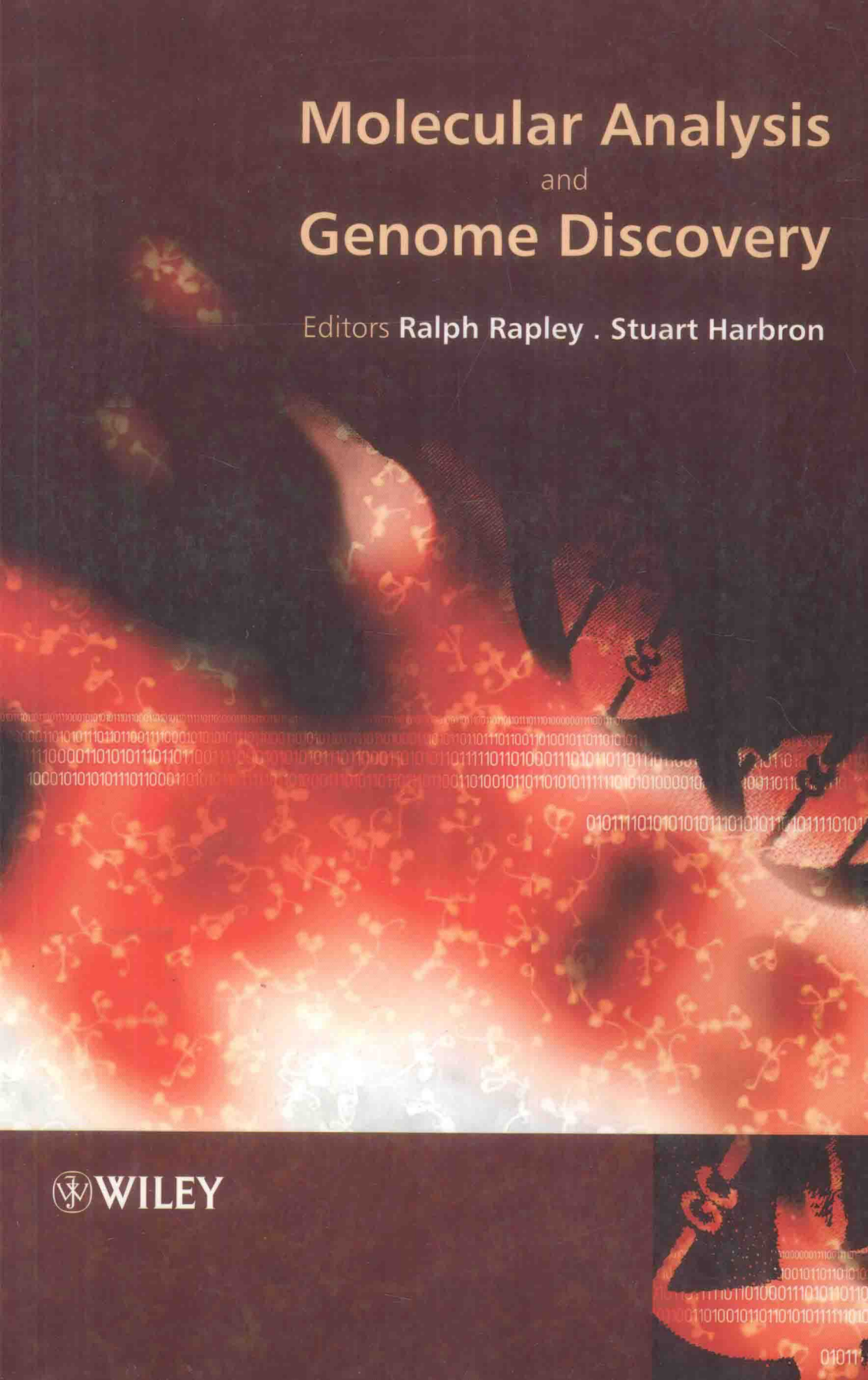


Molecular Analysis and Genome Discovery

Editors **Ralph Rapley . Stuart Harbron**

 **WILEY**



Molecular Analysis and Genome Discovery

Edited by

Ralph Rapley

University of Hertfordshire, UK

and

Stuart Harbron

The Enzyme Technology Consultancy, UK



JOHN WILEY & SONS, LTD

Copyright © 2004

John Wiley & Sons Ltd,
The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England
Telephone (+44) 1243 779777
Email (for orders and customer service enquiries):
cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com
or www.wiley.com

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770620.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons Inc.,
111 River Street, Hoboken, NJ 07030, USA
Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA
Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany
John Wiley & Sons Australia Ltd,
33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd,
2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd,
22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Molecular analysis and genome discovery/edited by Ralph Rapley and Stuart Harbron.
p.; cm.

Includes bibliographical references and index.

ISBN 0-471-49847-5 (hardback : alk. paper) – ISBN 0-471-49919-6 (pbk. : alk. paper)

1. Molecular diagnosis. 2. Genomics. 3. Proteomics. 4. Pharmacogenomics. 5. Polymerase chain reaction.
6. DNA microarrays.

[DNLM: 1. Pharmacogenetics—methods. 2. Biological Markers. 3. Drug Design. 4. Genetic Techniques.
5. Pharmaceutical Preparations—metabolism. 6. Quantitative Structure-Activity Relationship. QV 38 M717
2004] I. Rapley, Ralph. II. Harbron, Stuart.

RB43. 7. M595 2004

615'. 7—dc22

2004001166

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0 471 49847 5 hardback
0 471 49919 6 paperback

Typeset in 10.5 on 13pt by Kolam Information Services, Pvt. Ltd, Pondicherry, India.

Printed and bound in Great Britain by TJ International, Padstow, Cornwall

This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Molecular Analysis and Genome Discovery

Preface

There can be no doubt that the face of diagnostics and drug discovery has changed beyond recognition over the last decade. Advances in the techniques of molecular analysis, and the ever-increasing use of automation, especially high throughput approaches, have paved the way for new, rapid and more reliable diagnostic tests. Completion of the Human Genome project has, in addition, been a startling achievement that can only accelerate the discovery of new genes and biomarkers for disease processes.

The genomics revolution promises to replace current prescription medicines with a new class of much more potent and efficacious medicines. The success of this revolution, the process of converting the claims of personalized medicine into an armamentarium specific for each patient, is dependent on continuing advances in molecular analysis and genome discovery.

This book aims to bring together these two aspects and show how the rapid advances in molecular analysis are bringing new tools and approaches to genome discovery. In combination, these provide the impetus for the development of new diagnostic techniques.

Following a historical overview of pharmacogenetics and pharmacogenomics by Werner Kalow, the first part of the book has detailed explanations of the latest techniques from experts in the field, copiously illustrated with examples from their own laboratories.

Chapters from Jörg Dötsch and Elaine Lyon present alternative approaches to quantitative and real-time PCR techniques and how they may be used in tumour and mutation detection.

Ivo Gut covers Genotyping and SNP analysis, and issues such as improving the economy of this approach are explored. Emerging techniques and the promise of high throughput SNP genotyping methods are presented.

Paal Andersen and Lars Larsen provide a comprehensive overview of mutation detection and screening. Methods for mutation scanning are also discussed.

Pyrosequencing is an exciting new approach to genomic analysis, and its principles and applications are explained and illustrated by Elahe Elahi and Mostafa Ronaghi.

Development of DNA-based microarrays have been a key area over the last decade, both from a research point of view and commercially. Their fabrication

and use are expertly covered in chapters from Magdalena Gabig-Ciminska and Andrzej Ciminski, and from Janette Burgess; a chapter from Jon Terrett's group focuses on the complementary chip-based proteomic technology.

Ciara O'Sullivan *et al.* discuss the powerful approach enabled by aptamer technology. Uniquely she relates how aptamers may be selected for use in both proteomic *and* therapeutic applications.

Kin-Ying To describes how a number of these molecular techniques may be applied to an analysis of differential gene expression.

The remaining chapters focus on the latest approaches to the identification of new target compounds, and Roberto Solari provides an excellent overview of the multi-stage drug discovery process.

Huang and Jong provide an interesting insight into potentially powerful approaches to analysing global features of microorganisms as they relate to the elucidation of new antibacterial agents.

The discovery of new target compounds based on an analysis of known structural motifs and elements is covered by Kan *et al.* and David Winkler, while Hudes, Menon and Golemis contribute a chapter investigating the impact of drugs on protein–protein interactions.

In compiling *Molecular Analysis and Genome Discovery* we have sought to combine both current and emerging approaches to analysing DNA, proteins and genomes, whilst not losing sight of the challenging outcomes required for successful drug discovery. In achieving this aim we have benefited from calling on the expertise of a distinguished and creative group of contributing authors; we have thoroughly enjoyed working with them.

**Ralph Rapley
Stuart Harbron**

Contributors

Andersen, Paal Skytt

Department of Clinical Biochemistry, Statens Serum Institut, Artillerivej 5, DK-2300 Copenhagen, Denmark

Baldrich Rubio, Eva

Department of Chemical Engineering, Universitat Rovira i Virgili, Avinguda Paisos Catalans 26, 43007 Tarragona, Spain

Barry, Richard

Oxford Glycosciences, Abingdon Science Park, Abingdon, UK

Burgess, Janette K.

Respiratory Research Group, Department of Pharmacology, Bosch Building, D05, University of Sydney, Sydney, NSW, Australia

Chen-Chen Kan

Keck Graduate Institute of Applied Life Sciences, 535 Watson Dr., Claremont, CA, USA, and Eidogen, 1055 East Colorado Blvd, Suite 550, Pasadena, CA 81106, USA

Ciminski, Andrzej

Department of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 80-952 Gdansk Wrzeszcz, Poland

Debe, Derek A.

Eidogen, 1055 East Colorado Blvd, Suite 550, Pasadena, CA 81106, USA

Dötsch, Jörg

Department of Pediatrics, University of Erlangen-Nürnberg, Germany

Elahi, Elahe

Faculty of Science, University of Tehran, Tehran, Iran

Gabig-Ciminska, Magdalena

Department of Biotechnology, Royal Institute of Technology (KTH), S-10691 Stockholm, Sweden, and Laboratory of Molecular Biology (affiliated with the University of Gdansk), Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Kladki 24, 80-822 Gdansk, Poland

Golemis, Erica A.

Division of General Science, Fox Chase Cancer Center, 8801 Burholme Ave., Philadelphia, PA 19111

Gut, Ivo Glynne

Centre National de Génotypage, Bâtiment G2, 2 rue Gaston Crémieux, CP 5721, 91057 Evry Cedex, France

Hambly, Kevin

Eidogen, 1055 East Colorado Blvd, Suite 550, Pasadena, CA 81106, USA

Homs, Mònica Campàs i

Department of Chemical Engineering, Universitat Rovira i Virgili, Avinguda Paisos Catalans 26, 43007 Tarragona, Spain

Hudes, Gary

Division of Medical Science, Fox Chase Cancer Center, 8801 Burholme Ave., Philadelphia, PA 19111, USA

Jong, Ambrose

Children's Hospital, Los Angeles, and the University of Southern California, Los Angeles, CA 90027, USA

Kalow, Werner

Faculty of Medicine, University of Toronto, Department of Pharmacology, 1 King's College Circle, Toronto, M5S 1A8, Canada

Kin-Ying To

Institute of BioAgricultural Sciences, Academia Sinica, Taipei 115, Taiwan

Larsen, Lars Allan

Wilhelm Johansen Centre for Functional Genome Research, Department of Medical Genetics, IMBG, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen, Denmark

Lyon, Elaine

ARUP Laboratories, Department of Pathology, University of Utah, 500 Chipeta Way, Salt Lake City, UT 84108, USA

Menon, Sanjay

Morphochem Inc., Monmouth Junction, NJ 08852, USA

O'Sullivan, Ciara K.

Department of Chemical Engineering. Universitat Rovira i Virgili, Avinguda Països Catalans 26, 43007 Tarragona, Spain, and Institució Catalana de Recerca i Estudis Avançats, Passeig Lluís Companys 23, 08010 Barcelona, Spain

Rascher, Wolfgang

Department of Pediatrics, University of Erlangen-Nürnberg, Germany

Ronaghani, Mostafa

Stanford Genome Technology Center, Stanford University, 855 California Ave., Palo Alto, CA 9434, USA

Schoof, Ellen

Department of Pediatrics, University of Erlangen-Nürnberg, Germany

Sheng-He Huang

Children's Hospital Los Angeles, and the University of Southern California, Los Angeles, CA 90027, USA

Solari, Roberto

Apax Partners Ltd, 15 Portland Place, London W1B 1PT

Soloviev, Mikhail

Oxford Glycosciences, Abingdon Science Park, Abingdon, UK

Terrett, Jon

Oxford Glycosciences, Abingdon Science Park, Abingdon, UK

Winkler, David A.

CSIRO Molecular Science, Private Bag 10, Clayton South MDC, Clayton 3169, Australia

Contents

Preface	xi
List of contributors	xiii
1 Pharmacogenetics and Pharmacogenomics: An Overview	1
<i>W. Kalow</i>	
The origins of pharmacogenetics	1
The initial progress of pharmacogenetics	2
Molecular genetic methods enlarged pharmacogenetics	3
Multifactorial variation	7
New problems and opportunities	9
Summary and conclusions	12
References	13
2 Quantitative TaqMan Real-time PCR: Diagnostic and Scientific Applications	17
<i>Jörg Dötsch, Ellen Schoof and Wolfgang Rascher</i>	
Summary	17
Introduction	18
Principles of TaqMan real-time PCR	18
Reliability and validation of TaqMan real-time PCR	19
Applications for TaqMan real-time PCR	20
Limitations and pitfalls in the use of TaqMan real-time PCR	23
Alternative real-time PCR methods	24
Future developments	25
Conclusions	25
References	26
3 Hybridization Probes for Real-time PCR	29
<i>Elaine Lyon</i>	
Introduction	29
Probe chemistry	29
Fluorescence resonance energy transfer	30
Hybridization probe design	31
FRET applications	33

Summary	39
References	39
4 An Overview of Genotyping and Single Nucleotide Polymorphisms (SNP)	43
<i>Ivo Glynne Gut</i>	
Introduction	43
Genotyping of single nucleotide polymorphisms	44
Methods for interrogating SNPs	47
Analysis formats	53
How can SNP genotyping be made more economic?	
Miniaturization – Multiplexing	59
Haplotyping	59
DNA methylation analysis – analysis of methylation variable positions (MVPs)	60
SNP genotyping and quantitation	60
Emerging methods	61
High-throughput SNP genotyping methods	62
The next generation	62
Conclusions	63
Acknowledgements	63
References	64
5 High-throughput Mutation Screening	71
<i>Paal Skytt Andersen and Lars Allan Larsen</i>	
Introduction	71
Automated DNA sequencing for mutation analysis	72
Methods for detection of specific mutations	73
Methods for mutation scanning	87
References	98
6 Determination of Nucleic Acid Sequences by Pyrosequencing	101
<i>Elahe Elahi and Mostafa Ronaghi</i>	
Introduction	101
General principles	102
Methodology	103
Applications	104
Future perspectives	107
References	109
7 An Introduction to DNA Chips	113
<i>Magdalena Gabig-Ciminska and Andrzej Ciminski</i>	
Introduction	113
DNA chip structure and operating principles	115
Achievements and future research directions	122

Application and potential use of DNA chips	124
Concluding remarks	125
References	126
8 Overview of Microarrays in Genomic Analysis	127
<i>Janette K. Burgess</i>	
Introduction	128
What is a microarray?	129
History	129
Production of a cDNA microarray	131
Sources of cDNA	134
Production of an oligonucleotide microarray	135
Array substrates	138
Commercial sources of arrays	138
Isolation of RNA	139
Amplification	140
Target labelling	140
Hybridization	142
Image acquisition	143
Image analysis	146
Data storage	147
Reporting standards	149
Applications to genomic analysis	150
The future?	158
References	158
9 Overview of Differential Gene Expression by High-throughput Analysis	167
<i>Kin-Ying To</i>	
Introduction	167
Differential display	168
cDNA-amplified fragment length polymorphism	168
Representational difference analysis	172
Subtractive hybridization	172
Expressed sequence tags	173
Serial analysis of gene expression	175
DNA microarrays and plant functional genomics	177
Discussion	187
References	188
10 Aptamers: Powerful Molecular Tools for Therapeutics and Diagnostics	191
<i>Eva Baldrich Rubio, Mònica Campàs i Homs and Ciara K. O'Sullivan</i>	
Introduction	191
Aptamer selection	192

	Aptamers for therapeutics	195
	Aptamers in analysis	199
	Aptamers for proteomic applications	206
	Perspectives and future outlook	208
	Acknowledgements	209
	Abbreviations	209
	References	210
11	Chip-based Proteomics Technology	217
	<i>Mikhail Soloviev, Richard Barry and Jon Terrett</i>	
	Introduction	217
	Protein separation and purification approaches	219
	Chip-based proteomics	220
	Protein arrays	225
	Content	228
	Microfluidics	243
	Other chip-based technologies in proteomics	244
	Conclusion	246
	References	247
12	Infectomics Overview: Holistic and Integrative Studies of Infectious Diseases	251
	<i>Sheng-He Huang and Ambrose Jong</i>	
	Introduction	251
	Investigating microbial infection with 'omic' approaches	252
	Dissecting the microbiome: ecological solutions to infectious diseases	262
	Animal models and gene knockin/knockout technologies	263
	Computational and mathematic tools for infectomics	264
	Prospects of infectomics	265
	References	267
13	The Drug Discovery Process	271
	<i>Roberto Solari</i>	
	Introduction	271
	The process	272
	Target discovery	274
	Target validation	276
	Discovery of chemical leads	277
	Hits to leads	282
	Lead molecular optimization	282
	Lead to development candidate	284
	Biological candidates	284
	Pre-clinical development	285
	Development issues for biological drugs	289
	Clinical trials	290
	Intellectual property	292

Summary	293
References	293
Appendix	294
14 Structure-based Drug Discovery	295
<i>Chen-Chen Kan, Kevin Hambly and Derek A. Debe</i>	
Overview of drug discovery and development	295
Principles of structure-based drug discovery	297
Case studies of structure-based drug discovery	301
Structure-based drug discovery in the post-genomic era	313
Acknowledgements	318
References	318
15 Protein Interaction-targeted Drug Discovery	323
<i>Gary Hudes, Sanjay Menon and Erica A. Golemis</i>	
Introduction	323
PPIs: classes, functional relationships and regulation	324
Protein interactions in disease-related signalling: categories of disruption	327
Selecting targets and screens	330
Progress in cancer treatment	334
Summary	341
References	341
16 Overview of Quantitative Structure–Activity Relationships (QSAR)	347
<i>David A. Winkler</i>	
Modelling structure–activity relationships	347
The QSAR method	349
Generation of descriptors	352
Descriptor selection	356
Structure–activity mapping	357
Validation and testing	361
Applications and conclusions	362
References	362
Index	369

1

Pharmacogenetics and Pharmacogenomics: An Overview

W. Kalow

Pharmacogenomics has recently arisen from the well established science of pharmacogenetics; in fact, both names are sometimes mixed or interchanged by students of abnormal drug effects. The current overview will therefore start with a discussion of pharmacogenetics, its origin and its coverage; this will be followed by a description of the new aspects which pharmacogenomics brings into the study of interaction between genes and the drugs given to patients.

The origins of pharmacogenetics

Pharmacogenetics emerged from the combination of the old sciences of genetics and pharmacology. Sir Archibald Garrod (1931) anticipated the occurrence of individual differences in human reaction to drugs and environmental chemicals in his book *Inborn Factors in Diseases*. J.B.S. Haldane (1949) studied biochemical individuality and also predicted the occurrence of unusual reactions to drugs. By that time, a few genetic differences in drug response had been seen.

Snyder (1932) had described the inheritance of a deficient ability of some people to taste phenylthiocarbamide (PTC), fundamentally a pharmacologic observation. Sawin and Glick (1943) noticed a genetic lack of atropine esterase in some rabbits, a deficiency that restricted their consumption of belladonna-containing plants. During World War II, the generally safe antimalarial drug primaquine

caused haemolysis in many American soldiers, but only in soldiers of African descent; later, the event was shown to be due to deficiency of the enzyme glucose-6-phosphate dehydrogenase (G6PD), a deficiency frequent in Africans, where the mutant tended to cumulate because it protected its carriers from malaria (Beutler 1993). The new exciting tuberculosis-fighting drug isoniazid was seen (Hughes *et al.* 1954) to have neurological effects in some people, those later shown to have a familial deficiency of its destroying enzyme, N-acetyltransferase. Also in the 1950s, the new muscle-paralysing drug succinylcholine, used during anaesthesia, killed some patients while most had no trouble; the cause was found to be a mutation which inactivated plasma cholinesterase, the succinylcholine-destroying enzyme (Kalow 1956).

Several physicians and scientists felt stimulated by these reports. A committee of the American Medical Association invited the geneticist Motulsky (1957) to write a paper entitled 'Drug Reactions, Enzymes, and Biochemical Reactions' for their journal. The geneticist Vogel (1959) in Germany coined the word 'Pharmacogenetics' in a paper describing 'Modern Problems of Human Genetics'. I was in the process of summarizing all pertinent findings in a book (Kalow 1962). Pharmacogenetics was now an established entity.

The initial progress of pharmacogenetics

As time went by, more pharmacogenetic discoveries were made. Denborough (Denborough and Lovell 1960; Denborough *et al.* 1962) reported the occurrence of excessively high body temperatures (hyperpyrexia), rigidity and death among various family members in response to general anaesthetics; this condition is now called malignant hyperthermia. Dundee *et al.* (1962) emphasized the paralytic effect of barbiturates in cases of hepatic porphyria. Aebi *et al.* (1961) observed in Switzerland cases with a genetic lack of catalase activity, cases as seen before only in Japan. Von Wartburg *et al.* (1965) reported genetic variation of alcohol dehydrogenase. Further pharmacogenetic cases were reported in the following years, but, in addition, some of the older reports were extended. For instance, Kalow *et al.* (1970) showed that malignant hyperthermia represented a biochemical defect in skeletal muscle that caused the muscle to respond abnormally to caffeine, thereby making *in vitro* tests and predictions of this condition possible.

All these cases represented relatively rare observations and therefore did not dramatically affect the practice of medicine. This changed with the report on familial failures of the oxidation of debrisoquine, a sympatholytic, antihypertensive drug. Dr R.L. Smith (1986), a medical investigator in England, took the drug personally for experimental purposes, but suffered a marked and extensive hypotensive episode. He therefore started to investigate the metabolism of the drug, found the metabolic defect, and reported the findings in an excellent publication (Mahgoub *et al.* 1977). A short time later, Eichelbaum *et al.* (1979) in Germany

reported that some people could not metabolize sparteine, an anti-arrhythmic drug. Subsequently, other investigators (Bertilsson *et al.* 1980; Inaba *et al.* 1980) found that the same liver enzyme metabolized both debrisoquine and sparteine. All measurements were based on the drug/metabolite ratio in urine.

The medical interest in this topic was stimulated by several facts. Oxidation is an important and widespread metabolism in human and animal bodies. The defect was not rare but obviously affected quite a few people; that is, it was a polymorphism, a frequently occurring genetic alteration. The enzyme was a cytochrome P450 (CYP2D6), an important class of drug-metabolizing enzymes. It was soon found that the enzyme metabolized many drugs, presently thought to be as many as 20 per cent of all clinically used drugs. Medline cites currently almost 2000 publications dealing with CYP2D6.

The usual clinical effect of the deficiency is an over-response, or toxic response, to any drug that is not metabolized by CYP2D6 (Meyer 2001). However, if a drug needs metabolic activation, an under-response, or lack of response, can also occur. For instance, the debrisoquine-metabolizing enzyme is also the one which converts codeine into morphine. If the enzyme lacks, this conversion does not take place, and codeine is not a pain-killing analgesic (Dayer *et al.* 1988).

Molecular genetic methods enlarged pharmacogenetics

As we can see, pharmacogenetics was in the beginning entirely concerned with functional differences in drug response or drug metabolism, that is, *phenotype* differences which could be directly observed or measured in family members. Pharmacogenetics became much enriched when new methods, described in the following chapters, allowed a direct look at the DNA of mutant genes, thereby determining *genotype*.

The use of the methods of molecular genetics disclosed many kinds of mutations (Stockley and Ray 2001), revealed as changes of nucleic acids. The most frequent variants are SNPs, Single Nucleotide Polymorphisms, indicating simple exchanges of two nucleotides; many genes carry several, often many SNPs (Grant and Phillips 2001 and see Gut; Chapter 4). For instance, about 130 SNPs have been described for G6PD, the enzyme glucose-6-phosphate dehydrogenase (<http://www.bioinf.org.uk/g6pd/>). However, some SNPs do not alter the amino acid of the derived protein, and therefore do not affect protein structure and function; they are called *silent*. Some mutations may change reaction rates, substrate binding, sometimes affecting only selected substrates. There are also other complexities. For instance, absence of an enzyme activity could be caused by any one of four kinds of mutations: frame shift, splicing defect, strip codon or gene deletion. In addition, genetic tests sometimes revealed gene duplication or even multiplication, thereby greatly decreasing the action of dependent drugs. Table 1.1 lists the number of alleles among members of the drug-metabolizing