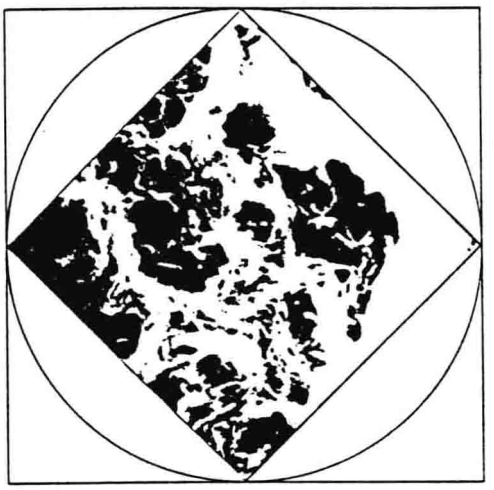
Molecular Genetics of Mycobacteria

Molecular Genetics of Mycobacteria



edited by

GRAHAM F. HATFULL

Pittsburgh Bacteriophage Institute and Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260

and

WILLIAM R. JACOBS, JR.

Howard Hughes Medical Institute, Albert Einstein College of Medicine, Bronx, NY 10461

ASM PRESS • WASHINGTON, D.C.

Cover: A microcolony of the Erdman strain of *Mycobacterium tuberculosis*, prepared on a microscope slide and stained with auramine-rhodamine T stain. M. *tuberculosis* forms characteristic "densely bunched and braided groups" (described by Robert Koch in his 1882 paper), which is often referred to as cording. This cording phenotype has been associated with virulence strains. Recent work showing that non-cording mutants are attenuated for growth has confirmed this observation (M. S. Glickman, J. S. Cox, and W. R. Jacobs, Jr., "A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of *Mycobacterium tuberculosis*," *Mol. Cell* **5**:717–727, 2000).

 ASM Press
American Society for Microbiology
1752 N Street, N.W.
Washington, DC 20036-2804

Library of Congress Cataloging-in-Publication Data

Molecular genetics of mycobacteria / edited by Graham F. Hatfull and William R. Jacobs, Jr.
p. cm.
Includes bibliographical references and index.
ISBN 1-55581-191-4
1. Mycobacteria. 2. Bacterial genetics. 3. Tuberculosis—Genetic aspects. .I. Hatfull, Graham F. II. Jacobs, William R., Jr.

QR82.M8 M64 2000
579.3'74135—dc21

00-028899

Molecular Genetics of Mycobacteria

Contributors

STOYAN BARDAROV
Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY 10461

ALAIN R. BAULARD
INSERM U447, Mécanismes Moléculaires de la Pathogénie Microbienne, Institut Pasteur de Lille, 1, rue du Pr. Calmette, 59019 Lille, France

AIMEE E. BELANGER
Department of Biological Sciences, University of Pittsburgh Medical School, Pittsburgh, PA 15260

JOHN T. BELISLE
Department of Microbiology, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, 200 West Lake St., Fort Collins, CO 80523

GURDYAL S. BESRA
Department of Microbiology and Immunology, The Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, United Kingdom

WILLIAM R. BISHAI
Center for Tuberculosis Research, Johns Hopkins School of Public Health, 615 N. Wolfe St., Baltimore, MD 21205

MIRIAM BRAUNSTEIN
Department of Microbiology and Immunology, Howard Hughes Medical Institute, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, NY 10461

R. BROSCH
Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex, France

PING CHEN
Center for Tuberculosis Research, Johns Hopkins School of Public Health, 615 N. Wolfe St., Baltimore, MD 21205

S. T. COLE
Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex, France

DESMOND M. COLLINS
AgResearch, Wallaceville Animal Research Centre, P.O. Box 40063, Upper Hutt, New Zealand

M. J. COLSTON
Division of Mycobacterial Research, The National Institute for Medical Research, London NW7 1AA, United Kingdom

NANCY D. CONNELL
Department of Microbiology and Molecular Genetics, New Jersey Medical School National Tuberculosis Center, University of Medicine and Dentistry of New Jersey, 185 South Orange Ave. MSB F605, Newark, NJ 07103

SUJOY K. DAS GUPTA
Department of Microbiology, Bose Institute, Calcutta-700009, India

E. O. DAVIS
Division of Mycobacterial Research, The National Institute for Medical Research, London NW7 1AA, United Kingdom

STEPHANIE S. DAWES
Molecular Biology Unit, South African Institute for Medical Research, and Department of Hematology and Molecular Medicine, University of the Witwatersrand Medical School, Johannesburg 2000, South Africa

KEITH M. DERBYSHIRE
Molecular Genetics Program, Wadsworth Center, David Axelrod Institute for Public Health, 120 New Scotland Ave., Albany, NY 12208

KEN DUNCAN
Glaxo Wellcome Research and Development, Gunnels Wood Road, Stevenage SG1 2NY, United Kingdom

K. EIGLMEIER
Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex, France

T. GARNIER
Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex, France

BRIGITTE GICQUEL
Unité de Génétique Mycobactérienne, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France

MANUEL GOMEZ
The TB Center, Public Health Research Institute, 455 First Ave., New York, NY 10016

JAMES GOMEZ
Center for Tuberculosis Research, Johns Hopkins School of Public Health, 615 N. Wolfe St., Baltimore, MD 21205

S. V. GORDON
Central Veterinary Laboratory, Woodham Lane, New Haw, Addlestone, Surrey KT15 3NB, United Kingdom

CHRISTOPHE GUILHOT
Unité de Génétique Mycobactérienne, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France

GRAHAM F. HATFULL
Pittsburgh Bacteriophage Institute and Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260

JULIA M. INAMINE
Mycobacteria Research Laboratories, Department of Microbiology, Center Avenue, Colorado State University, Fort Collins, CO 80523-1677

WILLIAM R. JACOBS, JR.
Howard Hughes Medical Institute, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, NY 10461

SHRUTI JAIN
Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260

LAURENT KREMER
Department of Microbiology and Immunology, The Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, United Kingdom

RUTH A. McADAM
Glaxo Wellcome Medicines Research Centre, Stevenage, Herts SG1 2NY, United Kingdom

VALERIE MIZRAHI
Molecular Biology Unit, South African Institute for Medical Research, and Department of Hematology and Molecular Medicine, University of the Witwatersrand Medical School, Johannesburg 2000, South Africa

K. G. PAPAVINASASUNDARAM
Division of Mycobacterial Research, The National Institute for Medical Research, London NW7 1AA, United Kingdom

CAREY PASHLEY
Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom

MARTIN S. PAVELKA, JR.
Department of Microbiology and Immunology, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642

SELWYN QUAN
Glaxo Wellcome Medicines Research Centre, Stevenage, Herts SG1 2NY, United Kingdom

HARVEY RUBIN
Division of Infectious Diseases, University of Pennsylvania Medical Center, Philadelphia, PA 19104-6073

JAMES C. SACCHETTINI
Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843-2128

ISSAR SMITH
The TB Center, Public Health Research Institute, 455 First Ave., New York, NY 10016

VALERIE SNEWIN
Department of Infectious Diseases and Microbiology, Imperial College School of Medicine, Norfolk Place, London W2 1PG, United Kingdom

GRAHAM STEWART
Department of Infectious Diseases and Microbiology, Imperial College School of Medicine, Norfolk Place, London W2 1PG, United Kingdom

NEIL G. STOKER
Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom

F. TEKAIA
Unité de Génétique Moléculaire des Levures, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France

AMALIO TELENTI
Division of Infectious Diseases and Institute of Microbiology, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland

ANIL K. TYAGI
Department of Biochemistry, University of Delhi South Campus, Benito Juarez Road, New Delhi-110021, India

E. YERAMIAN
Unité de Physiochimie des Macromolécules Biologiques, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France

DOUGLAS YOUNG
Department of Infectious Diseases and Microbiology, Imperial College School of Medicine, Norfolk Place, London W2 1PG, United Kingdom

YING ZHANG
Department of Molecular Microbiology and Immunology, Johns Hopkins University School of Hygiene and Public Health, 615 N. Wolfe St., Baltimore, MD 21205

years ago, prompting an explosion of activity in subsequent years. As a consequence, we have a deeper understanding of the biology of the mycobacteria and have perhaps acquired a new respect for these organisms. As our knowledge of the mycobacteria expands over the next dozen years, we look forward to new and more effective means of controlling the diseases that they cause.

We would like to express our thanks to each of our fellow authors and to the staff of ASM Press, especially Greg Payne and Ellie Tupper, without whose ceaseless patience and considerable skill this book would never have been a reality.

GRAHAM F. HATFULL
WILLIAM R. JACOBS, JR.

Contents

Contributors / vii

Preface / x

Introduction. *Mycobacterium tuberculosis*: a Once Genetically Intractable Organism / 1
WILLIAM R. JACOBS, JR.

GENOMES AND GENETIC EXCHANGE

1 Genomics, Biology, and Evolution of the *Mycobacterium tuberculosis* Complex / 19
R. BROSCH, S. V. GORDON, K. EIGLMEIER, T. GARNIER, F. TEKAIA, E. YERAMIAN, AND S. T. COLE

2 Molecular Genetics of Mycobacteriophages / 37
GRAHAM F. HATFULL

3 Plasmids in Mycobacteria / 55
CAREY PASHLEY AND NEIL G. STOKER

4 Mycobacterial Transposons and Their Applications / 69
RUTH A. McADAM, SELWYN QUAN, AND CHRISTOPHE GUILHOT

5 General Recombination / 85
M. J. COLSTON, E. O. DAVIS, AND K. G. PAPAVINASASUNDARAM

6 DNA Transfer in Mycobacteria: Conjugation and Transduction / 93
KEITH M. DERBYSHIRE AND STOYAN BARDAROV

GENE EXPRESSION

7 Determinants of Mycobacterial Gene Expression / 111
MANUEL GOMEZ AND ISSAR SMITH

8 Gene Expression: Reporter Technologies / 131
ANIL K. TYAGI, SUJOY K. DAS GUPTA, AND SHRUTI JAIN

9 Mycobacterial Transcription Regulation in Stationary Phase / 149
PING CHEN, JAMES GOMEZ, AND WILLIAM R. BISHAI

METABOLISM

10 DNA Replication / 159
VALERIE MIZRAHI, STEPHANIE S. DAWES, AND HARVEY RUBIN

11 Genetics of Mycolic Acid Biosynthesis / 173
LAURENT KREMER, ALAIN R. BAULARD, AND GURDYAL S. BESRA

12 Genetics of Cell Wall Biosynthesis / 191
AIMEE E. BELANGER AND JULIA M. INAMINE

13 Genetics of Protein Secretion / 203
MIRIAM BRAUNSTEIN AND JOHN T. BELISLE

14 Genetics of Mycobacterial Metabolism / 221
MARTIN S. PAVELKA, JR.

15 Genetics of Drug Resistance in *Mycobacterium tuberculosis* / 235
YING ZHANG AND AMALIO TELENTI

GENETIC STRATEGIES

16 Mutants and Mutagenesis / 257
NANCY D. CONNELL AND AIMEE E. BELANGER

17 Genetics of Mycobacterial Virulence / 265
DESMOND M. COLLINS AND BRIGITTE GICQUEL

18 Genetic Strategies for Vaccine Development / 279
VALERIE SNEWIN, GRAHAM STEWART, AND DOUGLAS YOUNG

19 Approaches to Tuberculosis Drug Development / 297
KEN DUNCAN AND JAMES C. SACCHETTINI

APPENDIXES

1 Some Common Methods in Mycobacterial Genetics / 313
Compiled by MICHELLE H. LARSEN

2 Internet Sites with Further Information on Mycobacterial Genetics / 321

3 Genome Organization of *Mycobacterium tuberculosis* H37Rv / 323

Index / 355

Mycobacterium tuberculosis: a Once Genetically Intractable Organism

WILLIAM R. JACOBS, JR.

> He had such a very full translation of Dr. Koch's famous paper made in English for me and presented it to me at Christmas. Surely I had never had a Christmas present that meant more to me than that big handwritten copybook! I read every word of it over and over again. Koch's paper on "The Etiology of Tuberculosis" is certainly one of the most, if not the most, important medical papers ever written, and a model of logic in application of the new experimental method to the study of disease.
>
> —Edward Livingston Trudeau (1915)

When I first read this passage, I was struck by the passion that this man had for Robert Koch's discovery. Indeed, E. L. Trudeau was a man of passion. In 1868, he graduated from Columbia Medical School (at that time, tuition was $300.00 for the three years plus a $5.00 matriculation fee). He watched his brother die of the dreaded disease consumption, and in 1871 he was himself diagnosed with consumption. It was a death sentence back then. Recognizing his fate, Trudeau reasoned that he might as well live his remaining life to the fullest, and decided to spend some time hunting and fishing at Paul Smith's camp in the Adirondack Mountains in upstate New York.

Edward Trudeau survived in that environment. He went back to New York City to bring his wife to Saranac Lake in the Adirondack Mountains. There they started a family; in fact, Garry Trudeau, author of the "Doonesbury" comic strip, is his great-grandson. Edward reasoned that there was healing power in the fresh air and sunlight of the mountains, and he set up the first "cure cottages" for the treatment of consumption—the dreaded tuberculosis.

Many notable people ended up visiting his "cure cottages." I am sure many of them asked the great doctor, "What causes consumption?" In those days the hypotheses included (1) the wrath of God, (2) an inherited trait, or (3) bad air. I have heard similar hypotheses for the cause of acquired immunodeficiency syndrome (AIDS). The knowledge of which hypothesis is correct is obviously critical, as the therapies to treat each hypothetical cause are radically different. The question in 1880 was, how could we find out?

Trudeau recognized that Robert Koch's paper on the "Aetiology of Tuberculosis" provided an experimental methodology and an intellectual basis (Koch's Postulate; see Table 1) to acquire the knowledge of the cause of tuberculosis. Trudeau was so inspired from reading this paper that he set up the first research laboratory in America to study tuberculosis. He went on to confirm Koch's work and published the famous "Rabbit Island" paper in 1897 (reviewed wonderfully by Frank Collins [1997]), demonstrating the curative effects of sunlight and fresh air.

I think Trudeau realized that knowledge of one's enemy (in this case the tubercle bacillus) is an essential first step for devising a battle plan to stop the disease. Similarly, the knowledge of the molecular bases (the genotypes) for the phenotypes (the properties or characteristics) which allow *Mycobacterium tuberculosis* to cause disease and escape therapies is essential in developing effective new strategies. The philosophical methodology by which we acquire the knowledge of what genotype is responsible for a particular phenotype is called the Molecular Koch's Postulate (Falkow, 1988) or Koch's Molecular Postulate (Table 1).

Essential to fulfilling Koch's Molecular Postulate is the ability to perform a gene transfer. I believe that the discovery of gene transfer in bacteria revolutionized biology during the 20th century. With the tools of transformation, conjugation, and transduction in bacteria, Koch's Molecular Postulate could be fulfilled and the knowledge of the basis for phenotypes for many other bacteria could be determined. Since bacteria and their phages, particularly *Escherichia coli*, are genetically tractable, easy to grow, and facile for any sort of biochemical or biophysical analyses, much of our basic understanding of molecular genetics, biochemistry, and cell biology comes from studies using those experimental systems. In contrast, while tuberculosis probably has the richest history of research of any disease, the inability to fulfill Koch's Molecular Postulate has severely curtailed advances. For example, in 1990, none of the enzyme targets for the antimycobacterial drugs isoniazid (INH), ethionamide (ETH), ethambutol, pyrazinamide, or para-salicylic acid were known. Similarly, in 1990, Koch's Molecular Postulate had not been validated for any virulent phenotype. The reason for being unable to fulfill the postulates lay not in the acquisition of mutants nor the ability to clone M. *tuberculosis* DNA. The problem was achieving the third condi-

Molecular Genetics of Mycobacteria
Edited by G. F. Hatfull and W. R. Jacobs, Jr.

TABLE 1 Koch's postulate[a] compared to Koch's molecular postulate

Koch's Postulate
To prove that tuberculosis was caused by the invasion of the bacilli, and the growth and multiplication of bacilli, it was necessary to: 1. isolate the bacilli from the body, 2. grow them in pure culture . . ., and 3. by administering the isolated bacilli to animals, reproduce the same moribund conditions (Koch, 1882)
Koch's Molecular Postulate
To prove that a phenotype in a mutant bacterium, such as drug resistance or virulence, is caused by a specific genotype, it is necessary to: 1. isolate a mutant with a defined altered phenotype, 2. clone the genotype from the mutant, and 3. upon introducing the cloned genotype into a wild-type bacterium, reproduce the same phenotype of the mutant bacterium

[a] Although many texts use "Postulates," I believe that this is incorrect. The singular form "postulate" should be used, as a postulate is defined as a self-evident truth. The truth in this case is that a bacillus is the cause of a disease. The three numbered phrases are conditions all of which must be met for the postulate to be asserted.

tion—gene transfer. There existed no way to transfer genes into the chromosome of the tubercle bacillus.

Over the past 15 years, this hurdle has finally been overcome. The goal of this chapter is to provide an account of the key steps it took to achieve gene transfer for *M. tuberculosis*. In contrast to this entire book, which aims to present the comprehensive state of the art of genetics for *M. tuberculosis*, this chapter will aim to provide the philosophy of genetic experimentation and models of logic in the application of previously unavailable experimental methods to the study of *M. tuberculosis*.

BACTERIAL GENETICS 101

I was a mathematics major in college, primarily because memorizing lots of seemingly unrelated facts was a struggle. I believed that if I simply memorized $A \times 1 = A$ and $A + 0 = A$, I could figure out all of Algebra. Similarly, I believe that there are only five principles you need to remember, and you can figure out all of bacterial genetics.

Principle 1. Bacterial genetics is just like New York City: anything that can happen, will happen, all at observed frequencies. The day I learned about the lactose operon, I was hooked on the power of bacterial genetics. I was awestruck how a simple bacterium could decide if it wanted to metabolize lactose and turn on a set of genes in response to that desire. What amazed me more was the way Jacob and Monod used mutants to elegantly define this process. Their use of mutants provided an unprecedented vision of a set of genes and the functions of genes in *E. coli*. The vision was remarkable, as it provided, with clarity, a cellular process. I remember thinking to myself, how did they make those mutants? The first time I saw the Bachmann genetic map (Bachmann and Low, 1980), I was amazed at how it defined the genome based on mutants that had been isolated. So many of the mutations were identified by using an amazing set of clever experiments to isolate the mutants. Each mutation told a story.

I wanted to learn how every geneticist had made his or her mutants. After all, bacterial genetics is the study of genes that make up bacteria. A bacterium is defined by its unique properties or characteristics, called phenotypes. Examples of phenotypes for bacteria include virulence or susceptibility to specific drugs. The best way to specifically define a phenotype is to contrast it to what it is not. Since phenotypes are derived from genes, mutations in genes that cause a loss or a change in a specific phenotype actually help to define the gene that encodes that phenotype. Thus, virulence of a bacterium can be defined by contrasting it with a mutant that has lost its virulence. Drug susceptibility of a bacterium is defined by contrasting it with a mutant that is drug resistant.

Bacterial genetics is really the study of mutants, as mutations cause a phenotype to be recognized. Mutations that cause altered phenotypes occur, but at low frequencies. The beauty of studying a bacterial population is that one can routinely deal with populations of a billion, a number at which mutational events usually occur. I don't think I really understood Darwin till the first day I walked down 5th Avenue in New York City. For in walking down 5th Avenue, I realized I was among the numbers of people in which variations could be readily observed. When you see the seemingly unfathomable numbers of people, you realize the existence of great variation. For example, you know that there has to be someone wearing Christmas tree socks. It doesn't matter what time of year—you know there has to be *someone* wearing Christmas tree socks. Darwin says variation is an essential part of evolution, and New York City, like a bacterial population, has variation in its population of individuals. The essential Darwinian truth is that in a population of bacteria—or individuals walking on 5th Avenue—all sorts of different types of people exist there before you ever decide to walk there. The variations pre-exist before you decide to observe them. The goal of a geneticist, one who studies mutants, is to find those rare mutants. This leads to the second Principle.

Principle 2. The secret to winning in bacterial genetics is to learn how to win at bacterial lotto. My first attempt at being a geneticist came when I was 8 years old when I started to collect Lincoln cents. My goal was to obtain one of each of the various types of Lincoln cents minted between 1909 and 1963. Mom and Dad gave me $25.00 to use, which I took to the bank and exchanged for 2,500 pennies. I took them home and examined the date and mintmarks on every cent. I would pull out the rare cents and exchange them for common cents. It's difficult to explain why this process is a therapeutic and tranquil exercise, but if you have ever enjoyed performing a mutant hunt of bacterial colonies with toothpicks, we share a kindred spirit. It must have been enjoyable because I returned to a different bank every day to get another 2,500 pennies. I soon had the whole neighborhood doing it. We'd sit by the pool, going through pennies. However, after I had made 207 trips to the bank I came to understand then a very cold and true reality—mutants are rare. I had gone through over 500,000 pennies and I still did not have my 1914-D or my 1909-S VDB. I realized that to find these rare mutants, I would have to go through a large number—greater than a billion pennies. Have you ever thought how big a number 1 billion is? True, we have surpassed a population of 6 billion individuals on planet earth, but I finally did the calculation. If I had examined a penny a second without sleeping for a billion seconds, I would have aged 31.7 years. I realized then that finding a 1914-D in a bank was like winning lotto.

Yet, to be a geneticist, you have to study rare mutants.

Thus, to be a geneticist you must win at bacterial lotto. So how do you win at lotto? Selection. You must find a way to select or screen for the rare bacterial mutant or a rare bacterial event. Luria and Delbrück found out how to win at bacterial lotto (Luria and Delbrück, 1943). They used phages to select for the rare phage-resistant mutants present at 1 in 1,000,000 cells. In fact, they realized they could use this method again and again to screen large numbers of different populations to prove Darwin was right. Lederberg and Tatum (1946) used the inability of two mutations to revert spontaneously at frequencies less than 1 trillion (it's sobering to consider the U.S. national debt is expressed in trillions) to discover that bacteria had sex with one another. Griffith's observation of transformation (1928) was sufficiently efficient to allow Avery, MacLeod, and McCarty to discover DNA was genetic material (Avery et al., 1944). Zinder and Lederberg, upon trying to observe conjugation in bacteria, discovered transduction—another win at bacterial lotto (Zinder and Lederberg, 1952). This theme is played over and over again in most genetic discoveries. Selecting for the rare mutants is the key to winning at lotto. Once isolated, the rare mutants are the keys to understanding bacterial phenotypes as it leads to the next principle.

Principle 3. To prove that a phenotype in a mutant bacterium, such as drug resistance or virulence, is caused by a specific genotype, it is necessary to (i) isolate a mutant with a defined altered phenotype; (ii) clone the genotype from the mutant; and (iii) upon introducing the cloned genotype into a wild-type bacterium, reproduce the same phenotype of the mutant bacterium. Contrast reveals the existence of a characteristic, a property, or a phenotype. Since most phenotypes are complex characteristics, mutations in a number of different genes can cause a similar phenotype. By identifying the specific genes, we can begin to unravel a complex phenotype. The fulfillment of Koch's Molecular Postulate simply results in the generation of strains that are different by a single genetic difference. By constructing such strains, it is possible to conclude that the phenotype is caused by a specific genotype. The acquisition of this knowledge is an essential step in developing a strategy to attack the phenotype. For example, the knowledge of a drug target is essential to develop drugs that can attack drug-resistant mutants. Similarly, knowledge of the genes required for pathogenesis should lead to a rational strategy for developing novel live-cell vaccine strains.

Principle 4. In a genetic experiment, you always get what you select for. It may not be what you want, but you always get what you select for. One of the greatest thrills in science is making a discovery. A great experiment often reveals surprises you never imagined—Nature's reward to the experimental scientist.

Principle 5. Imagination is more important than knowledge. Albert Einstein made this conclusion as he reasoned that knowledge tells us what *was* and imagination tells us what *will be*.

These simple principles are a framework which I have found useful in thinking about genetic strategies. I hope they provide an entertaining and useful framework for the reader as I describe some of my lab's discoveries as a result of applying genetic approaches to M. *tuberculosis*.

DEVELOPMENT OF TRANSFORMATION SYSTEMS FOR MYCOBACTERIA

Having established the need for gene transfer for providing the knowledge, one has to ask, why wasn't gene transfer developed for M. *tuberculosis* or its close relative BCG before now? It wasn't because investigators had failed to try (Bloch et al., 1959; Tarnok and Bonicke, 1970). It is likely that many other investigators failed to report failed experiments.

M. *tuberculosis* is a difficult organism to work with. First, it is a pathogen that requires Biosafety Level 3 containment, as aerosols could transmit infectious particles. Second, M. *tuberculosis* grows very slowly, requiring 3 to 4 weeks to form colonies from single cells. Often mutant cells require significantly more time to grow than wild-type cells. (We have some mutants that require 9 weeks to form colonies from single cells. I never knew how bad my aseptic technique was until I started working with these bacteria.) Since each transformation attempt would take at least 1 month to obtain results before analysis, developing a transformation system was a highly unattractive project for any graduate student who wanted to complete a doctoral thesis in a reasonable time. Third, M. *tuberculosis* is covered with a unique lipid-rich cell envelope which is difficult to transform with DNA and also causes the cells to clump, thereby making it difficult to obtain colonies from single clones. Last, since not even a fast-growing *Mycobacterium* had ever been transformed, no genetic tools had been developed for M. *tuberculosis*.

Phages, viruses that infect bacteria, played a key role in the development of molecular biology and genetics for *E. coli*. We reasoned that mycobacteriophages could play a similar role for mycobacteria. The first mycobacteriophage was isolated by Gardner and Weisner in 1947, using *Mycobacterium smegmatis* as a host. Froman reported the first infection of M. *tuberculosis* by the mycobacteriophage D29 in 1954 (Froman et al., 1954). Numerous groups subsequently reported the isolation of phages that infected both M. *smegmatis* and M. *tuberculosis*. Tokunaga described the first transfection system for M. *tuberculosis* in 1967 (Tokunaga and Nakamura, 1967), and the first transduction in M. *smegmatis* was reported in 1970 (Sundar Raj and Ramakrishnan, 1970). Despite these successes, attempts at transducing genetic material into M. *tuberculosis* were universally unsuccessful. Armed with the knowledge that a large variety of phages existed which infected M. *tuberculosis* and the belief that phage infection is the most efficient means of delivering DNA into any bacterial cells, we set out to develop genetic systems based on these mycobacteriophages.

GENERATION OF SHUTTLE PHASMIDS: MYCOBACTERIOPHAGE CLONING VECTORS

Fulfilling Koch's Molecular Postulate was our goal. Mutants of M. *tuberculosis* existed and, by using recombinant DNA technology, mycobacterial DNA could be easily cloned into *E. coli*. The limitation was how to deliver this cloned foreign DNA into mycobacteria. While cloning into plasmids might have been the most direct route, no one had ever described transformation of any plasmid into any mycobacteria, particularly M. *tuberculosis*.

However, we reasoned that we could make phage-cloning systems using mycobacteriophages. We had collected a large number of phages from Wilbur Jones at the Centers for Disease Control and from our own soil samples (e.g., Bxb1 came from my back yard in the Bronx). DNAs were isolated from these phages and analyzed to test whether the phages packaged by a cohesive-end mechanism or by headfull packing. Using phage DNAs, we developed a very robust transfection system for M. *smegmatis* protoplasts (Jacobs et al., 1987). The goal was then to introduce foreign DNA into a mycobacteriophage genome and obtain replication of that foreign DNA in a mycobacterial cell.

This is where the principles of genetics come into play. I *imagined* that somewhere within a mycobacteriophage there had to exist a nonessential region into which foreign DNA could be inserted. To find that region, I reasoned I could *duplicate the New York City scenario* and introduce an *E. coli* plasmid into every site within the mycobacteriophage genome. If such a library could be made, I reasoned we could *win at lotto* by transfecting the library of plasmid insertions into M. *smegmatis* and looking for plaques. The only recombinant molecules that should yield a plaque would be those into which the *E. coli* plasmid had been inserted into a nonessential region of the phage.

A key observation that had made this approach possible was that mycobacterial DNA is generally not expressed in *E. coli*, since the promoters of organisms with high guanosine and cytosine content generally are not recognized by *E. coli* RNA polymerase. (This observation I painfully learned from five unsuccessful years of attempting to express *Mycobacterium leprae* genes in *E. coli* for my Ph.D. studies [Clark-Curtiss et al., 1985; Jacobs et al., 1986a, 1986b].) I reasoned that we could clone an entire mycobacteriophage genome into an *E. coli* plasmid and the resulting recombinant would not kill *E. coli*. Such a construct could not be easily made with *E. coli* or *Salmonella* phages because the phage genomes would be expressed, resulting in the synthesis of genes that would be toxic to the *E. coli*. In contrast, the mycobacteriophage would be an inert DNA molecule. To our delight, the construction of a library of partially digested mycobacteriophage DNA into an *E. coli* cosmid was readily achievable. This strategy had generated a *New York City scenario*, and we transfected this library into M. *smegmatis* protopasts to *win at lotto*.

I vividly remember the day I succeeded in transfecting this library into M. *smegmatis* protoplasts, which yielded plaques. Very pleased with my own brilliance in achieving my anticipated results, I was humbled by Nature's wonder when the first 10 phage plaques I examined all yielded wild-type phage with no plasmid insert. At that moment, I came to appreciate that *in a genetic experiment you always get what you select for. It may not be what you want, but you always get what you select for.*

What I had failed to anticipate was that recombination can occur in mycobacterial cells. Transfection of the library likely resulted in the introduction of numerous DNA molecules into every cell, and wild-type phage could easily be regenerated if the region of the plasmid insertion in one molecule was replaced by the wild-type chromosomal DNA fragment from the other. This hypothesis was found to be true, when I screened the plaques for the presence of the *E. coli* plasmid and found that only 1 in 400 of the plaques contained the *E. coli* plasmid. Eureka! I knew this one plaque represented a novel chimeric vector, which I named a shuttle phasmid. These molecules were part phage, part plasmid, and could shuttle between *E. coli* and mycobacteria. Not only could their DNA be readily introduced into M. *smegmatis* protoplasts, but once packaged into mycobacteriophage particles they could be readily used to infect a diverse set of mycobacterial strains including BCG and M. *tuberculosis*. Thus, foreign cloned DNA was introduced and replicated in both fast- and slow-growing mycobacteria for the first time (Jacobs et al., 1987).

These shuttle phasmids provided a means to systematically develop a transformation system for mycobacteria. Since the same DNA molecule could be propagated in either *E. coli* or mycobacteria, we could ask the question whether mycobacterial DNA was restricted when introduced into mycobacteria. The *E. coli*-propagated DNA was not restricted, since shuttle phasmid DNA from *E. coli* yielded a comparable number of plaque-forming units as shuttle DNA isolated from M. *smegmatis*-propagated shuttle phasmids. Therefore the failure to obtain transformants was not because the DNA was being restricted.

This work had been done with the phage TM4 obtained from Patrick Brennan's group (Timme and Brennan, 1984). I had chosen this phage because it shared common properties with bacteriophage lambda—it was 50 kb in length, possessed cohesive ends, and was thought to be a temperate phage of *Mycobacterium avium*. Although it does form slightly turbid plaques, we have never been able to demonstrate that it could lysogenize any mycobacteria. (The property of being able to lysogenize mycobacteria has proven to be exceedingly fortuitous and allowed us to develop specialized transducing phages which will be discussed later.)

I reasoned we could use such a phage to stably introduce foreign DNA into *E. coli*. Thus, I began screening all the phages I had for the hallmark characteristic of a temperate phage—the ability to site-specifically integrate into a mycobacterial chromosome. I found two highly related phages that possessed this property: L1 and L5. Based on this property, Graham Hatfull reasoned that one of these phages represented the optimal phage species to determine the entire sequence of L5, which has provided an incredibly rich source of tools and knowledge for the molecular biologist (see below). Using the L1 phage, Scott Snapper, a shared graduate student of Barry Bloom's and mine, was successful in constructing the first temperate shuttle phasmids. These shuttle phasmids, in addition to being able to replicate in *E. coli* as plasmids and in mycobacteria as phages, could stably introduce foreign DNA into a mycobacterial chromosome. Using these shuttle phasmids, we demonstrated that we could introduce and express a foreign gene conferring kanamycin resistance into mycobacteria for the first time. These shuttle phasmids provided the knowledge needed to develop the first transformation systems for mycobacteria (Snapper et al., 1988).

M. SMEGMATIS mc^{2}155: A SURROGATE HOST FOR ANALYSIS OF *M. TUBERCULOSIS GENES*

M. *smegmatis* has been an attractive model organism to work with because it can readily be grown in simple media and has been the preferred host for phage infections. When I first obtained ATCC 607, designated mc^{2}1 in my culture collection, we discovered it was a mixed population that gave three unique colonial morphologies when plated on the *Streptomyces* Regeneration Media R5. (In case you're wondering about the origin of the name mc^2, it's in honor of Albert Einstein.) The predominant colonial morphotype was an orange rough colonial morphology designated mc^{2}6. The two other morphotypes were orange smooth and white rough. All three colonial morphotypes gave rise to their respective morphotypes, and thus I stocked them as clones. Subsequent work was all performed with the predominant morphotype—the orange rough morphotype designated mc^{2}6. This strain retained all the properties that had been ascribed to ATCC 607 and thus was an attractive experimental organism.

Armed with the knowledge that (i) mc^{2}6 possessed no ability to restrict *E. coli*-propagated DNA, (ii) the kanamycin resistance gene functioned in mc^{2}6, and (iii) kanamycin selection did not give rise to high numbers of spontaneous kanamycin-resistant mutants, we reasoned we should be able

TABLE 2 Surrogate hosts for analysis of M. *tuberculosis* genes

Characteristic	*E. coli*	M. *smegmatis*	BCG
Time required to form colonies	8 h	3 to 4 days	3 to 4 weeks
Recognizes M. *tuberculosis* promoters	Not usually	Yes	Yes
Glycosylates M. *tuberculosis* proteins	Not usually	Yes	Yes
Second SecA pathway	No	Yes	Yes
Susceptible to INH, ETH, and ETB[a]	No	Yes	Yes
Possesses mycolic acids	No	Yes	Yes
Replicates M. *tuberculosis* phages	No	Yes	Yes
Grows at 42°C	Yes	Yes	No
Containment required when cloning in M. *tuberculosis* genes	BSL2[b]	BSL2[c]	BSL3
Replicates pAL5000 episomes	No	Yes[c]	Yes
Integrates L5 *attP* vectors	No	Yes	Yes

[a] ETB, ethambutol.
[b] BSL2, Biosafety Level 2.
[c] Refers specifically to mc^2155 because it is an efficient plasmid transformation mutant and mc^2155 containing 40-kb inserts of M. *tuberculosis* DNA fragments was shown to be avirulent in SCID mice.

to develop a plasmid transformation system. Tobias Kieser, of the John Innes Institute, had made a marvelous library in which he had randomly incorporated an *E. coli* plasmid containing a kanamycin resistance gene into the plasmid pAL5000. Plasmid pAL5000 was particularly attractive, as Brigitte Gicquel's laboratory had completed the DNA sequence (Rauzier et al., 1988) and generously provided it. Tobias generously shared his library with us, and we attempted to obtain transformants in M. *smegmatis*.

Despite numerous attempts, however, we were unsuccessful in ever obtaining transformants following the introduction of the library into protoplasts of mc^26 under conditions with which we knew we were achieving reasonably high levels of transfection. We reasoned that there were two possibilities for our failures: either the protoplasts were unable to regenerate, or the plasmid failed to replicate in mc^26. To eliminate the regeneration variable, we took advantage of a newly developed methodology—electroporation. After a few unsuccessful attempts, Scott Snapper was able to obtain three independent transformants following electroporation with the pAL5000::Kan library. We *imagined* that we had *won at prokaryotic lotto* and obtained a mutant that could allow for replication of the pAL5000 replicon. To test this hypothesis, we simply grew the transformant in nonselective media to cure the plasmid. A clone of the cured strain, designated mc^2155, was electroporated with the plasmid that had been retrieved from our initial transformant. To our delight, mc^2155 routinely yielded $>10^4$ transformants per μg of DNA while mc^26 yielded no more than a few, all of which always had acquired the efficient plasmid transformation (*ept*) phenotype.

While improved protocols have allowed researchers to routinely obtain $>10^7$ transformants per μg of plasmid DNA, the nature of the mutation still remains a mystery (Snapper et al., 1990). We do know that the *ept* mutation does not (i) debilitate a restriction system, (ii) improve DNA uptake abilities of the cell, nor (iii) improve integration of homologous DNA substrates. Thus it likely affects replication of the pAL5000 replicon. No matter what the reason is for its efficient plasmid transformation phenotype, it makes M. *smegmatis* a highly attractive surrogate host for the analysis of mycobacterial genes and particularly those of M. *tuberculosis*.

The M. *smegmatis* strain mc^2155 has proven to be a workhorse for the molecular analysis of genes from mycobacteria and mycobacteriophages. First, it grows relatively fast (approximately 10 times faster than M. *tuberculosis* or BCG), yielding colonies in 2 to 4 days compared to 3 to 4 weeks for M. *tuberculosis*. M. *smegmatis* mc^2155 grows well-dispersed in media with Tween and thus is an excellent host for phage propagations. In addition, it grows on a simple defined medium consisting of glucose and salts or any number of media commonplace to microbiological laboratories. Unlike M. *tuberculosis*, M. *smegmatis* is nonpathogenic and can be safely used in a standard Biosafety Level 2 laboratory. The recent demonstration that mc^2155 containing cosmid-size genomic fragments of M. *tuberculosis* does not cause disease in immunocompromised mice has made it safe to analyze M. *tuberculosis* genes in Biosafety Level 2 containment facilities.

As a genetic host, mc^2155 is readily transformable with integration-proficient plasmids and pAL5000-based episomal plasmids, and it readily undergoes homologous recombination, allowing for the generation of targeted gene disruptions. Numerous studies have shown that genes from M. *tuberculosis* and M. *leprae* are readily expressed in M. *smegmatis*, allowing for the analysis of mycobacterial promoters and expression sequences. Past studies as described in chapter 4 have shown that auxotrophic mutants can be readily isolated in M. *smegmatis* and that genes can be transferred from one strain to another by either conjugation or transduction. Furthermore, numerous studies have shown that mutations in the genes of M. *smegmatis* can be readily complemented with genes from M. *tuberculosis* or M. *leprae*, thus proving mc^2155 to be a valuable surrogate host. Importantly, multicopy expression of M. *tuberculosis* genes in M. *smegmatis* has provided a means to identify the previously unknown targets of isoniazid (INH) (Banerjee et al., 1994), ethionamide (ETH) (Banerjee et al., 1994), ethambutol (Belanger et al., 1996; Telenti et al., 1997), and pyrazinamide (Zimhony et al., submitted). Clearly, M. *smegmatis* will continue to play an important role in the analysis of myco-

bacterial genes, and the determination of the genome sequence of mc²155 (recently initiated at The Institute for Genome Research) will greatly enhance future studies of mycobacteria.

GENETIC STRATEGIES FOR UNVEILING MECHANISMS FOR ELUCIDATING DRUG ACTION AND DRUG RESISTANCES

The Pathogen-Drug Analogy

A disease-causing bacterium is to a host what a drug is to a bacterium. A disease-causing bacterium enters a host and interacts with host cells, eliciting a number of events which eventually cause disease and possibly the death of the host. Similarly, a drug enters a bacterium and binds to a target, which initiates a set of events that will possibly lead to the death of the bacterium. M. *tuberculosis* enters a human host, usually by inhalation, invades a susceptible cell, multiplies, and causes a disease state that can lead to death. For it to cause disease, the M. *tuberculosis* cell must possess a number of properties that enable it to be effective, such as the ability to invade a mammalian cell, the ability to multiply in the mammalian cell, and the ability to avoid immune killing functions. The pathogen's success is also dependent upon which host the cell enters. Some hosts may be naturally resistant to a pathogen or may possess a genetic trait that precludes some process required for the pathogen's success.

Most of modern-day molecular pathogenesis is focused on defining the specific properties of pathogens by identifying the pathogen's functions by means of generating mutants that have lost a specific function. The knowledge of drug action, although much simpler, is analogous to understanding the pathogenic properties of the pathogen. Drugs first enter cells and then bind to a target molecule that is usually an enzyme. Once the drug is bound, certain events have to occur before the pathogen's cells die. Some drugs such as penicillin and INH are inactive on nondividing cells and are only active on actively dividing cells. In addition, just as a host may have a genetic trait that prevents expression of a pathogenic property, bacterial cells might have a mutation that confers resistance to the killing properties of a particular drug.

Resistance to a drug is a phenotype that can be mediated by a number of different genotypes. Resistance can be mediated by different mutations in the same gene, but also by mutations in different genes. For example, the gene encoding the target can acquire a mutation that causes a change in the target so that the target no longer binds the drug. Alternatively, the gene encoding the target can acquire mutations that cause it to be overexpressed, and this results in a titration of the killing action of the drug. If the drug requires an enzymatic activity to be activated, mutations that alter this activity would mediate resistance. Alternatively, resistance can be mediated by a mutation that affects the transport of the drug into the cell. The knowledge of the various genotypes that mediate drug resistance defines the different gene products that function in the drug action and drug resistance. Knowledge of the functions encoded by these gene products can lead to knowledge of (i) the drug's target, (ii) the gene functions that interact with the drug, and (iii) the molecular events that lead to the death of a bacterial cell. This knowledge is essential to develop strategies to better kill a bacterial cell and overcome resistance mechanisms.

Identifying Drug-Resistance Alleles by Analogy

The discovery of antibiotics and their application to treat bacterial infections represents one of the most successful accomplishments in the history of medical research. However, bacteria are masters of the New York City and Winning at Lotto principles—it's Darwinian genetics. Bacteria have repeatedly shown an amazing ability to adapt to the lethal effects of antibiotics by developing mutations conferring drug resistance. M. *tuberculosis* has been no exception, as reviewed in chapter 15. Drug resistance can occur by many different mechanisms, all of which elucidate some characteristic of the drug action. Indeed, drug resistance is a phenotype that must be defined by a mutated genotype, and to prove that this genotype is associated with drug resistance, Koch's Molecular Postulate must be fulfilled. Therefore, most of our knowledge of drug action comes from studying model organisms for which genetic systems exist and biochemistry can be readily performed.

Drug resistance is most often mediated by mutations that cause amino acid substitutions in the enzyme target that binds the drug. Since drug targets are often essential enzymes that are highly conserved molecules, specific mutations could be concluded to be associated with drug resistance based on analogy. The first mutations in M. *tuberculosis* found to be associated with drug resistance all were identified by analogy to mutations found to associate with drug resistance in *E. coli*, including resistance to rifampin, streptomycin, and fluroquinolones (see chapter 15). It is interesting to note that resistance to streptomycin could be mediated through mutations in the rRNA in yeast mitochondria, an organelle that has a single rRNA. Since M. *tuberculosis* was found to have a single rRNA gene, several groups postulated, and observed, that M. *tuberculosis* could mediate streptomycin resistance by mutations that mapped to the gene encoding the rRNA (see chapter 4). This was subsequently verified by gene transfer experiments in M. *smegmatis* (Sander et al., 1996).

Importantly, since bacterial genes can have natural polymorphisms, the demonstration that specific alleles are associated with drug resistance requires that a mutated allele be transferred and that the resistance genotype be associated with the drug resistance phenotype. To date, Koch's Molecular Postulate has never been fulfilled for any point mutation for any drug resistance determinant directly in M. *tuberculosis*, due to the difficulty in performing allelic exchanges in M. *tuberculosis*. Surprisingly, the number of polymorphisms found in M. *tuberculosis* genes has been significantly lower (100 to 1,000 times lower) than in other organisms (Sreevatsan et al., 1997), and thus the identification of a polymorphism often reflects a selected phenotype. This observation clearly strengthens, but does not prove, that a mutation causes a drug resistance phenotype.

Identification of Genes Conferring INH Resistance by Fulfilling Koch's Molecular Postulate

Isoniazid (INH) is a very simple molecule that was found in 1952 to be highly active against M. *tuberculosis* at MICs of 0.01 to 0.1 μg/ml (Bernstein et al., 1952; Fox, 1952; Steenken and Wolinsky, 1952). Despite its simple structure, its mechanism of action had been impossible to elucidate in the absence of being able to fulfill Koch's Molecular Postulate, as INH is inactive on genetically tractable organisms such as *E. coli* or *Bacillus subtilis*. Indeed, numerous studies

had been published which suggested that INH-resistant mutants were readily isolated in the laboratory. In fact, it is interesting that in the very same issue of the *American Review of Tuberculosis* that reports the discovery of antituberculosis action of INH, there are three brief reports describing the isolation of INH-resistant mutants of *M. tuberculosis* and *Mycobacterium bovis* (Middlebrook, 1952; Panzy et al., 1952) and *Mycobacterium ranae* (Szybalski and Bryson, 1952). In retrospect, INH action might have been predicted to be complex, as early studies revealed great difficulty in isolating and maintaining subcultures of *M. tuberculosis* from INH-treated patients (Barnett et al., 1953; Collard et al., 1953; Middlebrook and Cohn, 1953). Notably, in the Middlebrook and Cohn study, 8 of 17 strains isolated from concentrated sputum from INH-treated patients failed to grow on the relatively minimal OA medium, while the remaining strains grew well, but were attenuated for growth in guinea pigs (Middlebrook and Cohn, 1953). Thus, these early studies revealed that INH resistance is often associated with a number of complex phenotypes that were difficult to analyze without knowing that a specific phenotype was caused by a specific genotype.

katG and INH Resistance

Middlebrook was the first to report that mutants of *M. tuberculosis* resistant to high levels of INH were defective for catalase-peroxidase activity (Middlebrook, 1954). These were from the class of INH-resistant mutants isolated from INH-treated patients. They retained their ability to grow on OA agar medium, but were attenuated for growth in guinea pigs. This study suggested that a mutation that conferred INH resistance could also confer loss of catalase activity. However, the proof that a loss of catalase activity conferred INH resistance would have to wait 38 years until Zhang et al. were able to clone the *katG* gene of *M. tuberculosis* by using PCR primers that shared homology with the *katG*-encoded catalase-peroxidase gene of *E. coli* (Zhang et al., 1992). Zhang et al. were able to perform a gene transfer and transfer INH susceptibility to an INH-resistant variant of mc^{2}155. In addition, they observed that two clinical isolates of *M. tuberculosis* that were resistant to high levels of INH had deletions of their *katG* genes (Zhang et al., 1992). In a later study, these investigators were able to transform *M. tuberculosis katG* mutants and showed that the transformants were returned to wild-type levels of susceptibility to INH (Zhang et al., 1993). These studies confirmed that the INH resistance phenotype was caused by a specific genotype, a mutation in the *katG* gene. These studies further supported the hypothesis of Winder (Winder et al., 1970) that INH is a prodrug that becomes activated by the catalase-peroxidase and that this activated INH inhibits some target enzyme. The inhibition of this enzyme then leads to the death of the mycobacterial cells.

inhA, a Gene Encoding a Target of INH

Early reports had suggested that many INH-resistant isolates of *M. tuberculosis* were found to be resistant to ethionamide (ETH), even though the patients had never received ETH. This suggested that INH and ETH shared a common target enzyme. To identify the gene encoding this enzyme, Asesh Banerjee isolated a mutant of *M. smegmatis*, mc^{2}651, that was resistant to both INH and ETH. (Again, this was the *New York City principle* as these mutants preexisted in the population at 1 per 10^8 cells, and we *won at lotto* by selecting for the mutants on plates with INH and ETH.) We identified a DNA fragment from mc^{2}651 that conferred coresistance to both INH and ETH when cloned on a multicopy plasmid into mc^{2}155. Further analysis revealed that a single gene, which we named *inhA*, conferred both INH and ETH resistance when cloned on a plasmid (Banerjee et al., 1994). Sequence analysis of the *inhA* gene from mc^{2}651 revealed the presence of a single mutation compared to the *inhA* gene from the parent mc^{2}155 strain. The wild-type allele (from the parent mc^{2}155 in this case) is designated *inhA*, and the mutant allele (from mc^{2}651) is designated *inhA1*. The *inhA1* allele caused a serine-to-alanine substitution at amino acid 94 of the InhA protein. It was tempting to conclude that this mutation was responsible for both INH and ETH resistance phenotypes, but this could not be readily concluded because the parental *inhA* gene conferred INH and ETH resistance when cloned on a multicopy plasmid vector.

This points out the necessity to rigorously execute all three criteria of Koch's Molecular Postulate, which specifies that a specific genotype must be transferred if we are to conclude that the genotype is responsible for a phenotype. The specific genotype in this case was a point mutation on the chromosome on mc^{2}651—the *inhA1* allele. To determine whether the *inhA1* allele was the cause of the INH and ETH resistance, it was necessary that the cloned *inhA1* allele be introduced and exchanged for the wild-type *inhA* allele in mc^{2}155, with the resulting recombinants becoming INH and ETH resistant. Alternatively, the reciprocal experiment—the transfer of the *inhA* allele into mc^{2}651, with conversion from INH and ETH resistance to INH and ETH susceptibility—would also prove that the *inhA1* allele confers INH and ETH resistance.

Unfortunately, the frequency of obtaining transformants with chromosomal DNA (1 in 10^6 cells) was less than the frequency of isolating spontaneous INH-resistant mutants (1 in 10^5 cells). Thus, to ensure that we could analyze allelic exchanges and not spontaneous mutations, a cotransformational analysis was performed where we physically linked a kanamycin resistance marker gene to a chromosomal DNA fragment that contained the wild-type *inhA* gene. Transformation of this fragment, followed by selection for kanamycin resistance, occurred at a frequency of 10^{-6}, compared to the frequency of spontaneous kanamycin resistance, which was less than 10^{-7}. In numerous experiments, we observed that transformation of mc^{2}651 yielded tens to hundreds of kanamycin-resistant colonies. Southern analysis revealed that the colonies had acquired the kanamycin resistance gene. In *M. smegmatis*, this preferentially occurs by a homologous recombination event, particularly using long linear recombination substrates. Since the kanamycin resistance gene was closely linked to the wild-type *inhA* gene, a large fraction of the recombinants would also acquire the unselected *inhA* allele following a double crossover event. In fact, 70 to 80% of the kanamycin-resistant colonies became susceptible to both INH and ETH, indicating that the crossover to incorporate the kanamycin resistance gene had also incorporated the *inhA* allele. This was readily proven to result from the acquisition of the wild-type *inhA* allele, as the *inhA* and *inhA1* could be readily distinguished by the acquisition of a unique restriction enzyme site on a PCR product. There was always 100% correlation of the acquisition of the restriction enzyme site (and hence the point mutation) with the conversion from INH and ETH resistance to susceptibility (Banerjee et al., 1994). Thus,