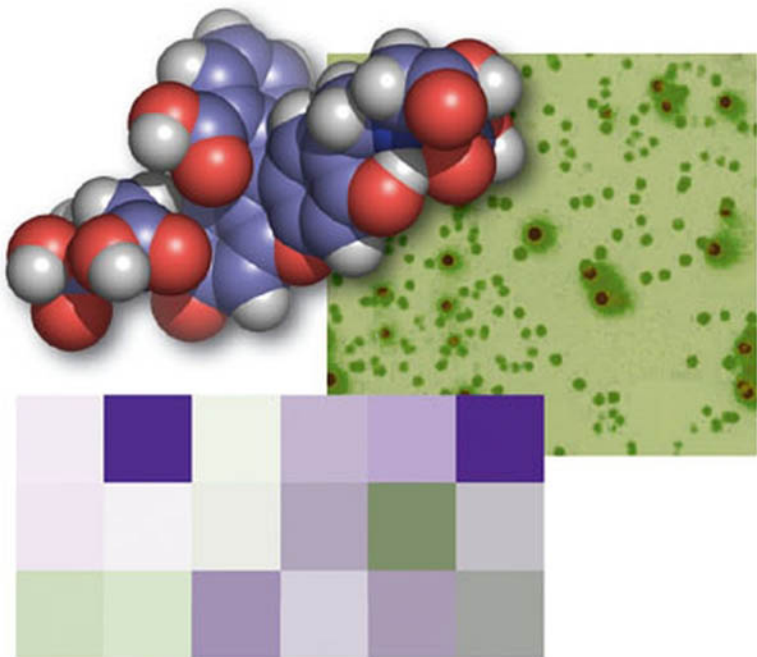


Edited by Jean-Louis Reymond

 WILEY-VCH

Enzyme Assays

High-throughput Screening,
Genetic Selection and Fingerprinting



Enzyme Assays

Edited by
Jean-Louis Reymond

Related Titles:

H. Bisswanger

Enzymkinetik

2000. ISBN 3-527-30096-1

K. Drauz, H. Waldmann (Eds.)

Enzyme Catalysis in Organic Synthesis A Comprehensive Handbook

2002. ISBN 3-527-29949-1

A. Pingoud et al.

Biochemical Methods A Concise Guide for Students and Researchers

2002. ISBN 3-527-30299-9

S. Brakmann, A. Schwienhorst (Eds.)

Evolutionary Methods in Biotechnology Clever Tricks for Directed Evolution

2004. ISBN 3-527-30799-0

H. Bisswanger

Practical Enzymology

2004. ISBN 3-527-30444-4

S. Lorkowski, P. Cullen (Eds.)

Analysing Gene Expression A Handbook of Methods. Possibilities and Pitfalls

2002. ISBN 3-527-30488-6

Enzyme Assays

High-throughput Screening,
Genetic Selection and Fingerprinting

Edited by
Jean-Louis Reymond



WILEY-
VCH

WILEY-VCH Verlag GmbH & Co. KGaA

Editor:

Prof. Dr. Jean-Louis Reymond

University of Berne
Department of Chemistry & Biochemistry
Freistrasse 3
3012 Berne
Switzerland

Cover illustration

The cover picture shows a cpk model of calcein (upper left), a fluorescent sensor useful for high-throughput screening of acylases, aminopeptidases, and proteases, as discussed in the Introduction.

The image on the right is a close-up view of an agar plate with colonies expressing mutant monoamine oxidases in the presence of (*S*)-alpha-methyl benzylamine as substrate and 3,3'-diaminobenzidine as sensor. Colony staining results from chemical oxidation of 3,3'-diaminobenzidine by the hydrogen peroxide produced in the enzyme oxidation, as discussed in Chapter 5.

The bottom grid shows a fingerprint of activity (color intensity) and enantioselectivity (purple = *R*-enantioselectivity, green = *S*-enantioselectivity) of *Bacillus thermocatenuatus* lipase (BTL2) on chiral ester substrates using the assay described in Chapter 1 and the color coding method in Chapter 10. The cover was based on a prototype by Peter Bernhardt.

All books published by Wiley-VCH are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

All books published by Wiley-VCH are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data:

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

Bibliographic information published by

Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the internet at <http://dnb.ddb.de>

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany

All rights reserved (including those of translation in other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Typesetting K+V Fotosatz GmbH, Beerfelden

Printing betz-druck GmbH, Darmstadt

Binding Litges & Dopf Buchbinderei GmbH, Heppenheim

Printed on acid-free paper

Printed in the Federal Republic of Germany

ISBN-13: 978-3-527-31095-1

ISBN-10: 3-527-31095-9

Contents

Preface XIII

List of Contributors XV

Introduction 1

Renaud Sicard and Jean-Louis Reymond

Enzyme Assays 1

Part I: The Chemistry of Enzyme Assays 7

Part II: Enzyme Assays and Genetic Selection 7

Part III: Enzyme Profiling 9

Enzyme Assays in Other Areas 11

How to Use this Book 11

Part I High-throughput Screening 15

**1 Quantitative Assay of Hydrolases for Activity and Selectivity
Using Color Changes** 17

Romas J. Kazlauskas

1.1 Overview 17

1.2 Direct Assays Using Chromogenic Substrates 18

1.3 Indirect Assays Using Coupled Reactions – pH Indicators 19

1.3.1 Overview of Quantitative Use of pH Indicator Assay 21

1.3.2 Applications 24

1.3.2.1 Searching for an Active Hydrolase
(Testing Many Hydrolases Toward One Substrate) 24

1.3.2.2 Substrate Mapping of New Hydrolases
(Testing Many Substrates Toward Hydrolase) 25

1.3.3 Comparison with Other Methods 26

1.4 Estimating and Measuring Selectivity 27

1.4.1 Estimating Selectivity without a Reference Compound 28

1.4.2 Quantitative Measure of Selectivity Using a Reference Compound
(Quick *E* and Related Methods) 30

1.4.2.1	Chromogenic Substrate	32
1.4.2.2	pH Indicators	33
1.4.3	Application	33
1.4.3.1	Substrate Mapping of Hydrolases	33
1.4.3.2	Screening of Mutants in Directed Evolution	33
1.4.4	Advantages and Disadvantages	36
	<i>References</i>	38
2	High-throughput Screening Systems for Assaying the Enantioselectivity of Enzymes	41
	<i>Manfred T. Reetz</i>	
2.1	Introduction	41
2.2	UV/Vis Spectroscopy-based Assays	42
2.2.1	Assay for Screening Lipases or Esterases in the Kinetic Resolution of Chiral <i>p</i> -Nitrophenyl Esters	43
2.2.2	Enzyme-coupled UV/Vis-based Assay for Lipases and Esterases	45
2.2.3	Enzymatic Method for Determining Enantiomeric Excess (EMDee)	46
2.2.4	UV/Vis-based Enzyme Immunoassay as a Means to Measure Enantiomeric Excess	47
2.2.5	Other UV/Vis-based <i>ee</i> -Assays	48
2.3	Assays Using Fluorescence	48
2.3.1	Umbelliferone-based Systems	48
2.3.2	Fluorescence-based Assay Using DNA Microarrays	51
2.3.3	Other Fluorescence-based <i>ee</i> -Assays	53
2.4	Assays Based on Mass Spectrometry (MS)	53
2.4.1	MS-based Assay Using Isotope Labeling	53
2.5	Assays Based on Nuclear Magnetic Resonance Spectroscopy	58
2.6	Assay Based on Fourier Transform Infrared Spectroscopy for Assaying Lipases or Esterases	62
2.7	Assays Based on Gas Chromatography	65
2.8	Assays Based on HPLC	68
2.9	Assays Based on Capillary Array Electrophoresis	69
2.10	Assays Based on Circular Dichroism (CD)	71
2.11	Assay Based on Surface-enhanced Resonance Raman Scattering	73
2.12	Conclusions	73
	<i>References</i>	74
3	High-throughput Screening Methods Developed for Oxidoreductases	77
	<i>Tyler W. Johannes, Ryan D. Woodyer, and Huimin Zhao</i>	
3.1	Introduction	77
3.2	High-throughput Methods for Various Oxidoreductases	78
3.2.1	Dehydrogenases	78

3.2.1.1	Colorimetric Screen Based on NAD(P)H Generation	78
3.2.1.2	Screens Based on NAD(P)H Depletion	79
3.2.2	Oxidases	80
3.2.2.1	Galactose Oxidase	80
3.2.2.2	D-Amino Acid Oxidase	82
3.2.2.3	Peroxidases	82
3.2.3	Oxygenases	85
3.2.3.1	Assays Based on Optical Properties of Substrates and Products	85
3.2.3.2	Assays Based on Gibbs' Reagent and 4-Aminoantipyrine	86
3.2.3.3	<i>para</i> -Nitrophenoxy Analog (pNA) Assay	87
3.2.3.4	Horseradish Peroxidase-coupled Assay	88
3.2.3.5	Indole Assay	89
3.2.4	Laccases	89
3.2.4.1	ABTS Assay	90
3.2.4.2	Poly R-478 Assay	90
3.2.4.3	Other Assays	90
3.3	Conclusions	91
	<i>References</i>	92
4	Industrial Perspectives on Assays	95
	<i>Theo Sonke, Lucien Duchateau, Dick Schipper, Gert-Jan Euverink, Joerd van der Wal, Huub Henderickx, Roland Bezemer, and Aad Vollebregt</i>	
4.1	Introduction	95
4.2	Prerequisites for an Effective Biocatalyst Screening in Chemical Custom Manufacturing	97
4.3	CCM Compliant Screening Methods Based on Optical Spectroscopy (UV/Vis and Fluorescence)	101
4.3.1	Optical Spectroscopic Methods Based on the Spectral Properties of the Product Itself	101
4.3.1.1	Example: Isolation of the D- <i>p</i> -Hydroxyphenylglycine Aminotransferase Gene	102
4.3.2	Optical Spectroscopic Methods Based on Follow-up Conversion of Product	104
4.3.2.1	Example: Fluorometric Detection of Amidase Activity by <i>o</i> -Phthaldehyde/Sulfite Derivatization of Ammonia	106
4.3.2.2	Example: Colorimetric Detection of Amidase Activity by Detection of Ammonia via Glutamate Dehydrogenase-coupled Assay	108
4.3.2.3	Example: Colorimetric Detection of Amino Amidase Activity Using Cu ²⁺ as Sensor for Amino Acids	112
4.4	CCM Compliant Screening Methods Based on Generic Instrumental Assays	114
4.4.1	Flow-injection NMR as Analytical Tool in High-throughput Screening for Enzymatic Activity	115

4.4.1.1	History	115
4.4.1.2	Current Practice	117
4.4.1.3	Practical Aspects	119
4.4.1.4	Example: Screening of a Bacterial Expression Library for Amidase-containing Clones	122
4.4.1.5	Example: Identification of a Phenylpyruvate Decarboxylase Clone	124
4.4.1.6	Example: Identification of Amidase Mutants with Improved Activity towards α -Methylphenylglycine Amide	125
4.4.2	Fast LC/MS for High-throughput Screening of Enzymatic Activity	126
4.4.2.1	Example: Screening of a Bacterial Expression Library for Amidase-containing Clones	127
4.4.2.2	Example: Screening of Enzymatic Racemase Activity	129
4.5	Conclusions	132
	<i>References</i>	133

Part II Genetic Selection 137

5 Agar Plate-based Assays 139

Nicholas J. Turner

5.1	Introduction	139
5.1.1	Directed Evolution of Enzymes: Screening or Selection?	139
5.1.2	General Features of Agar Plate-based Screens	141
5.2	Facilitated Screening-based Methods	143
5.2.1	Amidase	143
5.2.2	Esterase	144
5.2.3	Glycosynthase	144
5.2.4	Galactose Oxidase	146
5.2.5	Monoamine Oxidase	147
5.2.6	P450 Monooxygenases	150
5.2.7	Carotenoid Biosynthesis	151
5.2.8	Biotin Ligase	153
5.3	<i>In vivo</i> Selection-based Methods	154
5.3.1	Glycosynthase	154
5.3.2	Prephenate Dehydratase/Chorismate Mutase	155
5.3.3	Terpene Cyclase	157
5.3.4	Tryptophan Biosynthesis	157
5.3.5	Ribitol Dehydrogenase	157
5.3.6	Inteins	158
5.3.7	Aminoacyl-tRNA Synthetase	159
5.4	Conclusions and Future Prospects	159
	<i>References</i>	160

6	High-throughput Screens and Selections of Enzyme-encoding Genes	163
	<i>Amir Aharoni, Cintia Roodveldt, Andrew D. Griffiths, and Dan S. Tawfik</i>	
6.1	Introduction	163
6.2	The Basics of High-throughput Screens and Selections	164
6.3	High-throughput Selection of Enzymes Using Phage Display	165
6.4	High-throughput Selection of Enzymes Using Cell Display	168
6.5	<i>In vivo</i> Genetic Screens and Selections	169
6.6	Screens for Heterologous Protein Expression and Stability	169
6.6.1	Introduction	169
6.6.2	Screening Methodologies for Heterologous Expression	171
6.6.3	Directed Evolution for Heterologous Expression – Recent Examples	173
6.7	<i>In vitro</i> Compartmentalization	174
6.8	IVC in Double Emulsions	177
6.9	Concluding Remarks	179
	<i>References</i>	179
7	Chemical Complementation	
	<i>Scott Lefurgy and Virginia Cornish</i>	183
7.1	Introduction	183
7.2	Complementation Assays	184
7.2.1	Introduction	184
7.2.2	Early Complementation Assays	184
7.2.3	Enzymology by Complementation	186
7.2.4	Directed Evolution by Complementation	188
7.3	Development of Chemical Complementation	191
7.3.1	Introduction	191
7.3.2	Three-hybrid Assay	192
7.3.2.1	Original Yeast Three-hybrid System	192
7.3.2.2	Dexamethasone–Methotrexate Yeast Three-hybrid System	194
7.3.2.3	Technical Considerations	196
7.3.2.4	Other Three-hybrid Systems	198
7.3.3	Chemical Complementation	198
7.3.3.1	Selection Scheme and Model Reaction	199
7.3.3.2	Results	202
7.3.3.3	General Considerations	203
7.3.3.4	Related Methods	203
7.4	Applications of Chemical Complementation	204
7.4.1	Introduction	204
7.4.2	Enzyme–Inhibitor Interactions	204
7.4.2.1	Rationale	205
7.4.2.2	Screen Strategy	205

7.4.2.3	Enzyme Library Screen	208
7.4.2.4	General Considerations	210
7.4.3	Glycosynthase Evolution	210
7.4.3.1	Rationale	211
7.4.3.2	Selection Scheme	212
7.4.3.3	Glycosynthase Assay	213
7.4.3.4	Directed Evolution	215
7.4.3.5	General Considerations	216
7.5	Conclusion	216
	<i>References</i>	217

8 Molecular Approaches for the Screening of Novel Enzymes

Valéria Maia de Oliveira and Gilson Paulo Manfio 221

8.1	Introduction	221
8.2	Use of Nucleic Acid Probes to Detect Enzyme-coding Genes in Cultivated Microorganisms	222
8.2.1	Current Knowledge and Applications	223
8.2.2	Limitations of Probe Technology and the Need for Innovative Approaches	224
8.3	The Microbial Metagenome: a Resource of Novel Natural Products and Enzymes	226
8.3.1	Accessing the Uncultivated Biodiversity: the Community DNA Concept	226
8.3.2	Unravelling Metabolic Function: the BAC Strategy	227
8.3.3	Analysis of Metagenomic Libraries: Activity versus Sequence-driven Strategy, Enrichment for Specific Genomes and Application of High-throughput Screening Methods	230
8.3.4	Follow-up of the Metagenome Harvest	233
8.4	Concluding Remarks	235
	<i>References</i>	236

Part III Enzyme Fingerprinting 239

9 Fluorescent Probes for Lipolytic Enzymes 241

Ruth Birner-Grünberger, Hannes Schmidinger, Alice Loidl, Hubert Scholze, and Albin Hermetter

9.1	Introduction	241
9.2	Fluorogenic and Fluorescent Substrates for Enzyme Activity	242
9.2.1	Triacylglycerol Lipase Activity Assay	245
9.2.2	Diacylglycerol Lipase Activity Assay	247
9.2.3	Cholesteryl Esterase Activity Assay	249
9.2.4	Phospholipase Activity Assay	250
9.2.5	Sphingomyelinase Activity Assay	252
9.3	Fluorescent Inhibitors for Quantitative Analysis of Active Enzymes and Functional Enzyme Fingerprinting	254

- 9.3.1 Lipase and Esterase Profiling 254
- 9.3.1.1 Microbial Lipases and Esterases 255
- 9.3.1.2 Porcine Pancreatic Lipase 257
- 9.3.1.3 Hormone-sensitive Lipase 257
- 9.3.2 Probing Biophysical Enzyme Properties 259
- 9.3.3 Affinity-based Proteome Profiling (ABPP) 262
- 9.3.3.1 Functionality-based Serine Hydrolase Profiling in Tissue Preparations and Cell Lines 263
- References* 267

10 Fingerprinting Methods for Hydrolases

Johann Grognux and Jean-Louis Reymond 271

- 10.1 Introduction 271
- 10.1.1 One Enzyme – One Substrate 273
- 10.1.2 Enzyme Activity Profiles 275
- 10.1.3 The APIZYM System for Microbial Strain Identification 276
- 10.2 Hydrolase Fingerprinting 278
- 10.2.1 Fingerprinting with Fluorogenic and Chromogenic Substrates 279
- 10.2.2 Fingerprinting with Indirect Chromogenic Assays 284
- 10.2.3 Cocktail Fingerprinting 287
- 10.3 Classification from Fingerprinting Data 289
- 10.3.1 Fingerprint Representation 290
- 10.3.2 Data Normalization 293
- 10.3.3 Hierarchical Clustering of Enzyme Fingerprints 295
- 10.3.4 Analysis of Substrate Similarities 297
- 10.4 Outlook 299
- References* 300

11 Protease Substrate Profiling

Jennifer L. Harris 303

- 11.1 Introduction 303
- 11.2 Functional Protease Profiling – Peptide Substrate Libraries 304
- 11.2.1 Solution-based Peptide Substrate Libraries 306
- 11.2.2 Solid Support-based Synthesis and Screening of Peptide Libraries 314
- 11.2.3 Genetic Approaches to Identifying Protease Substrate Specificity 320
- 11.3 Identification of Macromolecular Substrates 322
- 11.3.1 Genetic Approach to the Identification of Macromolecular Substrates 323
- 11.3.2 Proteomic Approaches to Identifying Protease Substrates 326
- 11.4 Conclusions 327
- References* 328

12	Enzyme Assays on Chips	333
	<i>Souvik Chattopadhyaya and Shao Q. Yao</i>	
12.1	Introduction	333
12.2	Immobilization Strategies	335
12.2.1	Covalent versus Noncovalent Immobilization	335
12.2.2	Site-specific versus Nonspecific Immobilization	336
12.2.3	Site-specific Immobilization of Peptides/Small Molecules	336
12.2.4	Site-specific Immobilization of Proteins	337
12.2.4.1	Intein-mediated Protein Biotinylation Strategies	338
12.2.4.2	Puromycin-mediated Protein Biotinylation	343
12.2.4.3	Immobilization of N-terminal Cysteine-containing Proteins	343
12.3	Microarray-based Methods for Detection of Enzymatic Activity	344
12.3.1	Enzyme Assays Using Protein Arrays	345
12.3.2	Enzyme Assays Using Peptide/Small Molecule Substrate Arrays	348
12.3.2.1	Proteases and Other Hydrolytic Enzymes	348
12.3.2.2	Kinases	351
12.3.2.3	Carbohydrate-modifying Enzymes	355
12.3.2.4	Other Enzymes	356
12.3.3	Enzyme Assays Using Other Types of Arrays	356
12.4	Conclusions	357
	<i>References</i>	359
	Subject Index	363

Preface

When I discuss an enzyme assay with a chemist, we spend our time devising a process that will turn an enzymatic reaction into a detectable signal. The challenge lies in the synthesis of the molecular elements involved in the assay and whether they will behave as expected. Enzyme assay design has elements of rational drug design if it requires docking an unnatural substrate into an enzyme's active site. An enzyme assay may also offer a testbed for a supramolecular functional device, serving to demonstrate its utility. Eventually new principles emerge that might change enzyme analytics altogether.

Then I turn to the biochemist or microbiologist, who sees the enzyme assay as one of many elements in a broader setup, such as the genetic selection of an active enzyme, or the study of its function and mechanism. We usually settle for a commercially available probe or couple the enzyme reaction to a biological system. Our attention focuses on the genetic design of the experiment or its biochemical interpretation. When it succeeds, we wonder with amazement at the results which we only very partly understand.

Finally I meet the industrial researcher, who is hard pressed for preparative performance within a short time window. Our discussion is narrowed down by tight specifications bound to the goals and methods. Nevertheless, the unaltered passion of the scientist keeps shining through. In addition, the products of industrial research and development are remarkable and vindicate the efforts of the entire community.

Romas Kazlauskas, Manfred Reetz, Huimin Zhao, Theo Sonke, Nick Turner, Dan Tawfik, Andrew Griffiths, Virginia Cornish, Valéria Maia de Oliveira, Gilson Paulo Manfio, Albin Hermetter, Jennifer Harris, and Yao Qin Shao have agreed to join forces with me to compose a book on enzyme assays. These authors belong to the world's leading figures in this area. I thank them and their co-authors for their time and efforts, which were essential to the project. I also thank my co-authors and students Johann Grognum and Renaud Sicard, and Elke Maase and Romy Kirsten at Wiley-VCH, for their precious help in editing.

The field of enzyme assays is evolving rapidly and touches an ever increasing number of applications. The present volume captures what we as authors believe is a fair coverage of the area at that point in time. We hope that the book will prove a useful source of information, inspiration, and references for its readers across chemistry and biology.

Berne, October 2005

Jean-Louis Reymond

List of Contributors

Amir Aharoni

The Weizmann Institute of Science
Department of Biological Chemistry
Rehovot 76100
Israel

Roland Bezemer

DSM Food Specialties
Analysis
PO Box 1
2600 MA Delft
The Netherlands

Ruth Birner-Grünberger

Graz University of Technology
Department of Biochemistry
Petersgasse 12/2
8010 Graz
Austria

Souvik Chattopadhyaya

National University of Singapore
Department of Biological Science
3 Science Drive 3
Singapore 117543
Republic of Singapore

Virginia Cornish

Columbia University
Department of Chemistry
3000 Broadway, MC 3111
New York, NY 10027
USA

Lucien Duchateau

DSM Pharma Chemicals –
Advanced Synthesis
Catalysis & Development
PO Box 18
6160 MD Geleen
The Netherlands

Gert-Jan Euverink

University of Groningen
BioExplore
Groningen Biomolecular Sciences
and Biotechnology Institute
PO Box 14
9750 AA Haren
The Netherlands

Andrew D. Griffiths

University Louis Pasteur
Institut de Science et d'Ingénierie
Supramoléculaires (ISIS)
CNRS UMR 7006
8 allée Gaspard Monge, BP 70028
67083 Strasbourg Cedex
France

Johann Grogg

University of Berne
Department of Chemistry &
Biochemistry
Freistrasse 3
3012 Berne
Switzerland

Jennifer L. Harris

The Genomics Institute of the
Novartis Research Foundation
10675 John Jay Hopkins Drive
San Diego, CA 92121
USA

and

The Scripps Research Institute
Department of Molecular Biology
The Scripps Research Institute
10550 North Torrey Pines Road
San Diego, CA 92121
USA

Huib Henderickx

DSM Resolve
PO Box 18
6160 MD Geleen
The Netherlands

Albin Hermetter

Graz University of Technology
Department of Biochemistry
Petersgasse 12/2
8010 Graz
Austria

Tyler W. Johannes

University of Illinois
Department of Chemistry
600 S. Mathews Ave
Urbana, IL 61801
USA

Romas J. Kazlauskas

University of Minnesota
Department of Biochemistry
Molecular Biology and Biophysics
and The Biotechnology Institute
1479 Gortner Avenue
Saint Paul, MN 55108
USA

Scott Lefurgy

Columbia University
Department of Chemistry
3000 Broadway, MC 3153
New York, NY 10027
USA

Alice Loidl

Graz University of Technology
Department of Biochemistry
Petersgasse 12/2
8010 Graz
Austria

Valéria Maia de Oliveira

Center of Chemistry, Biological
and Agricultural Research
CPQBA/UNICAMP, CP 6171
CEP 13081-970
Campinas, SP
Brazil

Gilson Paulo Manfio

Natura Inovação e Tecnologia
de Produtos Ltda.
CEP 07750-000
Cajamar, SP
Brazil

Manfred T. Reetz

Max-Planck-Institut für
Kohlenforschung
Kaiser-Wilhelm-Platz 1
45470 Mülheim/Ruhr
Germany

Jean-Louis Reymond

University of Berne
Department of Chemistry &
Biochemistry
Freistrasse 3
3012 Berne
Switzerland

Cintia Roodveldt

The Weizmann Institute of Science
Department of Biological Chemistry
Rehovot 76100
Israel

Dick Schipper

DSM Food Specialties
Analysis
PO Box 1
2600 MA Delft
The Netherlands

Hannes Schmidinger

Graz University of Technology
Department of Biochemistry
Petersgasse 12/2
8010 Graz
Austria

Hubert Scholze

Graz University of Technology
Department of Biochemistry
Petersgasse 12/2
8010 Graz
Austria

Renaud Sicard

University of Berne
Department of Chemistry &
Biochemistry
Freistrasse 3
3012 Berne
Switzerland

Theo Sonke

DSM Pharma Chemicals –
Advanced Synthesis
Catalysis & Development
PO Box 18
6160 MD Geleen
The Netherlands

Dan S. Tawfik

The Weizmann Institute of Science
Department of Biological Chemistry
Rehovot 76100
Israel

Nicholas J. Turner

University of Manchester
School of Chemistry
Oxford Road
Manchester M13 9PL
UK

Sjoerd van der Wal

DSM Resolve
PO Box 18
6160 MD Geleen
The Netherlands

Aad Vollebregt

DSM Anti-Infectives
DAI Innovation
PO Box 425
2600 AK Delft
The Netherlands

Ryan D. Woodyer

University of Illinois
Department of Chemistry
600 S. Mathews Avenue
Urbana, IL 61801
USA

Shao Q. Yao

National University of Singapore
Department of Biological Science and
Department of Chemistry
3 Science Drive 3
Singapore 117543
Republic of Singapore

Huimin Zhao

University of Illinois
Department of Chemical
and Biomolecular Engineering
600 S. Mathews Avenue
Urbana, IL 61801
USA

Introduction

Renaud Sicard and Jean-Louis Reymond

An enzyme assay is a test for enzyme function. The enzyme assay probes the chemistry of a single catalytic step in an enzyme and makes it return an answer, which may be a light signal or color change in the sample, or a biological selection event, or both. How to achieve this is left to the experimenter, who can, and usually must, combine various chemical insights and intuitions to arrive at a working assay system. It is a molecular game with plenty of degrees of freedom, but strict demands on efficacy. Ideally, the assay should be simple and free of mistakes – that is, no false positives or false negatives. Success is also rated in terms of which actual reaction is being assayed, some being more difficult than others, and in terms of ease of implementation, which often reduces to the price and availability of the reagents necessary to perform the assay.

Fortunately, the design and utilization of enzyme assays serve a useful purpose. Enzyme assays are indispensable tools for enzyme discovery and enzyme characterization. The present book aims to reflect the tremendous developments that have taken place in these areas over the last 10 years, particularly with regard to high-throughput screening assays and array experiments with multiple substrates. These developments have been discussed in several review articles [1].

The driving force for the invention of new enzyme assays comes in large part from the field of enzyme discovery and engineering [2]. In these areas of investigation enzyme assays are used to identify active enzymes from microorganism collections or randomly generated enzyme mutant libraries. This approach has been found to be very practical for discovering industrially useful catalysts. Enzyme engineering has led to an increased acceptance and utilization of enzymes for manufacturing, in particular in the area of fine chemicals synthesis [3].

Enzyme Assays

What an enzyme assay does to visualize enzyme function is equivalent to what structural analysis tools do for visualizing enzyme structure. However, while

one understands structure intuitively through its three-dimensional representation, there is no unified representation of function. Function can be described as a list of qualitative statements, or as a series of values for suitably defined parameters. For small molecules such as drugs, function correlates well with structure, and predictive quantitative structure–activity relationship (QSAR) models allow one to reduce function to structural elements. For macromolecules, however, the relationship between structure and function is blurred by complexity, and structural analysis delivers at best a crude insight into molecular function. This is particularly true for enzymes, where insignificant alterations in either protein or substrate structure can produce dramatic changes at the level of function, whether it is catalytic activity, selectivity or regulation of the enzyme. In this case the description of molecular function becomes largely independent of structure (Figure 1). Experience has shown that the functional information delivered by an enzyme assay on thousands of mutants is much more useful in showing how an enzyme could be improved than a detailed structure of a single enzyme.

While structural determination methods use physical principles, enzyme assays are mostly born out of chemical principles. Enzyme assay technology builds on classical bioorganic chemistry, and starts with a detailed analysis and understanding of an enzyme’s reaction mechanism and the chemical properties of substrates and products. Engineering of substrate structure or the use of chemical sensors then allows the catalytic reaction to be translated into an observable signal. The assay design largely depends on intuition to formulate for each enzyme a working principle capable of turning enzymatic turnover into a signal.

In enzyme discovery and engineering the assay is used to select improved enzyme variants from pools of enzymes or enzyme mutants. The assay is critical in these experiments because “you get what you screen for”. This adage summarizes the outcome of many experiments: the product of a selection procedure is only as good as the selection principle used. The detailed chemistry of the assay involved is therefore a key parameter for ensuring success in isolating the desired enzyme.

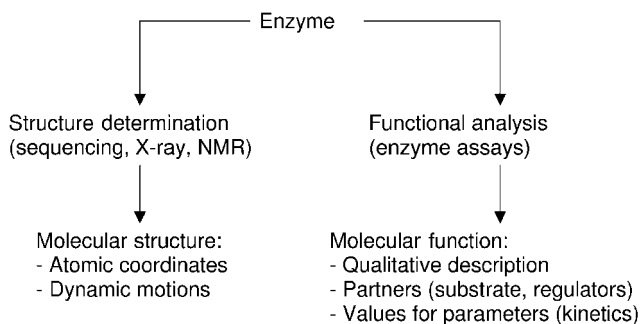


Fig. 1 Enzyme assays as tools for functional analysis.

The principles and applications of enzyme assays are reviewed in this book. In Part I the chemistry of enzyme assays is discussed, in Part II the assays used in the context of genetic selection are covered, and in Part III multisubstrate assays for biochemical characterization of enzymes are discussed. Before even starting into these applications one should remember that positive hits from high-throughput screening assays must always be confirmed by an independent method before concluding that a new enzyme has been discovered (Figure 2).

There are many ways to connect the conversion of a substrate into a product with an observable signal (Figure 3). Enzyme activity can often be detected by the action of the enzyme on its natural substrate. An enzyme activity might lead to heat production if the reaction is exothermic, or induce a macroscopic change in the reaction medium, such as the clearing of an insoluble polymer substrate, or the precipitation of a reaction product. It is also possible to follow reaction turnover using standard analytical methods such as chromatography (gas chromatography or high-performance liquid chromatography) and mass spectrometry, or by nuclear magnetic resonance (NMR) spectroscopy. Several application examples of such methods, in particular with respect to assays for measuring enantioselectivity, are discussed by Manfred Reetz in Chapter 2 and by Theo Sonke and the DSM group in Chapter 4. Direct high-throughput screening assays for enantioselectivity are particularly important in the context of fine chemical synthesis because enantioselectivity is almost always the property being pursued in the course of developing a new catalyst. Electrochemical monitoring of enzyme activity is typically used for glucose-sensing mediated by glucose oxidase [4], and has recently been applied for the lipase cutinase using a hydroqui-

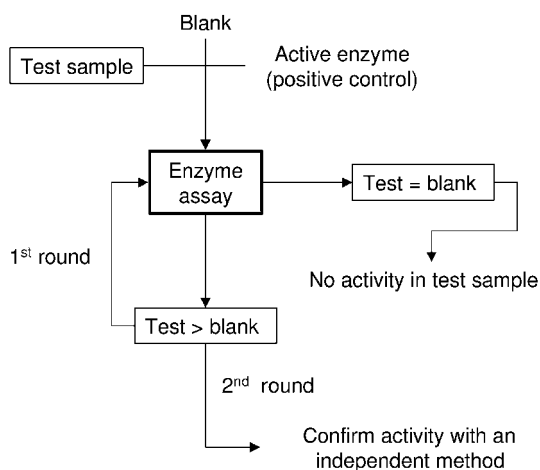


Fig. 2 Principles of high-throughput screening enzyme assays. The signal produced by the assay on the test sample must be checked against a blank sample (medium only) and against an active enzyme as positive control. A positive identification must be repeated, and then confirmed by an independent method.

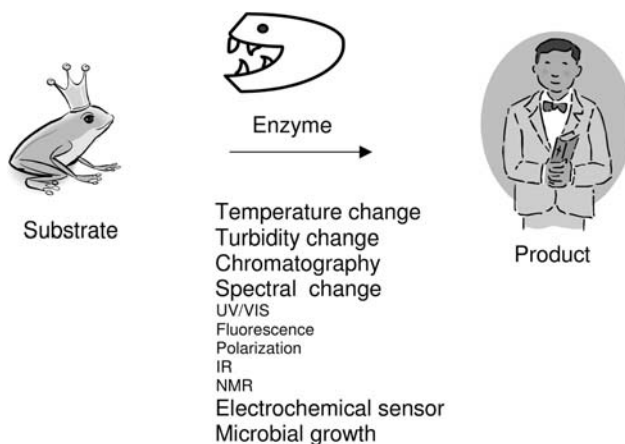


Fig. 3 Signals for enzyme assays produced from enzymatic reactions. Note that observable spectral changes may occur either directly due to structural differences between substrate and product (e.g. Chapters 2, 4, 5, 9, 11 and 12), or indirectly, for example

through a chemical indicator system (e.g. Chapter 1), by processing of the reaction product by secondary enzymes or reagents (e.g. Chapters 3, 6 and 10), or via the induction of gene expression by the reaction product (Chapter 7).

none monobutyrate ester substrate covalently linked to the surface of a gold electrode [5]. Microbial growth as a signal allows one to pick active colonies growing on substrates as carbon source, and also occurs upon genetic selection (Part II).

The largest group of enzyme assays are those that induce recordable changes in light absorbency or fluorescence in the assay medium. The simplest approach relies on colorimetric or fluorimetric chemosensors that respond to product formation or substrate consumption, such as pH indicators (Chapter 1). Such assays are particularly useful because they allow one to work with the substrate of synthetic interest. There are a number of strategies for inducing signals indirectly upon enzymatic turnover, as illustrated by the following examples.

In the copper-calcein assay in Figure 4 [6], an amidase releases a free amino acid as reaction product from the corresponding amide as substrate. Amino acids are strong chelators for metal ions, in particular Cu^{2+} ions, while amino acid amides are not. The assay is based on a complex of Cu^{2+} and the commercially available fluorescein derivative calcein (3), in which the calcein fluorophore is quenched by the metal ion. The deacetylation of *N*-acetyl-L-methionine (1) by acylase I induces a fluorescence increase because the free amino acid reaction product L-methionine (2) chelates Cu^{2+} , which releases free calcein, which regains its fluorescence. The copper-calcein assay can also be used to assay aminopeptidases and proteases using as substrate amino acid amides and bovine serum albumin, respectively. Following a similar principle, indirect product detection in a chemical transformation can also be realized by means of an

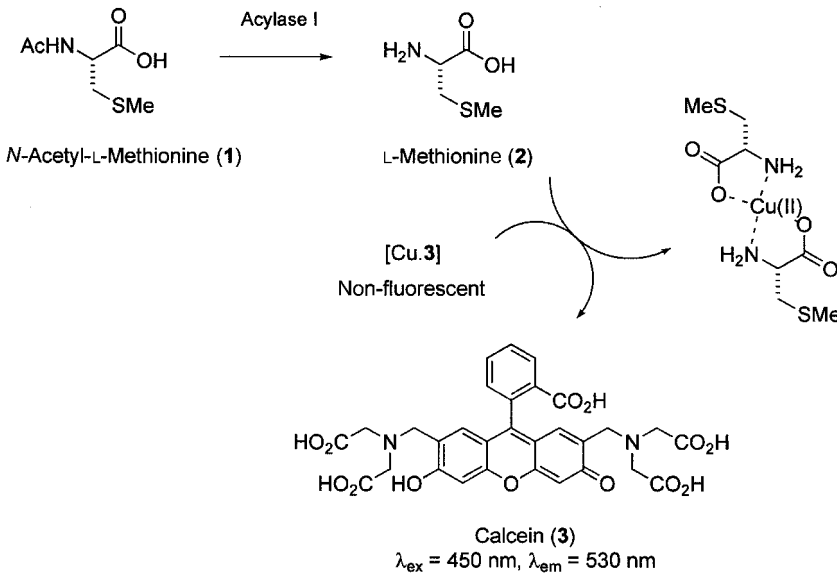


Fig. 4 An indirect fluorogenic assay for acylase I. Calcein (3) is a commercially available inexpensive fluorescein derivative. The assay is also suitable for other amino acid-releasing enzymes, such as aminopeptidases and proteases, when using the appropriate substrate.

immunoassay using an antibody capable of differentiating product from substrate [7]. In these immunoassays the product-selective antibody plays the same role as the Cu^{2+} ion in the amidase assay above.

Another elegant indirect assay by Matile and coworkers is based on vesicles containing a concentrated, autoquenched solution of fluorescein (Figure 5) [8]. The vesicles are equipped with synthetic pores for fluorescein. The pores are plugged by the enzyme substrate, but not by the reaction product. Reaction progress results in unplugging of the pore, which leads to diffusion of fluorescein outside the vesicles and an increase in fluorescence. The assay has been demonstrated for fructose bis-phosphate aldolase, alkaline phosphatase, galactosyltransferase, DNA exonuclease III, DNA polymerase I, RNase A, apyrase, heparinase I, hyaluronidase, papain, ficin, elastase, subtilisin, and pronase.

The most frequently used enzyme assays involve fluorogenic and chromogenic substrates. A synthetic substrate is designed such that the enzyme turns a nonfluorescent or colorless appendage of the substrate into a fluorescent or colored product. Thus, a color or fluorescent signal is created out of a dark or colorless solution by the direct action of the enzyme. This principle is realized by substrates with cleavable ethers or esters of electron-poor conjugated aromatic phenols. The conjugate bases of these phenols show very strong color and fluorescence properties not present in the protonated, alkylated or acylated derivatives. These include the well-known yellow nitrophenolate (4), the blue fluores-

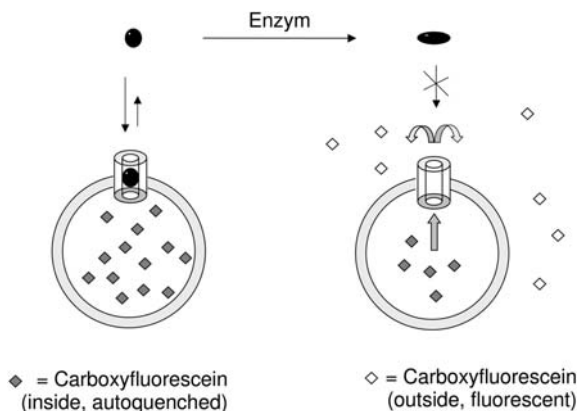


Fig. 5 A general fluorescence enzyme assay using synthetic pores. Product turnover unplugs the pores, which allow fluorescein to diffuse to the outside of the vesicles and become fluorescent.

cent umbelliferone anion (5), the red fluorescent resorufin anion (6) and the green fluorescent fluorescein anion (7) (Figure 6). Many fluorogenic and chromogenic enzyme substrates are commercially available and serve as reference substrates for hydrolytic enzymes (see Chapter 1).

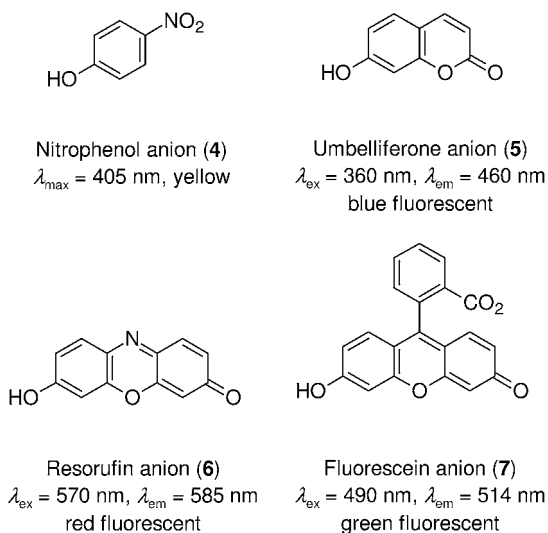


Fig. 6 Acidic conjugated electron-poor phenols used in fluorogenic and chromogenic enzyme substrates. The corresponding neutral phenols are generally colorless and nonfluorescent.

Part I: High-throughput Screening

The significance of an enzyme assay and its successful application depends on its chemical and analytical design. Part I discusses enzyme assays tailored to the problem of enzyme discovery, which requires high-throughput screening potential for relevant chemical transformations. In the context of fine chemical synthesis this means the ability to screen for enantio- and stereoselectivity of the targeted reactions.

In Chapter 1, Romas J. Kazlauskas describes the use of a colorimetric pH indicator together with reference fluorogenic substrates to carry out efficient high-throughput screening of esterolytic enzymes with chiral substrates. The method allows stereoselectivity information to be obtained directly from high-throughput screening with any substrate of synthetic interest.

In Chapter 2, Manfred T. Reetz reviews enzyme assays for screening enantioselective reactions. Analysis of isotopically labeled *pseudo*-enantiomeric mixtures by MS and NMR provides a practical approach for screening kinetic resolutions of racemic mixtures or the deracemization of prochiral substrates. For the case of asymmetric induction where a chiral product is formed from an achiral and nonprochiral substrate, the situation is more complex and requires indirect sensing of product chirality by enantioselective sensors.

In Chapter 3, Tyler W. Johannes, Ryan D. Woodyer, and Huimin Zhao review fluorogenic and chromogenic systems for redox enzymes. These assays are critical because redox enzymes have a particularly important and yet largely untapped potential for industrial applications. For example alkane monooxygenases can perform selective hydroxylation reactions on hydrocarbons that are simply not accessible at all to chemical catalysts [9]. In addition many chemical redox reagents are expensive, toxic, and difficult to handle, implying that economical enzyme replacements should be possible in almost all cases.

The best test bed for enzyme assays occurs in an industrial context, where practical catalysts need to be developed rapidly and applied in large-scale production. In Chapter 4, Theo Sonke, Lucien Duchateau, Dick Schipper, Gert-Jan Euverink, Sjoerd van der Wal, Huub Henderickx, Roland Bezemer, and Aad Vollebregt report their own experiences at the Dutch company DSM, where indirect colorimetric assays and high-throughput direct analyses such as HPLC and NMR have been used. This industrial contribution highlights the importance of screening for enantioselectivity, as also discussed in Chapters 1 and 2.

Part II: Genetic Selection

Enzyme assays play a central role in the context of microbial screening and directed evolution experiments. In this field the catalysis signal is used as the selection criterion to accept or reject single genes or microbial colonies in the hope of isolating enzyme mutants with desirable catalytic properties. The genetic diversity undergoing selection through the enzyme assay consists either in

enzyme mutants generated artificially, or in biodiversity collections, such as gene libraries from the metagenome or microorganism collections (Figure 7).

Screening preferentially describes experiments in which the assay signal is used to direct an external device to pick individual active enzymes or enzyme-producing genes, such as manual picking from agar plates or microtiter plates or the use of fluorescence-activated cell sorting. The term “genetic selection” best describes systems where the expression of an active enzyme is linked to cell survival without external signal processing.

In Chapter 5, Nicholas J. Turner reviews the design and application of enzyme assays in the context of selecting active enzymes by colony picking on agar, which is the most common screen used for microbial cultures. The chapter discusses how much can be achieved quickly by implementing straightforward chemical reaction principles in a microbiological context. A critical overview of genetic selection methods used to isolate active enzymes is also presented.

Random mutagenesis protocols such as gene shuffling [10], error-prone polymerase chain reaction (PCR) [11], and the later improvements or variations of these methods [12] readily allow on the order of 10^{12} mutants of a given enzyme to be generated in a single experiment. However, high-throughput screening experiments in microtiter plates or even on agar plate can only test a few tens of thousands of mutants for catalytic activity. In recent years several groups have invented methods to allow efficient screening of such large numbers of mutants.

In Chapter 6, Amir Aharoni, Cintia Roodveldt, Andrew D. Griffiths, and Dan S. Tawfik provide a general overview of screening methods applicable in the context of both functional genomics and directed evolution. The authors discuss the critical problem of choosing the right expression system for a given enzyme, and the implementation of selection pressure that is able to distinguish between

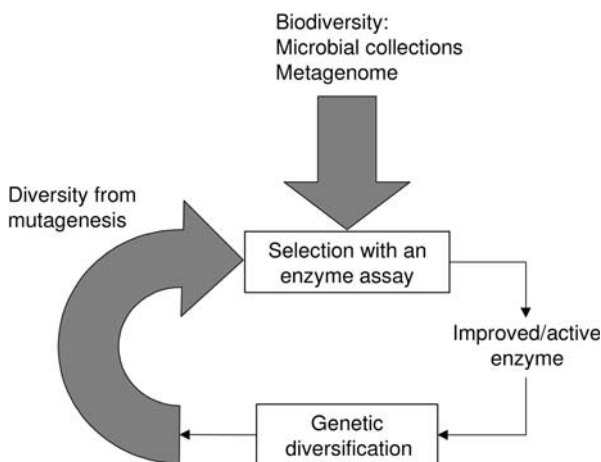


Fig. 7 Selection of active enzymes from genetic libraries and biodiversity.

protein expression levels and protein function. A variety of high-throughput screening approaches for enzymes, such as phage-display and fluorescence-activated cell sorting, are reviewed, including the author's own elegant emulsion-based compartmentalization system for screening large genetic libraries.

In Chapter 7, Scott Lefurgy and Virginia Cornish review high-throughput selection methods by chemical complementation. The chapter includes an excellent review of genetic selection experiments that can be used to perform directed evolution, and emphasizes chemical complementation by the yeast three-hybrid system. In this system a synthetic chemical inducer of dimerization (CID) acts as a tether between a DNA-binding domain and a transcription activation domain. The CID either serves as a substrate cleavable by the enzyme, or is the product of an enzyme coupling. The experiment is set up such that activation or deactivation of gene expression in the presence of an enzyme cleaving or forming the CID is conditional for cell survival, allowing genetic selection to take place. Chemical complementation is demonstrated by various examples, including β -lactamases and glycosynthase enzymes.

In recent years microbiologists studying biodiversity have come to realize that the natural environment, in particular biotopes under extreme conditions, harbor a very large number of diverse microbes. In Chapter 8, Valéria Maia de Oliveira and Gilson Paulo Manfio present an overview of screening methods in the context of exploiting the genetic biodiversity available in microbial collections and in environmental DNA. Environmental DNA is recovered by direct PCR amplification and includes genetic material from noncultivable microbes, which is considered to be the vast majority (>99%), and is collectively called the metagenome [13]. In addition to screening for expressed enzyme activity in such libraries, it is also possible to analyze gene sequences for conserved sequence patterns indicative of certain enzyme activities.

Part III: Enzyme Fingerprinting

An enzyme assay is a tool designed to visualize enzyme function. In its simplest expression, the resulting picture of enzyme activity is a single pixel in two colors (e.g. white=no activity, black=activity, with respect to the assay being performed). The picture can adopt higher levels of definition if the number of pixels is augmented, or if a color shading is allowed for each pixel. This can be realized by combining several different assays for the same enzyme into an array, and by obtaining quantitative rather than qualitative data from each assay. The resulting pictures of enzyme function are called activity profiles, or fingerprints (Figure 8).

The notion of fingerprint is associated with the possibility of using the activity profile as an identification mark for an enzyme or enzyme-containing sample, which is the prerequisite for all diagnostic applications of enzyme assays. Any device capable of recording an enzyme fingerprint is the equivalent of a camera for taking pictures of enzyme function. As for screening, enzyme assays for fin-

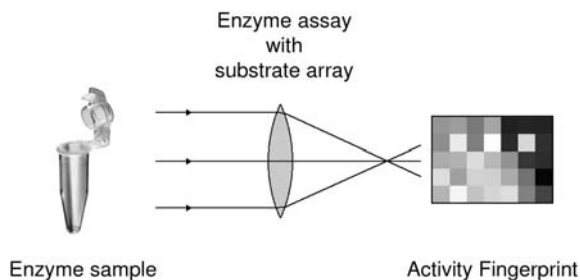


Fig. 8 Enzyme activity fingerprinting.

gerprinting must be applicable in high-throughput. The goal here is to collect the enzyme activity data simultaneously on many different substrates or in many different reaction conditions.

In Chapter 9, Ruth Birner-Grünberger, Hannes Schmidinger, Alice Loidl, Hubert Scholze, and Albin Hermetter discuss assays used for the identification and biochemical study of lipases and esterases. These include fluorogenic substrates specifically designed for targeted hydrolases. The authors also review active-site labeling probes, which are used to covalently tag active enzyme for later identification by gel electrophoresis and mass spectrometry. Such active-site labeling probes have established themselves as useful reagents for the discovery of new disease-related enzymes.

In Chapter 10, Johann Grognum and Jean-Louis Reymond report a series of practical methods for recording activity fingerprints of enzymes, mostly in the case of hydrolytic enzymes such as lipases, esterases, and proteases. Enzyme fingerprinting involves recording a reproducible image of the reactivity profile of an enzyme, such that the images obtained from different enzymes can be used for functional classification. The principle derives from multi-enzyme profiling as used for phenotyping in microbiology and medical diagnostics. Methods of fingerprinting include arrays of indirect fluorogenic substrates acting by a common mechanism of fluorescence release, and substrate cocktail reagents, which allow recording of an activity fingerprint in a single experiment. Data acquisition and statistical analysis techniques leading to functional classification of enzymes are presented.

In Chapter 11, Jennifer L. Harris reviews the application of fluorogenic peptide substrate libraries for large-scale profiling of proteases, a method which is used to define the substrate specificities and the actual natural substrates of proteases. The chapter reviews a number of protease profiling methods and experiments. Protease profiling has proven to be an indispensable tool for the biochemical study of these enzymes. The concept of positional scanning peptide libraries is central to surveying the entire sequence space of peptide substrates within a reasonable experimental effort.

The promise of fingerprinting lies not only in multiparametric analysis for studying enzymes, but also in possible applications in the area of diagnostics

and quality control. To realize this promise it will be necessary to develop miniaturized technologies rivaling DNA-chip technology. Several groups have shown that enzyme fingerprinting experiments can be incorporated into microarrays for highly parallel assays. An overview of miniaturized parallel enzyme assay technologies is presented by Souvik Chattopadhyaya and Shao Q. Yao in Chapter 12. Many of the future prospects for further development in enzyme fingerprinting reside in the exploration of microarray approaches, which will reveal how to produce reliable fingerprints with diagnostic value.

Enzyme Assays in Other Areas

Enzyme assays are used in many other areas of investigation outside the scope of this book. For example, they are used routinely in the practice of medical diagnosis; clinical testing includes tests for enzymes that act as disease markers [14]. Enzyme assays are key components of bioanalytical systems, including signal amplification in the enzyme-linked immunosorbent assay (ELISA) [15], genetic analysis using PCR [16], and gene sequencing using DNA polymerases [17]. Enzyme assays play a critical role in drug discovery, where they are used to test enzymes as drug targets against potential inhibitors [18]. Enzyme assays furthermore form a core technology used for imaging, where they serve to localize active enzymes inside living cells or whole organisms [19]. In these imaging applications enzyme assays are used to complement immunofluorescence staining using fluorescence-labeled antibodies against cellular proteins [20]. Imaging in live organisms also includes the methods of magnetic resonance imaging (MRI) [21] and positron emission tomography (PET) [22].

Recent chemical developments in imaging include the introduction of reagents incorporating red or near-infrared (NIR) chromophores/quencher pairs. For example in Figure 9 peptide substrate **8** could serve for imaging proteases [23]. Further development of enzyme-specific fluorescent probes are also of importance in the context of imaging. A recent example is the aminocoumarin substrate **9**, which can be used to visualize monoamine oxidases (MAO) A and B by forming the fluorescent product **10** [24]. In another example, the fluorescent resonance energy transfer (FRET) substrates **11** and **12** have been used to image phospholipases in zebrafish larvae [25]. This area of imaging has been reviewed recently [26].

How to Use this Book

The chapters that make up this book have been written by different authors and hence have different styles. Taken together, the book covers a large part of all enzyme assay advances in recent years, while also citing classical work. The index contains the names of enzymes for which an assay is described or at least cited, which will facilitate searches for a specific problem. The index also con-

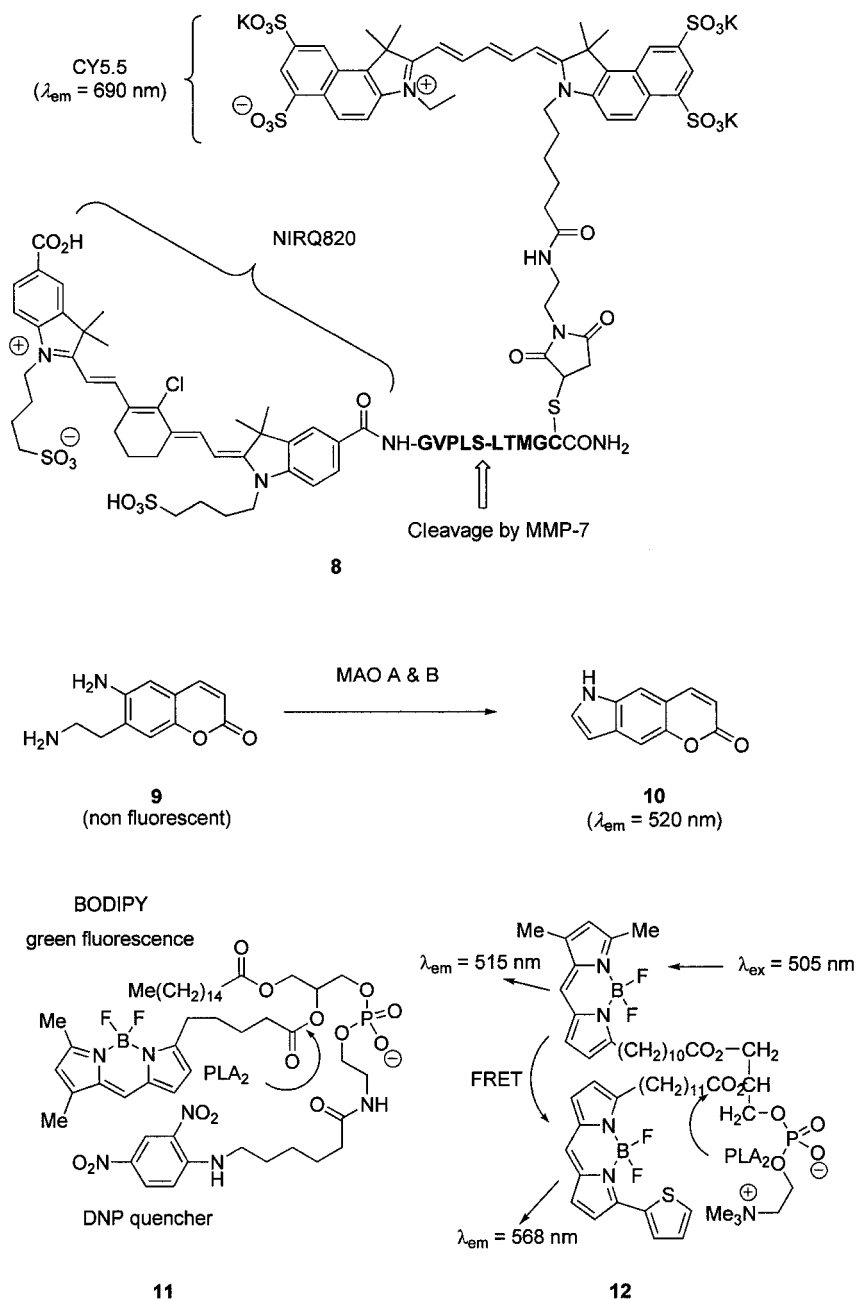


Fig. 9 Example of fluorescent probes for enzyme imaging.

tains names of reagents and substrates. A number of experimental procedures are included. The book should serve as a useful reference to the original literature and as the basis for course material on the subject of enzyme assays.

References

- 1 a) M. T. Reetz, *Angew. Chem.* **2001**, *113*, 292–320; b) M. T. Reetz, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 284–310; c) D. Wahler, J.-L. Reymond, *Curr. Opin. Chem. Biol.* **2001**, *5*, 152–158; d) D. Wahler, J.-L. Reymond, *Curr. Opin. Biotechnol.* **2001**, *12*, 535–544; e) J.-P. Goddard, J.-L. Reymond, *Trends Biotechnol.* **2004**, *22*, 363–370; f) J.-P. Goddard, J.-L. Reymond, *Curr. Opin. Biotechnol.* **2004**, *15*, 314–322; g) M. T. Reetz, High-throughput screening of enantioselective industrial biocatalysts, in *Evolutionary Methods in Biotechnology*, eds S. Brakmann, A. Schwienhorst, Wiley-VCH, Weinheim, **2004**; h) H. Lin, V. W. Cornish, *Angew. Chem. Int. Ed.* **2002**, *41*, 4402–4425; i) M. Uttamchandani, D. P. Walsh, S. Q. Yao, Y. T. Chang, *Curr. Opin. Chem. Biol.* **2005**, *9*, 4–13; j) F. H. Arnold, *Nature* **2001**, *409*, 253–257; k) F. H. Arnold, P. L. Wintrode, K. Miyazaki, A. Gershenson, *Trends Biochem. Sci.* **2001**, *26*, 100–106.
- 2 a) K. Hult and P. Berglund, *Curr. Opin. Biotechnol.* **2003**, *14*, 395–400; b) G. J. Williams, A. S. Nelson, A. Berry, *Cell Mol. Life Sci.* **2004**, *61*, 3034–3046; c) D. E. Robertson, B. A. Steer, *Curr. Opin. Chem. Biol.* **2004**, *8*, 141–149.
- 3 a) A. Schmid, J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts, B. Witholt, *Nature* **2001**, *409*, 258–268; b) S. Panke, M. Held, M. Wubbolts, *Curr. Opin. Biotechnol.* **2004**, *15*, 272–279.
- 4 a) G. Leegsma-Vogt, M. M. Rhemrev-Boom, R. G. Tiessen, K. Venema, J. Korf, *Biomed. Mater. Eng.* **2004**, *14*, 455–464; b) S. A. Jaffari, A. P. Turner, *Physiol. Meas.* **1995**, *16*, 1–15.
- 5 W.-S. Yeo, M. Mrksich, *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 3299–3301.
- 6 a) G. Klein, J.-L. Reymond, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 1771–1773; b) K. E. S. Dean, G. Klein, O. Renaudet, J.-L. Reymond, *Bioorg. Med. Chem. Lett.* **2003**, *10*, 1653–1656.
- 7 a) D. S. Tawfik, B. S. Green, R. Chap, M. Sela, Z. Eshhar, *Proc. Natl Acad. Sci. USA* **1993**, *90*, 373–377; b) G. MacBeath, D. Hilvert, *J. Am. Chem. Soc.* **1994**, *116*, 6101–6106; c) P. Geymayer, N. Bahr, J.-L. Reymond, *Chem. Eur. J.* **1999**, *5*, 1006–1012; d) C. Gauchet, B. Mohar, A. Valleix, P. Y. Renard, S. Meunier, F. Taran, A. Wagner, J. Grassi, C. Mioskowski, *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 124–127; e) F. Taran, P. Y. Renard, C. Creminon, A. Valleix, Y. Frobort, P. Pradelles, J. Grassi, C. Mioskowski, *Tetrahedron Lett.* **1999**, *40*, 1891–1894; f) S. Meunier, F. Taran, C. Mioskowski et al., *ChemBioChem* **2004**, *5*, 832–840; g) M. Matsushita, K. Yoshida, N. Yamamoto, P. Wirsching, R. A. Lerner, K. D. Janda, *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 5984–5987.
- 8 a) G. Das, P. Talukdar, S. Matile, *Science* **2002**, *298*, 1600–1602; b) N. Sorde, G. Das, S. Matile, *Proc. Natl Acad. Sci. USA* **2003**, *100*, 11964–11969.
- 9 a) Z. Li, J. B. van Beilen, W. A. Duetz, A. Schmid, A. de Raadt, H. Griengl, B. Witholt, *Curr. Opin. Chem. Biol.* **2002**, *6*, 136–144; b) W. A. Duetz, J. B. van Beilen, B. Witholt, *Curr. Opin. Biotechnol.* **2001**, *12*, 419–425; c) M. Alcalde, E. T. Farinas, F. H. Arnold, *J. Biomol. Screen.* **2004**, *9*, 141–146; d) M. W. Peters, P. Meinhold, A. Glieder, F. H. Arnold, *J. Am. Chem. Soc.* **2003**, *125*, 13442–13450; e) P. C. Cirino, F. H. Arnold, *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 3299–3301.
- 10 a) W. Stemmer, *Proc. Natl Acad. Sci. USA* **1994**, *91*, 10747–10751; b) A. Cramer, S. A. Raillard, E. Bermudez, W. P. C. Stemmer, *Nature* **1998**, *391*, 288–291; c) J. A. Kolkman, W. P. C. Stemmer, *Nat. Biotechnol.* **2001**, *19*, 423–428.

- 11 a) H. Zhao, L. Giver, Z. Shao, J.A. Affholter, F.H. Arnold, *Nat. Biotechnol.* **1998**, *16*, 258–261; b) H. Tao, V.W. Cornish, *Curr. Opin. Chem. Biol.* **2002**, *6*, 858–864; c) V. Sieber, C.A. Martinez, F.H. Arnold, *Nat. Biotechnol.* **2001**, *19*, 456–460.
- 12 C. Neylon, *Nucleic Acids Res.* **2004**, *32*, 1448–1459.
- 13 R.I. Amann, W. Ludwig, K.H. Schleifer, *Microbiol. Rev.* **1995**, *59*, 143–169.
- 14 a) Y. Li, C. Ronald Scott, N.A. Chamoles, A. Ghavami, B.M. Pinto, F. Turecek, M.H. Gelb, *Clin. Chem.* **2004**, *50*, 1785–1796; b) N.M. Verhoeven, B. Roos, E.A. Struys, G.S. Salomons, M.S. van der Knaap, C. Jakobs, *Clin. Chem.* **2004**, *50*, 441–443.
- 15 G.J. Doellgast, M.X. Triscott, G.A. Beard, J.D. Bottoms, T. Cheng, B.H. Roh, M.G. Roman, P.A. Hall, J.E. Brown, *J. Clin. Microbiol.* **1993**, *31*, 2402–2409.
- 16 S.M. Johnson, K.A. Simmons, D. Pappagianis, *J. Clin. Microbiol.* **2004**, *42*, 1982–1985.
- 17 M.M. Albà, *Genome Biol.* **2001**, *2*, reviews3002.1-reviews3002.4.
- 18 A.D. Rodrigues, J.H. Lin, *Curr. Opin. Chem. Biol.* **2001**, *5*, 396–401.
- 19 a) N. J. Emptage, *Curr. Opin. Pharmacol.* **2001**, *1*, 521–525; b) R.J. Gillies, *J. Cell. Biochem. (Suppl.)* **2002**, *39*, 231–238.
- 20 a) D.S. Smith, M.H. Al-Hakiem, J. Landon, *Ann. Clin. Biochem.* **1981**, *18*, 253–274; b) C.-H. Tung, N.-H. Ho, Q. Zeng, Y. Tang, F.A. Jaffer, G.L. Reed, R. Weissleder, *ChemBioChem* **2003**, *4*, 897–899.
- 21 J. Hirsch, *J. Clin. Invest.* **2003**, *111*, 1440–1443.
- 22 M.E. Phelps, *Proc. Natl Acad. Sci. USA* **2000**, *97*, 9226–9233.
- 23 W. Pham, Y. Choi, R. Weissleder, C.-H. Tung, *Bioconjugate Chem.* **2004**, *15*, 1403–1407.
- 24 G. Chen, D. J. Yee, N.G. Gubernator, D. Sames, *J. Am. Chem. Soc.* **2005**, *127*, 4544–4545.
- 25 S.A. Farber, M. Pack, S.Y. Ho, I.D. Johnson, D.S. Wagner, R. Dosch, M.C. Mullins, H.S. Hendrickson, E.K. Hendrickson, M.E. Halpern, *Science* **2001**, *292*, 1385–1388.
- 26 a) R.G. Blasberg, J.G. Tjuvajev, *J. Clin. Invest.* **2003**, *111*, 1620–1629; b) P.A. Binz, M. Muller, C. Hoogland, C. Zimmermann, C. Pasquarello, G. Corthals, J.C. Sanchez, D.F. Hochstrasser, R.D. Appel, *Curr. Opin. Biotechnol.* **2004**, *15*, 17–23.

Part I
High-throughput Screening

1

Quantitative Assay of Hydrolases for Activity and Selectivity Using Color Changes

Romas J. Kazlauskas

1.1 Overview

Color changes are the most convenient way to follow enzymatic reactions, but most enzymatic reactions do not directly cause a color change. (Reactions that consume or generate NAD(P)H are a notable exception, making them among the most widely used enzyme assays.) Chromogenic substrates conveniently introduce a color change into many reactions, but at the cost of using an artificial rather than a true substrate. One of the best solutions is to use the true substrate and then to couple the reaction of the substrate to a color change reaction. Many such coupled reaction assays have been developed. The focus in this chapter will be on reactions linked to pH changes, which are conveniently detected using pH indicators.

For many applications, substrate selectivity is the characteristic that distinguishes a useful enzyme from a non-useful one. To measure substrate selectivity, one must compare the reactions of two substrates. There are two ways to do this using color change reactions – separately or in the same reaction mixture. Measuring the rates of reactions of two substrates separately and then comparing these rates is the simplest way to measure selectivity. However, this approach only yields an estimate of selectivity. Enzyme selectivity stems from differences in the binding of the two substrates to the active site and from differences in the reaction rates of the two substrates. Measuring the reaction rate of each substrate separately may overlook differences in the binding of the two substrates to the active site.

Measuring selectivity by simultaneously measuring the reaction rates of both in the same reaction mixture seems at first to be a very special case, which would rarely apply to real world examples. However, by using a reference compound and two measurements, this approach can be applied to a wide range of compounds and, unlike separate measurements, can yield accurate selectivity values. The idea is to first measure the selectivity of the one true substrate and the reference compound. This reference compound is chosen to be convenient

to detect and may be a chromogenic substrate. Reaction of the true substrate may be detected by one color (e.g. a pH indicator), while reaction of the reference compound is detected by another color (the chromophore released upon reaction). This competitive experiment gives the true selectivity for this pair of substrates. The second step is to measure the selectivity of the second substrate toward the same reference compound. This yields a second true selectivity. Finally, since the reference compound was the same in both experiments, dividing the two gives the correct selectivity for the two true substrates. We refer to such colorimetric methods where the substrate selectivity is determined not by direct competition of the substrates with each other but in a two-step procedure by comparing reaction of each substrate with a reference compound, as “Quick selectivity methods.” This chapter will provide examples of using this Quick selectivity approach to measure enantioselectivity, diastereoselectivity as well as substrate selectivity.

1.2

Direct Assays Using Chromogenic Substrates

The simplest and most convenient assays are those involving a chromogenic substrate, that is, one where the reaction yields a product of a different color from the starting material. The most common chromogenic substrates to assay hydrolase activity are the *p*-nitrophenyl derivatives (Figure 1.1). In each case the *p*-nitrophenyl derivatives mimic the natural substrate of the hydrolase while adding a *p*-nitrophenyl moiety at the cleavage site. For lipases, phospholipases and esterases, these are esters of *p*-nitrophenol (e.g. 1, 2) [1]; for proteases, these are amides of *p*-nitroaniline (e.g. 3, 4) [2], and for glycosidases, these are glycosides of *p*-nitrophenol (e.g. 5) [3].

Other common chromogenic substrates are resorufin or 1-naphthol derivatives (Figure 1.2). Resorufin analogs have been used previously to spectrophotometrically monitor the activity of enzymes including galactosidases [4], cellulases [5], lipases [6], proteases, esterases, and phospholipases [7] and are particularly useful for detecting activity at low enzyme concentrations due to the intense absorption of resorufin, with an extinction coefficient approximately $70\,000\text{ M}^{-1}\text{ cm}^{-1}$. The 1-naphthol derivatives are used for activity staining of gels (e.g. 6). Hydrolysis yields 1-naphthol (7), which reacts with a diazo compound (Fast Red 8) to yield red-brown insoluble precipitate (9). Other chromogenic substrates are discussed in Chapters 9 to 11.

The disadvantage of chromogenic substrates is that they only mimic the true substrate of interest. One is never sure whether the true substrate will behave the same way as the chromogenic mimic.

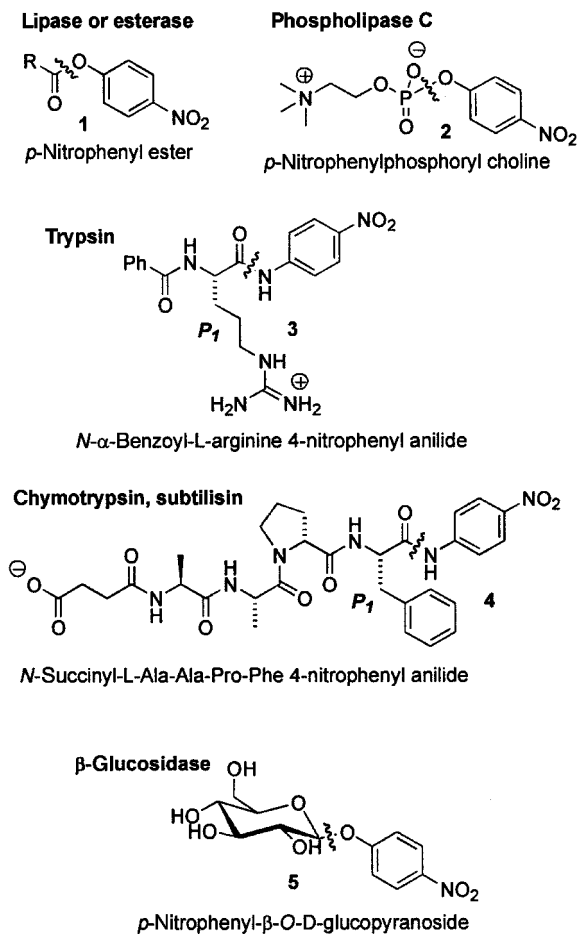


Fig. 1.1 Typical *p*-nitrophenyl derivatives for chromogenic assay of hydrolase activity. Each substrate includes a *p*-nitrophenyl moiety in the normal substrate in such a way that hydrolysis releases *p*-nitrophenol or *p*-nitroaniline, which can be detected by their yellow color. The wavy line marks the bond cleaved during hydrolysis.

1.3 Indirect Assays Using Coupled Reactions – pH Indicators

To detect reaction of a substrate lacking a chromogenic moiety, it is necessary to convert a product of the reaction into a colored compound. For example, hydrolysis of an acetate ester by a lipase or esterase can be detected using a coupled enzyme assay (see Chapter 2) [8]. Several sequential enzymatic reactions convert acetate to citrate accompanied by the formation of NADH, which absorbs at 340 nm. Another example, suitable for monitoring enzyme-catalyzed transesteri-

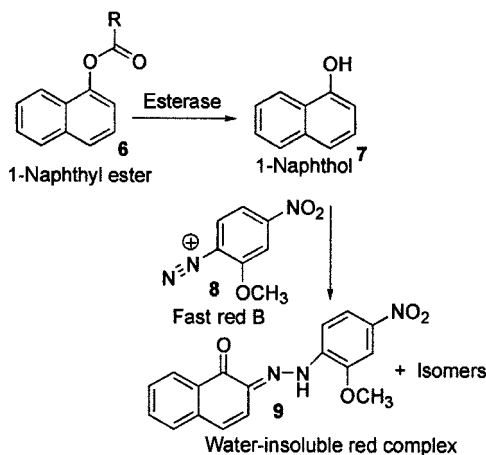


Fig. 1.2 Action of an esterase releases 1-naphthol (7) which reacts with the diazo compound Fast Red B (8) to give a water-insoluble azo dye (9).

fications in organic solvent, detects the acetaldehyde released upon transesterification of an alcohol (11) with a vinyl ester (10) to form ester (12) and acetaldehyde [9]. The acetaldehyde reacts with 4-hydrazino-7-nitro-2,1,3-benzoxadiazole (13) to yield a strongly fluorescent compound (14) (Figure 1.3). Both of the examples above are limited to certain substrates: acetate esters in the first case, transesterifications with vinyl esters in the second case.

A more general assay relies on pH indicators. Hydrolysis of an ester at neutral pH, for example, solketal butyrate (15) (butanoate ester of 2,2-dimethyl-1,3-dioxolane-4-methanol) to form solketal (16) below, releases a proton (Scheme 1.1). Detecting this proton release with a pH indicator allows one to monitor a wide range of ester hydrolyses. Researchers have used pH indicators to monitor the progress of enzyme-catalyzed reactions that release or consume protons since the 1940s [10, 11]. For example, researchers have monitored reactions catalyzed by amino acid decarboxylase [12], carbonic anhydrase [13], cholinesterase [14], hexokinase [15, 16], and ester hydrolysis by proteases [17]. In some cases, researchers used a pH indicator assay qualitatively, but in other cases, the color change was proportional to the number of protons released. This quantitative use requires a calibration with additional experiment or a careful choice of reaction conditions.

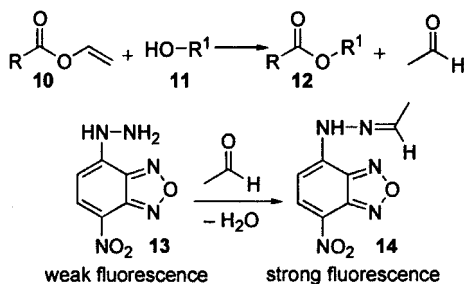
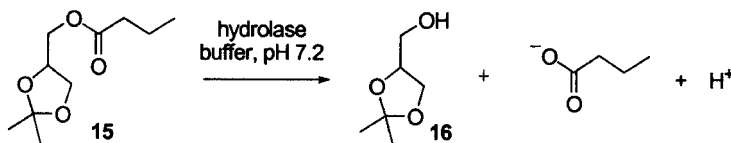


Fig. 1.3 Assaying the acylation of an alcohol with a vinyl ester in organic solvent relies on the detection of acetaldehyde.



Scheme 1.1

1.3.1

Overview of Quantitative Use of pH Indicator Assay

The most important variables in the pH indicator assay are the choices of buffer and pH indicator [18]. Both the buffer and the indicator must have the same affinity for protons (pK_a values within 0.1 unit of each other) so that the relative amount of buffer protonated and indicator protonated stays constant as the pH shifts during the reaction. A difference in pK_a of 0.3 units causes an 8% error when the pH changes by 0.1 unit [12]. In a typical assay, the pH changes by 0.05 units (10% hydrolysis of the substrate), thus, differences in pK_a values can lead to nonlinear and inaccurate rates. If different pK_a values cannot be avoided, accurate results can still be obtained by using calibration experiments [10] or a more complex equation [12].

If the pH indicator is the only species in the reaction solution that reacts with the released protons, then the rate of hydrolysis is equal to the rate of formation of the protonated form of the indicator. The color change in the solution reveals the protonation of the indicator as shown in Eq. (1), where $\Delta A/dt$ is the change in absorbance of the solution, $\Delta \epsilon$ is the difference in the extinction coefficient of the protonated and nonprotonated forms of the indicator and l is the path-length of the solution.

$$\text{rate} = \Delta[\text{indicator} \cdot \text{H}^+]/\text{time} = \frac{\Delta A/dt}{\Delta \epsilon \cdot l} \quad (1)$$

In practice, solutions containing only indicator are difficult to use because the color changes too readily due to small variation in conditions. For this reason, researchers usually add some buffer to the reaction mixture. Under these conditions, the rate of hydrolysis equals the rate of formation of the sum of the protonated indicator and the protonated buffer (Eq. 2).

$$\text{rate} = \Delta[\text{indicator} \cdot \text{H}^+]/\text{time} + \Delta[\text{buffer} \cdot \text{H}^+]/\text{time} \quad (2)$$

When the pK_a values of the indicator and buffer are the same, then released protons partition between the indicator and buffer according to the ratio of their concentrations (Eq. 3) [6, 7, 19]. The highest sensitivity (largest dA/dt) occurs with less buffer. The protons released add to either the buffer, giving no color change, or to the indicator, giving a color change. Thus, lowering the buffer concentration or increasing the indicator concentration increases the sensitivity

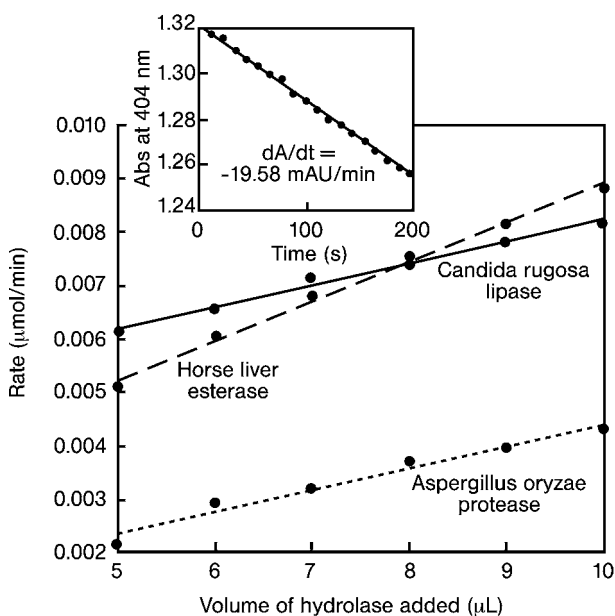
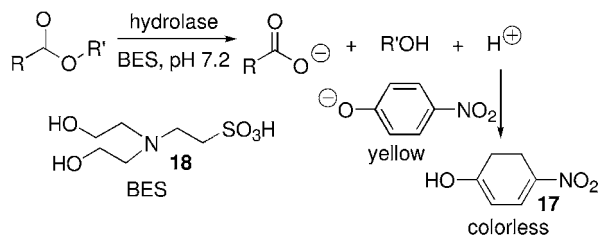


Fig. 1.4 4-Nitrophenol (17) as a pH indicator to detect hydrolysis of esters. Hydrolysis of an ester releases a proton, which protonates the yellow 4-nitrophenoxide ion. The graph shows experimental data for hydrolysis of solketal butyrate (15) by several hydro-

lases. The inset shows the linear loss of yellow color as a function of time, which yields the rate of reaction using Eq. (3). The graph shows that the measured rate of hydrolysis increases with increasing enzyme amount as expected. (Data from [18]).

of the assay. When the buffer to indicator concentration is high, the $(1 + [\text{buffer}]/[\text{indicator}])$ term can be replaced with $([\text{buffer}]/[\text{indicator}])$:

$$\text{rate} = \frac{\Delta A/dt}{\Delta \epsilon \cdot l} + \frac{[\text{buffer}]}{[\text{indicator}]} \frac{\Delta A/dt}{\Delta \epsilon \cdot l} = \frac{\Delta A/dt}{\Delta \epsilon \cdot l} \left(1 + \frac{[\text{buffer}]}{[\text{indicator}]} \right) \quad (3)$$

Since most hydrolases have maximal activity near neutral pH, we developed the assay for pH 7.2. As a pH indicator, we chose 4-nitrophenol 17 (Figure 1.4). The similarity of its pK_a (7.15) to the pH of the reaction mixture ensures that changes in pH give a large and linear color change [6]. The large difference in the extinction coefficients of the protonated and deprotonated forms (200 versus

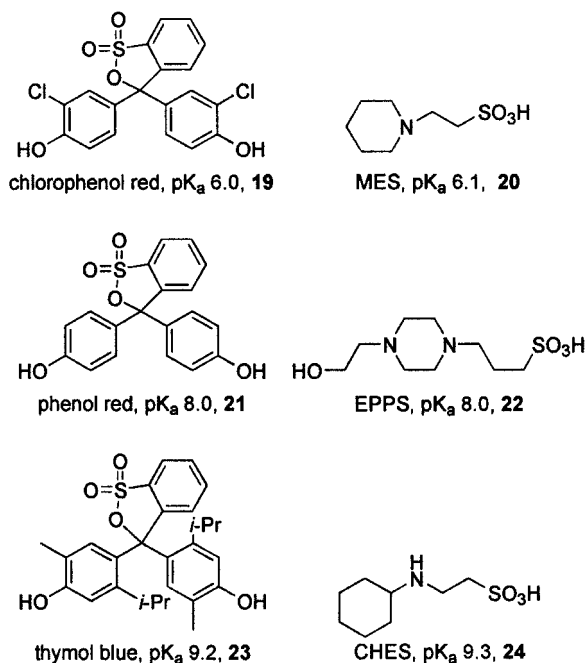


Fig. 1.5 Buffer/indicator pairs with similar pK_a values that may be suitable for screening at pH 6, 8, or 9.2.

$18000 \text{ M}^{-1} \text{ cm}^{-1}$ at 404 nm) gives good sensitivity. Finally, nitrophenols bind less to proteins than some polyaromatic indicators [20]. The concentration of the pH indicator should be as high as possible to maximize sensitivity (see Scheme 1.1). In our assay, the high initial absorbance of 4-nitrophenoxide/4-nitrophenol limited the concentration to 0.45 mM. (The pathlength in a 96-well plate depends on the volume of the solution in the well since the light passes from the top of the plate through the solution. Thus, the maximum indicator concentration varies with the solution volumes. With other acid–base indicators, poor water solubility can also limit the maximal concentration.) This concentration gave a starting absorbance of ~ 1.2 where the accuracy is still not compromised by low light levels. As a buffer, we chose BES (*N,N*-bis[2-hydroxyethyl]-2-aminoethanesulfonic acid) (**18**) because its pK_a (7.15) [21] is identical to that of 4-nitrophenol (**17**). We chose a buffer concentration of approximately 5 mM as a compromise between low buffer concentrations to maximize sensitivity (see Scheme 1.1), and high buffer concentrations to ensure accurate measurements and small pH changes throughout the assay (<0.05 pH units for 10% hydrolysis at our conditions). The small pH changes are important because kinetic constants can change with changing pH.

Other buffer/indicator pairs may be suitable for screening at other pHs. For example, at pH 6 chlorophenol red (**19**) (pK_a 6.0) and MES (2-[*N*-morpholino]-

ethanesulfonic acid, pK_a 6.1) (**20**) may be suitable; at pH 8 phenol red (**21**) (pK_a 8.0) and EPPS (*N*-[2-hydroxyethyl]piperazine-*N'*-[3-propanesulfonic acid], pK_a 8.0) (**22**) may be suitable; at pH 9 thymol blue (pK_a 9.2) (**23**) and CHES (2-[*N*-cyclohexylamino]ethanesulfonic acid, pK_a 9.3) (**24**) may be suitable (Figure 1.5).

Suitable substrate concentrations ranged from 0.5 to 2 mM, typically 1 mM. At substrate concentration below 0.5 mM, the absorbance changes are too small to be detected accurately. For example, hydrolysis of 5% of a 0.25 mM substrate concentration at our standard conditions (pH 7.2, 0.45 mM 4-nitrophenol (**17**), 5 mM BES (**18**)), changes the absorbance by only 0.005 absorbance units. Solubility in water sets the upper limit of substrate concentration because spectrophotometric measurements require clear solutions. Typical organic substrates dissolve poorly in water, so we added organic cosolvent – 7 vol% acetonitrile. For very insoluble substrates, we prepared clear emulsions using detergents. The assay tolerates small changes in reaction conditions, such as the addition of 7% acetonitrile. Indeed, the pK_a of 4-nitrophenol (**17**) changes only slightly from 7.15 to 7.17 upon addition of 10% ethanol. This result suggests that cosolvent concentrations below 10% do not compromise the accuracy of the assay. Also, small amounts of salts present in the hydrolase solutions (buffer salts in commercial hydrolase preparations, 2 mM $CaCl_2$ in the protease solutions) did not affect the accuracy.

1.3.2

Applications

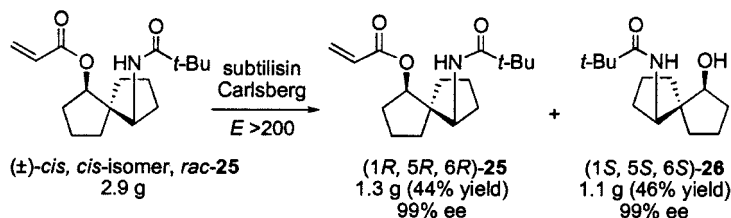
The pH indicator-based assay is a quick way to measure whether an enzyme accepts a particular substrate. This measurement quickly eliminates unsuitable enzymes or unsuitable substrates from more complicated subsequent experiments. It is most useful with unusual substrate/enzyme combinations where only a few are likely to work.

1.3.2.1 Searching for an Active Hydrolase

(Testing Many Hydrolases Toward One Substrate)

One application of this assay is for a rapid survey of commercial enzymes to find those that accept an unnatural or unusual substrate. In one project, we screened a library of 100 commercial hydrolases to find the two that catalyzed hydrolysis of hindered spiro compound **25** that was useful as a chiral auxiliary [22]. This screening identified several candidate hydrolases. Subsequent confirmation of this activity and optimization of the reaction led to a subtilisin-catalyzed preparative scale resolution of this auxiliary (Scheme 1.2).

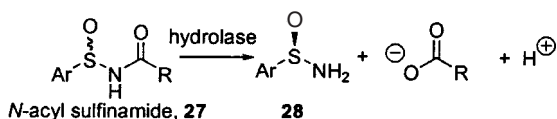
In a second project, we screened a library of 72 commercial hydrolases (lipases, esterases, and proteases) towards solketal butyrate (**15**) (see Scheme 1.1), a chiral building block [18]. Most hydrolases reacted with this less hindered substrate so this initial screen eliminated only 20 of the 72 candidate enzymes from the next stage of screening. In a third project, we again assayed a library of 100 hydrolases to find the few that catalyzed hydrolysis of the amide link in *N*-acyl



Scheme 1.2

sulfonamides (**27**) to form sulfonamides (**28**) [23]. Although hydrolysis of amides usually does not work in this assay because it does not release a proton, the *N*-acyl sulfonamide (**27**) was a special case. The sulfonamide (**28**) released is not a strong base like an amine and does not take up the proton released from the ionization of the acid, so the pH indicator assay was suitable for this special amide hydrolysis (Scheme 1.3).

Other researchers recently used similar pH indicator assays to measure the activity of glycosyl transferase [24], haloalkane dehalogenase [25], and nitrilase [26].



Scheme 1.3

1.3.2.2 Substrate Mapping of New Hydrolases (Testing Many Substrates Toward Hydrolase)

Another application of this assay is mapping the substrate range of newly discovered hydrolases, such as those from thermophiles. Advances in microbiology and molecular biology have made hundreds of new enzymes available from unusual microorganisms such as thermophiles. These enzymes may be useful for organic synthesis because they allow the use of higher temperatures, where reactions are faster and substrates are more soluble [27]. In addition, these enzymes may also tolerate other unusual conditions such as high concentrations of organic solvents. To apply these new enzymes in organic synthesis, one needs some idea of their substrate range and selectivity. The pH indicator-based assay is a quick way to map the substrate range of new enzymes. Recently, we, and others, used this pH indicator assay to map over 20 different esterases using the substrate library of more than 50 esters (Figure 1.6) [28–30].

This substrate mapping first grouped each substrate/esterase combination as slow (≤ 1 mmol ester hydrolyzed mg^{-1} protein min^{-1}), good (1–10) or very good (> 10). The 19 esterases and 31 substrates gave a total of 589 substrate/esterase combinations. Examination of these data yielded several conclusions. The best substrates were activated esters, which were also chemically the most reactive in the library: vinyl esters (**30**) and phenyl esters (**32**). Among the vinyl and ethyl

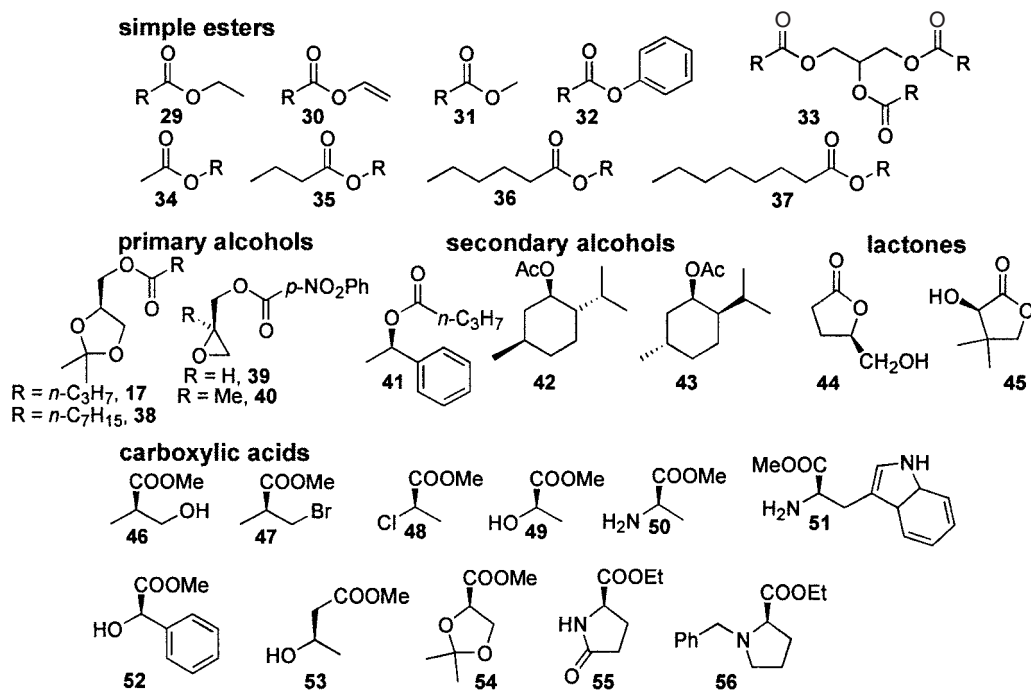


Fig. 1.6 Examples of esters used to survey the substrate selectivity of esterases.

esters with different chain lengths, most of the esterases appear to favor the hexanoate (36) or octanoate (37), while one esterase (E018b) appeared to strongly favor acetate esters (34). Esters with a sterically hindered acyl group (vinyl pivaloate 30, $R=t$ -butyl or vinyl benzoate 30, $R=Ph$) and polar esters, e.g. methyl 2-hydroxyacetate 31, $R=CH_2OH$, were usually poor substrates. Three pairs of esterases (E010/E011, E013/E014, E019/E020) showed similar reaction rates with all substrates and therefore might be very similar or even identical enzymes. Thus, this initial substrate mapping provided a good overview of substrate range and ideas for the further refinement of these general conclusions. This further refinement usually involves a quantitative measure of substrate selectivity, where one includes reference compounds in the screening (see below).

1.3.3

Comparison with Other Methods

There are several advantages with this screening method using pH indicators. First, it is hundreds of times faster than conventional screening. The 96-well format allows the analysis of large numbers of samples simultaneously. The method is quantitative, unlike screening for hydrolytic activity by thin-layer chromatography (TLC). Second, since the entire reactions and analyses take

place in the microplate wells, workup and analysis by gas chromatography (GC), high-performance liquid chromatography (HPLC) or nuclear magnetic resonance (NMR) is avoided. Third, it requires hundreds to thousands of times less substrate (typically 20 mg per well) and hydrolase (we used between 0.6 and 35 mg protein per well). Fourth, this assay measures the hydrolysis of any ester, not just chromogenic esters. The most important rule of screening is “You get what you screen for,” so the ability to screen the target compound, not an analog of the target compound, is an important advantage.

There are also few disadvantages with our screening method. This assay is approximately seven times less sensitive than one using hydrolysis of 4-nitrophenyl esters (e.g. **1**) (Figure 1.1). For example, if the rate of hydrolysis towards a nonchromogenic ester and a 4-nitrophenyl ester were identical, then our assay would require seven times more hydrolase to observe the same change in absorbance. The assay with 4-nitrophenyl esters releases one molecule of 4-nitrophenol (**17**) (53% of these will be deprotonated at pH 7.2), while our assay protonates one 4-nitrophenoxide for every 12 protons released. Second, this assay requires clear solutions. To obtain clear solutions with water-insoluble substrates, experimentation is sometimes required to find the best cosolvent or emulsion conditions. Third, the pH indicator method requires a reaction that generates or consumes protons, so it cannot be used for most amide hydrolysis or glycoside hydrolysis.

1.4 Estimating and Measuring Selectivity

One of the most useful characteristics of enzymes is their selectivity. The true selectivity of an enzyme toward a pair of substrates is the ratio of the specificity constants ($k_{\text{cat}}/K_{\text{m}}$) for each substrate [31]. For example, the enantioselectivity (or enantiomeric ratio, E) of an enzyme is the ratio of the specificity constants for the enantiomers (Eq. 4).

$$\text{Enantiomeric ratio} = E = \frac{(k_{\text{cat}}/K_{\text{m}})_{\text{fast enantiomer}}}{(k_{\text{cat}}/K_{\text{m}})_{\text{slow enantiomer}}} \quad (4)$$

Other examples of selectivity are substrate selectivity (e.g. selectivity for different acyl chain length among esters), regioselectivity (e.g. selectivity for one ester of several esters in a nucleoside derivative), and diastereoselectivity (e.g. selectivity for a cis isomer over a trans isomer). It is rarely convenient to measure selectivity by measuring the kinetic parameters for each substrate. Such measurement requires multiple kinetic studies, which are slow and tedious. In addition, combining four such measured values, each with an uncertainty of measurement, yields selectivities with large uncertainties.

The best methods to measure selectivities are competitive experiments. A mixture of both substrates competes for the enzyme active site and is converted to product. Measuring the amounts of each product formed reveals the selectivity.

One such method is the enantioselectivity determination method developed by C. J. Sih's group [32]. This method is the "gold standard" to which other methods are compared. To measure *E*, researchers run a test resolution, work up the reaction and measure two of the following: enantiomeric purity of the starting material (ees), enantiomeric purity of the product (eep), or conversion (c). The difficulty with this method is mainly the measurement of enantiomeric purity, which can be time-consuming. Screening hundreds of commercial enzymes or cultures of microorganisms by this method is difficult without laboratory automation.

1.4.1

Estimating Selectivity without a Reference Compound

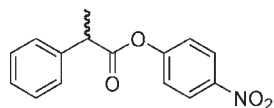
In some cases, a qualitative estimate of selectivity is sufficient. In these cases, the measured rates of reaction in two experiments are simply compared (Eq. 5). For example, we estimated the enantioselectivity of several lipases toward 4-nitrophenyl-2-phenylpropanoate (**57**) by measuring the initial rates of hydrolysis of each enantiomer (Table 1.1). The true *E*-values came from a competitive experiment where the enantiomeric purity of the products and starting materials

Table 1.1 True enantioselectivity and estimated enantioselectivity of several hydrolases toward 4-nitrophenyl 2-phenylpropanoate (**57**). (All data from [34]).

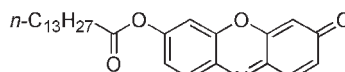
Hydrolase	True <i>E</i> ^{a)}	Estimated <i>E</i> ^{b)}	Quick <i>E</i> ^{c)}
<i>Pseudomonas cepacia</i> lipase	29	20	29
<i>Candida rugosa</i> lipase	3.5	1	3.5
Purified <i>Candida rugosa</i> lipase	>100	40	210
Porcine pancreatic lipase	1.1	1.4	1.4
<i>Candida antarctica</i> lipase A	1.4	4	2.3

- a) True *E* was measured by hydrolysis to approximately 40% conversion of a racemic sample of 4-nitrophenyl 2-methylpropanoate (**57**) with the indicated hydrolase followed by measurement of the enantiomeric purity of the product and remaining starting material by high-performance liquid chromatography using a chiral stationary phase.
- b) Quick *E*-values were measured resorufin tetradecanoate (**58**) as a reference compound.
- c) Estimated *E* was measured by comparing the initial rate of hydrolysis of pure enantiomers of 4-nitrophenyl 2-methylpropanoate (**57**) separately. Initial rates were measured by monitoring the release of 4-nitrophenol (**17**) spectrophotometrically.

In all cases, hydrolysis of (*S*)-**57** was favored.



4-nitrophenyl 2-phenylpropanoate, **57**



resorufin tetradecanoate, **58**

were measured. The estimated E came from a simple comparison of the initial rate of reaction for the two enantiomers separately. In this case, a pH indicator was not needed to detect hydrolysis because hydrolysis yields the yellow *p*-nitrophenoxide directly. The agreement is qualitatively good since the estimated E identified enantioselective and nonenantioselective enzymes. However, there is clearly no quantitative agreement. (The Quick E -values, which do agree quantitatively with the true E -values will be discussed below.)

$$\text{estimated } E = \frac{\text{rate of fast enantiomer measured separately}}{\text{rate of slow enantiomer measured separately}} \quad (5)$$

In another example, we estimated the enantioselectivity of the 52 hydrolases toward solketal butyrate (**15**) by separately measuring the initial rates of hydrolysis of the pure enantiomers. We used pH indicators to measure the rates of hydrolysis and then presented the ratio of these rates as an estimated enantioselectivity [33]. A few typical results are shown in Table 1.2. The true enantioselectivity and the estimated enantioselectivity agreed to within a factor of 2.3. This qualitative screening identified horse liver esterase as a new hydrolase for the resolution of solketal butyrate (**15**) with modest enantioselectivity.

In a last example, estimated selectivities differed so much from the true selectivities that they were useless. In this case we estimated the diastereoselectivity of 91 commercial hydrolases towards the *cis*-(2*S*, 4*S*) and *trans*-(2*R*, 4*S*) dioxolanes (**59a/b**) by measuring the initial rates of hydrolysis of the two pure diastereomers separately. The ratio of the two rates is the estimated diastereoselectivity (Table 1.3). Three hydrolases – cholesterol esterase, α -chymotrypsin and sub-

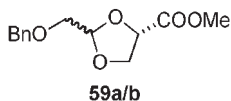
Table 1.2 True enantioselectivity and estimated enantioselectivity of several hydrolases toward solketal butyrate (**15**). (All data from [18]).

Hydrolase	True $E^a)$	Estimated $E^b)$
Horse liver esterase	15	12
<i>Rhizopus oryzae</i> lipase	5.0	11
Cutinase	5.0	2.3
<i>Candida rugosa</i> lipase	3.0	4.9
Esterase E013	1.0	1.0

- a) True E was measured by hydrolysis to approximately 40% conversion of a racemic sample of solketal butyrate (**15**) with the indicated hydrolase followed by measurement of the enantiomeric purity of the product and remaining starting material by gas chromatography using a chiral stationary phase.
- b) Estimated E was measured by comparing the initial rate of hydrolysis of pure enantiomers of solketal butyrate (**15**) separately. Initial rates were measured using a pH indicator (4-nitrophenol **17**) to detect release of acid upon hydrolysis of solketal butyrate (**15**).

With the exception of horse liver esterase, all enzymes favored (*R*)-solketal butyrate.

Table 1.3 True diastereoselectivity, estimated diastereoselectivity and Quick *D* of several hydrolases toward diastereomers of methyl (4*S*)-2-benzyloxy-1,3-dioxolane-4-carboxylate (**59a/b**). (All data from [37]).



Hydrolase	True <i>D</i> ^{a)}	Estimated <i>D</i> ^{b)}	Quick <i>D</i>
Cholesterol esterase	13	160	17
Subtilisin	6.3	> 100	4.4
α -Chymotrypsin	7.7	> 100	6.4

a) True *D* was measured by hydrolysis to approximately 40% conversion of a mixture of diastereomers (**59a/b**) with the indicated hydrolase followed by measurement of the diastereomeric purity of the product and remaining starting material by proton NMR.

b) Estimated *D* was measured by comparing the initial rate of hydrolysis of pure diastereomers of **59a** and **59b** separately. Initial rates were measured by monitoring the release of 4-nitrophenol (17) spectrophotometrically.

All hydrolases favored the *trans* stereoisomer (**59b**).

tilisin from *Bacillus licheniformis* – showed excellent (>100) estimated diastereoselectivity. Unfortunately, the true diastereoselectivity was much lower (4.4–17), so this estimated selectivity screening was not accurate enough to identify selective enzymes.

The main advantages of measuring estimated selectivity are speed and simplicity, while the main disadvantage is the risk that the estimate is so far off from the true value that it is misleading. Typical screening and data workup times for a library of 100 enzymes are several hours. In practice, we often include an estimated selectivity screening before the quantitative screen involving a reference compound (see below) to group the substrates into fast-, medium-, and slow-reacting. This information helps choose a good reference compound.

1.4.2

Quantitative Measure of Selectivity Using a Reference Compound (Quick *E* and Related Methods)

These estimated selectivities are only estimates because they do involve competition between the substrates. By measuring initial rates of the two substrates separately, we eliminate competitive binding between the two diastereomers. Both k_{cat} and K_{m} contribute to the overall selectivity of an enzyme, thus eliminating competitive binding can lead to inaccuracies, as the example in Table 1.3 above showed. To understand the contribution of binding to selectivity, imagine a hypothetical case where both substrates have the same k_{cat} values, but differ-

ent K_m values. In a competitive experiment, the enzyme will bind and transform the better binding substrate. The reaction is selective. However, if the hydrolysis of the two substrates occurs in separate vessels, the result depends on the substrate concentration. At saturating amounts of substrate, both reactions will proceed at the same rate and estimated selectivity will incorrectly indicate that the reaction is nonselective. At substrate concentrations well below K_m , the estimated selectivity will be close to the true selectivity. One can imagine other combinations of reaction conditions and kinetic parameters that can lead to either overestimation or underestimation of the true selectivity.

Measuring the true selectivity requires a competitive experiment. Adding two substrates to the reaction mixture creates a competitive experiment, but complicates analysis. One must be able to distinguish the reaction of each substrate individually. Methods like HPLC or GC can distinguish the two substrates, but are much slower than colorimetric methods. An alternative is to use one true substrate, whose hydrolysis can be measured by the pH indicator and a second, chromogenic reference compound, whose hydrolysis can be measured by formation of a color different from the pH indicator. Such an experiment would yield the correct selectivity for true substrate 1 and the reference compound (Eq. 6).

$$\frac{\text{substrate 1}}{\text{reference}} \text{selectivity} = \frac{\text{rate of substrate 1 reaction}}{\text{rate of reference reaction}} \cdot \frac{[\text{reference}]}{[\text{substrate 1}]} \quad (6)$$

In a second experiment, a second substrate competes against the same reference compound to yield the correct selectivity for true substrate 2 and the reference compound (Eq. 7).

$$\frac{\text{substrate 2}}{\text{reference}} \text{selectivity} = \frac{\text{rate of substrate 2 reaction}}{\text{rate of reference reaction}} \cdot \frac{[\text{reference}]}{[\text{substrate 2}]} \quad (7)$$

Finally, dividing these two selectivities yields the desired selectivity of substrate 1 versus substrate 2 (Eq. 8).

$$\frac{\text{substrate 1}}{\text{substrate 2}} \text{selectivity} = \frac{\frac{\text{substrate 1}}{\text{reference}} \text{selectivity}}{\frac{\text{substrate 2}}{\text{reference}} \text{selectivity}} \quad (8)$$

We define Quick selectivity methods (Quick *E* for enantioselectivity, Quick *D* for diastereoselectivity, Quick *S* for substrate selectivity) as methods that, instead of letting the substrates compete directly against one another, do so indirectly by competing each substrate against a reference compound.

1.4.2.1 Chromogenic Substrate

In the first Quick *E* experiments, we used chromogenic substrates and a chromogenic reference compound [34]. Hydrolysis of pure enantiomers of 4-nitrophenyl 2-phenylpropanoate (**57**) liberates the yellow *p*-nitrophenoxide ion (**7**). The increase in absorbance at 404 nm revealed the initial rates of hydrolysis of each enantiomer, but the ratio of these rates gave only an enantiomeric ratio (see Table 1.1). The ratio of rates over- or underestimated *E* by as much as 70%. To reintroduce competition, we added resorufin tetradecanoate (**58**) as a reference compound. We monitored the initial rates of hydrolysis of 4-nitrophenyl (*S*)-2-phenylpropanoate ((*S*)-**57**) at 404 nm and the reference compound **58** at 572 nm in the same solution (Figure 1.7).

Two experiments, one with (*S*)-enantiomer of **57** competing against the reference compound **58** and the second with the (*R*)-enantiomer of **57** competing against the reference compound **58** yielded the data to calculate the Quick *E*-values according to Eqs (6) to (8). These Quick *E*-values agreed with the values measured by the slower methods involving direct competition of the enantiomers (*R*)- or (*S*)-**57** and analysis of the enantiomers by HPLC (Table 1.1). We measured low (*E*=1.4), average (*E*=27), and excellent (*E*=210) enantioselectivities correctly by this technique. Each hydrolysis experiment requires 30 s; thus, the measurement time for *E* was only 1 min.

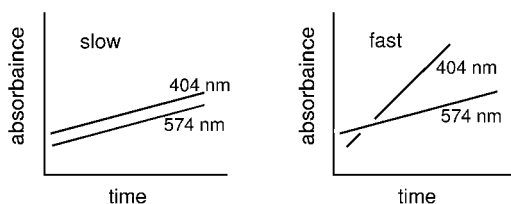
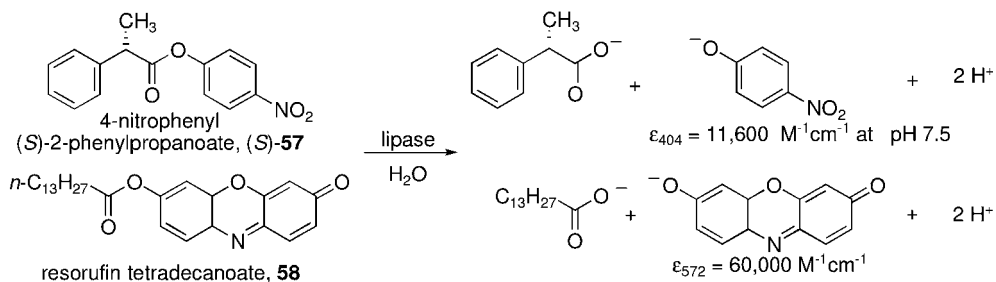


Fig. 1.7 The first step of the Quick *E* measure of enantioselectivity of lipases toward 4-nitrophenyl 2-phenylpropanoate (**57**). Lipase-catalyzed hydrolysis of 4-nitrophenyl (*S*)-2-phenylpropanoate ((*S*)-**57**) and the reference compound, resorufin tetradecanoate (**58**), yields yellow and pink chromophores,

respectively. The solution turns deep orange if both substrates are hydrolyzed; pink if only the reference compound is hydrolyzed. The second step of the Quick *E* is the same, except that it uses the (*R*)-enantiomer of the chiral ester. Equations (6) to (8) yield the enantioselectivity.

1.4.2.2 pH Indicators

To apply the Quick *E* ideas to nonchromogenic substrates, we used pH indicators to detect the hydrolysis of the substrate. The reference compound remained a chromogenic resorufin ester, acetate (**60**). We first demonstrated this to measure diastereoselectivity of the dioxolane stereoisomers **59a/b** in Table 1.2 (Figure 1.8).

The Quick *D* measurements agreed with the true values determined by NMR using a direct competitive hydrolysis of the two diastereomers.

1.4.3

Application

1.4.3.1 Substrate Mapping of Hydrolases

Mapping the substrate selectivity of hydrolases using Quick *S* screening was dramatically faster than traditional methods. For example, the mapping of the acyl chain length selectivity of a group of esterases from thermophiles confirmed the high acetyl preference of one esterase (E018b) (Figure 1.9).

We determined the true selectivity of the thermophile esterases toward different acyl chain length using a competitive experiment usually using resorufin acetate (**60**) as the reference compound. However, accurate measurement of the selectivities requires that the substrate and the reference compound react at comparable rates to measure both reaction rates accurately. When the substrate hydrolysis was much slower than resorufin acetate (**60**) hydrolysis, we replaced resorufin acetate (**60**) with the slower reacting resorufin pivaloate (**62**) or resorufin isobutyrate (**63**) as reference compounds and extended the measurement time. The trends for the true selectivities were similar to the estimated selectivities. Among the vinyl esters, the hexanoate or octanoate was the best substrate for all except E018b, where the acetate was the best substrate. We confirmed this high acetyl selectivity of E018b by a direct competition between vinyl butyrate and acetate monitored by ¹H-NMR. The resonances of both substrate vinyl esters, and the products, butyric and acetic acid, were monitored over time and revealed a 17-fold acetyl preference. In comparison the Quick *S* measurements showed a 24-fold acetyl preference. Surprisingly, acetyl esterase from orange peel showed only a fourfold preference for acetyl.

1.4.3.2 Screening of Mutants in Directed Evolution

An important application of Quick *E* screening was search for more enantioselective mutants in a library of randomly mutated esterases. For this application, the ability to measure enantioselectivity accurately was important to find those mutants that only increased enantioselectivity moderately. To increase the enantioselectivity of PFE toward hydrolysis of methyl 3-bromo-2-methylpropanoate (MBMP) (**64**) ($E_{WT}=12$ favoring (*S*)-**64**), we used random mutagenesis by an *Escherichia coli* mutator strain [35]. Screening of 288 crude cell lysates with Quick *E* revealed that most of the mutants had enantioselectivities near that of the wild

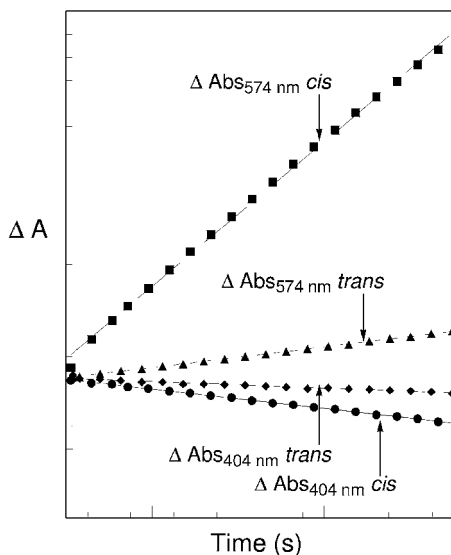
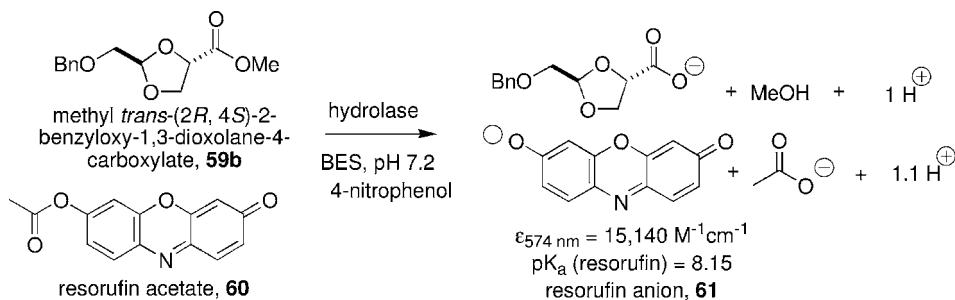


Fig. 1.8 The first selectivity step of measuring the Quick *D* diastereoselectivity of a hydrolase toward the 2*R* and 2*S* diastereomers of methyl (4*S*)-benzyloxy-1,3-dioxolane-4-carboxylate (**59a/b**). Enzyme-catalyzed hydrolysis of *trans*-(2*R*, 4*S*) diastereomers (**57b**) in the presence of the reference compound, resorufin acetate (**60**), releases protons which produce a decrease in the yellow absorbance of the pH indicator, and a brilliant pink chromophore, resorufin anion **61**. Both of the absorbance changes, at 404 nm and 574 nm respectively, can be monitored simultaneously. The second step of the Quick *D* uses the *cis*-(2*S*, 4*R*) diaster-

eomer (**59a**). The graph shows experimental data for the Quick *D* measurement with bovine pancreatic protease (Quick *D*=12.8). The selectivity ratio of the *trans*-(2*R*, 4*S*) diastereomer (**59b**) versus resorufin acetate (**60**) is 1.64 (Eq. 6) while the selectivity ratio of the *cis*-(2*S*, 4*S*) diastereomer (**59a**) versus resorufin acetate (**60**) is 0.128 (Eq. 7). The ratio of the two selectivity ratios equals the diastereoselectivity, 12.8 (Eq. 8). Conditions in the well during assay: 2.0 mM *cis* or *trans* dioxolane methyl ester (**59**), 0.1 mM resorufin acetate (**60**), 4.65 mM BES (**18**), 0.434 mM 4-nitrophenol (**17**), 0.86% Triton X-100, 7% acetonitrile. (Data from [34]).

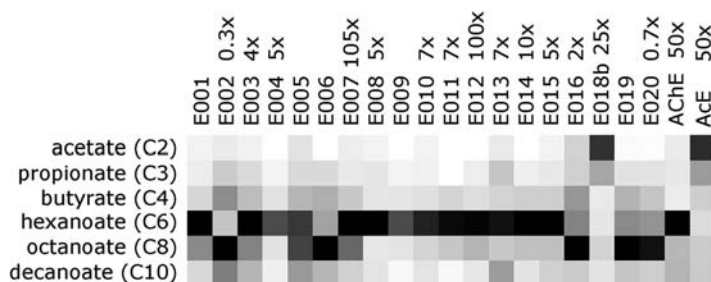


Fig. 1.9 Acyl length selectivity of thermophile esterases toward vinyl esters. Darker squares correspond to higher activity. Most esterases favor either vinyl hexanoate or octanoate, but E018b and AcE (acetyl esterase from orange peel) favor vinyl acetate. To show

selectivity clearly, the absolute enzyme activities were scaled as indicated. This grayscale array representation was created as described by Reymond and coworkers [38]. AcHE=acetylcholine esterase. (Data from [30]).

type (Quick $E=12$), but one mutant, MS6-31, showed significantly higher enantioselectivity (Quick $E=21$) (Figure 1.10 and Table 1.4). DNA sequencing revealed a C₆₈₉T transition, which changes Thr230 to isoleucine. This mutant showed $E=19$ in a scale-up reaction in good agreement with Quick E -value. For this and the Thr230Pro mutant in Table 1.4, the differences between Quick E and the true E are within experimental error even though we measured Quick E -values on cell lysates, but measured endpoint method values on purified enzyme.

The next generation of this directed evolution project involved focusing mutations closer to the substrate-binding site [36]. The much more enantioselective mutants at positions 28 and 121 identified in this project also highlighted a weakness of the Quick E screening method. The Quick E -values proved more difficult to measure because the rate reaction of the slow enantiomer was so

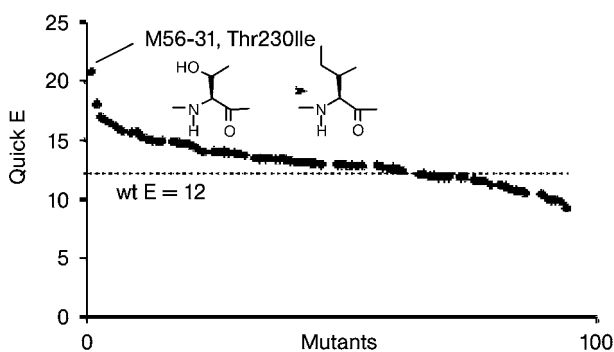
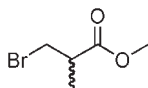


Fig. 1.10 Mutants of *Pseudomonas fluorescens* esterase generated by random mutagenesis of the entire protein and ordered from highest to lowest enantioselectivity. The flat central part of the curve represents colonies where there was little or no change in enantioselectivity because they either contained no mutations or mutations that did not affect enantioselectivity.

Table 1.4 Enantioselectivities of wild-type and selected PFE mutants toward methyl 3-bromo-2-methylpropanoate (MBMP) (**64**).methyl 3-bromo-2-methylpropanoate, **64**

Mutants	Quick E^a	True E^b
Wild type	12	12 ± 1
Thr230Ile	21	19 ± 2
Thr230Pro	19	17
Trp28Leu	100 ± 20	58 ± 7
Trp28Phe	39	32 ± 1
Trp28Tyr	23	29 ± 1
Val121Ser	104 ± 13	61 ± 1
Val121Met	76	36 ± 1

- a) Colorimetric measurement in microplates using *p*-nitrophenol (**17**) as the pH indicator and resorufin acetate (**60**) or resorufin isobutyrate (**63**) as the reference compound. Wild type and all mutants favored (*S*)-**64**.
- b) Small-scale reactions (50 mmol) where the enantiomeric purity of the products and starting materials were measured by gas chromatography on a chiral stationary phase. Standard errors are shown for those measurements that were repeated several times.

slow that it was difficult to measure accurately. First, we increased the concentration of substrate, but solubility limited how much we could add. We could not simply add more enzyme because then the reference compound reacted too quickly. Our solution was a slower reacting reference compound – resorufin isobutyrate (**63**) – in place of resorufin acetate (**60**) (Figure 1.11). This reference compound **63** reacted about seven times more slowly with *Pseudomonas fluorescens* esterase than the acetate **60**. With new reference compound **63**, we could measure the higher enantioselectivities (Table 1.4). The errors for measuring the higher enantioselectivity remained higher than that for low enantioselectivity. For this application, this accuracy was sufficient to identify the highly enantioselective mutants. For applications that require measuring high enantioselectivities accurately with Quick E , we are developing even slower-reacting reference compounds.

1.4.4

Advantages and Disadvantages

The main advantage of Quick E is speed, which comes from four differences between Quick E , which involves competition with a reference compound, and true E methods, which involve direct competition of the two enantiomers. First,

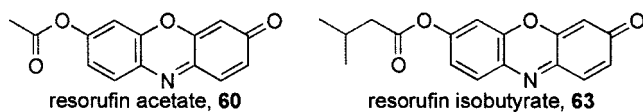


Fig. 1.11 Resorufin isobutyrate (**63**) reacts approximately seven times slower than resorufin acetate (**60**), which makes it a better reference compound for slow-reacting substrates.

and most important, the Quick *E* screening eliminates the need to measure enantiomeric purity. Measuring enantiomeric purity for substrate mapping would require several different analytical methods and probably several different chiral HPLC and GC columns. Each measurement would require at least an hour. Second, Quick *E* screening measures only initial rates (>5% conversion) so there is no need to wait for the reaction to reach 30 or 40% conversion. Third, the colorimetric methods allow us to follow 96 reactions simultaneously using a microplate reader. Fourth, the small scale of the reaction (typically 0.1 mmol) allows us to use more enzyme per mole of substrate and thus the reaction is faster. These advantages all contribute to making Quick *E* a much faster method than true *E* methods. Quick *E* is ideal for rapidly measuring the enantioselectivity of large numbers of samples and it is based on the same equations and assumptions as the true *E* methods.

One disadvantage of Quick *E* is that it requires pure enantiomers, albeit in small amounts. In many cases, small amounts are available from preparative HPLC with chiral stationary phases or other methods. Another disadvantage is the need for clear aqueous solutions of substrate. Dissolving hydrophobic substrates in aqueous solution, even with the help of surfactants, is sometimes difficult.

Measuring high enantioselectivity is challenging for both the true *E* methods and the Quick *E* methods. For the true *E* methods, measuring high enantioselectivity requires the slow-reacting enantiomer to be detected accurately. For the Quick *E* methods, measuring high enantioselectivity requires the rates of reaction of the slow-reacting enantiomer and a reference compound to be measured simultaneously. This measurement will normally require a slower reacting reference compound. Another limit to measuring high enantioselectivity with Quick *E* is the enantiomeric purity of the starting slow enantiomer. If the slow-reacting enantiomer is contaminated with the fast-reacting one, then the measured Quick *E*-value will be lower than the true *E*.

Acknowledgments

I am grateful to Lana E. Janes, who developed the pH indicator screening and Quick *E* during her PhD studies at McGill University and to the coworkers who tested, refined, and applied these methods to screening projects in biocatalysis: A. C. Löwendahl, A. Man Fai Liu, N. Somers, D. Yazbeck, V. Magloire, G. Hors-

man, P. Mugford, and K. Morley. The Quick *E* research was funded by the Petroleum Research Fund administered by the American Chemical Society and the Natural Sciences and Engineering Research Council (Canada).

References

- 1 S. Kurioka, M. Matsuda, *Anal. Biochem.* **1976**, *75*, 281–289.
- 2 L. D. Graham, K. D. Haggett, P. A. Jennings, D. S. Le Brocque, R. G. Whittaker, *Biochemistry* **1993**, *32*, 6250–6258.
- 3 T. L. Arnaldos, M. L. Serrano, A. A. Calderon, R. Munoz, *Phytochem. Anal.* **1999**, *10*, 171–174.
- 4 J. Hofmann, M. Seretz, *Anal. Chim. Acta* **1984**, *163*, 67–72.
- 5 G. G. Guilbault, A. N. J. Heyn, *Anal. Lett.* **1967**, *1*, 163–171.
- 6 T. Vorderwülbecke, K. Lieslich, H. Erdmann, *Enzyme Microb. Technol.* **1992**, *14*, 631–639.
- 7 D. N. Kramer, G. G. Guilbault, *Anal. Chem.* **1964**, *36*, 1662–1663.
- 8 M. Baumann, R. Sturmer, U. T. Bornscheuer, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 4201–4204.
- 9 M. Konarzycka-Bessler, U. T. Bornscheuer *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 1418–1420.
- 10 M. J. Wajzer, *C. R. Hebd. Seances Acad. Sci.* **1949**, *229*, 1270–1272.
- 11 R. A. John in *Enzyme Assays*, eds R. E. Szent-Györgyi, M. J. Danson, IRL, Oxford, **1992**, pp. 81–82.
- 12 R. M. Rosenberg, R. M. Herreid, G. J. Piazza, M. H. O'Leary, *Anal. Biochem.* **1989**, *181*, 59–65.
- 13 B. H. Gibbons, J. T. Edsall, *J. Biol. Chem.* **1963**, *238*, 3502–3507.
- 14 O. H. Lowry, N. R. Roberts, M.-L. Wu, W. S. Hixon, E. J. Crawford, *J. Biol. Chem.* **1954**, *207*, 19–37.
- 15 R. A. Darrow, S. P. Colowick, *Meth. Enzymol.* **1962**, *Vol. V*, 226–235.
- 16 R. K. Crane, A. Sols, *Meth. Enzymol.* **1960**, *Vol. I*, 277–286.
- 17 R. G. Whittaker, M. K. Manthey, D. S. Le Brocque, P. J. Hayes, *Anal. Biochem.* **1994**, *220*, 238–243.
- 18 L. E. Janes, A. C. Löwendahl, R. J. Kazlauskas, *Chem. Eur. J.* **1998**, *4*, 2324–2331.
- 19 R. G. Khalifah, *J. Biol. Chem.* **1971**, *246*, 2561–2573. An appendix includes a derivation of the pH dependence of the buffer factor, *Q*.
- 20 E. Banyai, in *Indicators*, ed. E. Bishop, Pergamon Press, Oxford, **1972**, p. 75.
- 21 R. J. Beynon, J. S. Easterby, *Buffer Solutions, The Basics*, IRL Press, Oxford, **1996**, p. 72.
- 22 P. F. Mugford, S. M. Lait, B. A. Keay, R. J. Kazlauskas, *ChemBioChem* **2004**, *5*, 980–987.
- 23 C. K. Savile, V. P. Magloire, R. J. Kazlauskas, *J. Am. Chem. Soc.* **2005**, *127*, 2104–2113.
- 24 C. Deng, R. R. Chen, *Anal. Biochem.* **2004**, *330*, 219–226.
- 25 a) P. Holloway, J. T. Trevors, H. Lee, *J. Microbiol. Methods* **1998**, *32*, 31–36;
b) H. Zhao, *Methods Mol. Biol.* **2003**, *230*, 213–221.
- 26 A. Banerjee, P. Kaul, R. Sharma, U. C. Banerjee, *J. Biomol. Screening* **2003**, *8*, 559–565.
- 27 I. N. Taylor, R. C. Brown, M. Bycroft, G. King, J. A. Littlechild, M. C. Lloyd, C. Praquin, H. S. Toogood, S. J. C. Taylor, *Biochem. Soc. Trans.* **2004**, *32*, 290–292.
- 28 D. C. Demirjian, P. C. Shah, F. Moris-Varas, *Top. Curr. Chem.* **1999**, *200*, 1–29.
- 29 A. M. F. Liu, N. A. Somers, R. J. Kazlauskas, T. S. Brush, F. Zocher, M. M. Enzelberger, U. T. Bornscheuer, G. P. Horsman, A. Mezzetti, C. Schmidt-Dannert, R. D. Schmid, *Tetrahedron Asymmetry* **2001**, *12*, 545–556.
- 30 N. A. Somers, R. J. Kazlauskas, *Tetrahedron Asymmetry* **2004**, *15*, 2991–3004.
- 31 A. Fersht, *Enzyme Structure and Mechanism*, 2nd edn, Freeman, New York, **1985**, pp. 103–106.

- 32 C. S. Chen, Y. Fujimoto, G. Girdaukas, C. J. Sih, *J. Am. Chem. Soc.* **1982**, *104*, 7294–7299.
- 33 Other researchers have also estimated enantioselectivity by separately measuring the rates of hydrolysis of the pure enantiomers. For example: a) G. Zandonella, L. Haalck, F. Spener, K. Faber, F. Paltauf, A. Hermetter, *Chirality* **1996**, *8*, 481–489; b) M. T. Reetz, A. Zonta, K. Schimossek, K. Liebeton, K.-E. Jaeger, *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 2830–2832.
- 34 L. E. Janes, R. J. Kazlauskas, *J. Org. Chem.* **1997**, *62*, 4560–4561.
- 35 G. P. Horsman, A. M. F. Liu, E. Henke, U. T. Bornscheuer, R. J. Kazlauskas, *Chem. Eur. J.* **2003**, *9*, 1933–1939.
- 36 S. Park, K. Morley, G. P. Horsman, M. Holmquist, K. Hult, R. J. Kazlauskas, *Chem. Biol.* **2005**, *12*, 45–54.
- 37 L. E. Janes, A. Cimpola, R. J. Kazlauskas, *J. Org. Chem.* **1999**, *64*, 9019–9029.
- 38 D. Wahler, F. Badalassi, P. Crotti, J.-L. Reymond, *Chem. Eur. J.* **2002**, *8*, 3211–3228.

2

High-throughput Screening Systems for Assaying the Enantioselectivity of Enzymes

Manfred T. Reetz

2.1

Introduction

Academic and industrial interest in the catalytic asymmetric preparation of enantiomerically pure or enriched organic compounds continues to grow rapidly, biocatalytic routes drawing increasing attention [1]. Indeed, already now a significant number of industrial enantioselective processes are based on the use of enzymes. In addition to progress in biotechnological engineering (e.g. reactor design) [1e], two important developments in the 1990s resulted in even greater interest in the use of enzymes in asymmetric catalysis, namely directed evolution of enantioselective enzymes [2] and metagenome DNA panning [3] (Figure 2.1). The evolutionary approach is based on the proper combination of molecular biological methods for random gene mutagenesis and gene expression [4] coupled with appropriate high-throughput screening systems to assess the enantiopurity of the chiral products [5]. Typically, thousands of samples arising from the catalytic action of the evolved enzyme variants on a given substrate of interest need to be assayed within a day or two. A similar analytical problem arises in the case of metagenome DNA panning in which large numbers of genes are collected in the environment followed by expression of the encoded potentially enantioselective enzymes in recombinant microorganisms.

The enantioselectivity of an enzyme catalyzing a given transformation is traditionally determined by the enantiomeric excess of the product (% *ee*) or, in the case of kinetic resolution of a racemate, by the selectivity factor *E*, which reflects the relative reaction rates of the two enantiomers [6]. Normally gas chromatography (GC) or high-performance liquid chromatography (HPLC) based on chiral columns is employed, but the conventional forms of these analytical tools can handle only a few dozen samples per day. For this reason medium- or preferably high-throughput *ee*-assays had to be developed, as in the new field of combinatorial asymmetric transition metal catalysis [5d, 7].

This chapter focuses on the most efficient and practical high-throughput *ee*-screening systems developed specifically to evaluate enantioselective enzymes.

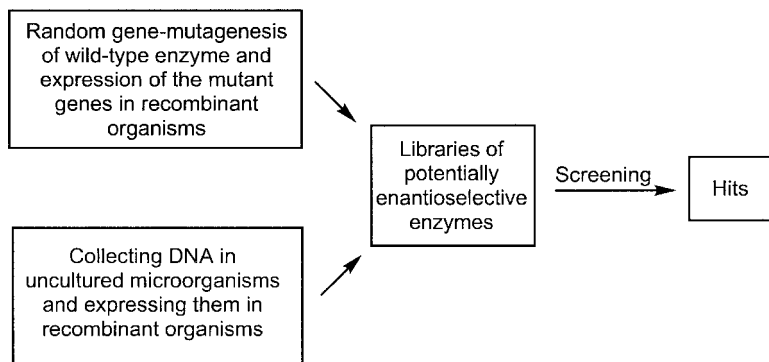


Fig. 2.1 Two sources of large libraries of potentially enantioselective enzymes.

A few other rapid *ee*-assays originally developed for combinatorial asymmetric transition metal catalysis, but adaptable to biocatalysis, are included as well. Due to space limitation only a few of the currently available *ee*-assays are illustrated by detailed protocols. Many of the *ee*-screening systems are complementary, and no single assay is truly universal. For general information concerning high-throughput *ee*-screening systems, the reader is referred to recent reviews [5]. Chapter 1 of this volume, focusing on chromogenic *ee*-assays of hydrolases, should also be consulted. Selection, as opposed to screening, has not been developed to date in the directed evolution of enantioselective enzymes [8]. However, a screening system based on differential cell growth has been described [9], which is a rare case of a colony-based *ee*-assay. All of the assays described in this chapter are carried out in the spatially addressable wells of microtiter plates following bacterial colony picking. If possible, it is advisable to apply an efficient prescreen for activity, which eliminates inactive mutants and thus makes it easier to perform the *ee*-assay.

2.2

UV/Vis Spectroscopy-based Assays

UV/Vis-based assays have a number of advantages, including the possibility of visual prescreening for activity on microtiter plates. More importantly, if a reliable UV/Vis signal arises as a consequence of an enzymatic reaction, commercially available UV/Vis plate readers can be used in screening thousands of mutant enzymes catalyzing reactions in the wells of microtiter plates [5].

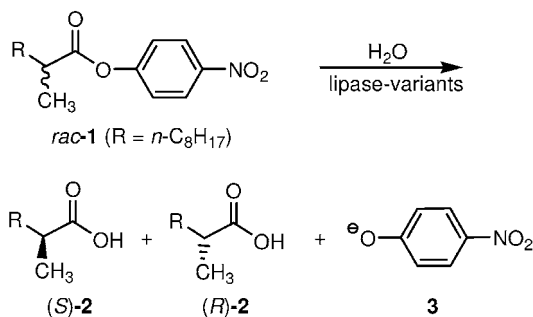
2.2.1

Assay for Screening Lipases or Esterases in the Kinetic Resolution of Chiral *p*-Nitrophenyl Esters

The first, though somewhat crude, high-throughput *ee*-assay used in the directed evolution of enantioselective enzymes was based on UV/Vis spectroscopy [2a]. This was restricted to the hydrolytic kinetic resolution of racemic *p*-nitrophenyl esters catalyzed by lipases or esterases. The specific goal was to evolve highly enantioselective mutants of the lipase from *Pseudomonas aeruginosa* as potential biocatalysts in the hydrolytic kinetic resolution of the chiral ester (*S*)/(*R*)-1. The wild type leads to an *E*-value of only 1.1 in slight favor of the (*S*)-acid (2). Lipase-catalyzed hydrolysis in buffered medium generates *p*-nitrophenolate (3), which shows a strong UV/Vis absorption at 405 nm (Scheme 2.1). Thus, reactions can be carried out on microtiter plates, a simple UV/Vis plate reader measuring absorption as a function of time (typically during the first 8 min). However, since the racemate delivers only information concerning the overall rate, the (*S*)- and (*R*)-substrates 1 were prepared and studied separately pairwise on 96-well microtiter plates. If the slopes of the absorption/time curves differ considerably, a hit is indicated (i.e. an enantioselective lipase variant has been identified which is then studied in detail in a lab-scale reaction using traditional chiral GC).

Figure 2.2 shows two experimental plots, illustrating the presence of a nonselective lipase (Figure 2.2a) and a hit (Figure 2.2b) [2a, 5h]. About 500–800 plots of this kind are possible per day. Using error-prone polymerase chain reaction (epPCR) [2a], saturation mutagenesis [2b], and DNA shuffling [2c], a total of 40000 lipase-variants were generated and screened in the model reaction [2]. Several enantioselective lipases variants were obtained, the best one showing an *E*-value of >51 [2c]. Moreover, reversal of enantioselectivity was achieved (*E* = 30 in favor of (*R*)-2), a process which again required only about 40000 mutants [2d].

The disadvantage of this assay has to do with the fact that a built-in chromophore is required (*p*-nitrophenol), yet *p*-nitrophenyl esters will never be used in real (industrial) applications. If directed evolution is being applied, the final out-



Scheme 2.1

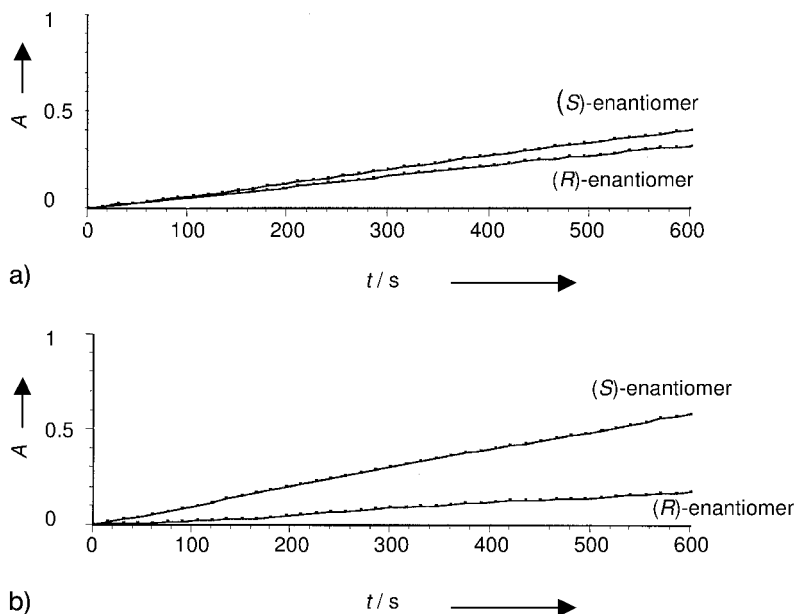


Fig. 2.2 Time course of the lipase-catalyzed hydrolysis of the (*R*)- and (*S*)-ester **1** measured with a UV/Vis plate reader [5 a, 5 h]. (a) Wild-type lipase from *P. aeruginosa*, (b) improved mutant in the first generation.

come may be influenced by the fact that the *p*-nitrophenyl ester and not the usual methyl or ethyl ester was chosen. Moreover, since the (*S*)- and (*R*)-substrates are tested separately pairwise, the enzyme does not compete for the two substrates, introducing some uncertainty. Nevertheless, the assay is useful in proof-of-principle studies.

Procedure 2.1: UV/Vis-based Screening of Lipases or Esterases in Kinetic Resolutions [2 a, 5 h]

Each colony of *P. aeruginosa* PABST7.1/pUCPL6A was picked with a colony picker (Genetix, UK) and resuspended in a well of a 96-deep-well microtiter plate filled with 1 mL of 2×LB (20 g L⁻¹ trypton, 10 g L⁻¹ yeast extract, 10 g L⁻¹ NaCl) containing 200 μg mL⁻¹ carbenicillin and 50 μg mL⁻¹ tetracycline [2 a, 5 h]. Cells were grown by shaking the plates overnight at 30 °C. Lipase expression was induced by addition of 0.4 mM IPTG and further shaking at 30 °C for 4–24 h. Cells were separated from the culture supernatant by centrifugation of the plates at 5000×*g* for 15 min. An aliquot of 50 μL was taken from each well and pipetted into two wells of a second microtiter plate each containing 10 μL of a 100 mM Tris-HCl pH 8.0 buffer, 30 μL A. dest. and 10 μL of the (*R*)- or (*S*)-enantiomer, respectively, of 2-methyldeca-

noic acid *p*-nitrophenyl ester (**1**) in 10% DMF. Plates were shaken for 10 s and the reactions monitored at 410 nm for 6–10 min at 30 °C using a Spectramax-8 channel photometer (Molecular Devices, Sunnyvale, CA, USA). Apparent *E*-values were determined from the ratio of the slopes of the linear part of the curves obtained for the (*R*)- and (*S*)-enantiomers.

Those clones showing an improved enantioselectivity as compared to the “parent” of the corresponding generation were further analyzed by reacting a racemic substrate mixture with culture supernatant and subsequent analysis of the reaction products by conventional chiral GC. The genes encoding lipases with improved enantioselectivity were sequenced (Eurogentec, Belgium). For this, the *Bam*HI/*Apa*I-fragments from pUCPL6A were cloned into pBluescriptII KS giving pMut5 which also served as the template for the next round of epPCR.

2.2.2

Enzyme-coupled UV/Vis-based Assay for Lipases and Esterases

In those cases in which the product of an enzymatic reaction under study can be transformed by another enzyme into a secondary product which gives rise to a spectroscopic signal, an enzyme-coupled assay is possible. This was first demonstrated using fluorescence as the spectroscopic detection method (see Section 2.3) [10a]. In a UV/Vis approach using this idea the hydrolase-catalyzed kinetic resolution of chiral acetates was studied using a high-throughput *ee*-assay based on an enzyme-coupled test, the presence of a fluorogenic moiety or a UV/Vis

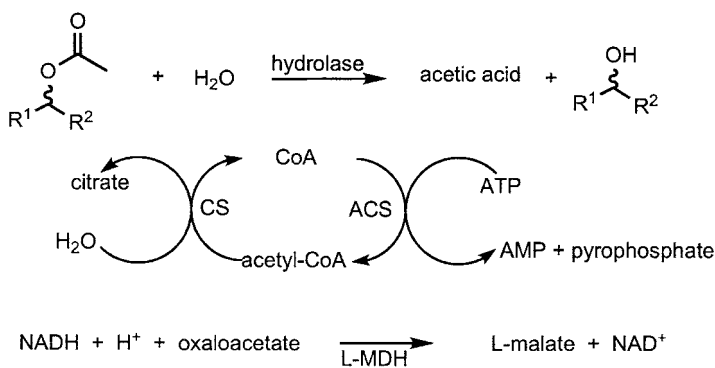


Fig. 2.3 Enzyme-coupled assay in which the hydrolase-catalyzed reaction releases acetic acid [11]. The latter is converted by acetyl-CoA synthetase (ACS) to acetyl-CoA in the presence of ATP and coenzyme A (CoA). Citrate synthase (CS) catalyzes the reaction between acetyl-CoA and oxaloacetate to give citrate. The oxaloacetate required for this reaction is formed from L-malate and

NAD^+ in the presence of L-malate dehydrogenase (L-MDH). Initial rates of acetic acid formation can thus be determined by the increase in adsorption at 340 nm due to the increase in NADH concentration. Use of optically pure (*R*)- or (*S*)-acetates allows the determination of the apparent enantioselectivity E_{app} .

chromophore not being necessary [11]. The assay is based on the idea that the acetic acid formed in the hydrolysis of a chiral acetate can be transformed stoichiometrically into NADH via a series of coupled enzymatic reactions using commercially available enzyme kits (Figure 2.3). The NADH is then easily detected UV-spectroscopically in the wells of microtiter plates using a standard plate reader. It has been estimated that about 13000 samples can be evaluated per day, although in practise this has not been demonstrated [11]. Using various commercially available hydrolases the kinetic resolution of (*S,R*)-1-methoxy-2-propylacetate was studied. The agreement between the apparent selectivity factor E_{app} and the actual value E_{true} determined by GC turned out to be excellent at low enantioselectivity ($E=1.4-13$), but less so at higher enantioselectivity (20% variation at $E=80$) [11]. This may cause problems when attempting to enhance enantioselectivity beyond $E=50$ in directed evolution studies.

Procedure 2.2: Enzyme-coupled UV/Vis-based Assay for Lipases and Esterases [11]

Spectrophotometric determination of NADH concentration is performed at 340 nm in the milliliter scale on an Ultrospec 3000 Photometer (Pharmacia Biotech Ltd., Uppsala, Sweden) and on the microliter scale using a fluorimeter (e.g. FLUOstar, BMG LabTechnologies, Offenburg, Germany). In order to determine the amount of acetic acid formed, a solution of *Pseudomonas fluorescens* esterase (PFE) (20 μ L, 2 mg mL⁻¹, unless stated otherwise) or the lipase B from *Candida antarctica* (CAL-B) (2 mg mL⁻¹) is added to a mixture of the test-kit components (150 μ L). The reactions are started by adding a solution of a chiral acetate (20 μ L) in sodium phosphate buffer (10 mM, pH 7.3). Mixtures of the test-kit with buffer or cell lysates of non-induced *E. coli* harboring the gene encoding recombinant PFE serve as controls. In a similar manner, reaction rates are determined by using optically pure (*R*)- and (*S*)-acetates separately. For reactions with crude cell extract, PFE is produced in microtiter plates similar to the published protocol for shake-flask cultivation. However, the cultivation volume is 200 μ L per well and cells are disrupted by two freeze-thaw cycles. Finally, cell debris is removed by centrifugation and the supernatants are used for the assay [11].

2.2.3

Enzymatic Method for Determining Enantiomeric Excess (EMDee)

A different UV/Vis-based approach is based on the idea that an appropriate enzyme selectively processes one enantiomer, giving rise to a UV/Vis signal [12]. A case in point concerns the determination of the enantiopurity of a library of chiral secondary alcohols such as **6** (in this particular study produced by asymmetric Et₂Zn-addition to benzaldehyde **4**), the (*S*)-enantiomer being oxidized selectively by the alcohol dehydrogenase from *Thermoanaerobium* sp. (Figure 2.4). The rate of this process can be monitored by a UV/Vis plate reader due to the

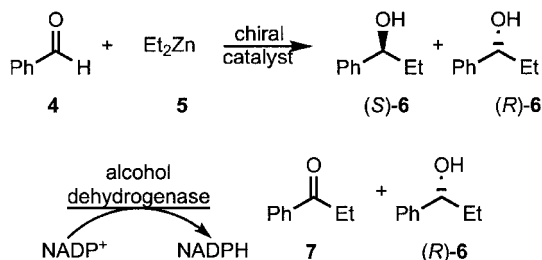


Fig. 2.4 Enzymatic method for determining enantiomeric excess (EMDee) in the case of 1-phenylpropanol produced by asymmetric addition of diethylzinc to benzaldehyde [12].

formation of NADPH (absorption at 340 nm), which relates to the quantity of the (*S*)-enantiomer present in the mixture. Although the screen was not specifically developed to evaluate chiral alcohols produced by an enzymatic process (e.g. reduction of ketones), this should be possible, provided an appropriate extraction process can be developed.

The success of the assay depends upon the finding that the rate of oxidation constitutes a direct measure of the *ee*-value. High throughput was demonstrated by analyzing 100 samples in a 384-well format by using a UV plate reader. Each sample contained 1 μmol of 1-phenylpropanol (**6**) in a volume of 100 μL . The accuracy of the *ee*-value amounts to $\pm 10\%$, as checked by independent GC determinations [12]. About 100 samples can be processed in 30 min, which works out at 4800 *ee* determinations per day. Of course, for each new alcohol to be assayed, the alcohol dehydrogenase needs to be selective. If this is not the case, a different (selective) alcohol dehydrogenase has to be found in order for the assay to function properly. Moreover, an accuracy of $\pm 10\%$ may not be sufficient in the late stages of an evolutionary project when attempting to increase the *ee*-value beyond 90%. Of course, this problem applies to any high-throughput *ee*-assay having such accuracy.

2.2.4

UV/Vis-based Enzyme Immunoassay as a Means to Measure Enantiomeric Excess

UV/Vis-based high-throughput screening of enantioselective catalysts is possible by enzyme immunoassays [13], a technology that is routinely applied in biology and medicine. As in the case of some of the other screening systems, this new assay was not developed specifically for enzyme-catalyzed processes. In fact, it was illustrated by analyzing (*R*)/(*S*)-mixtures of mandelic acid (**9**) generated by enantioselective Ru-catalyzed hydrogenation of benzoyl formic acid (**8**) (Figure 2.5). By employing an antibody that binds both enantiomers, it is possible to measure the concentration of the reaction product, thereby allowing the yield to be calculated. The use of an (*S*)-specific antibody then makes the determination of the *ee* possible (Figure 2.5).

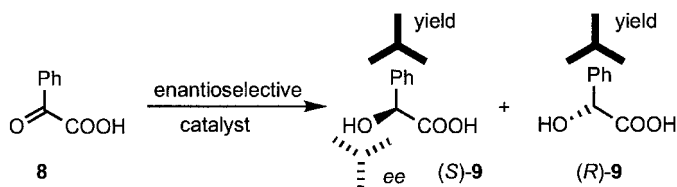


Fig. 2.5 High-throughput screening of enantioselective catalysts by competitive enzyme immunoassays [13]. The solid antibody recognizes both enantiomers, and the hatched antibody is (*S*)-specific, enabling the determination of yield and *ee*.

The success of this assay depends upon the availability of specific antibodies, and indeed these can be raised to almost any compound of interest. Moreover, simple automated equipment comprising a plate washer and plate absorbance reader is all that is necessary. About 1000 *ee* determinations are possible per day, precision amounting to $\pm 9\%$ [13].

2.2.5

Other UV/Vis-based *ee*-Assays

Several other high-throughput *ee*-assays based on the use of UV/Vis spectroscopy have been described [5]. Most prominent are colorimetric tests used in the evaluation of enantioselective lipases and esterases [14]. Since the hydrolysis of an ester leads to a change in pH, quantification is possible by the use of an appropriate pH indicator. These UV/Vis-based assays are described in detail in Chapter 1.

2.3

Assays Using Fluorescence

Assays based on fluorescence, either in the measurement of enzyme activity or in the quantification of enantioselectivity, all have a very useful property, namely a high degree of sensitivity [5]. This allows the use of very dilute substrate concentrations and extremely small amounts of catalysts. Basically, there are two different approaches. One involves the use of a substrate of interest to which a fluorescent-active (or potentially active) moiety is covalently attached. The second approach makes use of a fluorescence-based sensor which gives rise to a signal as a consequence of the enzyme-catalyzed reaction of a substrate of interest.

2.3.1

Umbelliferone-based Systems

A fluorescence-based *ee*-assay using umbelliferone (**13**) as the built-in fluorophore has been developed which works for several different types of enzymatic reactions [5 c, 10]. In an initial study the system was used to monitor the hydro-

lytic kinetic resolution of chiral acetates [10a], for example **10** (Figure 2.6). It is based on a sequence of two coupled enzymatic steps that converts a pair of enantiomeric alcohols formed by the asymmetric hydrolysis under study (e.g. (*R*)- and (*S*)-**11**) into a fluorescent product (e.g. **13**). In the first step (*R*)- and (*S*)-**10** are subjected separately to hydrolysis in reactions catalyzed by a mutant enzyme (lipase or esterase), a catalytic antibody, or in principle a synthetic catalyst compatible with the system. The goal of the assay is to measure the enantioselectivity of this kinetic resolution [10a]. The relative amount of (*R*)- and (*S*)-**11** produced after a given reaction time is a measure of the enantioselectivity and can be ascertained rapidly, but not directly. Two subsequent chemical transformations are necessary. First the enantiomeric alcohols (*R*)- and (*S*)-**11** are oxidized separately by horse liver alcohol dehydrogenase (HLDH) to the ketone **12**, from which the fluorescent final product umbelliferone (**13**) is released in each case by the catalytic action of bovine serum albumin (BSA). Thus, by measuring the fluorescence of **13** for the (*R*)- and the (*S*)-substrate separately, the relative amounts of (*R*)- and (*S*)-**11** can be determined.

Thirty different esterases and lipases were tested. The rate of release of **13** in the wells of standard microtiter plates was monitored by fluorescence. Control experiments ensured that the apparent rate of umbelliferone (**13**) release is directly proportional to the rate of acetate hydrolysis. The predicted and observed *E* and *ee*-values (as checked by standard HPLC assay on a chiral phase) were found to lie within $\pm 20\%$ [10a]. Only in one case was a larger discrepancy observed, a result that was believed to be caused by the occurrence of an unusual-

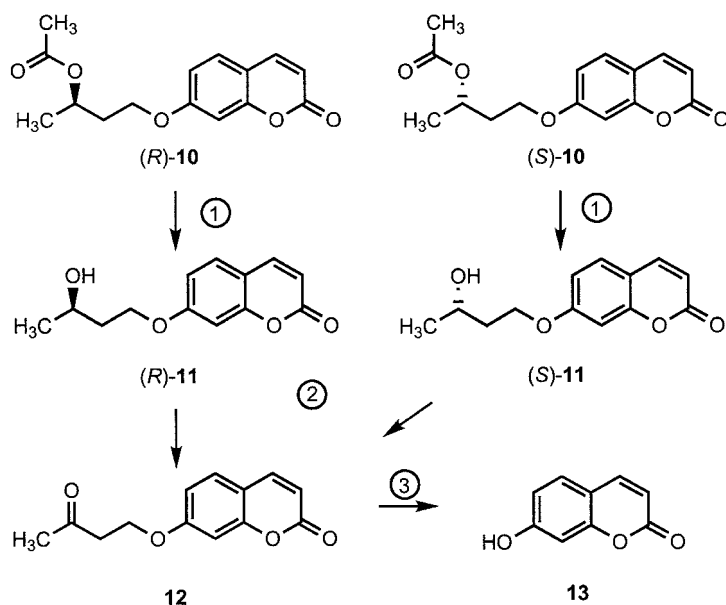


Fig. 2.6 Fluorescence-based assay for enantioselectivity of ester hydrolysis [10a].

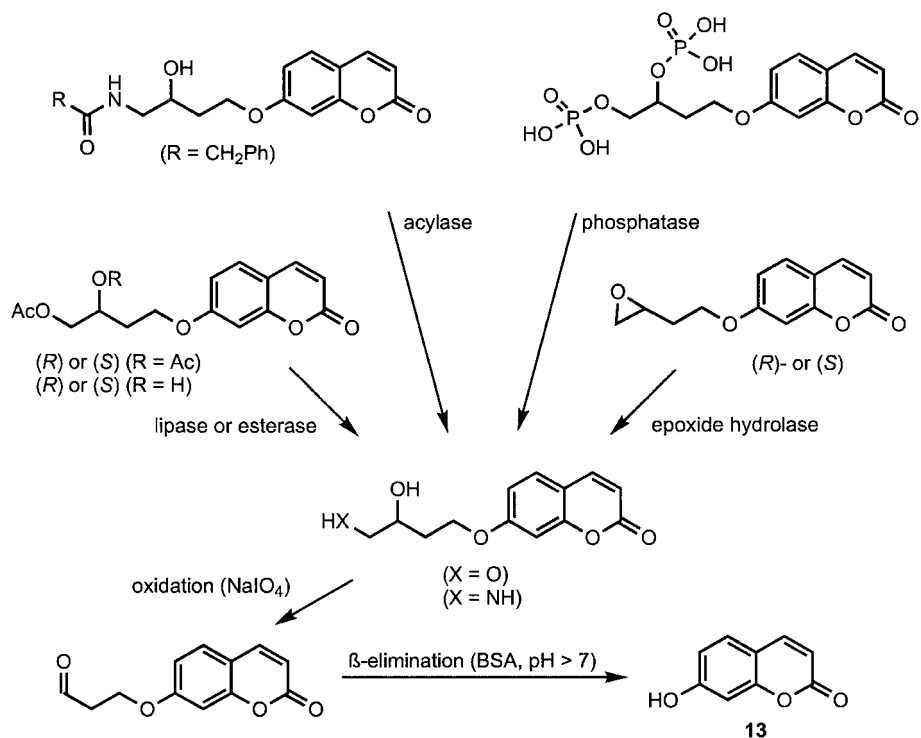


Fig. 2.7 Periodate-coupled fluorogenic assay for hydrolases based on the generation of umbelliferone **13** [10].

ly low K_m for one of the enantiomers. Thus, since the test can be carried out on 96-well microtiter plates, high throughput is possible.

Of course, the inherent disadvantage noted earlier for some of the colorimetric tests also applies here, namely the fact that the optimization of a potential catalyst is focused on a specific substrate **10** modified by incorporation of a probe, in this case the fluorogenic moiety **13**. However, one can expect the test to be useful in directed evolution projects in which proof-of-principle is the goal. Moreover, this kind of approach can be used in very practical applications, namely as a pretest for the activity of enzymes.

The basic idea has been extended beyond the *ee*-assay for lipases and esterases to include such enzymes as acylases, epoxide hydrolases, alcohol dehydrogenases, aldolases and phosphatases [5c, 10]. In many cases the assay makes use of NaIO₄-mediated oxidation of a diol-moiety followed by base-catalyzed β -elimination to generate the fluorescent umbelliferone (Figure 2.7). Again, the most likely area of application is prescreening, although adaption to include high-throughput *ee*-determination may be possible in some cases.

Procedure 2.3: Umbelliferone-based Assay [10 a]

Fluorescence measurements: Substrates were diluted from a 10 mM stock solution in 50% aq. MeCN. All buffers and solutions were prepared using MilliQ-deionized H₂O. Enzymes were diluted from 1 mg mL⁻¹ stock solutions in PBS (PBS=10 mM aq. phosphate, 160 mM NaCl, pH 7.4). BSA was diluted from a 40 mg mL⁻¹ stock solution in either 20 mM borate (pH 8.8) (assays with **10**). Reactions were initiated by addition of lipase or esterase to a solution containing substrate, HLDH, and NAD⁺. The 200- μ L assays were followed in individual wells of round-bottom polypropylene 96-well plates (Costar) with a Cytofluor-II fluorescence-plate reader (Perseptive Biosystems, filters λ_{ex} 360 \pm 20 nm, λ_{em} 460 \pm 20 nm). For each assay, fluorescence was related to the umbelliferone (**13**) concentration according to a calibration curve with pure **13** in the same buffers containing BSA. The reference maximum rate with alcohol substrates was taken from the linear portion of the curves (data points 0–5000 s). For assays with acetates, the linear portions of the curves (data points 15000–30000 s) were used to derive the reaction rates.

Preparative assays: Solid enzyme (1 mg) weighed into a 1.5-mL Eppendorf tube was dissolved in 20 mM aq. borate (pH 8.8; 200 μ L). Then, a solution of racemic acetate (**10**) (50 μ L; 50 mg mL⁻¹ in MeCN) was added and the resulting suspension agitated on a plate stirrer at room temperature. To follow the reaction, 2 μ L aliquots were taken and diluted with hexane/*i*-PrOH 1:1. Then, this solution (20 μ L) was injected on a Chiralpak AS column (Daicel, 0.45 \times 22 cm, flow 1.0 mL min⁻¹, hexane/*i*-PrOH 1:2): t_{R} 10.9 ((*R*)-**10**), 23.8 ((*S*)-**10**), 12.1 ((*R*)-**11**), 13.1 ((*S*)-**11**) min. The integral of the peaks recorded by UV at 325 nm was used to calculate the conversion and *ee* of released alcohol [10 a].

2.3.2

Fluorescence-based Assay Using DNA Microarrays

A very different fluorescence-based *ee*-assay makes use of DNA microarrays [15]. This type of technology had previously been employed to determine relative gene expression levels on a genome-wide basis as measured by the ratio of fluorescent reporters. In the case of the *ee*-assay, chiral amino acids were used as model compounds. Mixtures of a racemic amino acid are first subjected to acylation at the amino function with formation of *N*-Boc-protected derivatives. The samples are then covalently attached to amine-functionalized glass slides in a spatially arrayed manner (Figure 2.8). In a second step the uncoupled surface amino functions are acylated exhaustively. The third step involves complete deprotection to afford the free amino function of the amino acid. Finally, in a fourth step two *pseudo*-enantiomeric fluorescent probes are attached to the free amino groups on the surface of the array. An appreciable degree of kinetic reso-

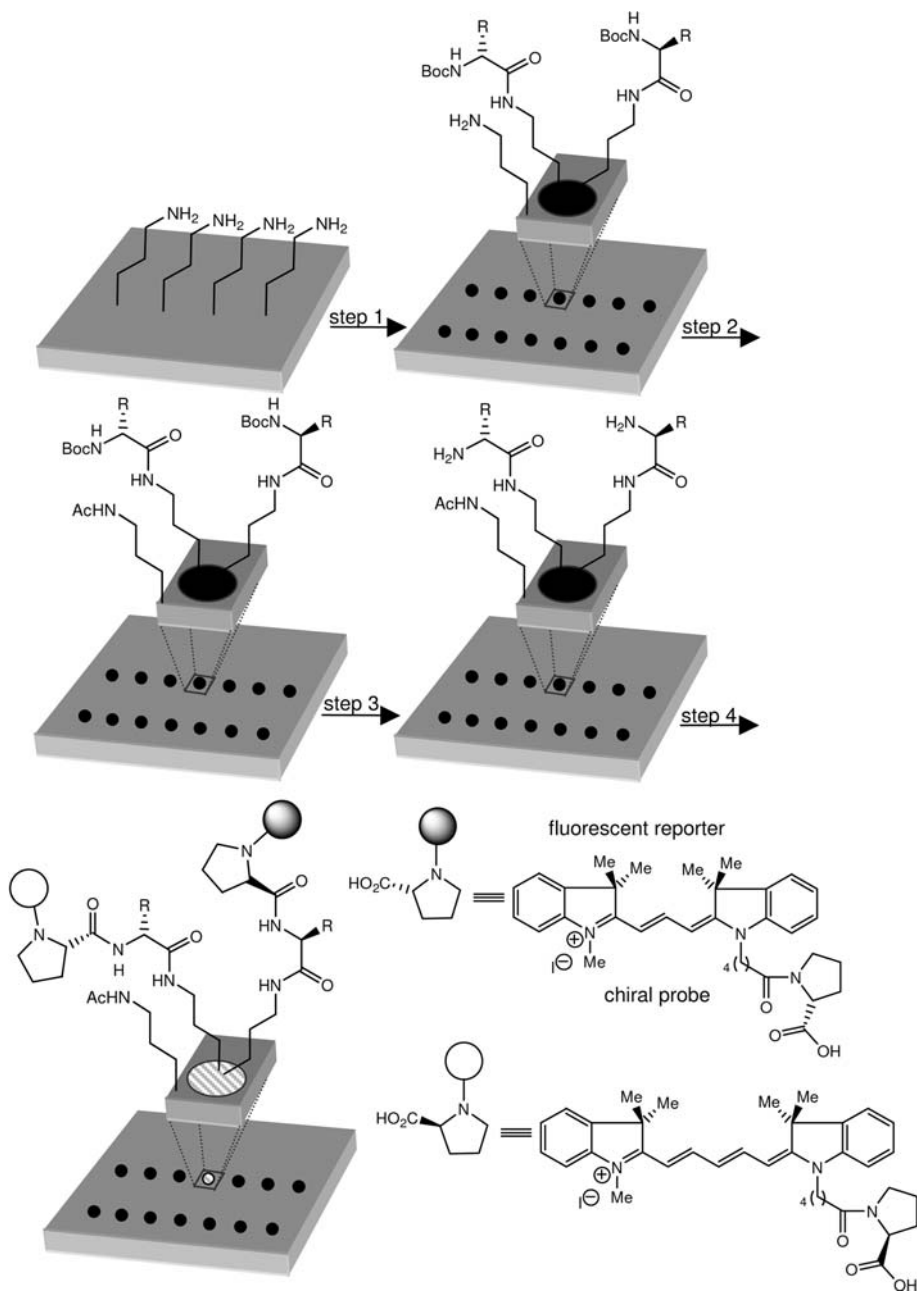


Fig. 2.8 Reaction microarrays in high-throughput *ee*-determination [15]. Reagents and conditions: step 1: BocNHCH(R)CO₂H, PyAOP, *i*Pr₂NEt, DMF; step 2: Ac₂O, pyridine; step 3: 10% CF₃CO₂H and 10% Et₃SiH in CH₂Cl₂, then 3% Et₃N in CH₂Cl₂; step 4: pentafluorophenyl diphenylphosphinate, *i*Pr₂NEt, 1:1 mixture of the two fluorescent proline derivatives, DMF, -20°C.

lution in the process of amide coupling is a requirement for the success of the *ee*-assay [15] (Horeau's principle [16]). In the present case the *ee*-values are accessible by measuring the ratio of the relevant fluorescent intensities. About 8000 *ee*-determinations are possible per day, precision amounting to $\pm 10\%$ of the actual value [15]. Although it was not explicitly demonstrated that this *ee*-assay can be used to evaluate enzymes (e.g. proteases), this should in fact be possible. So far this approach has not been extended to other types of substrates.

2.3.3

Other Fluorescence-based *ee*-Assays

Several other *ee*-assays relying on use of fluorescence have been described, although their general utility in the high-throughput evaluation of enantioselective enzymes remains to be demonstrated in most cases [5, 17]. A high-throughput *ee*-assay based on capillary electrophoresis (CE) [18], fluorescence being the detector system is described in Section 2.9.

2.4

Assays Based on Mass Spectrometry (MS)

Since enantiomers have identical mass spectra, the relative amounts of the (*R*)- and (*S*)-forms present in a given sample and therefore the *ee*-value cannot be measured by conventional MS techniques. However, special MS-based approaches have been described which do in fact allow *ee*-determination [5]. In one method [19] two conditions have to be met: (1) A mass-tagged chiral derivatization agent is applied to the mixture, and (2) in the process of derivatization a significant degree of kinetic resolution needs to occur (Horeau's principle [16]). The relative amounts of mass-tagged diastereomers can then be measured by MS simply by integrating the appropriate peaks, the uncertainty in the *ee*-value amounting to $\pm 10\%$ [19]. High-throughput application (e.g. in enzyme catalysis) has not been demonstrated, but this should be possible. Isotope labeling forms the basis of another MS-based *ee*-assay which has been used several times in the directed evolution of enantioselective enzymes [20] (Section 2.4.1).

2.4.1

MS-based Assay Using Isotope Labeling

In this MS-based approach derivatization reactions are not necessary, which makes the overall process easy [20]. It makes use of isotope-labeled *pseudo*-enantiomers or *pseudo-meso*-compounds. This practical method is restricted to studies involving kinetic resolution of racemates and desymmetrization of prochiral compounds bearing reactive enantiotopic groups (Figure 2.9) [20].

The products of these transformations are *pseudo*-enantiomers differing in absolute configuration and in mass, integration of the MS peaks and data process-

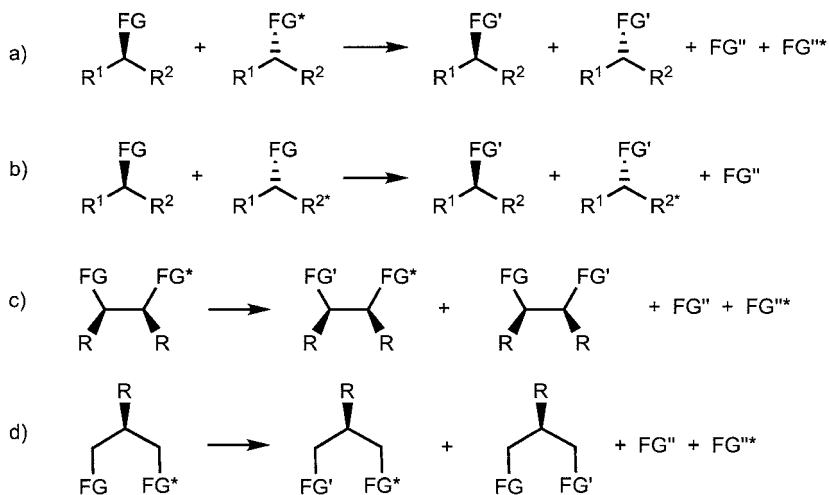
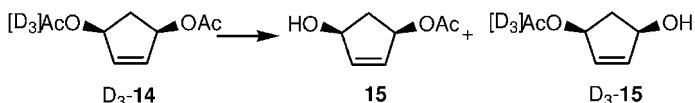


Fig. 2.9 (a) Asymmetric transformation of a mixture of *pseudo*-enantiomers involving cleavage of the functional groups (FG) and labeled FG*. (b) Asymmetric transformation of a mixture of *pseudo*-enantiomers involving either cleavage or bond formation at the FGs; isotopic labeling at R^2 is indicated by the asterisk. (c) Asymmetric transformation of a *pseudo*-meso substrate involving cleavage of the FGs and labeled FG*. (d) Asymmetric transformation of a *pseudo*-prochiral substrate involving cleavage of the FGs and labeled FG* [20a, b].

ing affording the *ee*- or *E*-values. Any type of ionization can be employed, but electrospray ionization (ESI) is used most commonly [20]. In some cases an internal standard is advisable if the determination of %conversion is necessary. The uncertainty in the *ee*-value amounts less than $\pm 5\%$. In the original version about 1000 *ee*-values could be measured per day [20a], but this has been increased to about 10000 samples per day as a result of a second-generation system based on an eight-channel multiplexed sprayer system [20b]. The method is illustrated here using lipase variants from *Bacillus subtilis*, produced by the methods of directed evolution and employed as catalysts in the desymmetrization of *meso*-1,4-diacetoxy-cyclopentene. The goal is to obtain enantioselective variants of this lipase which are expressed in *E. coli* [20d].

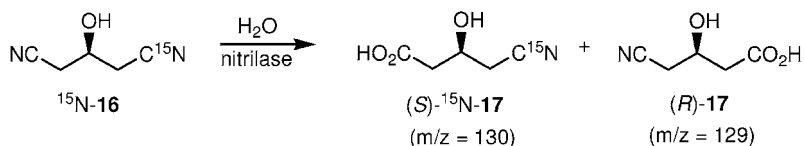
This particular substrate and the MS-based *ee*-assay have also been used in another study concerning the assembly of designed oligonucleotides (ADO) as a new recombinant method in directed evolution [20e]. Accordingly, the D_3 -labeled *pseudo*-meso-compound **14** is used as the substrate, the two products of



Scheme 2.2

the asymmetric transformation being non-labeled and the D_3 -labeled *pseudo*-enantiomers **15** and D_3 -**15** which are easily distinguished by ESI/MS because they differ by three mass units (Scheme 2.2).

The MS assay has also been applied successfully in the directed evolution of enantioselective epoxide hydrolases acting as catalysts in the kinetic resolution of chiral epoxides [20f]. Moreover, Diversa has applied the MS-based method in the desymmetrization of a prochiral dinitrile (1,3-dicyano-2-hydroxypropane) catalyzed by mutant nitrilases [21]. In this industrial application one of the nitrile moieties was labeled with ^{15}N as in ^{15}N -**16**, which means that the two *pseudo*-enantiomeric products (*S*- ^{15}N -**17** and (*R*)-**17** differ by one mass unit (Scheme 2.3). This is sufficient for the MS system to distinguish between the two products.



Scheme 2.3

As a note of caution, in the case of kinetic resolution the MS measurements must be performed in the appropriate time window (near 50% conversion). If this is difficult to achieve due to possibly different amounts or activities of the mutants from well to well, the system needs to be adapted in terms of time resolution. This means that samples for MS evaluation need to be taken as a function of time. Finally, it is useful to point out the possibility of multi-substrate *ee*-screening using the MS-based assay, which allows for enzyme fingerprinting with respect to the enantioselectivity of several substrates simultaneously.

Procedure 2.4: MS-based Assay Using Isotope Labeling [20 a, b]

In this protocol the desymmetrization of *meso*-1,4-diacetoxy-cyclopentene using labeled D_3 -**14** is described. Bacterial colonies on agar plates are collected and placed individually in the deep wells of microtiter plates (96-format) containing LB medium with the aid of an appropriate robot (e.g., Colony Picker Q-Pix commercially available from Genetix). Up to 10000 colonies can be handled per day. The LB medium is added using an eight-channel dispenser (e.g. Dispenser Multidrop DW from Labsystems).

The preparation of the reaction solutions and the process of carrying out the lipase-catalyzed reactions in the wells of 96-format microtiter plates are also fully automated. The preparation of the reaction solutions in the deep wells (2.2 mL) of microtiter plates (96-format) is performed using a pipette robot (e.g. Genesis supplied by Tecan GmbH, Maennedorf, Switzerland). Accordingly, pipette scripts are programmed (Software Gemini supplied by

Tecan) for robotically filling the wells with buffer and substrate solutions. This will vary, depending upon the particular substrate and enzyme. In the present case, which is typical for lipase-catalyzed desymmetrization, the reaction solutions are composed of 125 μL phosphate buffer (10 mM; pH 7.5), 50 μL supernatant and 25 μL substrate $\text{D}_3\text{-14}$ solution (0.1 M in dimethylsulfoxide). The reactions are allowed to take place in the wells of 96-format microtiter plates at room temperature in an incubator while shaking for 24 h. In order to activate all of the modules of the robot, the software Facts (supplied by Tecan) is used. The pipette robot consists of a workstation with spaces for 12 microtiter plates, a robot arm for transport and a carousel for storing the reaction plates as well as a 96-fold pipette module (Figure 2.10) [20 a, b].



Fig. 2.10 Pipette robot *Genesis* (Tecan GmbH, Maennedorf, Switzerland) with integrated carousel (right) [5 h].

Following lipase-catalyzed desymmetrization reactions of the substrate (e.g. $\text{D}_3\text{-14}$) in the wells of 96-format microtiter plates, an extraction step is necessary prior to ESI/MS analysis. In the present case ethyl acetate was used in the extraction step (200 μL reaction solution and 200 μL solvent). The samples need to be diluted prior to ESI/MS measurements using standard solution (methanol/10 mM NaOAc, 4:1 plus undeuterated *meso*-1,4-diacetoxy-cyclopentene as an internal standard necessary for the determination of conversion). This process is controlled by the software Facts. Figure 2.11 shows the required extraction process schematically. Four modules are controlled simultaneously: The robot arm (RoMa), the carousel for storing the microtiter plates, the 96-pipette system (TeMo) as well as the eightfold pipette head (Gemini). Iteration occurs within 12 min.

The control of the HPLC pump as well as the autosampler and the MS is ensured by the Software Masslynx 3.5 (Micromass). Following optimization of the measurement conditions, a list of process measurements is set up

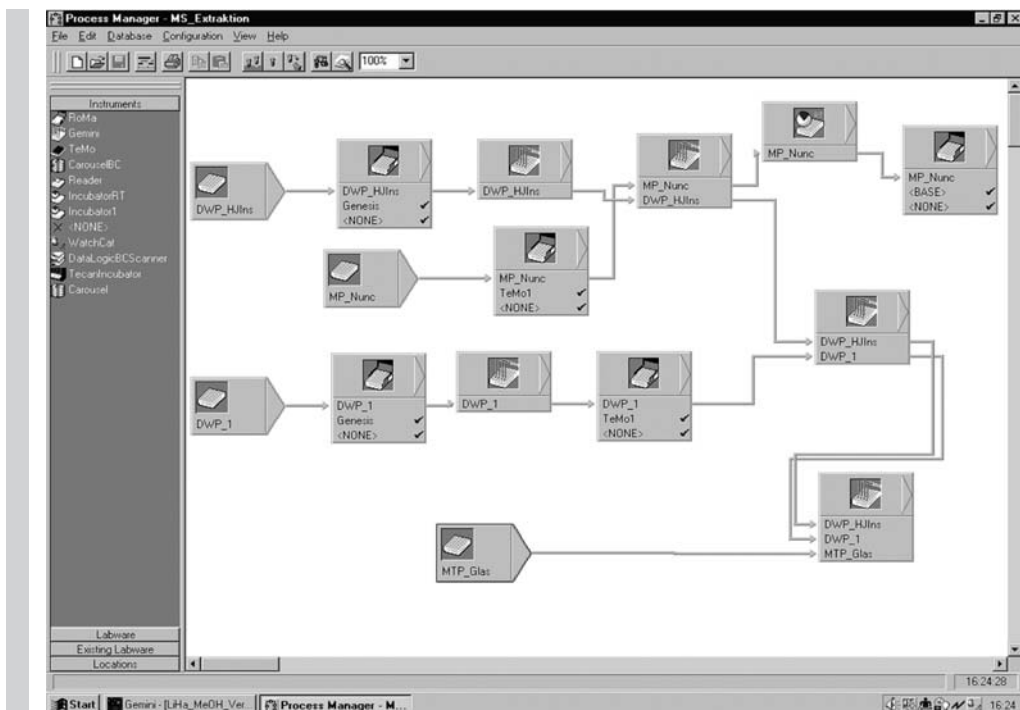


Fig. 2.11 Control of the individual modules by the software Facts (Tecan GmbH). In order to optimize the sequence of events, the processes are “interlocked” timewise [5 h].

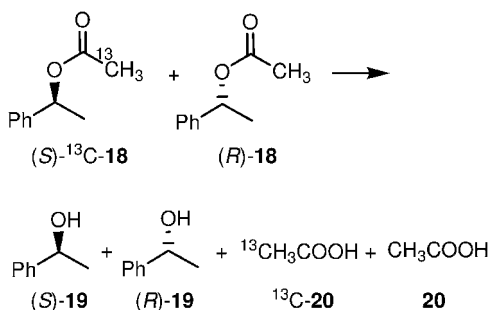
(sample list) and the desired HPLC and MS steps are called upon. Following a measurement the ESI source is automatically brought to room temperature (shut down). Using a 96-microtiter plate 576 samples can be processed per measurement. The chromatograms are integrated by the software package Quanlynx and Openlynx (Micromass) and transformed into an Excel[®] (Microsoft, Unterschleissheim, Germany) table. The use of a macro allows the calculation of the absolute intensities and therefore the $\epsilon\epsilon$ as well as the conversion. The E -values in the case of kinetic resolution are automatically calculated by the formula derived by Sih [6]. Data processing is possible by the use of the Openlynx Browser (Micromass). The overall process occurs continuously and makes possible the analysis of up to 10000 samples per day, provided the eight-channel multiplexed sprayer system (e.g. from Micromass) is used. It is also possible to use 384-format microtiter plates [20a, b].

2.5

Assays Based on Nuclear Magnetic Resonance Spectroscopy

For decades nuclear magnetic resonance (NMR) measurements were considered to be slow processes, but recent advances in the design of flow-through NMR cells have allowed the method to be applied in combinatorial chemistry [22]. These technological improvements were then applied to the development of two different NMR-based high-throughput *ee*-assays for evaluating the products of enzyme- or transition metal-catalyzed reactions [23]. In one version classical derivatization using a chiral reagent or NMR shift agent is parallelized and automated using flow-through cells, about 1400 *ee*-measurements being possible per day with a precision of $\pm 5\%$. In the second embodiment, illustrated here in detail, a principle related to that of the MS system described in Section 2.4 is applied. Chiral or *meso*-substrates are labeled so as to produce *pseudo*-enantiomers or *pseudo-meso*-compounds which are then used in the actual screen. Application is thus restricted to kinetic resolution of racemates and desymmetrization of prochiral compounds bearing reactive enantiotopic groups as illustrated previously for the MS system (Figure 2.9, Section 2.4.1).

A particularly practical form of this assay utilizes ^1H NMR spectroscopy, ^{13}C -labeling being used to distinguish between the (*R*)- and (*S*)-forms of a chiral compound under study. Essentially any carbon atom in the compound of interest can be labeled (Figure 2.9), but methyl groups in which the ^1H signals are not split by ^1H , ^1H coupling are preferred because the relevant peaks to be integrated are the singlet arising from the CH_3 -group of one enantiomer and the doublet of the ^{13}C -group of the other. A typical example which illustrates the method concerns the lipase- or esterase-catalyzed hydrolytic kinetic resolution of *rac*-1-phenylethyl acetate, derived from *rac*-1-phenyl ethanol (**19**) [23]. However, the acetate of any chiral alcohol or the acetamide of any chiral amine can be used. Labeling can be carried out in any position of a compound, as in (*S*)- ^{13}C -**18**. The synthesis is straightforward, since it simply involves acylation of the (*S*)-alcohol using commercially available ^{13}C -labeled acetyl chloride. Then a 1:1 mixture of labeled and nonlabeled compounds (*S*)- ^{13}C -**18** and (*R*)-**18** is prepared which simulates a racemate. It is used in the actual catalytic hydrolytic kinetic



Scheme 2.4

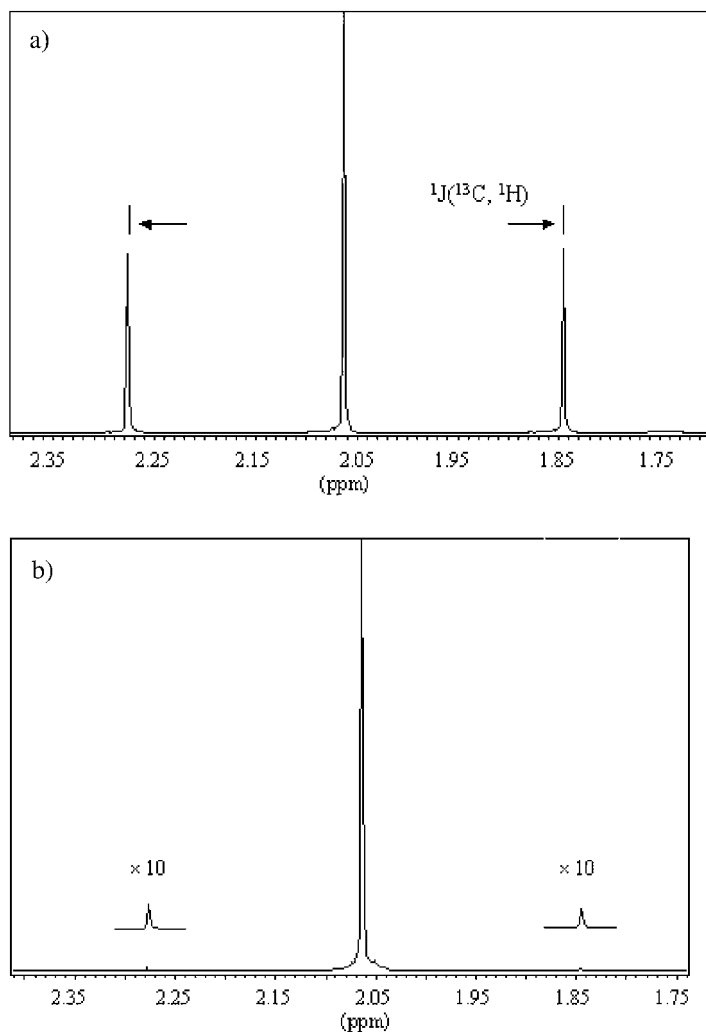


Fig. 2.12 Expanded region of the ^1H NMR spectra of (a) racemic mixture of (S) - ^{13}C -18/ (R) -18 and (b) (R) -18 alone [23 a].

resolution, which affords a mixture of true enantiomers (S) -19 and (R) -19 as well as labeled and non-labeled acetic acid ^{13}C -20 and 20, respectively, together with unreacted starting esters. At 50% conversion (or at any other point of the reaction) the ratio of (S) - ^{13}C -18 to (R) -18 correlates with the enantiomeric purity of the nonreacted ester, while the ratio of ^{13}C -20 to 20 reveals the relative amounts of (S) -19 and (R) -19, respectively (Scheme 2.4).

Figure 2.12a shows an excerpt of the ^1H NMR spectrum of a “racemic” mixture of (S) - ^{13}C -18 and (R) -18 featuring the expected doublet of the ^{13}C -labeled methyl group and the singlet of the nonlabeled methyl group. Figure 2.12b dis-

plays the singlet of the non-labeled methyl group of (*R*)-**18**, including the ^{13}C -satellites due to the presence of natural ^{13}C in the sample [23].

The exact ratio of the two *pseudo*-enantiomers is accessible by simple integration of the respective peaks, which provides the *ee*-value. The quantitative analysis can be accomplished automatically by suitable software such as AMIXTM (available from Bruker Biospin GmbH, Rheinstetten, Germany). The presence of naturally occurring ^{13}C in the nonlabeled (*R*)-substrate is automatically considered in the data processing step. As demonstrated by control experiments, the agreement with the corresponding *ee*-values obtained by independent GC analysis is excellent, the correlation coefficient amounting to 0.9998 [23].

Since the *ee*-value in an actual kinetic resolution depends on the degree of conversion, the selectivity factor *E* needs to be ascertained, which is possible if the conversion can be measured. In the present system this can be accomplished by automatic integration of the corresponding methine signals of the unreacted substrate ester at 5.9 ppm and the product alcohol at 4.9 ppm. Then the *E*-value can be estimated according to the method of Sih [6]. In other cases an internal standard may be more appropriate. The precision in the *ee*-values amounts to $\pm 2\%$ as checked by independent GC analysis. In the case of the first version of the high-throughput *ee*-assay based on traditional derivatization using chiral reagents such as Mosher's acid chloride, the same equipment and software can be used [23]. Again, about 1400 samples can be handled per day, precision in the *ee*-value in these cases being $\pm 5\%$ [23]. Thus, these two NMR-based *ee*-screening systems are practical, precise, and rather general.

In an extension of this method, a high-throughput NMR-based *ee*-assay using chemical shift imaging has been developed. The 19-capillary system allows the enantiomeric purity of 5600 samples to be determined within one day, the precision being $\pm 6.4\%$ [24]. In this NMR-tomographic method any commercially available NMR spectrometer can be used. In order to inject samples from 384-format deep-well microtiter plates into the bundle automatically, an in-house developed pipetting robot with a modified tray needs to be used.

Procedure 2.5: High-throughput NMR Assay [23]

All enzymatic reactions are performed in the wells of microtiter plates (96-format) in water (as in lipase-catalyzed hydrolytic reaction of (*S*)- ^{13}C -**18**/*(R)*-**18**), which is followed by a standard automatic extraction step. Depending upon the particular substrate to be assayed and upon the type of solvent used, it may be necessary to remove the solvent. However, this is often not necessary. In the case of enzymatic reactions in organic medium, solvent extraction is not required. For NMR analysis such solvents as CDCl_3 , D_6 -DMSO or D_2O are used. A minimum of about 6 μmol of substrate/product per mL of solvent is needed. Although the flow-through cell system does not need too much solvent (about 1 L in 24 h), the solvents can be mixed with the undeuterated form in 1:9 ratios in order to reduce costs [23].

Several flow-through NMR cells are commercially available, for example, the BEST™ NMR system (Bruker Biospin GmbH) as described here or the VAST™ NMR system (Varian). The schematic description is shown in Figure 2.13. In addition to the flow-through cell and the NMR spectrometer (300 MHz), the system requires an autosampler, for example, a Gilson 215 autosampler [23].

In the present case the hydrolytic kinetic resolution of the acetate of racemic 1-phenyl-ethanol (**19**) using a 1:1 mixture of the *pseudo*-enantiomers (*S*)-¹³C-**18** and (*R*)-**18** in water is carried out in the wells of microtiter plates (e.g. 96-format), followed by extraction (pipette robot) using 300 μL of CDCl₃. For storage the resulting organic layers are placed in the wells of microtiter plates (e.g. 96-format). The samples are then transferred to the autosampler of the BEST NMR system and analyzed using the high-speed mode as described below. The samples are taken by the movable needle and transferred into the first valve system. At the same time washing solution (CDCl₃) is introduced by the Dilutor 402 into the six-port selection valve. Injection occurs via the injection port, whereby the washing solution is first fed into the second six-port selection valve followed by the sample to be measured. The washing solution is then transferred rapidly via tubings into the flow-through cell by a hydraulic impulse. Immediately thereafter the sample follows which is separated from the washing solution by a small air gap. The washing solution is pumped through the flow-through cell, and once the sample has entered it, pumping is stopped and the NMR measurement is automatically initiated (maximum of four scans). During this time the washing solution is stored behind the cell [23].

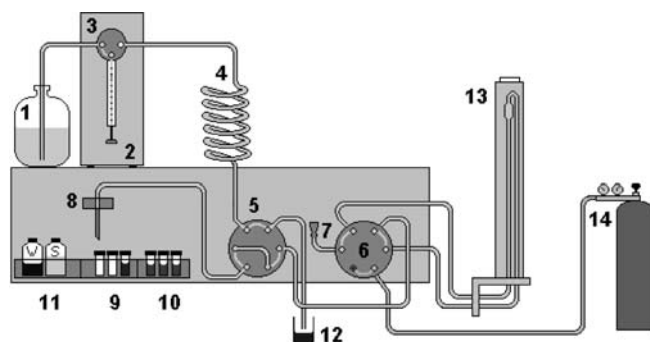


Fig. 2.13 Schematic representation of the BEST system (Bruker Biospin GmbH) [23 a]. (1) Bottle with transporting liquid; (2) Dilutor 402 single syringe (5 mL) with 1100 μL tube; (3) Dilutor 402 three-way valve; (4) Sample loop (250–500 μL); (5) Six-way valve (standard version)

(standard version) injecting sample; (7) Injection port; (8) XYZ needle; (9) Rack for sample vials; (10) Rack for recovering vials; (11) Rack for washing fluids and waste bottle (three glass bottles); (12) External waste bottle; (13) Flow probe with inner lock container; (14) Inert gas pressure bottle for drying process.

Since the same solvent is used for all samples, NMR locking and shimming is necessary only once at the beginning of the process. However, since the shim may not be constant, locking and shimming should be repeated after about every 10th sample. Following the NMR measurement, the sample and the washing solution are flushed out of the system by automatic pumping in the reverse direction. During the NMR measurement of one sample the next one is prepared by the autosampler. About 1400 samples can be handled per day. The analysis of the spectral data occurs with the aid of appropriate software, for example, Software AMIX (Bruker Biospin GmbH). For this purpose the region of the spectrum to be integrated needs to be defined. The data that are accumulated as a result of appropriate NMR peak integration are transferred onto Excel spreadsheets. The *ee*- or *E*-values are readily tabulated with the help of a macro [23].

2.6

Assay Based on Fourier Transform Infrared Spectroscopy for Assaying Lipases or Esterases

The basic idea of utilizing isotopic labeling for distinguishing *pseudo*-enantiomers in the kinetic resolution of chiral compounds and in the desymmetrization of prochiral substrates bearing reactive enantiotopic groups (Sections 2.4 and 2.5) can also be applied using Fourier transform infrared spectroscopy (FTIR) as the detection system in some cases [25]. Since FTIR spectroscopy is a cheap analytical technique available in almost all laboratories, the *ee*-assay has great potential. Moreover, it is of particular interest in the analysis of enzymatic reactions because the *ee*- or *E*-values can be measured directly in culture supernatants without time-consuming workup procedures. Of course, it is restricted to substrates of the type shown in Figure 2.9 which, however, must contain infrared (IR)-active functional groups. If all prerequisites are fulfilled, up to 10000 *ee*-values can be measured per day with an accuracy of $\pm 7\%$, making this a particularly practical and economical method for evaluating large libraries of potentially enantioselective enzymes [25].

When applying this *ee*-assay, the “best” position at which isotopes are introduced needs to be determined first. For the purpose of illustration the lipase- or esterase-catalyzed kinetic resolution of esters is considered here, although the method is not restricted to this type of reaction. ^{13}C -labeling of carbonyl groups is ideal for several reasons: Carbonyl groups provide intensive vibrational bands in an IR spectrum, allowing for easy and precise determination of the concentration of the compounds by applying Lambert-Beer’s law. In the spectral region between 1600 and 1800 cm^{-1} , which is typical for carbonyl stretching vibrations, almost no absorptions of other functional groups appear, eliminating interferences with other vibrational bands. The ^{13}C -labeled compounds can be easily prepared because reactive reagents with ^{13}C -labeled carbonyl groups like 1- ^{13}C -acetyl chloride are commercially available. Finally, the absorption maxima of the

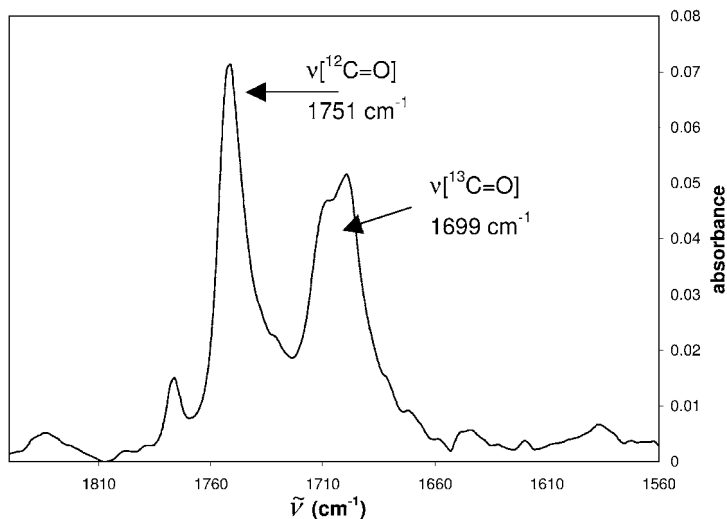


Fig. 2.14 Part of a Fourier transform infrared spectroscopy (FTIR) spectrum of a 1:1 mixture of (*R*)-**18** and (*S*)-¹³C-**18** [25].

carbonyl stretching vibration is shifted by 40–50 cm^{-1} to lower wave numbers by introducing a ¹³C-label, which prevents the overlap of the two carbonyl bands.

An example concerns the kinetic resolution of 1-phenylethyl acetate (**18**), previously used to illustrate the NMR-based *ee*-assay (see Section 2.5). The optimal way to proceed in the FTIR-based *ee*-assay is to apply ¹³C-labeling in the carbonyl moiety (i.e. to prepare a *pseudo*-racemate comprising a 1:1 mixture of (*S*)-¹³C-**18** and (*R*)-**18**). Figure 2.14 shows part of the FTIR spectrum 1:1 mixture of (*R*)-**18** and (*S*)-¹³C-**18**, illustrating the anticipated shift of the respective carbonyl stretching vibration which allows quantification of the *pseudo*-enantiomers [25].

It is necessary to apply Lambert-Beer's law in the calculation of the concentrations of the *pseudo*-enantiomeric substances, and for this purpose the molar coefficients of absorbance need to be determined. Thus, solutions of (*R*)-**18** and (*S*)-¹³C-**18** in cyclohexane at different concentrations have to be prepared. After measuring the corresponding absorbances at the absorption maxima of the carbonyl stretching vibration, it is possible to calculate the molar coefficients of absorbance by applying Lambert-Beer's law $E = \epsilon \times c \times d$ (Figure 2.15).

It is now possible to exploit the FTIR spectra of different synthetic mixtures of the labeled and non-labeled enantiomeric compounds. After applying an automated baseline correction to the spectra and correcting the absorbance of one enantiomer in the synthetic mixtures by the absorbance of the other enantiomer at this position, the accuracy of the *pseudo*-enantiomeric system based on 1-phenylethyl acetate is excellent, specifically within $\pm 3\%$ in comparison to the *ee*-values determined by chiral GC [25].

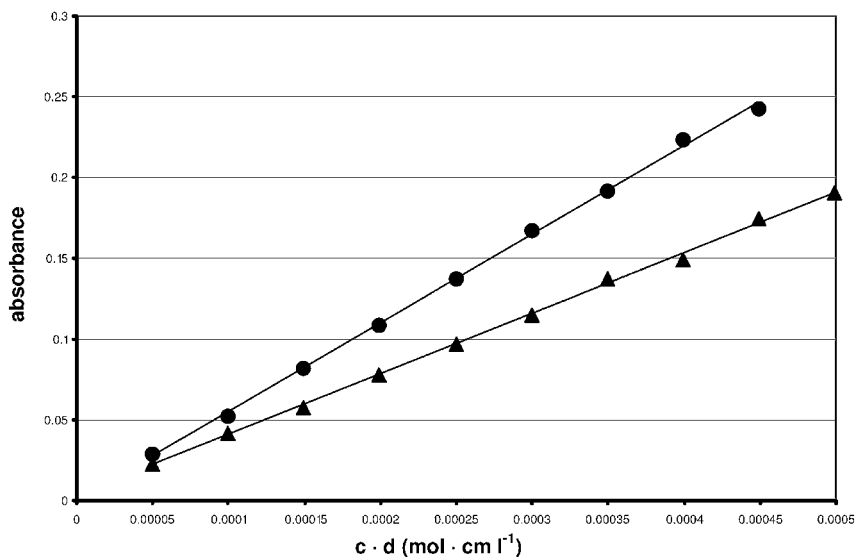


Fig. 2.15 Determination of the molar coefficients of absorbance of (R)-18 (circles, absorption maximum: 1751 cm^{-1}) and (S)- ^{13}C -18 (triangles, absorption maximum: 1699 cm^{-1}) by linear regression [25].

Using commercially available high-throughput screening (HTS)/FTIR systems, high-throughput measurements are easily possible. The analysis can be performed on a Tensor 27 FTIR spectrometer coupled to an HTS-XT system which is able to analyze the samples on 96- or 384-well microtiter plates. The plates are equipped with a silicon plate for IR transmittance. A great advantage of this system is the fact that the ϵ -values can be measured in culture supernatants, which is not possible in the case of the MS- or NMR-based assays. This means that the extraction step can be omitted.

In order to make this simple ϵ -assay possible, Bruker has already coupled the microplate stacking device TWISTER 1 to its microplate reader [25]. In this combination which is controlled by OPUS[®] software, 40 IR microplates can be measured automatically. In order to perform the sample loading at high throughput, the autosampler Microlab 4000 SP can be used. Both formats (96 and 384) of the Bruker silicon microplates are suitable for automatic loading of different types of samples (proteins, cells, culture media) [25]. At least 9000 samples can be handled per day.

Procedure 2.6: High-throughput ϵ -Determination Based on FTIR [25]

As a first step the molar coefficients of absorbance of the labeled and unlabeled substrate (in this case **18**) need to be determined. After preparation of a stock solution (0.200 M) of (R)-1-phenylethyl acetate ((R)-**18**) and (S)-1-

phenylethyl)-1-¹³C-acetate ((*S*)-¹³C-18) in cyclohexane, the solutions are diluted with cyclohexane to concentrations of 0.180, 0.160, 0.140, 0.120, 0.100, 0.080, 0.060, 0.040 and 0.020 M, respectively (total volume: 1 mL). The absorbance of the resulting samples is measured with an FTIR spectrometer at the corresponding absorption maxima of the carbonyl stretching vibration ((*R*)-18: 1751 cm⁻¹, (*S*)-¹³C-18: 1699 cm⁻¹) with a thickness of the layers of 25.0 μm. The measurements are carried out with 32 scans and a resolution of 4 cm⁻¹. The molar coefficients of absorbance are determined by linear regression. The correlation coefficients are in both cases better than 0.995 [25]. The analysis of synthetic mixtures of the *pseudo*-enantiomers of 1-phenylethyl acetate is performed under the same conditions at a concentration of 0.10 M.

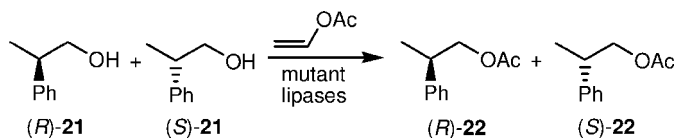
The high-throughput *ee*-measurements are performed on a Tensor 27 spectrometer in connection with a HTS-XT system provided by Bruker Optik GmbH. Aliquots of 3 μL of each supernatant mixture are transferred onto a 384-well microtiter plate equipped with a silicon plate for IR transmittance. Every sample is measured with a resolution of 8 cm⁻¹ and a scan number of 10, so that the total time for the analysis of each sample amounted to 8.9 s, allowing a throughput of >9000 samples per day. The resulting spectra were analyzed with the software Opus[®] and Opus Lab[®]. The first 14 samples are used for calibration purposes, while the remaining probes are taken as unknown samples. To evaluate the accuracy of the system, the *ee*-value of each mixture is independently determined by chiral GC analysis [25].

2.7

Assays Based on Gas Chromatography

It is well known that conventional chromatography (GC) using chiral stationary phases can only handle a few dozen *ee*-determinations per day. However, if medium- or high-throughput is desired, speed-up using automation is possible, although generally at the expense of precision. This disadvantage may be of no serious concern, since the purpose is to identify hits which can then be analyzed by a more precise system. Indeed, in some cases GC can be modified so that in optimal cases about 700–800 exact *ee*- and *E*-determinations are possible per day [26]. Such medium-throughput may suffice in certain applications. The first such case concerned the lipase-catalyzed kinetic resolution of the chiral alcohol (*R*)- and (*S*)-21 with formation of the acylated forms (*R*)- and (*S*)-22 [26a] (Scheme 2.5). Thousands of mutants of the lipase from *Pseudomonas aeruginosa* were created by epPCR for use as catalysts in the model reaction and were then screened for enantioselectivity [26a].

Although in principle it is possible to simply use several GC instruments each equipped with a sample manager and a PC, this is really not efficient because it is expensive, and at the same time data-handling becomes more diffi-



Scheme 2.5

cult. The first successful construction consisted of two GC instruments (e.g. GC instruments and data bus (HP-IB) are commercially available from Hewlett-Packard, Waldbronn, Germany), one prep-and-load sample manager (PAL) (commercially available from CTC, Schlieren, Switzerland) and one PC (Figure 2.16) [26 a]. The instruments are connected to the PC via a standardized data bus (HP-IB) (commercially available from Hewlett-Packard) which controls pressure, temperature, etc. and handles other data such as those of the detector. A wash station as well as a drawer system with a maximum of eight microtiter plates are included. The sample manager is attached to the unit in such a way as to reach both injection ports. Since the sample manager can inject samples from 96- or 384-well microtiter plates, over 3000 samples can be handled without manual intervention. The software Chemstation[®] (commercially available from Hewlett-Packard) enables additional programs (macros) to be applied before and after each analytical run. Such a macro controls the sample manager, each position on the microtiter plate being labeled via the sequence table. Another macro ensures analysis following each sample run in a specified manner (i.e. the peaks of the chiral compound **21** are analyzed quantitatively). The analytical data are transferred to an Excel spreadsheet via DDE (Dynamic Data Exchange) (commercially available from Microsoft, Unterschleissheim, Germany) in table form or in microtiter format, allowing for a rapid overview [26 a]. Final-

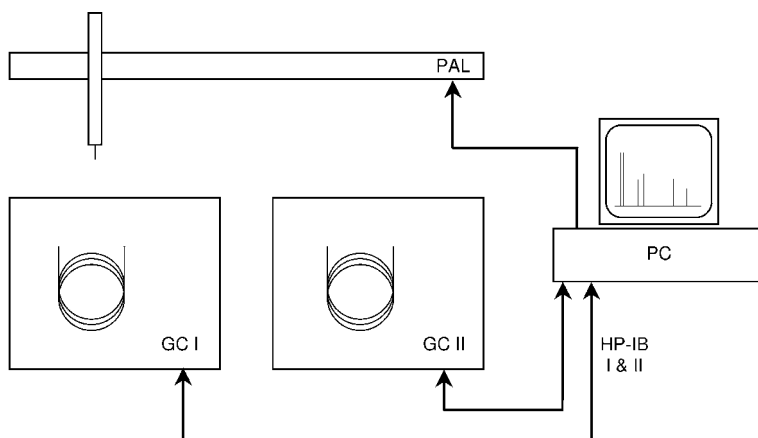


Fig. 2.16 Optimized unit for *ee*-determination based on two gas chromatography instruments driven by one PC [26 a].



Fig. 2.17 Picture of optimized unit for ee -determination based on three gas chromatography instruments driven by one PC [26c].

ly, the setup includes H_2 -guards which monitor the hydrogen concentration in the ovens; at concentrations exceeding 1% (potentially explosive at >4% H_2) the system responds and automatically switches to nitrogen as the carrier gas [26a].

Upon using a stationary phase (2,3-di-*O*-ethyl-6-*O*-*tert*-butyldimethylsilyl- β -cyclodextrin) complete separation of (*R*)- and (*S*)-**21** (but not of (*R*)/(*S*)-**22**) is possible within 3.9 min. Since the unit comprises two simultaneously operating GC units, about 700 exact ee -determinations of (*R*)/(*S*)-**21** are possible per day. The corresponding values for the conversion and the selectivity factor E (or s) are likewise automatically provided in microtiter format which means that the ee of (*R*)/(*S*)-**22** is also accessible. Of course, every new substrate has to be optimized anew using commercially available chiral stationary phases [26a].

More recently, this system has been extended to include three GC instruments operated by one PC (Figure 2.17). Experience gained in the initial GC system has shown that it is better to use one sample manager for each GC instrument, and not a single one for two or even three such instruments. The new system was applied successfully in the analysis of chiral products arising from catalysis of mutant Baeyer-Villiger monooxygenases [26b]. In this directed evolution project about 12 000 samples were screened.

Procedure 2.7: Medium-throughput GC-based ee -Assay [26a]

The GC system is described here using the lipase-catalyzed model reaction of *rac*-**21**. In principle mutants of any lipase can be used, but for the purpose of illustration lipase mutants from *Pseudomonas aeruginosa* were used [26a].

Mutants from *P. aeruginosa* lipase were obtained by epPCR as previously described [2]. Samples of the supernatants containing the mutant enzymes are brought into the glass reaction wells of a 96-membered microtiter plate

and lyophilized (freeze-dried), a process that has also been described [2a, b]. Then a mixture of the substrate *rac*-2-phenyl-1-propanol (**21**) (15 μ l) and the acylating agent (vinyl acetate, 30 μ l) in toluene (125 μ l) containing about 15 vol.% water (30 μ l) is added. It is important to note that normal plastic microtiter plates are not suitable because of the organic solvent used. The 96 wells are closed by covering them with an aluminum foil on which a silicone lid and an aluminum plate are placed and screw-fastened. Then the whole microtiter plate is placed on a shaking incubator for 24 h. Following centrifugation samples of 15 μ L of the reaction solutions are transferred into the wells of another glass microtiter plate together with 170 μ L of toluene. The GC measurements are performed using these mixtures [26a].

The sample manager comprises the HTS/PAL system (high-throughput-screening prep-and-load) [26a] and was incorporated in such a way that samples can be robotically taken from the wells of a collection of microtiter plates, injection into the two injection ports being possible. Accordingly, the sample manager is operated on the basis of microliter syringes, the cannulas being regularly cleaned in the wash station. The system includes drawers for storing microtiter plates. The final data are made available in the form of Excel tables or in microtiter format by the program Excel. In evaluating the GC chromatograms, calibration factors have to be determined first [26a].

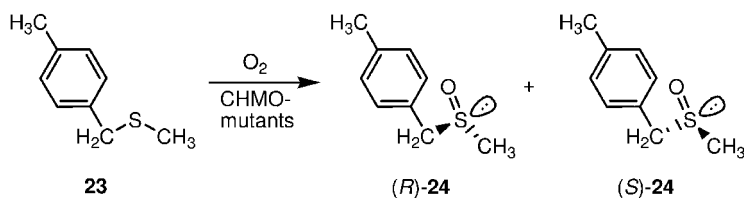
2.8

Assays Based on HPLC

The problems arising when developing a medium-throughput *ee*-assay based on HPLC are similar to those surrounding GC assays (Section 2.7). As before, speed is accomplished at the expense of precision, but systems can in fact be implemented which allow the identification of hits, which then need to be analyzed more precisely using conventional HPLC, using longer columns and of course longer reaction times. Nevertheless, only medium throughput can be achieved, meaning the evaluation of 300–800 samples per day (which suffices in some applications).

The first medium-throughput HPLC-based *ee*-assay was developed to identify hits in the directed evolution of enantioselective cyclohexanone monooxygenases (CHMO) as catalysts for the asymmetric O₂-mediated oxidation of prochiral thio-ethers such as **23** (Scheme 2.6) [27].

A commercially available HPLC instrument (Shimadzu LC 2010) was equipped with a sample manager and the appropriate software (Lab Solutions) to handle 96- or 384-well microtiter plates [27]. Since enantiomeric separation of (*R*)-**24**/*(S)*-**24** has to be fast and efficient, exploratory experiments were performed using a variety of different chiral stationary phases, solvents and conditions. An efficient system turned out to be benzoylated cellulose as the stationary phase with a mixture of *n*-heptane and ethanol as the mobile phase. For rapid analysis short columns



Scheme 2.6

50 mm in length and 4.5 mm in diameter were used. This system allows for at least 800 *ee*-determinations per day. Unfortunately, the *E. coli*-based expression system produces small amounts of indole, which leads to an overlap with the HPLC peak of (*S*)-**24**. Although this can be considered in the quantitative evaluation, the *ee*-values accessible under these conditions are not precise and were consequently used only to identify hits. These were subsequently analyzed precisely by a conventional HPLC setup using a longer column (250 mm) [27].

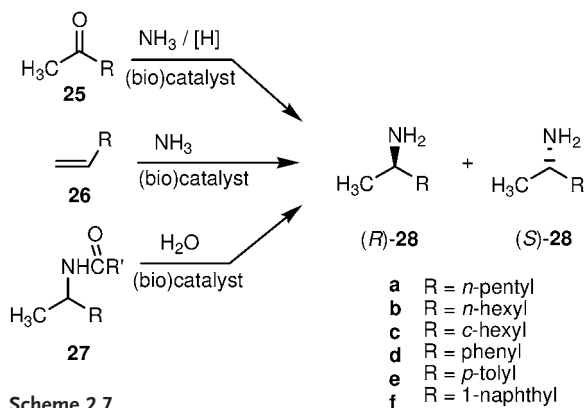
2.9

Assays Based on Capillary Array Electrophoresis

The chromatographic techniques described above may well serve as practical assays in some cases, but truly high-throughput encompassing thousands of *ee*-values per day is outside of the realm of these assays. In sharp contrast, capillary array electrophoresis (CAE) has been modified to allow the high-throughput determination of enantioselectivity [18]. It is well known that traditional capillary electrophoresis (CE) in which the electrolyte contains chiral selectors, such as β -cyclodextrin derivatives (β -CDs), can be used to determine the enantiomeric purity of a given sample. The conventional forms of this analytical technique allow for only a few dozen *ee* determinations per day. However, because of the analytical demands arising from the Human Genome Project, *inter alia*, CE has been revolutionized so that efficient techniques for instrumental miniaturization are now available, making super-high-throughput analysis of biomolecules possible for the first time.

Two different approaches have emerged, namely CAE [28] and CE on microchips (also called CAE on chips) [29]. Both techniques can be used to carry out DNA sequence analyses and/or to analyze oligonucleotides, DNA-restriction fragments, amino acids or PCR products. Many hundred thousand and more analytical data points can be accumulated per day.

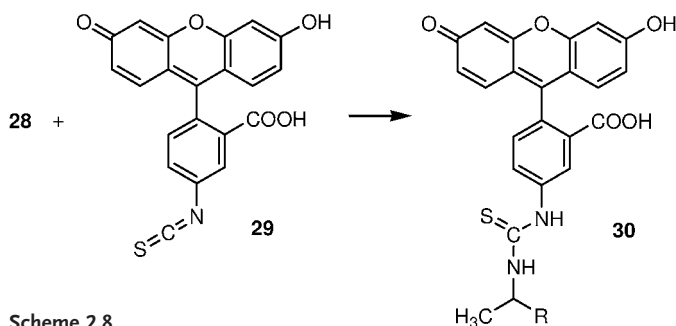
In the case of CAE, commercially available instruments have been developed that contain a high number of capillaries in parallel (e.g. the 96-capillary unit MegaBACE[®], which consists of six bundles of 16 capillaries; Amersham Pharmacia Biotech, Freiburg, Germany). The system can therefore address a 96-well microtiter plate. Each capillary is about 50 cm long. This system was adapted as a super-high-throughput analytical tool for *ee* determination [18]. In this study chiral amines of the type **28**, which are of importance in the synthesis of phar-



maceutical and agrochemical products, were used as the model substrates. They are potentially accessible by enzymatic reductive amination of ketones **25**, addition of ammonia to activated olefins **26** or hydrolysis of acetamides **27** (the reverse reaction also being possible) (Scheme 2.7).

The conditions for conventional CE assay of the amines **28** were first optimized using various β - and β -CD derivatives as chiral selectors [18]. In order to enable a sensitive detection system, specifically laser-induced fluorescence detection (LIF), the amines were derivatized with fluorescein-isothiocyanate (**29**) leading to fluorescence-active compounds **30** (Scheme 2.8). Although extensive optimization was not carried out (only six CD derivatives were tested), in all cases satisfactory baseline separation was accomplished.

The next step involved the use of compounds **28c/30c** as the model substrates for CAE analysis using an instrument of the kind MegaBACE[®]. Known enantiomeric mixtures of the amine **28c** were transformed into the fluorescence-active derivative **30c**. The latter samples were then analyzed by CAE. A special electrolyte having a high viscosity had to be developed. It is composed of 40 mM CHES pH 9.1/6.25 mM γ -CD 5:1 diluted with a buffer containing linear polyacrylamide. The MegaBACE instrument was operated at a potential of -10 kV/



Scheme 2.8

8 μA and a sample injection potential of $-2\text{ kV}/9\text{ s}$. Under these conditions, baseline separation is excellent. The agreement between ee -values of (*R*)/(*S*)-mixtures of **30c** determined by CAE and those of the corresponding (*R*)/(*S*)-mixtures of **28c** as measured by GC turned out to be excellent [18].

The enantiomer separation of (*R*)/(*S*)-**30c** on the MegaBACE instrument required about 19 min. This means that even though the conditions are far from optimized, the automated 96-array system provides more than 7000 ee determinations in a single day [18]. In related cases optimization resulted in shorter analysis times for enantiomer separation so that a daily throughput of 15000–30000 ee determinations is realistic. Such super-high-throughput screening for enantioselectivity is not readily possible by any other currently available technology. In view of the possibility of chiral selector optimization and the fact that CAE has many advantages such as extremely small amounts of samples, essentially no solvent consumption, absence of high-pressure pumps and valves as well as high durability of columns, this CAE assay is ideally suited for high-speed ee -determination. However, to date only chiral amines have been analyzed successfully, which means that more research/development is necessary when considering other types of compounds. A very cheap variation is on the horizon, namely CAE on microchips [18].

2.10

Assays Based on Circular Dichroism (CD)

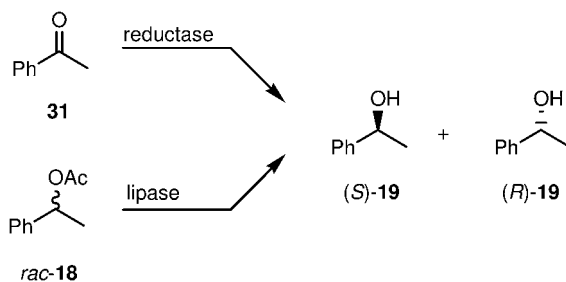
For certain applications circular dichroism (CD) is ideally suited as a high-throughput ee -assay, although to date it has been used in enzymatic reactions only once. An alternative to HPLC employing chiral columns which separate the enantiomers of interest is the use of normal columns that simply separate the starting materials from the enantiomeric products, the ee of the mixture of enantiomers then being determined by CD spectroscopy. Recently, it was shown that this classical method can be applied in high-throughput ee -screening [30, 31]. The method is based on the use of sensitive detectors for HPLC which determine in a parallel manner both the CD ($\Delta\epsilon$) and the UV absorption (ϵ) of a sample at a fixed wavelength in a flow-through system. The CD signal depends only on the enantiomeric composition of the chiral products, whereas the absorption relates to their concentration. Thus, only short HPLC columns are necessary [30, 31]. Upon normalizing the CD value with respect to absorption, the so-called anisotropy factor g is obtained:

$$g = \frac{\Delta\epsilon}{\epsilon}$$

It is thus possible to determine the ee -value without recourse to complicated calibration.

The fact that the method is theoretically valid only if the g -factor is independent of concentration and if it is linear with respect to ee has been emphasized

repeatedly [31]. However, it needs to be pointed out that these conditions may not hold if the chiral compounds form dimers or aggregates, because such enantiomeric or diastereomeric species would give rise to their own particular CD effects. Although such cases have yet to be reported, it is mandatory that this possibility be checked in each new system under study [31]. In work concerning the directed evolution of enantioselective enzymes there was the need to develop fast and efficient ways to determine the enantiomeric purity of these compounds, which can be produced enzymatically either by reduction of the prochiral ketone (e.g. **31**) using reductases or by kinetic resolution of *rac*-acetates (e.g. **18**) by lipases (Scheme 2.9). In both systems the CD approach is theoretically possible. In the former case, a liquid chromatography column would have to separate the educt **31** from the product (*S*)/(*R*)-**19**, whereas in the latter case (*S*)/(*R*)-**19** would have to be separated from (*S*)/(*R*)-**18** [31].



Scheme 2.9

Due to the fact that acetophenone (**31**) has a considerably higher extinction coefficient than 1-phenylethanol (**19**) at a similar wavelength (near 260 nm), separation of starting material from product was absolutely necessary, which was accomplished using a relatively short HPLC column based on a reversed-phase system. In preliminary experiments using enantiomerically pure product **19** the maximum value of the CD signal was determined [31]. Mixtures of **19** having different enantiomer ratios (and therefore *ee*-values) were prepared and analyzed precisely by chiral GC in control experiments. The same samples were studied by CD, resulting in the compilation of *g*-values. Upon plotting the *g*-against the *ee*-values, a linear dependency was in fact observed with a correlation factor of $r=0.99995$, which translates into the following simple equation for enantioselectivity:

$$ee = 3176.4 g - 8.0$$

It was then demonstrated that the *g*-factors have no dependency on concentration in the medium used.

Although complete optimization was not carried out, separation of **31** from (*S*)/(*R*)-**19** was in fact accomplished using reversed-phase silica as the column

material and methanol/water as the eluant. In view of the results concerning the dependency of the *g*-factor on concentration (see above), aggregation can be excluded in this protic medium. In the HPLC setup the mixture is fully separated within less than 90 s. Using the JASCO-CD-1595 instrument (Jasco International Co., Ltd., Tokyo) in conjunction with a robotic autosampler, it is possible to perform about 700–900 exact *ee* determinations per day [31]. Thus, this system is practical.

2.11

Assay Based on Surface-enhanced Resonance Raman Scattering

A novel method for assaying the activity and stereoselectivity of enzymes at *in vitro* concentrations is based on surface-enhanced resonance Raman scattering (SERRS) using silver nanoparticles [32]. Turnover of a substrate leads to the release of a surface targeting dye, which is detected by SERRS. In a model study lipase-catalyzed kinetic reduction of a dye-labeled chiral ester was studied [32]. It is currently unclear how precise the method is when identifying mutants which lead to *E*-values higher than 10. However, as a pretest for activity the assay is well suited.

2.12

Conclusions

This chapter covers the most important medium- and high-throughput *ee*-assays that can be used to evaluate the enantioselectivity of enzymes. No single assay is truly universal, and many are complementary. The currently most practical high-throughput systems which are potentially adaptable to many different enzymes and substrates are based on MS, NMR or CD. They also show the highest degrees of precision (better than $\pm 6\%$), which is important in the late stages of directed evolution studies. Using these assays typically 1500–8000 *ee* determinations are realistically possible per day. If in a given study medium-throughput suffices, meaning 300–800 samples per day, then adapted forms of GC or HPLC may solve the screening problem, although success will depend upon the particular nature of the product under consideration. Thus, taken together the *ee*-assays described herein can solve many analytical problems in the directed evolution of enantioselective enzymes, although not all of them have actually been tested under “operating conditions.” The development of new *ee*-assays continues to be a fascinating and rewarding area of research [33].

References

- 1 a) H.G. Davies, R.H. Green, D.R. Kelly, S.M. Roberts, eds, in *Biotransformations in Preparative Organic Chemistry: The Use of Isolated Enzymes and Whole Cell Systems in Synthesis*, Academic Press, London, **1989**; b) C.H. Wong, G.M. Whitesides, *Enzymes in Synthetic Organic Chemistry, Tetrahedron Organic Chemistry Series*, Vol. 12, Pergamon, Oxford, **1994**; c) K. Drauz, H. Waldmann, *Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook*, Vols I–III, 2nd edn, VCH, Weinheim, **2002**; d) K. Faber, *Biotransformations in Organic Chemistry*, 4th edn, Springer, Berlin, **2000**; e) A. Liese, K. Seelbach, C. Wandrey, *Industrial Biotransformations*, Wiley-VCH, Weinheim, **2000**.
- 2 a) M.T. Reetz, A. Zonta, K. Schimossek, K. Liebeton, K.-E. Jaeger, *Angew. Chem.* **1997**, *109*, 2961–2963; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 2830–2832. b) K. Liebeton, A. Zonta, K. Schimossek, M. Nardini, D. Lang, B.W. Dijkstra, M.T. Reetz, K.-E. Jaeger, *Chem. Biol.* **2000**, *7*, 709–718. c) M.T. Reetz, S. Wilensek, D. Zha, K.-E. Jaeger, *Angew. Chem.* **2001**, *113*, 3701–3703; *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 3589–3591. d) D. Zha, S. Wilensek, M. Hermes, K.-E. Jaeger, M.T. Reetz, *Chem. Commun. (Cambridge, UK)* **2001**, 2664–2665. e) M.T. Reetz, *Pure Appl. Chem.* **2000**, *72*, 1615–1622. f) M.T. Reetz, K.-E. Jaeger, *Chem.-Eur. J.* **2000**, *6*, 407–412. g) M.T. Reetz, *Tetrahedron* **2002**, *58*, 6595–6602. h) M.T. Reetz, Changing the enantioselectivity of enzymes by directed evolution, in *Methods in Enzymology*, Vol. 388 (Protein Engineering), eds D.E. Robertson, J.P. Noel, Elsevier Academic Press, San Diego, **2004**, pp. 238–256. g) M.T. Reetz, *Proc. Natl Acad. Sci. USA* **2004**, *101*, 5716–5722.
- 3 For example: a) S.F. Brady, C.J. Chao, J. Handelsman, J. Clardy, *Org. Lett.* **2001**, *3*, 1981–1984. b) P. Hugenholtz, B.M. Goebel, N.R. Pace, *J. Bacteriol.* **1998**, *180*, 4765–4774. c) S.B. Bintrim, T.J. Donohue, J. Handelsman, G.P. Roberts, R.M. Goodman, *Proc. Natl Acad. Sci. USA* **1997**, *94*, 277–282. d) G. DeSantis, Z. Zhu, W.A. Greenberg, K. Wong, J. Chaplin, S.R. Hanson, B. Farwell, L.W. Nicholson, C.L. Rand, D.P. Weiner, D.E. Robertson, M.J. Burk, *J. Am. Chem. Soc.* **2002**, *124*, 9024–9025.
- 4 a) F.H. Arnold, *Nature* **2001**, *409*, 253–257. b) K.A. Powell, S.W. Ramer, S.B. del Cardayré, W.P.C. Stemmer, M.B. Tobin, P.F. Longchamp, G.W. Huisman, *Angew. Chem.* **2001**, *113*, 4068–4080; *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 3948–3959. c) J.D. Sutherland, *Curr. Opin. Chem. Biol.* **2000**, *4*, 263–269. d) S. Brakmann, K. Johnsson, *Directed Molecular Evolution of Proteins (or How to Improve Enzymes for Biocatalysis)*, Wiley-VCH, Weinheim, **2002**; e) F.H. Arnold, G. Georgiou, Directed enzyme evolution: screening and selection methods, in *Methods and Molecular Biology*, Vol. 230, Humana Press, Totowa, NJ, **2003**.
- 5 Reviews: a) M.T. Reetz, *Angew. Chem.* **2001**, *113*, 292–320; *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 284–310; b) M.T. Reetz, *Angew. Chem.* **2002**, *114*, 1391–1394; *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 1335–1338; c) D. Wahler, J.-L. Reymond, *Curr. Opin. Biotechnol.* **2001**, *12*, 535–544; d) S. Dahmen, S. Bräse, *Synthesis* **2001**, 1431–1449; e) M.T. Reetz, in *Methods in Molecular Biology*, Vol. 230, eds F.H. Arnold, D. Georgiou, Humana Press, Totowa, NJ, **2003**, pp. 259–282; f) M.T. Reetz, *Methods in Molecular Biology*, Vol. 230, eds F.H. Arnold, G. Georgiou, Humana Press, Totowa, NJ, **2003**, pp. 283–290; g) M.T. Reetz, Screening for enantioselective enzymes, in *Enzyme Functionality – Design, Engineering, and Screening*, ed. A. Svendsen, Marcel Dekker, New York, **2004**, pp. 559–598; h) M.T. Reetz, High-throughput screening of enantioselective industrial biocatalysts, in *Evolutionary Methods in Biotechnology*, eds S. Brakmann, A. Schwienhorst, Wiley-VCH, Weinheim, **2004**, pp. 113–141.
- 6 C.-S. Chen, Y. Fujimoto, G. Girdaukas, C.J. Sih, *J. Am. Chem. Soc.* **1982**, *104*, 7294–7299.

- 7 a) C. Gennari, U. Piarulli, *Chem. Rev.* **2003**, *103*, 3071–3100; b) M. L. Snapper, A. H. Hoveyda, in *Combinatorial Chemistry*, ed. H. Fenniri, Oxford University Press, New York, **2000**, p. 433; c) J. G. de Vries, A. H. M. de Vries, *Eur. J. Org. Chem.* **2003**, 799–811.
- 8 M. J. Dröge, C. J. Rüggeberg, A. M. van der Sloot, J. Schimmel, R. S. Dijkstra, R. M. D. Verhaert, M. T. Reetz, W. J. Quax, *J. Biotechnol.* **2003**, *4*, 19–28.
- 9 M. T. Reetz, C. J. Rüggeberg, *Chem. Commun. (Cambridge, UK)* **2002**, 1428–1429.
- 10 a) G. Klein, J.-L. Reymond, *Helv. Chim. Acta* **1999**, *82*, 400–406; b) E. Leroy, N. Benschel, J.-L. Reymond, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2105–2108; c) D. Lagarde, H.-K. Nguyen, G. Ravot, D. Wahler, J.-L. Reymond, G. Hills, T. Veit, F. Lefevre, *Org. Process Res. Dev.* **2002**, *6*, 441–445; d) F. Badalassi, D. Wahler, G. Klein, P. Crotti, J.-L. Reymond, *Angew. Chem.* **2000**, *112*, 4233–4236; *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 4067–4077; e) J.-L. Reymond, D. Wahler, *ChemBioChem* **2002**, *3*, 701–708.
- 11 M. Baumann, R. Stürmer, U. T. Bornscheuer, *Angew. Chem.* **2001**, *113*, 4329–4333; *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 4201–4204.
- 12 P. Abato, C. T. Seto, *J. Am. Chem. Soc.* **2001**, *123*, 9206–9207.
- 13 F. Taran, C. Gauchet, B. Mohar, S. Meunier, A. Valleix, P. Y. Renard, C. Crémignon, J. Grassi, A. Wagner, C. Mioskowski, *Angew. Chem.* **2002**, *114*, 132–135; *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 124–127.
- 14 a) L. E. Janes, R. J. Kazlauskas, *J. Org. Chem.* **1997**, *62*, 4560–4561; b) L. E. Janes, A. C. Löwendahl, R. J. Kazlauskas, *Chem. Eur. J.* **1998**, *4*, 2324–2331; c) A. M. F. Liu, N. A. Somers, R. J. Kazlauskas, T. S. Brush, F. Zocher, M. M. Enzelberger, U. T. Bornscheuer, G. P. Horsman, A. Mezzetti, C. Schmidt-Dannert, R. D. Schmid, *Tetrahedron Asymmetry* **2001**, *12*, 545–556; d) F. Moris-Varas, A. Shah, J. Aikens, N. P. Nadkarni, J. D. Rozzell, D. C. Demirjian, *Bioorg. Med. Chem.* **1999**, *7*, 2183–2188.
- 15 G. A. Korbil, G. Lalic, M. D. Shair, *J. Am. Chem. Soc.* **2001**, *123*, 361–362.
- 16 A. Horeau, A. Nouaille, *Tetrahedron Lett.* **1990**, *31*, 2707–2710.
- 17 E. R. Jarvo, C. A. Evans, G. T. Copeland, S. J. Miller, *J. Org. Chem.* **2001**, *66*, 5522–5527.
- 18 M. T. Reetz, K. M. Kühling, A. Deege, H. Hinrichs, D. Belder, *Angew. Chem.* **2000**, *112*, 4049–4052; *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 3891–3893.
- 19 J. Guo, J. Wu, G. Siuzdak, M. G. Finn, *Angew. Chem.* **1999**, *111*, 1868–1871; *Angew. Chem. Int. Ed. Engl.* **1999**, *38*, 1755–1758.
- 20 a) M. T. Reetz, M. H. Becker, H.-W. Klein, D. Stöckigt, *Angew. Chem.* **1999**, *111*, 1872–1875; *Angew. Chem. Int. Ed. Engl.* **1999**, *38*, 1758–1761; b) W. Schrader, A. Eipper, D. J. Pugh, M. T. Reetz, *Can. J. Chem.* **2002**, *80*, 626–632; c) M. T. Reetz, M. H. Becker, D. Stöckigt, H.-W. Klein, patent application DE-A 19913858.3, PCT/EP 00/02121 and WO 00/58504; d) A. Funke, A. Eipper, M. T. Reetz, N. Otte, W. Thiel, G. van Pouderoyen, B. W. Dijkstra, K.-E. Jaeger, T. Eggert, *Biocatal. Biotransform.* **2003**, *21*, 67–73; e) D. Zha, A. Eipper, M. T. Reetz, *ChemBioChem* **2003**, *4*, 34–39; f) F. Cedrone, S. Niel, S. Roca, T. Bhatnagar, N. Ait-Abdelkader, C. Torre, H. Krumm, A. Maichele, M. T. Reetz, J. C. Baratti, *Biocatal. Biotransform.* **2003**, *21*, 357–364; g) M. T. Reetz, C. Torre, A. Eipper, R. Lohmer, M. Hermes, B. Brunner, A. Maichele, M. Bocola, M. Arand, A. Cronin, Y. Genzel, A. Archelas, R. Furstoss, *Org. Lett.* **2004**, *6*, 177–180.
- 21 G. DeSantis, K. Wong, B. Farwell, K. Chatman, Z. Zhu, G. Tomlinson, H. Huang, X. Tan, L. Bibbs, P. Chen, K. Kretz, M. J. Burk, *J. Am. Chem. Soc.* **2003**, *125*, 11476–11477.
- 22 a) M. J. Shapiro, J. S. Gounarides, *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *35*, 153–200; b) H. Schröder, P. Neidig, G. Rossé, *Angew. Chem.* **2000**, *112*, 3974–3977; *Angew. Chem. Int. Ed.* **2000**, *39*, 3816–3819; c) C. L. Gavaghan, J. K. Nicholson, S. C. Connor, I. D. Wilson, B. Wright, E. Holmes, *Anal. Biochem.* **2001**, *291*, 245–252; d) E. MacNamara, T. Hou, G. Fisher, S. Williams, D. Raftery, *Anal. Chim. Acta* **1999**, *397*, 9–16.

- 23 a) M. T. Reetz, A. Eipper, P. Tielmann, R. Mynott, *Adv. Synth. Catal.* **2002**, *344*, 1008–1016; b) M. T. Reetz, P. Tielmann, A. Eipper, R. Mynott, patent application DE-A 10209177.3.
- 24 M. T. Reetz, P. Tielmann, A. Eipper, A. Ross, G. Schlotterbeck, *Chem. Commun. (Cambridge, UK)* **2004**, 1366–1367.
- 25 P. Tielmann, M. Boese, M. Luft, M. T. Reetz, *Chem. Eur. J.* **2003**, *9*, 3882–3887.
- 26 a) M. T. Reetz, K. M. Kühling, S. Wilensek, H. Husmann, U. W. Häusig, M. Hermes, *Catal. Today* **2001**, *67*, 389–396; b) M. T. Reetz, B. Brunner, T. Schneider, F. Schulz, C. M. Clouthier, M. M. Kayser, *Angew. Chem.* **2004**, *116*, 4167–4170; *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 4075–4078; c) T. Schneider, Dissertation, Ruhr-Universität Bochum, Germany, 2004.
- 27 M. T. Reetz, F. Daligault, B. Brunner, H. Hinrichs, A. Deege, *Angew. Chem.* **2004**, *116*, 4170–4173; *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 4078–4081.
- 28 a) X. C. Huang, M. A. Quesada, R. A. Mathies, *Anal. Chem.* **1992**, *64*, 2149–2154; b) H. Kambara, S. Takahashi, *Nature (London)* **1993**, *361*, 565–566; c) N. J. Dovichi, *Electrophoresis* **1997**, *18*, 2393–2399; d) G. Xue, H. Pang, E. S. Yeung, *Anal. Chem.* **1999**, *71*, 2642–2649; e) S. Behr, M. Mätzig, A. Levin, H. Eickhoff, C. Heller, *Electrophoresis* **1999**, *20*, 1492–1507.
- 29 a) D. J. Harrison, K. Fluri, K. Seiler, Z. Fan, C. S. Effenhauser, A. Manz, *Science (Washington)* **1993**, *261*, 895–897; b) S. C. Jacobson, R. Hergenroder, L. B. Koutny, R. J. Warmack, J. M. Ramsey, *Anal. Chem.* **1994**, *66*, 1107–1113; c) L. D. Hutt, D. P. Glavin, J. L. Bada, R. A. Mathies, *Anal. Chem.* **1999**, *71*, 4000–4006; d) D. Schmalzing, L. Koutny, A. Adourian, P. Belgrader, P. Matsudaira, D. Ehrlich, *Proc. Natl Acad. Sci. USA* **1997**, *94*, 10273–10278; e) S. C. Jacobson, C. T. Culbertson, J. E. Daler, J. M. Ramsey, *Anal. Chem.* **1998**, *70*, 3476–3480; f) S. Liu, H. Ren, Q. Gao, D. J. Roach, R. T. Loder, Jr., T. M. Armstrong, Q. Mao, I. Blaga, D. L. Barker, S. B. Jovanovich, *Proc. Natl Acad. Sci. USA* **2000**, *97*, 5369–5374; g) S. R. Wallenborg, C. G. Bailey, *Anal. Chem.* **2000**, *72*, 1872–1878; h) I. Rodriguez, L. J. Jin, S. F. Y. Li, *Electrophoresis* **2000**, *21*, 211–219.
- 30 a) K. Ding, A. Ishii, K. Mikami, *Angew. Chem.* **1999**, *111*, 519–523; *Angew. Chem. Int. Ed. Engl.* **1999**, *38*, 497–501; b) R. Angelaud, Y. Matsumoto, T. Korenaga, K. Kudo, M. Senda, K. Mikami, *Chirality* **2000**, *12*, 544–547.
- 31 M. T. Reetz, K. M. Kühling, H. Hinrichs, A. Deege, *Chirality* **2000**, *12*, 479–482.
- 32 B. D. Moore, L. Stevenson, A. Watt, S. Flitsch, N. J. Turner, C. Cassidy, D. Graham, *Nature Biotechnol.* **2004**, *22*, 1133–1138.
- 33 *Note added in proof:* A highly efficient UV/Vis-based high-throughput *ee*-assay for alcohols has recently been developed which makes use of two enantioselective enzymes for modifying the chiral products: Z. Li, L. Bütikofer, B. Witholt, *Angew. Chem.* **2004**, *116*, 1730–1734; *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 1698–1702.

3

High-throughput Screening Methods Developed for Oxidoreductases *

Tyler W. Johannes, Ryan D. Woodyer, and Huimin Zhao

3.1 Introduction

The interest in using enzymes as synthetic chemistry tools continues to grow rapidly [1, 2]. Among various classes of enzymes, oxidoreductases represent a highly versatile class of biocatalysts for specific reduction, oxidation and oxy-functionalization reactions, and are currently used for the production of a wide variety of chemical and pharmaceutical products. Examples of such products include *L-tert*-leucine, 6-hydroxynicotinic acid, 5-methylpyrazine-2-carboxylic acid, (*R*)-2-hydroxyphenoxypropionic acid, and steroids [3]. Oxidoreductases are generally employed in whole-cell biotransformations or fermentation-based processes because they require expensive cofactors and coenzymes (e.g. NAD(H) or NADP(H)) to donate or accept the chemical equivalents for reduction or oxidation. To facilitate their applications *in vitro*, a large number of cofactor regeneration systems have been developed [4, 5].

As with most enzyme biocatalysts, the discovery and engineering of oxidoreductases with improved stability, catalytic activity, and substrate specificity are highly desirable for many applications. Thanks to recent advances in recombinant DNA technology, new enzymes can now be isolated directly from microorganisms that are difficult to cultivate by high-throughput screening of expressed libraries of environmental DNA [6], whereas improved enzymes with desired functions may be engineered rapidly by directed evolution approaches [7]. Identifying new or improved oxidoreductases by these routes requires the development of high-throughput screening methods that are sensitive, efficient, and simple to implement.

This chapter focuses on various high-throughput screening methods recently developed for the discovery and directed evolution of oxidoreductases. For the purposes of this chapter, oxidoreductases have been classified into four categories: dehydrogenases, oxidases, oxygenases, and laccases. Each section begins by

* Please find a list of abbreviations at the end of the chapter.

discussing the general uses for each oxidoreductase, then explains the details of various high-throughput screening assays, and ends with specific examples of how these methods have been used to date.

3.2

High-throughput Methods for Various Oxidoreductases

3.2.1

Dehydrogenases

Dehydrogenases catalyze the removal and transfer of hydrogen from a substrate in an oxidation-reduction reaction and are widely used to reduce carbonyl groups of aldehydes or ketones and carbon/carbon double bonds. The reduction reactions carried out by dehydrogenases often introduce chirality into a prochiral center, which may represent a key step in pharmaceutical and fine chemical synthesis. The majority of dehydrogenase enzymes require nicotinamide cofactors (NAD(P) and NAD(P)H), while other cofactors such as pyrroloquinoline quinone (PQQ) and flavins (FAD, FMN) are encountered more rarely [1]. The production or depletion of the nicotinamide cofactors has enabled scientists to measure the kinetics of several dehydrogenases. These reactions can be monitored colorimetrically by combining the redox reaction with a dye-forming reaction. High-throughput screening methods based on this approach have been developed and used to identify dehydrogenase variants with improved characteristics in various directed evolution studies.

3.2.1.1 Colorimetric Screen Based on NAD(P)H Generation

The absorbance of NAD(P)H at 340 nm is commonly used to measure the activity of dehydrogenases. Unfortunately, this approach is generally not suitable for high-throughput methods due to background noise created by cell lysates and the 96-well plate itself. Special plates designed so that they do not absorb in the UV range can be used, but tend to be very expensive when considering a large-scale screening effort. Colorimetric assays resolve most of these difficulties and are quite amenable to high-throughput screening methods. Thus an indirect method requiring either a synthetic compound or a secondary enzyme must be applied. Tetrazolium salts such as nitroblue tetrazolium (NBT) (**1**) are useful because they can be reduced to formazan dyes such as **2**, which absorb light in the visible region. These reactions are essentially irreversible under biological conditions and the increase in color can be easily monitored visually on filter discs or on a standard 96-well plate reader. A cascade reaction leading to the formation of a colored formazan links the production of NAD(P)H to the catalytic activity of a dehydrogenase.

An NBT/phenazine methosulfate (PMS) assay has been developed that works well under several different high-throughput formats. Dehydrogenase activity on

nitrocellulose membranes, electrophoretic gels, and in liquid phase has been successfully monitored using the NBT/PMS assay. The dehydrogenase cofactor NAD(P)H reacts with NBT (1) in the presence of PMS (3) to produce an insoluble blue-purple formazan (2) (Figure 3.1). Other transferring agents besides PMS may be used such as the enzyme diaphorase or Meldola's Blue (8-dimethylamino-2-benzophenoxazine, 4) [8]. Compared with PMS, Meldola's Blue has a superior proton transfer rate and is far less sensitive to light. In addition, the cost of Meldola's Blue is approximately US\$4 per gram, while the price of PMS is nearly US\$16 per gram (Sigma catalog, 2004, St. Louis, MO, USA). Despite these disadvantages, PMS is used as the primary transferring agent in the NBT colorimetric assay. For the liquid-phase NBT/PMS assay, 0.1% gelatin is often added to the assay mix to help prevent precipitation of the insoluble blue-purple formazan (2) [9]. When scanned spectrophotometrically (range 400–700 nm), the blue-purple formazan (2) shows a maximal absorption at 555 nm.

By measuring the increase in formazan (2) production at 580 nm in a microplate reader, the NBT/PMS assay has been used to identify *E. coli* 6-phosphogluconate dehydrogenase (6PGDH) variants with higher activity and thermostability [10]. This assay can also be used to screen dehydrogenase variants on a nitrocellulose membrane. Briefly, a library of variants is transferred to a nitrocellulose filter followed by lysis of the colonies by washing in a lysozyme solution. A NBT/PMS cocktail is then sprayed onto the membrane to detect dehydrogenase activity. Active mutants produce a blue halo on the nitrocellulose filters. This approach has been successfully used by Holbrook and coworkers to produce a lactate dehydrogenase that no longer requires the expensive activator fructose-1,6-bisphosphate (FBP) [11]. The same group also used this approach to create new substrate specificities for lactate dehydrogenase [12].

The NBT/PMS assay has been used to identify novel alcohol dehydrogenases (ADH) in a high-throughput format without enzyme purification [13]. This new screening approach uses bioinformatics, polymerase chain reaction (PCR) techniques, and the direct *in vitro* expression of enzymes to rapidly detect and characterize novel ADHs. A pool of 18 novel thermoactive ADHs with a broad substrate range was discovered and characterized.

The compound *p*-rosaniline (5) can also be used as an alternative to NBT (1). No transferring agents are necessary when using *p*-rosaniline (5), however bisulfite must be added to help convert the dye into its rose-colored form. Dehydrogenase activity can be monitored by observing the formation of a red color on solid agar media and this allows the rapid screening of thousands of colonies [14].

3.2.1.2 Screens Based on NAD(P)H Depletion

NAD(P)H consumption can be monitored either directly or indirectly, similar to NAD(P)H generation in the previous section (see Section 3.2.1.1). Enzyme kinetics can be directly monitored by observing the decrease in absorbance at 340 nm as NAD(P)H is depleted. Coupling NAD(P)H depletion to a liquid- or

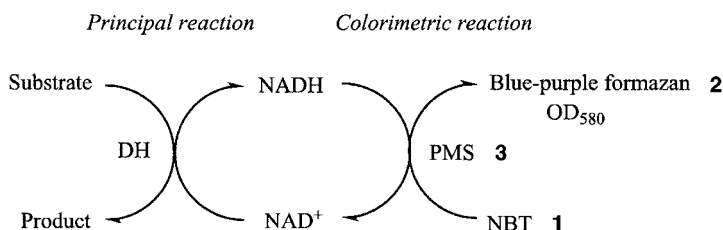


Fig. 3.1 Colorimetric assay for dehydrogenase activity. DH=dehydrogenase; PMS=phenazine methosulfate; NBT=nitroblue tetrazolium.

solid-phase colorimetric assay such as NBT/PMS (described in Section 3.2.1.1) allows for an indirect method for monitoring enzyme activity. Residual NAD(P)H can be titrated with NBT (1) in the presence of PMS (3) to produce an insoluble blue-purple formazan 2 (Figure 3.1).

High-throughput methods involve use of a microplate reader to measure the decrease in absorption at 580 nm or screening bacterial colonies on agar plates by visually identifying white spots on a purple background. Instead of focusing on NAD(P)H, the oxidized form NAD(P)⁺ provides an alternative. The basis for this method relies on the instability of NAD(P)H in a strong alkali environment. Under these conditions NAD(P)H breaks down to form highly fluorescent products [15]. Recently this method has been adapted to a high-throughput screening method using microtiter plates [16]. In general, any oxidoreductase that utilizes either NADH or NADPH as a cofactor can potentially be screened using this method.

3.2.2

Oxidases

Oxidases catalyze the removal of hydrogen from substrates and transfer it to molecular oxygen to form either hydrogen peroxide or water. These fundamental synthetic reactions have made oxidases a prime target for directed evolution. Some oxidases are currently being used on a large scale, such as the food antioxidant D-glucose oxidase, but unfortunately the vast majority of oxidases remain economically unattractive due to their poor biocatalytic performance such as low catalytic activity [1]. Thus, several high-throughput screening methods have been developed for directed evolution studies of these oxidases with the goal of overcoming the limitations of naturally occurring enzymes.

3.2.2.1 Galactose Oxidase

Galactose oxidases catalyze the oxidation of primary alcohols to yield aldehydes and hydrogen peroxide. These synthetically useful chemical transformations have made galactose oxidases an attractive target of directed evolution in recent years [17, 18]. Standard chemical methods for oxidizing primary alcohols rely

on environmentally unfriendly heavy metals and organic solvents, whereas galactose oxidase provides an environmentally friendly alternative. Galactose oxidase has broad substrate specificity as well as strict regioselectivity and stereoselectivity which make it widely applicable in chemical synthesis, diagnostics, and biosensors [17]. For example, the enzyme can be used to detect a disaccharide tumor marker, Gal-GalNAc, in certain types of colon cancer [19]. Galactose oxidase isolated from its natural host has insufficient expression and activity to be economically used in a large-scale industrial process. Liquid- and solid-phase high-throughput screening methods have been developed to help analyze galactose oxidase libraries to identify variants with improved expression and specific activity.

HRP/ABTS/ H_2O_2 Assay

Galactose oxidase activity can be monitored by coupling the production of hydrogen peroxide (H_2O_2) to a horseradish peroxidase (HRP)/2,2'-azinobis(3-ethylbenzthiazoline)-6-sulfonic acid (ABTS)/ H_2O_2 reaction [19]. The enzyme HRP catalyzes the polymerization of ABTS (**6**) to produce a soluble end-product (**7**) that is green in color and can be measured spectrophotometrically at 405 nm (Figure 3.2). The high solubility of ABTS (**6**) in aqueous solutions combined with its high extinction coefficient and reproducibility make this assay well suited for a 96-well plate format.

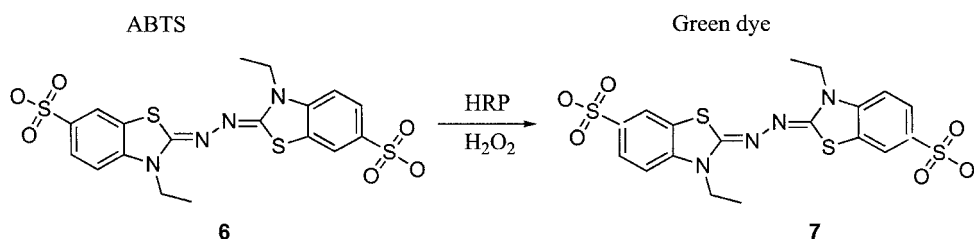


Fig. 3.2 Colorimetric assay for oxidase activity based on the HRP/ABTS/ H_2O_2 reaction. HRP=horseradish peroxidase; ABTS=2,2'-azinobis(3-ethylbenzthiazoline)-6-sulfonic acid.

Other advantages of ABTS (**6**) are its low toxicity, stability at high temperature (for sterilization), and the resulting green dye is relatively stable once oxidized. However, problems such as autooxidation, instability, and a lack of sensitivity have been reported [20]. This method has been used to create a galactose oxidase enzyme with enhanced thermostability and improved expression in *E. coli* [17]. It has also been successfully used to screen fungal sugar oxidases [21].

HRP/4CN/ H_2O_2 Assay

In addition to traditional 96-well plate screening methods, a relatively new technology involving digital imaging shows great potential in screening large enzyme libraries. The efficiency of digital imaging technology provides a distinct

advantage over its traditional counterpart [18]. This system combines simple well-known colorimetric activity assays with single-pixel imaging spectroscopy on a solid-phase screening format. The assay is performed in the solid phase, and 4-chloronaphthol (4CN, **7**) is used to generate an insoluble compound (**8**) which absorbs visible light maximally around 550 nm. As the color develops on the solid support due to oxidase activity, digital imaging equipment is used to track the absorbance versus time for each pixel in the image. Delgrave and co-workers used this technology to isolate a galactose oxidase variant that showed a 16-fold increase in activity and a threefold lower K_m compared with the wild-type enzyme [18].

3.2.2.2 D-Amino Acid Oxidase

The flavoprotein D-amino acid oxidase (D-AO) catalyzes the deamination of D-amino acids to their corresponding α -imino acids. This reaction has drawn significant attention recently due to its potential to deaminate cephalosporin C in a sequential enzymatic route to form 7-amino cephalosporanic acid [22]. 7-Amino cephalosporanic acid is a basic building block used in the synthesis of many important semi-synthetic cephalosporin antibiotics. Moreover, D-AO has been used to resolve racemic amino acid mixtures and prepare α -keto acids. A high-throughput screening method based on the oxidation of the chromogen, *o*-dianisidine (**9**), has been used to detect the activity of D-AO and identify new microbial D-AO producers [20].

Peroxidase/*o*-Dianisidine Assay

The flavin cofactor FADH₂ is oxidized by molecular oxygen to form oxidized FAD and H₂O₂. The product H₂O₂ can then be coupled with HRP and a chromogen to form a colored compound. ABTS (**6**) was found to be unsuitable with D-AO due to problems such as autooxidation, instability, and a lack of sensitivity. Another chromogen *o*-dianisidine (**9**) has been successfully used as an alternative to ABTS (**6**) [20]. HRP reacts with *o*-dianisidine (**9**) and H₂O₂ to form a red-colored compound (**10**) under acidic conditions. The *o*-dianisidine reagent (**9**) has the added benefit of being a competitive inhibitor of catalase, an enzyme that competes *in vitro* with HRP for the substrate H₂O₂. This assay might be adaptable to screening other oxidases by simply changing to the particular oxidase's substrate.

3.2.2.3 Peroxidases

Peroxidases catalyze the oxidation of a wide variety of molecules utilizing H₂O₂ as the oxidant. Many peroxidases have an iron protoporphyrin IX prosthetic group in the active site, while others have a vanadium or, in the case of some bacterial peroxidases, no metal cofactor [23]. Peroxidases catalyze reactions that are potentially useful in both industrial and biotechnological applications. For example, HRP has been used as a reporter in diagnostic assays, biosensors, and

histochemical staining [24]. HRP has also been suggested as a catalyst for chemical synthesis as it catalyzes polymerization and dehydration of aromatic compounds, heteroatom oxidations, and epoxidation reactions [24, 25]. Furthermore, peroxidases have shown potential in bioremediation for removal of phenols and aromatic amines [26]. However, many peroxidases are not stable enough under the oxidative conditions created by high peroxide concentration and there are many substrates that peroxidases do not react with. Therefore, a number of activity assays have been developed that can be used in a high-throughput fashion for the discovery and engineering of new and improved peroxidases.

ABTS and *o*-Dianisidine Assays

As mentioned above in the galactose oxidase and D-AO sections, ABTS (**6**) and *o*-dianisidine (**9**) can both be used to assay peroxidases. ABTS (**6**) is oxidized by peroxidases in the presence of H₂O₂ to form a radical cation with intense green color (**7**) as shown in Figure 3.2, whereas the oxidation of *o*-dianisidine (**9**) results in the formation of a red dye (**10**). These types of assays are carried out similarly to those described in the oxidase section, except that hydrogen peroxide is now provided as a substrate.

ABTS screening has been used to find cytochrome *c* peroxidase mutants with altered specificity [27]. Using a combination of DNA shuffling and saturation mutagenesis, Iffland and coworkers were able to obtain variants of the cytochrome *c* peroxidase with a 70-fold increased specificity toward ABTS (**6**) compared with the natural substrate (cytochrome *c*) [27]. The ABTS assay has also been used to screen a library of HRP variants created by random mutagenesis [28], which resulted in a mutant that is more thermally stable, resistant to a variety of chemical and physical denaturants, and has a higher activity towards other organic substrates. The ABTS-based assay was also utilized to find mutants of HRP that can be functionally expressed in *E. coli* by directed evolution [29], to find mutants with 174-fold higher thermostability and 100-fold higher oxidative stability in *Coprinus cinereus* heme peroxidase by directed evolution [30], and to enhance the peroxidase activity of horse heart myoglobin by directed evolution [31]. The usefulness of the *o*-dianisidine (**9**) assay has been demonstrated in determining HRP mutants with expanded substrate activity [28]. The convenience of the high-throughput *o*-dianisidine (**9**) assay was further demonstrated when it was used to measure the total plasma peroxidase activity to assess the clinical severity of sickle cell anemia and citrulline therapy [32].

TMB Assay

Among the numerous colorimetric assays available for peroxidases, one of the most sensitive ones is based on the colorless compound 3,3',5,5'-tetramethylbenzidine (TMB) (**11**). Upon oxidation of this substrate by the peroxidase, TMB (**11**) forms a blue charge-transfer complex (**12**) initially, followed by a stable yellow final product (**13**) (Figure 3.3) [33].

A typical cell lysate screen is performed by monitoring the production of the blue charge-transfer product (**12**) at 650 nm on a microplate reader after mixing

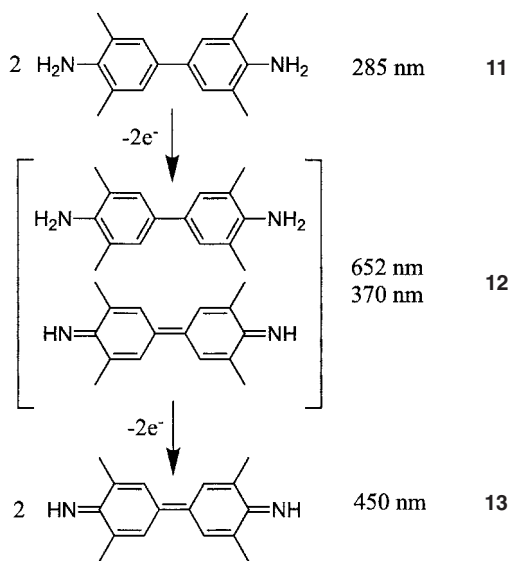


Fig. 3.3 Colorimetric assay for peroxidase activity based on 3,3',5,5'-tetramethylbenzidine (TMB).

a solution of the TMB (**11**) substrate, H_2O_2 , and the cell lysate. This assay is commonly used in HRP-based biosensors to determine the effects of mutations on peroxidases, and would be a reasonable assay for directed evolution of a peroxidase.

Guaiacol Assay

The guaiacol peroxidase assay is another colorimetric assay that is amenable to high-throughput screening. In this case, the colorless guaiacol (2-methoxyphenol, **14**) is oxidized to the phenoxy radical (**15**) in the presence of the peroxidase and H_2O_2 . The phenoxy radical subsequently undergoes polymerization to form tetraguaiacol (**16**), which is brown and can be measured at 470 nm. This assay can be performed in microplate format or in solid media format. Removal of background activity can be achieved by the addition of a small amount of ascorbate, which will scavenge the phenoxy radicals until all the ascorbate has reacted. The guaiacol assay has been successfully applied in identifying cytochrome *c* peroxidase mutants with altered substrate specificity from a library of mutants created by combinatorial mutagenesis [34]. Additionally, cytochrome *c* peroxidase mutants were created by directed evolution and screened for activity with this nonnatural substrate with a final mutant displaying a 1000-fold change in substrate preference [35].

MNBDH Assay

4-(*N*-Methylhydrazino)-7-nitro-2,1,3-benzoxadiazole (MNBDH, **17**) has been demonstrated to be a good substrate for peroxidase screening. The hydrazine portion of MNBDH (**17**) is oxidized by peroxidases in the presence of hydrogen per-

oxide to yield the strongly fluorescent nitroaniline (**18**) [36]. When measured with a fluorescent plate reader, this assay gives a very sensitive readout and is superior to chromogenic assays such as ABTS (**6**). However, application of this assay in directed evolution or discovery of peroxidases has yet to be demonstrated.

3.2.3

Oxygenases

There are several families of oxygenase enzymes whose members fall into one of two groups: those that introduce one oxygen atom into their substrate are generically referred to as monooxygenases, while those that introduce two oxygen atoms are referred to as dioxygenases. These enzymes typically catalyze the hydroxylation of nonactivated carbons or epoxidations of organic substrates and therefore have potential applications in synthetic chemistry, environmental remediation, toxicology, and gene therapy [37]. Examples of oxygenases are the omnipresent cytochrome P450 heme monooxygenases, di-iron enzymes such as methane and toluene monooxygenases, and Rieske nonheme iron dioxygenases such as naphthalene dioxygenase [38]. Molecular oxygen often serves as the oxidant, while reducing equivalents typically come from NAD(P)H-dependent reductase proteins via electron transfer.

Since reducing equivalents are transferred from a separate protein in most oxygenases (excluding those with a fused reductase domain), they typically function as large multimeric protein complexes. As such, oxygenases are often not stable enough or inadequately active towards the substrate of interest *in vitro*. Therefore, protein engineering of mono- and dioxygenases and the discovery of new oxygenases has received much interest in the hopes of discovering or evolving oxygenases for use in organic synthesis or biotechnology. Discovery of new or improved oxygenases from large random libraries requires an efficient and sensitive high-throughput screening method for desired enzyme function. Many such screening methods have been developed and used. In this section some of these methods and their applications will be discussed. Assays based on NADPH depletion have been discussed previously. However, these assays can yield misleading results with oxygenases, since uncoupling of the reductase and oxygenase activity can occur, especially for unnatural substrates.

3.2.3.1 Assays Based on Optical Properties of Substrates and Products

Many substrates of mono- and dioxygenases are aromatic and upon oxidation become colored or have a shifted absorbance spectrum. In other cases, the reaction products will autooxidize further to form colored products under the proper conditions. Although this is not applicable to all oxygenase products, it is the most convenient method of assaying oxygenase activity in a high-throughput manner.

2-Hydroxybiphenyl-2-monooxygenase (HpbA), in addition to its natural activity, catalyzes at low efficiency the regioselective *o*-hydroxylation of many different 2-substituted phenols to form the corresponding catechols. Random muta-

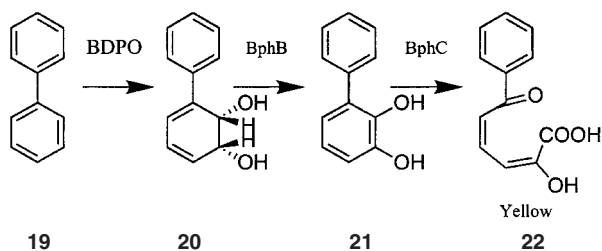


Fig. 3.4 Colorimetric assay for biphenyl dioxygenase (BDPO) activity. BphB=dihydrodiol dehydrogenase; BphC=2,3-dihydroxybiphenyl-1,2-dioxygenase.

genesis of HpbA was used to create mutants with enhanced monooxygenase activity with 2-substituted phenolic substrates [39]. Screening was accomplished by the formation of color when the reaction products autooxidize [39]. Using this method, mutants were discovered with greater than eightfold improvement in catalytic efficiency of hydroxylation of 2-methoxyphenol and a fivefold improvement in turnover rate with 2-*tert*-butylphenol [39].

Biphenyl (19) dioxygenase (BPDO) was screened in a somewhat different manner. BPDO is important industrially for the degradation of polychlorinated biphenyls (PCBs) and directed evolution was utilized to broaden its substrate specificity [40]. Screening was achieved by coexpressing dihydrodiol dehydrogenase (BphB) and 2,3-dihydroxybiphenyl-1,2-dioxygenase (BphC), which converted the BPDO reaction product (20) to a diol (21) and finally into a ring *meta* cleavage product (22) that was yellow in color (Figure 3.4). This method was first developed to screen a BPDO library created by DNA shuffling for enhanced PCB degradation capabilities [41] and later used to obtain dioxygenase variants with enhanced activity towards monocyclic aromatic hydrocarbons such as toluene and benzene [42]. In another example, hybrid genes of catechol-2,3-dioxygenase were screened for enhanced thermostability based on the color of the product [43, 44]. In this case, the substrate, catechol, is colorless, while the product, 2-hydroxy-3-oxo-1-phenylbut-3-en-1-ylideneacetic acid, is bright yellow. Using this method of screening, mutant dioxygenases with 26-fold enhancement in thermostability were isolated. Finally, mutants of toluene 4-monooxygenase were screened for enhanced activity on *o*-cresol and *o*-methoxyphenol based on the formation of the red-brown autooxidation products [45]. Mutants created by saturation mutagenesis were identified that had sevenfold and twofold higher catalytic rate with *o*-methoxyphenol and *o*-cresol as substrates, respectively [45].

3.2.3.2 Assays Based on Gibbs' Reagent and 4-Aminoantipyridine

4-Aminoantipyridine (4AAP, 23) and 2,6-dichloroquinone-4-chloroimide (Gibbs' reagent, 24) are very similar in the products they are able to detect. They both react strongly with ortho- and meta-substituted phenolic compounds to produce

strongly colored products (e.g. **25**). Additionally, they react well with *para*-substituted compounds if the substituent is a halide or alkoxy moiety to form colored products (e.g. **26**) [46]. Since many oxygenases produce such compounds, these are useful assays for measuring their activity. Additionally, dioxygenase *cis*-dihydrodiol products can be assayed by this method if the *cis*-dihydrodiol is first converted to a phenol by treatment with acid or *cis*-dihydrodiol dehydrogenase. Both of the reagents form adducts with the products, which can then be measured with a spectrophotometer. Both reagents are added at the end of the biotransformation (end-point assay) and the colored product develops in several minutes [47]. In the case of 4AAP (**23**) this occurs under basic conditions in the presence of potassium persulfate or similar oxidant.

Although 4AAP (**23**) has infrequently been used for high-throughput screening, Gibbs' reagent (**24**) has been applied in directed evolution experiments. The Gibbs' assay was used to isolate toluene dioxygenases that accept 4-methylpyridine as a substrate. After several rounds of mutagenesis and screening, a mutant in which the stop codon was replaced with a threonine was found to have sixfold higher activity with 4-methylpyridine and enhanced activity towards the natural substrate [48].

3.2.3.3 *para*-Nitrophenoxy Analog (pNA) Assay

The chemical methods available for linear chain hydrocarbon oxidation are energy intensive and hazardous to the environment. Several monooxygenases such as the P450s can catalyze alkane hydroxylation, but at slow rates. An ingenious assay involving the release of *para*-nitrophenolate from a *para*-nitrophenoxy analog (pNA, e.g. **25**) of a linear chain hydrocarbon has been developed and utilized. When the analogs are hydroxylated by the oxygenase, an unstable hemiacetal (**26**) is formed that dissociates into an aldehyde and *para*-nitrophenolate (**27**); the latter is bright yellow (Figure 3.5). The obvious disadvantage of this

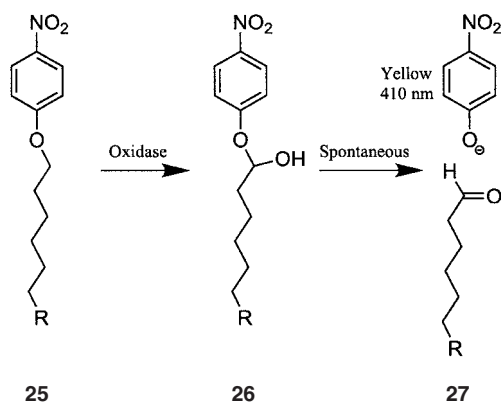


Fig. 3.5 Colorimetric assay for the activity of P450 monooxygenases.

assay is that the pNA (25) substrate is not the desired substrate and may result in the evolved enzyme only acting on pNA (25). However, when compared with other alternatives, this is the easiest assay for long-chain hydrocarbon hydroxylation by monooxygenases. The assay is accomplished by the addition of pNA (25) and NADPH (or use of a regenerative system) and then monitoring the increase in absorbance at 410 nm.

P450 monooxygenase BM-3 mutants were screened using an octane pNA (25). Typically this monooxygenase only has activity on longer chain hydrocarbons (at least 12 carbons), but after two rounds of mutagenesis and screening with the analog, the activity of this enzyme on octane was improved fivefold [49]. Another group used site-specific saturation mutagenesis and pNA screening to create mutants of the same enzyme with similarly enhanced activity toward smaller substrates [50]. Recently this P450 was optimized for activity in organic cosolvents by directed evolution using a pNA assay [51]. Activity of the P450 was increased 10-fold in the presence of 2% tetrahydrofuran (THF) and sixfold in the presence of 25% DMSO. Overall, this assay has proved to be very useful for screening monooxygenases that act on long-chain hydrocarbons.

3.2.3.4 Horseradish Peroxidase-coupled Assay

The use of HRP/ H_2O_2 to detect the reaction products of oxygenases has shown promise [52]. This assay can be used when the products of the oxygenase reaction on aromatic substrates, e.g. naphthalene (28), are or can be converted into hydroxylated aromatic compounds. These hydroxylated aromatic compounds (e.g. naphthol, 29) are then converted by oxidation into colored or fluorescent compounds (e.g. 30 in the presence of HRP and H_2O_2) (Figure 3.6). Mono- and dihydroxylated aromatic compounds form dimers (e.g. 30) and polymers in this assay and, depending on their structure, they will display different colorimetric and fluorescent properties, allowing determination of the desired product. This assay can be carried out in cell strains expressing HRP either on solid phase or in liquid assay by adding H_2O_2 to the solid media or cell culture. The positive mutants will then display an enhanced fluorescent or colorimetric signal.

This assay has been suggested for use in screening toluene dioxygenase in combination with *cis*-dihydrodiol dehydrogenase [52]. Additionally, this assay was used to screen for improved activity of P450cam on naphthalene (28) using H_2O_2 as the oxidant. Mutants were found with a 20-fold improvement in activity with differences in regioselectivity of hydroxylation in some cases [53].

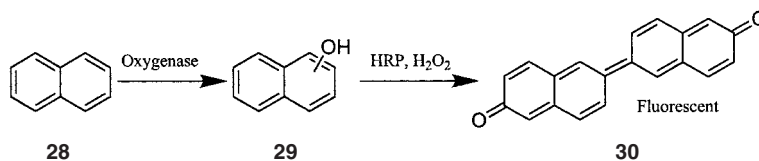


Fig. 3.6 Horseradish peroxidase (HRP)-based fluorimetric assay for oxygenase activity.

3.2.3.5 Indole Assay

Indole (**31**) can be oxidized by mono- and dioxygenases to various 2 and 3 position hydroxyl and epoxide indoles. Upon exposure to air, the generated compounds further oxidize and dimerize to form indigo (**32**) and indirubin (**33**) (Figure 3.7). Both of the formed compounds are intensely colored and are easily monitored colorimetrically.

This assay, coupled to site-specific saturation mutagenesis, was used to enhance the activity of P450 BM-3 on indole (**31**) from a level that was too low to detect up to a turnover rate of 2.7 s^{-1} [54]. This assay was recently used to broaden the substrate specificity of toluene dioxygenase such that it would accept *p*-xylene as a substrate. One round of mutagenesis and screening resulted in a 4.3-fold activity improvement on *p*-xylene as well as enhancements with several other unnatural substrates [55]. This method of assaying oxygenases is well suited for removing inactive clones from a library or for evolution of altered substrate specificity.

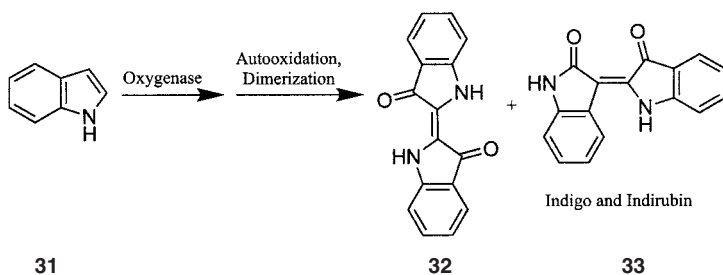


Fig. 3.7 Indole-based assay for oxygenase activity.

3.2.4

Laccases

Laccases (EC 1.10.3.2) belong to the blue copper oxidase family and are multi-copper enzymes capable of oxidizing phenols, polyphenols, substituted phenols, diamines, anilines, and other similar compounds. Electrons are removed one at a time from the substrate by a type 1 blue copper ion and transferred to a trinuclear copper cluster and molecular oxygen is utilized as the electron acceptor [56]. The product radical can undergo further oxidation catalyzed by the laccase or it may undergo a nonenzymatic reaction such as hydration or polymerization [56]. Laccases are common enzymes in nature, but are primarily produced by fungi and plants.

Laccases are currently receiving a lot of attention as industrial enzymes with proposed applications including textile dye bleaching, pulp bleaching, effluent detoxification, and bioremediation [57]. However, a common hurdle that must be overcome with these enzymes is their low expression level in native hosts. Additionally, laccases can catalyze the oxidation of polycyclic hydrocarbons (very

toxic pollutants) in the presence of primary substrates such as 1-hydroxy benzothiazole (HBT) [58, 59]. These primary substrates are often toxic, expensive, and result in side-reactions [60]. Therefore, high-throughput screening applied to the discovery and/or directed evolution of laccases is critical to overcome these hurdles. Several screening methods have been developed and are described below.

3.2.4.1 ABTS Assay

The ABTS (6) assay is a very flexible assay applicable to many oxidases, peroxidases as well as laccases. Again the production of the green radical cation (7) is observed spectrophotometrically. This assay has been utilized to improve the functional expression of a fungal laccase in *Saccharomyces cerevisiae* [61]. Directed evolution was utilized to improve protein expression by eightfold to the highest level yet reported with an additional increase of 22-fold in the turnover number. The ABTS assay has also been used to screen for functional expression of another laccase from *Trametes versicolor* into the heterologous host *Yarrowia lipolytica* [62]. This assay is well suited for development of functional selection and enhanced stability screening, and should prove useful in the future.

3.2.4.2 Poly R-478 Assay

Poly R-478 is a polymeric dye that can act as a surrogate substrate for lignin degradation by laccases in the presence of a mediator such as HBT. Upon oxidation by laccases the dye is decolorized, making it easy to follow the activity. The activity of laccases on this substrate has been correlated to the oxidation of polycyclic aromatic hydrocarbons (PAHs) [63]. The assay is performed simply by mixing the water-soluble polymeric dye with the library members to be screened and measuring the decrease in absorbance at 520 nm. Notably, this assay has been utilized recently to discover novel laccases from cultivable fungal strains [64] and to find PAH-degrading strains [63].

3.2.4.3 Other Assays

Several other assays have been described for laccases that have not yet seen significant application as high-throughput screening methods [65, 66]. In one such assay, anthracene is converted to 9,10-anthroquinone by the laccase enzyme, followed by subsequent chemical reduction to 9,10-anthrahydroquinone, which is orange in color and can be measured at 419 nm. Another assay is the production of triiodide in acidic conditions from sodium iodide [66]. In this assay the activity of the laccase can be monitored at 353 nm as the formed triiodide is yellow. Both of these assays require the use of mediators like HBT or ABTS (6) and might be useful in the directed evolution of laccases that have the desired activity in their absence by screening with progressively less mediator.

3.3

Conclusions

High-throughput screening methods for oxidoreductases represent a key step in identifying variants with improved catalytic function by directed evolution as well as discovering new oxidoreductases by expression of environmental DNA. In this chapter, we have discussed many effective high-throughput screening methods for various oxidoreductases and their successful applications in discovering and engineering new or improved oxidoreductases. The resulting oxidoreductases are not only important for many potential applications such as biocatalysis, bioremediation, diagnostics, and gene therapy, but also are critical to increase our understanding of cellular metabolism and protein structure/function relationships. No doubt in the years to come, these high-throughput screening methods will be used increasingly to isolate oxidoreductases with new or improved characteristics.

Acknowledgements

We thank Biotechnology Research and Development Consortium (BRDC) and Office of Naval Research for supporting our work on the characterization and engineering of oxidoreductases including a phosphite dehydrogenase and arylamine oxidases.

Abbreviations

4AAP	4-aminoantipyrine
4CN	4-chloronaphthol
ABTS	2,2-azino-bis(3-ethylbenzthiazoline)-6-sulfonic acid
ADH	alcohol dehydrogenase
BPDO	biphenyl dioxygenase
D-AO	D-amino oxidase
DH	dehydrogenase
HpbA	2-Hydroxybiphenyl 2-monoxygenase
HRP	horseradish peroxidase
MNBBDH	4-(N-methylhydrazino)-7-nitro-2,1,3-benzoxadiazole
NBT	nitroblue tetrazolium
NAD(H)	nicotinamide adenine dinucleotide (reduced form)
NADP(H)	nicotinamide adenine dinucleotide phosphate (reduced form)
PCR	polymerase chain reaction
PMS	phenazine methosulfate
PQQ	pyrroloquinoline quinone
pNA	<i>para</i> -nitrophenoxy analog
TMB	3,3',5,5'-tetramethylbenzidine

References

- 1 K. Faber, *Biotransformations in Organic Chemistry: a Textbook*, 4th edn, Springer, Berlin, **2000**.
- 2 A. Schmid, J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts, B. Witholt, *Nature* **2001**, *409*, 258–268.
- 3 S. K. Powell, M. A. Kaloss, A. Pinkstaff, R. McKee, I. Burimski, M. Pensiero, E. Otto, W. P. Stemmer, N. W. Soong, *Nat. Biotechnol.* **2000**, *18*, 1279–1282.
- 4 H. Zhao, W. A. van der Donk, *Curr. Opin. Biotechnol.* **2003**, *14*, 583–589.
- 5 W. A. van der Donk, H. Zhao, *Curr. Opin. Biotechnol.* **2003**, *14*, 421–426.
- 6 J. Handelsman, M. R. Rondon, S. F. Brady, J. Clardy, R. M. Goodman, *Chem. Biol.* **1998**, *5*, R245–249.
- 7 C. Schmidt-Dannert, *Biochemistry* **2001**, *40*, 13125–13136.
- 8 D. C. Demirjian, P. C. Shah, F. Moris-Varas, *Top. Curr. Chem.* **1999**, *200*, 1–29.
- 9 J. Fibla, R. Gonzalezduarte, *J. Biochem. Biophys. Methods* **1993**, *26*, 87–93.
- 10 K. M. Mayer, F. H. Arnold, *J. Biomol. Screening* **2002**, *7*, 135–140.
- 11 S. J. Allen, J. J. Holbrook, *Protein Eng.* **2000**, *13*, 5–7.
- 12 A. S. ElHawrani, R. B. Sessions, K. M. Moreton, J. J. Holbrook, *J. Mol. Biol.* **1996**, *264*, 97–110.
- 13 G. Ravot, D. Wahler, O. Favre-Bulle, V. Cilia, F. Lefevre, *Adv. Synth. Catal.* **2003**, *345*, 691–694.
- 14 T. Conway, G. W. Sewell, Y. A. Osman, L. O. Ingram, *J. Bacteriol.* **1987**, *169*, 2591–2597.
- 15 N. O. Kaplan, S. P. Colowick, C. C. Barnes, *J. Biol. Chem.* **1951**, *191*, 461–472.
- 16 G. E. Tsotsou, A. E. G. Cass, G. Gilardi, *Biosens. Bioelectron.* **2002**, *17*, 119–131.
- 17 L. H. Sun, I. P. Petrounia, M. Yagasaki, G. Bandara, F. H. Arnold, *Protein Eng.* **2001**, *14*, 699–704.
- 18 S. Delagrave, D. J. Murphy, J. L. R. Pruss, A. M. Maffia, B. L. Marrs, E. J. Bylina, W. J. Coleman, C. L. Grek, M. R. Dilworth, M. M. Yang, D. C. Youvan, *Protein Eng.* **2001**, *14*, 261–267.
- 19 G. Y. Yang, A. M. Shamsuddin, *Histol. Histopathol.* **1996**, *11*, 801–806.
- 20 M. Gabler, M. Hensel, L. Fischer, *Enzyme Microb. Technol.* **2000**, *27*, 605–611.
- 21 H. J. Danneel, M. Ullrich, F. Giffhorn, *Enzyme Microb. Technol.* **1992**, *14*, 898–903.
- 22 H. D. Conlon, J. Baqai, K. Baker, Y. Q. Shen, B. L. Wong, R. Noiles, C. W. Rausch, *Biotechnol. Bioeng.* **1995**, *46*, 510–513.
- 23 B. Valderrama, M. Ayala, R. Vazquez-Duhalt, *Chem. Biol.* **2002**, *9*, 555–565.
- 24 N. C. Veitch, A. T. Smith, *Adv. Inorg. Chem.* **2001**, *51*, 107–162.
- 25 S. Colonna, N. Gaggero, C. Richelmi, P. Pasta, *Trends Biotechnol.* **1999**, *17*, 163–168.
- 26 J. A. Nicell, J. K. Bewtra, N. Biswas, C. C. Stpierre, K. E. Taylor, *Can. J. Civil Eng.* **1993**, *20*, 725–735.
- 27 A. Iffland, S. Gendreizig, P. Tafelmeyer, K. Johnsson, *Biochem. Biophys. Res. Commun.* **2001**, *286*, 126–132.
- 28 B. Morawski, S. Quan, F. H. Arnold, *Biotechnol. Bioeng.* **2001**, *76*, 99–107.
- 29 Z. Lin, T. Thorsen, F. H. Arnold, *Biotechnol. Prog.* **1999**, *15*, 467–471.
- 30 J. R. Cherry, M. H. Lamsa, P. Schneider, J. Vind, A. Svendsen, A. Jones, A. H. Pedersen, *Nat. Biotechnol.* **1999**, *17*, 379–384.
- 31 L. Wan, M. B. Twitchett, L. D. Eltis, A. G. Mauk, M. Smith, *Proc. Natl Acad. Sci. USA* **1998**, *95*, 12825–12831.
- 32 W. H. Waugh, *J. Pediatr. Hematol. Oncol.* **2003**, *25*, 831–834.
- 33 P. D. Josephy, T. Eling, R. P. Mason, *J. Biol. Chem.* **1982**, *257*, 3669–3675.
- 34 M. Wilming, A. Iffland, P. Tafelmeyer, C. Arrivoli, C. Saudan, K. Johnsson, *Chembiochem* **2002**, *3*, 1097–1104.
- 35 A. Iffland, P. Tafelmeyer, C. Saudan, K. Johnsson, *Biochemistry* **2000**, *39*, 10790–10798.
- 36 J. Meyer, A. Buldt, M. Vogel, U. Karst, *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 1453–1455.
- 37 D. R. Nelson, T. Kamataki, D. J. Waxman, F. P. Guengerich, R. W. Estabrook, R. Feyereisen, F. J. Gonzalez, M. J. Coon, I. C. Gunsalus, O. Gotoh, et al. *DNA Cell Biol.* **1993**, *12*, 1–51.

- 38 C. S. Butler, J. R. Mason, *Adv. Microb. Physiol.* **1997**, *38*, 47–84.
- 39 A. Meyer, A. Schmid, M. Held, A. H. Westphal, M. Rothlisberger, H. P. Kohler, W. J. van Berkel, B. J. Witholt, *Biol. Chem.* **2002**, *277*, 5575–5582.
- 40 H. Suenaga, M. Goto, K. Furukawa, *J. Biol. Chem.* **2001**, *276*, 22500–22506.
- 41 T. Kumamaru, H. Suenaga, M. Mitsuo-ka, T. Watanabe, K. Furukawa, *Nat. Biotechnol.* **1998**, *16*, 663–666.
- 42 H. Suenaga, M. Mitsuo-ka, Y. Ura, T. Watanabe, K. Furukawa, *J. Bacteriol.* **2001**, *183*, 5441–5444.
- 43 A. Okuta, K. Ohnishi, S. Harayama, *Gene* **1998**, *212*, 221–228.
- 44 M. Kikuchi, K. Ohnishi, S. Harayama, *Gene* **1999**, *236*, 159–167.
- 45 Y. Tao, A. Fishman, W. E. Bentley, T. K. Wood, *J. Bacteriol.* **2004**, *186*, 4705–4713.
- 46 C. R. Otey, J. M. Joern, *Methods Mol. Biol.* **2003**, *230*, 141–148.
- 47 J. M. Joern, T. Sakamoto, A. Arisawa, F. H. Arnold, *J. Biomol. Screen.* **2001**, *6*, 219–223.
- 48 T. Sakamoto, J. M. Joern, A. Arisawa, F. H. Arnold, *Appl. Environ. Microbiol.* **2001**, *67*, 3882–3887.
- 49 E. T. Farinas, U. Schwaneberg, A. Glieder, F. H. Arnold, *Adv. Synth. Catal.* **2001**, *343*, 601–606.
- 50 Q. S. Li, U. Schwaneberg, M. Fischer, J. Schmitt, J. Pleiss, S. Lutz-Wahl, R. D. Schmid, *Biochim. Biophys. Acta* **2001**, *1545*, 114–121.
- 51 T. Seng Wong, F. H. Arnold, U. Schwaneberg, *Biotechnol. Bioeng.* **2004**, *85*, 351–358.
- 52 H. Joo, A. Arisawa, Z. Lin, F. H. Arnold, *Chem. Biol.* **1999**, *6*, 699–706.
- 53 H. Joo, Z. Lin, F. H. Arnold, *Nature* **1999**, *399*, 670–673.
- 54 Q. S. Li, U. Schwaneberg, P. Fischer, R. D. Schmid, *Chemistry* **2000**, *6*, 1531–1536.
- 55 L. M. Newman, H. Garcia, T. Hudlicky, S. A. Selifonov, *Tetrahedron* **2004**, *60*, 729–734.
- 56 C. F. Thurston, *Microbiol-Uk* **1994**, *140*, 19–26.
- 57 N. Duran, M. A. Rosa, A. D'Annibale, L. Gianfreda, *Enzyme Microb. Technol.* **2002**, *31*, 907–931.
- 58 C. Johannes, A. Majcherczyk, A. Huttermann, *Appl. Microbiol. Biotechnol.* **1996**, *46*, 313–317.
- 59 M. A. Pickard, R. Roman, R. Tinoco, R. Vazquez-Duhalt, *Appl. Environ. Microbiol.* **1999**, *65*, 3805–3809.
- 60 M. Alcalde, T. Bulter, *Methods Mol. Biol.* **2003**, *230*, 193–201.
- 61 T. Bulter, M. Alcalde, V. Sieber, P. Meinhold, C. Schlachtbauer, F. H. Arnold, *Appl. Environ. Microbiol.* **2003**, *69*, 987–995.
- 62 C. Jolivalt, C. Madzak, A. Brault, E. Caminade, C. Malosse, C. Mougin, *Appl. Microbiol. Biotechnol.* **2005**, *66*, 450–456.
- 63 J. A. Field, E. de Jong, G. Feijoo Costa, J. de Bont, A. *Appl. Environ. Microbiol.* **1992**, *58*, 2219–2226.
- 64 L. L. Kiiskinen, M. Ratto, K. Kruus, *J. Appl. Microbiol.* **2004**, *97*, 640–646.
- 65 M. Alcalde, T. Bulter, F. H. Arnold, *J. Biomol. Screen.* **2002**, *7*, 547–553.
- 66 F. Xu, *Appl. Microbiol. Biotechnol.* **1996**, *59*, 221–230.

4

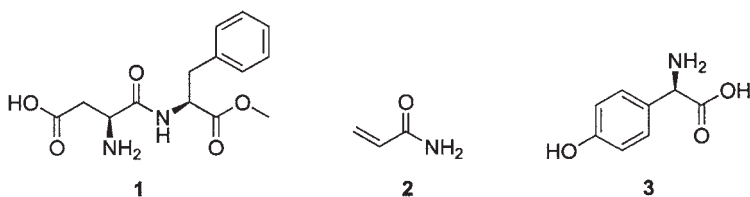
Industrial Perspectives on Assays

Theo Sonke, Lucien Duchateau, Dick Schipper, Gert-Jan Everink, Sjoerd van der Wal, Huub Henderickx, Roland Bezemer, and Aad Vollebregt

4.1

Introduction

Although microbial cells and enzymes have been used by humans for millennia to produce a variety of foodstuffs, it was not until the twentieth century that biocatalysis was introduced in the chemical industry. Since that time it has grown to become an established technology for chemical manufacturing. Examples of biocatalytic processes currently successfully operated at industrial scale are the production of the low-calorie sweetener aspartame (1) catalyzed by the protease thermolysin [1, 2], the production of the bulk chemical acrylamide (2) using a nitrile hydratase [3, 4], and the production of *D-p*-hydroxyphenylglycine (3) using microbial cells containing a hydantoinase and a carbamoylase [5, 6].



The recent need for more sustainable processes has driven the implementation of biocatalysis in industry further, in many cases by replacing existing chemical processes. An example of such second-generation “green processes” is the chemo-enzymatic synthesis of the semi-synthetic β -lactam antibiotic cephalixin (6). In this new process concept the side-chain *D*-phenylglycine (4a), activated as amide (4b) or ester (4c), is coupled to the β -lactam nucleus 7-amino-desacetoxycephalosporanic acid (7-ADCA, 5) by a penicillin G acylase (Figure 4.1) [7]. This process fully exploits the advantages of a biocatalyst in an industrial setting. The coupling reaction of activated side-chain and nucleus is done in water at ambient temperature, thereby eliminating the use of halogenated

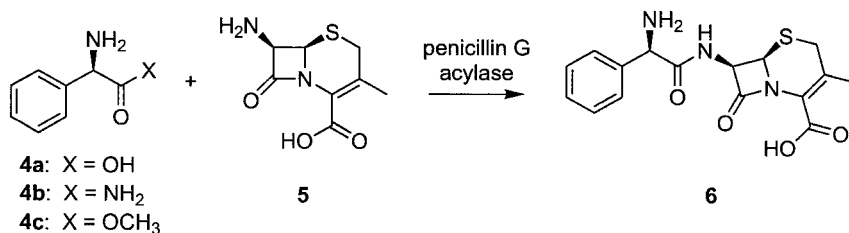


Fig. 4.1 Penicillin G acylase-catalyzed synthesis of the semi-synthetic β -lactam antibiotic cephalosporin (6).

solvents and low temperatures. The strict selectivity of the biocatalyst contributes to the high efficiency and increased sustainability of this chemo-enzymatic process making protection of the two building blocks and deprotection of the coupling product superfluous [8, 9]. Altogether, over 140 processes using enzymes in isolated form or microbial cells have been commercialized up to now, most of which deliver small-scale chiral products for applications in the pharmaceutical sector [10, 11].

Chemical custom manufacturing (CCM) is the industry segment in which the fine chemicals industry produces a compound on demand for a client (usually a pharmaceutical company) in a one-to-one relation (as opposed to multi-client products). In this highly dynamic business area, speed and responsiveness are among the most important key success factors. General timelines for CCM are given in Table 4.1. It can be seen that proof of principle including a sample often needs to be delivered to the customer within just three months.

In CCM the use of biocatalysis for the production of (chiral) fine chemicals is still limited compared with, for instance, resolution via diastereomer crystallization or chromatography, asymmetric organometal catalysis, or conventional

Table 4.1 Typical timeline for process development within chemical custom manufacturing.

Time	Status	Remark
0	Pharmaceutical customer request	Chemical route exists but usually high costs Quest for alternatives begins
2 weeks	Selection of 2–3 routes	Based on rough cost price estimates Plan and quote to customer After customer's positive response, experimental work begins
2 months	Feasibility demonstrated	Final choice made More accurate cost price calculated
3 months	Lab sample delivered for approval by customer	Based on specs or use tests

chemistry starting from the chiral pool [12, 13]. The main reason for this difference is the long lead time to find and implement a biocatalytic process compared with a chemical process, despite the fact that biocatalysts in general show excellent enantio-, regio- and stereoselectivity, which, in principle, often enables the design of shorter and more cost-efficient processes.

In order to be able to implement a biocatalytic reaction step successfully in such a product, extremely strict timelines to (bio)catalyst identification and subsequent process development need to be met. Although the three months' timeline and approach schematized in Table 4.1 is an average scenario and deviations frequently occur, it clearly demonstrates the challenge for (bio)catalysis in CCM: the time available for (bio)catalyst identification and (small-scale) process development is extremely short.

In this chapter we will deal with the development of enzymatic assays and activity-based screenings from an industrial perspective. We will focus solely on CCM, and try to describe how this specialized branch of the chemical industry must cope with the demanding timelines for biocatalyst identification. In the first part, a number of prerequisites for the ideal screening will be discussed, including one of the most important ones: integral screening time. In the second part of this chapter, these general principles will be exemplified by a number of assay/screening methods developed for DSM's chemical custom manufacturing business unit DSM Pharma Chemicals (DPC).

4.2

Prerequisites for an Effective Biocatalyst Screening in Chemical Custom Manufacturing

Identification of a biocatalyst for a certain chemical conversion requires input of sufficiently high-quality biodiversity to start from. In the last decade tremendous progress has been made in the generation of such biodiversity. Construction of expression libraries directly from environmental DNA is applied in many laboratories as a source of diversity. Use of this so-called metagenome now enables identification of interesting biocatalysts from microorganisms that could not be cultivated in laboratories before, increasing the biodiversity dramatically [14]. The rapidly increasing quantity of genome sequence information, together with improved software tools to predict the function of the encoded proteins, offers another source for new enzymes. This genome sequence information can be transferred into ready-to-screen enzymatic platforms.

Directed evolution is another relatively recent approach to the creation of biodiversity. Numerous formats ranging from simple error-prone polymerase chain reaction (epPCR) [15, 16] of a single gene to shuffling of large families of genes [17, 18] and even whole genomes [19, 20] have been published in the last 10 years to access the desired sequence space. These molecular biology tools have become quite robust, which makes the generation of sufficiently high-quality biodiversity no longer a limiting factor in many biocatalyst identification pro-

grams. This is also true in CCM if the inputs for the screening (for example expression libraries and enzymatic platforms) have been generated proactively to meet the short timelines in this industrial segment.

Screening for biocatalysts has also attracted considerable attention from academia and industry in the last decade, undeniably resulting in significant progress in this field [21, 22]. Nevertheless, the development and implementation of a new screening within the limited time available in CCM is still a significant challenge. The following criteria are pivotal for successful screening in this demanding field.

Screening for Real Conversion

One of the most commonly used statements in screening is “You get what you screen for”. In the open and patent literature, numerous artificial substrates have been described for colorimetric or fluorometric detection of a wide range of enzymatic activities, but use of this type of substrate bears the inherent risk that a biocatalyst identified does not catalyze the desired reaction. This can be caused by steric incompatibility (artificial substrates are often larger than the actual ones) or catalytic incompatibility (certain types of artificial substrates are more active than the actual substrates). Although improved artificial substrates have recently been reported [23], screening for the real conversions still results in the most reliable hits. Furthermore, synthesis of the artificial substrate can be difficult and time consuming, which will delay the outcome of the screening, thereby making it even harder to identify the desired biocatalyst on time.

Robust and Reliable

The ideal screening is robust and reliable. It must not suffer from matrix effects, such as small differences in pH, salt concentration, and residual growth medium. Furthermore, interexperimental differences should be small. We noticed that the robustness of screening protocols sharply decreases with increasing numbers of handling (e.g. centrifugation, pipetting, and filtration). Therefore, simple screening protocols are preferred whenever possible. Altogether, the screening method must clearly discriminate true positive clones without delivering a large number of false positives.

Quantitative

To be able to discriminate positive clones the applied screening protocol must result in a clear change of measured response. However, different types of projects require different levels of accuracy. Screening a (metagenomic) library for new biocatalysts, for example, places less strict requirements on accuracy than

screening in a directed evolution program to improve the catalytic performance of an enzyme. In the first example, the scenario is more or less black and white – the few positive clones that are expected should give a signal that is sufficiently different from the background signal. The exact intensity of the signal of the positive clones is not very useful, however, since clones containing the same gene will often have a different genetic constitution that can severely influence the expression of the genes on the insert fragment. Therefore, an assay that delivers semi-quantitative results is sufficient for this type of screening. In the case of a directed evolution project for improved enzymes the situation is different. In this case, the enzymatic reaction is tuned so that the conversion with the wild-type clones amounts to about 10% conversion. Consequently, these wild-type clones will already give a (low) signal, requiring more quantitative screening methods. The fact that all clones to be screened contain the (mutated) gene(s) of interest in an identical genetic constitution is another reason why quantitative screenings are advantageous in this case. As the development of highly quantitative screenings can be time consuming, the accuracy aimed for should be tuned to the needs of the specific application.

Sensitive

As the most reliable results are obtained by mimicking industrial conditions in the screening reaction as far as possible, substrate concentrations are usually high (> 50 mM). Consequently, the sensitivity of the analytical method is typically not an issue. Where the substrates or products are poorly soluble in water the situation is different, and a sensitive analytical method is required. Using mass spectrometry (MS) positive clones can be identified starting from substrate concentrations as low as 0.5 mM. Addition of a cosolvent enabling higher substrate concentrations is another possibility, assuming that it is not interfering with the screening.

Integral Screening Time

This is the most important factor for screenings in CCM. It can be defined as the time needed from the start of the screening development to the generation of the first screening results. As will have become clear from the introductory section, only screening methods with a short integral screening time are compliant with CCM timelines. At first glance, ultra-high-throughput screening (UHTS, in this chapter defined as a screening with a throughput of > 10⁴ samples a day) formats look suitable for application within CCM as they are characterized by an extremely short analysis time per sample. However, this is certainly not true in all cases, especially when the development of UHTS methods is time consuming (Figure 4.2).

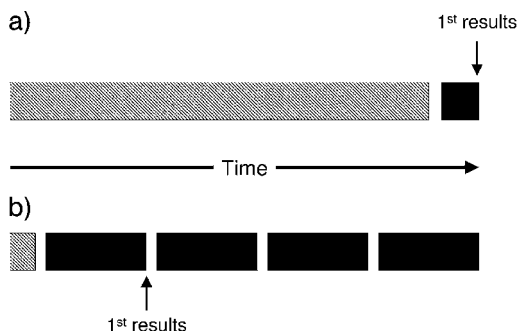


Fig. 4.2 Graphical representation of the effect of a short integral screening time on the time 1st results are generated. (a) An ultra-high-throughput screening with a long development time (typical for reader assays without colorimetric or fluorometric model substrates). (b) A generic assay that combines

a much shorter development time with lower throughput in the actual assay (typical for instrumental assays like flow-injection NMR). The gray boxes represent the development time for the assay, and the black boxes the time needed for the actual measurements of all samples in a screening.

Reader assays (both UV/Vis and fluorescence) are appropriate, in principle, for use in UHTS. A 96-well microtiter plate can be measured in less than 30 s, leading to a theoretical throughput of 10^4 – 10^5 samples a day in the case of an end-point measurement. Unfortunately, this throughput can only be achieved with simple screening protocols, which are only possible when the conversion of substrate into product leads to strong changes in molar extinction coefficient or fluorescence. As this is quite often not the case with the actual substrates and/or products and use of artificial substrates is undesirable, derivatization of substrates and/or products into highly absorbing or fluorescent compounds is necessary as an additional step. These derivatization reactions not only lead to much more complicated screening protocols and consequently longer screening times (less throughput), but also to much longer development times as they require careful tuning of the derivatization conditions (e.g. reagent concentration, temperature, pH, and reaction time) and solutions for matrix effects. Overall, this will result in a relative long integral screening time (Figure 4.2a), which makes such a screening method less attractive for applications in CCM.

Figure 4.2b shows a situation that is better suited for use in CCM. In this case, implementation of a screening method with very short development time and sufficient throughput (not necessarily ultra high throughput) generates the first screening results in a shorter period of time, leading to quicker hit identification and start of application research as well as sample preparation for the customer. The characteristics needed (short development time and sufficient throughput) are offered by generic assays, as for example based on nuclear magnetic resonance (NMR) or MS. These types of assays are characterized by the fact that they generate distinct signals for many chemical functionalities enabling straightforward differentiation between substrate and product of the enzymatic reaction to be screened. Combined with their relatively good tolerance

to fluctuations in the matrix, these instrumental assay techniques are quite powerful.

A different approach is the proactive development of less generic assays for frequently requested chemical group conversions, as for example the synthesis of amino acids or for reactions producing ammonia as a side-product. In this case the challenge is to find reagents that are specific for these molecules and generate a clearly distinguishable signal in the derivatization reaction. Furthermore, this derivatization reaction must not be interfered by other chemical groups in these molecules, which will be different in each customer's request.

In the next sections of this chapter examples will be discussed from each of these CCM compliant screening approaches.

4.3

CCM Compliant Screening Methods Based on Optical Spectroscopy (UV/Vis and Fluorescence)

As stated in the previous section, reader assays are characterized by a very short measuring time per sample. However, often this does not translate into a short integral screening time as the product formed in the screening reaction lacks a chromophore or fluorophore and consequently needs to be chemically derivatized before it can be detected in a reader with sufficient sensitivity and selectivity. In other cases the product of the screening reaction can be detected in a reader by a coupled enzymatic assay. In such an assay an enzymatic reaction cascade leads to the formation of a reader detectable compound (for instance NAD^+/NADH). In both approaches, the overall screening protocols are quite laborious and their development is often time consuming.

Nevertheless, reader assays can be successfully applied within the CCM context, for instance when the formation of the (side)-products results in a strong change in molar extinction coefficient (Section 4.3.1), or when the (side)-products formed belong to a category of compounds for which a derivatization method or a coupled enzymatic assay has been developed beforehand (Section 4.3.2).

4.3.1

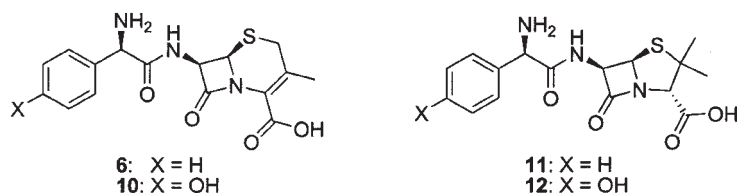
Optical Spectroscopic Methods Based on the Spectral Properties of the Product Itself

The possibility of implementing screening methods based solely on changes in spectral properties upon conversion of substrate into product is small. Nevertheless, the great potential of this method justifies the efforts that are needed to investigate whether it is possible. Screening requests for transformations of molecules that contain a suitable chromophoric or fluorophoric group can best be started by recording the optical (UV/Vis/fluorescence) spectra of substrate and product. In this way it can be established whether the product absorbs and/or

emits light of a specific wavelength. By including the appropriate blanks in these first experiments, interference of other compounds in the matrix (e.g. the reaction buffer and cell-derived compounds) can be assessed. If these experiments have a positive outcome, a simple screening procedure can be developed in a few days.

4.3.1.1 Example: Isolation of the D-*p*-Hydroxyphenylglycine Aminotransferase Gene [24]

One of the process concepts for the production of enantiomerically pure α -H- α -amino acids is the conversion of α -keto acids to the desired amino acids by aminotransferases [25, 26]. For the production of the D-enantiomers of α -H- α -amino acids, the “standard” aminotransferases require equimolar amounts of D-Ala, D-Glu, or D-Asp as amino donor, which makes this process less attractive as these compounds are rather expensive. A more attractive aminotransferase process to enantiomerically pure D- α -H- α -amino acids is possible with an enzyme that can use cheaply available L-Glu (8) and/or L-Asp as amino donor instead of their D-congeners. Such a stereoinverting aminotransferase had been described before in relation to the production of D-*p*-hydroxyphenylglycine (3) and D-phenylglycine (4a, Figure 4.1) [27, 28]. These amino acids are produced as side-chains in the manufacture of semi-synthetic β -lactam antibiotics such as cephalexin (6), cephadroxy (10), ampicillin (11), and amoxycillin (12) [9, 29]. The reaction catalyzed by this aminotransferase (HpgAT) is shown in Figure 4.3.



We decided to identify the HpgAT gene by screening bacterial expression libraries including that of *Pseudomonas putida* NCIMB 12565, which had been reported to contain HpgAT activity [27]. In this case neither a PCR-based approach nor the synthesis of the complete gene were possible, as the sequence of the HpgAT gene had not been published. Recording of the optical spectra of D-*p*-hydroxyphenylglycine (3) and *p*-hydroxyphenylglyoxylate (9) showed that the latter compound has a specific absorbance at about 300–350 nm (Figure 4.4).

Based on this observation, we designed the following simple screening strategy.

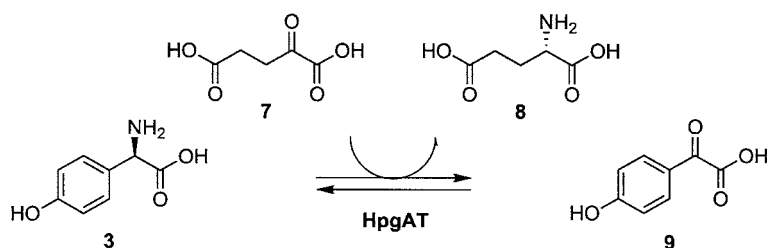


Fig. 4.3 Reaction catalyzed by the stereoinverting D-*p*-hydroxyphenylglycine aminotransferase (HpgAT) from *P. putida* NCIMB 12565.

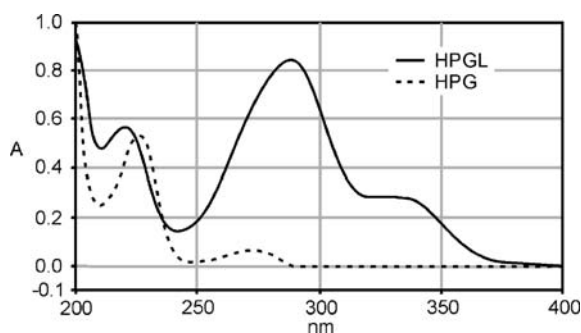


Fig. 4.4 Optical spectra of *p*-hydroxyphenylglycine (HPG, dashed line) and *p*-hydroxyphenylglyoxylate (HPGL, black line) in 45 mM potassium phosphate buffer, pH 7.0. Concentration of both compounds was 55 μ M.

Procedure 4.1: UV Screening of Aminotransferase Reactions

Colonies from a gene library were cultivated in microtiter plates containing 150 μ L of a rich medium containing the appropriate antibiotic. After 16–20 h at 28 $^{\circ}$ C, cells were harvested by centrifugation, washed once with 50 mM potassium phosphate buffer, pH 7.0, and resuspended in 180 μ L reaction mix (100 mM potassium phosphate, pH 7.0, 15 mM α -ketoglutarate (7), 0.1 mM pyridoxal-phosphate (PLP) and 0.5% v/v Triton X-100). The screening reaction was started by addition of D-*p*-hydroxyphenylglycine (3) to a final concentration of 5 mM, after which the $A_{340\text{nm}}$ in each well was monitored for 20 min using an Optimax microtiter plate reader (Molecular Devices, Sunnyvale, CA, USA).

As expected, clones containing the desired HpgAT gene showed significant increase in $A_{340\text{nm}}$ relative to negative wells, as can be seen in Figure 4.5. Integral screening time in this specific case was less than one month due to the very short development time and modest analysis time (20 minutes per 96-well microtiter plate).

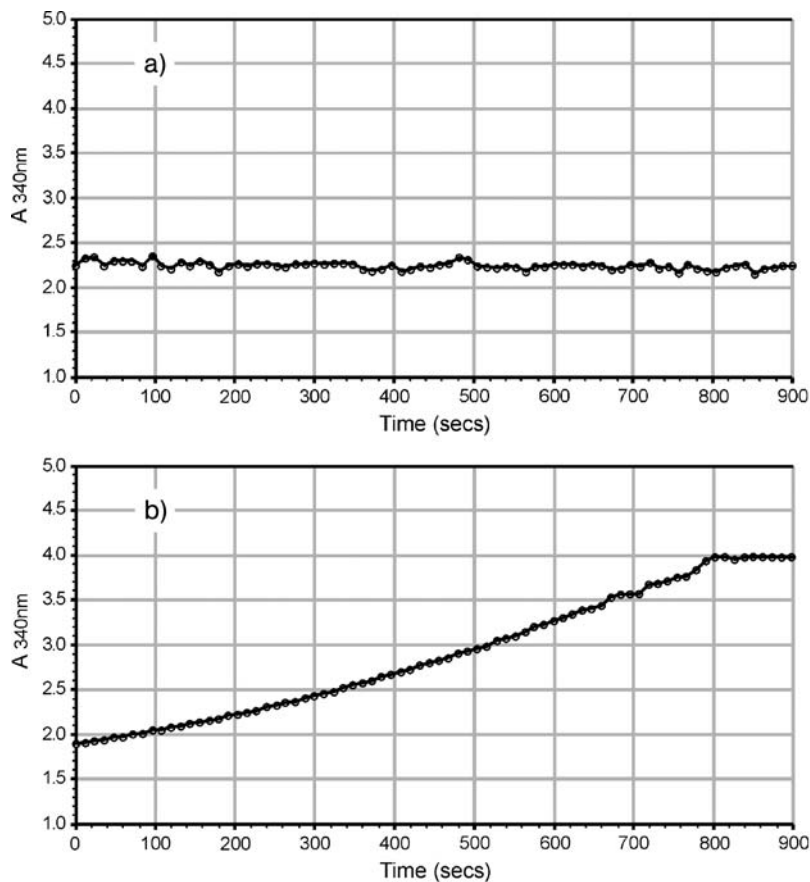


Fig. 4.5 Typical progress curves of (a) a negative *E. coli* clone and (b) an *E. coli* clone containing the HpgAT gene from *P. putida* ATCC NCIMB 12565.

4.3.2

Optical Spectroscopic Methods Based on Follow-up Conversion of Product

As the majority of the industrially relevant conversions are not accompanied by clear changes in optical signal, direct reader-based assays are frequently not an option. In many cases, however, a reader assay can still be developed by linking the actual screening reaction to a subsequent conversion that leads to a signal-generating compound. This can be realized in various ways. In a first approach, a chemical reagent or complexing agent reacts with a functional group of the (side)-products. As these derivatization reactions often require harsher conditions (especially pH) than the enzyme reaction to be screened, these types of assays are typically end-point measurements. Interference of the measurement by other compounds that contain the same functional group as the product is the

major complicating factor in these derivatization-based reader assays. A second approach is based on the subsequent conversion of one of the (side)-products of the screening reaction in an enzymatic reaction cascade, in which the final step is a conversion of a reader-detectable compound (for instance NAD^+/NADH). The major advantage of this second approach is the higher selectivity of the enzyme compared with the chemical derivatization reagent, thereby strongly reducing the number of potentially interfering compounds. Furthermore, the mild conditions needed for the enzymatic cascade more often enable measurements in a kinetic mode.

Although possible in principle, the long development time of these kinds of optical spectroscopic assays hampers their application in CCM. This is due to the large number of parameters that have to be optimized and tuned, such as concentration of the reactants and conditions (pH, temperature, and reaction time) of the different steps as well as their order. Proactive development of robust reader assays for a number of frequently encountered types of molecules ("group assays") can solve this problem and makes these potentially attractive assays CCM timelines compliant.

This approach is further elaborated in the following examples relating to synthetic routes for enantiomerically pure amino acids, which are important chemical building blocks for the pharmaceutical and agrochemical industry [30]. A significant number of the CCM market requests fall within this category. A large number of different chemo-enzymatic process concepts for the production of enantiomerically pure amino acids (Figure 4.6) have been developed and

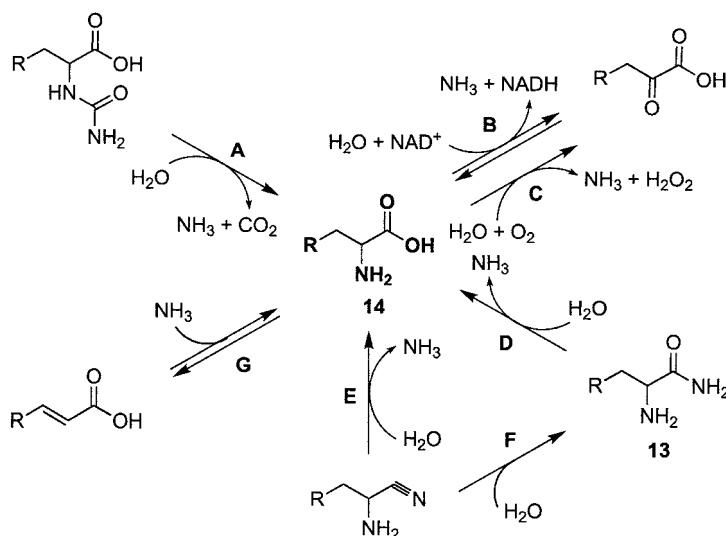


Fig. 4.6 Schematic overview of different enzymatic reactions to enantiomerically pure amino acids. (A) carbamoylase; (B) amino acid dehydrogenase; (C) amino acid oxidase; (D) amidase; (E) nitrilase; (F) nitrile hydratase; (G) ammonia lyase.

commercialized [31], and ammonia is involved in the majority of these concepts either as side-product (amidases, carbamoylases, nitrilases, or amino acid oxidases) or as substrate (for instance when using amino acid dehydrogenases and ammonia lyases). We set out to develop reader assays for amino acids as well as for ammonia and primarily focused on amidases as these are excellent tools for the production of enantiopure α -H- and α,α -disubstituted α -amino acids [30].

During development of the assays for ammonia, we noticed consistently high backgrounds when crude extracts or cell suspensions of microorganisms were used. This is caused by the growth media, which often contain yeast extract and peptone. Growth of microorganisms in these media involves the production of ammonia from the peptides and amino acids, as these media contain a surplus of nitrogen over carbon. Therefore, it may be necessary to harvest the cells by centrifugation followed by washing with a suitable buffer before starting the actual screening in order to get a sufficiently low background.

4.3.2.1 Example: Fluorometric Detection of Amidase Activity by *o*-Phthalaldehyde/Sulfite Derivatization of Ammonia

One of the reader assays we developed for the quantitative determination of ammonia is based on the derivatization reaction of ammonia with *o*-phthalaldehyde (15) and sulfite, which results in a fluorescent derivatization product (Figure 4.7). The exact structure of this fluorescent product is not known.

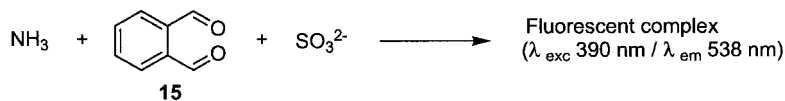


Fig. 4.7 *o*-Phthalaldehyde/sulfite derivatization reaction of ammonia leading to a fluorescent product.

We modified the original flow-injection method [32] into an analytical procedure that can be run in an automated pipetting robot in combination with a microtiter plate fluorescence reader instrument. Using this hardware combination large numbers of enzymatic samples can be screened. The method shows considerable selectivity for ammonia over amino acids and amino acid amides, which makes the method also suitable for monitoring conversion in amidase reactions.

The *o*-phthalaldehyde/sulfite assay was successfully applied in the screening of an epPCR mutant library of a bacterial amidase gene. This screening targeted mutants with improved activity towards a specific α -H- α -amino acid amide (13, Figure 4.6), without affecting this biocat's absolute enantioselectivity and good process stability.

Procedure 4.2: Amidase Assay by Spectroscopic Ammonia Detection

E. coli colonies containing the mutant library were cultivated in 96-wells microtiter plates containing 150 μL of a rich medium supplemented with the appropriate antibiotic and inducer. After 18 h at 37 $^{\circ}\text{C}$, 50 μL cell suspension was mixed with 50 μL of a 20% racemic α -H- α -amino acid amide solution (pH 8). After 2 h of incubation at 55 $^{\circ}\text{C}$, the reaction was stopped by adding 200 μL 1 M H_3PO_4 , after which the microtiter plates were stored at 4 $^{\circ}\text{C}$ before ammonia was assayed. The stopped assay mixture and calibration solutions (*vide infra*) were diluted by mixing 20 μL with 250 μL 0.1 M H_3BO_3 , pH 9.5. From this mixture, 10 μL was mixed with 45 μL of a 50 mM o-phthalaldehyde solution (in 25% methanol, 75% 0.53 M H_3BO_3 , pH 9.5) and 45 μL 50 mM Na_2SO_3 in a 384-well microtiter plate. After 30 min at 55 $^{\circ}\text{C}$, fluorescence was determined (excitation at 390 nm, emission at 538 nm) in a Fluorstar fluorimeter (BMG laboratories GmbH, Offenburg, Germany). As the signal obtained after this incubation at 55 $^{\circ}\text{C}$ still slightly increased in time, it was essential to measure all plates after the same derivatization time.

For quantification it is a prerequisite that the calibration samples resemble the gene library samples as much as possible. We achieve this by imitating a realistic amidase reaction in the same biological matrix as present in the screened library samples (that is, as conversion proceeds we get increasing amino acid and ammonia concentrations and correspondingly equimolar decreasing amino acid amide concentrations). We call this a complementary calibration procedure. All calibration samples were prepared in 0.67 M H_3PO_4 and subjected to the same procedure as the mutant library samples.

To establish the standard deviation of this screening method, the whole screening protocol was executed with a few microtiter plates with *E. coli* cells containing the wild-type amidase gene. The standard deviation of the enzyme activity was less than 15%.

In this specific screening, 30 microtiter plates each containing 96 different random clones resulted in a dataset from 2880 enzyme assays. The absolute data values (Figure 4.8) were corrected for the amount of cells in the assay by dividing the fluorescence by the optical density ($\text{OD}_{600\text{nm}}$) of the cultures.

The resulting specific activities were compared with the activity of the wild-type clones, which resulted in the identification of 41 possible hits for further testing. These mutants were rescreened on a somewhat larger scale. In this second round eight positive mutants were confirmed. Chiral high-performance liquid chromatography (HPLC) analysis of the mutants proved that the enantioselectivity of the amidase was not lost in the mutants.



Fig. 4.8 Data of an *o*-phthalaldehyde/sulfite screening: activities of 2880 clones from an amidase epPCR library normalized for the cell density of the cultures (OD_{600nm}). Wild-type level is shown as a dashed line.

4.3.2.2 Example: Colorimetric Detection of Amidase Activity by Detection of Ammonia via Glutamate Dehydrogenase-coupled Assay [33]

A second reader assay we developed to measure the ammonia concentration in high-throughput mode is based on the enzymatic oxidation of NADH by glutamate dehydrogenase (EC 1.4.1.3) [34]. In this reaction glutamate dehydrogenase converts its substrates, α -ketoglutarate (7) and ammonia, into L-Glu (8) thereby oxidizing NAD(P)H to NAD(P)⁺ (see Figure 4.9). The oxidation of NAD(P)H can be followed in a spectrophotometer at 340 nm or in a fluorimeter (excitation at 339 nm, emission at 460 nm). If the concentration of ammonia to be determined is lower than the concentration of α -ketoglutarate and NAD(P)H, the decrease in NAD(P)H is directly related to the concentration of ammonia in the sample. It can be calculated using the extinction coefficient of NAD(P)H at 340 nm, $6.22 \times 10^3 \text{ M}^{-1} \text{ cm}^{-1}$.

As glutamate dehydrogenase is very specific for ammonia, high concentrations of amino acids and amino acid amides do not interfere with the measurements. This makes this assay very suitable for the quantification of ammonia in samples from different enzymatic conversions as depicted in Figure 4.6, as long as neither NAD(P)⁺ nor the reduced form are cofactors in such reactions. Another advantage of this method is the stable signal obtained after the glutamate

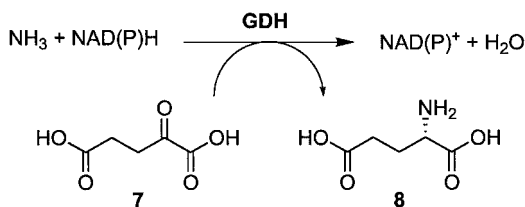


Fig. 4.9 The glutamate dehydrogenase (GDH) reaction, coupling the amount of ammonia in a sample to oxidation of NAD(P)H.

dehydrogenase reaction has gone to completion. This makes the glutamate dehydrogenase incubation step less vulnerable than the *o*-phthalaldehyde (**15**)/sulfite derivatization method described in Section 4.3.2.1. Since screening projects for biocatalysts are usually carried out with crude extracts or cell suspensions of microorganisms, these suspensions may contain high NADH oxidase activity. This will lead to an overestimation of the ammonia concentration because NADH is also oxidized by these enzymes. To prevent this, NADH oxidase activity (as well as other disturbing enzyme activities that may be present in this ammonia assay), can be inactivated by acid treatment of cells, debris and/or proteins, followed by removal of the precipitate via centrifugation, before assaying the ammonia concentration. In many cases, this inactivation will not require an additional pipetting step as the addition of the acid also stops the actual screening reaction, which is a step in many screening protocols (especially for improved enzyme variants) anyhow. Furthermore, this acid treatment will remove all particulates from the screening reaction mixture, preventing noise in the spectrophotometric assay.

We used the glutamate dehydrogenase ammonia assay in a screening project to identify amino acid amide racemases (ARs) [33]. In conjunction with an enantioselective amidase, these enzymes enable the synthesis of enantiomerically pure α -H- α -amino acids (**14**) from the corresponding racemic α -H- α -amino acid amides

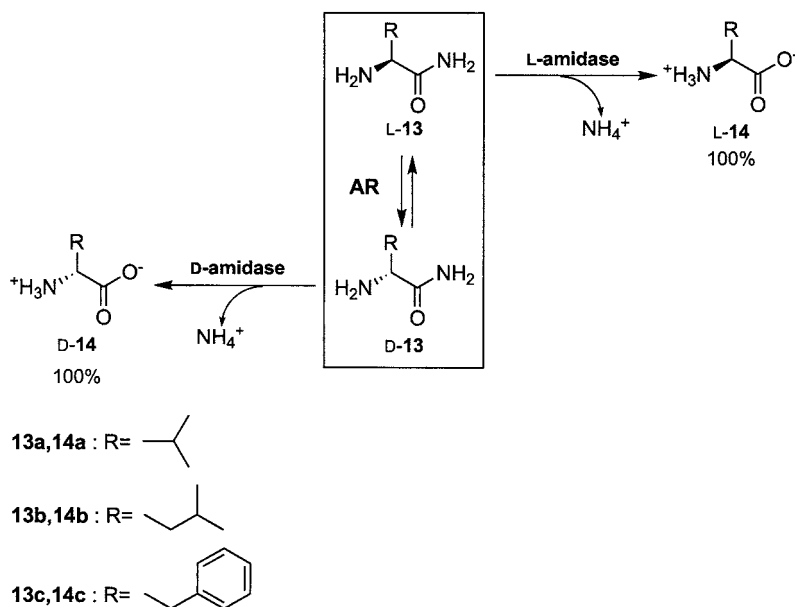


Fig. 4.10 Dynamic kinetic resolution (DKR) of α -H- α -amino acid amides by combining an enantioselective amidase with an amino acid amide racemase (AR), leading to 100% theoretical yield of the desired amino acid enantiomer. The AR reaction is boxed.

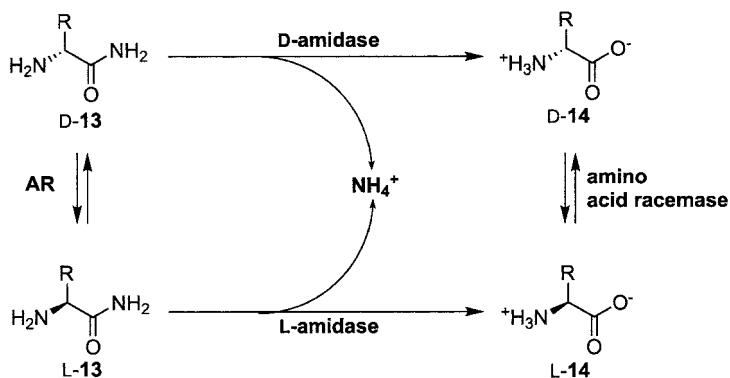
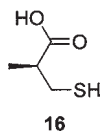


Fig. 4.11 Schematic representation of possible reaction pathways using a *D-a-H-a*-amino acid amide as screening substrate for an amino acid amide racemase (AR). Besides the desired pathway via the AR and L-amidase, hits identified may alternatively contain a D-amidase, potentially together with an

amino acid racemase. Both pathways lead to formation of equimolar amounts of *L-a-H-a*-amino acid and ammonia, which is subsequently detected using the glutamate dehydrogenase assay after the screening reaction has been stopped with phosphoric acid.

(13) in a dynamic kinetic resolution (DKR) with 100% theoretical yield (Figure 4.10). Without ARs this amidase process is a kinetic resolution with an inherent maximum yield of the desired amino acid stereoisomer of only 50%.

The method for this screening for ARs is shown in Figure 4.11. Expression libraries in *E. coli* are incubated with a *D-a-H-a*-amino acid amide (D-13) as substrate, which is converted into the corresponding *L-a-H-a*-amino acid amide (L-13) by AR-positive *E. coli* clones only. As the assay mixture also contains the strictly selective L-aminopeptidase from *Pseudomonas putida* ATCC 12633 [30, 35], the *L-a-H-a*-amino acid amide formed is hydrolyzed to the *L-a-H-a*-amino acid (L-14) under formation of ammonia, which will subsequently be determined using the glutamate dehydrogenase assay. As can be seen in Figure 4.11, this screening approach will not only identify AR-positive clones, but also *D*-selective or nonselective amidase-containing clones, since they will directly hydrolyze the *D-a-H-a*-amino acid amide screening substrate. Therefore, more specific follow-up assays were needed to distinguish *D*-amidase from AR-positive clones. This was done on an HPLC system in which all four components of the reaction mixture, namely the *a-H-a*-amino acid amide (13) and *a-H-a*-amino acid (14) enantiomers, were quantified after derivatization with *o*-phthalaldehyde (15) in combination with *D*-3-mercapto-2-methylpropionic acid (16) as chiral reagent [36].



Procedure 4.3: Enzyme-coupled Assay for Amino Acid Amide Racemase via NADH Detection

The actual AR screening was executed as follows. Bacterial expression libraries in *E. coli* were cultivated in 96-well microtiter plates containing 200 μL of a rich medium with the appropriate antibiotic. The cultures were grown at 700 rpm and 28–30 $^{\circ}\text{C}$ for 2 days, before cells were harvested by centrifugation (10 min at 1500 $\times g$), washed twice with 50 mM Tris–HCl buffer, pH 7.5, and resuspended in 50 μL of 50 mM Tris–HCl buffer, pH 7.5. The screening reaction was started by the addition of 50 μL of substrate solution containing a mixture of 140 mM D-valine amide (D-13a), D-leucine amide (D-13b) and D-phenylalanine amide (D-13c) each in 50 mM Tris–HCl buffer, pH 7.5 and 2 mM MnCl_2 . After 20 h incubation in a microtiter plate shaker at 30 $^{\circ}\text{C}$ and 700 rpm, 2 μL of help solution containing the L-aminopeptidase from *P. putida* ATCC 12633 [30, 33] was added to be sure that all L- α -H- α -amino acid amide (L-13) formed in the screening reaction and not yet converted by the endogenous *E. coli* aminopeptidase, was converted into ammonia and L- α -H- α -amino acid (L-14). After an additional 1.5 h incubation at 30 $^{\circ}\text{C}$, all reactions were stopped by the addition of 100 μL of 0.15 M HCl. After removal of the cell debris by centrifugation (10 min at 1500 $\times g$), 7 μL of the clear supernatant was transferred to new microtiter plates containing 93 μL of glutamate dehydrogenase (GDH) reagent per well. This GDH reagent contained (per 100 mL) 43 mg of NADH, 116 mg of α -ketoglutarate (7), 11.8 mg of ADP and 1200 U of GDH (Sigma, St. Louis, MO, USA) in 150 mM Tris–HCl buffer, pH 8.0. After 15 min incubation at 37 $^{\circ}\text{C}$, potential positive clones were determined by measuring the $A_{340\text{nm}}$ in each well using a Spectramax 250 plus

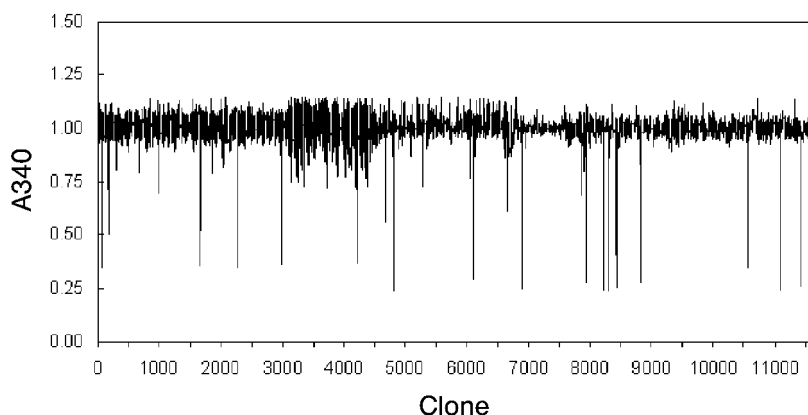


Fig. 4.12 Data of a glutamate dehydrogenase screening: decrease of $A_{340\text{nm}}$ of 11 000 clones from a gene library of *O. anthropi* expressed in *E. coli*. Absolute values obtained have been normalized to the median of each microtiter plate.

microtiter plate reader (Molecular Devices, Sunnyvale, CA, USA). The $A_{340\text{nm}}$ values obtained in the AR screening of the *Ochrobactrum anthropi* IA gene library are shown in Figure 4.12.

In order to reduce plate-to-plate variation within the $A_{340\text{nm}}$ data, the following calculation method was used to identify positive hits. From each 96 data points from a 96-well microtiter plate the average $A_{340\text{nm}}$ and standard deviation were calculated.

Subsequently, the standard deviation was multiplied by a factor of 3 and subtracted from the average. The resulting number was considered as the cut-off value. Clones with an $A_{340\text{nm}}$ lower than this cut-off value were regarded as potentially positives and selected for further screening. In this way the effect of random noise and plate-to-plate variation was diminished.

Based on this criterion, 32 potential positive clones were selected out of the 11272 clones screened. In the follow-up assays, 22 of the 32 potential positives showed clear α -H- α -amino acid formation, one of which contained the desired AR gene, as this clone clearly showed L-leucine (L-14b) formation from D-leucine amide (D-13b). The other 21 clones expressed no AR but were most likely either D-selective or nonselective amidases.

4.3.2.3 Example: Colorimetric Detection of Amino Amidase Activity Using Cu^{2+} as Sensor for Amino Acids [37]

Besides assays based on ammonia detection, we also developed reader assays for the detection of reactions involving amino acids. One of the requirements for such an assay was that it had to be applicable in screening of amino amidases (amidases requiring an α -amino group for activity). This implied that formation of a small amount of α -amino acid (product) in a matrix with a usually much higher concentration of α -amino acid amide (substrate) had to result in a clearly detectable change of spectral or fluorescent properties, in spite of the fact

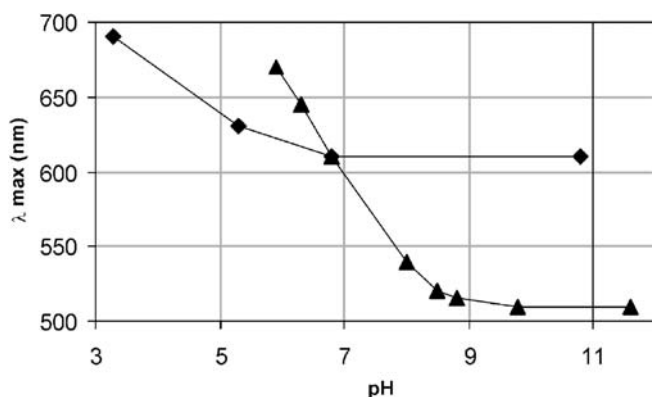


Fig. 4.13 Influence of pH on λ_{max} of the Cu^{2+} complex of valine amide (triangles) and valine (diamonds).

that both compounds contain a primary amine group. Furthermore, interference of the assay by ammonia, which is formed as side-product in equimolar amounts, had to be excluded.

Recently we described a method for the determination of amino amidase activity that meets these requirements [37]. This method is based on the use of Cu^{2+} alone as metal sensor, which distinguishes this method from earlier methods that use Cu^{2+} or Ni^{2+} as metal sensors in conjunction with a fluorescent macrocyclic metal chelator [38, 39]. It appears that the copper complexes of an α -amino acid and an α -amino acid amide have different spectral properties at alkaline pH (Figure 4.13).

Below pH 8, α -amino acid amides form bidentate complexes with Cu^{2+} , wherein two Cu-O and two Cu-N bonds are present, resulting in a blue complex (λ_{max} 620–680 nm). At a pH higher than 8, a violet-red complex is formed (λ_{max} 510 nm). The structures of both complexes (the blue as well as the violet-red one) have been described before [40]. The Cu^{2+} complex of α -amino acids, on the other hand, stays blue upon a pH change from 3.3 to 10.8, although there is a shift in λ_{max} from 690 to 620 nm [37]. The above observations were the basis for the newly developed amino amidase assay method, as this activity will then be indicated by a color change from violet-red (α -amino acid amide) to blue (α -amino acid) provided that the pH of the color assay mixture is higher than 8 (Figure 4.14).

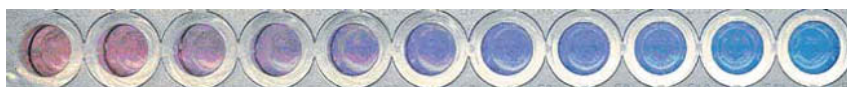


Fig. 4.14 Colors of complementary calibration solutions of α -methylvaline amide and α -methylvaline after complexation with Cu^{2+} . Color changes from violet-red (outside left well containing 43 mM α -methylvaline amide) to blue (outside right well containing 43 mM α -methylvaline and 43 mM ammonia). From left to right, wells contain an increasing concentration of α -methylvaline and ammonia (4.3 mM per well), and an equi-

molar decreasing concentration of α -methylvaline amide.

(Reprinted from *Analytical Biochemistry*, Vol. 330, A. L. L. Duchateau, M. G. Hillemans-Crombach, A. van Duijnhoven, R. Reiss, T. Sonke, A colorimetric method for determination of amino amidase activity, pp. 362–364, Copyright (2004), with permission from Elsevier).

We successfully applied this assay method in monitoring amino amidase activity in a large number of fractions resulting from enzyme purification studies. α -Methylvaline amide (**17a**) was used as substrate, which was converted into α -methylvaline (**18a**) by the desired amino amidase (Figure 4.15).

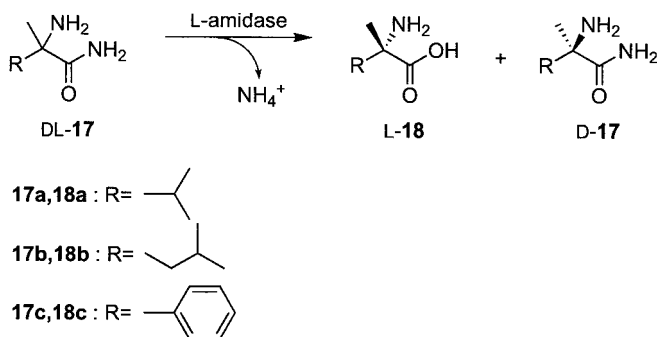


Fig. 4.15 Conversion of DL- α,α -disubstituted α -amino acid amides by an L-amidase.

Procedure 4.4: Colorimetric Amino Amidase Assay Using Cu^{2+}

An aliquot of 50 μL of each fraction was mixed with 100 μL of a 130 mM DL- α -methylvaline amide solution, pH 8.0, and 50 μL 400 mM HEPES-KOH, pH 8.0 buffer. After 1 h incubation at 37°C, the amino amidase reaction was stopped by the addition of 100 μL 1 M H_2SO_4 . After removal of particulates by centrifugation, 70 μL of the clear stopped reaction samples were added to 74 μL of an aqueous 20 mM CuSO_4 solution followed by the addition of 56 μL of a 43:5 mixture of 0.4 M borate buffer pH 9.4 and 10 M KOH. After mixing for 10 s, the absorbance was determined at 620 nm in an Optimax microtiter plate reader (Molecular Devices).

Quantification of the α -methylvaline formed was possible using a complementary calibration curve approach (Section 4.3.2.1). In this specific case such curve was prepared by mixing increasing amounts of a solution containing 130 mM ammonia and 130 mM α -methylvaline with correspondingly decreasing amounts of a 130 mM α -methylvaline amide solution (total concentration amino acid and amino acid amide: 130 mM). Just before measurement, 220 μL of each calibration solution was mixed with 220 μL 1 M H_2SO_4 and 220 μL 0.2 M HEPES-KOH, pH 9.0 to obtain the identical matrix as in the stopped reaction mixtures.

The results obtained showed that fractions containing the desired amino amidase were easily distinguishable by this simple spectrophotometric assay. Furthermore, validation of this Cu^{2+} method by measuring the amino amidase-catalyzed conversion of α -methylvaline amide by this method and HPLC/MS showed that the reader assay delivers quantitative data [37].

4.4

CCM Compliant Screening Methods Based on Generic Instrumental Assays

The short integral screening time required in CCM can also be met by use of generic assays. These assays, which are mostly instrumental, combine extremely

broad applicability with short development time. Therefore, the introduction of these methods can in most cases be done reactively after receiving a question from a customer. This clearly distinguishes them from assay methods requiring proactive development to be compliant with CCM timelines, as is the case with reader assays in which chemical derivatization or enzymatic cascade reactions (Section 4.3.2) are needed to visualize the product. Generic instrumental assays, for instance, are based on NMR and MS

4.4.1

Flow-injection NMR as Analytical Tool in High-throughput Screening for Enzymatic Activity

NMR spectroscopy is a well-established analytical technique that can yield structural and quantitative information about molecules present in solution. Originally, it was mainly applied in the synthetic field and for physicochemical research; nowadays there are numerous new applications in biology and medicine. The historical drawbacks of low sensitivity and overlapping signals have been overcome by the use of higher field strengths with their intrinsically higher sensitivity and better dispersion of resonances. Recently, the advent of so-called cryo-probeheads again increased the sensitivity by a factor of four.

Minimal sample pretreatment and better capability of water suppression also make samples of biological origin amenable for routine analysis. In enzymatic conversions substrate and product are usually different chemical entities with different magnetic properties, so they yield different NMR spectra. Due to the quantitative nature of NMR spectroscopy, conversions can easily be monitored, measuring both substrate and product.

The availability of classical sample changers makes recording of about 100 spectra per day possible. To increase the number of measurements per day, different methods of sample introduction have been developed and the application of flow-through probes has increased the number by one order of magnitude, so that ^1H -NMR spectroscopy can now be applied to medium/high-throughput screening. In the literature, screening in drug discovery and development, combinatorial chemistry libraries and body fluids is well described [41–45].

4.4.1.1 History

Convinced by the potential of NMR spectroscopy in the analysis of large series of similar samples, until then limited in our practice to 60 per night, we were thinking about increasing the throughput by alternative sample introduction techniques. Liquid Chromatography/NMR (LC/NMR) had been developed in the late 1980s and further developed in the 1990s and the use of sample loops made it possible to store HPLC fractions in those loops and later introduce and measure stored fractions in a flow-through probe. In 1995 we carried out a first trial at the Bruker site with biological samples, manually introduced into the sample loops. Using the software developed for LC/NMR, it turned out to be

- 1 = bottle for transport liquid
2 = single syringe diluter

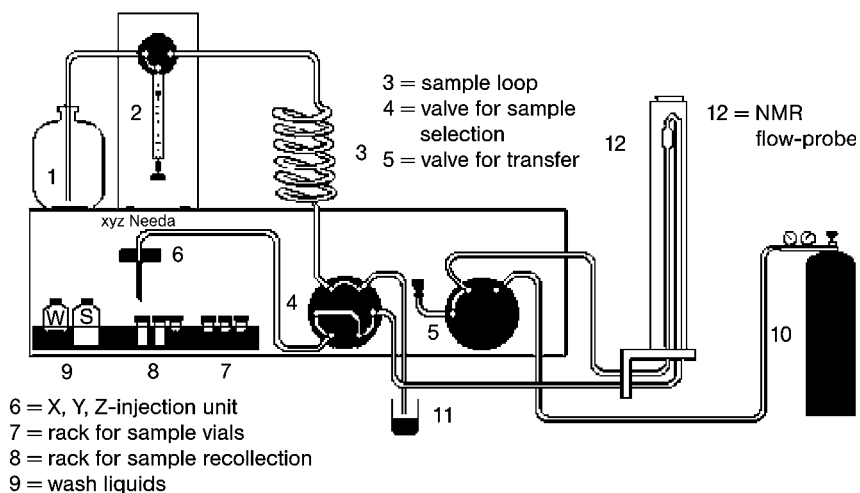


Fig. 4.16 Schematic overview of the BEST NMR system based on the Gilson XL215 autosampler.

possible to measure those samples sequentially. Although the speed at that time was still limited due to practical problems of the equipment (thin capillaries restricting the flow), the test was successful in that it proved to be possible to automatically introduce and measure NMR spectra of samples for which the only treatment had been the addition of some D_2O as a lock solvent [46].

Bruker biospin continued the development, supported by our real-life samples and suggestions. They published together with the Nicholson group in open literature about the principle in 1997 and soon thereafter commercialized their product (Bruker Efficient Sample Transport (BEST) NMR) using a Gilson 215 as a liquid handler and sample changer for vials and well plates [47].

Our first flow-through system (500 MHz) in 1998 was able to change and measure a sample in 3.5 min, which meant a total of 400 samples per day, already 4 times more than the classical sample changers could deliver. Further improvement, by optimizing each step in the process, turned out to be possible in the years thereafter to about 1 min per sample. A crucial element in this was the use of a heatable transfer capillary, which, after careful calibration, made it possible to get the sample to the desired temperature before introduction to the flow-cell, so that the normal temperature equilibration of 4–5 min could be skipped and the measurement started immediately.

For a more detailed description of the principles of flow-through systems on the market the reader is referred to the review of Stockman et al. [45]. A schematic overview of the BEST NMR set-up based on the Gilson XL215 autosampler is given in Figure 4.16.

4.4.1.2 Current Practice

There are several reasons why NMR, and in particular flow-injection NMR, is suitable for screening large libraries for new or improved biocatalysts. These are explained below.

Direct Measurement of Both Substrate and Product

No artificial substrates are needed for detection purposes. As indicated earlier, this prevents erroneous conclusions, as the artificial substrate is not representative of the aimed substrate, especially in studies to improve the catalytic properties of a specific enzyme. Derivatization reactions are also unnecessary, thus obviating the need for additional steps to the screening.

Intrinsically Quantitative

Substrate and product in most cases have similar NMR properties, such as relaxation behavior, as long as the size of the molecule does not change dramatically. As a consequence the response factor, which is unity under controlled conditions, also applies under screening conditions.

Short Development Time

Finding out whether a particular (enzymatic) conversion can be detected by NMR is easily done by measuring the spectra of the intended substrate and product under the expected conversion conditions. This will take less than 1 h. If successful, the actual screening can start, if not, conditions like pH or temperature can be changed to see if the desired separation of the resonances can be obtained. In general it is not necessary to separate all the resonances of the substrate and the product, since one is sufficient.

Where no product is available yet, simulation of the NMR spectrum by specific computer programs gives an indication whether the screening in principle would be successful. In those screenings, manual inspection of the individual spectra is needed to find actual positive hits. This is easily done by scrolling through the spectra at the expected region of interest.

No Calibration Necessary

Because substrate and product generally have similar molecular weights, integrals of substrate and product resonances can be compared more or less quantitatively, without applying the normally needed long relaxation delay (about 30 s between scans). This way, the actual measuring time can be limited and in practice can be made the same as the time needed for the preparation of the next sample.

No Critical Sample Pretreatment

The standard sample pretreatment is centrifugation, sampling the supernatant with a robotized system (e.g. 350 μL) into a deep-well plate, and adding a constant amount (50 μL) of a solution containing D_2O , EDTA (50 mM), and a known amount of maleic acid. In this way, a lock signal is supplied, and the metal ions, present from the cell-free extracts, are chelated to prevent broadening of the resonances. Moreover, addition of this solution supplies an internal standard to monitor unexpected disappearance of the substrate. In the case of screening for new biocatalysts, it is generally not necessary to stop the screening reaction, for any positive hit is welcome, irrespective of the response in a specific incubation time. When the screening has been set up to improve the enzyme, the reaction should be stopped, for example by adding acid or base, at a moment that the reaction of the wild-type enzyme is around 10% of conversion. In that case improvements can easily be detected.

Relatively Matrix-independent

NMR spectra are in general not very sensitive to small changes in sample conditions. Most substrate incubations are done between pH 5 and 7, where the influence of pH on the NMR spectrum is normally negligible. If the reaction is stopped by the addition of acid or base, care must be taken that the resulting pH is far enough from the $\text{p}K_{\text{a}}$ values of the ionizable groups in the molecule, that is below pH 1 or above pH 12.

No Reference Material Necessary

To observe the conversion of a substrate to an expected product, it is best to have a sample of the product available. If this is not possible, as pointed out under “short development time”, the spectra have to be inspected visually to find the resonances of the product. Once a positive hit has been found, the resonance positions of the product are known and the remainder of the spectra can be evaluated automatically.

Screening for Multiple Substrates/Products Possible

With flow-injection NMR it is possible to screen more than one substrate to be converted at a time. The boundary condition for this is that each substrate and product has at least one signal that does not overlap with the others, so that conversions for each substrate can be calculated. By measuring the individual spectra from potential substrates, clever combinations can be chosen. In our practice, up to four different substrates have been analyzed for conversion in one round of screening, saving time and effort needed for both the substrate in-

cubation and the measurement. The template for the calculation of the conversions can easily be adapted to report the conversions for individual substrates.

Unexpected Conversions Detectable

Another advantage of NMR spectroscopy is that, being a heuristic technique *par excellence*, it is not only the expected conversion that can be measured, but also conversions that were not intended. These can be conversions of other functional groups present or follow-up reactions of primary products.

Robust

NMR spectroscopy in general is a technique without direct interaction between the sample and the instrument during the measurement. There is no wear and no direct contact between the sample and measuring device. In practice, therefore, the downtime is very low. As long as the cryogenic liquids are replenished, the instrument will be in operation. If an electronic device fails, replacement with that item from the manufacturer's stock will solve the problem. For the flow-injection NMR system a specific area that can be problematic is the flow through the capillaries and the flow-cell. As the inlet capillary is thinnest (0.5 mm diameter), possible clogging occurs either there or in the injection port. In those cases, reversal of the flow and flushing with sufficient amounts of cleaning liquid will help. Taking care that centrifugation of the microtiter plates is done correctly and the injection needle is adjusted correctly (so that the needle does not touch the debris present) will in practice prevent clogging.

4.4.1.3 Practical Aspects

To operate the flow-through system routinely, a number of issues need to be dealt with.

Temperature Equilibration

As most enzyme assays are water-based, a number of spectroscopic problems associated with NMR measurements in water (i.e. temperature equilibration and suppression of the huge water resonance) have to be solved. The temperature of NMR probes is normally actively stabilized at a few degrees above room temperature. As a consequence, a new sample from the sample changer has to equilibrate from room temperature to the temperature of the probehead. This takes about 4–5 min. By preheating the sample on the fly before entering the flow-cell, this time can be saved, thus increasing the number of samples measured. This preheating is done in the capillary between the liquid handler and

the probehead. Bruker developed a heatable transfer capillary that, after careful calibration, does the job. This calibration must be done with actual samples and using the intended dispensing speed on the fly of a series of measurements. When the calibration is correct, the sample will be at the desired temperature and the measurement can start immediately after sample introduction.

Adjustment of Field Homogeneity

In contrast to classical NMR tubes, the flow-cell has a constant volume and wall thickness. Furthermore, the measuring coil is fixed to the outside of the cell. Consequently, each sample experiences the same surroundings, which results in an equal homogeneity of the magnetic field for all samples. In practice this means that the magnetic homogeneity after initial adjustment does not need to be changed or at most by minor adjustment of the Z and Z2 shims. These minor adjustments can be done during the dummy scans (scans without data acquisition, needed to acquire a steady state).

Water Suppression

Water suppression in automation is not trivial. Success is dependent on shimming and stable temperature. The advantage of screening with flow-injection NMR is, as mentioned earlier, that samples are very similar and that the environment inside the probe is more or less constant. In practice this means that initial shimming and determining the correct offset is sufficient to reach a good level of water suppression. Software manipulation of the data can cosmetically improve the spectrum, so that automated processing is possible. If resonances of interest are (too) close to the water resonance, increasing or decreasing the sample temperature will shift the residual water resonance to a more favorable position.

Sensitivity and Detection Limit

The detection limit on a Bruker 500 MHz machine is in the higher micromolar range. By taking more scans this can be lowered, but then the throughput rapidly decreases. When screening for new biocatalysts, substrate concentrations are used that mimic industrial application conditions whenever possible, which corresponds to concentrations of more than 50 mM. That means that in the case of a conversion of 1%, 0.5 mM should be detectable, which in practice is possible. Only when the resonances of the product are highly overlapped by those of background compounds, conditions should be optimized (pH, temperature) to allow detection.

Back Mixing

An essential prerequisite for applying flow-injection NMR in general is that samples do not mix in the flow-cell. Separation of samples is accomplished by adding cleaning segments and small gaps with air to the sample train that is injected. The amount of back mixing is influenced by factors such as aspiration and dispensing speed, amount of cleaning liquid, number and size of trails, and composition of sample. Some samples can be very sticky, such as solutions containing sugars and polyols. Trial and error with different conditions will allow optimization of the method to be used. When screening large libraries (e.g. 10000 samples or more) for a specific enzymatic activity, speed is limiting. Two factors balance the overall speed possible: sample preparation and data acquisition. The total acquisition time should be equal to the time needed for the preparation of the next sample. This time plus the time that is needed for the actual dispensing of the sample then determines the repetition time. In the case of sufficient sensitivity, sample preparation and sample dispensing become limiting. However, only a few positive hits at most are to be expected and mixing becomes less critical and some mixing can be tolerated. Even 10% of mixing is acceptable, because samples will then report 10% of the activity of the previous sample following a hit and can therefore easily be recognized. An automatic routine in a spreadsheet program can also correct for carryover from the previous experiment.

Samples can be recovered, but this limits the speed to an unacceptable low level for screening purposes.

Automation (Data Acquisition)

To handle a large number of samples, the NMR manufacturers provide standard software so that the operation of a flow probe does not differ from that of a classical sample changer [45].

Automation (Data Processing and Handling)

To keep track of the fast acquisition of NMR data, further processing of the data (Fourier transformation, phase correction, baseline correction and integration) needs to be completely automated. Modern NMR software supplies commands that can do automatic phase- and baseline correction that can also be implemented in macros. One method is to run such a macro each time a measurement has finished, the other, more reliable way is to determine the correct phase settings for the first experiment of a series (for instance a microtiter plate) and to run the processing for the series with these settings after the measurement of a microtiter plate has been finished. Automatic baseline correction is easiest when, after initial baseline correction of the complete spectrum, selected regions of interest are corrected with optimized parameters for a specific type of region. Integrals can be set

automatically once they have been determined for a single experiment. In practice the settings are constant for all similar experiments.

To calculate conversions, integrals are put into a text file, which can then be imported into a spreadsheet on a separate computer. There, calculations can be performed by macros that can easily be adapted for individual experiments. For evaluation purposes, a graphical representation of the conversion in a microtiter plate (see Figure 4.20) and a list sorted on highest conversion can be added to a report.

Cleaning

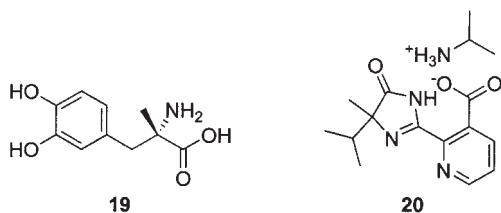
As crude enzyme preparations contain some residual biomass even after centrifugation, precautions have to be taken to prevent microbial growth in the flow system. Adding sodium azide to the cleaning and pushing liquids and flushing the last sample with one or more dummy samples with cleaning liquid is sufficient for routine operation. Regular cleaning with surfactant solution, alcohol, ammonia, and diluted nitric acid is done to remove traces of biomass and sticky components.

Capacity

NMR machines can run more or less unattended for weeks, as long as the regular filling intervals for liquid nitrogen and helium are maintained. The Gilson 215 liquid handler can contain 10 microtiter plates and most of the screenings are done in 96-well format. With a repetition time of 1 min, new samples have to be added every 16 h, which during the weekends has to be carried out four times when large screenings are done. Switching to a 384-well format, composed of four 96-well plates, limits this number to one at most. Deep-well plates of this format can contain 360 μL liquid, which is sufficient for operation with a 4 mm flow-cell. The use of a 3 mm cryo flow-cell can be done with 160 μL and can thus limit the amount of sample needed.

4.4.1.4 Example: Screening of a Bacterial Expression Library for Amidase-containing Clones

As well as α -H- α -amino acids, α,α -disubstituted α -amino acids are also a group of industrial relevance, as is exemplified by the use of L- α -methyl-3,4-dihydroxyphenylalanine (L- α -methyldopa, **19**) as an antihypertensive drug [48, 49] and α -methylvaline (**18a**, Figure 4.15) as an intermediate for the herbicide Arsenal (**20**), currently marketed as its isopropylamine salt by BASF, and related herbicides [50].



These disubstituted amino acids can be produced in enantiomerically pure form by an amidase process that is quite similar to the process for the production of α -H- α -amino acids [30]. In this case, however, ketones are used as starting material instead of aldehydes and, since the L-aminopeptidase from *P. putida* ATCC 12633 is not active for this class of amides, a different amidase is applied in the resolution step [30].

The screening assay presented here was intended to identify new amidases with activity towards α,α -disubstituted α -amino acid amides. For this purpose, α -methyl leucine amide (**17b**) was used as substrate, which will be converted into α -methyl leucine (**18b**) by positive clones (Figure 4.15). The actual gene library screening was performed as follows.

Procedure 4.5: Amidase Screening by NMR

96-Well microtiter plates filled with rich medium (200 μL per well) supplemented with the appropriate antibiotic, were inoculated with transformants from a bacterial gene library in *E. coli*. After incubation for 48 h at 22 $^{\circ}\text{C}$ at a shaking speed of 600 rpm, 100 μL of culture was transferred from each well to a wide well plate containing 300 μL DL- α -methyl leucine amide (67.5 mM, pH 8.0) and incubated for 24 h at 40 $^{\circ}\text{C}$. Subsequently, the microtiter plates were centrifuged for 10 min at 1000 $\times g$ to pellet the cells. From the supernatant in each well, 300 μL was transferred to deep-well 96-well microtiter plates each containing 50 μL D₂O with 40 g L⁻¹ maleic acid, 20 g L⁻¹ EDTA, and 1 g L⁻¹ DSS (2,2-dimethyl-2-silapentane-5-sulfonate sodium salt) (pH of this solution was adjusted to 6.5). ¹H-NMR measurements were performed at 500 MHz on a Bruker Avance spectrometer using a 4 mm flow-cell (active volume 120 μL) equipped with standard Bruker software (XWINNMR version 3.1 and ICONNMR 2.6). Aspiration and dispensing speed were set to 8 mL min⁻¹ and the sample volume was 300 μL . The total measurement time was 25 s (eight scans with two dummy scans), equalling the simultaneous preparation time of the next sample. The overall repetition time was about 60 s. Processing (Fourier transformation, phase correction, baseline correction, and integration) was done per microtiter plate, using parameters optimized for the first sample of the microtiter plate. After a quick visual check of the spectra, the integral values were exported to a text file, which was imported into an Excel file, designed to cal-

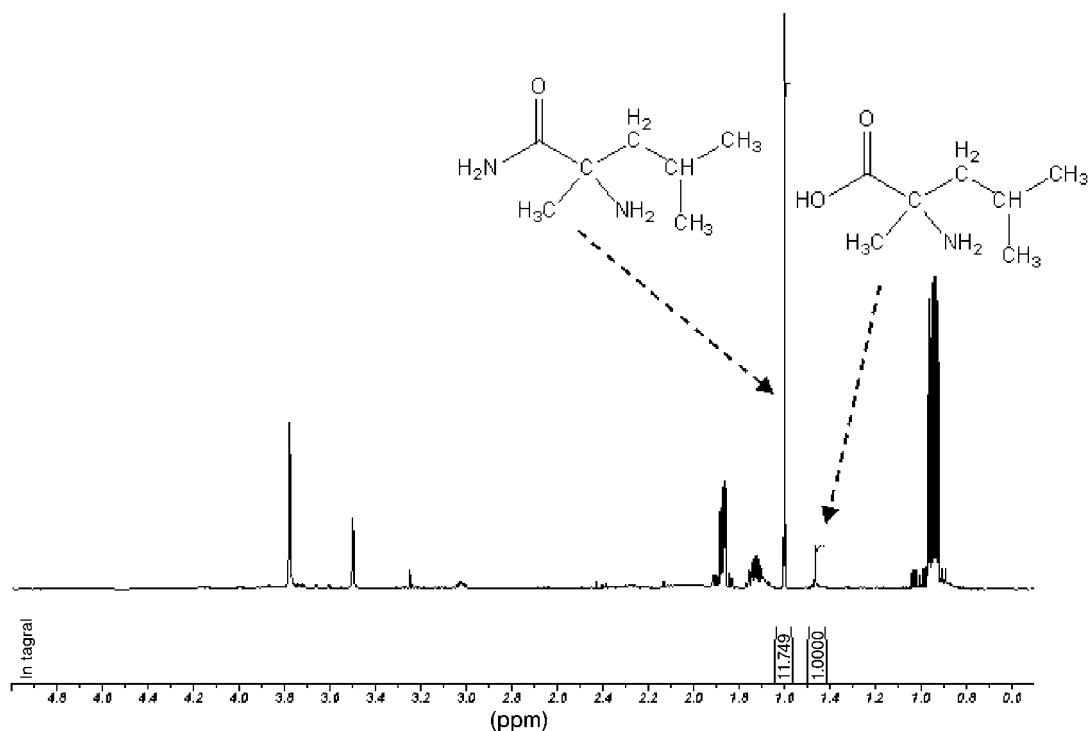


Fig. 4.17 $^1\text{H-NMR}$ spectrum of a positive hit in the screening for an amidase that hydrolyzes DL- α -methyl leucine amide.

calculate the conversions. For each microtiter plate, conversion figures and graphical representation of the results were displayed and plotted.

In total 120 microtiter plates (11 520 samples) were screened in this way, which led to the identification of two positive clones (see also Section 4.4.2.1). A spectrum of one of these positive clones is given in Figure 4.17.

4.4.1.5 Example: Identification of a Phenylpyruvate Decarboxylase Clone

Phenylpyruvate decarboxylases catalyze the decarboxylation of phenylpyruvate (21) resulting in the formation of phenylacetaldehyde (22) (Figure 4.18).

Interestingly, this enzyme can also catalyze the asymmetric acyloin condensation of phenylpyruvate with various aldehydes [51, 52]. In a similar set-up as

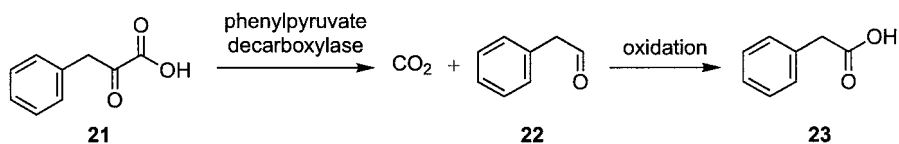


Fig. 4.18 Natural reaction catalyzed by the enzyme phenylpyruvate decarboxylase.

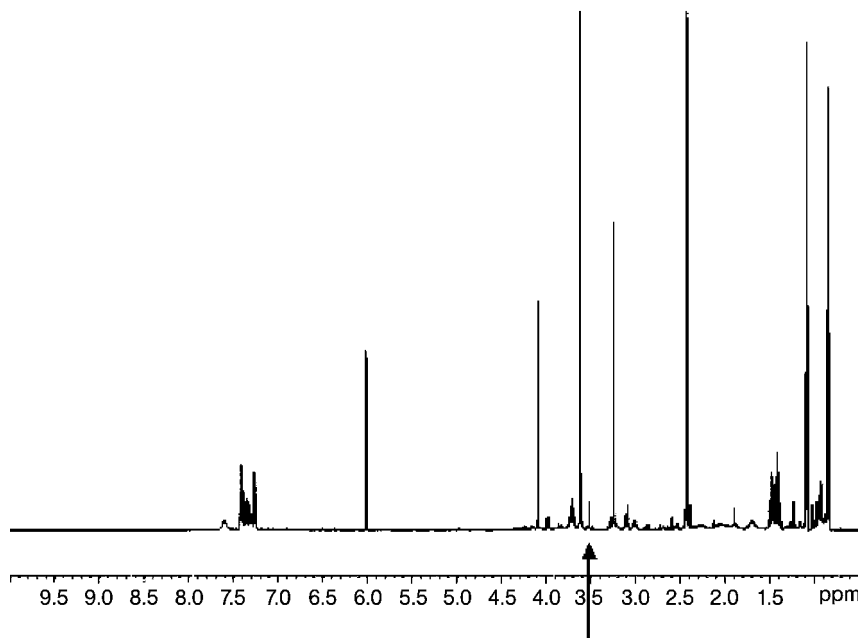


Fig. 4.19 Spectrum of a positive hit in the phenylpyruvate decarboxylase screening. This spectrum indicates the formation of phenylacetic acid (resonance indicated by the arrow) rather than phenylacetaldehyde, most likely because of (chemical) oxidation of the screening product formed.

that used in the amidase screening described in Section 4.4.1.4, but with 120 mM phenylpyruvate as substrate, a bacterial gene library in *E. coli* was screened and assayed with flow-injection NMR to isolate the phenylpyruvate decarboxylase gene from this microorganism. On visual inspection of the spectra it turned out that there were no hits, as formation of phenylacetaldehyde could be demonstrated in none of the wells. However, a few spectra could be found where the oxidation product of this compound (phenylacetic acid, **23**) was detected (Figure 4.19). Apparently, the primary aldehyde product had been oxidized into the carboxylic acid. Further examination of these clones by nucleotide sequence determination followed by subcloning proved that these hits did indeed contain the desired phenylpyruvate decarboxylase gene.

4.4.1.6 Example: Identification of Amidase Mutants with Improved Activity towards α -Methylphenylglycine Amide

Using a combination of different directed evolution techniques such as epPCR and saturated mutation primer PCR DNA shuffling [53], the specific activity of the L-amidase from *Ochrobactrum anthropi* NCIMB 40321 towards DL- α -methylphenylglycine amide (**17c**, Figure 4.15) could be improved by approximately

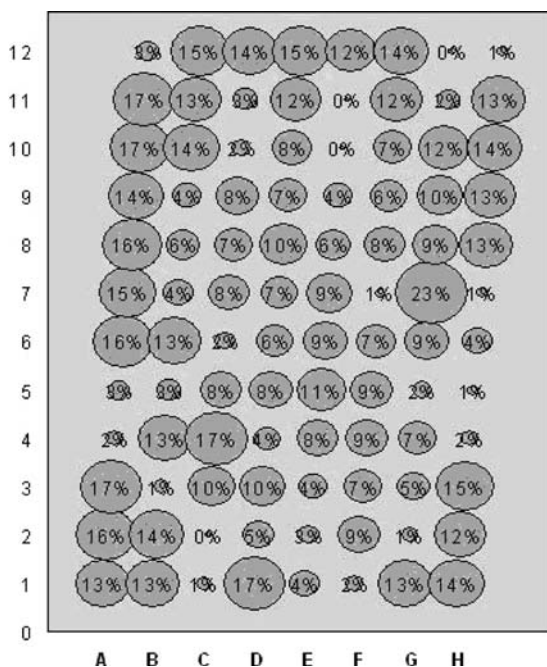


Fig. 4.20 Graphical representation of the results of one microtiter plate obtained in the screening for an amidase with improved activity towards α -methylphenylglycine amide (expressed as % conversion).

500%. Also in this case the screening method used was identical to that in example 4.4.1.4, now using DL- α -methylphenylglycine amide as substrate. In Figure 4.20 a graphical representation of the results of a typical microtiter plate is given.

4.4.2

Fast LC/MS for High-throughput Screening of Enzymatic Activity

Among the techniques developed for high-throughput screening, mass spectrometry offers good potential to act as a generic screening method and has gained increased attention in the past decade [54, 55]. Some high-throughput MS screening methods for biocatalysts have been described, without combination with chromatography. For example MALDI MS methods have been developed and optimized to quantify low-molecular-mass substrates and products of enzyme-catalyzed reactions [56, 57]. Methods based on electrospray ionization mass spectrometry (ESI MS) have been developed for the determination of conversion [58] and enantioselectivity [59, 60] of biocatalytic reactions, in which 10^3 – 10^4 biocatalysts can be screened per day. Despite the high throughput that can be realized by the above-mentioned flow-injection analysis (FIA) MS systems, by-products may not be detected in this mode [61] and screening for

enantioselective (bio)catalysts necessitates the use (and consequently the synthesis) of *pseudo*-enantiomers (chiral compounds differing in absolute configuration and isotopic labeling only) or *pseudo*-prochiral compounds [59]. The development of simple, reliable LC/MS interfaces, most notably ESI and atmospheric pressure chemical ionization (APCI), has spurred the development and acceptance of LC/MS methods, such that, nowadays, there are many laboratories that routinely use LC/MS as the primary analytical method.

The LC system is used to fractionate a preparation and the mass spectrometer is positioned as a detector, monitoring the effluent. With regard to fast LC/MS methods for the analysis of small molecules (mol wt <1000) in biological matrices, an overview of these methods has recently been published [62]. In this review article, fast LC/MS methods for analytes in, for example, blood, plasma, serum, urine, and saliva, are described. Both gradient and isocratic methods are discussed, whereby 'fast' is defined as exhibiting a cycle time of less than 5 min. For the determination of enantioselectivity, fast chiral HPLC/MS methods are reported using commercial chiral stationary phases (CSP) [63–65]. Despite the fact that numerous LC/MS applications for blood and plasma are described in the literature, relatively few studies deal with LC/MS applications for medium to high-throughput screening of biocatalysts [66, 67].

Because of its sensitivity, specificity, and medium to high throughput capability, LC/MS is often the method of choice for screening unpurified biotransformations where low conversion numbers are expected. The throughput required with a large "random" screening population dictates for both the HPLC and mass spectrometer a specific set-up. Fast chromatographic run times per sample are required, which necessitate the use of short column lengths, accompanied by sufficient efficiency. The latter may be realized by using either particulate column packings with small particles (for instance 2 or 3 μm) or silica rods. With respect to the mass spectrometric part, both LC/MS and LC/MS/MS can be run in both positive and negative ion modes.

4.4.2.1 Example: Screening of a Bacterial Expression Library for Amidase-containing Clones [68]

This LC/MS biocatalyst screening was set up in our laboratory for validation of the hits found with the flow-injection NMR screening assay described in Section 4.4.1.4. The screening assay was intended to determine amidase activity towards *a*-methyl leucine amide (**17b**, Figure 4.15) as example for *a,a*-disubstituted *a*-amino acid amides, so the product of this reaction was *a*-methyl leucine (**18b**, Figure 4.15).

Procedure 4.6: LC/MS Assay of Amidase Activity

The two 96-well plates out of 120 containing hits according to the flow-injection NMR screening assay, were centrifuged for 10 min at 1000 \times g. From the supernatant (cleared reaction mixture) in each well, 5 μL was trans-

ferred to shallow 96-well microtiter plates each containing 245 μL acetonitrile/water (50/50) supplemented with NH_4OH (0.5 weight %, pH 10.8). The LC/MS system used for the screening of the two plates consisted of an Agilent 1100 system with a binary pump, a well-plate sampler with hand-held controller and an Agilent 1100 mass-selective detector (MSD). For HPLC, a 50×4.6 mm SpeedRod column (Merck) was used in combination with an isocratic mobile phase (1.5 weight % acetonitrile in 0.1% formic acid at a flow-rate of 4 mL min^{-1} at room temperature). This system operates well with 0.3–1 μL injections and a whole-plate injection program.

Analysis time per well is 20 s, resulting in a cycle time of 32 min for a 96-well plate. The mass spectrometer was operated in the selected ion mode (SIM) at m/z 146, that is the $[\text{M}+\text{H}]^+$ ion of the acid, and at m/z 145, that is the $[\text{M}+\text{H}]^+$ ion of the acid amide.

To ensure quantitative LC/MS analysis, use of a complementary calibration procedure as described in Section 4.3.2.1 is essential. The complementary calibration mixtures were prepared by mixing a solution of DL-*a*-methyl leucine amide HCl (67.5 mM) and a solution of L-*a*-methyl leucine and NH_4OH (67.5 mM each) in increasing ratios covering a range of 0–2.54 mM L-*a*-methyl leucine (corresponding to 0–3.75% conversion of the L-*a*-methyl leucine amide substrate). The starting concentration of the amino acid amide in the 0% conversion calibration sample corresponds to the amount of amino acid amide used as substrate in the screening experiments. A conversion of 50% means that all of the L-amino acid amide is converted into L-amino acid and ammonia. As solvent for both solutions, the same rich medium as applied in the screening protocol was used in which *E. coli* had been cultured over 24 h at 20 °C. After removing the cells by centrifugation, the supernatant was diluted 4 times with Milli-Q water (Millipore, Bedford, MA, USA) before dissolving the calibration curve components. The chromatograms obtained for the calibration solutions by SIM are given in Figure 4.21.

In all calibrations carried out during the LC/MS screening assay, linear regression analysis indicated that the correlation coefficient of each single calibration curve was better than 0.994. The coefficient of variation for seven calibrations was of the order of 5%. The detection limit of conversion that could be determined was about 0.1% conversion in the case of *a*-methyl leucine amide. For quantitation, the ratio of the acid and the amide (positive ion detection) was taken. By this method of internal standardization, the effect of possible evaporation is corrected. The two hits identified by flow-injection NMR were also found by LC/MS assay. The results are shown in Table 4.2. The percentage conversion found by using LC/MS was somewhat lower than those measured by flow-injection NMR, however, in both assays the two positive hits could be clearly identified among the whole sample population.

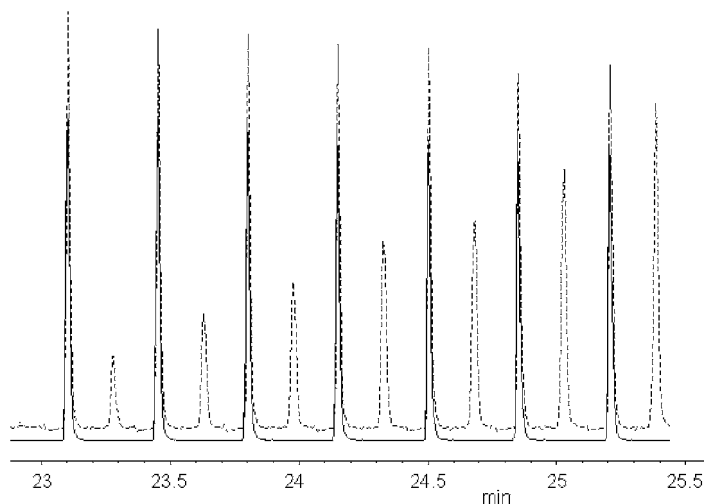


Fig. 4.21 Chromatograms of complementary calibration mixtures recorded in SIM mode (m/z 145 is $[M+H]^+$ of α -methyl leucine amide (black line); m/z 146 is $[M+H]^+$ of α -methyl leucine (dashed line)). The abundance plots are normalized to full scale.

Table 4.2 Amidase hit identification (expressed as % conversion of DL- α -methyl leucine amide) in a bacterial gene library by flow-injection analysis NMR (FIA NMR) and LC/MS.

Plate number	Well number	FIA NMR (% conv.)	LC/MS (% conv.)
77	E4	3.7	3.3
97	B5	3.4	2.2

4.4.2.2 Example: Screening of Enzymatic Racemase Activity [66]

As a second example we describe the use of LC/MS for the screening of enzymatic racemase activity. In this specific case, clones from an expression library were screened for a racemase for α -amino- ϵ -caprolactam (ACL, **24**) using a fast chiral LC separation method combined with ion-spray MS as the detection technique. The determination of the conversion in a racemase reaction necessitates

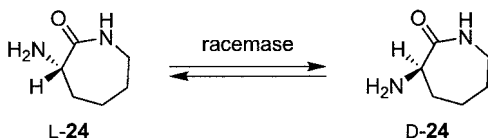


Fig. 4.22 Reaction catalyzed by an α -amino- ϵ -caprolactam (ACL) racemase.

the use of a method to distinguish both enantiomers of the substance, as one enantiomer serves as substrate for the enzyme, while the other is the product of the reaction (Figure 4.22).

Procedure 4.7: LC/MS Assay for Racemase

Using a Crownpak CR (+) column and an aqueous perchloric acid solution as mobile phase, baseline separation of the ACL (**24**) enantiomers was obtained. Since perchloric acid is a nonvolatile acid, this type of mobile phase is not suited for interfacing with a mass spectrometer.

To circumvent this problem, the mineral acid may be substituted by a short-chain organic acid as mobile phase additive. In the literature, the use of trifluoroacetic acid and acetic acid as mobile phases for Crownpak columns has been reported. In our case it is necessary to protonate the ACL enantiomers in order to obtain sufficient retention and resolution on the crownether stationary phase. As the pK_a of ACL is 7.8, an acidic pH is necessary to transform the enantiomers into their ammonium form. A 0.1% aqueous solution of formic acid (measured pH 2.6) was chosen as eluent instead of the nonvolatile perchloric acid solution. The enantioselectivity and resolution found by using the formic acid eluent were comparable with those of the perchloric acid mobile phase. By setting the flow-rate to 1.5 mL min^{-1} , a run time of 1 min per sample was obtained. Using overlapping injections, a total analysis time of 56 min per 96-well microtiter plate could be achieved. With respect to the LC/MS interface, the suitability of both the turbo ion spray and the heated nebulizer interface for the ionization of ACL were evaluated. It turned out that the sensitivity for ACL was of the same order of magnitude for both interfaces. However, in the heated nebulizer interface severe peak broadening occurred, which resulted in overlapping peaks of the enantiomers at a flow-rate of 1.5 mL min^{-1} . As peak broadening in the heated nebulizer interface is flow-dependent, flow splitting was tried. At lower flow rates the peak broadening improved, but sensitivity was lost as this type of interface is a mass-sensitive interface.

It was noticed that peak broadening in the turbo ion-spray interface was minimal and that the peak separation obtained by LC/UV could be maintained in this MS interface. Ion-spray MS in the positive ion mode was therefore chosen as detection technique for measurement of the $[M+H]^+$ ion of the ACL enantiomers at m/z 129. The *D*-ACL enantiomer showed hardly any retention on the column and coeluted with components arising from the biological matrix. This coelution caused some ionization suppression of the signal of the *D*-ACL enantiomer. The m/z 129 signal showed no contribution from the biological matrix. Figure 4.23 shows selected ion chromatograms of racemic ACL with and without a biological matrix.

The signal for the *D*-enantiomer was suppressed by about 20%. On the other hand, the *L*-enantiomer was hardly influenced by the biological ma-

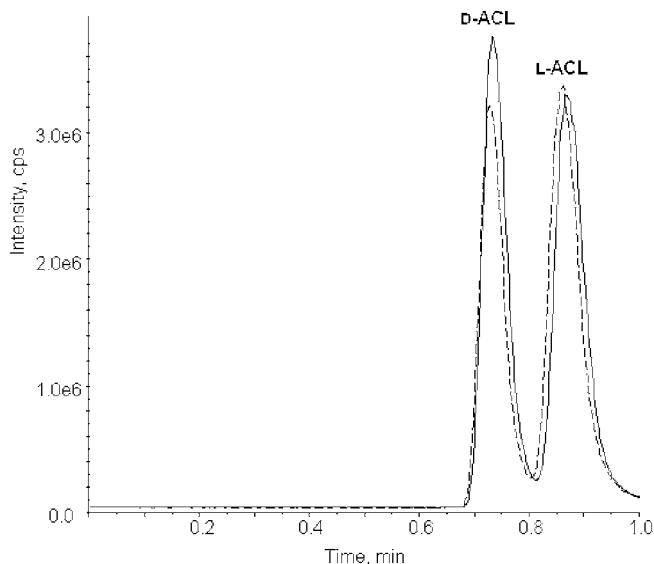


Fig. 4.23 Selected ion chromatograms (m/z 129) showing the influence of the biological matrix on the signal intensity of the both ACL enantiomers. Chromatogram (black line) showing the analysis of a racemic mixture of ACL without biological matrix and chromatogram (dashed line) with biological matrix.

(Reprinted from *Journal of Chromatography A*, Vol. 1020, H.J.W. Henderickx, A. L. L. Duchateau, P. C. Raemakers-Franken, Chiral liquid chromatography-mass spectrometry for high-throughput screening of enzymatic racemase activity, pp. 69–74, Copyright (2003), with permission from Elsevier).

trix. As the goal of the assay was the measurement of the product of the racemase reaction, that is the D-ACL enantiomer, the difference in suppression between the two enantiomers was not a problem. In the preparation of the calibration standards, a racemase reaction was mimicked in a biological matrix that was comparable with the matrix of the expression library samples. The concentration of the calibration samples covered the range of 0–50% conversion of L-ACL, with 50% conversion implying the presence of equal amounts of D-ACL and L-ACL. Figure 4.24 shows a typical intra-day calibration graph, which was measured in quadruplicate every day.

The nonlinearity of the calibration curve is caused by the limited dynamic range of ion-spray ionization. Ionization suppression and ion signal saturation account for this effect [69]. For the calibrated concentration range, the intra-day coefficient of variation (four replicate injections of the calibration samples) was typically around 8%. The inter-day coefficient of variation was about 15%. The sensitivity of the method allowed a theoretical detection limit of 0.1% conversion. In practice this level could not be attained as the substrate of the reaction already contained 0.5% D-ACL. Because of this impurity of the substrate, the detection limit for the actual screening was set to 1% conversion.

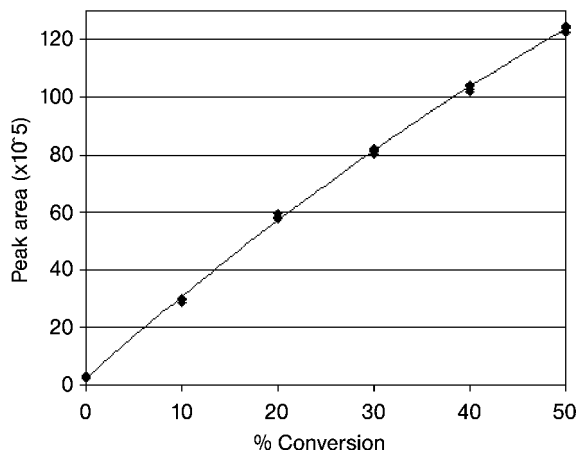


Fig. 4.24 Typical intra-day calibration graph for D-ACL, showing four replicate data points per calibration level. Optimal fit is obtained with a quadratic curve $y = -0.0106x^2 + 2.9546x + 2.3098$. The coefficient of regression is 0.9992.

(Reprinted from *Journal of Chromatography A*, Vol. 1020, H. J. W. Henderickx, A. L. L. Duchateau, P. C. Raemakers-Franken, Chiral liquid chromatography-mass spectrometry for high-throughput screening of enzymatic racemase activity, pp. 69–74, Copyright (2003), with permission from Elsevier).

In total almost 12 000 clones from an expression library were analyzed on a single Crownpak column. No significant change in column performance was noticed after screening the total expression library.

4.5 Conclusions

Although timelines in chemical custom manufacturing are extremely short, the two approaches outlined in this chapter will enable on-time identification of new or improved biocatalysts. The first approach requires proactive development of assay methods that are normally not suitable for this demanding business field because of their long development time. This situation is typical for reader assays requiring set-up of chemical derivatization or enzymatic cascade reactions. Such chemical or enzymatic follow-up reactions are needed, as it is essential to assay the truly desired industrial conversion instead of using artificial colorimetric or fluorometric substrates. The second approach is based on use of instrumental assay techniques such as NMR and MS. As these kind of techniques are very broadly applicable (“generic”), leading to rather short development times, the CCM timelines can be met without any proactive development. We are convinced that by these two screening approaches one of the major hurdles for further implementation of biocatalysis in CCM, in many cases offering the cheapest and most environmentally benign solution, can be cleared.

Acknowledgments

The authors want to gratefully acknowledge Pieter Grijpstra (BioExplore), Rob van Blijswijk, Gabriëlle v. Bruggen, Richard Kerkman, Ruud Luiten, and Charlotte van Winden (DSM Anti-Infectives, DAI Innovative), Paul Groen and Marjan Zeeman (DSM Food Specialties, Analysis), and Friso van Assema, Math Boesten, Sandra Ernste, Miriam Hillemans-Crombach, Ilse Maes, Wilco Peeters, Elly Raemakers-Franken and Anja Wagemans (DSM Pharma Chemicals, Advanced Synthesis, Catalysis & Development) for their valuable experimental and theoretical contributions. Ben Dassen and Bernard Kaptein (DSM Pharma Chemicals, Advanced Synthesis, Catalysis & Development) are gratefully acknowledged for their essential organic chemical contribution, and Prof. Lubbert Dijkhuizen (Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen), Roel Bovenberg (DSM Anti-Infectives), Birgit Schulze (DSM Food Specialties) and Marcel Wubbolts (DSM Pharma Chemicals) for their stimulating discussions.

References

- 1 S. Hanzawa, in *Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysis, and Bioseparation*, Vol. 1, eds M. C. Flickinger, S. W. Drew, John Wiley & Sons, New York, **1999**, pp. 201–210.
- 2 K. Oyama, in *Chirality in Industry*, eds A. N. Collins, G. N. Sheldrake, J. Crosby, John Wiley & Sons, Chichester, **1992**, pp. 237–247.
- 3 T. Nagasawa, H. Yamada, *Pure Appl. Chem.* **1990**, *62*, 1441–1444.
- 4 H. Yamada, M. Kobayashi, *Biosci. Biotech. Biochem.* **1996**, *60*, 1391–1400.
- 5 J. Ogawa, S. Shimizu, in *Stereoselective Biocatalysis*, ed. R. N. Patel, Marcel Dekker, New York, **2000**, pp. 1–21.
- 6 R. Olivieri, E. Fascetti, L. Angelini, L. Degen, *Biotechnol. Bioeng.* **1981**, *23*, 2173–2183.
- 7 E. J. A. X. v. d. Sandt, E. de Vroom, *Chim. Oggi* **2000**, *18*, 72–75.
- 8 Organisation for Economic Co-operation and Development (OECD), *Report: The Application of Biotechnology to Industrial Sustainability*, **2001**.
- 9 A. Bruggink, P. D. Roy, in *Synthesis of β -Lactam Antibiotics – Chemistry, Biocatalysis & Process Integration*, ed. A. Bruggink, Kluwer Academic Publishers, Dordrecht, **2001**, pp. 12–54.
- 10 A. Liese, K. Seelbach, C. Wandrey, *Industrial Biotransformations*, Wiley-VCH, Weinheim, Germany, **2000**.
- 11 A. J. J. Straathof, S. Panke, A. Schmid, *Curr. Opin. Biotechnol.* **2002**, *13*, 548–556.
- 12 J. Crosby, in *Chirality in Industry*, eds A. N. Collins, G. N. Sheldrake, J. Crosby, John Wiley & Sons, Chichester, **1992**, pp. 1–66.
- 13 A. M. Rouhi, *Chem. Eng. News* **2004**, *82*, 47–62.
- 14 M. R. Rondon, P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, R. M. Goodman, *Appl. Environ. Microbiol.* **2000**, *66*, 2541–2547.
- 15 L. Giver, A. Gershenson, P.-O. Freskgard, F. H. Arnold, *Proc. Natl Acad. Sci. USA* **1998**, *95*, 12809–12813.
- 16 J. C. Moore, F. H. Arnold, *Nat. Biotechnol.* **1996**, *14*, 458–467.
- 17 A. Crameri, S.-A. Raillard, E. Bermudez, W. P. C. Stemmer, *Nature* **1998**, *391*, 288–291.
- 18 J. E. Ness, M. Welch, L. Giver, M. Bueno, J. R. Cherry, T. V. Borchert, W. P. C. Stem-

- mer, J. Minshull, *Nat. Biotechnol.* **1999**, *17*, 893–896.
- 19 R. Patnaik, S. Louie, V. Gavrilovic, K. Perry, W.P.C. Stemmer, C.M. Ryan, S.B. del Cardayré, *Nat. Biotechnol.* **2002**, *20*, 707–712.
- 20 Y.-X. Zhang, K. Perry, V.A. Vinci, K. Powell, W.P.C. Stemmer, S.B. del Cardayré, *Nature* **2002**, *415*, 644–646.
- 21 M.T. Reetz, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 284–310.
- 22 D. Wahler, J.L. Reymond, *Curr. Opin. Biotechnol.* **2001**, *12*, 535–544.
- 23 G. Klein, J.L. Reymond, *Helv. Chim. Acta* **1999**, *82*, 400–407.
- 24 C.A. Townsend, M. Gunsior, U. Mueller, F.B.J. v. Assema, T. Sonke, WO 02/34921 to DSM N.V., DSM Biotech and Johns Hopkins University, **2002**.
- 25 D.J. Ager, I.G. Fotheringham, T. Li, D.P. Pantaleone, R.F. Senkpeil, *Enantiomer* **2000**, *5*, 235–243.
- 26 P.P. Taylor, D.P. Pantaleone, R.F. Senkpeil, I.G. Fotheringham, *Trends Biotechnol.* **1998**, *16*, 412–418.
- 27 W.J. J. van den Tweel, R.L.H.P. Ogg, J.A.M. d. Bont, EP 0315786 to Stamicarbon B.V., 1989.
- 28 S. Wiyakrutta, V. Meevootisom, *J. Biotechnol.* **1997**, *55*, 193–203.
- 29 M.A. Wegman, M.H.A. Janssen, F. van Rantwijk, R.A. Sheldon, *Adv. Synth. Catal.* **2001**, *343*, 559–576.
- 30 T. Sonke, B. Kaptein, W.H.J. Boesten, Q.B. Broxterman, J. Kamphuis, F. Formaggio, C. Toniolo, F.P.J.T. Rutjes, H.E. Schoemaker, in *Stereoselective Biocatalysis*, ed. R.N. Patel, Marcel Dekker, New York, **2000**, pp. 23–58.
- 31 A.S. Bommarius, M. Schwarm, K. Drauz, *Chimia* **2001**, *55*, 50–59.
- 32 Z. Genfa, P.K. Dasgupta, *Anal. Chem.* **1989**, *61*, 408–412.
- 33 T. Sonke, W.H.J. Boesten, G.J.W. Euverink, P. Grijpstra, P.C. Raemakers-Franken, WO 03/106691 to DSM N.V., **2003**.
- 34 H.U. Bergmeyer, H.-O. Beutler, in *Methods of Enzymatic Analysis*, ed. H.U. Bergmeyer, Academic Press, New York, **1985**, pp. 454–461.
- 35 H.F.M. Hermes, T. Sonke, P.J.H. Peters, J.A.M. Van Balken, J. Kamphuis, L. Dijkhuizen, E.M. Meijer, *Appl. Environ. Microbiol.* **1993**, *59*, 4330–4334.
- 36 A.L.L. Duchateau, H. Knuts, J.M.M. Boesten, J.J. Guns, *J. Chromatogr. A* **1992**, *623*, 237–245.
- 37 A.L.L. Duchateau, M.G. Hillemans-Crombach, A. van Duijnhoven, R. Reiss, T. Sonke, *Anal. Biochem.* **2004**, *330*, 362–364.
- 38 K.E.S. Dean, G. Klein, O. Renaudet, J.L. Reymond, *Biorg. Med. Chem. Lett.* **2003**, *13*, 1653–1656.
- 39 G. Klein, J.L. Reymond, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 1771–1773.
- 40 F. Karczynski, H. Lapkowska, H. Ratajczyk, K. Klinert, *Chemia* **1977**, *2*, 60–64.
- 41 J.T. Brindle, H. Antti, E. Holmes, G. Tranter, J.K. Nicholson, H.W.L. Bethel, S. Clarke, P.M. Scofield, E. Mcillikghan, D.E. Mosedale, D.J. Grainger, *Nature Med.* **2002**, *8*, 1439–1445.
- 42 C. Dalvit, E. Ardini, M. Flocco, G.P. Fogliatto, N. Mogelli, M. Veronesi, *J. Am. Chem. Soc.* **2003**, *125*, 14620–14625.
- 43 P.A. Keifer, S.H. Smallcombe, E.H. Williams, K.E. Salomon, G. Mendez, J.L. Belletire, C.D. Moore, *J. Comb. Chem.* **2000**, *2*, 151–171.
- 44 J.C. Lindon, E. Holmes, J.K. Nicholson, *Anal. Chem.* **2003**, *75*, 385A–391A.
- 45 B.J. Stockman, K.A. Farley, D.T. Angwin, *Methods Enzymol.* **2001**, *338*, 230–246.
- 46 D. Schipper, M. Spraul, personal communication, **1995**.
- 47 M. Spraul, M. Hofmann, M. Ackermann, A.W. Nicholls, S.J.P. Dammert, J.N. Haselden, J.P. Shockcor, J.K. Nicholson, J.C. Lindon, *Anal. Commun.* **1997**, *34*, 339–341.
- 48 A. Kleemann, J. Engel, D. Reichert, B. Kutscher, *Pharmaceutical Substances: Syntheses, Patents, Applications*, 3rd edn, Thieme, Stuttgart, **1999**, pp. 1213–1215.
- 49 D.F. Reinhold, R.A. Firestone, W.A. Gaines, J.M. Chemerda, M. Slettinger, *J. Org. Chem.* **1968**, *33*, 1209–1213.
- 50 C.D.S. Tomlin, *The Pesticide Manual*, 13th edn, British Crop Protection Council, Alton, Hampshire, **2003**, pp. 555–556.

- 51 Z. Guo, A. Goswami, K.D. Mirfakhrae, R.N. Patel, *Tetrahedron Asymmetry* **1999**, *10*, 4667–4675.
- 52 Z. Guo, A. Goswami, V.B. Nanduri, R.N. Patel, *Tetrahedron Asymmetry* **2001**, *12*, 571–577.
- 53 R.A.L. Bovenberg, R. Kerkman, WO 03/010183 to DSM N.V., **2003**.
- 54 C. Enjalbal, J. Martinez, J.-L. Aubagnac, *Mass Spectrom. Rev.* **2000**, *19*, 139–161.
- 55 V. Swali, G. J. Langley, M. Bradley, *Curr. Opin. Chem. Biol.* **1999**, *3*, 337–341.
- 56 M.-J. Kang, A. Tholey, E. Heinzle, *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1972–1978.
- 57 M.-J. Kang, A. Tholey, E. Heinzle, *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1327–1333.
- 58 S.A. Raillard, Y.H. Chen, C. Krebber, WO 00/048004 to Maxygen, Inc., **2000**.
- 59 M.T. Reetz, M.H. Becker, H.-W. Klein, D. Stöckigt, *Angew. Chem. Int. Ed. Engl.* **1999**, *38*, 1758–1761.
- 60 W. Schrader, A. Eipper, D.J. Pugh, M.T. Reetz, *Can. J. Chem.* **2002**, *80*, 626–632.
- 61 J.-L. Aubagnac, M. Amblard, C. Enjalbal, G. Subra, J. Martinez, P. Durand, P. Renaut, *Comb. Chem. High Throughput Screening* **1999**, *2*, 289–296.
- 62 P. R. Tiller, L.A. Romanyshyn, U.D. Neue, *Anal. Bioanal. Chem.* **2003**, *377*, 788–802.
- 63 R. Bakhtiar, F.L.S. Tse, *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1128–1135.
- 64 L. Ramos, R. Bakhtiar, T. Majumdar, M. Hayes, F. Tse, *Rapid Commun. Mass Spectrom.* **1999**, *13*, 2054–2062.
- 65 Y.-Q. Xia, D.Q. Liu, R. Bakhtiar, *Chirality* **2002**, *14*, 742–749.
- 66 H. J. W. Henderickx, A. L. L. Duchateau, P. C. Raemakers-Franken, *J. Chromatogr. A* **2003**, *1020*, 69–74.
- 67 C. Preisig, G. Byng, *J. Mol. Catal. B: Enzym.* **2001**, *11*, 733–741.
- 68 S. J. v. d. Wal, unpublished work, **2001**.
- 69 A. P. Bruins, *J. Chromatogr. A* **1998**, *794*, 345–357.

Part II
Genetic Selection

5

Agar Plate-based Assays

Nicholas J. Turner

5.1

Introduction

5.1.1

Directed Evolution of Enzymes: Screening or Selection?

The directed evolution of enzymes has rapidly become established as a powerful strategy for changing the properties of enzymes in a targeted manner. Such changes can make enzymes more suitable for use as biocatalysts for practical applications or alternatively engender new catalytic function [1–6]. Directed evolution combines two independent areas of technology: (1) the generation and expression of random genetic libraries in a suitable host system and (2) methods for screening or selecting variants of interest that possess the desired characteristic (e.g. appropriate substrate specificity, enantioselectivity, catalytic activity, stability under process conditions, etc.). Several methods now exist for producing libraries of genes, some of which are in the public domain and hence can be freely used by academic and industrial laboratories around the world. Truly vast gene libraries ($>10^{12}$ variants) can now be routinely prepared, raising the prospect of increasing the possibility of identifying important mutations which can lead to beneficial improvements. In contrast, however, the poor availability of generally applicable techniques for identifying active variants of interest still constitutes a bottleneck which inevitably restricts the range of enzymes that can be successfully subjected to directed evolution approaches. The developments of novel analytical tools and protocols for finding the “needle in the haystack” is an area of active research, particularly in terms of procedures that allow very high-throughput assessment of variants.

It is important to be clear about the difference between screening and selection, and the author has chosen to adopt the nomenclature outlined by Hilvert and coworkers [7] (Figure 5.1). Random screening involves examining each variant of the library in turn to check for a particular characteristic (e.g. enantioselectivity) without any prior evaluation of the library. For example, it may be

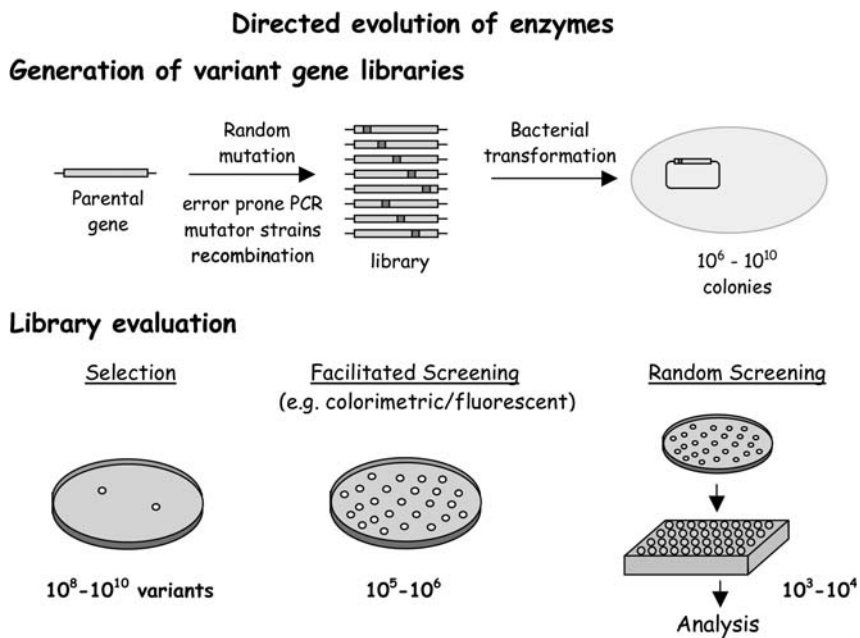


Fig. 5.1 General strategies for the directed evolution of enzymes.

necessary to pick individual clones from an agar plate, either manually or robotically using a colony picker, grow each clone in a 96-well microtiter plate, and then subject each well to analysis via liquid chromatography/mass spectrometry (LC/MS) or gas chromatography/mass spectrometry (GC/MS). Clearly random screening is a laborious process and severely limits the size of libraries that can be practically screened.

Facilitated screening reduces significantly the numbers of variants to be screened by associating the desired characteristic with a phenotype that is easy to recognize, for example a colorimetric or fluorescent reaction. Also included here, although not discussed in detail below, would be zone-clearing assays, in which the activity of a hydrolytic enzyme is detected when secreted by a colony because of the formation of a zone of clearing in the turbid solid medium that contains the substrate. In general the size of the halo is a good indication of the enzyme activity secreted by the growing colony. This type of assay has been most effectively used when screening for lipases and other hydrolytic enzymes [8]. Facilitated screening of colonies on agar plates is reasonably high throughput, allowing several hundred thousand clones to be screened per round of evolution.

Selection, by comparison, is quite different and involves directly linking the evolved enzyme activity to a vital survival characteristic of the cell (e.g. the production of an amino acid in a mutant lacking that particular precursor and hence normally unable to grow when plated-out on minimal media). Selection

has its origins in genetic complementation and has been widely used historically for improving antibiotic resistance genes. In some cases, groups have reported using selection as a primary assay, in which very large numbers of clones are examined, followed by facilitated screening as a secondary assay in which the number of clones is more manageable (see Chapter 7 for a more in-depth treatment of selection strategies based upon chemical complementation).

5.1.2

General Features of Agar Plate-based Screens

The aim of this chapter is to present a selected overview of various agar plate-based screening methods that have been developed and to illustrate how they have been coupled with random mutagenesis regimes to achieve the directed evolution of enzymes. The examples given are broadly classified as belonging to either facilitated screening or selection methods.

In general agar plate-based screens are attractive options for high-throughput screening for the following reasons:

- They can typically be used to evaluate from about 10^5 – 10^6 (facilitated screening) up to 10^{12} (selection) colonies per round of directed evolution. A typical standard size Petri dish can accommodate about 2000–3000 colonies when plated-out at high density for facilitated screening such that individual clones can be reliably picked.
- “Hit” colonies can be rapidly picked, replated at lower density and then grown in liquid culture at small scale to confirm the activity in a complementary assay (e.g. by LC/MS or GC/MS).
- A wide range of characteristics can be screened for by adjusting the assay conditions, for example, substrate specificity, thermostability, resistance to product inhibition (see below).
- Once established they are inexpensive to operate, requiring minimal equipment.

However there are still some factors that currently limit their more widespread application including:

- A requirement for general skills in molecular biology/microbiology that would be outside the expertise of some laboratories.
- A need to identify specific enzymatic transformations that result in either a colored/fluorescent product (facilitated screening) or the production of a metabolite that leads to improved levels of growth in the host cell (screening).

Although there are some important differences between the various agar plate-based methods reported in the literature and discussed below, in general the overall procedure is based upon the protocol shown in Figure 5.2. First the library of transformants is plated-out onto a nitrocellulose membrane which is placed on top of an agar plate. Diffusion of nutrients through the membrane allows the colonies to grow as efficiently as if they were directly plated-out onto

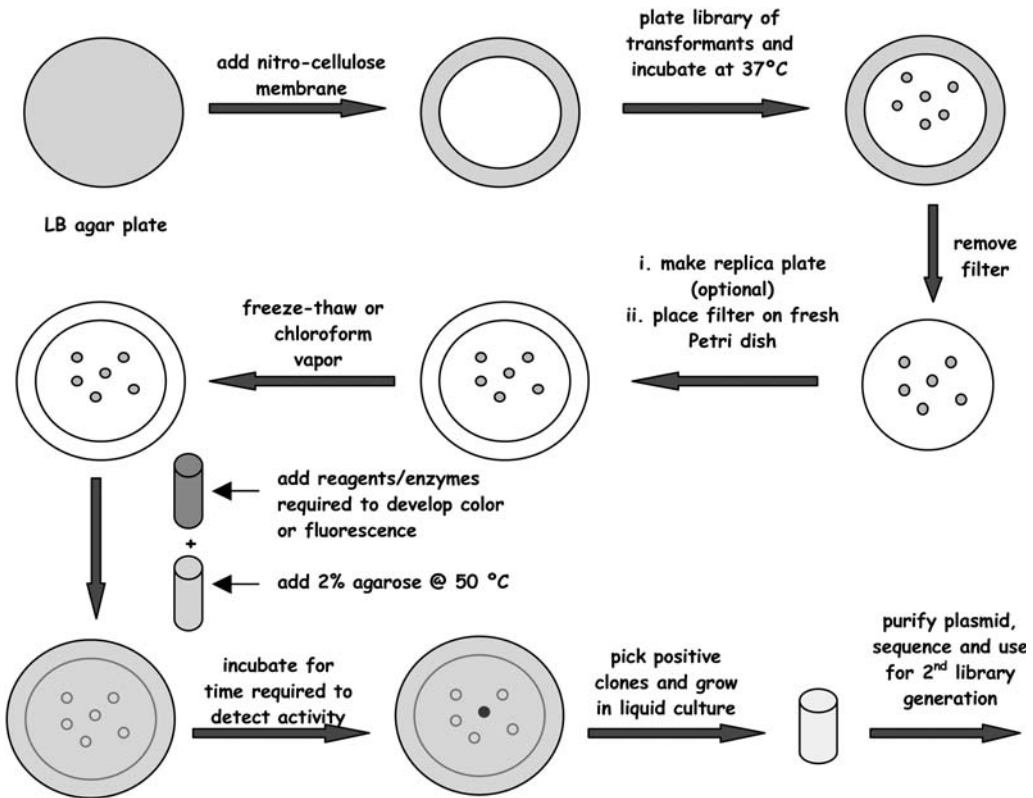


Fig. 5.2 Protocol for screening or selection of colonies on agar plates.

the agar. At this stage any reagents (e.g. IPTG (isopropyl- β -D-thiogalactopyranoside)) necessary to induce the enzyme activity of interest can be added to the agar. Once sufficient growth of the colonies has occurred the membrane is then lifted prior to assay. Many groups opt to make a replica plate at this point in order to ensure that there is an archive of the library of colonies. This step is especially important if the assay method is toxic to the cells resulting in no surviving population that can ultimately be picked. The advantage of having the colonies plated-out on a membrane is that it is easy to introduce additional screening criteria such as heat treatment of the membrane to try to identify variants that are more thermostable.

Prior to carrying out the assay of the colonies it is usually necessary to partially lyse the cells in order to release the intracellular enzyme activity. Lysis is typically carried by either using a carefully controlled freeze-thaw procedure or by immersing the membrane in a vapor of an organic solvent such as chloroform. A further option is to add lysozyme to ensure partial degradation of the cell wall. Thereafter the cocktail of substrates/enzymes necessary to generate the colored/fluorescent

product is added together with a solution of agarose which cools rapidly upon contact and helps to prevent diffusion of the color generated within individual colonies. The membranes are then incubated for the required length of time in order for individual colonies to be visualized by whichever method of detection is being employed. Active clones are then carefully picked and restreaked on fresh agar plates until single clones are obtained. These colonies can then be grown in liquid culture on a small scale (typically 100 mL volume) to generate enough cells to confirm the activity using an alternative assay (e.g. LC/MS or GC/MS). Provided that the activity is confirmed then the plasmid DNA can be isolated and the gene sequenced to identify the locations of the specific mutations. Clearly in the context of directed evolution the objective is then to subject any improved variants to subsequent rounds of random mutagenesis coupled with screening in order to obtain further improvements.

Selection is inherently simpler than screening, in that the library of clones is simply plated-out on the membrane suspended on the agar plate which typically would contain a minimal media formulation in order to encourage selection of those colonies which are able to grow under the selection conditions. In principle it is possible to examine large numbers of colonies at this stage since one is simply trying to identify clones that grow relative to those that do not. Having identified vigorously growing colonies the procedure is essentially the same as above in that these individual colonies can be picked and subsequently processed in an analogous manner.

The remainder of this chapter is devoted to a presentation of selected examples of both facilitated screening and selection strategies using agar plate solid-phase assays. The organization of the material is according to specific enzyme classes in order to encourage the reader to understand the underlying basis for the screening/selection method that has been devised and how it relates to the specific transformation that is occurring.

5.2

Facilitated Screening-based Methods

5.2.1

Amidase

Asano and coworkers have carried out the directed evolution of a D-amino acid amidase from *Ochrobactrum anthropi* SV3 in order to improve both its thermostability and catalytic activity [9]. They employed a previously developed plate-based colorimetric screen in order to select variants of interest [10]. Random mutagenesis of the *daaA* gene using error-prone polymerase chain reaction (epPCR), followed by transformation of the plasmid library into *E. coli* and growth on nitrocellulose filters gave a library of variant clones. The colonies were then lysed and heat treated for 15 min at a specific temperature (45–65 °C) to select thermostable mutants. In order to detect activity the filter was soaked

in a solution containing the substrate (D-phenylalanine amide), phenol, D-amino acid oxidase, peroxidase and 4-aminoantipyrine. Colonies that developed a pink/red color were picked and replated to give single clones. This assay exploits the capture of hydrogen peroxide generated from the D-amino acid oxidase reaction and conversion, in the presence of peroxidase, into a colored product. Using this plate-based assay they carried out two rounds of directed evolution, initially involving 10 000 mutants, which led to the identification of variant B29, whose thermostability was increased by about 5 °C. This gene was subjected to a second round of evolution (40 000 colonies screened) at a temperature of 50 °C, resulting in the identification of clone BFB40. This variant had mutations at K278M and E303V and had an optimum temperature shifted higher by about 5 °C. Furthermore, the V_{\max} was increased by about threefold relative to the wild-type enzyme, whereas the apparent K_m value was similar.

5.2.2

Esterase

Bornscheuer and coworkers have reported an effective screen for esterase activity based upon a combined growth selection/pH indicator assay [11, 12]. Libraries of the esterase from *Pseudomonas fluorescens* (PFE) were firstly generated using the Epicurian Coli[®] XL1-Red mutator strain (available from Stratagene, La Jolla, CA, USA) and subsequently transformed into the *E. coli* host expression strain. These libraries were assayed on minimal media agar plates supplemented with pH indicators (neutral red and crystal violet), thus allowing the identification of active esterase variants by formation of a red color caused by a reduction in pH due to the released acid. An additional selection criterion was introduced by using the corresponding glycerol ester because release of glycerol as a carbon source facilitated growth on the minimal media. Using this approach, one double mutant (A209D/L181V) of PFE was identified which was able to hydrolyze the target substrate with moderate enantioselectivity (25% *ee* for the remaining ester corresponding to *E*~5) but significantly enhanced activity compared to the wild-type enzyme.

5.2.3

Glycosynthase

“Glycosynthase” is a term coined by Withers and coworkers to describe enzymes derived from glycosidases that have considerably reduced hydrolytic activity and correspondingly higher glycosyl transfer activity towards sugar acceptors [13]. Glycosynthases are generated by mutating the active site catalytic residue (glutamate) to an alanine. Such enzymes become hydrolytically inactive but can transfer an activated donor sugar, typically a glycosyl fluoride (1), to a sugar acceptor (2) (Figure 5.3).

In order to improve the activity of the glycosynthase from *Agrobacterium* sp., Withers and coworkers have reported directed evolution of the gene using a

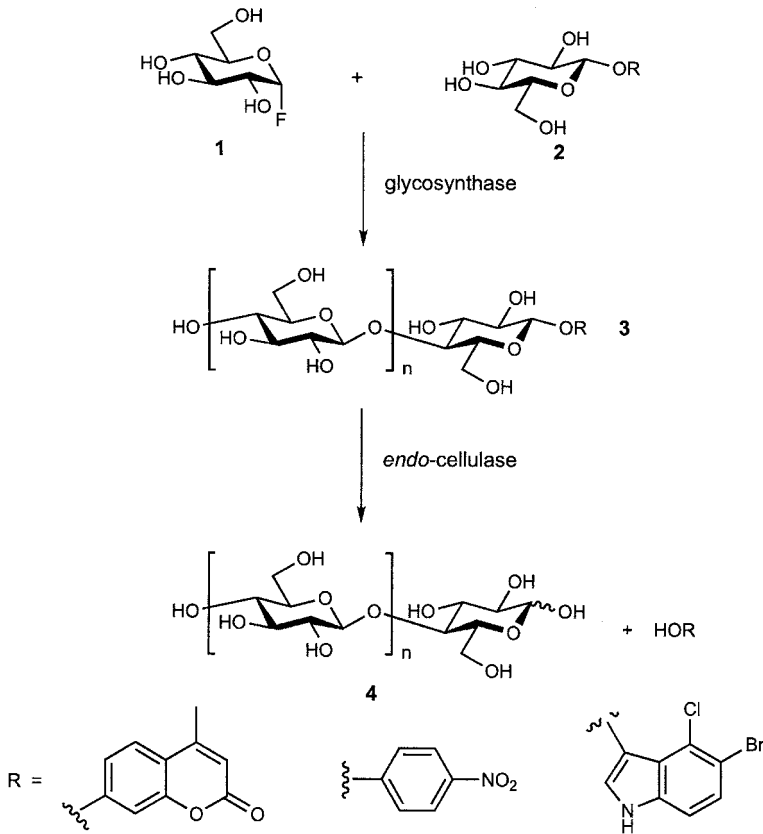


Fig. 5.3 Coupled enzyme assay for detection of glycosynthase activity.

novel two-plasmid screening system [14]. In their system, the glycosynthase gene is located on one plasmid and the reporter gene, in this case *Cel5A*, which codes for an *endo-cellulase*, is placed under separate replication on a second plasmid. The *endo-cellulase* enzyme can only catalyze release of the fluorescent umbelliferone from 3 to yield 4 after prior transfer of a glucose residue to the substrate. In order to improve the contrast of the fluorescent colonies they introduced a replica plating step by transferring the colonies to a filter paper. A small library of variants, in which residue 385 was fully randomized, was screened (about 136 colonies), resulting in the identification of a number of mutants (E385G, E385S, E385Cys, E385Ala) that possessed comparable glycosynthase activity. Recently, in order to identify mutants with significantly enhanced activity, the same group has reported applying this screen to larger libraries of glycosynthase variants [15]. Interestingly the group also decided to convert the “two-plasmid” system to a one-plasmid equivalent in order to facilitate some of the manipulation steps, in particular transformation and selection of reagents for induction. Through three rounds of directed evolution, involving

screening of about 20000 variants, they were able to identify a mutant (2F6) which possessed 27-fold improvement in catalytic efficiency (k_{cat}/K_m).

5.2.4

Galactose Oxidase

Delagrave and coworkers (Kairos Scientific) have reported a high-throughput digital imaging screen to evolve the enzyme galactose oxidase from *Fusarium* sp. [16]. Their system, known commercially as “Kcat”, combines single-pixel digital imaging spectroscopy with solid-phase screening of colonies on agar plates in order to select variants with improved activity (Figure 5.4). Transformed cells were plated-out on a polyester membrane placed on top of an agar plate. After overnight growth the membrane was transferred to another agar plate containing ampicillin in order to induce galactose oxidase. Thereafter, lysis of the microcolonies (radii <0.4 mm) was carried out by transferring the membrane to a chloroform vapor chamber for 45 s. Membranes containing the lysed colonies were then transferred to assay plates containing a cocktail of the substrate, 1% agarose, 4-chloronaphthol, and peroxidase to generate the colored colonies indicative of galactose oxidase activity. Using the insoluble galactose-containing polymer guar as substrate they screened about 100000 colonies and found a number of mutants that had greater activity than the wild-type enzyme. The most active variant (Cys383Ser) showed a 16-fold improvement in activity, partly due to

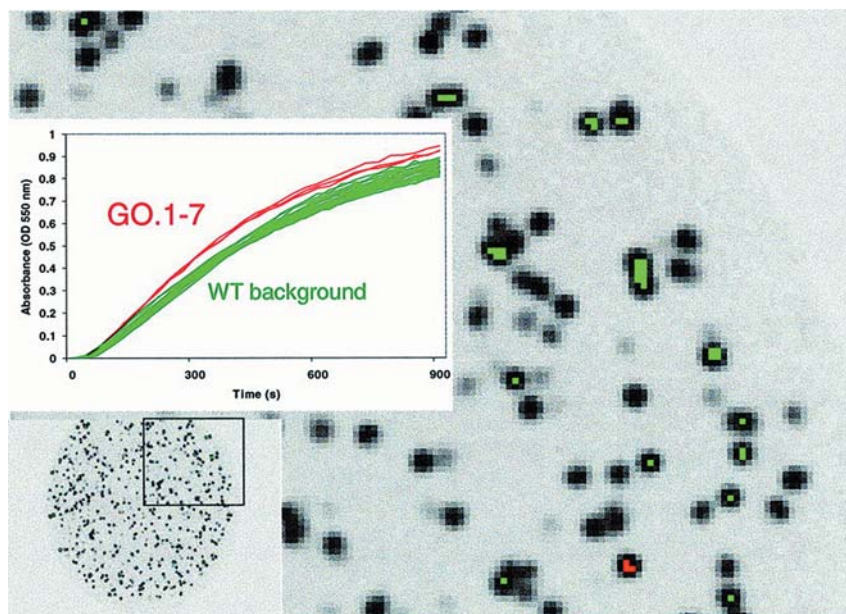


Fig. 5.4 Digital imaging of galactose oxidase variants expressed in *E. coli*.

a threefold improvement in K_m . The authors estimate that it should be possible to screen up to 100 000 mutants in a day using this technology.

The same group has reported the use of the Kcat technology to evolve the enzyme β -glucosidase from *Agrobacterium* sp. [17]. By simultaneously assaying against two different substrates (glucose and galactose) tagged with spectrally distinct chromogenic reporters they were able to identify enzyme variants that showed specificity for either glucosyl or galactosyl substrates.

5.2.5

Monoamine Oxidase

Turner and coworkers have recently developed a novel method for the efficient synthesis of enantiomerically pure α -amino acids and amines by deracemization of the corresponding racemic mixtures. Their approach relies upon coupling enzyme-catalyzed enantioselective oxidation of a chiral amine (e.g. **5**) with nonselective chemical reduction of the intermediate imine **6** in a one-pot process. Provided that the enantioselectivity of the oxidase enzyme is high, then only seven catalytic cycles are required to transform a racemic mixture into a single enantiomer (Figure 5.5). In order to identify enzymes with appropriate substrate specificity and enantioselectivity towards target substrates of interest they have developed an agar plate-based screening assay to evolve the enzyme monoamine oxidase N (MAO-N) from *Aspergillus niger* [18]. The principle of the assay is similar to that described above by Asano and coworkers (Section 5.2.1) and Youvan and coworkers (Section 5.2.4) in that it relies upon capture of the hydrogen peroxide generated from the oxidase reaction in the presence of the substrate diaminobenzidine to generate colored colonies.

In their work a plasmid harboring the MAO-N gene was transformed into *Escherichia coli*, and used to express the wild-type enzyme which was found to have very low, but detectable activity towards α -methylbenzylamine, the amine chosen for the model studies. Interestingly there was also evidence that the wild-type enzyme was inherently (*S*)-enantioselective although the intrinsic rates were very low. Random mutagenesis of the gene, using the XL1-Red mutator strain, yielded a library of variants (about 150 000) that were screened against (*S*)- α -methylbenzylamine (**5**) as substrate using the agar plate colorimetric assay.

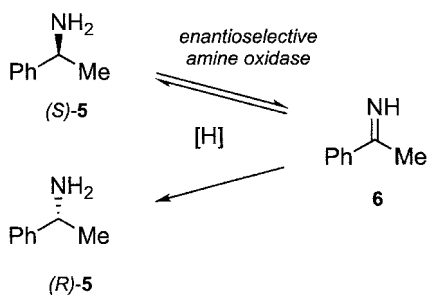


Fig. 5.5 Deracemization of racemic chiral amines using an enantioselective amine oxidase in combination with a non-selective chemical reducing agent.

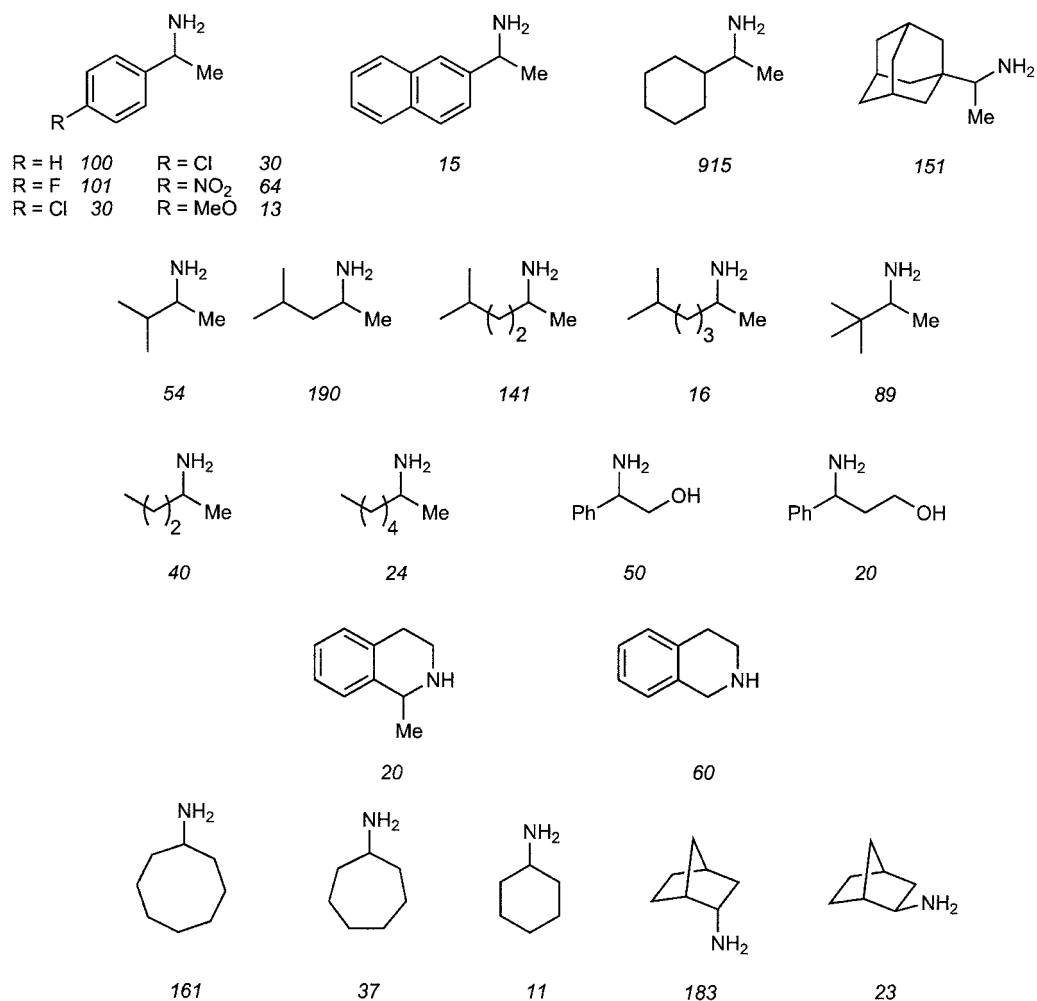


Fig. 5.6 Substrate specificity of Asn336Ser variant of MAO-N (numbers beneath structures denote relative activities).

The frequency of mutation was adjusted such that individual clones possessed about 1–2 nucleotide mutations per gene. Approximately 30 clones were selected at this stage, on the basis of their activity in the screen, and amongst these one clone in particular was found to have very high (*S*)-selectivity. Small-scale growth of this clone and partial purification of the amine oxidase revealed that this variant possessed approximately 47-fold greater activity towards *α*-methylbenzylamine (5) compared with the wild-type enzyme. This variant was sufficiently catalytically active to be used in combination with ammonia-borane to deracemize *α*-methylbenzylamine (5), yielding the (*R*)-enantiomer in 77% yield and 93% *ee*.

Sequencing of this variant revealed that it possessed a single mutation (Asn336Ser). This variant amine oxidase was purified to homogeneity and examined for its reactivity towards a panel of about 50 different chiral amines. Interestingly, more than 60% of these structurally different amines were found to be substrates, some with greater activity than *α*-methylbenzylamine (5) itself (Figure 5.6). For those substrates where the individual enantiomers were available, the enantioselectivity was determined and in all cases found to be highly (*S*)-selective ($E > 10$). By contrast, the wild-type enzyme was active towards only about 15% of the substrates, confirming that the Asn336Ser variant possessed a dramatically different substrate specificity [19].

An attractive feature of directed evolution is that variants selected from the first round of evolution can form the starting point for further rounds of mutagenesis/selection in the anticipation that additional mutations might lead to further changes in an additive manner. Further cycles can be carried out rapidly (3–4 weeks) in an automated fashion in order to fine tune the enzyme towards substrates of specific interest. Figure 5.7 illustrates such an approach involving

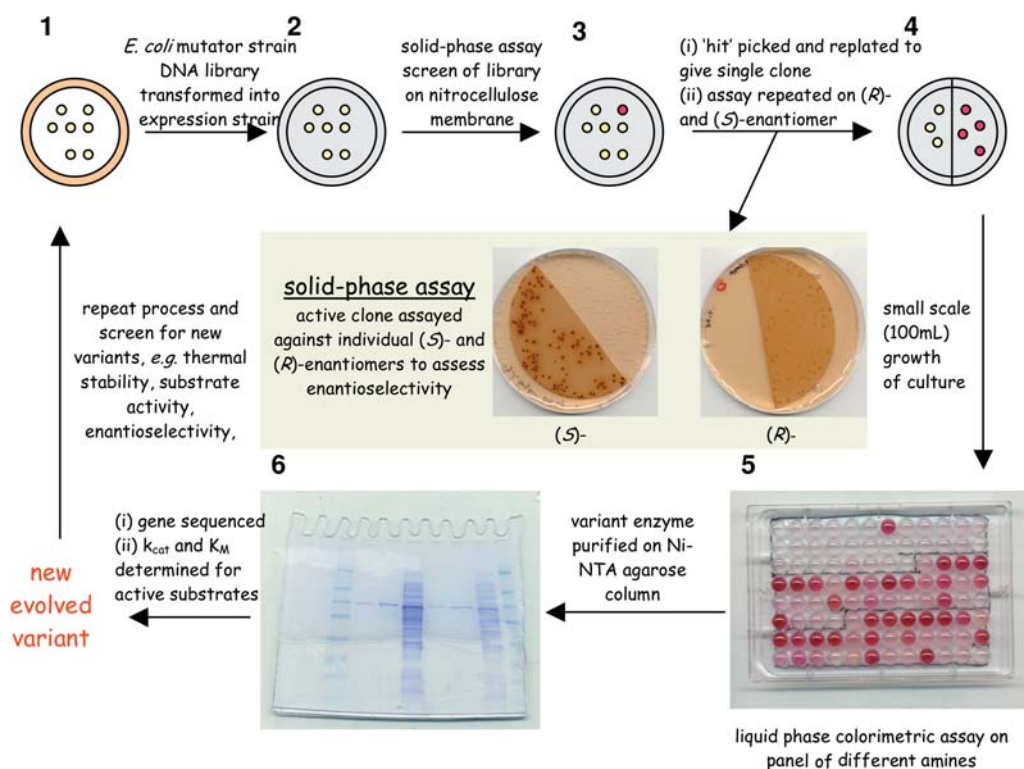


Fig. 5.7 Iterative directed evolution of an amine oxidase using random mutagenesis coupled with a hierarchical screening strategy.

(1) initial passage through the mutator strain, to generate new libraries (about 10 000–300 000 clones) which are then plated-out (2) and screened (3) for specific characteristics. Any identified “hits” are then picked (4) and assessed against a panel of substrates (5), in 96-well microtiter plate format, in order to roughly define the substrate range. Variants of particular interest are subsequently purified to homogeneity, via Ni-chelation chromatography (6) and then fully characterized against specific chiral amines of interest, including determination of k_{cat} and K_{m} values. Using this approach, Turner and coworkers recently reported the identification of a new variant (Asn336Ser/Ile246Met) that is able to deracemize racemic cyclic secondary amines on a preparative scale [20].

5.2.6

P450 Monooxygenases

Arnold and coworkers have reported a fluorescent plate-based assay for the directed evolution of P450cam from *Pseudomonas putida* [21]. In their system a library of about 200 000 P450cam variants were coexpressed with an engineered horseradish peroxidase (HRP) by cotransformation of two plasmids into *E. coli*. After incubation for 16 h, the colonies were replicated using a nitrocellulose membrane and transferred onto a fresh agar plate containing naphthalene and hydrogen peroxide. Colonies were then screened for activity using fluorescence digital imaging which indicated formation of the oxidized and coupled product (Figure 5.8). Three clones with enhanced fluorescence relative to the wild-type enzyme were picked and grown for further study. A second round of evolution led to the identification of a variant with about 20-fold improvement in activity.

Arnold and coworkers have extended the underlying principle of this digital imaging screen, using HRP as a coupling enzyme to generate colored/fluorescent products derived from the initial hydroxylation reaction, in order to screen for other monooxygenase activities [22]. For example toluene dioxygenase, which catalyzes the conversion of aromatic substrates (e.g. chlorobenzene) to the corre-

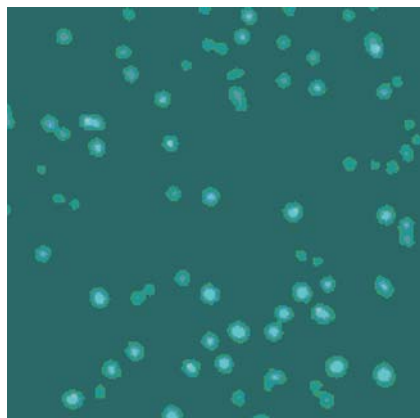


Fig. 5.8 Fluorescence digital image of a plate containing colonies of *E. coli* BL21 (DE3) cells expressing cytochrome P450cam and horseradish peroxidase and containing naphthalene and hydrogen peroxide.

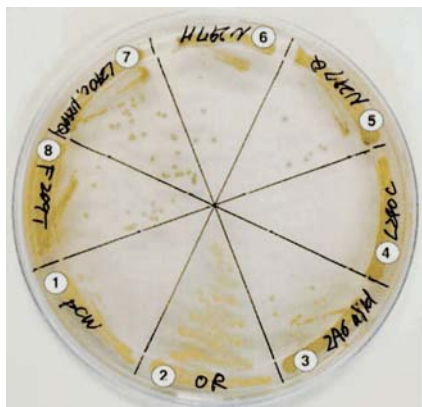


Fig. 5.9 *E. coli* colonies expressing variant cytochrome P450 2A6 indicating indigo production.

sponding 3-substituted-catechol, generates products possessing a red-brown color ($\lambda=500$ nm) in the presence of HRP. They note that the colors generated by the coupling reaction are sensitive to the site of hydroxylation and hence may be used to identify variant enzymes with altered regiospecificities.

Guengerich and coworkers have devised an alternative system for screening P450 monooxygenase activity based upon production of the blue pigment indigo. This screening method is based upon their observation that during expression of recombinant cytochrome P450 2E1 some of the colonies developed a blue coloration which was subsequently shown to be indigo [23]. They proposed that the indigo arose from dimerization of 3-hydroxyindole, itself derived from P450-mediated hydroxylation of indole, which is a breakdown product of tryptophan in bacteria. Subsequently they exploited this colorimetric screening method for the directed evolution of human cytochrome P450 2A6 (Figure 5.9) [24].

Random libraries of the gene were expressed in *E. coli* and screened for their ability to hydroxylate indole and hence generate blue colonies. One mutant, Phe209The, showed a reduced rate of hydroxylation for indole compared with the wild-type enzyme but a $k_{cat} \sim 13$ -fold higher for hydroxylation of coumarin at the 7-position. Li and coworkers have also employed this indigo screening system, in their case to evolve the fatty acid hydroxylase P450 BM-3 from *Bacillus megaterium* [25]. They identified a number of variants, in particular the triple mutant Phe87Val, Leu188Gln, Ala74Gly, which were able to hydroxylate indole. Significantly the wild-type P450 BM-3 is unable to carry out the equivalent hydroxylation reaction.

5.2.7

Carotenoid Biosynthesis

Carotenoid biosynthesis in a noncarotenogenic microorganism such as *E. coli* requires extension of the general terpenoid pathway with the genes for geranylgeranyl diphosphate synthase (*crtB*) and phytoene synthase (*crtE*) for the pro-

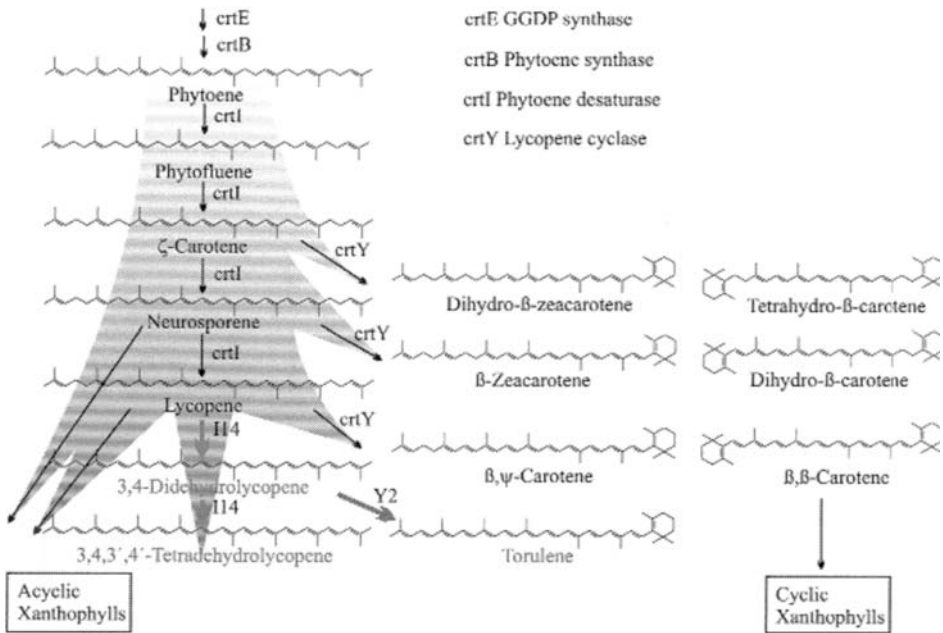


Fig. 5.10 Biosynthetic pathway leading to C-40 carotenoids.

duction of the C-40 carotenoid phytoene (Figure 5.10). Subsequent desaturation by phytoene desaturase (*crtI*) and further modifications by, for example, cyclases, hydroxylases, and ketolases, result in the production of different carotenoids.

Arnold and coworkers have reported molecular breeding of these acyclic carotenoid pathways [26]. By employing a two-plasmid-based screen, genes that produced the carotenoid precursors that serve as substrates for the target enzyme were cloned on a pACYC184-derived plasmid. Genes for the enzymes that were subjected to evolution *in vitro* were cloned on a pUC19-derived plasmid. A library of desaturases, generated by *in vitro* homologous recombination (DNA shuffling) of the genes from *E. herbicola* and *E. uredovora*, was then transformed into phytoene synthesizing *E. coli* JM101 harboring pAC-*crtEEU-crtBEU*. Colonies were transferred to nitrocellulose membranes, which provided a white background for visual screening of the clones based on color. Approximately 10000 colonies were screened with about 30% appearing white as a result of inactivation of the desaturase. Twenty colonies turned yellow, indicating the presence of carotenoids with fewer conjugated double bonds than lycopene. In addition, one pink clone (I14) was isolated suggesting the introduction of additional double bonds into lycopene by this mutant. Analysis of cell extracts by high-pressure liquid chromatography (HPLC) allowed full characterization of the reaction products.

Liao and coworkers have reported analogous experiments on the directed evolution of carotenoid-producing pathways [27]. Geranylgeranyl diphosphate (GGPP) synthase is an important rate-controlling enzyme in the production of carotenoids in *E. coli*. They constructed a library of GGPP synthase variants from *Archaeoglobus fulgidus*, which was expressed in *E. coli* and screened on agar plates. Mutants were picked on the basis of the orange color generated, which is characteristic of production of the final product astaxanthin. Eight mutants were isolated and the optimal one found to increase lycopene production by 100%. Sequencing of the variants indicated that the mutations were clustered in four “hot spots”.

The same group has applied directed evolution to the improvement of phytoene desaturase from *Rhodobacter sphaeroides* [28]. In this example they screened for libraries of variants that were able to synthesize lycopene (pink colonies) rather than neurosporene (yellow colonies).

5.2.8

Biotin Ligase

Heinis and coworkers have described a variation on the standard method for screening on agar plates which they term “colony filter screening” [29]. *E. coli* cells expressing the enzyme of interest are grown on a porous master filter that receives nutrients from an agar plate by diffusion in the normal manner. The enzyme is expressed in the cytoplasm of the cells and hence does not diffuse in the master filter. This porous master filter, containing the bacterial colonies, is then placed on top of a second filter (reaction filter). Enzymes are released from the bacterial cells by freeze-thawing and can then convert suitable substrates on the reaction filter. As a model enzyme they studied biotin ligase (BirA), which catalyzes the forma-

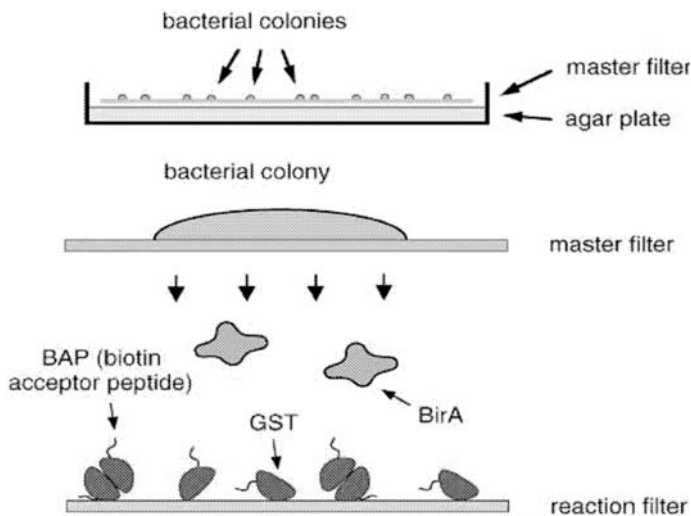


Fig. 5.11 Filter sandwich assay for biotin ligase.

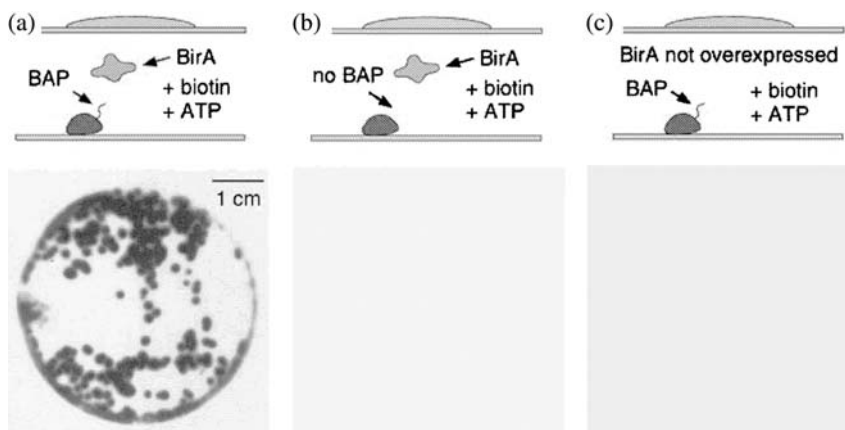


Fig. 5.12 X-ray films exposed to the reaction filters treated with streptavidin/HRP and the chemiluminescence detection kit with appropriate controls.

tion of biotinyl-5'-adenylate from biotin and ATP. The enzymatic activity of the secreted enzyme is assayed as follows and shown further below (Figure 5.11).

The reaction filter is coated with the substrate, in this case the biotin acceptor peptide (BAP), which is biotinylated if it comes into contact with the BirA ligase secreted from the master filter. Finally the reaction filter is washed, and product formation detected with streptavidin/HRP and a chemiluminescent substrate.

Figure 5.12a shows an X-ray image of the filter. Importantly, the biotin ligase released by the bacterial cells does not diffuse, permitting the identification of individual colonies expressing the desired enzyme activity. Figure 5.12b,c shows appropriate controls in which no BAP or no birA were present.

5.3

In vivo Selection-based Methods

5.3.1

Glycosynthase

Cornish and coworkers have devised an alternative to the Withers' assay described above (Section 5.2.3), namely their yeast three-hybrid chemical complementation assay, in order to carry out directed evolution of the *Humicola insolens* glycosynthase [30]. Chemical complementation detects enzyme-catalyzed bond-forming or bond-breaking reactions based upon coupling of two small molecule ligands *in vivo*. The heterodimeric small molecule reconstitutes a transcriptional activator, turning on the transcription of a downstream reporter gene (Figure 5.13). The assay is high throughput since it can be run as a growth selection

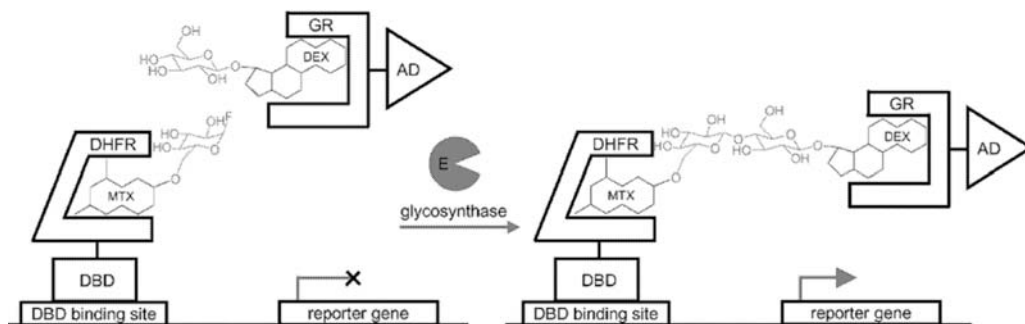


Fig. 5.13 Chemical complementation in which the heterodimeric small molecule reconstitutes a transcriptional activator, turning on the transcription of a downstream reporter gene. In this case a dexamethasone (Dex)–methotrexate (Mtx) yeast three-hybrid system is used.

whereby only cells containing functional enzyme survive. In this example they linked the glycosynthase activity of the Glu197Ala mutant of *H. insolens* to transcription of a *LEU2* reporter gene, making cell growth dependent on glycosynthase activity in the absence of leucine. A library, in which position 197 was saturated, was screened using this method resulting in the identification of a mutant (Glu197Ser) possessing fivefold greater catalytic activity.

5.3.2

Prephenate Dehydratase/Chorismate Mutase

Fotheringham and coworkers have described an interesting and powerful selection method to improve the production of L-phenylalanine (7) in *E. coli* [31]. One of the key enzymes in the conversion of glucose to L-phenylalanine (7) is the bifunctional enzyme prephenate dehydratase/chorismate mutase, which catalyzes the conversion of prephenic acid (8) to phenylpyruvic acid (9) (Figure 5.14). This enzyme, which is encoded by the *pheA* gene, is subject to strong feedback inhibition by the product of the reaction, L-phenylalanine (e.g. inhibition of prephenate dehydratase activity is almost complete at concentrations of L-phenylalanine >1 mM). A library of about 1000 *pheA* variants generated by treatment with nitrous acid was screened on agar plates impregnated with 10–20 mM β -2-thienyl-L-alanine (10), which is an analog of L-phenylalanine and is toxic to the cells. A number of colonies showed improved growth, implying greater production of L-phenylalanine in order to counteract the toxic effect of β -2-thienyl-L-alanine (10), and four in particular proved to have greater resistance to feedback inhibition. Sequencing of these variants revealed that all mutations occurred within the region specified by codons 304–310. Some mutations resulted in specific amino acid changes (e.g. Gln306Leu, Gly309Cys) whereas others involved excision of residues 307–310 and 304–306.

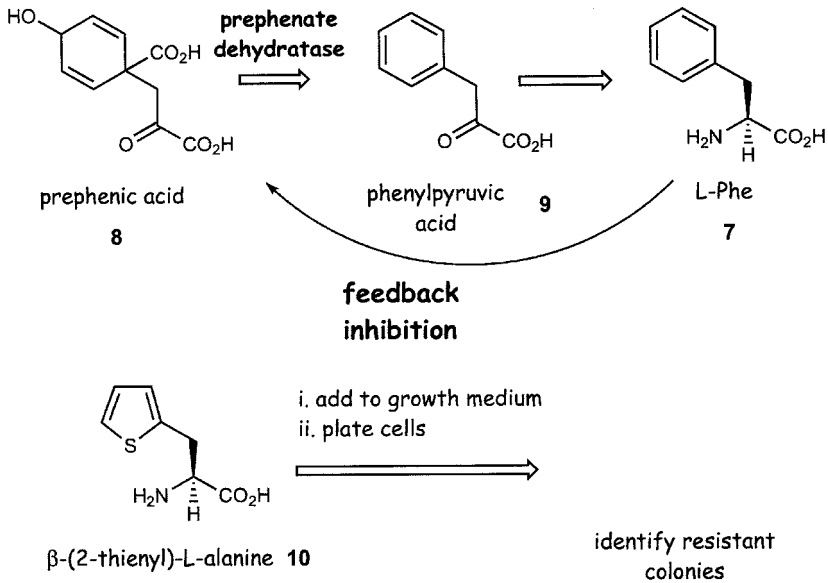


Fig. 5.14 Selection for feedback inhibition resistant variants of prephenate dehydratase.

Hilvert and coworkers have made extensive use of genetic selection methods to probe the structure and mechanism of chorismate mutase from *Methanococcus jannaschii* [32, 33]. They constructed an *E. coli* strain (KA12) in which the genes for the bifunctional enzymes chorismate mutase/prephenate dehydratase and chorismate mutase/prephenate dehydrogenase were deleted and replaced by the monofunctional versions of the dehydratase and dehydrogenase. Random gene libraries were generated and introduced into this strain and the colonies were then grown on minimal agar media lacking added phenylalanine and tyrosine. Colonies that grew under this regime were indicative of having evolved chorismate mutase activity (Figure 5.15).

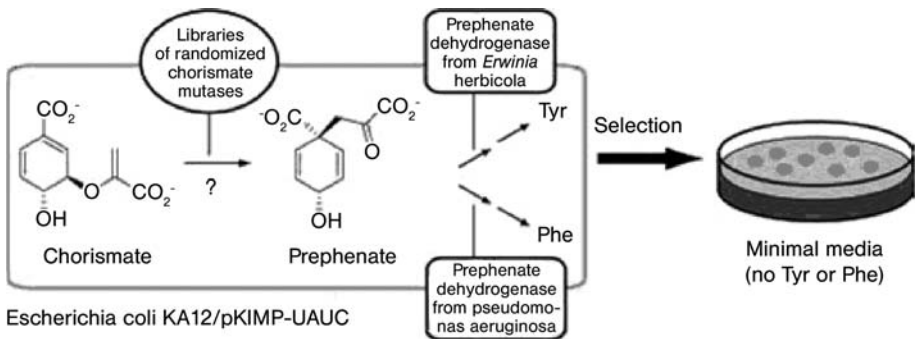


Fig. 5.15 Selection for chorismate mutase activity in *E. coli*.

5.3.3

Terpene Cyclase

Hart and coworkers have employed an *in vivo* selection method to evolve the enzyme cycloartenol synthase, which catalyzes the conversion of oxidosqualene to cycloartenol via the intermediate lanosteryl cation [34]. The wild-type enzyme yields cycloartenol and parkeol in a ratio of 99:1. In order to select for variant enzymes able to produce lanosterol they used a yeast mutant that was deficient in lanosterol production. Growing this culture in the presence of limiting amounts of ergosterol led to the isolation of a spontaneous mutant with lower ergosterol requirements. This mutant had a single amino acid change (Ile481-Val) and was shown to produce the three sterols cycloartenol:lanosterol:parkeol in a ratio of 54:25:21.

5.3.4

Tryptophan Biosynthesis

The two enzymes *N*-[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) isomerase (HisA) and phosphoribosylanthranilate (PRA) isomerase (TrpF) catalyze mechanistically related reactions in two different metabolic pathways, namely in the biosynthesis of histidine and tryptophan respectively. The key transformation is the isomerization of an aminoaldose to the corresponding aminoketose. Both enzymes have a similar (β)₈-barrel structure, a common structural motif that is associated with “progenitor enzymes” that constitute a number of superfamilies of enzymes (e.g. enolase, crotonase). Jürgens and coworkers have used random mutagenesis and selection to identify a number of HisA variants from *Thermatoga maritima* that were able to catalyze the TrpF reaction while still retaining some HisA activity [35]. In order to select appropriate variants they used an *E. coli* strain that lacked the *trpF* gene on its chromosome and hence was unable to grow without the addition of tryptophan. This strain was transformed with a plasmid library containing randomly mutated *hisA* genes generated by DNA shuffling. After growth for 3 days on a medium lacking tryptophan, three colonies were selected. Of the mutations identified in the variants, the Asp127Val change appeared particularly important in imparting TrpF activity.

5.3.5

Ribitol Dehydrogenase

Homsibrandeburgo and coworkers have reported obtaining a mutant ribitol dehydrogenase from *Klebsiella aerogenes* that has improved xylitol dehydrogenase activity [36]. The host strain containing the ribitol dehydrogenase gene was grown under selective pressure to improve growth on xylitol as the sole carbon source.

5.3.6

Inteins

Liu and coworkers have reported the directed evolution of an intein (self-splicing protein element) using an elegant selection/screening based method (Figure 5.16) [37]. The idea is based upon the development of inteins as molecular switches for activation of proteins in response to small molecules. Inteins are attractive in this respect in that their insertion into a protein blocks the target protein's function until splicing occurs. Initially they began with the *Mycobacterium tuberculosis* RecA intein using *Saccharomyces cerevisiae* as the screening host. Insertion of a natural ligand-binding domain into this minimal intein resulted in destruction of the splicing activity. In order to restore the activity they linked protein splicing to cell survival or fluorescence in the presence

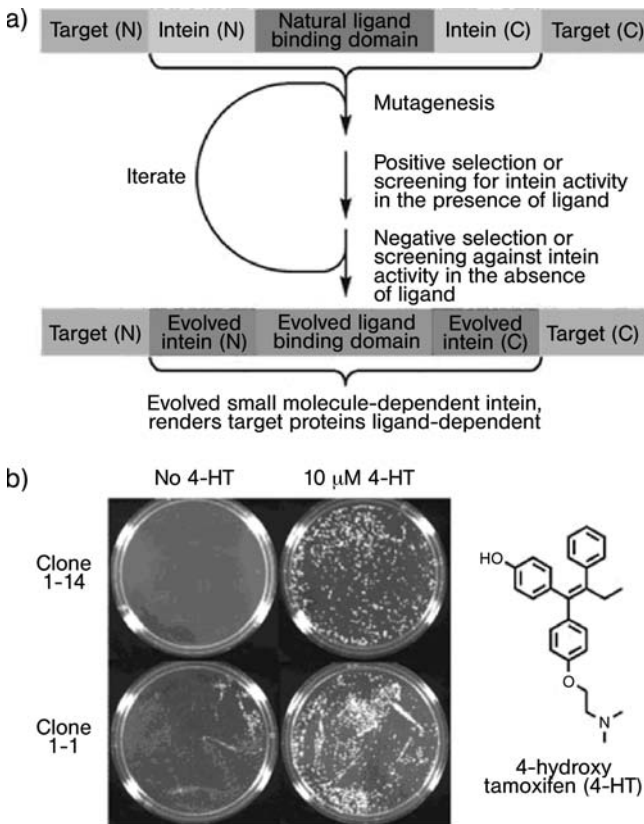


Fig. 5.16 (a) Strategy for the directed evolution of a ligand-dependent intein. (b) *Saccharomyces cerevisiae* cells, expressing kanamycin-resistance protein, grown in the presence and absence of 10 mM 4-hydroxytamoxifen (4-HT).

of the small molecule 4-hydroxytamoxifen (4-HT). Several rounds of random mutagenesis coupled with selection/screening led to efficient splicing activity in a range of different systems.

5.3.7

Aminoacyl-tRNA Synthetase

The Schultz group has for a number of years been pursuing their ambitious goal of developing strategies for the introduction of unnatural amino acids into proteins. Expanding the amino acid composition of proteins beyond the 20 natural building blocks could potentially lead to novel enzyme activities by incorporating unnatural functional groups into specific parts of the protein (e.g. the active site of an enzyme). Recently this group has reported a combined genetic selection and screen that allows the rapid evolution of aminoacyl-tRNA synthetase activity [38]. The basis of their method is to use an “orthogonal” aminoacyl-tRNA synthetase and tRNA pair that cannot interact with any of the naturally occurring synthetase/tRNA pairs in *E. coli*. A chloroamphenicol-resistance gene was used as a primary selection method to select for highly active synthetase variants and an amplifiable fluorescence reporter gene was used in combination with fluorescence-activated cell sorting (FACS) as a secondary screen to identify variants with changes in amino acid specificity. Through iterative rounds of evolution three new variants of a tyrosyl-tRNA synthetase were identified which allowed selective incorporation of amino-, isopropyl-, and allyl-containing tyrosine analogs into a protein.

5.4

Conclusions and Future Prospects

It is generally accepted that the major limiting factor in the application of directed evolution approaches is the availability of a suitable high-throughput screen/selection system for the enzyme activity of interest. In this respect, agar plate-based screens are not different and although the past 10 years has witnessed a substantial increase in the number of systems available, there still remains a need for a wider range of screens and genetic selection procedures to expand the range of enzyme activities that can be addressed. For example in the area of P450-mediated hydroxylation, the screening assays described above are either highly product specific or rely upon screening for a surrogate activity (e.g. hydroxylation of indole and subsequent formation of indigo) rather than directly on the substrate of interest. Thus there is a need for more general hydroxylation assays that can be carried out on agar plates which would undoubtedly increase the application of this important class of enzymes. Among the other classes of reactions that are difficult to screen for on agar plates one could also include reduction of carbon-carbon double bonds, sulfoxidation, dehalogenation, carbon-carbon bond formation (e.g. using aldolases, oxynitrilases, transke-

tolases) and many more. Similarly, in the area of selection using genetic complementation, the systems described above are generally confined to primary metabolism (e.g. amino acids, carbohydrates, nucleotides) or involve the production of colored compounds in secondary metabolic pathways (e.g. carotenoids). Given the inherent power of selection in terms of throughput then undoubtedly developments in these areas will have a major impact on enzyme evolution work and in this context it is interesting to note that new systems based upon transcriptional activation are starting to be developed and have the potential to be more generally applicable to a wide range of enzyme-catalyzed reactions.

References

- 1 N. J. Turner, *TIBTECH* **2003**, *21*, 474–478.
- 2 M. Alexeeva, R. Carr, N. J. Turner, *Org. Biomol. Chem.* **2003**, *1*, 4133–4137.
- 3 H. Tao, V. W. Cornish, *Curr. Opin. Chem. Biol.* **2002**, *6*, 858–864.
- 4 F. Valetti, G. Gilardi, *Nat. Prod. Rep.* **2004**, *21*, 490–511.
- 5 S. Lutz, W. M. Patrick, *Curr. Opin. Biotechnol.* **2004**, *15*, 291–297.
- 6 H. Zhao, K. Chockalingam, Z. Chen, *Curr. Opin. Biotechnol.* **2002**, *13*, 104–110.
- 7 S. V. Taylor, P. Kast, D. Hilvert, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 3310–3335.
- 8 A. Schwienhorst, Advanced screening strategies for biocatalyst discovery, in *Directed Molecular Evolution of Proteins*, eds S. Brakmann, K. Johnsson, Wiley-VCH, Weinheim, **2002**, pp. 159–175.
- 9 H. Komeda, N. Ishikawa, Y. Asano, *J. Mol. Cat. B: Enzym.* **2003**, *21*, 283–290.
- 10 H. Komeda, Y. Asano, *Eur. J. Biochem.* **2000**, *267*, 2028–2035.
- 11 U. T. Bornscheuer, J. Altenbuchner, H. H. Meyer, *Bioorg. Med. Chem. Lett.* **1999**, *7*, 2169–2173.
- 12 U. T. Bornscheuer, J. Altenbuchner, H. H. Meyer, *Biotechnol. Bioeng.* **1998**, *58*, 554–559.
- 13 L. F. Mackenzie, Q. P. Wang, R. A. J. Warren, S. G. Withers, *J. Am. Chem. Soc.* **1998**, *120*, 5583–5584.
- 14 C. Mayer, D. L. Jakeman, M. Mah, G. Karjala, L. Gal, R. A. J. Warren, S. G. Withers, *Chem. Biol.* **2001**, *8*, 437–443.
- 15 Y.-W. Kim, S. S. Lee, R. A. J. Warren, S. G. Withers, *J. Biol. Chem.* **2004**, *279*, 42787–42793.
- 16 S. Delgrave, D. J. Murphy, J. L. Rittenhouse Pruss, A. M. Maffia III, B. L. Marrs, E. J. Bylina, W. J. Coleman, C. L. Grek, M. R. Dilworth, M. M. Yang, D. C. Youvan, *Protein Eng.* **2001**, *14*, 261–267.
- 17 E. J. Bylina, W. J. Coleman, C. L. Grek, M. M. Yang, D. C. Youvan, *Proc SPIE-The International Society for Optical Engineering* **2000**, 3926, 186–191.
- 18 M. Alexeeva, A. Enright, M. J. Dawson, M. Mahmoudian, N. J. Turner, *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 3177–3180.
- 19 R. Carr, M. Alexeeva, A. Enright, T. S. C. Eve, M. J. Dawson, N. J. Turner, *Angew. Chem. Int. Ed. Engl.* **2003**, *39*, 4955–4958.
- 20 R. Carr, M. Alexeeva, M. J. Dawson, V. Gotor-Fernández, C. E. Humphrey, N. J. Turner, *ChemBioChem* **2005**, *6*, 637–639.
- 21 H. Joo, Z. Lin, F. H. Arnold, *Nature* **1999**, *399*, 670–673.
- 22 H. Joo, A. Arisawa, Z. Lin, F. H. Arnold, *Chem. Biol.* **1999**, *6*, 699–706.
- 23 E. M. J. Gillam, A. M. A. Aguinardo, L. M. Notley, D. Kim, R. G. Mundkowski, A. A. Volkov, F. H. Arnold, P. Souček, J. J. Devoss, F. P. Guengerich, *Biochem. Biophys. Res. Commun.* **1999**, *265*, 469–472.
- 24 K. Nakamura, M. V. Martin, F. P. Guengerich, *Arch. Biochem. Biophys.* **2001**, *395*, 25–31.

- 25 Q.-S. Li, U. Schwaneberg, P. Fischer, R. D. Schmid, *Chem. Eur. J.* **2000**, *6*, 1531–1536.
- 26 C. Schmidt-Dannert, D. Umeno, F. H. Arnold, *Nat. Biotechnol.* **2000**, *18*, 750–753.
- 27 C.-w. Wang, M.-K. Oh, J. C. Liao, *Biotechnol. Prog.* **2000**, *16*, 922–926.
- 28 C.-w. Wang, J. C. Liao, *J. Biol. Chem.* **2001**, *276*, 41161–41164.
- 29 C. Heinis, S. Melkko, S. Demartis, D. Neri, *Chem. Biol.* **2002**, *9*, 383–390.
- 30 H. Lin, H. Tao, V.W. Cornish, *J. Am. Chem. Soc.* **2004**, *126*, 15051–15059.
- 31 J. Nelms, R. M. Edwards, J. Warwick, I. Fotheringham, *Appl. Environ. Microbiol.* **1992**, *58*, 2592–2598.
- 32 P. Kast, M. Asif-Ullah, N. Jiang, D. Hilvert, *Proc. Natl Acad. Sci. USA* **1996**, *93*, 5043–5048.
- 33 K. J. Woycechowsky, D. Hilvert, *Eur. J. Biochem.* **2004**, *271*, 1630–1637.
- 34 E. A. Hart, L. Hua, L. B. Darr, W. K. Wilson, J. H. Pang, S. P. T. Matsuda, *J. Am. Chem. Soc.* **1999**, *121*, 9887–9888.
- 35 C. Jürgens, A. Strom, D. Wegener, S. Hettwer, M. Wilmanns, R. Sterner, *Proc. Natl Acad. Sci. USA* **2000**, *97*, 9925–9930.
- 36 M. I. HomisBrandeburgo, M. H. Toyama, S. Maragoni, S. J. Ward, J. R. Giglio, B. S. Hartley, *J. Prot. Chem.* **1999**, *18*, 489–495.
- 37 A. R. Buskirk, Y.-C. Ong, Z. J. Gartner, D. R. Liu, *Proc. Natl Acad. Sci. USA* **2004**, *101*, 10505–10510.
- 38 S. W. Santoro, L. Wang, B. Herberich, D. S. King, P. G. Schultz, *Nat. Biotechnol.* **2002**, *20*, 1044–1048.

6 High-throughput Screens and Selections of Enzyme-encoding Genes

Amir Aharoni, Cintia Roodveldt, Andrew D. Griffiths, and Dan S. Tawfik

6.1 Introduction

The availability of vast gene repertoires from both natural (genomic and cDNA libraries) and artificial sources (gene libraries) demands the development and application of novel technologies that enable the screening or selection of large libraries for a variety of enzymatic activities. This chapter describes recent developments in the selection of enzyme-coding genes for directed evolution and functional genomics. We focus on high-throughput screening (HTS) approaches that enable selection from large libraries ($>10^6$ gene variants) with relatively modest means (i.e. nonrobotic systems), and on *in vitro* compartmentalization (IVC) in particular.

The basis of all screening and selection methodologies is a linkage between the gene, the enzyme it encodes, and the product of the activity of that enzyme (Figure 6.1). The difference between screening and selection is that screening is performed on individual genes or clones and requires some spatial organization of the screened variants on agar plates (Chapter 5), microtiter plates (Chapters 1, 2 and 4), arrays, or chips (Chapters 11 and 12), whereas selections act simultaneously on the entire pool of genes. The focus of this chapter (and of Chapter 7) is selection techniques, although formally speaking, several of the techniques described are in fact screening methods. The techniques described herein are applicable in the context of both functional genomics and directed evolution.

Directed enzyme evolution has been used in the past two decades as a powerful approach for generating enzymes with desired properties. Enzyme variants have been evolved for catalytic activity under extreme conditions such as high temperatures, acidic and alkaline environments, and organic solvents [1, 2] and with improved catalytic activities and new substrate specificities [3, 4]. Directed evolution experiments are based on the principles of natural Darwinian evolution, and therefore consist of two major steps: (1) creation of genetic diversity in the target gene in the form of gene libraries; and (2) an effective selection of the library for the desired catalytic activity. A large variety of methods for creation of genetic diversity are currently available (for recent reviews see [5, 6]).

However, the typical library size is still many orders of magnitude larger than the number of protein variants that can be screened. The same restriction applies to cDNA and genomic libraries derived from natural sources (e.g. environmental libraries) the diversity of which is almost unlimited. Further, while methods for creating gene diversity are generic, the screens for activity need to be tailored for each enzyme and reaction. The “bottleneck” for most enzyme isolation endeavors is therefore the availability of a genuinely high-throughput screening or selection for the target activity.

6.2

The Basics of High-throughput Screens and Selections

Screening and selection methodologies should meet the following demands: (1) They should be, if possible, directly for the property of interest – “you get what you select for” is the first rule of directed evolution [7]. Thus, the substrate should be identical, or as close as possible to the target substrate, and product detection should be under multiple turnover conditions to ensure the selection of effective catalysts. (2) The assay should be sensitive over the desired dynamic range. The first rounds of any evolution experiments demand isolation with high recovery – all improved variants, including those that exhibit only several-fold improvement over the starting gene, should be recovered. The more advanced rounds must be performed at higher stringency to ensure the isolation of the best variants. A limited dynamic range seems to be the drawback of most selection approaches. (3) The procedure should be applicable in a high-throughput format.

Numerous assays enable the detection of enzymatic activities in agar colonies or crude cell lysates by the production of a fluorophore or chromophore (Chapter 5; see also [8–10]). Assays on agar-plated colonies typically enable the screening of $>10^4$ variants in a matter of days, but they are often limited in sensitivity: soluble products diffuse away from the colony and hence only very active variants are detected. Assays based on insoluble products have higher sensitivity, but their scope is rather limited (for example, see [11]). The range of assays that are applicable for crude cell lysates is obviously much wider, but their throughput is rather restricted: in the absence of sophisticated robotics that are usually unavailable to academic laboratories, only 10^3 – 10^4 variants are typically screened [12]. These low-to-medium throughput screens have certainly proved effective for the isolation of enzyme variants with improved properties as described in a number of recent reviews [4, 13, 14]. However, a far more efficient sampling of sequence space is required for the isolation of rare variants with dramatically altered phenotypes.

Whilst high-throughput selections for binding activity have become abundant [15–17], enzymatic selections remain a challenge. The main obstacle is that catalytic screens or selections demand a linkage between the gene, the encoded enzyme, and multiple product molecules (Figure 6.1) [18, 19]. This chapter focuses on methodologies that enable the selection of large libraries, typically well over 10^6 variants, for enzymatic activities. This throughput is still beyond the reach

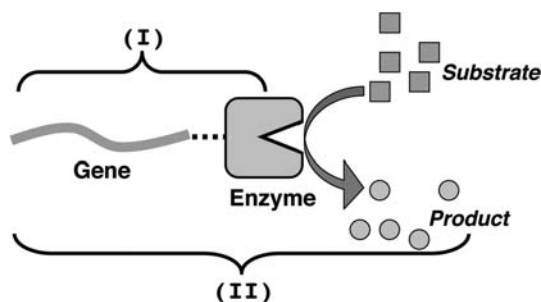


Fig. 6.1 The basis of all selection and screens for enzyme-coding genes, whether in a directed evolution or functional genomics context, is the linkage between the gene, the enzyme it encodes, and the products of the activity of that enzyme. The first link, between the enzyme and its coding gene (step I) can be achieved in a variety of ways, including the cloning and expression of the

library in living cells, phage display, ribosome and mRNA peptide display, and by cell-free translation in emulsion droplets (IVC). The second link, between the product of the enzymatic activity and the linking gene (step II), is usually harder to obtain. A variety of methods that provide the above link in a genuinely high-throughput format are discussed in the text.

of high-throughput technologies that are based on two-dimensional arrays and robotics (microplates, chips, etc.).

6.3

High-throughput Selection of Enzymes Using Phage Display

The display of the screened enzyme on the surface of bacteriophages has several advantages: first, the phage provides the link between the gene and the protein it encodes (Figure 6.1, step I); second, although library size is limited by transformation, 10^7 transformants can be obtained quite easily, and 10^{11} – 10^{12} are not beyond reach; third, display on the surface allows unhindered accessibility of the substrate and reaction conditions of choice (e.g. buffer, pH, metals). The main challenge with display systems is maintaining the linkage between the enzyme and the products of its activity (Figure 6.1, step II).

Previously, selections for enzymatic activities of either catalytic antibodies or enzymes displayed on phage were performed indirectly, by binding to transition state analogs (TSAs) or suicide inhibitors (for a recent review see [20]). In recent years, direct selections for product formation under single [21, 22] as well multiple turnovers have been achieved. A notable example is the selection of a synthetic antibody library for alkaline phosphatase activity [23]. Hydrolysis of a soluble substrate generated a product which acts as an electrophilic reagent and couples onto the phage particle that displays the catalytically active antibody variant. These phage particles were then captured by affinity chromatography to the product. Two rounds of selection from an initial library of $>10^9$ antibody variants yielded an antibody with catalytic efficiency (k_{cat}/K_m) that is 1000-fold

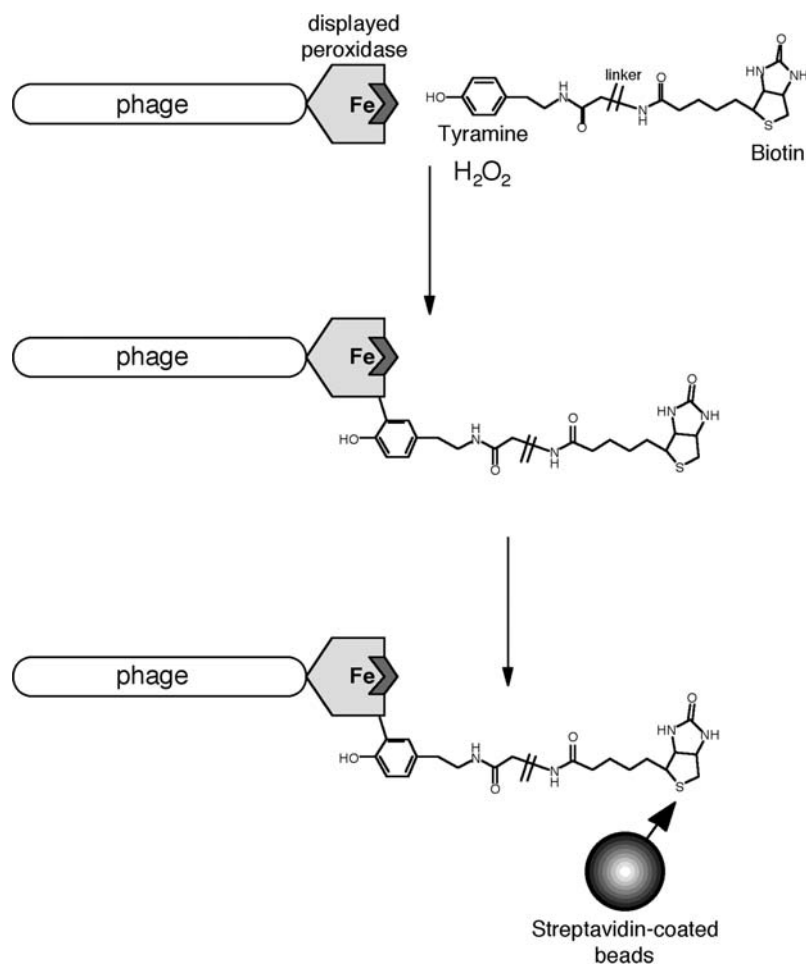


Fig. 6.2 Selection of phages displaying catalytic antibodies with peroxidase activity [24]. The selection was based on oxidation of the peroxidase substrate tyramine that is known to undergo rapid condensation with phenolic side-chains on protein surfaces upon oxida-

tion. The tyramine substrate was linked to biotin through a linker. As a result, phage particles displaying active antibodies were covalently modified and labeled with biotin, and could then be isolated by capture onto a streptavidin affinity support.

higher than other alkaline phosphatase-like antibodies generated by immunization with a transition state analog. This high catalytic efficiency was mainly the result of an increase in k_{cat} value, demonstrating the importance of selection for multiple turnovers.

Another example regards the selection of catalytic antibodies with peroxidase activity. Selection was based on a biotin-linked tyramine substrate being oxidized and coupled to the antibody's binding site. Phages displaying active antibody catalysts were isolated by binding to streptavidin (Figure 6.2) [24].

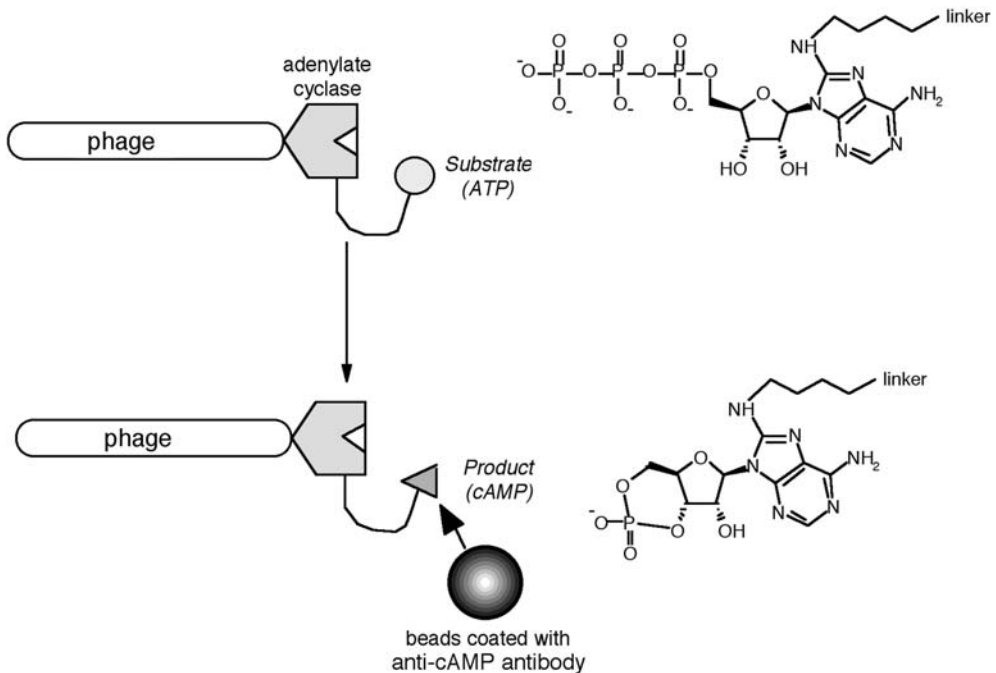


Fig. 6.3 Selection of enzymes by phage display and anti-product antibodies [27]. The ATP substrate was attached to a maleimide cross-linker and subsequently to phage particles displaying the selected enzymes. The proximity of the enzyme allows conversion of substrate to product on phage. Phage particles displaying active adenylate cyclase (the product) were then captured using an antibody that binds cAMP but not ATP.

Another strategy for selection by phage display is based on linking the substrate to the phage particle expressing the target enzyme. Active enzyme variants transform the phage-linked substrate to product, which remains attached to the phage, and the phage can be isolated by affinity chromatography to the product [21, 22, 25, 26]. The strategy was recently applied for a model selection of phages displaying adenylate cyclase from a large excess of phages that do not display the enzyme. The enzymatic conversion of a chemically linked ATP into cAMP was selected using immobilized anti-cAMP antibodies, and an enrichment factor of about 70-fold was demonstrated (Figure 6.3) [27]. A selection for DNA polymerases was also described [28, 29]. In this case, the substrate (a DNA primer) was also covalently attached via a flexible tether to the phage coat protein. Active variants were selected by virtue of elongation of the linked primer, first by incorporation of the target nucleotides, and finally by a biotinylated nucleotide that was used to capture the phage onto avidin-coated beads. Using this strategy, a DNA-dependent DNA polymerase was evolved to efficiently incorporate rNTPs and thereby function as a DNA-dependent RNA polymerase [28].

The same selection system was used to evolve DNA polymerases that incorporate 2'-O-methyl ribonucleosides with high efficiency [29].

Several groups applied a phage selection strategy termed "catalytic elution" for enzymes the catalytic activity of which depends on cofactor binding [21, 30]. Following protein expression on the phage surface, the cofactor (e.g. a metal) is removed and the catalytically inactive phages are bound to an immobilized substrate. The cofactor is then added, and phages displaying an active enzyme are eluted by conversion of the substrate into product.

6.4

High-throughput Selection of Enzymes Using Cell Display

The application of cell surface display for directed evolution has gained much momentum. In particular, screening by FACS (fluorescence-activated cell sorter) has yielded a number of highly potent binding proteins such as antibodies [15, 31]. Cell surface display can also be used to select for catalysis and has many features in common with phage display. However, sorting for enzymatic activity was thus far achieved only in few particular cases where the product of the enzymatic reaction could be captured within bacterial cells [32] or on their surface [33, 34]. The protease OmpA was displayed on the surface of *E. coli*, and a fluorescence resonance energy transfer (FRET) substrate was added that adheres to the surface of the bacteria. Enzymatic cleavage of the substrate releases a quencher group, and the resulting fluorescent cells were isolated by FACS. This approach allowed the screening of a library of $\sim 10^6$ OmpA variants and the subsequent isolation of a variant exhibiting a 60-fold improvement in catalytic activity [33]. A complementary approach was developed by Kolmar and co-workers, whereby the products of a surface-displayed enzyme (an esterase) was covalently linked to the cell surface using peroxidase-activated tyramide conjugation [18].

Another potentially useful screening approach is based on the immobilization of enzyme-displaying cells on beads [35]. The cells are adsorbed to polyacrylamide beads pre-equilibrated with growth medium to form a bead population containing, on average, a single cell per bead. The beads are immobilized on a solid glass support and cells are allowed to grow and form microcolonies while utilizing the medium retained within the hydrogel matrix. The microbead colonies are then equilibrated with a chromogenic or fluorogenic substrate, and the beads are screened under the microscope. This technique was recently demonstrated by enriching *E. coli* cells expressing β -lactamase on their surface from a large excess of other cells.

6.5

***In vivo* Genetic Screens and Selections**

Genetic, or *in vivo*, selections were the tools used for most early directed evolution experiments [36], but as the scope of targets for evolution widened, the utility of this approach became limited. After all, these selections are usually based on the evolving activity complementing an auxotroph strain in which an enzyme was knocked-out. Thus, broadly speaking, the target for evolution (substrate, reaction, etc.) needs to parallel an already existing enzyme. Nevertheless, genetic screens can still be useful, for example in the identification of promiscuous enzyme activities [37, 38]. Furthermore, the scope of genetic selections has been significantly widened by the application of the three-hybrid system to link the catalytic activity of an enzyme to the transcription of a reporter gene in yeast cells, as described in Chapter 7.

6.6

Screens for Heterologous Protein Expression and Stability

6.6.1

Introduction

The high levels of expression of stable and functional proteins remain the bottleneck for many enzyme applications. Conventional approaches for production of soluble and active proteins in heterologous expression systems include low-temperature expression, promoters with different strengths, modified growth media, and a variety of solubility-enhancing fusion tags (reviewed in [39–42]). More recently, a series of vectors and fusion partners that can be screened for high-level functional expression of a target protein have been developed [43]. These approaches are often successful but, at the end of the day, they cannot modify the intrinsic properties of the protein. Thus, even when a soluble fraction of the protein is obtained, the protein may still be inactive, or aggregate, during purification and storage.

An alternative to expression optimization is protein engineering by rational design (reviewed in [44]). There are now several examples of proteins that have been stabilized by the introduction of mutations with small yet cumulative stabilizing effects [45–47]. It is largely unknown, however, how the stability of a protein is encoded in its sequence, and how individual amino acid changes contribute to stability [48], not to mention that most of the interesting targets for structural and mechanistic studies are obviously of unknown structure and are therefore not amenable to rational design.

This section focuses on the application of directed evolution to obtain efficient heterologous expression of proteins in *E. coli* and yeast, by increasing the stability and solubility of the protein in the host's environment (Table 6.1) (for reviews on related issues, see [49, 50]).

Table 6.1 Recent examples of directed evolution of enzymes for heterologous expression, solubility or stability.^{a)}

Target protein	Library origin	Screening system	Results ^{b)}		Ref.
			Property	Improvement	
1. Galactose oxidase from fungi	EP/StEP	Liquid bacterial cultures, assayed for the native enzymatic activity	Functional expression, long-term stability (residual activity in presence of CuSO ₄ and catalase)	18-fold, 60% higher long-term stability, and twofold higher k_{cat}/K_m	[87]
2. Human cytochrome P450 (1A2) and bacterial P450 (BM3)	SHIPREC (16% homology)	CAT preselection, and screening of colonies with 1A2's substrate	Functional expression	From <1.5% to 14% soluble protein in the cytosolic fraction	[88]
3. N-Carbamyl-D-amino acid amidohydrolase from <i>Agrobacterium tumefaciens</i>	DS	Filter screening of colonies for native activity, after incubation at high temperature, or with oxidizing reagent	Thermal and oxidative stability (residual activity after heat or oxidizer)	Eightfold (incubation at 70 °C) and 16-fold (incubation with H ₂ O ₂)	[89]
4. Methyl transferase (MT), nucleoside diphosphate kinase (NDP-K), and tartrate dehydratase β (TD- β) from <i>P. aerophilum</i>	DS	Fluorescence screening of <i>E. coli</i> colonies displaying correctly folded protein (GFP reporter)	Functional expression	From insoluble to 50–95% soluble expression. Crystal structure determination for NDP-K	[57]
5. Barnase from <i>Bacillus amyloliquefaciens</i>	17 positions diversified by oligo-nucleotide assembly	Proteolytic selection by phage display, affinity selection for binding to ligand	Thermodynamic stability (free energy of folding)	Similar or lower than wild-type barnase	[90]
6. Laccase from <i>Myceliophthora thermophila</i> fungus (MtL)	EP/ <i>in vivo</i> shuffling in yeast/ <i>in vivo</i> gap repair/StEP	Screening yeast cultures for native enzymatic activity	Functional expression	8–9-fold higher expression and 15–30-fold higher k_{cat}	[53]

Table 6.1 (continued)

Target protein	Library origin	Screening system	Results ^{b)}		Ref.
			Property	Improvement	
7. Glucose dehydrogenase (GlcDH) from <i>Bacillus</i> species	FS of three wild-type genes and one improved mutant	Filter-based screening and selection for enzymatic activity after incubation at high temperature	Thermal stability in the absence of NaCl	415-fold higher $t_{1/2}$ at high temperature (66 °C)	[67]
8. Mammalian serum para-oxonase (PON1)	FS of four mammalian PON1 genes/DS	Agar colony screens for a promiscuous esterase activity	Functional expression	>100-fold in soluble, active protein. Crystal structure determination of the first PON family member	[68, 91]
9. Lipase B from <i>Candida antarctica</i> (CALB)	<i>In vitro</i> and <i>in vivo</i> FS with two thermostable bacterial lipases B	Screening yeast cultures for native enzymatic activity at ambient temperature	Thermal stability and activity	12-fold higher $t_{1/2}$ at 45 °C, 11% higher T_m , 16-fold higher k_{cat}/K_m	[92]
10. Phosphotriesterase from <i>Pseudomonas diminuta</i> (PTE)	EP or BA/DS	Agar colony screens for a promiscuous esterase activity	Functional expression	~20-fold higher content in soluble fraction, owing to increased stability of the apo-enzyme	[70]

- a) The examples shown (in chronological order) use established methods of sequence diversification in either isolation or combination: error-prone PCR (EP); base analogs-induced mutagenesis (BA); saturation mutagenesis (SM); DNA shuffling (DS); family shuffling (FS); staggered extension process (StEP); sequence homology-independent protein recombination (SHIPREC).
- b) Note that the varying property was not part of the screen or selection in some cases. The improvement factor is based on the best clone(s) reported in terms of functional expression, stability, or solubility.

6.6.2

Screening Methodologies for Heterologous Expression

Heterologous expression is performed primarily in *E. coli*, although yeast and insect cells are becoming increasingly popular, particularly for eukaryotic proteins that require posttranslational processing [51, 52]. Considerations of transformation efficiency, stability of plasmid DNA, and growth rate, make *E. coli* and *S. cerevisiae* the best-suited organisms for directed evolution experiments [53].

Directed evolution in the laboratory, like natural evolution, involves two key steps: generation of genetic diversity and selection for function. In the laborato-

ry, diversity in the gene of interest is typically created by random mutagenesis or family shuffling (for recent reviews see [5, 54, 55]). Whilst the methods of creating genetic diversity are generally applicable, the selection system needs to be tailored, or modified, for each target protein and aim – a selection for higher expression may not be suitable for altering substrate selectivity and vice versa. Broadly speaking, genetic screens or selections for heterologous expression can be done in two ways: (1) screening/selecting for the activity of a reporter protein, or (2) screening/selecting for the protein's own function. Examples of both approaches are discussed below.

A variety of generic “C-fusion” approaches have been developed that rely on expression of the target protein as an N-terminal fusion to a reporter protein with a selectable function. An insoluble target protein leads to the aggregation of the reporter protein and loss of its function [49]. As an example, colonies overexpressing green fluorescent protein (GFP) fused to a soluble protein can be easily distinguished owing to folding and fluorophore formation of the GFP – whereas those expressing an insoluble protein do not fluoresce [56, 57].

Another generally applicable method relies on chloramphenicol acetyltransferase (CAT) fusion, whereby selection for solubility of the target protein is performed by antibiotic resistance [58]. Using this methodology, Arnold and coworkers converted the membrane-associated, insoluble human P450 cytochrome (1A2) into a fairly soluble protein by recombination of 1A2 with a distant bacterial homolog (Table 6.1) [88]. The resulting library of hybrid proteins was first selected by fusion to CAT, in order to isolate all in-frame and correctly folded variants. A subsequent screen for active enzyme variants was performed with a fluorogenic P450 substrate.

In the β -Gal complementation assay, a small α -fragment of β -galactosidase (LacZ) of about 50 amino acids is fused to the C-terminus of the target protein, restoring – in those cases where the target protein remains soluble – the β -galactosidase activity of a truncated LacZ (LacZ- Ω) by complementation in trans [59]. Other potentially generic methodologies are based on screening by cellular stress responses to misfolded proteins [60, 61], “proteolytic selection” by phage display [62], or “protein stability increased by directed evolution” (Proside) [48, 63], the last two technologies being based on the principle that infectivity of the phage is coupled to protease resistance of the protein variants. This assumption in turn relies on the observation that proteolysis resistance can be used as a marker for folding [64].

Because the “C-fusion” techniques do not depend on the function of the target protein, they are generally applicable, and particularly advantageous when attempting to evolve new folds, or in cases where the target protein has no assigned function. But a screen for soluble expression can turn into a major drawback. The blind fashion by which evolution acts (“you get what you select for” is the first rule of directed evolution [7]) implies that selection for soluble expression may be accompanied by significant changes in activity that may even render the soluble protein nonfunctional [57]. To drive the selection of soluble and stable variants without compromising the protein's functional properties, the selection system must be based on the protein's own activity. A large variety of assays are available for screening enzyme libraries [9, 65, 66]. Indeed, functional

assays have been used in a number of cases. As an example, Baik and co-workers have evolved glucose dehydrogenase (GlcDH), to increase thermal stability 400-fold by screening a library generated by family shuffling for the enzyme's activity (Table 6.1, entry 7) [67]. Several other examples are listed in Table 6.1 (entries 1–3, 5–9).

6.6.3

Directed Evolution for Heterologous Expression – Recent Examples

Recent examples of directed evolution for heterologous expression illustrate the power and versatility of this approach in facilitating structural and mechanistic studies. Directed evolution has thus far been the key to two crystal structures of general interest: a hyperthermophile nucleotide diphosphate kinase (NDP-K) [57], and the first PON family member [68]. Using a GFP “C-fusion” screen, Waldo and coworkers evolved the essentially insoluble NDP-K from *P. aerophilum*, into a 90% soluble and active variant that enabled the determination of its crystal structure (Table 6.1, entry 4) [57].

Serum paraoxonases (PONs) provide a particularly interesting target for directed evolution. PONs comprise a family of enzymes that play a key role in organophosphate (OP) detoxification and in preventing atherosclerosis. PONs are widely spread in mammals and other vertebrates, as well as in invertebrates. Three subfamilies of PON are known (PON1, 2 and 3) to show around 60% sequence homology, but the structure of not a single family member was known prior to the directed evolution of PON recombinant variants. Attempts to crystallize serum-purified PON1 resulted in the crystallization of a contaminant protein, instead of PON1 [69]. Moreover, although PONs exhibit a range of hydrolytic activities towards esters, phosphotriesters, and lactones, their native or physiological substrate is unknown.

Family shuffling and screening for esterase activity led to the first PON1 and PON3 variants that express in a soluble and active form in *E. coli* (Table 6.1, entry 8). However, in the absence of knowledge of the native substrate, the screen was based not on PON's native function, but on its promiscuous esterase and phosphotriesterase activities. Interestingly, whilst evolution of PON1 led to a variant with wild type-like enzymatic properties and high *E. coli* expression, evolution of PON3 by the very same method led to a mutant with increased esterase activity relative to wild-type PON3, in addition to higher expression.

The promiscuous esterase activity was also used for evolving a bacterial phosphotriesterase (PTE) [70] (Table 6.1, entry 10). In this case, libraries created by random mutagenesis of the PTE gene were screened for higher esterase activity. However, the attempt to evolve this promiscuous activity led to a variant with a 20-fold increase in functional expression, and enzymatic properties that are essentially identical to wild-type PTE.

From a methodological point of view, these works highlight the advantages and drawbacks of using screens based on a promiscuous activity, particularly when the native substrate is unknown (e.g. in PON's case), when a facile screen

for the native substrate is not available, or when the detection of the native activity falls out of the dynamic range of the screen (e.g. in PTE's case). The snag is that this approach can obviously lead to variants with increased promiscuous activity as well as, or instead of, increased expression.

Some proteins, especially eukaryotic ones, are notoriously difficult to express in heterologous systems. Arnold and coworkers [53] have succeeded in obtaining high functional expression levels for fungal MtL laccase – a glycoprotein – along with higher activity, by directed evolution [53] (Table 6.1, entry 6). For this purpose, they used a yeast host, *Saccharomyces cerevisiae*, and a high-throughput screen based on the enzyme's native activity. Both *in vitro* and *in vivo* shuffling were used to achieve random recombination of the isolated variants between rounds. In addition, polymerase chain reaction (PCR) and gap repair tools [71, 72] were used to ensure the recombination of neighboring mutations in a site-directed manner.

Directed evolution also provided some general insights into the factors that dictate protein solubility, stability, and assembly in non-native environments. It is widely assumed that protein expression levels are determined by the “solubility” or “stability” of the expressed protein. This rationale is reflected in most strategies that have been designed to improve overexpression yields. However, recent works point to a key factor that is often overlooked. The main determinant of (active) overexpression for many proteins is defined by the thermodynamic and kinetic properties of an intermediate, whether an apo-enzyme, pro-protein, monomeric state, or a folding (on- or off-pathway) intermediate, rather than by the stability and solubility of the final state [73]. Because these expression bottlenecks are largely unidentified, the design of stabilizing mutations may prove a daunting task. Evolution, which is directed only by the phenotypic outcome of the process and requires no knowledge of the protein's structure and folding pathway, seems like an effective solution to this problem. Care must be taken, however, that selection is performed in a controlled manner that improves functional expression but does not alter the protein's function and structure.

6.7

In vitro Compartmentalization

In vitro compartmentalization (IVC) is based on water-in-oil emulsions, where the water phase is dispersed in the oil phase to form microscopic aqueous compartments. Each droplet contains, on average, a single gene, and serves as an artificial cell in allowing for transcription, translation, and the activity of the resulting proteins, to take place within the compartment. The oil phase remains largely inert and restricts the diffusion of genes and proteins between compartments. The droplet volume (~ 5 fL) enables a single DNA molecule to be transcribed and translated [74], as well as the detection of single enzyme molecules [75]. The high capacity of the system ($>10^{10}$ in 1 mL of emulsion), the ease of preparing emulsions, and their high stability over a broad range of temperatures, render IVC an attractive system for enzyme high-throughput screening.

IVC makes it easy to co-compartmentalize genes and the proteins they encode, but the selection of an enzymatic activity requires a link between the desired reaction product and the gene (Figure 6.1). One possible selection format is to have the substrate, and subsequently the product, of the desired enzymatic activity physically linked to the gene. Enzyme-encoding genes can then be isolated by virtue of their attachment to the product while other genes, which encode an inactive protein, carry the unmodified substrate. The simplest applications of this strategy lie in the selection of DNA-modifying enzymes where the gene and substrate comprise the same molecule. Indeed, IVC was first applied for the selection of DNA methyltransferases (MTases) [74]. Selection was performed by extracting the genes from the emulsion and subjecting them to digestion by a cognate restriction enzyme that cleaves the nonmethylated DNA [74, 76, 77]. Other applications include the selection of restriction endonucleases [78] and DNA polymerases [79, 80]. The selection of DNA polymerases was based on the fact that inactive variants failed to amplify their own genes and therefore disappeared from the library pool.

These modes of selection are obviously restricted not only in the scope of the selected enzymatic activities but also by the stoichiometry – typically one gene, and hence one substrate molecule (or several substrate sites), is present per droplet, together with 10–100 enzyme molecules. Despite these restrictions, several new and interesting enzyme variants have been evolved by IVC. A variant of *HaeIII* MTase was evolved with up to 670-fold improvement in catalytic efficiency for a nonpalindromic target sequence (AGCC) and ninefold improvement for the original recognition site (GGCC) [77]. Active *FokI* restriction nuclease variants were also selected by IVC from a large library of mutants with low residual activity [78]. And, DNA polymerase variants with increased thermal stability [79], and the ability to incorporate a diverse range of bases including fluorescent dye-labeled nucleotides have been evolved [80].

The first application of IVC beyond DNA-modifying enzymes was demonstrated by a selection of bacterial phosphotriesterase variants [75]. The selection strategy was based on two emulsification steps: in the first step, microbeads, each displaying a single gene and multiple copies of the encoded protein variant, were formed by translating genes immobilized to microbeads in emulsion droplets and capturing the resulting protein via an affinity tag. The microbeads were isolated and re-emulsified in the presence of a modified phosphotriester substrate (Figure 6.4). The product and any unreacted substrate were subsequently coupled to the beads. Product-coated beads, displaying active enzymes and the genes that encode them, were detected with fluorescently-labeled anti-product antibodies and selected by FACS. Selection from a library of $>10^7$ different variants led to the isolation of a variant with a very high k_{cat} value ($>10^5 \text{ s}^{-1}$). Microbead display libraries formed by IVC can also be selected for binding activity [81] (for other binding selection by IVC see [82, 83]).

Some of the IVC selection modes take advantage of the fact that this system is purely *in vitro*, and can allow selection for substrates, products, and reaction conditions that are incompatible with *in vivo* systems. However, the cell-free translation must be performed under defined pH, buffer, ionic strength, and

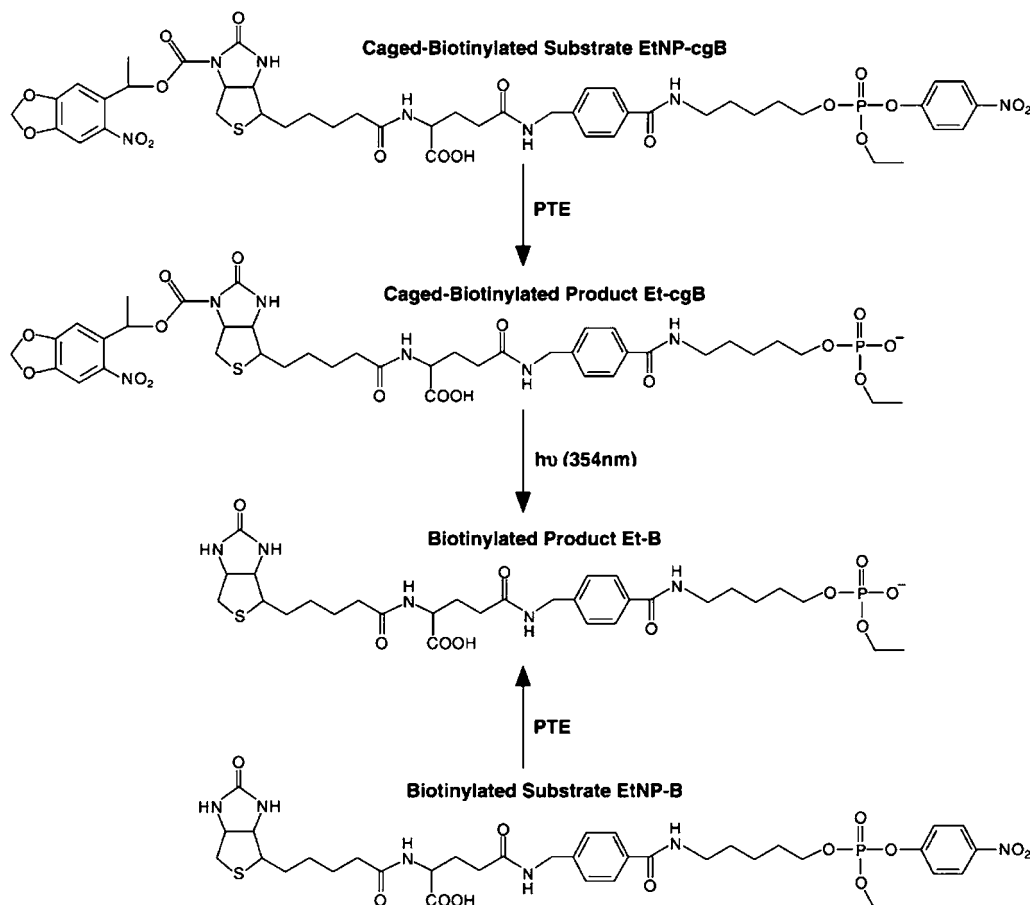


Fig. 6.4 The application of caged substrates for selection by *in vitro* compartmentalization (IVC) [75]. Paraoxon (the substrate of the target enzyme) was modified by substituting an ethyl group with a linker connected to caged biotin. The hydrolysis of the resulting substrate (EtNP-cgB) by the selected enzyme (PTE, or *Pseudomonas diminota* phosphotriesterase; see text) gives *p*-nitrophenol and the corresponding phos-

phodiester Et-cgB. Irradiation at 354 nm releases the caging group and carbon dioxide to yield the uncaged biotinylated substrate (EtNP-B) or product (Et-B). A similar chemistry can be used in other screens or selections, where the reaction can take place between a soluble enzyme and soluble substrate, and the resulting product is subsequently immobilized for detection by anti-product antibodies [93].

metal ion composition. In the selection for phosphotriesterase using IVC described above [19], translation is completely separated from catalytic selection by using two sequential emulsification steps, allowing selection for catalysis under conditions which are incompatible with translation. However, it is also possible to use a single emulsification step, and to modify the content of the droplets without breaking the emulsion once translation is completed. There are cur-

rently several ways of modulating the emulsion content without affecting its integrity. These include the delivery of hydrophobic substrates through the oil phase, reduction of the droplet's pH by delivery of acid, and photoactivation of a substrate contained within the aqueous droplets [19]. More recently, a nanodroplet delivery system has been developed that allows the transport of various solutes, including metal ions, into the emulsion droplets. This transport mechanism was applied for the selection of DNA nuclease inhibitors. Inactive DNA nucleases were co-compartmentalized with a gene-library, and once translation has been completed, the nuclease was activated by delivery of nickel or cobalt ions. Genes encoding nuclease inhibitors survived the digestion and were subsequently amplified and isolated. Selection was therefore performed directly for inhibition, and not for binding of the nuclease [84].

6.8 IVC in Double Emulsions

The need to link the product to the enzyme-coding gene complicates and restricts the scope of selection especially for non-DNA-modifying enzymes. Recently, an alternative strategy has been developed based on compartmentalizing and sorting single genes, together with the fluorescent product molecules generated by their encoded enzymes. The technology makes use of double, water-in-oil-in-water (w/o/w) emulsions that are amenable to sorting by FACS (Figure 6.5). This circumvents the need to tailor the selection for each substrate and reaction, and allows the use of a wide variety of existing fluorogenic substrates (e.g. see Chapter 9). Recently it has been shown that the making and sorting of w/o/w emulsion droplets does not disrupt the content of the aqueous droplets of the primary w/o emulsions. Further, sorting by FACS of w/o/w emulsion droplets containing a fluorescent marker and parallel gene enrichments have been demonstrated [85]. More recently, w/o/w emulsions were applied for the directed evolution of two different enzymatic systems: new variants of serum paraoxonase (PON1) with thiolactonase activity (Aharoni et al., personal communication), and new enzyme variants with β -galactosidase activity (Mastrobatista et al., personal communication) were selected from libraries of $>10^7$ mutants. The β -galactosidase variants were translated *in vitro* as with previously described IVC selections [19, 74]. In the case of PON1, intact *E. coli* cells in which the library variants were expressed, were emulsified and FACS sorted, thus demonstrating the applicability of double emulsions for single-cell phenotyping and directed enzyme evolution.

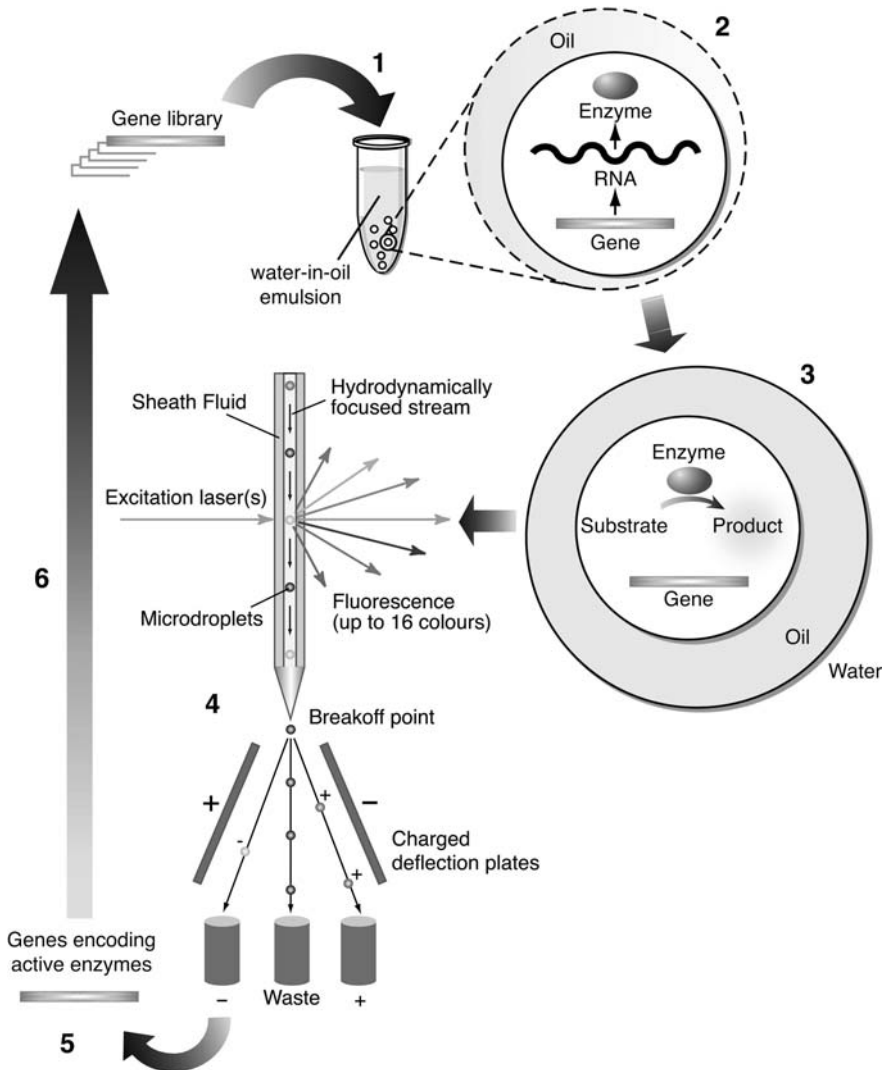


Fig. 6.5 Selections by flow sorting of double emulsion microdroplets in using a fluorescence-activated cell sorter (FACS). A library of genes, each encoding a different enzyme variant, is dispersed to form a water-in-oil (w/o) emulsion with typically one gene per aqueous microdroplet (1). The genes are transcribed and translated within their microdroplets (2), using either *in vitro* (cell-free) transcription/translation, or by compartmentalizing single cells (e.g. bacteria into which the gene library is cloned) in the microdroplets. Proteins with enzymatic

activity convert the nonfluorescent substrate into a fluorescent product and the w/o emulsion is converted into a water-in-oil-in-water (w/o/w) emulsion (3). Fluorescent microdroplets are separated from nonfluorescent microdroplets (or microdroplets containing differently colored fluorochromes) using FACS (4). Genes from fluorescent microdroplets, which encode active enzymes, are recovered and amplified (5). These genes can be recompartimentalized for further rounds of selection (6).

6.9

Concluding Remarks

The past few years have seen significant progress in the development and application of high-throughput screens for enzymatic activities. Most significantly, a variety of methods have become available for the direct selection of the target enzymatic activity under multiple turnover conditions. The maturity of this area is also reflected in the shift from model selections (i.e. demonstrating the isolation of an existing enzyme-coding gene from a large excess of irrelevant genes) to library selections, and the subsequent isolation of a variety of new enzyme variants [6]. Cytoplasmic, periplasmic (and periplasmic-anchored [86]) expression, phage and cell display, cell-free translation in emulsion droplets (IVC) and other *in vitro* display technologies such as ribosome or mRNA peptide display [16], were all shown to provide an effective means of linking genes to their encoded proteins (Figure 6.1, step I). However, linking the gene to the products of the desired enzymatic activity (Figure 6.1, step II) still needs to be specifically tailored for each enzyme, reaction, and substrate.

The use of fluorogenic substrates and sorting by FACS appears to be the most promising avenue. Modern FACS machines routinely analyze and sort $>10^7$ events per hour. Fluorescence is a sensitive, versatile, and general signal that can detect a huge range of enzymatic reactions. And, unlike selections performed in bulk, FACS allows the stringency and recovery from the selection to be fine-tuned [18, 34]. Thus, the combination of FACS, display technologies, and compartmentalization in double emulsions holds much potential in the areas of directed enzyme evolution and functional genomics.

References

- 1 F. H. Arnold, P. L. Wintrode, K. Miyazaki, A. Gershenson, *Trends Biochem. Sci.* **2001**, *26*, 100–106.
- 2 H. Tao, V. W. Cornish, *Curr. Opin. Chem. Biol.* **2002**, *6*, 858–864.
- 3 H. Lin, V. W. Cornish, *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 4402–4425.
- 4 P. A. Dalby, *Curr. Opin. Struct. Biol.* **2003**, *13*, 500–505.
- 5 C. Neylon, *Nucleic Acids Res.* **2004**, *32*, 1448–1459.
- 6 S. Lutz, W. M. Patrick, *Curr. Opin. Biotechnol.* **2004**, *15*, 291–297.
- 7 C. Schmidt-Dannert, F. H. Arnold, *Trends Biotechnol.* **1999**, *17*, 135–136.
- 8 U. T. Bornscheuer, *Curr. Opin. Biotechnol.* **2002**, *13*, 543–547.
- 9 J. P. Goddard, J. L. Reymond, *Trends Biotechnol.* **2004**, *22*, 363–370.
- 10 J. P. Goddard, J. L. Reymond, *Curr. Opin. Biotechnol.* **2004**, *15*, 314–322.
- 11 V. Khalameyzer, I. Fischer, U. T. Bornscheuer, J. Altenbuchner, *Appl. Environ. Microbiol.* **1999**, *65*, 477–482.
- 12 M. L. Geddie, L. A. Rowe, O. B. Alexander, I. Matsumura, *Methods Enzymol.* **2004**, *388*, 134–145.
- 13 S. Panke, M. Held, M. Wubbolts, *Curr. Opin. Biotechnol.* **2004**, *15*, 272–279.
- 14 N. J. Turner, *Trends Biotechnol.* **2003**, *21*, 474–478.
- 15 A. Hayhurst, G. Georgiou, *Curr. Opin. Chem. Biol.* **2001**, *5*, 683–689.
- 16 P. Amstutz, P. Forrer, C. Zahnd, A. Pluckthun, *Curr. Opin. Biotechnol.* **2001**, *12*, 400–405.
- 17 T. T. Takahashi, R. J. Austin, R. W. Roberts, *Trends Biochem. Sci.* **2003**, *28*, 159–165.

- 18 S. Becker, H.U. Schmoldt, T.M. Adams, S. Wilhelm, H. Kolmar, *Curr. Opin. Biotechnol.* **2004**, *15*, 323–329.
- 19 A.D. Griffiths, D.S. Tawfik, *Curr. Opin. Biotechnol.* **2000**, *11*, 338–353.
- 20 A. Fernandez-Gacio, M. Uguen, J. Fastrez, *Trends Biotechnol.* **2003**, *21*, 408–414.
- 21 H. Pedersen, S. Holder, D.P. Sutherlin, U. Schwitter, D.S. King, P.G. Schultz, *Proc. Natl Acad. Sci. USA* **1998**, *95*, 10523–10528.
- 22 S. Atwell, J.A. Wells, *Proc. Natl Acad. Sci. USA* **1999**, *96*, 9497–9502.
- 23 S. Cesaro-Tadic, D. Lagos, A. Honegger, J.H. Rickard, L.J. Partridge, G.M. Blackburn, A. Pluckthun, *Nat. Biotechnol.* **2003**, *21*, 679–685.
- 24 J. Yin, J.H. Mills, P.G. Schultz, *J. Am. Chem. Soc.* **2004**, *126*, 3006–3007.
- 25 S. Demartis, A. Huber, F. Viti, L. Lozzi, L. Giovannoni, P. Neri, G. Winter, D. Neri, *J. Mol. Biol.* **1999**, *286*, 617–633.
- 26 J.L. Jestin, P. Kristensen, G. Winter, *Angew. Chem. Int. Ed. Engl.* **1999**, *38*, 1124–1127.
- 27 H. Strobel, D. Ladant, J.L. Jestin, *J. Mol. Biol.* **2003**, *332*, 1–7.
- 28 G. Xia, L. Chen, T. Sera, M. Fa, P.G. Schultz, F.E. Romesberg, *Proc. Natl Acad. Sci. USA* **2002**, *99*, 6597–6602.
- 29 M. Fa, A. Radeghieri, A.A. Henry, F.E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 1748–1754.
- 30 I. Ponsard, M. Galleni, P. Soumillion, J. Fastrez, *ChemBioChem* **2001**, *2*, 253–259.
- 31 M.J. Feldhaus, R.W. Siegel, L.K. Opresko, J.R. Coleman, J.M.W. Feldhaus, Y.A. Yeung, J.R. Cochran, P. Heinzelman, D. Colby, J. Swers, C. Graff, H.S. Wiley, K.D. Wittrup, *Nat. Biotechnol.* **2003**, *21*, 163–170.
- 32 Y. Kawarasaki, K.E. Griswold, J.D. Stevenson, T. Selzer, S.J. Benkovic, B.L. Iverson, G. Georgiou, *Nucleic Acids Res.* **2003**, *31*, e126.
- 33 M.J. Olsen, D. Stephens, D. Griffiths, P. Daugherty, G. Georgiou, B.L. Iverson, *Nat. Biotechnol.* **2000**, *18*, 1071–1074.
- 34 M. Olsen, B. Iverson, G. Georgiou, *Curr. Opin. Biotechnol.* **2000**, *11*, 331–337.
- 35 A. Freeman, N. Cohen-Hadar, S. Abramov, R. Modai-Hod, Y. Dror, G. Georgiou, *Biotechnol. Bioeng.* **2004**, *86*, 196–200.
- 36 B.G. Hall, *Biochemistry* **1981**, *20*, 4042–4049.
- 37 K. Yang, W.W. Metcalf, *Proc. Natl Acad. Sci. USA* **2004**, *101*, 7919–7924.
- 38 B.G. Miller, R.T. Raines, *Biochemistry* **2004**, *43*, 6387–6392.
- 39 S.C. Makrides, *Microbiol. Rev.* **1996**, *60*, 512–538.
- 40 P. Braun, J. LaBaer, *Trends Biotechnol.* **2003**, *21*, 383–388.
- 41 G. Marsischky, J. LaBaer, *Genome Res.* **2004**, *14*, 2020–2028.
- 42 J. Pearlberg, J. LaBaer, *Curr. Opin. Chem. Biol.* **2004**, *8*, 98–102.
- 43 D.A. Berthold, P. Stenmark, P. Nordlund, *Protein Sci.* **2003**, *12*, 124–134.
- 44 V.G. Eijsink, A. Bjork, S. Gaseidnes, R. Sirevag, B. Synstad, B. van den Burg, G. Vriend, *J. Biotechnol.* **2004**, *113*, 105–120.
- 45 P.L. Wintrode, F.H. Arnold, *Adv. Protein Chem.* **2000**, *55*, 161–225.
- 46 M. Lehmann, C. Loch, A. Middendorf, D. Studer, S.F. Lassen, L. Pasamontes, A.P. van Loon, M. Wyss, *Protein Eng.* **2002**, *15*, 403–411.
- 47 S. D'Amico, C. Gerday, G. Feller, *J. Mol. Biol.* **2003**, *332*, 981–988.
- 48 A. Martin, V. Sieber, F.X. Schmid, *J. Mol. Biol.* **2001**, *309*, 717–726.
- 49 G.S. Waldo, *Curr. Opin. Chem. Biol.* **2003**, *7*, 33–38.
- 50 T.J. Magliery, L. Regan, *Eur. J. Biochem.* **2004**, *271*, 1595–1608.
- 51 E.T. Boder, K.D. Wittrup, *Methods Enzymol.* **2000**, *328*, 430–444.
- 52 D.W. Colby, B.A. Kellogg, C.P. Graff, Y.A. Yeung, J.S. Swers, K.D. Wittrup, *Methods Enzymol.* **2004**, *388*, 348–358.
- 53 T. Bulter, M. Alcalde, V. Sieber, P. Meinhold, C. Schlachtbauer, F.H. Arnold, *Appl. Environ. Microbiol.* **2003**, *69*, 987–995.
- 54 F.H. Arnold, G. Georgiou (eds), *Directed Enzyme Evolution: Screening and Selection Methods*, Humana Press, Clifton, NJ, **2003**.
- 55 E.T. Farinas, T. Bulter, F.H. Arnold, *Curr. Opin. Biotechnol.* **2001**, *12*, 545–551.
- 56 G.S. Waldo, B.M. Standish, J. Berendzen, T.C. Terwilliger, *Nat. Biotechnol.* **1999**, *17*, 691–695.

- 57 J. D. Pedelacq, E. Piltch, E. C. Liong, J. Berendzen, C. Y. Kim, B. S. Rho, M. S. Park, T. C. Terwilliger, G. S. Waldo, *Nat. Biotechnol.* **2002**, *20*, 927–932.
- 58 K. L. Maxwell, A. K. Mittermaier, J. D. Forman-Kay, A. R. Davidson, *Protein Sci.* **1999**, *8*, 1908–1911.
- 59 W. C. Wigley, R. D. Stidham, N. M. Smith, J. F. Hunt, P. J. Thomas, *Nat. Biotechnol.* **2001**, *19*, 131–136.
- 60 S. A. Lesley, J. Graziano, C. Y. Cho, M. W. Knuth, H. E. Klock, *Protein Eng.* **2002**, *15*, 153–160.
- 61 I. Vostiar, J. Tkac, C. F. Mandenius, *J. Biotechnol.* **2004**, *111*, 191–201.
- 62 P. Kristensen, G. Winter, *Fold. Des.* **1998**, *3*, 321–328.
- 63 V. Sieber, A. Pluckthun, F. X. Schmid, *Nat. Biotechnol.* **1998**, *16*, 955–960.
- 64 D. A. Parsell, R. T. Sauer, *J. Biol. Chem.* **1989**, *264*, 7590–7595.
- 65 A. Aharoni, A. D. Griffiths, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **2005**, *9*, 210–216.
- 66 U. T. Bornscheuer, M. Pohl, *Curr. Opin. Chem. Biol.* **2001**, *5*, 137–143.
- 67 S. H. Baik, T. Ide, H. Yoshida, O. Kagami, S. Harayama, *Appl. Microbiol. Biotechnol.* **2003**, *61*, 329–335.
- 68 M. Harel, A. Aharoni, L. Gaidukov, B. Brumshtein, O. Khersonsky, R. Meged, H. Dvir, R. B. Ravelli, A. McCarthy, L. Toker, I. Silman, J. L. Sussman, D. S. Tawfik, *Nat. Struct. Mol. Biol.* **2004**, *11*, 412–419.
- 69 A. Fokine, R. Morales, C. Contreras-Martel, P. Carpentier, F. Renault, D. Rochu, E. Chabriere, *Acta Crystallogr. D Biol. Crystallogr.* **2003**, *59*, 2083–2087.
- 70 C. Roodveldt, D. S. Tawfik, *Protein Eng. Des. Sel.* **2005**, *18*, 51–58.
- 71 D. Muhlrads, R. Hunter, R. Parker, *Yeast* **1992**, *8*, 79–82.
- 72 H. Ma, S. Kunes, P. J. Schatz, D. Botstein, *Gene* **1987**, *58*, 201–216.
- 73 W. C. Wigley, M. J. Corbo, T. D. Cutler, P. H. Thibodeau, J. Oldan, M. G. Lee, J. Rizo, J. F. Hunt, P. J. Thomas, *Nat. Struct. Biol.* **2002**, *9*, 381–388.
- 74 D. S. Tawfik, A. D. Griffiths, *Nat. Biotechnol.* **1998**, *16*, 652–656.
- 75 A. D. Griffiths, D. S. Tawfik, *EMBO J.* **2003**, *22*, 24–35.
- 76 Y. F. Lee, D. S. Tawfik, A. D. Griffiths, *Nucleic Acids Res.* **2002**, *30*, 4937–4944.
- 77 H. M. Cohen, D. S. Tawfik, A. D. Griffiths, *Protein Eng. Des. Sel.* **2004**, *17*, 3–11.
- 78 N. Doi, S. Kumadaki, Y. Oishi, N. Matsuura, H. Yanagawa, *Nucleic Acids Res.* **2004**, *32*, e95.
- 79 F. J. Ghadessy, J. L. Ong, P. Holliger, *Proc. Natl Acad. Sci. USA* **2001**, *98*, 4552–4557.
- 80 F. J. Ghadessy, N. Ramsay, F. Boudsocq, D. Loakes, A. Brown, S. Iwai, A. Vaisman, R. Woodgate, P. Holliger, *Nat. Biotechnol.* **2004**, *22*, 755–759.
- 81 A. Sepp, D. S. Tawfik, A. D. Griffiths, *FEBS Lett.* **2002**, *532*, 455–458.
- 82 M. Yonezawa, N. Doi, Y. Kawahashi, T. Higashinakagawa, H. Yanagawa, *Nucleic Acids Res.* **2003**, *31*, e118.
- 83 N. Doi, H. Yanagawa, *FEBS Lett.* **1999**, *457*, 227–230.
- 84 K. Bernath, S. Magdassi, D. S. Tawfik, *J. Mol. Biol.* **2005**, *345*, 1015–1026.
- 85 K. Bernath, M. Hai, E. Mastrobattista, A. D. Griffiths, S. Magdassi, D. S. Tawfik, *Anal. Biochem.* **2004**, *325*, 151–157.
- 86 B. R. Harvey, G. Georgiou, A. Hayhurst, K. J. Jeong, B. L. Iverson, G. K. Rogers, *Proc. Natl Acad. Sci. USA* **2004**, *101*, 9193–9198.
- 87 L. Sun, I. P. Petrounia, M. Yagasaki, G. Bandara, F. H. Arnold, *Protein Eng.* **2001**, *14*, 699–704.
- 88 V. Sieber, C. A. Martinez, F. H. Arnold, *Nat. Biotechnol.* **2001**, *19*, 456–460.
- 89 K. H. Oh, S. H. Nam, H. S. Kim, *Biotechnol. Prog.* **2002**, *18*, 413–417.
- 90 J. S. Pedersen, D. E. Otzen, P. Kristensen, *J. Mol. Biol.* **2002**, *323*, 115–123.
- 91 A. Aharoni, L. Gaidukov, S. Yagur, L. Toker, I. Silman, D. S. Tawfik, *Proc. Natl Acad. Sci. USA* **2004**, *101*, 482–487.
- 92 W. C. Suen, N. Zhang, L. Xiao, V. Madison, A. Zaks, *Protein Eng. Des. Sel.* **2004**, *17*, 133–140.
- 93 D. S. Tawfik, B. S. Green, R. Chap, M. Sela, Z. Eshhar, *Proc. Natl Acad. Sci. USA* **1993**, *90*, 373–377.

7

Chemical Complementation

Scott Lefurgy and Virginia Cornish

7.1

Introduction

Directed evolution of enzymes designed *de novo* and from existing enzyme scaffolds has the potential to produce catalysts for many useful chemical reactions. Efforts in directed evolution to date have produced a few cases in which significantly increased enzyme activity, increased thermostability, or modification of substrate specificity have been obtained [1]. Notably, many of these successful examples of directed evolution have used complementation strategies as the basis for selection of improved variants, in which the function of a protein is linked to cell survival. While this approach has the advantage of being a selection rather than a screen, a major limitation is that it can only be applied to reactions naturally carried out by the cell that are inherently selectable. In order to exploit complementation to evolve enzymes for a broad range of chemical reactions, general assays must be developed.

Recently, our laboratory reported a complementation assay for enzyme catalysis that is reaction-independent [2]. This assay has been called “chemical complementation” because it combines the power of genetic selection with the diversity of synthetic chemistry. Here, enzyme catalysis is linked to transcription of an essential gene through a small molecule yeast three-hybrid system, in which the reaction chemistry is specified in the linker of a small molecule chemical inducer of dimerization (CID). In addition to serving as the substrate for the enzyme, the CID dimerizes a DNA-binding domain and transcription activation domain to reconstitute an artificial transcription factor and activate expression of a reporter gene. Since the chemistry of the linker (i.e. the substrate) can change while the transcription readout remains constant, the assay has the advantage of being reaction-independent. This system has the potential to assay for numerous enzymatic activities that had previously been inaccessible to selection methods. To demonstrate the flexibility of this assay, we have applied it to such diverse reactions as β -lactam bond cleavage and glycosidic bond formation [3].

In this chapter, we first present a brief overview of the historical context of complementation methods, followed by an outline of the development of chemical complementation and its adaptation to detect enzyme activity. Finally, two applications of chemical complementation are described: the study of enzyme–inhibitor interactions and directed evolution of a glycosynthase.

7.2

Complementation Assays

7.2.1

Introduction

The most widely used genetic method for determination of protein function is the complementation assay. This strategy consists of modifying a cell to knock-out an essential protein and then rescuing it from death by introducing a compensatory protein on a plasmid. In the context of genetics, the protein can be derived from the genomic DNA of another organism, as in the case of gene cloning. The power of this assay lies in its ability to select from a library of thousands or millions of variants only those proteins that possess the desired function. Since the generation of variability in DNA and protein sequences is now routine, researchers have begun to apply complementation to enzymology and directed evolution. The only caveat is that a suitable selection system must be devised that (1) can adequately distinguish functional and nonfunctional proteins, and that (2) can also be tuned to different levels of stringency, depending on the desired outcome. The major limitation to this method has been the reliance on naturally selectable or screenable phenotypes. Despite these parameters, however, many promising results have emerged from complementation approaches.

7.2.2

Early Complementation Assays

Complementation assays have historically been the tools of geneticists. In a now classic experiment, Beadle and Tatum identified enzymes responsible for the biosynthesis of essential metabolites (Figure 7.1) [4]. They introduced random mutations throughout the chromosomal DNA of the ascomycete *Neurospora crassa* by treating the organism with X-rays. Approximately 2000 individual mutated strains were then screened for growth in both rich and minimal medium, where the minimal medium contained only nutrients, such as biotin, that *Neurospora* cannot make itself. Three mutant strains were identified that had growth rates indistinguishable from the wild-type strain in rich media, but grew poorly or not at all in minimal medium. By adding back individual metabolites to the minimal medium and then measuring the cell growth rates, the three strains were shown to be unable to synthesize vitamin B6, vitamin B1, and

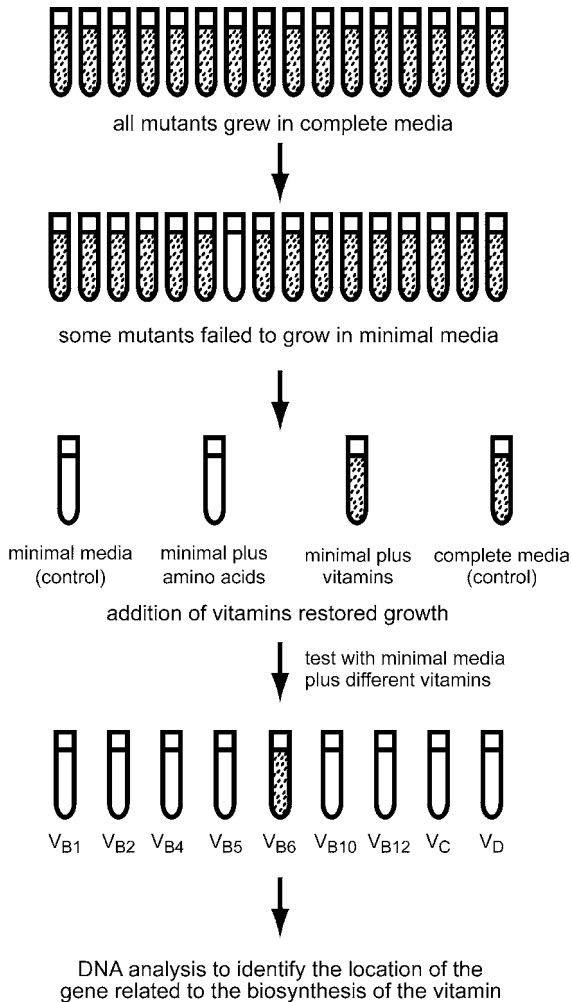


Fig. 7.1 A genetic method to identify genes responsible for the biosynthesis of essential metabolites. Mutant ascomycetes generated by X-ray irradiation were identified by failure to grow on minimal medium. Restoration of growth was conferred by the addition of a

single essential metabolite (vitamin B6). This result pointed toward a putative pathway affected by the mutation and subsequently led to the identification of genes involved in biosynthesis [4].

p-aminobenzoic acid, respectively. This basic approach of mutating the DNA encoding the proteins and then assaying for some detectable phenotype such as cell growth is the basis for genetic analysis. Most of the genes whose function is known have been identified through this type of method, and the functions of thousands of proteins have been assayed in this manner. The contribution of Beadle and Tatum was to point out that the mutant strains could be screened

based on enzymatic activity, as opposed to just cell death. Their method, however, still required several years of laborious study to identify the enzymes encoded by the genes they had mutated.

In the 1950s Benzer introduced the strategy of complementation, which provided a general solution to the problem of gene identification [5]. In genetic complementation, a pool of plasmid DNA encoding fragments of wild-type chromosomal DNA is introduced into the mutant cell line and then the DNA fragment that complements the mutation is identified based on the recovery of the wild-type phenotype. In the *Neurospora* example above, plasmid DNA that encoded fragments of the wild-type *Neurospora* genome would be introduced into the mutant *Neurospora* deficient in vitamin B6 biosynthesis. Only mutant *Neurospora* with a copy of the wild-type vitamin B6 biosynthetic gene would be able to grow in the absence of vitamin B6, allowing the plasmid DNA encoding the vitamin B6 biosynthetic gene to be distinguished readily. The recent availability of several complete genome sequences further simplifies genetic complementation because the plasmid DNA sequence can be compared to the complete genome sequence. In addition, it has allowed several labs to construct libraries of plasmid DNA not just encoding random fragments of chromosomal DNA, but instead explicitly encoding every open reading frame expressed in a given organism or cell line [6].

7.2.3

Enzymology by Complementation

While the cloning of proteins based on function continues to be the major application of complementation assays, they have more recently been shown to be useful tools for enzymology. The advantage of such an approach lies in the rapid assay of millions of mutants simultaneously. The advent of modern molecular biology has facilitated the generation of large libraries of protein variants. Since these libraries are genetically encoded, cell-based assays that link function of the enzyme to cell survival can be exploited for global analysis of the roles of amino acids in protein structure and function.

One aspect of protein structure that has been investigated by complementation is hydrophobic core packing. The observation that substitutions in the protein core are destabilizing led to the theory that optimal packing of side-chains is required for protein folding. While this hypothesis can be derived from first principles, it is difficult to test experimentally. However, Lim and Sauer addressed this problem by using a complementation assay to select properly folded lambda repressor proteins from a collection of mutants with all possible combinations of the core residues [7]. They first generated a library in which seven hydrophobic amino acids in the protein core were randomized by cassette mutagenesis. Then functional repressors were selected from the library based on their ability to prevent infection of *Escherichia coli* by a repressor-defective phage. Substitutions that resulted in tightly packed cores, with hydrophobicities similar to that of the wild-type, produced active repressors. Since this experi-

ment examined every possible set of core residues, the authors were able to draw general conclusions about the rules and patterns of protein core packing. Their findings indicate that hydrophobicity, rather than steric constraints, play the most important role in determining core stability. This example illustrates an experimental question that could not be addressed by site-directed mutagenesis, but was well suited to a combinatorial mutagenesis/selection approach.

Another question of interest to enzymologists is how electrostatic interactions, hydrogen bonding, and conformational constraint contribute to catalysis. Hilvert and coworkers used a complementation approach to assess the role of an active site arginine (R90) in chorismate mutase [8]. This enzyme catalyzes the first committed step in the synthesis of aromatic amino acids. They constructed a mutant *E. coli* strain whose chorismate mutase gene was inactive, and showed that the introduction of the active gene on a plasmid restored growth on a medium lacking phenylalanine and tyrosine. When all possible mutants of R90 were assayed by complementation, only the enzyme containing arginine was selected as active. This result confirmed the requirement of positive charge in that region of the active site. To assess the conformational constraint on this charge, a library was constructed in which both R90 and neighboring cysteine residue (C88) were completely randomized. The selection from this library produced several double mutants that were active (Table 7.1).

Table 7.1 Kinetic parameters for selected chorismate mutase mutants [9].

Chorismate mutase	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($\text{M}^{-1} \text{s}^{-1}$)	$k_{\text{cat}}/k_{\text{uncat}}$	$k_{\text{cat}}/K_m k_{\text{uncat}}$ (M^{-1})
WT	46	67	690000	4.0×10^6	6.0×10^{10}
R90G	0.00027	150	1.8	2.3×10^1	1.6×10^5
R90K	—	—	31	—	2.7×10^6
C88S/R90K	0.29	4300	67	2.5×10^4	5.8×10^6
C88K/R90S	0.32	1900	170	2.8×10^4	1.5×10^7

The R90L mutation was shown to be active when paired with a compensatory mutation at position 88 to serine, alanine, or glycine. Surprisingly, a lysine at position 88 could restore function when position 90 was changed to serine, showing that the placement of the cation is somewhat flexible. Kinetic analysis of these mutants showed that the removal of the cation in the R90G mutant severely diminished catalytic activity, while restoration of the cation by lysine at position 88 or 90 improved this activity by 1000-fold [9]. This result demonstrated that alternative active-site architectures are not only possible, but can be accessed experimentally. Once again, the use of a selection allowed the examination of many more mutants than could be tested by traditional screening methods. Remarkably, this study and the previous one stand alone as examples of complementation applied to enzymology. Given the power of genetic selections, it is clear that this strategy is underutilized.

7.2.4

Directed Evolution by Complementation

Complementation assays have proven particularly powerful tools for directed evolution of proteins with new functions. In the last section, we saw examples of protein folding and activity being linked to cell survival. These kinds of assays also lend themselves well to the selection of enzymes with new or improved properties. In a directed evolution experiment, plasmids that can complement the cellular defect are isolated from a starting library, mutagenized, and submitted to additional rounds of selection. This process proceeds iteratively until the desired activity is attained or convergence of the protein sequence is reached. Increasingly sophisticated methods for generating variation in the library have been developed in order to sample a wider array of sequences. In addition, selective pressures can be applied to the cells to increase the stringency of the selection, and thus glean only the most active members of the mutant library. This process attempts to recapitulate Darwinian evolution in a test tube, and has demonstrated the capacity to produce enzymes with novel and useful properties.

The earliest example of directed evolution by complementation was essentially a systematic observation of artificial selection of bacteria. Hall used a *lacZ*-deletion strain of *E. coli* to select for spontaneous mutants that could grow with lactose as the sole carbon source [10]. The mutations all fell on the opposite side of the chromosome in a single gene, which was termed *ebg* (for evolved β -galactosidase). He first identified two phenotypes that arose from the initial round of selection: class I, which grew on lactose; and class II, which grew slowly on lactose but well on lactulose. He subjected the class I mutants to selection on lactulose and obtained class IV mutants, which grew well on both substrates. Genetic analysis suggested that class IV mutants possessed a combination of the class I and II mutations. The demonstration that enzymes with new functions could be obtained through “directed evolution” by a process that mimicked natural selection was an important step forward for the field. It is remarkable that this kind of analysis was carried out in the absence of genetic engineering methods that we now take for granted.

While Hall relied on naturally occurring mutations, chemical and enzymatic methods of mutagenesis soon came into use for generating libraries of random mutants. Typically, reagents like sodium bisulfite and nitrous acid were used to chemically modify single-stranded DNA, such that misincorporations would occur when the complementary strand was synthesized, resulting in transitions or transversions [11]. Unfortunately, the resulting libraries were often biased, producing many mutations at “hot spots” and failing to mutagenize other positions. An important step forward was the generation of statistically random DNA libraries through the use of chemically synthesized oligonucleotides [12]. In 1990, Knowles and colleagues devised a selection system to find active triosephosphate isomerase (TIM) mutants from a pool of random variants [13]. They started with a known mutant (E165D) that was 1000-fold less active than wild-

type. Then they generated a library of TIM mutants by using oligonucleotide primers that had been doped with non-wild-type bases throughout the length of the sequence, such that each member of the library had an average of one or two mutations, but retained the original E165D mutation. This library was then introduced into an *E. coli* strain that lacked the TIM gene. Selections for functional enzymes were carried out on media that contained glycerol, lactose, or both. The wild-type could grow with glycerol as the sole carbon source, whereas the E165D mutant could only grow on lactose (and grew very slowly on glycerol). Mutants of E165D with increased activity were selected on glycerol-only media. Six “pseudo-revertants” that contained additional mutations were characterized and found to possess higher catalytic efficiency than the starting mutant enzyme. These new mutations were found to cluster near the active site. The best mutant, E165D/S96P, had a catalytic efficiency of $6 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$, which was 19-fold higher than the E165D mutant alone. By comparison, chemical mutagenesis only identified a single mutant with improved activity (S96P).

In addition to increasing catalytic activity, directed evolution has been used to change the substrate specificity of an enzyme. For example, Yano and coworkers used a complementation assay to convert an aspartate aminotransferase to a branched-chain aminotransferase [14]. First, they engineered an *E. coli* strain with a knockout in the wild-type branched-chain aminotransferase gene and showed that this strain was not viable unless the growth medium was supplemented with Val, Ile, and Leu. The wild-type aspartate aminotransferase gene was mutated and recombined using DNA shuffling and then introduced into the *E. coli* selection strain. The authors point out that DNA shuffling, a PCR-based method, could be used both to introduce point mutations and to generate recombined gene products. Five rounds of mutagenesis and selection were carried out. At each round, 10^6 – 10^7 mutants could be examined, and the stringency of the selection could be readily increased by changing the concentration of 2-oxovaline in the growth media, the expression level of the enzyme, or the incubation time. After the final round, a mutant enzyme with 13 point mutations was isolated that had a 10^5 -fold increase in $k_{\text{cat}}/K_{\text{m}}$ for β -branched amino acids and a 30-fold decrease in $k_{\text{cat}}/K_{\text{m}}$ for aspartate (Tables 7.2 and 7.3). The improvement in $k_{\text{cat}}/K_{\text{m}}$ here is impressive and likely reflects the importance of

Table 7.2 Evolution of branched-chain aminotransferase activity from enzyme aspartate aminotransferase [14].

Residue No.	7	34	37	41	126	139	142	269	282	293	297	311	326	363	381	387	397
Wild-type	E	N	I	K	K	S	N	A	R	A	N	S	M	S	A	V	M
AV5A-1	V	D	M			G	T		C	V	S		K	G	V		
AV5A-7	V	D	M	N	R	G	T	T		V	S	G				L	L
AV5A-7(6)		D	M			G	T				S					L	

Table 7.3 Kinetic parameters of evolved branched-chain aminotransferases [14].

Substrate	Wild-type			AV5A-7			AV5A-7 ₆		
	k_{cat} (s ⁻¹)	K_{m} (mM)	$k_{\text{cat}}/K_{\text{m}}$ (M ⁻¹ s ⁻¹)	k_{cat} (s ⁻¹)	K_{m} (mM)	$k_{\text{cat}}/K_{\text{m}}$ (M ⁻¹ s ⁻¹)	k_{cat} (s ⁻¹)	K_{m} (mM)	$k_{\text{cat}}/K_{\text{m}}$ (M ⁻¹ s ⁻¹)
L-Valine	–	–	<0.001	46	420	110	–	–	14
2-Oxovaline	0.0057	100	0.057	28	3.8	7.4×10^3	60	33	1.8×10^3
L-Aspartate	550	4.5	1.2×10^5	220	68	3.2×10^3	70	11	6.4×10^3
Oxaloacetate	800	0.035	2.3×10^7	650	0.52	1.3×10^6	1400	0.46	3.0×10^5

being able to modulate the stringency of a selection – first to detect poor catalysts and finally to demand high catalytic efficiency.

The authors also address the question of how hard it is to alter specificity. By testing each of the 13 mutations individually, they showed that only six point mutations are actually required for a 10^5 -fold increase in activity toward a new substrate. Interestingly, their results also showed that all six mutations are beneficial on their own – raising the possibility that each position in a protein can be randomized independently. In a related application, the Diversa Corporation (San Diego, CA) has applied a mutagenesis and selection technique, termed Gene Site Saturation MutagenesisTM (GSSM), to evolve proteins that have improved thermostability and stereoselectivity [15]. In this approach, each position in the protein is randomized independently. With library sizes approaching 10^8 variants, it is actually possible to randomize all positions in the protein in the same experiment.

On the challenging problem of changing the substrate specificity of an enzyme, directed evolution approaches have shown great promise. For example, Schultz and colleagues have used genetic selections to generate aminoacyl-tRNA synthetases that accept unnatural amino acids as substrates [16]. In addition, Pabo and coworkers have adapted a bacterial two-hybrid assay [17] to evolve zinc finger variants with defined DNA-binding specificities [18]. In our opinion, the most successful examples of directed evolution to change substrate specificity have employed complementation assays.

Since the advent of catalytic antibodies, the hope has been that complementation assays could be used to improve the efficiency of those first-generation catalysts. However, antibodies have proven difficult to express in *E. coli* and *Saccharomyces cerevisiae* and thus difficult to improve by selection in these organisms [19–21]. With recent successes in the *de novo* computational design of enzymes [22–25], it seems we may be able to return to this basic approach but with better-behaved protein scaffolds. Computational design could be used to generate the first-generation catalyst, and selection via a complementation assay could then be employed to improve the activity of that catalyst. To achieve this lofty goal, however, general complementation assays, which can link enzyme catalysis to cell growth in a way that is reaction-independent, will be needed. In

the next section, we describe a system based on the yeast three-hybrid assay that fulfills this requirement.

7.3 Development of Chemical Complementation

7.3.1 Introduction

As we saw in the previous section, complementation assays are very powerful, but limited to a few reactions that are inherently screenable or selectable. What is needed is an assay that has the power of a complementation growth selection but where the chemistry can be adapted to a reaction of interest. Thus, several laboratories have sought to develop high-throughput assays for enzyme catalysis that are general and can be applied to a broad range of chemical reactions [26–32]. Our laboratory recently reported such an assay, which uses the yeast three-hybrid assay to link enzyme catalysis to reporter gene transcription *in vivo* [2, 33]. We called this assay “chemical complementation” because it resembles a complementation assay in that enzyme activity is required for the cellular phenotype, but the chemistry being complemented is controlled by the chemical linker between two small-molecule ligands and thus can be readily varied. Chemical complementation detects enzyme catalysis of bond formation or cleavage reactions based on covalent coupling of two small molecule ligands *in vivo*. The heterodimeric small molecule reconstitutes a transcriptional activator, turning on the transcription of a downstream reporter gene. Bond formation is detected as activation of an essential reporter gene; bond cleavage is detected as repression of a toxic reporter gene. The assay can be run on a large library of enzymes as a growth selection where only cells containing a functional enzyme survive. A number of well-characterized reporters for both positive and negative selections, as well as screens, have been reported [34].

The assay can be extended to new chemistry simply by synthesizing small molecule heterodimers with different chemical linkers as the enzyme substrates. All of the other components of the readout system remain constant while the chemistry changes. This flexibility extends this assay to a wide range of reactions, with a variety of substrate specificities and regio- and stereoselectivities. A distinct advantage of this general method is that a new assay would not have to be developed for each new enzyme target.

In this section, we describe the development of chemical complementation. Since chemical complementation hinges on the integrity of the transcriptional readout system, we initially set out to devise a robust yeast three-hybrid assay. We then adapted this three-hybrid assay into a reaction-independent enzyme assay, using β -lactamase as a model enzyme. We begin by considering three-hybrid assays generally, focusing on the dexamethasone-methotrexate (Dex-Mtx) system that was developed in our laboratory. Then we present the basic chemi-

cal complementation system and consider technical improvements made to the assay since its inception.

7.3.2

Three-hybrid Assay

Obviously, the key to chemical complementation working well is the small molecule three-hybrid system. Several three-hybrid systems and CIDs had already been reported when we began designing the chemical complementation assay. These systems supported the feasibility of our strategy. Nonetheless, we decided to begin by designing a new CID pair to provide a yeast three-hybrid system that would simplify the synthesis of the small-molecule dimerizer and give a strong transcriptional readout (Figure 7.2).

7.3.2.1 Original Yeast Three-hybrid System

Licitra and Liu reported the first yeast three-hybrid assay [35]. This assay was an extension of the yeast two-hybrid assay reported by Fields and Song for the high-throughput detection of protein–protein interactions [36]. The yeast two-hybrid assay is based on the observation that eukaryotic transcriptional activators can be dissected into two functionally independent domains, a DNA-binding domain (DBD) and a transcription activation domain (AD), and that hybrid transcriptional activators can be generated by mixing and matching these two domains [37, 38]. It seems that the DNA-binding domain only needs to bring the activation domain into the proximity of the transcription start site, suggesting that the linkage between the DBD and the AD can be manipulated without disrupting activity [39]. Thus, the linkage in the yeast two-hybrid assay is the non-covalent bond between the two interacting proteins. In the yeast three-hybrid assay, the linkage between the DBD and AD is a small-molecule chemical inducer of dimerization (CID) [40]. The CID consists of two ligands that are covalently linked, and that bind to their respective receptor proteins. One receptor is fused to the DBD and the other to the AD, such that when the CID binds to both receptors at the same time, the complete transcription factor is reconstituted (Figure 7.3).

In Licitra and Liu's yeast three-hybrid system, the transcription factor Gal4 was split into its DBD and AD. The DBD of Gal4 was fused to the ligand-binding domain of rat glucocorticoid receptor (GR), which binds to ligand 1 (dexamethasone, Dex) with nanomolar affinity. The AD of Gal4 was fused to yeast


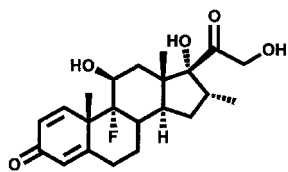
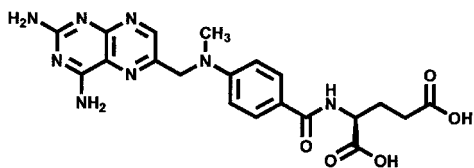


Fig. 7.2 Molecules used for yeast three-hybrid systems. Monomeric small molecules (1–4) that bind with high affinity to their respective receptor proteins are frequently employed in yeast three-hybrid systems.

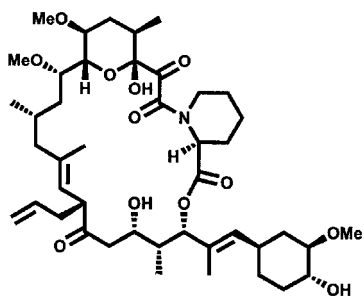
Heterodimeric small molecules (5–7) that can reconstitute hybrid transcriptional activators are constructed by connecting monomers with a long, flexible linker (see text).



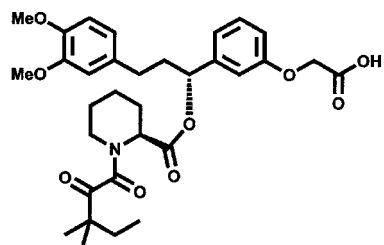
1



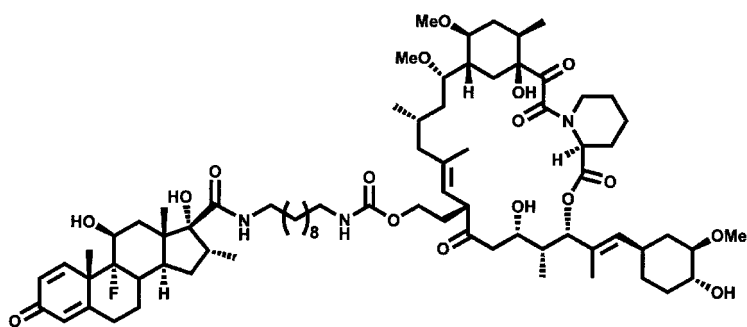
2



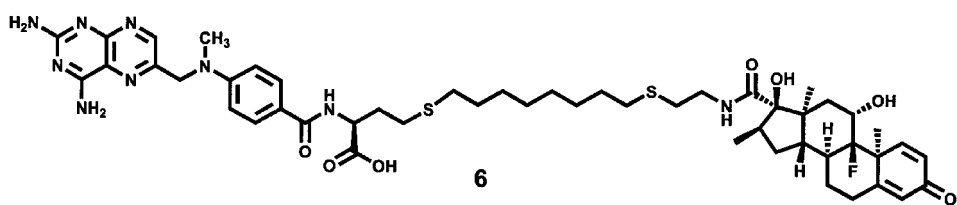
3



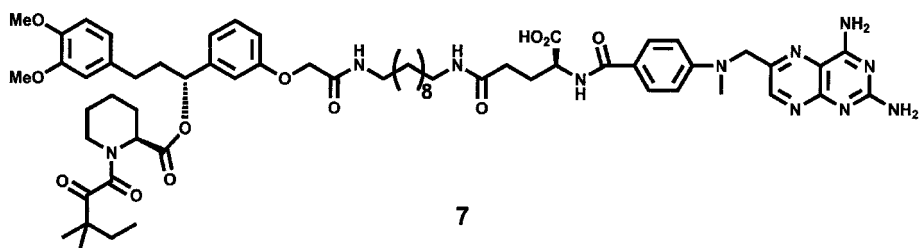
4



5



6



7

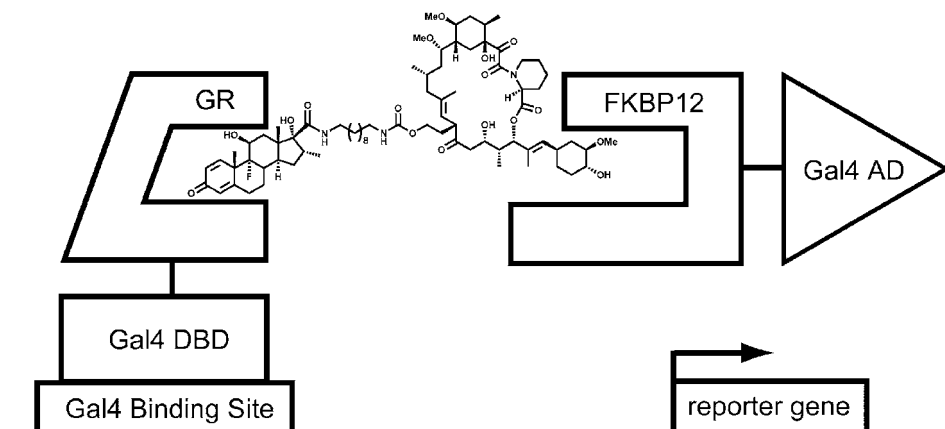


Fig. 7.3 Dex-FK506 yeast three-hybrid system. Heterodimeric ligand **5** (Dex-FK506) bridges a DNA-binding protein–receptor protein chimera (Gal4 DBD-GR) and a transcription activation protein–receptor protein chimera (Gal4 AD-FKBP12), effectively reconstituting a transcriptional

activator and stimulating transcription of a reporter gene. The levels of reporter transcription serve as an indicator of the efficiency of Dex-FK506-induced protein dimerization [35]. AD=activation domain; GR=glucocorticoid receptor; DBD=DNA-binding domain.

FK506 binding protein (FKBP12), which binds to the ligand **3** (FK506) with nanomolar affinity. Both **1** and **3** can be chemically modified without disrupting binding to their respective receptors, so heterodimer **5** (Dex-FK506) served as the CID. The reporter gene consisted of the Gal4 promoter placed upstream of the *LEU2* gene. The use of *LEU2* as a reporter gene linked reconstitution of the Gal4 transcription factor to survival of a leucine-auxotrophic yeast strain. In a proof of principle experiment, Licitra and Liu selected FKBP12 from a Jurkat cDNA library fused to the Gal4 AD. Despite its promise, there were two drawbacks to the Dex-FK506 yeast three-hybrid assay. First, polyketide **3** is a complex natural product – expensive to obtain in large quantities, requiring several steps for chemical modification, and sensitive to many standard chemical transformations. This synthetic complexity was considered a major drawback for chemical complementation, where the linker between **1** and **3** would need to be a complex substrate rather than a simple commercial linker. Second, the use of a Gal4-based system precluded the use of the regulatable *GAL* promoter for expression of the three-hybrid components or the enzyme in the eventual chemical complementation system. In our system, we sought to address these issues.

7.3.2.2 Dexamethasone–Methotrexate Yeast Three-hybrid System

The yeast three-hybrid assay is based on the interaction between two small molecule ligands and their respective protein receptors. In theory, any ligand–receptor pair could serve in this capacity, but certain pairs have advantages for this

system. The ideal ligand–receptor pair would have a very low dissociation constant to increase the lifespan and concentration of the three-hybrid complex and thus increase the levels of reporter gene transcription. The ideal ligand would be small, membrane-permeable, and readily derivatized without disrupting its binding to the receptor. Its receptor would be monomeric and able to be expressed as a fusion protein. Methotrexate and dihydrofolate reductase (DHFR) meet these criteria perfectly. Ligand **2** (methotrexate, Mtx) is commercially available and can be synthesized readily from simple starting materials. DHFR fusion proteins have been used for a variety of biochemical applications [41, 42] due to the picomolar dissociation constant of 2 for DHFR [43]. These properties made Mtx-DHFR an ideal choice for replacement of FK506-FKBP12 in the original yeast three-hybrid system. For the second pair, we chose to retain the dexamethasone–glucocorticoid receptor pair from Liu’s system. The rat glucocorticoid receptor binds **1** with a K_D of 5 nM, and mutants of GR with increased affinity for **1** have been isolated [44]. The commercially available steroid **1** has been used extensively as a cell-permeable small molecule to regulate the activity and nuclear localization of GR fusion proteins *in vivo* [45]. The heterodimer **6** (Dex-Mtx) thus served as a synthetically tractable heterodimer that could easily be modified to add the specialized linker for chemical complementation.

Another important consideration for yeast three-hybrid systems is the relative expression of the receptor–fusion proteins. The *GAL* promoter is the most tightly regulated promoter in yeast and can be induced to high levels with galactose, completely repressed by glucose, and tuned with varying ratios of galactose and glucose [46]. Of course, this promoter could not be used in previous yeast three-hybrid systems that employed the Gal4 system because Gal4 is part of the *GAL* promoter system. Controllable expression of the protein fusions offers a number of advantages, particularly for molecular evolution. The enzyme can be expressed at high levels during early rounds of selection to obtain a low level of the desired activity. The enzyme expression levels can then be decreased during later rounds of selection to discriminate the most active enzyme variants. Inducible systems also offer a rapid means to check for false positives, by rescreening for the ability to activate reporter genes in the absence of inducer. In order to make a yeast three-hybrid system that would permit the use of the *GAL* promoter, we employed the orthogonal LexA-B42 system similar to that of Brent and coworkers [47]. The *E. coli* repressor LexA served as the DBD, while the *E. coli* B42 “acid-blob” domain served as the AD. Neither of these components is known to interfere with the yeast regulatory systems. The selective markers on the LexA and B42 plasmids were changed from ampicillin-resistance to kanamycin-resistance in anticipation of their use in chemical complementation (*vide infra*). This modification eventually allowed the use of a different marker on the enzyme vector to facilitate its isolation following the selection. Finally, *lacZ* and *LEU2* were used for the reporter genes so that the assay could either be run as a screen or a selection.

Having made these improvements to the system, we showed that heterodimer **6** can activate *lacZ* transcription *in vivo* [48]. On the basis of previous studies

showing that *lacZ* transcription levels correlate with the strength of protein–protein interactions in the yeast two-hybrid assay [47], we expected β -galactosidase activity to be a good indicator of Dex-Mtx-induced protein dimerization. Using standard β -galactosidase activity assays we determined that optimal transcription of *lacZ* occurred at concentrations of **6** greater than 1 μ M. Control experiments established that *lacZ* transcription was dependent on **6**. Only background levels of β -galactosidase activity were detected when **6** was omitted. Competition of **6** with a 10-fold excess of **2** reduced Dex-Mtx-dependent *lacZ* transcription to near background levels. A ten-fold excess of ligand **1**, however, did not affect Dex-Mtx-dependent *lacZ* transcription, and higher concentrations of **1** were toxic to the yeast cells. This result may be due to differences in cell permeability between **1** and **6** or may suggest that LexA-DHFR, but not B42-GR, is the limiting reagent. In addition, when either or both receptors were deleted, only background levels of *lacZ* transcription were detected. The fully assembled system produced a strong transcriptional response in the presence of **6** and provided a robust system in which to read out enzymatic activity in chemical complementation.

7.3.2.3 Technical Considerations

Since the introduction of the yeast two-hybrid assay in 1989, there have been a number of improvements to the basic system. Different DNA-binding and activation domains have been introduced. Convenient vectors with different bacterial drug-resistance markers, yeast origins of replication, and yeast auxotrophic markers are now commercially available. An array of common yeast genetic markers have been introduced as reporters, making it possible to test large pools of protein variants (about 10^6) using growth selections. In addition to the simple activator system, reverse- and split-hybrid systems have been developed to detect the disruption of protein–protein interactions, and most recently a transcriptional repressor-based system was reported. These improvements to the yeast two-hybrid system are directly applicable to the three-hybrid system, and have been given detailed treatment in several recent reviews [34, 47]. The CIDs are an additional complexity of the yeast three-hybrid system and bring their own challenges. Two important considerations that have been examined experimentally in our laboratory are the effect of the CID linker on transcription activation and the species-dependence of the receptor proteins [49].

To carry out a high-throughput screen with the Dex-Mtx yeast three-hybrid assay, it proved necessary to stabilize the transcription readout. The original Dex-Mtx system showed both a high number of false positives and false negatives, and an inconsistency in the levels of small-molecule-induced transcription activation. It is generally assumed that this variability in the transcription readout comes from the use of multiple 2μ plasmids. For the Dex-Mtx yeast three-hybrid system, the DBD fusion, AD fusion, and reporter gene are all on 2μ plasmids, and for the chemical complementation system a fourth 2μ plasmid would then be used to express the enzyme. Not only can these plasmids recombine with

one another since they have large regions of sequence identity, but also the distribution of plasmid copies likely varies in each cell. The common remedy is to integrate the genes of interest into the chromosome. However, doing so decreases the number of copies of each gene from as many as 60 to one or two (haploid or diploid), which may not be sufficient for a strong transcription readout. Thus, all combinations of the DBD, AD, and reporter gene were integrated into the chromosome, and the resulting strains were tested for their levels of small-molecule-induced *lacZ* transcription [50]. The AD and reporter gene were found to be the limiting reagents, such that the DBD and any one other component could be integrated without affecting the levels of transcription activation significantly. Use of a Dex-Mtx yeast three-hybrid strain in which the DBD and reporter gene were integrated reduced the false positive rate from 20% to 3%, a significant improvement. Interestingly, integration of the DBD actually increased the levels of transcription activation, which arguably is consistent with optimizing the concentration of DBD bound both to DNA and to the CID. For all of the applications carried out in our laboratory to date, it has proven necessary to integrate one or more components of the yeast three-hybrid system into the chromosome, and it is not simply predictable which integrated strain will be optimal for a given application.

Despite the widespread use of yeast two- and three-hybrid assays, very little is known about how transcription activation levels correlate with the strength of the binding interaction. Golemis and coworkers undertook a systematic study of the relationship between protein–protein interaction and the yeast two-hybrid readout using homo- and heterodimers for which the dissociation constants had been determined *in vitro* [47]. From their studies they concluded that in the yeast two-hybrid system it is not possible to detect interactions with dissociation constants above 1 μM . To extend this work to the yeast three-hybrid system, we set out to examine the correlation between the strength of the binding interaction and the transcription readout using a series of mutants of a well-studied ligand–receptor pair [51]. We chose to carry out these studies in the Mtx-DHFR yeast three-hybrid system described above. The sensitivity of the three-hybrid assay should be a function of the ligand–receptor pair being used as an anchor. Since **2** binds to DHFR with low picomolar affinity, the use of **2** as an anchor should facilitate the detection of relatively weak interactions at the other end. To characterize the sensitivity of the Mtx yeast three-hybrid system, we considered the interaction between FKBP12 and **4** (synthetic ligand for FKBP12, SLF). The nanomolar dissociation constant of this pair left ample room for the design of FKBP12 mutants with reduced but measurable affinity. First, we demonstrated that **7** (Mtx-SLF) could activate *lacZ* transcription in a yeast three-hybrid system consisting of LexA-DHFR and B42-FKBP12. Then, to obtain FKBP12 variants with a wide range of binding affinities for **4**, we designed point mutants of FKBP12 on the basis of previous biochemical studies and analysis of the high-resolution crystal structure of FKBP12 with **4** and **3** [52, 53]. We then determined the dissociation constants for **4** and the FKBP12 mutant variants by using a fluorescence polarization assay [54]. Finally, we inserted these FKBP12

point mutants into the yeast three-hybrid assay and measured the levels of transcription activation using a *lacZ* reporter gene. This experiment showed that binding affinity in the yeast three-hybrid system correlates with transcription, but only over a dynamic range of an order of magnitude. Furthermore, this study demonstrated that the yeast three-hybrid system has a K_D cutoff of 50 nM when using the Mtx-DHFR anchor. Interestingly, another study, in which the Mtx-DHFR anchor is used to screen a cDNA library for drug targets, estimates the detection limit to be low micromolar [55]. Since the readout is affected by the placement of the linker on the ligand and the folding and localization of the receptor, there may not be a simple correspondence between affinity and transcription for every interacting pair.

7.3.2.4 Other Three-hybrid Systems

The appeal of high-throughput screening methods to detect protein–small molecule interactions has prompted the development of yeast three-hybrid systems that use alternative strategies for transcription activation. Since any ligand–receptor pair can potentially be used in the three-hybrid system, some groups have looked to exploit increasingly high-affinity interactions as anchors. GPC Biotech introduced a drug–target interaction into a yeast three-hybrid system anchored by Mtx-DHFR. They then used this system to screen cDNA libraries for cyclin-dependent kinases (CDKs) that bind selectively to a given CDK inhibitor [55, 56]. Peterson and coworkers developed a yeast three-hybrid system anchored by the femtomolar biotin–streptavidin interaction [57]. Taking this approach to the extreme, Johnsson and colleagues designed a system based on covalent bond formation between the CID and the DBD, using the human DNA-repair enzyme *O*⁶-alkylguanine-DNA alkyltransferase (hAGT) and its substrate, *O*⁶-methylguanine [58]. The DBD-hAGT fusion served as one “receptor” and DHFR-AD as the other, while the CID consisted of **2** linked to *O*⁶-benzylguanine.

Another interesting possibility is the use of a small molecule to activate transcription, without the need for a protein AD. Mapp and coworkers have produced a system in which the CID consists of **2** linked to derivatives of the small molecule isoxazolidine, which interacts directly with the transcription machinery to activate reporter gene expression. Three-hybrid approaches have generated much interest as high-throughput screening tools, but the validation of these methods will require their application to practical problems of protein–small molecule interactions.

7.3.3

Chemical Complementation

To implement chemical complementation, we needed a robust small molecule yeast three-hybrid assay and an enzyme–substrate pair that could be programmed into the assay. Our optimized Dex-Mtx yeast three-hybrid system produced a strong small-molecule-dependent transcription readout. All we needed

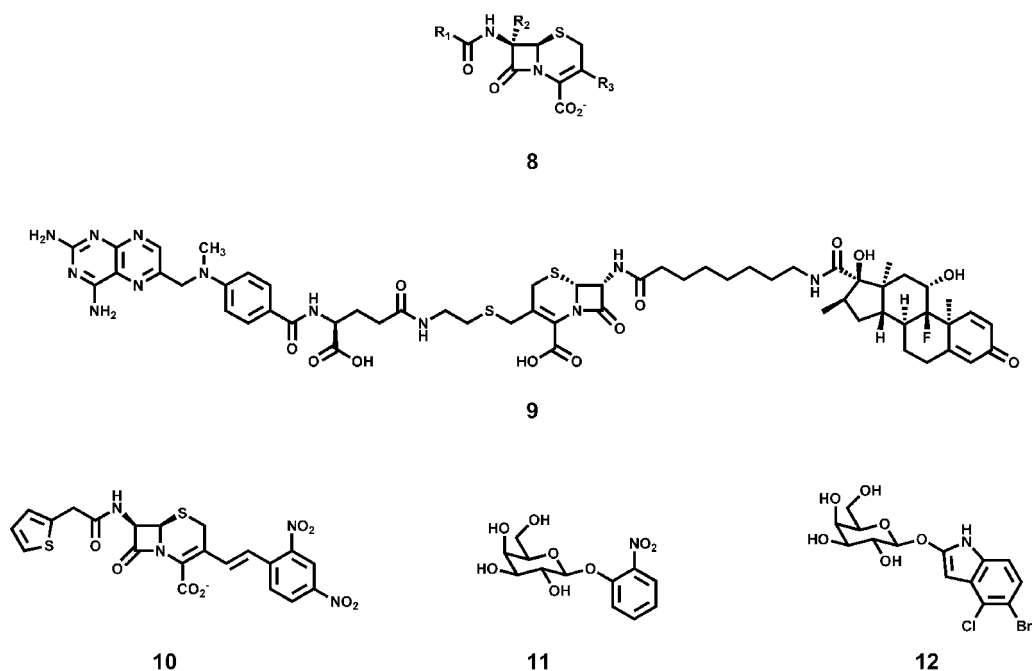


Fig. 7.4 Molecules used in the development of chemical complementation. The cephem core β -lactam (**8**) is a substrate for cephalosporinases. Mtx-Cephem-Dex (**9**) (MCD) is both a substrate for cephalosporinases and a heterodimeric ligand for the Dex-Mtx yeast

three-hybrid system. Nitrocefina (**10**) is a chromogenic cephalosporinase substrate. *o*-Nitrophenyl- β -D-galactopyranoside (**11**) (ONPG) and X-Gal (**12**) are chromogenic substrates of the β -galactosidase reporter gene (LacZ) [2].

to do was introduce the enzyme on a plasmid and the substrate in the linker of **6**. Having developed the assay around a model reaction, we then established it as a high-throughput screen for enzyme activity (Figure 7.4).

7.3.3.1 Selection Scheme and Model Reaction

To adapt our Dex-Mtx yeast three-hybrid system to read out enzyme catalysis of bond cleavage reactions, we replaced the linkage between Mtx and Dex in **6** with the substrate for the reaction and added an enzyme as a fourth component to the system. We chose cephalosporin hydrolysis by the *Enterobacter cloacae* P99 cephalosporinase as a simple cleavage reaction [59] to demonstrate the selection strategy (Figure 7.5). Cephalosporins are β -lactam antibiotics, and cephalosporinases are the bacterial resistance enzymes that hydrolyze and, therefore, inactivate these antibiotics. The P99 cephalosporinase is well-characterized biochemically and structurally [60, 61] and the synthesis of cephem compounds (the family described by **8**) is established [62]. We chose to incorporate Mtx and Dex at the C3' and C7 positions, respectively, of the cephem core (R_3 and R_1 of **8**).

Cleavage of the β -lactam bond in the cephem moiety of heterodimer **9** (Mtx-Cephem-Dex, MCD) results in expulsion of the leaving group at the C3' position, effectively breaking the bond between Mtx and Dex (Figure 7.5 a). Thus, heterodimer **9** should reconstitute the transcriptional activator, causing transcription of the *lacZ* reporter gene in the yeast three-hybrid assay. When the cephalosporinase enzyme is expressed, however, the cephem linkage should be cleaved, and protein dimerization and transcription of the reporter gene should be disrupted (Figure 7.5 b).

The next step was to add the enzyme to the yeast three-hybrid system and find conditions where we could observe enzyme-dependent transcription. To this end, the yeast three-hybrid strain was re-engineered so that the expression of the transcriptional activator fusion proteins and the enzyme could be regulated independently. The LexA-DHFR and B42-GR proteins were placed under the control of the fully regulatable *GAL1* promoter; the P99 cephalosporinase, under the repressible *MET* promoter. In addition, the enzyme was expressed

Fig. 7.5 Chemical complementation links enzyme catalysis to reporter gene transcription. Cephalosporin hydrolysis provides a simple cleavage reaction to demonstrate the complementation strategy.

(a) Cephalosporin hydrolysis by a cephalosporinase enzyme. Cephalosporinases are serine-protease enzymes and catalyze the hydrolysis of cephalosporin antibiotics **8** by means of an acyl-enzyme intermediate.

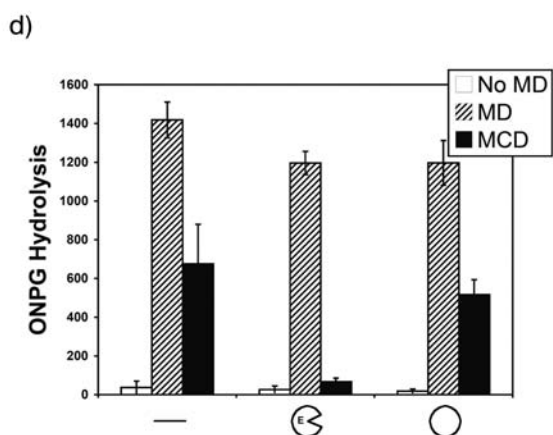
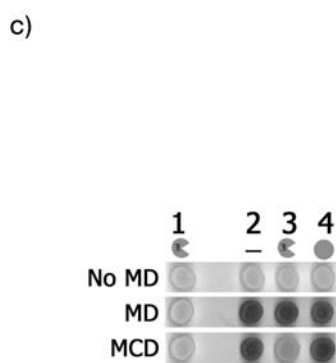
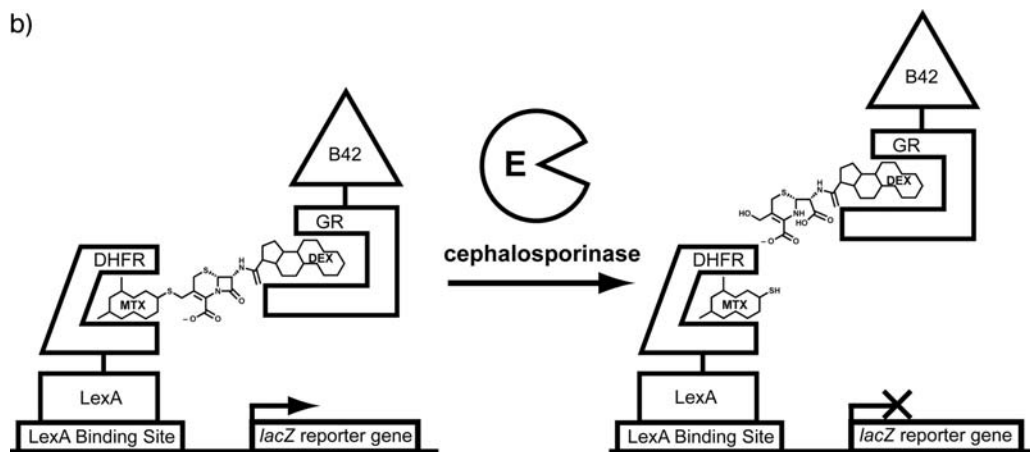
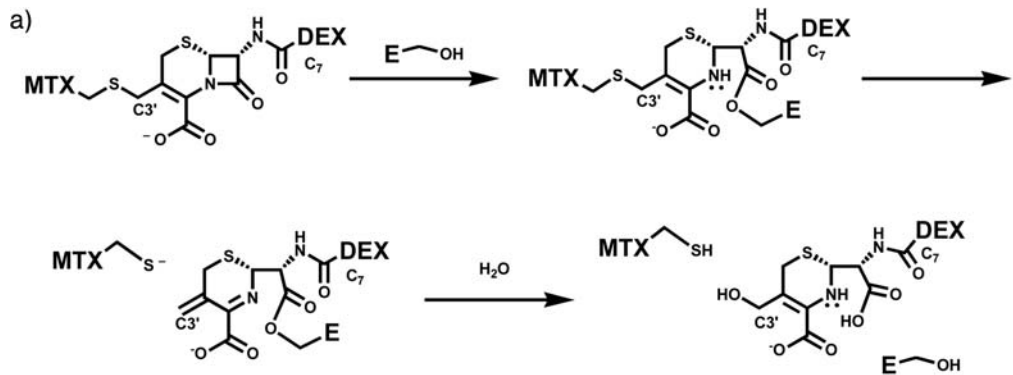
Hydrolysis of the β -lactam bond in **9** results in expulsion of the leaving group at the C3' position of the cephem core, effectively breaking the bond between Mtx and Dex.

(b) Cephalosporin hydrolysis by the cephalosporinase enzyme disrupts transcription of a *lacZ* reporter gene. Substrate **9** dimerizes a LexA DNA-binding domain-dihydrofolate reductase (LexA-DHFR) and a B42 activation domain-glucocorticoid receptor (B42-GR) fusion protein, activating transcription of a *lacZ* reporter gene. Addition of active cephalosporinase enzyme results in cleavage of substrate **9** and disruption of *lacZ* transcription.

(c) X-Gal plate assays of cephalosporinase-dependent Mtx-Cephem-Dex (MCD)-induced *lacZ* transcription. Yeast strains containing a *lacZ* reporter gene were grown on indicator plates containing **12** with or without Mtx-linker-Dex molecules as indicated. Columns 1–4 correspond to yeast strains containing

a LexA DNA-binding domain fusion protein, a B42 activation domain fusion protein, and enzyme, as follows: Column 1, LexA-DHFR, B42, P99 cephalosporinase. Strains in column 1 lack GR and are used as negative controls. Column 2, LexA-DHFR, B42-GR, no enzyme; column 3, LexA-DHFR, B42-GR, P99 cephalosporinase; and column 4, LexA-DHFR, B42-GR, P99 S64A cephalosporinase. The rows correspond to individual X-Gal plates, which have different small molecules as indicated: No MD, No substrate; MD, 1 μ M **6**; MCD, 10 μ M **9**.

(d) *o*-Nitrophenyl- β -D-galactopyranoside (ONPG) liquid assays. Yeast strains expressing the LexA-DHFR and B42-GR fusion proteins and containing *lacZ* reporter gene and expressing no enzyme (left), P99 cephalosporinase (center), or P99 S64A cephalosporinase (right) were grown in liquid culture and assayed for β -galactosidase activity with **11** as a substrate. The liquid culture contained small molecules as indicated. The assays were done in triplicate. Hydrolysis rates of **11** are reported as $\text{nmol min}^{-1} \text{mg}^{-1}$ of total protein, and the error bars for the specific activity correspond to the standard error. Strains containing the active P99 cephalosporinase showed an eightfold decrease in the level of *lacZ* transcription relative to strains containing the inactive S64A variant [2].



from a plasmid containing a spectinomycin-resistance marker, whereas the transcriptional activator fusion proteins were expressed from vectors containing kanamycin-resistance markers, to facilitate isolation of the plasmid encoding the enzyme at the end of a screen or selection. Finally, no ampicillin-resistance markers were used because these markers encode β -lactamase enzymes, which, if expressed, could cleave the MCD substrate. First, we established independently that the cephalosporinase was being expressed in an active form in the yeast cells by using **10** (nitrocefin), a known chromogenic substrate for the P99 cephalosporinase [23]. We then developed conditions where expression of the P99 cephalosporinase disrupted MCD-mediated *lacZ* transcription (Figure 7.5c, column 3). These results were confirmed by using quantitative assays in liquid culture with **11** (*o*-nitrophenyl- β -D-galactopyranoside, ONPG) and demonstrated that the three-hybrid assay could be used to detect cephalosporinase activity (Figure 7.5d).

7.3.3.2 Results

To confirm that the observed activation of the reporter gene by heterodimer **9** was truly enzyme-dependent, we carried out a number of control experiments. Using standard β -galactosidase assays both on plates and in liquid culture, we measured the levels of *lacZ* transcription in yeast strains expressing different LexA and B42 fusion proteins, enzymes, and a *lacZ* reporter gene (Figure 7.5c,d). Transcription in each of these strains was tested in the presence of no small molecule, **6** (MD), or **9** (MCD). As can be seen, *lacZ* transcription in the strain expressing LexA-DHFR and B42-GR is small-molecule-dependent (Figure 7.5c, column 2). Addition of the wild-type P99 cephalosporinase disrupts this small-molecule-induced transcription activation when the cells are grown in the presence of **9**, but not **6** (Figure 7.5c, column 3, and 7.5d). Importantly, expression of the P99 cephalosporinase has little effect on the levels of Mtx-Dex-activated *lacZ* transcription (Figure 7.5c, row 2, and 7.5d). Another important control was to show that the disruption of *lacZ* transcription was due to turnover of the cephem linkage and not merely due to sequestration of substrate **9** by the cephalosporinase enzyme. To address this question, we compared the activity of the wild-type P99 cephalosporinase in this assay with that of the inactive mutant S64A (Figure 7.5c, column 4, and 7.5d). No detectable change in the level of *lacZ* transcription was observed for cells expressing the inactive enzyme. Together, these results established that the change in transcription of the *lacZ* reporter gene is due to enzyme-catalyzed turnover of substrate **9**.

One criterion for a complementation assay is its ability to select enzymes with a desired activity from a large cDNA or mutant library. To determine the usefulness of our system for high-throughput screening, an experiment was carried out to isolate active enzymes from a large pool of inactive mutants. The yeast selection strain was transformed with a mixture of plasmids containing the inactive S64A mutant (95%) or the wild-type P99 cephalosporinase (5%). The clones were grown on medium containing **9** and **12** (X-Gal) for blue-white

screening. Five blue and five white colonies were analyzed: all five blue colonies contained the expected inactive mutant enzyme, while four of five white colonies contained the wild-type active enzyme. This result indicated that the assay has a low false-positive rate, demonstrating that chemical complementation can be used reliably to screen libraries of proteins *in vivo* on the basis of catalytic activity.

7.3.3.3 General Considerations

An important parameter for any enzyme assay is its ability to detect enzymes of varying catalytic activity. In an evolution experiment, for example, one might wish to select for poor catalysts in early rounds and then increase the stringency to detect better catalysts in later rounds. For chemical complementation, the key question was whether differences in transcription of the reporter correlated with the catalytic efficiency of the enzyme. To address this question, mutant P99 cephalosporinases with a range of catalytic activity were constructed and assayed. *LacZ* reporter gene activation was indeed found to correlate with catalytic activity, and the system was able to distinguish between enzymes spanning three orders of magnitude in k_{cat}/K_m [63]. This dynamic range makes the assay useful for many applications, including directed evolution and enzymology. To improve this dynamic range, switching to a growth selection strategy will be necessary. In such a system, the reporter gene would be an enzyme that catalyzes the synthesis of an essential metabolite. The advantages of this system are twofold. First, small changes in transcription are amplified into large differences in growth advantage, allowing detection of slight differences in enzymatic activity. Second, the assay can be tuned to detect enzymes in a particular catalytic range by providing more or less of the essential metabolite in the growth medium. Alternatively, changing the expression level of the enzyme could produce a similar effect. Efforts are currently underway in our laboratory to develop conditions that improve the dynamic range of the assay.

7.3.3.4 Related Methods

Since a general assay for enzyme activity would be an invaluable tool for drug discovery, directed evolution and proteomics, many groups have attempted to develop such a system. Some of these approaches are also based on the yeast three-hybrid, but use different selection strategies. Benkovic and colleagues have reported a three-hybrid system in bacteria in which the presence of an active enzyme results in transcription of the arabinose operon. Enzyme activity is linked to transcription by using unreacted substrate to compete with a CID for binding to the yeast three-hybrid receptor fusion proteins [30]. In this example, scytalone dehydratase (SD) can process a scytalone analog, rendering the substrate unable to bind to the yeast three-hybrid components. Thus, when an active enzyme is present, the three-hybrid is reconstituted and transcription of the arabinose operon ensues.

In a totally different strategy, Peterson and coworkers have developed a three-hybrid-based system to detect tyrosine kinases. In this system, the peptide substrate of the kinase is covalently fused to the DBD by in-frame translation and takes the role of the CID. Recognition and phosphorylation of the peptide by an active kinase converts the peptide into a ligand recognized by an SH3 domain, which acts as a receptor [64]. Further, this group has reported a system that makes transcription dependent on the enzymatic biotinylation of a protein substrate [65]. Here, the CID consists simply of biotin, which the enzyme BirA covalently links to a peptide fused to the AD. The biotinylated peptide then binds to the DBD–streptavidin fusion to reconstitute the transcriptional activator.

Additional methods that use small molecules to control transcription have been put forth, including the “chemical complementation” assay of Doyle and coworkers [66]. Kodadek and colleagues [67] have developed a complementation system for selecting peptides that bind to proteins of interest using a negative selection, where binding of a peptide to its target protein creates a transcriptional repressor. While these methods have not been applied to enzyme catalysis, they point to a growing interest in the field of high-throughput assays that are based on complementation strategies.

7.4 Applications of Chemical Complementation

7.4.1 Introduction

To explore the flexibility and power of chemical complementation, we have used the system to study inhibitor interactions with a β -lactamase and to improve the catalytic activity of a glycosynthase. In the first example, the system was used to screen a library of mutant β -lactamases for their interaction with third-generation cephalosporin antibiotics, in order to dissect residues that contribute to antibiotic resistance. In the second example, we added carbohydrate chemistry to the repertoire of this assay and carried out a selection for glycosynthase enzymes with improved catalytic function. In this section, we will present the results of these two applications, which exemplify the ease with which this system can be adapted to new chemistry and used to answer complex biochemical questions.

7.4.2 Enzyme–Inhibitor Interactions

Although chemical complementation was designed to read out the activity of an enzyme, a logical extension of this system would be to identify or evaluate inhibitors of an enzyme. This approach is particularly attractive as a means of drug discovery using a library of small molecules to screen for an enzyme in-

hibitor. Alternatively, the effect of a single inhibitor on the function of an enzyme could be probed systematically by screening it against a series of mutant enzymes. In contrast to the classical approach of generating point mutants and correlating them with altered binding, chemical complementation can rapidly screen hundreds of thousands of targeted or random mutants in a functional assay to qualitatively assess the effectiveness of an inhibitor.

7.4.2.1 Rationale

β -Lactam antibiotics are the largest class of antibacterial agents administered worldwide, and the continuing development of resistance by bacterial strains to these drugs is a major public health concern. Resistance is most often mediated by the expression of β -lactamases and cephalosporinases, which inactivate the antibiotics by hydrolyzing the β -lactam bond [68]. The β -lactamase reaction proceeds through a covalent acyl-enzyme intermediate, which in the case of resistant enzymes is rapidly hydrolyzed to regenerate the catalyst. Therefore, significant resources have been devoted to the development of highly functionalized β -lactam antibiotics that can trap these enzymes in the acyl-enzyme state, thus making them unavailable for β -lactam hydrolysis. Mutant β -lactamases that confer resistance to third-generation cephalosporins such as **13** (cefotaxime) and **15** (ceftazidime) have recently been isolated. Structural studies of these mutants suggest that residues contacting the C7 substituent of the cephalosporin are important determinants of resistance [69, 70]. In particular, mutating tyrosine 221 (Y221) of *E. coli* cephalosporinase AmpC to alanine or glycine appears to remove a steric constraint on the acyl-enzyme intermediate, allowing it to adopt a favorable conformation for deacylation. Additionally, biochemical evidence has shown that mutations at glutamic acid 217 (E217) and alanine 220 (A220) may also contribute to cephalosporin resistance [71].

In order to understand the molecular basis of antibiotic resistance, we sought to use chemical complementation to measure the effect of inhibitors on β -lactamases containing mutations at positions that contact the C7-substituent [72]. Since genetic assays for β -lactamase activity already exist, we also used this experiment as a way to evaluate the precision and accuracy of chemical complementation for high-throughput screening.

First, we made a library of enzymes in which position 221 was mutated to all 20 amino acids and compared the inhibition profiles obtained by chemical complementation with the *in vitro* activity of these mutants. Subsequently, we used chemical complementation to screen a library in which E217, A220, and Y221 were all randomized for enzymes that were active against third-generation cephalosporins.

7.4.2.2 Screen Strategy

In order to determine the inhibition profiles of mutant β -lactamases, we first set out to demonstrate that the assay could distinguish active, inactive, susceptible and resistant enzymes. Since the third-generation cephalosporins **13–16**

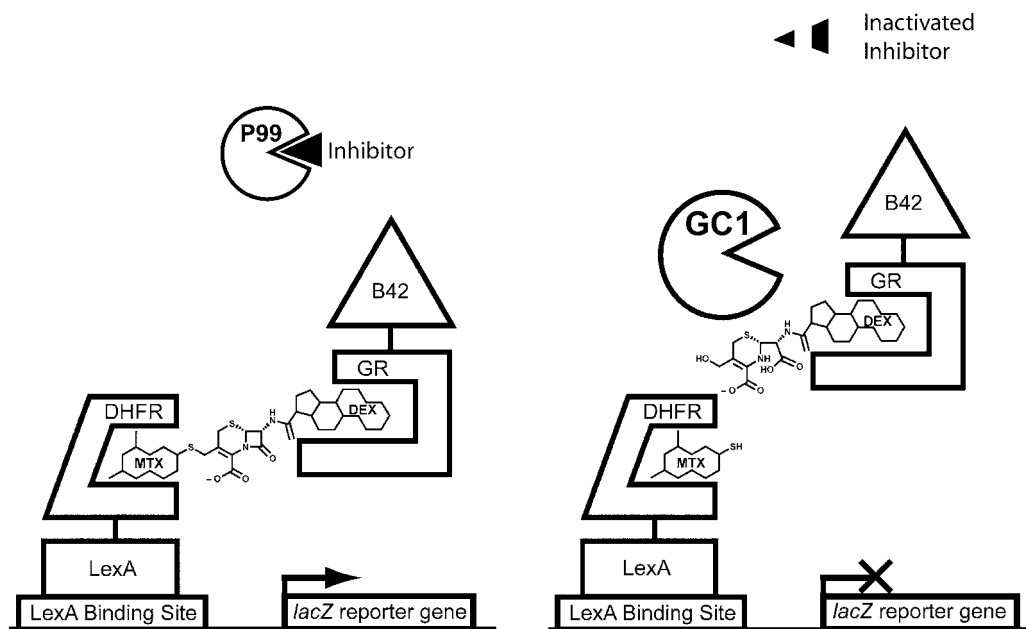


Fig. 7.6 Chemical complementation as a screen for enzyme inhibition. The wild-type β -lactamase (left) is unable to cleave the cephem bond in heterodimer **9** due to the presence of an inhibitor bound in its active site, resulting in transcription of the *lacZ*

reporter gene. Mutant β -lactamases (right) that can cleave or avoid the inhibitor can be identified using the chemical complementation assay because the enzyme cleaves the cephem bond of heterodimer **9**, turning off transcription of the *lacZ* reporter gene.

form a long-lived acyl-enzyme intermediate with the β -lactamase, it was envisioned that these inhibitors would also block cleavage heterodimer **9** by P99 cephalosporinase, resulting in an increase in transcription of the *lacZ* reporter gene. Further, mutant β -lactamases, such as the GC1 variant cephalosporinase, that are not inhibited by the third-generation cephalosporins would be able to cleave heterodimer **9**, and thus could be detected based on a decrease in transcription of the *lacZ* reporter gene (Figure 7.6). This strategy should be general for enzymes that proceed through a covalent enzyme intermediate.

Initially, to establish the feasibility of this screen, we determined the ability of chemical complementation to distinguish the inhibition profiles of wild-type P99 cephalosporinase and a known resistance mutant GC1 (Figure 7.7). Both of these enzymes were able to cleave heterodimer **9**, resulting in background levels of *lacZ* transcription. When the third-generation cephalosporins **13** (cefotaxime) or **14** (cefuroxime) were added to the growth medium, cells expressing the P99 cephalosporinase showed increased transcription of the *lacZ* reporter, indicating that the enzyme was being inhibited. Cells expressing the GC1 variant, however, showed relatively little change in *lacZ* transcription, indicating that this enzyme was available to cleave heterodimer **9** and thus not inhibited. By contrast, as ex-

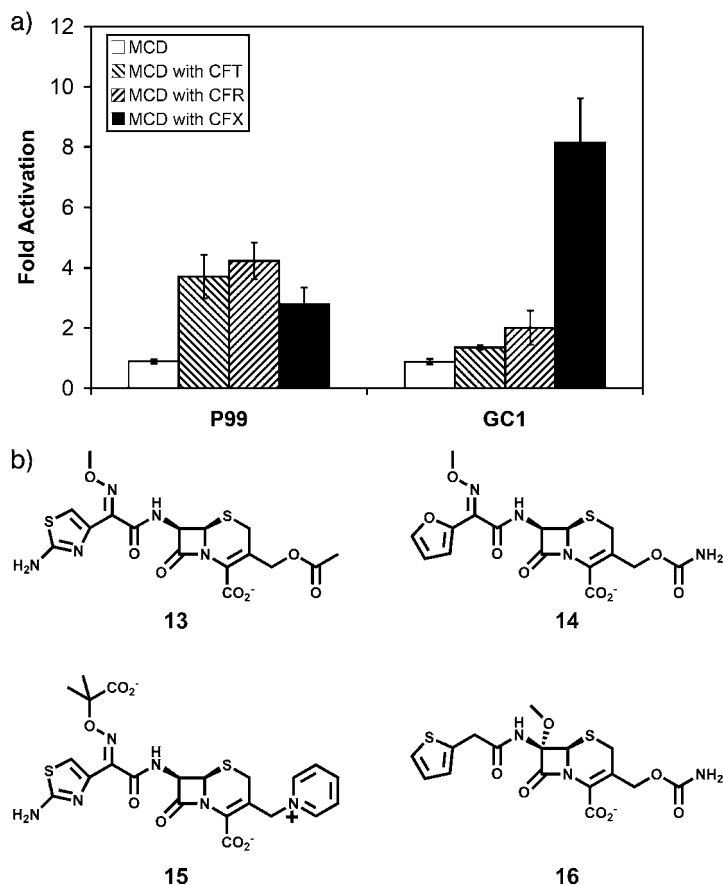


Fig. 7.7 Inhibition profiles of two β -lactamases with cephalosporins. Chemical complementation can distinguish the inhibition of the P99 cephalosporinase by a series of second- and third-generation cephalosporins from that of the extended spectrum GC1 variant.

(a) *o*-Nitrophenyl- β -D-galactopyranoside (ONPG) liquid assay. Screening strains expressing either P99 or GC1 are assayed for *lacZ* transcription in liquid medium containing 5 μ M heterodimer **9** in the presence or absence of 100 μ M **13** (cefotaxime, CFT), **14** (cefuroxime, CFR), or **16** (cefoxitin, CFX). Signals are reported as fold-activation

relative to basal transcription of the *lacZ* reporter gene in the absence of heterodimer **9**. P99 cephalosporinase and GC1 variant both effectively cleave heterodimer **9** in the absence of any added inhibitors, reducing the *lacZ* signal down to background levels.

P99 cephalosporinase is clearly inhibited by all three inhibitors, as shown by the four-fold increase in its *lacZ* transcription signal, while GC1 variant is only inhibited by **16**. Error bars represent the standard error for three separate transformants.

(b) Molecules described in this study. Cephalosporin **15** (ceftazidime, CFZ) was not tested.

pected, the presence of cephamycin **16** (cefoxitin) strongly inhibited both enzymes, resulting in large increases in transcription. Thus, chemical complementation could successfully classify enzymes into three phenotypes: inactive, active and inhibited, or active and resistant. This convincingly showed that the assay could be used as an inhibitor screen.

7.4.2.3 Enzyme Library Screen

Having established the feasibility of the inhibitor screen, we then carried out a pilot screen at position Y221 followed by a high-throughput screen of residues contacting the C7 substituent of the cephalosporin antibiotic. The Y221 library was characterized thoroughly to evaluate the precision and accuracy of the chemical complementation assay. The high-throughput assay then extended the assay to a large scale and explored different positions in the cephalosporin binding pocket.

For the Y221 screen, position 221 of the P99 cephalosporinase was randomized to all 20 amino acids by using degenerate oligonucleotide primers bearing the NNS codon (N=G, A, T, C; S=G, C). The library was transformed into the yeast selection strain and plated under nonselective conditions. Ninety-five randomly chosen transformants were arrayed in a microtiter plate and assayed in the chemical complementation *lacZ* screen for transcription with heterodimer **9** in the presence and absence of cephalosporin **13**. Based on this assay, each clone was assigned a phenotype: active against heterodimer **9** and resistant to **13**; active, but susceptible to inhibition by **13**; or inactive. The plasmids were isolated from each clone and sequenced in the region containing position 221 so that the phenotypes and mutations could be correlated (Figure 7.8). As expected, the Tyr and Ala mutants produced susceptible and resistant phenotypes, respectively, consistent with previous studies. Surprisingly, mutants with Phe, Leu, Val, Cys, His, Lys, Arg, and Ser all showed resistant phenotypes, in contrast to previous studies suggesting a small, hydrophobic residue was required at this position for efficient deacylation of **13**. To determine whether the screen had correctly predicted the *in vitro* activity of the mutant enzymes, a few mutants from each phenotype were expressed, purified, and assayed for activity against **13**. All of the mutants deemed resistant showed a substantial increase in their k_{cat} values with **13**, compared with wild-type P99 cephalosporinase, ranging from a 430-fold increase with Y221L to over 2700-fold increase with Y221A. In addition, the K_{m} values of these mutants were increased by about 10-fold over the wild-type enzyme, leading to an overall improvement of 100-fold in their catalytic efficiencies ($k_{\text{cat}}/K_{\text{m}}$). By contrast, mutants that were phenotypically susceptible to **13**, such as Y221T, showed only modest 30-fold increases in k_{cat} over wild-type. Inactive mutant Y221E was substantially impaired against the heterodimer **9** substrate, with a 100-fold decrease in $k_{\text{cat}}/K_{\text{m}}$ compared with wild-type, despite its 400-fold increase in k_{cat} for **13**. In every case where a mutant was characterized *in vitro*, the chemical complementation assay had accurately predicted the effect of the mutation. Since redundancy was built into the

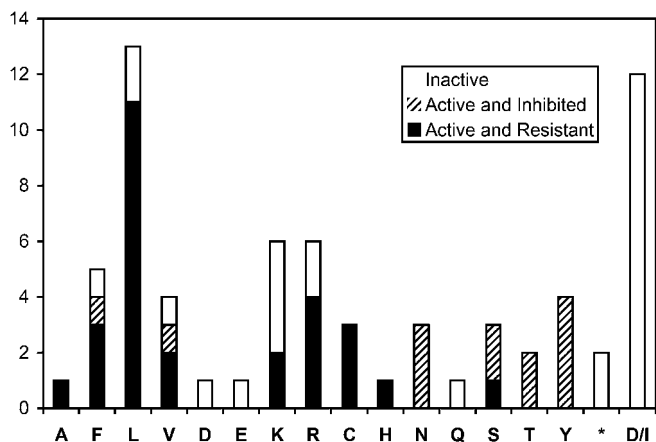


Fig. 7.8 Y221X mutants of P99 cephalosporinase characterized using the chemical complementation *lacZ* screen. Mutations are grouped as hydrophobic amino acids (A, F, L, V), negatively and positively charged amino acids (D, E, K, R), polar amino acids (C, H, N, Q, S, T, Y), stop codons (*), or deletions and insertions (D/I). Based on the signals obtained from the liquid *o*-nitrophenyl- β -D-galactopyranoside (ONPG) data

using the chemical complementation *lacZ* screen in the presence of 5 μ M heterodimer **9**, each mutant was categorized as an active or inactive β -lactamase. Active β -lactamases were further categorized as inhibited by or resistant to cephalosporin **13** on the basis of the same assay carried out in the presence of 100 μ M **13**. No mutants containing G, I, M, P, or W were obtained out of 85 colonies screened.

assay and most mutants were assayed multiple times, the precision of the assay can also be estimated as a ratio of the number of correct results to total results, 74%. This estimate of 74% is conservative because the sequencing results showed that many of the false positives/negatives had additional mutations in the coding region. Thus, the Y221 library showed the chemical complementation assay to be robust and even revealed some surprising resistance mutations at this well-studied position.

Confident of the accuracy of the chemical complementation assay, a high-throughput screen was then undertaken to survey the effects of additional mutations in the region of the enzyme contacting the C7 substituent of the cephalosporin. A library in which positions 217, 220, and 221 of P99 were randomized using NNS codons was constructed and a random subset of 960 clones was screened with **13** in the chemical complementation assay as described above. Mutants that showed resistant profiles in this screen were characterized *in vitro* and several were found to have increased activity toward **13**. Overall, the A220N/Y221H/N226D gave both the highest k_{cat} and highest K_m obtained for any mutant tested, with a k_{cat} that was 5600-fold over that obtained for wild-type P99 cephalosporinase and a K_m that was 100-fold over that obtained for wild-type P99 cephalosporinase. Although the library was not screened comprehensively, an improvement in catalytic efficiency of two orders of magnitude was

observed for the best mutant, D88N/D217V/A220F/Y221A. Surprisingly, a few mutants in which tyrosine was retained at position 221 also showed up to 77-fold increases in catalytic efficiency toward **13**. This result clearly demonstrated the ability of chemical complementation to screen mutant libraries and identify improved catalysts.

7.4.2.4 General Considerations

The requirement of yeast cell permeability to cephalosporins led us to make two critical modifications to the system. First, the growth medium was supplemented with 1% dimethyl sulfoxide (DMSO) to slightly permeabilize the yeast cell membrane to the inhibitors. Next, the *PDR1* gene was knocked-out in the selection strain. *Pdr1p* positively regulates the expression of *PDR5*, *SNQ2*, and *YOR1*, which encode ABC transporters that work on a variety of small molecules [73–75]. *PDR1* knockout strains have been used in a number of studies to increase the permeability of yeast to various small molecules [76]. Together, these two modifications to the assay permitted the observation of enzyme inhibition by cephalosporins *in vivo*.

Like any high-throughput assay, chemical complementation is prone to false positives and false negatives, which result from clonal variation in the yeast three-hybrid components. Conveniently, secondary screening to identify these artifacts simply consists of multiple repetitions of the chemical complementation *lacZ* screen in the presence and absence of heterodimer **9**. Clones that fail to activate *lacZ* transcription in the presence of the CID are considered false positives for β -lactamase activity; conversely, clones that activate transcription in the absence of the CID are regarded as false negatives. In the Y221 library screen, 10 of the 95 colonies tested in this secondary screen were found to be false positives and eliminated. Since this method of secondary screening is enzyme-independent, it should be generally applicable for chemical complementation. The use of secondary screens for the assessment of yeast transcription assays is treated in greater detail elsewhere [34].

7.4.3

Glycosynthase Evolution

One of the major goals of enzyme research is the successful design of enzyme catalysts for any desired reaction. There are exciting recent advances in computational modeling [23], but the goal of designing protein sequences *de novo* that will have a specified activity is still beyond our reach. Directed evolution, on the other hand, seeks to improve catalysts by creating large libraries of protein variants and selecting for those that have improved activity. The bottleneck to enzyme discovery by this method is the design of a suitable selection system. Chemical complementation offers a way to evolve an enzyme for which there is no intrinsic selection, such as a glycosynthase.

7.4.3.1 Rationale

Chemical synthesis of carbohydrates is presently limited by the need for differentially protected intermediates and reactant-dependent coupling yields and stereo-control. Enzymes, with their exquisite control of both regio- and stereo-chemistry, provide a promising alternative to traditional small molecule chemistry for the synthesis of oligosaccharides. Glycosyltransferases, the natural enzymes responsible for the synthesis of oligo- and polysaccharides, and glycosidases, the enzymes that normally hydrolyze carbohydrates, have both been used for carbohydrate synthesis [77–80]. The use of glycosyltransferases, however, is limited by the need for nucleotide diphosphate glycosyl donors, which are relatively more expensive and difficult to synthesize than glycosidase substrates. Oligosaccharide synthesis using glycosidases, not surprisingly, suffers from low yields since the enzyme also catalyzes the hydrolysis of the desired product.

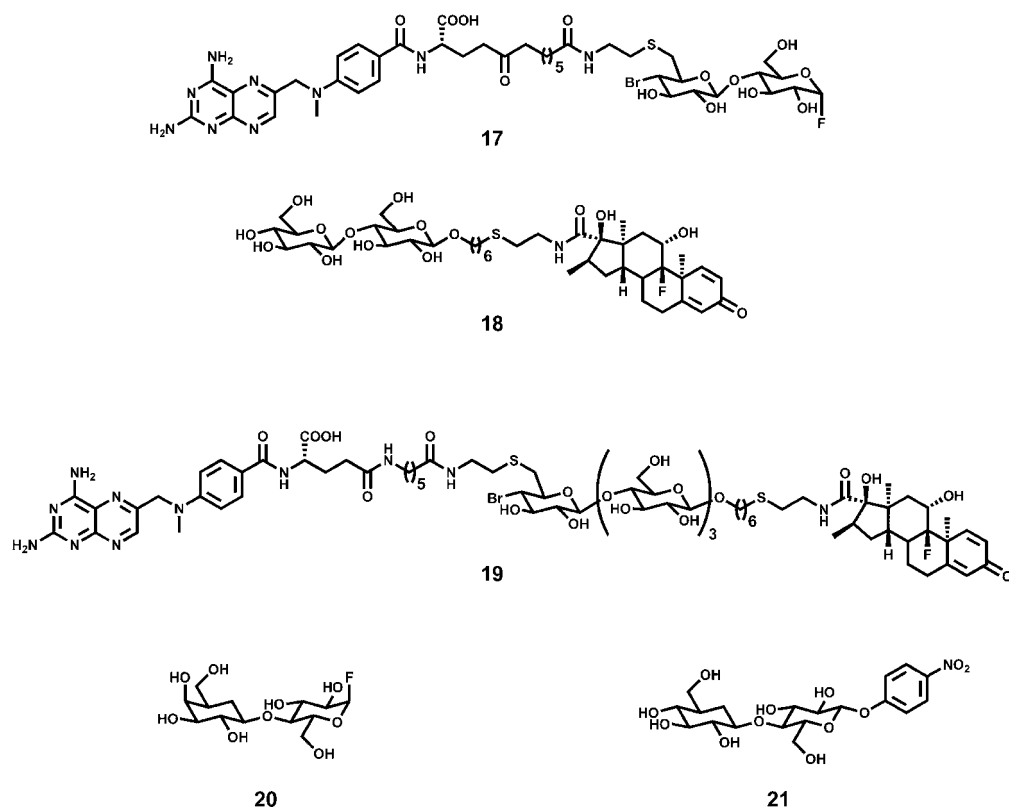


Fig. 7.9 Molecules used for glycosynthase evolution. α -Glycosyl fluoride donor **17** and cellobiose acceptor **18** are linked by the Cel7B:E197A glycosynthase to make heterodimer **19**, which can reconstitute the

Dex-Mtx yeast three-hybrid system [3]. The catalytic activity of evolved glycosynthases is measured *in vitro* using commercially available analogous substrates **20** and **21**.

Thus, alternative methods are being sought for enzyme-catalyzed carbohydrate synthesis.

Recently, Withers and coworkers demonstrated that retaining glycosidases can be engineered to glycosynthases simply by mutating the nucleophilic glutamate residue at the base of the active site to a small hydrophobic residue and using an α -glycosyl fluoride as the donor substrate [81]. Mutation of the active-site nucleophile to a small, hydrophobic residue both accommodates the α -glycosyl fluoride donor and inactivates the hydrolytic activity of the enzyme, allowing the reaction to proceed in the reverse direction. This result opened a new route for carbohydrate synthesis, and several retaining glycosidases have already been successfully converted to glycosynthases using this strategy [82–88]. Directed evolution would offer an obvious route to improve the activity of these enzymes and modify their substrate specificities, except that there is no intrinsic way to screen or select for glycosynthase activity. Thus, we sought to bring chemical complementation to bear on this important class of enzymes for carbohydrate synthesis (Figure 7.9).

7.4.3.2 Selection Scheme

Chemical complementation is a powerful method because to look at a new reaction class, all that needs to be changed is the bond linking the two halves of the CID. For the glycosynthase reaction, we chose to model the linker after the disaccharide substrates used by the E197A mutant of Cel7B from *Humicola insolens* [89]. Cel7B is an endoglucanase that catalyzes the hydrolysis of β -1,4-linked glucosidic bonds in cellulose with retention of stereochemistry at the anomeric center [90]. The E197A mutant has been shown to be a glycosynthase [89, 91]. We envisioned that chemical complementation could detect glycosynthase activity as formation of a bond between donor **17** (methotrexate–disaccharide–fluoride, Mtx-Lac-F) and acceptor **18** (dexamethasone–disaccharide, Dex-Cel) in a Dex-Mtx yeast three-hybrid system (Figure 7.10) [3]. For the acceptor compound **18**, cellobiose (the natural substrate of the glycosynthase) was used as the disaccharide with **1** attached at the anomeric position. For the α -glycosyl fluoride donor **17**, however, cellobiosyl fluoride could not be used, since it could also act as an acceptor and thus self-polymerize. Therefore, an epimer of cellobiose fluoride, lactosyl fluoride, was chosen that differed only in the stereochemistry at the 4' position. Ligand **2** was installed at the 6' position of the galactose unit to facilitate the chemical synthesis and because the crystal structure of Cel7B complexed with a substrate analog suggests that this position is exposed and, therefore, that the addition of **2** would not interfere with substrate binding. To validate the choice of substrates for this enzyme system, heterodimer **19** (Mtx-Lac-Cel-Dex) was synthesized *in vitro* from the donor and acceptor substrates **17** and **18** using a commercial preparation of the Cel7B:E197A glycosynthase. The complete CID activated transcription of the *LEU2* reporter gene in the Dex-Mtx three-hybrid system, demonstrating that the designed small molecules **17** and **18** were suitable for use in the glycosynthase assay.

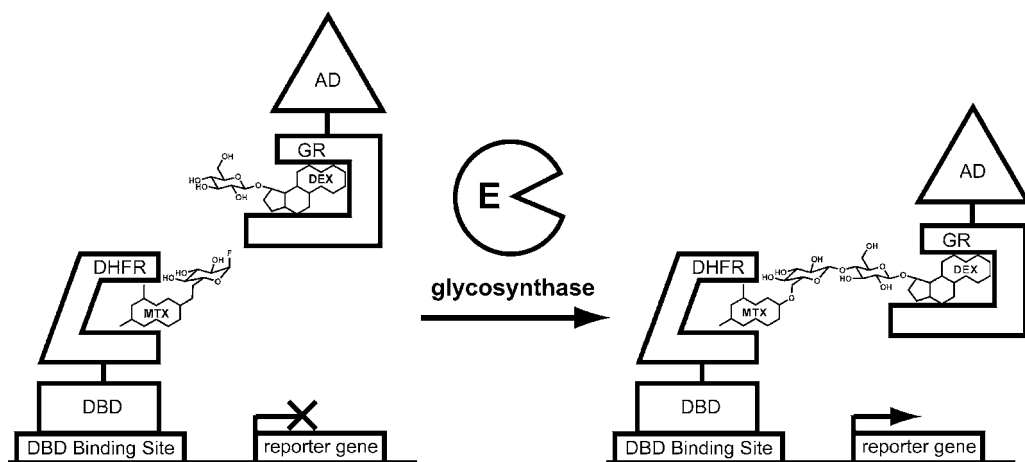


Fig. 7.10 Chemical complementation as a high-throughput assay for glycosynthase activity. Chemical complementation detects enzyme catalysis of bond formation or cleavage reactions based on covalent coupling of two small molecule ligands. The heterodimeric small molecule reconstitutes

a transcriptional activator, turning on the transcription of a downstream reporter gene. Here, a Dex-Mtx yeast three-hybrid system is used. Glycosynthase activity is detected as formation of a glycosidic linkage between the Mtx α -fluoride donor **17** and Dex cellobiose acceptor **18** [3].

7.4.3.3 Glycosynthase Assay

Once the CID for the glycosynthase was established, we could determine if chemical complementation could detect the glycosynthase activity of the Cel7B:E197A enzyme. Using standard conditions for a *LEU2* growth assay, we showed that expression of Cel7B:E197A glycosynthase in the presence of **17** and **18** conferred a growth advantage to the yeast three-hybrid selection strain (Figure 7.11) [3], presumably because the Cel7B:E197A glycosynthase catalyzed the synthesis of heterodimer **19**. Several control experiments were carried out to confirm that the growth advantage is indeed caused by the catalytic activity of the Cel7B:E197A glycosynthase. First, we showed that transcription activation required both the donor and the acceptor substrates. The yeast three-hybrid selection strain expressing the Cel7B:E197A glycosynthase was grown in the presence of no small molecule, 10 μ M **18** alone, 10 μ M **17** alone, or 10 μ M **17** and 10 μ M **18** in media lacking the appropriate auxotrophs and leucine. Activation of the *LEU2* reporter gene results in an increase in cell growth and hence in the OD₆₀₀ for the cell culture. Cell growth was at background levels for cells grown with no small molecule or with only **18** and near background levels with only **17**. Cells grown in the presence of both **17** and **18** showed a clear growth advantage.

Next, we showed that transcription activation is dependent on the catalytic activity of the Cel7B:E197A glycosynthase. Cells expressing the wild-type Cel7B glycosidase, which should be able to bind the two substrates, but should be much less efficient at product synthesis, showed no growth advantage in the

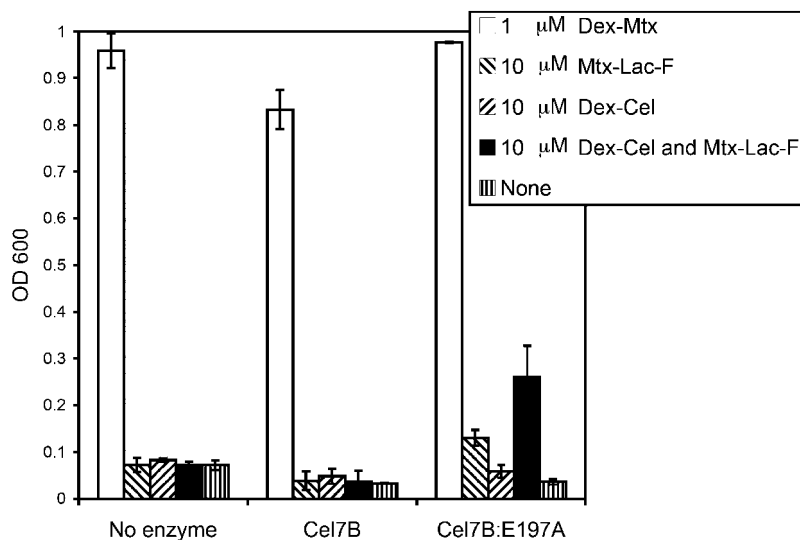


Fig. 7.11 Chemical complementation links Cel7B:E197A glycosynthase activity to *LEU2* reporter gene transcription *in vivo*. Yeast cells containing the DBD-DHFR and AD-GR fusion proteins and the *LEU2* reporter gene, expressing either no enzyme, Cel7B glycosidase, or Cel7B:E197A glycosynthase were grown in selective medium with or without the substrates **17** (methotrexate-lactose fluoride, Mtx-Lac-F) and **18** (dexamethasone-cellobiose, Dex-Cel). Small molecule dimeri-

zer **6** (Dex-Mtx), with a methylene linker, was used as a positive control. Cells expressing no enzyme or Cel7B glycosidase were used as negative controls. Activation of the *LEU2* reporter gene and hence cell growth in the absence of leucine is reported here as the OD₆₀₀ after 6 days of growth at 30°C. The growth assays were carried out in triplicate, and the error bars correspond to the standard error [3].

presence of **17** and **18**. Furthermore, cells expressing the Cel7B glycosidase were indistinguishable from cells expressing no enzyme. Together, these data suggest that the Cel7B:E197A glycosynthase activity can be detected using chemical complementation.

To demonstrate the ability of chemical complementation to carry out a high-throughput selection for glycosynthase activity, the growth advantage conferred by the Cel7B:E197A glycosynthase was used to select the glycosynthase from a pool of inactive variants. A mock library containing 100:1 Cel7B glycosidase to Cel7B:E197A glycosynthase was used to transform the yeast three-hybrid selection strain. Here, the Cel7B glycosidase was used as the “inactive” control. First, the transformed cells were plated under nonselective conditions to determine the transformation efficiency and ten random colonies were analyzed to establish the integrity of the library. All ten colonies were found to express the Cel7B glycosidase, as would be expected from the 100:1 library ratio, based on colony PCR and restriction digestion. Then the library was plated on selective media lacking the appropriate auxotrophs and leucine and containing substrates **17** and **18**. The

10 largest colonies were analyzed by colony PCR and restriction digestion. Of these 10 colonies, eight contained the Cel7B:E197A glycosynthase, and two contained the Cel7B glycosidase, which corresponds to an enrichment of 400-fold after a single round of selection. This result established that the system could select active glycosynthases from a large library and thus was suitable for directed evolution.

7.4.3.4 Directed Evolution

Having demonstrated that chemical complementation could select for glycosynthase activity, we applied the *LEU2* selection to the directed evolution of Cel7B with a E197 saturation library (Figure 7.11). Position 197 was randomized to all 20 amino acids using cassette mutagenesis with an NNK codon (N = G,A,T,C; K = T,G). This library was transformed into the yeast three-hybrid selection strain and plated on selective media lacking the appropriate auxotrophs and leucine and containing substrates **17** and **18**. The 96 largest colonies from the selection plate were arrayed in microtiter plates for secondary screening to eliminate false positives. In the secondary screen, growth of the selected mutants was monitored under selective conditions with and without the two substrates. Only those colonies that showed better growth with the two substrates were considered true positives and subjected to further characterization. Of the 96 colonies picked, 35 were active in the secondary screen. The Cel7B genes from those 35 positives were PCR amplified and sequenced. Three mutants, Ala, Gly, and Ser, occurred most frequently and represent 80% of the positive clones. In addition, the Leu mutant was isolated twice and the Pro, Asp, and Thr mutants were all isolated once. Of the 10 Ala mutants isolated, three of them also had a spontaneous N196D mutation. To confirm that these mutants are indeed glycosynthases, plasmids for individual mutants were isolated, transformed back into the yeast three-hybrid selection strain, and resubjected to the secondary screen with or without small molecules. All of these variants again showed a growth advantage with substrates **17** and **18**, with the N196D/E197A variant growing faster than the E197A variant.

The E197A, E197S and N196D/E197A variants were all subsequently overexpressed in yeast and purified to allow determination of their glycosynthase activities. These three variants were assayed *in vitro* using a standard glycosynthase assay with the substrates **20** (*α*-lactosyl fluoride) and **21** (*p*-nitrophenyl- β -cellobioside) [91]. These two substrates are structurally similar to **17** and **18**, respectively. The original Cel7B:E197A glycosynthase has a specific activity of 8 ± 2 mol product min^{-1} mol^{-1} enzyme. The E197S variant isolated from this selection shows a significant improvement in activity, with a specific activity of 40 ± 5 mol min^{-1} mol^{-1} . The N196D/E197A variant, somewhat surprisingly, is within experimental error of E197A, with a specific activity of 7 ± 1 mol min^{-1} mol^{-1} . Thus, we were able to increase the activity of the glycosynthase five-fold in this first directed evolution experiment. We are presently expanding our efforts toward screening larger libraries through DNA shuffling, by which we expect to find mutants of significantly higher activity.

7.4.3.5 General Considerations

For any high-throughput assay, a secondary screen is important for ruling out false positives. The secondary screen for glycosynthase activity simply consisted of the chemical complementation growth assay, repeated in both the presence and absence of small molecules **17** and **18**. Only those clones that exhibited small-molecule-dependent growth were considered to be true positives. Additionally, it is important to demonstrate that cell growth is enzyme-dependent and not due to random mutation in the strain. This was easily accomplished by isolating the plasmid of the putative active enzyme and retransforming it into the selection strain to repeat the assay.

The number of mutants that can be screened or selected in chemical complementation is presently limited by the transformation efficiency of yeast. While this was not a concern for the small library considered here, the largest libraries that can be considered are on the order of 10^5 – 10^7 . For the E197 saturation library, the transformation efficiency was 10^6 colonies per microgram of DNA, showing that the yeast three-hybrid selection strain is not impaired in its transformation efficiency. Yet, for larger library sizes (10^8 – 10^{10}), the system will have to be moved to *E. coli*, and we have made efforts in that direction [92].

7.5

Conclusion

As we have seen, complementation approaches are a powerful means to select active enzymes from a large library of proteins. Here, we have presented chemical complementation as a reaction-independent assay that links enzyme activity to cell survival. The desired chemistry is programmed into a small molecule such that bond cleavage or bond formation reconstitutes a yeast three-hybrid transcriptional activator that drives expression of an essential reporter gene. The readout system, Mtx-Dex, LexA-DHFR, B42-GR and the reporter gene can all remain constant while the chemistry changes. Thus, all that needs to be changed for each new reaction is the Mtx-substrate-Dex, or Dex-substrate and Mtx-substrate molecules synthesized in the lab, and the enzyme library. This assay could be used to engineer glycosyltransferases, aldolases, esterases, amidases, and “diels-alderases,” all with a variety of substrate specificities and regio- and stereoselectivities. By converting the assay to a coupled enzyme assay, it may even be possible to detect oxidases and reductases.

In addition to providing a powerful selection for the evolution of enzymes with new activities, there should be many uses for a reaction-independent high-throughput assay for enzyme catalysis. The assay can be used to study enzyme function, either to test hundreds of mutants to identify amino acids important for the catalytic activity of an enzyme or hundreds of different molecules to determine the substrate specificity of an enzyme. Likewise, the assay could be applied to drug discovery by screening libraries of small molecules on the basis of inhibition of enzyme activity and a change in transcription of the reporter gene.

A distinct advantage is that a new assay would not have to be developed for each new enzyme target.

Finally, this assay should be particularly well suited to proteomics. A battery of Mtx-Dex substrates with different substrates as linkers could be prepared and then used to screen cDNA libraries to identify enzymes that fall into common families, such as glycosidases or aldolases. Because mammalian, as well as yeast, three-hybrid assays are standard, the assays could be carried out in the endogenous cell line, ensuring correct posttranslational modification of the proteins. The key to all of these applications is a robust assay for enzymatic activity. Chemical complementation promises to be a powerful high-throughput assay for a wide variety of enzymatic reactions.

References

- 1 H. Tao, V.W. Cornish, *Curr. Opin. Chem. Biol.* **2002**, 6, 858–864.
- 2 K. Baker, C. Bleczinski, H. Lin, G. Salazar-Jimenez, D. Sengupta, S. Krane, V.W. Cornish, *Proc. Natl Acad. Sci. USA* **2002**, 99, 16537–16542.
- 3 H. Lin, H. Tao, V.W. Cornish, *J. Am. Chem. Soc.* **2004**, 126, 15051–15059.
- 4 G.W. Beadle, E.L. Tatum, *Proc. Natl Acad. Sci. USA* **1941**, 27, 499–506.
- 5 S. Benzer, *Sci. Am.* **1962**, 206, 10–84.
- 6 P. Uetz, L. Giot, G. Cagney, T.A. Mansfield, R.S. Judson, J.R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, J.M. Rothberg, *Nature* **2000**, 403, 623–627.
- 7 W.A. Lim, R.T. Sauer, *Nature* **1989**, 339, 31–36.
- 8 P. Kast, M. Asif-Ullah, N. Jiang, D. Hilvert, *Proc. Natl Acad. Sci. USA* **1996**, 93, 5043–5048.
- 9 P. Kast, C. Grisostomi, I.A. Chen, S. Li, U. Krengel, Y. Xue, D. Hilvert, *J. Biol. Chem.* **2000**, 275, 36832–36838.
- 10 B.G. Hall, *Genetics* **1978**, 89, 453–465.
- 11 R.M. Myers, L.S. Lerman, T. Maniatis, *Science* **1985**, 229, 242–247.
- 12 J.D. Hermes, S.M. Parekh, S.C. Blacklow, H. Koster, J.R. Knowles, *Gene* **1989**, 84, 143–151.
- 13 J.D. Hermes, S.C. Blacklow, J.R. Knowles, *Proc. Natl Acad. Sci. USA* **1990**, 87, 696–700.
- 14 T. Yano, S. Oue, H. Kagamiyama, *Proc. Natl Acad. Sci. USA* **1998**, 95, 5511–5515.
- 15 G. DeSantis, K. Wong, B. Farwell, K. Chatman, Z. Zhu, G. Tomlinson, H. Huang, X. Tan, L. Bibbs, P. Chen, K. Kretz, M.J. Burk, *J. Am. Chem. Soc.* **2003**, 125, 11476–11477.
- 16 S.W. Santoro, L. Wang, B. Herberich, D.S. King, P.G. Schultz, *Nat. Biotechnol.* **2002**, 20, 1044–1048.
- 17 S.L. Dove, J.K. Joung, A. Hochschild, *Nature* **1997**, 386, 627–630.
- 18 J.K. Joung, E.I. Ramm, C.O. Pabo, *Proc. Natl Acad. Sci. USA* **2000**, 97, 7382–7387.
- 19 S.A. Lesley, P.A. Patten, P.G. Schultz, *Proc. Natl Acad. Sci. USA* **1993**, 90, 1160–1165.
- 20 J.A. Smiley, S.J. Benkovic, *Proc. Natl Acad. Sci. USA* **1994**, 91, 8319–8323.
- 21 Y. Tang, J.B. Hicks, D. Hilvert, *Proc. Natl Acad. Sci. USA* **1991**, 88, 8784–8786.
- 22 J. Kaplan, W.F. DeGrado, *Proc. Natl Acad. Sci. USA* **2004**, 101, 11566–11570.
- 23 M.A. Dwyer, L.L. Looger, H.W. Hellinga, *Science* **2004**, 304, 1967–1971.
- 24 C.A. Voigt, S.L. Mayo, F.H. Arnold, Z.G. Wang, *Proc. Natl Acad. Sci. USA* **2001**, 98, 3778–3783.
- 25 B. Kuhlman, G. Dantas, G.C. Ireton, G. Varani, B.L. Stoddard, D. Baker, *Science* **2003**, 302, 1364–1368.
- 26 A.D. Griffiths, D.S. Tawfik, *EMBO J.* **2003**, 22, 24–35.

- 27 H. Pedersen, S. Holder, D.P. Sutherlin, U. Schwitter, D. S. King, P.G. Schultz, *Proc. Natl Acad. Sci. USA* **1998**, *95*, 10523–10528.
- 28 D. S. Tawfik, A.D. Griffiths, *Nat. Biotechnol.* **1998**, *16*, 652–656.
- 29 M.J. Olsen, D. Stephens, D. Griffiths, P. Daugherty, G. Georgiou, B.L. Iverson, *Nat. Biotechnol.* **2000**, *18*, 1071–1074.
- 30 S.M. Firestine, F. Salinas, A.E. Nixon, S.J. Baker, S.J. Benkovic, *Nat. Biotechnol.* **2000**, *18*, 544–547.
- 31 G. Xia, L. Chen, T. Sera, M. Fa, P.G. Schultz, F.E. Romesberg, *Proc. Natl Acad. Sci. USA* **2002**, *99*, 6597–6602.
- 32 F.J. Ghadessy, J.L. Ong, P. Holliger, *Proc. Natl Acad. Sci. USA* **2001**, *98*, 4552–4557.
- 33 S.W. Michnick, F.X. Valois, *Proc. Natl Acad. Sci. USA* **2002**, *99*, 16513–16515.
- 34 B.T. Carter, H. Lin, V.W. Cornish, in *Directed Molecular Evolution of Proteins*, eds K. Brakmann, K. Johnsson, Wiley-VCH, Weinheim, **2002**.
- 35 E. Licitra, J. Liu, *Proc. Natl Acad. Sci. USA* **1996**, *93*, 12817–12821.
- 36 S. Fields, O. Song, *Nature* **1989**, *340*, 245–246.
- 37 R. Brent, M. Ptashne, *Cell* **1985**, *43*, 729–736.
- 38 I. Hope, K. Struhl, *Cell* **1986**, *46*, 885–894.
- 39 S.J. Triezenberg, R. Kingsbury, S. McKnight, *Genes Dev.* **1988**, *2*, 718–729.
- 40 D.M. Spencer, T.J. Wandless, S.L. Schreiber, G.R. Crabtree, *Science* **1993**, *262*, 1019–1024.
- 41 M.T. Ryan, H. Muller, N. Pfanner, *J. Biol. Chem.* **1999**, *274*, 20619–20627.
- 42 J.N. Pelletier, F.X. Campbell-Valois, S.W. Michnick, *Proc. Natl Acad. Sci. USA* **1998**, *95*, 12141–12146.
- 43 S.P. Sasso, R.M. Gilli, J.C. Sari, O.S. Rimet, C.M. Briand, *Biochim. Biophys. Acta* **1994**, *1207*, 74–79.
- 44 P.K. Chakraborti, M.J. Garabedian, K.R. Yamamoto, S.S. Simons Jr., *J. Biol. Chem.* **1991**, *266*, 22075–22078.
- 45 E.J. Rebar, C.O. Pabo, *Science* **1994**, *263*, 671–673.
- 46 M. Johnston, *Microbiol. Rev.* **1987**, *51*, 458–476.
- 47 J. Estojak, R. Brent, E.A. Golemis, *Mol. Cell. Biol.* **1995**, *15*, 5820–5829.
- 48 H. Lin, W.M. Abida, R.T. Sauer, V.W. Cornish, *J. Am. Chem. Soc.* **2001**, *122*, 4247–4248.
- 49 W.M. Abida, B.T. Carter, E.A. Althoff, H. Lin, V.W. Cornish, *ChemBioChem* **2002**, *3*, 887–895.
- 50 K. Baker, D. Sengupta, G. Salazar-Jimenez, V.W. Cornish, *Anal. Biochem.* **2003**, *315*, 134–137.
- 51 K.S. de Felipe, B.T. Carter, E.A. Althoff, V.W. Cornish, *Biochemistry* **2004**, *43*, 10353–10363.
- 52 O. Futer, M.T. DeCenzo, R.A. Aldape, D.J. Livingston, *J. Biol. Chem.* **1995**, *270*, 18935–18940.
- 53 D.A. Holt, J.I. Luengo, D.S. Yamashita, H.J. Oh, A.L. Konilian, H.K. Yen, L.W. Rozamus, M. Brandt, M.J. Bossard, M.A. Levy, D.S. Eggleston, J. Liang, L.W. Schultz, T.J. Stout, J. Clardy, *J. Am. Chem. Soc.* **1993**, *115*, 9925–9938.
- 54 P.D. Braun, K.T. Barglow, Y.M. Lin, T. Akompong, R. Briesewitz, G.T. Ray, K. Haldar, T.J. Wandless, *J. Am. Chem. Soc.* **2003**, *125*, 7575–7580.
- 55 F. Becker, K. Murthi, C. Smith, J. Come, N. Costa-Roldan, C. Kaufmann, U. Hanke, C. Degenhart, S. Baumann, W. Wallner, A. Huber, S. Dedier, S. Dill, D. Kinsman, M. Hediger, N. Bockovich, S. Meier-Ewert, A.F. Kluge, N. Kley, *Chem. Biol.* **2004**, *11*, 211–223.
- 56 S. Lefurgy, V.W. Cornish, *Chem. Biol.* **2004**, *11*, 151–153.
- 57 S.L. Hussey, S.S. Muddana, B.R. Peterson, *J. Am. Chem. Soc.* **2003**, *125*, 3692–3693.
- 58 S. Gendreizig, M. Kindermann, K. Johnsson, *J. Am. Chem. Soc.* **2003**, *125*, 14970–14971.
- 59 M. Page, *The Chemistry of the β -Lactams*, Chapman & Hall, Glasgow, **1992**.
- 60 M. Galleni, J.M. Frere, *Biochem. J.* **1988**, *255*, 119–122.
- 61 E. Lobkovsky, P.C. Moews, H. Liu, H. Zhao, J.M. Frere, J.R. Knox, *Proc. Natl Acad. Sci. USA* **1993**, *90*, 11257–11261.
- 62 W. Durckheimer, F. Adam, G. Fischer, R. Kirrstetter, *Adv. Drug Res.* **1988**, *17*, 61–234.
- 63 D. Sengupta, H. Lin, S.D. Goldberg, J.J. Mahal, V.W. Cornish, *Biochemistry* **2004**, *43*, 3570–3581.

- 64 D. D. Clark, B. R. Peterson, *ChemBioChem* **2003**, *4*, 101–107.
- 65 S. Athavankar, B. R. Peterson, *Chem. Biol.* **2003**, *10*, 1245–1253.
- 66 B. Azizi, E. I. Chang, D. F. Doyle, *Biochem. Biophys. Res. Commun.* **2003**, *306*, 774–780.
- 67 Z. Zhang, W. Zhu, T. Kodadek, *Nat. Biotechnol.* **2000**, *18*, 71–74.
- 68 J. M. Ghuysen, *Annu. Rev. Microbiol.* **1991**, *45*, 37–67.
- 69 M. Nukaga, S. Kumar, K. Nukaga, R. F. Pratt, J. R. Knox, *J. Biol. Chem.* **2004**, *279*, 9344–9352.
- 70 R. A. Powers, E. Caselli, P. J. Focia, F. Prati, B. K. Shoichet, *Biochemistry* **2001**, *40*, 9207–9214.
- 71 Z. Zhang, Y. Yu, J. M. Musser, T. Palzkill, *J. Biol. Chem.* **2001**, *276*, 46568–46574.
- 72 B. T. Carter, H. Lin, S. D. Goldberg, E. A. Althoff, J. Raushel, V. W. Cornish, *ChemBioChem* **2005**, in press.
- 73 T. Delaveau, A. Delahodde, E. Carvajal, J. Subik, C. Jacq, *Mol. Gen. Genet.* **1994**, *244*, 501–511.
- 74 H. B. van den Hazel, H. Pichler, M. A. do Valle Matta, E. Leitner, A. Goffeau, G. Daum, *J. Biol. Chem.* **1999**, *274*, 1934–1941.
- 75 A. Nourani, M. Wesolowski-Louvel, T. Delaveau, C. Jacq, A. Delahodde, *Mol. Cell. Biol.* **1997**, *17*, 5453–5460.
- 76 J. Kato-Stankiewicz, I. Hakimi, G. Zhi, J. Zhang, I. Serebriiskii, L. Guo, H. Edamatsu, H. Koide, S. Menon, R. Eckl, S. Sakamuri, Y. Lu, Q. Z. Chen, S. Agarwal, W. R. Baumbach, E. A. Golemis, F. Tamanoi, V. Khazak, *Proc. Natl Acad. Sci. USA* **2002**, *99*, 14398–14403.
- 77 S. L. Flitsch, *Curr. Opin. Chem. Biol.* **2000**, *4*, 619–625.
- 78 H. J. Gijzen, L. Qiao, W. Fitz, C. H. Wong, *Chem. Rev.* **1996**, *96*, 443–474.
- 79 K. M. Koeller, C. H. Wong, *Chem. Rev.* **2000**, *100*, 4465–4494.
- 80 K. M. Koeller, C. H. Wong, *Nature* **2001**, *409*, 232–240.
- 81 P. Sears, C. H. Wong, *Science* **2001**, *291*, 2344–2350.
- 82 L. F. Mackenzie, Q. Wang, R. A. J. Warren, S. G. Withers, *J. Am. Chem. Soc.* **1998**, *120*, 5583–5585.
- 83 M. Fajjes, J. K. Fairweather, H. Driguez, A. Planas, *Chemistry* **2001**, *7*, 4651–4655.
- 84 G. J. Davies, S. J. Charnock, B. Henrissat, *Trends Glycosci. Glycotechnol.* **2001**, *13*, 105–120.
- 85 J. K. Fairweather, R. V. Stick, S. G. Withers, *Aust. J. Chem.* **2000**, *53*, 913–916.
- 86 C. Mayer, D. L. Zechel, S. P. Reid, R. A. Warren, S. G. Withers, *FEBS Lett.* **2000**, *466*, 40–44.
- 87 O. Nashiru, D. L. Zechel, D. Stoll, T. Mohammadzadeh, R. A. Warren, S. G. Withers, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 417–420.
- 88 A. Trincone, G. Perugino, M. Rossi, M. Moracci, *Bioorg. Med. Chem. Lett.* **2000**, *10*, 365–368.
- 89 S. Fort, V. Boyer, L. Greffe, G. J. Davies, O. Moroz, L. Christiansen, M. Schulein, S. Cottaz, H. Driguez, *J. Am. Chem. Soc.* **2000**, *122*, 5429–5437.
- 90 C. Schou, G. Rasmussen, M. B. Kaltoft, B. Henrissat, M. Schulein, *Eur. J. Biochem.* **1993**, *217*, 947–953.
- 91 V. M. Ducros, C. A. Tarling, D. L. Zechel, A. M. Brzozowski, T. P. Frandsen, I. von Ossowski, M. Schulein, S. G. Withers, G. J. Davies, *Chem. Biol.* **2003**, *10*, 619–628.
- 92 E. A. Althoff, V. W. Cornish, *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 2327–2330.

8

Molecular Approaches for the Screening of Novel Enzymes

Valéria Maia de Oliveira and Gilson Paulo Manfio

8.1

Introduction

Microorganisms have proved to be an exceptionally rich source of useful products, and they will continue to be so in the future. Such products vary enormously, both in terms of structural complexity and biological activity. Although microbes have played an important role as a source of known primary metabolites, such as amino acids and vitamins, the quest for novel compounds has focused on the products of secondary metabolism, which are considered to be nonessential compounds.

Exploitation of microorganisms as a source of chemical compounds began in the 1940s as the significance of the discovery of penicillin was realized [1]. During this period, screening programs were mainly directed towards the discovery of antimicrobial activities, and were successful in identifying many new antibiotics, such as cephalosporins and tetracyclins, and compounds for the treatment of cancer and fungal infections. More recently, the trend has been the quest for compounds having pharmacological activities, such as antihypertensive and hypocholesteremic agents, and immunomodulators [2]. The number of reported microbial metabolites with non-antibiotic biological activities has increased enormously since the early 1970s, and by 1990 it had already exceeded the reported number of antibiotics [3].

Regardless of the type of activity sought, selectivity and sensitivity are two key issues to consider in the choice of a screening strategy. Selectivity is necessary to allow detection of a single compound (or enzymatic reaction) in the presence of many others, while sensitivity allows for detection of very low concentrations of the target compound.

Advances in analytical and information technology, and the rapid progress of whole-genome sequence projects, now provide access to a much wider range of new molecular targets, including human diseases and specific microbial groups, and new opportunities for the exploitation of the molecular diversity contained in microbial cells, including metabolites, enzymes, and enzymatic reactions.

The current chapter will supplement information provided in other chapters that focused mainly on traditional and novel screening strategies and on methods based upon the detection of products formed by enzymatic reactions and pathways, by providing an introduction to the application of gene probes and metagenomic screening in the search for enzyme-coding genes and surrogate-host expression of enzymatic reactions directly from environmental DNA.

8.2

Use of Nucleic Acid Probes to Detect Enzyme-coding Genes in Cultivated Microorganisms

DNA probe technology fulfills both of the required qualities of a good screening program: selectivity and sensitivity. A probe is a labeled fragment of nucleic acid (DNA or RNA) with a nucleotide sequence complementary to that of a target gene sequence of interest. A gene is a fragment of DNA that contains genetic information for making a specific peptide, protein, or structural ribonucleic acid (rRNA, tRNA). Under standardized conditions, probe and single-stranded target gene sequence hybridize, forming heteroduplexes (double-stranded fragments comprising DNA strands from different sources, that is, a strand of probe DNA annealed to a strand from the target DNA). These are maintained as stable double-stranded DNA by hydrogen bonds held between the complementary bases. The selectivity or specificity (stringency) of such a hybridization reaction can be controlled by manipulating parameters that affect the stability of the heteroduplex formed, mainly temperature, denaturant (e.g. formamide) and salt concentration. Other parameters, such as pH, probe size and concentration, may also affect the reaction.

Stringency is a term used to denote the conditions in which hybridization experiments are carried out. Under high stringent hybridization conditions, that is, high temperature and/or denaturant and low salt concentrations, only heteroduplexes that have a high percentage homology between probe and target nucleic acid sequence are maintained under stable annealing conditions, and thus detected as a positive hybridization signal.

A positive hybridization reaction is recognized by means of a detectable label attached to the hybridized probe molecules. The nucleic acid probe may be labeled directly or indirectly. In the former, a label is covalently bound to the probe, whereas in the latter an unlabeled reporter group attached to the probe is detected by a labeled reporter-binding molecule. Thus, detection of the probe–target heteroduplex can be based upon detecting the direct signal (e.g. radioactivity, fluorescence, enzymatic reaction) from the labeled deoxynucleotides present in the probe sequence, or from a secondary reaction of the labeled reporter-binding molecule (such as an enzyme-linked or fluorescently-labeled antibody) with the reporter moieties in the probe. Initially, radioactive isotopes were used as markers, and these are still preferred in research studies that require high sensitivity. However, problems related to safety, lack of stability

and waste disposal led to the development of non-radioactive labeling and detection systems, several of which are now commercially available. These comprise a wide choice of labels, including enzymes (e.g. alkaline phosphatase [4] and horseradish peroxidase [5]), fluorescent molecules (e.g. fluorescein [6]), biotin, and digoxigenin [7].

The probe sequence may be originated from a DNA fragment cloned at random, with unknown function. This is generally applicable only to the species it was derived from. Alternatively, a probe may be a cloned fragment of a known gene, which confers a specific characteristic to a group of organisms, such as genes that code for the production of enzymes and proteins, toxins, secondary metabolites, and antibiotic resistance responses, amongst others.

Yet a probe may be a synthetic oligonucleotide sequence (14–40 nucleotides in length) based upon a DNA sequence of interest. The oligonucleotides offer the advantage of rapidly hybridizing with the target nucleic acid sequence, making the hybridization reaction much faster compared with the use of longer probes [8]. In addition, they can be highly selective, such that, under high stringency conditions, oligonucleotide probes may detect a single base pair difference in the target nucleic acid sequence [9]. A disadvantage of a short probe is that it offers lower detection sensitivity, since only a small number of labels or reporter groups can be directly attached to the probe fragment. A basic requirement for the use of oligonucleotide probes in the search for specific genes is to have previous knowledge of the DNA sequences that encode for the proteins of interest.

8.2.1

Current Knowledge and Applications

Nucleic acid probes may be used to detect homologous gene sequences (gene sequences from different organisms derived from a common ancestor by speciation; high sequence similarity may be found in closely related microbial species) in isolated microbial strains, in DNA extracted from environmental samples, or in the screening of cloned fragments in a genomic library. Probes may be designed to target different taxonomic levels from species up to families. The use of nucleic acid probes may facilitate an enzyme screening program, since they can be designed to be highly specific, and stringency conditions may be adjusted if necessary.

The probe technology can be applied to the study of enzymes in different ways. DNA probes can be used: (a) to determine the presence and location of an enzyme-coding gene in the genome of an organism; (b) to evaluate the distribution of an enzyme-coding gene among different microbial taxa, acting as a marker for epidemiological studies [10]; (c) as taxonomic markers in genetically-based typing schemes [11]; and (d) in the search for novel, but similar enzyme-coding sequences in microbial isolates [12, 13]. Applications of enzyme-targeted probes come mainly from studies on resistance genes from clinically significant microorganisms.

Kaufhold and co-workers [10] screened clinical isolates of *Enterococcus faecalis* and *E. faecium* for high-level gentamicin resistance using dot-blot hybridization assays and non-radioactive oligonucleotide probes, targeting the structural gene coding for the bifunctional aminoglycoside-modifying 6'-acetyltransferase-2''-phosphotransferase (6'AAC-2''APH) enzyme. They showed that isolates with and without high-level gentamicin resistance could be discerned unequivocally. In a search for genes coding for multiple resistance to antibiotics in enterococci, the use of probes derived from staphylococcal β -lactamase and gentamicin resistance genes indicated their occurrence and chromosomal location in all enterococcal strains tested [14]. Teran et al. [11] constructed a probe targeting the gene coding for an aminoglycoside 6'-N-acetyltransferase able to modify amikacin, cloned from a plasmid isolated from a clinical strain of *Enterobacter cloacae*. Hybridization assays revealed a high specificity of the probe towards gene sequences of strains harboring related enzymes, indicating its potential application for epidemiological studies and discovery of novel sequences of related enzymes. Weingart and Hooke [13] used a probe targeting the gene that codes for a non-haemolytic phospholipase C secreted by *Burkholderia cepacia* Pc224c, a potential virulence factor, to screen other *B. cepacia* strains in hybridization experiments. Nine out of the 10 *B. cepacia* strains tested reacted with the probe, suggesting the prevalence of similar sequences in other isolates.

Shiffman et al. [12] used hybridization probes prepared from isopenicillin-N-synthase (IPNS) genes of *Streptomyces clavuligerus*, *S. jumonjinensis*, and *S. lipmanii* to detect analogous genes in other streptomycetes, including both penicillin and cephalosporin producers and non-producers. All producer strains responded positively in the hybridization assays and several non-producer strains also responded to one or to both probes, suggesting that IPNS-related genes may be more prevalent in *Streptomyces* than previously thought.

An innovative use of the probe technology was shown by Hönerlage et al. [15]. These authors demonstrated, by using whole-cell hybridization, the presence of messenger RNA (the mRNA is the single-stranded ribonucleic acid template that derives information from the DNA for the protein synthesis) of the gene coding the major extracellular protease (*nprM*) in growing cells of *Bacillus megaterium* germinated from heat-activated spores in soil. These results demonstrated the applicability of nucleic acid-probing techniques to localize microbial-mediated enzymatic activities in soil, targeting the probe to the expressed mRNA, instead of targeting the gene itself.

8.2.2

Limitations of Probe Technology and the Need for Innovative Approaches

Cultured soil microorganisms are the most common resources for biotechnological exploration of novel natural products and processes [16]. These microbial compounds include a wide range of traditionally industrially important enzymes, such as cellulases, amylases, proteases, and lipases. Other microbial enzymes are important for academic research and the development of biotechno-

logical applications, including restriction enzymes and nucleic acid polymerases, which are routinely used in recombinant DNA technology.

Traditionally, enzymatic activities have been accessed by methods involving isolation and culturing of the organisms from environmental samples, mainly soil. Isolates and/or fermentation broths are then screened for the desired activities. Most enzymes used industrially have been discovered in this way. Despite the apparent success of this approach, severe limitations have been recognized. Some biotransformation and biocatalytic reactions still remain impractical, due to deficiencies associated with the enzymes, such as low expression rate or intracellular location. The discovery of new or improved activity profiles often requires novel enzymes, and the culturing approach appears to recover mostly known microorganisms, so that known enzymes are rediscovered again and again.

The high “rediscovery rate” of natural products, mainly secondary metabolites, from cultivated organisms in the last 20 years led to the idea that soil microorganisms had been “mined out” for new products [17]. However, by the early 1990s, scientists had realized that the direct visualization of microorganisms in a natural sample by staining and microscopy yielded a population count one to four orders of magnitude higher than that measured by culturing from the same sample [18–20]. This means that only 1–10% of the microbial diversity is retrievable from an environmental sample by current culturing techniques.

The “great plate counting anomaly” may be explained by several hypotheses. One possibility is that “uncultured” cells are microorganisms that are phylogenetically similar or identical to the cultured minority but in a physiological state that makes them recalcitrant to culturing, usually due to adverse conditions, for example starvation or temperature stresses [21]. Another hypothesis is that the remaining cells represent novel lineages of bacteria that are phylogenetically distinct from the cultured members of the community, and cannot be cultivated using current standard media. In addition, all the methods and media available in a standard microbiology laboratory are selective to a certain extent, and simply cannot reproduce the conditions necessary to recover the vast majority of bacteria from an environmental sample. Certain microorganisms can only live in consortia, where one depends upon the metabolic products produced by the other in order to survive and multiply. Indeed, there is a huge number of novel microbial types in natural samples yet to be discovered by microbiologists.

Studies of DNA sequence information from organisms in soil microhabitats and their gene expression under different conditions may provide clues for designing new and improved culturing methods that resemble their natural niches. However, the development of techniques for culturing the enormous diversity of soil microbiota will be slow and tedious, and will require more knowledge about the physiology of the unknown microbes than is currently at hand [22–24]. Alternative approaches that offer more direct, global, and rapid methods for accessing the genetic diversity of the environmental samples, overcoming the limitations imposed by traditional culturing techniques, will greatly enhance the discovery of novel properties, allowing access to the genetic resources of uncultured microorganisms.

8.3

The Microbial Metagenome: a Resource of Novel Natural Products and Enzymes

8.3.1

Accessing the Uncultivated Biodiversity: the Community DNA Concept

Novel culture-independent molecular biology-based methods developed over the last decades, for example fluorescence *in situ* hybridization (FISH) and 16S rDNA clone libraries, triggered a series of ecological studies that drastically changed the emerging view of microbial diversity. These techniques may involve the direct detection of microorganisms in samples by using probes (FISH) or the extraction of the collective genomes (DNA material) of all microorganisms present in a habitat at a certain time point: the community DNA or “metagenome” [25].

Subsequent polymerase chain reaction (PCR)-based studies, in which a specific gene or DNA sequence is amplified millions of times using the metagenome as a template, have revealed that there is a spectacular diversity amongst uncultured microbes. In this type of study, amplification of the gene or DNA segment of interest via PCR is followed by cloning and determination of its sequence. In brief, cloning means inserting a foreign DNA sequence, by means of recombinant DNA techniques, into a vector (usually a plasmid) containing genetic markers (e.g. antibiotic resistance genes) for the selection of the vector–insert assembly. The vector is, in turn, introduced into a host cell, which is then multiplied in culture and plated onto selective media (rich culture medium containing the antibiotic for selection of the marker) in order to produce millions of cells harboring the foreign DNA (Figure 8.1).

The cloning of ribosomal genes (rDNA) in order to identify uncultivated microorganisms from environmental samples, and to map their phylogenetic relationships on the basis of comparisons between the rDNA sequences obtained and sequence data from organisms represented in public databases, was first introduced by Pace and colleagues [26]. This approach has subsequently been adopted as a standard tool to analyze the diversity of many different environments, demonstrating an astonishing diversity of new microbial groups that had never been found by cultivation [19, 20, 27–30].

As the molecular ecological studies progress, it becomes evident that many of the novel lineages are widely distributed geographically and often represent abundant components of microbial populations [20, 31]. However, information derived from molecular microbial ecology studies still allow us to gain restricted knowledge on the metabolic potential of the new microbial groups detected in the environment, since isolation and culturing efforts are still very much behind for most of the taxa discovered [32].

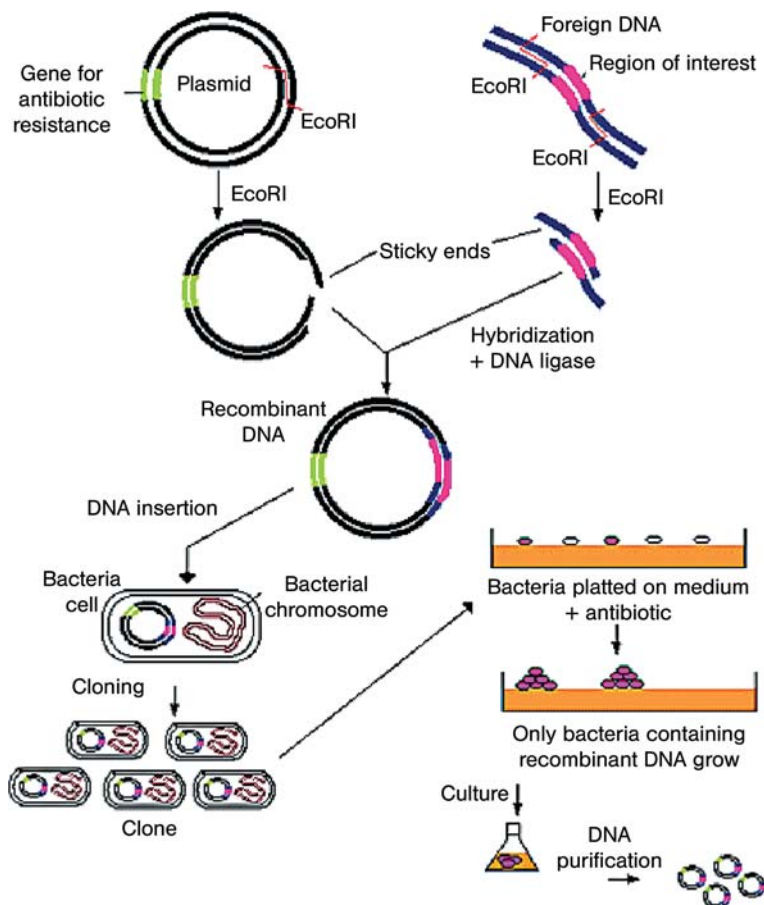


Fig. 8.1 Scheme illustrating the basic steps of the cloning technique, a procedure commonly used in molecular biology laboratories.

8.3.2

Unravelling Metabolic Function: the BAC Strategy

A molecular strategy that offers an alternative for exploring the metabolic potential of microorganisms that are not recovered by traditional culturing techniques (the uncultured microorganisms) involves the use of bacterial artificial chromosome (BAC) vectors for cloning very large DNA fragments (20 kb up to >500 kb), carrying entire gene clusters or even a complete biochemical pathway from the environmental samples [25, 33, 34]. The strategy involves cloning these large DNA fragments originating from environmental community DNA (metagenome) and analyzing the resulting complex metagenomic libraries in a search for novel enzymatic activities (Figure 8.2) [35].

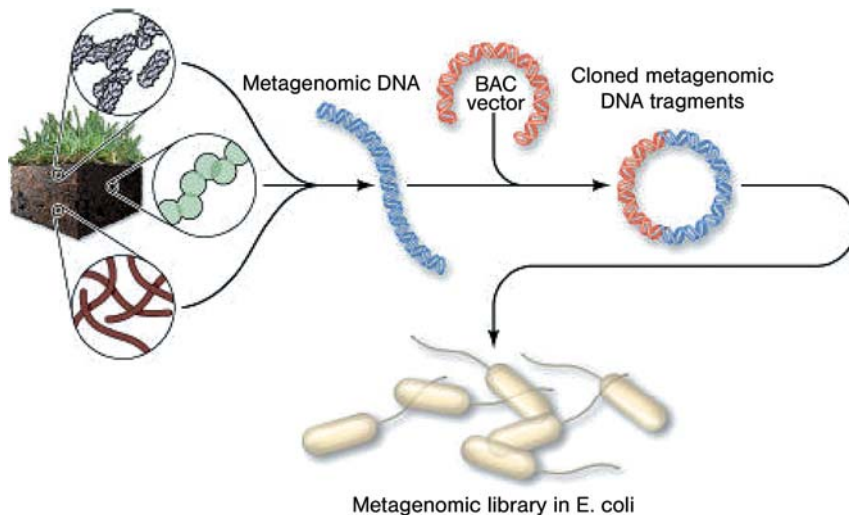


Fig. 8.2 Assembly of a metagenomic library. Large DNA fragments from uncultivated soil microorganisms are ligated into BAC vectors and inserted into *Escherichia coli* host cells. (After [35], reprinted with permission from the American Society for Microbiology. *ASM News*, January 2002, pp. 20–24).

BACs are specialized cloning systems derived from the *Escherichia coli* F plasmid that regulate their own replication at a copy number of one or two per cell. BACs also have features typical of most cloning vectors, including a multiple-cloning site and rapid clone selection mechanisms [36]. Originally developed for chromosome mapping and analysis of the human genome, BACs can maintain DNA inserts exceeding 300 kilobase pairs [37], and allow expression of prokaryotic genes in a heterologous bacterial host [38], that is, a bacterial strain of a different species from that of which the genes were originally cloned, which may have different genetic expression and regulation systems. These vectors are capable of stably maintaining large fragments of DNA inserts in *E. coli*, and are amongst the most commonly used vectors in bacterial and eukaryotic whole-genome projects [36, 37]. Furthermore, BAC clones allow genomic studies of certain prokaryotes, such as many Archaea, whose genomes were previously unclonable due to their instability in high copy number plasmids [36].

The low copy number of BACs inside the host cell is a desirable mechanism for avoiding genetic rearrangements between cloned inserts. However, an important disadvantage of BACs for functional screening studies is its low yield of DNA, both for vectors alone as well as for constructs carrying genomic inserts. Wild et al. [39] described a new class of BAC vectors that retain all the advantages of low copy BAC vectors, but are endowed with a conditional and tightly controlled amplification system based on the *oriV* origin of replication, which is in turn dependent on the TrfA replication protein. This vector system allows a

yield of ~100 copies of vector per host cell when conditionally induced with L-arabinose. Amplifiable clones and libraries facilitate high-throughput DNA sequencing and enzyme activity-based screening, due to higher levels of gene expression. This system can also be further regulated by glucose and fucose (copy number inhibitors), enhancing the stability of BAC maintenance in the host cell and making the system robust and reliable.

Complex genomic libraries that contain large DNA inserts in BAC vectors (up to 40 kb, and sometimes more than 100 kb) have been constructed from marine environments and provided a significant insight into the genomic potential of uncultivated planktonic marine archaea [40–43]. The environmental cloning approach has also been successfully used for studying soil microbial communities [34, 44, 45]. However, as the success of the metagenomic approach is totally dependent on the purity and clonability of the extracted DNA, the presence of high molecular weight inhibitors such as humic and fulvic acids in soil samples makes this environment a considerably more difficult substrate for preparing metagenomic libraries than aquatic samples. A number of protocols for isolating DNA from soil have been evaluated [34, 40, 46, 47] and they basically fall into two categories.

In the first approach, named the direct method, microbes are lysed in the soil matrix. Soil is suspended directly in lysis buffer and successively treated with detergent and enzymes. Most protocols in this category involve vigorous mechanical agitation and the use of glass beads to break up microbial cells (bead beating). The resulting DNA is subsequently extracted using phenol/chloroform and size-fractionated on agarose gel. The advantages of this approach are high DNA yields and a relatively good representation of the different types of microorganisms present in the sample, due to the low chance of failure to lyse strong cell walls. However, the size range of resulting DNA fragments is from 1 kb up to 50 kb, due to mechanical shearing.

In the second approach, named the indirect method, microorganisms are physically separated from the soil particles and colloids before lysis. This method limits the inhibiting activity of organic and inorganic soil components over DNA extraction and purification procedures, and can be directed to significantly reduce mechanical stress on released genomic DNA. This has been the approach of choice of many studies exploring the soil metagenome for novel products with biological activity. Bacterial cells are separated from the soil using a blender, and subsequently recovered by centrifugation by using a gradient-making matrix. The soil biomass is then blocked into agarose and lysis performed on the immobilized cells. After that, the genomic DNA is partially digested using restriction endonucleases and size-fractionated by pulsed-field gel electrophoresis (PFGE). High molecular weight DNA fragments are then recovered from agarose blocks by electroelution, and subsequently ligated into BAC vectors in order to produce the metagenomic libraries. An advantage of the indirect method is a reduction of contaminating non-microbial and free soil components in DNA preparation. However, disadvantages include loss of microorganisms during preparation of the bacterial fraction from the soil matrix, and

consequently low DNA recoveries. In addition, microbial species having robust cell walls, for example Gram-positive bacteria, may escape lysis and be underrepresented in the metagenomic DNA due to the low mechanical stress employed in this approach.

When using the BAC strategy for cloning bacterial genomes, it is common to obtain gene expression of the BAC clones, since the foreign DNA is also from prokaryotic origin. The expression of heterologous DNA from *Bacillus cereus* in an *E. coli*-harboring BAC system was detected in a reasonable frequency, showing that the low copy number BAC vector could be used to express foreign DNA in *E. coli* [38].

As pointed out above, BAC libraries may offer a powerful tool for accessing the microbial diversity in the environment within a wider perspective, allowing the analysis of functional genes of different members of the microbiota, specially the uncultivated ones. In that sense, a broader range of compounds with biological activity may be obtained simultaneously from a metagenomic BAC library, when compared with the traditional strategy of search for novel natural products, based upon isolation, culturing, and screening of pure microbial cultures (Figure 8.3) [48].

8.3.3

Analysis of Metagenomic Libraries: Activity versus Sequence-driven Strategy, Enrichment for Specific Genomes and Application of High-throughput Screening Methods

The analysis of BAC libraries for enzymatic activities, followed by subsequent screening of genes of interest, may be conducted using two distinct strategies. One of the strategies is based upon insert sequence similarity, whereas the other is based upon expression of the enzyme activity, usually in an established host system such as *E. coli* [48]. The latter, named functional-driven analysis, offers the potential of accessing totally unknown genomic sequences and detecting only enzymes that are active. However, this strategy may be compromised by problems related to the “heterologous expression” of the genes, that is, expression is driven by transcriptional and translational enzymes from the host cell (*E. coli*) in a context where the genes and regulatory sequences are originated from another microorganism. Heterologous expression may yield low levels or absence of expression of the target molecule, or its transport through the cell may be affected, compromising the detection of the synthesized molecule in the assays.

Another limitation of this approach is that it requires all of the genes involved with the enzyme function to be in a cluster and not disrupted in the cloned insert. It also depends on the availability of an assay for the function of interest capable of being carried out efficiently on large clone libraries, since the frequency of active clones may be low for the activity sought after. Improved systems for heterologous gene expression are being developed using “shuttle vectors”, which are vectors with cloned origins of replication from different hosts,

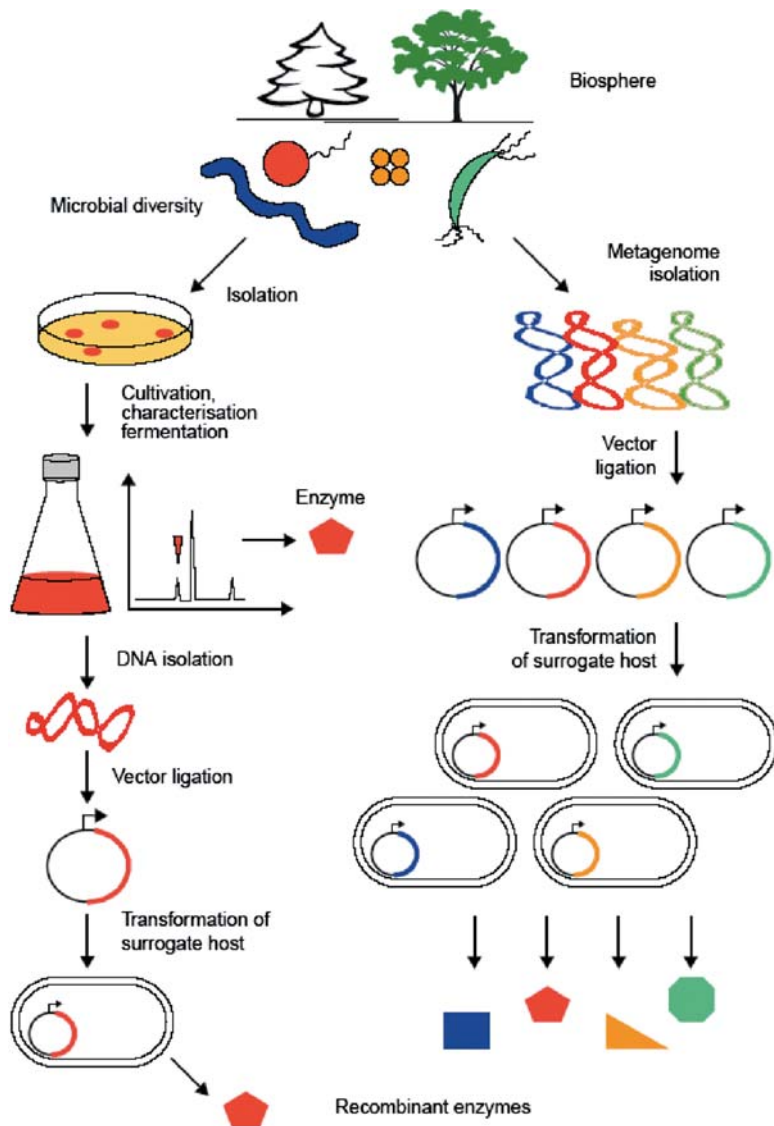


Fig. 8.3 Scheme comparing culturing (left) and metagenome (right) strategies in the search for novel enzymes. In culture-dependent strategies, only a fraction of the microbial diversity can be isolated and grown on pure culture plates. After characterization and fermentation of the cultured microorganisms, products can be obtained directly or following isolation, cloning and expression of the genes involved. The uncultured

microorganisms can only be accessed by means of their genomes. Thus, using the metagenome strategy, the collective genomes of all microorganisms in a given habitat may be isolated and cloned in a host system for subsequent screening and possible expression of a wide range of potentially novel products. (After [48], reprinted with permission from Elsevier. *Current Opinion in Biotechnology*, 2002, 13, pp. 572–577).

enabling replication on different host species systems. This facilitates screening of the metagenomic DNA in diverse hosts by enabling rapid transfer of cloned inserts to multiple expression hosts [49].

In addition to addressing the challenge of detecting rare active clones in large clone libraries, efforts are being directed towards the design of highly sensitive assays and robotic screening systems for the rapid and efficient detection of low levels of activity in a large number of samples.

The method of screening of BAC libraries based on sequence similarity, named sequence-driven analysis, is more conservative, since the oligonucleotide sequences used to identify the target genes, either by PCR or via hybridization methods (probes), must be designed on the basis of conserved amino acid sequence motifs in a way that guarantees a high matching probability with the unknown sequences retrieved from the environment. This is an attractive strategy because any genes presenting moderate to high homology with the DNA sequence used to screen the library (either probe or PCR primers) can be detected independently on their expression in the host cell.

Actually, the ideal strategy in broad screening studies of BAC libraries searching for different biological activities is to combine both approaches based on activity and sequence similarity. They should complement each other allowing one to fully explore the natural microbial diversity available in a given habitat.

One of the frustrations associated with the analysis of metagenomic libraries is the low frequency of clones of a desired characteristic. To increase the proportion of specific clones in a library, various strategies have been designed to enrich for the sequences of interest prior to cloning. A simple enrichment strategy is to select for the GC content of the genomes. Many bacteria that have high GC content in their DNA, such as the actinomycetes, are of great interest for biotechnological applications. In order to find such bacteria, DNA is extracted from the soil and subjected to ultracentrifugation in order to separate the portion of the community with high GC content DNA. This is not a very refined strategy, but may increase the representation of certain genomes in the library. A more elaborate strategy is based on the use of bromodeoxyuridine (BrdU) and enriches for the metabolically active organisms in a given habitat. In this strategy, BrdU is amended to soil samples together with selective substrates that discriminate among the members of the microbial community, enriching metagenomic libraries for those bacteria that grow on the added nutrient. Metabolically active bacteria will incorporate the labeled nucleotide into their DNA, which is then recovered by immunocapture [50, 51]. This strategy can be used to enrich for organisms that grow on substrates such as starch, cellulose, and proteins, in order to find amylases, cellulases, and proteases, respectively, or other enzymes of interest.

The feasibility of metagenomic libraries as a method for screening new compounds of interest is only possible when an integrated approach is used, which combines high-throughput screening techniques, such as miniaturized enzymatic assays using microplates and plate readers, and DNA microarray hybridization technology, with broad-scale sequencing and bioinformatics. The latter

involves the development and use of databases and softwares for DNA and protein sequence analyses to accelerate the functional and comparative analysis of the cloned genes or operons. The technology of broad-scale automated DNA sequencing, as well as the use of databases and software tools for organization, analysis, and comparison of sequences, was developed in response to the growing demand from whole-organism genome-sequencing projects. These developments enabled the assembly of genomes from several prokaryotic and eukaryotic organisms, accumulating an enormous amount of sequence and functional genomic data that are beginning to be explored to unravel novel functions from microbial metagenomes.

The DNA microarray hybridization technology, initially developed for application in medical research and clinical diagnosis, makes it easier to study thousands of genes simultaneously, and has been used for the identification of specific genes, comparison of different genomes and monitoring of gene expression [52, 53].

DNA microarrays consist of thousands of unique DNA sequences, the “probes,” each attached at a known location to a small, solid support, such as glass or nylon. The targets are fluorescent or radioactive-labeled total mRNA, DNA derived from mRNA by reverse transcriptase PCR (RT-PCR), or genomic DNA extracted from microorganisms from a given sample. The targets are hybridized to complementary probes of the array and are identified and quantified by the intensity of the resulting signal. Using sophisticated, commercially available instruments, up to 300,000 distinct oligonucleotides can be spotted onto a 1.28 cm² glass array [54].

In the context of metagenomic library screening, the hybridization targets are the DNA of the BAC clones, which are spotted onto the high-density support. The probes, oligonucleotides directed to a gene coding for a specific biological activity, are labeled and used to hybridize the arrays containing the BAC clones in the search for homologous sequences. This high-throughput screening strategy has not yet been fully explored in the analysis of complex metagenomic libraries. A study conducted using a microarray constructed from a cosmid library of a groundwater microcosm indicated that this approach can be used to access bacterial diversity [55].

8.3.4

Follow-up of the Metagenome Harvest

As mentioned before, the screening of metagenomic libraries opened up perspectives for the discovery of new natural products. Using the BAC strategy it became possible to clone complete biosynthetic pathways, and capture, express, and detect new enzymes from libraries constructed from environmental DNA. Recent studies involving the screening for novel biocatalysts have demonstrated the value of metagenome cloning in that sense. Yet, the production of such compounds in a genetically defined system, such as *E. coli*, makes subsequent genetic manipulation and selection of cloned genes an easier task.

Rondon et al. [34] showed that BAC clones, obtained using metagenomic DNA isolated from agricultural soil, expressed several biological functions in *E. coli*, including the production of amylases, lipases, and antibacterial agents. They also demonstrated that the sequences coding for those compounds were new, corroborating the hypothesis that BAC libraries of environmental DNA constitute a valuable source of new genes. Using the same BAC library, Gillespie and co-workers [56] achieved the expression and isolation of pigmented compounds having broad-spectrum antibiotic activity, named turbomycin A and turbomycin B. They demonstrated that these compounds were produced from the interaction of indole, normally secreted by the *E. coli* host, with homogentisic acid synthesized by the cloned gene product of environmental origin, an enzyme sharing extensive sequence similarity with members of the 4-hydroxyphenylpyruvate dioxygenase (4HPPD) family of enzymes. These results demonstrated successful heterologous expression of DNA extracted directly from soil as a means to access previously uncharacterized small organic compounds, serving as an example of a chimeric pathway for obtaining novel chemical compounds.

Other screening studies of metagenomic libraries allowed the detection of chitinase [57], 4-hydroxy butyrate dehydrogenase [44], lipase [58], protease [59], polyketide synthase [60, 61], and nitrilases [62]. A broad study recently conducted by Voget et al. [63], combining a culture-dependent strategy with metagenomic library screening and sequencing, resulted in the identification of several enzyme-coding genes of biotechnological value, including agarases, cellulases, one α -amylase, pectate lyases, lipases, and a putative stereoselective amidase.

Most studies have exploited metagenomic DNA from environmental samples, including soil, sediments, fresh water, and oceans. However, recent studies have shown that metagenomic libraries from uncultivated bacterial symbionts of beetles and sponges may also harbor biotechnologically relevant products, such as novel type I polyketide synthase (PKS) genes related to a putative pederin PKS system from an uncultured bacterial symbiont of *Paederus fuscipes* beetles [64]. Metagenomic analysis of small-insert gene banks may also be a source of novel enzymes. Clone libraries constructed using DNA extracted from soil and sediments, and from enrichment cultures inoculated with these, enabled the recovery of *E. coli* clones with distinct amidase activity profiles [65].

The great majority of the above-mentioned studies have employed activity-based assays to screen for new enzymes, mainly because this strategy may lead to completely new gene sequences, which may be not detected when using the sequence similarity-driven strategy. So far, most functional screens performed have detected activity by using simple plate or liquid assays. However, expression-based screens may also be conducted by means of gene capture experiments complementing growth-deficient mutants on selective media with metagenomic DNA [66]. Though microarray results are still at an early stage, this technique may well become another high-throughput tool broadly applicable to the analysis of metagenomic libraries [55].

8.4 Concluding Remarks

The BAC strategy may be broadened by introducing BAC plasmids into other bacterial hosts, such as *Bacillus*, *Pseudomonas*, or *Streptomyces*, in an attempt to overcome some of the limitations imposed by heterologous gene expression. This kind of approach should use vectors that allow DNA exchange between *E. coli* and the alternative host. Sosio et al. [67] developed BAC vectors harboring actinomycete gene clusters that could replicate autonomously in *E. coli* hosts or stably integrate into *Streptomyces* genome by site-specific recombination. The *E. coli*–*Streptomyces* artificial chromosomes (ESAC) will permit the analysis of previously genetically non-characterized actinomycetes. Optimization of the expression of integrated genes may yield novel production hosts for antibiotics, enzymes or other important metabolites of actinomycete origin.

The use of the BAC strategy allows for the screening of functional genes with the analysis of 16S ribosomal RNA gene sequences in the metagenomic libraries. This may represent the first steps towards linking phylogeny and function, and may help to elucidate the ecological roles of individual components of microbial communities.

Recent advances have clearly demonstrated that metagenomic cloning and screening is a promising technology for enzyme research. High-throughput screening methods, such as those described elsewhere in this book, make it possible to test thousands of clones rapidly. The sensitivity of modern biological activity assays, especially the ones conducted on a nanoscale, allows the identification of clones that produce or export small amounts of an active molecule.

In addition, the use of *E. coli* as the host cell enhances the power of the metagenome strategy, since this bacterium is commonly used in industrial fermentation and optimized production processes can be rapidly developed. This means that many of the development stages for commercial production of potentially useful compounds are already established well before the genes are cloned, offering an advantage over compounds derived directly from “wild-type” organisms, and the difficulties associated to the development of production and downstream purification processes.

All the innovative approaches discussed herein, applied to the discovery of new enzymes from uncultured microorganisms from environmental samples, mainly soils, are beginning to be used to explore other habitats, for example the microbiota of insects [68], marine organisms [64, 69], and other prokaryotic groups, such as Archaea [70]. Also, extensive large-scale metagenomic projects have been focusing upon the oceans, where a wealth of previously unknown gene sequences have been retrieved and may be a promising source of new compounds and enzymes [43].

Acknowledgments

The authors are grateful to Dr. Fabiano L. Thompson for his revision of the final draft and to the American Society of Microbiology and Elsevier for permission to use published material.

References

- 1 J. Berdy, *Adv. Appl. Microbiol.* **1974**, *18*, 309–406.
- 2 R. L. Hamill, Screens for pharmacologically active fermentation products, in *Bioactive Microbial Products: Search and Discovery*, eds J. D. Bu'Lock, L. J. Nisbet, D. J. Winstanely, Academic Press, New York, **1982**.
- 3 S. Omura, *J. Ind. Microbiol.* **1992**, *10*, 135–156.
- 4 M. Renz, C. Kurz, *Nucleic Acids Res.* **1984**, *12*, 3435–3444.
- 5 M. S. Urdea, B. D. Warner, J. A. Running, M. Stempien, J. Clyne, T. Horn, *Nucleic Acids Res.* **1988**, *16*, 4937–4956.
- 6 R. Amann, L. Krumholz, D. A. Stahl, *J. Bacteriol.* **1990**, *172*, 762–770.
- 7 D. A. Stahl, R. Amann, Development and applications of nucleic acid probes, in *Nucleic Acid Techniques in Bacterial Systematics*, eds E. Stackebrandt, M. Goodfellow, John Wiley & Sons, Chichester, **1991**.
- 8 R. N. Bryan, J. L. Ruth, R. D. Smith, J. M. Le Bon, Diagnosis of clinical samples with synthetic oligonucleotide hybridization probes, in *Microbiology*, ed. L. Leive, American Society for Microbiology, Washington DC, **1986**.
- 9 R. B. Wallace, J. Shaffer, R. F. Murphy, J. Bonner, T. Hirose, K. Itakura, *Nucleic Acids Res.* **1979**, *6*, 3543–3557.
- 10 A. Kaufhold, A. Podbielski, T. Horaud, P. Ferrieri, *Antimicrob. Agents Chemother.* **1992**, *36*, 1215–1218.
- 11 F. J. Teran, J. E. Suarez, M. C. Mendoza, *Antimicrob. Agents Chemother.* **1991**, *35*, 714–719.
- 12 D. Shiffman, M. Mevarech, S. E. Jensen, G. Cohen, Y. Aharonowitz, *Mol. Gen. Genet.* **1998**, *214*, 562–569.
- 13 C. L. Weingart, A. M. Hooke, *Curr. Microbiol.* **1999**, *38*, 233–238.
- 14 L. B. Rice, G. M. Eliopoulos, C. Wennersten, D. Goldmann, G. A. Jacoby, R. C. Moellering Jr., *Antimicrob. Agents Chemother.* **1991**, *35*, 272–276.
- 15 W. Hönerlage, D. Hahn, J. Zeyer, *Arch. Microbiol.* **1995**, *163*, 235–241.
- 16 A. T. Bull, A. C. Ward, M. Goodfellow, *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 573–606.
- 17 H. Zaehner, H. P. Fiedler, The need for new antibiotics: possible ways forward, in *Fifty Years of Antimicrobials: Past Perspectives and Future Trends*, ed. J. J. Russell, Cambridge University Press, Cambridge, **1995**.
- 18 R. L. Ferguson, E. N. Buckley, A. V. Palumbo, *Appl. Environ. Microb.* **1984**, *47*, 49–55.
- 19 R. I. Amann, W. Ludwig, K. Schleifer, *Microbiol. Rev.* **1995**, *59*, 143–169.
- 20 P. Hugenholtz, B. M. Goebel, N. R. Pace, *J. Bacteriol.* **1998**, *180*, 4765–4774.
- 21 D. B. Kell, A. S. Kaprelyants, D. H. Weichart, C. R. Harwood, M. R. Barer, *Antonie van Leeuwenhoek* **1998**, *73*, 169–187.
- 22 S. H. Zinder, *Environ. Microbiol.* **2002**, *4*, 14–15.
- 23 J. A. Papin, N. D. Price, S. J. Wiback, D. A. Fell, B. O. Palsson, *Trends Biochem. Sci.* **2003**, *28*, 250–258.
- 24 G. W. Tyson, J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, J. F. Banfield, *Nature* **2004**, *428*, 37–43.
- 25 J. Handelsman, M. R. Rondon, S. F. Brady, J. Clardy, R. M. Goodman, *Chem. Biol.* **1998**, *5*, R245–249.
- 26 N. R. Pace, D. A. Stahl, D. J. Lane, G. J. Olsen, *Adv. Microb. Ecol.* **1986**, *9*, 1–55.
- 27 S. J. Giovannoni, T. B. Britschgi, C. L. Moyer, K. G. Field, *Nature* **1990**, *345*, 60–63.

- 28 D. M. Ward, R. Weller, M. M. Bateson, *Nature* **1990**, *345*, 63–65.
- 29 N. R. Pace, *ASM News* **1996**, *62*, 463–470.
- 30 P. Hugenholtz, C. Pitulle, K. L. Hershberger, N. R. Pace, *J. Bacteriol.* **1998**, *180*, 366–376.
- 31 E. F. DeLong, *Curr. Opin. Genet. Dev.* **1998**, *8*, 649–654.
- 32 M. S. Rappé, S. A. Connon, K. L. Vergin, S. J. Giovannoni, *Nature* **2002**, *418*, 630–633.
- 33 M. R. Rondon, R. M. Goodman, J. Handelsman, *Trends Biotechnol.* **1999**, *17*, 403–409.
- 34 M. R. Rondon, P. R. August, A. D. Bettermann, S. F. Brady, I. A. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, R. M. Goodman, *Appl. Environ. Microb.* **2000**, *66*, 2541–2547.
- 35 M. S. Osburne, T. H. Grossman, P. R. August, I. A. MacNeil, *ASM News* **2000**, *66*, 411–417.
- 36 U.-J. Kim, B. W. Birren, T. Slepak, V. Mancino, C. Boysen, H.-L. Kang, M. I. Simon, H. Shizuya, *Genomics* **1996**, *34*, 213–218.
- 37 H. Shizuya, B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiri, M. Simon, *Proc. Natl Acad. Sci. USA* **1992**, *89*, 8794–8797.
- 38 M. R. Rondon, S. J. Raffel, R. M. Goodman, J. Handelsman, *Proc. Natl Acad. Sci. USA* **1999**, *96*, 6451–6455.
- 39 J. Wild, Z. Hradecna, W. Szybalski, *Genome Res.* **2002**, *12*, 1434–1444.
- 40 J. L. Stein, T. L. Marsh, K. Y. Wu, H. Shizuya, E. F. DeLong, *J. Bacteriol.* **1996**, *178*, 591–599.
- 41 O. Beja, M. T. Suzuki, E. V. Koonin, L. Aravind, A. Hadd, L. P. Nguyen, R. Villacorta, M. Amjadi, C. Garrigues, S. B. Jovanovich, R. A. Feldman, E. F. DeLong, *Environ. Microbiol.* **2000**, *2*, 516–529.
- 42 M. Y. Galperin, *Environ. Microbiol.* **2004**, *6*, 543–545.
- 43 J. C. Venter, K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y.-H. Rogers, H. O. Smith, *Science* **2004**, *304*, 66–74.
- 44 A. Henne, R. Daniel, R. A. Schmitz, G. Gottschalk, *Appl. Environ. Microb.* **1999**, *65*, 3901–3907.
- 45 I. A. MacNeil, C. L. Tiong, C. Minor, P. R. August, T. H. Grossman, K. A. Loiacono, B. A. Lynch, T. Phillips, S. Narula, R. Sundaramoorthi, A. Tyler, T. Aldredge, H. Long, M. Gilman, D. Holt, M. S. Osburne, *J. Mol. Microbiol. Biotechnol.* **2001**, *3*, 301–308.
- 46 J. Zhou, M. A. Bruns, J. M. Tiedje, *Appl. Environ. Microbiol.* **1996**, *62*, 316–322.
- 47 A. E. Berry, C. Chiocchini, T. Selby, M. Sosio, E. M. Wellington, *FEMS Microbiol. Lett.* **2003**, *6*, 15–20.
- 48 P. Lorenz, K. Liebeton, F. Niehaus, J. Eck, *Curr. Opin. Biotechnol.* **2002**, *13*, 572–577.
- 49 A. Martinez, S. J. Kolvek, C. L. T. Yip, J. Hopke, K. A. Brown, I. A. MacNeil, M. S. Osburne, *Appl. Environ. Microb.* **2004**, *70*, 2452–2463.
- 50 E. Urbach, K. L. Vergin, S. J. Giovannoni, *Appl. Environ. Microb.* **1999**, *65*, 1207–1213.
- 51 J. Borneman, *Appl. Environ. Microb.* **1999**, *65*, 3398–3400.
- 52 O. S. Lee, K. H. Lee, *Curr. Opin. Biotechnol.* **2000**, *11*, 171–175.
- 53 N. L. van Hal, O. Vorst, A. M. van Houwelingen, E. J. Kok, A. Peijnenburg, A. Aharoni, A. J. van Tunen, J. Keijer, *J. Biotechnol.* **2000**, *78*, 271–280.
- 54 R. J. Lipshutz, S. P. Fodor, T. R. Gingeras, D. J. Lockhart, *Nat. Genet.* **1999**, *21*, 20–24.
- 55 J. L. Sebat, F. S. Colwell, R. L. Crawford, *Appl. Environ. Microb.* **2003**, *69*, 4927–4934.
- 56 D. E. Gillespie, S. F. Brady, A. D. Bettermann, N. P. Cianciotto, M. R. Liles, M. R. Rondon, J. Clardy, R. M. Goodman, J. Handelsman, *Appl. Environ. Microb.* **2002**, *68*, 4301–4306.
- 57 M. T. Cottrell, J. A. Moore, D. L. Kirchman, *Appl. Environ. Microb.* **1999**, *65*, 2553–2557.
- 58 A. Henne, R. A. Schmitz, M. Bomeke, G. Gottschalk, R. Daniel, *Appl. Environ. Microb.* **2000**, *66*, 3113–3116.

- 59 R. Gupta, Q.K. Beg, P. Lorenz, *Appl. Environ. Microb.* **2002**, *59*, 15–32.
- 60 K. T. Seow, G. Meurer, M. Gerlitz, E. Wendt-Pienkowski, C. R. Hutchinson, J. Davies, *J. Bacteriol.* **1997**, *179*, 7360–7368.
- 61 A. Ginolhac, C. Jarrin, B. Gillet, P. Robe, P. Pujic, K. Tüphile, H. Bertrand, T. M. Vogel, G. Perrière, P. Simonet, R. Nalin, *Appl. Environ. Microb.* **2004**, *70*, 5522–5527.
- 62 M. J. Burk, Discovery of novel enzymes from natural diversity: development of nitrilases for production of pharmaceutical intermediates, in *The Pharmaceutical Fine Chemicals and Biomolecule Manufacturing Report*. PharmaVentures Inc., Oxford, **2002**.
- 63 S. Voget, C. Leggewie, A. Uesbeck, C. Raasch, K. E. Jaeger, W. R. Streit, *Appl. Environ. Microb.* **2003**, *69*, 6235–6242.
- 64 J. Piel, D. Hui, N. Fusetani, S. Matsunaga, *Environ. Microbiol.* **2004**, *6*, 921–927.
- 65 E. M. Gabor, E. J. de Vries, D. B. Janssen, *Environ. Microbiol.* **2004**, *6*, 948–958.
- 66 A. Majernik, G. Gottschalk, R. Daniel, *J. Bacteriol.* **2001**, *183*, 6645–6653.
- 67 M. Sosio, F. Giusino, C. Cappellano, E. Bossi, A. M. Puglia, S. Donadio, *Nat. Biotechnol.* **2000**, *18*, 343–345.
- 68 M. Ohkuma, T. Kudo, *Appl. Environ. Microb.* **1996**, *62*, 461–468.
- 69 W. Fenical, *Trends Biotechnol.* **1997**, *15*, 339–341.
- 70 A. H. Treusch, A. Kletzin, G. Raddatz, T. Ochsenreiter, A. Quaiser, G. Meurer, S. C. Schuster, C. Schleper, *Environ. Microbiol.* **2004**, *6*, 970–980.

Part III
Enzyme Fingerprinting

9

Fluorescent Probes for Lipolytic Enzymes

Ruth Birner-Grünberger, Hannes Schmidinger^{}, Alice Loidl, Hubert Scholze, and Albin Hermetter*

9.1

Introduction

Serine hydrolases comprise lipases, other carboxyl esterases (EC 3.1.1.x) and thioesterases (EC 3.1.2.x), as well as many proteases (EC 3.4.x.x). Lipases are involved in degradation and thus mobilization of lipids inside and outside cells [1]. Unlike other hydrolases that are active in the aqueous phase, lipases are activated at the hydrophobic/hydrophilic interface of oil/water or detergent/water systems, where lipolytic substrates usually form equilibria between monomeric, micellar, and higher aggregated states [2]. Lipases show various specificities for fatty acyl chain length, positional and optical isomers [3]. They may also possess activity towards short-chain carboxylic esters and thus accept a wide range of substrates. The hydrolytic reaction is reversible, and if the amount of water is limiting (e.g. in the presence of organic solvents), the enzymes effectively catalyze inter- and transesterification reactions. Thus their enantio-, chemo-, and stereoselectivity make them important tools in organic synthesis.

Development of new or improved enzymes is essential for making progress in biocatalysis. Determination of enzymatic activities is required during isolation, purification, and characterization of lipases, as well as for characterization of natural lipase sources originating from animals, plants, and microorganisms, or screening of recombinant proteins. Since lipases play an important role in human (patho-) physiology (obesity and cardiovascular disease), determination of lipase activities is also important in medical diagnosis and research [4].

The 3D structures of crystallized lipases show a common alpha/beta hydrolase fold as well as a so-called “nucleophilic elbow” to which the catalytic serine is bound [5, 6]. The amino acid sequence around the nucleophilic serine contains the typical G-X-S-X-G motif. Most lipases contain a lid controlling the access of substrates to the hydrophobic active site. The same structural features are found

^{*} Co-first author.

in esterases, except for the lid. Therefore, most esterases do not show interfacial activation. However, it has to be emphasized that the borderline between lipases and esterases is not well defined and some exceptions exist. For the same reasons, analytical discrimination and classification of lipolytic enzymes is difficult.

Fluorescent substrate analogs with different chemical structure are useful tools for the characterization of individual enzymes. The fluorescence emitted from these probes can either be measured spectroscopically in cuvettes or microtiter plates, or visualized on surfaces such as thin-layer chromatography (TLC) plates or polyacrylamide gels using an imaging device (CCD camera or laser scanner). Important factors for substrate and inhibitor design are the detailed chemical structure, overall polarity and stereochemistry of the probe. In addition, the “quality” of the hydrophobic–hydrophilic interface critically influences the apparent reactivities. Typical systems for substrate inhibitor solubilization include synthetic detergents, phospholipids, and/or proteins. Albumin complexes with fluorescent triacylglycerols can be lyophilized, and in solid form they can be used as instant substrates. It is important to note that the optimal condition for substrate solubilization may vary between the individual enzymes. In some cases effects on activity and stereopreference of lipases have been observed [7].

9.2

Fluorogenic and Fluorescent Substrates for Enzyme Activity

Methods for lipase detection and screening have recently been reviewed [3, 8]. Titrimetric, radioactive, and spectrophotometric methods are most commonly used in practice because the required instrumentation is part of the standard equipment of biochemical laboratories. However, most of these techniques suffer from several shortcomings such as poor reproducibility, lack of sensitivity, production of radioactive waste, or time-consuming procedures. Titrimetric methods are not very sensitive. In addition, problems might arise if the reaction medium contains detergent or organic solvent. Nevertheless, they generally serve as a reference lipase assay because they are simple, accurate, and fairly reproducible [9]. Radiolabeled substrates are highly specific but require separation of the released free fatty acids by solvent extraction [10]. Colorimetric methods involving the use of copper salts [11] or rhodamine [12] for detection of the released fatty acids also depend on separation of substrate and reaction products. Although the latter methods are rather time-consuming, they have the advantage to be more sensitive than titrimetry. In addition, they are based on (quasi) natural substrates.

p-Nitrophenyl esters of various fatty acids differing in chain length are useful as chromogenic substrates since the hydrolytic release of *p*-nitrophenol can be measured continuously due to its absorption at 410 nm [13, 14]. The 1,2-di-*O*-butylglycerol-3-*p*-nitrophenylcarbonate has been proposed as a synthetic substrate for the determination of pancreatic lipases [15]. Major limitations of this assay, however, are its low specificity and spontaneous hydrolysis of the substrate [16]. Similar photometric methods utilize carboxylic acid esters of 2,4-dini-

trophenol [17] or α -naphthol [18]. Commercially available 1,2-*O*-dilauryl-*rac*-glycero-3-glutaric acid resorufin ester is widely used for the spectrophotometric determination of lipase activity in serum. The method is based on the absorption of released resorufin. It is very sensitive but rather unspecific because its chemical structure in the *sn*-3 position is very different from that of natural triacylglycerols [3]. Another continuous but indirect spectrophotometric assay is based on the metachromatic properties of the cationic dye safranin. Upon lipid hydrolysis the net negative charge at the lipid/water interface increases as a consequence of fatty acid formation, leading to a change in safranin absorption [19].

The advantage of fluorimetric assays is their high sensitivity and the ability to continuously monitor the reaction kinetics. Indirect methods are based on the increase of rhodamine B fluorescence if this dye binds free fatty acids released from triacylglycerols. This assay is also applicable to lipases embedded in an agarose gel [20]. Another technique measures binding of a fluorescent fatty acid (e.g. 11-(dansylamino)undecanoic acid or α -parinaric acid) released from a labeled ester to a fatty acid-binding protein [21, 22]. These methods are not very reliable if the solvent contains detergents, high salt concentrations, albumin, or membrane fractions.

A bioluminescent assay employs a fatty acid-dependent luminescent bacterial mutant that requires free myristic acid as a substrate for light production. This method was originally developed for measuring lipase activity using trimyristin as a substrate for lipases or 1- α -dimyristoyl phosphatidylcholine as a substrate for phospholipase A and C activity [23, 24]. For phospholipase C, the detection is based on a cascade reaction in which the mixture contains an excess of lipase which rapidly hydrolyzes the diacylglycerol generated from phosphatidylcholine. Its application for detection of free myristic acid released from cholesteryl myristate by bovine pancreas cholesterol esterase is also possible [25]. Concentrations down to 5 nM of myristic acid could be detected. Although the assay has the advantage of continuous measurement, it is limited by its dependence on myristic acid and problems might arise with complex biological samples.

Defined fluorogenic lipase substrates may contain the label in the acyl or the alcohol moiety. The latter situation is given in acyl esters of 4-methylumbelliferone [26–29]. Hydrolysis gives rise to the formation of the free alcohol which in contrast to the esterified form is highly fluorescent. The respective activity assays are highly sensitive but not specific. In addition, the substrates are prone to spontaneous hydrolysis. An analogous substrate is pyrenemethyl laurate. Upon hydrolysis a shift in fluorescence wavelength is observed from 475 nm (substrate excimer emission) to 375 nm (monomer emission) [30].

Phospholipase activities can be measured using phospholipid substrates containing fluorescent fatty acids or head groups. The labels include dansyl (5-dimethylaminonaphthalene-1-sulfonyl)- [31], parinaroyl- [32], NBD (7-nitro-benz-2-oxa-1,3-diazole)- [33] pyrene, or other residues [34, 35]. These methods are explained in more detail in Section 9.2.4.

Triacylglycerols containing a fatty acyl chain labeled with pyrene at the ω position 1 (Figure 9.1) have been proposed as lipase substrates [36]. However this

sensitive assay requires separation of the released pyrene fatty acids. In our laboratory, a double-labeled triglyceride analog **2** and **3** (Figure 9.1) was developed containing a fluorophore (pyrene) and a quencher in the same molecule. This lipid is the basis of a continuous lipase assay for complex biological samples [37]. The method is described in detail in Section 9.2.1.

To overcome the specificity problems associated with synthetic substrates, Beisson and coworkers [38] established a lipase assay using fluorescent triacylglycerols as substrates that contain the aliphatic polyene fatty acid parinaric acid (triglycerides purified from *Parinari glaberrimum* seed oil). Since these lipids contain high concentrations of the fluorophore, their fluorescence is low due to self-quenching. An increase in fluorescence is observed upon release of free parinaric acid followed by dilution in an excess of mixed detergent micelles. However, this method has a major drawback as parinaric acid is highly susceptible to oxidation by atmospheric oxygen. Thus, the assay has to be performed in the presence of antioxidants under an argon or nitrogen atmosphere.

Enzyme-induced degradation of triacylglycerols may involve the sequential formation of diacylglycerol, then monoacylglycerol and finally glycerol. Each step is accompanied by the release of free fatty acid. The ratio of these products depends on substrate specificity of the lipase of interest. For defined determination of triacylglycerol and diacylglycerol lipase activity alkylglycerol derivatives have been introduced [39, 40]. A selective fluorimetric assay for the determination of diacylglycerol hydrolase activity has been developed in our laboratory. It is described in detail in Section 9.2.2.

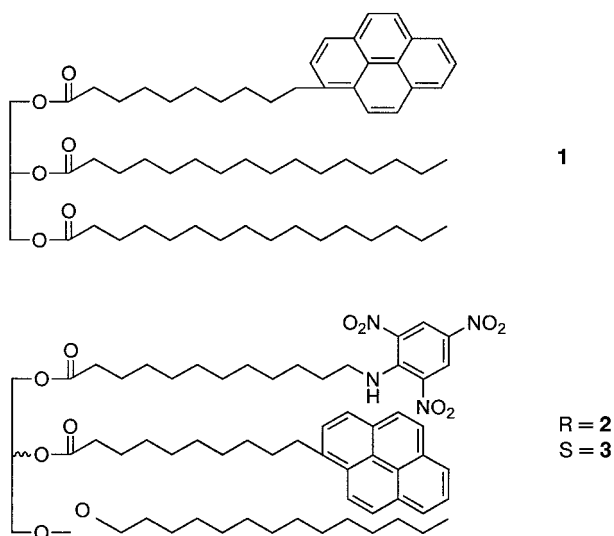


Fig. 9.1 Chemical structures of lipase substrates. **1**: Triacylglycerol with a pyrene decanoic acid label; **2** and **3**: fluorogenic alkyl-diacylglycerol substrates ($R = 2$, $S = 3$). Note that the alkyl chain is not cleavable by lipases.

Fluorescence assays for cholesteryl esterase activity will be discussed in Section 9.2.3. These enzymes may also show triacylglycerol hydrolase activity and thus, defined substrates are mandatory for characterization of enzyme function.

Lipid hydrolases are not only important in lipid metabolism, they are also important enzymes of the signal transduction machinery generating lipid second messengers (fatty acids, lysophospholipids, sphingolipids, etc.). These enzymes are highly regulated and in most cases specific substrates are needed for activity measurements. Fluorescence assays for phospholipase activity are described in Section 9.2.4. Established methods for sphingomyelinase activity measurement include assays for acid sphingomyelinase [41], high-throughput techniques [42, 43] and others [44] which all require radioactive substrates. A novel nonradioactive approach makes use of biotinylated substrate conjugates for purification of enzyme products by bioaffinity capture on streptavidin/agarose beads prior to analysis by electrospray ionization mass spectrometry [45]. Fluorescence assays for sphingomyelinases have been developed by Samet and Barenholz [46] and our laboratory [47], and are explained in detail in Section 9.2.5.

9.2.1

Triacylglycerol Lipase Activity Assay

A continuous fluorescence method for the determination of lipase activities and their stereopreference towards enantiomeric triglycerides was developed in our laboratory [7, 37]. Enantiomeric *O*-alkyl-diacylglycerols **2** and **3** (Figure 9.1) containing pyrene as a fluorescent reporter and a trinitrophenylamino (TNP) residue as a fluorescence quencher were synthesized. The fluorophore and the quencher were covalently bound to the ω -end of the acyl chains in position *sn*-2 and *sn*-3(1), respectively. Since the emission spectrum of pyrene overlaps to a large extent with the absorption spectrum of TNP, fluorescence is efficiently suppressed due to fluorescence resonance energy transfer (FRET) in the intact molecule. From studies on excimer fluorescence of 1,2- and 1,3-dipyreneacylalkylglycerols it is known that these compounds show an extended conformation in organic solvent and detergent micelles. Therefore, their hydrophobic chains are not in close vicinity to each other [48]. Even if we assume such a conformation for the self-quenched lipids, TNP and pyrene must still be located within the critical distance required for FRET between this donor–acceptor pair since effective pyrene quenching is observed in the same media.

Upon lipase- or chemically induced hydrolysis of the fluorogenic substrates, fluorophore and quencher are spatially separated from each other. Thus, the release of the fatty acyl chains leads to fluorescence dequenching, and, under substrate saturation conditions ($\geq 2 \mu\text{M}$), the rate of lipolysis can be measured from the time-dependent increase in fluorescence intensity.

Procedure 9.1: Triacylglycerol lipase Assay

Substrate (6 nmol) dissolved in 25 μL tetrahydrofuran is added to a solution of 0.4 mg fatty acid free bovine serum albumin (BSA) (fraction V) in 3 mL aqueous buffer (e.g. 0.1 M Tris-HCl, pH 7.4), 0.04 μM Triton X-100, or phosphate-buffered saline (PBS), 0.04 μM Triton X-100, under stirring at 37°C. The final substrate concentration is 2 μM . The solubilized substrate is stored overnight at 4°C for equilibration before use. Enzyme samples are added to 3 mL of substrate solution and rates of lipolysis are determined from the continuous increase in fluorescence intensity at 378 nm (excitation 342 nm, slit width 10 nm each) on a spectrofluorometer for several minutes at 37°C. For the quantitative determination of very large numbers of samples (several hundreds) in a short time a fluorescence plate reader has already been used successfully.

Most lipase assays have in common the use of detergents as solubilizers of the natural, radioactive, or chromogenic substrates that are rather hydrophobic. Thus, the respective assays give an information about lipase activities on detergent-substrate mixed micelles, but not under physiological conditions. It is therefore our main concern to maintain “native conditions” when analysing lipases using our substrates. This may become especially important if lipases are to be determined in the presence of lipoproteins (of blood plasma) or in living cells. We found that the fluorogenic alkyldiacylglycerols were readily dispersed in aqueous buffer if fatty acid-free albumin was present. Under such conditions, activities of isolated lipoprotein lipase (LPL) and postheparin plasma could be determined within minutes in a highly reproducible manner. Short- and long-chain homologs as well as optical isomers of the fluorogenic alkyldiacylglycerols were hydrolyzed by pancreatic lipase, hepatic lipase, LPL, and lipases in postheparin plasma at highly different rates, depending on the substrate and the enzyme [37]. Thus a useful set of enantiomeric and/or homologous substrates in combinations may be applied to the selective determination of one lipase in a mixture of lipases (e.g. hepatic and lipoprotein lipase in postheparin plasma for medical diagnosis).

Stereopreference of microbial lipases, namely *Chromobacterium viscosum* (CVL), *Candida rugosa* (CRL), *Pseudomonas* sp. (PSL) and *Rhizopus arrhizus* (RAL) lipase, depended, in general, on how the substrate was solubilized in the reaction medium [7]. All lipases under investigation preferentially hydrolyzed the *sn*-1 acyl ester bond of alkyldiacylglycerols if the lipid analog was dispersed in aqueous buffer (0.1 M Tris-HCl, pH 7.4) containing BSA (0.13 mg mL⁻¹). In mixtures of water and ethanol (1/1, v/v), however, the opposite result was found and the microbial enzymes showed higher activity toward the *sn*-3 acyl ester bond. Different stereopreferences were observed with the different lipases if the substrates were solubilized by amphiphiles (micelles of 10 mM *N*-dodecyl-*N,N*-dimethylammonio-3-propane-sulfonate in 0.1 M Tris-HCl, pH 7.4). However, in the micellar system the activity towards the *sn*-3 enantiomers was always higher

than that towards the protein complex. In contrast to the microbial enzymes, LPL from bovine milk showed a different behavior because it preferred the *sn*-1 enantiomers under all solubilization conditions tested. The stereopreference of LPL for the *sn*-1 acyl isomer of the fluorogenic alkyldiacylglycerols is in agreement with its *sn*-1 specificity for triacylglycerols containing aliphatic side-chains [49]. Thus, the combined results demonstrate that great care should be taken when selecting the solubilization system, since it appears to be crucial for the stereospecificity of many lipolytic enzymes. In general, biological solubilizers such as phospholipids or albumin should be preferred over detergents or, even worse, organic solvents

Since perylene fluorescence can also be quenched by TNP, similar substrates using perylene instead of pyrene as a fluorophore and TNP as a quencher were prepared in our laboratory [50]. Mixtures of these enantiomeric fluorogenic alkyldiacylglycerols, selectively labeled with pyrene or perylene, can be used for a dual-wavelength assay of lipase activity and stereoselectivity. Absorption and emission maxima of both labels are clearly separated. Thus, hydrolysis of the respective enantiomeric substrates can be determined simultaneously, and the difference in the rates of hydrolysis can be taken as a measure for the stereopreference of a lipase. The major drawback of the method, however, is that hydrolysis rates measured with perylene-substituted lipids are generally lower than those obtained with pyrene analogs. Thus, the apparent differences in activity are a result of the selectivity of the enzyme for both types of labels.

9.2.2

Diacylglycerol Lipase Activity Assay

Diacylglycerol lipase activity is usually determined by quantification of the release of [^3H] or [^{14}C] fatty acid from radioactively labeled diacylglycerol or acylalkylglycerol [39, 40, 51–53]. Acylalkylglycerol as a substrate has the advantage that only one defined product (alkylglycerol) is formed, which cannot be degraded by monoacylglycerol lipases.

A fluorimetric assay using a fluorescently labeled alkylacylglycerol derivative, namely 1-NBD-dodecanoyl-2-*O*-hexadecanoyl-*sn*-glycerol **4** (NBD-DAG) as substrate was established in our laboratory (unpublished) (Figure 9.2). The ether bond in *sn*-2 position of the alkylacylglycerol is not cleavable by (lipase-catalyzed) hydrolysis. Therefore, only the fluorescently labeled fatty acid in the *sn*-1 or *sn*-3 position of glycerol is released by the enzyme.

We have used this fluorescent substrate to measure diacylglycerol lipase activity of isolated hormone sensitive lipase (HSL), which was provided by C. Holm, Lund, Sweden (unpublished data) [54]. For solubilization of the substrate, phospholipids instead of detergents were used, since the latter inhibit HSL activity towards triacylglycerols and, to a minor extent, towards diacylglycerols too [55]. After the reaction, the fluorescent fatty acid (NBD-dodecanoic acid), together with the remaining substrate, is extracted quantitatively by organic solvent at pH 2–3 by one extraction step. Traces of acid and salts can be removed in a sin-

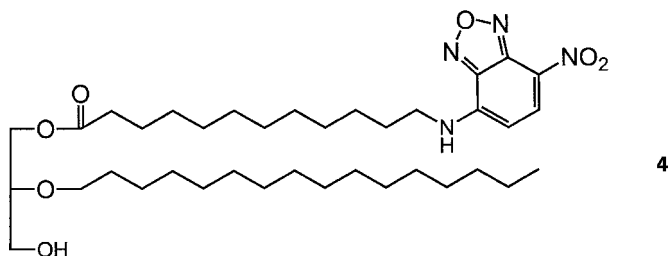


Fig. 9.2 Fluorescent substrate for diacylglycerol lipase activity assays.

gle washing step. The lipids are separated by TLC on silica gel and fluorescence of the individual spots is analyzed by means of a CCD camera.

Procedure 9.2: Diacylglycerol Lipase Assay

A solution (70 μL) of 10 mg mL^{-1} phosphatidylcholine/phosphatidylinositol (PC from fresh egg yolk, PI from soy beans, both from Sigma, St. Louis, MO, USA) 3/1 (w/w) in dichloromethane was mixed with a solution of 40 nmol of NBD-DAG in a glass vial. The organic solvent was removed under a gentle stream of nitrogen and then under vacuum. Next, 1.5 mL of 100 mM potassium phosphate buffer pH 7.0 were added. The lipid was dispersed by vortexing at high speed several times followed by sonification at 50 W under ice-cooling for 4 min using a 4 mm ultrasonification tip. Finally, 0.5 mL of a 20% BSA solution in 100 mM potassium phosphate buffer pH 7.0 were added. Substrate (2 mL) containing 20 μM lipid is obtained, which is ready for use and stable for at least several hours at room temperature. For activity measurements, 100 μL substrate solution were mixed with the HSL sample in buffer (20 mM potassium phosphate, 1 mM EDTA, 1 mM dithioerythreitol and 0.02% defatted BSA, pH 7.0) and filled up to a total volume of 200 μL . The reaction was performed in Eppendorf tubes at 37 $^{\circ}\text{C}$ and the reaction was stopped by addition of 25 μL 1 N HCl and 1 mL chloroform/methanol 2/1 (v/v) after incubation for 30 min. The two-phase system was vigorously vortexed for 15 min. After centrifugation (2 min at 12000 rpm) the water phase was removed quantitatively using a thin glass pipette. The organic phase was washed once with 500 μL water/methanol 4/1 (v/v), and the organic solvent was removed under a gentle stream of nitrogen. The resultant lipid residue was dissolved in 10 μL chloroform/methanol 2/1 (v/v) and quantitatively applied onto a 10 $\text{cm} \times 10 \text{ cm}$ TLC plate (Merck Silica gel 60, Merck, Darmstadt, Germany). The TLC plate was developed with chloroform/methanol 9/1 (v/v) as solvent. The fluorescence intensity of the free acids were quantified using a CCD camera (Herolab FH EASY gel documentation system) at an excitation of 365 nm. The signal was 1 to 16 times integrated according to illumination times of 0.04–0.64 s,

depending on fluorophore concentration. Data acquisition and processing was carried out using the software EasyWin32. Calibration of fluorescence intensity versus fluorophore concentration on TLC plates was performed using defined amounts of NBD-dodecanoic acid.

The detection limit for NBD-dodecanoic acid was 0.2 pmol. The lowest amount of HSL which could be detected was 0.2 ng. The specific HSL activity was $1.25 \pm 0.11 \mu\text{mol min}^{-1} \text{mg}^{-1}$. The K_m was $4 \mu\text{M}$ for NBD-monoacyl-mono-alkylglycerol 4.

9.2.3

Cholesteryl Esterase Activity Assay

The most commonly used cholesteryl esterase assays are based on radioactive cholesteryl [^{14}C]oleate emulsions as substrates [56, 57]. A fluorimetric method has been established by Pozdnev et al. [58], using emulsified cholesteryl-*o*-coumarate as substrate. Its hydrolysis yields *o*-coumaric (*trans*-2-hydroxycinnamic) acid, which is detected fluorimetrically after shifting the pH to 10.4. About $1 \mu\text{g}$ of pancreatic cholesteryl esterase could be detected by the authors if the enzyme was incubated with the substrate for 15 min. However, the fluorescence of *o*-coumaric acid is pH dependent and does not allow continuous measurement at low or neutral pH. Since the structure of the substrate is very different from that of aliphatic cholesteryl acylesters, substrate specificity is a major concern.

A BodipyTM fatty acid cholesterol ester 5 (Figure 9.3) is commercially available (Molecular Probes, Eugene, Oregon, USA). We have used this fluorescent substrate for determination of cholesteryl esterase activity of HSL (unpublished data). The aqueous substrate suspension was prepared the same way as described for NBD-alkylacylglycerol (see Section 9.2.2) except that the lipid film and the buffer were kept at 37°C while vortexing to facilitate lipid dispersion in water. The experimental procedures were similar compared to the assays using NBD-DAG.

The detection limit was 0.1 pmol for Bodipy-dodecanoic acid, and 1 ng for HSL. The specific activity was $8.0 \pm 0.8 \text{ pmol min}^{-1} \mu\text{g}^{-1}$, and K_m was $8 \mu\text{M}$ for the Bodipy fatty acid cholesterol ester. Compared with the activity data obtained with the NBD diacylglycerol analog, the degradation of the cholesterol ester was found to be 150 times slower. In contrast, only sevenfold lower enzyme activities were reported when cholesteryl[^3H]oleate instead of mono-[^3H]oleoyl-2-*O*-

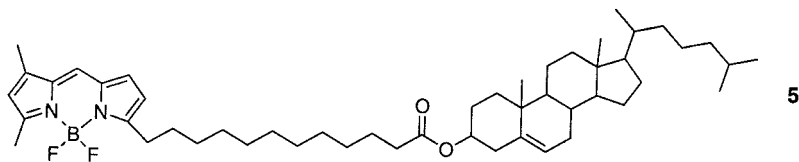


Fig. 9.3 Fluorescent substrate for cholesterol esterase activity assay.

oleoylglycerol was the substrate [40]. It has to be taken into account, however, that the fluorescent lipids differ in structure and labeling (Bodipy versus NBD).

9.2.4

Phospholipase Activity Assay

Phospholipase activities have been measured using phospholipids containing fluorescent fatty acids or labeled head groups. A wide range of substrates is commercially available (Molecular Probes).

1-Acyl-2-parinaroylphosphatidylcholine has been used as a substrate in a discontinuous assay for phospholipase A₂ [32]. As already mentioned above, parinaric acid has the advantage of being a natural, aliphatic fatty acid fluorophore. However, it has to be emphasized that it is very susceptible to oxidation.

For measurement of phospholipase activity of milk LPL the fluorescent, head group-labeled phospholipid *N*-dansyl phosphoethanolamine was incorporated into very low density lipoproteins (VLDL) by mild bath sonification [31]. Lipid hydrolysis resulted in formation of dansyl-lysophosphoethanolamine, which after binding to albumin showed a threefold increase in fluorescence intensity and a 20 nm blue shift in the emission wavelength. In addition to fluorescence intensity, an increase in fluorescence polarization of the dansyl label and energy transfer from albumin tryptophan to the label are observed. Despite the simplicity of this assay, one has to keep in mind that head group labeling changes the structure and polarity of the phospholipids. Thus these substrates are only suitable for prescreening of enzymatic activities and not for determination of substrate preference.

The fluorescent phospholipid 1-acyl-2-NBD-hexanoyl phosphocholine (C6-NBD-PC) was used as a substrate for porcine pancreatic phospholipase A₂ (PLA₂) (phosphatidylcholine 2-acylhydrolase, EC 3.1.1.4) and bovine milk LPL [33]. Hydrolysis of C6-NBD-PC by either enzyme resulted in a greater than 50-fold fluorescence enhancement with no shift in the emission maximum at 540 nm. A prerequisite for the observed change in fluorescence intensity is a sufficiently high NBD-phospholipid concentration leading to intermolecular self-quenching of the substrate. Identification of hydrolysis products showed cleavage at the *sn*-1 and *sn*-2 glycerol positions by LPL and PLA₂, respectively. The same substrate was used for determination of phospholipase activity in potato leaves [59] and in a discontinuous assay for lipolytic activity in fish gut [60]. Similar NBD-labeled analogs of phosphatidylethanolamine, phosphatidylglycerol, and phosphatidic acid were evaluated as substrates for phospholipases by Moreau [61]. NBD-hexanoyl- and NBD-dodecanoyl-glycerophosphocholines are commercially available. If lipids are labeled with fatty acyl chains containing polar fluorophores like NBD it has to be taken into account that the label "loops back" to the head group region in phospholipid bilayers and, as a consequence, perturbs the bilayer packing, which in turn influences enzyme activities [62].

Phospholipid analogs with pyrene-labeled acyl chains are perhaps the best membrane probes for monitoring biochemical and biophysical processes includ-

ing phospholipase-mediated lipid degradation. Continuous assays are based on the different emission maxima of the substrate observed at high label concentration (excimer fluorescence at about 470 nm) and low label concentration (monomeric fluorescence at 380–400 nm in bilayers/micelles). 1-Palmitoyl-2-(ω -pyrenehexanoyl)-*sn*-glycero-3-phosphocholine, -ethanolamine, -glycerol, -serine, and -phosphoric acid monomethylester were used as substrates to determine the influence of lipid head groups on pancreatic PLA₂ activity [34]. When high concentrations of the lipids were solubilized in water, mostly pyrene excimer fluorescence was observed. After release of the *sn*-2-fatty acid, the latter was bound to albumin and showed monomer emission. Except for the synthetic phosphatidylmethanol, which was optimally hydrolyzed in the presence of 3.5 mM CaCl₂, fastest hydrolysis of the other pyrene phospholipids was obtained at 2 mM Ca²⁺. Sodium cholate inhibited hydrolysis of the ethanolamine and serine lipids, whereas a slight (1.4- to 2.0-fold) activation was observed for the -choline, -glycerol, and -monomethylester derivatives. Whereas snake venom PLA₂ preferred phosphatidylcholine as a substrate, the porcine and bovine pancreatic and porcine intestinal PLA₂ preferred acidic phospholipids in the following order: monomethylester \geq phosphatidylglycerol \geq phosphatidylserine.

A continuous fluorescence assay utilizes the intramolecular excimer formation in 1,2-bis(4-(1-pyrenyl)butanoyl)-*sn*-glycero-3-phosphocholine as substrate for PLA₂ activity measurement [35]. The adjacent pyrene fluorophors in the same molecule form excited state dimers emitting at 480 nm. Upon lipid hydrolysis one fatty acid is released and pyrene monomer emission increases. The assays based on bis-pyrene labeled substrates show higher robustness with respect to variation in substrate, detergent, or albumin concentration. Thus they are the preferred system for assaying phospholipases in complex biological samples. However, they do not allow discrimination between phospholipase A₁ and A₂ activity.

A commercially available fluorogenic substrate for PLA₂ (Molecular Probes) contains a Bodipy-FLTM-labeled *sn*-2 acyl chain and a head group labeled with 2,4-dinitrophenol which is a quencher of Bodipy-FL emission. Release of the labeled acyl chain eliminates the intramolecular quenching by the dinitrophenyl residue and leads to an increase of fluorescence intensity.

For phospholipases C (PLC) and D (PLD), enzyme-coupled assays are available (Molecular Probes). Activities are monitored indirectly using Amplex Red reagent, which fluoresces when oxidized by H₂O₂ and horseradish peroxidase (POD). Phospholipase C-mediated hydrolysis of phosphatidylcholine leads to formation of phosphocholine and diacylglycerol. Phosphocholine is subsequently converted to choline and inorganic phosphate by alkaline phosphatase. PLD hydrolyzes phosphatidylcholine yielding choline and phosphatidic acid. Choline is oxidized by choline oxidase to betaine and H₂O₂. Finally, H₂O₂ reacts with the Amplex Red reagent in the presence of horseradish peroxidase to generate fluorescent resorufin. Since phosphocholine is also a product of sphingomyelin hydrolysis, the same reagent can be used for sphingomyelinase activity measurement. However, one has to bear in mind that due to the indirect measurement, PLC and sphingomyelinase activities cannot be distinguished.

9.2.5

Sphingomyelinase Activity Assay

Sphingomyelinases (Smases) under “*in vitro*” conditions catalyze the hydrolysis of sphingomyelin, giving rise to the formation of biologically active ceramide and phosphocholine [63]. Smase assays are performed using labeled lipids including radioactive or fluorescent sphingomyelin. In many cases Triton X-100 is used for lipid solubilization. After the enzymatic reaction, substrates and products have to be separated, e.g. by thin-layer chromatography (TLC), before quantification.

A fluorescent substrate analog, namely lissamine rhodamine dodecanoyl sphingosyl phosphocholine, was used to determine the enzymatic activities of acidic and neutral sphingomyelinases in crude extracts of HL-60 cells [46]. Both enzymes had similar K_m and V_{max} values, but different values with respect to temperature and detergent stability which was higher for acidic Smase in Triton X-100 (inactivation at >5 mM; ≤ 0.4 mM for the neutral Smase) at 37°C . Neutral Smase, but not acidic Smase, displayed different activities towards natural bo-

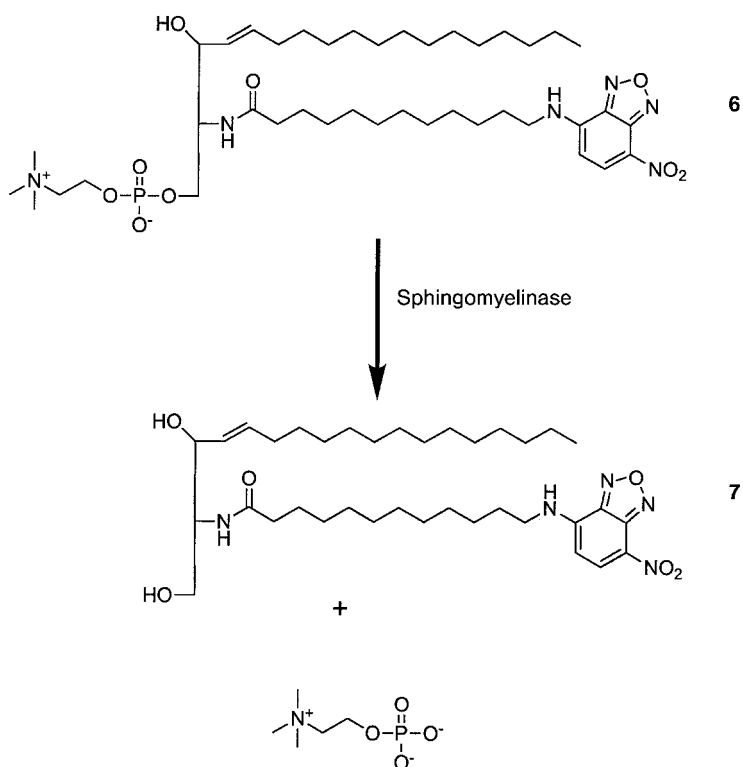


Fig. 9.4 Fluorescent substrate for sphingomyelinase activity assay 6 and its product after cleavage 7.

vine sphingomyelin and a fluorescent analog, suggesting a higher substrate specificity of neutral Smase.

Another fluorescence method for the fast and accurate determination of sphingomyelinases in biological samples has been developed in our laboratory [47]. The assay is based on a fluorescent sphingomyelin analog 6 (Figure 9.4) carrying fluorescent NBD-dodecanoic acid instead of an aliphatic acyl chain at the amide nitrogen atom. The fluorescent substrate is hydrolyzed by Smases to form fluorescent ceramide 7, which can be separated from the remaining substrate using TLC on silica gel. The fluorescence intensity pattern obtained on the TLC plate can accurately be determined using a CCD camera. Typically, a large number of samples can be analyzed simultaneously. Smases from freshly prepared cellular homogenates as well as from commercial sources can be quantified. This assay is suitable as a routine method for determination of Smase activity in pure enzymes and crude enzyme preparations as well as in complex biological samples (e.g. arterial smooth muscle cells). The detection limit is 0.2 pmol of the NBD-fluorophore. The linearity range lies between 0.2 pmol and at least 2000 pmol label.

Procedure 9.3: Sphingomyelinase Assay

A solution of NBD-sphingomyelin 6 in tetrahydrofuran (20 nmol in 30 μ L) is injected with a Hamilton syringe into 2 ml of the appropriate reaction buffer under stirring at room temperature. For determination of acid Smase activity, the buffer is either 250 mM sodium acetate, 1 mM EDTA, pH 5.0, or 20 mM HEPES, 1 mM $MgCl_2$, pH 7.4, for neutral Mg^{2+} -dependent Smase activity, or 20 mM Hepes, pH 7.4, in the case of neutral Mg^{2+} -independent Smase activity. The final substrate concentration is 10 μ M. After addition of an enzyme preparation (e.g. 1.6 units of acid Smase from human placenta) or a biological sample (e.g. cell homogenate of smooth muscle cells corresponding to 20 μ g protein) in a total volume of 200 μ L, the sample is incubated for 1 h at 37°C. The reaction is stopped by adding 600 μ L $CHCl_3/CH_3OH$ 2/1 (v/v) for lipid extraction. The solvent of a 100 μ L aliquot of the organic phase is removed under N_2 , the residue is redissolved in 10 μ L $CHCl_3$ and the resultant solution is quantitatively applied onto a silica TLC plate. The labeled lipids are separated using $CHCl_3/CH_3OH/H_2O$ 65/25/4 (v/v/v) as a mobile phase. Under these conditions, NBD-SM 6 and NBD-ceramide 7 appear as single spots at $R_F=0.15$ and $R_F=0.85$, respectively. The fluorescence intensities are determined with a CCD camera (Herolab) at excitation wavelength of 365 nm using an appropriate software (EasyWin) for data acquisition and processing. Absolute amounts of ceramide per pmol are obtained from a calibration plot of the amount of NBD-ceramide (0.2–20 pmol) versus fluorescence intensity.

9.3

Fluorescent Inhibitors for Quantitative Analysis of Active Enzymes and Functional Enzyme Fingerprinting

Specific and even less specific inhibitors are useful tools in enzyme biochemistry and biophysics provided the mechanisms underlying enzyme function are known [64]. In this context, *p*-nitrophenyl- or fluor- esters of alkylphosphonic acids turned out to be useful suicide inhibitors of serine hydrolases. They have been applied to investigate enzyme–lipid interactions with respect to substrate and stereoselectivity [65–70] and they have been established as powerful probes for enzyme analytics. One technique is based on measurement of *p*-nitrophenol release from *p*-nitrophenyl esters by the active serine of the enzyme [71–73], however, large amounts of enzyme are required to generate measurable quantities of the analyte. Fluorescent labeling of the inhibitor circumvents this limitation because of the much higher sensitivity of fluorescence. The fluorescent inhibitors form covalent and stoichiometric (1/1 mol/mol) complexes even with low abundant enzymes in complex samples that can easily be visualized after separation by 1D or 2D gel electrophoresis. For separation of hydrophobic proteins, chromatography is the method of choice. The fluorescent inhibitors can be used to detect, discover and quantify the absolute amount (mol) of active lipase, esterase, or protease (depending on inhibitor) in virtually any sample ranging from an electrophoretically pure protein to a complex tissue preparation.

9.3.1

Lipase and Esterase Profiling

Lipases and esterases are responsible for extracellular and intracellular lipid degradation. On the other hand they belong to the most widely used enzymes in the biotechnology, organic chemistry, and pharmaceutical industries for racemic resolution of synthetic alcohols or carboxylic acids and transesterification reactions [74–84]. For applications in biocatalysis and industrial processes cheap enzyme preparations are preferred, which are often contaminated with other than the desired lipolytic activities leading to unwanted side-reactions and lower enantioselectivity [85]. Thus knowledge of the enzyme composition of complex proteomes and commercially available protein preparations is important but difficult to determine. To solve the problem, we have developed specific fluorescent inhibitors that covalently and stoichiometrically bind to lipolytic enzymes even in the presence of an excess of other proteins. The following sections are devoted to this novel class of compounds which have meanwhile been established as powerful probes for qualitative and quantitative determination of lipases and esterases in active form.

9.3.1.1 Microbial Lipases and Esterases

Berman et al. were the first to use short-chain fluoro-alkylphosphonates containing pyrene or NBD to characterize acetylcholinesterases from microbial sources [86, 87]. Long- and short-chain fluorescent alkylphosphonates recognizing lipases and esterases were developed in our laboratory [88–90]. Some of these compounds contained a fluorescent dialkylglycero moiety, thus mimicking triacylglycerol substrates of these enzymes.

Scholze et al. used pyrene- and perylene-tagged *p*-nitrophenyl esters of alkyl phosphonic acid (Figure 9.5) to detect and quantify the active protein fractions in crude *Rhizomucor miehei* lipase (RML), purified *Rhizopus oryzae* lipase (ROL) and purified esterase B from *Burkholderia gladioli* (Est B) preparations [88]. The underlying enzyme–inhibitor complexes were prepared using Triton X-100 for solubilization of the hydrophobic inhibitors.

The fluorescently labeled *p*-nitrophenyl esters of 1,2-di-*O*-alkyl-glycero-3-alkylphosphonic acids (**8**, **9**) and alkoxy-alkylphosphonic acid (**10**) (Figure 9.5) were successfully applied as covalent inhibitors for lipases (**8**, **9**) and lipases and esterases (**10**). The ether bonds in the *sn*-2 and *sn*-1(3) positions of glycerol in these compounds are not cleavable by lipase-catalyzed hydrolysis. The *sn*-3(1) positions containing the *p*-nitrophenyl ester bonds are the specific reaction sites for the lipolytic enzymes. Nucleophilic substitution of the *p*-nitrophenyl residue by the active serine in the catalytic site of a lipase or esterase leads to the forma-

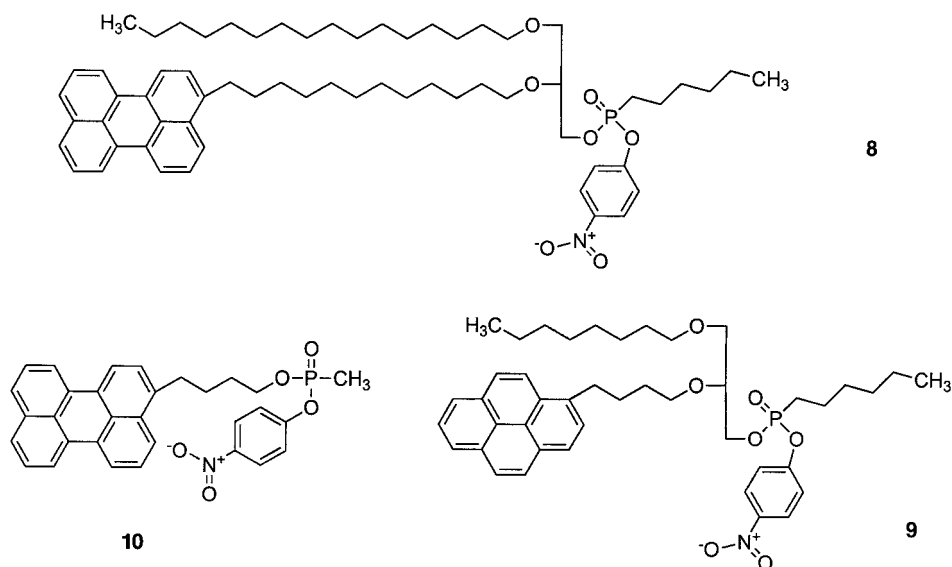


Fig. 9.5 Chemical structures of fluorescent lipase and esterase inhibitors:

8: long-chain triacylglycerol-analogue inhibitor;

9: short-chain triacylglycerol-analogue inhibitor;

10: single-chain inhibitor.

tion of a stable and stoichiometric (1/1) complex of lipid and protein. *p*-Nitrophenol is a very good leaving group, which facilitates the reaction. The glycerophosphonates contain two chiral centers, the *sn*-2 carbon of glycerol and the phosphorus atom. The configuration at the phosphorus is crucial for inhibitor reactivity, the S_P enantiomer being a much better substrate than the R_P lipid. For enzyme analytics, inhibitors racemic at phosphorus are also suitable because the lipid is always used in excess and the presence of the less reactive R_P compound has no influence on the reaction of the S_P lipid. Reactivity also depends on the configuration of the chiral glycerol C-atom, but to a lesser extent than the stereochemistry at phosphorus. Inhibitor–protein complexes are stable and can thus be detected individually after separation by SDS gel electrophoresis. During electrophoresis the surplus inhibitor is removed and the fluorescence of the bound inhibitor can selectively be detected. In conjunction with calibration of the fluorescence signal intensity, the moles of inhibitor–enzyme complexes and therefore the amount of active enzyme mass can be quantified in any protein or tissue preparation [88].

In addition, we developed fluorescent 1-*O*-alkyl-2-acylamino-*sn*-glycero-hexylphosphonic acid-*p*-nitrophenyl esters derivatives bearing a perylene or NBD label at the ω -position of the *sn*-2 acyl-chain. Two single-chain inhibitors containing the same fluorophores were synthesized too. Three highly homologous enzymes, namely *Rhizopus oryzae* lipase, *Pseudomonas cepacia* lipase, and *Pseudomonas* species lipase were investigated using this probe. Despite their overall structural similarities, these enzymes showed different reactivities. Obviously the enzymes must differ with respect to the structures of their active sites resulting in different function [90].

Recently, 19 different lipase and esterase preparations, mostly from microbial sources, were screened in our laboratory using a new library based on NBD-labeled inhibitors (Schmidinger et al., unpublished data). The respective long- and short-chain phosphonates were synthesized employing two labeling strategies. Esterase inhibitors were labeled at their alcohol moiety, whereas the lipase inhibitors were labeled in ω -position of the alkylphosphonic acid residue. Enzyme preparations were probed with the inhibitors in the Tris–HCl/Triton X-100 buffer system described above, separated by 1D SDS-PAGE and identified using a laser scanner.

Procedure 9.4: Lipase Fluorescence Tagging and Analysis

In a typical experiment, 10 μ L of the inhibitor stock solution (0.1 mM in CHCl_3) and 2 μ L of a Triton X-100 solution (10 mM in CHCl_3) were added to an Eppendorf tube, and the organic solvent was removed under a gentle stream of argon. The enzymes were dissolved in 0.1 M Tris–HCl buffer at their pH optimum and 20 μ L of the resultant enzyme stock solution was added to the solvent-free inhibitor Triton X-100 mixture (final concentrations: 50 μ M inhibitor and 1 mM Triton X-100). After short vigorous shak-

ing, the mixture was incubated at 37 °C on an Eppendorf shaker for 2 h. Subsequently, SDS-loading buffer was added to the samples, heated to 95 °C for 5 min, and then subjected to 1D SDS-PAGE. The fluorescent inhibitor–enzyme complexes were visualized using laser excitation at 488 nm and the appropriate emission filters at 530 nm.

The fluorescent background in gels was reduced when NBD instead of pyrene or perylene was the fluorophore. The short-chain inhibitors preferably interacted with esterases whereas the long-chain inhibitors were only recognized by lipases. One of the most selective inhibitors of the library was the cholesterol ester derivative. It solely reacted with the cholesterol esterase fraction in enzyme preparations from bovine or porcine pancreas. Recognition of glycerolipids by the lipolytic enzyme was highly stereoselective. 1,2-Di-*O*-hexadecyl-*sn*-glycero-3-phosphonate was a much better inhibitor of cholesterol esterase than 2,3-di-*O*-hexadecyl-*sn*-glycero-1-phosphonate. Finally, enzyme activity mainly depended on the alkyl phosphonic acid chain length and the overall hydrophobicity of the alcohol moiety.

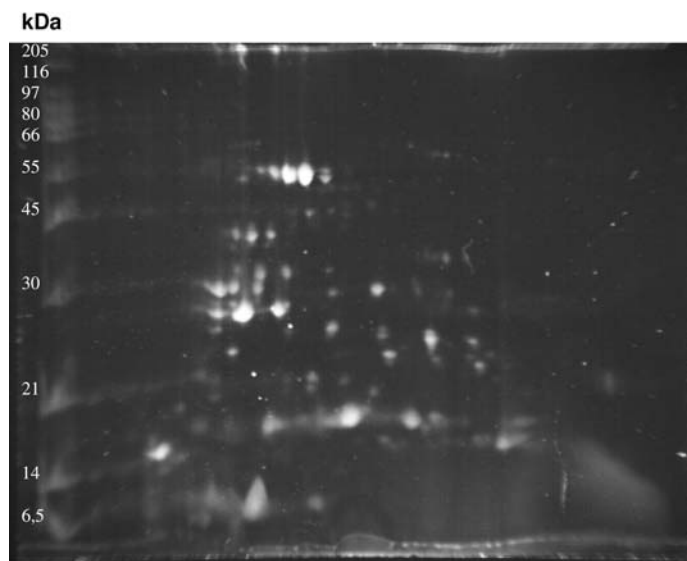
9.3.1.2 Porcine Pancreatic Lipase

Lipase preparations from porcine pancreas (ppL) are among the cheapest and most widely used lipases for biotransformation of natural and non-natural substrates. Sixteen per cent of all lipase reactions are performed with such preparations. Their composition, which was largely unknown, was investigated in our laboratory [91] using fluorescent alkylphosphonate inhibitors (8–10) (Figure 9.5). The enzyme–inhibitor complexes were prepared as described above using Tris–HCl buffers containing 1 mM Triton X-100 or 3 mM taurodeoxycholic acid to solubilize the inhibitors.

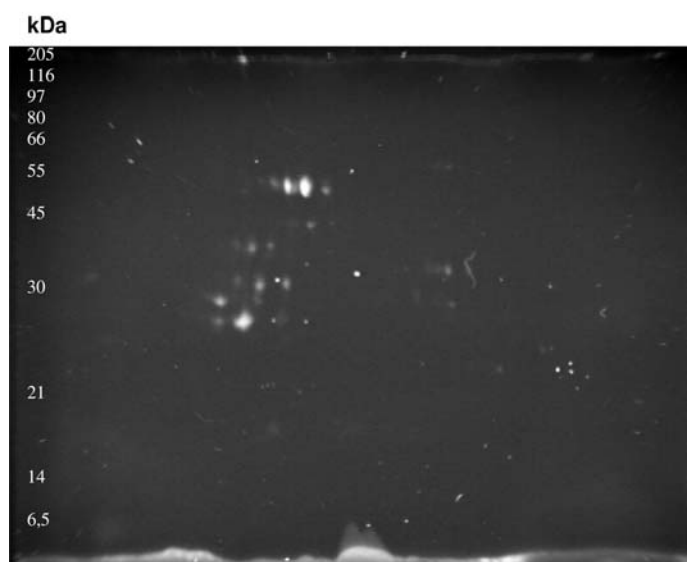
Crude ppL preparations contain many esterolytic and lipolytic components. Some of them, such as cholesterol esterase, pancreatic lipase, α -chymotrypsin or chymotrypsinogen, are known. Pure ppL itself is not active on the glycerolipid inhibitors but reacts only with the single-chain inhibitors. We found that commercial ppL preparations contain a large number of lipolytic enzymes, and confirmed by 2D gel electrophoresis that purified ppL itself exists in different glycosylated forms (horizontal protein ladder at 50 kD) (Figure 9.6).

9.3.1.3 Hormone-sensitive Lipase

Hormone-sensitive lipase (HSL), which is important for fatty acid mobilization in fat cells, can also be functionally analyzed using fluorescent phosphonates, but under very specific reaction conditions. HSL exhibits some unusual properties compared with other lipases, including the reported intolerance to detergents. If Triton X-100 is used to solubilize the inhibitors in the aqueous medium, the detergent concentration has to be low to minimize inactivation (≤ 1 mM Triton X-100). Pure HSL was a recombinant enzyme from baculovirus-



a)



b)

Fig. 9.6 2D electrophoretic separation of pure PPL labeled with compound **10**. Pure PPL (Sigma) was separated by 2D electrophoresis (first dimension: nonlinear pH gradient from 3 to 10, second dimension: 12.5% SDS-PAGE) after preincubation with **10** in the taurodeoxycholic acid system.

Proteins labeled by compound **10** (Figure 9.5) (b) and total proteins stained with Sypro RubyTM (Molecular Probes) (a) are shown in the printout of a fluorescence gel documentation system. Excitation wavelength was 254 nm.

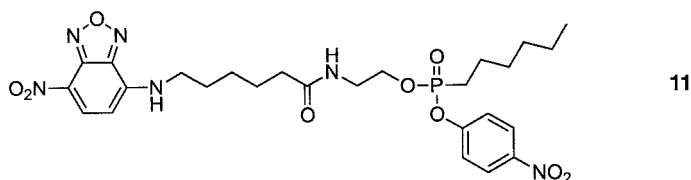


Fig. 9.7 Structure of compound 11.

insect cell systems and was a kind gift from Holm, Lund, Sweden. The enzyme was dissolved in aqueous buffer containing 5 mM sodium phosphate, pH 7.40, 1 mM dithioerythreitol (DTE), 0.2% C₁₃E₁₂ (polydisperse preparation of alkyl polyoxyethylenes with the average composition specified by the manufacturer; C, alkyl carbons; E, oxyethylene units) and 50% glycerol. Aliquots (20 μ L) of HSL containing 0.2 mg mL⁻¹ protein were stored at -70 °C. Shortly prior to use the enzyme solution was diluted with buffer (20 mM potassium phosphate, 1 mM EDTA, 1 mM dithioerythreitol and 0.02% defatted BSA, pH 7.0).

For HSL analysis, compounds **8** and **10** (Figure 9.5) and compound **11** (Figure 9.7) were used as covalent inhibitors. HSL reacted much faster with the single-chain inhibitors **10** and **11** compared with the triacylglycerol analog **8**. This result is in line with its significant esterase activity. The fraction of active enzyme in the recombinant HSL preparation studied was determined using **10** as an inhibitor and found to be 95% of total protein mass. Compound **10** inhibited both the diacylglycerol hydrolase and the cholesterol esterase activity of HSL, as determined using Bodipy-FL cholesterol ester **5** and NBD-monoacyl-monoalkylglycerol **4** as substrates, respectively.

9.3.2

Probing Biophysical Enzyme Properties

A new inhibitor containing 6-acryloyl-2-dimethylaminonaphthalene (acrylodan) as a solvent-sensitive fluorophore was synthesized in our laboratory to probe polarity and conformational changes in the active site of lipases. Dramatic effects were observed when the enzymes were exposed to different reaction media that had previously been shown to affect enzyme activity and stereoselectivity [7].

The enzymes under investigation (*Rhizopus oryzae* lipase, ROL, *Pseudomonas cepacia* lipase, PCL, and *Chromobacterium viscosum* lipase, CVL) were labeled with this lipid followed by removal of excess inhibitor by dialysis. Fluorescence spectra of the inhibitor–enzyme complexes were then measured in different reaction media, such as 0.1 M Tris–HCl, pH 7.40 with or without 2 mM Triton X-100, or a 1:1 mixture of 0.1 M Tris–HCl, pH 7.40 with ethanol (Figure 9.8). Inhibitor complexes of CVL and PCL showed nearly identical emission maxima at 489 and 488 nm, which were 18 and 17 nm red-shifted relative to the ROL complex, respectively. Therefore, the CVL- and PCL-bound labels seemed to be

more accessible to water compared with ROL under identical conditions. In detergent-free aqueous solution the emission maximum of the CVL complex was blue-shifted, indicating that the fluorophore experienced a more hydrophobic environment under these conditions. The spectra of the PCL–lipid complex were nearly unaffected by the different solvents, indicating that the enzyme did not undergo dramatic conformational changes. Thus PCL is very stable in the different solvents.

Compound **9** (Figure 9.5) and its *sn*1-glycero-phosphonyl derivative were used to study the interactions of ROL with the individual alkyl segment of the triglyceride analog [92]. For this purpose, fluorescence resonance energy transfer (FRET) between the pyrene-labeled acyl chain of the inhibitor with the tryptophans of the inhibited enzymes was determined using the following protocol.

Procedure 9.5: FRET Lipase Assay

To a solution of 50 μg ROL in 450 μL buffer (0.1 M Tris–HCl, pH 7.4) 50 μL of inhibitor solubilized with Triton X-100 (10 mM) in buffer (0.1 M Tris–HCl, pH 7.4) were added, followed by incubation for 10 h at 4 °C. For FRET measurements the probe was excited at 295 nm (absorption maximum of tryptophan) and the emission intensities at 340 nm (tryptophan, FRET donor) and 378 nm (pyrene, FRET acceptor) were determined. Measurements were performed at 30 °C using a Shimadzu spectrofluorometer RF-540. Enzyme–lipid conjugates in solvent systems other than Triton X-100 micelles were obtained according to the following procedure. A 50 μL aliquot of a solution of inhibitor (40 μM) solubilized with Triton X-100 (10 mM) in Tris–HCl buffer (0.1 M, pH 7.4) was added to a solution of ROL (175 μL , 7.4 nmol mL⁻¹) in buffer (0.1 M Tris–HCl, pH 7.4). The samples were incubated overnight at 4 °C, diluted with Tris–HCl (0.1 M, pH 7.4) to a final volume of 1 mL, followed by dialysis (Servapor dialysis tubes from Serva, Heidelberg, Germany) for 24 h against Tris–HCl (0.1 M, pH 7.4), Tris–HCl (0.1 M, pH 7.4) containing *N*-dodecyl-*N,N*-dimethylammonio-3-propanesulfonate (SB-12, 3.2 mg mL⁻¹), or ethanol:H₂O (1:1, v/v).

FRET was observed with both stereoisomeric inhibitors, indicating that protein tryptophans and the covalently bound fluorescent inhibitor were within the crucial FRET distance of 25 Å reported for this donor–acceptor pair. Pyrene fluorescence emission upon tryptophan excitation increased with increasing inhibitor-to-protein ratios, reaching a maximum around 1. FRET efficiencies between protein tryptophans and fluorescent inhibitor were the same for both stereoisomers ($R_C S_P$, $S_C S_P$). The similar extent of FRET observed in complexes of ROL with either stereoisomer points to similar interfluorophoric distances and/or similar orientation of pyrene and lipase tryptophans in both systems. Thus the *sn*-2 alkyl chain in the protein-bound inhibitors appears to be accommodated in the same binding region of the lipase irrespective of stereochemistry at glycerol.

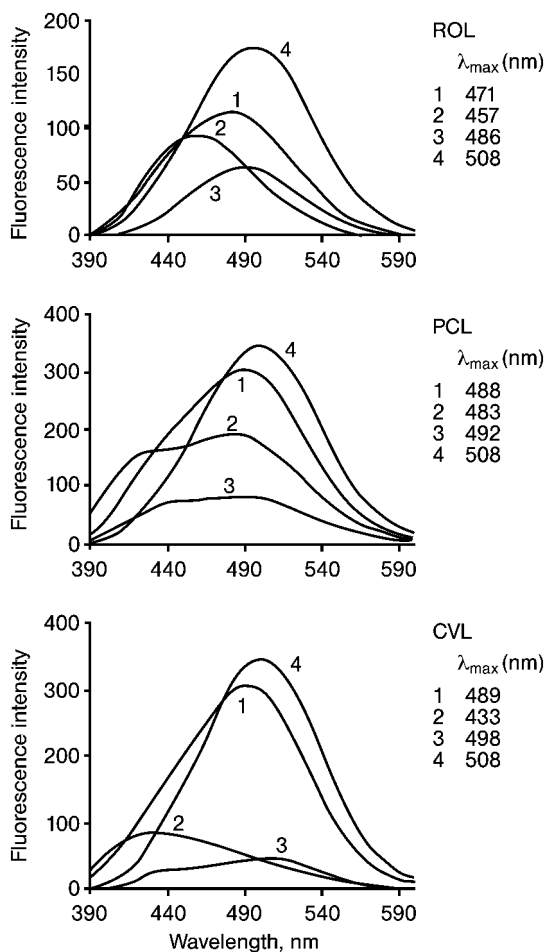


Fig. 9.8 Solvent effects on fluorescence emission spectra. *Rhizopus oryzae* lipase (ROL) ($0.2 \mu\text{M}$) was incubated with a 10-fold molar excess of solvent-sensitive inhibitor in 2 mM Triton X-100, 0.1 M Tris-HCl (pH 7.4) at 4°C for 3 h. Incubation times for *Chromobacterium viscosum* lipase (CVL) and *Pseudomonas cepacia* lipase (PCL) were 30 h. Unreacted inhibitor was removed by dialysis against 2 mM Triton X-100 in 0.1 M Tris-HCl (pH 7.40 at 37°C) (1). The solvent was changed by further dialysis against 0.1 M

Tris-HCl (pH 7.40) (2), followed by dialysis against Tris-HCl/ethanol (1:1, by vol.) (3) (each dialysis step was carried out for 24 h at 4°C in the dark). Spectra (4) represent fluorescence emission spectra of the inhibitor without enzyme in 2 mM Triton X-100. Spectra were recorded at 25°C (excitation at 370 nm). The means of three independent experiments are presented. Standard deviations were lower than 5% in all cases.

However, interfluorophore distances and thus lipid–protein interaction very much depend on the solvent. In Tris–HCl very efficient FRET between the lipase and the inhibitor was found, whereas in ethanol/H₂O no FRET was observed. In SB-12 micelles the extent of FRET was lower compared with aqueous buffer but still detectable. The decrease in FRET efficiency observed in the latter two systems reflects profound changes in protein conformation, which are also associated with changes in enzyme activity since ROL is neither active in ethanol:water (1:1, v/v) nor in the presence of SB-12 micelles. Thus the active enzyme form can be associated with a compact conformation whereas in the presence of ethanol or detergent the inactive proteins exhibit a more extended conformation.

9.3.3

Affinity-based Proteome Profiling (ABPP)

The elucidation of the genetic maps of many microorganisms, fungi, vertebrates, and the completion of the Human Genome Project paved the way to examine RNA abundance of different states of the organism or tissue under investigation. Therefore, two major multiplexing techniques have been well established which are DNA microarrays and real-time PCR. Those methods give insight into the transcriptome of the system investigated, but nevertheless, differences in protein expression, modification and activity cannot directly be deduced from those data, since RNA levels do not necessarily correlate with protein abundance and, even more importantly, protein or enzyme activity. Researchers are now challenging the proteome with methods like 2D gel electrophoresis followed by mass spectrometry for protein abundance examination, yeast two-hybrid screens for protein–protein interaction and various other methods for abundance-based proteome profiling. Again, the amount of proteins present at a certain state of a cell might not correlate with enzyme activities responsible for the metabolic fluxes, cell management, and signal transduction. Therefore, elucidation of changes in protein activity is the ultimate goal of proteomics.

The detection of protein activity can be facilitated using affinity labels. Basically, an affinity label is a molecule consisting of (1) a recognition site targeting a certain enzyme species, (2) a properly positioned reactive site which forms a covalent bond with the target, and (3) a tag for visualization and or purification of the covalently bound target [93–96]. Considering serine hydrolase and some protease enzymes, fluorescently labeled *p*-nitrophenyl- and fluoro-alkylphosphonates meet the above prerequisites. Based on their polarity and recognition moieties they are excellent baits to profile serine hydrolase activities in complex proteomes.

9.3.3.1 Functionality-based Serine Hydrolase Profiling in Tissue Preparations and Cell Lines

In our laboratory, the fluorescent phosphonic acid ester **11** (Figure 9.7) was used as a substrate-analogous suicide inhibitor of lipases and esterases to identify lipolytic enzymes in mouse adipose tissue (Birner-Gruenberger et al., unpublished results). Compound **11** was able to recognize a broad range of isolated lipases and esterases, demonstrating its usefulness as a bait for a variety of lipolytic enzymes. A photometric *p*-nitrophenyl laurate assay was employed to measure the initial lipolytic activity in homogenates of white and brown murine adipose tissue. Activity of the tissue homogenates was completely abolished upon addition of a large excess of **11**, indicating that all lipolytic enzymes active towards *p*-nitrophenyl laurate were inhibited using **11**.

For proteomic analysis, homogenates of white and brown mouse adipose tissue (0.5 mg mL^{-1} protein) were incubated with $20 \text{ }\mu\text{M}$ of compound **11** and 1 mM Triton X-100 in 10 mM Tris-HCl, pH 7.4, 0.25 M sucrose at 37°C for 2 h. The inhibitor-enzyme complexes were resolved by 2D gel electrophoresis, detected with a laser scanner and compared with the whole protein stain Sypro RubyTM (Molecular Probes) (Figure 9.9). Spots showing NBD fluorescence were cut out and after tryptic in-gel digest the isolated peptides were analyzed by nano-HPLC/MS/MS. Identified proteins were esterolytic enzymes involved in fatty acid metabolism, as well as known lipases, namely monoglyceride lipase, hormone-sensitive lipase,

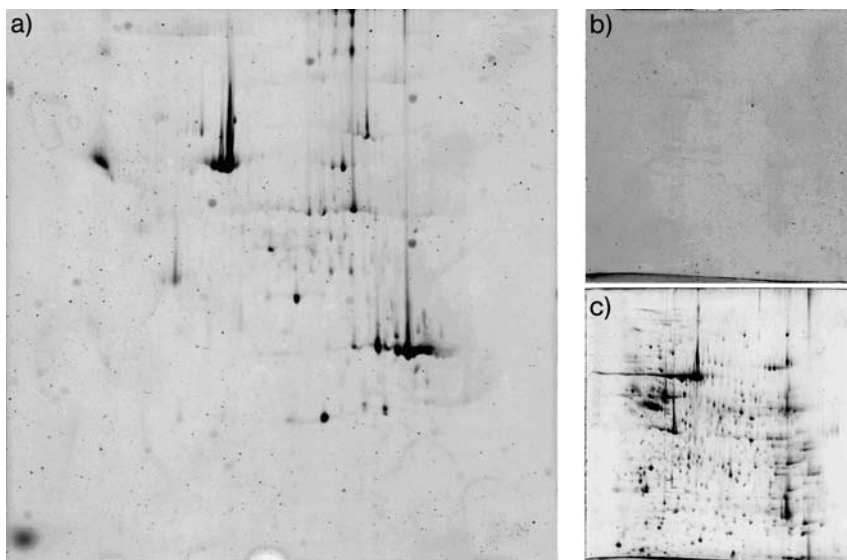


Fig. 9.9 Lipolytic proteome of mouse adipose tissue. White mouse adipose tissue was incubated with compound **11**, separated by 2D gel electrophoresis (first dimension: nonlinear pH gradient from 3 to 10,

second dimension: 10% SDS-PAGE). Proteins labeled by **11** are shown in (a) as compared to the unlabeled sample (b) and the whole protein sample stained by Sypro RubyTM (c).

and triacylglycerol hydrolase. In addition, so far uncharacterized potential lipases were identified. As a control, some of the identified proteins were transiently expressed in COS 7 cells and labeled with the inhibitor to ensure that the identified proteins were targets inhibited by compound 11.

Patricelli et al. reported on the visualization of serine hydrolase activity in COS 7 cells transfected with fatty acid amide hydrolase and in liver and testis tissue preparations from rats. They compared the same phosphonate-based inhibitor in three modifications: a tetramethylrhodamine and a fluorescein derivative (FP-PEG-fluorescein/tetramethylrhodamine) for direct visualization and a biotin-labeled analog (FP-PEG-biotin) [97].

Procedure 9.6: Proteome Profiling with Fluorophosphonates

The inhibitor was added to 50 μL of the chosen proteome (1 mg mL⁻¹ protein) in 50 mM Tris-HCl pH 8 for the biotin derivative or 50 mM Tris-HCl, pH 7.0, 100 mM NaCl for the fluorescent derivatives, respectively. For comparison of the tetramethylrhodamine and the biotin derivative, the final inhibitor concentration in the sample was 5 μM , whereas for the work with the fluorescent derivatives a final concentration of 2 μM was chosen to give best results. The samples were allowed to stand at room temperature for 1 h prior to quenching with an equal volume of 2 \times SDS loading buffer (125 mM Tris-HCl, pH 6.75, 20% glycerol, 4% SDS, 10% 2-mercaptoethanol, 0.005% bromophenol blue). An aliquot (0.025 mL, 0.018 mg protein) of the solution was removed and separated on a 10% SDS gel. For the FP-PEG-Biotin treated proteomes, labeling was visualized using western blot standard protocols. Fluorescently labeled samples were visualized in-gel on a HitachiFMBio Iie flatbed fluorescence scanner (MiraiBio, Alameda, CA, USA) with excitation provided by the 532 nm line of a 50 mW neodymium-doped yttrium-aluminium-garnet (Nd:YAG) laser. A 605 nm bandpass filter was used to detect FP-PEG-tetramethylrhodamine while a 505 nm bandpass filter was used for FP-PEG-fluorescein.

Using this procedure, 100 amol (equivalent to 100 amol tetramethylrhodamine) of fatty acid amid hydrolase could be detected reliably with the respective inhibitor. This sensitivity is more than 100 times better than the detection limit of 15 fmol for the biotin-based inhibitors [98]. They also multiplexed their proteome analysis by employing inhibitors featuring two spectrally different fluorophores (tetramethylrhodamine and fluorescein). Therefore, a rat liver tissue preparation and its counterpart originating from testis were probed with one of the inhibitors described above, samples were combined, separated on 1D SDS-PAGE and imaged simultaneously. As a result, two different tissues were successfully analyzed within the same lane of a gel.

Jessani et al. used a similar approach profiling enzyme activity of the secreted, soluble and membrane proteomes of several human cancer cell lines

[99]. Standard conditions for fluorophosphonate proteome reactions were as described above, in short proteomes were adjusted to a final protein concentration of 1 mg mL^{-1} and treated with 1 or $4 \text{ }\mu\text{M}$ rhodamine-coupled fluorophosphonate inhibitor at room temperature for 1 h. After labeling, a portion of each proteome sample was treated with peptide-*N*-glycosidase F (PNGaseF) (New England Biolabs, Beverly, MA, USA) to provide deglycosylated proteomes. Reactions were quenched with one volume of standard $2\times$ SDS-PAGE loading buffer (reducing), separated by SDS-PAGE (10–14% acrylamide), and visualized in-gel with a Hitachi FMBio Iie flatbed fluorescence scanner (MiraiBio) as described above. Integrated band intensities (normalized for volume) were calculated for the labeled proteins.

For each enzyme activity, 4–6 data points were generated from independent labeling reactions conducted on 2 or 3 independently prepared proteomic samples. These data points were averaged to provide the level of each enzyme activity in each cell line. The activity levels of each enzyme were compared across the cell lines by using the Tukey's honestly significant difference test, where *P*-values >0.05 were considered statistically significant.

The results indicate that the inhibitors specifically and quantitatively detect serine hydrolases that depict invasiveness of cancer cell lines, and in addition, that those enzymes are preferably located in the secreted and membrane proteome of the cell lines. Cluster analysis of all serine hydrolase activities as determined by this approach revealed that the cancer cell lines used in this study segregated into three main groups: a melanoma cluster, a breast carcinoma cluster, and an invasive cancer cluster which were in turn in good correlation with the origin and phenotype-driven classification.

The same fluorescent suicide inhibitors were taken to screen inhibitory effects of noncovalent binding inhibitors in complex proteomes [100]. The method is based on the assumption that the fluorescent alkylphosphonate suicide inhibitor can only bind to the active site of an enzyme if this site is not occupied by a reversible binding inhibitor. A strong reversible inhibitor would diminish the fluorescence intensity of an enzyme when probed with the fluorescent “reporter inhibitor”, as binding of the latter is largely retarded. This approach was tested using mouse tissue proteomes from heart and brain which were probed with a large variety of noncovalent binding trifluoromethylketone, α -keto-pyridyloxazole inhibitors and the tetramethylrhodamine-labeled fluoroalkylphosphonate (FP-rhodamine), respectively.

Procedure 9.7: Competitive Proteomic Profiling of Noncovalent Binding Inhibitors with FP-Rhodamine

Mouse tissues were Dounce-homogenized in Tris buffer (50 mM Tris-HCl, pH 8.0) with 320 mM sucrose and membrane proteomes were isolated by centrifugation at 4°C ($22000\times g$), washed, resuspended in Tris-HCl buffer, and adjusted to a protein concentration of 1 mg mL^{-1} . Proteomes were pre-

incubated with inhibitors (10–100 000 nM, DMSO stocks) for 10 min and then treated with FP-rhodamine (100 nM) at room temperature for 10 min (final DMSO concentration 2%). Reactions were quenched with SDS-PAGE loading buffer, subjected to SDS-PAGE and visualized in-gel using a flatbed fluorescence scanner (MiraBio). Labeled proteins were quantified by measuring integrated band intensities (normalized for volume); control samples (DMSO alone) were considered 100% activity and inhibitor-treated samples were expressed as the percentage of remaining activity. Potent inhibitors (IC_{50} values <10 nM) were also tested at 0.5, 1, and 5 nM with proteomes adjusted to $0.1 \text{ mg protein mL}^{-1}$, so that the estimated concentration of target enzymes was kept at least fivefold below the inhibitor concentration. IC_{50} values were determined from dose–response curves from three trials at each inhibitor concentration using Prism software (GraphPad).

For two enzymes, fatty acid amide hydrolase and triacylglycerol hydrolase, the IC_{50} values measured by this approach were compared with the K_I values obtained with the same noncovalent binding inhibitors using [^{14}C]oleamide and *p*-nitrophenyl laurate as substrates, respectively [101]. The values of these two parameters correlated within a range of 5- to 20-fold for most of the reversible binding inhibitors under the assay condition described above. Nevertheless, this method seems to be rather powerful for qualitative prescreening of possible inhibitory effects of such inhibitors in complex proteomes.

For the sake of completeness it should be stated that biotin tagged affinity probes are also used to profile complex proteomes. Labeling of proteomes with such affinity probes is done similarly to methods described above for the fluorescently labeled derivatives. To visualize targets, proteins are electroblotted on membranes after separation on 1D SDS-PAGE, incubated with avidin–horseradish peroxidase conjugates and, subsequently, detected with luminol H_2O_2 under standard western blotting reaction conditions [98, 102, 103].

In summary, fluorescently labeled alkyl phosphonic acid esters are potent and extremely versatile tools for the rapid identification and classification of lipase and esterase activity in defined protein samples as well as complex proteomes. They have been used successfully for different applications such as the rapid and convenient identification of changes in activity of distinct enzymes in a certain proteome, the quantitative and qualitative determination of active enzymes therein and identification of potential targets for noncovalent inhibitors using the fluorescent inhibitors as reporter molecules.

Acknowledgment

This work was supported by the Austrian Federal Ministry for Education, Science and Culture (GOLD: Genomics of Lipid-associated Disorders) project (<http://gold.uni-graz.at/>) which is a joint GEN-AU (GENome research in Austria) project (<http://www.gen-au.at/>).

References

- 1 H. Wong, M.C. Schotz, *J. Lipid Res.* **2002**, *43*, 993–999.
- 2 I. Panaitov, R. Verger, in *Physical Chemistry of Biological Interfaces*, eds A. Baszkin, W. Norde, Marcel Dekker, New York, Basel, **2000**, pp. 359–400.
- 3 F. Beisson, A. Tiss, C. Riviere, R. Verger, *Eur. J. Lipid Sci. Technol.* **2000**, *102*, 133–153.
- 4 N.W. Tietz, *Clin. Chim. Acta* **1997**, *257*, 85–98.
- 5 D.L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S.J. Remington, I. Silman, J. Schrag, *Protein Eng.* **1992**, *5*, 197–211.
- 6 J.D. Schrag, M. Cygler, *Methods Enzymol.* **1997**, *284*, 85–107.
- 7 G. Zandonella, L. Haalck, F. Spener, K. Faber, F. Paltauf, A. Hermetter, *Eur. J. Biochem.* **1995**, *231*, 50–55.
- 8 R. Gupta, P. Rath, N. Gupta, S. Bradoo, *Biotechnol. Appl. Biochem.* **2003**, *37*, 63–71.
- 9 P. Desnuelle, M.J. Constantin, J. Baldy, *Bull. Soc. Chim. Biol. (Paris)* **1955**, *37*, 285–290.
- 10 P. Belfrage, M. Vaughan, *J. Lipid Res.* **1969**, *10*, 341–344.
- 11 W.G. Duncombe, *Biochem. J.* **1963**, *88*, 7–10.
- 12 O. Hirayama, H. Matsuda, *Agric. Biol. Chem.* **1972**, *36*, 1831–1833.
- 13 N. Gupta, P. Rath, R. Gupta, *Anal. Biochem.* **2002**, *311*, 98–99.
- 14 U.K. Winkler, M. Stuckmann, *J. Bacteriol.* **1979**, *138*, 663–670.
- 15 P. Ciuffreda, A. Loseto, A. Manzocchi, E. Santaniello, *Chem. Phys. Lipids* **2001**, *111*, 105–110.
- 16 A. Kademi, N. Ait-Abdelkader, L. Fakhreddine, J. Baratti, *Appl. Microbiol. Biotechnol.* **2000**, *54*, 173–179.
- 17 E.W.J. Mosmuller, J.D.H. Van Heemst, C.J. Van Delden, M.C.R. Franssen, J.F.J. Engbersen, *Biocatalysis* **1992**, *5*, 279–287.
- 18 J. Whitaker, *Clin. Chim. Acta* **1973**, *44*, 133–138.
- 19 A. Rawlyer, P. A. Siegenthaler, *Biochim. Biophys. Acta* **1989**, *1004*, 337–344.
- 20 J.F. Jette, E. Ziomek, *Anal. Biochem.* **1994**, *219*, 256–260.
- 21 A.M. Rogel, W.L. Stone, F.O. Adebo-
nojo, *Lipids* **1989**, *24*, 518–525.
- 22 D.C. Wilton, *Biochem. J.* **1991**, *276*
(Pt 1), 129–133.
- 23 S. Ulitzur, J.W. Hastings, *Proc. Natl
Acad. Sci. USA* **1979**, *76*, 265–267.
- 24 S. Ulitzur, *Biochim. Biophys. Acta* **1979**,
572, 211–217.
- 25 K.W. Cho, *Misaengmul Hakhoechi* **1993**,
31, 532–536.
- 26 N. Prim, M. Sanchez, C. Ruiz, F.I. Javier
Pastor, P. Diaz, *J. Mol. Catal. B Enzym.*
2003, *22*, 339–346.
- 27 P. Diaz, N. Prim, F.I.J. Pastor, *Biotech-
niques* **1999**, *27*, 693–700.
- 28 T.J. Jacks, H.W. Kircher, *Anal. Biochem.*
1967, *21*, 279–285.
- 29 T. De Laborde de Monpezat, B. De Jeso,
J.L. Butour, L. Chavant, M. Sancholle,
Lipids **1990**, *25*, 661–664.
- 30 A. Negre, R. Salvayre, A. Dagan, S. Gatt,
Biochim. Biophys. Acta **1989**, *1006*, 84–88.
- 31 J.D. Johnson, M.R. Taskinen, N. Mat-
suoka, R.4L. Jackson, *J. Biol. Chem.*
1980, *255*, 3461–3465.
- 32 C. Wolf, L. Sagaert, G. Bereziat, *Biochem.
Biophys. Res. Commun.* **1981**, *99*, 275–283.
- 33 L.A. Wittenauer, K. Shirai, R.L. Jackson,
J.D. Johnson, *Biochem. Biophys. Res.
Commun.* **1984**, *118*, 894–901.
- 34 T. Thuren, J.A. Virtanen, R. Verger, P.K.
Kinnunen, *Biochim. Biophys. Acta* **1987**,
917, 411–417.
- 35 H.S. Hendrickson, P.N. Rauk, *Anal. Bio-
chem.* **1981**, *116*, 553–558.
- 36 A.E. Negre, R.S. Salvayre, A. Dagan, S.
Gatt, *Clin. Chim. Acta* **1985**, *149*, 81–88.
- 37 M. Duque, M. Graupner, H. Stutz,
I. Wicher, R. Zechner, F. Paltauf, A.
Hermetter, *J. Lipid Res.* **1996**, *37*, 868–876.
- 38 F. Beisson, N. Ferte, J. Nari, G. Noat, V.
Arondel, R. Verger, *J. Lipid Res.* **1999**, *40*,
2313–2321.
- 39 B.A. Wolf, *Methods Mol. Biol.* **1998**, *105*,
167–174.
- 40 C. Holm, G. Olivecrona, M. Ottosson,
Methods Mol. Biol. **2001**, *155*, 97–119.
- 41 T. Levade, M. Leruth, D. Graber, A.
Moisand, S. Vermeersch, R. Salvayre,
P.J. Courtroy, *J. Lipid Res.* **1996**, *37*,
2525–2538.

- 42 A. G. Barbone, A. C. Jackson, D. M. Ritchie, D. C. Argentieri, *Methods Enzymol.* **2000**, *311*, 168–176.
- 43 D. F. Hassler, R. M. Laethem, G. K. Smith, *Methods Enzymol.* **2000**, *311*, 176–184.
- 44 B. Liu, Y. A. Hannun, *Methods Enzymol.* **2000**, *311*, 164–167.
- 45 X. Zhou, F. Turecek, C. R. Scott, M. H. Gelb, *Clin. Chem.* **2001**, *47*, 874–881.
- 46 D. Samet, Y. Barenholz, *Chem. Phys. Lipids* **1999**, *102*, 65–77.
- 47 A. Loidl, R. Claus, H. P. Deigner, A. Hermetter, *J. Lipid Res.* **2002**, *43*, 815–823.
- 48 E. Grund, G. Zandonella, F. Paltauf, A. Hermetter, *J. Fluorescence* **1994**, *4*, 365–366.
- 49 F. Paltauf, E. Wagner, *Biochim. Biophys. Acta* **1976**, *431*, 359–362.
- 50 G. Zandonella, L. Haalck, F. Spener, K. Faber, F. Paltauf, A. Hermetter, *Chirality* **1996**, *8*, 481–489.
- 51 T. Xia, R. A. Coleman, *Biochim. Biophys. Acta* **1992**, *1126*, 327–336.
- 52 P. W. Majerus, S. M. Prescott, *Methods Enzymol.* **1982**, *86*, 11–17.
- 53 H. Tornqvist, P. Belfrage, *Methods Enzymol.* **1981**, *71*, 646–652.
- 54 C. Holm, J. A. Contreras, *Methods Mol. Biol.* **1999**, *109*, 165–175.
- 55 C. Holm, T. Osterlund, *Methods Mol. Biol.* **1999**, *109*, 109–121.
- 56 N. J. Haley, S. Fowler, C. de Duve, *J. Lipid Res.* **1980**, *21*, 961–969.
- 57 R. A. Coleman, E. B. Haynes, *Biochim. Biophys. Acta* **1983**, *751*, 230–240.
- 58 V. F. Pozdnev, K. S. Planutis, A. I. Tochilkin, *Bioorg. Khim.* **1991**, *17*, 1347–1351.
- 59 R. A. Moreau, *Plant Sci.* **1986**, *47*, 1–9.
- 60 M. S. Izquierdo, R. J. Henderson, *Fish Physiol. Biochem.* **1998**, *19*, 153–162.
- 61 R. A. Moreau, *Lipids* **1989**, *24*, 691–699.
- 62 D. Huster, P. Müller, K. Arnold, A. Herrmann, *Biophys. J.* **2001**, *80*, 822–831.
- 63 F. M. Goni, A. Alonso, *FEBS Lett.* **2002**, *531*, 38–46.
- 64 D. A. Campbell, A. K. Szardenings, *Curr. Opin. Chem. Biol.* **2003**, *7*, 293–303.
- 65 C. M. Taylor, C. J. O'Connor, J. L. Chamberlain, C. Q. Sun, *J. Mol. Catal. B Enzym.* **2001**, *15*, 15–22.
- 66 M. L. M. Mannesse, J.-W. P. Boots, R. Dijkman, A. T. Slotboom, H. T. W. M. van der Hijden, M. R. Egmond, H. M. Verheij, G. H. de Haas, *Biochim. Biophys. Acta Lipids Lipid Metab.* **1995**, *1259*, 56–64.
- 67 R. Hamilton, B. Walker, B. J. Walker, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1655–1660.
- 68 M. P. Eglhoff, F. Marguet, G. Buono, R. Verger, C. Cambillau, H. van Tilbeurgh, *Biochemistry* **1995**, *34*, 2751–2762.
- 69 H. A. Berman, D. F. Olshefski, M. Gilbert, M. M. Decker, *J. Biol. Chem.* **1984**, *260*, 3462–3468.
- 70 A. M. Brzozowski, U. Derewenda, Z. S. Derewenda, G. G. Dodson, D. M. Lawson, J. P. Turkenburg, F. Bjorkling, B. Høge-Jensen, S. A. Patkar, L. Thim, *Nature* **1991**, *351*, 491–494.
- 71 D. Rotticci, T. Norin, K. Hult, M. Martinelle, *Biochim. Biophys. Acta* **2000**, *1483*, 132–140.
- 72 P. Stadler, G. Zandonella, L. Haalck, F. Spener, A. Hermetter, F. Paltauf, *Biochim. Biophys. Acta Lipids Lipid Metab.* **1996**, *1304*, 229–244.
- 73 H. Scheib, J. Pleiss, P. Stadler, A. Kovac, A. P. Potthoff, L. Haalck, F. Spener, F. Paltauf, R. D. Schmid, *Protein Eng.* **1998**, *11*, 675–682.
- 74 C. M. Carvalho, M. R. Aires-Barros, J. M. Cabral, *Biotechnol. Bioeng.* **1999**, *66*, 17–34.
- 75 F. Zoicher, N. Krebsfanger, O. J. Yoo, U. T. Bornscheuer, *J. Mol. Catal. B: Enzym.* **1998**, *5*, 199–202.
- 76 H. Yang, E. Henke, U. T. Bornscheuer, *Tetrahedron Asymmetry* **1999**, *10*, 957–960.
- 77 R. T. Otto, H. Scheib, U. T. Bornscheuer, J. Pleiss, C. Syltatk, R. D. Schmid, *J. Mol. Catal. B: Enzym.* **2000**, *8*, 201–211.
- 78 A. Musidowska-Persson, U. T. Bornscheuer, *Tetrahedron Asymmetry* **2003**, *14*, 1341–1344.
- 79 A. Musidowska-Persson, U. T. Bornscheuer, *J. Mol. Catal. B: Enzym.* **2002**, *19–20*, 129–133.
- 80 S. Hari Krishna, M. Persson, U. T. Bornscheuer, *Tetrahedron Asymmetry* **2002**, *13*, 2693–2696.

- 81 U.T. Bornscheuer, M. Pohl, *Curr. Opin. Chem. Biol.* **2001**, *5*, 137–143.
- 82 U.T. Bornscheuer, *Curr. Opin. Chem. Biol.* **2002**, *13*, 543–547.
- 83 U.T. Bornscheuer, C. Bessler, R. Srinivas, S.H. Krishna, *Trends Biotechnol.* **2002**, *20*, 433–437.
- 84 M. Baumann, B.H. Hauer, U.T. Bornscheuer, *Tetrahedron Asymmetry* **2000**, *11*, 4781–4790.
- 85 R.D. Schmid, R. Verger, *Angew. Chem. Weinheim Bergstr. Ger.* **1998**, *110*, 1694–1720.
- 86 H.A. Berman, P. Taylor, *Biochemistry* **1978**, *17*, 1704–1713.
- 87 H.A. Berman, D.F. Olshefski, M. Gilbert, M.M. Decker, *J. Biol. Chem.* **1984**, *260*, 3462–3468.
- 88 H. Scholze, H. Stutz, F. Paltauf, A. Hermetter, *Anal. Biochem.* **1999**, *276*, 72–80.
- 89 O.V. Oskolkova, A. Hermetter, *Biochim. Biophys. Acta* **2002**, *1597*, 60–66.
- 90 O.V. Oskolkova, R. Saf, E. Zenzmaier, A. Hermetter, *Chem. Phys. Lipids* **2003**, *125*, 103–114.
- 91 R. Birner-Grunberger, H. Scholze, K. Faber, A. Hermetter, *Biotechnol. Bioeng.* **2004**, *85*, 147–154.
- 92 G. Zandonella, P. Stadler, L. Haalck, F. Spener, F. Paltauf, A. Hermetter, *Eur. J. Biochem.* **1999**, *262*, 63–69.
- 93 A.E. Speers, B.F. Cravatt, *ChemBioChem* **2004**, *5*, 41–47.
- 94 B.F. Cravatt, E.J. Sorensen, *Curr. Opin. Chem. Biol.* **2000**, *4*, 663–668.
- 95 G.C. Adam, E.J. Sorensen, B.F. Cravatt, *Mol. Cell Proteomics* **2002**, *1*, 781–790.
- 96 D.A. Campbell, A.K. Szardenings, *Curr. Opin. Chem. Biol.* **2003**, *7*, 296–303.
- 97 M.P. Patricelli, D.K. Giang, L.M. Stamp, J.J. Burbaum, *Proteomics* **2001**, *1*, 1067–1071.
- 98 Y. Liu, M.P. Patricelli, B.F. Cravatt, *Proc. Natl Acad. Sci USA* **1999**, *96*, 14694–14699.
- 99 N. Jessani, Y. Liu, M. Humphrey, B.F. Cravatt, *Proc. Natl Acad. Sci USA* **2002**, *99*, 10335–10340.
- 100 D. Leung, C. Hardouin, D.L. Boger, B.F. Cravatt, *Nat. Biotechnol.* **2003**, *21*, 687–691.
- 101 D.L. Boger, H. Sato, A.E. Lerner, B.J. Austin, J.E. Patterson, M.P. Patricelli, B.F. Cravatt, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 265–270.
- 102 D. Kidd, Y. Liu, B.F. Cravatt, *Biochemistry* **2003**, *40*, 4005–4015.
- 103 G.C. Adam, B.F. Cravatt, E.J. Sorensen, *Chem. Biol.* **2001**, *8*, 81–95.

10 Fingerprinting Methods for Hydrolases

Johann Grognum and Jean-Louis Reymond

10.1 Introduction

Enzyme activity fingerprinting consists in measuring the activity of a single enzyme under a series of conditions, in particular a series of different substrates (Figure 10.1) [1]. The data set resulting from this high-content assay can be defined as a fingerprint if the measurement is easily repeatable and reproducible, which calls for practical and simple procedures. Activity fingerprinting is possible on the basis of any enzyme assay adaptable to a range of different substrates. All data points should be measured simultaneously using the same enzyme solution. An enzyme activity fingerprint may provide sufficient information to identify a specific enzyme by its activity and distinguish it from other

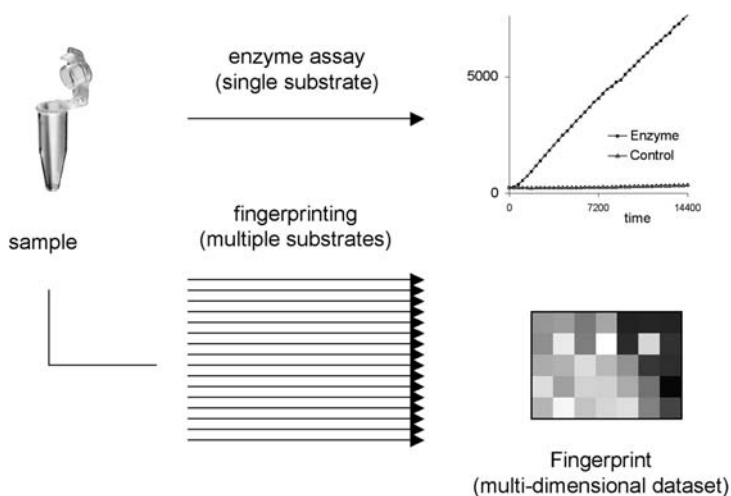


Fig. 10.1 Principles of enzyme fingerprinting.

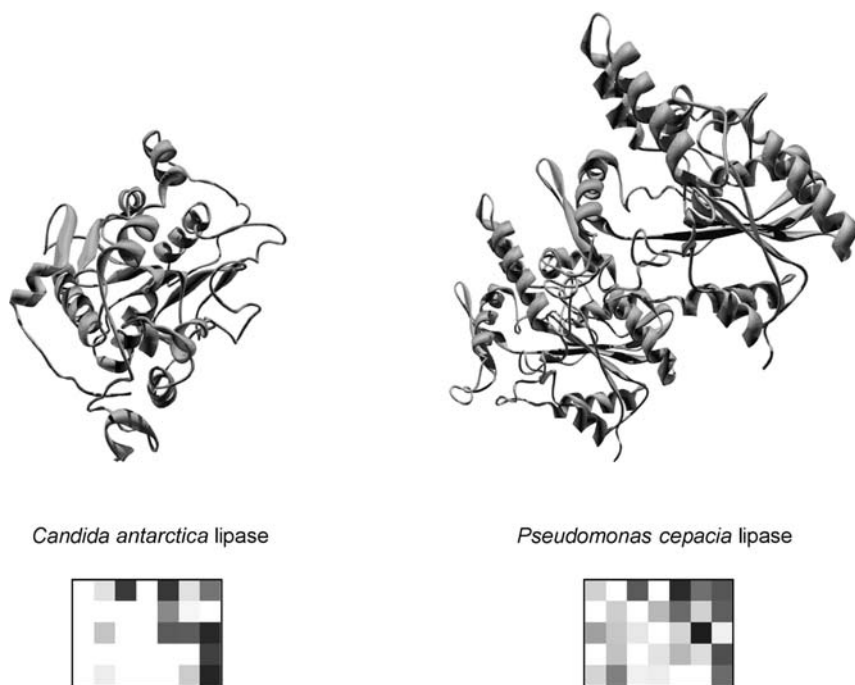


Fig. 10.2 Images of protein structure (upper) and activity fingerprints (lower) for two microbial lipases. The 3D structures were generated from Protein Database (PDB) data. The activity fingerprints are

taken with the polyol acetate substrate array in Figure 10.14. Both pairs of images are different. Note that the functional information in the fingerprint cannot be inferred from the protein structure.

closely related enzymes. This functional identification might serve as a tool for diagnostic and quality control of enzyme-containing samples [2].

The data points in an enzyme fingerprint can be used to compose an image by associating each data point with a pixel. The resolution of this image depends on the number of pixels and on their accuracy, and its meaning is given by the association of each pixel with a particular enzyme substrate. The fingerprint image conveys information about the enzyme's reactivity, much as the set of atom coordinates or their representation conveys information about an enzyme's structure (Figure 10.2). Most importantly, one cannot obtain one from the other because our current theoretical understanding of enzyme structure–activity relationships has only very limited predictive value. Indeed predicting the detailed effect of a structural modification in either enzyme (mutations) or substrate (chemical analog) on the reactivity of the enzyme–substrate pair is almost impossible. The complexity of the problem is illustrated by the fact that a single mutation in an enzyme may have only a barely detectable influence on its structure, but might change its reactivity and selectivity completely. Similarly, minor structural changes in a substrate can completely change its reactivity towards a given enzyme.

Enzyme assays based on microarrays such as those described in Chapter 12 are large-scale profiling experiments similar to fingerprinting. For example peptide arrays prepared by SPOT synthesis are used to profile the activity of proteases, kinases, and phosphatases, which serves to define their substrate specificity and identify their biological function by defining the optimal substrate [3]. These experiments use arrays of thousands of substrates derived from combinatorial or automated parallel synthesis. Positional scanning libraries form the basis of the analysis of protease substrate specificity, as discussed in detail in Chapter 11. However such experiments are often not easy to reproduce due to the complexity of the reagents and size of the data sets produced.

The present chapter reviews fingerprinting methods based on small sets of enzyme substrates, where emphasis is placed on reproducibility and simplicity of the measurements. The aim is to describe a photographic method of taking pictures of reactivity landscapes of enzymes, and using these pictures to analyze function. The principle of enzyme fingerprinting derives from microorganism activity fingerprinting, which is used to as a taxonomic tool in microbiology and in medical diagnostic applications [4]. Enzyme activity fingerprints may be used in the area of quality control and medical diagnostics, where rapid functional identification of enzymes may be needed as an indicator. Enzyme fingerprinting also implicitly addresses the fundamental question of catalytic promiscuity: provided that a pair of enzymes displays identical fingerprints across a series of substrates, are these two enzymes actually identical from the functional point of view, to the degree that their entire reactivity profile against any substrate will be identical?

10.1.1

One Enzyme – One Substrate

In the traditional view of biochemistry, each enzyme is associated with catalysis on a specific substrate, and this holds particularly true for enzymes of metabolic cycles which process single intermediates with high specificity. A large number of enzyme assays are enzyme-coupled assays, whereby the reaction product of the enzyme to be detected is subsequently converted by secondary enzymes, which are part of the detection reagent, to ultimately produce a detectable signal. Most enzyme-coupled assays relay product processing to a redox enzyme interconverting NAD^+ and NADH or NADP^+ and NADPH . The latter reaction can be followed spectrophotometrically at 340 nm, a wavelength compatible with almost any type of container, including polystyrene microtiter plates. An example of an enzyme-coupled assay relevant in the context of enzyme fingerprinting is the enzyme-coupled detection system for acetic acid, in which a series of three highly specific enzymes are used (Figure 10.3). First, acetyl-CoA synthase (ACS) activates acetic acid to acetyl-CoA with ATP. Second, acetyl-CoA is condensed with oxaloacetate to form citrate by citrate synthase (CS). Finally, oxaloacetate is produced from L-malate by L-malate dehydrogenase (L-MDH). Each of these three enzymes is highly specific for its substrate. This assay system for acetic acid was used by

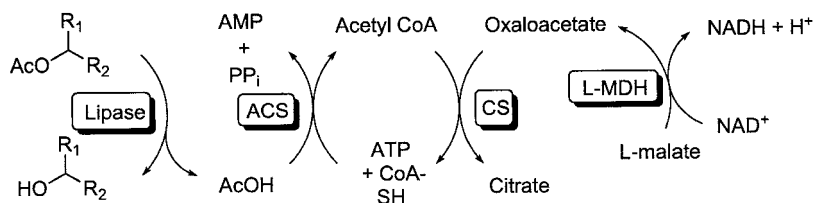


Fig. 10.3 Enzyme-coupled detection of acetic acid via NAD/NADH at 340 nm allows lipase activities to be monitored. The lipase may accept a variety of substrates, while the detection enzymes ACS (acetyl-CoA synthase), CS (citrate synthase) and L-MDH (L-malate dehydrogenase) are very specific for their substrates.

Bornscheuer et al. to follow lipases and esterases activities on various substrates [5], in particular enantiomeric pairs, as also discussed in Chapter 2, and could form the basis for fingerprinting experiments for lipases.

Reference assays associating a single substrate with an enzyme serve to define activity units. Unit measurements are carried out under defined conditions of medium, temperature, and substrate concentration. One unit generally corresponds to $1 \mu\text{mol}$ of product formed per minute under specified assay conditions. It is recommended to use M s^{-1} under standard conditions [6]. Activity units are used to compare different enzyme samples, for example during isolation and purification. All commercial enzymes are sold together with an indication of activity in units per milligram protein, whereby the reference assay substrate and conditions are described. The activity units in many lipases and esterases are defined with reference to the hydrolysis of chromogenic nitrophenyl esters, and those of proteases by reference to the hydrolysis of amino acid nitroanilides, which are also chromogenic. Alcohol dehydrogenases are measured by spectrophotometric recording of the reduction of the cofactor NAD(P)^+ to NAD(P)H in the presence of a reference alcohol substrate, for example ethanol for horse liver alcohol dehydrogenase. Glycosidases units refer to the hydrolysis of chromogenic nitrophenyl glycosides as substrates, for example β -D-nitrophenyl-galactoside for β -galactosidase.

For many enzymes the reference assay defining units is performed under this enzyme's optimal conditions and sometimes using the only accepted substrate. In this situation the unit determination provides a fair estimate of the enzyme's catalytic power in an absolute sense. However the application of enzymes to chemical synthesis has shown that many enzymes accept a wide variety of substrates and may operate under very different conditions (pH, solvent, temperature), each of which corresponds to a certain activity and selectivity pattern of the enzyme. For such enzymes, which include many hydrolytic enzymes, activity units may not reflect the enzyme's performance. This is the case, for example, for lipases as discussed above, because their reference nitrophenyl ester substrates are often quite poor substrates compared with their natural triglyceride substrates.

10.1.2

Enzyme Activity Profiles

Enzymes are intrinsically complex catalysts, as evidenced by their folded three-dimensional structure. The smallest enzymes are in the range of 20 kDa, which corresponds to 150 amino acids. The polypeptide is always much larger than the catalytic site itself and imparts essential functions in addition to catalysis, including stability, allosteric and biochemical regulation, and cellular localization. Primary sequences often allow a gene product to be assigned to a given enzyme class due to the occurrence of conserved catalytic site arrangements in certain enzymes. For example lipases almost always display a catalytic triad involving a nucleophilic serine within a G-X-S-X-G conserved motif often located in a buried type II' turn [7]. Three-dimensional structures may also allow a partial functional assignment classification. For example, the occurrence of an eight-stranded β -barrel (TIM barrel) is often indicative of enzymatic activities. However structural data do not provide detailed information on an enzyme's catalytic behavior, which must be measured directly.

The description of an enzyme's catalytic function from the chemical standpoint includes activity changes in response to changing reaction conditions (temperature, pH, medium) and substrate structure (chemo- and stereospecificity). These data can be defined as the activity profile of the enzyme, which thus consists in a series of multidimensional data sets. The simplest form of activity profile consists of parametric profiles using a single substrate, which are used to extract mechanistic information. The substrate/concentration profile can lead to the so-called Michaelis–Menten parameters of K_m and k_{cat} following a simple kinetic model [8]. Most enzymes also display a specific pH/rate profile which provides information about the catalytic mechanism, for example the role of aspartate and glutamate residues in acidic xylanases [9]. pH/rate profiles often indicate which reaction step is rate-limiting, for example in protein tyrosine phosphatases [10]. Temperature profiles can be interpreted in terms of enzyme denaturation (melting) [11], temperature-dependent activity [12], or using the Arrhenius equation to determine activation parameters. Arrhenius plots for enzymic rates may be nonlinear and their interpretation quite problematic [13], in particular for membrane-bound enzymes [14], this due in part to the temperature dependence of substrate affinity [15].

The most significant activity profiles concern variations in substrate structure, and this will be the main topic of this chapter. Substrate profiles, if recorded in simple and reproducible manner, may provide entirely specific and reproducible patterns that can be used to identify an enzyme or enzyme containing sample, the functional equivalent of a fingerprint. Substrate profiling allows the determination of the substrate preference of an enzyme, which is of practical interest for biocatalysis applications. In the area of biochemistry, the determination of preferred substrates helps to elucidate the cellular function of enzymes, in particular in the case of proteases, kinases, and phosphatases through the identification of the target peptide sequences of each enzyme as discussed in Chapters 11 and 12.

10.1.3

The APIZYM System for Microbial Strain Identification

The idea of using a substrate profile as a fingerprint for sample identification dates back to the 1960s, with the invention of the APIZYM system by Bussi ere et al. as a tool for microorganism classification and identification. The system, first named Auxotab, involved an array of 16 different enzyme substrates [16], and included chromogenic substrates for lipases and esterases, aminopeptidases, chymotrypsin, trypsin, phosphatases, sulfatases, and β -galactosidases. The assays were formulated as a combined set on a filter-paper format, and carried out in parallel on crude microorganism cultures to assess the presence/absence of the corresponding enzymes. The APIZYM system was then developed from this discovery by Daniel Monget at Biomerieux and commercialized [17], and became a popular tool in microbiology. This system uses a series of 19 enzyme assays characteristic of 19 enzyme activities and one blank as reference. It has been used broadly to

Table 10.1 Enzymes and substrates assayed by the parallel assay APIZYM system.

No.	Enzyme	Substrate	pH	Color ^{a)}
1	None	None		b)
2	Alkaline phosphatase	2-Naphthyl phosphate	8.5	purple
3	Esterase C4	2-Naphthyl butyrate	7.1	purple
4	Lipase C8	2-Naphthyl caprylate	7.1	purple
5	Lipase C14	2-Naphthyl myristate	7.1	purple
6	Leucine aminopeptidase	L-Leucyl 2-naphthylamide	7.5	orange
7	Valine aminopeptidase	L-Valyl-2-naphthylamide	7.5	orange
8	Cystine aminopeptidase	L-Cystyl-2-naphthylamide	7.5	orange
9	Trypsin	N-Benzoyl-DL-arginine 2-naphthylamide	8.5	orange
10	Chymotrypsin	N-Benzoyl-DL-phenylalanine 2-naphthylamide	7.1	purple
11	Acid phosphatase	2-Naphthyl phosphate	5.4	purple
12	Phosphoamidase	Naphthol AS bis-phosphodiamide	5.4	blue
13	α -Galactosidase	6-Bromo-2-naphthyl- α -D-galactopyranoside	5.4	purple
14	β -Galactosidase	2-Naphthyl- β -D-galactopyranoside	5.4	purple
15	β -Glucuronidase	Naphthol AS bis- β -D-glucuronide	5.4	blue
16	α -Glucosidase	2-Naphthyl- α -D-glucopyranoside	5.4	purple
17	β -Glucosidase	6-Bromo-2-naphthyl- β -D-glucopyranoside	5.4	purple
18	N-Acetyl- β -glucosaminidase	1-Naphthyl-N-acetyl- β -D-glucosaminide	5.4	brown
19	α -Mannosidase	6-Bromo-2-naphthyl- α -D-mannopyranoside	5.4	purple
20	α -Fucosidase	2-Naphthyl- α -L-fucoside	5.4	purple

A fibrous material (filter paper) is impregnated with substrate added as an alcohol solution, then with a pH stabilizer (Tris-HCl > pH 7.0, or Tris-maleate), and then put in contact for 2–4 h at 37°C with the biological sample to be analyzed.

- a) Coloration observed after reaction with a solution of Fast Blue BB (N-(4-amino-2,5-diethoxyphenyl)-benzamide) in 25% aqueous Tris-HCl containing 10% weight of lauryl sulfate.
- b) Control with biological sample only.

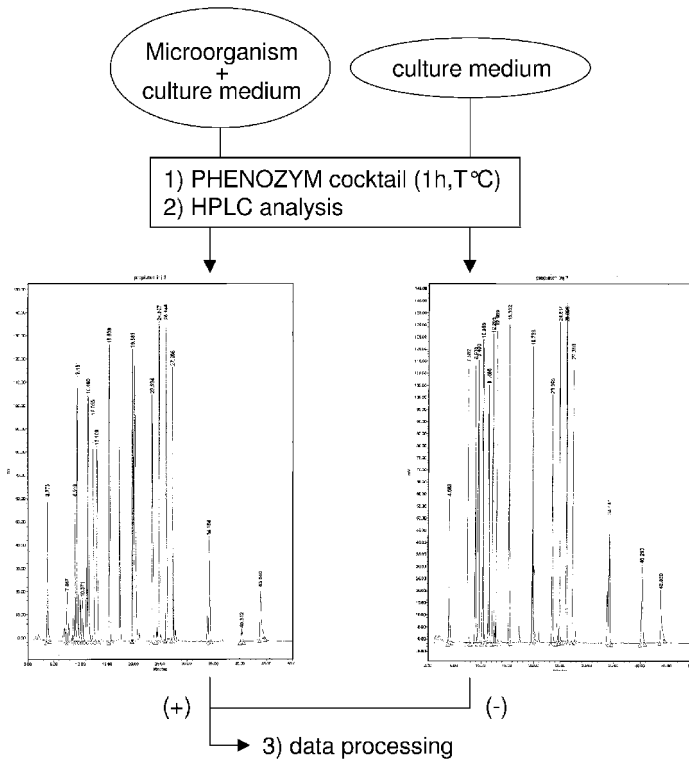


Fig. 10.4 Principle of PHENOZYM microorganism profiling. The cocktail contains 16 labeled enzyme substrates (see Table 10.2). Data processing involves determination of percentage conversion of each substrate in the positive (culture) and negative (medium alone) assay.

characterize microorganisms (Table 10.1) [18]. The apparent activities are classified qualitatively from enzymic tests using labels such as α for a strong activity and β for a weak activity, and interpreted in terms of the presence/absence of the corresponding enzymes [19]. Newer, more exhaustive variations on this theme have been developed. The current format includes 32 different assays and comes in miniaturized, preformed plates [20]. The APIZYM system demonstrates that enzyme activity data allow a functional classification of microorganisms.

A more compact version of the APIZYM can be realized in a single assay using a cocktail containing different enzyme substrates, called PHENOZYM (Figure 10.4 and Table 10.2) [21]. This format is applicable to extremophilic microorganisms under their optimal growth conditions. Labeled substrates for 16 different enzyme types are combined in this cocktail. The assay involves a single reaction followed by determination of substrate consumption by high-performance liquid chromatography (HPLC) analysis. This allows a rapid identification of multiple enzyme activities, and is compatible with a diversity of growth media and reaction conditions (pH, temperature), including those for thermo-

Table 10.2 The PHENOZYM cocktail for profiling thermophilic microorganisms.

No.	Enzyme	Substrate	t_R ^{a)}
1	Phosphatase	4-NP-phosphate ^{b)}	3.94
2	Amylase	4-NP- α -D-hexa-(1,4)-glucopyranoside ^{c)}	7.49
3	β -Galactosidase	4-NP- β -D-galactopyranoside	8.98
4	α -Galactosidase	4-NP- α -D-galactopyranoside	9.50
5	β -Glucosidase	4-NP- β -D-glucopyranoside	10.41
6	β -Glucuronidase	4-NP- β -D-glucuronide	11.53
7	<i>N</i> -Acetyl- β -glucosaminidase	4-NP- <i>N</i> -acetyl- β -D-glucosaminide	12.39
8	α -Glucosidase	4-MU- α -D-glucopyranoside ^{b)}	13.13
		4-Methylumbelliferone	17.38
9	α -Mannosidase	4-MU- α -D-mannopyranoside	15.49
10	α -Fucosidase	4-NP- α -L-fucopyranoside	19.83
		4-Nitrophenol	20.12
		4-Nitroaniline	20.35
11	Valine aminopeptidase	L-Valine- <i>para</i> -nitroanilide	23.44
12	Leucine aminopeptidase	L-Leucine- <i>para</i> -nitroanilide	24.83
13	Chymotrypsin	L-Phenylalanine-4-nitroanilide	26.12
14	Trypsin	<i>N</i> - α -Benzoyl- α -L-arginine-4-nitroanilide	27.33
15	Esterase C4	5-(4-Nitrophenoxy)-2-hydroxy-pentyl butanoate ^{d)}	40.23
16	Lipase C8	3-(Umbelliferyl)-2-methyl-2-hydroxypropyl octanoate ^{d)}	43.85

a) Analysis conditions: Vydac 218TP54 RP-C18 column, 0.4×22 cm, elution 1.5 mL min^{-1} , gradient water-acetonitrile + 0.1% TFA, detection by UV at 300 nm.

b) NP: nitrophenyl; MV: methylumbelliferyl.

c) This substrate gave mostly nitrophenol. Intermediates corresponding to hydrolysis between the glucosyl units were also detected in small amounts.

d) The 1,2-diol hydrolysis products were not detected and are apparently further degraded under the assay conditions.

philic microorganisms. The functional profiles can be used for a functional classification of the different microbial strains, resulting in a “phylo-enzymatic” tree consistent with the known phylogenetic classification of the strains.

10.2

Hydrolase Fingerprinting

The APIZYM and PHENOZYM analyses identify a microorganism on the basis of the enzyme panel it expresses. The principle of enzyme fingerprinting also extends to identifying single enzymes on the basis of their activity profiles against a series of substrates. Substrates for fingerprinting should be sufficiently reactive to return an activity with most enzymes within an enzyme class, yet show differential reactivity to distinguish between different enzymes. The activity detection should also be simple and reproducible to facilitate analysis.

10.2.1

Fingerprinting with Fluorogenic and Chromogenic Substrates

Fluorogenic and chromogenic substrates provide a very direct, selective, and flexible tool to design enzyme-specific assays. Optimally such substrates react only in contact with the targeted enzyme or enzyme class, while being unreactive otherwise. The indirect release of a fluorescent or colored phenolate (X^-) by β -elimination from a ketone or aldehyde intermediate provides a general strategy for fluorogenic and chromogenic substrates for a broad variety of enzymes (Figure 10.5). This principle is suitable for the assay of alcohol dehydrogenases (1) [22], and to measure retro-aldolization in catalytic antibodies (e.g. 2) [23] and transaldolases (e.g. 3) [24]. Similar catalytic antibody assays have been reported, releasing either resorufin from 4 [25], or bromonaphthol from 5 [26]. In the latter case reaction of the released bromonaphthol with brilliant blue forms an insoluble blue precipitate suitable for direct detection in agar plates.

Assays using a similar β -elimination strategy to release a fluorescent or colored phenolate have been reported for lipases (6) [27], transketolase (7) [28], Baeyer-Villiger monooxygenases (8) [29], β -lactamase (9) [30], and a fluorescent probe for NADPH and NADH (10) [31]. Indirect release of umbelliferone (7-hy-

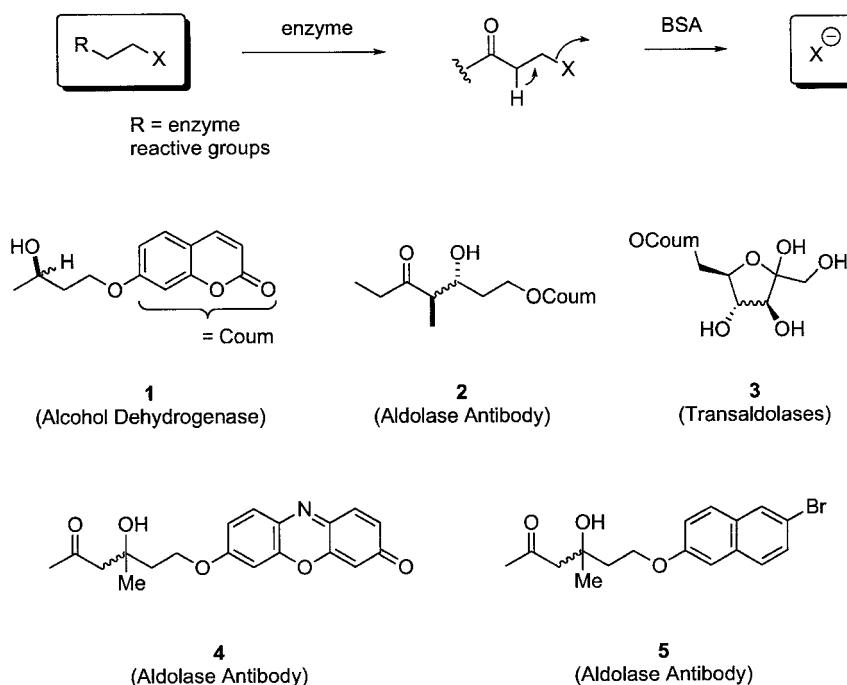


Fig. 10.5 Fluorogenic and chromogenic substrates using secondary β -elimination. The bonds cleaved by the primary enzyme being assayed are marked as wobble lines.

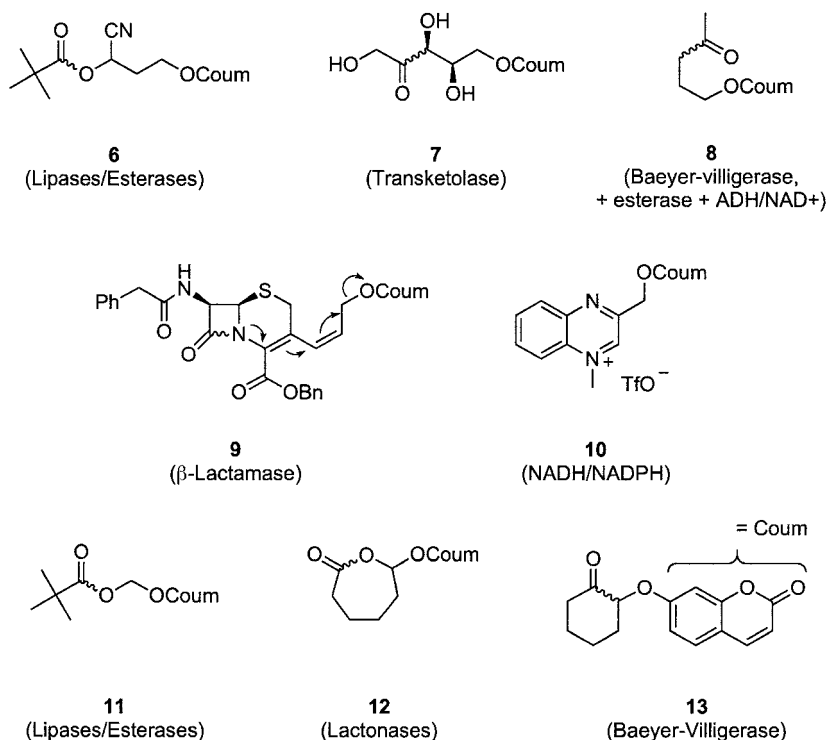


Fig. 10.6 Indirect release fluorogenic substrates. The bonds cleaved by the primary enzyme being assayed are marked as wobble lines.

droxycoumarin) is also possible via an intermediate hemi-acetal (Figure 10.6), which leads to assays for lipases and esterases (e.g. **11**) [32], lactonases (e.g. **12**) and Baeyer-Villiger monooxygenases (e.g. **13**) [33].

The β -elimination can also be coupled with periodate oxidation of an intermediate 1,2-diol or 1,2-aminoalcohol, thereby allowing the assay of enzymes releasing these as primary products (Figure 10.7) [34]. The strategy is suitable for lipases and esterases (e.g. **14–15**) [35], epoxide hydrolases (e.g. **16**) [36], phosphatases (e.g. **17**) [37], and acylases and proteases (e.g. **18**) [38].

In a similar manner release of the fluorescent 6-methoxy-naphthaldehyde (**20**) from alcohol precursors can be used in a blue fluorescent assay for alcohol dehydrogenase (**19**) [39], aldolase antibodies and proline (**21**) [25], or via the periodate cleavage for esterases or lipases (**22**), and epoxide hydrolases (**23**) (Figure 10.8) [1].

Such substrates do not show any significant background reaction in the absence of specific enzymes, and can be prepared in many variations by simple synthetic steps. A series of mono-acetates, diacetates, cyclic carbonates, and epoxides derived from optically pure versions of fluorogenic and chromogenic 1,2-diols related to **14–16** were prepared and used for fingerprinting hydrolases [1]. The fingerprinting experiment involved incubation of a single enzyme with all the differ-

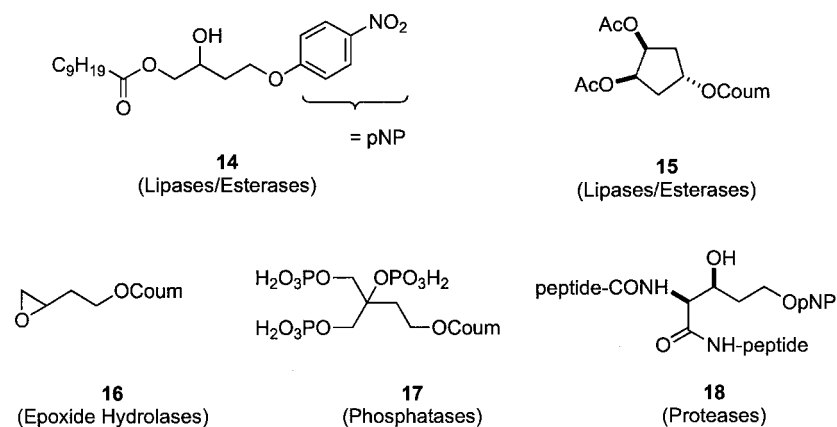
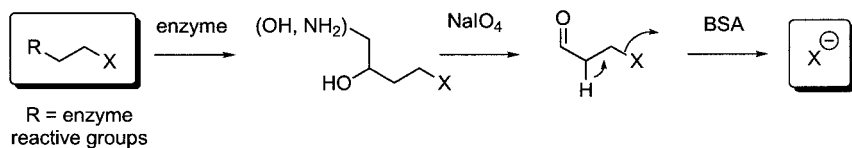


Fig. 10.7 Fluorogenic and chromogenic substrates for periodate-coupled enzyme assays with β -elimination.

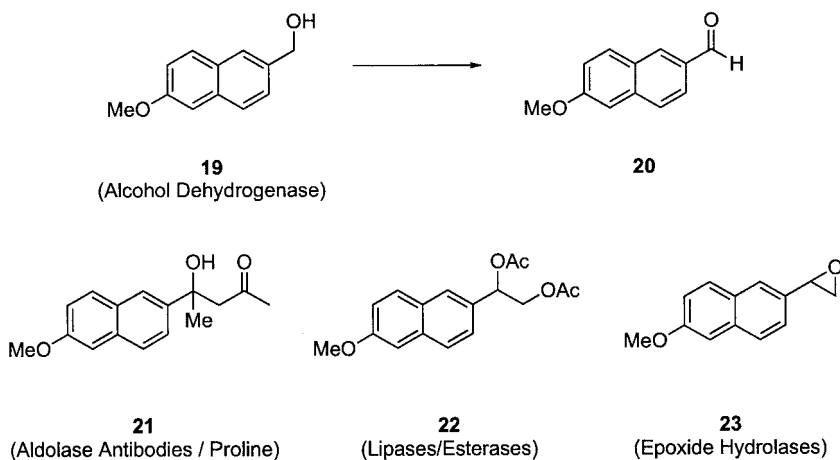


Fig. 10.8 Naphthaldehyde release assays.

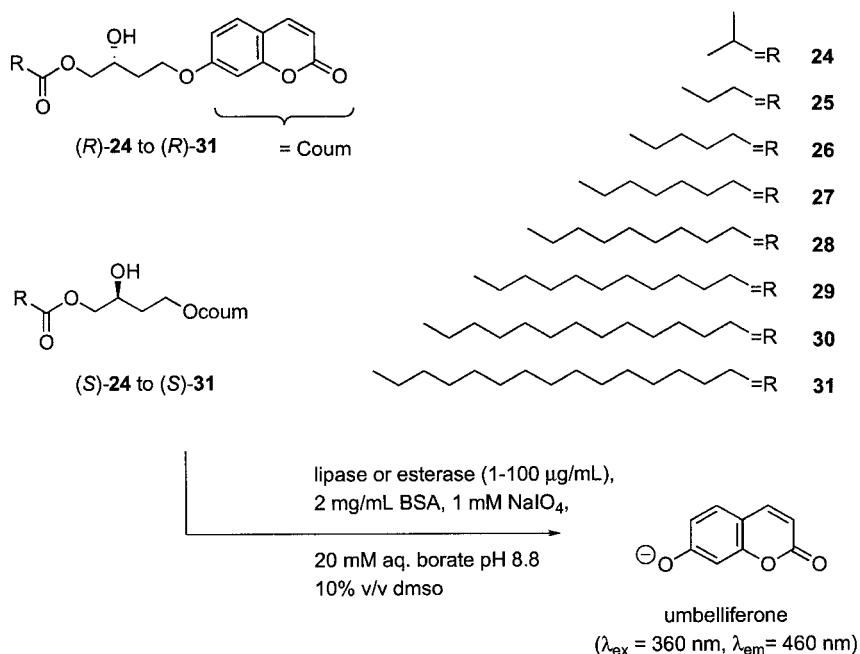


Fig. 10.9 Lipase fingerprinting using a series of fluorogenic monoacylglycerol equivalents with different acyl chain length. The data from this array are shown in Figure 10.19.

ent substrates in parallel in the same microtiter plate, and following the evolution of fluorescence or color simultaneously for all substrates. This fingerprinting procedure for the measurement is repeated at will for each enzyme, and provides reproducible activity profiles across the entire array of substrates. The key advantages of this array are the negligible background reaction without enzyme, the common chromophore/fluorophore release chemistry and the simple signal acquisition using a microtiter plate reader. The array of fluorogenic monoacylglycerol analogs shown in Figure 10.9 was used to analyze the acyl chain length selectivity of lipases and esterases, as discussed below [40].

Esters of nitrophenol and umbelliferone are often used as reference substrate to define activity units of esterases and lipases. These substrates usually show a high level of nonspecific hydrolysis in the absence of enzyme, and therefore cannot be used for fingerprinting. We have recently found that the problem of nonspecific hydrolysis can be circumvented in the case of umbelliferyl esters by using the substrates in a solid-supported format [41]. Umbelliferyl esters (32–47) and oxymethyl ethers (48–51) (Figure 10.10) can be adsorbed onto silica gel plates from dichloromethane solutions to form a homogeneously impregnated layer. The silica gel plates thus prepared are then treated with the enzyme-containing solution in buffer (Figure 10.11). Under these conditions only active enzymes induce a fluorogenic hydrolysis reaction, while buffer and noncatalytic

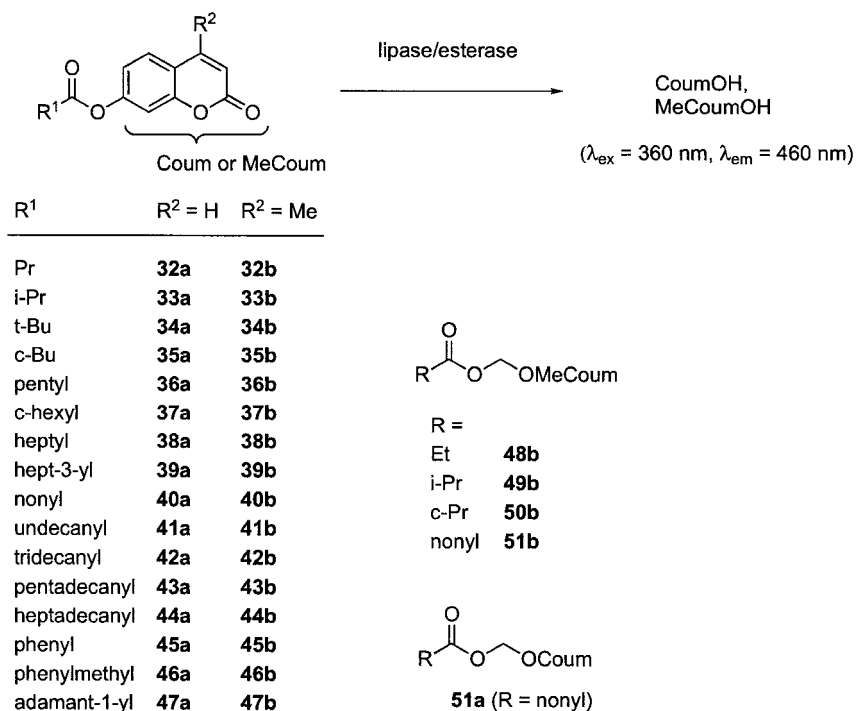


Fig. 10.10 Umbelliferol esters used for fingerprinting on solid support.

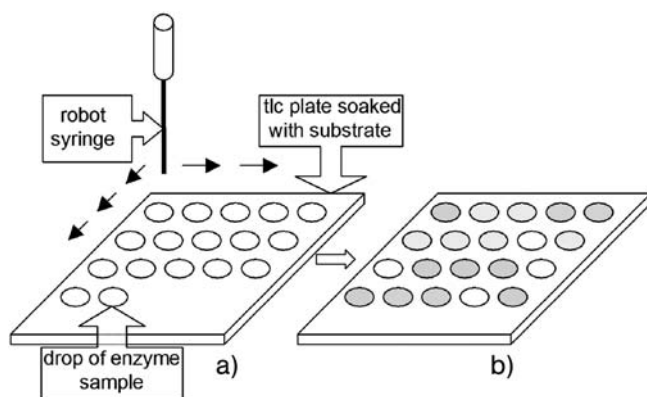


Fig. 10.11 Principle of solid-supported assay.

proteins have no effect. This solid-supported format is advantageous for fingerprinting because only small volumes of enzyme solution are required (1 μ L per assay), and simple esters of umbelliferone are prepared in a single synthetic step, or are directly available commercially.

Procedure 10.1: Esterase/lipase Fingerprinting on Impregnated Silicagel Plates

General procedure for the synthesis of esters 32a/b–47a/b: A solution of umbelliferone or 4-methyl-umbelliferone (1.23 mmol) in anhydrous tetrahydrofuran (THF) (3 mL) was treated with NaH (118 mg, 55% suspension in oil, 2.3 equiv.). After 30 min at 25 °C, the reaction was cooled to 0 °C and the acyl chloride (1.85 mmol, 1.5 equiv.) was added as a solution in dry THF (1 mL). After 2 h at 25 °C, the reaction mixture was poured into aqueous 1 M HCl (50 mL) and extracted with CH₂Cl₂ (2×50 mL). The organic phase was dried over Na₂SO₄, evaporated and the residue was purified by flash chromatography to give the pure esters.

Fingerprinting: Silica gel thin-layer chromatography (TLC) glass-plates sil G-25 (Macherey-Nagel, Dueren, Germany) (silica gel 40–63 μm, surface 550 m² g⁻¹, layer 0.25 mm) were soaked in a solution of substrate (**32a/b–51a/b**, 2 mM in CH₂Cl₂) for 2 min and dried. Aliquots (1 μL) of enzyme samples in phosphate-buffered saline (PBS, 10 mM phosphate, 160 mM NaCl, pH 7.4, with 30% v/v glycerol) were spotted at 25 °C. Product formation was recorded after 2–24 h using a fluorescence microtiter plate reader ($\lambda_{\text{ex}}=360$ nm, $\lambda_{\text{em}}=460$ nm).

10.2.2

Fingerprinting with Indirect Chromogenic Assays

The use of a chromogenic or fluorogenic signal allows multiple quantitative analyses in microtiter plates. However the use of tagged substrates limits the range of structures available to compose the substrate array for fingerprinting. In addition, many chromogenic and fluorogenic substrates are not available commercially, so that the composition of a substrate array for fingerprinting requires a large synthetic effort. A practical alternative consists in using an indirect chromogenic or fluorogenic assay that returns a signal upon reaction progress with an untagged, commercially available substrate. Multisubstrate profiling of lipases and esterases has been realized using a colorimetric assay based on pH indicators, as described by Kazlauskas et al. [42]. Hydrolysis of an ester substrate releases a carboxylic acid, which results in a drop of pH in the assay solution. This pH change can be detected colorimetrically with nitrophenolate as indicator (yellow to colorless upon decreasing pH). In the so-called Quick *E* variation described in Chapter 1 [43], the reaction rates of separate enantiomers are determined by the pH method in the presence of a competing resorufin ester substrate. Activity profiling with this method across an array of substrates was used to describe the substrate range of four esterases.

Indirect fingerprinting of lipases and esterases has also been demonstrated with an assay using the principle of back titration with adrenaline (**52**) (Figure 10.12) [44]. In this assay a periodate-resistant substrate, for example a phosphate of carboxylic ester of a 1,2-diol or its epoxide precursor, is converted to a periodate-sen-

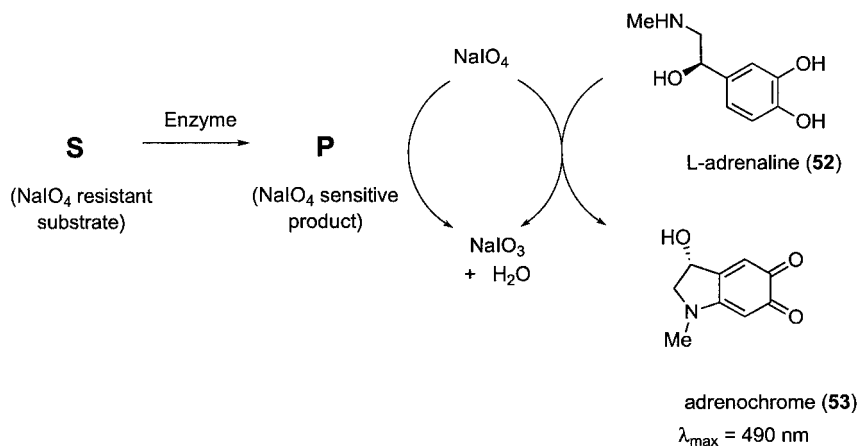


Fig. 10.12 Indirect chromogenic endpoint assay using back titration of periodate with adrenaline.

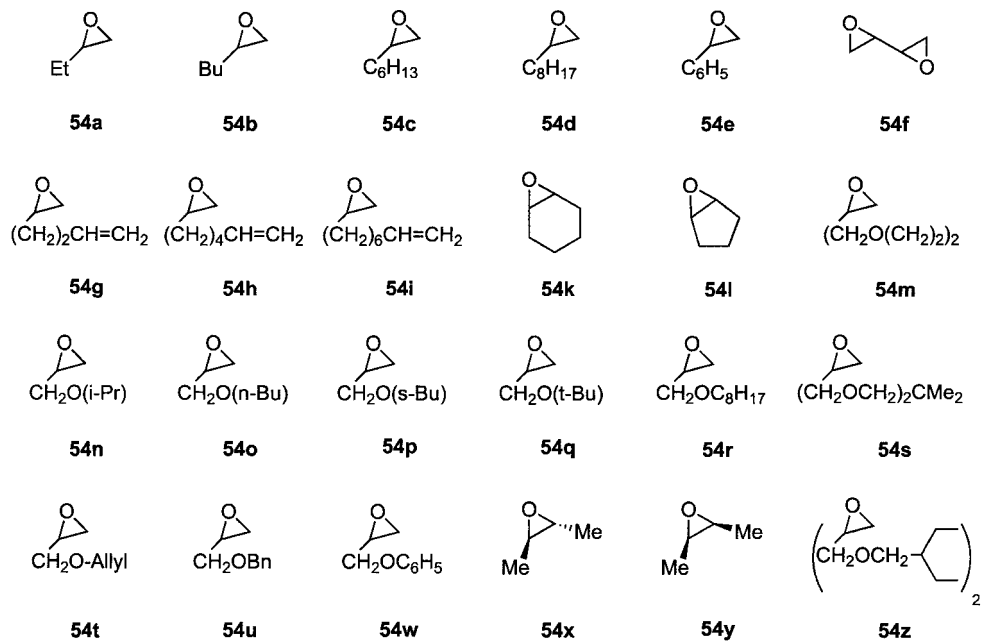


Fig. 10.13 Array of epoxides used for fingerprinting epoxide hydrolases using the assay described.

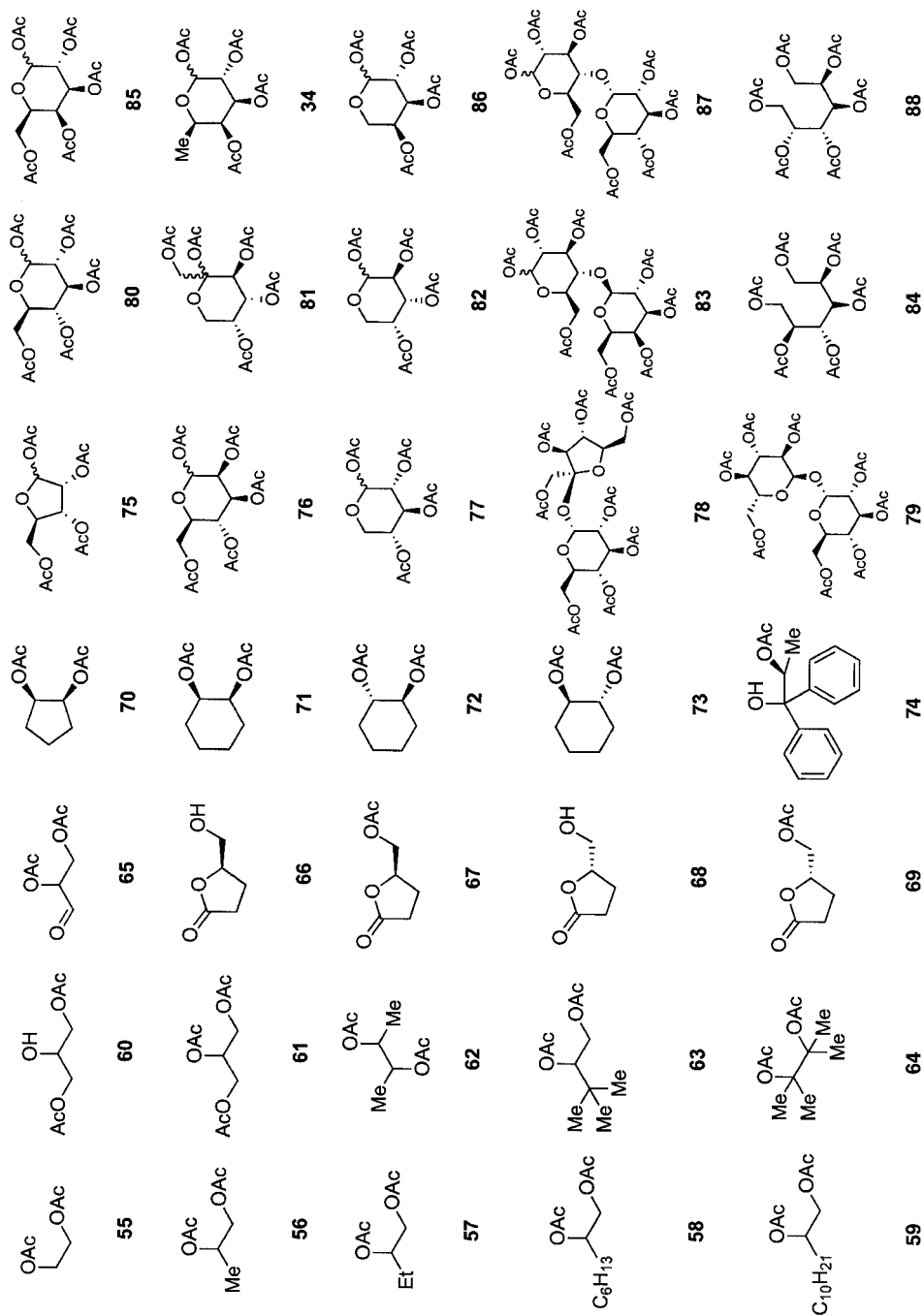


Fig. 10.14 Array of polyol acetates for lipase/esterase fingerprinting.

sitive reaction product, that is a 1,2-diol, upon enzymatic transformation, that is hydrolysis by an esterase or an epoxide hydrolase. The test solution is treated with a measured amount of sodium periodate, which reacts with the oxidizable functional groups in the products. The unreacted periodate reagent is then revealed by addition of adrenaline (52), which undergoes an instantaneous oxidation with periodate to give adrenochrome (53), a cationic orthoquinone dye with a red absorption maximum in the visible spectrum. This colorimetric assay provides off-the-shelf endpoint assays for lipases using vegetal oils as substrates, phytases using phytic acid as substrate, and epoxide hydrolases using epoxides as substrates. In the latter case, simply plating out a series of epoxides (54a–z) provides an array suitable for fingerprinting of epoxide hydrolase reactivity (Figure 10.13). The assay can be adapted for fingerprinting hydrolases using an array of acetates (55–88) derived from commercially available polyols such as carbohydrates and glycols (Figure 10.14) [45]. The resulting activity fingerprints were used to identify unusual reactivities in newly discovered microbial esterases. A related indirect assay based on chelation of copper by a fluorescent chelate or an amino acid is available for screening proteases and peptidases with their natural substrates [46], and could form the basis for similar fingerprinting arrays.

Procedure 10.2: Lipase/esterase Fingerprinting Using the Adrenaline Test for Enzymes

Substrates were diluted from 10 mM stock solutions in water/acetonitrile mixtures. Enzymes were diluted from 10 mg mL⁻¹ stock solutions in aqueous borate buffer (50 mM, pH 8.0). NaIO₄ was added as a freshly prepared 10 mM stock solution in water. Adrenaline (as HCl salt) was added as a 15 mM stock solution in water. Assays (0.1 mL) were conducted in individual wells of 96-well flat-bottom half-area polystyrene microtiter plates (Costar, Cambridge, MA, USA). Conditions: (1) Enzyme in 50 mM aqueous borate pH 8.0, 1 mM NaIO₄, 1 mM substrate, 60 min at 37 °C; (2) 1.5 mM adrenaline, 5 min, 26 °C. The optical density (OD) was recorded using a Spectramax 190 Microplate Spectrophotometer (Molecular Devices, Ismaning/München, Germany; $\lambda = 490$ nm). Commercial enzyme samples were tested at 1 mg mL⁻¹, proprietary esterases and lipases samples were used at one-tenth of crude enzyme extracts filtered through size-exclusion chromatography columns.

10.2.3

Cocktail Fingerprinting

Recording good fingerprinting data requires a simple yet reliable assay procedure. Optimally the reagent should be easy to assemble and use, and it should be possible to check its composition at any time. In addition, fingerprinting should be based on readily available instruments and adaptable to a broad range of reaction conditions, such as to be available to any laboratory working with en-

zymes. A remarkably straightforward solution to this problem is realized in the form of substrate cocktails, which are mixtures of enzyme substrates chosen such that either the substrates, or the reaction products, or both, are easily separable, selectively detectable, and precisely quantifiable after reaction.

The principle of cocktail fingerprinting was first demonstrated with a cocktail of 20 different lipase substrates tagged with a strong UV chromophore, and such that the reaction products are separated by HPLC analysis [47]. The cocktail composition can be checked at any time by HPLC or any spectroscopic analysis method, and the assay involves only a single measurement, which ensures that the relative product distribution is completely conserved for two measurements under identical conditions. The cocktail reagent was used to record the activity fingerprints of 40 different lipases and esterases. All substrates were mono-octanoyl esters of 1,2-diol, which ensured that most substrates showed significant reactivity with the enzymes. Thus, the cocktail ensured activity detection of all enzymes tested, while returning sufficient different reactivity information to uniquely identify each enzyme by a particular reactivity pattern.

Cocktail fingerprinting is also suitable for the analysis of protease reactivities [48]. In this case the challenge consists in reducing the enormous diversity of possible peptide sequences to a limited yet representative set of peptides spanning a sufficient range for ensuring broad and differentiated reactivity with various proteases while being separable by HPLC. A series of five hexapeptides including all 25 possible dipeptides between (1) acidic, (2) basic, (3) hydrophobic, (4) aromatic, and (5) small and polar amino acids was realized following a domino-game assembly (Figures 10.15 and 10.16). The peptides were fluorescence

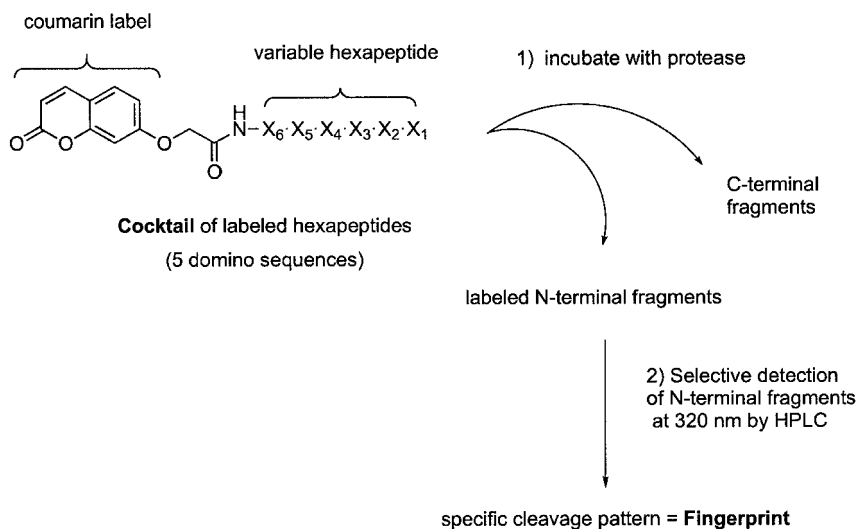


Fig. 10.15 Cocktail fingerprinting of protease with N-terminal labeled hexapeptide substrates. Sequences are selected by domino design.

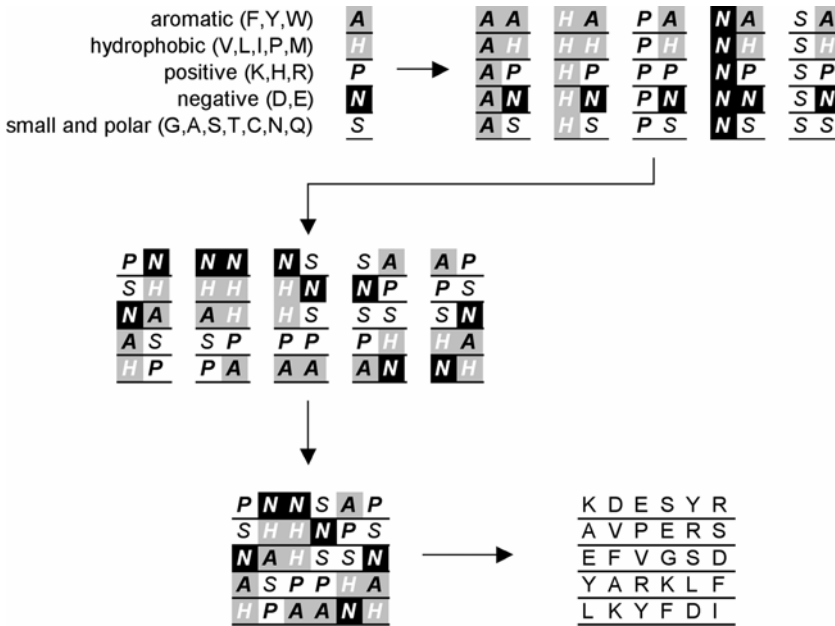


Fig. 10.16 Domino design of peptide cocktail. All 25 possible types of dipeptide (upper right) serve as domino pieces to assemble five hexapeptides. Only one possible solution of the domino game is shown and realized as an actual sequence

(lower right). Note that the domino solution selected also realizes 16 of 25 possible 1,3-arrangements of amino acid types. The missing 1,3-arrangements are AXA, HXA, HXH, PXP, PXS, NXP, NXN, SXA, SXS.

labeled at the N-terminus, resulting in 31 possible measurable fragments, including the five substrates and their N-terminal cleavage products. The resulting cocktail reagent reacts with a broad range of proteases, and returns a specific cleavage pattern in each case (Figure 10.18).

10.3 Classification from Fingerprinting Data

The fingerprints provided by the methods above return enzyme-specific or sample-specific activity patterns. The information that can be extracted from such patterns depends on data accuracy, the variability between different samples, and the actual structure of the substrates. In general the fingerprint data consist either of a series of initial reaction rates measured for each substrate in the presence of the enzyme, or of a series of percentages of conversion observed for each substrate after a fixed incubation time. The first step in processing fingerprinting data is to extract this reaction rate information from the primary signal recorded following appropriate calibration curves with reference products. Trans-

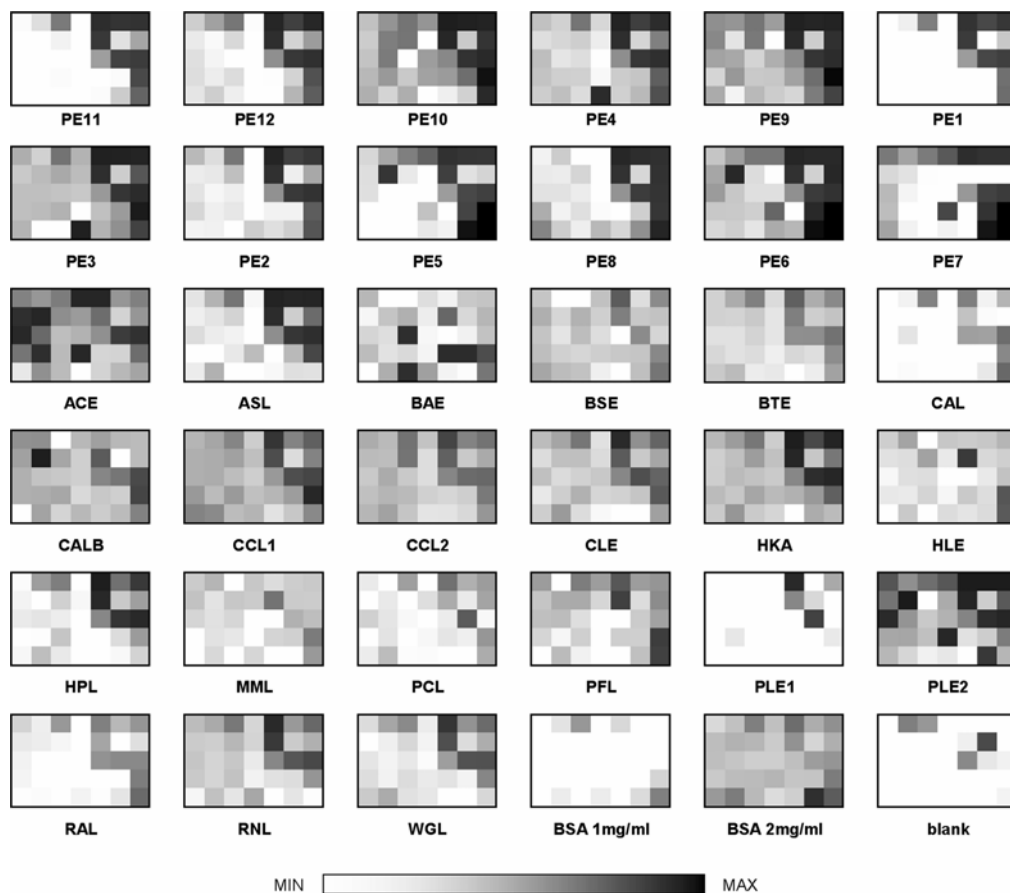


Fig. 10.17 Activity of esterases and lipases towards polyacetates in Figure 10.14 as determined by the adrenaline test. Each grayscale square corresponds to one ester substrate according to the layout, from white (0%, no activity) to black (100%, maximum signal), after correction from blank values.

formation of the recorded signal to a chemically significant value such as a reaction rate or percentage conversion is not a formal requirement if one is only interested in enzyme similarity by multivariate analysis (see below).

10.3.1

Fingerprint Representation

A fingerprint data set can be represented in a visual format to compare multiple enzymes and substrates simultaneously. This is done conveniently using grayscale or multicolor array displays associating each square in an array with an enzyme–substrate combination identifiable by its coordinates. The color scale

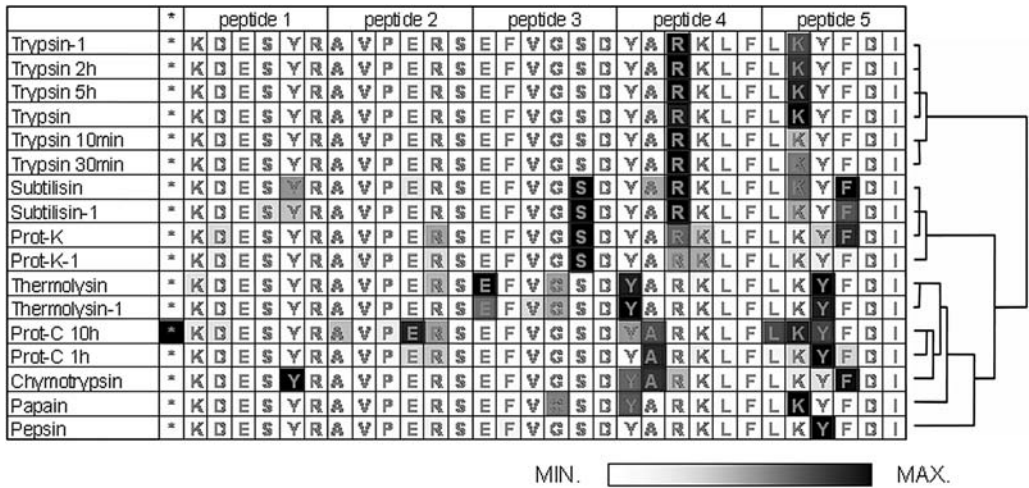


Fig. 10.18 Grayscale-coded protease fingerprints from domino-peptide cocktail. The relative distribution of products is shown by coloring the P1 position of cleavage of each detected N-terminal coumarin-labeled fragment relative to the most abundant fragment detected (shown in black) using the indicated scale. Unreacted substrates are not color coded. Proteases are ordered according to the hierarchical clustering

shown at right (Ward's clustering using squared Euclidean distances with product percentages as variables. The variables were not normalized).

* is the (7-coumaryl)oxycetyl label and the capital letters are standard amino acid codes. C-termini of the peptide substrates are CONH₂ (peptide 1, 3, 5) or COOH (peptide 2, 4).

should be adapted for each fingerprint such that the full color scale is used to span reactivity scales from zero to the maximum reaction rate or conversion observed in the fingerprint. Rectangular grids representing the substrate array associated with an enzyme are a convenient option for visual representation. Such grids can be created directly from a data table by conversion to the portable gray map (.pgm) or portable pixel map (.ppm) file formats. Grayscales are suitable for representation of a single data point per square in the array. The representative series of gray-scale fingerprints in Figure 10.17 was obtained with the carbohydrate acetates array in Figure 10.14 using the indirect adrenaline test described in Figure 10.12. The advantage of the array layout is that similarities between enzymes can be quickly assessed by a visual inspection of the data. Such comparisons are not possible if the data are represented as table numbers.

One can also represent a series of fingerprints in a combined array where each line corresponds to a different enzyme fingerprint, and each column corresponds to a different substrate in the array. Such combined arrays can be integrated into a table, giving additional information about enzymes and substrates, and the different fingerprints can be ordered according to data clustering as discussed below. The clustered data set of protease cleavage patterns derived from the domino cock-

tail is shown as a representative example of such data display (Figure 10.18). The contrast in these grayscales can be augmented by using multicolor scales as available from many graphic programs, for example the blue–green–yellow–red scale used for coding temperature ranges. One of the most useful multicolor scales combines color shading with color intensity by using gradual desaturation of blue, green, and red channels, which translates into a white–yellow–orange–red–black color scale with increasing data values.

Selectivities between pairs of substrates (enantiomers, stereoisomers, or analogs) can be represented using a two-dimensional color code [1]. In this format one square in the array represents data for the substrate pair by associating one color channel to the first substrate, a second color channel to the second substrate, and the third color channel to the average value of the first two channels. Complete selectivity for either substrate appears as a pure color, and the absence of se-

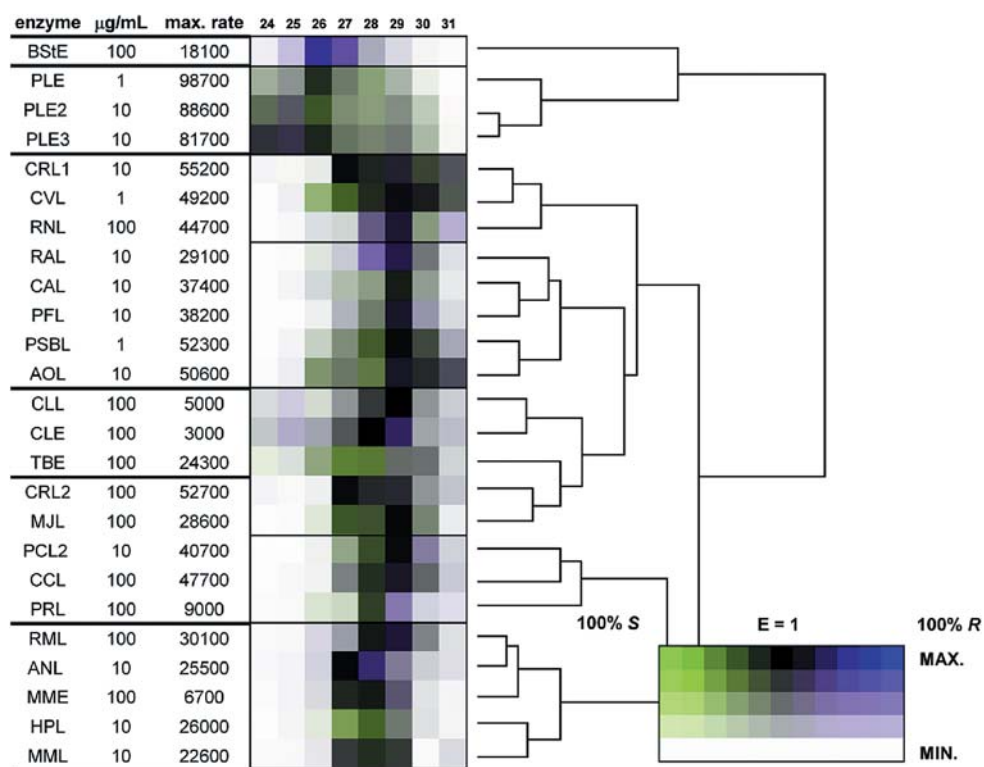


Fig. 10.19 Lipase/esterase fingerprints and hierarchical tree. Each line represents a different enzyme, and each column represents a different substrate. Each colored square represents the reaction rates of two enantiomeric lipase substrates (measured separately) relative to the

maximum rate observed with the corresponding enzyme (line), which is given in pM s^{-1} (color key at lower right). Hierarchical agglomerative clustering was carried out using Ward's method on the basis of Euclidean distances.

lectivity (equal rates for each of the two substrates) appears as gray. This color scale is very convenient for visual analysis. Taking into account the occurrence of color-blindness in red–green contrast, the best choice of colors is to use the green (G) and blue (B) channel of the RGB code for the substrate values, and the red channel (R) for the average, thus producing a purple-to-green contrast map. An example of color-coded representation of a data set is shown in Figure 10.19, showing the data set of lipase and esterase fingerprints obtained against the array of enantiomeric fluorogenic monoacyl glycerol analogs in Figure 10.9.

Procedure 10.3: Data Treatment for Representation of Colored Selectivity Arrays

The file format used to generate the colored selectivity arrays is the portable pixel map (.ppm) format. Each grid position is first assigned three whole numbers corresponding to the RGB color code between 0 (zero intensity, maximum activity) and 255 (maximum intensity, no activity) as follows: the first number is set according to the activity observed with the (*R*)-enantiomer (or the first of two given stereoisomers), the second number according to the activity observed with the (*S*)-enantiomer (or second stereoisomer), and the third number is simply the mathematical average of the first two numbers. Thus a grid with *X* columns and *Y* lines is coded with 3*X* columns and *Y* lines of whole numbers between 0 and 255. The grid of numbers is saved as comma separated value (.csv) file. This file is then opened in a text editor and the following three (or four) lines are inserted at the top of the file:

```
P3
# (optional line with identifier)
X Y
255
```

where *X* is the number of columns in the array and *Y* the number of lines in the array. The file is then saved from the text editor in the portable pixel map format by simply adding the “.ppm” suffix. This file is opened by image-processing software and can be resized and saved in a different format (e.g. .bmp). In this format it is possible to permute the order of the three columns corresponding to the RGB code, and to obtain the color shading orange-to-blue (when G is the average of R and B) and a green-to-purple (when R is the average of G and B, as shown in Figure 10.19).

10.3.2

Data Normalization

Data processing should be kept to a minimum, paying attention to staying as close to the actual chemistry taking place in the flask as possible. Raw output data recorded during the experiment are expressed in relative fluorescence units

in the case of fluorescence assays, or peak areas when processing HPLC spectra. Such primary data are already suitable for similarity analysis between substrates and between enzymes without any manipulation. Nevertheless one usually converts this primary output into chemically meaningful information such as substrate conversions (%) or reaction rates ($\mu\text{M s}^{-1}$) by means of a calibration curve. With or without such conversion, the unspecific signal recorded in blank assays without enzyme should be subtracted from the primary signal.

If the different substrates in the array display very different chemical reactivities, the rate data can be converted to relative rate accelerations over background to report the enzyme-specific contribution only. If there is no observable background reaction, one can calculate the relative rate in relation to a reference nonselective chemical catalyst (Eq. 1). Reporting the enzyme-specific rate acceleration rather than the absolute reaction rate allows a stronger statistical weight to be assigned to unreactive substrates, such as esters of hindered alcohols, where even a small conversion by the enzyme is noteworthy.

$$V(i)_{\text{rel}} = V(i)_{\text{obs}}/V(i)_{\text{ref}} \quad (1)$$

where $V(i)_{\text{obs}}$ is the observed reaction rate or conversion of substrate i ; $V(i)_{\text{ref}}$ is the reaction rate or conversion of substrate i with a reference nonselective chemical catalyst; $V(i)_{\text{rel}}$ is the relative rate acceleration of substrate i over reference.

For the purpose of enzyme similarity analysis, we usually normalize the data such that the sum of all reaction rates or conversions observed with a given enzyme across the different substrates in the array is set to a constant value of 1 or 100% (Eq. 2). In this manner two samples of the same enzyme at two different concentrations or at two different reaction times should appear similar. This normalization enables to conserve the actual ratio between each substrate across the total enzyme reactivity against the array. This allows one to compare biocatalysts independently of their concentration in the test samples. Such a selectivity analysis circumvents the need to define enzyme activity units for each enzyme sample being analyzed. The lipase selectivity data set (Figure 10.19) was normalized using this method.

$$V(i)_{\text{rel}} = V(i)_{\text{obs}} / \sum_{i=1}^n V(i) \quad (2)$$

where $\sum_{i=1}^n V(i)$ is the sum of the reaction rates of all the substrates for a given enzyme.

Although we have not tested all possible data treatment approaches, we found that the following two data correction methods must be avoided in fingerprint data analysis. First, reporting the data for each substrate relative to the maximum rate or conversion observed in the fingerprint, as is done for the color-coded representation, provides unreliable results in terms of statistical comparisons, probably due to the automatic propagation of any error relating to this maximum peak (Eq. 3). Second, the standard variable normalization methods in

multivariate analysis, which consists in transforming each variable such that its average is 0 and its standard deviation is 1, must be avoided because all substrates receive equal statistical weight, in particular substrates with very low conversion where the observed small variations in reaction rate or conversion reflect measurement inaccuracies rather than actual data.

$$V(i)_{\text{rel}} = V(i)/V_{\text{max}} \quad (3)$$

where V_{max} is the maximum reaction rate observed for an enzyme over the array. Correction of rate data by nonlinear mathematical operations (e.g. logarithmic scales) can be considered when the different reaction rates or conversions comprising the fingerprint span several orders of magnitude yet are all reliably measured. This correction should be considered after normalization of all observed rates of conversion to a constant sum as discussed above, such as to preserve a data set independent of catalyst concentration.

10.3.3

Hierarchical Clustering of Enzyme Fingerprints

In multivariate analysis, each enzyme is considered as an observation and positioned as a point in a multidimensional space whose dimensions correspond to the reaction rates observed with each substrate in the array, defined as variables

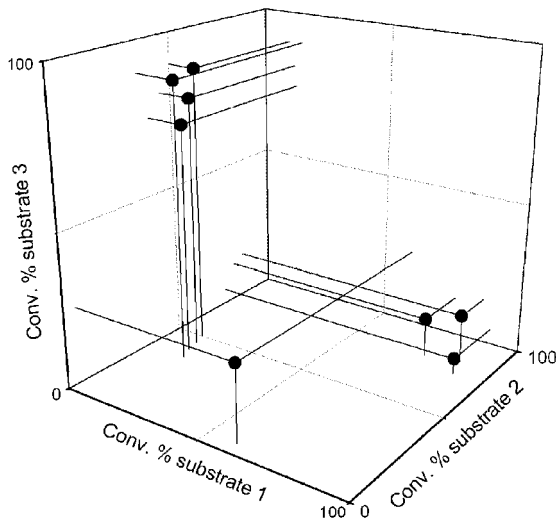


Fig. 10.20 Multivariate analysis of enzyme rate data. Each substrate defines a dimension, and the enzyme is positioned in the n -dimensional space using the observed reaction rate or conversion with each substrate as coordinate. Hierarchical cluster analysis compares distances in this n -dimensional space.

(Figure 10.20). Similarity analysis is based on comparing distances between different enzymes (observations) in this multidimensional space or the different substrates (variables) [49]. This comparison depends on the relative weight assigned to the variables as discussed in the previous section.

Hierarchical clustering is an automated procedure for grouping different observations consisting in a set of measured variables according to their similarity (geometrical proximity in the variable space). A number of distance measures and algorithms can be chosen for clustering. For the case of enzyme fingerprints, we have used the agglomerative technique based on Ward's method which is the most commonly used in statistics [50]. We employed Euclidean and squared Euclidean distances as a measure of similarity. The latter is actually recommended [51], but both resulted in the formation of chemically meaningful clusters. A representative example is the data set in Figure 10.19. The classification proposed by Ward clustering correctly represents the intuitive reactivity pattern analysis operated visually, and groups esterases operating on short-chain acyl groups at the beginning of the list, and lipases active on long-chain acyl groups at the end of this list. The cluster structure is shown in the hierarchical tree.

Principal component analysis (PCA) reduces the dimensionality by combining correlated variables linearly into one factor whilst maximizing the variance captured in this factor. For example, PCA identifies if a set of points in a three-

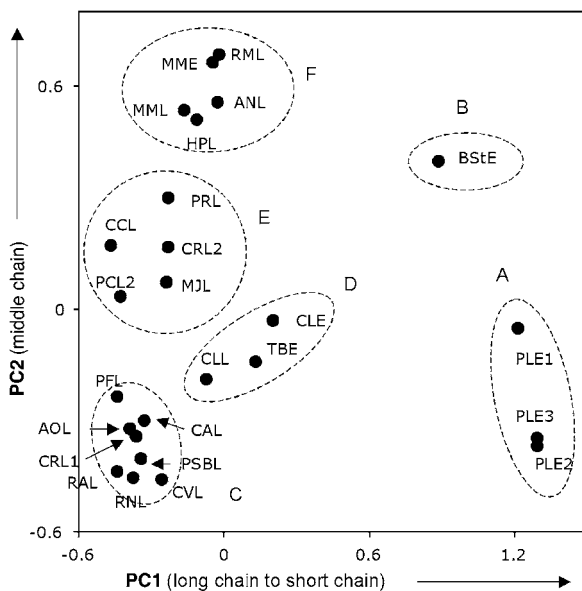


Fig. 10.21 Principal component analysis (PCA) of lipases and esterases from fingerprint data. The groups A–F identified visually (dashed circles) correspond to those obtained by agglomerative clustering, with the exception of CRL2 and MJL in cluster E, which are assigned to cluster D (see Figure 10.19).

dimensional space are grouped in the same plane (two dimensions) or along the same line (one dimension) and determines the corresponding axes, with coefficients for each variable. In most data sets more than 60% of the total variance is explained by the first two principal components, allowing a 2D representation of the intrinsic structure of the clusters. For example, the lipase and esterase data set discussed above can be represented in such a 2D plot, giving a more graphical layout to the data (Figure 10.21).

10.3.4

Analysis of Substrate Similarities

The chemical composition of the substrate series composing a fingerprinting array is central to the analysis method, and the structures of these substrates contain the chemical essence of the measurement. As pointed out above, fingerprints based on a limited set of substrates must incorporate substrates displaying a significant reactivity with the targeted enzyme class, such as to avoid “zero” fingerprints devoid of any reactivity information. Most substrates should therefore be variations on a structural type known to be favored by the enzyme class under investigation.

Any substrate array can be investigated for the functional relevance of its substrates once a data set has been acquired. In particular, the key question is whether the different substrates in the array all return differential reactivity information across the different enzymes tested. The PCA discussed above answers part of the question. Indeed if each substrate in the array behaves differently, the variability of the data set spans as many dimensions as substrates. However, if any two substrates display a similar reactivity pattern across the different enzymes, the diversity observed will be reduced to a lower number of dimensions. In the data set of lipase and esterase fingerprints discussed above, the first two principal components explain 68% of the diversity observed between the different enzymes. A bar diagram representation of the principal component coefficients indicates how each substrate contributes to the principal components of the data set (Figure 10.22).

Substrate similarities can also be directly investigated by hierarchical clustering or PCA of substrates as observations against enzymes as variables. The data set is first normalized against substrates, with the sum of all reaction rates observed for each substrate across all enzymes being set to a constant of 100%. The results of clustering can be visualized as a hierarchical tree, or by direct rendering of the symmetrical distance matrix, which indicates functional similarities between substrates, as illustrated for the lipase data set (Figure 10.23). Pairs of enantiomers appear very similar in their reactivity against the different enzymes with the exception of the butyryl and hexanoyl esters, suggesting that this particular fingerprinting set could be reduced to a lower number of substrates to return essentially the same analysis results in terms of differentiation between enzymes.

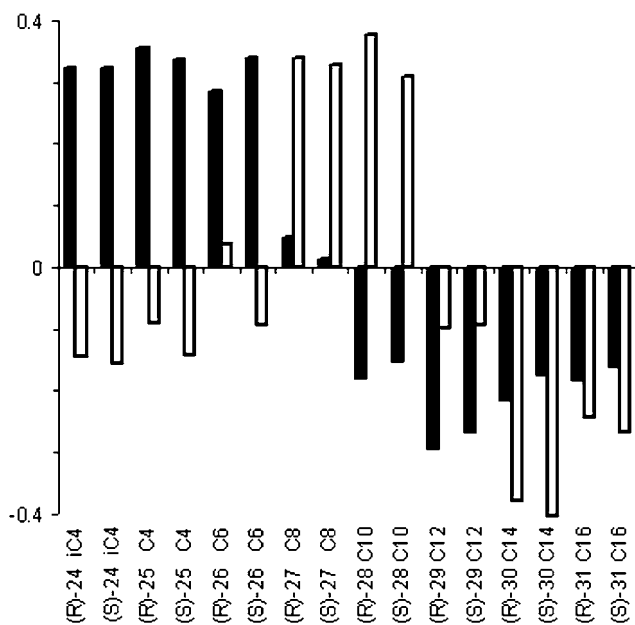


Fig. 10.22 Coefficients per substrate of the first two principal components (PC), which account for 68% of observed variance: PC1=46% (black bars), PC2=22% (white bars).

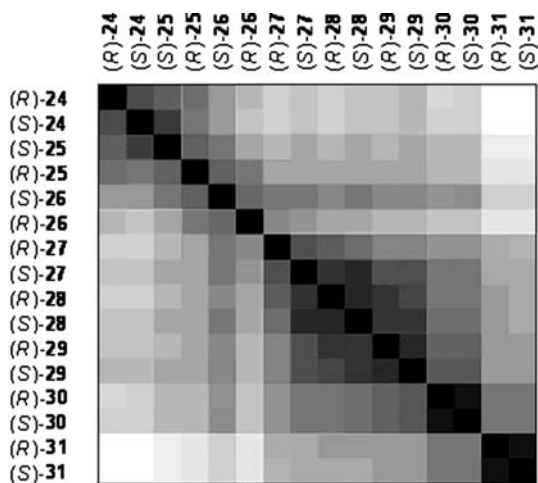


Fig. 10.23 Euclidean distance matrix between substrates 24–31 in the 25-dimensional space of enzyme reaction rate, rendered in grayscale (black = distance 0, white = maximum distance). For each substrate, reactivity for each enzyme was

expressed as per cent of the total reactivity observed with this substrate across all enzymes measured. Substrates were clustered by agglomerative clustering using the group-average method.

10.4 Outlook

Enzyme fingerprints are images of the reactivity profile of an enzyme across the chemical structural space of a given substrate type. The resolution of these images depends on the number and structural diversity of the substrates composing the fingerprinting array. Fingerprinting arrays composed of small sets of substrates (<20) are readily assembled and used for fingerprinting by parallel measurements in microtiter plates or in the form of cocktails analyzed by a separative analysis such as HPLC. These arrays or cocktails can be easily reproduced and therefore used as reference measurements of activity profiles. Cocktails can be made fluorogenic by using FRET labels, thereby providing general activity detection and functional identification reagents for enzyme classes such as lipases and proteases [52].

The reactivity profiles derived from enzyme fingerprints provide sufficient information for a functional classification and identification of closely related enzymes within a class. Enzyme fingerprinting can be used as microbial strain identification in medical diagnostics, as exemplified by the APIZYM system. This analysis principle may be extendable to quality control and medical diagnostics in a broadened context. Enzyme fingerprints may also prove useful for enzyme engineering. The most functionally diverse enzymes of a given collection, for example a microbial strain collection, could be identified by fingerprinting and selected as representative members for testing new substrate types, or as most favorable starting points for gene shuffling experiments to create new enzymes with novel selectivities. Fingerprinting might also enable a systematic analysis of the effect of directed evolution on reactivity landscapes to reveal the effect of single mutations on reactivity in a broader context. Finally, the data from enzyme fingerprinting experiments should provide a rich source of information to complement structural information and lead to a refined understanding of structure–activity relationships in enzymes.

Acknowledgment

This work was financially supported by the University of Berne, the Swiss National Science Foundation, and Protéus SA, Nîmes, France.

References

- 1 a) D. Wahler, F. Badalassi, P. Crotti, J.-L. Reymond, *Angew. Chem. Int. Ed.* **2001**, *40*, 4457–4460; b) D. Wahler, F. Badalassi, P. Crotti, J.-L. Reymond, *Chem. Eur. J.* **2002**, *8*, 3211–3228.
- 2 J.-L. Reymond, D. Wahler, *ChemBioChem* **2002**, *3*, 701–708.
- 3 U. Reimer, U. Reineke, J. Schneider-Mergener, *Curr. Opin. Biotechnol.* **2002**, *13*, 315–320.
- 4 a) E. Gruner, A. Von Graevenitz, M. Altwegg, *J. Microbiol. Methods* **1992**, *16*, 101–118; b) P. Garcia-Martos, P. Marin, J.M. Hernandez-Molina, L. Garcia-Agudo, S. Aoufi, J. Mira, *Mycopathologia* **2001**, *150*, 1–4.
- 5 M. Baumann, R. Stürmer, U.T. Bornscheuer, *Angew. Chem.* **2001**, *113*, 4329–4333; *Angew. Chem. Int. Ed.* **2001**, *40*, 4201–4204.
- 6 Nomenclature Committee of the International Union of Biochemistry (NC-IUCB), Recommendation 1978, *Eur. J. Biochem.* **1979**, *97*, 319–320.
- 7 M. Kordel, U. Menge, G. Morelle, H. Erdmann, R.D. Schmid, Comparative analysis of lipases in view of protein design in *Lipases: Structure, Mechanism and Genetic Engineering*, GBF Monographs vol. 16, eds L. Alberghina, R.D. Schmid, R. Verger, VCH, Weinheim, **1991**.
- 8 J. Kraut, *Science* **1988**, *242*, 533–540.
- 9 M.D. Joshi, G. Sidhu, I. Pot, G.D. Brayer, S.G. Withers, L.P. McIntosh, *J. Mol. Biol.* **2000**, *299*, 255–279.
- 10 J.M. Denu, J.E. Dixon, *Proc. Natl Acad. Sci. USA* **1995**, *92*, 5910–5914.
- 11 D.J. Hei, D.S. Clark, *Biotechnol. Bioeng.* **1993**, *42*, 1245–1251.
- 12 Y.-X. Fan, P. McPhie, E.W. Miles, *J. Biol. Chem.* **2000**, *275*, 20301–20307.
- 13 a) S. Kunugi, H. Hirohara, N. Ise, *Eur. J. Biochem.* **1982**, *124*, 157–163; b) D.G. Truhlar, A. Kohen, *Proc. Natl Acad. Sci. USA* **2001**, *98*, 848–851.
- 14 a) G. Lenaz, A.M. Sechi, G. Parenti-Castelli, L. Landi, E. Bertoli, *Biochem. Biophys. Res. Commun.* **1972**, *49*, 536–542; b) M.L. Puterman, N. Hrboticky, S.M. Innis, *Anal. Biochem.* **1988**, *170*, 409–420.
- 15 J.R. Silvius, B.D. Read, R.N. McElhaney, *Science* **1978**, *199*, 902–904.
- 16 J. Bussièrè, A. Foucart, L. Colobert, C.R. Acad. *Seances Paris* **1967**, *264 D*, 415–417.
- 17 D. Monget, P. Nardon, French patent FR 2341865, Paris, **1976**, US patent 4277561, **1981**.
- 18 M.W. Humble, A. King, I. Philipps, *J. Clin. Pathol.* **1977**, *30*, 275–277.
- 19 a) E. Gruner, A. Von Graevenitz, M. Altwegg, *J. Microbiol. Methods* **1992**, *16*, 101–118; b) P. Garcia-Martos, P. Marin, J.M. Hernandez-Molina, L. Garcia-Agudo, S. Aoufi, J. Mira, *Mycopathologia* **2001**, *150*, 1–4.
- 20 <http://www.biomerieux.com>
- 21 R. Sicard, J.-P. Goddard, M. Mazel, C. Audiffren, L. Fourage, G. Ravot, D. Wahler, F. Lefèvre and J.-L. Reymond, *Adv. Synth. Catal.* **2005**, *347*, 987–996.
- 22 a) G. Klein, J.-L. Reymond, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1113–1116; b) G. Klein, J.-L. Reymond, *Helv. Chim. Acta* **1999**, *82*, 400–407.
- 23 a) N. Jourdain, R. Perez-Carlón, J.-L. Reymond, *Tetrahedron Lett.* **1998**, *39*, 9415–9418; b) R. Pérez Carlón, N. Jourdain, J.-L. Reymond, *Chem. Eur. J.* **2000**, *6*, 4154–4162.
- 24 E.M. Gonzalez-Garcia, V. Hélaïne, G. Klein, M. Schuermann, G. Sprenger, W.-D. Fessner, J.-L. Reymond, *Chem. Eur. J.* **2003**, *9*, 893–899.
- 25 B. List, C.F. Barbas, R.A. Lerner, *Proc. Natl Acad. Sci. USA* **1998**, *95*, 15351–15355.
- 26 F. Tanaka, L. Kerwin, D. Kubitz, R.A. Lerner, C.F. Barbas, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2983–2986.
- 27 E. Leroy, N. Bensenl, J.-L. Reymond, *Adv. Synth. Catal.* **2003**, *345*, 859–865.
- 28 A. Sevestre, V. Hélaïne, G. Guyot, C. Martin, L. Hecquet, *Tetrahedron Lett.* **2003**, *44*, 827–830.
- 29 M.C. Gutierrez, A. Slegers, H.D. Simpson, V. Alphand, R. Furstoss, *Org. Biomol. Chem.* **2003**, *1*, 3500–3506.
- 30 W. Gao, B. Xing, R.Y. Tsien, J. Rao, *J. Am. Chem. Soc.* **2003**, *125*, 11146–11147.

- 31 C. A. Roeschlaub, N. L. Maidwell, M. Reza Rezai, P. G. Sammes, *Chem. Commun.* **1999**, 1637–1638.
- 32 E. Leroy, N. Bensele, J.-L. Reymond, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2105–2108.
- 33 R. Sicard, L. S. Chen, A. J. Marsaioli, J.-L. Reymond, *Adv. Synth. Catal.* **2005**, *347*, 1041–1050.
- 34 F. Badalassi, D. Wahler, G. Klein, P. Crotti, J.-L. Reymond, *Angew. Chem. Int. Ed.* **2000**, *39*, 4067–4070.
- 35 a) D. Lagarde, H.-K. Nguyen, G. Ravot, D. Wahler, J.-L. Reymond, G. Hills, T. Veit, F. Lefevre, *Org. Process. R. & D.* **2002**, *6*, 441–445; b) E. Nyfeler, J. Grognum, D. Wahler, J.-L. Reymond, *Helv. Chim. Acta* **2003**, *86*, 2919–2927; c) J. Grognum, D. Wahler, E. Nyfeler, J.-L. Reymond, *Tetrahedron Asymmetry* **2004**, *15*, 2981–2989.
- 36 F. Badalassi, G. Klein, P. Crotti, J.-L. Reymond, *Eur. J. Org. Chem.* **2004**, 2557–2566.
- 37 E. M. Gonzalez-Garcia, J. Grognum, D. Wahler, J.-L. Reymond, *Helv. Chim. Acta* **2003**, *86*, 2458–2470.
- 38 F. Badalassi, H.-K. Nguyen, P. Crotti, J.-L. Reymond, *Helv. Chim. Acta* **2002**, *85*, 3090–3098.
- 39 J. Wierzchowski, W. P. Dafeldecker, B. Holmquist, B. L. Vallee, *Anal. Biochem.* **1989**, *178*, 57–62.
- 40 J. Grognum, J.-L. Reymond, *ChemBioChem* **2004**, *5*, 826–831.
- 41 P. Babiak, J.-L. Reymond, *Anal. Chem.* **2005**, *77*, 373–377.
- 42 A. M. F. Liu, N. A. Somers, R. J. Kazlauskas, T. S. Brush, F. Zocher, M. M. Enzelberger, U. T. Bornscheuer, G. P. Horsman, A. Mezzetti, C. Schmidt-Dannert, R. D. Schmid, *Tetrahedron Asymmetry* **2001**, *12*, 545–556.
- 43 a) L. E. Janes, R. J. Kazlauskas, *J. Org. Chem.* **1997**, *62*, 4560–4561; b) L. E. Janes, A. C. Löwendahl, R. J. Kazlauskas, *Chem. Eur. J.* **1998**, *4*, 2324–2331.
- 44 D. Wahler, J.-L. Reymond, *Angew. Chem. Int. Ed.* **2002**, *41*, 1229–1232.
- 45 D. Wahler, O. Boujard, F. Lefevre, J.-L. Reymond, *Tetrahedron* **2004**, *60*, 703–710.
- 46 a) G. Klein, J.-L. Reymond, *Angew. Chem. Int. Ed.* **2001**, *40*, 1771–1773; b) K. E. S. Dean, G. Klein, O. Renaudet, J.-L. Reymond, *Bioorg. Med. Chem. Lett.* **2003**, *10*, 1653–1656.
- 47 J.-P. Goddard, J.-L. Reymond, *J. Am. Chem. Soc.* **2004**, *126*, 11116–11117.
- 48 Y. Yongzheng, J.-L. Reymond, *MolBiosys* **2005**, *1*, 57–63.
- 49 C. Chatfield, A. J. Collins, *Introduction to Multivariate Analysis*, Chapman and Hall, London, **1983**.
- 50 B. S. Everitt, *Cluster Analysis*, Edward Arnold, London, **1993**.
- 51 B. S. Everitt, S. Landau, M. M. Leese, *Cluster Analysis*, Edward Arnold, London, **2001**.
- 52 Y. Yongzheng, PhD thesis, University of Berne, **2005**.

11

Protease Substrate Profiling

Jennifer L. Harris

11.1

Introduction

For all of the coordination and energy that an organism puts into the process of protein production, the degradation of proteins is an equally important and critical process. The specialized enzymes that carry out the hydrolysis of peptide bonds are commonly referred to as proteases (also known as peptidases and proteinases). With more than 500 proteases encoded in the human genome, proteases are one of the largest classes of enzymes [1, 2]. The fundamental activity of a protease is the hydrolysis of peptide bonds. Proteases can be divided into two general groups: exopeptidases that cleave terminal peptide bonds and endopeptidases that cleave internal peptide bonds. Proteases are further classified on the basis of catalytic mechanism into five classes, serine, threonine, cysteine, aspartyl, and metalloproteases [1]. In the human genome, metalloproteases are the largest class, with approximately 186 members followed by serine proteases with 176 members, cysteine proteases with 143 members and threonine and aspartyl proteases with just 27 and 21 members respectively [2].

Historically, proteases were recognized as nonselective, promiscuous enzymes that were responsible for indiscriminately degrading dietary proteins. However, with the commitment that is inherent in the irreversibility of the peptide hydrolysis reaction carried out by proteases coupled with the exquisite substrate discrimination shown by many proteases, it is now widely accepted that this enzyme class plays crucial roles in the initiation and regulation of biological pathways. Indeed, proteases have been implicated in almost all aspects of life and death – from fertilization, development, differentiation, homeostasis, immunity, cell migration, cell activation, wound healing, to cell death. Likewise, from a therapeutic standpoint, the modulation of proteolytic activity offers considerable promise for the treatment of human diseases, among them cardiovascular disease, lung disease, cancer, stroke, inflammation, arthritis, asthma, osteoporosis, neurodegeneration, and infectious disease. It has been estimated that approximately 14% of human proteases are under investigation as drug targets, this in addition to the proteases targeted from infectious diseases [3].

With the complete sequencing of multiple genomes, information on the repertoire of proteases can be established; however, large gaps still remain in our understanding of the biological role of most proteases. These gaps include limited characterization of enzyme active sites, limited understanding of the substrates and pathways that specific proteases regulate or are regulated by, and redundancy of function among proteases. In contrast to genomics, where the changes in the content or amount of cellular DNA or RNA can be readily examined, monitoring translational and posttranslational dynamics of functional proteins on a genome-wide level is a more difficult challenge. This challenge of profiling the functional activity of proteases has led to the development of multiple tools and technologies (reviewed in [4–8]). This chapter will focus on the advantages, disadvantages, and application of a variety of tools that have been developed for the identification of protease substrates.

11.2

Functional Protease Profiling – Peptide Substrate Libraries

One crucial characteristic of a protease is the ability to discriminate among many potential substrates, termed the substrate specificity of the protease. The substrate specificity of a protease is determined by multiple factors that include the temporal and spatial expression of the protease, temporal and spatial expression of potential substrates, activation of the protease by posttranslational modification, availability of essential cofactors or adaptor proteins, and the presence of endogenous inhibitors. But fundamentally, the protease substrate specificity is determined by the make-up of the substrate-binding pockets that comprise the active site [9, 10] (Figure 11.1). Identifying the preferred substrate cleavage

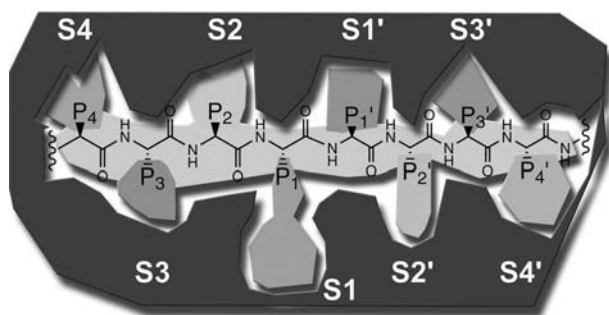


Fig. 11.1 Schematic of protease-substrate subsite binding. The substrate is shown in gray and the protease is shown in black. According to Berger and Schechter [104] nomenclature, the nonprime side substrate sites are labeled P1, P2, ... P_n from the scissile bond towards the N-terminus of the

substrate and the corresponding enzyme sites are designated S1, S2, ... S_n. The prime side substrate sites are labeled P1', P2', ... P_n' from the scissile bond towards the C-terminus of the substrate and the corresponding enzyme sites are designated as S1', S2', ... S_n'.

sequence of a protease can have a large impact on the ability to characterize the underlying biology of the protease as well as aid in the drug discovery process.

Traditionally, the substrate specificity of a protease is determined by digestion of standard proteins such as gelatin, casein, hemoglobin, or insulin B-chain. Following digestion, the fragments are separated by chromatography or electrophoresis and the cleaved sequences are identified typically by mass spectrometry or Edman degradation. An example of this method is demonstrated by Bleux et al. [11] where the substrate specificity of a wheat gluten aspartyl protease was determined through digestion of the 30-amino-acid oxidized insulin B-chain peptide and determination of the cleavage sites using Edman degradation. Using this method, the authors were able to demonstrate that the wheat gluten aspartyl protease preferred to cleave peptide bonds between or next to large hydrophobic amino acids [11]. While cleavage of standard proteins and peptides is widely used, the main limitation is in the number of sequences that can be sampled in a given protein or peptide substrate. Because of this, the results from these studies are useful but only provide a limited analysis of the substrate specificity determinants.

In addition to the digestion of standard proteins by proteases, preparation and evaluation of synthetic peptides with systematic substitutions in substrate sequence sites have been used to define the specific substrate subsites. Examples of this approach are found widely in the literature and allow for the quantification of specific contributions of amino acids in the substrate sequence to the substrate specificity, expressed as the k_{cat}/K_m kinetic term [12, 13]. Taking this approach, Tsu and Craik were able to determine that the serine collagenase 1 from the fiddler crab recognized a wide range of basic, hydrophobic, and polar P1 amino acids [14]. They demonstrated this through the kinetic analysis of the protease against 15 peptide substrates of the form succinyl-Ala-Ala-Pro-Xaa-*p*NA, where *p*NA represents the *para*-nitroanilide chromogenic leaving group that can be monitored spectrophotometrically upon cleavage by the protease.

The major drawback of techniques using single protein or peptide substrates is that only limited substrate sequences can be sampled by the protease. To illustrate, determination of a tetrapeptide sequence of the 20 naturally occurring amino acids would require the systematic synthesis and testing of 160 000 (20^4) distinct tetrapeptide sequences. This number increases exponentially with an eight-amino-acid sequence requiring over 20 billion peptide sequences. Beyond the systematic examination of 1–3 substrate subsites, the synthesis and testing of single proteins or peptides is simply not practical. To address this limitation and to overcome the inefficiencies of producing and screening single substrates, combinatorial chemistry and genetic techniques have been developed for a more exhaustive determination of the substrate specificity of proteases.

11.2.1

Solution-based Peptide Substrate Libraries

The introduction of combinatorial synthesis techniques has allowed for the facile production of large peptide libraries with which protease activity can be investigated [15]. While the ease of peptide production has increased the number of peptides one can make, the challenge still remains on how to assay, deconvolute, and interpret the results of combinatorial peptide libraries. Because of this, considerable effort has been put into the development of mixture-based screening for the identification of protease substrate specificity determinants.

Positional scanning synthetic combinatorial libraries (PSSCLs) developed by Houghten et al. [16] represent a solution-based method for the determination of substrate specificity of proteases. The major advantage of PSSCLs is that they exploit the power of combinatorial peptide synthesis, in which all possible combinations of amino acids can be represented, while simplifying the time and effort for library production, analysis, and deconvolution. The PSSCL concept can be illustrated by the creation of a tetrapeptide library of 20 amino acids for 160 000 possible sequences (20^4). In a one-position fixed tetrapeptide PSSCL, the library is synthesized in 80 discrete pools. Each pool consists of 8000 peptide substrates where one position is defined and the other three positions contain an equimolar mixture of 20 amino acids ($20 \times 20 \times 20 \times 1 = 8000$). In other words, the same 160 000 different peptide substrates are assayed four different times, each time probing a different subsite of the peptide independent of the other three positions. The PSSCL synthesis and deconvolution method has been used to identify individual active molecules in a variety of assays, including the identification of peptide inhibitors of prohormone convertase 1 and 2 by Apletalina et al. [17].

The application of PSSCL to protease substrates was elaborated by Thornberry et al. who combined the use of PSSCLs with 7-aminocoumarin-linked peptide substrates for the rapid and quantitative determination of the substrate specificity of the caspase family of serine proteases [18, 19]. Two of the main advantages of peptide coumarin substrates, first described by Zimmerman et al. [20], are that (1) the fluorescence of the coumarin group is masked when the 7-amino group is conjugated to a peptide substrate and (2) the register of cleavage is defined (Figure 11.2a). Both of these advantages rest on the fact that only upon cleavage of the peptide–anilide bond does the coumarin group undergo a shift in the excitation and emission wavelengths with a concomitant increase in fluorescence over the uncleaved peptide coumarin substrate.

Caspases are cysteine proteases that require Asp at the P1 site of their substrates. The requirement for P1-Asp was exploited for the synthesis of coumarin substrate libraries by linking the carboxylic acid of the aspartic acid side-chain to a solid support to facilitate the solid-phase synthesis of the PSSCL [18] (Figure 11.2b). Cleavage of the library from the solid support resulted in regeneration of the aspartic acid and allowed for solution-based screening of the library. While the P1 position was held constant as P1-Asp, the P2, P3, and P4 side-

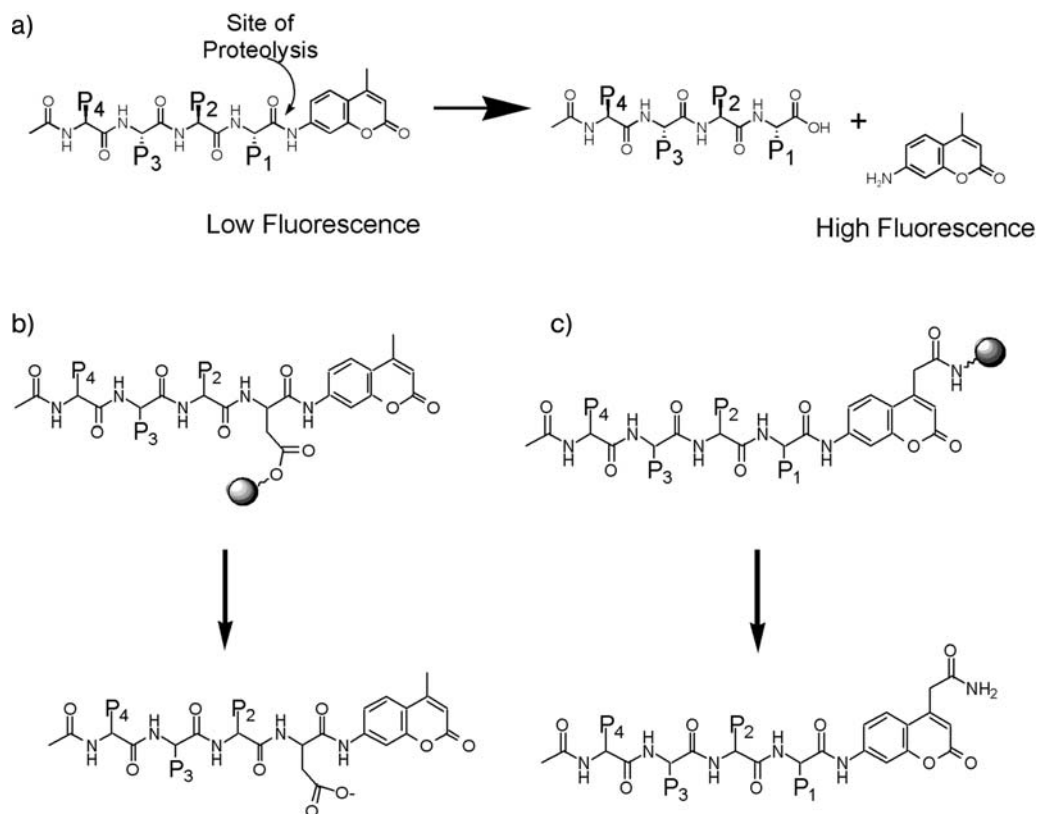


Fig. 11.2 7-Aminocoumarin substrates. (a) Structure of 7-aminocoumarin substrate. Upon cleavage at the anilide bond there is a concomitant increase in fluorescence of the coumarin. (b and c) Comparison of solid-phase synthesis strategies for coumarin substrates linked to the resin through the P1-aspartic acid (b) or through modification of coumarin (c).

chains were varied and included the 20 natural amino acids, with the exception of cysteine and methionine, and the inclusion of two non-natural amino acids norleucine and *D*-alanine for a total of 8000 substrates (20^3) in 60 pools of 400 substrates per pool. The library was used to determine the substrate specificity of the caspase family of cysteine proteases and the serine protease granzyme B [19]. The results from their analysis showed that the caspase family could be separated into three classes based on their substrate specificity as determined from the substrate library. All three classes showed a preference for Asp in the P1 position and Glu in the P3 position. The largest differences in preference between the enzymes were observed to be in the P4 position. Class I caspases showed a preference for large hydrophobic amino acids in the P4 position and His in the P2 position. Caspases in class I included those enzymes associated with inflammation such as caspases 1, 4, and 5. The class II caspases showed a

distinct preference for Asp in the P4 position and include caspases 2, 3, and 7. The consensus sequence of Asp-Glu-Xaa-Asp is found in many of the downstream macromolecular substrates cleaved during apoptosis, suggesting that members of class II are involved during the effector phase of apoptosis. Finally, the class III members, caspases 6, 8, and 9, show a P4 preference for aliphatic amino acids. The resulting consensus site of (Ile/Leu/Val)-Glu-Xaa-Asp resembles the zymogen activation site for the class II caspases and implicates the class III enzymes as upstream activators of the apoptotic protease cascade.

The major limitation of the method described above is that the synthesis of the libraries requires connection to the solid support through the P1-Asp side-chain (Figure 11.2b). Only a fraction of the natural amino acid side-chains have functionality that would allow them to be covalently linked to a solid support. To overcome this limitation, Harris et al. developed a modified 7-aminocoumarin that contains a carboxylic acid group rather than methyl group in the 4-position of the coumarin [21, 22]. The coumarin could then be directly linked to the solid support through its carboxylic acid moiety, allowing for greater diversity in the peptide portion of the molecule (Figure 11.2c). These libraries allowed for complete examination of the P1 functionality and could be combined in multiple formats for substrate preparation and screening. For example, the most simple tetrapeptide library that allows for complete examination of the specificity determinants in a minimum number of wells is the one-position fixed library [23, 24]. In this format, only one position in the tetrapeptide is fixed at any one time and the three other positions contain an equimolar mixture of all the amino acids used in the library. For the serine protease thrombin in a one-position fixed tetrapeptide coumarin library of 19 amino acids (cysteine is not used and methionine is replaced by the isosteric amino acid norleucine), the major determinant is for the basic amino acids, Arg and Lys in the P1 position and Pro or Ala in the P2 position (Figure 11.3). Thrombin does not show a significant preference for amino acids in the P3 and P4 positions in this library format.

Analysis of these results with thrombin brings up one major drawback to the format of positional scanning libraries where only one position is examined in any given pool of substrates – the inability to observe cooperativity or dependence between subsites in the substrate. Clearly, the format of a one-position fixed PSSCL only allows for a gross overview of substrate specificity determinants. To accomplish a more comprehensive evaluation of the substrate specificity of a protease requires reducing the complexity of the mixtures and increasing the number of fixed positions. For example, in a two-position fixed tetrapeptide PSSCL, the same 130 321 (19^4) substrates are monitored six different times, each time with two different positions fixed in the substrate. Rather than 6859 (19^3) substrates per pool, the complexity is reduced to 361 (19^2) substrates per pool, but the number of pools to synthesize and assay has increased from 76 ($19+19+19+19$) to 2166 ($19 \times 19 \times 6$ for the six different combinations of fixing two positions in the substrate).

Another advantage of the two-position fixed format over the one-position fixed format is that the concentration of individual substrates in the pool can be in-

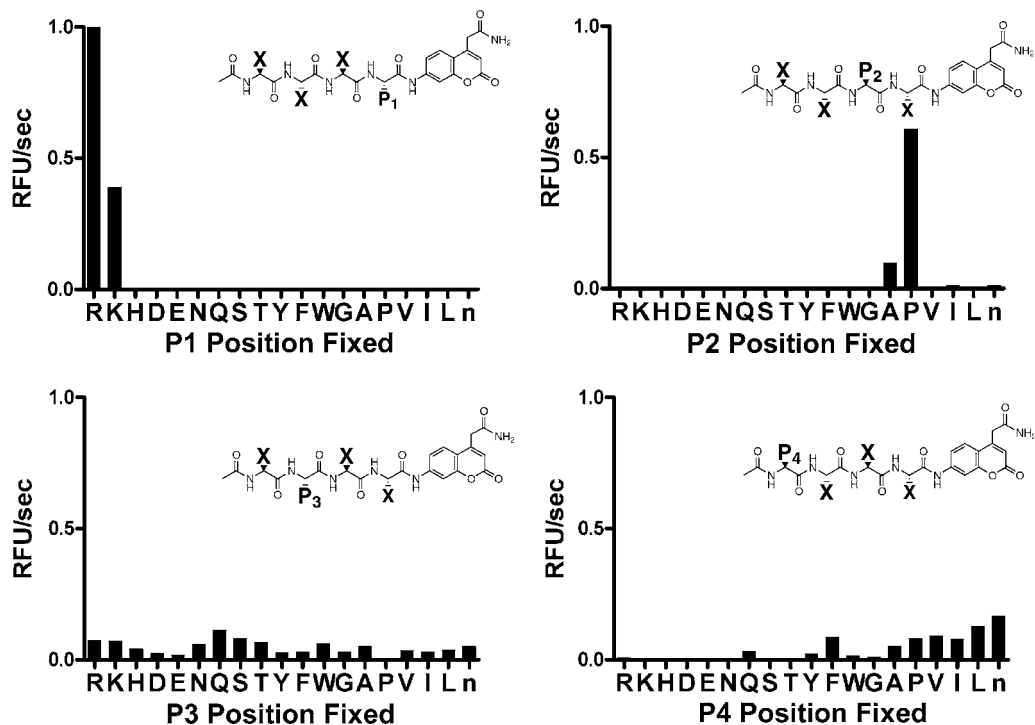


Fig. 11.3 Profiling data from thrombin in a 1-position-fixed tetrapeptide positional scanning synthetic combinatorial library (PSSCL). The x-axis represents the fixed amino acid, the other three nonfixed positions contain an equimolar mixture of 19 amino acids

for a total of 6859 substrates per well (represented by each bar on the histogram). The y-axis represents the normalized rate of coumarin released from the mixture of substrates by thrombin in relative fluorescence units per second (RFU s^{-1}).

creased from approximately $0.01 \mu\text{M}$ in the one-position fixed library to $0.25 \mu\text{M}$ in the two-position fixed library. This is because the concentration of the pooled substrates is limited by the overall solubility of all substrates in the pool, and therefore the lower complexity of the two-position fixed library allows for higher concentrations of individual substrates in the pool, resulting in a higher signal-to-noise ratio. These substrate concentrations are still well below the typical K_m values for these substrates, allowing the rate of hydrolysis to be proportional to the specificity constant, k_{cat}/K_m . The results with thrombin now clearly show a preference for P4 aliphatic amino acids (Figure 11.4) [25].

One major disadvantage of coumarin substrates or any peptide substrate that relies on the cleavage of a peptide/latent leaving group bond, is that the latent leaving group occupies the prime site, preventing examination of prime site/peptide interactions. This limits the utility of these libraries to examining only the nonprime specificity of proteases. Also, while coumarin peptide libraries

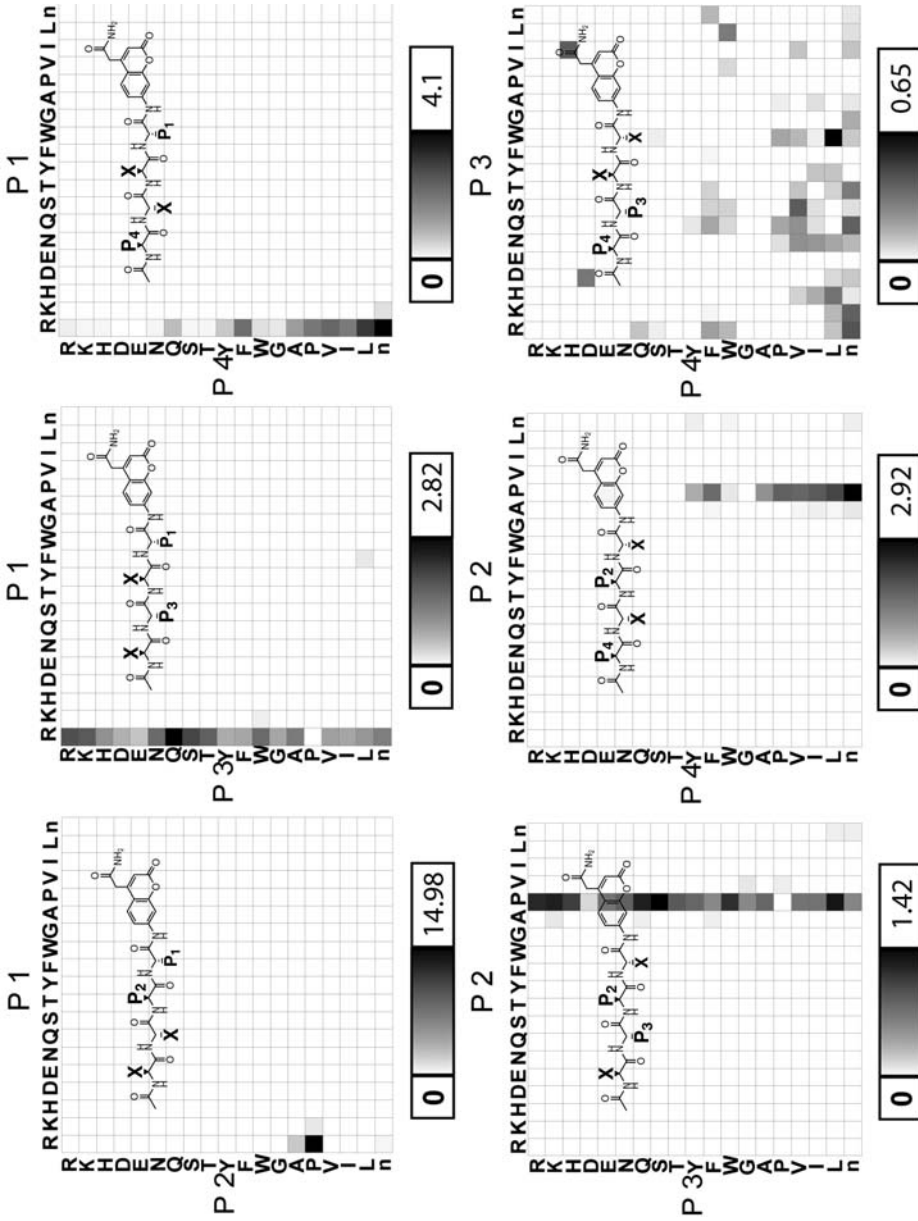


Fig. 11.4 Profiling data from thrombin in a 2 position-fixed tetrapeptide PSSCL. The x- and y-axes represent the fixed amino acid, the other two nonfixed positions contain an equimolar mixture of 19 amino acids for a total of 361 substrates assayed in each well (represented by one square on the 2D array). Each 2D array depicts, by relative intensity, reaction velocities in RFU s^{-1} .

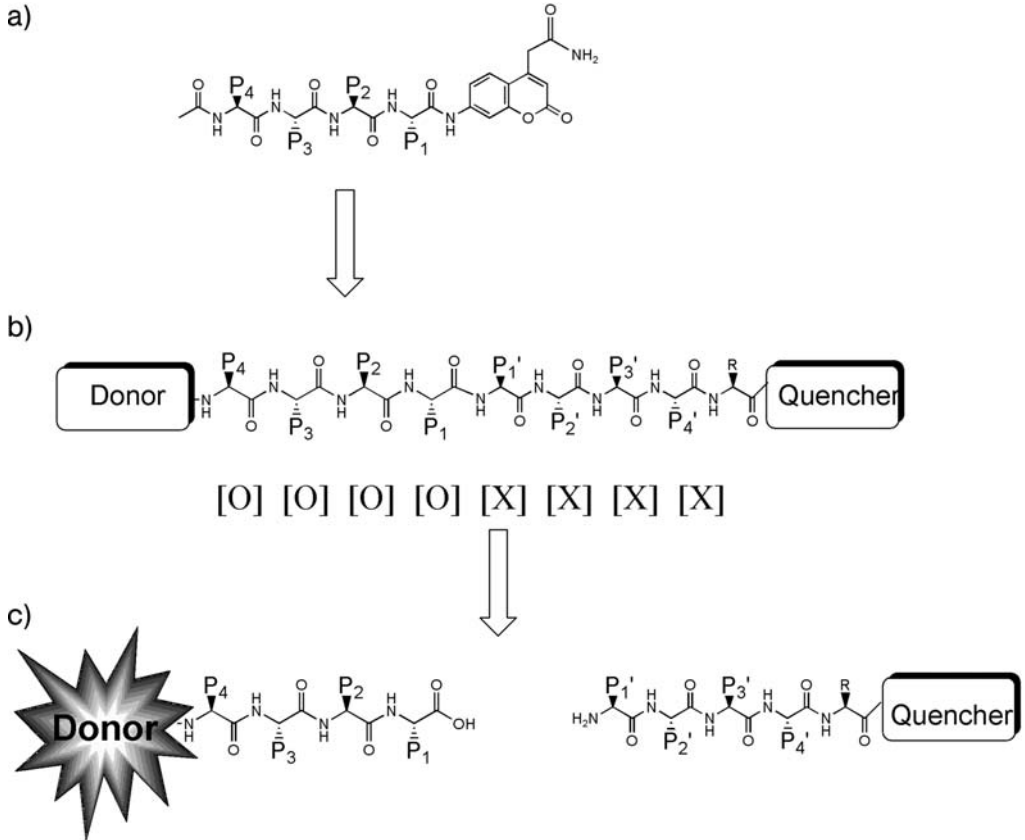


Fig. 11.5 Schematic representation of design of octapeptide donor quencher library.

(a) The nonprime side specificity is determined by the coumarin libraries. (b) The nonprime side sequence is fixed to bias

for cleavage between the P1 and P1' positions of the substrate. (c) Upon cleavage by the protease, the quencher is separated from the fluorescent donor group and the increase in fluorescence can be directly monitored.

have been shown to be effective against some aspartyl proteases [26], in general, these substrates show little activity toward enzymes that have strong requirements for prime site interactions such as aspartyl- and metallo-proteases.

An iterative strategy that incorporates the nonprime information identified from the PSSCL coumarin libraries in a donor-quencher PSSCL format has been recently demonstrated by Shipway et al. for the identification of prime side substrate determinants (Figure 11.5) [25, 27]. After identifying the nonprime side substrate specificity of the serine protease prostaticin as P4-Lys P3-His P2-Tyr P1-Arg, a prime side library was synthesized where the P1' to P4' positions were randomized in a PSSCL format and the P1–P4 positions were held constant as Lys-His-Tyr-Arg. Construction of a one-position fixed PSSCL in this format addressed the issue of catalytic register of the substrate by satisfying the

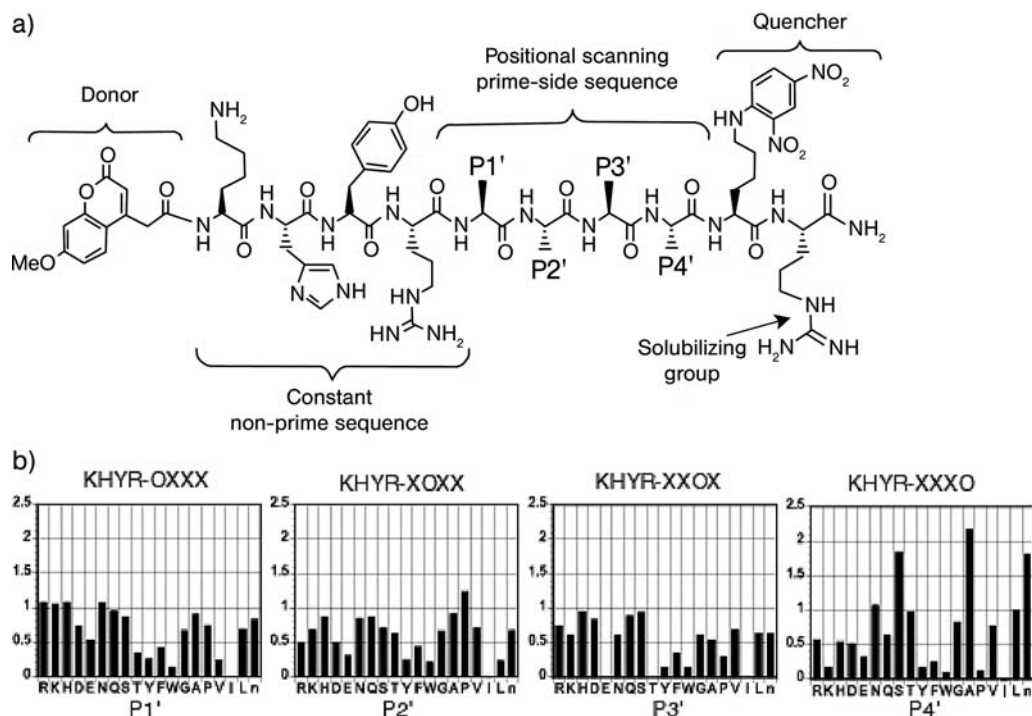


Fig. 11.6 Profiling data from prostatic acid with donor-quencher library. (a) Structure of donor-quencher substrate library. (b) Results of prostatic acid activity in the 1 position-fixed P1'–P4' substrate libraries. Each graph represents a position in the P1'–P4' sub-

strate sequence with an 'O' representing the fixed position and an 'X' representing variable positions. The identity of the amino acid in the fixed position is indicated on the x-axis. The y-axis is the reaction velocity in RFU s⁻¹. (Taken from [27]).

prime side positions and biasing for cleavage after the internal Arg amino acid (Figure 11.6). The results of prostatic acid in the library showed no activity for P1'-Ile and helped explain the inability of the enzyme to autoactivate through processing of its prodomain at the sequence Pro-Gln-Ala-Arg▼Ile-Thr-Gly-Gly [27].

Another interesting application of PSSCLs to protease profiling was reported by Mason et al. for the identification of substrate specificity determinants of two deubiquitin hydrolase (DUB) family members, isopeptidase T (isoT) and UCH-L3 [28]. DUBs recognize and cleave the isopeptide bond created when the C-terminus of ubiquitin is conjugated to the ϵ -amine of a lysine residue. A PSSCL was synthesized based on a C-terminal fragment of ubiquitin, Gly-Gly-Arg-Leu-Arg-Leu-Val-Leu-phenyl-¹²C₆ isocyanate, conjugated to the ϵ -amine of the lysine in a one-position fixed positional scanning library of the sequence Ac-Arg-Leu-Xaa₄-Xaa₃-Xaa₂-Xaa₁-Lys-Gln-Leu-Glu-Asp-Gly-Arg. The kinetic cleavage rate of the branched peptide library by DUBs was quantitatively determined by monitoring the appearance of free Gly-Gly-Arg-Leu-Arg-Leu-Val-Leu-phenyl-¹²C₆ isocyanate as compared to

the ^{13}C -labeled internal standard Gly-Gly-Arg-Leu-Arg-Leu-Val-Leu-phenyl- $^{13}\text{C}_6$ isocyanate by mass spectrometry. This study showed that DUBs indeed show distinct specificity for their substrates as demonstrated by the 10-fold preference of UCH-L3 for basic amino acids over other amino acids in the Xaa₃ position.

Another method that has been used quite extensively to address the prime side specificity of proteases is Edman degradation of peptide mixtures [29–32]. For instance, Cantley's lab has demonstrated iterative design and synthesis of peptide substrates coupled with Edman degradation to determine the optimal substrate specificity of proteases. In particular, a primary dodecamer peptide library is synthesized of the form acetyl-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-amide, where Xaa represents a degenerate position consisting of all natural amino acids except cysteine. To ensure equimolar incorporation of all 19 amino acids, the concentrations of the individual amino acids are adjusted to compensate for unequal coupling rates as determined by Ostresh et al. [33]. Only the substrates that have been cleaved by the protease will have free N-termini and the whole mixture can be sequenced and the amounts and identities of the prime side amino acids (i.e. C-terminal to the cleaved bond) can be quantitated using Edman degradation. One cycle of Edman degradation and sequencing reveals the preferred P1' amino acids, two cycles reveals the preference at P2' and so on. A second library can then be synthesized that examines the nonprime side specificity. The prime side optimal substrate sequence as determined by the primary library is fixed and degenerate positions are synthesized in the nonprime side positions.

The N-termini of the peptides in this library are not acylated, but rather a biotin molecule is added to the C-terminus of each peptide. Once the library is digested with the protease of interest, streptavidin is used to capture the C-termini of both the cleaved and the uncleaved peptides leaving only the N-terminal fragments of the library, which can then be sequenced using Edman degradation. Using this approach the authors were able to profile multiple metalloproteinases including matrix metalloprotease-2 (MMP-2) [32] and the anthrax lethal factor metalloprotease [34].

One shortcoming with this iterative library design strategy is that the substrates with cleavable sequences may be relatively rare and therefore may not give a significant signal over the background in the primary library. To overcome this, the authors suggest the design of biased libraries where some of the positions in the substrate are held constant, for example a caspase library would have the sequence acetyl-Asp-Glu-Val-Asp-Xaa-Xaa-Xaa-Xaa-Ala-Lys-Lys-amide to bias for cleavage between the Asp-Xaa bond. This second approach requires a priori knowledge of some substrate specificity information for the protease of interest. Another potential drawback that may complicate the execution of this strategy is the potential for misdirected cleavage. This can be especially true for serine proteases of the chymotrypsin fold, where the majority of the substrate specificity determinants are in the nonprime rather than the prime side position of the substrate and therefore fixing the prime side positions may be insufficient to direct the cleavage register of the substrate.

11.2.2

Solid Support-based Synthesis and Screening of Peptide Libraries

The fundamental advantage of the mixture-based approach arises from the fact that substrates are synthesized and assayed as mixtures and this allows for the examination of a large amount of substrate space in a minimal number of assays. However, this also inevitably leads to the fundamental shortcoming of the mixture-based approach: the results of the screen represent an average of the activity on the multiple substrates in the pool. Combinatorial libraries that rely on the synthesis and screening of pools of substrates rather than individual substrates trade off the reduction in the effort of preparing and screening the library for the increase in the effort needed to interpret, deconvolute and validate the results of the screen. In many cases this trade-off is a favorable one because the preparation of large numbers of individual substrates is a daunting task. However, as parallel synthesis techniques improve for single substrate synthesis [35, 36], the next challenge will be to increase the capacity to assay individual substrates.

Inspired by DNA chips, microarrays have been successful in miniaturizing an ever-widening range of protein- and peptide-based assays (reviewed in [37]). An interesting application to protease substrate microarrays involved the exploration of the subsite specificity of the outer membrane protease T (OmpT) from *E. coli* by Dekker et al. [38]. In this study, spot synthesis [39] was used to produce an 80-peptide donor-quencher array on 8 mm spots of cellulose paper that had been derivatized with a polyoxyethylene glycol (PEG) linker. The peptides in the library were of the form acetyl-Dap(dnp)-Ala-Arg-Arg-Ala-Lys(Abz)-Gly-PEG where each position in the P2–P2' sequence Ala-Arg-Arg-Ala was systematically replaced with one of the 20 naturally occurring amino acids. Upon incubation of the cellulose-bound peptides with OmpT, the peptide spots that were recognized and cleaved by OmpT increased in fluorescence signal due to the removal of the N-terminal portion of the peptide that contained the quenching dinitrophenyl group. The length of the PEG linker, which corresponded in length to approximately 200 carbon–carbon bonds, was shown to be essential for OmpT activity on the cellulose-bound substrates. Direct coupling of the substrates to cellulose support produced substrates that were active against trypsin but were presumably not accessible to the larger detergent-solubilized OmpT protease. The results from the spot assay revealed a broad tolerance for amino acids at the P2 position, Lys>Arg at the P1 position, Lys>Ile>R in the P1' position, and Ile=Val>Ala in the P2' position of the substrate. While the general features of optimal and suboptimal substrate preference are present in the results of the cellulose-bound peptide array, the relative substrate specificity observed on the array does not correlate well with the $k_{\text{cat}}/K_{\text{m}}$ determined on the select number of substrates that were tested in solution. For example, based on the results from the cellulose-bound peptide array, P2'-Ile should be preferred over P2'-Ala, however the $k_{\text{cat}}/K_{\text{m}}$ for the P2'-Ile substrate showed a greater than 1000-fold decrease in activity over the P2'-Ala substrate. The authors speculate that this

may be due to a dominant influence of k_{cat} versus $k_{\text{cat}}/K_{\text{m}}$ for substrates immobilized on the cellulose array. Another complicating factor that could contribute to inconclusive results is the lack of register for the site of cleavage within the donor-quencher peptide.

Another substrate microarray approach was reported by Salisbury et al. on the synthesis and evaluation of microarrays of coumarin substrates in a miniaturized format representing 800 data points in an area less than 1.7 cm² [40]. In this study, they linked individually synthesized alkoxyamine-functionalized coumarin substrates to 4-carboxybenzaldehyde derivatized BSA-covered glass slides via a chemoselective oxime ligation. A spatially arrayed collection of substrates was produced where the P4 position in the substrate was held constant as alanine, P1 was constant as lysine, and P2 and P3 consisted of 19 amino acids for a total of 361 individual substrates. Upon treatment of the array with a protease, susceptible substrates are cleaved at the peptide anilide bond thus allowing detection of the now highly fluorescent coumarin linked to the array. The relative $k_{\text{cat}}/K_{\text{m}}$ for the individual substrates can be determined based on the amount of fluorescence detected for each of the substrates on the array. In this assay, thrombin showed a preference for P2 proline and broad tolerance of multiple amino acids at the P3 position. To address the correlation of solid phase cleavage versus solution-based cleavage, four single substrates were assayed in solution and compared with the microarray results. In this case, the relative ranking of substrate cleavage on the solid-surface microarray was predictive of substrate cleavage in solution, but distinct differences were observed in suboptimal substrates. For example, Ac-Ala-Phe-Ser-Lys-ACC was shown to be hydrolyzed in solution but not on the microarray.

While these microarray techniques for monitoring protease substrate specificity solve the issue of assay miniaturization, they still require greater effort for the parallel synthesis of individual substrates than do combinatorial synthesis schemes. Methods that combine the use of solid-phase based portion-mixing (also known as split and mix) with the ability to detect single substrates have proven to be powerful tools [41–43]. One example of the portion-mixing solid-phase assay approach has been developed by Meldal et al. [43] through the use of a bead-based intramolecular fluorescence quenching assay to select for optimal peptide substrates. An important aspect of this solid phase-based assay is the development of biocompatible solid supports, such as polyethylene glycol-poly-(*N,N*-dimethylacrylamide) copolymer (PEGA) resin, that allow for the libraries to be screened directly on the beads. The general approach for this strategy requires that the peptide be flanked by an internally quenched fluorescence pair, typically Abz as the fluorescent donor on the C-terminal portion of the peptide nearest the bead linker, and 3-nitro-tyrosine as the fluorescent quencher located near the N-terminal portion of the peptide. The synthesis of an exhaustive set of tetrapeptide sequences containing each of the 20 amino acids at all four positions using portion-mixing requires only 84 synthesis steps (20×4 coupling steps + 4 deprotection steps), whereas a two-position fixed PSSCL of 20 amino acids requires 19200 synthesis steps (400×6×4 coupling steps +

400×6×4 deprotection steps), and finally the synthesis of all 160 000 individual substrates requires 1280 000 steps (160 000×4 coupling steps + 160 000×4 deprotection steps).

Following synthesis, each bead contains a unique peptide sequence that upon incubation with the protease of interest can be cleaved by removing the N-terminal 3-nitro-tyrosine quencher, thus causing the bead to be fluorescent when placed under UV light. The fluorescent beads are then manually chosen and separated from the nonfluorescent beads. The peptide identity is identified directly from the bead using Edman degradation. Hydrolysis of the peptide on the bead is terminated before it is complete, typical values are 20–80% total hydrolysis of the peptide, therefore a portion of the uncleaved peptide remains and both the full-length sequence and the cleavage site within the peptide can be identified by this method. The ratio of cleaved to uncleaved peptide is then used to quantify an approximate rate of hydrolysis for each substrate. Because this is a selection rather than a screening method, the statistical analysis of preferred amino acids in the subsites requires the selection and sequencing of a large number of beads.

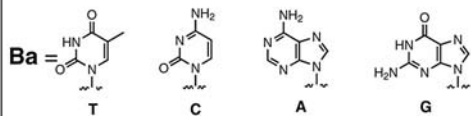
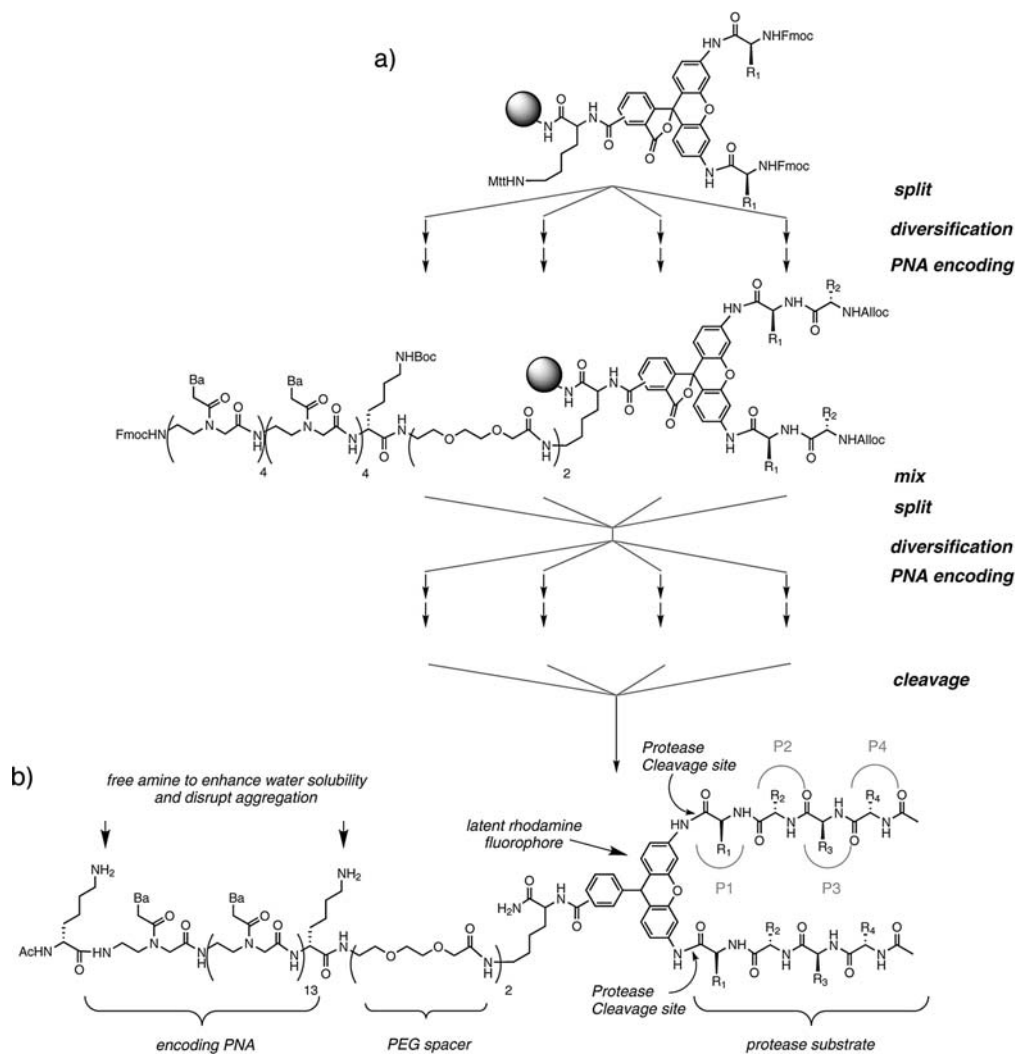
This method has been applied to many proteases, including subtilisin Carlsberg, cruzain, and papain [44–46] and, in general, it identifies good substrate sequences that can then be used to further characterize the enzyme. However, as with other solid-phase selection techniques, a direct correlation is not always seen between the activity of substrates on the solid support and the activity of the substrates in solution. Several reasons could account for this discrepancy, the simplest being that a more extensive set of beads would need to be selected and sequenced to improve the statistical correlation and be representative of the true substrate specificity of the protease. Another potential reason for discrepancies is that additional factors may influence the interaction of the protease with the support-bound substrate. This was indeed observed when screening papain against a library of 20 natural amino acids. The majority of the sequences selected from the assay contained acidic amino acids at P4', P5', and P6' [45]. However, the equivalent preference of activity of these peptides did not translate when the peptides were tested in solution. The authors determined that this was an artifact and likely due to an increased local concentration of the enzyme for beads with negatively charged residues because of the positive charge of papain at the pH of the assay. The authors followed this up by designing a library that was absent of negatively charged amino acids. These problems notwithstanding, bead-based methods remain a powerful means of identifying the substrate specificity of proteases.

The major complicating factor for methods that utilize solid support assay techniques is that the potential interaction of the protease to the solid support may bias the interpretation of the substrate specificity information and may not reflect the true substrate specificity of the protease. To address this issue and combine the advantages of portion mixing synthesis with solution-based screening methods, Winssinger et al. recently reported the use of peptide nucleic acid (PNA)-encoded rhodamine substrates [47]. The PNA-encoding strategy is ideal

for the generation of peptide substrate libraries as it allows for the rapid generation of libraries by split and mix synthesis. The substrates are synthesized using a bifunctional linker to the solid support with orthogonal protecting groups that allow for cosynthesis of the peptide and PNA chains (Figure 11.7). The synthesis of the reported library of 192 substrates relied on the alternating use of Fmoc-protected PNA monomers and Alloc-protected amino acids [48]. Proteolysis of the bond connecting the peptide substrate to the latent rhodamine fluorophore changes the electronic properties of this fluorophore, resulting in a large increase in fluorescence that can be detected at $\lambda_{\text{ex}}=488$ nm and $\lambda_{\text{em}}=530$ nm (Figure 11.8) [49]. Proteolysis is measured by incubating the PNA-encoded library of substrates with the protease or biological sample of interest. The fluorescence signal from the library is resolved by hybridization to an oligonucleotide microarray. Location of hybridization is dictated by the sequence of the PNA that is linked to the sequence of the substrate peptide (Figure 11.8).

While other protease microarray methods have been reported [38, 40], the PNA-encoding system has the advantage that the proteolysis can be carried out in solution. This is important in order to exclude the effects of nonspecific interactions of the enzymes with the surface and offers better control of substrate/analyte concentration and reaction condition. While only demonstrated on a limited set of 192 tetrapeptide substrates, experiments with the purified enzymes thrombin and caspase 3 established that the microarray method produced reliable results for the determination of protease substrate specificity. The activity profile of blood samples from healthy donors was compared to blood samples from an individual treated with warfarin, a frequently used drug for oral anticoagulant therapy. More significantly, the PNA-encoded substrate libraries were applied to clinical blood samples to compare protease activity of patients on warfarin therapy. Warfarin decreases the level of active prothrombin, factor X, IX, VII, and protein C by blocking their vitamin K-dependent carboxylation of glutamic acid moieties that is essential for their synthesis and function [50]. The normal blood sample showed a profile Arg in the P1 position and Pro in the P2 position – a profile indicative of thrombin activity. A loss of this thrombin profile was observed in blood from patients treated with warfarin, indicating a decrease in thrombin-mediated signal transduction.

One limitation of the PNA-encoded substrates is that detection of proteolysis is dependent on the interaction of the encoded PNA oligomer with the DNA oligomer attached to the microarray. Our knowledge of the properties that define PNA:DNA and PNA:PNA interactions is currently incomplete and requires increased understanding to aid in the design of large substrate libraries. Furthermore, development of cell-permeable probes will be required to capture many dynamics of protein function.



R₁	R₂	R₃	R₄
Asp (D) = CCGT	Phe (F) = AGC	Asp (D) = AGT	Asp (D) = CAA
Arg (R) = GGGT	Val (V) = CGA	Arg (R) = ACA	Arg (R) = TAG
Leu (L) = GGCA	Thr (T) = CAG	Thr (T) = GTA	Nle (n) = CTT
	Pro (P) = GAC	Pro (P) = CAA	Pro (P) = AGT

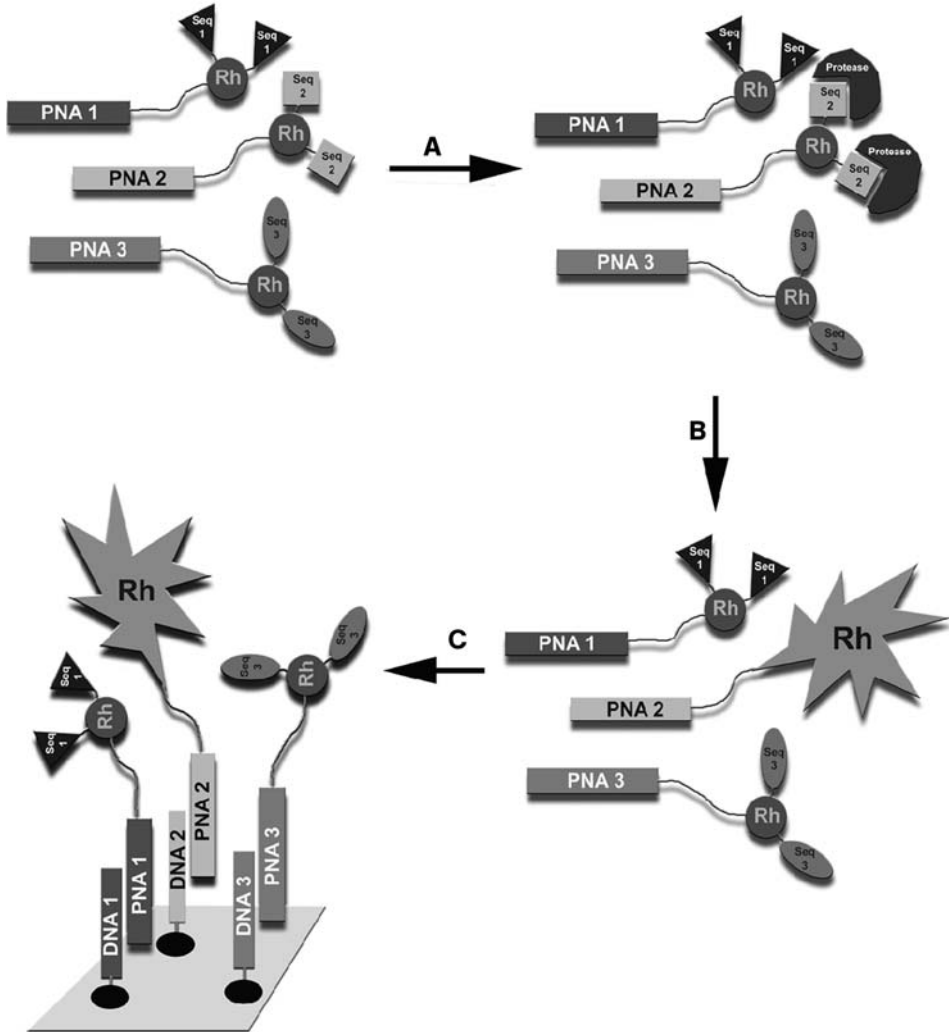


Fig. 11.8 Profiling with the PNA-encoded rhodamine-substrate library. A: Following split and mix synthesis, the rhodamine substrate library is present in solution as a mixture. B: Incubation of substrates with protease or lysate releases selected peptides

from rhodamine scaffold, thus increasing the intrinsic fluorescence of the rhodamine group. C: Active substrates are spatially resolved through hybridization to complementary oligonucleotides on a microarray and visualized with fluorescence detection.

Fig. 11.7 Peptide nucleic acid (PNA)-encoded rhodamine substrates. (a) Synthesis of the PNA-encoded rhodamine library using split and combine chemistry. (b) General structure of the 192 member PNA-encoded

rhodamine substrate library where each of the amino acids in the peptide sequence on the rhodamine scaffold is encoded by a specific sequence of PNA monomers as shown in the lower boxes. (Adapted from [47]).

11.2.3

Genetic Approaches to Identifying Protease Substrate Specificity

The use of peptide libraries displayed on the surface proteins of filamentous bacteriophage has become a widely adopted technique for optimizing peptide or protein sequences [51]. The major advantage of bacteriophage display is that the protein or peptide displayed on the surface of the bacteriophage is directly linked to the genetic material of the phage. Therefore, the identity of the protein sequence on the surface of the phage can be determined by simply sequencing the phage DNA. The application of phage display peptide libraries to protease substrate optimization was first developed by Matthews and Wells [52]. The typical approach to making phage display libraries is the insertion of randomized sequences into protein pIII (gene III product) followed by a protein or peptide sequence that can be used as an affinity tag. Upon incubation of the phage library with the protease of interest, the phage that are displaying substrate sequences susceptible to cleavage by the protease will lose their affinity tag. The phage displaying sequences refractory to cleavage by the protease will retain their affinity tag and can be trapped preferentially by binding to an affinity matrix. The phage with susceptible sequences can be used to infect *E. coli* cells and be amplified for another round of substrate selection or isolated for DNA sequencing of individual clones (Figure 11.9).

The substrate specificity of multiple proteases has been determined using substrate phage display and typically results in the identification of substrates with good catalytic efficiencies [53–59]. For example, Coombs et al. reported the identification of the substrate consensus sequence Ser-Ser-(Tyr or Phe)-Tyr▼Ser-(Gly or Ser) for the serine protease prostate-specific antigen (PSA) [55]. They reported k_{cat}/K_m values as high as $3100 \text{ M}^{-1} \text{ s}^{-1}$ as compared to $2.0 \text{ M}^{-1} \text{ s}^{-1}$ for the sequence Gln-Gln-Leu-Leu▼His-Asn found in semenogelin, a physiological substrate of PSA.

Tenzer et al. reported the use of subtractive substrate phage display to identify specific proteolytic activity in the cytosol of cancer cells in response to treatment with an antiproliferative regiment consisting of staurosporine and ionizing radiation [60]. In this library, a randomized gene encoding seven amino acids and a C-terminal 6×His affinity tag were fused within the T7 phage capsid protein Xb. Fusion within this phage protein allows for the expression of 10 copies of the putative substrate on the surface of the phage. The phage were then bound to a Ni^{2+} affinity matrix and were first incubated with cytosolic extracts from untreated cells to release the treatment-independent substrates from the matrix. The same matrix was then treated with cytosolic extract from cancer cells treated with the antiproliferation regiment, releasing the treatment-dependent proteolytic substrates.

While differences in substrate sequences were observed between treated and untreated cell lysates, there was a lack of agreement between the substrate sequences in two independent experiments. Despite this, several sequences such as Trp-Ala-Val-Glu-Ile-Asp-Phe and Ser-Glu-Ala-Gly-Ile-Ser-Ala identified from

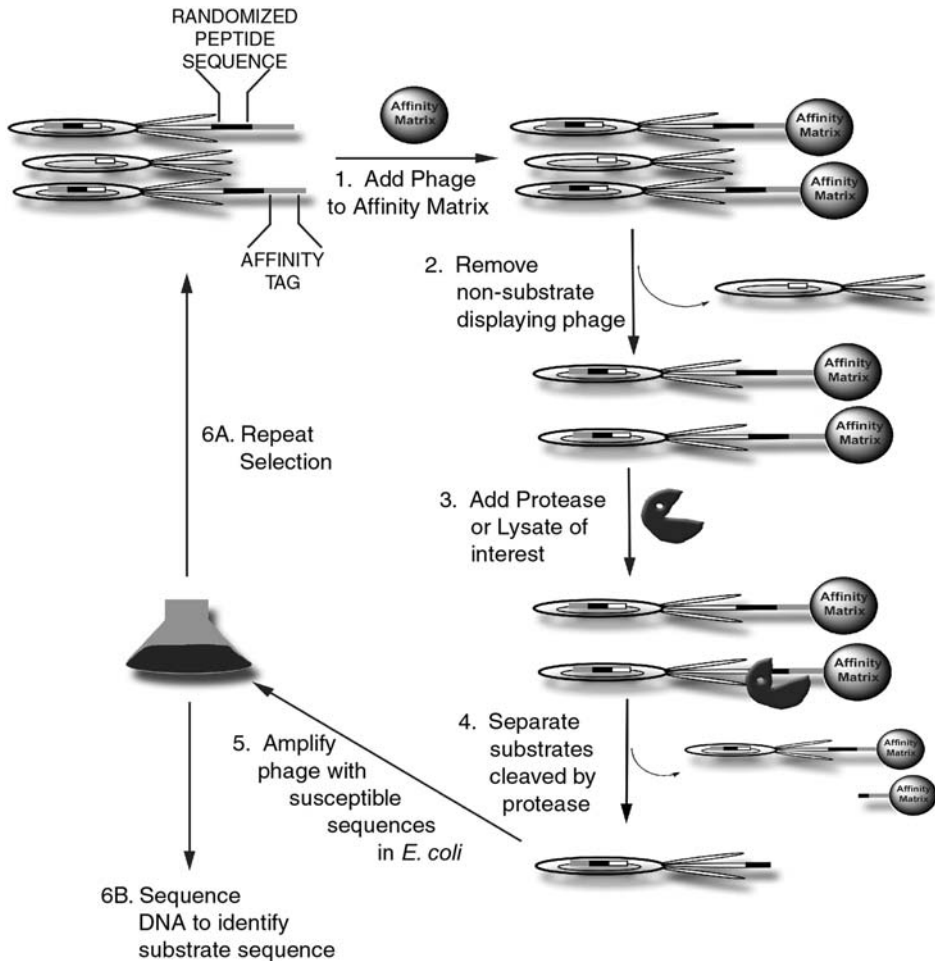


Fig. 11.9 Schematic representation of substrate phage.

one of the experiments were synthesized and further investigated. While no difference in cleavage from cytosolic extracts of treated versus untreated cells was observed for the Ser-Glu-Ala-Gly-Ile-Ser-Ala peptide, a significant increase in cleavage of Trp-Ala-Val-Glu-Ile-Asp-Phe for treated versus untreated cells was observed. The authors went on to show that this substrate was likely being cleaved by a caspase that was activated upon treatment of the cells with the antiproliferative regimen.

The major limitation to substrate phage is that the size and diversity of the phage library are ultimately determined by the transformation efficiency and biological constraints of the *E. coli* host. If only a small fraction of the possible sequences can be sampled in the selection, then the optimal sequence may be missed. By design, substrate phage is a selection rather than a screening pro-

cess so typically only the optimal substrates are identified and information as to the suboptimal sequences can be inferred, but ultimately remain unaddressed with this method. The catalytic alignment of the cleavage site can also be cryptic and often requires the incorporation of known specificity determinants to bias for cleavage at a particular site or the iterative production and testing of multiple phage libraries [58, 59]. As with the solid-phase approaches, substrate optimization by phage display can also be subject to artifactual results due to the context of the peptide sequence within the phage construct. Furthermore, as elaborated by Sharkov et al., under typical selection conditions the enzyme is present at nanomolar concentrations, whereas the phage particles are typically present at low picomolar concentrations [61]. Under these conditions the kinetics of hydrolysis are more representative of a single turnover reaction rather than steady state conditions and therefore may not enrich for substrates with increased $k_{\text{cat}}/K_{\text{m}}$. Indeed, several studies have shown a lack of agreement between substrate phage-enriched clones and the k_{cat} and K_{m} values of the peptides derived from the clones [55, 62–64].

11.3

Identification of Macromolecular Substrates

Perhaps the most useful information to know about a protease is the identity of the downstream proteins that it cleaves. By knowing the physiological substrates of a protease, one can more completely understand the pathways in which the protease is operating and determine the extent of impact that modulating the protease activity would have on a therapeutic or pathological process. Information from synthetic or genetic peptide libraries gives a rapid overview of the proximal substrate cleavage site that can provide insight into the determinants of the substrate binding site as well as sensitive tools for assaying the protease. While the information from the substrate specificity of a protease can be used to identify a list of potential substrates through the bioinformatic examination of the proteome [65], much effort is still needed to validate the proteins as actual substrates of the protease. Indeed, the determinants of the proximal substrate cleavage site of a protease identified by *in vitro* methods may not accurately represent the determinants for macromolecular substrates that a protease will encounter in a biological setting. The traditional biochemical approach to identifying protease substrates has been the serial candidate approach that is often tedious and likely to only yield partial or inadequate identification of protease substrates. Even when a protease is shown to cleave a protein, the functional significance of that cleavage may be poorly understood. Direct identification of protease substrates on a genome-level in a biological context is still one of the major challenges in protease biology.

11.3.1

Genetic Approach to the Identification of Macromolecular Substrates

Genetic modification of protease function has been an important approach to identifying downstream protease substrates and furthering our understanding of the biological role of specific proteases. Typically, loss of function mutations in proteases results in the accumulation of specific proteins that lead to the identification of those proteins as substrates for the protease and confirmation that the protease is acting in the pathway. For example, the familial disease haemophilia A is due to loss-of-function mutations in factor VIII that lead to diminished proteolytic activity of factor IXa [66]. Human hereditary disorder for early onset of Alzheimer's disease can be linked to gain-of-function mutations in presenilins-1 and -2 [67, 68] as well as mutations in the amyloid precursor protein substrate itself [69].

Targeted deletion of proteases in mice has also proven useful in the identification and validation of protease substrates [70–73]. For example, glucagon-like peptide-1 (GLP-1(7–36)-amide) is a gut hormone that enhances the glucose-dependent secretion of insulin from pancreatic β -cells [74]. GLP-1(7–36)-amide is rapidly degraded to an inactive protein by the removal of the two N-terminal amino acids to generate GLP-1(9–36)-amide. Marguet et al. generated mice deficient in dipeptidyl peptidase IV and showed that there was an increase in the level of circulating GLP-1(7–36)-amide with a consequent increase in the insulinotropic effect [75].

This knockout strategy for the identification of substrates can be time consuming and because of the existence of compensatory or redundant pathways may sometimes lead to inconclusive results. The question usually remains as to whether a particular substrate is a direct or indirect substrate of the genetically targeted protease because many proteases operate in cascades and activation of downstream proteases may in fact be responsible for the substrate cleavage. Also, unless an obvious phenotype is observed that can guide the researcher to examine candidate proteins as potential substrates, identifying the *in vivo* substrates by this approach can be tedious and may overlook other relevant substrates. This method is good for confirming hypotheses about the protease and substrate interactions, but rarely are new substrates identified. Methods that allow for the proteome to be queried in a systematic fashion would greatly complement *in vivo* genetic approaches.

One technology that is emerging as a tool to rapidly generate and validate proteases and protease substrates in biology is RNA interference (RNAi) [76–79]. For example, Li et al. used RNAi to demonstrate the role of the de-ubiquitinating protease HAUSP in regulating the antiproliferative effects of the p53 tumor suppressor protein [79]. The ubiquitin ligase Mdm2 is responsible for ubiquitinating p53 thus targeting p53 for destruction by the proteasome. HAUSP can reverse the destruction of p53 through direct binding and de-ubiquitination of the protein. In support of this interaction, a partial reduction of HAUSP in cells by HAUSP-directed RNAi led to the destabilization of p53. When a more

complete ablation of HAUSP was achieved through amplified treatment of cells with RNAi, the unexpected result of p53 stabilization was observed. The authors were able to show that Mdm2 was also a substrate of HAUSP and that when levels of HAUSP were completely reduced, Mdm2 was destabilized and no longer able to ubiquitinate p53, thus leading to the stabilization of p53. This study illustrates the subtleties in the dynamics of protease–substrate interactions that can be observed with tools like RNAi but that would be completely overlooked by targeted deletion of the gene.

A method that attempts to address cleavage of all proteins in a proteome in an unbiased manner is small pool cDNA expression screening originally developed by Lustig et al. [80] (Figure 11.10). In this method, a cDNA library is constructed from the cells, tissues or organisms. The clones from a cDNA library are combined in pools of approximately 10–100 clones per pool. The pools of cDNA are then transcribed/translated *in vitro* in the presence of a labeling agent, typically [³⁵S]methionine. The labeled protein pools are then incubated with the protease of interest and subsequently separated on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Pools that show a shift in the labeled proteins with treatment of the protease are transformed into *E. coli* and pooled in less complex pools of approximately 1–10 clones per pool. The *in vitro* transcription/translation and cleavage assay is then iterated until the single cleavage susceptible clone is isolated. The cDNA is then simply sequenced to reveal the identity of the cleaved protease substrate.

Using this method Kothakota et al. were able to systematically investigate 1000 cDNA pools of 100 clones per pool (a total of 100 000 clones) to identify substrates of caspase-3 [81]. In three different pools a 65 kDa protein was reduced to a 48 kDa protein. Deconvolution of the pool and sequencing of the clone revealed the protein to be the partial sequence of gelsolin from residues 142–731. Based on the *in vitro* determined substrate specificity of caspase-3 [19], the cleavage site in gelsolin was identified as Asp-Gln-Thr-Asp▼Gly. The functional significance of gelsolin cleavage during apoptosis was then examined and shown to play an important role in effecting the morphological changes of apoptosis through the severing of actin filaments. Additional confirmation of this role was identified by a delayed onset of membrane blebbing and DNA fragmentation in neutrophils isolated from gelsolin-deficient mice.

While this method provides a rapid and direct approach to the identification of downstream targets of proteases, there are some limitations. First, the output of the method is only as good as the quality of the cDNA library input – partial sequences lacking an initiation site or empty sequences will decrease the diversity of the cDNA library. Also, abundant transcripts may be overrepresented in the collection and rare transcripts may not be represented at all. Hence, not all proteins will be reliably expressed in the *in vitro* transcription/translation mixture – proteins may require membrane insertion, posttranslational modifications, or the presence of cofactor or adaptor proteins. All of these factors can contribute to a decrease in representation in the screen and indeed, only 10–40% of the clones in a pool will typically result in labeled proteins [82]. Also, as

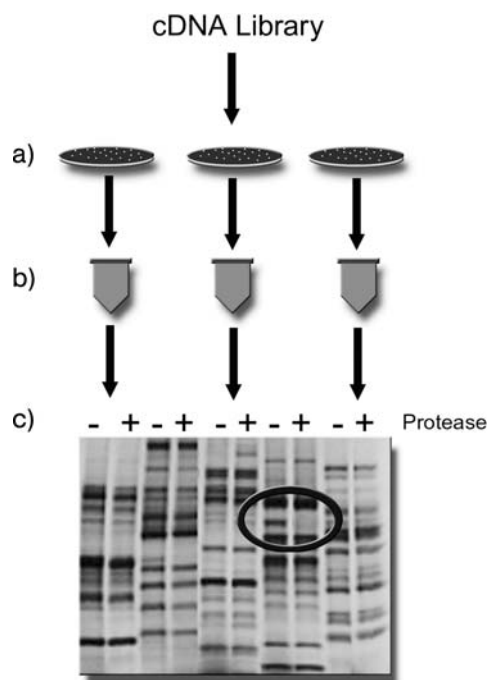


Fig. 11.10 Schematic representation of small-pool cDNA expression screening. (a) The cDNA library of interest is transformed into *E. coli* to produce single clones. (b) The DNA from approximately 100 clones are pooled into a single vial and the encoded protein is produced by *in vitro* transcription and translation in the presence of [³⁵S]methionine. (c) The protein mixtures are then incubated with or without the protease of interest. The proteins are then resolved

on SDS-PAGE and visualized by autoradiography. The cDNA pools that correspond to lanes that show a disappearance or appearance of bands upon incubation with the protease (an example is circled) are again transformed into *E. coli*. The process of *in vitro* transcription translation and protease incubation is repeated with the single clones to isolate the putative protease substrate. The protease substrate is identified by sequencing the isolated cDNA clone.

with most other proteomic techniques, cleavage of the identified protein may not be due to direct cleavage of the protease of interest, but rather through the activation of a downstream protease in the *in vitro* transcription/translation lysate.

Another genetic approach that attempts to identify macromolecular substrates is exosite scanning introduced by Overall et al. An exosite is a domain or site that is peripheral to the primary catalytic site of a protease, but nonetheless can be an important determinant of substrate specificity [83, 84]. To identify new substrates of MMP-2, McQuibban et al. [85] fused the exosite domain of MMP-2 to the β -Gal activator domain to use as bait in a yeast two-hybrid screen. The screen resulted in the identification of monocyte chemoattractant protein-3 (MCP-3) as a protein that interacted with the exosite domain of MMP-2. Further

investigation using full-length MMP-2 showed that MCP-3 is indeed cleaved by MMP-2 at a rate that exceeds that of gelatin cleavage. MMP-2 cleavage of MCP-3 was shown to be functionally relevant as a broad-spectrum CC-chemokine receptor antagonist in *in vitro* assays. Mouse models of inflammation further supported this role with the observation that MMP-2-cleaved MCP-3 protein significantly eliminated monocyte infiltration in a dose-dependent manner. While exosite scanning has been shown to facilitate the identification of substrates for some proteases, the method is not generally applicable to all proteases. The method is currently limited to proteases that contain identifiable exosite domains and the fundamental limitations of the yeast two-hybrid approach also apply.

11.3.2

Proteomic Approaches to Identifying Protease Substrates

In contrast to the genetic approach of exosite scanning for identifying protein substrates through binding interactions, several groups have taken a similar proteomic approach to identify substrates using protease catalytic domains with active site mutations [86, 87]. For example, Beresford et al. used affinity chromatography with an inactive variant of the serine protease granzyme A, the active site Ser was mutated to Ala, to isolate potential substrates of granzyme A [86]. Granzyme A is a serine protease that is released from the granules of cytotoxic lymphocytes to facilitate the elimination of tumor and virus-infected cells through a caspase-independent mechanism. Using the affinity chromatography approach the authors were able to identify the endoplasmic reticulum (ER)-associated 270–450 kDa SET complex that among other proteins contained tumor suppressor NM23-H1 which was shown to be a granzyme A-activated DNase [86, 88, 89]. Identification of the substrates of granzyme A provided an increased understanding of the molecular mechanism that granzyme A uses to eliminate tumor and virus cells that evade caspase-mediated apoptosis.

The major limitation of this approach is that by removal of the active site residue(s) the interaction of the protease with the substrate relies only on the ground state binding, approximated by K_m , and does not take into account the catalytic step, approximated by k_{cat} . Specificity of an enzyme for a substrate is dependent on the ability of the enzyme to utilize substrate-binding energy to promote catalysis, k_{cat}/K_m . By only relying ground state binding to identify substrates, efficiently catalyzed substrates could easily be overlooked in favor of tight binding substrates that may be less efficiently hydrolyzed by the protease.

An interesting solution to the identification of protease substrates is the use of selective tagging of proteolyzed substrates. Wang et al. suggested the use of peptide nucleophiles for identifying substrates of serine and cysteine proteases [90]. Serine and cysteine proteases undergo a two-step mechanism for the hydrolysis of protein substrates. A stable acyl- or thioacyl-enzyme intermediate is formed with the N-terminal portion of the substrate. In this method, a peptide nucleophile tag is designed in which the peptide portion binds to the prime

side subsites of the enzyme, placing the nucleophile in position to attack the acyl–enzyme intermediate, thereby incorporating the peptide nucleophile tag to the C-terminus of the N-terminal portion of the substrate. Using biotin in the nucleophile tag allows for selective isolation of the substrates using immobilized streptavidin. The authors showed that this method was effective for labeling bovine serine albumin cleavage products of the cysteine protease cathepsin S and the serine protease human neutrophil elastase. They also showed that pro-interleukin-1 β could be selectively labeled by caspase-1.

An obvious limitation to the method includes the requirement for design of selective prime side peptide nucleophile tags so that only substrates for the protease of interest are tagged. Also, since the nucleophile tag is competing with water only a small percentage of the substrate will be tagged, leading to potential issues of detection. While this approach is promising, the general application of this method to identifying novel substrates in complex biological systems has yet to be demonstrated.

Traditional proteome analysis techniques couple high-resolution protein separation with protein identification and quantification. Typical approaches for protein separation include two-dimensional PAGE, high-performance liquid chromatography, and capillary electrophoresis. Protein identification and quantification are usually achieved by mass spectrometry of stable isotope-incorporated protein mixtures or Edman degradation-based sequencing. These powerful proteomic approaches are beginning to be applied to the challenging field of enzyme substrate identification [65, 91, 92]. Of particular interest, Tam et al. reported the use of an isotope-coded affinity tag (ICAT) coupled with multidimensional liquid chromatography and mass spectrometry to identify proteins shed from the surface of MDA-MB-231 breast cancer cells upon transfection with membrane type 1-matrix metalloproteinase [93]. Proteins that differed in size or abundance between the transfected and untransfected cells were identified using tandem mass spectrometry. While this is still an emerging field, these techniques offer promise for identifying the natural substrates of proteases thereby more readily establishing the physiological function of proteases in biology.

11.4 Conclusions

Multiple tools have been developed or are in the process of being developed to facilitate the identification of protease substrates in terms of both identifying the optimal substrate specificity of a protease as well as identifying the physiologically relevant macromolecular substrates of a protease. The application of this knowledge can have far-reaching implications into the characterization of the protease of interest.

A practical application of substrate specificity information is in the production of sensitive substrates to monitor protease activity for screening for small molecule modulators. Moreover, in the absence of leads from screening efforts, utili-

zation of the optimal substrate specificity can be used as a starting point for inhibitor design. Indeed, it is typical for efforts in protease drug discovery to utilize the subsite substrate specificity at some point in the inhibitor development process.

Other applications of protease substrate specificity information include the visualization of specific protease activity in lysates, cells, tissues, or whole organisms. For example, near infrared (NIR) fluorochromes coupled to optimal peptide substrate sequences have been used to visualize the activity of proteases in whole animals (reviewed in [94]). Low absorbance and low autofluorescence of NIR by tissues allows light to travel 5–6 cm, making NIR well suited for imaging of intact organisms [95]. The ability to view specific proteolysis *in vivo* will enable the diagnosis of diseases and the real time assessment of drug treatment [96, 97].

Substrate specificity has also been used to selectively target drugs or retroviruses to specific tissues or sites of disease [98–101]. As an example, several groups have taken advantage of the selective upregulation of the PSA serine protease in prostate cancer to target cytotoxic drugs to the cancer cells [102, 103]. Conjugation of the cytotoxic drug doxorubicin to the optimal peptide substrate of PSA results in a drug that is not toxic unless activated by PSA cleavage. This strategy should allow for an increased therapeutic window for cytotoxic chemotherapy.

Proteases form the cornerstone of many biological processes. While much progress has been made over the century of protease research, issues of redundancy and selectivity still brand proteases as a demanding enzyme class. Systematic tools that allow for functional profiling of substrate specificity will be an important mechanism to characterize, classify, and differentiate proteases and protease families. Some of the tools for identifying optimal and suboptimal peptide substrates as well as tools for the discovery of physiological substrates have been presented.

References

- 1 A. J. Barrett, N. D. Rawlings, J. Fred Woessner (eds), *Handbook of Proteolytic Enzymes*, Elsevier Academic Press, London, 2004.
- 2 X. S. Puente, L. M. Sanchez, C. M. Overall, C. Lopez-Otin, *Nat. Rev. Genet.* 2003, 4, 544–558.
- 3 C. Southan, *Drug Discov. Today* 2001, 6, 681–688.
- 4 C. Lopez-Otin, C. M. Overall, *Nat. Rev. Mol. Cell Biol.* 2002, 3, 509–519.
- 5 P. L. Richardson, *Curr. Pharm. Des.* 2002, 8, 2559–2581.
- 6 W. M. Frederiks, O. R. F. Mook, *J. Histochem. Cytochem.* 2004, 52, 711–722.
- 7 A. Baruch, D. A. Jeffery, M. Bogyo, *Trends Cell Biol.* 2004, 14, 29–35.
- 8 N. Jessani, B. F. Cravatt, *Curr. Opin. Cell Biol.* 2004, 8, 54–59.
- 9 J. J. Perona, C. S. Craik, *Protein Sci.* 1995, 4, 337–360.
- 10 L. Hedstrom, *Chem. Rev.* 2002, 102, 4501–4523.
- 11 W. Bleulx, K. Brijs, S. Torrekens, F. Van Leuven, J. A. Delcour, *Biochim. Biophys. Acta* 1998, 1387, 317–324.

- 12 S. Butenas, M. Kalafatis, K.G. Mann, *Biochemistry* **1997**, *36*, 2123–2131.
- 13 H.R. Stennicke, M. Renuatus, M. Meldal, G.S. Salvesen, *Biochem. J.* **2000**, *350*, 563–568.
- 14 C.A. Tsu, C.S. Craik, *J. Biol. Chem.* **1996**, *271*, 11563–11570.
- 15 D.J. Maly, L. Huang, J.A. Ellman, *ChemBioChem* **2002**, *3*, 16–37.
- 16 C. Pinilla, J.R. Appel, P. Blanc, R.A. Houghten, *BioTechniques* **1992**, *13*, 901–902, 904–905.
- 17 E. Apletalina, J. Appel, N.S. Lamango, R.A. Houghten, I. Lindberg, *J. Biol. Chem.* **1998**, *273*, 26589–26595.
- 18 T.A. Rano, T. Timkey, E.P. Peterson, J. Rotonda, D.W. Nicholson, J.W. Becker, K.T. Chapman, N.A. Thornberry, *Chem. Biol.* **1997**, *4*, 149–155.
- 19 N.A. Thornberry, T.A. Rano, E.P. Peterson, D.M. Rasper, T. Timkey, M. Garcia-Calvo, V.M. Houtzager, P.A. Nordstrom, S. Roy, J.P. Vaillancourt, K.T. Chapman, D.W. Nicholson, *J. Biol. Chem.* **1997**, *272*, 17907–17911.
- 20 M. Zimmerman, E. Yurewicz, G. Patel, *Anal. Biochem.* **1976**, *70*, 258–262.
- 21 J.L. Harris, B.J. Backes, F. Leonetti, S. Mahrus, J.A. Ellman, C.S. Craik, *Proc. Natl Acad. Sci. USA* **2000**, *97*, 7754–7759.
- 22 D.J. Maly, F. Leonetti, B.J. Backes, D.S. Dauber, J.L. Harris, C.S. Craik, J.A. Ellman, *J. Org. Chem.* **2002**, *67*, 910–915.
- 23 J.L. Harris, P.B. Alper, J. Li, M. Rechsteiner, B.J. Backes, *Chem. Biol.* **2001**, *8*, 1131–1141.
- 24 W.W. Raymond, S.W. Ruggles, C.S. Craik, G.H. Caughey, *J. Biol. Chem.* **2003**, *278*, 34517–34524.
- 25 H. Petrassi, J. Williams, J. Li, C. Tumanut, B. Backes, J. Harris, Submitted for publication.
- 26 D.S. Dauber, R. Ziermann, N. Parkin, D.J. Maly, S. Mahrus, J.L. Harris, J.A. Ellman, C. Petropoulos, C.S. Craik, *J. Virol.* **2002**, *76*, 1359–1368.
- 27 A. Shipway, H. Danahay, J.A. Williams, D.C. Tully, B.J. Backes, J.L. Harris, *Biochem. Biophys. Res. Commun.* **2004**, *324*, 953–963.
- 28 D.E. Mason, J. Ek, E.C. Peters, J.L. Harris, *Biochemistry* **2004**, *43*, 6535–6544.
- 29 A.J. Birkett, D.F. Soler, R.L. Wolz, J.S. Bond, J. Wiseman, J. Berman, R.B. Harris, *Anal. Biochem.* **1991**, *196*, 137–143.
- 30 J.R. Petithory, F.R. Masiarz, J.F. Kirsch, D.V. Santi, B.A. Malcolm, *Proc. Natl Acad. Sci. USA* **1991**, *88*, 11510–11514.
- 31 D. Arnold, W. Keilholz, H. Schild, T. Dumrese, S. Stevanovic, H.-G. Rammensee, *Eur. J. Biochem.* **1997**, *249*, 171–179.
- 32 B.E. Turk, L.L. Huang, E.T. Piro, L.C. Cantley, *Nat. Biotechnol.* **2001**, *19*, 661–667.
- 33 J.M. Ostresh, J.H. Winkle, V.T. Hamashin, R.A. Houghten, *Biopolymers* **1994**, *34*, 1681–1689.
- 34 B.E. Turk, T.Y. Wong, R. Schwarzenbacher, E.T. Jarrell, S.H. Leppla, R.J. Collier, R.C. Liddington, L.C. Cantley, *Nat. Struct. Mol. Biol.* **2004**, *11*, 60–66.
- 35 B.J. Backes, J.L. Harris, F. Leonetti, C.S. Craik, C.J.A. Ellman, *Nat. Biotechnol.* **2000**, *18*, 187–193.
- 36 J.E. Sheppeck II, H. Kar, L. Gosink, J.B. Wheatley, E. Gjerstad, S.M. Loftus, A.R. Zubiria, J.W. Janc, *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2639–2642.
- 37 M.F. Templin, D. Stoll, M. Schrenk, P.C. Traub, C.F. Vohringer, T.O. Joos, *Trends Biotechnol.* **2002**, *20*, 160–166.
- 38 N. Dekker, R.C. Cox, R.A. Kramer, M.R. Egmond, *Biochemistry* **2001**, *40*, 1694–1701.
- 39 R. Frank, *Tetrahedron* **1992**, *48*, 9217–9232.
- 40 C.M. Salisbury, D.J. Maly, J.A. Ellman, *J. Am. Chem. Soc.* **2002**, *124*, 14868–14870.
- 41 A. Furka, F. Sebestyen, M. Asgedom, G. Dibo, *Int. J. Pept. Protein. Res.* **1991**, *37*, 487–493.
- 42 Y. Duan, R.A. Laursen, *Anal. Biochem.* **1994**, *216*, 431–438.
- 43 M. Meldal, I. Svendsen, K. Breddam, F.-I. Auzanneau, *Proc. Natl Acad. Sci. USA* **1994**, *91*, 3314–3318.
- 44 M.A. Juliano, E. Del Nery, J. Scharfstein, M. Meldal, I. Svendsen, A. Walmsley, L. Juliano, in *Peptides 1996, Proceedings of the 24th European Peptide Symposium, Edinburgh, Sept. 8–13, 1996*, **1998**, pp. 511–512.

- 45 P. M. St. Hilaire, M. Willert, M. A. Juliano, L. Juliano, M. Meldal, *J. Comb. Chem.* **1999**, *1*, 509–523.
- 46 P. M. St. Hilaire, L. C. Alves, S. J. Sander-son, J. C. Mottram, M. A. Juliano, L. Juliano, G. H. Coombs, M. Meldal, *ChemBioChem* **2000**, *1*, 115–122.
- 47 N. Winssinger, R. Damoiseaux, D. C. Tully, B. H. Geierstanger, K. Burdick, J. L. Harris, *Chem. Biol.* **2004**, *11*, 1351–1360.
- 48 F. Debaene, L. Mejias, J. L. Harris, N. Winssinger, *Tetrahedron* **2004**, *60*, 8677–8690.
- 49 S. P. Leytus, W. L. Patterson, W. F. Mangel, *Biochem. J.* **1983**, *215*, 253–260.
- 50 B. Furie, B. C. Furie, *Cell* **1988**, *53*, 505–518.
- 51 G. P. Smith, J. K. Scott, *Methods Enzymol.* **1993**, *217*, 228–257.
- 52 D. J. Matthews, J. A. Wells, *Science* **1993**, *260*, 1113–1117.
- 53 S.-H. Ke, G. S. Coombs, K. Tachias, M. Navre, D. R. Corey, E. L. Madison, *J. Biol. Chem.* **1997**, *272*, 16603–16609.
- 54 S.-H. Ke, G. S. Coombs, K. Tachias, D. R. Corey, E. L. Madison, *J. Biol. Chem.* **1997**, *272*, 20456–20462.
- 55 G. S. Coombs, R. C. Bergstrom, J.-L. Pellequer, S. I. Baker, M. Navre, M. M. Smith, J. A. Tainer, E. L. Madison, D. R. Corey, *Chem. Biol.* **1998**, *5*, 475–488.
- 56 Z. Q. Beck, L. Hervio, P. E. Dawson, J. H. Elder, E. L. Madison, *Virology* **2000**, *274*, 391–401.
- 57 H. R. Maun, C. Eigenbrot, R. A. Lazarus, *J. Biol. Chem.* **2003**, *278*, 21823–21830.
- 58 T. Takeuchi, J. L. Harris, W. Huang, K. W. Yan, S. R. Coughlin, C. S. Craik, *J. Biol. Chem.* **2000**, *275*, 26333–26342.
- 59 J. L. Harris, E. P. Peterson, D. Hudig, N. A. Thornberry, C. S. Craik, *J. Biol. Chem.* **1998**, *273*, 27364–27373.
- 60 A. Tenzer, B. Hofstetter, C. Sauser, S. Bodis, A. P. Schubiger, C. Bonny, M. Pruschy, *Proteomics* **2004**, *4*, 2796–2804.
- 61 N. A. Sharkov, R. M. Davis, J. F. Reidhaar-Olson, M. Navre, D. Cai, *J. Biol. Chem.* **2001**, *276*, 10788–10793.
- 62 M. M. Smith, L. Shi, M. Navre, *J. Biol. Chem.* **1995**, *270*, 6440–6449.
- 63 L. Ding, G. S. Coombs, L. Strandberg, M. Navre, D. R. Corey, E. L. Madison, *Proc. Natl Acad. Sci. USA* **1995**, *92*, 7627–7631.
- 64 L. S. Hervio, G. S. Coombs, R. C. Bergstrom, K. Trivedi, D. R. Corey, E. L. Madison, *Chem. Biol.* **2000**, *7*, 443–452.
- 65 T. Lohmueller, D. Wenzler, S. Hagemann, W. Kiess, C. Peters, T. Dandekar, T. Reinheckel, *Biol. Chem.* **2003**, *384*, 899–909.
- 66 D. J. Bowen, *Mol. Pathol.* **2002**, *55*, 1–18.
- 67 M. Citron, D. Westaway, W. Xia, G. Carlson, T. Diehl, G. Levesque, K. Johnson-Wood, M. Lee, P. Seubert, A. Davis, D. Kholodenko, R. Motter, R. Sherrington, B. Perry, H. Yao, R. Strome, I. Lieberburg, J. Rommens, S. Kim, D. Schenk, P. Fraser, P. S. G. Hyslop, D. J. Selkoe, *Nat. Med.* **1997**, *3*, 67–72.
- 68 C. De Jonghe, M. Cruts, E. A. Rogaeva, C. Tysoe, A. Singleton, H. Vanderstichele, W. Meschino, B. Dermaut, I. Vanderhoeven, H. Backhovens, E. Vanmechelen, C. M. Morris, J. Hardy, D. C. Rubinsztein, P. H. St. George-Hyslop, C. Van Broeckhoven, *Hum. Mol. Genet.* **1999**, *8*, 1529–1540.
- 69 J. Hardy, D. J. Selkoe, *Science* **2002**, *297*, 353–356.
- 70 P. Li, H. Allen, S. Banerjee, S. Franklin, L. Herzog, C. Johnston, J. McDowell, M. Paskind, L. Rodman, J. Salfeld, E. Towne, D. Tracey, S. Wardwell, F.-Y. Wei, W. Wong, R. Kamen, T. Seshadri, *Cell* **1995**, *80*, 401–411.
- 71 G. P. Shi, J. A. Villadangos, G. Dranoff, C. Small, L. Gu, K. J. Haley, R. Riese, H. L. Ploegh, H. A. Chapman, *Immunity* **1999**, *10*, 197–206.
- 72 B. M. Steiglitz, M. Ayala, K. Narayanan, A. George, D. S. Greenspan, *J. Biol. Chem.* **2004**, *279*, 980–986.
- 73 U. Sahin, G. Weskamp, K. Kelly, H.-M. Zhou, S. Higashiyama, J. Peschon, D. Hartmann, P. Saftig, C. P. Blobel, *J. Cell Biol.* **2004**, *164*, 769–779.
- 74 L. B. Knudsen, *J. Med. Chem.* **2004**, *47*, 4128–4134.
- 75 D. Marguet, L. Baggio, T. Kobayashi, A.-M. Bernard, M. Pierres, P. F. Nielsen, U. Ribel, T. Watanabe, D. J. Drucker, N. Wagtmann, *Proc. Natl Acad. Sci. USA* **2000**, *97*, 6874–6879.

- 76 G. J. Hannon, J. J. Rossi, *Nature* **2004**, *431*, 371–378.
- 77 S. Tanida, T. Joh, K. Itoh, H. Kataoka, M. Sasaki, H. Ohara, T. Nakazawa, T. Nomura, Y. Kinugasa, H. Ohmoto, H. Ishiguro, K. Yoshino, S. Higashiyama, M. Itoh, *Gastroenterology* **2004**, *127*, 559–569.
- 78 J. L. Contreras, M. Vilatoba, C. Eckstein, G. Bilbao, A. Thompson, D. E. Eckhoff, *Surgery* **2004**, *136*, 390–400.
- 79 M. Li, C. L. Brooks, N. Kon, W. Gu, *Mol. Cell* **2004**, *13*, 879–886.
- 80 K. D. Lustig, K. L. Kroll, E. E. Sun, M. W. Kirschner, *Development* **1996**, *122*, 4001–4012.
- 81 S. Kothakota, T. Azuma, C. Reinhard, A. Klippel, J. Tang, K. Chu, T. J. McGarry, M. W. Kirschner, K. Koths, D. J. Kwiatkowski, L. T. Williams, *Science* **1997**, *278*, 294–298.
- 82 V. L. Cryns, Y. Byun, A. Rana, H. Mellor, K. D. Lustig, L. Ghanem, P. J. Parker, M. W. Kirschner, J. Yuan, *J. Biol. Chem.* **1997**, *272*, 29449–29453.
- 83 B. Steffensen, U. M. Wallon, C. M. Overall, *J. Biol. Chem.* **1995**, *270*, 11555–11566.
- 84 C. M. Overall, E. Tam, G. A. McQuibban, C. Morrison, U. M. Wallon, H. F. Bigg, A. E. King, C. R. Roberts, *J. Biol. Chem.* **2000**, *275*, 39497–39506.
- 85 G. A. McQuibban, J.-H. Gong, E. M. Tam, C. A. G. McCulloch, I. Clark-Lewis, C. M. Overall, *Science* **2000**, *289*, 1202–1206.
- 86 P. J. Beresford, C.-M. Kam, J. C. Powers, J. Lieberman, *Proc. Natl Acad. Sci. USA* **1997**, *94*, 9285–9290.
- 87 J. M. Flynn, S. B. Neher, Y.-I. Kim, R. T. Sauer, T. A. Baker, *Mol. Cell* **2003**, *11*, 671–683.
- 88 P. J. Beresford, D. Zhang, D. Y. Oh, Z. Fan, E. L. Greer, M. L. Russo, M. Jaju, J. Lieberman, *J. Biol. Chem.* **2001**, *276*, 43285–43293.
- 89 Z. Fan, P. J. Beresford, D. Y. Oh, D. Zhang, J. Lieberman, *Cell* **2003**, *112*, 659–672.
- 90 Y. Wang, D. Rasnick, J. Klaus, D. Payan, D. Broemme, D. C. Anderson, *J. Biol. Chem.* **1996**, *271*, 28399–28406.
- 91 L. Guo, J. R. Eisenman, R. M. Mahimkar, J. J. Peschon, R. J. Paxton, R. A. Black, R. S. Johnson, *Mol. Cell. Proteomics* **2002**, *1*, 30–36.
- 92 A. J. Bredemeyer, R. M. Lewis, J. P. Malone, A. E. Davis, J. Gross, R. R. Townsend, T. J. Ley, *Proc. Natl Acad. Sci. USA* **2004**, *101*, 11785–11790.
- 93 E. M. Tam, C. J. Morrison, Y. I. Wu, M. S. Stack, C. M. Overall, *Proc. Natl Acad. Sci. USA* **2004**, *101*, 6917–6922.
- 94 M. Funovics, R. Weissleder, C.-H. Tung, *Anal. Bioanal. Chem.* **2003**, *377*, 956–963.
- 95 R. Weissleder, C.-H. Tung, U. Mahmood, A. Bogdanov Jr., *Nat. Biotechnol.* **1999**, *17*, 375–378.
- 96 C. Bremer, C.-H. Tung, R. Weissleder, *Natl. Med.* **2001**, *7*, 743–748.
- 97 A. Wunder, C.-H. Tung, U. Mueller-Ladner, R. Weissleder, U. Mahmood, *Arthritis Rheum.* **2004**, *50*, 2459–2465.
- 98 P. L. Carl, P. K. Chakravarty, J. A. Katzenellenbogen, M. J. Weber, *Proc. Natl Acad. Sci. USA* **1980**, *77*, 2224–2228.
- 99 G. M. Dubowchik, R. A. Firestone, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3341–3346.
- 100 P. M. Loadman, M. C. Bibby, J. A. Double, W. M. Al-Shakhaa, R. Duncan, *Clin. Cancer Res.* **1999**, *5*, 3682–3688.
- 101 R. M. Schneider, Y. Medvedovska, I. Hartl, B. Voelker, M. P. Chadwick, S. J. Russell, K. Cichutek, C. J. Buchholz, *Gene Ther.* **2003**, *10*, 1370–1380.
- 102 S. R. Denmeade, A. Nagy, J. Gao, H. Lilja, A. V. Schally, J. T. Isaacs, *Cancer Res.* **1998**, *58*, 2537–2540.
- 103 D. DeFeo-Jones, V. M. Garsky, B. K. Wong, D.-M. Feng, T. Bolyar, K. Haskel, D. M. Kiefer, K. Leander, E. McAvoy, P. Lumma, J. Wai, E. T. Senderak, S. L. Motzel, K. Keenan, M. Van Zwieten, J. H. Lin, R. Freidinger, J. Huff, A. Oliff, R. E. Jones, *Nat. Med.* **2000**, *6*, 1248–1252.
- 104 A. Berger, I. Schechter, *Philos. Trans. R. Soc. Lond. B* **1970**, *257*, 249–264.

12

Enzyme Assays on Chips

Souvik Chattopadhyaya and Shao Q. Yao

12.1

Introduction

Enzyme assays are essential tools for the discovery and characterization of enzymes [1]. In recent years, much effort has been expended on developing miniaturized assays that can reproducibly analyze enzyme activities with high accuracy and be integrated into microarray-based formats for potential high-throughput screenings [2]. This is of paramount importance, especially in the postgenomic era, where the research focus is shifting from protein identification to protein functional analysis and from the study of single proteins to that of multiple proteins. The adaptation of assays possessing high-throughput capabilities aims to analyze thousands of samples in parallel with very small sample volumes, while retaining a good degree of detection sensitivity. In addition, many of these assays are scalable and amenable to automation, making them operationally simple, rapid, and extremely robust.

Traditionally, the 96-well microtiter plate-based assays, such as ELISA, have been the staple of enzymatic screenings. Early attempts at array-based high-throughput screening focused mainly on the use of peptide libraries generated *in situ* using methods such as SPOT synthesis [3] and light-directed parallel synthesis using photolithographic methods [4]. The field witnessed a significant advancement in 1995 through the pioneering efforts of Brown and coworkers [5], who generated cDNA microarray by directly spotting cDNA onto a glass slide, thus making the technology readily accessible to most research laboratories at an affordable price. Protein microarrays were introduced 5 years later by MacBeath and Schreiber [6], who robotically spotted proteins onto glass slides and used them to study protein–protein and protein–ligand interactions. For the first time, they also demonstrated the array-based detection of kinase activities using a microarray immobilized with known substrates of kinases. In the same year, Zhu et al. analyzed 122 predicted yeast kinases for their phosphorylation against 17 protein substrates in nanowells, which were fabricated on a standard glass slide treated with polydimethylsiloxane (PDMS) (1) [7]. The authors identified

27 novel kinases with unexpected tyrosine kinase activities. In essence, these pioneering studies underlined the potential of protein microarray and related microarray-based technologies, and showed for the first time that “delicate” biomolecules such as proteins may retain their native activities in a miniaturized format, making it possible for future developments of microarray-based, high-throughput enzyme-profiling methods.

Conceptually, enzyme-profiling assays on chips may be carried out with any of the three microarray-based platform technologies – namely the protein array, the peptide array, and the small molecule array [2]. In a protein array, where enzymes are immobilized on a glass slide, the critical issue is to develop strategies that allow the simultaneous assessment of the enzymatic activity of individual proteins while minimizing cross-contamination from neighboring proteins [8]. Alternatively, enzyme profiling can be done with a peptide microarray in which potential peptide substrates of enzymes are immobilized [6]. More recently, small molecule microarrays have been developed [9]. Here, small molecules, in most cases fluorogenic substrates of enzymes, are immobilized on the slide. Either way, the biggest challenge of these technologies is to develop methods that translate enzymatic activities quantitatively into some detectable readouts, such as fluorescence. Despite the runaway success of their predecessor technology – the DNA microarray, the development of microarray technologies based on proteins, peptides, and small molecules has been much more challenging, especially when these arrays are used for enzyme-profiling experiments. This is primarily due to the fact that successful enzymatic catalysis requires a number of criteria to be satisfied, including the proper folding/orientation of the immobilized enzyme/substrate, buffer composition of the reaction, the presence of cofactors, pH, temperature, and so forth. Coincidentally, in a microarray-based, enzyme-profiling assay, the loss of enzyme activity often accompanies the immobilization of enzymes onto a solid surface. Nevertheless, the high-throughput aspect of microarray-based technologies, together with the economy of reagents they provide, often offsets these limitations. Consequently the field has experienced climactic developments since the original work of MacBeath and Schreiber [6].

This chapter serves as a treatise on the current state of chip-based enzymatic assays. It includes detailed treatments of (1) site-specific immobilization of proteins in a microarray, which is the prerequisite for successful enzymatic assays, (2) detection schemes available to track enzyme activities, and (3) on-chip assays for enzyme identification, profiling and kinetic measurements. Finally, we also highlight recent advances in the adaptation of *in vitro* selection schemes such as Expression Display in chip-based formats, which could potentially expand the scope of enzymatic assays. Though the major focus has been on assays utilizing glass slides, helpful insights from other formats have been provided, where applicable, to enable readers to have a global picture of the field.

12.2 Immobilization Strategies

12.2.1 Covalent versus Noncovalent Immobilization

One of the first and key steps for successful on-chip enzyme assays is the immobilization of proteins or ligands (small molecules/peptides) [10]. For small molecules, their immobilization is relatively straightforward and has been reviewed recently [11]. The immobilization of proteins and peptides, however, is more complicated as these molecules have multiple reactive groups. Nevertheless, various immobilization strategies, both covalent and noncovalent, have been developed.

Noncovalent capture of proteins can be carried out through direct physical adsorption onto hydrophobic surfaces such as nitrocellulose (2) [12]. This is possible because proteins have a large surface area, allowing sufficient hydrophobic/electrostatic interactions and their subsequent adsorption/immobilization to the surface. Alternatively, they may be immobilized by the use of specific capture agents [13]. Antibody arrays in which monoclonal antibodies are immobilized on poly-L-lysine (3)-coated glass or nitrocellulose membranes by noncovalent adsorption, are examples of this type [13a, 13b]. Polyclonal antibodies may also be used but they often result in high background due to cross-reactivity. With the antibody array, antigenic ligands, for example enzymes, may then be specifically captured on the slide for subsequent biological assays. Besides antibodies, other capture molecules such as peptides [13c], functionalized dendrimers [13d], photoaptamers [13e], affinity ligands derived from the IgG-binding domain of protein A [13f], antibody mimics like TRINECTIN [13e] and protein domains [13h], have also been used for such immobilization. Methods such as electrospray ionization (ESI) and ink-jet technology have been used to noncovalently tether proteins without the need for capture molecules [14, 15]. The common feature of these methods, especially in the case of the ESI method, is that it allows further miniaturization of existing microarray-based methods by generating spots of biomolecules which are only a few micrometers in size, thereby potentially increasing the throughput, as well as the sensitivity of the assays. Polyacrylamide (4) gel packets [16], agarose (5) thin films [17] and hydrogels (6) [18] have also been developed for noncovalent immobilization of proteins. The 3D architecture of gel-based immobilization methods allows a greater amount of protein to be immobilized while maintaining them in an aqueous environment provided by the gel. However, major limitations of these arrays have been difficulties in fabrication, the drying of gel surfaces leading to variable performance over time and long incubation time required for the assay, especially for low-abundance proteins since the gel presents a diffusion barrier to protein movement. Peptides, on the other hand, are not typically immobilized to a solid surface via noncovalent methods, as they do not get adsorbed strongly to the surface due to their limited surface area. Instead, they are immobilized to the solid surface by covalent methods [19]. This will be discussed in detail later.

The covalent immobilization of proteins has been mediated by the use of a variety of solid surfaces modified with chemicals such as aldehydes (7) [6], epoxides (8) [7], or even photo-crosslinking agents (9) [20]. Recently, Morozov et al. reported the fabrication of “dry arrays” which make use of electrospray deposition to deliver dry proteins followed by covalent linkage to a dextran (10) grafted surface [21]. Some of these methods have been detailed in the following paragraphs where appropriate.

12.2.2

Site-specific versus Nonspecific Immobilization

As alluded to earlier, one of the prerequisites for a successful microarray-based enzyme assay is that the immobilized molecules, either proteins (in the case of a protein array), peptides (in a peptide array), or small molecules (in a small molecule array), have to be in a correct and uniform orientation. Randomly immobilized biomolecules exhibit relatively weaker activities/affinities to their targets compared with their solution-phase counterparts. This may be due to non-productive orientations, resulting in steric blockage of the enzyme active site or the structural element recognized by the enzyme (for small molecules and peptides), or the effect of the surrounding microenvironment near the surface [22]. In order to circumvent this in a microarray-based assay, site-specific tethering methods are typically employed to generate microarrays having uniformly oriented proteins, peptides, or small molecules. The immobilization strategies as mentioned above (Section 12.2.1), both covalent and noncovalent, lead to random immobilization and hence are not preferred in an enzyme assay.

In order to expand the utilities of microarray-based technologies such that enzyme assays could be routinely carried out in a highly quantitative, sensitive, and reproducible fashion, we and others have been focusing on developing methods that allow site-specific and uniform immobilization of biomolecules on a variety of solid surfaces, including glass slides.

12.2.3

Site-specific Immobilization of Peptides/Small Molecules

Small molecules, which in most cases are obtained by chemical synthesis, are typically anchored site-specifically onto the glass slide with unique chemical handles introduced during the synthesis. Since the introduction of the first small molecule microarray in 1999 [23], a number of mild and selective immobilization methods have been developed, based on different chemistries including the Michael reaction [23], the reaction between a primary alcohol and silyl chloride (11) [24], the 1,3-dipolar cycloaddition [25], the reaction between a phenolic derivate and diazobenzylidene (12) [26], the Staudinger ligation [27], and the reaction between an amine and *N*-hydroxysuccinimide (NHS) (13) ester [28]. In all cases where the site-specific immobilization was used, small molecule ligands were presented on the surface in a uniform orientation independent of

their diverse chemical structures, thus ensuring their optimal interactions with enzymes in solution.

Compared with small molecules, the site-specific immobilization of peptides is more challenging, because peptides contain multiple reactive functional groups. The first example of site-specific peptide immobilization used in a microarray-based enzyme assay was reported by Falsey et al., in which the authors chemically synthesized peptides containing an N-terminal cysteine and site-specifically immobilized them onto a glass slide functionalized with glyoxylic acid (**14**) [29]. An alternate strategy developed by Mrksich and coworkers focused on the use of self-assembled monolayers (SAMs) for peptide immobilization [30a, 30b]. To achieve this, a gold surface was modified with polyethylene glycol (PEG) (**15**) and alkanethiol (**16**) to generate the SAMs. Peptides containing a cyclopentadiene (**17**) moiety at their N-termini were chemically synthesized by solid-phase peptide synthesis, and site-specifically coupled to the quinone (**18**) groups on the SAM surface via the Diels-Alder reaction. The inert SAM surface was found to prevent nonspecific protein adsorption, thus minimizing background signals generated in downstream enzyme assays. Alternatively, the same group developed a strategy to immobilize N-terminal cysteine peptides onto maleimide (**19**)-derivatized SAM surfaces [30c]. In a more recent study, Olivier et al. anchored glyoxyl (**20**)-containing peptides onto semicarbazide (**21**)-derivatized slides via the chemoselective α -oxo-semicarbazone ligation reaction [31].

Our group has developed two approaches for the site-specific immobilization of peptides in a microarray format [32]. In the first strategy, N-terminally biotinylated peptides were spotted on avidin-functionalized slides to generate an immobilized array of peptides [32a]. Avidin is an extremely stable protein and the reaction with its natural ligand takes place almost instantaneously, thus doing away with the long incubation time required in other immobilization methods. It also acts as a molecular layer that minimizes nonspecific bindings of the slide surface. In the second strategy [32a], by utilizing the chemoselective native chemical ligation reaction [33], peptides containing an N-terminal cysteine were covalently immobilized onto slides functionalized with thioester (**22**) groups. In order to minimize nonspecific bindings of the glass surface, a molecular layer of PEG (**15**) was incorporated on the slide surface before attaching the thioester (**22**) moieties. The use of the PEG layer was advantageous as, unlike bovine serum albumin (BSA) (**23**) or other macromolecular “cushions”, it did not prevent the immobilized peptides from interacting with incoming targets. We subsequently used these peptide arrays to detect various enzymatic activities [32b, 32c].

12.2.4

Site-specific Immobilization of Proteins

Of the various site-specific immobilization methods used in a protein microarray, those that are rapid, minimize rigorous protein purifications and allow control over the density of immobilized proteins at each spot to ensure reproducibility, are highly desirable. In the first ever report of such site-specific tethering, Zhu

et al. immobilized 5800 yeast proteins, all of which were expressed with a glutathione-*S*-transferase-polyhistidine (GST-His₆) tag at their N-termini, onto a nickel-nitrilotriacetic acid (Ni-NTA) (24)-coated slide to generate the “yeast proteome array” [22 b]. However, the noncovalent His₆-Ni-NTA interaction is not ideal due to its relatively weak binding, and often susceptible to interference by commonly used chemicals and salts [34]. Consequently, only a limited number of downstream screening techniques were compatible with this type of protein array [2].

Several groups, including our own, have developed alternative approaches that allow stable and site-specific immobilization of proteins [35–38]. Mrksich and coworkers immobilized cutinase-fused proteins onto phosphonate (25)-containing SAMs [35]. Johnsson et al. developed a site-specific method to covalently immobilize hAGT-fused proteins onto modified glass surfaces [36]. Similarly, Walsh et al. used Sfp phosphopantetheinyl transferase to mediate site-specific covalent immobilization of target proteins fused to the peptide carrier protein (PCP) excised from a nonribosomal peptide synthetase (NRPS) [37]. We recently developed intein-mediated approaches to site-specifically biotinylate proteins both *in vitro* and *in vivo* (Figure 12.1) and subsequently immobilized them onto avidin-coated slides to generate the corresponding protein array [38]. In the subsequent paragraphs, we will discuss our recent extensions of this work [39], and at the same time list some of the key advantages associated with our approaches. Selected protocols for protein microarray generation using our approaches have also been included for the convenience of interested readers.

12.2.4.1 Intein-mediated Protein Biotinylation Strategies

By taking advantage of the extremely high affinity between biotin (26) and avidin or streptavidin ($K_d \sim 10^{-15}$ M), we developed intein-mediated approaches to express recombinant proteins which can then be site-specifically biotinylated and are therefore suitable for protein microarray generation (Figure 12.1a) [38]. Intein-mediated protein expression, originally developed for easy and effective purification of fusion proteins on chitin (27) columns [40], had previously been used to modify proteins with a number of chemical tags [41]. Our biotinylation strategies may be carried out either *in vitro*, *in vivo*, or in a cell-free expression system (methods A, B, and C in Figure 12.1a, respectively). They are highlighted by the following key aspects: (1) proteins are site-specifically biotinylated at their C-termini, leading to their subsequent immobilization on avidin-functionalized surfaces in a uniform orientation; (2) unlike other fusion protein approaches, our approaches do not introduce any extraneous macromolecular tag at the end of protein biotinylation, thus minimizing the potential perturbation to the protein's biological activity; (3) the biotin/avidin interaction is one of the strongest noncovalent interaction known, and is stable under most stringent conditions, making it ideal for protein immobilization; (4) each molecule of avidin binds four molecules of biotin and thereby allows for uniform protein density at each spot; and (5) the various formats applicable to our protein biotinylation strategies (*in vitro*, *in vivo* or cell-free) allow an easy access to desired

biotinylated proteins from crude cellular lysates (or mixtures of unpurified proteins) for subsequent protein immobilization and microarray generation. We have shown that endogenous nonbiotinylated proteins present in the cell lysate can be washed away in an efficient and highly parallel fashion (thousands of different protein spots could be processed simultaneously on a single glass slide), so that protein purification and immobilization are essentially carried out in a single step to generate functional protein microarrays [38b]. This is true because of the rare occurrence of naturally biotinylated proteins in the cell, and the highly specific and strong nature of biotin/avidin interaction, which can withstand extremely stringent washing/purification conditions otherwise impossible with other affinity tags [38a].

In our *in vitro* strategy (Figure 12.1a; method A), the protein of interest was fused through its C-terminus to an intein, which contains a chitin-binding domain as an affinity tag [38a]. To biotinylate the protein *in vitro*, the host cell overexpressing the protein of interest was first lysed and the lysate containing the intein fusion protein was loaded onto a column packed with chitin (27) beads. Following addition of a thiol-cleaving reagent (for example, cysteine-biotin (28), inset in Figure 12.1a), the fusion protein underwent a on-column self-cleavage reaction, catalyzed by the fused intein, to generate a protein having a reactive α -thioester (22) group at its C-terminus. The thioester (22) moiety was subsequently quenched by the thiol side-chain from the added cysteine-biotin (28), resulting in a thioester-linked intermediate that spontaneously rearranged to form a native peptide bond and generate the target protein which was site-specifically biotinylated at its C-terminus [38a]. We have recently shown that this strategy is capable of biotinylating a variety of proteins from different biological sources in 96-well formats [38b], making it possible for future high-throughput generation of a large number of biotinylated proteins needed in a protein microarray.

We also successfully carried out the intein-mediated strategy to biotinylate proteins *in vivo* in both bacterial and mammalian systems [38b]. Early attempts of *in vivo* protein biotinylation had relied on fusing proteins at the N- or C-termini with a 15-amino-acid peptide, the AvitagTM, which was subsequently biotinylated by biotin ligase, a 35.5 kDa monomeric enzyme encoded by the *birA* gene in *E. coli* [42]. Biotin ligase catalyzed the transfer of biotin (26) to the ϵ -amino group of a specific lysine residue within the Avitag. Unfortunately, *in vivo* biotinylation of proteins mediated by biotin ligase is often inefficient due to a limit amount of biotin ligase in the cells, and is highly cytotoxic [42]. In our approach (Figure 12.1a, method B), the simple addition of the cell-permeable cysteine-biotin (28) probe to the culture media containing cells expressing the target protein, followed by a brief incubation, resulted in a substantial biotinylation of the protein inside the cells. Further optimizations of the cell growth and *in vivo* biotinylation conditions led to an increased level (90–95%) of protein biotinylation in the cells. For both systems, apart from endogenous biotinylated proteins, the only other biotinylated protein was the target protein. More recently, we have found that the efficiency of intein-mediated protein biotinylation, both *in vitro* and *in vivo*, depends greatly on the intein fused to target pro-

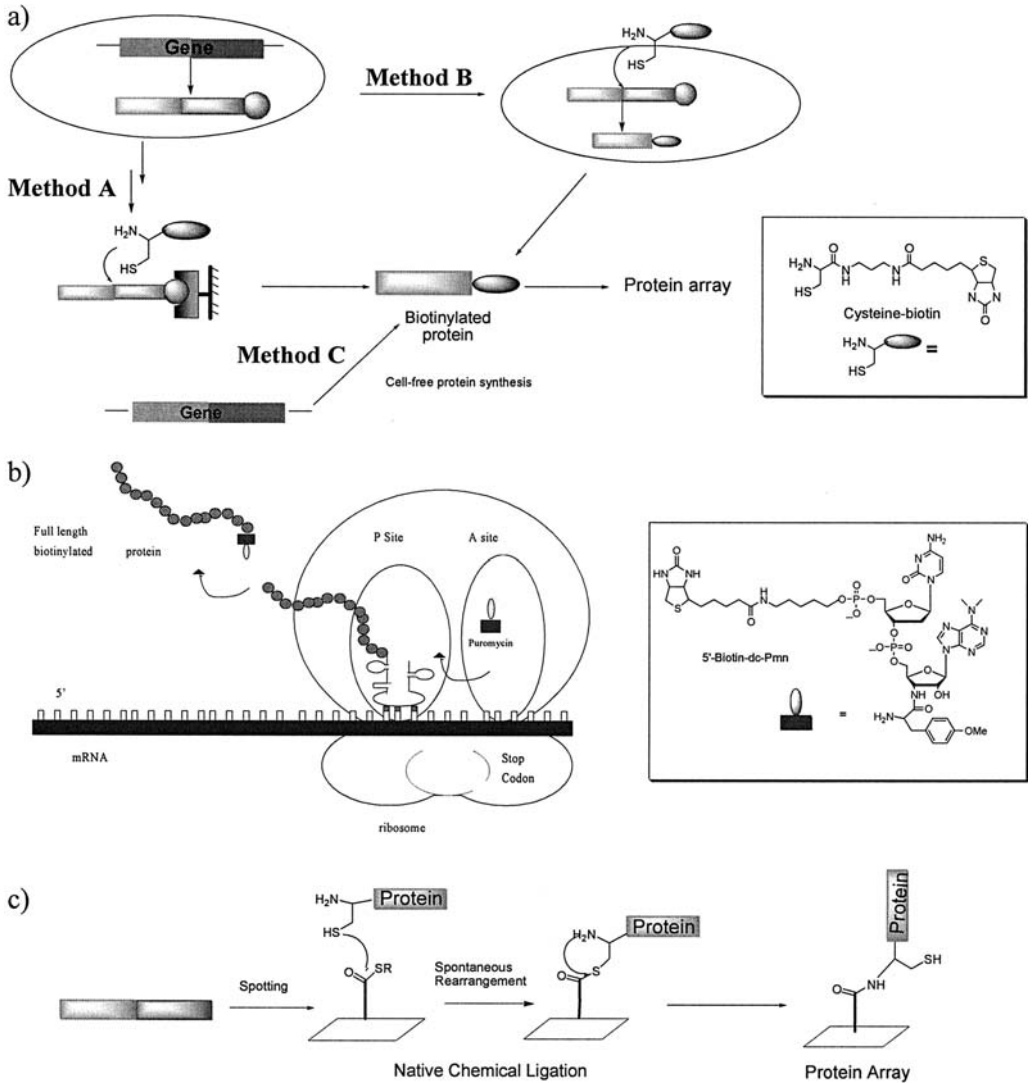


Fig. 12.1 (a) Methods for intein-mediated site-specific biotinylation of proteins using *in vitro* (method A), *in vivo* (method B) and cell-free (method C) systems. (b) *In vitro* protein biotinylation using Puromycin (37) analogs. (c) Immobilization of N-terminal cysteine proteins on thioester (22) slides via native chemical ligation reaction.

tein – in as much as 2- to 10-fold improvement in protein biotinylation may be achieved by the simple switch in the intein used [28].

The intein approach has also been extended to a cell-free protein synthesis system (Figure 12.1a; method C) [38b]. A cell-free system has many advantages

over both the *in vitro* and *in vivo* methods described [43]. Potentially a large number of proteins could be simultaneously expressed in a matter of hours in 96- or 384-well formats using commercially available, cell-free protein translation systems. Cellular toxicities due to the overexpression of certain proteins, possible degradation by endogenous proteases and formation of inclusion bodies by proteins can be all together avoided as well.

Protocol 12.1: Intein-mediated Protein Biotinylation and Immobilization

1. Generate constructs for expressing intein fusion proteins.^{a)}
2. Site-specifically biotinylate proteins as described in the subsequent steps.
 - **In vitro:**
 - i. Transform the plasmid into a suitable host strain that bears the T7 RNA polymerase gene under an inducible promoter; induce protein expression under optimal conditions.^{b)}
 - ii. Harvest cells by centrifugation (5000×g, 15 min, 4 °C). Discard supernatant. Store pellets at -20 °C or immediately proceed to cell lysis.
 - iii. Resuspend pellet from 1 L culture into 50 mL cold lysis buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM EDTA, 0.1% Triton X-100).^{c)} Lyse cells by sonication or using French press.^{d)}
 - iv. Clarify lysate by centrifugation (20000×g, 30 min, 4 °C) and collect the supernatant.
 - v. Pack the desired volume of chitin beads into a column (3 mL of beads is sufficient for protein purification from 200 mL of culture).
 - vi. Prior to loading of the crude cell lysate, pre-equilibrate the column with 10 column volumes of column buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, and 1 mM EDTA) at 4 °C.^{e)}
 - vii. Load the clarified cell lysate onto the column at a flow rate of 0.5 mL min⁻¹.
 - viii. Wash the column with 30 column volumes of column buffer at a flow rate of 2 mL min⁻¹ to remove all traces of contaminating proteins.
 - ix. Quickly flush the column with 3 bed volumes of column buffer containing 30 mM cysteine-biotin (28); stop the flow and incubate the column overnight at 4 °C.^{f)}
 - x. Elute biotinylated target protein using column buffer or a specific buffer for long-term storage of proteins.
 - xi. Desalt with a NAP-5 column, if necessary, before proceeding to spotting.

- **In vivo** biotinylation in bacterial cells:
 - i. The initial steps of cloning, transformation, and induction of protein expression are essentially the same as the *in vitro*-based method.
 - ii. Following induction, add 2-mercaptoethanesulfonic acid (MESNA) (29) and cysteine-biotin (28) to the induced bacterial culture to a final concentration of 30 mM and 3 mM, respectively; incubate at 4 °C for 24 h with constant shaking.
 - iii. Harvest cells by centrifugation (6000×g, 15 min, 4 °C).
 - iv. Wash the cell pellet at least twice with phosphate-buffered saline (PBS) to remove excessive unreacted cysteine-biotin (28).
 - v. Resuspend the cell pellet in 1 mL lysis buffer for protein extraction. Cell pellet can also be stored at –20 °C without any significant degradation of the biotinylated protein.
 - vi. Lyse the cells by sonication on ice at 50% duty, 20% power in five treatments of 30 s each with 30 s cooling interval or by using French press.
 - vii. Remove cell debris by centrifugation (20 000×g, 20 min, and 4 °C) and collect clarified cell lysate (supernatant).
 - vii. The cell lysate can be spotted directly onto the avidin slide without any further treatment.
- 3. Prepare a source plate of the pure biotinylated proteins dissolved in PBS (pH 7.4) or of clarified cell lysates containing the biotinylated protein.
- 4. Spot the proteins onto avidin-coated slides⁸⁾ using an ESI SMA-Arrayer with a spot spacing of 220 μm and incubate for 10 min in a slide hybridization chamber; wash slides with PBS extensively to remove unbound proteins and air dry the slides.
- 5. Incubate slides with labeled antibody (or mixture of antibodies) for 1 h; wash with PBST (PBS with 0.1% Tween-20 followed by water; air dry and scan the slides using the arrayWoRxTM (Applied Precision, WA, USA) scanner.

Notes

- a) It is imperative to choose the appropriate vector for cloning that best suits the needs of the study. Detailed procedures can be found at the New England BioLabs website (www.neb.com).
- b) ER2566 is available from New England BioLabs for expression. Other strains [BL 21 (DE3) and its derivatives] can also be used for expression.
- c) Other nonionic detergents (0.1–0.2% Tween 20, protease inhibitors (phenylmethylsulfonyl fluoride (PMSF) (30), pepstatin (31), leupeptin (32), and reducing agents like 1 mM Tris (2-carboxyethyl) phosphine hydrochloride (TCEP) (33)/2,4',4-trichloro-2-hydroxydiphenyl ether (TCCP) (34) can be included to stabilize target proteins. The presence of thiol reagents (β -mercapto-

- ethanol (35), 1,4-dithiothreitol (DTT) (36), cysteine will cause premature cleavage of fusion protein, resulting in a loss of target protein before affinity purification. As such, thiol compounds should be strictly avoided in all steps to maximize target protein recovery.
- d) Lysozyme binds and digests chitin and should be avoided during lysis. If alternate methods for lysis are not available, mild treatment with lysozyme ($10\text{--}20\ \mu\text{g mL}^{-1}$, $4\ ^\circ\text{C}$, 1 h) can be used.
 - e) All purification steps should be carried out at $4\ ^\circ\text{C}$ to ensure stability of fusion protein.
 - f) Several factors like the amino acid residue at the cleavage site, duration, pH, and temperature during cleavage may affect the cleavage efficiency and hence the final yield of protein. For proteins which do not cleave effectively, longer time (40 h) at a higher temperature ($16\text{--}23\ ^\circ\text{C}$) and pH 9.0 may be used.
 - g) Detailed protocols for slide preparation have been described previously [38a].

12.2.4.2 Puromycin-mediated Protein Biotinylation

Alternatively, we reported a cell-free strategy which utilizes puromycin (37)-containing small molecules to site-specifically biotinylate proteins at their C-termini (Figure 12.1b) [39a]. Puromycin (37) is an aminonucleoside antibiotic produced by *Streptomyces alboniger*. As puromycin (37) resembles the 3' end of the aminoacyl-tRNA, it competes with ribosomal protein synthesis by blocking the action of the peptidyl transferase, leading to inhibition of protein synthesis in both prokaryotic and eukaryotic ribosomes. It was previously found that, at low concentrations, puromycin and its analogs act as non-inhibitors of the ribosomal protein synthesis, and get incorporated at the C-terminus of the newly synthesized protein [44]. Our approach thus exploited a similar phenomenon for protein biotinylation. Using this newly developed method, we showed biotinylated proteins could be obtained in a matter of hours using plasmids or PCR products as DNA templates, and that this method is compatible with other high-throughput cloning/proteomics methods such as the GatewayTM (Invitrogen, Carlsbad, MA, USA) cloning strategy [39a].

12.2.4.3 Immobilization of N-terminal Cysteine-containing Proteins

In yet another complementary approach, we generated N-terminal cysteine-containing proteins using the pTWIN vectors (Figure 12.1c) [45]. These vectors allow the expression of target proteins with the self-cleavable modified *Ssp DnaB* intein having a chitin (27)-binding domain fused at their N-termini. Briefly, the recombinant protein was engineered by standard PCR-based methods and subsequently expressed to have a cysteine residue as its N-terminus. Following purification of the fusion protein on chitin (27) beads, on-column cleavage of the intein tag was initiated by adjusting the pH of the column buffer. The eluted protein, having an N-terminal cysteine residue, could be site-specifically immobilized onto thioester (22)-functionalized slides via the chemoselective native chemical ligation reaction [33]. Only the terminal cysteine residue reacts with the thioester to form a stable peptide bond; other reactive side-chains, including internal cysteines, are tolerated.

12.3

Microarray-based Methods for Detection of Enzymatic Activity

Most of the microarray-based screening methods developed so far have relied on the detection of protein functions by virtue of the noncovalent binding of a protein with its ligand(s) thereby limiting the screening to protein–protein, protein–ligand, or antigen–antibody interactions [2]. These approaches had essentially precluded enzymes, which play vital roles in all cellular processes and metabolic transformations, from such high-throughput studies. Nonetheless, recent advances in bioassay development have provided sensitive, selective tools for the potential high-throughput identification and characterizations of a variety of enzymes.

The screening of enzymes on the basis of their catalytic activities is advantageous as it allows for intrinsic signal amplification due to multiple turnovers of enzymes, leading to extremely high detection sensitivity. Enzyme assays not only detect enzymes but also can be used to profile enzymes in terms of their substrate preference to generate “fingerprint” profiles that are unique for each enzyme [46]. Such analysis reveals the type of chemical entities accepted by the enzyme as its potential substrates, and thereby helps in better understanding its catalytic mechanism and properties. Alternatively, the pattern generated for an unknown enzyme using a set of substrates may possibly be used to obtain its identity. In a biological context, it is this ability of an enzyme to discriminate among potential substrates that makes it an essential component in maintaining the fidelity of most cellular processes. Within the cellular environment, while the substrate specificity of an enzyme is the overriding principle that dictates substrate turnovers, other parameters like the presence of cofactors, spatial and temporal localization of enzymes and substrates, as well as their concentrations, may regulate enzyme activities. Insights into the substrate preference of an enzyme under different sets of assay conditions would not only provide invaluable information for identification of physiological substrates and help in the dissection of complex biological pathways, but also aid in the design of potent, selective inhibitors.

Different types of enzyme assays have been developed which are based on a variety of analytical techniques [46]. Most high-throughput enzyme assays including those based on microarray technologies, however, employ colorimetric methods because of their operational simplicity, reliability, and greater sensitivity for low enzyme turnover reactions (especially when fluorescence-based detection methods are used). In most cases, fluorogenic substrates used in these assays are highly selective and allow real-time monitoring of enzyme activity by providing a direct correlation between enzyme activity and the signal. Fluorescence as opposed to radioactivity or chemiluminescence is the method of choice for detection since radioactive methods are considered unsafe and in most cases only serve as endpoint sensors and require a long exposure time. Chemiluminescence-based methods have a limited dynamic range of signal detection and are less compatible for potential multiplexing experiments. By overcoming

many of these limitations, fluorescence-based methods provide a high signal-to-noise ratio, and more importantly, are compatible with standard microarray scanners, thereby keeping instrumentation costs low.

12.3.1

Enzyme Assays Using Protein Arrays

Traditionally, enzymatic assays have been carried out in a microtiter plate format, where enzymes are mixed separately with their substrates in individual wells, and their activities read out by the measurement of substrate turnovers. The wells functioned as individual compartments to confine each enzymatic assay and prevent its contamination from neighboring reactions. This is obviously not possible in a protein microarray, where thousands of protein spots are immobilized on a single flat surface. Without the existence of any physical barrier to prevent reagents used in enzyme assays, including enzyme substrates, from freely diffusing across the whole array surface, most microtiter plate-based enzyme assays could not be adapted to a protein array format. To address this problem, we have recently developed a novel activity-based strategy that allows sensitive detection of enzymes present in a protein microarray (Figure 12.2a) [8]. We made use of small molecule probes derived from mechanism-based suicide inhibitors (Figure 12.2b). An extensive use of these probes has previously been reported only in gel-based methods for the global analysis of enzyme expression and functions [47]. By covalent modifications of target enzymes in a highly selective manner, these probes facilitate the easy identification and/or purification of proteins from complex proteomes. Several types of probes have been developed that target different classes of proteins on the basis of their enzymatic activities. For example, fluorophosphonate/fluorophosphates (38) derivatives have been developed to selectively profile serine hydrolases, including serine proteases [48]. For cysteine proteases, different classes of chemical probes have been reported, including probes containing α -halo (39) or (acyloxy)methyl ketone (40) substituents [49]. Other known activity-based probes include sulfonate ester (41)-containing probes that target a few different classes of enzymes, as well as probes conjugated to *p*-hydroxymandelic (42) acid that specifically label protein phosphatases and proteases [50]. More recently, affinity-based approaches have also been developed to profile other classes of enzymes, including kinases [51a] and metalloproteases [51b, 51c], which were previously inaccessible by activity-based profiling approaches [51]. We have demonstrated that these probes are ideal tools for microarray applications [8]. Under appropriate conditions, incubation of these probes with an enzyme array leads to its reaction with the immobilized enzymes by virtue of their activity against the inhibitors. Because the probes are pre-labeled with a fluorescent dye, the formation of covalent enzyme–inhibitor adducts renders the enzymes detectable. Since different mechanism-based inhibitors, either highly specific or general in terms of their reactivity, are known, the strategy would allow for the identification of specific enzymes or most enzymes belonging to the same class. Furthermore, we

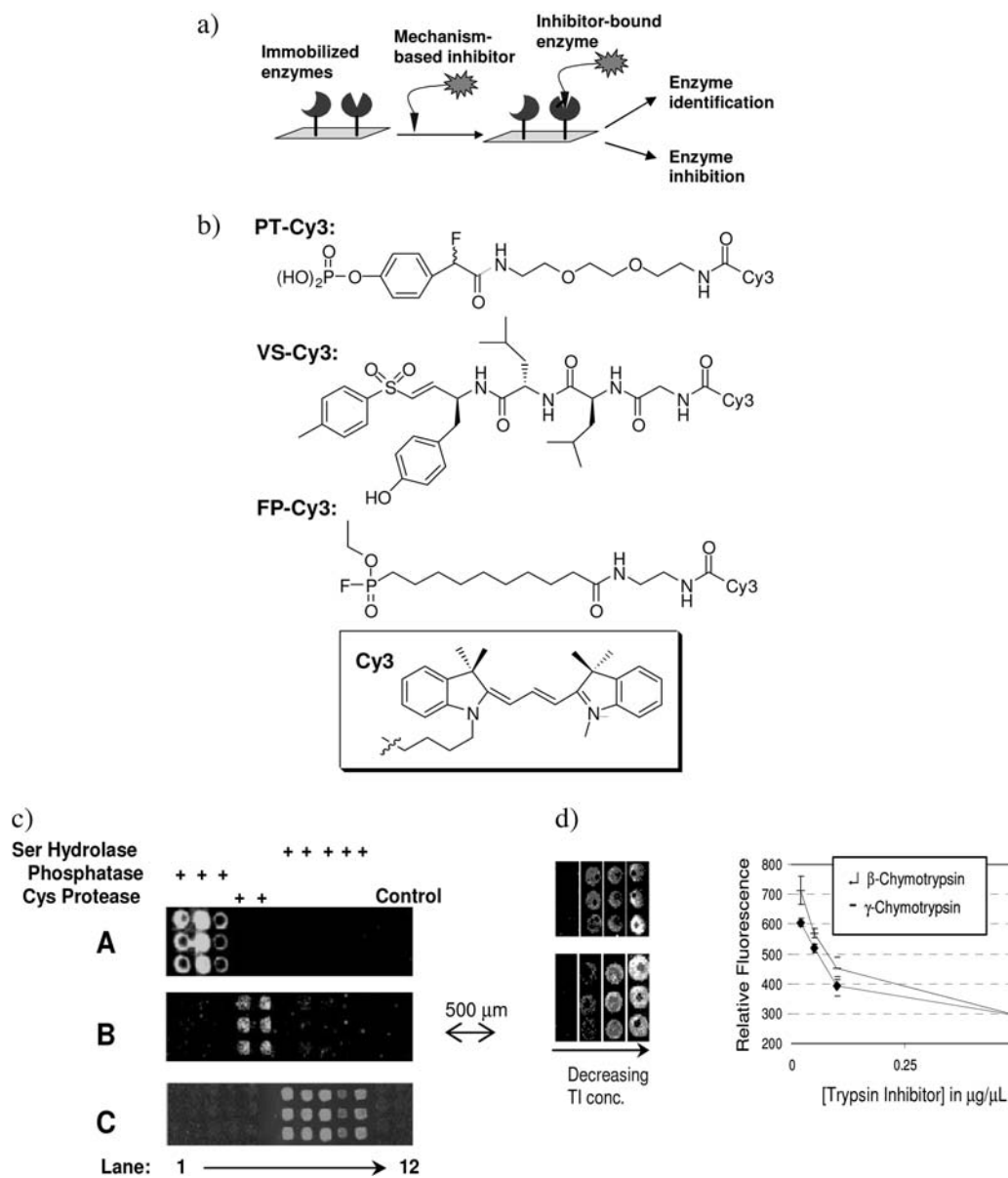


Fig. 12.2 (a) Principle of activity-based detection of enzymes in a microarray. (b) Mechanism-based probes used for detection of phosphatases (PT-Cy3), cysteine (VS-Cy3) and serine (FP-Cy3) proteases. (c) Enzymatic screening using PT-Cy3 (panel A), VS-Cy3 (panel B), and FP-Cy3 (panel C) of 12 commercial enzymes on a microarray. (d) Competitive inhibition study of trypsin inhibitor (TI) and FP-Cy3 for binding to α - and β -chymotrypsin.

have established that the probes react in a time- and concentration-dependent fashion on the array, thus allowing for quantitative kinetic data to be obtained for the complete characterization of enzyme activity. Our strategy was successfully carried out with a protein microarray immobilized with different classes of enzymes, including phosphatases, cysteine proteases, serine hydrolases (Figure 12.2c), using a panel of different mechanism-based probes [8]. Two different approaches were employed on the enzyme array. First, direct labeling of the enzymes was effected by the application of the probes to the array, which allowed for high-throughput identification and assignment of protein function by virtue of their enzymatic activities. Second, inhibition studies were carried out using potential enzyme inhibitors, either general or specific, to simultaneously probe and intercompare the efficacy of such inhibitors towards different enzymes. Rapid profiling of enzymes provided by this approach empowers researchers to uncover either broad- or narrow-range inhibitors that regulate a variety of enzyme activities. Along the same vein, activators for enzymes may also be sought which may either turn on, or increase the turnover of enzymes. Such inhibitors and activators, once profiled and characterized, may then be used for confirmatory *in vivo* studies and in the development of novel therapeutics. In our studies, we have used trypsin inhibitor (TI) that selectively targets serine proteases in order to demonstrate the utility of this strategy (Figure 12.2d).

Protocol 12.2: Enzyme Assay Using Mechanism-based Probes

1. Prepare each mechanism-based inhibitor probe^{a)} to 2 μM using 0.5 μL of 200 μM probe stock solution^{b)} (100 \times) in 49 μL of Tris buffer (50 mM, pH 8), and 0.5 μL BSA (23) (1% w/v). BSA serves as a blocking agent to prevent nonspecific binding.
2. Apply 50 μL of this freshly prepared mixture to each printed slides, using the cover slip method, and incubate for 30 min in the dark.^{c)}
3. Rinse away unreacted probe using repeated washes with distilled water.
4. Wash slides further with PBS containing Tween-20 (0.2% v/v) for 15 min on a shaker.
5. Rinse with distilled water and dry slides under a stream of nitrogen or using a centrifuge fitted with necessary adapters. Leave slides to air dry, if necessary, until completely dry.
6. Scan the slides using an arrayWoRxTM (Applied Precision, WA, USA) microarray scanner at $\lambda_{\text{ex/em}} = 548/595 \text{ nm}$.

Notes

- a) The chemical synthesis of the probes had been described previously [8, 49b, 49e, 50c].
- b) Probe stock solutions were prepared in dimethyl sulfoxide (DMSO) (43) and stored at -20°C .

- c) In case of expensive reagents, where it is desirable to minimize volumes, glass cover slips are used. For a 22×60 mm cover slip, a 50 μ L of reagent is typically sufficient to cover the slide. Two different methods can be used to apply the reagent to the glass surface. The reagent can be applied to the glass surface followed by the cover slip or vice versa. Both methods work equally well, but one ought to use the method that produces the most uniform layer of reagent across the slide surface, without introducing bubbles or air in the space between the glass and cover slip. Use slide hybridization chambers for incubation. Cover slips can be slid off the surface after the reaction is complete or by shaking the slide in a water bath until the cover slip floats off.

12.3.2

Enzyme Assays Using Peptide/Small Molecule Substrate Arrays

12.3.2.1 Proteases and Other Hydrolytic Enzymes

Proteases, enzymes that catalyze hydrolysis of peptide bonds, comprise approximately 2% of the encoded genes in organisms whose genomes have been decoded. These enzymes regulate diverse physiological processes ranging from digestion, fertilization to cell migration and apoptosis and are therefore naturally linked to disease states like cancer, AIDS and other inflammatory, immunological, cardiovascular, and neurodegenerative disorders including Alzheimer's disease. As such, proteases must be highly discriminating for substrate cleavage sequences. A number of approaches – such as the use of fluorescence resonance energy transfer (FRET) peptides [52], substrate phage libraries [53], positional-scanning libraries [54], mixture-based and oriented peptide libraries [55, 56] – have been described to determine the substrate specificity of proteases (see Chapter 11). However these methods do not possess the degree of throughput normally provided by microarray-based technologies. Consequently, the complete specificity profile of a protease, as well as the kinetic evaluation of all of its potential substrates, is not easily attainable.

To address these shortcomings, a number of microarray-based protease profiling methods have been developed [18, 57–59]. By using a solid support for peptide immobilization, Kiyonaka et al. studied protease activities using fluorogenic peptides trapped in a 3D supramolecular hydrogel [18]. Cleavage of a pentapeptide 5-dimethylaminonaphthalene-1-(*N*-2-aminoethyl)sulfonamide (DANSen) (**44**) conjugate by lysyl endopeptidase (LEP) was monitored as a function of the increase in fluorescence intensity of DANSen (**44**). The method was extended to profile other proteases such as V8 and chymotrypsin, as well as to screen for LEP inhibitors like *N*^α-tosyl-lysine chloromethylketone (TLCK) (**45**). Gosalia and Diamond recently described the use of a liquid-phase microarray that utilizes nanoliter sample volumes for protease screenings [57]. For the purpose of array fabrication, glycerol droplets were spotted on glass slides to form individualized “reaction centers”. Homogeneous enzyme assays were then assembled onto the array by simple aerosol deposition of reagents, thus doing away with the need for elaborate surface modifications which are common in other chip-based enzyme assays. Such fluid phase reactions are advantageous as it is possible to tailor optimized reaction conditions at each individual position on the array. Using

fluorogenic peptide substrates, the authors were able to detect multiple enzymes on a single array and demonstrated that the approach is amenable to enzymatic profiling and inhibition studies in a high-throughput manner. A chromogenic assay for detecting hydrolases was described by Park and Clark [58]. In this method, hydrolases were encapsulated in sol-gel microstructures to create the “solzyme” array and hydrolysis was monitored by the color change of a generic indicator, bromothymol blue (46). A panel of 20 different hydrolases was studied. Activities of the encapsulated enzymes were found to be consistent with those in solution-based assays. The array was also used to probe the enzyme specificity and inhibition using different peptide substrates.

Ellman et al. recently developed an elegant approach for potential high-throughput profiling of proteases in a microarray format [59]. They made use of combinatorial libraries of peptidyl coumarin derivatives immobilized site specifically in a microarray. In this method, cleavage of the peptides was monitored as a function of the increase in fluorescence due to the release of coumarin by hydrolysis of the peptide anilide bond. 7-Amino-4-carbamoylmethyl coumarin (ACC) (47) was used as the reporter. In all, 361 different fluorogenic peptides were arrayed on a single slide, in which P1-lysine and P4-alanine positions of the peptides were kept constant while the P2 and P3 positions randomized with all 19 amino acids (excluding cysteine). The authors showed that using this method it is possible to generate the complete specificity profile of a given protease, in this case thrombin, and obtain k_{cat}/K_m values which were determined to concur well with those from solution-phase results.

We independently developed a similar approach for high-throughput identifications of hydrolytic enzymes including esterases, lipases, proteases, and epoxide hydrolases [9]. In our method (Figure 12.3a), a fluorogenic coumarin (47) derivative was suitably modified to generate a series of substrates that were immobilized on a glass slide to yield a small molecule array. The coumarin derivative served as a bifunctional handle to immobilize the substrates via a carboxyl group and also had an electron-donating hydroxyl/amino group for attaching the enzyme recognition head that, in turn, was tailored to target a particular enzyme class or differentiate between enzymes with altered substrate specificities within a given class. For example (Figure 12.3b), compounds 1 and 2 targeted epoxide hydrolases and esterases, respectively, with each containing an enzyme recognition head joined to the hydroxyl moiety of the coumarin (47) derivative via a linker. Compounds 3a, 3b, and 4 were designed to target phosphatases and proteases. All conjugates were highly resilient to nonspecific hydrolysis and almost nonfluorescent when the enzyme recognition head was bound to the coumarin. Upon enzymatic cleavage of the head/substrate, the fluorescent coumarin was “unmasked” thereby allowing for enzymatic activity to be translated to fluorescent readouts (Figure 12.3c). The release of the highly fluorescent coumarin was either direct (proteases/phosphatases) or indirect (esterases, epoxide hydrolases) via the formation of a vicinal diol or amino alcohol as the primary reaction product, which was then processed by *in situ* oxidation and spontaneous β -elimination.

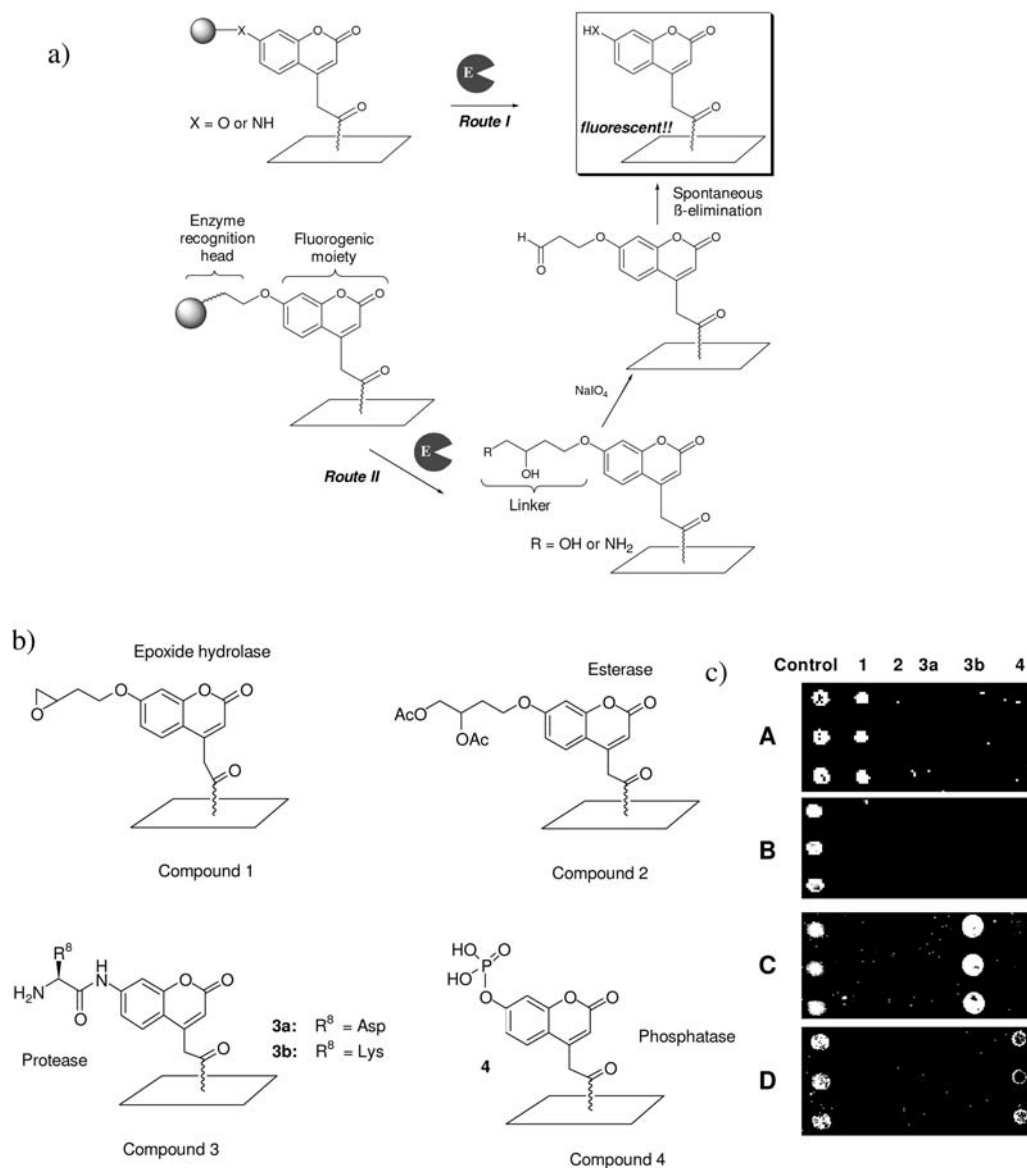


Fig. 12.3 (a) Strategies for detection of hydrolytic enzymes on microarrays (route I for proteases and phosphatases, route II for epoxide hydrolases and esterases). (b) Chemical structure of the fluorogenic substrates (1–4) used for the detection of the different hydrolases. (c) Fluorescence profiles generated on microarray following treatment with (A) epoxide hydrolase from *Rhodococcus rhodochrous*, (B) acetylcholine esterase from *Electrophorus electricus*, (C) trypsin, (D) alkaline phosphatase. 7-Hydroxy-4-carboxymethylcoumarin (47) was used as control for all the experiments.

It should be noted that our method was inspired by the extensive and elegant work carried out by Reymond et al., in which they showed similar approaches could be used in a microplate format to generate “fingerprints” of a variety of enzymes [60]. Besides our studies, the use of fluorescently tagged affinity ligand as a means to quantitatively measure on-chip bindings of reversible and irreversible inhibitors to proteases has been reported by Miyake and coworkers [61].

Protocol 12.3: Detecting Hydrolytic Enzymes Using Fluorogenic Substrate Arrays

1. Immobilize compounds 1–4 on amine functionalized slides^a).
2. Treat each slide with the 50 μL of desired enzyme (1 mg mL^{-1}) in 20 mM borate buffer pH 8.8 containing 1 mM sodium periodate (48) and 2 mg mL^{-1} BSA (23) (for esterases and epoxide hydrolase) or 50 mM Tris–HCl, pH 8.0 (for alkaline phosphatase and trypsin)^b).
3. Incubate the slides for 4–8 h, as required. Rinse slides with distilled water and dry under a stream of nitrogen or use a centrifuge with appropriate adapters. Leave slides to air, if necessary, until completely dry.
4. Scan the slides using an arrayWoRxTM (Applied Precision, WA, USA) microarray scanner at $\lambda_{\text{ex/em}} = 360/457 \text{ nm}$ to assess the release of the highly fluorescent coumarin.

Notes

- a) The carboxyl moiety of the coumarin (47) was first activated by NHS (13)/*N,N'*-dicyclohexylcarbodiimide (DCC) (49)/*N,N*-diisopropylethylamine (DIEA) (50) followed by direct spotting on the slides.
- b) The presence of periodate (48) and BSA (23) in the reaction mix does not interfere with enzymatic activity but aids in the efficient sodium periodate oxidation of the 1,2-diol/1,2-amino alcohol linker and the subsequent β -elimination.

12.3.2.2 Kinases

Protein phosphorylation mediated by kinases represents one of the most prevalent posttranslational modifications, reflected in as many as one-third of all eukaryotic gene products. Phosphorylation events are involved in the intricate control of cellular signal transduction pathways and form an essential facet of the regulatory mechanism in the cell. Consequently, kinases are believed to be highly “drugable” proteins. Likewise with other enzymes, their substrate specificity is governed by the primary sequence around the phosphorylation site located within the target protein. Methods including those based on peptide phage display [62], on-bead peptide libraries [63], or SPOT technology [64] for profiling kinases have been known for some time. In recent years, however, peptide microarrays in which potential peptide substrates of kinases are immobilized on a glass slide, have become the preferred method to study kinase activities and identify their substrate sequences.

The earliest functional kinase assay in a microarray format was in the example of kemptide (51) phosphorylation catalyzed by cAMP-dependent protein kinase A (PKA) [6]. In a more extensive study, Zhu et al. tested 119 of the 122 putative yeast kinases against a panel of 17 different proteins immobilized in a nanowell format [7]. The group was able to identify as many as 27 proteins possessing novel tyrosine kinase activities. However, the rate-limiting step in the use of proteins as substrates for such studies remains in the production of a large number of pure and biologically active proteins. Peptides as kinase substrates was first used by Falsey et al. who showed the incorporation of ^{33}P into an immobilized peptide by p60^{c-src} tyrosine kinase [29]. Alternatively, Houseman et al. developed self-assembled monolayer (SAM)-based peptide microarrays for the rapid and quantitative evaluation of kinase activity [30a]. In their work, peptide-cyclopentadiene conjugates were immobilized, in a site-specific manner, onto hydroquinone (52)-derivatized surfaces. Following incubation with the kinase, [γ - ^{32}P]-ATP and varying amounts of the inhibitors – quertin (53), tyrphostin A47 (54), and PP1 (55), the radioactivity incorporated into the peptides was measured and inhibition constants (K_i) calculated. A direct determination of K_i values was possible in this study as data acquired were a true reflection of the equilibrium binding between inhibitors and the kinase since, in an immobilized format, a large excess of the inhibitors relative to peptides precludes any competitive binding of the peptides to the enzyme. A recent addition to the identification of peptidic kinase substrates has been the use of “phospho-site” collection arrays [65]. These high-density arrays contain libraries of peptides derived from annotated phosphorylation sites in human proteins. Following synthesis by SPOT technology, the peptides were released from the cellulose membrane and immobilized chemo-selectively via an N-terminus aminoxyacetyl moiety. Fluorescently tagged antibodies or autoradiography-based detection was used to profile different kinases including PKA, CK2, Abl-tyrosine kinase, and NEK-6. The method also allowed the study of key events in kinase recognition and regulation [64a]. The phosphorylation of cytoplasmic domains of human membrane proteins by CK2 and the selectivity, subsite specificity and cross-reactivity of generic antiphosphoamino antibodies have also been investigated using this method [64b].

In order to obviate the drawbacks of radioactivity-based detection methods used in the above-mentioned studies without the loss in sensitivity, our group developed an alternate detection scheme based on the use of fluorescently labeled antiphosphoamino antibodies (Figure 12.4a) [32a]. The utility of this approach was shown in several proof-of-concept experiments using the peptides ALRSLG and YIYGSFK, which are substrates for PKA and p60^{c-src} kinase, respectively [32b]. The peptides were immobilized site-specifically onto avidin and/or thioester (22) slides, treated with the cognate kinase and probed with the antibody. The antibody detected not only the phosphorylated amino acids with high fidelity but also permitted the study of concentration-dependent phosphorylation events in a quick and convenient format. An alternate, fluorescence-based method in the detection of phosphorylated peptides/proteins has also become commercially available in recent years that make use of the dye ProQ Dia-

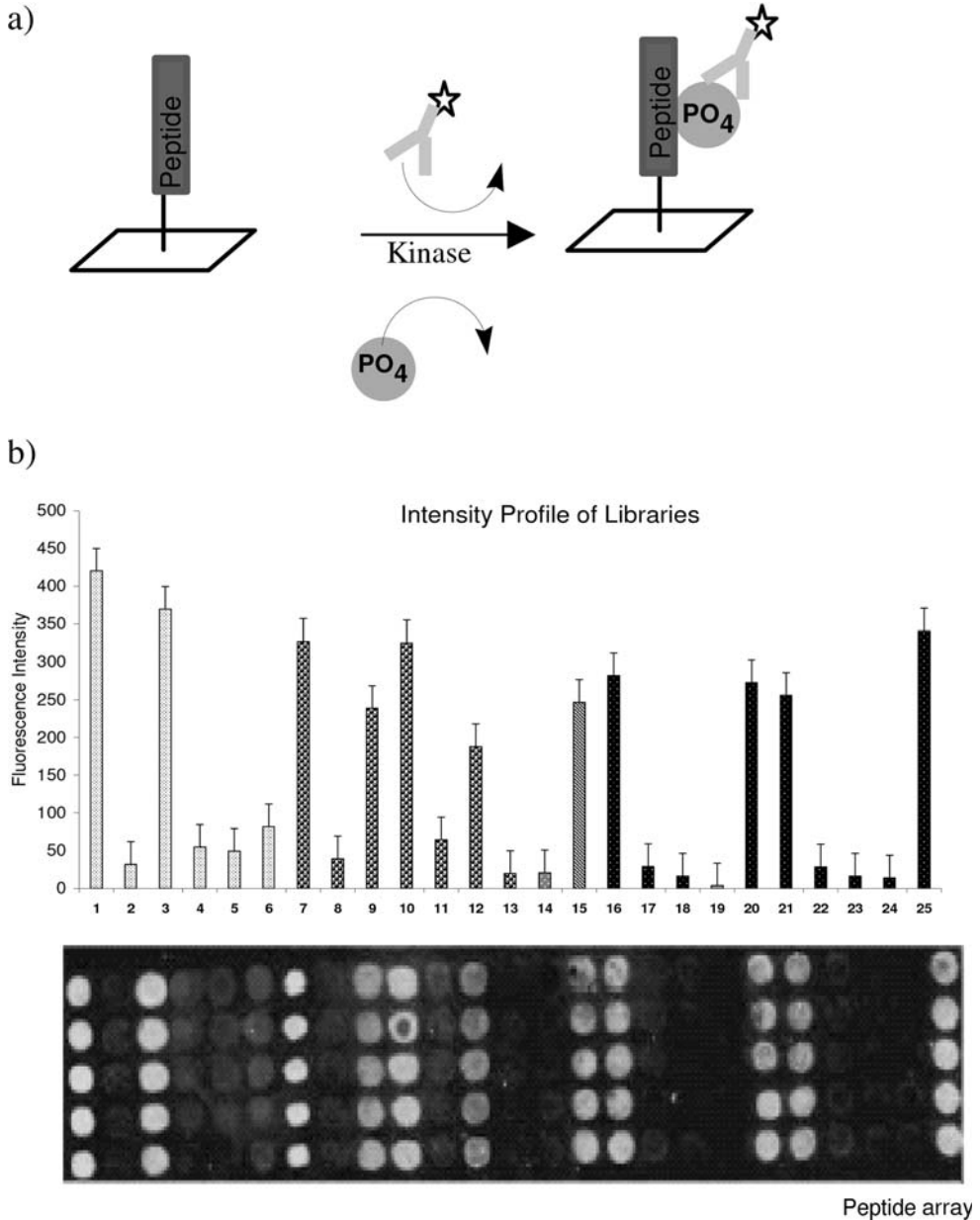


Fig. 12.4 (a) Overall strategy for on-chip detection of kinase substrates using antiphosphoamino antibodies. (b) Intensity profiles and image obtained with the different combinatorial libraries following treatment with p60^{C-SRC} kinase: 1–6 Deletion libraries, 7–13 Alanine-scanning libraries,

16–25 Positional-scanning libraries, 14 Full combinatorial mixture peptide libraries, 15 Original putative substrate. All peptides were synthesized with a CCG linker at the N-terminus and immobilized on PEGylated thioester (22) slides.

mond™ (56) [66]. To further expand the utility of microarray-based techniques in the profiling of kinase substrate specificity, we developed a combinatorial method to characterize kinase substrates [32c].

Traditional microarrays employ a “one-spot one-compound” approach for the direct identification of positive hits. This, however, limits the number of compounds that can be studied in an array to what may be synthesized, thus severely limiting throughput. In order to study a greater diversity of compounds more rapidly and efficiently, it is imperative that alternative strategies like combinatorial peptide synthesis be suitably employed together with array-based methods to facilitate the development of next-generation peptide microarrays. We have shown that, by spotting libraries with various combinations of peptide sequences, it is possible to draw conclusions about positive hits or substrate specificity without generating large numbers of peptide sequences individually [32c]. In our studies, starting with the putative substrate YIYGSK, three different classes of combinatorial libraries – alanine scanning (wherein alanine is systematically substituted in substrate sequences), deletion (where the progressive truncation of residues flanking the phosphorylation site establishes the minimum sequence length required), and positional scanning (where each position in the peptide sequence is individually and systematically “scanned”) libraries were synthesized and immobilized site-specifically via an N-terminal CGG linker onto PEGylated thioester slides (Figure 12.4b). After kinase treatment and detection with the fluorescently labeled antiphosphotyrosine antibody, a unique fingerprint was generated that showed the requirement of minimum of six amino acids backbone with isoleucine and phenylalanine at –1 and +3 positions, respectively, as being critical for phosphorylation.

Measurements of kinetic parameters showed that the kinase reaction occurred in both concentration- and time-dependent manners, thereby demonstrating that antibody-based detection can potentially be extended to the screening of inhibitors/agonists of any kinase with high throughput in a microarray format. Thus, the augmentation of combinatorial strategies for rapid diversity production provides for an even more rapid means of generating and applying peptide microarrays in high-throughput studies.

Protocol 12.4: Detecting Kinases Using Peptide Substrate Arrays

1. The synthesis of the peptides, their immobilization onto a microarray and fluorescence labeling of the antibodies was done as previously described^{a)} [32].
2. Apply 50 μ L of the PKA (2 U in 25 mM Tris, pH 7.4, 15 mM $MgCl_2$, 1 mM DTT (36), 2 mM EGTA, 100 μ M ATP) or p60^{C-SRC} (2 U in 25 mM, pH 7.4, 35 mM $MgCl_2$, 7 mM $MnCl_2$, 0.5 mM EGTA, 100 μ M ATP enzyme solutions.

3. Incubate for different time periods depending on kinase activity. For p60^{c-src} and PKA we applied the kinases on to the slides for 5–6 h in order to obtain good fluorescence before detection.
4. Probe slides with the corresponding labeled antibody for 1–2 h. Antibody solutions are made in PBS with 1% BSA (23). Wash slides with PBST and water; dry and scan using the arrayWoRxTM (Applied Precision, WA, USA) scanner.

Notes

- a) An N-terminal cysteine or biotin (26) may be added to virtually any synthetic peptide. This makes it possible to apply peptides synthesized from various combinatorial strategies, such as positional-scanning libraries, onto an array format. Within reasonable lengths, under 10 amino acid residues, Fmoc synthesis yields sufficiently pure peptides that may be directly applied to an array without extensive purification (also dependent on the nature of the residues within peptide and quality of synthesis).

12.3.2.3 Carbohydrate-modifying Enzymes

Sugars and their conjugates, for example glycoproteins, mediate important cellular events such as cell–cell communication, cell adhesion, signal transduction, the attachment of microbes to cells during infection, and so forth. A major impediment in glycobiology had been the lack of convenient and effective tools for the synthesis and analysis of these complex and structurally diverse molecules [67]. Significant advances in chemical and enzymatic methods for the solid phase synthesis of complex carbohydrates in the past few years have eased the development of tools to explore new vistas in carbohydrate biology and have spawned a new ‘omic level discipline’ – “glycomics” [68].

On the technology front, carbohydrate arrays have been investigated in a series of reports [69]. The usefulness of these arrays for functional enzymatic studies has also been demonstrated [30 b, 30 c, 69]. Houseman and Mrksich reported a carbohydrate microarray in which the Diels-Alder reaction was used for carbohydrate immobilization. They subsequently used the “carbochip” to profile the binding specificities between rhodamine (57)-labeled lectins and different monosaccharides, as well as to characterize the substrate specificity of β -1,4-galactosyltransferase and also quantitatively estimate the inhibitory concentration of methyl α -mannoside (58) for concanavalin A in competitive binding studies [30 b]. In a related study, the same group profiled the binding specificity of lectins against an array of monosaccharide-thiol conjugates immobilized onto a SAM derivatized with maleimide (19) groups [30 c]. Park et al. reported the fabrication of carbohydrate arrays by immobilization of maleimide (19)-terminated carbohydrates on thiol-coated glass slides [70]. Using these arrays, the group demonstrated both qualitative and quantitative bindings of lectins to α -, β -, and N-linked carbohydrates. They were also able to track the glycosylation reactions mediated by β -1,4-galactosyltransferase and α -1,3-fucosyltransferase, thereby showing that, besides aiding in the identification and characterization of novel carbohydrate-binding proteins, the carbohydrate array may be equally amenable for assaying and profiling carbohydrate-modifying enzymes *en masse*.

12.3.2.4 Other Enzymes

Lee et al. performed solid-phase enzyme reactions in the presence of *S*-adenosyl-L-methionine (**59**) to identify substrates for arginine *N*-methyltransferases [71]. The proteins used in the study were isolated from the human brain and subsequently arrayed, at high density, onto polyvinylidene fluoride (PVDF) (**60**) membranes. Other enzymes, such as horseradish peroxidase, alkaline phosphatase (AP) and β -D-glucuronidase, have been studied by Arenkov et al. [72]. The enzymes were applied to colorimetric substrates immobilized in 3D gel pads that produced color or fluorescent precipitates with the progress of enzymatic reaction. The concentration-dependent enzymatic dephosphorylation of ELF-97 phosphate substrate (**61**) by AP, as well as the enzyme's inhibition by the stereospecific noncompetitive inhibitor L-phenylalanine (**62**), was also measured.

The activity of alcohol dehydrogenase, enolase, and pyruvate kinase has been investigated using nanowells [73a]. The nanowell format has been used by Angenendt et al. to profile the inhibition of β -galactosidase activity but the enzymes used in this study were generated *in situ* by cell-free protein translation systems [73b]. Seeberger and coworkers used a small molecule array of aminoglycoside mimetics (**63**) to screen for compounds that inhibits the activity of enzymes responsible for antibiotic resistance in bacteria – the aminoglycoside acetyltransferases [74]. In their screenings, the authors identified 6'- β -Ala-guanidinoribostamycin (**64**) as a tight binder that hampers the activity of these enzymes and thereby showed that this kind of a microarray-based approach can potentially allow the discovery of improved antibiotics that evade current modes of antibiotic resistance in bacteria.

12.3.3

Enzyme Assays Using Other Types of Arrays

Schultz and colleagues reported the development of a peptide nucleic acid (PNA)-tagged, small molecule array that in conjunction with the standard DNA microarray technology can be used to monitor the levels and activities of enzymes on a proteomic scale [75]. Using the split-pool combinatorial method, a small molecule library based on mechanism-based inhibitors of cysteine proteases was synthesized, with each member of the library having a PNA tag. The PNA tags did not alter the activity/selectivity of these small molecules and, at the same time, served as molecular “barcodes” which could be used to decode the library by hybridization to a suitable DNA microarray. In proof-of-concept experiments, an equimolar mixture of six cysteine protease inhibitors, having an acrylamide moiety as the mechanism-based “warhead” which is highly specific towards cysteine proteases, was incubated against individual cathepsins [74a]. The resulting enzyme-inhibitor adducts were fractionated by size-exclusion filtration to remove any free, unreacted small molecule probes. The protein fractions, with PNA-tagged small molecules still covalently attached to them, were then hybridized to a GenFlexTM oligonucleotide microarray. Fluorescence signals generated on the array, as a result of hybridization, enabled the identifi-

cation of small molecules that are highly specific, and thus most reactive, towards the target cathepsin. Alternatively, the utility of the method to monitor enzymatic activity on a proteomic scale was demonstrated by the detection and characterization of caspase-3 activity in crude extracts of cells undergoing apoptosis triggered by granzyme B [74b]. Since the enzyme activity governs its binding to the small molecule, the approach may potentially be used to guide the design of drugs against diseased phenotypes.

Our own inputs into the area of genome-wide identification of enzyme activities using the DNA microarray technology have been through the development of a technique termed “Expression Display” (Figure 12.5a), which combines the advantages of three different techniques: ribosome display, DNA microarray, and activity-based enzyme profiling [76]. In this method, a mixture of proteins was expressed, in a single vessel, from their cDNA library by ribosome display which ensured that each protein member was linked and thus encoded by its own mRNA sequence, via the formation of a stable mRNA–ribosome–protein ternary complex.

Subsequent functional selections using an activity-based small molecule probe (Figure 12.5b) enriched only complexes that contain proteins having the desired enzymatic activity. The identity of these enzymes was then decoded in a high-throughput manner, following hybridization of their tethered mRNA to a DNA microarray.

So far we have demonstrated that this approach can be used to isolate and identify, in a single experiment, representative yeast tyrosine phosphatases from a large pool of unrelated proteins (Figure 12.5c). We envision that in future this work can be extended to the high-throughput screening of other enzyme classes.

12.4 Conclusions

As new roles of enzymes emerge [77], many novel applications of enzyme assays can be envisioned. For instance, an on-chip metabolic engineering approach has been recently demonstrated for the optimization of biochemical pathways [78]. Enzyme assays have matured from mere attempts at the identification of individual enzymes to providing the complete fingerprints in terms of their substrate specificities, kinetic measurements and high-throughput identifications of inhibitors. Such fingerprint analysis could be used to solve the issue of functional convergence of enzymes on a broad scale. On the industrial front, such analysis would enable the batch identification and quality control of enzyme-catalyzed processes. Rapid developments in enzymatic assays will continue to be fueled by the identification of new enzymes stemming from the growth in genome information which has provided a sequence-based framework for data mining. Challenges remain abundant as many of the solution-based enzyme assays are difficult to be executed at high throughput, in a miniaturized

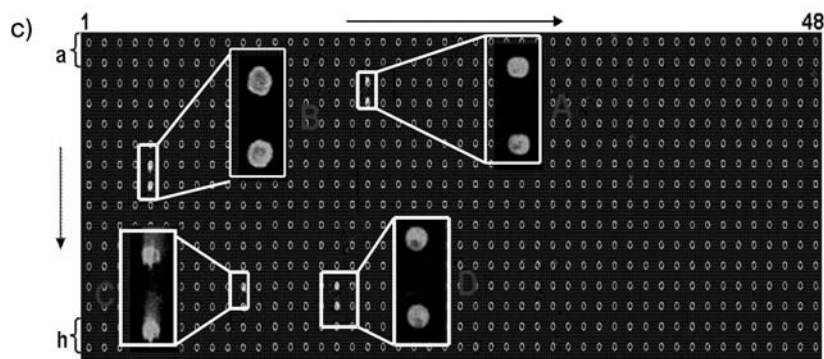
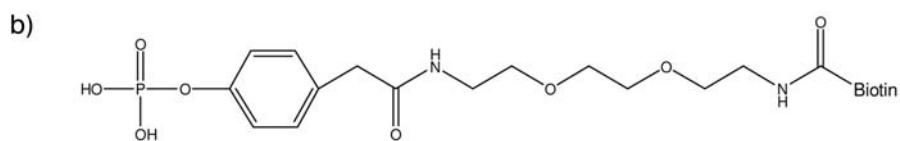
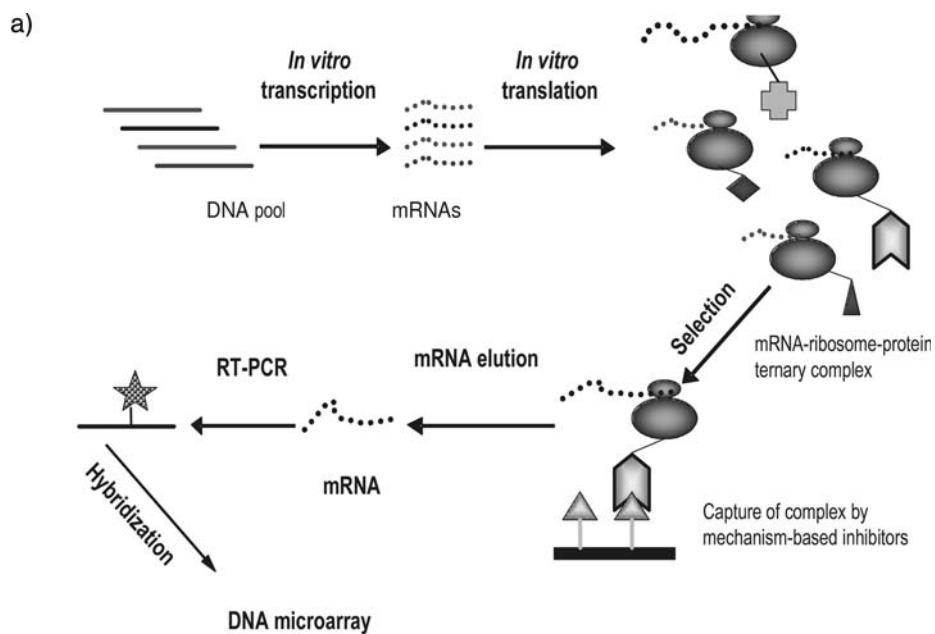


Fig. 12.5 (a) Schematic representation of Expression Display. (b) Structure of the activity-based probe specific for identification of protein tyrosine phosphatases (PTPs). (c) Hybridization of Cy3-labeled reverse transcripts following *in vitro* selection using

activity-based probe onto a “decoding” DNA microarray containing 384 yeast open reading frames (in duplicates). The spotting pattern is indicated by white circles. Zoomed regions represent positive hits.

format with sufficient levels of accuracy. This is because miniaturized assays need to be highly robust, reproducible and extremely sensitive, which is not fulfilled in many solution-based enzyme assays.

Though we have tried informing the readers briefly about the current state of enzyme assays conducted in other formats, we have consciously overlooked the use of microfluidic devices as the means for miniaturization in high-throughput assays [79]. We envisage, however, that such “lab-on-chip” devices, with their sophisticated liquid-handling procedures that allow the precise control of nanoliter-scale liquid flow, should lend them particularly amenable to microarray-based assays for potential hand-free, ultrahigh-throughput studies of enzymes.

Acknowledgments

Funding support was provided by the National University of Singapore (NUS) and the Agency for Science, Technology and Research (A*STAR) of Singapore. S.Q.Y. is the recipient of the 2002 Young Investigator Award (YIA) from the Biomedical Research Council (BMRC).

References

- 1 J. L. Reymond, D. Wahler, *ChemBioChem* **2002**, *3*, 701–708.
- 2 G. Y. J. Chen, M. Uttamchandani, R. Y. P. Lue, M. L. Lesaicherre, S. Q. Yao, *Curr. Top. Med. Chem.* **2003**, *3*, 705–724.
- 3 R. Frank, *Tetrahedron* **1992**, *48*, 9217–9232.
- 4 S. P. Fodor, J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu, D. Solas, *Science* **1991**, *251*, 767–773.
- 5 M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **1995**, *270*, 467–70.
- 6 G. MacBeath, S. L. Schreiber, *Science*, **2000**, *289*, 1760–1763.
- 7 H. Zhu, J. F. Klemic, S. Chang, P. Bertone, A. Casamayor, K. G. Klemic, D. Smith, M. Gerstein, M. A. Reed, M. Synder, *Nat. Genet.* **2000**, *26*, 283–289.
- 8 G. Y. J. Chen, M. Uttamchandani, Q. Zhu, G. Wang, S. Q. Yao, *ChemBioChem* **2003**, *4*, 336–339.
- 9 Q. Zhu, M. Uttamchandani, D. Li, M. L. Lesaicherre, S. Q. Yao, *Org. Lett.* **2003**, *5*, 1257–1260.
- 10 S. Y. D. Yeo, R. C. Panicker, L. P. Tan, S. Q. Yao, *Comb. Chem. High Throughput Screening* **2004**, *7*, 213–221.
- 11 M. Uttamchandani, D. P. Walsh, S. Q. Yao, Y. T. Chang, *Curr. Opin. Chem. Biol.* **2005**, *9*, 4–13.
- 12 a) T. O. Joos, M. Schrenk, P. Höpfl, K. Kröger, U. Chowdhury, D. Stoll, D. Schörner, M. Durr, K. Herick, S. Rupp, K. Sohn, H. Hämmerle, *Electrophoresis* **2000**, *21*, 2641–2650; b) H. Ge, *Nucleic Acids Res.* **2000**, *28*, e3.
- 13 a) R. M. de Wildt, T. Mundy, B. D. Gorick, I. M. Tomlinson, *Nat. Biotechnol.* **2000**, *18*, 989–994; b) B. B. Haab, M. J. Dunham, P. O. Brown, *Genome Biol.* **2001**, *2*, R4; c) J. L. Naffin, Y. Han, H. J. Olivos, M. M. Reddy, T. Sun, T. Kodadek, *Chem. Biol.* **2003**, *10*, 251–259; d) R. Bentes, C. M. Niemeyer, D. Wöhrle, *Chem-BioChem* **2001**, *2*, 686–694; e) www.somalogic.com; f) www.affibody.com; g) www.phylos.com; h) A. Espejo, J. Cote, A. Bednarek et al., *Biochem. J.* **2002**, *367*, 697–702.
- 14 O. Zheng, Z. Takáts, T. A. Blake, B. Gologan, A. J. Guymon, J. M. Wiseman, J. C. Oliver, J. Davisson, R. G. Cooks, *Science* **2003**, *301*, 1351–1354.

- 15 A. Roda, M. Guardigli, C. Russo, P. Pasini, M. Baraldini, *Biotechniques* **2000**, *28*, 492–496.
- 16 D. Guschin, G. Yershov, A. Zaslavsky, A. Gemmell, V. Shick, D. Proudnikov, P. Arenkov, A. Mirzabekov, *Anal. Biochem.* **1997**, *250*, 203–211.
- 17 V. Afanassiev, V. Hanemann, S. Wolfl, *Nucleic Acids Res.* **2000**, *28*, e66.
- 18 S. Kiyonaka, K. Sada, I. Yoshimura, S. Shinkai, N. Kato, I. Hamachi, *Nature Mat.* **2004**, *3*, 58–64.
- 19 R. C. Panicker, X. Huang, S. Q. Yao, *Comb. Chem. High Throughput Screening* **2004**, *7*, 547–556.
- 20 a) J. C. Miller, H. Zhou, J. Kwekei, R. Cavallo, J. Burke, E. D. Butler, B. S. Teh, B. B. Haab, *Proteomics* **2003**, *3*, 56–63; b) I. Caelen, H. Gao, H. Sigrist, *Langmuir* **2002**, *18*, 2463–2467.
- 21 N. V. Avseenko, T. Y. Morozova, F. I. Ataulkhanov, V. N. Morozov, *Anal. Chem.* **2002**, *74*, 927–933.
- 22 a) R. A. Vijayendran, D. E. Leckband, *Anal. Chem.* **2001**, *73*, 471–480; b) H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein, M. Synder, *Science* **2001**, *293*, 2101–2105.
- 23 G. MacBeath, A. N. Koehler, S. L. Schreiber, *J. Am. Chem. Soc.* **1999**, *121*, 7967–7968.
- 24 P. J. Hergenrother, K. M. Depew, S. L. Schreiber, *J. Am. Chem. Soc.* **2000**, *122*, 7849–7850.
- 25 F. Fazio, M. C. Bryan, O. Blixt, J. C. Paulson, C.-H. Wong, *J. Am. Chem. Soc.* **2002**, *124*, 14397–14402.
- 26 D. B. Seeman, S. B. Park, A. N. Koehler, S. L. Schreiber, *Angew. Chem. Int. Ed.* **2003**, *42*, 2376–2379.
- 27 M. Köhn, R. Wacker, C. Peters, H. Schröder, L. Soulère, R. Breinbauer, C. M. Niemeyer, H. Waldmann, *Angew. Chem. Int. Ed.* **2003**, *42*, 5830–5834.
- 28 M. Uttamchandani, D. P. Walsh, S. M. Khersonsky, X. Huang, S. Q. Yao, Y. T. Chang, *J. Comb. Chem.* **2004**, *6*, 862–868.
- 29 J. R. Falsey, M. Renil, S. Park, S. Li, K. S. Lam, *Bioconj. Chem.* **2001**, *12*, 346–353.
- 30 a) B. T. Houseman, J. H. Huh, S. J. Kron, M. Mrksich, *Nat. Biotechnol.* **2002**, *20*, 270–274; b) B. T. Houseman, M. Mrksich, *Chem. Biol.* **2002**, *9*, 443–454; c) B. T. Houseman, E. S. Gawalt, M. Mrksich, *Langmuir* **2003**, *19*, 1522–1531.
- 31 C. Olivier, D. Hot, L. Huot, N. Olivier, O. El-Mahdi, C. Gouyette, T.-H. Dinh, H.-G. Masse, Y. Lemoine, O. Melnyk, *Bioconj. Chem.* **2003**, *14*, 430–439.
- 32 a) M. L. Lesaichere, M. Uttamchandani, G. Y. J. Chen, S. Q. Yao, *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2079–2083; b) M. L. Lesaichere, M. Uttamchandani, G. Y. J. Chen, S. Q. Yao, *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2085–2088; c) M. Uttamchandani, E. W. S. Chan, G. Y. J. Chen, S. Q. Yao, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2997–3000.
- 33 P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776–779.
- 34 L. R. Paborsky, K. E. Dunn, C. S. Gibbs, J. P. Dougherty, *Anal. Biochem.* **1996**, *234*, 60–65.
- 35 C. D. Hodneland, Y. S. Lee, D. H. Min, M. Mrksich, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5048–5052.
- 36 N. Kindermann, M. George, N. Johnsson, K. Johnsson, *J. Am. Chem. Soc.* **2003**, *125*, 7810–7811.
- 37 J. Yin, F. Liu, X. Li, C. T. Walsh, *J. Am. Chem. Soc.* **2004**, *126*, 7754–7755.
- 38 a) M. L. Lesaichere, R. Y. P. Lue, G. Y. J. Chen, Q. Zhu, S. Q. Yao, *J. Am. Chem. Soc.* **2002**, *124*, 8768–8769; b) R. Y. P. Lue, G. Y. J. Chen, Y. Hu, Q. Zhu, S. Q. Yao, *J. Am. Chem. Soc.* **2004**, *126*, 1055–1062.
- 39 a) L. P. Tan, G. Y. J. Chen, S. Q. Yao, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5735–5738; b) L. P. Tan, R. Y. P. Lue, G. Y. J. Chen, S. Q. Yao, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 6067–6070; c) S. Y. D. Yeo, R. Srinivasan, G. Y. J. Chen, S. Q. Yao, *Chem. Eur. J.* **2004**, *10*, 4664–4672.
- 40 a) S. R. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. B. Kucera, C. A. Hiryonen, J. J. Pelletier, H. Paulas, M.-Q. Xu, *Gene* **1997**, *192*, 277–281; b) S. R. Chong, G. E. Montello, A. Zhang, E. J. Cantor, W. Liao, M.-Q. Xu,

- J. Benner, *Nucleic Acids Res.* **1998**, *26*, 5109–5115.
- 41 a) T. Tolbert, C.-H. Wong, *J. Am. Chem. Soc.* **2000**, *122*, 5421–5428; b) S. Y. D. Yeo, R. Srinivasan, M. Uttamchandani, G. Y. J. Chen, Q. Zhu, S. Q. Yao, *Chem. Commun.* **2003**, *23*, 2870–2871; c) R. Srinivasan, S. Q. Yao, S. Y. D. Yeo, *Comb. Chem. High Throughput Screening* **2004**, *7*, 597–604.
- 42 J. E. Cronan, K. E. Reed, *Methods Enzymol.* **2000**, *326*, 440–458.
- 43 M. Y. He, M. J. Taussig, *Nucleic Acids Res.* **2001**, *29*, e73.
- 44 a) N. Nemoto, E. Miyamoto-Sato, H. Yanagawa, *FEBS Lett.* **1999**, *462*, 43–46; b) E. Miyamoto-Sato, N. Nemoto, K. Kobayashi, H. Yanagawa, *Nucleic Acids Res.* **2000**, *28*, 1176–1182; c) Y. Kawahashi, N. Doi, H. Takashima, C. Tsuda, Y. Oishi, R. Oyama, M. Yonezawa, E. Miyamoto-Sato, H. Yanagawa, *Proteomics* **2003**, *3*, 1236–1243; d) S. R. Starck, R. W. Roberts, *RNA* **2002**, *8*, 890–903; e) S. R. Starck, H. M. Green, J. Alberola-Ila, R. W. Roberts, *Chem. Biol.* **2004**, *11*, 999–1008.
- 45 a) R. Y. P. Lue, S. Y. D. Yeo, L. P. Tan, M. Uttamchandani, G. Y. J. Chen, S. Q. Yao, *Protein Microarrays*, ed. M. Schemm, Jones and Bartlett Publishers, Sudbury, MA, **2004**; b) R. Y. P. Lue, G. Y. J. Chen, Q. Zhu, M. L. Lesaichere, S. Q. Yao, *Methods Mol. Biol.* **2004**, *264*, 85–100.
- 46 J. P. Goddard, J. L. Reymond, *Trends Biotechnol.* **2004**, *22*, 363–370.
- 47 Y. Hu, X. Huang, G. Y. J. Chen, S. Q. Yao, *Mol. Biotechnol.* **2004**, *28*, 63–76.
- 48 a) Y. Liu, M. P. Patricelli, B. F. Cravatt, *Proc. Natl Acad. Sci. USA* **1999**, *96*, 14694–14699; b) D. Kidd, Y. Liu, B. F. Cravatt, *Biochemistry* **2001**, *40*, 4005–4015.
- 49 a) L. Faleiro, R. Kobayashi, H. Fearnhead, Y. Lazebnik, *EMBO J.* **1997**, *16*, 2271–2281; b) M. L. Liau, R. C. Panicker, S. Q. Yao, *Tetrahedron Lett.* **2003**, *44*, 1043–1046; c) D. Greenbaum, K. F. Medzihradzky, A. Burlingame, M. Bogyo, *Chem. Biol.* **2000**, *7*, 569–581; d) T. Nazif, M. Bogyo, *Proc. Natl Acad. Sci. USA* **2001**, *98*, 2967–2972; e) G. Wang, M. Uttamchandani, G. Y. J. Chen, S. Q. Yao, *Org. Lett.* **2003**, *5*, 737–740.
- 50 a) G. C. Adam, E. J. Sorensen, B. F. Cravatt, *Nat. Biotechnol.* **2002**, *20*, 805–809; b) L. C. Lo, T. L. Pang, C. H. Kuo, Y. L. Chiang, H. Y. Wang, J. J. Lin, *J. Proteome Res.* **2002**, *1*, 35–40; c) Q. Zhu, X. Huang, G. Y. J. Chen, S. Q. Yao, *Tetrahedron Lett.* **2003**, *44*, 2669–2672; d) Q. Zhu, A. Girish, S. Chattopadhyaya, S. Q. Yao, *Chem. Commun.* **2004**, *13*, 1512–1513.
- 51 a) M. C. Hagenstein, J. H. Mussnug, K. Lotte, R. Plessow, A. Brockhinke, O. Kruse, N. Sewald, *Angew. Chem. Intl. Ed.* **2003**, *42*, 5635–5638; b) A. Saghatelian, N. Jessani, A. Joseph, M. Humphrey, B. F. Cravatt, *Proc. Natl Acad. Sci. USA* **2004**, *101*, 10000–10005; c) E. W. S. Chan, S. Chattopadhyaya, R. C. Panicker, X. Huang, S. Q. Yao, *J. Am. Chem. Soc.* **2004**, *126*, 14435–14446.
- 52 S. Mizukami, K. Kikuchi, T. Higuchi, Y. Urano, T. Mashima, T. Tsuruo, T. Nagano, *FEBS Lett.* **1999**, *453*, 356–360.
- 53 D. J. Matthews, J. A. Wells, *Science* **1993**, *260*, 1113–1117.
- 54 a) B. J. Backes, J. L. Harris, F. Leonetti, C. S. Craik, J. A. Ellman, *Nat. Biotechnol.* **2000**, *18*, 187–193; b) J. L. Harris, P. B. Alper, J. Li, M. Rechsteiner, B. J. Backes, *Chem. Biol.* **2001**, *8*, 1131–1141.
- 55 B. E. Turk, L. C. Cantley, *Methods* **2004**, *32*, 398–405.
- 56 B. E. Turk, L. L. Huang, E. T. Piro, L. C. Cantley, *Nat. Biotechnol.* **2001**, *19*, 661–667.
- 57 D. N. Gosalia, S. L. Diamond, *Proc. Natl Acad. Sci. USA* **2003**, *100*, 8721–8726.
- 58 C. B. Park, D. S. Clark, *Biotechnol Bioeng.* **2002**, *78*, 229–235.
- 59 C. M. Salisbury, D. J. Maly, J. A. Ellman, *J. Am. Chem. Soc.*, **2002**, *124*, 14868–14870.
- 60 a) D. Wahler, F. Badalassi, P. Crotti, J. L. Reymond, *Angew. Chem. Intl. Ed.* **2001**, *40*, 4457–4460; b) R. P. Carlon, N. Jourdain, J. L. Reymond, *Chem. Eur. J.* **2000**, *6*, 4154–4162; c) G. Klein, J. L. Reymond, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1113–1116.
- 61 J. Eppinger, D. P. Funeriu, M. Miyake, L. Denizot, J. Miyake, *Angew. Chem. Intl. Ed.* **2004**, *43*, 3806–3810.

- 62 G. Cassirini, *FEBS Lett.* **1992**, 307, 66–70.
- 63 K. S. Lam, M. Renil, *Curr. Opin. Chem. Biol.* **2002**, 6, 353–358.
- 64 W. R. G. Dostmann, M. S. Taylor, C. K. Nickl, J. E. Brayden, R. Frank, W. J. Tegge, *Proc. Natl Acad. Sci. USA* **2000**, 97, 14772–14777.
- 65 a) M. Schutkowski, U. Reimer, S. Panse, L. Dong, J. M. Lizcano, D. R. Alessi, J. Schneider-Mergener, *Angew. Chem. Int. Ed.* **2004**, 43, 2671–274; b) S. Panse, L. Dong, A. Burian, R. Carus, M. Schutkowski, U. Reimer, J. Schneider-Mergener, *Mol. Divers.* **2004**, 8, 291–299; c) J. M. Lizcano, M. Deak, N. Morrice, A. Kieloch, C. J. Hastie, L. Dong, M. Schutkowski, U. Reimer, D. R. Alessi, *J. Biol. Chem.* **2002**, 271, 27839–27849.
- 66 K. Martin, T. H. Steinberg, L. A. Cooley, K. R. Gee, J. M. Beechem, W. F. Patton, *Proteomics* **2003**, 3, 1244–1255.
- 67 P. Sears, C.-H. Wong, *Science* **2001**, 291, 2344–2350.
- 68 a) O. J. Plante, E. R. Palmacci, P. H. Seeberger, *Science* **2001**, 291, 1523–1527; b) D. M. Ratner, E. W. Adams, M. D. Disney, P. H. Seeberger, *ChemBioChem* **2004**, 5, 1375–1383.
- 69 a) S. Fukui, T. Feizi, C. Galustian, A. M. Lawson, W. Chai, *Nature Biotechnol.* **2002**, 20, 1011–1017; b) D. Wang, S. Liu, B. J. Trummer, C. Deng, A. Weng, *Nature Biotechnol.* **2002**, 20, 275–281.
- 70 S. Park, M.-R. Lee, S.-J. Pyo, I. Shin, *J. Am. Chem. Soc.* **2004**, 126, 4812–4819.
- 71 J. Lee, M. T. Bedford, *EMBO Rep.* **2002**, 3, 268–273.
- 72 P. Arenkov, A. Kukhtin, A. Gemmell, S. Voloshchuk, V. Chupeeva, A. Mirzabekov, *Anal. Biochem.* **2000**, 278, 123–131.
- 73 a) H. R. C. Dietrich, J. Knoll, L. R. van den Doel, G. W. K. van Dedem, P. A. S. Daran-Lapujade, L. J. van Vliet, R. Moerman, J. T. Pronk, I. T. Young, *Anal. Chem.* **2004**, 76, 4112–4117; b) P. Angenendt, L. Nyarsik, W. Szaflarski, J. Glökler, K. H. Nierhaus, H. Lehrach, D. J. Cahill, A. Lueking, *Anal. Chem.* **2004**, 76, 1844–1849.
- 74 M. D. Disney, S. Magnet, J. S. Blanchard, P. H. Seeberger, *Angew. Chem. Int. Ed.* **2004**, 43, 1591–1594.
- 75 a) N. Winessinger, J. L. Harris, B. J. Backes, P. G. Schultz, *Angew. Chem. Int. Ed.* **2001**, 40, 3152–3155; b) N. Winessinger, S. Ficarro, P. G. Schultz, J. L. Harris, *Proc. Natl Acad. Sci. USA* **2002**, 99, 11139–11144.
- 76 H. Yi, G. Y. J. Chen, S. Q. Yao, *Angew. Chem. Int. Ed.* **2005**, 44, 1048–1053.
- 77 D. A. Hall, H. Zhu, X. Zhu, T. Royce, M. Gerstein, M. Synder, *Science* **2004**, 306, 482–484.
- 78 G. Y. Jeol, G. Stephanopoulos, *Science* **2004**, 304, 428–431.
- 79 J. Wang, *Electrophoresis* **2002**, 23, 713–718.

Subject Index

a

ABTS 81 ff., 90
 acetylcholinesterases 255
 acetyl-CoA synthetase 45
 acrylamide 265, 365
 acrylodan 260
 actin 324
 acylase 4, 50
 adenylate cyclase 167
 affinity tag 175, 320, 327, 339
 agar 79, 140 ff., 163, 171, 279
 agarase 234
 agarose 229, 243, 245, 335
 alcohol dehydrogenase 50
 aldolase 50, 159, 216 f.
 alkaline phosphatase 5, 165 f., 223, 251,
 276, 350 f., 356
 amidase 4 f., 105 ff., 112 ff., 122 ff., 143,
 216, 234
 amidohydrolase 170
 amikacin 224
 amino acid 20
 – amide racemase (AR) 109 f.
 – dehydrogenase 105
 – oxidase 105
α-amino- ϵ -caprolactam (ACL) 129 f.
 7-amino cephalosporanic acid 82
 aminoacyl-tRNA synthetase 159, 190
 4-aminoantipyrine (4AAP) 86 f., 144
 7-aminocoumarin 306 ff.
 7-aminodesacetoxycephalosporanic acid
 (7-ADCA) 95
 aminoglycoside 6'-N-acetyl-transferase
 224
 aminoglycoside acetyltransferase 356
 aminoglycoside-modifying 6'-acetyltrans-
 ferase-2''-phosphotransferase 224
 aminopeptidase 276, 278
 aminotranferase 103, 189
 ammonia lyase 105 f.

amoxicillin 102
 ampicillin 102, 146, 195, 202
 Amplex Red 251
 amylase 224, 232, 234, 278
 anthracene 90
 9,10-anthrahydroquinone 90
 9,10-anthroquinone 90
 antibiotic cephalixin 95
 antibiotic resistance 205, 223, 226, 356
 antibody 5, 11, 47 f., 165 ff., 175 f., 190,
 222, 279, 335, 342, 344, 352 ff.,
 apoptosis 308, 324, 326, 348, 357
 apyrase 5
 AR 110 f.
 archaea 228 f., 235
 arginine N-methyltransferases 356
 astaxanthin 153
 asymmetric 41 f., 47, 49, 54 f., 68, 124
 ATP 154, 167, 273, 352, 354

b

BAC 227, 230 ff., 235
 Baeyer-Villiger monooxygenase 67
 barnase 170
 benzol formic acid 47
 biotin 154, 166, 175, 184, 204, 223, 245,
 265, 267, 327, 337 f., 355
 – acceptor peptide (BAP) 134
 – ligase 153, 339
 – streptavidin 198
 biphenyl dioxygenase (BPDO) 86
 bodipy 249, 251, 260
 bovine pancreatic protease 34
 bovine serum albumin (BSA) 49
 bromothymol blue 349
 BSA 246, 248, 259, 315, 337, 347, 351,
 355

c

cAMP 352
Candida 246
 – *antarctica* (CAL) 46
 – *antarctica* lipase 28
 – *rugosa* lipase 28
 capillary array electrophoresis (CAE) 69
 carbamoylase 95, 105f.
 carbonic anhydrase 20
 caspase 306ff., 313, 317, 321, 324, 326f., 357
 catalase 82, 170
 cDNA 192, 198, 202, 217, 323ff., 333, 357
 cefotaxime 205ff.
 ceftazidime 207
 cellobiose 211ff.
 cellulase 18, 145, 224, 232, 234
 cellulose 68
 cephadroxyl 102
 cephalixin 96, 102
 cephalosporin 199, 204, 207ff., 221, 224
 – antibiotics 82
 cephalosporinase 199f., 202f., 205
 cephem 199f., 202, 206
 chemical complementation 154f., 183
 chemiluminescence 344
 CHES (2-[*N*-cyclohexylamino]ethanesulfonic acid) 24
 chloramphenicol acetyltransferase 172
 4-chloronaphthol 82, 146
 6 chlorophenol red 23
 cholesterol esterase 30
 cholinesterase 20
 chorismate mutase 155f., 187
 chromatography 254, 287, 305, 326
 chromogenic 17ff., 31ff., 147, 242, 274, 276, 279, 284, 305, 349
 chymotrypsin 257, 276, 278, 313, 346, 348
α-chymotrypsin 30
 circular dichroism (CD) 71
 cluster 189, 227, 235, 230, 266
 – analysis 265, 295
 clustering 291f., 295, 297
 cocktail 277f., 287ff., 299
 cofactor 77f., 80, 82, 168, 274, 304, 324, 344
 combinatorial 305f., 309, 314f., 349, 353ff.
 – mutagenesis 187
 copper 113, 242, 287
 – oxidase 89
 copper-calcein 4
 coumarin 279, 291, 306ff., 315, 349ff.

crotonase 157
 cruzain 316
 crystal violet 144
 cutinase 338
 cycloartenol synthase 157
α-cyclodextrin (*α*-CD) 69
 cytochrome 83f., 151, 172
 – P 450 85
 cytoplasm 153

d

d-3-mercapto-2-methylpropionic acid 110
 D-AO 82f.
 decarboxylase 20, 124f.
 dehydratase 155f., 170
 dehydrogenase 45ff., 77ff., 86ff., 106, 108f., 111, 156, 171, 173, 234, 273f., 279, 356
 deracemization 147
 dexamethasone 154, 191, 194, 212
 dextran 336
 dianisidine 82f.
 2,6-dichloroquinone-4-chloroimide 86
 digoxigenin 223
 dihydrofolate reductase 195, 200
 2,3-dihydroxybiphenyl-1,2-dioxygenase (BpHC) 86
 dinitrophenol 251
 dioxygenase 87ff., 150, 234
 directed evolution 54f., 67f., 73, 83f., 86, 97, 99, 139ff., 149, 151f., 154, 158f., 163f., 168f., 171ff.
 DNA exonuclease III 5
 DNA polymerase 5, 175
 DNA shuffling 43, 83, 86, 125, 152, 157, 171
D-phenylalanine 111
D-phenylglycine 95
D-*p*-hydroxyphenylglycine 95, 102f.
 – aminotransferase 102f.

e

E. coli 187f., 190, 195, 205, 228, 230, 234, 320f., 324, 339
 – clones 110
 Edman 305, 313, 316, 327
 EDTA 248, 253f., 259, 341
 elastase 5, 327
 electrophoresis 53, 229, 254, 256, 259, 263, 324, 327
 ELISA 11, 333
 enantioselective 41ff., 53f., 62, 68, 72f., 147

enantioselectivity 27 ff., 35 f., 42 f., 45 f.,
48 f., 55, 65, 69, 71, 126 f., 149
 endoglucanase 212
 endonuclease 175, 229
 enolase 157, 356
 epoxide 284, 287
 – hydrolase 50, 280, 285, 287, 349 ff.
 esterase 18 f., 25 f., 29 f., 33, 43 ff., 58, 62,
144, 168, 171, 173, 216, 241 ff., 254 ff.,
263, 267, 274 ff., 282 ff., 292 f., 296 f.,
349 ff.
 exosite 325 f.
 Expression Display 334, 357 f.

f

FACS 168, 175, 177 ff.
 FAD 78, 82
 family shuffling 172 f.
 fatty acid 242 ff., 259, 263, 265 f.
 ficin 5
 fingerprinting 239 ff.
 flavoprotein D-amino acid oxidase
(D-AO) 82
 fluorescein 223, 264 f.
 fluorescence-activated cell sorting
(FACS) 159
 fluorogenic 45, 50, 177, 179, 242, 279,
284, 293, 299, 334, 344, 348 f., 351
 – P450 172
 fluorophosphonate 264 f., 345
 FMN 78
 formazan 78 ff.
 FRET 11, 168, 245, 260, 262, 299, 348
 fructose bis-phosphate aldolase 5
 fungal 81, 173, 221
 – laccase 90

g

galactose oxidase 80 ff. 146
 galactosidase 18, 188, 196, 199, 202 f., 274,
276, 278, 355 f.
 GC 232 f.
 genetic complementation 141
 gentamicin 224
 geranyl-geranyl diphosphate synthase
(*crtB*) 151
 GFP 170, 172
 glucosidase 276, 278
 glycosidase 18, 211 ff., 217, 265, 274
 glycosyl fluoride 211 f., 144
 glycosyl transferase 25, 211, 216
 glycosynthase 144 f., 154 f., 184, 204, 210,
213

granzyme A 326
 granzyme B 307, 357
 guaiacol 84

h

haloalkane dehalogenase 25
 heparinase I 5
 heterologous 228, 230, 234 f.
 hexokinase 20
 hormone sensitive lipase 247, 259, 263
 HPLC 32, 49, 56, 68 f., 72, 81 ff., 114,
150 f., 263, 277, 288, 299
 hyaluronidase 5
 hydantoinease 95
 hydratase 105
 4-hydrazino-7-nitro-2,1,3-benzoxadiazole 20
 2-hydroxybiphenyl-2-monooxygenase
(HpbA) 85 f.
 3-hydroxyindole 151
 hydroxylation 86 ff., 150 f., 159
 4-hydroxytamoxifen 158 f.

i

imaging 11, 60, 81 f., 146, 150, 242
 immobilization 334 f., 348, 354 f.
 immobilized 167 f., 175 f.
 immunoassay 5, 47 f.
in vitro compartmentalization (IVC) 163
 indicator 17, 19, 22, 284
 indigo 89, 151, 159
 indirubin 89
 indole 69, 89, 159
 inhibitor 82, 165, 177, 184, 198, 204, 229,
242, 254, 304, 306, 328, 342 f., 344 ff.,
356 f.
 intein 157, 338 f., 341, 343
 isopenicillin-N-synthase 224
 IVC 165, 174 ff.

k

kanamycin 158
 α -ketoglutarate 103, 108, 111
 kinase 170, 198, 204, 273, 275, 333 f., 345,
351, 354, 356
 knockout 184, 189, 210, 323

l

laccase 77, 89, 170, 173
 β -lactam antibiotics 102
 lactamase 191, 202, 204 ff., 224, 279
 LacZ 172
 L-aminopeptidase 122
 lectins 355

- lipase 18f., 25, 43ff., 58, 62, 65, 67, 72f.,
140, 171, 224, 234, 241ff., 254ff., 260,
262f., 267, 272, 274ff., 280ff., 292ff., 349
- lipoprotein lipase 246
- lyase 234
- lysozyme 343
- m**
- magnetic resonance imaging (MRI) 11
- mammalian serum paraoxonase (PON1)
171
- mass spectrometry 245, 262, 305, 312, 327
- meso*-1,4-diacetoxy-cyclopentene 54ff.
- metabolite 184f., 203, 221, 223, 225, 235
- metagenome 41f., 97, 226f.
- 2-methoxyphenol 86
- α*-methyl leucine amide 123f., 127ff.
- α*-methylbenzylamine 147ff.
- 2-methyldecanoic acid 45
- α*-methylphenylglycine amide 125f.
- 4-methylpyridine 87
- methyltransferases (MTases) 175
- α*-methylvaline 113f.
- amide 113f.
- microarray 51f., 232ff., 262, 273, 314f.,
317, 319, 333ff., 344ff., 358
- microchips 69
- MNBDH 84
- monoamine oxidase (MAO) 11, 147
- monooxygenase 68, 85ff., 150f., 279f.
- mutagenesis 8, 33, 35, 41, 43, 83ff., 141,
143, 147, 157, 159, 171, 173, 186, 188ff.,
215
- mutation 35, 84, 139, 143f., 148f., 153,
169, 174, 184, 188ff., 208f., 323, 325
- mutator strain 33, 144, 150
- n**
- N'-[5'-phosphoribosyl]formimino]-5-aminoi-
midazole-4-carboxamide ribonucleotide
(ProFAR) isomerase (HisA) 157
- N*-acetyl-L-methionine 4
- N*-acyl sulfinamide 25
- NAD(P)/NAD(P)H 17, 19, 45 ff, 77ff., 85,
88, 101, 105, 108, 111, 273ff., 279
- naphthalene 88, 150
- dioxygenase 85
- naphthol 18, 20, 88
- NBD 243, 247, 249f. 253, 255ff., 263
- NBT/PMS 80
- neutral red 144
- NHS 336, 351
- nitrilase 25, 55, 105f., 234
- nitrile 105
- hydratase 95
- nitroaniline 85
- nitroblue tetrazolium (NBT) 78
- nitrocefin 199, 202
- nitrocellulose 79, 141, 143, 150, 152, 335
- nitrophenol 23, 242, 254f., 278, 282, 284
- 4-nitrophenol 22, 24, 27ff.
- 4-nitrophenyl 2-methylpropanoate 28
- nitrous acid 155
- NMR 30, 33, 58ff., 100, 115ff., 125,
127ff., 132
- nuclease 177
- nucleotide diphosphate kinase (NDP-K)
173
- o**
- o*-cresol 86
- o*-dianisidine 83
- o*-hydroxylation 85
- oligonucleotide 54, 69, 188f., 208, 210,
223f., 232f., 317, 319, 356
- OmpA 168
- OmpT 314
- oxidase 77, 80, 82, 90, 106, 109, 144,
147ff., 170, 216
- oxidoreductase 77f., 80, 91
- oxovaline 189
- oxygenase 77, 85, 89
- oxynitrilase 159
- p**
- P450 87ff., 150f., 159, 170
- pancreatic lipase 242, 257
- papain 5, 316
- para*-nitrophenolate 87
- PCR 8, 11, 43, 45, 65, 67, 69, 79, 97, 106,
108, 125, 143, 171, 174, 189, 214f., 226,
232ff., 262, 343
- PEG 265, 314, 337, 353f.
- PEG-fluorescein 264f.
- penicillin 221, 224
- G acylase 95f.
- peptide 204, 288, 303f., 306, 313, 323,
227f., 333ff., 348, 356
- peroxidase 81ff., 90, 144, 146, 150, 166,
168, 223, 251, 267, 356
- PFE 33, 36
- pH indicator 20ff., 31ff., 48, 144
- phage display 165, 167f., 172, 320, 322,
351
- phenotype 140, 184ff., 208, 266, 323, 357
- phenylacetaldehyde 124f.

phenylalanine 143, 155 f.
 1-phenylethanol 72
 phosphatase 50, 273, 275 f., 278, 346 f.,
 349 f., 357 f.
 phosphoamidase 276
 phospholipase 11, 18, 224, 243, 245, 250
 phosphoribosylanthranilate (PRA)
 isomerase (TrpF) 157
 phosphotriesterase 173, 175 f.
 phthalaldehyde 106 ff.
 phytoene desaturase 152 f.
 phytoene synthase (*crtE*) 151
 PLA₂ 250 f.
 PLC 251 f.
 PLD 251
 PLP 103
p-nitroaniline 18 f.
p-nitrophenol 18 f., 36, 176
p-nitrophenolate 43
 poly R-478 90
 polychlorinated biphenyl (PCB) 86
 polyketide synthase 234
 polymerase 43, 79, 97, 143, 167 f., 225 f.,
 341
 PON 173
 PON1 171
 porcine pancreatic lipase 28
 positional scanning library 273, 308, 312,
 353, 355
 positron emission tomography (PET) 11
 prephenate dehydratase 156
 prochiral 54 f. 58, 62, 68, 77, 127
 profiling 254, 262, 266, 277 f., 284, 303,
 334, 345, 347, 349, 351, 353, 355, 357
 pronase 5
p-rosaniline 79
 protease 18, 20, 25, 95, 224, 232, 234,
 241, 254, 262, 273 ff., 287 ff., 291, 299,
 303, 341 f., 345 ff.
 proteasome 323
 proteome 254, 262 ff., 322 ff., 338, 345
Pseudomonas 235, 246, 256, 260
 – *aeruginosa* 65, 67
 – *cepacia* lipase 28
 – *diminota* phosphotriesterase 176
 – *diminuta* 171
 – *fluorescens* esterase (PFE) 35 f., 46, 144
 – *putida* 102, 110, 150
 PSSCL 306, 308 f., 311 f., 315
 puromycin 340, 343
p-xylene 89
 pyrroloquinoline quinine (PQQ) 78

q

quencher 244 f., 247, 251, 311 f., 314 ff.
 Quick *D* 30 f., 33 f.
 Quick *E* 28 ff., 36 ff. 284
 Quick *S* 31

r

racemase 111, 129 ff.
 radioactive 222 ff., 233, 242, 245 f., 249,
 252, 344
 radioactivity 222, 247, 344, 352
 random screening 139 f.
 rat glucocorticoid receptor 192, 195
 resorufin 18, 28, 33 ff., 243, 251, 279, 284
 – tetradecanoate 32
 rhodamine 242 f., 252, 264 ff., 316 f., 319,
 355
 ribitol 157
 RNAi 323 f.
 RNase A 5

s

safranine 243
 SAM 337 f., 352, 355
 scytalone dehydratase 203
 SDS-PAGE 257 f., 265 ff., 324 f.
 shuffling 174, 189, 215, 299
 solubility 174
 solketal 20, 22, 25, 29
 solubility 23, 36, 81, 169 ff., 309
 spectinomycin 202
 sphingomyelinase 245, 252 f.
 SPOT 273, 333, 351
 starch 232
 staurosporine 320
 streptavidin 166, 204, 245, 313, 327, 338
 streptavidin/HRP 154
 subtilisin 5, 25, 30, 316
 sulfinamide 25
 suppressor 323, 326
 synthetase 159, 338

t

terpene cyclase 157
 2-*tert*-butylphenol 86
 tetracyclins 221
 thermolysin 95
 thermophile 25, 33, 35
 thermostability 79, 81, 83, 86, 141, 144,
 183, 190
 thermostable 142 f., 171
 thiolactonase 177

- three-hybrid 154f., 168, 183, 191f., 194, 198ff., 210ff.
 - thrombin 308f., 315, 317, 349
 - thymol blue 24
 - TLC 241f., 252, 284
 - TMB 83f.
 - transaldolase 279
 - transferase 338, 343
 - transketolase 159, 279
 - translation 204, 324f., 356
 - translational 217, 230, 304, 351
 - triacylglycerol 242ff., 256
 - hydrolase 266
 - lipase 245
 - triosephosphate isomerase 188
 - trypsin 276, 278, 314, 346f., 350f.
 - tryptophan 157
 - two-hybrid assay 190, 192, 196
 - tyrosine 156, 159
- u**
- ubiquitin ligase 323
 - umbelliferone 6, 48ff., 145, 278f., 282
- v**
- vitamin 184ff., 221, 317
- y**
- yeast three-hybrid 192
 - yeast two-hybrid 325f.