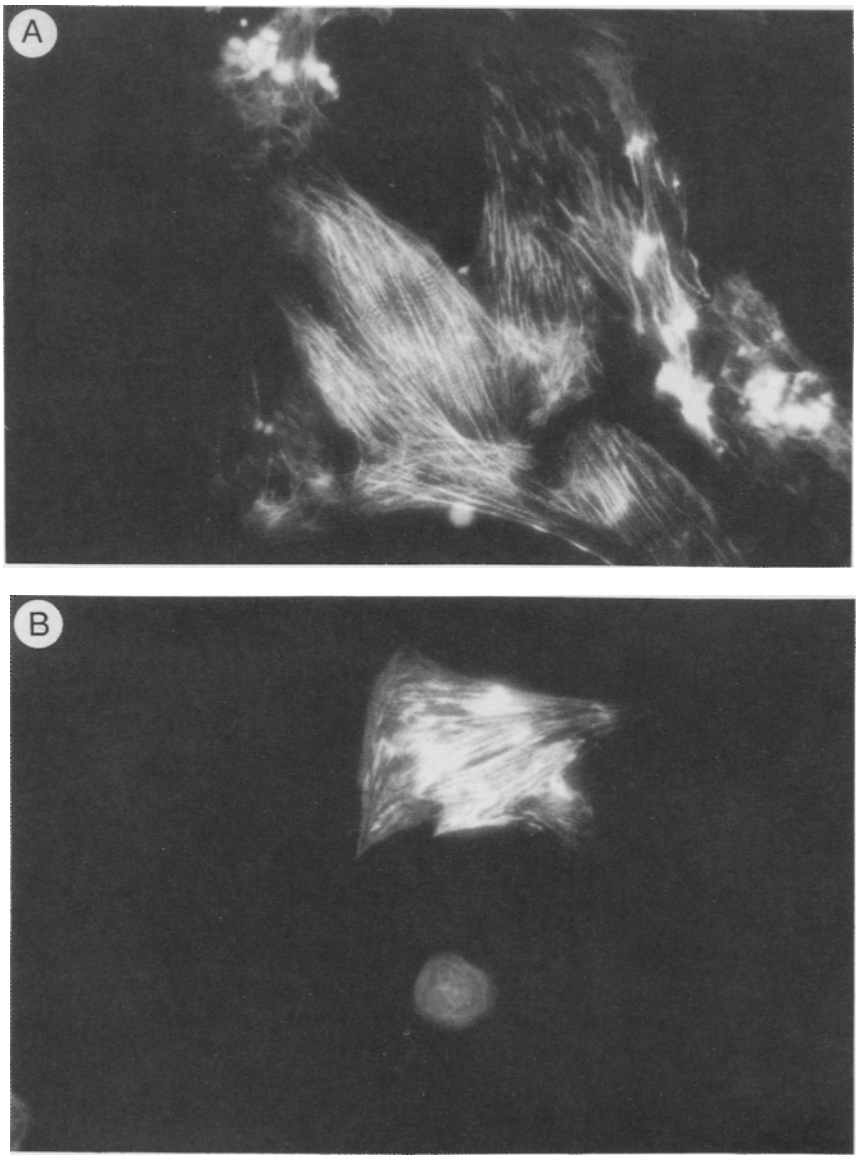


FIG. 2 Examples of immunostained atrial and ventricular cardiac myocytes. (A) Isolated myocytes from the atrial segment of a day-9.5 mouse embryo heart tube. Primary antibody was a 1:500 dilution of the MLC-2a antibody (22). Secondary antibody was a 1:50 dilution of FITC-conjugated donkey anti-rabbit antisera. (B) Isolate myocytes from the ventricular segment of a day-9.5 mouse embryo heart tube. Primary antibody was a 1:50 dilution of the MLC-2v antibody (23). Secondary antibody was the same as that used in immunostaining for MLC-2a above.

body, and Figure 2B shows ventricular cells detected by using the MLC-2v primary antibody. Both primary antibodies were visualized after incubation with an FITC-conjugated secondary antibody.

Embryonic Stem Cell Differentiation as Model for in Vitro Development

The differentiation of embryonic stem (ES) cells into embryoid bodies (EBs) results in "embryo-like" structures that express markers of endodermal, ectodermal, and mesodermal differentiation and have been used as a model system to study hematopoiesis, vasculogenesis, and angiogenesis (25–30). Additionally, EBs express a variety of muscle markers such as atrial and ventricular myosin light chain 2v (22, 31), myosin heavy chain, and tropomyosin (32–36), which suggests that this system may be useful as a model to study early cardiogenesis. *In vitro* differentiation of ES cells results in a proportion of EBs that demonstrate spontaneous contractile activity by day 8 in suspension cultures (31). Furthermore, it has been shown that the temporal activation of specific muscle genes in developing EBs mimicks that of embryonic myogenesis (31, 32–36). Thus, the embryonic stem cell system serves as a model for early cardiac muscle commitment and differentiation. It follows that purified cardiac cell progenitors from developing embryoid bodies would be an invaluable tool in pursuing questions involving gene regulation or genetic manipulations such as homologous recombination experiments that may otherwise be embryonic lethal in transgenic mice.

Embryonic Stem Cell Culture

Cells that have been successfully used for the following method or purification are the D3 embryonic stem cell line (37), and are maintained and propagated in high-glucose Dulbecco's modified Eagle's medium (Life Technologies, Inc., Grand Island, NY) supplemented with 15% (v/v) heat-inactivated fetal calf serum (Sigma), 2 m*M* L-glutamine, and 0.1 m*M* 2-mercaptoethanol (Sigma). Cells can be maintained in the undifferentiated state either by the addition of purified recombinant mouse leukemia inhibitory factor (LIF; Esgro, Life Technologies, Inc.) at 1000 units/ml to the culture medium or by culture on confluent feeder layers of mitomycin C-treated primary mouse embryonic fibroblasts. Under these conditions, the majority of the ES cells (>95%) possess an undifferentiated phenotype as assessed by visual inspection under phase-contrast microscopy. Cells are maintained at 37° C in a humidified atmosphere of 10% CO_2/90% air. Monolayers are passaged by trypsinization at 70–80% confluence. *In vitro* differentiation is initiated by harvesting ES cell monolayers with trypsin–EDTA and transferring dissociated single-cell suspensions to bacterial petri dishes containing growth medium as described above, excluding LIF. If ES cells

are grown on feeder layers, differential cell culture substrate attachment is performed to remove feeder cells (35). To ensure single-cell suspension, the harvested cells are passaged through a small-bore plastic pipette; in some cases, additional trituration by passage through a fine-bore Pasteur pipette may be required. On the fourth day of culture, the serum concentration is raised to 20% and the medium changed every other day, with replacement of half of the medium on the in-between days. Under these conditions, cells aggregate to form three-dimensional structures or "embryoid bodies," which increase in complexity as differentiation progresses. In differentiation cultures, 40–70% of EBs should exhibit spontaneous contractile activity, usually first identified after 1 week in culture.

Purification of Embryoid Body-Derived Cardiac Myocytes

The following method results in an approximate 500- to 1000-fold increase in purity of cardiac lineage muscle cells compared to simple enzymatic dispersion of the cells from embryoid bodies, each of which contains only a small fraction of cardiac myocytes. For purification of cardiac lineage cells, ES cells and induction of EB development is performed as above. However, on the fourth day after initiating EB formation the EBs are transferred either to new dishes or, for large-scale cultures, to spinner flasks. For large-scale cultures, approximately 3.5×10^7 ES cells are incubated in 1 liter of medium. Spinner flask cultures should be kept in a 37°C warm room and receive EB growth medium supplemented with 25 m*M* HEPES daily. The total volume of the culture should be adjusted to maintain a neutral pH.

Embryoid bodies that have been allowed to develop for 9–14 days have been successfully used for this isolation procedure and can yield up to 2×10^6 cells/liter of culture. To harvest cardiac lineage muscle cells, EBs are allowed to settle in several 50-ml conical centrifuge tubes, washed with 50 ml of PBS, and incubated for 4 hr at 37°C in 8 ml of PBS (per 2 ml of packed EBs) containing collagenase (3 mg/ml) (Worthington) and bovine serum albumin (20 mg/ml). The EBs are incubated with an equal volume of growth medium for 5–10 min and collected by centrifuging at 200 *g* for 5 min at room temperature. The pellet is resuspended in growth medium and remaining aggregates allowed to settle by gravity, followed by centrifugation of the supernatant at 200 *g* for 5 min at room temperature. After a final resuspension in growth medium, cells are purified via a discontinuous Percoll (Pharmacia, Piscataway, NJ) density gradient. The gradients are made by first preparing a stock solution of Percoll with a density of 1.110 g/ml by diluting Percoll with 10× ADS buffer (nine parts commercially available Percoll to one part 10× ADS). A bottom layer, 1.090 g/ml, and a top layer, 1.070 g/ml, Percoll solution are then made by diluting stock with 1× ADS buffer. For the 1.070-g/ml solution, mix an equal volume of 1× ADS and stock Percoll. For the 1.090-g/ml solution, mix one part 1× ADS and three parts stock Percoll. Four gradient tubes are set up for each 1 ml of original packed EBs. Put 4 ml of 1.070-g/ml Percoll in a 15-ml conical tube and carefully underlay

with 1 ml of 1.090-g/ml Percoll. Layer the cells in 2 ml of 1× ADS buffer on top of the gradients. Centrifuge at 3000 *g* for 30 min at room temperature (T6000B; Sorvall, DuPont Medical Products, Wilmington, DE). Cells that migrate to the interface are myocytes and are removed by aspirating the top Percoll solution down to just above the interface. Myocytes are removed, along with some of the Percoll just above and below the cells, with a wide-bore pipette and are transferred to a 50-ml conical tube. Following a washing step with 50 ml of medium, cells are plated in medium containing 20% fetal calf serum on gelatin and laminin-coated glass coverslips or chamber slides (Nunc) as stated above.

Cardiac lineage muscle cells are allowed to attach for 24–36 hr prior to performing experiments. The expected yield of cells is approximately 82×10^5 cells for a 250-ml culture. Cells harvested by this method result in 60–98% beating cells. The identification of cells as cardiac is determined by immunostaining 4% (w/v) paraformaldehyde-fixed cells using antibodies specific for the atrial or ventricular myosin light chain 2v as described above.

Microinjection of Cardiac Myocytes

A distinct advantage of isolated cardiac myocytes is their usefulness for microinjection experiments to investigate the effects of introducing specific signaling molecules (i.e., antibodies, proteins, etc.) and expression vectors into single cells (38–40). At present, microinjection of cardiac cells has largely been performed in isolated rat neonatal myocytes because purification protocols yield more cells from rat heart as compared to mouse heart. Nonetheless, microinjection of rat neonate myocytes serves as an excellent assay system with which to examine the role of various molecules and expression vectors in normal and abnormal functioning of the cardiac myocyte. The following method is a modification of the original isolation procedure of Simpson and Shoshana (41) and describes the purification of rat neonate cardiac myocytes. In addition, details of the microinjection procedure of these cells are described.

Isolation of Rat Neonatal Cardiac Myocytes

Hearts from 1- to 2-day-old Sprague-Dawley rats are recovered (the method reported here is for 100 neonates yielding $1-2 \times 10^6$ cells/animal) and the ventricles are pooled and trisected into 1× ADS buffer on ice. The myocytes are dispersed in a spinner flask by digestion with a combination of collagenase type II (0.5 mg/ml) and pancreatin (1.0 mg/ml) in 1× ADS buffer for 6 min at 37° C. The enzyme solution is removed with a pipette and discarded because it contains blood cells and other cellular debris. Thirty-five milliliters of fresh enzyme solution is added to the tissue and incubated an additional 20 min at 37° C. The tissue is allowed to settle and the enzyme mixture containing isolated cells removed and transferred to a 50-ml conical centrifuge tube containing 2 ml of newborn calf serum (to inactivate the enzymes).

Fresh enzyme solution is added to the flask and incubated for another 20 min at 37° C. The 50-ml conical tube is centrifuged at 500 *g* for 5 min at room temperature, after which the supernatant is aspirated and 4 ml of newborn calf serum added to resuspend the pellet. This mixture is then transferred to a clean 50-ml conical tube and placed in a 37° C, 10% CO_2 incubator with the lid loosened to allow gas exchange. The process of adding fresh enzyme solution and dispersion of cells is repeated three more times (a total of five digestions that are saved) or until the tissue is completely digested. Each 4 ml of resuspended cells is pooled together and centrifuged at 500 *g* for 5 min at room temperature. The supernatant is aspirated, and cells are resuspended in 24 ml of ADS buffer.

Purification of Cardiac Myocytes

Myocytes are purified using a discontinuous Percoll density gradient similar to that described above for embryoid bodies except that the top layer is 1.060 g/ml and the bottom layer is 1.080-g/ml Percoll (1.060-g/ml Percoll is made by combining 90 ml of Percoll stock plus 110 ml of 1× ADS; 1.080-g/ml Percoll is made by combining 130 ml of Percoll stock with 70 ml of 1× ADS). Four milliliters of 1.060-g/ml Percoll is placed in a 15-ml polystyrene tube. For 100 hearts, twelve 4-ml gradients are required. Using a 5-ml pipette, slowly pipette 3 ml of the 1.080-g/ml Percoll mix in the bottom of the tube so that the top layer rises without mixing with the bottom layer. Transfer 2 ml of cells onto each gradient, being careful to make a sharp interface between the cells and the top Percoll. Centrifuge the tubes at 2500 *g* for 30 min at room temperature in a table-top centrifuge. Gently aspirate the top layer of Percoll to just above the myocyte layer. Remove the myocytes and combine the cells of all tubes. Cells are split between two 50-ml conical tubes and each washed with 45 ml of supplemented DMEM/199, centrifuged, and resuspended in 20 ml of supplemented DMEM/199. Purified cardiac myocytes are plated on gelatin/laminin-pretreated chamber slides, coverslips, or tissue culture dishes as described above for cardiac lineage muscle cells from embryoid bodies. For microinjection, cells should be plated at a density of approximately 20,000 cells/cm^2.

Preparation of Cells for Microinjection

Immediately prior to injection, cells are treated for at least 1 hr with 20 m*M* 2,3-butanedione monoxomine (BDM), which prevents hypercontraction after influx of calcium caused by puncture of the cell membrane and increases long-term viability after microinjection. The cells remain in this medium throughout the microinjection procedure and 4–6 hr after injection. Wash cells and incubate in appropriate BDM-free medium until harvest.

Microinjection of Cardiac Myocytes

The following microinjection protocol has been performed using the Eppendorf ECET microinjector 5242 system, the ECET micromanipulator 5170 system, the

ECET Femtotips, and a Zeiss axioverted microscope. Microinjector pressure settings may need to be empirically determined; however, the following settings have been successfully used for injection into rat neonate myocytes (38–40). Purge pressures (P1) must be over 3000 hPa. Injection pressures (P2) will vary from injectate to injectate and with the bore size of the needle, but initial experiments may be started at 200 hPa for automatic injections. Holding pressures (P3), which regulate the continuous flow of injectate out of the needle, should be lower than P2 when carrying out automatic injections (75 hPa) and much higher than P2 when carrying out manual injections (400 hPa). The injection time should be between 0.3 and 0.5 sec.

The injectate can be loaded in the capillary needle by backloading or by using Eppendorf microloaders. Backloading is achieved by putting the blunt end of the needle into the injectate, allowing loading of the needle by capillary action. If using Eppendorf microloaders, load about 1.2 μl. Injection buffers consist of 50 m*M* HEPES–40 m*M* NaCl (pH 7.4) or 100 m*M* KCl–5 m*M* KH_2PO_4 (pH 7.4) for plasmid DNA and 0.5× PBS for protein injections. Concentrations of DNA and protein should be optimized for each study. However, experiments may be initiated with 0.1–1.0 μg of DNA per microliter and between 4–10 μg of protein per microliter. DNA samples should be precipitated, pelleted, washed with 70% (v/v) ethanol, resuspended in injectant buffer, and stored at 4°C. Both DNA and protein samples should be centrifuged in a microcentrifuge at 14,000 rpm at room temperature to pellet particulate material prior to injection.

Acknowledgments

The authors thank the NHLBI (K.R.C.: HL36139, HL45069, and HL46345) and the AHA (K.R.C.: 91-022170) for financial support. S.W.K. is a recipient of an Individual NIH National Research Award.

References

1. K. R. Chien, H. Zhu, K. U. Knowlton, W. Miller-Hance, M. van Bilsen, T. X. O'Brien, and S. M. Evans, *Annu. Rev. Physiol.* **55,** 77 (1993).
2. M. Kirby, *Pediatr. Res.* **21,** 219 (1987).
3. A. Korestsky, *Am. J. Physiol.* **262C,** 261 (1992).
4. K. R. Chien, *Science* **260,** 916 (1993).
5. H. A. Rockman, R. S. Ross, A. N. Harris, K. U. Knowlton, M. E. Steinhelper, L. J. Field, J. J. Ross, and K. R. Chien [published erratum appears in *Proc. Natl. Acad. Sci. U.S.A.* **88,** 9907 (1991)], *Proc. Natl. Acad. Sci. U.S.A.* **88,** 8277 (1991).
6. H. A. Rockman, S. P. Wachhorst, L. Mao, and J. J. Ross, *Am. J. Physiol.* **266,** H2468 (1994).
7. H. A. Rockman, K. U. Knowlton, J. R. Ross, Jr., and K. R. Chien, *Circulation* **87,** 14 (1993).

8. H. A. Rockman, S. Ono, R. S. Ross, L. R. Jones, M. Karimi, V. Bhargava, J. J. Ross, and K. R. Chien, *Proc. Natl. Acad. Sci. U.S.A.* **91,** 2694 (1994).
9. J. K. Gwathmey, L. Copelas, R. MacKinnon, F. J. Schoen, M. D. Feldman, W. Grossman, and J. P. Morgan, *Circ. Res.* **61,** 70 (1987).
10. R. Barbee, B. Perry, R. Re, and J. Murgo, *Am. J. Physiol.* **263,** R728 (1992).
11. C. A. Milano, P. C. Dolber, H. A. Rockman, R. A. Bond, M. E. Venable, L. F. Allen, and R. J. Lefkowitz, *Proc. Natl. Acad. Sci. U.S.A.* **91,** 10109 (1994).
12. C. A. Milano, L. F. Allen, H. A. Rockman, P. C. Dolber, T. R. McMinn, K. R. Chien, T. D. Johnson, R. A. Bond, and R. J. Lefkowitz, *Science* **264,** 582 (1994).
13. J. H. Omens, H. A. Rockman, and J. W. Covell, *Am. J. Physiol.* **266H,** 1169 (1994).
14. T. Osborn, R. Bashey, T. Moore, and V. Fischer, *J. Mol. Cell. Cardiol.* **19,** 581 (1987).
15. J. J. Hunter, N. Tanaka, H. A. Rockman, J. J. Ross, and K. R. Chien, *J. Biol. Chem.* **270,** 23173 (1995).
16. E. Dyson, H. Sucov, S. W. Kubalak, G. W. Schmid-Schonbein, F. Delano, R. M. Evans, J. J. Ross, and K. R. Chien, *Proc. Natl. Acad. Sci. U.S.A.* **92,** 7386 (1995).
17. H. M. Sucov, E. Dyson, C. L. Gumeringer, J. Price, and K. R. Chien, *Genes Dev.* **8,** 1007 (1994).
18. J. M. Schnee, S. J. Fuller, S. W. Kubalak, and K. R. Chien, unpublished observations (1994).
19. M. P. Davies, R. H. An, P. A. Doevendans, S. W. Kubalak, K. R. Chien, and R. S. Kass, *Circ. Res.* **78,** 15 (1996).
20. K. Sturm and P. P. L. Tam, *Methods Enzymol.* **225,** 164 (1993).
21. S. W. Kubalak and K. R. Chien, unpublished observations (1994).
22. S. W. Kubalak, W. C. Miller-Hance, T. X. O'Brien, E. Dyson, and K. R. Chien, *J. Biol. Chem.* **269,** 16961 (1994).
23. K. Iwaki, V. P. Sukhatme, H. E. Shubeita, and K. R. Chien, *J. Biol. Chem.* **265,** 13809 (1990).
24. E. Harlow and D. Lane, "Antibodies: A Laboratory Manual." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1988.
25. W. Risau, H. Sariola, H.-G. Zerwes, J. Sasse, P. Ekblom, R. Kemler, and T. Doetschman, *Development* **102,** 471 (1988).
26. M. H. Lindenbaum and F. Grosveld, *Genes Dev.* **4,** 2075 (1990).
27. R. M. Schmitt, E. Bruyns, and H. R. Snodgrass, *Genes Dev.* **5,** 728 (1991).
28. M. V. Wiles and G. Keller, *Development* **111,** 259 (1991).
29. R. Wang, R. Clark, and L. V. Bautch, *Development* **114,** 303 (1992).
30. G. Keller, M. Kennedy, T. Papayannopoulou, and M. V. Wiles, *Mol. Cell. Biol.* **13,** 473 (1993).
31. W. C. Miller-Hance, M. LaCorbiere, S. J. Fuller, S. M. Evans, G. Lyons, C. Schmidt, J. Robbins, and K. R. Chien, *J. Biol. Chem.* **268,** 25244 (1993).
32. J. Robbins, J. Gulick, A. Sanchez, P. Howles, and T. Doetschman, *J. Biol. Chem.* **265,** 11905 (1990).
33. J. Robbins, T. Doetschman, W. K. Jones, and A. Sanchez, *Trends Cardiovasc. Med.* **2,** 44 (1992).
34. D. F. Wieczorek, P. Howles, and T. Doetschman, "The Dynamic State of Muscle Fibers." Walter de Gruyter, New York, 1990.
35. A. Sanchez, W. K. Jones, J. Gulick, T. Doetschman, and J. Robbins, *J. Biol. Chem.* **266,** 22419 (1991).

36. M. Muthuchamy, L. Pajak, P. Howles, T. Doetschman, and D. F. Wieczorek, *Mol. Cell. Biol.* **13,** 3311 (1993).
37. T. C. Doetschman, H. Eistetter, M. Katz, W. Schmidt, and R. Kemler, *J. Embryol. Exp. Morphol.* **87,** 27 (1985).
38. H. E. Shubeita, J. Thorburn, and K. R. Chien, *Circulation* **85,** 2236 (1992).
39. A. Thorburn, J. Thorburn, S.-Y. Chen, S. Powers, H. E. Shubeita, J. R. Feramisco, and K. R. Chien, *J. Biol. Chem.* **268,** 2244 (1993).
40. V. J. LaMorte, J. Thorburn, D. Absher, A. Spiegel, J. H. Brown, K. R. Chien, J. R. Feramisco, and K. U. Knowlton, *J. Biol. Chem.* **269,** 13490 (1994).
41. P. Simpson and S. Shoshana, *Circ. Res.* **50,** 101 (1982).

This Page Intentionally Left Blank

Index

R

S

T