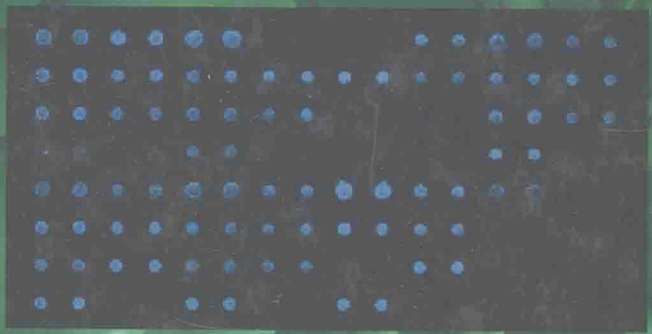
METHODS IN MOLECULAR BIOLOGY™ 310

Chemical Genomics

Reviews and Protocols

Edited by

Edward D. Zanders



HUMANA PRESS

METHODS IN MOLECULAR BIOLOGY™

Chemical Genomics

Reviews and Protocols

Edited by

Edward D. Zanders, PhD

CamBP Ltd., Cambridge, UK

HUMANA PRESS TOTOWA, NEW JERSEY

999 Riverview Drive, Suite 208
Totowa, New Jersey 07512

www.humanapress.com

This publication is printed on acid-free paper. ∞
ANSI Z39.48-1984 (American Standards Institute)

Permanence of Paper for Printed Library Materials.
Cover illustration: Background: Imantinib co-crystalized with c-Abl kinase domain. Used with the permission of Dr. P. M. Dean, CSO, DeNovo Pharmaceuticals, Histon, Cambridge, UK. Foreground: Image of protein array. Figure 4 from Chapter 14, "Fabrication of Protein Function Microarrays for Systems-Oriented Proteomic Analysis," by Jonathan M. Blackburn and Darren J. Hart. Used with the permission of the authors.

Production Editor: Tracy Catanese
Cover design by Patricia F. Cleary

For additional copies, pricing for bulk purchases, and/or information about other Humana titles, contact Humana at the above address or at any of the following numbers: Tel.: 973-256-1699; Fax: 973-256-8341; E-mail: orders@humanapr.com; or visit our Website: www.humanapress.com

Printed in the United States of America. 10 9 8 7 6 5 4 3 2 1

e-ISBN: 1-59259-148-6

Library of Congress Cataloging in Publication Data

Chemical genomics : reviews and protocols / edited by Edward D. Zanders.
p. ; cm. -- (Methods in molecular biology ; v. 310)
Includes bibliographical references and index.
ISBN 1-58829-399-8 (alk. paper)
1. Genomics. 2. Proteomics. 3. Pharmacogenetics. 4. Biochemical genetics. 5. DNA microarrays. 6. Combinatorial chemistry.
[DNLM: 1. Molecular Probes--chemical synthesis--Laboratory Manuals. 2. Protein Array Analysis--methods--Laboratory Manuals. 3. Combinatorial Chemistry Techniques--Laboratory Manuals. 4. Genomics--methods--Laboratory Manuals. QU 25 C517 2005] I. Zanders, Edward D. II. Series.
QH431.C45197 2005
572.8'6--dc22

2005013118

Preface

Chemical genomics is an exciting new field that aims to transform biological chemistry into a high-throughput industrialized process, much in the same way that molecular biology has been transformed by genomics. The interaction of small organic molecules with biological systems (mostly proteins) underpins drug discovery in the pharmaceutical and biotechnology industries, and therefore a volume of laboratory protocols that covers the key aspects of chemical genomics would be of use to biologists and chemists in these organizations. Academic scientists have been exploring the functions of proteins using small molecules as probes for many years and therefore would also benefit from sharing ideas and laboratory procedures. Whatever the organizational backgrounds of the scientists involved, the challenges of extracting the maximum human benefit from genome sequencing projects remains considerable, and one where it is increasingly recognized that chemical genomics will play an important part.

Chemical Genomics: Reviews and Protocols is divided into two sections, the first being a series of reviews to describe what chemical genomics is about and to set the scene for the protocol chapters. The subject is introduced by Paul Caron, who explains the various "flavors" of chemical genomics. This is followed by Lutz Weber and Philip Dean who cover the interaction between organic molecules and protein targets from the different perspectives of laboratory experimentation and *in silico* design. The protocols begin with the methods developed in Christopher Lowes' laboratory (Roque et al.) for what could be described as a classical example of chemical genomics, namely the design of small molecules as affinity ligands for specific protein families. The theme is continued with detailed protocols for *in silico* docking by Jongejan et al. that highlights the importance of computational approaches to protein–small molecule interactions. The remaining protocols are directed towards the aim of producing highly diverse collections of proteins, carbohydrates, and small molecules for use in arrays containing large numbers of molecules. This high-throughput approach to screening for interaction between small and large biological molecules is the essence of chemical genomics. The chapters by Ryu, Doyle, Murphy, Sawasaki, Endo, Kohno, and Hoyt cover methods for the production of proteins and carbohydrates using different expression systems. Webster and Oxley give a protocol for analyzing the proteins using mass spectrometry. The techniques for arraying these proteins and carbohydrates on solid supports are detailed in the chapters by Blackburn, Marik, and Wang. Finally

an in vivo method for identifying small molecule–protein interactions is described by Khazak et al. using the yeast two-hybrid system.

Although we recognize that no single book on chemical genomics can be totally comprehensive in its coverage, we hope that the protocols here, in covering the key elements of the subject, will be of genuine use to the wide variety of scientists in this rapidly expanding field.

Edward D. Zanders

Contributors

JONATHAN M. BLACKBURN • *Department of Biotechnology, University of the Western Cape, Cape Town, South Africa; Procognia Ltd, Maidenhead, UK*
PAUL R. CARON • *Head of Informatics, Vertex Pharmaceuticals, Cambridge MA*
MIN CHEN • *The State Key of Microbial Technology, School of Life Science, Shandong University, Jinan, Shandong, People's Republic of China*
PHILIP M. DEAN • *Chief Scientific Officer, De Novo Pharmaceuticals Ltd, Cambridge, UK*
SHARON A. DOYLE • *Proteomics Group, DOE Joint Genome Institute, Walnut Creek, CA*
YAETA ENDO • *Cell-Free Science and Technology Research Center, Ehime University, Matsuyama, Japan*
IWAN J. P. DE ESCH• *Division of Medicinal Chemistry, Leiden/Amsterdam Center for Drug Research (LACDR), Faculty of Sciences, Vrije Universiteit, Amsterdam, The Netherlands*
ERICA A. GOLEMIS • *Division of Basic Sciences, Fox Chase Cancer Center, Philadelphia, PA*
MUDEPPA D. GOUDA • *Cell-Free Science and Technology Research Center, Ehime University, Matsuyama, Japan*
CHRIS DE GRAAF • *Division of Molecular Toxicology, Leiden/Amsterdam Center for Drug Research (LACDR), Faculty of Sciences, Vrije Universiteit, Amsterdam, The Netherlands*
GEETA GUPTA • *Institute of Biotechnology, University of Cambridge, Cambridge, UK*
DARREN J. HART • *High Throughput Group, Grenoble Outstation, European Molecular Biology Laboratory, Grenoble, France*
JON HOYT • *Department of Cell Biology, Institute of Chemistry and Cell Biology, Boston, MA*
ALDO JONGEJAN • *Division of Medicinal Chemistry, Leiden/Amsterdam Center for Drug Research (LACDR), Faculty of Sciences, Vrije Universiteit, Amsterdam, The Netherlands*
TAKAYASU KAWASAKI • *Cell-Free Science and Technology Research Center, Ehime University, Matsuyama, Japan*
VLADIMIR KHAZAK • *Director of Biology, NexusPharma, Langhorne, PA*
RANDALL W. KING • *Department of Cell Biology, Institute of Chemistry and Cell Biology, Boston, MA*

Toshiyuki Kohno • *Laboratory of Structural Biology, Mitsubishi Kagaku Institute of Life Sciences (MITILS), Tokyo, Japan*
Kit S. Lam • *Division of Hematology & Oncology, Department of Internal Medicine, UC Davis Cancer Center, University of California, Davis, CA*
Rob Leurs • *Division of Medicinal Chemistry, Leiden/Amsterdam Center for Drug Research (LACDR), Faculty of Sciences, Vrije Universiteit, Amsterdam, The Netherlands*
Hanfen Li • *Department of Biochemistry, The Ohio State University, Columbus, OH*
Steven Lin • *Department of Biochemistry, The Ohio State University, Columbus, OH*
Shaoyi Liu • *Columbia Genome Center, Columbia University College of Physicians & Surgeons, New York, NY*
Christopher R. Lowe • *Institute of Biotechnology, University of Cambridge, Cambridge, UK*
Jan Marik • *Division of Hematology & Oncology, Department of Internal Medicine, UC Davis Cancer Center, University of California, Davis, CA*
Michael B. Murphy • *Proteomics Group, DOE Joint Genome Institute, Walnut Creek, CA*
David Oxley • *Proteomics Research Group, Babraham Institute, Cambridge, UK*
Ana Cecília A. Roque • *Institute of Biotechnology, University of Cambridge, Cambridge, UK*
Kang Ryu • *Department of Biochemistry, The Ohio State University, Columbus, OH*
Tatsuya Sawasaki • *Cell-Free Science and Technology Research Center, Ehime University, Matsuyama, Japan*
Dhaval Shah • *Columbia Genome Center, Columbia University College of Physicians & Surgeons, New York, NY*
Jun Shao • *Department of Biochemistry, The Ohio State University, Columbus, OH*
Jing Song • *The State Key of Microbial Technology, School of Life Science, Shandong University, Jinan, Shandong, People's Republic of China*
Kazuyuki Takai • *Cell-Free Science and Technology Research Center, Ehime University, Matsuyama, Japan*
Yuzuru Tozawa • *Cell-Free Science and Technology Research Center, Ehime University, Matsuyama, Japan*
Takafumi Tsuboi • *Cell-Free Science and Technology Research Center, Ehime University, Matsuyama, Japan*

NICO P. E. VERMEULEN • *Division of Molecular Toxicology, Leiden/Amsterdam Center for Drug Research (LACDR), Faculty of Sciences, Vrije Universiteit, Amsterdam, The Netherlands*
DENONG WANG • *Carbohydrate Microarray Laboratory, Departments of Genetics, Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA*
PENG GEORGE WANG • *Department of Biochemistry, The Ohio State University, Columbus, OH*
RUOBING WANG • *Carbohydrate Microarray Laboratory, Departments of Genetics, Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA*
WEI WANG • *The State Key of Microbial Technology, School of Life Science, Shandong University, Jinan, Shandong, People's Republic of China*
LUTZ WEBER • *President, NexusPharma, Langhorne, PA*
JUDITH WEBSTER • *Proteomics Research Group, Babraham Institute, Cambridge, UK*
WEN YI • *Department of Biochemistry, The Ohio State University, Columbus, OH*
EDWARD D. ZANDERS • *CamBP Ltd, Cambridge, UK*

Contents

I

Reviews

1

Introduction to Chemical Genomics

Paul R. Caron

1. Introduction

Small-molecule drugs are a cost-effective way to treat and prevent disease. A study by the Slone Institute published in 2002 estimated that over 50% of the adult population in the United States used at least one pharmaceutical drug during the preceding week. The positive impact of small-molecule drugs on health care has been well documented ***(1,2)***.

The discovery of novel drugs has traditionally been a combination of clever science, brute force, and good fortune. With the advent of high-throughput screening technology, combinatorial chemistry, and the completion of the human genome sequence in the late 1990s, the hope was that technology could address the brute-force aspect and the genome sequence would provide insights into the underlying science, and good fortune would continue. Although there are some exceptions, productivity in the industry overall has gone down. Some of this is owing to higher regulatory standards and more difficult therapeutic areas, but a significant portion is the result of the lack of well-validated targets to apply the technology to. The industry portfolio of pharmaceutical targets of approx 500 in 1996 ***(3)*** has not been significantly expanded.

The availability of the human genome sequence and novel biological tools, such as siRNA, antisense, knockouts, and transgenics, suggests that over time, the physiological function of many of the genes in the genome may be deciphered. However, the time frame for this may be much greater than most people anticipated. For comparison, the first bacterial genome sequence was completed in 1995 ***(4)***, and although we may be able to now classify the majority of the genes by biochemical function, we don't know most of their physiological roles. One

From: *Methods in Molecular Biology, vol. 310: Chemical Genomics: Reviews and Protocols*
Edited by: E. D. Zanders © Humana Press Inc., Totowa, NJ

conceivable way to speed up the molecular dissection of the biology underlying various disease states is to use small-molecule compounds that specifically inhibit individual targets.

There are several key factors required to be successful when using small molecules to explore biology.

- The relative selectivity of the chemical probes that will be used must be known.
- The correlations between the cellular readout(s) used and the pathway or phenotype that is being assessed must be independently validated.
- All data must be fully integrated, allowing the user to navigate through biological pathways and supporting literature, assay results, and detailed information on compounds.

2. Different Flavors of Chemical Genomics

As with any emerging field, there are often differences of opinion on terminology among researchers, sometimes subtle, sometimes not. I will attempt to capture and describe the major variants and apologize if I inadvertently leave out any major themes, or end up misrepresenting some differences in trying to summarize the field.

2.1. Chemical Genetics

Chemical genetics, as described by Schreiber et al. in 1999 *(5)*, refers to the use of small molecules to induce alterations in gene products in mammalian systems, in a manner similar to using mutations. This approach became feasible through the combination of high-throughput cellular assays and diverse libraries of compounds. The ability to perform genetic screens in cellular assays vastly increases the throughput—traditionally a key limitation when studying higher organisms. It also allows the separation of effects in somatic cells from those in development.

It is critical to this chemical genetic approach to have a library of compounds that have a high probability of being relatively selective; otherwise, the ability to interpret the results becomes at least as complex as deciphering highly polygenetic phenotypes. To address this, diversity-oriented synthesis has been proposed to provide arrays of complex small molecules that are easily synthesized. The natural-product basis for many of the molecules and their complexity are believed to contribute to their cellular potency and selectivity *(6)*. This chemical genetic approach has been applied to identify novel inhibitors of alpha-tubulin and histone deactylation *(7)*.

As in classical genetics, a chemical genetic approach involves screening with probes that potentially could interact with any target in the genome, while trying to identify specific phenotypes. An alternate approach, termed *reverse chemical*

genetics, is analogous to introducing specific gene disruptions. Here, compounds that are known to specifically interact with a given target are used in broad phenotype screens to help identify the physiological role of that target.

2.2. Reverse Chemical Genetics

The key to a reverse chemical genetic approach is to have a one-to-one link between the small-molecule compound and the target of interest. This can be achieved by optimization of chemical reagents by thorough profiling against other potential targets, or alternatively by altering the target itself, to introduce changes that can be exploited for specificity. This approach has been most broadly applied to members of the protein kinase family, where specific changes can be made to the residues surrounding the active site that don't significantly alter the affinities or kinetics for natural substrates, but now allow the binding of specific inhibitor analogs *(8)*. Replacement of the wild-type copies of a given gene with these engineered mutants allows these compounds to be used to inhibit the function of the gene in cellular assays as well as in adult animals. The in vivo assay, complete with the complications of pharmacokinetics and pharmacodynamics, closely mimics the effect that would be expected from dosing an animal with a selective inhibitor against the wild-type target.

Additionally, introducing mutations into the adenosine triphosphate (ATP)-binding sites of targets allows the binding of labeled ATP analogs; these can then be used to trace biochemical pathways at the molecular level by looking directly at phosphorylated substrates, thus furthering the link between the target and the observed phenotype.

A comparison between standard genetics and chemical genetics is shown in **Fig. 1** (*see also* **Table 1**).

2.3. Screening

Screening of large sets of compounds, often assembled to be quite diverse, is often one of the first steps in a drug discovery project. The assay used for the initial screen will both help define the likelihood of getting potent hits, as well as form the foundation for the follow-up path. Assays that are more physiological require that active molecules pass through additional filters depending on the assay, such as transversing the cell wall, serum binding, bioavailability, and metabolic stability. These factors tend to decrease the hit rate, but result in molecules with better overall properties. The downside to this approach is that further compound optimization may be hindered by a lack of knowledge about the molecular target(s) of the initial hits. Biochemical-based target screening is likely to yield hits that have a clearer path to optimization, but are at higher overall risk because the link between the target and the desired physiological

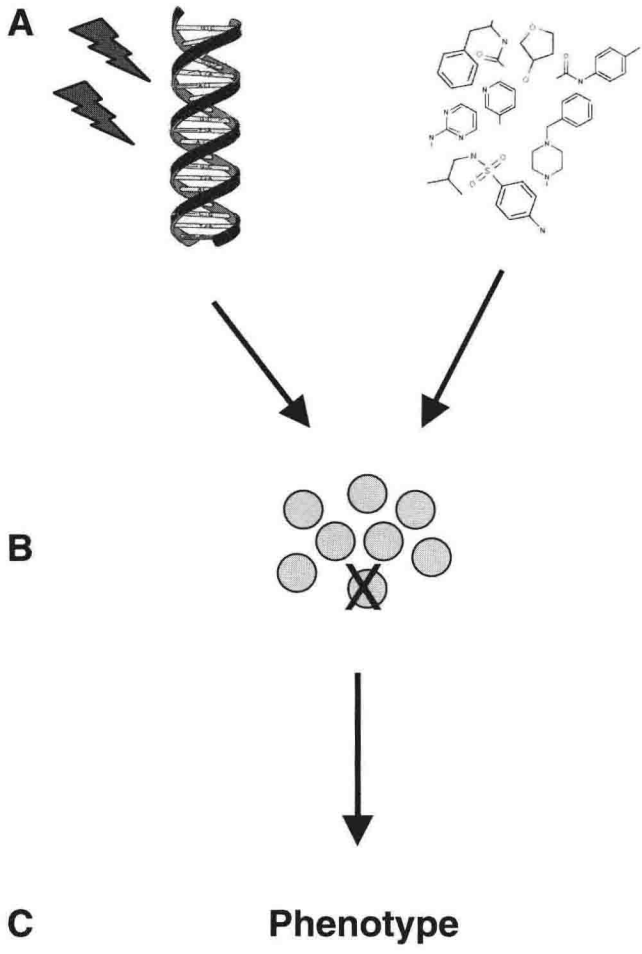


Fig. 1. Comparison of genetic and chemical genetics. In the traditional genetics approach genes are mutated resulting in missing or altered protein products. In a chemical genetic approach, compounds bind to specific proteins modulating their normal physiological functions. The throughput at steps A, B, C is used in Table 1 to help define the different chemical genetic approaches.

changes may not be strong and these compounds may be far from having the desired physical properties.

Smaller compound sets and focused compound libraries can be used to screen more broadly for physiological phenotypes. In the typical high-content screening experiment, the effects of compounds on cellular assays are captured

Table 1
Approaches to Chemical Genetics

		Approach	Goal	Step 1	Step 2	Step 3
Genetics	AcB	Randomly perturbate the system looking for changes in a specific phenotype	Target/pathway identification	Assemble large diverse chemical library	Assay for specific phenotype	Identify targets of compounds that induce the phenotype
Reverse Genetics	baC	Broad search for phenotype associated with perturbations of specific genes	New role for target	Select target	Find specific inhibitor	Broadly assay for phenotypes
Screening	bAc	Search for compounds which modulate a target with a known phenotype	Compounds/drugs for validated targets	Select target with presumed physiological role	Assemble large diverse chemical library	Screen for modulators of target
Genomics	ABC	Screen all targets for specific modulators, then try to identify phenotypes for each target	Target/pathway identification	Assemble large diverse chemical library	Screen for modulators of any targets	Broadly assay for phenotypes
Profiling	AC	Broadly screen for phenotypes associated with compounds, regardless of the target	Link between compounds and phenotype—efficacy or toxicity	Assemble large diverse chemical library	Broadly assay for phenotypes	
Chemo-genomics	BaC	Use information on targets to identify specific modulators, then look for phenotypes	Compounds/drugs for novel targets	Select targets	Identify inhibitor	Broadly assay for phenotypes

The steps in the second column refer to **Fig. 1A–C**. An uppercase letter indicates that the number of compounds, targets, or assays in the approach represents a large set. A lowercase letter is use to denote a small, focused set.