Methods in Molecular Biology 1263

Springer Protocols

Jonathan E. Hempel
Charles H. Williams
Charles C. Hong *Editors*

Chemical Biology

Methods and Protocols

Methods in Molecular Biology

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Chemical Biology

Methods and Protocols

Edited by

Jonathan E. Hempel

Department of Medicine - Cardiovascular Medicine, Vanderbilt Institute of Chemical Biology, Vanderbilt University School of Medicine, Nashville, TN, USA

Charles H. Williams

Department of Cell and Developmental Biology, Vanderbilt Institute of Chemical Biology, Vanderbilt University School of Medicine, Nashville, TN, USA

Charles C. Hong

Department of Medicine - Cardiovascular Medicine, Department of Pharmacology, Department of Cell and Developmental Biology, Vanderbilt Institute of Chemical Biology, Vanderbilt University School of Medicine, Nashville, TN, USA

Humana Press

Editors
Jonathan E. Hempel
Department of Medicine - Cardiovascular Medicine
Vanderbilt Institute of Chemical Biology
Vanderbilt University School of Medicine
Nashville, TN, USA

Charles H. Williams
Department of Cell and Developmental Biology
Vanderbilt Institute of Chemical Biology
Vanderbilt University School of Medicine
Nashville, TN, USA

Charles C. Hong
Department of Medicine - Cardiovascular Medicine
Department of Pharmacology
Department of Cell and Developmental Biology
Vanderbilt Institute of Chemical Biology
Vanderbilt University School of Medicine
Nashville, TN, USA

ISSN 1064-3745 ISSN 1940-6029 (electronic)
ISBN 978-1-4939-2268-0 ISBN 978-1-4939-2269-7 (eBook)
DOI 10.1007/978-1-4939-2269-7
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2014959096

© Springer Science+Business Media New York 2015
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Humana Press is a brand of Springer
Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The emergence of chemical biology represents the culmination of multifactorial forces in the fields of chemistry, biology, pharmacology, and medicine at the turn of the twenty-first century. Revolutions in genomic sequencing and robotic automation led to improved access to enabling technologies for academic researchers, and molecular biology methods for analyzing and manipulating intricate environmental cellular responses fostered new approaches to understanding biological systems. However, the sequencing of the human genome seemingly posed more questions than provided solutions to therapeutic hypotheses and strategies, and thus the genesis of chemical biology sought to employ the tools of chemistry to illuminate the complex underpinnings of cellular function.

To develop novel chemical tools, high-throughput screening (HTS) platforms and concepts traditionally utilized by the pharmaceutical and biotechnology industries for target-centric molecular discovery have been embraced for in vitro- and in vivo-based compound screening. In these chemical genetic systems, the fundamental ability of discrete chemicals to bind to and modulate the function of proteins leads to a phenotypic alteration. Most importantly, in contrast to biochemical HTS assays that screen for molecular binding to a protein with a therapeutic hypothesis, chemical genetic assays do not necessarily preselect for target identity, thus requiring the rate-limiting step of chemical biology: target identification.

With this compilation of methods in chemical biology, we seek to enable the discovery of novel chemical biology tools by providing readers with an array of techniques ranging from initial chemical genetic screening to target identification through the central theme of molecules. We have specifically organized the book into four parts to highlight essential components of the chemical biology tool discovery process. Part I details platforms for molecular discovery in in vitro cellular systems, and Part II provides in vivo chemical genetic screening protocols organized roughly in increasing order of organism complexity. These methods constitute a broad sampling of current state-of-the-art biological systems and phenotypic readouts for chemical genetic screening in chemical biology.

The unifying theme of chemicals in chemical biology necessitates the methods described in Part III, in which compounds are isolated, purified, selected, analyzed, and profiled to create their biological value. And the functional protein targets responsible for their phenotypes can be ascertained through methods described in Part IV for target identification. These four parts, taken together, describe processes for developing molecular tools to dissect biological function, and while no single prescription exists, our aim is to improve the success rate of this field through the dissemination of detailed, experiential knowledge.

We are extremely grateful for the tireless efforts of all of our authors throughout the writing and editing stages of this project; their willingness to share their expertise and experience single-handedly creates the value of this book. We also thank Dr. John M. Walker and David Casey for their help, guidance, and responsiveness and for the opportunity to compile this work.

Nashville, TN, USA

Jonathan E. Hempel
Charles H. Williams
Charles C. Hong

Contents

Preface v
Contributors xi

PART I IN VITRO CHEMICAL GENETIC SCREENING

1 Identification of Therapeutic Small-Molecule Leads in Cultured Cells Using Multiplexed Pathway Reporter Readouts 3
Ozlem Kulak, Kiyoshi Yamaguchi, and Lawrence Lum

2 Applying the Logic of Genetic Interaction to Discover Small Molecules That Functionally Interact with Human Disease Alleles 15
Ari D. Brettman, Pauline H. Tan, Khoa Tran, and Stanley Y. Shaw

3 Construction and Application of a Photo-Cross-Linked Chemical Array 29
Yasumitsu Kondoh, Kaori Honda, and Hiroyuki Osada

4 High Content Screening for Modulators of Cardiac Differentiation in Human Pluripotent Stem Cells 43
Sean Spiering, Herman Davidovics, Paul J. Bushway, Mark Mercola, and Erik Willems

5 Small-Molecule High-Throughput Screening Utilizing *Xenopus* Egg Extract 63
Matthew R. Broadus, P. Renee Yew, Stephen R. Hann, and Ethan Lee

PART II IN VIVO CHEMICAL GENETIC SCREENING

6 Fission Yeast-Based High-Throughput Screens for PKA Pathway Inhibitors and Activators 77
Ana Santos de Medeiros, Grace Kwak, Jordan Vanderhooft, Sam Rivera, Rachel Gottlieb, and Charles S. Hoffman

7 A Method for High-Throughput Analysis of Chronological Aging in *Schizosaccharomyces pombe* 93
Jessica Stephan and Ann E. Ehrenhofer-Murray

8 Protocols for the Routine Screening of Drug Sensitivity in the Human Parasite *Trichomonas vaginalis* 103
Manal J. Natto, Anthonius A. Eze, and Harry P. de Koning

9 Chemical Genetic Screens Using *Arabidopsis thaliana* Seedlings Grown on Solid Medium 111
Thanh Theresa Dinh and Xuemei Chen

10 Small-Molecule Screening Using *Drosophila* Models of Human Neurological Disorders 127
Mickael Poidevin, Feiran Zhang, and Peng Jin

11 High-Throughput Small-Molecule Screening in *Caenorhabditis elegans* 139
Sunitha Rangaraju, Gregory M. Solis, and Michael Petrascheck

12 Whole-Organism Screening for Modulators of Fasting Metabolism Using Transgenic Zebrafish 157
Philipp Gut and Didier Y.R. Stainier

13 High Content Screening for Modulators of Cardiovascular or Global Developmental Pathways in Zebrafish 167
Charles H. Williams and Charles C. Hong

PART III SMALL MOLECULES IN CHEMICAL BIOLOGY

14 Extraction Methods of Natural Products from Traditional Chinese Medicines 177
Jingyu Zhang, Jianying Han, Ayokunmi Oyeleye, Miaomiao Liu, Xueting Liu, and Lixin Zhang

15 Bioassay-Guided Identification of Bioactive Molecules from Traditional Chinese Medicines 187
Jianying Han, Jingyu Zhang, Wenni He, Pei Huang, Ayokunmi Oyeleye, Xueting Liu, and Lixin Zhang

16 NMR Screening in Fragment-Based Drug Design: A Practical Guide 197
Hai-Young Kim and Daniel F. Wyss

17 Practical Strategies for Small-Molecule Probe Development in Chemical Biology 209
Jonathan E. Hempel and Charles C. Hong

18 Principal Component Analysis as a Tool for Library Design: A Case Study Investigating Natural Products, Brand-Name Drugs, Natural Product-Like Libraries, and Drug-Like Libraries 225
Todd A. Wenderski, Christopher F. Stratton, Renato A. Bauer, Felix Kopp, and Derek S. Tan

19 Small-Molecule Library Screening by Docking with PyRx 243
Sargis Dallakyan and Arthur J. Olson

PART IV SMALL MOLECULE TARGET IDENTIFICATION

20 Fluorous Photoaffinity Labeling to Probe Protein-Small Molecule Interactions 253
Weigang Huang and Qisheng Zhang

21 Identification of the Targets of Biologically Active Small Molecules Using Quantitative Proteomics 263
Glòria Vendrell-Navarro, Andreas Brockmeyer, Herbert Waldmann, Petra Janning, and Slava Ziegler

22 Drug Affinity Responsive Target Stability (DARTS) for Small-Molecule Target Identification 287
Melody Y. Pai, Brett Lomenick, Heejun Hwang, Robert Schiestl, William McBride, Joseph A. Loo, and Jing Huang

23 Chemical Genomic Profiling via Barcode Sequencing to Predict Compound Mode of Action 299
Jeff S. Piotrowski, Scott W. Simpkins, Sheena C. Li, Raamesh Deshpande, Sean J. McIlwain, Irene M. Ong, Chad L. Myers, Charlie Boone, and Raymond J. Andersen

24 Image-Based Prediction of Drug Target in Yeast 319
Shinsuke Ohnuki, Hiroki Okada, and Yoshikazu Ohya

Index *329*

Contributors

RAYMOND J. ANDERSEN • *Department of Chemistry, University of British Columbia, Vancouver, BC, Canada*
RENATO A. BAUER • *Tri-Institutional PhD Program in Chemical Biology, Memorial Sloan Kettering Cancer Center, New York, NY, USA*
CHARLIE BOONE • *Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada*
ARI D. BRETTMAN • *Center for Systems Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; Cardiology Division, Massachusetts General Hospital, Boston, MA, USA*
MATTHEW R. BROADUS • *Department of Cell and Developmental Biology, Vanderbilt University Medical Center, Nashville, TN, USA*
ANDREAS BROCKMEYER • *Abt. Chemische Biologie, Max-Planck-Institut für molekulare Physiologie, Dortmund, Germany*
PAUL J. BUSHWAY • *Muscle Development and Regeneration Program, Sanford-Burnham Medical Research Institute, La Jolla, CA, USA*
XUEMEI CHEN • *Department of Botany and Plant Sciences, Institute of Integrative Genome Biology, University of California, Riverside, CA, USA; Howard Hughes Medical Institute, University of California, Riverside, CA, USA*
SARGIS DALLAKYAN • *Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA*
HERMAN DAVIDOVICS • *Muscle Development and Regeneration Program, Sanford-Burnham Medical Research Institute, La Jolla, CA, USA*
RAAMESH DESHPANDE • *Department of Computer Science and Engineering, University of Minnesota-Twin Cities, Minneapolis, MN, USA*
THANH THERESA DINH • *Department of Botany and Plant Sciences, Institute of Integrative Genome Biology, University of California, Riverside, CA, USA; Laboratory of Immunology and Vascular Biology, Department of Pathology, Stanford University of Medicine, Stanford, CA, USA*
ANN E. EHRENHOFER-MURRAY • *Institut für Biologie, Humboldt-Universität Berlin, Berlin, Germany*
ANTHONIUS A. EZE • *College of Medical, Veterinary and Life Sciences, Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, UK; Department of Medical Biochemistry, University of Nigeria, Enugu, Nigeria*
RACHEL GOTTLIEB • *Department of Biology, Boston College, Chestnut Hill, MA, USA*
PHILIPP GUT • *Gladstone Institutes and University of California, San Francisco, CA, USA*
JIANYING HAN • *Chinese Academy of Sciences Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China; University of Chinese Academy of Sciences, Beijing, China*
STEPHEN R. HANN • *Department of Cell and Developmental Biology, Vanderbilt University Medical Center, Nashville, TN, USA*
WENNI HE • *Chinese Academy of Sciences Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China*

JONATHAN E. HEMPEL • *Department of Medicine - Cardiovascular Medicine, Vanderbilt Institute of Chemical Biology, Vanderbilt University School of Medicine, Nashville, TN, USA*
CHARLES S. HOFFMAN • *Biology Department, Boston College, Chestnut Hill, MA, USA*
KAORI HONDA • *Chemical Biology Research Group, RIKEN Center for Sustainable Resource Science, Wako, Saitama, Japan*
CHARLES C. HONG • *Department of Medicine - Cardiovascular Medicine, Department of Pharmacology, Department of Cell and Developmental Biology, Vanderbilt Institute of Chemical Biology, Vanderbilt University School of Medicine, Nashville, TN, USA*
JING HUANG • *Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA, USA; Department of Molecular and Medical Pharmacology, University of California Los Angeles, Los Angeles, CA, USA*
PEI HUANG • *Chinese Academy of Sciences Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China; University of Chinese Academy of Sciences, Beijing, China*
WEIGANG HUANG • *Division of Chemical Biology and Medicinal Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*
HEEJUN HWANG • *Department of Molecular and Medical Pharmacology, University of California Los Angeles, Los Angeles, CA, USA*
PETRA JANNING • *Abt. Chemische Biologie, Max-Planck-Institut für molekulare Physiologie, Dortmund, Germany*
PENG JIN • *Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA*
HAI-YOUNG KIM • *Merck Research Laboratories, Kenilworth, NJ, USA*
YASUMITSU KONDOH • *Antibiotics Laboratory, Wako, Saitama, Japan; Chemical Biology Research Group, RIKEN Center for Sustainable Resource Science, Wako, Saitama, Japan*
HARRY P. DE KONING • *College of Medical, Veterinary and Life Sciences, Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, UK*
FELIX KOPP • *Molecular Pharmacology and Chemistry Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA*
OZLEM KULAK • *Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX, USA*
GRACE KWAK • *Department of Biology, Boston College, Chestnut Hill, MA, USA*
ETHAN LEE • *Department of Cell and Developmental Biology, Vanderbilt University Medical Center, Nashville, TN, USA; Vanderbilt Institute of Chemical Biology, Vanderbilt University Medical Center, Nashville, TN, USA; Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA*
SHEENA C. LI • *RIKEN Center for Sustainable Resource Science, Wako, Saitama, Japan*
MIAOMIAO LIU • *Chinese Academy of Sciences Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China; University of Chinese Academy of Sciences, Beijing, China; Eskitis Institute, Griffith University, Brisbane, QLD, Australia*
XUETING LIU • *Chinese Academy of Sciences Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China*
BRETT LOMENICK • *Department of Molecular and Medical Pharmacology, University of California Los Angeles, Los Angeles, CA, USA*
JOSEPH A. LOO • *Department of Biological Chemistry, University of California Los Angeles, Los Angeles, CA, USA*
LAWRENCE LUM • *Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX, USA*

William McBride • *Department of Radiation Oncology, University of California Los Angeles, Los Angeles, CA, USA*
Sean J. McIlwain • *Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI, USA*
Ana Santos de Medeiros • *Department of Biology, Boston College, Chestnut Hill, MA, USA*
Mark Mercola • *Muscle Development and Regeneration Program, Sanford-Burnham Medical Research Institute, La Jolla, CA, USA; Department of Bioengineering, University of California, San Diego, CA, USA*
Chad L. Myers • *Department of Computer Science and Engineering, University of Minnesota-Twin Cities, Minneapolis, MN, USA*
Manal J. Natto • *Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom*
Shinsuke Ohnuki • *Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba Prefecture, Japan*
Yoshikazu Ohya • *Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba Prefecture, Japan*
Hiroki Okada • *Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba Prefecture, Japan*
Arthur J. Olson • *Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA*
Irene M. Ong • *University of Wisconsin-Madison, Great Lakes Bioenergy Research Center, Madison, WI, USA*
Hiroyuki Osada • *Antibiotics Laboratory, Wako, Saitama, Japan; Chemical Biology Research Group, RIKEN Center for Sustainable Resource Science, Wako, Saitama, Japan*
Ayokunmi Oyeleye • *Forestry Research Institute of Nigeria, Jericho, Ibadan, Nigeria*
Melody Y. Pai • *Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA, USA*
Michael Petrascheck • *Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA, USA; Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA, USA; Department of Molecular and Cellular Neuroscience, The Scripps Research Institute, La Jolla, CA, USA*
Jeff S. Piotrowski • *University of Wisconsin-Madison, Great Lakes Bioenergy Research Center, Madison, WI, USA*
Mickael Poidevin • *Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA*
Sunitha Rangaraju • *Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA, USA; Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA, USA; Department of Molecular and Cellular Neuroscience, The Scripps Research Institute, La Jolla, CA, USA*
Sam Rivera • *Biology Department, Boston College, Chestnut Hill, MA, USA*
Robert Schiestl • *Department of Environmental Health Sciences, University of California Los Angeles, Los Angeles, CA, USA; Department of Radiation Oncology, University of California Los Angeles, Los Angeles, CA, USA; Department of Pathology, University of California Los Angeles, Los Angeles, CA, USA*
Stanley Y. Shaw • *Center for Systems Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; Cardiology Division, Massachusetts General Hospital, Boston, MA, USA*
Scott W. Simpkins • *Department of Computer Science and Engineering, University of Minnesota-Twin Cities, Minneapolis, MN, USA*

GREGORY M. SOLIS • *Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA, USA; Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA, USA; Department of Molecular and Cellular Neuroscience, The Scripps Research Institute, La Jolla, CA, USA*

SEAN SPIERING • *Muscle Development and Regeneration Program, Sanford-Burnham Medical Research Institute, La Jolla, CA, USA*

DIDIER Y.R. STAINIER • *Max-Planck Institute for Heart and Lung Research, Bad Nauheim, Germany*

JESSICA STEPHAN • *Institut für Biologie, Humboldt-Universität Berlin, Berlin, Germany*

CHRISTOPHER F. STRATTON • *Tri-Institutional PhD Program in Chemical Biology, Memorial Sloan Kettering Cancer Center, New York, NY, USA*

DEREK S. TAN • *Molecular Pharmacology and Chemistry Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA; Tri-Institutional Ph.D. Program in Chemical Biology, and Tri-Institutional Research Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA*

PAULINE H. TAN • *Center for Systems Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA*

KHOA TRAN • *Center for Systems Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA*

JORDAN VANDERHOOFT • *Department of Biology, Boston College, Chestnut Hill, MA, USA*

GLÒRIA VENDRELL-NAVARRO • *Abt. Chemische Biologie, Max-Planck-Institut für molekulare Physiologie, Dortmund, Germany*

HERBERT WALDMANN • *Abt. Chemische Biologie, Max-Planck-Institut für molekulare Physiologie, Dortmund, Germany; Lehrbereich Chem. Biologie, Fakultät Chemie, Technische Universität Dortmund, Dortmund, Germany*

TODD A. WENDERSKI • *Molecular Pharmacology and Chemistry Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA*

ERIK WILLEMS • *Muscle Development and Regeneration Program, Sanford-Burnham Medical Research Institute, La Jolla, CA, USA*

CHARLES H. WILLIAMS • *Department of Cell and Developmental Biology, Vanderbilt Institute of Chemical Biology, Vanderbilt University School of Medicine, Nashville, TN, USA*

DANIEL F. WYSS • *Merck Research Laboratories, Kenilworth, NJ, USA*

KIYOSHI YAMAGUCHI • *Division of Clinical Genome Research, Advanced Clinical Research Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan*

P. RENEE YEW • *Department of Molecular Medicine, Institute of Biotechnology, The University of Texas Health Science Center at San Antonio, San Antonio, TX, USA*

FEIRAN ZHANG • *Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA*

JINGYU ZHANG • *Chinese Academy of Sciences Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China*

LIXIN ZHANG • *Chinese Academy of Sciences Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China; Chinese Academy of Sciences, Institute of Microbiology, Beijing, China*

QISHENG ZHANG • *Division of Chemical Biology and Medicinal Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*

SLAVA ZIEGLER • *Abt. Chemische Biologie, Max-Planck-Institut für molekulare Physiologie, Dortmund, Germany*

Part I

In Vitro Chemical Genetic Screening

Chapter 1

Identification of Therapeutic Small-Molecule Leads in Cultured Cells Using Multiplexed Pathway Reporter Readouts

Ozlem Kulak, Kiyoshi Yamaguchi, and Lawrence Lum

Abstract

The rapid expansion of molecular screening libraries in size and complexity in the last decade has outpaced the discovery rate of cost-effective strategies to single out reagents with sought-after cellular activities. In addition to representing high-priority therapeutic targets, intensely studied cell signaling systems encapsulate robust reference points for mapping novel chemical activities given our deep understanding of the molecular mechanisms that support their activity. In this chapter, we describe strategies for using transcriptional reporters of several well-interrogated signal transduction pathways coupled with high-throughput biochemical assays to fingerprint novel compounds for drug target identification agendas.

Key words Small-molecule screening, RNAi, Luciferase assay, Wnt, TP53, Kras, Dot blotting

1 Introduction

Phenotypic screens in cultured cells incorporating large molecular libraries constitute a workhorse discovery platform that has been successfully used for gene discovery in diverse cellular processes. Unlike in vitro strategies that are typically designed for interrogating an isolated mechanism, in vivo approaches can measure a multitude of cellular phenomena that manifest as changes in a given endpoint readout. Thus, cellular reporters can be exploited to identify unanticipated mechanisms of action that can be targeted for therapeutic goals or unwanted activities associated with a given chemical reagent.

Current approaches aimed at capturing all cellular responses to a given genetic or chemical perturbation (a systems biology-based perspective) are not cost-effective solutions for screening large molecular libraries. For example, genome-scale expression profiling strategies reveal transcriptional changes in response to a given

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_1, © Springer Science+Business Media New York 2015

intervention that can then be used to infer the affected cell biological process. Whereas such methods can potentially better inform chemical selection processes at the initial screening step, they are slow to progress for screening purposes and beyond the economic reach for a minimally sized screening library. The additional requirement for computational infrastructure to establish functional relationships from such large datasets further imposes limitations to general accessibility.

Collapsing complex cellular phenomenon into the activity of a limited number of reporters enables cost-effective fingerprinting of large molecular libraries [1, 2]. These reporters can be selected for their sensitivity to a broad range of perturbations or for their specificity for a particular cellular process. The fingerprints of each chemical can then be matched to those of reference reagents targeting cellular components with assigned cellular roles to identify shared modes of action. In this manner, desirable and unwanted cellular targets can be defined early in the molecular library screening process, thus ultimately yielding a more robust collection of candidate genes or small molecules.

In this chapter, we build on a strategy previously used to interrogate the Wnt and Hedgehog signal transduction pathways with large chemical and siRNA libraries [3–5]. The approach incorporates luciferase-based reporters for several intensely studied signal transduction pathways that can be deployed in a single screening platform or sequentially to delineate chemical/gene activity (Fig. 1). We describe strategies for maximizing information recovery from this approach using novel reagents with selective activity against different luciferase enzymes (one secreted and two intracellular) as well as high-throughput biochemical analysis of cellular lysates expended for reporter-based activities by dot-blotting, a technique that enables Western blot analysis of high-density protein sample arrays (*see* **Note 1**).

2 Materials

2.1 Cell Culture and Reporter Constructs

1. HCT116 cells, colorectal cancer cell line (ATCC).
2. Dulbecco's modified Eagle medium (DMEM): Prepare full medium with 10 % fetal bovine serum (FBS), 100 units/mL penicillin, and 100 μg/mL streptomycin.
3. Phosphate-buffered saline (PBS).
4. 0.25 % Trypsin-EDTA.
5. 8XTCF-Renilla luciferase (RL) reporter construct is generated by inserting Tcf response elements and minimal promoter from STF (Addgene plasmid 12456) into the pGL4.71 vector (Promega).

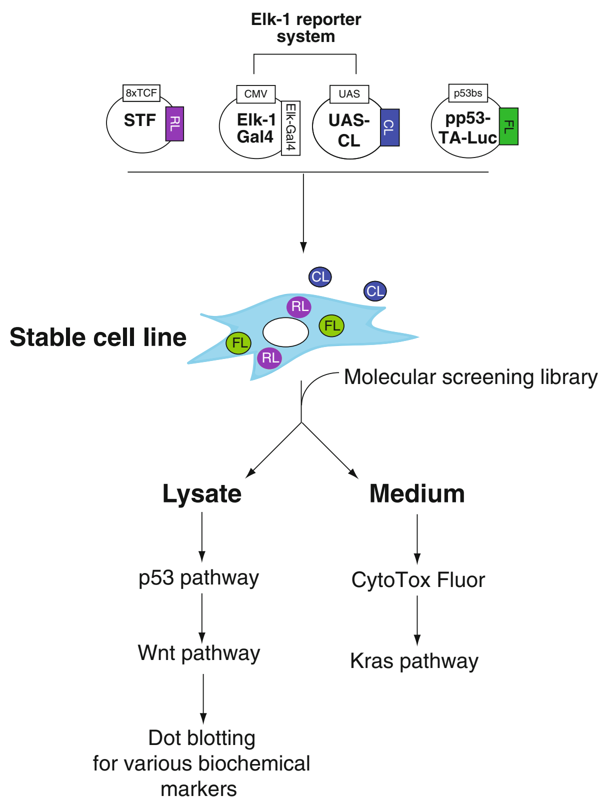


Fig. 1 A multiplexed luciferase reporter and biochemical platform for fingerprinting molecular libraries. A cell line transiently or stably harboring reporter plasmids for monitoring Wnt/β-catenin, p53, and Ras activity forms the basis for screening large molecular libraries to identify novel points of therapeutic intervention or detecting off-targeting effects of reagents. *RL Renilla* luciferase, *FL* firefly luciferase, *CL Cypridina* luciferase. SuperTopFlash reporter incorporates TCF/LEF-binding elements, thus reporting Wnt/β-catenin pathway activity [12]. The Elk-1 reporter system measures an output of Ras signaling and the pp53-TA-Luc plasmid reports TP53 activity. The Cytotox Fluor assay monitors the release of a cytoplasmic protease from cells with compromised membrane integrity. Molecular screening library reagents include small molecules and pooled siRNAs

6. pp53-TA-Luc plasmid (Agilent Technologies).
7. Elk1-Gal4 and UAS-CL vector (Elk-1 reporter system, Agilent Technologies).

2.2 Transient Transfection

1. Reporter DNA stock solution: Prepare 1 mL of a DNA reporter stock solution by combining 300 μL of p53-Firefly luciferase (FL), 300 μL of 8XTCF-Renilla luciferase (RL), 300 μL of Elk1-Gal4, 60 μL UAS-*Cypridina* luciferase (CL), and 40 μL of water to achieve a final 5:5:5:1 ratio of reporters. The final total DNA concentration of this stock solution is 0.96 mg/mL.
2. Fugene 6 Transfection Reagent (Promega).

2.3 Luminescence Detection

1. 96-Well white solid plates.
2. *Cypridina* luciferase (CL) assay reagents.
3. Dual-Glo Luciferase Reagent (Promega): Contains 5× passive lysis buffer, luciferase assay reagent II, and Stop & Glo Reagent.

2.4 Dot Blotting

1. Nitrocellulose membrane.
2. Gel blotting paper.
3. PBS-Tween (PBS-T): Add Tween-20 to PBS to a final concentration of 0.1 %.
4. Blocking buffer: Add nonfat milk to a final concentration of 5 % in PBS-T.
5. Horseradish peroxidase (HRP)/infrared dye-conjugated secondary antibodies.
6. P53 antibody (Santa Cruz) and β-actin antibody (Sigma).
7. Chemiluminescence detection kit (Thermo Scientific, ABC Scientific, Amresco, etc.).

3 Methods

The presented protocol simultaneously monitors the activity of three cancer-relevant cellular processes: the p53, Kras, and Wnt signal transduction pathways using a luciferase-based transient transfection protocol (*see* **Notes 2** and **3**).

3.1 Reporter Transient Transfection

The following protocol is written for a small-scale chemical screen that interrogates a library consisting of 1,000 chemical features (*see* **Note 4**). To improve signal uniformity across high-density cell culture plates, cells are transfected in culture dishes in bulk and replated into 96-well plates the following day.

3.1.1 Transfecting Cells in Bulk

Use 96-well conical bottom PCR plates to prepare multiple transfection mix pools. Using a 96-well plate will facilitate ease of pipetting and organization overall. In order to screen a 1,000-compound library, use 72 transfection mix pools (six transfection mix pools per plate) and transfect cells sufficient to plate twelve 10 cm^2 dishes (in other words 72 wells of the PCR plate) (*see* **Note 5**).

1. Dilute reporter DNA stock solution to 0.02 μg/μL in DMEM. For transfecting twelve 10 cm^2 dishes use 3.6 mL of diluted DNA reporter stock solution (*see* **Note 6**).
2. Dilute the transfection reagent Fugene 6–60 μL/mL in DMEM for generating 3,600 μL of diluted Fugene 6 solution.
3. Add 50 μL of the diluted DNA reporter stock solution per well into 72 wells of PCR plate. Add 50 μL of the diluted transfection reagent into each of the wells. Incubate for 10 min (*see* **Note 7**).

4. During the 10 min required to complete the formation of the DNA/transfection reagent mixes, prepare the cells for transfection (*see* Subheading 3.1.1, **step 4**) (*see* **Note 8**).
5. Wash nine 10 cm^2 plates of 80–90 % confluent HCT116 cells with 10 mL of PBS and then harvest cells using 1 mL of 0.25 % trypsin followed by incubation at 37 °C for 1 min for transfecting sufficient cells for plating 12×10 cm^2 plates. Neutralize the suspended cells with 10 mL of medium per plate.
6. Transfer 90 mL of trypsinized and neutralized cells into two 50 mL conical tubes and centrifuge at $130 \times g$ for 5 min, remove the supernatant, and resuspend the cell pellets in 90 mL per tube of full culture medium.
7. Add 7 mL of cells to each 10 cm^2 culture dish; prepare 12 dishes.
8. Add 600 μL of transfection mix dropwise to each of the twelve 10 cm^2 plates containing HCT116 cells. Repeat this step for the 12 plates.
9. Incubate the cells treated with transfection mix at 37 °C in an incubator with 5 % CO_2 for 24 h.

3.2 Replating Transiently Transfected Cells into High-Density Cell Culture Plates

1. Wash the twelve 10 cm^2 plates of transfected cells (now at 80–90 % confluency) with 10 mL of PBS and then harvest the cells with 1 mL of 0.25 % trypsin.
2. Transfer the suspended and neutralized cells into three 50 mL conical tubes and centrifuge at ~$130 \times g$ for 5 min, remove the supernatant, and resuspend the cell pellet in 120 mL medium.
3. Using a cell counter, generate a cell solution with 1.0×10^5 cells/mL to yield a 360 mL cell suspension. Plate 1.0×10^4 cells by adding 100 μL of cell solution to each well of a 96-well cell culture plate using a microplate liquid dispenser. The cell suspension should be sufficient in this case for plating 36×96-well plates.
4. Incubate the plates at 37 °C in an incubator with 5 % CO_2 for 4 h to allow the cells to re-adhere.
5. The compounds to be tested should be at a stock concentration of 250 μM in order to reach a final concentration of 2.5 μM. If not, dilute the chemicals to this concentration by using vehicle (DMSO in this case here).
6. Transfer 1 μL per well of the chemicals to each well of the 96-well plate containing the cells. As each chemical will be tested in triplicate, add the same chemicals to two additional 96-well plates containing cells.
7. Treat the cells with the compounds for 36 h in a standard CO_2 cell incubator at 37 °C. This incubation time suggested here is primarily based upon prior experiences with luciferase

reporter-based chemical screens in cultured cells that yielded successful outcomes. This parameter can be adjusted to accommodate the speed of cell doubling for a given cell line of interest (which will dilute out the transfected reporter) as well as the ability to robustly detect the effects of a positive chemical control.

3.3 Luciferase Assays

After 36 hrs luciferase activities are measured. The endpoint should of course be optimized for the specific cell line and readouts used (*see* **Notes 9** and **10**). In our study, a 36-h incubation time should yield a robust signal (*see* **Note 11**). This study incorporates multiple reporters, one secreted into the culture medium, and two others expressed in the cell cytoplasm, so obtain measurements from both the culture medium and cellular lysate (*see* **Note 12**).

3.3.1 Detection of CL Activity in the Culture Medium

1. Transfer 20 μL of culture medium from the assay plated in Subheading 3.2 to a white opaque 96-well plate using a liquid handler.
2. Add 20 μL of *Cypridina* luciferin assay buffer to each well.
3. Add 10 μL of *Cypridina* luciferin substrate to each well and detect CL activity immediately using a luminometer (*see* **Notes 13–15**). Note that the suggested manufacturer's protocol for the detection CL differs from a typical luciferase detection system (say for FL) in that it requires the addition of buffer to samples prior to the addition of the substrate.

3.3.2 Detection of FL and RL Activity

1. Remove the culture medium and then lyse the cells. Add 30 μL/well of 1× passive lysis buffer to each well of cells using a microplate liquid dispenser and place on a platform rocker set at a medium rocking speed for 5 min at room temperature.
2. Add 10 μL of the luciferase assay reagent II per well and immediately measure FL activity using the luminometer (*see* **Note 16**).
3. After measuring FL activity, this signal will be simultaneously blocked and RL substrate added by the addition of 10 μL of Stop & Glo Reagent per well. After measuring RL activity retain the lysate for additional biochemical assays.

3.4 Biochemical Assays

To increase content recovery, a biochemical component that complements the luciferase reporter-based readouts is incorporated in the overall drug discovery platform. The following protocol describes how the lysate from a single 96-well plate is transferred to nitrocellulose for Western blotting using a 96-well dot-blotting device (a dot blotter).

1. Pre-wet a 90 × 130 mm nitrocellulose membrane in PBS for 30 s. Place the membrane squarely onto sheets of blotting paper cut to the same size pre-wetted with PBS. Clamp the nitrocellulose and blotting paper assembly between the upper

and lower modules of the dot blotting apparatus (BioRad and Millipore both sell popular models). By tightening the screws on all four corners of the apparatus, individual watertight chambers for accommodating 96 different samples are generated. Attach the apparatus to a vacuum source using appropriate tubing.

2. Pre-rinse the membrane by adding 100 μL of PBS to each well using a multichannel pipette and applying vacuum at 5 PSI.
3. Prepare the lysate subjected to luciferase activity measurements for binding to nitrocellulose by adding 60 μL of 1× passive lysis buffer to each well of lysate using the microplate liquid dispenser and then transferring 30 μL of lysate onto the nitrocellulose membrane using the 96-channel liquid handler. A tip touch to the side of the well will improve volume-dispensing consistency across all wells (programmable with some 96-well liquid handlers).
4. Allow the lysate to slowly filter through the membrane by gravity for 10 min and then briefly apply vacuum to the manifold (5 PSI) to complete the transfer.
5. Wash the nitrocellulose membrane with 200 μL of PBS again by applying vacuum filtration.
6. Remove the membrane from the apparatus and incubate in blocking buffer for 30 min at room temperature to prevent nonspecific antibody binding.
7. Wash with PBS-T for 10 min three times.
8. Incubate the nitrocellulose membrane with mouse p53 antibody diluted to 1:1,000 in PBS-T for 60 min at room temperature.
9. Wash with PBS-T for 10 min three times.
10. Incubate with anti-mouse secondary antibody conjugated with horseradish peroxidase (HRP) in PBS-T.
11. Incubate the nitrocellulose membrane with rabbit β-actin antibody diluted to 1:10,000 in PBS-T for 60 min at room temperature.
12. Wash with PBS-T for 10 min three times.
13. Incubate with anti-rabbit secondary antibody conjugated with an infrared fluorescent dye (IRDye 800CW, for example) diluted in PBS-T for 60 min at RT in the dark.
14. Wash with PBS-T for 10 min three times.
15. Incubate the membrane with chemiluminescence detection reagents for 1 min to detect HRP-generated signal. Acquire both HRP- and infrared dye-generated signals using a Li-COR Odyssey Fc instrument (Fig. 2) (*see* **Note 1**). A predefined optimal wavelength of excitation will be used to detect the respective infrared dye conjugated to the secondary antibody. In the case of the IRDye 800CW secondary, an excitation wavelength of 800 nM is used.

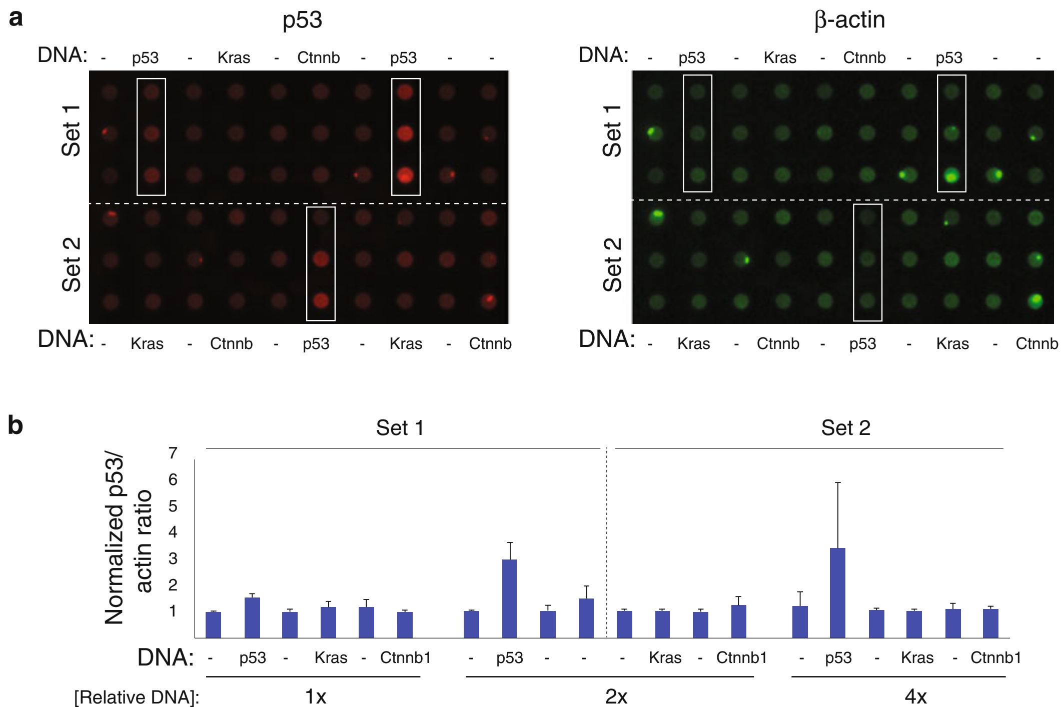


Fig. 2 Increasing content recovery by coupling luciferase-based assays with high-throughput biochemical readouts. The same cell line subjected to the luciferase-based assay protocol (HCT116 cells) is evaluated here for its reliability in reporting a biochemical readout (p53 expression) by dot blot analysis. Cells transfected with indicated expression construct and pathway reporters in a 96-well culture plates were lysed 48 h post-transfection. Following luciferase activity measurements (data not shown), protein from spent lysates was immobilized on nitrocellulose using a liquid handler and filtration manifold. (**a**) p53 and β-actin protein levels detected using protein-specific antibodies, infrared fluorescent dye-coupled secondary antibodies (with emissions at 680 and 800 nM), and the Li-COR imaging system. Columns of lysate corresponding to cells transfected with p53 DNA are *boxed*. (**b**) Quantification of the p53 to β-actin protein ratio

3.5 Interpreting Screening Results and Improving the Next Screen

Statistical means to identify outlier phenomenon rank-order screening results are based on reproducibility but not their potential biological relevance. Nevertheless, identifying robust effects of protein perturbation by genetic or chemical means is typically the first step towards achieving end-point screening goals. Frequently, a standard deviation threshold of the normalized data relative to the mean is used to identify outliers. The selection of the threshold and the number of hits to be considered for further analysis can be guided by the ability of the algorithm to return (a) known components (controls), (b) useful effects based on secondary counter-screen results/experience, and (c) an economically feasible counter-screen strategy. Below, we briefly discuss issues/strategies that could improve the overall primary dataset such that the challenges just discussed can be minimized [6].

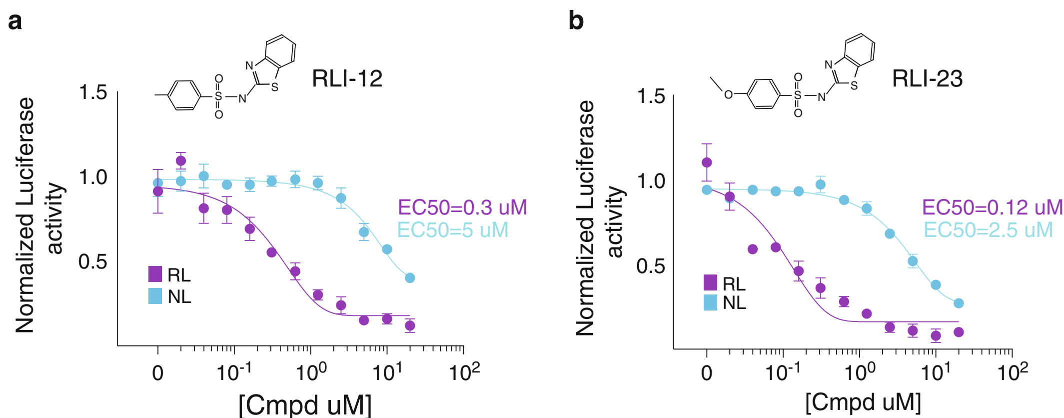


Fig. 3 Chemically mediated minimization of luciferase enzyme cross talk in multiplexed screening platforms. The RL inhibitors 12 and 23 (RLI12 and RLI23) exhibit >10-fold selectivity for RL over NL. These compounds could be used to eliminate inadvertent release of cytoplasmic RL into the medium where the typically secreted NL activity is found. The RLI compounds were identified as false positives from a high-throughput chemical screen for novel Wnt pathway inhibitors [3, 8]

A consistent problem with high-throughput screens that rely on high-density multiwell plates is known as an "edge effect." This effect is associated with a frequency of outlier results from cells evaluated in wells found on the edge of the plate that is higher than that observed from the remaining wells. To limit the contribution of this phenomenon to the overall selection of hits from a screen, a number of computational approaches can be taken (*see* [7] for example).

Improvements in the selection of potentially meaningful reagents from a molecular library could also be found in the increase of content return from the initial screening effort. This can be accomplished by utilizing new technologies that afford even greater multiplexing capability than what is described above. For example, luciferase enzymes engineered to emit a signal at a given wavelength could be used to detect specific pathway signals simply by altering the spectral detection parameters of a luminometer. At the same time, the discovery of enzymes that generate luminescence using novel substrates could be leveraged for multiplexing exercises when combined with specific luciferase inhibitors such as those that target RL [8] (Fig. 3).

4 Notes

1. Combining multiplexed luciferase assays with a dot blotting strategy can greatly expand the number of data points used to generate signatures for each chemical feature found in a screening library. From a single well three dot blots can be produced using the protocol provided above. A Li-COR Odyssey Fc imaging system also provides an additional opportunity to test

three antibodies simultaneously or sequentially from the same blot using different infrared- and HRP-coupled secondary antibodies. For this protocol we chose p53 antibody that detects a protein with a wide dynamic range, and β-actin antibody that serves as a loading control.

2. Stable cell lines harboring p53, Kras, and Wnt pathway reporters can be employed in lieu of transient transfection strategies. Creating a stable cell line that harbors all four DNA plasmids can be challenging. Multiple cell lines each harboring one or two reporters provide an alternate strategy that may increase the success rate of generating such reporter cell lines. For example, one cell line can harbor the p53 and Wnt reporters while another can be used to monitor the Kras pathway. Regardless of the approach, these cell lines likely would provide a more robust screening platform by mitigating transient transfection-associated stochasticism.
3. A variety of luciferase-based reporters for monitoring diverse cell biological processes are commercially available. For example, luciferase-based reporters of other pathways such as the BMP and Notch pathways can be used for similar multiplexed luciferase screening projects.
4. We describe here strategies for multiplexing luciferase reporters of various cellular pathways using a small chemical library but it can be easily adapted for large cDNA or siRNA libraries [1].
5. We suggest achieving a large volume of transfection mix by combining a series of smaller transfection mix reactions using an optimized smaller scale protocol provided by the manufacturer to maintain consistency in transfection efficiency. The use of 96-well conical bottom PCR plates greatly facilitates the preparation of multiple transfection mixes given the accessibility of the samples to multichannel pipettes.
6. Although more labor intensive and tedious, a multichannel pipettor can also be used to dispense cells into 96-well plates and/or transferring lysate to the dot blotting apparatus. Although not readily accessible to all, a liquid handler with a 96-channel liquid dispensing head would facilitate most liquid transfer procedures. For larger scale screens (beyond 1,000 samples for example), this instrument becomes a necessity.
7. For this protocol we employed Fugene 6 as the transfection reagent, so dilutions of DNA stocks and transfection reagent should be done in cell culture medium (DMEM) lacking FBS and penicillin/streptomycin as recommended by supplier.
8. If unable to complete the cell preparation within 10 min of incubation time of DNA/transfection reagent, then neutralize the lipid complex formation reaction by adding full medium (DMEM/10 % FBS/1 % penicillin/streptomycin). Now the trypsinization procedure can be safely completed.

9. The selection of a small-molecule library relates to its intended purpose: a library with a wider chemical space increases the probability to find a drug-like small molecule although small libraries like natural product libraries can be employed to uncover new biology [9, 10].
10. The selection of the luciferase enzyme readout can greatly influence the labor, cost, and instrument sensitivity requirements for a given screen. Whereas firefly luciferase is the most commonly used enzyme in commercially available constructs, the advent of new luciferase reporter systems that do not require ATP (as in the case of RL-based reactions) or that incorporate enzymes with greater activity promises to improve the reproducibility and cost-effectiveness of luciferase-based research platform, respectively.
11. An important note regarding the selection of luciferase reporters in chemical screens is the susceptibility of FL to chemical inhibition that could give rise to false positives [11]. In our experience, enzymes such as *Renilla*, *Gaussia*, *or* Nanoluc luciferase (RL, GL, and NL, respectively) that utilize coelenterazine (a larger substrate than luciferin) are less susceptible to chemical inhibition. Thus, given the opportunity to select or design a reporter construct with a chemical screen in mind, avoidance of FL would be advised.
12. For high-throughput screens, *Z*-prime statistical analysis can provide an indication of overall assay robustness based on the average and standard deviation of a positive and negative control. *Z*-prime values above 0.5 indicate a robust and acceptable assay.
13. The most frequently used multiplexing luciferase enzymes are FL and RL given the availability of plug-and-play kits for measuring their activities in sequence. A standard protocol typically entails adding luciferin to cell lysates to reveal levels of FL activity followed by a quenching reagent that is simultaneously deposited with coelenterazine to yield a secondary RL signal. With the addition of other luciferases that can be secreted such as GL or that use yet another substrate such as *Cypridina* luciferase (CL), the number of approaches available for generating high-content data using luciferase enzymes continues to grow.
14. Other reporter systems can be incorporated into this protocol to increase the content of this experiment. For example, the CytoTox Fluor assay can be used to monitor cellular toxicity by sampling the levels of an intracellular protease released into the culture medium. The addition of the CytoTox assay reagent does not influence the activity of the CL reporter.

15. Commercially available CL assay kits consist of two parts, assay buffer and reconstituted substrate. Since for a screen a large number of plates are used, after addition of assay buffer to cell media, media and assay buffer-containing plates can be stored at 4 °C while the rest of the plates are processed.
16. In a luciferase assay a flash of light is generated that decays rapidly after the enzyme and substrates are combined. For this reason, the use of "flash kits" requires rapid measurement within 5 min upon substrate addition and since a screen requires processing large numbers of plates, a luminometer coupled with an automated plate stacker is also required unless luciferase kits optimized for extended signal duration are used.

Acknowledgment

This work was supported by the NIH (R01 CA168761), CPRIT (RP130212), and the Welch Foundation (I-1665). This work was also supported by the fund from Japan Society for the Promotion of Science (JSPS) for the "Institutional Program for Young Researcher Overseas Visits."

References

1. Kulak O, Lum L (2013) A multiplexed luciferase-based screening platform for interrogating cancer-associated signal transduction in cultured cells. J Vis Exp 77:e50369
2. Potts MB et al (2013) Using functional signature ontology (FUSION) to identify mechanisms of action for natural products. Sci Signal 6:ra90
3. Chen B et al (2009) Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. Nat Chem Biol 5:100–107
4. Jacob LS et al (2011) Genome-wide RNAi screen reveals disease-associated genes that are common to Hedgehog and Wnt signaling. Sci Signal 4:ra4
5. Tang W et al (2008) A genome-wide RNAi screen for Wnt/beta-catenin pathway components identifies unexpected roles for TCF transcription factors in cancer. Proc Natl Acad Sci U S A 105:9697–9702
6. Birmingham A et al (2009) Statistical methods for analysis of high-throughput RNA interference screens. Nat Methods 6:569–575
7. Zhong R et al (2013) SbacHTS: spatial background noise correction for high-throughput RNAi screening. Bioinformatics 29:2218–2220
8. Lum LD, Kulak O (2012) Multiplexed luciferase reporter assay systems. US Patent #12/908,754
9. Hasson SA, Inglese J (2013) Innovation in academic chemical screening: filling the gaps in chemical biology. Curr Opin Chem Biol 17:329–338
10. Frearson JA, Collie IT (2009) HTS and hit finding in academia: from chemical genomics to drug discovery. Drug Discov Today 14:1150–1158
11. Sheridan C (2013) Doubts raised over 'read-through' Duchenne drug mechanism. Nat Biotechnol 31:771–773
12. Veeman MT et al (2003) Zebrafish prickle, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. Curr Biol 13:680–685

Chapter 2

Applying the Logic of Genetic Interaction to Discover Small Molecules That Functionally Interact with Human Disease Alleles

Ari D. Brettman, Pauline H. Tan, Khoa Tran, and Stanley Y. Shaw

Abstract

Despite rapid advances in the genetics of complex human diseases, understanding the significance of human disease alleles remains a critical roadblock to clinical translation. Here, we present a chemical biology approach that uses perturbation with small molecules of known mechanism to reveal mechanistic and therapeutic consequences of human disease alleles. To maximize human applicability, we perform chemical screening on multiple cell lines isolated from individual patients, allowing the effects of disease alleles to be studied in their native genetic context. Chemical screen analysis combines the logic of traditional genetic interaction screens with analytic methods from high-dimensionality gene expression analyses. We rank compounds according to their ability to discriminate between cell lines that are mutant versus wild type at a disease gene (i.e., the compounds induce phenotypes that differ the most across the two classes). A technique called compound set enrichment analysis (CSEA), modeled after a widely used method to identify pathways from gene expression data, identifies sets of functionally or structurally related compounds that are statistically enriched among the most discriminating compounds. This chemical:genetic interaction approach was applied to patient-derived cells in a monogenic form of diabetes and identified several classes of compounds (including FDA-approved drugs) that show functional interactions with the causative disease gene, and also modulate insulin secretion, a critical disease phenotype. In summary, perturbation of patient-derived cells with small molecules of known mechanism, together with compound-set-based pathway analysis, can identify small molecules and pathways that functionally interact with disease alleles, and that can modulate disease networks for therapeutic effect.

Key words Chemical screen, Functional genomics, Chemical genetics, Chemical genomics, Drug repurposing, Genetic interaction

1 Introduction

The growing catalogue of genetic variants that influence human disease risk has highlighted two related challenges: how to understand the biological function of a risk allele, and how to translate genetic and functional insights into new therapies. Targeted mutagenesis, RNA interference, and genome editing represent genetic approaches for dissecting genotype-phenotype correlations.

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_2, © Springer Science+Business Media New York 2015

Here, we present a complementary chemical biology strategy that applies the logic of synthetic genetic interaction.

Our approach is based upon chemical perturbation using small molecules with defined mechanisms to elucidate the functional and therapeutic implications of human disease alleles. If a small molecule causes a distinct phenotype in the presence versus absence of a disease allele, then the small molecule, or its protein target(s), may be inferred to have a functional connection with the disease gene and the phenotype. This rationale is borrowed from the notion of genetic interaction in classical genetics, in which mutations in two different genes interact to produce a phenotype that is unexpected based on the phenotype of each mutation in isolation. A common implementation of genetic interaction is synthetic lethal screening in model organisms, in which two mutations (each nonlethal in isolation) lead to lethality when coincident in the same organism (Fig. 1). Genetic interaction in model organisms has been a powerful tool to discover functional relationships among genes or their gene products [1, 2].

We have adapted this logic to identify small molecules, genes, or pathways that functionally interact with disease alleles. By analogy to genetic interaction screens, the first "hit" is a mutation in a gene of interest, such as a gene that influences disease susceptibility. The second "hit" is a small molecule with characterized mechanism(s), such as an FDA-approved drug or a tool compound (e.g., a kinase inhibitor). A synthetic phenotype can then be observed when a small molecule causes a qualitatively or quantitatively distinct phenotype in the presence of a wild-type

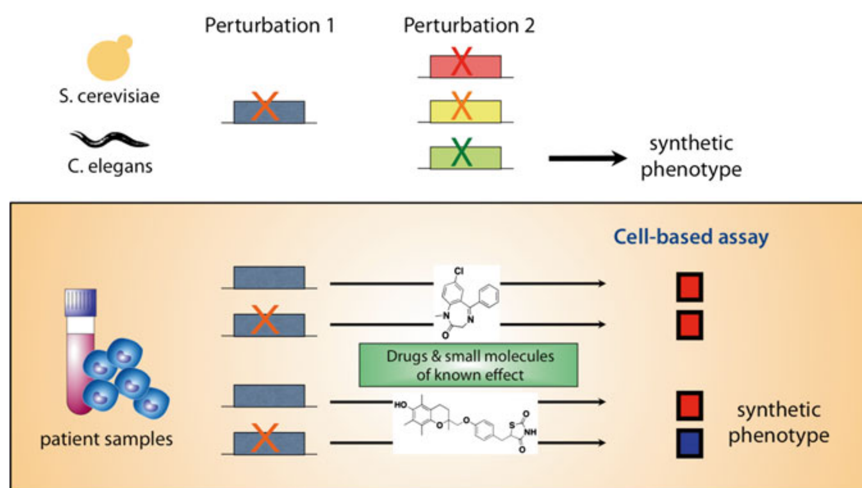


Fig. 1 Analogy between genetic interaction screens in model organisms and chemical:genetic interaction screen in patient-derived cells. In the latter, the second "hit" is provided by a small molecule with characterized mechanism; the synthetic phenotype is manifest as an assay phenotype that is distinct in mutant versus wild-type cells

versus mutant allele at the disease gene (Fig. 1). While operationally these experiments constitute a chemical screen, our annotated small molecules function as quasi-genetic perturbations. By screening FDA-approved drugs as part of a known bioactives collection, new therapeutic hypotheses can be developed that "repurpose" existing drugs and can be rapidly tested in proof-of-concept studies in humans.

To enable a synthetic interaction approach in patient-focused studies, we first developed a framework to analyze the results of chemical perturbation of multiple patient-derived cell isolates and rank compounds according to the degree that their induced phenotypes differ between cells mutant versus wild type at a disease gene. The resulting data set structurally resembles a gene expression dataset, in which multiple cell lines that belong to distinct classes (e.g., mutant versus wild type at a disease gene) are represented in columns, and each row corresponds to a different small molecule (rather than genes, in the case of gene expression) (Fig. 2). The values depicted within each cell of this data matrix (and thus the features that characterize each cell line) are assay phenotypes induced by a small molecule (rather than expression values of individual genes).

Taking advantage of this analogy between chemical screening and gene expression datasets, we developed methods to analyze chemical screening data using statistical methods used to examine gene expression data. For instance, identifying compounds that cause different assay phenotypes between mutant versus wild-type cell lines is analogous to identifying genes whose expression most differs between two classes of cell lines.

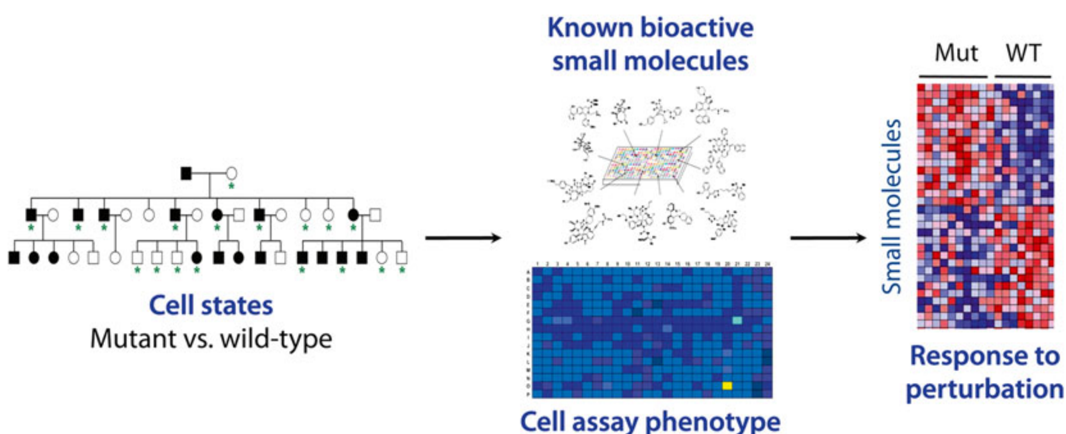


Fig. 2 Overview of chemical genetic interaction screen in patient-derived cells. The resulting dataset resembles that of gene expression. In the heatmap, individual cell lines (belonging to mutant and wild-type classes) are depicted in *columns*, and *rows* represent small molecules. The content of each cell is the quantitative assay phenotype for each small molecule in each cell line

Second, we developed an approach that identifies sets of related compounds that are statistically enriched among the most discriminating compounds (i.e., those compounds that induce phenotypes that differ most between mutant versus wild-type cell lines). We call this method compound set enrichment analysis (CSEA) [3], after the gene set enrichment method (GSEA) on which it is based (http://www.broadinstitute.org/gsea/index.jsp) [4, 5]. GSEA is widely applied to gene expression data to identify sets of functionally related genes that are coordinately up- or downregulated across a class distinction (even if changes for individual genes are statistically modest). Screened compounds are first ranked according to quantitative difference between mutant versus wild-type cells; CSEA then tests if a prespecified set of compounds *S* are randomly distributed throughout the ranked list, or are enriched at the top or bottom (as would be expected if members of set *S* can discriminate between mutant and wild-type classes) (Fig. 3). CSEA calculates a Kolmogorov-Smirnov-like statistic by walking down the ranked list, and increasing a running-sum statistic whenever a member of set *S* is encountered, and decreasing the running-sum statistic whenever a compound that is not in set *S* is encountered. The enrichment score (ES) is defined as the greatest deviation

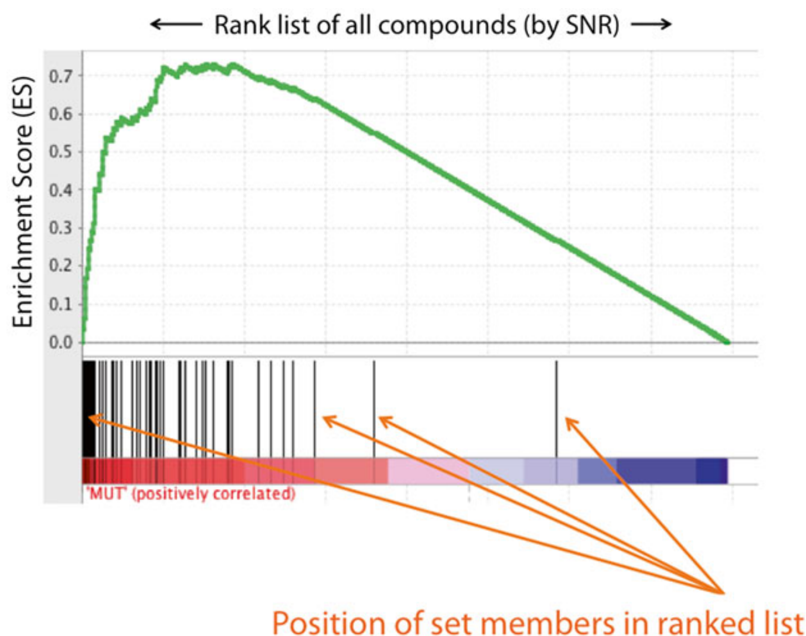


Fig. 3 Sample graphical output of CSEA. The algorithm steps through the ranked list of compounds (ranked according to SNR); at each position, the enrichment score increases if a member of the compound set is encountered, and decreases if a set member is not encountered. *Bottom panel*: The *red:blue horizontal bar* represents the ranked compound list (ranked by SNR); each *vertical line* represents the position of a member of the compound set within the ranked list. In this example, members of the compound set are highly enriched among compounds with the highest SNR; the enrichment score is 0.73

from zero (either positive or negative) achieved by the running-sum statistic (Fig. 3), and the normalized enrichment score (NES) adjusts the enrichment score for the number of compounds in a set. To help evaluate statistical significance, CSEA calculates a permutation *p*-value for the enrichment of each compound set by randomly permuting class assignments (i.e., which cell lines are mutant or wild type, preserving the number of cell lines in each class) 1,000 times, calculating the enrichment score for each permutation, and generating a null distribution from these permutations [5]. While we apply CSEA here to a screen across multiple cell lines belonging to two classes, CSEA may also be applied to traditional chemical screens in a single cell line. In this case, the screen results are inputted to CSEA as a ranked list based on assay *Z*-scores [6, 7]; to calculate statistical significance, CSEA randomly generates 1,000 compound sets with the same number of compounds as the query set, and generates a null distribution from the enrichment scores for these permutated compound sets. Compound sets can be defined by membership in the same metabolic pathway, or the same drug class, or any other shared property. Rather than choosing compound "hits" individually, CSEA identifies promising groups of functionally related compounds, increasing confidence in hit selection, and providing structural and/or functional insights into screen results. CSEA also allows statistical significance to be ascertained for compound sets, even if individual compound effects are statistically modest.

This overall screening and analytic approach was applied to patient-derived cells from a family pedigree whose members were diagnosed with maturity-onset diabetes of the young type 1 (MODY1), a form of monogenic type 2 diabetes due to highly penetrant loss-of-function mutations in the orphan nuclear hormone receptor HNF4α [8–10]. Despite the monogenic cause of MODY1, how mutations in HNF4α lead to impaired insulin secretion and diabetes remains poorly understood. Selecting a surrogate cell line for screening involves balancing physiologic fidelity, and the accessibility and availability of cells. We opted to screen Epstein-Barr virus-transformed lymphoblasts, primarily because the ubiquity of these cell lines in association with clinical cohorts (generally created as a renewable source of DNA) makes them an attractive cell resource for high-throughput screens [11]. As a cellular phenotype, we selected a commercially available assay for cellular ATP content for two main reasons: (a) ATP is a key intracellular sensor in the pancreatic β-cells that can help initiate a series of ionic fluxes that ultimately lead to insulin release; (b) lymphoblasts grow in clumps and are only partially adherent, and the ATP assay used does not involve any wash steps that could result in cell loss. Note that for our synthetic genetic interaction screen, interesting interactions between the disease gene (HNF4α) and a small molecule or its target are revealed regardless of the direction of the small

molecule's effect on assay phenotype; that is, we do not require a specific direction of effect.

We screened lymphoblasts from 18 members of a MODY1 family (10 with the diabetes-causing HNF4α mutation and 8 without). CSEA identified several classes of small molecules that interact with HNF4α (including a series of fatty acids that likely physically bind HNF4α). Several small molecules showed a synthetic interaction with HNF4α genotype in both human lymphoblasts and a murine pancreatic β-cell model, indicating that a subset of interactions between HNF4α and the small molecule (or its protein target) are conserved across lymphoblasts and β-cells [3]. Analysis of the pathways modulated by discriminating compound sets supported a functional connection between the causative disease gene, HNF4α, and "metabolism-excitation coupling" (a pathway crucial for insulin secretion in pancreatic β-cells) [12]. As validation, some of the drugs identified in our screen also modulated insulin secretion from β-cells, a critical disease phenotype in MODY1 [3]. None of the drugs identified in our study have been studied in association with MODY1, and none were approved for diabetes-related indications. These data together demonstrate how perturbation with small molecules of known mechanism, together with compound-set-based pathway analysis, can identify pathways that functionally interact with disease alleles, and that can modulate disease networks for therapeutic effect. More broadly, this approach identifies small molecules that induce phenotypes that are dependent on the presence of disease alleles, and thus reveals the functional consequences of disease alleles in the native genetic context of cells from individual patients.

Recently, the Cancer Therapeutics Response Portal has catalogued both allele- and lineage-specific effects of 354 small molecules on cell viability in 242 genomically defined cancer lines [13]. As datasets incorporating genomic and chemical screening data become more widely available, chemical:gene synthetic interaction analysis will yield mechanistic and therapeutic insights for a variety of diseases and susceptibility genes.

In this chapter, we describe a protocol for the systematic perturbation of patient-derived cell lines using small-molecule probes, which is both scalable to high-throughput workflow and generalizable to a variety of assays. In the protocol presented here, multiple patient-derived lymphoblastoid cell lines either mutant or wild type at a defined genetic locus (e.g., HNF4α) are perturbed by an annotated chemical library. After sufficient incubation, cells are subjected to a phenotypic assay, in this case a luminescence-based readout of cellular ATP content, that aims to quantify the effect of compounds on oxidative phosphorylation, viability, or other relevant traits. The effect of each compound is expressed as a metric that reflects the difference in compound-induced phenotypes between mutant and wild-type cells. These ratios are then

ranked. The resulting rank list is further analyzed using CSEA (described above). The pattern of enriched compound sets, mechanisms of which are extensively annotated, may provide insights into the function of the gene in question.

2 Materials

1. Chemical library in 96- or 384-well format (*see* **Note 1**).
2. 384-Well assay plates: Solid white plates are recommended for luminescence-based assays whereas black plates are recommended for fluorescence assays.
3. Pin-transfer robot with 96-well and 384-well pin-tool, to add compounds to assay plates, such as the CyBI-Well Vario robot (CyBio US) (*see* **Note 2**).
4. CellTiter-Glo (Promega Corporation) assay for cellular ATP. Depending on the experimental question, different assays may be selected (*see* **Note 3**).
5. Plate reader with stacker: For example, an Analyst HT plate reader (LJL Biosystems, Molecular Devices) but many other comparable devices are available.
6. Liquid handler that can add cells and reagents rapidly to a 384-well assay plate (e.g., Multidrop Combi, Thermo Scientific).
7. Lymphoblastoid cell lines (LCLs) (Coriell Cell Repositories).
8. Cell culture medium: For the LCLs, culture medium is RPMI medium 1640 supplemented with 10 % fetal bovine serum, 2 mM L-glutamine, and 1 % penicillin/streptomycin (10,000 U/mL) solution.
9. GenePattern software available at http://www.broadinstitute.org/cancer/software/genepattern.
10. KNIME software available at https://www.knime.org/knime.

3 Methods

3.1 Cellular Assay

While many of the specific steps are unique to the assay described in the protocol, certain general principles apply to many cell-based phenotypic assay screens.

1. Thaw the LCLs from frozen stocks in 5 mL of culture medium in 6-well cluster plates or T25 flasks. Expand the cells, which grow in suspension, until the desired number of cells is attained. Count the cells daily and dilute with medium or passage cells as needed to maintain a concentration of 100,000–300,000 cells/mL (*see* **Note 4**).

2. Plate the LCLs in 384-well assay plates in 40 μL of medium at a density of 300,000 cells/mL (using a liquid-handling device such as a Multidrop Combi).
3. Pin-transfer compound stocks (~50–100 nL) from the compound library using a CyBi-Well Vario robot. Test each compound dose in each cell line in at least two replicates (*see* **Note 5**).
4. Incubate the plates at 37 °C and 5 % CO_2 for 40 h.
5. Remove the assay plates from the incubator and allow the plates to equilibrate to room temperature for approximately 10 min (*see* **Note 6**).
6. Vortex each plate briefly (30 s), moving the plate across the vortexer in a "Z" pattern (*see* **Note 7**).
7. Prepare the CellTiter-Glo reagent according to the manufacturer's recommendations and allow the reagent to equilibrate to room temperature (*see* **Note 8**).
8. Add 40 μL of CellTiter-Glo to each well of the assay plates (an equal volume as culture media). Place the plate on a vibrating platform for 2 min.
9. Allow the plate to incubate at room temperature for 10 min to stabilize the luminescent signal.
10. Read luminescence values using a microplate reader (*see* **Note 9**).

3.2 Data Analysis

1. Raw screening data from the plate reader and files mapping library wells to compound identity are loaded into KNIME for data transformation and analysis. These steps can be done manually in a spreadsheet program such as Excel, but using KNIME (or other similar programs, such as Pipeline Pilot) is more rapid for larger screens with less chance for human error.
2. The baseline distribution of luminescence values for DMSO-treated control wells is calculated utilizing Statistics and Data View KNIME modules.
3. After data processing, small molecules' effects are assessed by calculating a *Z*-score, which expresses each small molecule's effect in units of standard deviation of the distribution of DMSO-treated wells. Several approaches to preprocessing of screening data have been published elsewhere [14, 15].
4. At this point, the screening dataset consists of multiple cell lines (mutant or wild type at a disease gene), and each cell line is described by thousands of features consisting of small-molecule *Z*-score values. In a gene expression experiment, each cell line is described by features consisting of gene expression values. Subsequent analysis steps will take advantage of the analogy to gene expression and use a publicly available suite of gene expression analysis tools to analyze our small-molecule data (GenePattern; publicly available at http://www.broadinstitute.

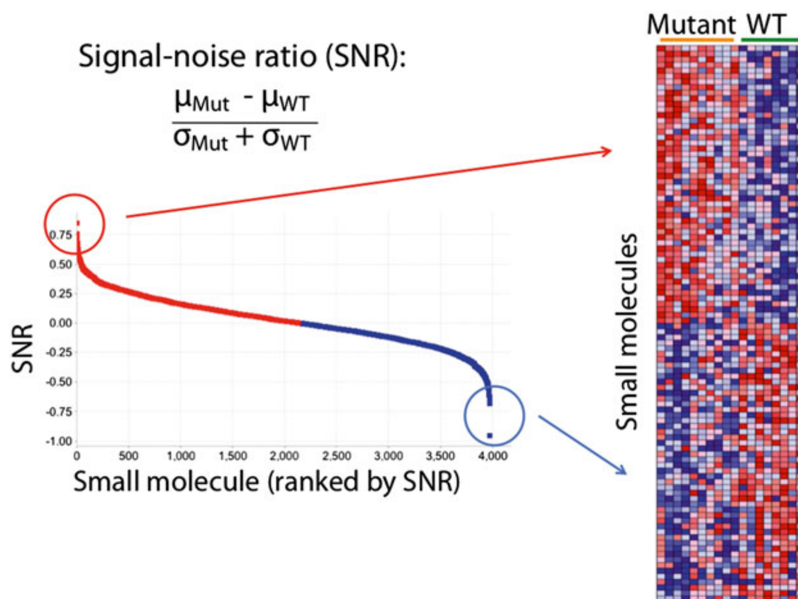


Fig. 4 Calculation of SNR and heatmap representation. The signal-to-noise ratio (SNR) represents the degree to which a compound induces a phenotype that is distinct in mutant versus wild-type cell lines. Compounds with the most positive and most negative SNR represent the compounds that best discriminate between the two classes of cells, and may be depicted in a heatmap (analogous to gene expression data)

org/cancer/software/genepattern) [16]. Other gene expression analysis tools that achieve the same purpose are widely available.

5. Run CSEA. CSEA calculates a "signal-to-noise ratio" (SNR) for each small molecule, which reflects the extent to which small molecule-induced assay values differ between mutant and wild type: $\mathrm{SNR} = (\mu_{\mathrm{MUT}} - \mu_{\mathrm{WT}})/(\sigma_{\mathrm{MUT}} + \sigma_{\mathrm{WT}})$, where μ_{MUT} and σ_{MUT} are the mean and standard deviation, respectively, of the small molecule's composite *Z*-scores in all mutant cell lines (with analogous definitions for μ_{WT} and σ_{WT} in wild-type cell lines) (Fig. 4). CSEA requires three input files: (a) a tab-delimited data table listing composite *Z*-scores for each small molecule in each cell line (*.gct file); (b) a "class assignment" file that specifies which cell lines belong to which class (mutant versus wild type; *.cls file); and (c) a compound set file listing the name of each set and what compounds belong to each set (*.gmx or *.gmt file; compound identifiers must match in this file and the *.gct file) (*see* **Notes 10** and **11**). For each compound set, CSEA generates an enrichment score, normalized enrichment score, a permutation *p*-value, a false discovery rate (FDR), and a visual depiction of where members of the compound set fall within the ranked list (Fig. 3).

6. The compound sets that are used in CSEA analysis can be curated according to structural or functional similarity. For instance, the majority of the sets used in our synthetic interaction analysis were based on the World Health Organization Anatomical Therapeutic Chemical Classification System (http://www.whocc.no/atc_ddd_index) which classifies drugs according to therapeutic use and chemical properties; additional sets can be curated based on membership in pathways based on Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/) and Ingenuity Pathway Analysis (Ingenuity Systems). Compound set files were formatted as described in the GSEA documentation.
7. To identify the most discriminating small molecules by SNR without performing CSEA, we used the ComparativeMarker Selection module of GenePattern; the algorithm requires the same file types as CSEA but without the compound set file. Results of ComparativeMarkerSelection can be viewed graphically using the ComparativeMarkerSelectionViewer module (including heatmap views), and the ExtractComparative MarkerResults module creates a table listing the SNR and *p*-values for all compounds that can be exported to Excel. These analyses identify the most discriminating compounds between mutant and wild-type classes as the small molecules with the greatest magnitude of SNR, either positive or negative (*see* **Note 12**).

4 Notes

1. The compound collections used in the MODY1 study [3] included the following libraries: Prestwick Chemical library of marketed drugs (Prestwick Chemical, 1,120 compounds, 2 mg/mL stock concentration); Spectrum Collection of known bioactives, including drugs, tool compounds, and natural products (MicroSource Discovery Systems, 2,000 compounds, 10 mM stock concentration); Institute of Chemistry and Cell Biology Bioactives collection (Enzo Life Sciences, 480 compounds, variable concentrations); BioMol-NT (Neurotransmitter) collection of neurotransmitter drugs and bioactives (Enzo Life Sciences, subset of 287 compounds, 10 mM stock concentration); and 86 discretes (variable sources, 10 mM stock concentration). In addition, other bioactive sets are commercially available, such as the LOPAC collection (1,280 compounds, Sigma-Aldrich), or collections focused on specific protein families or pathways (e.g., Screen-Well compound libraries, Enzo Life Sciences).

2. While a pin-transfer robot was used in the cited study, liquid transfer from compound stock plates for low- to medium-throughput chemical screens can be conducted by hand using a precision manual pin-tool. Models are available in 48- to 1,536-well formats through manufacturers such as V&P Scientific, Inc.
3. The choice of phenotypic assay is one of the most flexible decisions of this protocol and depends on the biological question being asked. In other investigations, instead of using a biochemical readout like ATP content, we have employed high-throughput epifluorescence microscopy and high-content image analysis software to quantify image-based phenotypes, such as numbers of intracellular organelles and co-localization of intracellular proteins. *See* ref. 7 for more details.
4. We observed that reproducibility of small-molecule phenotypes was increased if the lymphoblast cells were maintained (through counting and dilution daily) at a concentration associated with exponential growth for at least 2 weeks from the initial thaw to harvest for screening [11]. Note that LCLs grow as non-adherent clumps.
5. Each compound plate is pinned into two identical plates containing cells, so that each compound is assayed in duplicate. Also, for every 6–8 compound plates, we include a plate containing only DMSO that is also pin-transferred in duplicate. The specific volume pinned is less important than the final concentration of compound (typically ~10 μM) and DMSO (ideally <0.4 % v/v), and depends on the pin set, volume in the 384-well pates, and other operational parameters.
6. To avoid systematic bias, cell lines that are mutant or wild type at the disease allele are processed in random order.
7. We found that this vortexing step significantly decreases the coefficient of variation of the CellTiter-Glo assay (to <10 %) [11]. We believe that the vortexing helps break up the clumps of LCLs, allowing more complete cell lysis by the CellTiter-Glo reagent.
8. We used the CellTiter-Glo reagent diluted 1:3 in phosphate-buffered saline to save on costs with comparable assay performance [3].
9. Because the luminescence values for the assay are time dependent, we coordinated the timing of addition of CellTiter-Glo addition with the time needed for each plate to be read by the plate reader. For instance, if the plate reader can read a plate every 30 s, then we added the CellTiter-Glo reagent to plates at 30-s intervals.

10. The table of small-molecule screening data (containing small-molecule composite *Z*-scores for each cell line screened) must be reformatted according to the requirements of whatever software is used for analysis. For GenePattern software, detailed descriptions of file formats are described within the online documentation (http://www.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data_formats). While our analysis used the default SNR metric, a *T*-statistic yielded similar results.
11. To apply CSEA to a conventional small-molecule screen in a single cell line, CSEA can be run in its "pre-ranked" mode. In this case, CSEA inputs include a rank-ordered list of small molecules and their assay scores (*.rnk file) and the file containing compound sets (*.gmx or *.gmt file).
12. By analogy to gene expression, the most discriminating small molecules for the class distinction (i.e., most positive and most negative SNR) can be used to define "small-molecule signatures" that are characteristic of two classes of cell lines.

Acknowledgements

This work was supported by contract HHSN268201000044C (NIH/NHLBI) to S.Y.S.; A.D.B. was supported by grant T32HL007208 (NIH), and P.A.T. was supported by grant T32CA79443 (NIH).

References

1. Lehner B et al (2006) Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. Nat Genet 38:896–903
2. St Onge RP et al (2007) Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nat Genet 39:199–206
3. Shaw SY et al (2011) Disease allele-dependent small-molecule sensitivities in blood cells from monogenic diabetes. Proc Natl Acad Sci U S A 108:492–497
4. Mootha VK et al (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 34:267–273
5. Subramanian A et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102: 15545–15550
6. Liu Y et al (2013) Autosis is a Na+, K+-ATPase-regulated form of cell death triggered by autophagy-inducing peptides, starvation, and hypoxia-ischemia. Proc Natl Acad Sci U S A 110:20364–20371
7. Shaw SY et al (2013) Selective modulation of autophagy, innate immunity, and adaptive immunity by small molecules. ACS Chem Biol 8:2724–2733
8. Fajans SS (1990) Scope and heterogeneous nature of MODY. Diabetes Care 13:49–64
9. Byrne MM et al (1995) Altered insulin secretory responses to glucose in subjects with a mutation in the MODY1 gene on chromosome 20. Diabetes 44:699–704
10. Vaxillaire M, Froguel P (2008) Monogenic diabetes in the young, pharmacogenetics and relevance to multifactorial forms of type 2 diabetes. Endocr Rev 29:254–264
11. Choy E et al (2008) Genetic analysis of human traits in vitro: drug response and gene expression

in lymphoblastoid cell lines. PLoS Genet 4:e1000287
12. Remedi MS, Nichols CG (2009) Hyperinsulinism and diabetes: genetic dissection of beta cell metabolism-excitation coupling in mice. Cell Metab 10:442–453
13. Basu A et al (2013) An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell 154: 1151–1161
14. Seiler KP et al (2008) ChemBank: a small-molecule screening and cheminformatics resource database. Nucleic Acids Res 36: D351–D359
15. Malo N et al (2006) Statistical practice in high-throughput screening data analysis. Nat Biotechnol 24:167–175
16. Reich M et al (2006) GenePattern 2.0. Nat Genet 38:500–501

Chapter 3

Construction and Application of a Photo-Cross-Linked Chemical Array

Yasumitsu Kondoh, Kaori Honda, and Hiroyuki Osada

Abstract

Chemical array technology is a powerful tool for high-throughput screening of small-molecule ligand-protein interactions. A chemical array is a collection of small-molecule compounds spotted and immobilized on a glass slide surface, providing a multiplex platform to identify small-molecule compounds binding to a protein of interest in high-throughput screening. Several research groups have developed a variety of methods for the immobilization of small molecules onto a solid matrix. We have developed a unique photo-cross-linked chemical array for immobilizing small molecules in a functional-group-independent manner. In this chapter, we describe in detail a protocol for the construction of a photo-cross-linked chemical array and its application for ligand screening by using a tag-fused protein.

Key words Chemical array, High-throughput screening, Photo-cross-linking, Photoaffinity linker, Trifluoromethylaryldiazirine, PEG, Proline, Ligand screening

1 Introduction

Chemical array screening is one of the most promising approaches for the discovery of bioactive ligands of a protein of interest [1–6]. A chemical array, which is a collection of small-molecule compounds spotted and immobilized on a glass slide, is used to discover ligands of a protein of interest among thousands of library compounds in a high-throughput manner. In chemical array screening, a protein fused with a tag (e.g., fluorochrome, red fluorescence protein, GST, His, or FLAG) is applied on the chemical array and protein molecules binding to immobilized compounds are detected by fluorescence (when tagged with a fluorochrome or red fluorescence protein) or by immunostaining with fluorochrome-labeled antibody (in the case of a nonfluorescent GST, His, or FLAG tag). Ligand screening using chemical array technology does not require structural or functional information about a protein because it is based on a binding assay technique. Therefore, this technology can be used to discover ligands of a protein of unknown structure or function.

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_3, © Springer Science+Business Media New York 2015

Multiple immobilization methods have been reported in earlier studies by several research groups. In general, compounds with reactive functional groups are attached to the surface of glass slides derivatized with certain functional groups by a selective coupling reaction [1, 5, 7, 8]. MacBeath et al. [1] immobilized small molecules containing free thiols to glass slides coated with maleimide groups by the Michael addition reaction. Park and Shin attached maleimide-linked sugars to a glass slide coated with thiol groups [7]. Hergenrother et al. activated glass slides with thionyl chloride and catalytic *N*,*N*′-dimethylformamide to give chlorinated slides, and attached compounds containing a primary alcohol to the slide surface [8]. Barnes-Seeman et al. used a glass slide derivatized with a diazobenzylidene moiety to attach phenols and carboxylic acids [9]. Several other functional-group-dependent immobilization methods have been reported. For additional details of other immobilization methods, excellent reviews are available [5, 10]. These immobilization techniques using selective coupling reactions require reactive functional groups of small molecules, and bias the orientation of the small molecules on the surface of glass slide.

In contrast, a nonselective, functional-group-independent immobilization method has the advantage of increasing molecular diversity of small molecules on the chemical array and the number of binding modes that a given protein molecule can sample. The chemical array described here includes photo-cross-linking for immobilization of compounds onto a glass slide in a functional-group-independent manner [11, 12]. Trifluoromethylaryldiazirine (TAD) [13] is used as a photoreactive group and is introduced onto a glass slide with a spacer such as polyethylene glycol (PEG). Upon UV irradiation, the photoreactive group undergoes photolysis, forming a highly reactive carbene species. The carbene species irreversibly binds to proximal small molecules present on the surface. This photo-cross-linking method does not require incorporation of specific functional groups into the compound library. Therefore, this method is suitable for screening complex natural products.

The length and hydrophilicity of the linker connecting TAD groups with the slide surface and the connection mode between the linker and the slide surface significantly influence the binding signal and background noise when using a protein sample [12]. The photoaffinity PEG linker (Fig. 1a), a combination of TAD and PEG as a spacer, resulted in optimal interaction between FKBP12 and FK506/rapamycin [12]. Recently, we have developed a longer photoaffinity linker (proline linker) (Fig. 1b), in order to enhance the binding signal. The photoaffinity proline linker is designed by inserting a proline helix [14] into the root of the PEG spacer. The distance between TAD and the slide surface is consequently increased, improving the accessibility of protein molecules to immobilized small molecules. This photoaffinity proline linker enhanced the binding signal of p38α MAP kinase and its inhibitor SB203580.

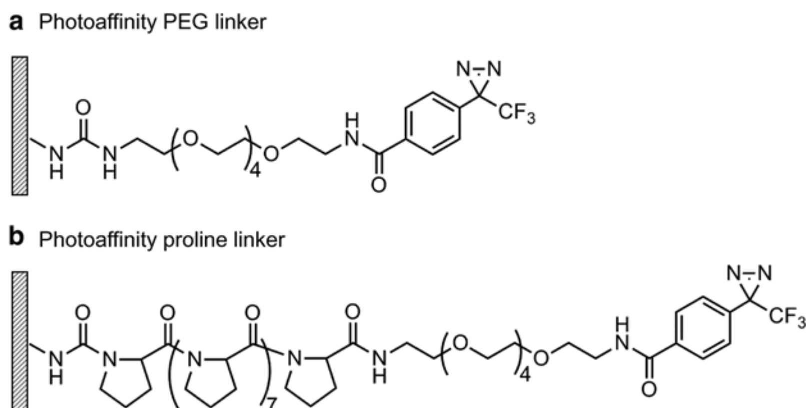


Fig. 1 Photoaffinity linkers. (**a**) Photoaffinity PEG linker; (**b**) photoaffinity proline linker

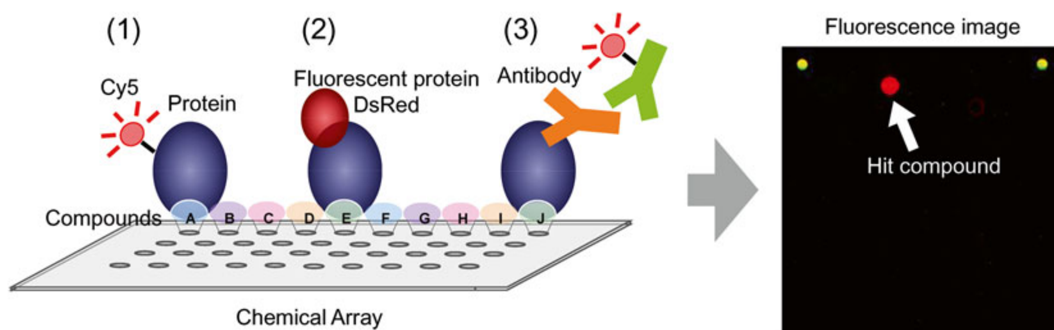


Fig. 2 Method of detection of protein bound to compounds on chemical array. (1) Method using fluorochrome Cy5-labeled protein. (2) Method using immunostaining with primary antibody and Cy5-labeled secondary antibody. (3) Method using a fluorescent DsRed-fused protein

In a chemical array, the ligand compounds are identified either using a protein that is directly labeled with a fluorochrome [15], a protein that is fused with a red fluorescent protein [16, 17], or a protein that is indirectly labeled via immunostaining [18, 19] (Fig. 2). For detection by immunostaining, a His-, GST-, or FLAG-tagged protein is used as a target analyte. Protein concentration and buffer conditions are important for optimal binding assay results. Since the conditions vary for different proteins, optimization is required for each protein under study. In this chapter, we describe a standard protocol for ligand screening by using His-, GST-, or FLAG-tagged protein. Here, we describe an updated method for preparation of photoaffinity linker-coated slides using two types of photoaffinity linkers, fabrication of chemical array and chemical array screening using tag-fused protein as a probe.

2 Materials

All solvents should be purchased at reagent grade and ultrapure Milli-Q grade water is used unless otherwise specified.

2.1 Photoaffinity Linker-Coated Slide Preparation

1. Amine-coated glass slide (*see* **Note 1**).
2. Slide-staining basket.
3. Square slide-staining chamber.
4. Figure-8 shaker.
5. Teflon® slide chamber with glass lid (Fig. 3) (*see* **Note 2**).
6. Spin dryer.
7. Sieves-dried reagent-grade *N*,*N*-dimethylformamide (DMF): Add 500 mL of DMF to 10 g of 4 Å molecular sieves and allow the mixture stand for at least 24 h before use.
8. Slide-activation solution: Add 1.28 g of *N*,*N*′-disuccinimidyl carbonate (10 mM) and 871 μL of *N*,*N*-diisopropylethylamine (10 mM) to 500 mL of sieves-dried reagent-grade DMF (for ~100 slides). Prepare prior to use.
9. Photoaffinity PEG linker (Fig. 1a) [12].

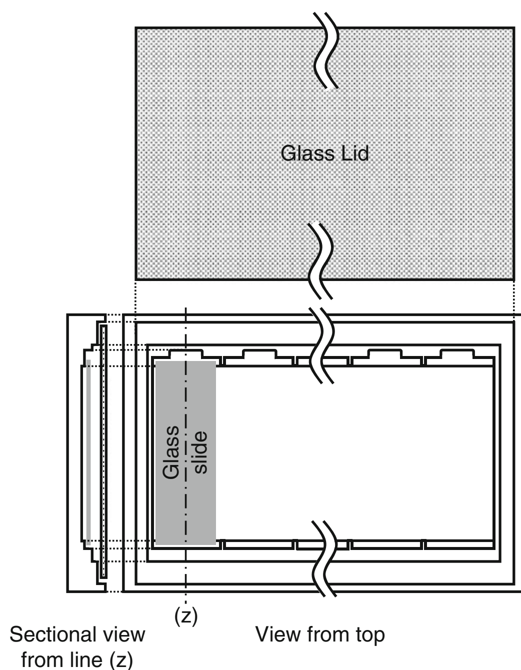


Fig. 3 Teflon® chamber with glass lid and glass slide

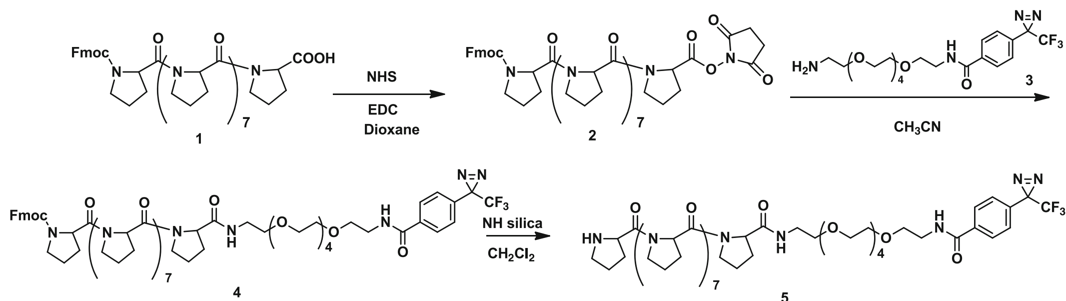


Fig. 4 Synthesis scheme of the proline linker

10. Photoaffinity proline linker (Fig. 1b): *See* Fig. 4 for the synthesis scheme. Briefly, to a solution of *N*-Fmoc-Proline9-OH (**1**, 6.00 g, 4.49 mmol) [14] in anhydrous dioxane (150 mL) is added *N*-hydroxysuccinimide (805 mg, 7.00 mmol) and *N*-(dimethyl-aminopropyl)-*N′*-ethylcarbodiimide (EDC) hydrochloride (1.34 g, 7.00 mmol). The solution is stirred for 5.5 h at room temperature (RT) to generate *N*-Fmoc-Proline9-O-succinimidate (**2**). A solution of H_2N-PEG6-diazirine (**3**, 2.49 g, 5.06 mmol) in anhydrous acetonitrile (10 mL) is slowly added to the solution (**2**, 5.10 g, 4.21 mmol) in acetonitrile (130 mL). The reaction mixture is stirred for 12 h at RT to prepare *N*-Fmoc-Proline9-PEG6-diazirine (**4**). NH-silica gel (100 g) is added to solution **4** (5.06 g, 3.18 mmol) in dichloromethane (250 mL), and the mixture is stirred for 12 h at RT under a nitrogen atmosphere. The suspension is placed on a NH-silica gel bed. Elution with $CHCl_3$:MeOH (10:1) gives the proline linker (**5**, 3.74 g, 2.74 mmol).
11. Photoaffinity linker solution: Add 936 mg of PEG or 2.60 g of proline linker (10 mM) and 1.65 mL of *N*-diisopropylethylamine to 190 mL sieves-dried reagent-grade DMF (for ~100 slides). Prepare prior to use.
12. Blocking solution: Add 30 mL ethanolamine (1 M) to 470 mL of reagent-grade DMF (for ~100 slides). Store at RT.
13. Desiccant: Add silica gel to a permeable plastic bag.
14. Sealer bag.
15. Vacuum sealer.

2.2 Fabrication of Chemical Array

1. Library of compounds.
2. 384-Well v-bottom polypropylene plate (*see* **Note 3**).
3. Chemical array-manufacturing apparatus (Fig. 5) (collaborative development with THK Co., LTD, Tokyo, Japan) (*see* **Note 4**).
4. Ultraviolet cross-linking chamber.
5. Ultrasonic cleaner.

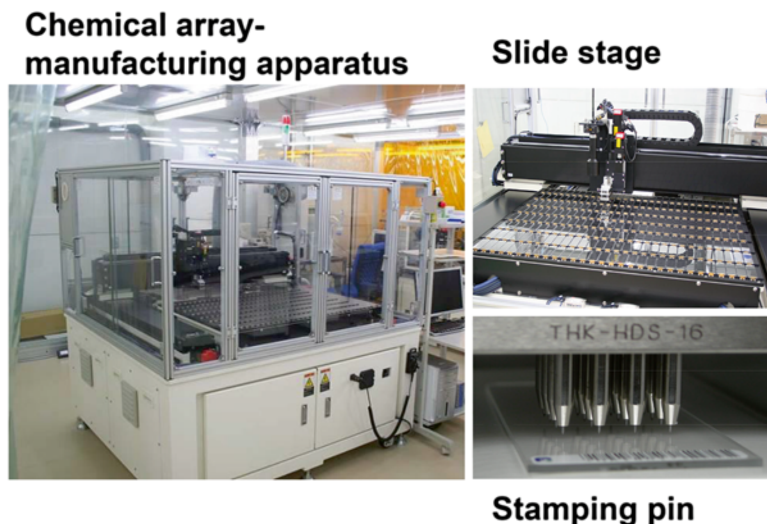


Fig. 5 Chemical array-manufacturing apparatus

6. Barcode printer.
7. Five-slide box.
8. Sealer bag.
9. Vacuum sealer.

2.3 Chemical Array Screening

1. Tris-buffered saline with Tween 20 (TBS-T): Add 200 mL of 1 M Tris–HCl (pH 8.0) (10 mM), 175.32 g of NaCl (150 mM), and 10 mL of Tween 20 (0.05 %) to 19.8 L of Milli-Q water.
2. TBS-T with skimmed milk (S-TBS-T): Add 0.5 g of skimmed milk powder (1 %) to 50 mL of TBS-T.
3. Purified tag-fused protein (His-, GST-, or FLAG-tag).
4. Gap cover glass.
5. Array chamber (*see* **Note 5**).
6. Circular slide-staining chamber.
7. Primary antibody: Monoclonal anti-His antibody produced in mouse, polyclonal anti-GST antibody IgG fraction produced in rabbit, or polyclonal anti-FLAG antibody produced in rabbit (*see* **Note 6**).
8. Secondary antibody: Cy5 goat anti-mouse IgG antibody or Cy5 goat anti-rabbit IgG antibody.
9. Microarray scanner (*see* **Note 7**).
10. Acquisition and analysis microarray software (*see* **Note 7**).
11. Photoshop software.

3 Methods

3.1 Photoaffinity Linker-Coated Slide Preparation

1. Set the glass slides in a slide-staining basket (~20 slides) and immerse them in the slide activation solution using a square slide-staining chamber.
2. Shake the chamber on a figure-8 shaker at RT for 2 h.
3. Wash the glass slides in the chamber with sieves-dried DMF four times, giving 10 min per wash. To exchange DMF during washes, lift up the slide basket with tweezers and discard the solution in the chamber. Pour fresh DMF into the chamber and immerse the basket. Execute this step quickly to avoid drying of the glass slide.
4. Put the glass slides on a Teflon® slide chamber facing down. Fill the space between the chamber and the glass slides with photo-affinity linker solution. Cover the chamber with a glass lid.
5. Incubate the slides in the Teflon® slide chamber overnight (for PEG linker) or over two nights (for proline linker) at RT in the dark.
6. Wash the glass slides by shaking them individually in beakers with ethanol three times, and with reagent-grade DMF once.
7. Place the glass slides directly into a slide-staining basket in a square slide-staining chamber filled with blocking solution.
8. Shake the chamber on a figure-8 shaker at RT for 1 h.
9. Wash the glass slides by shaking them individually in beakers—twice with ethanol and twice with Milli-Q water. Dry the slides for 1 min with a spin dryer.
10. Package the dried slides in a box. Place them in sealer bags with desiccant and seal them with a vacuum sealer. Store the sealed slides at −20 °C until use.

3.2 Fabrication of Chemical Array

1. Dissolve the compounds to be placed on the array at a concentration of 2.5 mg/mL in DMSO (*see* **Note 8**).
2. Transfer 10 μL of each compound solution to individual wells of a 384-well polypropylene plate (*see* **Note 9**).
3. Centrifuge the plates at $400 \times g$ for 30 min in a benchtop centrifuge.
4. Incubate the plates for 3 h at 35 °C in a humidified incubator (*see* **Note 10**). Place the plates on a stacker on the chemical array-manufacturing apparatus.
5. Warm the linker-coated glass slides to RT and place them on the chemical array-manufacturing apparatus stage.
6. Array compounds in the desired format (Fig. 6) by aspirating a small amount of compound solution from a 384-well polypropylene plate with stamping pins and by spotting a droplet onto

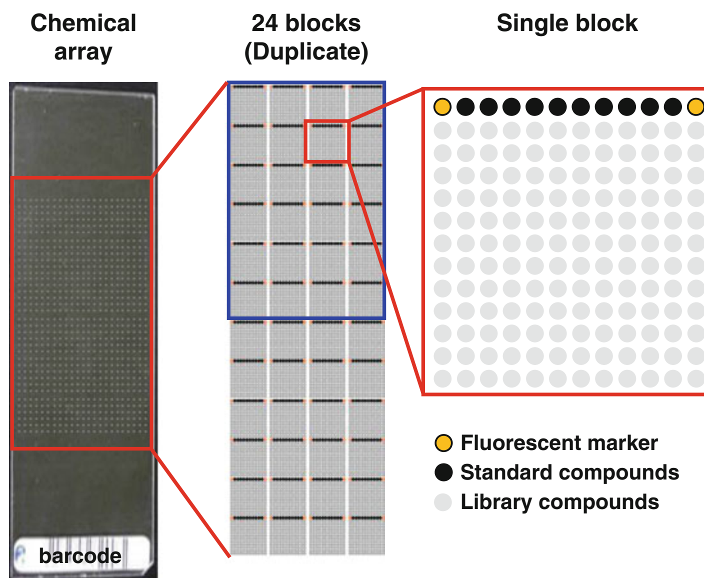


Fig. 6 Chemical array, formatted with 24 blocks in duplicate. Each block contains 144 library compounds, 10 standard compounds, and 2 fluorescent markers

glass slides (2–200 slides). Wash and dry the stamping pins (*see* **Note 11**), aspirate compounds from subsequent wells, and spot onto glass slides. Repeat this spot-and-wash movement until all compounds in the plates have been transferred (*see* **Note 12**). For a small-scale array, *see* **Note 13**.

7. Collect the glass slides and cross-link them with 365 nm ultraviolet light at 4 J/cm^2 in an ultraviolet cross-linking chamber (*see* **Note 14**).
8. Set the glass slides in the slide-staining basket. Immerse the basket in a square slide-staining chamber filled with DMSO, and then immerse the basket containing DMSO filled chamber in an ultrasonic cleaner filled with water. Wash the glass slides by sonicating three times for 5 min. Exchange solvents in the chamber sequentially and repeat washes using DMF, acetonitrile, THF, dichloromethane, ethanol, and Milli-Q water in sequence. Keep the temperature of the ultrasonic cleaner water bath under 35 °C by adding ice, if necessary.
9. Rinse the glass slides for about 10 s individually in a beaker with Milli-Q water. Dry the slides for 1 min using a spin dryer.
10. Print barcodes and place them on the edge of the arrays (Fig. 6). For smaller scale arrays, a hand-written label can be used.
11. Package the dried slides in five-slide boxes. Put them in sealer bags and seal them using a vacuum sealer. Store the sealed slides at −20 °C until use.

3.3 Chemical Array Screening

1. Warm the chemical array to RT. Put the array in a plastic case filled with S-TBS-T and incubate the array for 1 h at RT on a figure-8 shaker. Prepare two arrays for one sample. One is for applying sample protein, while the other is to be used as a reference without applying sample protein.
2. Wash the array with TBS-T three times for 5 min each on a figure-8 shaker. Dry the array for 1 min using a spin dryer.
3. Dilute the purified, tag-fused protein to a concentration of 1–5 μM in S-TBS-T (*see* **Note 15**). For a reference array, skip Subheading 3.3, **steps 3**–7.
4. Place the dried array on the array chamber facing up and cover it with a gap cover glass. Fill the space between the array and the cover glass with ~50 μL of diluted sample solution.
5. Place water between the array and the chamber surface. Close the lid of the chamber and incubate for 1 h in a 30 °C incubator (*see* **Note 16**).
6. Fill a circular slide-staining chamber with TBS-T. Shake the array gently in a chamber and remove the cover glass slowly (*see* **Note 17**).
7. Wash the array in a plastic case filled with TBS-T three times for 5 min on a figure-8 shaker. Dry the array for 1 min using a spin dryer.
8. Dilute the primary antibody (1:100) in S-TBS-T.
9. Place the dried array on the array chamber facing up and cover the chamber with a gap cover glass. Fill the space between the array and the cover glass with ~50 μL of diluted primary antibody solution.
10. Repeat Subheading 3.3, **steps 5**–7.
11. Dilute the secondary antibody 1:100 with S-TBS-T.
12. Place the dried array on a chamber facing up and cover the chamber with gap cover glass.
13. Fill the spaces between the array and cover glass with ~50 μL of diluted secondary antibody solution.
14. Repeat Subheading 3.3, **steps 5**–7.
15. Scan the array using the Cy5 channel (excitation: 635 nm, emission: 655–695 nm) on a microarray scanner (*see* **Note 18**).
16. Use microarray software (GenePix Pro7) to quantify the fluorescence signal from each compound spot.
17. Compare the fluorescence signal between the sample and reference arrays, and calculate significant differences in signal intensity between the arrays by using a spreadsheet software such as Microsoft Excel. First of all, calculate real fluorescence intensity I of each spot by subtracting spot local background

noise B from spot fluorescence signal F ($I = F - B$). Then, reference intensity I_r is subtracted from sample intensity I_s to calculate ΔI ($\Delta I = I_s - I_r$). The average and standard deviation (SD) of intensity for all spots on the array is also calculated. Finally, calculate the significance by calculating the difference between ΔI and the average of total spots on the array. The compounds with a ΔI larger than the average and having a difference >1 SD are identified as a positive. Create a list of positive compounds.

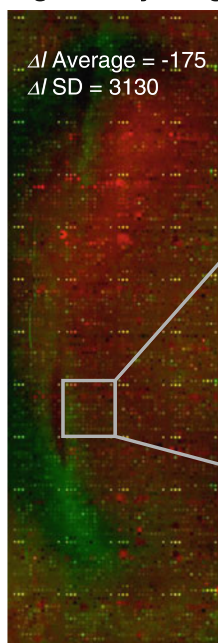
18. Merge the image files of sample and reference arrays using an image analysis software such as Photoshop. Display the sample image using the red color and the reference image using the green color. Merge the two images (Fig. 7).

19. Compare the list of positive compounds and the merged image to determine the final list of positive compounds (Fig. 7). Check the corresponding spot of the listed compounds in the

Positive list

Block	Row	Column	Compound	I_s	I_r	ΔI	ΔI - ave	Ave + nSD
30	1	8	SB203580	25778	2	25776	25951	8.3
30	1	9	SB203580	19815	21	19794	19969	6.4

Merged array image



Block 30 array image

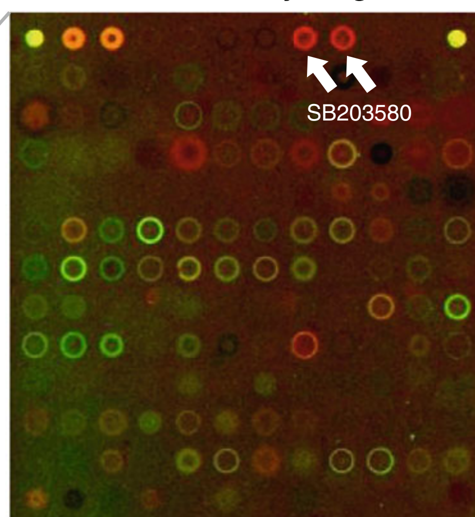


Fig. 7 List of positive compounds and a representative merged image. Sample protein: His-p38α MAP kinase. The listed compound SB203580, an inhibitor of p38α MAP kinase, is indicated by a *red* color on the image (*arrows*)

merged images, and if the spot is indicated by red color, it is confirmed as a final positive compound. In contrast, if the merged image has a corresponding yellow/green color spot or when the red color is caused by scratches/dust, it is designated as false positive.

4 Notes

1. High-density amine-coated glass slide (Matsunami Glass Industries, Ltd.) is recommended.
2. Designed and produced collaboratively with Fuji Giken.
3. Only X6004 384-well plates from Molecular Devices can be used by our chemical array-manufacturing apparatus.
4. For a small number of chemicals, a P2 pipetman can be used.
5. A typical hybridization chamber can be used.
6. It is possible to use a primary antibody specific for the sample protein. We use antibody against tags because they are versatile for use with many kinds of protein.
7. In our laboratory, the microarray scanner GenePix 4300A and its analysis software GenePix Pro7 (Molecular Devices, LLC, California, USA) are used for microarray data acquisition and analysis.
8. Our compound library stocks exist in solution at a concentration of 10 mg/mL. We usually dilute these stock solutions to 2.5 mg/mL to prepare compounds for the array. Molar concentrations may also be used for compound solution preparation. The recommended concentration is 5–10 mM.
9. Automated transfer robotics can be used if compound samples are abundant.
10. Moisturizing the compound solution stabilizes the size and shape of the spots on arrays produced by our chemical array-manufacturing apparatus.
11. The stamping pin washing step is very important for reducing carryover of compounds. We usually wash the pins successively in DMSO, ethanol, and Milli-Q water using sonication.
12. Our chemical array-manufacturing apparatus enables loading of 7,488 spots on a single array. The usual format in our laboratory is 96 marker spots, 480 standard compound (e.g., biotin, His peptide, and FLAG peptide) spots, and 6,912 library spots (Fig. 5). Standard compounds are used to confirm the completeness of an array or are used as a positive control for some tagged proteins.

13. For a small number of compounds, use a P2 pipetman to spot compound solutions. Aspirate 0.02 μL and spot the droplet onto glass slide.
14. If the joule unit is unsupported in the cross-linking chamber being used, calculate the required intensity by the equation "W = J/s" (W: watt, J: joule, s: second).
15. Try various concentrations for the protein between 1 and 5 μM to optimize conditions. We decide concentration by (1) histogram peak of pixel signal intensity of array surface, and (2) polarization and intensity of background signal. (1) Using the GenePix Pro7 software, the histogram of pixel signal intensity can be checked. We decide the screening concentration by estimating the intensity of the highest peak in the histogram. The ideal intensity is between 3,000 and 5,000. (2) If the microarray scanning software is different, the only indicator of the best conditions will be background polarization. If background is too high, the concentration must be low, but if background is very low and concentration and volume of sample stock solution are enough, screening concentration can be >5 μM.
16. Array screening temperature can be at RT. However, we recommend using a 30 °C incubator if the temperature at the bench is not constant and is subject to fluctuations.
17. Do not let the cover glass scratch the surface of the array. It is preferable to let the cover glass slide off due to its own weight by tilting it downward.
18. The recommended label for secondary antibody is Cy5 (or other fluorophores with similar excitation/emission wavelengths). A shorter wavelength channel (e.g., Cy3, FITC) causes higher autofluorescence of compounds, which interfere with antibody-derived fluorescence.

References

1. MacBeath G, Koehler AN, Schreiber SL (1999) Printing small molecules as microarrays and detecting protein-ligand interactions en masse. J Am Chem Soc 121:7967–7968
2. Kuruvilla FG et al (2002) Dissecting glucose signalling with diversity-orienter synthesis and small-molecule microarrays. Nature 416: 653–657
3. Ma H, Horiuchi KY (2006) Chemical microarray: a new tool for drug screening and discovery. Drug Discov Today 11:661–668
4. Uttamchandani M, Wang J, Yao SQ (2006) Protein and small molecule microarrays: powerful tools for high-throughput proteomics. Mol Biosyst 2:58–68
5. Vegas AJ, Fuller JH, Koehler AN (2008) Small-molecule microarrays as tools in ligand discovery. Chem Soc Rev 37:1385–1394
6. Stanton BZ et al (2009) A small molecule that binds Hedgehog and blocks its signaling in human cells. Nat Chem Biol 5:154–156
7. Park S, Shin I (2002) Fabrication of carbohydrate chips for studying protein-carbohydrate interactions. Angew Chem Int Ed 41: 3180–3182
8. Hergenrother PJ, Depew KM, Schreiber SL (2000) Small-molecule microarrays: covalent attachment and screening of alcohol-containing small molecules on glass slides. J Am Chem Soc 122:7849–7850

9. Barnes-Seeman D et al (2003) Expanding the functional group compatibility of small-molecule microarrays: discovery of novel calmodulin ligands. Angew Chem Int Ed 42: 2376–2379
10. Walsh DP, Chang YT (2004) Recent advances in small molecule microarrays: applications and technology. Comb Chem High Throughput Screen 7:557–564
11. Kanoh N et al (2003) Immobilization of natural products on glass slides by using a photoaffinity reaction and the detection of protein-small-molecule interactions. Angew Chem Int Ed 42:5584–5587
12. Kanoh N et al (2006) Photo-cross-linked small-molecule microarrays as chemical genomic tools for dissecting protein-ligand interactions. Chem Asian J 1:789–797
13. Brunner J, Senn H, Richards FM (1980) 3-Trifluoromethyl-3-phenyldiazirine. A new carbene generating group for photolabeling reagents. J Biol Chem 255:3313–3318
14. Sato S et al (2007) Polyproline-rod approach to isolating protein targets of bioactive small molecules: isolation of a new target of indomethacin. J Am Chem Soc 129:873–880
15. Minagawa S et al (2011) Cyclic lipopeptide antibiotics bind to the N-terminal domain of the prokaryotic Hsp90 to inhibit the chaperone activity. Biochem J 435:237–246
16. Miyazaki I et al (2008) Robust and systematic drug screening method using chemical arrays and the protein library: identification of novel inhibitors of carbonic anhydrase II. Biosci Biotechnol Biochem 72: 2739–2749
17. Miyazaki I et al (2010) A small-molecule inhibitor shows that pirin regulates migration of melanoma cells. Nat Chem Biol 6: 667–673
18. Burger M et al (2012) Crystal structure of the predicted phospholipase LYPLAL1 reveals unexpected functional plasticity in spite of close relationship to acyl protein thioesterases. J Lipid Res 53:43–50
19. Zimmermann TJ et al (2013) Boronic acid inhibitors of acyl protein thioesterase 1 and 2. ChemBioChem 14:115–122

Chapter 4

High Content Screening for Modulators of Cardiac Differentiation in Human Pluripotent Stem Cells

Sean Spiering, Herman Davidovics, Paul J. Bushway, Mark Mercola, and Erik Willems

Abstract

Chemical genomics has the unique potential to expose novel mechanisms of complex cellular biology through screening of small molecules in in vitro assays of a biological phenotype of interest, followed by target identification. In the case of disease-specific assays, the cellular proteins identified might constitute novel drug targets, and the small molecules themselves might be developed as drug leads. In cardiovascular biology, a chemical genomics approach to study the formation of cardiomyocyte, vascular endothelial, and smooth muscle lineages might contribute to therapeutic regeneration. Here, we describe methods used to develop high content screening assays implementing multipotent cardiovascular progenitors derived from human pluripotent stem cells and have identified novel compounds that direct cardiac differentiation.

Key words Human pluripotent stem cells, Small molecules, Small RNAs, High content screening, Cardiac differentiation

1 Introduction

The ability to generate pluripotent stem cells (PSC) from humans has provided the unprecedented opportunity to study human development and disease in vitro [1, 2]. Efforts over the past decade to define the mechanisms underlying cardiomyocyte generation from pluripotent stem cells have led to the commercial production of human cardiomyocytes from pluripotent stem cells and have shed light on fundamental developmental mechanisms that may underlie the etiology of congenital heart disease [3, 4]. Moreover, these endeavors might lead to strategies for therapeutic regeneration of adult hearts, which retain a modest ability to self-renew following injury [5].

To probe the biology of PSC differentiation to the cardiac lineage in a large-scale and unbiased way, we have developed phenotypic high throughput assays that allow the simultaneous screening of thousands of bioactive compounds. The large amount

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_4, © Springer Science+Business Media New York 2015

of data points in such screens accelerates the discovery of novel biological mechanisms relevant for stem cell differentiation. A typical screening assay consists of large-scale expansion of PSC, which are differentiated in bulk to enrich for a progenitor of interest. Progenitor enriched cultures are then seeded into a high throughput plate format, after which small molecules can be added at any time. The effects of compounds are visualized through a reporter system, usually based on phenotype specific promoters driving a fluorescent or luciferase reporter, or through staining with fate specific antibodies. High throughput plate readers or microscope systems are then implemented to collect data. Consecutive steps comprise hit verification, analysis of the biological mode of action and target identification of the discovered small molecules.

Our screening campaigns using differentiating PSC cultures have led to the discovery of two important classes of molecules that drive the conversion of mesoderm to cardiac mesoderm via TGFβ inhibition and cardiac mesoderm to cardiomyocytes through Wnt inhibition [6–9]. Stem cell-based screening approaches are however not limited to cardiac differentiation, for example small molecules targeting endoderm differentiation have been identified using similar screening methods [10, 11]. Here, we provide methods for chemical genomics applied to cardiovascular biology, which has already yielded insights into basic differentiation mechanisms and reagents to generate large numbers of pure cardiomyocytes. Such molecules might moreover be useful as tools to probe whether the target proteins can be engaged to enhance the limited regenerative potential of the adult heart [1, 4, 12].

2 Materials and Equipment

2.1 Reagents

1. Human PSC (hPSC) Growth medium: Knockout Dulbecco's Modified Eagle's medium (KO DMEM, Invitrogen) supplemented with 2 mM L-glutamine with 20 % Knockout Serum Replacement (KOSR, Invitrogen), 0.1 mM nonessential amino acids (NEAA), 50 U/mL penicillin, 50 mg/mL streptomycin, 0.1 mM beta-mercaptoethanol, and 8 ng/mL basic fibroblast growth factor (bFGF). An alternative medium can be used to maintain hPSC in culture (*see* **Note 1**).
2. hPSC Differentiation medium: Stem Pro 34 with the included supplement and 2 mM L-glutamine, 0.1 mM NEAA, 50 U/mL penicillin, 50 mg/mL streptomycin, 0.1 mM beta-mercaptoethanol, and 5 ng/mL ascorbic acid.
3. Mouse embryonic fibroblast (MEF) medium: DMEM High Glucose (4.5 mg/mL), 10 % Fetal Bovine Serum (FBS), 2 mM L-glutamine, 50 U/mL penicillin, 50 mg/mL streptomycin.

4. Serum-free medium (SFM): 75 % Iscove's Modified Dulbecco's Medium (IMDM), 25 % Ham's-F12 medium complemented with 2 mM L-glutamine, 50 U/mL penicillin, 50 mg/mL streptomycin, 1 % of the serum replacing B27 supplement without vitamin A (GIBCO), 0.5 % of the serum replacing N2 supplement (GIBCO), 0.5 mM ascorbic acid, 0.05 % Bovine Serum Albumin (BSA), and 0.46 mM 1-thioglycerol. This medium is stable for about a week and is ideally made fresh for every use.
5. 1× TrypLE Express dissociation reagent (GIBCO).
6. Collagenase IV: diluted in KO DMEM at 1 mg/mL or 1.5 mg/mL (concentration depends on the application).
7. 83.3 μg/mL Growth factor reduced Matrigel (BD Biosciences): diluted in ice cold KO DMEM.
8. 0.1 % Gelatin solution: diluted in sterile water.
9. 10 ng/μL Activin A (R&D Systems): diluted in Phosphate Buffered Saline (PBS) with 0.1 % BSA (*see* **Note 2**).
10. 25 ng/μL Bmp4 (R&D Systems): diluted in PBS containing 0.1 % BSA and 4 mM HCl (*see* **Note 2**).
11. 25 ng/μL bFGF: diluted in KO DMEM.
12. 10 mM Inhibitor of Wnt Response (IWR): 53AH [6] or the analog IWR-1 diluted in dimethylsulfoxide (DMSO).
13. 30 μM Triiodothyronine (T3): diluted in DMEM high glucose (*see* **Note 3**).
14. Opti-MEM (GIBCO) (optional, if screening siRNA instead of small molecules) (*see* **Note 4**).
15. Lipofectamine RNAiMax (Invitrogen) (optional, if screening siRNA or miRNA instead of small molecules) (*see* **Note 4**).
16. 10 mM Thiazovivin: diluted in DMSO.
17. Ultralow attachment 6-well plates (CoStar).
18. 0.1 % gelatin/Matrigel solution: add 300 μL of Matrigel solution per mL of 0.1 % gelatin.
19. 4 % Paraformaldehyde (PFA) in PBS: add 4 g of PFA to 100 mL of PBS while heating the solution to 65 °C until PFA is completely dissolved. Filter with a 0.20 μm filter before use.
20. Blocking buffer: 1× PBS supplemented with 5 % FBS and 0.1 % Triton X-100.
21. Antibody staining buffer: 1× PBS supplemented with 5 % FBS and 0.5 % Triton X-100.
22. 1× PBS (sterile, without magnesium and calcium).
23. Antibodies: MYH6 clone MF20 (supernatant, Hybridoma Bank), α-Actinin (ascites solution, Sigma), PDGFRA directly

labeled with the fluorochrome allophycocyanin (APC) (saline solution, R&D Systems), anti-mouse Alexa 488 or 568 (2 mg/mL, Molecular Probes) (*see* **Note 5**).

24. Optical black 384-well plates.
25. 2× Trypan blue solution.

2.2 Cell Lines

1. Our preferred human embryonic stem cell line (hESC) H9 for screening carries a MYH6-mCherry reporter, and nuclear PGK1-H2B-GFP reporter [13]. MYH6 is a cardiomyocyte specific gene; by coupling its promoter to a fluorescent reporter we can thus visualize cardiomyocytes being formed. The PGK1-H2B-GFP reporter yields a nuclear GFP signal in every cell driven by the ubiquitous PGK1 promoter, which allows cell counting for toxicity analysis. These reporters thus facilitate high content analysis downstream (*see* **Note 6**). Other cell lines are equally suitable, as we have successfully achieved 384-well assays with hESC from the H7 line as well as human induced pluripotent stem cell lines. In case the cell line does not have a reporter, a MYH6 reporter can be inserted easily if needed (we have made lentiviral reporters available on www.addgene.org, plasmids 21228 or 21229, also see Kita-Matsuo et al. [13]) or immunostaining can be implemented for the readout (*see* Subheading 3.5.2).
2. Mouse embryonic fibroblasts (MEFs), various commercial sources are available.

2.3 Equipment

1. Laminar flow cabinet with stereoscope.
2. Biosafety cabinet.
3. Automated cell counter, type Countess (Invitrogen) or TC20 (Bio-Rad) or equivalent.
4. Flow cytometer, type FACSCanto (BD Biosciences) or LSR Fortessa (BD Biosciences) or equivalent.
5. Liquid handler, type Star (Hamilton Robotics) or equivalent.
6. Incubator at 37 °C with 5 % CO_2.
7. High content microscope, type InCell (GE Healthcare), Opera (Perkin Elmer), Celigo (Brooks), or equivalent.

3 Methods

We have built hESC/hIPSC differentiation assays to study cardiac differentiation that have several main advantages over classic differentiation protocols (Fig. 1). They are completely serum-free, which focuses the small molecule biology on the differentiation of the cells, rather than on effects of serum components. Secondly, we miniaturized the assay to allow simultaneous screening of thousands of small molecules or small RNAs (*see* **Note 7**).

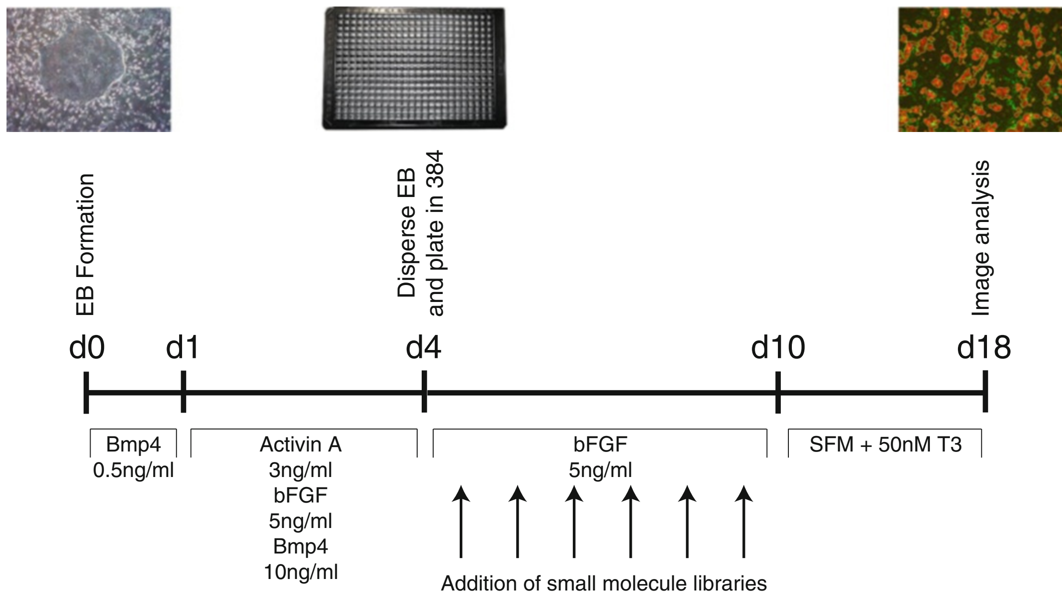


Fig. 1 Schematic overview of the cardiac differentiation assay. hESC/hIPSC are differentiated with Activin A and Bmp4 in the form of embryoid bodies from day 0 to day 4 to maximize the formation of MESP1+/PDGFRA+ cells. EB are then dissociated to single cells and plated into optical 384-well plates. At this stage small molecules can be added at any desired time or concentration. At day 10 of differentiation, the cells are exposed to T3 to maximize the MYH6 response. At day 18 the plates are processed and imaged on an automated microscope and analyzed automatically

3.1 hESC/hIPSC Culture

Human ESC or IPSC lines can be maintained in the pluripotent state by growing the cells on Matrigel in the presence of MEFs. We describe routine maintenance methods, which we find most suitable for screening purposes. However, the cells can also be maintained feeder-free (*see* **Note 1**).

1. Coat 6-well plates with 1 mL of growth factor reduced Matrigel, overnight at 4 °C. One 6-well plate typically yields enough cells to fill two full 384-well plates in Subheading 3.4.
2. Plate MEFs in hPSC growth medium at 250,000 cells per well of a 6-well plate, allowing the MEFs to attach overnight before seeding hESC/hIPSC the following day (*see* **Note 8**).
3. Prepare hESC/hIPSC for culture from a frozen vial or from a previous culture. One 6-well plate typically yields enough cells to fill two full 384-well plates in Subheading 3.4.
4. For recovery of a frozen stock, thaw a frozen vial of cells by plating them in 3 mL of growth medium in the presence of 2 μM Thiazovivin for increased recovery (*see* **Note 9**). If splitting from a previous culture, incubate hESC/hIPSC cultures with 1 mg/mL collagenase IV solution for 5 min at 37 °C after removing the growth medium. After the incubation, replace the collagenase IV with growth medium and mechanically

dissociate the colonies (by slicing the colonies into small pieces with a 2 mL serological pipet) under a stereoscope and plate at the desired density (*see* **Note 10**).

5. Maintain the hESC/hIPSC in a 37 °C incubator, changing the growth medium daily until cells are confluent. When confluent, cells can be used for freezing, expansion, or differentiation.

3.2 Bulk Differentiation of MESP1/PDGFRA Positive Progenitors

For bulk differentiation of hESC/hIPSC into $MESP1^+/PDGFRA^+$ progenitor enriched cultures, we make use of an embryoid body (EB) differentiation step, which is based on a previously described protocol [14]. We pre-passage the hESC/hIPSC onto Matrigel spiked gelatin-coated dishes to facilitate colony removal for EB formation (*see* **Note 11**).

1. Coat 6-well plates with 1 mL per well of 0.1 % gelatin/Matrigel solution. Coating can be done overnight at 37 °C (*see* **Note 12**). Once coating is completed, aspirate the coating solution until the plate is completely dry (*see* **Note 13**). Two 6-well plates will provide enough cells for one full 384-well plate in Subheading 3.4.
2. Seed MEFs on plates at 250,000 cells per well in hPSC growth media and allow MEFs to settle overnight before seeding hESC/hIPSCs (*see* **Note 14**). Check the following day to ensure attachment of MEFs is sufficient before seeding hESC/hIPSCs. If the MEFs did not attach well, do not continue with these plates, it is best to start over.
3. Split hESC/hIPSC as described in Subheading 3.1, **steps 3** and **4**, but now divide the colonies from one confluent well of a 6-well plate to 4 wells of a gelatin/Matrigel coated 6-well plate containing 3 mL of hPSC growth medium per well to allow maintenance of pluripotency.
4. Allow colonies to grow for 3–4 days, aspirate hPSC growth medium daily and replace with 4 mL of fresh hPSC growth medium.
5. After 3–4 days the colonies have grown significantly, but should not yet be touching each other. At this stage, remove the hPSC growth medium and incubate the cultures with 1 mL of a 1.5 mg/mL collagenase IV solution for 15 min at 37 °C (*see* **Note 15**).
6. Once the colonies are detached from the plate and are floating in collagenase IV, add 1 mL of hPSC growth medium per well to neutralize the collagenase IV.
7. Collect the floating colonies and place them into a 15 mL (or 50 mL) conical tube and allow them to settle by gravity pelleting (*see* **Note 16**). Rinse the plate with additional hPSC growth medium to collect remaining colonies and add to the conical tube.

8. Once pelleted, wash the hESC/hIPSC colonies in 2 mL of hPSC growth medium and again allow the colonies to pellet by gravity. Perform an additional wash in 1 mL of hPSC differentiation medium.
9. Start differentiation for the screening assay (this is day 0 of the assay, the outline of the assay is detailed in Fig. 1) by plating the colonies of one well of a 6-well plate into one well of an ultralow attachment 6-well plate containing 2 mL of hPSC differentiation medium. Add 0.5 ng/mL of Bmp4 to the cultures. Incubate the floating colonies at 37 °C overnight, allowing them to form sphere-like structures, known as embryoid bodies (EB).
10. At Day 1 of differentiation, aspirate the medium containing the EB from each well with a 5 mL serological pipet and transfer to a 15 mL conical tube. Pellet the EB by gravity. Remove the hPSC differentiation medium and add 2 mL of fresh hPSC differentiation media per well containing 10 ng/mL Bmp4, 3 ng/mL Activin A, and 5 ng/mL bFGF. Return the 2 mL containing the EB to each well of the low attachment plates (*see* **Note 2**). Do not continue beyond this point if major cell death is seen or if no EB are formed.
11. At Day 3, collect EBs by gravity pelleting and refresh hPSC differentiation media again including 10 ng/mL Bmp4, 3 ng/mL Activin A, and 5 ng/mL bFGF (*see* **Note 2**).
12. At Day 4 of differentiation, the EB are ready to be dissociated for plating into the 384-well format (*see* Subheading 3.3). Alternatively, the EB can be kept for continued differentiation in EB form to test new lots of Activin A and Bmp4 (*see* **Note 17**).

3.3 Cell Preparation and Quality Control of Bulk MESP1+/PDGFRA+ Cultures

Before moving ahead with small molecule screening we routinely perform quality control tests on day 4 EB cells.

3.3.1 EB Dissociation and Viability Assessment

1. At day 4 of differentiation aspirate the medium containing EB and collect 3 wells from a 6-well plate into one 15 mL conical tube by gravity pelleting. Aspirate the hPSC differentiation medium and wash the EB with 5 mL of 1× PBS.
2. After pelleting again, resuspend the EB in 2 mL of TrypLE per 15 mL conical tube and incubate the 15 mL tubes at 37 °C for 5 min (*see* **Note 18**).
3. Dissociate the EB gently by repetitive resuspending with a 1 mL micropipet tip (*see* **Note 19**).
4. Once dissociated to a single cell suspension (*see* **Note 20**), add 6 mL of MEF medium to each 15 mL conical tube containing 2 mL of TrypLE.

5. Pellet single cells by a 5 min centrifugation at 300 × *g*.
6. Aspirate the supernatant and resuspend the cells in 1 mL of hPSC differentiation medium per 15 mL tube.
7. Remove possible clumps by running the single cell suspension from all 15 mL tubes through a 40 μm cell strainer placed on a 50 mL conical tube. Single cells will flow through and the clumps will remain on the strainer. Discard the strainer at this point. Use the flow through for Subheadings 3.3.2 and 3.4.
8. Prepare the samples for counting by mixing 10 μL of the cell suspension from Subheading 3.3.1, **step 7** with 10 μL of trypan blue. Use 10 μL of this solution to manually or automatically count the cells (*see* **Note 21**). Cell viability at this stage is an important quality control checkpoint. If the cells do not have a viability of 90 % or more, we do not proceed with screening (*see* **Note 22**).

3.3.2 PDGFRA Expression

At this step the cultures can also be verified for PDGFRA expression as an extra quality control step for the first 4 days of differentiation.

1. Pellet 10^5 cells of the cell suspension by centrifugation at 300 × *g* for 5 min, and resuspend in 100 μL of MEF medium, here used as staining buffer.
2. Add PDGFRA antibody at 4 μL per 10^5 cells and incubate for 20 min on ice.
3. Wash the cells three times in 1× PBS, and analyze the cells on a flow cytometer. Ideally the day 4 cultures should have at least 30 % of $PDGFRA^+$ cells for a reliable screen. Lower yields of $PDGFRA^+$ cells reduce the dynamic range of cardiac induction in the assay.

3.4 Differentiation of MESP1 Enriched Cultures in 384-Well Plates

After checking the viability of the cells and determining incidence of $PDGFRA^+$ cells, the assay is continued in the 384-well format for high throughput screening purposes.

1. Coat optical 384-well plates with 25 μL of 0.1 % gelatin (this can be done before Subheading 3.3). Dispensing of the gelatin solution can either be done with a 16-channel pipettor (*see* **Note 23**) or using a liquid handler for larger scale. Incubate the coated plates for at least 1 h at 37 °C before seeding the cells. Aspirate the gelatin solution using an 8-channel aspirator device.
2. Seed cells in hPSC differentiation medium including 5 ng/mL of bFGF, at a density of 17,000 cells per well of a 384-well plate, in a total volume of 75 μL per well (*see* **Notes 22** and **24**).
3. Add small molecules at the desired concentration directly into the wells prepared in Subheading 3.4, **step 2** (various approaches

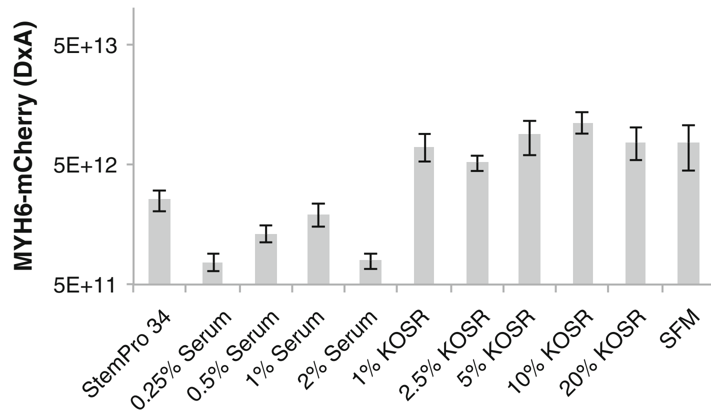


Fig. 2 Effects of media and serum on the differentiation assay. Differentiation was quantified after the indicated media were used from day 10 of differentiation, the stage when cardiomyocytes form. The use of serum at indicated concentrations suppresses cardiac differentiation compared to StemPro medium. In contrast, use of a range of KOSR containing media or a serum free medium increased differentiation over the level observed StemPro medium

for compound addition exist) (*see* **Note 25**). Compounds can be added any time between day 4 and end of assay at day 18 to probe different time windows of differentiation (*see* **Notes 26–30**). siRNAs can also be transfected in at day 4 (*see* **Note 3**).

4. Incubate the plates in a 37 °C incubator until the end of assay (Day 18) (*see* **Note 31**).
5. Depending on the experimental design, compounds can be washed out any day by removing the hPSC differentiation medium and adding 75 μL of fresh hPSC differentiation medium per well.
6. At day 10 of differentiation, remove the hESC differentiation medium and add 75 μL of SFM including 100 nM of T3 (*see* **Notes 2** and **32**) (Figs. 2 and 3). No further media changes are needed until the end of assay at day 18.
7. By day 14 of differentiation, MYH6 levels should already be elevated and should be visible as a fluorescent signal when using a reporter cell line. We however keep the cultures for 4 more days to allow a further increase of the MYH6 signal, which results in an improved dynamic range for quantification.
8. At day 18 the assay has completed and the plates are processed for imaging (*see* Subheading 3.5).

3.5 Plate Processing (and Immunostaining if Needed)

When the assay is completed, process the plates for imaging. There are several ways to handle the plates depending on the type of high content microscopes available and depending on whether a reporter line was used or not.

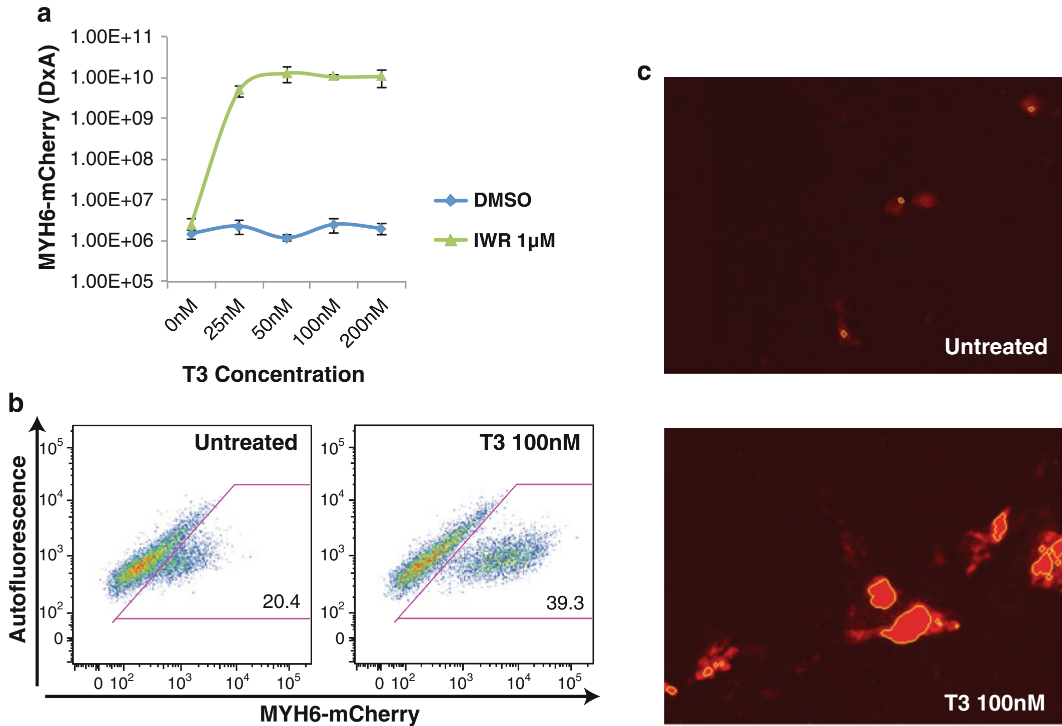


Fig. 3 Triiodothyronine (T3) boosts MYH6 expression. The synthetic analog T3 boosts the signal of a MYH6 reporter, but only when cardiomyocyte fate is induced by IWR, as assayed by high content imaging (**a**). Flow cytometry analysis confirms enhanced MYH6 expression per cell (**b**). Use of T3 enhances image analysis, as the signal intensity is boosted over tenfold (**c**), shown using the identical image settings for comparison

3.5.1 When Reporter Lines Were Used

1. In the preferred case of a reporter cell line, such as MYH6-mCherry, PGK1-H2B-GFP hESC, which gives a cytoplasmic red and nuclear green fluorescence, the cells can be imaged either live or following fixation. For live imaging, remove the SFM medium from the plate and replace with 25 μL of 1× PBS. Several high content imaging microscopes such as the Celigo can image whole plates very quickly (30–45 min), thus allowing live imaging of numerous plates in a short time frame (*see* **Note 33**).
2. However, when a larger screen of ten plates or more is performed on advanced high content microscopes, remove the SFM medium and fix the cells by adding 25 μL of a 4 % paraformaldehyde (PFA) solution to the wells for 15 min (*see* **Notes 33** and **34**). Remove the PFA and wash the cells three times with 1× PBS. After washing, add 50 μL of 1× PBS to the wells. (For longer term storage 50 % glycerol can be used instead of 1× PBS) (*see* **Note 35**).

3.5.2 If No Reporter Lines Were Used

1. Fix the cells as in Subheading 3.5.1.
2. After fixation, the cell membrane is permeabilized with the detergent Triton X-100 and the cells are blocked to prevent nonspecific antibody binding. To achieve both in one step, add 25 μL of blocking buffer to each well, for 1 h at room temperature, shaking the plates while incubating.
3. Wash once with 25 μL of 1× PBS.
4. Incubate the cells with primary antibody, overnight at 4 °C while shaking. Add 25 μL of antibody staining buffer per well, using a 1–100 dilution for the MF20 antibody or 1–500 for α-Actinin.
5. The next day, wash the wells three times in 1× PBS.
6. Incubate the cells with secondary antibody and a nuclear stain for 90 min at room temperature while shaking. Stain the cells as follows: add 25 μL of antibody staining buffer per well including 4 μg/mL of secondary antibody labeled with Alexa 488 or 568 and 1 μg/mL of DAPI.
7. Perform three washes with 25 μL 1× PBS.
8. Add 50 μL of 1× PBS to the plates (alternatively, 50 % glycerol can be used when storing the plates) (*see* **Note 35**).
9. Plates are now ready to be imaged. Examples of reporter and immunostaining are shown in Fig. 4.

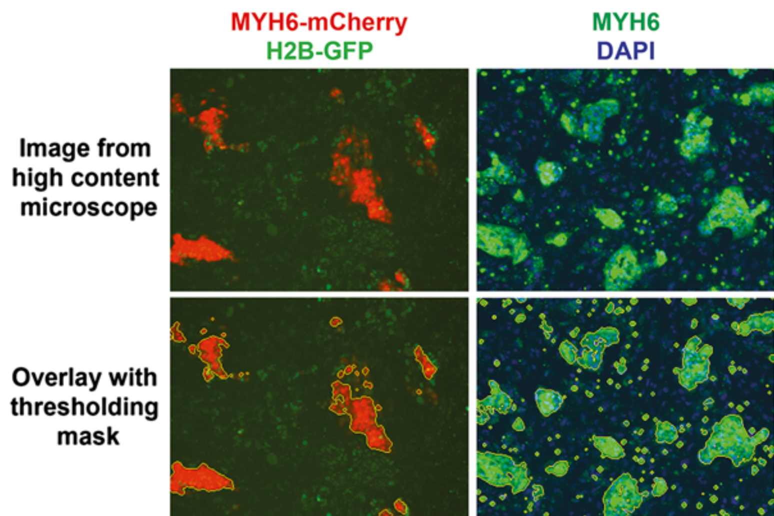


Fig. 4 Image examples of the assay readout. Cells engineered to express a fluorescent reporter can be imaged directly (*left panels*), relying on the red (cardiac MYH6-mCherry) and green (nuclear H2B-GFP) fluorescent signals. Alternatively, immunostaining for MYH6 with an Alexa 488-conjugated secondary antibody can be used to visualize the cardiomyocytes (*right panels*). Counterstaining with DAPI indicates nuclei. A thresholding algorithm demarcates areas of fluorescent cardiac cells (the mask boundary is outlined in *yellow*) in both cases

3.6 Imaging and Data Analysis

To quantify cardiac induction, we rely on a high content (HC) imaging approach to image as much of the well surface as possible. The dynamic range of an HC screening approach can be considerably greater than that of a plate reader assay [12, 15]. We here describe the typical flow using an InCell 1000 instrument, yet it is similar for other HC imaging instruments. Images are typically acquired in three color channels (red, green, blue) using filter sets appropriate for the fluorochromes (*see* **Note 33**).

1. Using a 10× objective (numerical aperture = 0.45), 9 fields/well are acquired to maximize imaging of the whole cell surface (*see* **Note 36**).
2. During acquisition, pixels are typically binned at 4 × 4 to reduce the file size (*see* **Note 37**).
3. Once images are collected we run a simple thresholding algorithm to quantify MYH6 expression and nuclear expression. As illustrated in Fig. 4, the MYH6 expression intensity is masked by selecting a certain threshold, and displaying a signal intensity above the background [16] (*see* **Note 38**).
4. After running the algorithm, different parameters can be collected for converting MYH6 expression into a numerical output. We typically report the overall area of the signal captured by the mask to estimate the number of cardiomyocytes formed. We also use the total integrated intensity of the reporter signal within the mask as an estimate of the level of expression per cell. We then multiply both numbers to generate a data output that reflects both the number of positive cells and the expression per cell (typical data are shown in Figs. 2 and 3a). Note that the values only estimate cell number and expression as cardiomyocytes typically grow in tight three-dimensional clusters (*see* **Notes 39** and **40**).
5. Independently, a nuclear count algorithm (also based on thresholding) on the H2B-GFP or DAPI images is run to estimate toxicity of compounds. Reduced levels of nuclei typically indicate toxic effects of the compounds, which aids in discerning toxic compounds from inhibitors when needed.

3.7 Notes

1. We initially developed this assay by maintaining hESC/hIPSC on Matrigel and MEF feeder layers. While the cells plated on gelatin for differentiation still require MEFs (*see* Subheading 3.2, **steps 1–3**), cells for routine maintenance can be grown feeder free on Matrigel using mTeSR medium or TeSR-E8 (Stem Cell Technologies).
2. These concentrations may vary from source to source (we use proteins from R&D Systems), and batch to batch and cell line to cell line and need to be titrated. Even though we have not seen much Activin A or Bmp4 batch-to-batch variation with

our cell lines, each new lot should be carefully titrated. We recommend running an Activin A/Bmp4 array type experiment with twofold doses around our recommended dosages. The EB do not need to be dissociated at day 4 for this purpose (*see* **Note 15**).

3. When we developed the assay, we found that the MYH6-reporter was not bright enough for automated imaging analysis, due to the fact that the cells grew as a single layer of cardiomyocytes in the dish. While we could see weak fluorescence coinciding with contraction by eye, the MYH6-mCherry intensity was not sufficient to quantitatively distinguish signal from background (Fig. 3c top panel). We found that addition of Triiodothyronine (T3), a synthetic thyroid hormone analog selectively boosts the MYH6 signal (Fig. 3a), only when the promoter is active (Fig. 3a) [17]. Flow cytometry clearly demonstrates that T3 causes an increase in the MYH6-mCherry signal, moreover allowing the detection of more cardiomyocytes (Fig. 3b). The impact of T3 for imaging yields a dramatically increased signal to background ratio so that specific signal can be readily detected by a thresholding algorithm (Fig. 3c).
4. Aside from small molecules, this assay is also suitable for small RNA screening. Various collections of siRNAs are available from different vendors and range from pathway specific to genome wide panels. siRNA transfection per well is performed by growing the cells in 65 μL of hESC differentiation medium, supplemented with 0.1 μL of RNAiMax, siRNAs at a desired concentration in water up to 5 μL and Opti-MEM added to complement the siRNAs up to 10 μL. After 24 h, the medium can be replaced with 75 μL of hPSC differentiation medium.
5. Other vendors provide similar antibodies that may replace the ones indicated. Careful titrations would be required when alternate options are used as the concentrations listed here are based on the antibodies listed.
6. The use of a reporter line has multiple benefits for the high content screening process: (a) assay development is facilitated as reporter expression can be followed in real time. (b) Assay controls are immediately visible before processing the plates as described. (c) Antibody-based read outs are costly and can be prohibitive when running larger screens.
7. hESC/hIPSC cultures in our hands are less consistent than mouse PSC, and therefore similar assays for mouse PSC have historically allowed for larger scale screens than the hESC/hIPSC assays. Our experience suggests that hits identified in either assay can be translated to the other. To perform a mouse assay, EB from a Myh6-GFP reporter line are formed in SFM by plating 50,000 cells per mL in non-coated dishes.

The assay similarly relies on the exposure to Activin A and Bmp4 from day 2 to day 4, after which the cells are plated into 384 wells at 8,000 cells per well. IWR then is a key switch to turn on cardiac fate at day 5, with cardiac induction seen by GFP expression at day 7.

8. Alternatively MEFs can be seeded simultaneously with hESC/hIPSC, but pre-plating of MEFs is recommended.
9. We typically freeze one confluent well of a 6-well plate into one 2 mL cryovial. Since recovery after thawing is never 100 %, we thaw one vial to one well of a 6-well plate.
10. We typically passage 1–6, meaning one confluent well of a 6-well plate is sufficient for 6 wells of a 6-well plate.
11. EB formation from intact colonies is much more efficient than from pieces of cut up colonies. To facilitate removal of colonies, we therefore switch Matrigel for a solution of 0.1 % gelatin and a low amount of Matrigel to coat plates, providing just enough attachment to maintain pluripotency.
12. Alternatively, plates can be generated the same day at least 2 h before seeding MEFs, again placing the coated plate at 37 °C. Overnight coating is however recommended.
13. To facilitate drying of the plates, we found that leaving plates open in a biosafety cabinet for about 1 h is sufficient. Drying is important, as we noted that hESC/hIPSC colonies tend to peel off after 48 h if the plates were not fully dried.
14. We use MEFs to preserve the pluripotency and compact nature of the PSC colonies for differentiation. We tend to get better differentiation results compared to hESC/hIPSC maintained on Matrigel in MEF conditioned hPSC growth medium or mTeSr.
15. Colonies should lift off easily. Otherwise the plate can be tapped gently or agitated to facilitate colony lifting. Manual dislodging may also help. We do not recommend incubation for longer times as we have found that longer exposure affects survival of EB. If needed, slightly increase collagenase IV concentration, but avoid concentrations above 2 mg/mL.
16. Gravity pelleting of colonies or EBs selectively pellets viable colonies/EB and avoids dead cells, which would be present if colonies were collected by centrifugation. The media containing the dead cells should however be aspirated immediately after the colonies or EB have settled at the bottom of the tube. Prolonged times of gravity pelleting will also allow dead cells to pellet.
17. To verify that EB differentiation progressed normally from day 0 to 4 or to optimize day 1–4, a differentiation control can be used. Day 4 EB are plated on 0.1 % gelatin coated dishes in

hPSC differentiation medium in the presence of 1 μM IWR and 5 ng/mL bFGF. Cardiac induction can be monitored by beating, reporter expression, immunostaining, RT-qPCR for cardiac markers such as TNNT2 or MYH6 or by flow cytometry with a SIRPA antibody [6, 18].

18. We prefer to use 15 mL conical Falcon tubes, but 50 mL conical tubes are also suitable. Note that the volume should not be increased above 5 mL (it is important to maintain the EB to TrypLE ratio), as this will affect the efficiency of dissociation with a 1 mL micropipet tip.

19. We do not recommend the use of serological pipets for dissociation, as the opening is too wide for efficient disruption of EB to single cells.

20. EB are large clumps that are clearly discernable by eye. As dissociation takes place, these clumps will dissolve and should disappear. Manual resuspension is essential to remove final clumps. If clumps or strands of cells are observed after this process, do not continue dissociation, but rather continue the process. Chunks will be removed subsequently using the cell strainer.

21. For cell count consistency between screens, we use an automated cell counter such as the Invitrogen Countess or the Bio-Rad TC20. Both count cells and determine their viability using trypan blue. Alternatively a hemocytometer can be used.

22. Cell preparations with lower viability may still differentiate properly, but in our hands, under these conditions the risk of failure is too high to warrant continuation of the screening efforts. We do use these lower viability preparations for smaller secondary assays.

23. We use automatic 16-channel pipets (such as the Matrix series from Thermo Scientific) as they allow repeated accurate dispensing of volumes down to 2 μL. Manual alternatives exist, but use of such pipets will increase row-to-row or column-to-column variation and we do not recommend these for screening purposes.

24. Cell number can be increased to ensure proper differentiation, and the number should be re-titrated when developing this assay, as counting can vary from counter to counter. If viability is lower than 90 % it is possible to increase cell number per well, but this is only recommended when running small-scale hit confirmation experiments.

25. For compound addition, we use an acoustical spotter (Echo 550, Labcyte Inc), which dispenses nanoliter volumes of compound directly into the wells of a 384-well plate, which already contain cells and medium in most cases. An alternative for nanoliter transfers are pintools, which also directly transfer

small molecules into the wells containing medium and cells. Both the acoustic spotter and pintool can also be used to spot the compounds in the gelatin coated 384-well plates before the cells and medium are added. Pintools and acoustic spotters are less accessible and an alternative approach using liquid split as hand-lers may be used to add compounds diluted in hESC differentiation medium, transferring the compounds in a volume of 2–5 μL to the cells. To achieve effective working concentrations, the compound libraries should be sub-diluted into hPSC differentiation medium to a 20–35× concentrated working solution.

26. Allow cells to settle to the bottom of the plate by gravity at room temperature. Centrifugation or incubation at 37 °C will give a less uniform distribution, and an even distribution is essential for optimal cardiac differentiation down the line.

27. The assay can be run in agonist mode to identify inducers of cardiac fate or in antagonist mode to identify molecules that block cardiac differentiation. In agonist mode, day 4 EB cells are not induced to form cardiomyocytes, and the added small molecules or siRNAs are expected to promote cardiac differentiation. In antagonist mode, however, the day 4 cells are exposed to 1 μM of IWR, which will induce cardiac differentiation efficiently. Small molecules or siRNAs added can then be used to block cardiac differentiation to probe the signaling or genetic cascades involved.

28. Positive control compounds to induce cardiac differentiation at day 4 include Wnt inhibitors [6] and additional treatment with the Nodal/TGFβ inhibitor SB-431542 [19].

29. The choice of libraries is an important upfront decision. We have screened collections of relatively uncharacterized molecules selected for chemical diversity (for example the Chembridge DiverSet collection). Such libraries are frequently used in target-based screens, which are designed to identify hits that elicit a particular biochemical activity (e.g., enzyme inhibition), often through a constrained molecular mechanism of action, against the target. These libraries are typically assembled based on their representation of chemical diversity space, comprising structures that have been deemed likely to engage the target of interest. However, since relatively little is known about the biological activities of these molecules, identification of the biological mechanism of action and the actual target typically requires a vast amount of work. Nevertheless, such collections are more likely to identify novel mechanisms and in our hands resulted in a completely novel TGFβ inhibitor [7, 9].

    An alternative approach to gain biological insight is to screen focused collections of small molecules that selectively

engage known sets of cellular proteins. We typically use StemSelect and Inhibitor Select collections from EMD/Millipore or the Lopac 1280 collection from Sigma. Although small molecules are rarely selective for a unique protein, following the known protein targets of screen hits in this case has proven an effective strategy for identifying cellular processes that control complex biological phenomena. For example, the use of small molecule pathway modulators facilitated the discovery of biological mechanisms that drive IPSC generation of cardiac differentiation [6, 20].

30. The number of replicates necessary to discern hits should be determined during assay development, and a useful discussion is in the reference by Zhang et al. [21]. We screen in triplicate to ascertain hits since dynamic range of the cardiac differentiation protocol is typically insufficient to screen without replicates.
31. Place extra water pans in the incubator to ensure a well-humidified environment to limit evaporation of media from the wells.
32. StemPro 34 supports differentiation of hESC/hIPSC to cardiomyocytes, but our experience is that the cardiomyocytes are less viable. Evaluating different media, we found that use of certain serum-free options, including Knock Out Serum replacer or B27/N2 supplement based media, yield healthy cardiomyocytes. Serum-free media also prevents fibroblast overgrowth, thus enhancing signal to background levels (Fig. 2).
33. Currently, there are many different high content imagers that can be used for reading out this assay. The relatively inexpensive Celigo platform images at lower resolution enabling quick whole well imaging with sufficient quality for image analysis. Other options include the InCell series from GE Healthcare or the Opera/Operetta series from Perkin Elmer, allowing higher resolution images at different magnifications. For quantifying the cytoplasmic fluorescent signal as described in this protocol, the Celigo, InCell, and Opera platforms provide similar signal to background and dynamic range.
34. Longer fixations are not recommended to avoid loss of signal from fluorescent proteins and increased background.
35. Alternatively 50 μL of 50 % glycerol in PBS can be added to the plates to preserve the fluorescent signal for several weeks.
36. Partial field or higher magnification acquisition works, but we prefer to capture as much of the well as possible as cardiac differentiation typically occurs in an unpredictable and unevenly distributed pattern.
37. This reduces resolution of the images, and thus overall quantity of data, and consequently speeds up processing time. It does

not affect quantification of the cytoplasmic stain. However, binning is not appropriate for higher resolution imaging.

38. We use a custom-built algorithm in the imaging package Cyteseer (Vala Sciences Inc.) to generate masks. The algorithms permit manual determination of the threshold to be used (based on principles described by Bushway et al. [16]). Most current HCS instruments include an image analysis package that can run similar thresholding operations.
39. Secondary assays using proper quantification methods such as flow cytometry should be used to quantify expression level and incidence of positive cells.
40. An alternative output to estimate cardiomyocyte yield is the count of MYH6 positive objects. We typically do not use this metric, instead prefer secondary assays that directly quantify incidence (*see* **Note 39**).

Acknowledgements

This work was supported by grants from the California Institute for Regenerative Medicine (CIRM) (RC00132) and NIH (HL113601, HL108176, and AG043917) to M.M. E.W. was supported by American Heart Association and CIRM training grants.

References

1. Murry CE, Keller G (2008) Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell 132:661–680
2. Mercola M, Colas A, Willems E (2013) Induced pluripotent stem cells in cardiovascular drug discovery. Circ Res 112:534–548
3. Bruneau BG (2008) The developmental genetics of congenital heart disease. Nature 451:943–948
4. Burridge PW et al (2012) Production of de novo cardiomyocytes: human pluripotent stem cell differentiation and direct reprogramming. Cell Stem Cell 10:16–28
5. Garbern JC, Lee RT (2013) Cardiac stem cell therapy and the promise of heart regeneration. Cell Stem Cell 12:689–698
6. Willems E et al (2011) Small-molecule inhibitors of the Wnt pathway potently promote cardiomyocytes from human embryonic stem cell-derived mesoderm. Circ Res 109: 360–364
7. Willems E et al (2012) Small molecule-mediated TGF-β type II receptor degradation promotes cardiomyogenesis in embryonic stem cells. Cell Stem Cell 11:242–252
8. Lanier M et al (2012) Wnt inhibition correlates with human embryonic stem cell cardiomyogenesis: a structure–activity relationship study based on inhibitors for the Wnt response. J Med Chem 55:697–708
9. Schade D et al (2012) Synthesis and SAR of b-annulated 1,4-dihydropyridines define cardiomyogenic compounds as novel inhibitors of TGFβ signaling. J Med Chem 55:9946–9957
10. Borowiak M et al (2009) Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. Cell Stem Cell 4:348–358
11. Chen S et al (2009) A small molecule that directs differentiation of human ESCs into the pancreatic lineage. Nat Chem Biol 5:258–265
12. Willems E, Bushway PJ, Mercola M (2009) Natural and synthetic regulators of embryonic stem cell cardiogenesis. Pediatr Cardiol 30:635–642
13. Kita-Matsuo H et al (2009) Lentiviral vectors and protocols for creation of stable hESC lines

for fluorescent tracking and drug resistance selection of cardiomyocytes. PLoS One 4:e5046

14. Yang L et al (2008) Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. Nature 453: 524–528
15. Bushway PJ, Mercola M, Price JH (2008) A comparative analysis of standard microtiter plate reading versus imaging in cellular assays. Assay Drug Dev Technol 6:557–567
16. Bushway P, Mercola M (2006) High-throughput screening for modulators of stem cell differentiation. Methods Enzymol 414: 300–316
17. Gustafson TA et al (1987) Thyroid hormone regulates expression of a transfected alpha-myosin heavy-chain fusion gene in fetal heart cells. Proc Natl Acad Sci U S A 84:3122–3126
18. Dubois NC et al (2011) SIRPA is a specific cell-surface marker for isolating cardiomyocytes derived from human pluripotent stem cells. Nat Biotechnol 29:1011–1018
19. Cai W et al (2013) Coordinate Nodal and BMP inhibition directs Baf60c-dependent cardiomyocyte commitment. Genes Dev 27:2332–2344
20. Li Z, Rana TM (2012) A kinase inhibitor screen identifies small-molecule enhancers of reprogramming and iPS cell generation. Nat Commun 3:1085
21. Zhang J, Chung T, Oldenburg K (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen 4:67–73

Chapter 5

Small-Molecule High-Throughput Screening Utilizing *Xenopus* Egg Extract

Matthew R. Broadus, P. Renee Yew, Stephen R. Hann, and Ethan Lee

Abstract

Screens for small-molecule modulators of biological pathways typically utilize cultured cell lines, purified proteins, or, recently, model organisms (e.g., zebrafish, *Drosophila*, *C. elegans*). Herein, we describe a method for using *Xenopus laevis* egg extract, a biologically active and highly tractable cell-free system that recapitulates a legion of complex chemical reactions found in intact cells. Specifically, we focus on the use of a luciferase-based fusion system to identify small-molecule modulators that affect protein turnover.

Key words *Xenopus* egg extract, *Xenopus laevis*, Cell-free, Small molecules, High-throughput screening, Protein turnover, Protein degradation

1 Introduction

Traditionally, small-molecule screening to identify potential therapeutic leads and/or biological tools have been performed using in vitro (purified components) or in vivo (cultured cells/whole organism) approaches. Each approach has its own strengths and weaknesses. The use of purified proteins simplifies the process considerably because one is sampling only molecules that directly bind and alter the activity of the protein being targeted. The major weakness of this approach, however, is that the biological consequences of inhibition/activation by the small molecule at the organismal level are less clear. Screening for phenotypic changes using cultured cells or whole organisms is obviously more biologically relevant, although manipulations are more complex. Lack of effects may be due to failure of compounds to pass through the plasma membrane, expulsion via efflux pumps, or cell death. Additionally, target identification remains a major hurdle.

The *Xenopus laevis* egg extract system overcomes some of the limitations of using purified proteins or cells/organisms for small-molecule screening by providing a cell-free, yet robust, biologically active system that can be readily manipulated [1].

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_5, © Springer Science+Business Media New York 2015

Because *Xenopus* egg extract lacks intact plasma membranes, small molecules are allowed unfettered access to putative targets. In addition, *Xenopus* egg extract contains all of the eukaryotic cellular machinery and complex signaling pathways required for the early development of an organism. Finally, large amounts of homogenous *Xenopus* egg extract can be prepared at one time, an important consideration for large-scale screens and reproducibility [1–3].

Xenopus egg extract is a homogenous mixture of cellular components including cytoplasmic proteins, cellular organelles, amino acids, and nucleotides at near physiological levels [4]. This system has been used to answer numerous biological questions regarding the cell cycle, cytoskeletal dynamics, signal transduction, apoptosis, nuclear assembly, nucleocytoplasmic transport, ubiquitin metabolism, and protein turnover [5–32]. While the versatility of the *Xenopus* egg extract system in reconstituting a large number of complex biological reactions is a major strength for small-molecule screening, different methodologies for extract preparation must be used to optimize the system for a particular pathway or biological event. Thus, the preparation methodology of *Xenopus* egg extract is a major consideration in performing a high-throughput screen to ensure that one has the best chance of identifying useful small molecules. Additional methods for *Xenopus* egg extract preparation have been described elsewhere [6, 10, 13, 25, 32–36].

Xenopus egg extract is a particularly robust system for studying protein turnover that lacks the potentially confounding influence of gene transcription. The method of *Xenopus* egg extract preparation described within this chapter is optimized for analyzing protein turnover of β-catenin, the key effector protein of the Wnt signaling pathway; also, we found that it supports the degradation of another Wnt component, Axin, as well as other signaling pathway proteins that are known to rapidly turn over [5, 37, 38]. The usefulness of *Xenopus* egg extract for studying key aspects of cytoplasmic Wnt pathway regulation is supported by multiple studies that identify important regulatory proteins/steps that contribute to β-catenin degradation [2, 3, 5, 38–44]. Significantly, the preparation of *Xenopus* egg extract described herein was successfully used to screen and identify small molecules that stimulate β-catenin turnover and inhibit Wnt signaling [2, 3].

In this chapter we provide a detailed method for using *Xenopus* egg extract preparations that are optimized for examining protein turnover. We take advantage of firefly luciferase (Luciferase), a protein normally stable in *Xenopus* egg extract that, when fused to proteins of interest, provides a simple and rapid readout of protein turnover. We describe herein how these Luciferase fusion proteins can be used to perform high-throughput (HTS) biochemical screens in *Xenopus* egg extract to identify biologically active small-molecule compounds.

2 Materials

2.1 Xenopus Egg Extract Preparation

1. 100 U/mL pregnant mare serum gonadotropin (PMSG): Stock is prepared fresh before injections by dilution of 1,000 units (U) of PMSG in 10 mL of purified deionized water.
2. Storage water: 40 L of 20 mM sodium chloride. Weigh out 46.72 g of sodium chloride into 40 L of deionized water.
3. 20× stock Marc's Modified Ringers (MMR): 100 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 40 mM potassium chloride, 2 M sodium chloride, 20 mM magnesium chloride, and 40 mM calcium chloride, pH 7.4. Weigh out 35.7 g of HEPES, 4.5 g of potassium chloride, 175.2 g of sodium chloride, 2.9 g of magnesium chloride, and 6.7 g of calcium chloride. Mix these into a total volume of 1.25 L purified deionized water. Once all is dissolved, adjust the pH of the solution to 7.4 with NaOH and fill to a final volume of 1.5 L with deionized water.
4. 750 U/mL human chorionic gonadotropin (HCG): HCG is prepared fresh before injections by dilution of 10,000 U of HCG in 13.3 mL of purified deionized water.
5. 2 % (w/v) cysteine solution: 8 g of cysteine is diluted into 400 mL of deionized water, and pH is adjusted to 7.7 with NaOH.
6. Leupeptin, pepstatin, aprotinin (LPA): 10 mg/mL leupeptin, 10 mg/mL pepstatin, and 10 mg/mL aprotinin. Dissolve 10 mg of leupeptin, 10 mg of pepstatin, and 10 mg of aprotinin in 1 mL of dimethyl sulfoxide (DMSO).
7. 10 mg/mL cytochalasin D: Dilute 10 mg of cytochalasin D into 1 mL of DMSO.
8. 10 mg/mL cycloheximide: Dilute 10 mg of cycloheximide into 1 mL of purified deionized water.

2.2 Luciferase-Fused Proteins

1. Several in vitro-transcription/translation kits are commercially available. We typically use a rabbit reticulocyte system in which the cDNA of interest is subcloned into the pCS2+ plasmid with transcription driven by the SP6 promoter.

2.3 Active Extract

1. 20× energy reaction (ER) mix: 20 mM adenosine triphosphate, 150 mM creatine phosphate, 20 mM magnesium chloride, and 600 μg/mL creatine phosphokinase. Weigh out 10.1 mg of adenosine triphosphate, 31.7 mg of creatine phosphate, 1.7 mg of magnesium chloride, and 600 μg creatine phosphokinase. Mix these into a total volume of 1 mL of purified deionized water. Divide ER mix into 50 μL aliquots and store at −80 °C until needed.

2.4 Z-Factor Scoring and Screening

1. White 96-well plate (*see* **Note 1**).
2. Small-molecule library of choice.
3. Luciferin reagent/commercially available kit to assess luciferase activity.

3 Methods

3.1 Preparation of Xenopus Egg Extract for Screening of Protein Turnover

As described above, this purification method is optimized for analyzing β-catenin protein turnover. The method described is for preparing extract from ten frogs. For larger or smaller preparations, the amount of buffer should be adjusted accordingly. Typically, each frog yields ~1 mL of extract with a protein concentration of ~50 mg/mL.

1. To induce frog egg production, female frogs are primed with 100 U of PMSG injected subcutaneously into the dorsal lymph sac using a 3 mL tuberculin syringe and 27 G needle.
2. Primed frogs are stored in 4 L of 20 mM NaCl at 18 °C for 5–10 days (*see* **Note 2**).
3. Prepare a 0.5× MMR solution to be used in the next step. This is performed by diluting a 20× MMR stock to make 40 L of a 0.5× MMR solution. Set up ten 4 L buckets (*see* **Note 3**). These buckets should be prepared and kept in a 16 °C incubator overnight prior to injecting frogs with HCG.
4. After 5–10-day incubation, inject the dorsal lymph sac of each primed frog using a 3 mL tuberculin syringe and 27 G needle with 750 U HCG. Each injected frog should be placed in a bucket containing 4 L of 0.5× MMR at 16 °C.
5. Allow the frogs to lay eggs for 15–16 h at 16 °C.
6. A day prior to injections, dilute 20× MMR to 4 L of a 1× MMR solution and 50 mL of a 0.1× MMR solution. The morning of the egg extract prep, prepare fresh 400 mL 2 % cysteine solution, pH 7. These solutions should be stored at 16 °C (*see* **Note 4**).
7. After the 15–16-h egg laying period, gently squeeze the abdomen and lower back of each frog to expel additional eggs, and place the frogs into a separate container of deionized water.
8. Remove the majority of the MMR from each bucket, leaving the eggs in the smallest volume possible. Make sure, however, that the eggs remain covered in MMR.
9. Remove poor-quality eggs with a plastic transfer pipet as these will decrease the quality of the overall extract. If greater than 10 % of the eggs appear poor in quality, the entire batch should be thrown away (*see* **Note 5**).

10. Combine cleared, high-quality eggs and remaining MMR in a 500 mL glass beaker.
11. Again, pour out the majority of the MMR keeping the eggs submerged.
12. Estimate volume of egg bed, and wash eggs by carefully adding twice the volume of 1× MMR along the inside of the beaker. Gently swirl the eggs and pour off debris and the majority of the MMR. Repeat this twice and continue to remove any poor-quality eggs.
13. To de-jelly the eggs, pour 100 mL of 2 % cysteine along the inside of the beaker. Swirl the beaker gently to mix and allow the eggs to settle at 16 °C for 5 min. Pour off the majority of the cysteine, keeping the eggs submerged. Repeat until the eggs appear tightly packed (*see* **Note 6**).
14. Wash off the cysteine by adding 1× MMR along the inside of the beaker, gently swirl, and again, pour off most of the solution. Repeat until the 1× MMR solution is no longer cloudy. Poor-quality eggs should be continually removed during this process.
15. Gently rinse the eggs with 30 mL of 0.1× MMR, and pour off the majority of solution.
16. Add LPA (10 μg/mL final) and cytochalasin D (20 μg/mL final) to the remaining 20 mL of 0.1× MMR.
17. Add the 0.1× MMR solution containing LPA and cytochalasin D to the washed eggs, swirl gently, and incubate at 16 °C for 5 min.
18. Transfer the eggs into prechilled 50 mL centrifuge tubes using a 25 mL pipet (*see* **Note 7**). After the eggs have settled, the excess buffer should be removed. The eggs should remain covered with buffer.
19. Pack the eggs by centrifugation at 400 × *g* for 60 s at 4 °C using a fixed-angle rotor.
20. Remove any excess buffer from the packed eggs.
21. Crush the eggs by spinning tubes at 15,000 × *g* for 5 min at 4 °C.
22. The egg extract will now be separated into three layers. The bottom and darkest layer contains yolk, pigmented granules, etc.; the middle layer contains cytoplasmic fraction (the desired material), and the top layer contains lipid-enriched material. In order to collect the cytoplasmic layer, the lipid layer must first be disrupted, which can be accomplished by piercing the lipid layer with a P1000 pipet tip so as to create a hole.
23. Using a new P1000 pipet tip, collect the cytoplasmic layer (straw-colored middle layer) through the hole in the lipid layer and transfer the collected cytoplasm to a new centrifuge tube on ice (*see* **Note 8**).

24. Spin the collected cytoplasmic layer at 15,000 × *g* for 10 min at 4 °C (*see* **Note 9**).
25. Again, collect the cytoplasmic layer into a new prechilled centrifuge tube and add LPA, cycloheximide, and cytochalasin D to final concentrations of 10 μg/mL each.
26. Dispense the extract into 100–1,000 μL aliquots and snap-freeze in liquid nitrogen for storage (*see* **Notes 10** and **11**).

3.2 Preparation of Recombinant In Vitro-Transcribed/Translated Luciferase-Fusion Proteins

It is important to be able to readily produce sufficient amounts of recombinant Luciferase-fused protein(s) in order to perform high-throughput screening. We have found that recombinant protein production by in vitro-transcription/translation (IVT), bacterial expression, or the *Sf9*/baculovirus systems all work well. Protein production by IVT is the quickest and easiest of the three, although the limited protein yield can be an issue. A much greater protein yield can be obtained using the bacterial or *Sf9*/baculovirus systems, but these are much more labor intensive.

1. Produce IVT protein(s) using commercially available kits (*see* **Note 12**).
2. Luminescence activity at this point should be assessed by measuring a small sample (typically 1 μL of IVT protein). This can be performed by mixing 1 μL of protein and 25 μL of luciferin reagent in a 96-well white plate. The IVT protein is then aliquoted and stored at −80 °C until used. The size of protein aliquots is determined based on considerations in Subheading 3.3 and **Note 13**.

3.3 Assessing Z-Factor Score

Assessment of a screen's *Z*-factor is important to ensure its usefulness and/or probability of finding small molecules. Thus, at this point it is important to optimize the screen in order to obtain the most effective *Z*-factor score and ensure the best chance of reliably identifying biologically active small molecules. The method described is for a 96-well plate. With appropriate scaling, however, this protocol can be modified for a 384-well format or alternative well formats.

1. Place a white 96-well plate on ice to cool. This step is important to inhibit the degradation reaction until setup is complete.
2. Quickly thaw frozen aliquots of *Xenopus* egg extract and 20× ER mix by rubbing the tubes between one's hands or gently swirling in a 30 °C water bath until only a small amount of frozen extract remains. Place the extract on ice. Add the ER mix (1× final) to the extract and mix by brief vortex pulses to generate the reaction mix. Place the reaction mix on ice.
3. Quickly thaw frozen recombinant Luciferase-fusion protein by rubbing the tubes between one's hands until only a small amount of frozen protein remains. Place the protein on ice.

4. Add the appropriate amount of luciferase-fusion protein to the reaction mix such that the relative luminescence units (RLU) will be approximately 10,000 RLU/μL (*see* **Note 13**).
5. Dispense 10 μL of the reaction mix plus Luciferase-fusion protein into each of the 96 wells on ice.
6. For *Z*-factor scoring, load vehicle and control in alternating wells, mimicking a checkerboard design. We typically use DMSO (vehicle, negative control) and MG132 (proteasome inhibitor, positive control) when screening proteins that are degraded in a proteasome-dependent manner. Compounds are added at ~500 μM (0.5 μL of each compound from 10 mM stocks) to respective wells. If using DMSO as vehicle, add an equal volume of the positive control (*see* **Note 14**).
7. Mix the plate by lightly shaking either by hand or vortexing at low speed, being careful not to eject liquid from the wells.
8. Incubate the plate at room temperature for a predetermined optimal period of time (*see* **Note 15**).
9. Stop the reaction by addition of 75 μL of luciferin reagent to each well.
10. Mix the plate by lightly shaking either by hand or vortexing at low speed, again being careful not to eject liquid from the wells.
11. Measure luciferase activity using a luminometer.
12. The *Z*-factor can be assessed by calculation as previously described [45]. A *Z*-factor score of 1 is ideal, a score between 1 and 0.5 indicates that the assay is excellent, a score between 0.5 and 0.0 indicates that the assay is weak, and a score ≤0 indicates that the assay is error prone and is, therefore, not reliable.

3.4 Screening for Small Molecules

1. After optimizing the screen for an effective *Z*-factor score, a small-molecule pilot screen can be performed under similar conditions. Perform the pilot screen with identical conditions used to achieve an optimized *Z*-factor score in Subheading 3.3. Load the same volume of compounds as was used for addition of controls when determining the *Z*-factor (*see* **Notes 14** and **16**). It is important to load both negative and positive controls (typically in triplicate) in order to assess the effectiveness of the screen. For proteasome-mediated degradation screens, we use DMSO and MG132 as negative and positive controls, respectively.
2. We have found that it is important to run a Luciferase-only control screen in order to identify compounds that directly inhibit/enhance Luciferase activity [46]. The enzymatic activity of the Luciferase protein requires ATP. Thus, it is possible that some compounds may alter the activity of the Luciferase protein by altering ATP levels.

3. Upon completion of the HTS screen, assess whether the screen ran optimally by comparing values of the negative and positive controls, which should reflect values that were observed when assessing the *Z*-factor for the screen.
4. Effective small molecules are those that increase or decrease the luminescence by >3 standard deviations. Screens should be repeated at least three times. Small molecules that repeatedly cause greater than a threefold change in standard deviation are likely to represent "true hits."

4 Notes

1. Plates with round- or flat-bottom wells work equally well.
2. It takes at least 5 days in order for priming to take full effect, and priming should last for 10 days. After 10 days the effect of priming is diminished, and a decreased amount of eggs are obtained.
3. The use of large-sized buckets containing multiple frogs increases the risk that a given frog might lay poor-quality eggs; in that case, a significant amount of time and effort will be required to separate poor-quality from high-quality eggs. This additional time increases the likelihood that high-quality eggs will lyse or otherwise degenerate. Thus, it is not worth the risk to use fewer tanks.
4. At this point, it is important to maintain the temperature at 16 °C throughout the remainder of the extract preparation. It is also important to work as rapidly as possible. As noted above, the longer the amount of time needed to process the eggs, the greater the likelihood of spontaneous egg lysis.
5. High-quality eggs will have a high dark-to-light contrast between the darkly pigmented animal hemisphere and the lightly colored vegetal hemisphere. Poor-quality eggs will appear stringy (immature eggs) or white and puffy (lysed eggs).
6. Eggs will become more compact as the jelly coat is removed, which will float above the eggs. Three cysteine treatments are usually required for full de-jellying to occur. Once the eggs have been de-jellied, they will become very fragile and prone to lyse, so it is important to swirl gently and ensure that eggs are not exposed to the air.
7. When pipetting, to prevent eggs from being exposed to air, first draw up some buffer before suctioning up the eggs.
8. For preparation of high-quality extract for β-catenin degradation, it is important to minimize the amount of lipid or pigmented layers transferred.

9. This clearing step may be repeated if the cytoplasmic layer still contains a significant amount of pigment or lipid material, which may result in proteolysis of β-catenin by non-Wnt pathways. Excessive spins, however, decrease the robustness of the extract to support β-catenin degradation mediated by Wnt components.

10. Do not freeze extract if you wish to maintain the translational capacity of the egg extract. Once frozen, *Xenopus* egg extract loses significant capacity to translate exogenously added mRNA. For more information, *see* [5, 47].

11. *Xenopus* egg extract, once prepared, is stable for long-term storage in liquid nitrogen. Extract can alternatively be stored at −80 °C; however, it should be used within 2 months.

12. To confirm that a protein of interest is produced using an IVT reaction, immunoblot analysis can be performed. Alternatively, proteins can be radiolabeled with [^{35}S]methionine and analyzed by SDS-PAGE/autoradiography.

13. We found that a readout of 10,000 RLU/μL provides a robust initial signal for monitoring changes in β-catenin protein turnover. It should be noted that the more dilute the extract, the less efficient the degradation reaction becomes. Thus, to maintain robustness of the degradation reaction, it is important to minimize the volume of reagents added to the extract. We found that diluting the volume of the extract more than 35 % significantly lowered the capacity of the extract to degrade β-catenin.

14. Because *Xenopus* egg extract is highly concentrated (~50 mg/mL), we find that small molecules need to be added in the μM range to be effective. Also, for small molecules dissolved in DMSO, it is important to add as little volume as possible. We found that adding more than 10 % DMSO will significantly inhibit the degradation reaction.

15. Different proteins will require different reaction times depending on the half-life of the protein. This is a key optimization step that should be properly assessed during *Z*-factor determination. The key is to identify the time in which the protein of interest has degraded by roughly half of its initial concentration. Working near this threshold will allow one to more readily identify small molecules that either inhibit or enhance degradation of the protein of interest. However, the bigger the difference/change between the positive and negative controls, the easier it is to achieve a robust *Z*-factor. Additionally, for longer incubation times it may be necessary to incubate the plate in a humidity chamber to prevent evaporation due to the small volume. This can be accomplished by placing damp paper towels within the bottom of a plastic box that can be closed

completely. The plate can then be set on the paper towels with the box lid closed during the incubation.

16. Arraying the small molecules themselves in a 96-well format significantly simplifies the transfer process.

Acknowledgments

We thank Laurie Lee for critical reading of the manuscript. M.R.B. is supported by a National Cancer Institute training grant (T32 CA119925). E.L. is supported by the National Institutes of Health (R01GM081635 and R01GM103926). R.Y. is supported by Award Number 8UL1TR000149 from the National Center for Advancing Translational Sciences and the CTRC P30 Cancer Center Support Grant from the National Cancer Institute (CA054174). S.R.H. is supported by the National Cancer Institute (P50 CA095103).

References

1. Hang BI et al (2012) Screening for small molecule inhibitors of embryonic pathways: sometimes you gotta crack a few eggs. Bioorg Med Chem 20:1869–1877
2. Thorne CA et al (2010) Small-molecule inhibition of Wnt signaling through activation of casein kinase 1alpha. Nat Chem Biol 6: 829–836
3. Thorne CA et al (2011) A biochemical screen for identification of small-molecule regulators of the Wnt pathway using Xenopus egg extracts. J Biomol Screen 16:995–1006
4. Ma L et al (1998) Corequirement of specific phosphoinositides and small GTP-binding protein Cdc42 in inducing actin assembly in Xenopus egg extracts. J Cell Biol 140: 1125–1136
5. Salic A, Lee E, Mayer L, Kirschner MW (2000) Control of beta-catenin stability: reconstitution of the cytoplasmic steps of the Wnt pathway in Xenopus egg extracts. Mol Cell 5:523–532
6. Murray AW (1991) Cell cycle extracts. Methods Cell Biol 36:581–605
7. Glotzer M, Murray AW, Kirschner MW (1991) Cyclin is degraded by the ubiquitin pathway. Nature 349:132–138
8. Tutter AV, Walter JC (2006) Chromosomal DNA replication in a soluble cell-free system derived from Xenopus eggs. Methods Mol Biol 322:121–137
9. Theriot JA et al (1994) Involvement of profilin in the actin-based motility of L. monocytogenes in cells and in cell-free extracts. Cell 76: 505–517
10. Maresca TJ, Heald R (2006) Methods for studying spindle assembly and chromosome condensation in Xenopus egg extracts. Methods Mol Biol 322:459–474
11. Shennan KI (2006) Xenopus egg extracts: a model system to study proprotein convertases. Methods Mol Biol 322:199–212
12. Kornbluth S, Yang J, Powers M (2006) Analysis of the cell cycle using Xenopus egg extracts. Curr Protoc Cell Biol. Chapter 11: Unit 11.11
13. Chan RC, Forbes DI (2006) In vitro study of nuclear assembly and nuclear import using Xenopus egg extracts. Methods Mol Biol 322:289–300
14. Dabauvalle MC et al (1991) Spontaneous assembly of pore complex-containing membranes ("annulate lamellae") in Xenopus egg extract in the absence of chromatin. J Cell Biol 112:1073–1082
15. Dabauvalle MC, Scheer U (1991) Assembly of nuclear pore complexes in Xenopus egg extract. Biol Cell 72:25–29
16. Masui Y, Markert CL (1971) Cytoplasmic control of nuclear behavior during meiotic maturation of frog oocytes. J Exp Zool 177:129–145
17. Forbes DJ, Kirschner MW, Newport JW (1983) Spontaneous formation of nucleus-like structures around bacteriophage DNA microinjected into Xenopus eggs. Cell 34:13–23

18. Lohka MJ, Masui Y (1983) Formation in vitro of sperm pronuclei and mitotic chromosomes induced by amphibian ooplasmic components. Science 220:719–721
19. Newport JW, Kirschner MW (1984) Regulation of the cell cycle during early Xenopus development. Cell 37:731–742
20. Mitchison T, Kirschner M (1984) Dynamic instability of microtubule growth. Nature 312:237–242
21. Blow JJ, Laskey RA (1986) Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of Xenopus eggs. Cell 47:577–587
22. Verma R et al (2004) Ubistatins inhibit proteasome-dependent degradation by binding the ubiquitin chain. Science 306:117–120
23. Yu H et al (1996) Identification of a novel ubiquitin-conjugating enzyme involved in mitotic cyclin degradation. Curr Biol 6: 455–466
24. Murray AW, Solomon MJ, Kirschner MW (1989) The role of cyclin synthesis and degradation in the control of maturation promoting factor activity. Nature 339:280–286
25. Murray AW, Kirschner MW (1989) Cyclin synthesis drives the early embryonic cell cycle. Nature 339:275–280
26. Lohka MJ, Hayes MK, Maller JL (1988) Purification of maturation-promoting factor, an intracellular regulator of early mitotic events. Proc Natl Acad Sci U S A 85:3009–3013
27. Pomerening JR, Kim SY, Ferrell JE Jr (2005) Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations. Cell 122:565–578
28. Lohka MJ, Masui Y (1984) Roles of cytosol and cytoplasmic particles in nuclear envelope assembly and sperm pronuclear formation in cell-free preparations from amphibian eggs. J Cell Biol 98:1222–1230
29. Finlay DR et al (1987) Inhibition of in vitro nuclear transport by a lectin that binds to nuclear pores. J Cell Biol 104:189–200
30. Newmeyer DD, Finlay DR, Forbes DJ (1986) In vitro transport of a fluorescent nuclear protein and exclusion of non-nuclear proteins. J Cell Biol 103:2091–2102
31. Newport J, Spann T (1987) Disassembly of the nucleus in mitotic extracts: membrane vesicularization, lamin disassembly, and chromosome condensation are independent processes. Cell 48:219–230
32. Newport J (1987) Nuclear reconstitution in vitro: stages of assembly around protein-free DNA. Cell 48:205–217
33. Cross MK, Powers M (2008) Obtaining eggs from Xenopus laevis females. J Vis Exp 18:e890
34. Cross MK, Powers M (2008) Preparation and fractionation of Xenopus laevis egg extracts. J Vis Exp 18:e891
35. Willis J et al (2012) Study of the DNA damage checkpoint using Xenopus egg extracts. J Vis Exp 69:e4449
36. Walter J, Sun L, Newport J (1998) Regulated chromosomal DNA replication in the absence of a nucleus. Mol Cell 1:519–529
37. Saito-Diaz K et al (2013) The way Wnt works: components and mechanism. Growth Factors 31:1–31
38. Lee E et al (2003) The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway. PLoS Biol 1:E10
39. Lee E, Salic A, Kirschner MW (2001) Physiological regulation of [beta]-catenin stability by Tcf3 and CK1epsilon. J Cell Biol 154:983–993
40. Seeling JM et al (1999) Regulation of beta-catenin signaling by the B56 subunit of protein phosphatase 2A. Science 283:2089–2091
41. Guger KA, Gumbiner BM (1995) Beta-Catenin has Wnt-like activity and mimics the Nieuwkoop signaling center in Xenopus dorsal-ventral patterning. Dev Biol 172:115–125
42. Cselenyi CS et al (2008) LRP6 transduces a canonical Wnt signal independently of Axin degradation by inhibiting GSK3's phosphorylation of beta-catenin. Proc Natl Acad Sci U S A 105:8032–8037
43. Jernigan KK et al (2010) Gbetagamma activates GSK3 to promote LRP6-mediated beta-catenin transcriptional activity. Sci Signal 3:ra37
44. Major MB et al (2007) Wilms tumor suppressor WTX negatively regulates WNT/beta-catenin signaling. Science 316:1043–1046
45. Zhang JH, Chung TD, Oldenburg KR (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen 4:67–73
46. Baldwin TO (1996) Firefly luciferase: the structure is known, but the mystery remains. Structure 4:223–228
47. Sive HL, Grainger RM, Harland RM (2000) Early development of Xenopus laevis: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Part II

In Vivo Chemical Genetic Screening

Chapter 6

Fission Yeast-Based High-Throughput Screens for PKA Pathway Inhibitors and Activators

Ana Santos de Medeiros, Grace Kwak, Jordan Vanderhooft, Sam Rivera, Rachel Gottlieb, and Charles S. Hoffman

Abstract

Features of the fission yeast *Schizosaccharomyces pombe* cAMP/PKA pathway make *S. pombe* particularly amenable for heterologous expression of cAMP pathway proteins such as $G\alpha_S$ subunits and their cognate adenylyl cyclases, PKA catalytic and regulatory subunits, and cyclic nucleotide phosphodiesterases. We have constructed two PKA-repressed reporters for use in high-throughput screens to detect compounds that elevate or reduce PKA activity, thus facilitating the discovery of both inhibitors and activators of these target proteins. Here, we describe steps to construct screening strains and to optimize and conduct these screens.

Key words Cyclic nucleotide phosphodiesterase, Adenylyl cyclase, PKA, $G\alpha_S$, Fission yeast, *Schizosaccharomyces pombe*, *fbp1*, High-throughput screen

1 Introduction

The fission yeast *Schizosaccharomyces pombe* detects and responds to glucose through a cAMP signaling pathway that is not essential for cell viability [1]. This allows one to construct strains that express a wide range of PKA activity and to identify compounds in high-throughput screens (HTSs) that alter PKA activity without killing the cells. PKA represses transcription of genes involved in gluconeogenesis and sexual development, including the *fbp1* gene whose expression can vary over a 200-fold range in a PKA-dependent manner [2, 3]. Most of the *S. pombe* PKA pathway genes were originally identified by their role in controlling transcription of an *fbp1-ura4* reporter, whose expression is required for uracil biosynthesis, but is toxic in cells exposed to the pyrimidine analog 5-fluoro-orotic acid (5FOA) [2]. Unlike wild-type strains, mutants with reduced PKA activity can form colonies on glucose-rich solid medium lacking uracil, but lose the ability to grow on 5FOA medium. Suppressor mutations and cloned genes

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_6, © Springer Science+Business Media New York 2015

that restore glucose-repression confer 5FOA-resistant (5FOA^R) growth to strains with low PKA activity [1]. The *fbp1-ura4* reporter can also be used in HTSs to detect small molecules that elevate PKA activity to allow 5FOA^R growth by a strain whose PDE activity is responsible for a low PKA, 5FOA-sensitive (5FOA^S) phenotype [4]. Such screens can be carried out in strains that produce cAMP or strains lacking adenylyl cyclase (AC) in which PKA is activated by exogenous cAMP or cGMP, allowing the detection of inhibitors of both cAMP- and cGMP-hydrolyzing PDEs [5]. We have used this approach to identify PDE4, PDE7, PDE8, and PDE11 inhibitors that are biologically active in cell culture [4, 6–8]. One could also use this screen to detect compounds that elevate the activity of a heterologously expressed AC or PKA protein.

While *S. pombe* strains carrying mutations that lower PKA activity can be detected by their ability to form colonies on solid medium lacking uracil [2], this phenotype is not sufficiently robust for HTSs to detect compounds that reduce PKA activity. Recently, we constructed an *fbp1*-driven GFP reporter that allows for small-molecule screens to detect compounds that stimulate PDE activity or inhibit AC (or their associated Gα_S) or PKA proteins [9]. Cell-based assays utilizing the *fbp1-ura4* and *fbp1*-GFP reporters allow for inexpensive HTSs for small-molecule modulators of cloned PKA pathway genes expressed in *S. pombe*.

Here, we describe the general process of creating and screening *S. pombe* strains that express cAMP pathway genes. This is done by PCR amplification of cloned genes and their introduction into *S. pombe* expression vectors through the transformation of host strains lacking the activity of interest. These strains are then used to create a screening strain that is subjected to assay optimization to allow the detection of either inhibitors or activators of these activities. Once optimized, these assays are then suitable for HTSs in a 384-well format.

2 Materials

Edinburgh Minimal Medium (EMM) can be stored at room temperature. Media containing 5FOA or cyclic nucleotides should be stored at 4 °C. Cyclic nucleotide-containing solutions should be used fresh, although storage for up to 1 month at 4 °C is acceptable.

2.1 Media Components

1. Synthetic complete mix lacking uracil: Combine 2 g each of adenine, alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and valine, with 4 g of leucine, 0.1 g of inositol, and 0.2 g of para-aminobenzoic acid. This mix is

stored as a dry powder at room temperature in a large enough container to allow mixing by vigorous shaking and should be made fresh annually.

2. 5FOA liquid medium: Dissolve 80 g of glucose, 1.45 g of yeast nitrogen base without amino acids and without ammonium sulfate, 5 g of ammonium sulfate, 0.4 g of 5FOA, 2 g of synthetic complete mix lacking uracil, and 50 mg of uracil in 950 mL of distilled H_2O. Dissolve by stirring under low heat. Filter-sterilize (*see* **Note 1**).
3. 5FOA solid medium: Dissolve 80 g of glucose, 1.45 g of yeast nitrogen base without amino acids and without ammonium sulfate, 5 g of ammonium sulfate, 0.4 g of 5FOA, 2 g of synthetic complete mix lacking uracil, and 50 mg of uracil in 450 mL of distilled H_2O. Dissolve by stirring under low heat. Filter-sterilize. Mix with 490 mL of water plus 20 g of Bacto agar that has been autoclaved in a 2 L flask (*see* **Note 1**).
4. EMM complete liquid medium: Dissolve 12.33 g of EMM without dextrose and 30 g of glucose in 970 mL of distilled H_2O. Add 150 mg of leucine and 75 mg each of adenine, histidine, lysine, and uracil (for selective media, do not add the supplement for which selection is based). Dissolve by mixing under low heat and filter-sterilize (*see* **Note 2**).
5. EMM complete solid medium: Dissolve 12.33 g of EMM without dextrose and 30 g of glucose in 470 mL of distilled H_2O. Add 150 mg of leucine and 75 mg each of adenine, histidine, lysine, and uracil (for selective media, do not add the supplement for which selection is based). Dissolve by mixing under low heat and filter-sterilize. Combine with 490 mL of water plus 20 g of Bacto agar that has been autoclaved in a 2 L flask.
6. Cyclic nucleotides: Make 10 mM cAMP and 5 mM cGMP stock solutions in both EMM and 5FOA growth media. Determine that cyclic nucleotides are fully dissolved before filter-sterilizing (*see* **Note 3**). Store at 4 °C for no more than 1 month.
7. Small molecules: Make stock solutions in DMSO (20–100 mM depending upon solubility). For yeast medium containing compounds, such as a positive control compound for an assay or HTS, place an Eppendorf tube containing medium into a beaker of boiling water and turn off heat. After 30 s to 1 min add compounds dissolved in DMSO. Mix well, but do not vortex (*see* **Note 4**).

2.2 General Equipment

1. 384-Well microtiter assay dishes: Clear sterile plates such as the Corning 3680 assay plate for 5FOA-based growth screens and black sterile plates with clear bottoms such as the Corning 3712 assay plate for GFP-based screens.

2. Multichannel pipettes (16 channels with 50 μL capacity): To deliver media and cells to wells of a 384-well microtiter plate or when making serial dilutions of compounds or cyclic nucleotide-containing medium in a microtiter plate.
3. Liquid handlers: For experiments requiring a large number of wells, use a liquid handler such as the Thermo Multidrop 384 or Wellmate microplate dispenser.

3 Methods

Construction of strains expressing cAMP pathway genes is straightforward for individuals with experience in yeast molecular genetics. While these methods are not technically demanding, one should consider collaborating with a yeast lab for the gene cloning and strain construction steps of the project. The genotype of the final screening strain will depend upon the goal of the screen (i.e., what is the target enzyme and is one screening for compounds that increase or decrease the target's activity?) (*see* **Note 5**).

3.1 Construct Yeast Strains Expressing the Gene(s) of Interest

1. PCR amplify the gene of interest using 80-mer oligonucleotides that consist of ~20 bases at the 3′ ends to amplify the gene and ~60 bases at the 5′ ends to target insertion into an expression vector. There are many *S. pombe* expression vectors that possess promoters of varying strength such as the strong *adh1* and *nmt1* promoters, the moderately active *nmt41* and *tif471* promoters, or the weakly active *lys7* promoter (Fig. 1) (*see* **Note 6**). PDE genes can also be directly inserted into the *S. pombe cgs2* PDE gene locus [4] (*see* **Note** 7).
2. Introduce PCR products carrying the gene of interest into an expression vector by gap-repair transformation in which the PCR product and linearized vector are co-transformed into a host *S. pombe* strain [10]. The genotype of the host will depend upon the type of gene that is being cloned (*see* **Note 8**).
3. Plate the transformed cells onto solid medium that is selective for the marker in the cloning vector. For example, *LEU2-marked* plasmids are selected for on EMM medium lacking leucine, while *lys2*-marked plasmids are selected for on EMM medium lacking lysine. In this way, only plasmid-carrying cells can form colonies.
4. Screen transformants based on the expected change in PKA activity (*see* **Notes 6** and **8**).
5. Rescue plasmids from yeast to *E. coli* [11] for amplification and purification: This is done by a simultaneous glass bead lysis of yeast transformants together with phenol-chloroform extraction of the nucleic acids that can be used to transform *E. coli*. Confirm by DNA sequence analysis that candidate plasmids carry the desired genes.

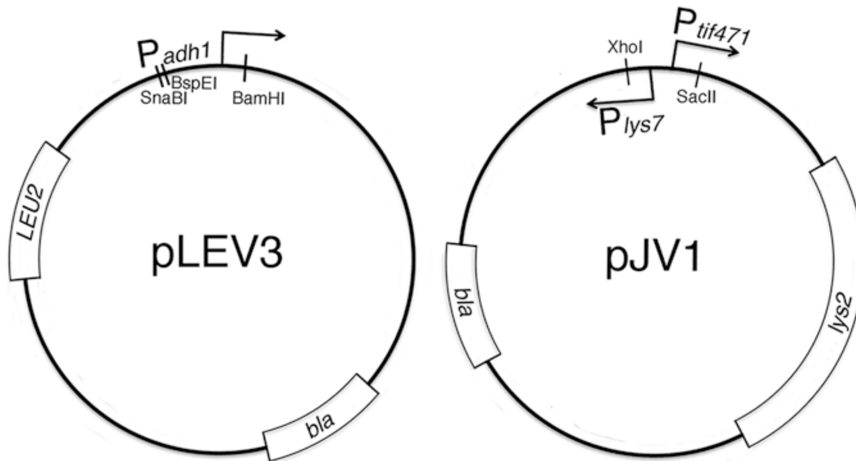


Fig. 1 *S. pombe* expression vectors for cAMP pathway genes. Plasmid pLEV3 [15] carries the *LEU2* selectable marker and is used to drive gene expression from the strong *adh1* promoter. Genes are inserted into BamHI-cut plasmid by gap repair transformation [10]. Once the gene of interest is cloned, the plasmid can be linearized with either SnaBI or BspEI, as long as these sites do not exist in the cloned gene, to direct insertion into the *adh1* locus of the *S. pombe* genome. Plasmid pJV1 is a derivative of the pRH3 cloning vector [16] carrying the moderate *tif471* and weak *lys7* promoters from plasmid pUL57 [17]. Genes are inserted in front of the *tif471* promoter using SacII-cut plasmid or the *lys7* promoter using XhoI-cut plasmid by gap repair transformation. Once the gene of interest is cloned, the plasmid can be linearized at any of several unique restriction sites in the *lys2* gene such as AgeI, BglII, NsiI, PstI, or SexAI to direct insertion into the *lys2* locus of the *S. pombe* genome

6. Introduce plasmids into the *S. pombe* chromosome of a host strain by linearizing within a portion of the plasmid that is homologous to the target site of integration and transforming a host strain (Fig. 1) (*see* **Note 9**).
7. Construct HTS strains by crosses and tetrad dissection to combine the gene or genes of interest with the *fbp1-ura4* or *fbp1-GFP* reporter and other mutations to facilitate the screen (*see* **Notes 10** and **11**). PKA activity of the host strain should be low for an *fbp1-ura4*-mediated HTS for PDE inhibitors, AC activators, or PKA activators and high for an *fbp1-GFP-mediated* HTS for PDE activators, AC inhibitors, $G\alpha_S$ inhibitors, or PKA inhibitors.
8. Pilot the HTS to optimize screening conditions as described in Subheading 3.2.

3.2 Pilot 5FOA Assays

3.2.1 For Strains Lacking AC Activity

This procedure identifies the optimal concentration of exogenously added cAMP or cGMP for use in an HTS for which the compound of interest elevates PKA activity by inhibiting a target PDE. This concentration should slightly increase the OD_{600} of a culture growing in 5FOA medium in a microtiter dish well. For example, in the absence of added cyclic nucleotide the culture should grow to an OD_{600} of less than 0.1, while in the presence of cyclic nucleotide it

should grow to an OD_{600} of 0.1–0.2. Under these conditions, PDE inhibition should result in an OD_{600} of ~1.2 [7, 8] (Fig. 2) (*see* **Note 12**).

1. Inoculate a 5 mL EMM complete liquid culture in an 18 × 150 mm culture tube from a plate of freshly growing cells. Grow overnight at 30 °C with shaking.
2. Determine the cell density using a hemocytometer. Subculture the cells into fresh EMM complete liquid medium that has been supplemented with a concentration of a cyclic nucleotide that will repress the *fbp1-ura4* reporter (this can be combined with a PDE inhibitor, if available, to reduce the amount of cyclic nucleotide required) targeting for a cell concentration of 1×10^7 cells/mL the following day (*see* **Notes 13** and **14**).
3. Once the cells have grown to exponential phase, pipet 25 μL of fresh 5FOA medium lacking cells into the wells of a 384-well microtiter dish (columns 2–16).
4. Pipet 75 μL of 5FOA medium containing a cyclic nucleotide (10 mM cAMP or 5 mM cGMP) into column 1 of the 384-well microtiter dish.
5. Carry out serial dilutions of the cyclic nucleotide by transferring 50 μL of medium from column 1 to column 2 using a multichannel pipette, and repeating through to column 14 (remove 50 μL of medium from column 14 after mixing and discard to leave 25 μL in the wells). Columns 15 and 16 serve as negative controls and should not contain cyclic nucleotides (*see* **Note 15**).
6. Centrifuge the cells and resuspend in 5FOA medium without cyclic nucleotides to a density of 3×10^5 cells/mL (*see* **Note 16**). Transfer to a sterile, empty Petri plate to allow pipetting using a multichannel pipette.
7. Pipet 25 μL of cells to each well. Avoid making bubbles in the wells that would interfere with the OD_{600} reading. Mix the cells into the medium by pipetting or using a microtiter plate vortexer after transferring cells to the plate.
8. Incubate the dishes for 48 h at 30 °C in a sealed container with wet paper towels to reduce evaporation, and stacked between two blank microtiter dishes to reduce condensation.
9. Resuspend cells by pipetting or by vortexing with a microtiter dish vortex. Measure the OD_{600} of wells using a plate reader (*see* **Note 17**).
10. An optimized assay is one in which the growth conditions prior to the assay and the initial cell density in the assay produce an OD_{600} of less than 0.2 in 5FOA medium lacking cyclic nucleotides and of approximately 1.2 in medium containing a high level of cAMP or cGMP (Fig. 2a).

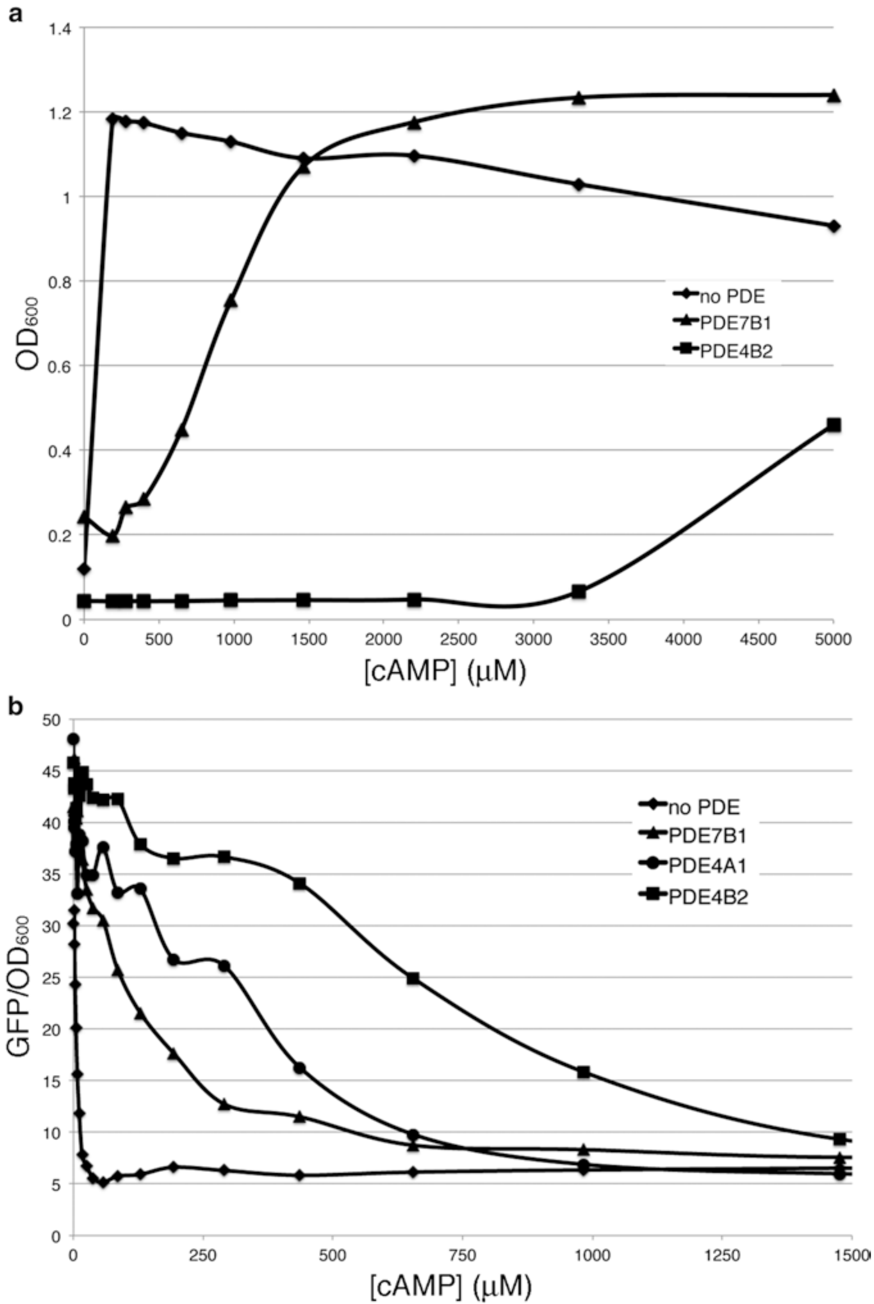


Fig. 2 cAMP response curves in strains lacking AC activity. (**a**) 5FOA growth response to exogenous cAMP in strains expressing no PDE activity, PDE7B1, or PDE4B2. These data indicate that PDE4B2 is more active than PDE7B1 in this system. (**b**) GFP expression in response to exogenous cAMP in strains expressing no PDE activity, PDE7B1, PDE4A1, or PDE4B2. These data indicate that PDE4B2 is more active than PDE4A1, which is more active than PDE7B1 in this system

3.2.2 For Strains Expressing an AC Gene or When Screening for PKA Activators

Strains that express the *S. pombe git2* AC (or a heterologously expressed AC, for which one is screening for activators) will be 5FOA^S if the PDE activity outweighs the AC activity. As such, these strains do not require the addition of cyclic nucleotides to the 5FOA growth medium for the HTSs. Cyclic nucleotides or PDE inhibitors are still required in the EMM medium used to grow cells prior to the screen to repress the *fbp1-ura4* reporter before the cells are transferred to 5FOA medium. The same is true for screens designed to identify PKA activators that increase PKA activity in a cAMP-independent manner. The following modified protocol is used to optimize the 5FOA assay for such HTSs.

1. Culture the cells according to **steps 1** and **2** described in Subheading 3.2.1. PKA must be activated in these cultures by the presence of a cyclic nucleotide or a known PDE inhibitor.
2. Pipet 25 μL 5FOA medium with (a) no cyclic nucleotide as a negative control, (b) 10 mM cAMP or 5 mM cGMP as a cyclic nucleotide positive control, or (c) 40 μM of a known PDE inhibitor in <0.5 % DMSO as a small-molecule positive control, into replicate wells of a 384-well microtiter dish.
3. Proceed with **steps 6–9** described in Subheading 3.2.1 to complete the 5FOA assay. Calculate the Z' factor to determine whether or not these conditions are suitable for an HTS (*see* **Notes 16** and **18**).

3.3 5FOA-Based HTS for PDE Inhibitors, AC Activators, or PKA Activators

The specific details for these HTSs depend on the equipment at the screening facility (*see* **Note 19**).

1. Grow the cells according to conditions established during assay optimization pilot studies (Subheading 3.2).
2. Pellet the cells to remove cyclic nucleotides and/or PDE inhibitors and resuspend in 5FOA medium (with or without cyclic nucleotides depending on the strain as determined in the pilot studies (Subheading 3.2)) at the desired cell density.
3. Reserve column 23 for negative, DMSO-pinned controls and column 24 for positive controls (medium containing a high concentration of a cyclic nucleotide or pinned with a known PDE inhibitor).
4. Deliver 30–50 μL of cells to wells of two microtiter dishes for each compound plate to be screened using an automated liquid handler (*see* **Note 20**).
5. Pin 100 nL of compounds from compound plates into wells.
6. Incubate the microtiter dishes for 48 h at 30 °C in a sealed container with moist paper towels to reduce evaporation in the wells and sandwiched between blank plates to reduce condensation on the plate lids.

7. Resuspend the cells by vortexing and measure the OD_{600} using a plate reader.
8. Determine the Z' factor for the assay and Z scores of individual compounds (*see* **Notes 18** and **21**).

3.4 Pilot GFP-Based Assay

In contrast to 5FOA growth-based HTSs, *fbp1-GFP*-based HTSs are suited for detecting molecules that reduce PKA activity to increase GFP expression. Thus, these screens can identify PDE activators, $G\alpha_S$ inhibitors, AC inhibitors, or PKA inhibitors. Such screens utilize strains whose PKA activity largely represses *fbp1-GFP* expression, such that a reduction in cyclic nucleotide levels or direct PKA inhibition is readily detected. While one can use exogenous cAMP or cGMP to regulate PKA (Fig. 2), it is less expensive to use mutations in the *S. pombe* cAMP pathway to modulate AC activity if the target protein is either a PDE or PKA [2, 3, 12]. Furthermore, target AC genes can be expressed in *S. pombe* in the presence or absence of $G\alpha_S$ proteins to create additional PKA pathway targets for inhibition. By controlling the level of PDE activity in our strains, we have detected both basal mammalian AC activity and GNAS1 ($G\alpha_S$)-stimulated AC activity suitable for HTS inhibitor screens (Kwak and Hoffman, unpublished). In these cases, one does not need to carry out cyclic nucleotide response curves. Simply identify a strain whose GFP signal is poised to increase upon lowering PKA activity (*see* **Note 22**). The following pilot screen will allow one to characterize a candidate screening strain and optimize the initial cell density in the assay to produce a consistent normalized GFP value with a low standard deviation.

1. Inoculate a 5 mL EMM complete liquid culture in an 18×150 mm culture tube from a plate of freshly growing cells. Grow overnight at 30 °C with shaking.
2. Count the cells using a hemocytometer. Dilute them into fresh medium to achieve a cell concentration of 1×10^7 cells/mL for the following day (*see* **Note 13**).
3. Collect the cells by centrifugation and resuspend in fresh EMM medium at a density of 4×10^6 cells/mL (*see* **Note 23**).
4. Transfer 50 μL of cells into wells of a 384-well microtiter dish (black walls, optical bottoms). Incubate at 30 °C in a sealed container with wet paper towels to reduce evaporation and sandwiched between two blank dishes to reduce condensation on the dish lids.
5. Read the OD_{600} and GFP signal after 24- and 48-h incubation and determine whether the OD_{600} value would suggest a need to increase or decrease the starting cell density (*see* **Note 24**). Divide the GFP signal by the OD_{600} value to generate a normalized GFP value (GFP/OD_{600}) as seen in Fig. 2b that is in the range of 10–15 units (*see* **Note 22**).

3.5 GFP-Based HTS for PDE Activators, AC Inhibitors, or PKA Inhibitors

1. Carry out the HTS using the growth conditions and initial cell density as determined in the pilot experiments in Subheading 3.4.
2. Deliver 30–50 μL of cells to duplicate microtiter dishes using an automated liquid handler (*see* **Note 20**).
3. Pin 100 nL of compounds or DMSO (negative control) to the dishes. Measure the fluorescence immediately to identify compounds that are themselves fluorescent.
4. Incubate microtiter dishes for 48 h at 30 °C in a sealed container with wet paper towels to reduce evaporation in the wells. Sandwich the dishes between two blank dishes to reduce condensation on the dish lids.
5. Measure OD_{600} and fluorescence of the wells to determine the normalized GFP signals. Use averages and standard deviations of negative (DMSO-pinned) and positive (if there is a positive control) control wells to calculate the Z' factor for the assay and Z scores for test compounds (*see* **Notes 18** and **21**).

4 Notes

1. Accurate weighing of components for 5FOA medium is crucial. Too little 5FOA allows weak growth of the screening strain. Too much 5FOA reduces $5FOA^R$ growth. Similarly, insufficient uracil can reduce $5FOA^R$ growth, while too much uracil allows strains with low PKA activity to grow in 5FOA medium.
2. Some commercially formulated Edinburgh Minimal Medium (EMM) contain heat-labile components, so note whether autoclaving or filter-sterilization should be used.
3. Cyclic nucleotides lose activity in liquid medium upon long-term storage. Short-term storage at 4 °C is acceptable, but do not store these media for more than 1 month before using.
4. Compounds in HTS libraries often display poor solubility in yeast growth media. After preheating the medium, gently mix by pipetting. Vortexing can cause compounds to come out of solution.
5. The general approach is to first clone the target gene of interest into an autonomously replicating *S. pombe* expression vector and then to integrate this plasmid into the *S. pombe* genome to reduce copy number and increase mitotic stability. *S. pombe* encodes a single AC (*git2/cyr1*), PDE (*cgs2/pde1*), PKA regulatory subunit (*cgs1*), and PKA catalytic subunit (*pka1*) gene. The host strain should lack the endogenous activity of interest to create a strain that is suitable for the identification of inhibitors or activators of the target protein (these *S. pombe* genes are not essential and deletion strains exist for all

four of these genes). In addition, the host for the initial cloning of the gene of interest should be a homothallic (h^{90}) strain (homothallic cells undergo mating-type switching to produce cells capable of mating and sexual development within a colony). High PKA activity inhibits mating and sporulation in homothallic strains, while low PKA activity shifts h^{90} cells from mitotic growth to mating and sporulation [5]. PDE expression from a plasmid can reduce cAMP levels (unless working with a cGMP-specific PDE) to increase mating in a host strain that lacks PDE activity. Similarly, PKA regulatory subunit genes could be detected by their ability to increase mating in a strain that expresses an unregulated PKA catalytic subunit gene. Conversely, AC and PKA genes can be detected by their ability to confer growth and reduce mating in colonies of an h^{90} strain that lacks the *git2* AC gene.

6. Expression vectors such as pJV1 and pLEV3 (Fig. 1), as well as those that utilize the strong *nmt1* or moderate *nmt41* promoters, can be obtained from our laboratory.
7. Screening strains must not express the *S. pombe* Cgs2 PDE if another PDE is the target enzyme. The target PDE can be expressed from the *cgs2* locus or from an integrated expression vector, which would allow more choices of promoters to vary the level of expression. In the latter case, *cgs2* can be inactivated by a disruption allele or the *cgs2-2* frameshift allele [13]. Do not use autonomously replicating plasmids to express the PDE or other target genes in the final screening strains as this produces greater cell-to-cell variation when compared to using single-copy integrated constructs.
8. Mating is detected by iodine staining of the colonies (invert a plate of colonies over a Petri plate lid containing crushed iodine for 30 s to 2 min). Confirm the presence of asci by microscopy of cells from iodine-stained colonies.
9. Linearized plasmids can recombine with chromosomal loci that are homologous to the linearized ends of the plasmid. Replica-plating transformants from selective medium (lacking leucine or lysine, for example) to nonselective medium (containing leucine or lysine) allows plasmid loss of autonomous, but not integrated, plasmids. Replica-plate to nonselective medium every 2–3 days for 1 week, and then back to selective medium. Colonies of cells that carry an integrated plasmid are easily identified after 2-day growth.
10. Tetrad dissections are performed on a specialized microscope that is equipped with a stage that can hold a Petri dish. If the equipment and expertise for such manipulations are not available on site or through a collaboration, spores can be obtained by zymolyase or glusulase treatment of acsi to produce isolated

spores and plated under dilute conditions to obtain individual colonies that can be screened for the desired genotype. Tetrad dissection is preferable as one can observe the phenotype of all four progeny from a single ascus to be certain of the genetic integrity of each of the strains produced.

11. In addition to the *fbp1-ura4* or *fbp1-GFP* reporters, HTS strains have several features to enhance screening. When screening for PDE inhibitors or activators, there are several mutations in cAMP pathway genes that can be incorporated to alter the level of cAMP synthesis [5]. If PDE activity is very low, one must delete the *git2* AC gene and use cAMP or cGMP to regulate PKA [5]. The *git2* deletion is also used in strains that express ACs from other organisms. Finally, deleting the *pap1* transcription factor gene increases sensitivity to 5FOA and may reduce efflux of compounds [14].

12. The cyclic nucleotide optimization assay is not required for screens in which PKA activation occurs due to inhibition of a PKA regulatory subunit or stimulation of an AC. In such assays, cyclic nucleotides are required during the growth of the culture in EMM medium to repress the *fbp1-ura4* reporter, but are not required in the 5FOA medium.

13. Growth rates of strains vary depending on the strain genotype and the growth medium. An average doubling time is ~3 h; however strains will fail to grow if diluted too much. Plan to grow cultures for no more than six doublings prior to initiating the screen. Cultures larger than 10 mL are grown in flasks that are five times the culture volume to allow aeration. Strains should grow to between 0.5×10^7 and 2×10^7 cells/mL to assure proper regulation of reporter expression. Until one has established reproducible culturing conditions for a given strain, it is wise to start multiple cultures with different initial cell densities.

14. To detect PDE inhibition, AC activation, or PKA activation by $5FOA^R$ growth, the *fbp1-ura4* reporter must be repressed prior to the start of the screen. Otherwise, preexisting Ura4 protein will kill cells even if PKA is subsequently activated. This can be accomplished by adding sufficient cAMP or cGMP (generally, one uses the nucleotide against which the PDE is least effective) to EMM complete liquid medium. Alternatively, one can use both a cyclic nucleotide and a known inhibitor of the PDE if an effective inhibitor is available. The concentration of the inhibitor depends upon its potency, but may be 10–20 μM. The final concentration of DMSO in the medium should be less than 0.5 %. Microscopic examination of the cells can determine if PKA is activated as this will lead to cells that are visibly longer than cells of this strain when growing in medium lacking cAMP or cGMP.

15. Multichannel pipettes allow one to set up replicate assays or to carry out dilutions to test more than one strain at a time. After transferring 50 μL from one column to the next, pipette gently to mix the medium. One set of tips can be used to generate a dilution series. Carefully remove bubbles from wells if made during pipetting.
16. Centrifugation is carried out at low speed (1,000 × *g* for 5 min) in a tabletop centrifuge. Diluting cells to 3×10^5 cells/mL produces a starting cell density of 1.5×10^5 cells/mL in the wells. Optimal initial densities vary from 0.5×10^5 to 2×10^5 cells/mL and should be determined for each strain. Too few cells can limit the 5FOA growth response, while too many cells can produce high OD_{600} values in the negative controls.
17. Cells will settle out of the medium and grow unevenly in the wells, leading to inconsistent OD_{600} values unless cells are resuspended before reading. Do not centrifuge plates as this also produces uneven cell distributions in the wells.
18. The Z' factor, which must be >0.5 for an HTS, is determined by subtracting three times the sum of the standard deviations of positive and negative controls, divided by the absolute value of the difference between means of the positive and negative controls, from 1:

$$Z' \text{ factor} = \frac{1 - 3(\sigma p + \sigma n)}{|\mu p - \mu n|}$$

19. A screening facility is not necessarily required to conduct these assays. If one is testing a small number of compounds, the assays can be carried out using standard or multichannel pipettes. If working with only a few hundred compounds, a manual pin tool such as the V&P Scientific VP 386 Multi-Blot replicator could be used for compound transfer. However, screening facilities are required when carrying out HTSs. They provide access to chemical libraries and the automated equipment needed to screen large numbers of compounds. They provide the technical support for optimizing screens and the computational support for analyzing the data.
20. The volume of the culture affects two variables. For 5FOA growth assays, larger cultures allow for more cell growth and thus a higher maximum OD_{600}, which could produce higher *Z* scores. However, since one typically pins 100 nL of compound into wells, a larger volume will also lead to a lower final concentration of compound, which may reduce the impact of exposure to the compound. Screening facilities may also have a preference regarding the culture volume based on the calibration of their pin tools. Finally, compounds are generally dissolved at a concentration of 5 mg/mL; thus the molarities vary as a

function of molecular weights. For example, a 250 Da compound will be at 20 mM, resulting in a screening concentration of 40 μM when pinned into a 50 μL culture.

21. *Z* scores are determined by taking the experimental sample value and subtracting the mean of the negative controls. This is then divided by the standard deviation of the negative controls. Candidate hits should have *Z* scores of at least 6 to be considered statistically significant. Depending upon the quality of the HTS, one may choose a larger cutoff to identify candidates.
22. As seen in Fig. 2, the dynamic range of the GFP/OD_{600} values is from ~6 units for fully repressed cells to 40–50 units for fully depressed cells. HTS strains should express from 10 to 15 units in this assay. Strains that express less than 10 units may be insensitive to small changes in cyclic nucleotide levels, while strains that express more than 15 units will have relatively high standard deviations in the negative controls.
23. Unlike the 5FOA assay, the GFP assay is permissive for growth of cultures; therefore a higher starting cell density is used. We obtain similar GFP/OD_{600} values after 48-h growth when starting with 2×10^6 to 5×10^6 cells/mL. Higher cell densities reduce the amount of growth by the culture after compound addition and could reduce the response to a compound. Lower cell densities can prevent the culture from growing to saturation after 48 h, and thus reduce the GFP signal and increase well-to-well variability.
24. A good starting density should produce an OD_{600} in the wells of 0.7–1.0 after 24 h and 1.2–1.6 after 48 h when using 50 μL cultures. Cells settle out of the medium during growth; therefore the GFP signal should be read via a bottom read of the plate.

Acknowledgments

This work was supported by NIH grant GM079662, the Peter Rieser Lectureship Fund, and a grant from Boston College to C.S.H.

References

1. Hoffman CS (2005) Glucose sensing via the protein kinase A pathway in *Schizosaccharomyces pombe*. Biochem Soc Trans 33:257–260
2. Hoffman CS, Winston F (1990) Isolation and characterization of mutants constitutive for expression of the *fbp1* gene of *Schizosaccharomyces pombe*. Genetics 124:807–816
3. Hoffman CS, Winston F (1991) Glucose repression of transcription of the *Schizosaccharomyces pombe fbp1* gene occurs by a cAMP signaling pathway. Genes Dev 5: 561–571
4. Ivey FD et al (2008) Development of a fission yeast-based high-throughput screen to identify

chemical regulators of cAMP phosphodiesterases. J Biomol Screen 13:62–71

5. Demirbas D et al (2011) Use of a *Schizosaccharomyces pombe* PKA-repressible reporter to study cGMP metabolising phosphodiesterases. Cell Signal 23:594–601
6. Alaamery MA et al (2010) New classes of PDE7 inhibitors identified by a fission yeast-based HTS. J Biomol Screen 15:359–367
7. Ceyhan O, Birsoy K, Hoffman CS (2012) Identification of biologically active PDE11-selective inhibitors using a yeast-based high-throughput screen. Chem Biol 19:155–163
8. Demirbas D et al (2013) A yeast-based chemical screen identifies a PDE inhibitor that elevates steroidogenesis in mouse Leydig cells via PDE8 and PDE4 inhibition. PLoS One 8: e71279
9. de Medeiros AS et al (2013) Use of PKA-mediated phenotypes for genetic and small-molecule screens in *Schizosaccharomyces pombe*. Biochem Soc Trans 41:1692–1695
10. Wang L et al (2004) Strategies for gene disruptions and plasmid constructions in fission yeast. Methods 33:199–205
11. Hoffman CS, Winston F (1987) A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. Gene 57:267–272
12. Welton RM, Hoffman CS (2000) Glucose monitoring in fission yeast via the gpa2 Gα, the git5 Gβ, and the git3 putative glucose receptor. Genetics 156:513–521
13. Wang L et al (2005) *Schizosaccharomyces pombe* adenylate cyclase suppressor mutations suggest a role for cAMP phosphodiesterase regulation in feedback control of glucose/cAMP signaling. Genetics 171:1523–1533
14. Ivey FD et al (2010) Activated alleles of the *Schizosaccharomyces pombe gpa2*$^{+}$ Gα gene identify residues involved in GDP-GTP exchange. Eukaryot Cell 9:626–633
15. Janoo RT et al (2001) Transcriptional regulators of the *Schizosaccharomyces pombe fbp1* gene include two redundant Tup1p-like corepressors and the CCAAT binding factor activation complex. Genetics 157:1205–1215
16. Hoffman RL, Hoffman CS (2006) Cloning the *Schizosaccharomyces pombe lys2*$^{+}$ gene and construction of new molecular genetic tools. Curr Genet 49:414–420
17. Mudge DK et al (2012) Use of a *ura5*$^{+}$-*lys7*$^{+}$ cassette to construct unmarked gene knock-ins in *Schizosaccharomyces pombe*. Curr Genet 58:59–64

Chapter 7

A Method for High-Throughput Analysis of Chronological Aging in *Schizosaccharomyces pombe*

Jessica Stephan and Ann E. Ehrenhofer-Murray

Abstract

The measurement of chronological life span (CLS) in *Schizosaccharomyces pombe* is traditionally performed by plating back aliquots of aging liquid cultures on solid medium and counting the number of colony-forming units (CFU). However, this method is labor and cost intensive and therefore not amenable to high-throughput screening. Here, we describe a simple method for CLS measurement using aging minicultures in microtiter plates and batch plate-back for the determination of culture viability. This assay can be used to screen a large number of strains, conditions, or compounds in parallel for effects on aging.

Key words Aging, Chronological life span, Longevity, Compound screen, Microtiter plate, High-throughput

1 Introduction

In the last two decades, tremendous advances have been made in understanding the molecular mechanisms underlying aging processes. Genetic pathways have been defined that, when mutated, cause lifespan extension in a wide range of organisms, including the target of rapamycin (TOR) [1] and the insulin/IGF-1 signaling pathway [2], suggesting a strong evolutionary conservation of the respective aging mechanisms. This suggests that chemical inhibitors of these or other aging pathways can be used as antiaging agents, perhaps even for human use, and there thus is an increasing interest in identifying such compounds. Due to the evolutionary conservation, one cost-effective option is to screen for lifespan-extending compounds using the unicellular eukaryotes *Saccharomyces cerevisiae* or *Schizosaccharomyces pombe* as model organisms [3, 4].

Screening large compound libraries necessitates the availability of a high-throughput aging assay. In principle, two types of aging or life span can be distinguished in yeast: replicative (RLS) and chronological (CLS) life span [3]. RLS describes the number of

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_7, © Springer Science+Business Media New York 2015

mitotic divisions a cell can undergo before terminal senescence and serves as a model for the aging of actively dividing cells like germ line cells and stem cells. RLS assays in *S. cerevisiae* and *S. pombe* typically are performed by manual separation of mother and daughter cells under the microscope [5, 6], a laborious process that is not amenable to high throughput. RLS measurement in *S. cerevisiae* using microfluidics devices has been described [7, 8], but this method awaits development for large-scale analysis.

CLS refers to the time a nondividing cell population can remain viable, as defined by their ability to reenter the cell cycle after a longer period of time in stationary phase, and CLS has been linked to the aging of differentiated somatic cells, for example neurons [9]. The traditional method to determine the CLS of budding and fission yeast is to measure the ability of individual cells to form a colony, which is referred to as the colony-forming unit (CFU) method. For this purpose, aliquots of the aging cultures are taken at regular intervals and serially diluted. Multiple dilutions are plated on full medium plates and incubated at 30 °C for 3–5 days. The forming colonies are then counted and used to calculate the number of colony-forming units per mL culture (CFU/mL) [3]. However, this methodology requires substantial amounts of material per data point and is time consuming, and it thus is not suitable for high-throughput approaches. A high-throughput method was described by Murakami et al., but it requires special honeycomb plates and a matching plate reader (Bioscreen C MBR machine) [10], which may not be easily available. We therefore have designed a simple, time- and cost-effective method of measuring CLS that enables the CLS measurement of many yeast cultures simultaneously using *S. pombe* as a model organism [11]. The use of *S. pombe* has the advantage that, unlike *S. cerevisiae*, it does not exhibit regrowth or "gasping" of aging cells [12], which complicates lifespan measurements.

In order to facilitate high-throughput screening, we have adopted the CFU method of lifespan determination to the format of a microtiter plate. Using this assay, we have been able to recapitulate the well-documented effect of dietary restriction (Fig. 1), the signaling kinase Sck2, and glucose-mediated nutrient signaling via the Git3/PKA pathway on lifespan extension in *S. pombe* [11, 12]. Furthermore, we have used the assay to screen a small library of 522 natural products for lifespan-extending compounds (Fig. 2) and in doing so have discovered several compounds with antiaging properties in yeast (Fig. 3) [11]. In brief, small aging cultures are grown in 96-well microtiter plates, and batches from these "minicultures" are taken at regular time intervals and spotted on agar plates using a replica device in order to determine the viability of the culture (Fig. 1). Thus, the regrowth in these batches, rather than individual colony counting, serves as a measurement for aging in the miniculture. Since the aging cultures and regrowth assays are

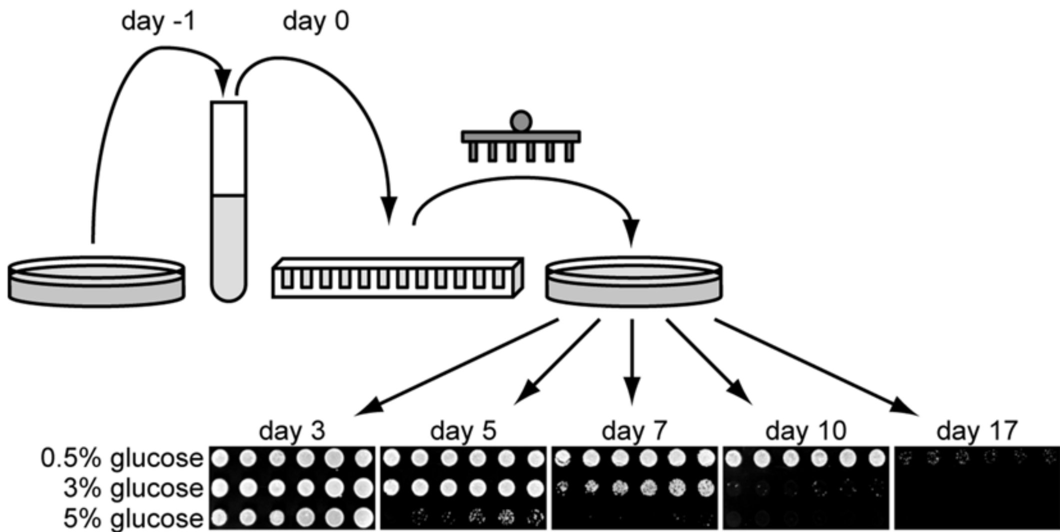


Fig. 1 High-throughput method for the measurement of CLS in *S. pombe. S. pombe* wild-type cells are streaked from cryogenic cultures on YES plates on day 3. On day 1, the cells are used to inoculate an overnight culture, which is used on day 0 to set up the aging cultures. On days 3, 5, 7, etc., the aging cultures are spotted onto YES plates using a replica tool and incubated at 30 °C for 2–3 days. Each spot represents 1 well. Decreased growth reflects a reduction in the ability of the cells in the aging culture to reenter the cell cycle and thus a shortening of CLS. Cells grown under dietary restriction (0.5 % glucose) are long-lived, whereas overnutrition (5 % glucose) accelerates aging

carried out in 96-well format, this method can easily be used to screen a large number of different conditions, compounds, or different strain backgrounds in a single experiment with a relatively low use of materials.

2 Materials

All media are prepared using ultrapure water and are sterilized before use by autoclaving or sterile filtration. All glassware and pipette tips are sterilized before use. Disposal of waste materials must be performed according to regulations.

2.1 Overnight Cultures of Aging Tester Strains

1. Cryogenic culture of the *S. pombe* strains to be analyzed, e.g., AEP1 (*h⁻ leu1-32 ura4-D18 his3-D3*) (*see* **Note 1**).
2. Yeast extract with supplements (YES) agar plates: 5 g of Bacto™ Yeast Extract (BD Becton, Dickinson and Company), 20 g of agar, 30 g of glucose, and 250 mg each of adenine, histidine, leucine, uracil, and lysine [13]. Transfer the dry chemicals into a 3 L Erlenmeyer flask, add 1 L of water, and autoclave during 20 min at 120 °C. Allow the medium to cool to approximately 60 °C before pouring 25 mL aliquots into plastic petri dishes (diameter 9 cm). The plates are stored at 4 °C before use.

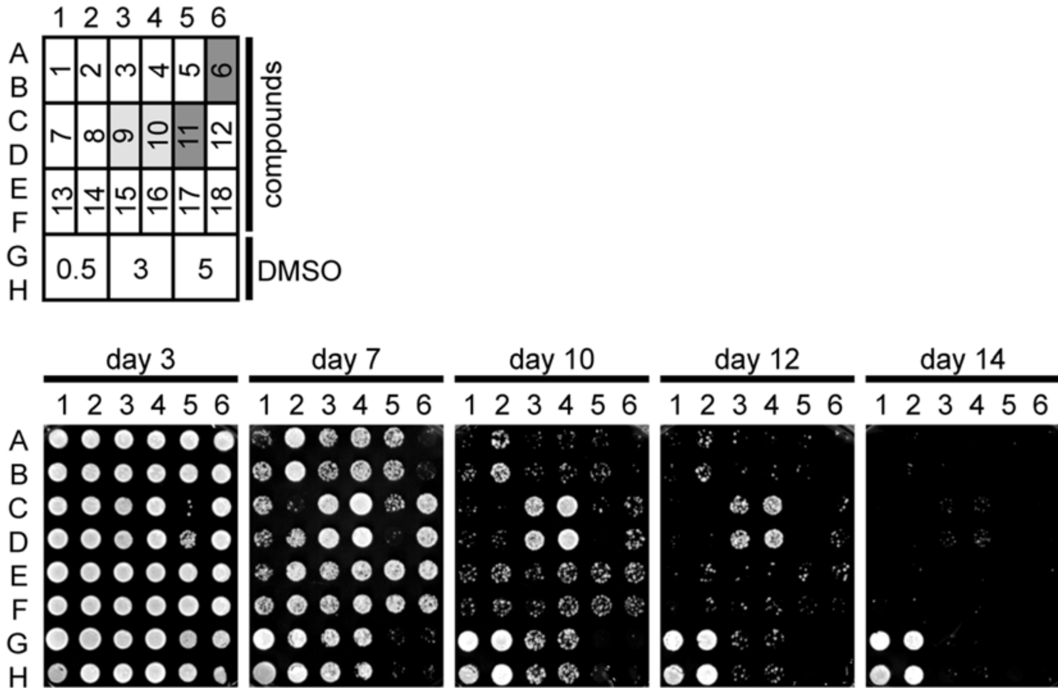


Fig. 2 Example of a screen for compounds that influence CLS. The wells of 96-well microtiter plates were inoculated with an *S. pombe* wild-type strain. The compounds (1–18) were tested in SD medium containing 3 % glucose. As controls, DMSO-treated cells were grown in standard SD medium (3 % glucose) as well as under dietary restriction and overnutrition conditions (0.5 % and 5 % glucose, respectively). At the indicated time points, the cells were spotted onto full medium agar plates and incubated at 30 °C for 2 days. *Light grey*: primary candidates that increased CLS, *dark grey*: primary candidates that resulted in a shortened life span

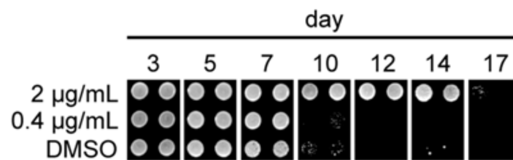


Fig. 3 Wortmannin as an example for a life-extending compound. *S. pombe* wild-type cells grown with 2 μg/mL wortmannin showed an extended life span compared to 0.4 μg/mL wortmannin or DMSO. Wortmannin likely decreases aging through the inhibition of TOR kinases

3. YES liquid medium: Reagents as above for YES plates, but without agar. Autoclave and store at room temperature.
4. Standard shaking incubator for incubation of liquid culture flasks at 30 °C.
5. Standard incubator for incubation of agar plates at 30 °C.

2.2 Preparation of Cultures for Aging Assay

1. 40 % glucose in water (w/v), autoclaved.
2. Supplemented synthetic dextrose (SD) medium (*see* **Note 2**): 6.7 g of Difco™ yeast nitrogen base without amino acids (BD

Becton, Dickinson and Company, Franklin Lakes, NJ, USA) [14]; depending on the auxotrophies of the *S. pombe* strains, add 150 mg each of adenine, uracil, histidine, and leucine (*see* **Note 3**). Transfer the dry chemicals to a 3 L Erlenmeyer flask and add 950 mL of water. Autoclave the medium and store it at room temperature. Before use, an adequate amount of an autoclaved 40 % glucose solution is added to the medium (*see* **Note 4**).

3. Compounds to be tested for their effect on CLS, dissolved in sterile-filtered DMSO at a concentration of approximately 0.2 mg/mL.
4. Multi-pipettes with eight channels: 0.5–10 μL and 30–300 μL.
5. Autoclavable 60 mL reagent reservoir with lid.
6. Sterile 96-well microtiter plates with U-based wells (*see* **Note 5**).
7. SILVERseal™ films (e.g., Greiner Bio-One International AG).
8. Standard spectrophotometer for determination of culture density by measurement of optical density at 600 nm (OD_{600}) and disposable plastic cuvettes.
9. Microtiter plate shaker combined with incubation chamber (e.g., TiMix5 with TH15, Edmund Buehler GmbH, Hechingen, Germany).

2.3 Measurement of CLS

1. Tabletop centrifuge with swing-out rotor and adapters for microtiter plates.
2. Replica plater for 96-well plate, 48 pins (e.g., Sigma R2383).
3. 100 % denatured ethanol in glass dish (*see* **Note 6**).
4. YES plates (*see* Subheading 2.1).
5. Standard incubator for incubation of agar plates at 30 °C.
6. Imaging system or camera for documentation.

3 Methods

Unless indicated otherwise, all experimental procedures are carried out at room temperature, and all *S. pombe* strains are grown at 30 °C. Sterile working procedures at all stages are important in order to avoid contamination of long-term aging cultures.

3.1 Overnight Cultures of Aging Tester Strains

1. Day-3: Directly streak yeast cells from cryogenic cultures on YES plates and grown for 2 days at 30 °C.
2. Day-1: Take a small amount (2–3 colonies) of the freshly grown yeast cells with an inoculation loop and use to inoculate 5 mL of liquid YES medium in a test tube or small Erlenmeyer flask for overnight cultures (Fig. 1).
3. Grow cultures overnight in an incubator shaker while shaking at 170 rpm at 30 °C.

3.2 Preparation of Aging Cultures

1. Day-0: Determine the OD_{600} of the overnight culture(s) (*see* **Note 7**). The OD_{600} should be between 2 and 4.
2. Using the reagent reservoir and a multi-pipette, dispense 150 μL of SD medium containing 3 % glucose into the appropriate wells of a microtiter plate (*see* **Note 8**) (Fig. 1).
3. Prepare controls for each half of a microtiter plate: SD medium with 0.5 % (dietary restriction), 3 % (normal), and 5 % glucose (overnutrition) (at least 2 wells per control).
4. Add 3 μL of 0.2 mg/mL compound stock (final compound concentration: 4 μg/mL) or 3 μL DMSO (as a control) to the appropriate wells (*see* **Note 9**). Test each compound/condition at least in duplicate.
5. Add 2 μL of overnight culture per well (*see* **Note 10**).
6. Seal the microtiter plates with SILVERseal film.
7. Shake the microtiter plates at 900 rpm and 30 °C (*see* **Note 11**).

3.3 Measurement of CLS

1. Day-3: Start CLS measurement 3 days after setting up the aging cultures (*see* **Note 12**).
2. Centrifuge the microtiter plates in the adaptors of the tabletop centrifuge for 2 min at 700 × *g* (*see* **Note 13**).
3. Remove the SILVERseal film and resuspend the cells by gently pipetting up and down with a multi-pipette.
4. Sterilize the replica plater by dipping in 100 % ethanol and flaming, and then allow a few moments for cooling.
5. Insert the replica tool into one-half of the microtiter plate of aging cultures, stir slightly, and remove swiftly in order for droplets of the cultures to adhere to the inoculation pins. Check visually for even droplets on all pins.
6. Place the replica tool gently on a YES plate to allow transfer of the culture droplets onto the plate. Check visually for even transfer (*see* **Note 14**).
7. Place the YES plates in the incubator and incubate at 30 °C for 2–3 days (*see* **Note 15**).
8. Clean the replica tool by rinsing in water and using an adequate brush. Dry with a paper towel.
9. Repeat **steps 1–8** of Subheading 3.3 with the second half of the microtiter plate if appropriate.
10. Seal the microtiter plate with a fresh SILVERseal film.
11. Return the microtiter plate to the shaking incubator, and shake at 900 rpm and 30 °C (*see* **Note 16**).
12. Incubate the spotted YES plates for 2 days at 30 °C and subsequently document the growth of the yeast cells by photographing the plates on a dark background (Figs. 1 and 2).

The density of cell growth per spot reflects the viability of the microtiter aging culture.

13. Repeat Subheading 3.3 at regular intervals (every 2–3 days).

4 Notes

1. *S. pombe* strains are available from public repositories, for instance the National Collection of Yeast Cultures (NCYC, www.ncyc.co.uk), or from the National BioResource Project (Yeast) (yeast.lab.nig.ac.jp/nig/index_en.html).
2. EMM (Edinburgh minimal medium) is the standard minimal medium for *S. pombe*, while the SD medium used here is the standard minimal medium for *S. cerevisiae*. However, previous studies have shown that nutrients are limiting in EMM, which causes lifespan shortening of *S. pombe* [4, 12]. Conversely, in SD medium, *S. pombe* CLS is increased by dietary restriction (0.5 % glucose) and shortened by overnutrition (5 % glucose) [11, 12], as is generally expected for dietary effects [1]. Therefore, it is advisable to use SD medium instead of EMM for *S. pombe* lifespan experiments.
3. Addition of supplements depends on the genotype of the strain used. For instance, the strain AEP1 (h^- *leu1-32 ura4-D18 his3-D3*) is unable to synthesize the amino acids leucine, uracil, and histidine due to mutations in the genes $leu1^+$, $ura4^+$, and $his3^+$. Accordingly, leucine, uracil, and histidine must be added to minimal growth medium.
4. Under normal conditions, the final glucose concentration is 3 %. For final use, prepare, e.g., 20 mL of EMM with 1.5 mL of 40 % glucose. For overnutrition or dietary restriction, 5 % (2.5 mL of 40 % glucose per 20 mL of EMM) and 0.5 % (250 μL) or 1 % glucose (500 μL) are used, respectively.
5. U-based microtiter plates are used in order to reduce settling of yeast cells.
6. The glass dish and ethanol are used to sterilize the replica plater device (*see* Subheading 2.3, **item 2**) and therefore should have the appropriate dimensions (approximately like a standard petri dish, diameter 9 cm). Fill with approximately 20 mL of denatured ethanol.
7. Take an aliquot of the culture and dilute with water or YES medium (e.g., 1:10). Use this dilution to measure the OD_{600} and calculate the OD_{600} of the undiluted culture. In general, the OD_{600} reading on the photometer should be between 0.05 and 0.6, because the relationship between cell number and OD_{600} reading is not linear outside of this range. If the final OD_{600} is too low, grow culture for a longer period, or retry

with a larger inoculum. In case that the final OD_{600} is too high (>4.0), it is recommended to start a new culture or to strongly dilute the culture and let the cells regrow, as a high OD_{600} indicates that the culture is in stationary phase. This measurement is mainly used if different strains are used in the same experiment to equilibrate the cell numbers in the starter aging culture.

8. In our hands, 150 μL is the optimal culture volume, because it allows sufficient movement and aeration in the well upon shaking. Furthermore, this culture volume is sufficient for long-term aging experiments.
9. For a compound of molecular weight 400 g/mol, this results in a final concentration of 10 μM. Concentrations between 10 and 50 μM are frequently chosen for substance screens [15]. We found that higher concentrations can lead to growth impairment and toxicity.
10. If several tester strains are used, make sure that the inoculation cultures have a similar density (OD_{600}). Adjust with liquid YES medium if necessary.
11. Seal and shake cultures promptly to avoid settling of cells.
12. In our experience, cultures of wild-type strains grow to stationary phase in 3 days. For life-shortening compounds or yeast strains with a reduced lifespan (e.g., *vma1Δ*), it may be advisable to start CLS measurement 1 day after setting up the aging cultures, because the viability drops earlier than in a wild-type strain. Conversely, some compounds can reduce the growth of yeast cells. In this case, the point of reaching stationary phase and total viability has to be determined individually and may take longer than the standard 3 days. For comparison, e.g., with control compounds/strains, this time point is designated "day 3."
13. Centrifugation of the plates is used to remove liquid drops from the seal, such that cross-contamination is avoided when the seal is removed.
14. It is advisable to make a duplicate of the transfer to a separate YES plate as a backup in case the transfer of a plate is uneven.
15. At the early stages of the aging experiment, 2 days of incubation is optimal. At later stages, cells regrow more slowly and require up to 3 days for colony formation.
16. Briefly shake the plate before placing in the incubator to avoid settling of the cells. When performing an aging experiment involving more than one microtiter plate, perform the steps with one plate at a time to avoid settling of the cells.

Acknowledgment

This work was supported by the Jürgen Manchot Foundation (to J.S.), the University of Duisburg-Essen, and the Humboldt-Universität zu Berlin.

References

1. Fontana L, Partridge L, Longo VD (2010) Extending healthy life span: from yeast to humans. Science 328:321–326
2. Kenyon CJ (2010) The genetics of ageing. Nature 464:504–512
3. Longo VD et al (2012) Replicative and chronological aging in Saccharomyces cerevisiae. Cell Metab 16:18–31
4. Roux AE et al (2010) Fission yeast and other yeasts as emergent models to unravel cellular aging in eukaryotes. J Gerontol A Biol Sci Med Sci 65:1–8
5. Mortimer RK, Johnston JR (1959) Life span of individual yeast cells. Nature 183: 1751–1752
6. Steffen KK, Kennedy BK, Kaeberlein M (2009) Measuring replicative life span in the budding yeast. J Vis Exp 28:e1209
7. Huberts DH et al (2013) Construction and use of a microfluidic dissection platform for long-term imaging of cellular processes in budding yeast. Nat Protoc 8:1019–1027
8. Zhang Y et al (2012) Single cell analysis of yeast replicative aging using a new generation of microfluidic device. PLoS One 7:e48275
9. Fabrizio P, Longo VD (2003) The chronological life span of Saccharomyces cerevisiae. Aging Cell 2:73–81
10. Murakami CJ et al (2008) A method for high-throughput quantitative analysis of yeast chronological life span. J Gerontol A Biol Sci Med Sci 63:113–121
11. Stephan J, Franke J, Ehrenhofer-Murray AE (2013) Chemical genetic screen in fission yeast reveals roles for vacuolar acidification, mitochondrial fission, and cellular GMP levels in lifespan extension. Aging Cell 12:574–583
12. Chen BR, Runge KW (2009) A new Schizosaccharomyces pombe chronological lifespan assay reveals that caloric restriction promotes efficient cell cycle exit and extends longevity. Exp Gerontol 44:493–502
13. Moreno S, Klar A, Nurse P (1991) Molecular genetic analysis of fission yeast Schizosaccharomyces pombe. Methods Enzymol 194:795–823
14. Sherman F (1991) Getting started with yeast. Methods Enzymol 194:3–21
15. Bedalov A et al (2001) Identification of a small molecule inhibitor of Sir2p. Proc Natl Acad Sci U S A 98:15113–15118

Chapter 8

Protocols for the Routine Screening of Drug Sensitivity in the Human Parasite *Trichomonas vaginalis*

Manal J. Natto, Anthonius A. Eze, and Harry P. de Koning

Abstract

Trichomonas vaginalis is a sexually transmitted protozoan parasite of humans. Treatment of trichomoniasis is almost completely dependent on the old drug metronidazole and is hampered by resistance. New drug development, like routine screening for drug resistance, has however been hampered by the lack of reliable screening protocols with sufficient throughput. Here we report on two separate in vitro protocols that use fluorescent dyes and allow for standardized drug sensitivity testing on the required scale.

Key words *Trichomonas vaginalis*, Trichomoniasis, Drug screening, High-throughput, Drug resistance, In vitro assay, Resazurin, Resorufin, Propidium iodide, Alamar Blue®

1 Introduction

With an estimated 248 million new infections per annum [1], trichomoniasis is one of the most prevalent sexually transmitted diseases. Although the symptoms of *Trichomonas vaginalis* infection are relatively mild, the pathology renders the patient more susceptible to other infections, particularly with human immunodeficiency virus (HIV) [2]. In addition, the infection is significantly associated with low birth rates and preterm delivery [3]. The drug of choice against trichomoniasis is metronidazole, with the closely related and more expensive tinidazole as sole backup [4]. Resistance to metronidazole is increasingly recognized as an important problem in treatment, especially as cross-resistance with tinidazole is not uncommon, leaving the infection untreatable [5, 6]. It is thus essential that new drugs against trichomoniasis are developed, preferably of a different chemical class than the established 5-nitroimidazoles. However, ab initio drug discovery requires high-capacity drug screening methods, using either in vivo animal models or in vitro cultures. As *T. vaginalis* is an exclusively human parasite, no animal models are available and, although culture media have been available for decades [7, 8], this did not result in viable drug screening protocols

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_8, © Springer Science+Business Media New York 2015

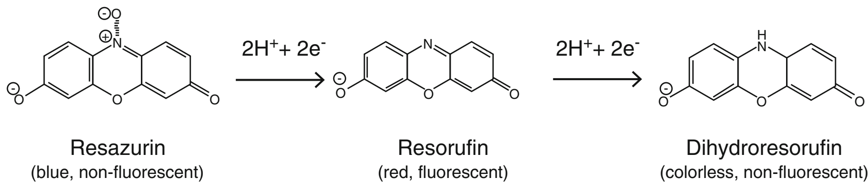


Fig. 1 Metabolism of resazurin (Alamar Blue®) by *Trichomonas vaginalis*. Resazurin is rapidly reduced to resorufin by the parasites, but is also reduced by the ascorbic acid in the medium, making it an unreliable indicator for the presence of *T. vaginalis* trophozoites. The ascorbate or a similar reducing agent is essential for trophozoite growth and thus cannot be omitted from the medium. Resorufin is stable in the medium but is rapidly further reduced by the trophozoites

and the standard procedure remained microscopic evaluation of drug-exposed cultures [9]. Although it has been suggested that the indicator dye resazurin (Alamar Blue®) could be used for the routine screening of compounds against *T. vaginalis* [10, 11], we showed recently that this dye is reduced to its red and fluorescent metabolite resorufin not just by the parasites, but also by the high levels of ascorbic acid present in the culture medium, making the readouts unreliable [12]. However, we also noticed that longer incubation of *T. vaginalis* with resazurin rendered the culture colorless, as resorufin was further reduced to dihydroresorufin (Fig. 1) and that this process, unlike the first reduction, was not performed by the medium but only by live parasites [12], fulfilling the requirements for a genuine viability indicator. Here, we describe in detail the culture of *T. vaginalis* and the protocol for the resorufin-based assay. One advantage of this protocol is that only trichomonads and possibly a few other amitochondriate protozoa such as *Giardia* spp. have the reductive potential for this reaction, thus generating a specific signal even when it is not a monoculture, as the case may be with primary clinical samples. In addition, we describe an alternative, less specific protocol suitable only for monocultures that allows a standardized fluorescent readout based on the number of parasites in the well. This second assay is based on the dye propidium iodide (PI), which becomes highly fluorescent upon binding with DNA. This assay has the advantage that it does not require incubation with live cells—rather, cells are permeabilized (we use digitonin for permeabilization) after a predetermined incubation time with drugs, allowing the PI to access the DNA; its fluorescence is proportional to the amount of DNA. Both assays were fully validated for 96-well plate formats [12] and should be adaptable to 384-well formats as well.

2 Materials

1. Modified Diamond's Medium (MDM): For 1 L, dissolve 20 g of Trypticase Peptone, 10.0 g of yeast extract, 5 g of maltose monohydrate, 1 g of L-ascorbic acid, 1 g of KCl, 1 g of $KHCO_3$, 1 g of KH_2PO_4, 0.5 g of K_2HPO_4, 0.1 g of $FeSO_4{\cdot}2H_2O$ in

850 mL of ultrapure water and adjust the pH to 6.3 with concentrated HCl. Adjust the volume to 900 mL, filter-sterilize with a 0.22 μm filter, and add 10 % heat-inactivated horse serum (100 mL) aseptically and mix. Aliquot the medium and store at −20 °C.

2. Resorufin stock solution, 500 μM: Dissolve 11.8 mg of resorufin sodium salt in 100 mL in phosphate-buffered saline (PBS), taking care to protect from light and filter-sterilize the solution before storage as aliquots at −20 °C (*see* **Notes 1–3**).
3. PI stock solution, 20 mM: Add 13.37 mg of PI to 1 mL of DMSO. Store at 4 °C.
4. Digitonin stock solution, 20 mM: Add 26.6 mg of digitonin to 1 mL of DMSO. Store at 4 °C.
5. PI/digitonin 10× solution: Add 90 μL of 20 mM PI and 200 μL of 20 mM digitonin to 19.71 mL of PBS (*see* **Note 4**). Aliquot and store at −20 °C protected from light (*see* **Note 1**).
6. 100× Test compound solutions: Prepare as DMSO solutions at 100× if solubility allows (*see* **Note 5**), and store at −20 °C. Dilute from this to 2× in MDM by adding 5 μL of 100× test compound in DMSO to 495 μL MDM, just before use.
7. Fluorescence plate reader: Determine fluorescence using an appropriate plate reader (*see* **Note 6**), at the end of each incubation period.

3 Methods

3.1 Culture Conditions and Parasite Maintenance

1. Grow *T. vaginalis* under anaerobic conditions throughout (*see* **Note 7**). Perform routine culture and strain maintenance in sterile 25 mL culture flasks that are filled completely and capped tightly to ensure anaerobic conditions. Culture at 37 °C.
2. Seed the culture flasks with 2×10^6 parasites (e.g., 1 mL of cell suspension or culture at 2×10^6 cells/mL should be added to 25 mL of fresh culture media), as determined by hemocytometer count. Passage every day (*see* **Note 8**).
3. When culturing in 96-well multi-well plates, the seeding density of *T. vaginalis* is 5×10^4 cells per well, added in 100 μL MDM from a suspension of 5×10^5 trophozoites/mL medium.
4. Seal the plates with Nescofilm or a similar product and insert into BD GasPak EZ pouches to create anaerobic conditions as this system produces CO_2 whilst absorbing O_2.
5. The wells on the outside of the plates should be filled with sterile distilled water in order to prevent damaging evaporation from the wells, leaving six rows of ten wells each for the assays (Fig. 2).

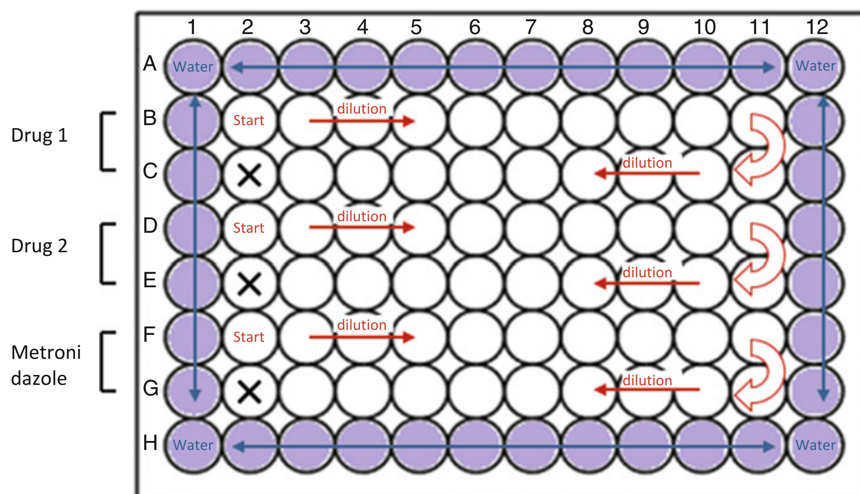


Fig. 2 Map of 96-well plates as used for either resazurin or the PI assay. The outer wells are filled with 200 μL of sterile distilled water (*shaded*). On this plate setup 200 μL of test compound or control drug, usually at 200 μM, is added to the wells labeled *Start*, and a doubling dilution is carried out over the next 18 wells (*white*). The final wells, marked with *multiplication sign*, do not receive any drug but are control for maximum growth of the *Trichomonas vaginalis* trophozoites

3.2 Setting Up a 96-Well Plate of Test Compounds and Controls

All procedures are performed at room temperature unless otherwise indicated. Perform all procedures with the culture and plates aseptically in a flow cabinet. Plates should be opaque and white for fluorescence. Our standard assay uses 19 doubling dilutions and a no-drug control (20 wells/compound) in order to generate highly accurate EC_{50} values over a broad range, and this procedure is given below, but other formats are possible and important, especially for high-throughput screening (*see* **Note 9**). The highest concentration of test compound in the assay may depend on solubility in the culture medium, availability, cost, etc. but we usually use 100 μM as the top concentration and will use that consistently in the example protocol described here. The basic layout of the plate for the 20-point assay is shown in Fig. 2.

1. Pipette 100 μL MDM to each well B3–C2 (19 wells, *see* Fig. 2). To the well marked “Start” (well B2) add 200 μL of 2× μM test compound 1 in MDM (200 μM in this example).
2. Take 100 μL of 200 μM test compound from well B2 and mix this with the 100 μL MDM in the next well (B4) by placing the tip just under the surface of the liquid and gently pipetting up and down several times (≥5×).
3. Repeat this procedure of transferring 100 μL to the next well, creating a doubling dilution series until well C3. From this well, 100 μL is discarded rather than carried over to well C2, which serves as the drug-free control. This creates a dilution

range from 200 μM in B2 to 0.76 nM in C3, with 100 μL in each well. Upon addition of 100 μL of *T. vaginalis* suspension to each well (see below) these concentrations will be halved.

4. Repeat Subheading 3.2, **steps 2–4**, for test compound 2 (wells D2–E3) and for the control drug, metronidazole (wells F2–G3) (*see* **Note 10**).

3.3 Exposure of *T. vaginalis* Trophozoites to the Different Drug Concentrations

1. Grow a 25-mL culture of *T. vaginalis* anaerobically for 24 h in MDM at 37 °C, seeded with 2×10^6 parasites (resulting in a suspension of 8×10^4 cells/mL).
2. Using a hemocytometer or similar implement, determine the cell density of the culture.
3. Dilute the cell culture to 5×10^5 trophozoites/mL with MDM.
4. Add 100 μL of this suspension to each well (i.e., wells B2–G11) (*see* Fig. 2).
5. Put the lid back on the plate and seal tightly with Parafilm® or similar.
6. Place the sealed plate in a GasPak® pouch and place in a 37 °C incubator for anaerobic culture.

3.4 Resorufin-Based Drug Sensitivity Assay

1 Remove plates from the incubator after a standardized period, usually close to 24 h (*see* **Note 11**).

2 At exactly the preset time of exposure to test compounds, add 20 μL of filter-sterilized 500 μM resorufin solution in PBS to each well, except those filled with water. Take care to minimize light exposure of the dye.

3 Place the plates back at 37 °C for 5 min and read the fluorescence in the plate reader, using a 544 nm filter for excitation and a 599 nm emission filter.

4 Analyze the data by nonlinear regression using an equation for a sigmoid curve with variable slope for the determination of EC_{50} values (50 % effective concentrations) (*see* **Note 12**).

3.5 Propidium Iodide-Based Drug Sensitivity Assay

1. This is an alternative to the resorufin-based assay, using the same plate setup as described in Subheading 3.3 (*see* **Note 13**).
2. Following incubation of the parasites with small molecules for exactly the predetermined time (usually 24 h), add 20 μL of the PI/digitonin mixture to each well.
3. The plate is placed back in the anaerobic pouch and incubated for a further 60 min at 37 °C to allow for complete permeabilization of the cells, as well as penetration and binding of the PI (*see* **Note 14**).
4. Read the fluorescence on a plate reader using 544 and 620 nm filters for excitation and emission.

5. Analyze the data by nonlinear regression using an equation for a sigmoid curve with variable slope for the determination of EC_{50} values (50 % effective concentrations) (*see* **Note 12**).

4 Notes

1. Resorufin and propidium iodide should be protected from light as much as possible. Cover the beaker in which the solution is made in aluminum foil and if possible weigh the powder out in dimmed light. Likewise, storage is in tubes covered in aluminum foil, aliquoted, and at −20 °C.
2. It is not necessary or advisable to have antibiotics in the *Trichomonas* culture as long as all solutions and media are sterile and all work is carried aseptically in a flow cabinet.
3. Keep one aliquot for immediate use at 4 °C to prevent repeated freeze-thaw cycles. It can be kept at 4 °C for 2 weeks.
4. Propidium iodide is prepared in solution together with digitonin, so that both are added together at 10× strength, reducing the amount of pipetting and increasing reproducibility. In this 10× solution the concentration of PI is 90 μM and the concentration of digitonin is 200 μM.
5. DMSO is a good solvent for many drug-like compounds. The main advantage is that 100 % DMSO solutions, like ethanol, are sterile. Making the compound up in aqueous buffer or media requires subsequent filter sterilization, with the consequent loss of volume and the potential for filtering out some of the compound in the process. This is true even if the buffer is sterile because the powder of the test compound usually is not. DMSO solutions can be aliquoted and stored at either 4 °C or −20 °C. It is essential to ensure that the final concentration of DMSO the parasites are exposed to is ≤1 % at all times.
6. We use a FLUOstar Optima (BMG Labtech) but any fluorescence plate reader with the right wavelength filters and sensitivity will do.
7. Although *T. vaginalis* is aero-tolerant, both the cultures and the test plates should be incubated anaerobically for the best and most reproducible results. However, it is not necessary to anaerobically add reagents or read the plates anaerobically.
8. Under optimized conditions *Trichomonas vaginalis* trophozoites grow very well, and rapidly. That also means that they overgrow rapidly, with deleterious effects on culture viability. At the cell density we use for seeding, it is highly advisable to passage the cells every 24 h to new culture medium and this ensures a viable log-phase culture. It is possible to seed at a lower density but we found the growth rate more variable and

the cell density after a standardized incubation time less predictable when we seeded at low density. It is advisable to check optimal growth and seeding densities for each strain.

9. The 20-point EC_{50} curve is most useful when testing a relatively small number of compounds of unknown potency against *T. vaginalis*. When used with a highest drug concentration of 100 μM it will produce well-defined sigmoid curves for activities between 1 nM and ~20 μM. However, it will not be necessary to include 19 doubling dilutions if the approximate EC_{50} values are known, and a single row of wells may suffice. Economies can also be achieved by using a fourfold dilution range over one row of wells, for instance. For high-throughput screening of libraries, a first screen could be undertaken at a single drug concentration (e.g., 10 μM) to identify a smaller compound set with a minimum potency (e.g., at least equal to metronidazole).
10. It is not necessary to have a metronidazole control on every plate—this would occupy one-third of the entire capacity even though the procedure is quite reproducible. We propose to have metronidazole included every fifth or tenth plate.
11. It is good to understand the issues underpinning the right duration for drug exposure. Some drugs may act slowly on the parasite, gradually inhibiting growth, rather than rapidly killing the cells. This may mean that in those cases a 24-h incubation is insufficient to observe the true dose-dependent action of the compound. On the other hand, a much longer incubation time risks spurious results due to overgrowth and depletion of the growth medium. We find that 24 h works well in most cases for *T. vaginalis*.
12. When calculating the EC_{50} values it is essential to use an equation with variable slope, as different drugs have much steeper dose–response relationships than others, leading to different slopes in the curve. It is also important to be aware of the limitations to extrapolation if the data does not define a complete sigmoid curve with well-defined minimum and maximum levels. It is not good practise to try to extrapolate an EC_{50} value when less than 50 % growth inhibition has occurred at the highest test compound concentration. If extrapolation seems justified by near-complete inhibition, make certain to determine the fluorescence value for complete inhibition by using an efficient positive control such as metronidazole, and enter this value into the equation; otherwise very inaccurate outcomes may result.
13. It is not possible to perform both the resorufin and propidium iodide assays on the same plates.
14. As described in reference 11, it is important to rigorously standardize the incubation time with digitonin/PI. Too short,

and the permeabilization is incomplete; too long and the fluorescence signal rapidly decreases, probably as a result of degradation of nucleic acids in the permeabilized cells. Both produce a suboptimal signal-to-noise ratio and, more importantly, a non-standardized incubation time generates potential plate-to-plate and inter-assay variation. Therefore, ensure that each individual plate is read the standard time after addition of the mixture.

Acknowledgements

This work was supported through a Research Fellowship from the Government of Saudi Arabia to M.J.N. (grant number S8306) and a Commonwealth scholarship to A.A.E.

References

1. Poole DN, McClelland RS (2013) Global epidemiology of *Trichomonas vaginalis*. Sex Transm Infect 89:418–422
2. McClelland RS et al (2007) Infection with *Trichomonas vaginalis* increases the risk of HIV-1 acquisition. J Infect Dis 195:698–702
3. Cotch MF et al (1997) *Trichomonas vaginalis* associated with low birth weight and preterm delivery. Sex Transm Dis 24:353–360
4. Bachmann LH et al (2011) *Trichomonas vaginalis* genital infections: progress and challenges. Clin Infect Dis 53(Suppl 3):S160–S172
5. Dunne RL et al (2003) Drug resistance in the sexually transmitted protozoan *Trichomonas vaginalis*. Cell Res 13:239–249
6. Miller MR, Nyirjesy P (2011) Refractory trichomoniasis in HIV-positive and HIV-negative subjects. Curr Infect Dis Rep 13: 595–603
7. Diamond LS (1957) The establishment of various trichomonads of animals and man in axenic cultures. J Parasitol 43:488–490
8. Clark CG, Diamond LS (2002) Methods for cultivation of luminal parasitic protists of clinical importance. Clin Microbiol Rev 15: 329–341
9. Giordani RB et al (2009) Anti-*Trichomonas vaginalis* activity of synthetic lipophilic diamine and amino alcohol derivatives. Biomed Pharmacother 63:613–617
10. Campos Aldrete ME et al (2005) A high-throughput colorimetric and fluorometric microassay for the evaluation of nitroimidazole derivatives anti-*Trichomonas* activity. Toxicol In Vitro 19:1045–1050
11. Duarte M et al (2009) A quantitative resazurin assay to determinate the viability of *Trichomonas vaginalis* and the cytotoxicity of organic solvents and surfactant agents. Exp Parasitol 123:195–198
12. Natto MJ et al (2012) Validation of novel fluorescence assays for the routine screening of drug susceptibilities of *Trichomonas vaginalis*. J Antimicrob Chemother 67:933–943

Chapter 9

Chemical Genetic Screens Using *Arabidopsis thaliana* Seedlings Grown on Solid Medium

Thanh Theresa Dinh and Xuemei Chen

Abstract

Genetic screening has been a powerful tool in identifying new genes in a pathway of interest (forward genetics) or attributing function to a particular gene via mutagenesis (reverse genetics). Small molecule-based chemical genetics is increasingly adapted in *Arabidopsis* research as a tool for similar purposes, i.e., to identify genes involved in certain biological processes and to dissect the biological roles of a gene. Chemical genetic screens have been successful in circumventing genetic redundancy to assign biological roles to a gene family as well as novel functions for well-known genes. Here, we describe how to screen *Arabidopsis* seedlings grown on solid medium with chemical compounds.

Key words Chemical genetics, Small molecules, Genetic screens, *Arabidopsis* seedlings, Solid medium

1 Introduction

Given its advantages, such as its small genome and plant size, high fecundity, and rapid generation time [1], *Arabidopsis* is an ideal genetic model organism. Genetic screens in *Arabidopsis* have vastly contributed to our knowledge of key regulatory genes in a diverse array of biological processes: from growth, flowering, and immunity to speciation. An example of the power of *Arabidopsis* genetics is the discovery of *LEAFY*, a gene that controls floral meristem identity [2, 3]. Homologues of *LEAFY* are found in all flowering species, and overexpression of *LEAFY* causes plants to ignore environmental cues and flower earlier [4]. This discovery in *Arabidopsis* is changing agriculture such that researchers overexpressed *LEAFY* in attempts to break the juvenile stage (to get earlier flowering) in apples [5].

Traditional genetics involves the induction of mutations in populations through mutagenesis and screening for mutants with desirable phenotypes. The most commonly used mutagenesis methods, such as ethyl methanesulfonate (EMS) and transfer (T)-DNA, have been highly successful in *Arabidopsis* genetics,

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_9, © Springer Science+Business Media New York 2015

aiding the discovery of genes and mechanisms governing numerous biological processes. However, these traditional genetic approaches have their limitations. For instance, EMS mutagenesis causes single-nucleotide changes so that phenotypes may not be observed if the mutation is in a non-imperative region of the gene. With T-DNA mutagenesis, the T-DNA has a tendency to integrate into euchromatin, thereby generating a bias for the types of genes that can be found. Further, this method can cause chromosomal rearrangements, such as inversions or deletions [6], which makes subsequent genetic analyses difficult.

In the past decade, plant researchers have begun using small molecules as a tool to interrogate biological processes. Small molecule or chemical genetic screening is not a new concept—the pharmaceutical industry and the mammalian biology field have been exploiting this tool for clinical and research purposes for decades. Chemical genetic screens exploit the vast chemical space to identify compounds that perturb a biological process, eventually leading to the discovery of genes governing the biological process. The advantages of a chemical genetic screen over a conventional genetic screen are the following: (1) it can perturb the activity of a protein through allosteric or competitive inhibition, such that the protein itself is not mutated, which allows investigations of other properties of the protein; (2) it has the ability to inhibit multiple family members functioning in the same genetic pathway, thereby circumventing the issue of genetic redundancy (a major problem with traditional mutagenesis methods); and (3) one can further probe the function of a gene by reversibly altering the chemical dosage or controlling the time of chemical exposure. Recent studies have mainly used very young seedlings grown in liquid medium for screening; however, hypoxia may be a problem when certain biological processes are being investigated, such as those involving reactive oxygen species. Further, growth in liquid medium under constant shaking is not natural for *Arabidopsis* and may in turn affect plant growth and development. Here, we describe a protocol to screen small molecules in a 96-well format using solid medium.

1.1 Primary Chemical Screening

Coming up with a design for a primary screen is the most important part of the project so that a large number of compounds can be screened efficiently. Several efficient screening methods have been reported. One involves the screening of chemicals that disturb plant proteins using a heterologous system such as yeast [7, 8]. This method was successfully used to identify pyrabactin and quinabactin, both of which are agonists for abscisic acid (ABA), a plant hormone, and both target the ABA receptor family [7, 8]. Other notable examples involved the use of a GFP-based reporter construct coupled with automated screening based on the germination and growth of free-living tobacco pollen to search for small molecules involved in endomembrane cycling; this method yielded

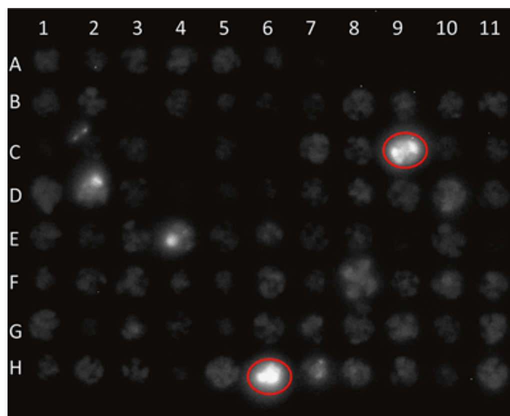


Fig. 1 An example of a primary screen using solid medium. Two *Arabidopsis* seeds from a luciferase reporter line [13] were placed into each well and seedlings were grown under continuous light. Column 1 contains all DMSO-treated seedlings (used as a negative control). Two potential hit compounds are *circled* in *red*

over 360 compounds [9, 10]! Several labs have also used etiolated seedlings in their studies that resulted in the identification of L-kynurenine as a competitive inhibitor of TAA1/TAR activity in ethylene-directed auxin biosynthesis and root growth [11] or a synthetic elicitor involved in disease resistance [12]. All of the aforementioned studies have used a robot equipped with an automated microplate hotel and a magnetically loaded 384-pin tool to automate the primary screen. However, this setup is cost prohibitive and not feasible in screens using older plants (the pin would stab the plant). Thus here, we will elaborate on a method that does not require a sophisticated robotic system and can be used to screen older plants. Although we have mainly used this system to identify small molecules that release DNA methylation [13, 14], it could easily be adapted to other biological questions. The primary screen is done with solid medium in a 96-well plate format (Fig. 1), a method that is more native to plant growth than liquid medium.

1.2 Generating a Reporter Line

Given the large chemical space that one desires to cover in a chemical genetic screen, it is crucial to choose an easily "scorable" phenotype representing the biological process of interest as the basis of the screen. Although morphological or physiological phenotypes can serve as the basis of the screen, it may be more desirable to screen against a reporter line in which the reporter gene's activity reflects the status of the biological process in question. One of the most important aspects of designing a chemical genetic screen is choosing an optimal reporter line that will streamline the initial screening, which is the most arduous part of the chemical genetic screen. We do not discuss how to design a reporter transgene, such as the regulatory elements to be appended to the reporter gene,

as the design varies depending on the biological process of interest. Here, we highlight the different commonly used reporter genes, green fluorescent protein (GFP) and variants, luciferase (LUC), and β-glucuronidase (GUS), and the screening facilities and efforts associated with each. A GFP-based chemical screen entails the extensive use of a fluorescence dissecting or confocal microscope. For LUC-based chemical screens, an LUC imaging dock and camera are needed, and the plants need to be treated with the LUC substrate a few minutes prior to imaging. GUS-based chemical screens have the advantage of requiring no special equipment, but are extremely laborious; GUS staining, the process of "visualizing" GUS activity in plants, entails tissue fixation, hours of incubation of plants with the substrate of the enzyme, and post-reaction washes. We highly recommend the LUC reporter, as chemical screening with this reporter entails the least labor and time as compared to the other two reporters. In Subheadings 2 and 3, we will mainly discuss LUC-based chemical screening. As a general note of caution with reporter transgenes, we recommend that the reporter construct be transformed into an *RNA-dependent RNA polymerase 6* (*rdr6*) mutant background to prevent posttranscriptional silencing [15–17] and multiple transgenic lines be screened to identify one in which the transgene has inserted into a single locus (this simplifies subsequent genetic analyses). Moreover, we recommend bulking up 2–3 mL of seeds prior to the screening and continually analyzing the genetic material during the screening process to ensure that the transgene remains expressed.

1.3 Secondary Chemical Screening and Further Confirmation

After obtaining "hit" compounds, additional assays with the compound(s) need to be done to verify the hits. At this point, one should obtain fresh compounds from the commercial vendor, verify the hits again by re-conducting the original assay, and perform additional assays such as dose and time series. The concentration series will discern the optimal concentration for the chemical. If a reporter gene such as LUC is used for the screen, LUC activity can be quantified to determine the IC_{50} of the chemical compound. The time series will determine the minimal duration of treatment necessary to observe the effects of the chemical, which can be informative. For instance, if changes in LUC activity are not seen for 2 days, cell division is probably required for the chemical to take effect. Or, if changes in LUC activity are quickly observed, the chemical must be rapidly uptaken and the effects of the chemical do not require cell division. In addition, it is important to determine that the changes in LUC activity caused by the chemical are attributed to changes in transgene expression rather than the chemical affecting the LUC protein itself.

1.4 Target Identification

Though it will not be discussed in detail, we should note that the two most common methods of identifying the target gene of a hit compound are the following: (1) traditional EMS mutagenesis

(screen for mutants that are resistant to the compound) followed by map-based cloning or (2) tag the compound, "pull down" its target, and use proteomics to identify the target protein. Although map-based cloning is laborious (a detailed protocol describing that method can be found here [18]), high-throughput sequencing technologies have greatly facilitated this process.

2 Materials

2.1 Primary Screening

1. 96-Well flat-bottom plates with lids.
2. Heat and stir block.
3. 500 mL or 1 L glass deep-dish cylindrical plate (autoclavable) (*see* **Note 1**).
4. Magnetic stir bar (autoclavable) (*see* **Note 2**).
5. 1,000 μL multichannel pipette (*see* **Note 3**).
6. 0.5× MS medium: Dissolve 2.165 g of Murashige and Skoog Basal Salt Mixture (MS) in deionized water, make the final volume to 1 L with deionized water, and autoclave.
7. 0.1 % agarose gel: Add 0.1 g of agarose to 100 mL of autoclaved H_2O. Microwave to dissolve the agar and leave to cool at room temperature prior to use (*see* **Note 4**).
8. 0.5× MS medium in 0.8 % agar: Dissolve 2.165 g of MS medium in ~900 mL of deionized water, and add deionized water to make up volume to 1 L. Add 8 g of agar to the solution and autoclave (*see* **Note 5**).
9. Cut 200 μL tips: Cut ½ cm off the tip with a sterile razor blade in a sterile environment.
10. 20 μL pipette.
11. Micropore tape (*see* **Note 6**).
12. Luciferin (the substrate of LUC): Prepare a solution of 1 mM luciferin in 0.01 % Triton X-100 (*see* **Note** 7).
13. Nasal spray bottle (*see* **Note 8**).
14. Library of chemical compounds (*see* **Note 9**).
15. Bleach.
16. 37 % HCl.
17. Rectangular container with an airtight lid.
18. Photonics Onyx Luminescence Dark Box.
19. Roper Pixis 1024B camera.
20. Compatible computer for processing.
21. Winview software.

2.2 Secondary and Additional Screening

1. 6-Well plates.
2. Direct-zol™ RNA MiniPrep w/TRI-Reagent® Kit (Zymo Research) (*see* **Note 10**).
3. iScript™ Advanced cDNA Synthesis Kit for RT-qPCR (Bio-Rad) (*see* **Note 11**).
4. Primers for *UBIQUITIN* (*UBQ*; internal loading control) and *LUC* (*LUC*) (*see* **Note 12**).
5. Hard-Shell® Low-Profile Thin-Wall 96-Well Skirted PCR Plates (Bio-Rad).
6. Microseal® "C" Optical Seals (Bio-Rad).
7. CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad).
8. Optional: Luciferase Activity System (Promega).

3 Methods

3.1 Primary Screening

Contamination is a major hindrance of the screening process. Therefore, it is highly advisable that all tools be sterile and the following steps should be performed in a sterile environment.

1. Place 500 μL of seeds containing the reporter transgene in a 1.5 mL Eppendorf tube and label the Eppendorf tube using a pencil with the seed genotype (*see* **Note 13**).
2. Leave the lids open and place the tubes in a rack.
3. Place the rack in a rectangular container.
4. Add 47 mL of bleach into a glass beaker and place it in the box along with the rack. The height of the beaker should be no more than the 80 % of the height of the container.
5. SLOWLY add 3 mL of HCl into the beaker containing 47 mL of bleach and place the mixture into the rectangular container (*see* **Note 14**).
6. Close the container with an airtight lid and let it sit for 6 h. This incubation time allows the seeds to be properly sterilized, thereby decreasing contamination.
7. After 6 h, remove the lid in a sterile environment.
8. Remove the rack containing the seeds and close lids of the tubes (containing seeds).
9. Discard the HCl and bleach solution accordingly.
10. Relabel the tubes containing the seeds with its genotype for proper identification using a permanent marker.
11. Add 1 mL of agar solution to one Eppendorf tube containing the sterilized seeds.
12. Mix thoroughly and let the seeds sit for at least 1 h (*see* **Note 15**).

13. Turn on a heat block to at least 40 °C but no more than 60 °C (*see* **Note 16**).
14. Place an autoclaved glass dish with a stir bar on the heat block.
15. Pour autoclaved 0.5× MS medium + agar into the glass dish.
16. Turn on the "stir" setting to low/medium.
17. Pipet 220 μL of 0.5× MS medium + agar into each well of a 96-well plate using a multichannel pipette.
18. Replace the lid and let the plate sit until the medium solidifies (*see* **Note 17**).
19. Repeat **steps 17** and **18** of Subheading 3.1 49 more times to generate a total of fifty 96-well plates (*see* **Note 18**).
20. Leave the plates in the sterile environment while the medium cools completely, usually about 30 min (*see* **Note 19**).
21. Using a 20 μL pipette and cut tip, pipet two seeds into each well (*see* **Note 20**).
22. Seal the plates with micropore tape.
23. Store the plates at 4 °C for 2 days to stratify the seeds.
24. Transfer the plates to a growth chamber with continuous light and incubate for 7 days (*see* **Note 21**).
25. On the seventh day, transfer the plates to a sterile environment.
26. Remove the micropore tape.
27. Add 100 μL of 0.5× MS medium (no agar) to the wells (*see* **Note 22**).
28. Add chemicals to a final concentration of 10–25 μM in each well using a multichannel pipette (*see* **Note 23**).
29. Label the plates AND lids with the same chemical plate number (*see* **Note 24**).
30. Seal the plates with micropore tape.
31. Place the plates in the growth chamber for 3 days.
32. On the third day, remove the micropore tape from the plates. At this point, sterility is no longer an issue.
33. Bring the plates into the LUC imaging room, which is often dark.
34. Focus the Roper Pixis camera prior to adding LUC substrate.
35. Turn off the light, and using only the computer as light source, turn the nasal spray bottle upside down and spray the plants in all the plates with the luciferin substrate solution and leave them in the dark (*see* **Note 25**).
36. After 5 min, remove one plate from the pile, remove its lid, and place it in the dark box.
37. In the Winview program, set the capture time to 5 min.
38. Save image for later analysis.

39. Process the rest of the plates.
40. In the meantime, while the next plate is being imaged (since the capture time is 5 min), with a permanent marker, make note of all the "hits" (*see* **Note 26**). An example of a plate with a "hit" is shown in Fig. 1.
41. Subject all the hits to another round of screening (repeat Subheading 3.1 with hits by "cherry picking") (*see* **Note 27**).
42. If a hit can be repeated, order the compound from the manufacturer and proceed to secondary screening in Subheading 3.2.

3.2 Secondary Screening and Additional Assays

3.2.1 Secondary Screening: Concentration and Time Series Analyses

1. Dilute the purchased hit compound according to the manufacturer's instructions (*see* **Note 28**).
2. Retest the compound for activity by repeating the assay (Subheading 3.1) with the newly obtained chemical (*see* **Note 29**).
3. If the chemical is still positive, it is ready to be subjected to concentration and time series analyses.
4. Prepare seeds in a 96-well format identical to Subheading 3.1.
5. Choose a minimum of five concentrations to test. For instance, if the primary screen was performed at 25 μM, perform the secondary screen with 0 μM (usually DMSO alone; column 1), 0.2 μM (column 2), 1 μM (column 3), 5 μM (column 4), 10 μM (column 5), 15 μM (column 6), 25 μM (column 7), and 50 μM (column 8). For the time series analysis, add the chemicals to the longest time point first (reverse order) so that after the shortest time point (2 h), plates can be imaged with all the time points present on one plate. Row 1 = 2 h; row 2 = 4 h; row 3 = 8 h; row 4 = 16 h; row 5 = 24 h; row 6 = 2 days; row 7 = 3 days (*see* **Note 30**). Please *see* Fig. 2a for an example.
6. Seal and place the plate(s) (each plate should contain a different chemical) into the growth chamber.
7. Assay the plate as directed in **steps 34–40** from Subheading 3.1. An example can be seen in Fig. 2a.
8. *Optional*: Quantify LUC activity using a luciferase activity measurement system.

3.2.2 Additional Assays: Checking Transcript Levels of the Reporter Construct

1. Prepare and autoclave 0.5× MS medium + agar (~250 mL/hit compound).
2. In a sterile environment, add 1–2 mL of medium into each well of a 6-well plate.
3. Once the medium has solidified, sow ~10 seeds into each well (*see* **Note 31**).
4. Seal the plate with micropore tape.
5. Store the plate at 4 °C for 2 days.

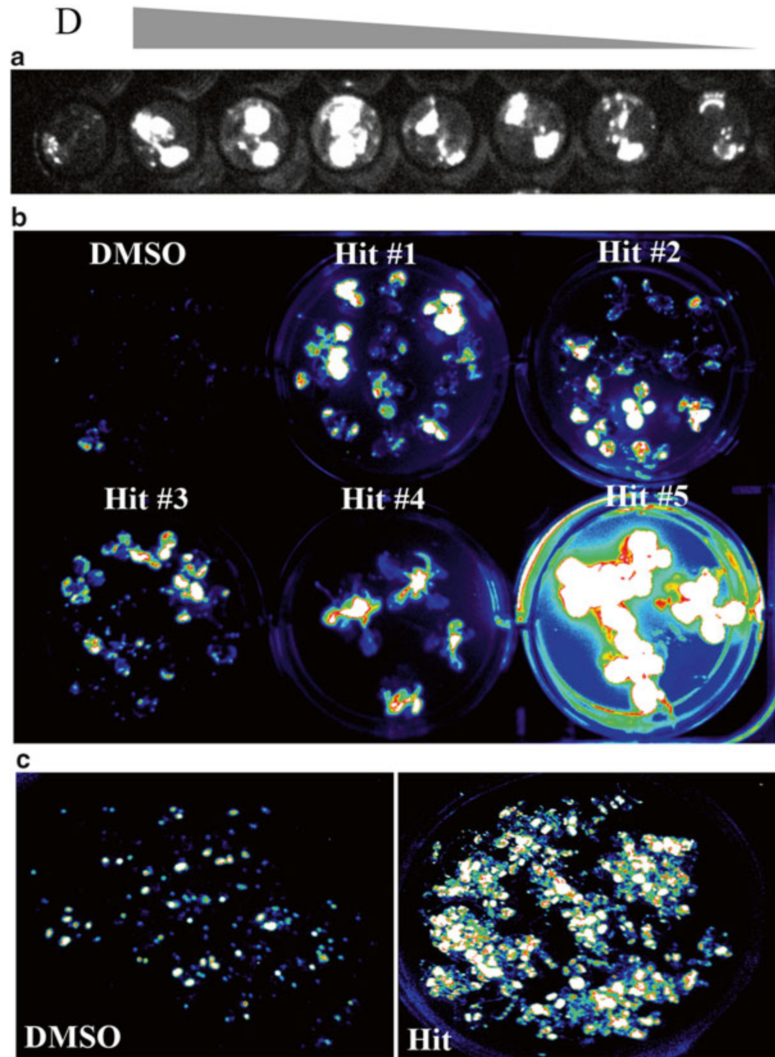


Fig. 2 Examples of secondary screens using solid medium. (**a**) The "hit" chemical was added at different concentrations in each well to ascertain the optimal chemical concentration. The *bright spots* represent seedlings with high luciferase activity. D = DMSO control. The *gray triangle* indicates decreasing amount of the chemical added. This is an example of one time point. Different time points of chemical treatment can be performed by increasing the number of rows. (**b**) An example of a 6-well plate layout. Not all seedlings had the same chemical uptake. Only select those with even, high activity for subsequent analysis. (**c**) An example how not to seed the plates. Plants on two 100 mm petri plates are shown. The seeds were not sown evenly spaced apart, so the LUC activity of the seedlings is very uneven (*right* "hit" panel)

6. Transfer the plate to a growth chamber with continuous light.
7. On the seventh day, add 200 μL of sterile 0.5× MS medium to each well.
8. Add DMSO or chemical to each well (*see* **Note 32**).
9. Transfer the plate back to the chamber and incubate for the optimal time determined in Subheading 3.2.1.
10. Assay the plate to check LUC activities (*see* **Note 33**). An example can be seen in Fig. 2b.
11. Add liquid nitrogen to an autoclaved mortar and pestle (*see* **Note 34**).
12. While the mortar and pestle are cooling down, remove seedlings from the plate.
13. Cut and remove the roots with a sterile razor blade.
14. Place the aerial tissue (minus the roots) in a mortar and pestle containing liquid nitrogen.
15. Grind the tissue until it turns into a pale green powder.
16. Transfer the ground tissue to an RNase-free 1.5 mL Eppendorf tube.
17. Add 100 μL of Trizol (*see* **Note 35**).
18. Follow the Direct-zol™ RNA MiniPrep with TRI-Reagent® Kit instructions to obtain total RNA (*see* **Note 36**).
19. Once total RNA is obtained, DNase-treated, and purified, proceed to cDNA synthesis with the iScript™ Advanced cDNA Synthesis Kit for RT-qPCR (*see* **Note 37**).
20. Perform Real-Time RT-PCR. In a 20 μL total reaction volume, add 2 μL of cDNA, 6 μL of molecular grade H_2O, 10 μL of Syber Green Master Mix, and 2 μL of primer (*UBQ* or *LUC*) into a Hard-Shell® Low-Profile Thin-Wall 96-Well Skirted PCR plate (*see* **Note 38**).
21. Seal the plate with a Microseal® "C" Optical Seal.
22. Perform triplicates of each sample using CFX96 Touch™ Real-Time PCR Detection System.

3.2.3 Additional Assays: Delineating Function in Biological Pathways of Interest

1. If mutants in genes controlling the biological process of interest exist, cross the LUC reporter gene into these mutants and perform the chemical assay as in Subheading 3.2.1 in the mutant backgrounds. This may reveal at which step in the pathway the chemical exerts its effect.
2. Observe results by testing LUC activity and comparing it to chemical-treated wild-type seedlings (*see* **Note 39**).
3. If the chemical has a previously characterized cellular target, order mutants through the Arabidopsis Stock Center. If the target is not known, prepare for target identification experiments.

4 Notes

1. One will use this glass dish to keep the medium warm (it will be placed on a heat block with a magnetic stir bar) while pipetting the medium into each well. Therefore, make sure to use a glass dish that has a diameter greater than the width of a multichannel pipette.
2. Autoclave the magnetic stir bar with the glass dish.
3. Since this is done by hand, we suggest purchasing a 12-channel (rather than 8-channel) pipette so that the pipetting only has to be done 8 rather than 12 times per plate.
4. Suspending the seeds in 0.1 % agar will allow for better dispersal of the seeds.
5. Traditional MS medium contains sucrose; however, this protocol eliminated sucrose to help prevent fungal contamination as the plates will be handled often. If wells are contaminated, it will slow down the screening process.
6. Micropore tape is a breathable paper tape that allows oxygen and carbon dioxide flow but keeps the plate free from airborne bacteria and contaminants. It also helps keep the plate environment moist, which aids plant growth.
7. 50 mL of substrate solution is normally prepared in a 50 mL Falcon tube, which is wrapped in aluminum foil (the substrate is light sensitive) and stored at 4 °C. The solution is stable at 4 °C for 2 weeks.
8. Purchase nasal spray from a local drugstore, dispose of the contents, clean thoroughly, and add the LUC substrate. Store solution at 4 °C.
9. Prepare the chemical compounds according to the manufacturer's instructions; they are usually already available in a plate format. Make a master stock solution set (usually 10 mM, but follow the manufacturer's suggestions) and a working set. Be careful of pipetting error. The master set should be rarely used. Store the master and working sets at −20 °C. Be careful with the master set—DO NOT thaw repeatedly as some compounds are not stable and may degrade over time. Remember to either order chemicals in a 96-well plate format or aliquot to a 96-well format as this will facilitate screening. Also, add DMSO to columns 1 AND 12. This will serve as controls or can be used as a substitute should some wells become contaminated.
10. Make sure to follow the manufacturer's instructions and perform the DNase I (RNase-free) treatment in the column. The kit does not have to be used. If there is already a setup in the lab for RNA purification and generating cDNA, it is acceptable to use a system one is already comfortable with.

11. Use the iQ™ SYBR® Green Supermix for the RT-qPCR in the absence of a preferred cDNA synthesis kit.
12. The primer sequences are as follows: UBQ-N: 5′-GGTGCTAAGAAGAGGAAGAAT-3′; UBQ-C: 5′-CTCCTTCTTTCTGGTAAACGT-3′; LUC-F: 5′-CTCCCCTCTCTAAGGAAGTCG-3′; LUC-R: 5′-CCAGAATGTAGCCATCCATC-3′.
13. There are other methods of seed sterilization such as incubation in 30 % bleach followed by a series of washes with dH_2O; however, the seeds must be used right away. With this method, several tubes of seeds (as well as different genotypes) can be sterilized at the same time and if kept sterile can be stored for a maximum of 2 months. Seed sterilization can be done prior to plate preparation. Also, use a pencil rather than marker to label the tubes as the ink will disappear during the sterilization process. This is especially important when sterilizing more than one genotype of seeds at the same time.
14. Be careful when adding HCl: it is extremely corrosive and dangerous. The solution will bubble so make sure to add the HCl slowly. Perform this step in a fume hood.
15. Letting the seeds sit in the agarose gel for at least an hour allows them to soak up water and expand, thereby making it easier to sow the seeds in the well.
16. Having the heat block at temperatures greater than 60 °C may cause the agar to burn and affect plant growth.
17. If pouring all the plates on one day and then planting on another day, wrap the plates in Saran wrap and store them at 4 °C. However, it is better to stagger the plating and screening over several days as it is less demanding physically. Also, the plates should only be stored at 4 °C for several days to prevent contamination.
18. This is the most that has been done in 1 day in our lab. Remember, it takes 10 min to process each plate (imaging and compiling the data), so processing 50 plates takes $10 \times 50 = 500$ min, or ~8.3 h.
19. Make sure that the plates are COMPLETELY cool before adding seeds. If the medium is still hot, the seeds may not germinate.
20. This is the most laborious part of the initial screening. Another technique to aliquot the seeds involves the use of a disposable transfer pipette.
21. This time period may vary depending on the growth conditions and is dependent on the user's needs. For example, if looking for genes involved in circadian rhythm, having the plants grown under continuous light would probably not be a good idea.

Chemicals should be added when the first two true leaves begin to emerge (such that only their tips can be seen). No matter what days are chosen, keep this variable constant throughout the entire experiment.

22. Apply the solution to the side of the well; avoid touching the plants.
23. Follow the manufacturer's recommendations for the screening concentration. This is calculated as the molar concentration of the chemical using the volume of the solid medium (220 μL), as the chemical should soak into the medium to be uptaken by the plant.
24. This is very important because if a plate is mislabeled, it will be difficult to identify the correct compound later. Also, label both lid and plate as the lid has to be removed during LUC imaging.
25. Luciferin is light sensitive, so it is best to have everything in the dark. Ensure that the leaves have been fully sprayed.
26. The entire leaf should have a uniform alteration of LUC activity. For instance, if screening for compounds that release LUC activity, the entire plant should display LUC activity whereas if screening for compounds that inhibit LUC activity, the entire plant should have low LUC activity.
27. Cherry picking is the process of going back to the original working set of compounds and removing an aliquot for retesting. Retest with the same concentration.
28. Chemicals are usually dissolved in DMSO; make sure to read the manufacturer's instruction PRIOR to dissolving the compounds as some chemicals may need to be dissolved in other organic solvents, such as methanol or dimethylformamide. Aliquot and store in different tubes so they do not have to be thawed often. Dissolve the chemical compound in as high of a concentration as possible as DMSO is toxic to plants, so the less amount added the better. Be careful of pipetting error. Usually, we make a 100 mM or 1 M stock solution. Once the compound is completely dissolved in DMSO, we further dilute it with water (right before we perform the experiment). However, it must be confirmed that the compound does not precipitate out of solution by leaving it in the tube with water for several days (the same length as the assay).
29. This is important as the chemical may have degraded or been altered in the primary screen.
30. Add DMSO to the same amount as the highest volume of chemical added. For instance if the highest volume is 3 μL, add 3 μL of DMSO.

31. Sow the seeds evenly to ensure even uptake of the chemical. Do not plant seeds densely (as in Fig. 2c), although the number of seeds planted can be more or less than 10.
32. Optimal concentrations are determined in Subheading 3.2.1.
33. It is important to obtain a uniform response of all plants in that well. If changes in LUC activity do not appear to occur evenly in all plants or if plants display hypoxia symptoms, try adding the chemical directly to the 0.5× MS medium + agar after it has cooled down. To test whether the medium is cool, touch the bottle against the inner wrist. If it burns, do not add chemical. After the medium has cooled down, add the chemical into the 6-well plate. In this case, grow the seedlings on a separate petri dish first and then transfer the seedlings to the 6-well plate for the given time. Assay LUC activity to check the activity distribution among plants in the same well. As in Fig. 2b, the chemical may not be uniformly taken up by the plants. Therefore, only select plants with increased luciferase activity for subsequent analyses. 100 or 150 mm plates can also be used for other analyses that require more tissue. However, be careful to sow the seeds such that they are evenly spaced (so the plants have room to grow and are not crowded; otherwise, the chemical will not be evenly available to each plant). *See* Fig. 2c as example of what *not* to do.
34. Use a different mortar and pestle for each sample.
35. The solution should turn pinkish-brown. If greenish-brown, add more Trizol. If pink, that means too much Trizol has been added.
36. If the lab already has a setup for RNA isolation, the kit is not necessary. Remember to perform DNase I treatment of total RNA.
37. It is not necessary to use the kit.
38. Make a master mix of the cDNA + H_2O in one tube and another master mix containing Syber Green Master Mix and primer in another tube. Aliquot the cDNA + H_2O first to each well, then Syber Green Master Mix, and primer. This helps to decrease pipetting error. Also, after adding the cDNA + H_2O mixture, turn the plate 180°, and then add the Syber Green mix. Change the tip and check the pipette after every three technical replicates.
39. If the biological pathway of interest has been implicated in floral development, chemicals can be applied to inflorescences and followed by phenotypic observation. Add chemical to 1 mL of water and 1 μL of Silwet. Vortex thoroughly. Using a 20 or 200 μL pipette, apply the solution onto an inflorescence bud drop-by-drop every day for 10 days. Observe the phenotype.

Acknowledgement

Research in the Chen lab is supported by funds from Howard Hughes Medical Institute, Gordon and Betty Moore Foundation (GBMF3046), National Institutes of Health (GM061146), and National Institutes of Food and Agriculture (2010-04209). T.T.D. was supported by the National Science Foundation ChemGen IGERT program (DGE0504249) and NIH NIAID, Molecular and Cellular Immunobiology (5T32 AI07290) fellowship.

References

1. Meyerowitz EM (1989) Arabidopsis, a useful weed. Cell 56:263–269
2. Schultz EA, Haughn GW (1991) LEAFY, a homeotic gene that regulates inflorescence development in Arabidopsis. Plant Cell 3: 771–781
3. Weigel D et al (1992) LEAFY controls floral meristem identity in Arabidopsis. Cell 69: 843–859
4. Weigel D, Nilsson O (1995) A developmental switch sufficient for flower initiation in diverse plants. Nature 377:495–500
5. Flachowsky H et al (2010) Overexpression of LEAFY in apple leads to a columnar phenotype with shorter internodes. Planta 231:251–263
6. Radhamony RN, Prasad AM, Srinivasan R (2005) T-DNA insertional mutagenesis in Arabidopsis: a tool for functional genomics. Electron J Biotechnol 8(1). http://www.ejbiotechnology.info/content/vol8/issue1/full/4/
7. Okamoto M et al (2013) Activation of dimeric ABA receptors elicits guard cell closure, ABA-regulated gene expression, and drought tolerance. Proc Natl Acad Sci U S A 110: 12132–12137
8. Park SY et al (2009) Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. Science 324:1068–1071
9. Robert S et al (2008) Endosidin1 defines a compartment involved in endocytosis of the brassinosteroid receptor BRI1 and the auxin transporters PIN2 and AUX1. Proc Natl Acad Sci U S A 105:8464–8469
10. Drakakaki G et al (2011) Clusters of bioactive compounds target dynamic endomembrane networks in vivo. Proc Natl Acad Sci U S A 108:17850–17855
11. He W et al (2011) A small-molecule screen identifies L-kynurenine as a competitive inhibitor of TAA1/TAR activity in ethylene-directed auxin biosynthesis and root growth in Arabidopsis. Plant Cell 23:3944–3960
12. Knoth C et al (2009) The synthetic elicitor 3,5-dichloroanthranilic acid induces NPR1-dependent and NPR1-independent mechanisms of disease resistance in Arabidopsis. Plant Physiol 150:333–347
13. Dinh TT et al (2013) Generation of a luciferase-based reporter for CHH and CG DNA methylation in *Arabidopsis thaliana*. Silence 4:1
14. Dinh TT, Gao L, Liu X, Chen X (2014) DNA topoisomerase IA promotes transcriptional silencing of transposable elements through DNA methylation and histone lysine 9 dimethylation in Arabidopsis. PLoS Genet 10: e1004446
15. Dalmay T et al (2000) An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell 101:543–553
16. Mourrain P et al (2000) Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell 101:533–542
17. Peragine A et al (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes Dev 18: 2368–2379
18. Dinh TT et al (2014) Genetic screens for floral mutants in *Arabidopsis thaliana*: enhancers and suppressors. Methods Mol Biol 1110: 127–156

Chapter 10

Small-Molecule Screening Using *Drosophila* Models of Human Neurological Disorders

Mickael Poidevin, Feiran Zhang, and Peng Jin

Abstract

Within the last decade, *Drosophila* has emerged as a premiere model system for the study of human neurodegenerative diseases, due to the realization that flies and humans share many structurally and functionally related gene families. Development of such disease models in the fly allows genetic approaches to be applied to address specific hypotheses concerning disease progression and to test candidate modifier genes. More recently these fly models have also been used for drug discovery. Here, we describe how to utilize the existing fruit fly models of human neurological disorders to identify small-molecule leads that could potentially be further developed for therapeutic use.

Key words *Drosophila melanogaster*, Fruit fly, Small molecules, Chemical screening, Human diseases

1 Introduction

Many features of fruit fly make it an attractive model to study basic biology. With a rapid life cycle (10 days at 25 °C), fruit fly has four distinct developmental stages: embryo, larva, pupa, and adult. It is estimated that 100,000 neurons are present in fly brain. These neurons form complex circuits and neuropil that mediate multifarious and complicated behavior such as flight navigation, aggression, grooming, feeding, learning and memory, sleep, and circadian rhythms. It is noteworthy that many drugs acting on the mammalian central nervous system have the same effects on *Drosophila* brain [1–8]. *Drosophila* has been a powerful model system due to the ease of manipulating the expression and function of its genes. Genetic tools developed in fly provide quick and easy ways to generate human disease models by mutation, expression, inactivation, or misexpression of ortholog genes in fly [9–13]. These advantages of fruit fly also make it a valuable model for primary high-throughput drug screening (HTS) as well as a quick validation platform for traditional HTS (Fig. 1). The use of fruit fly in

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_10, © Springer Science+Business Media New York 2015

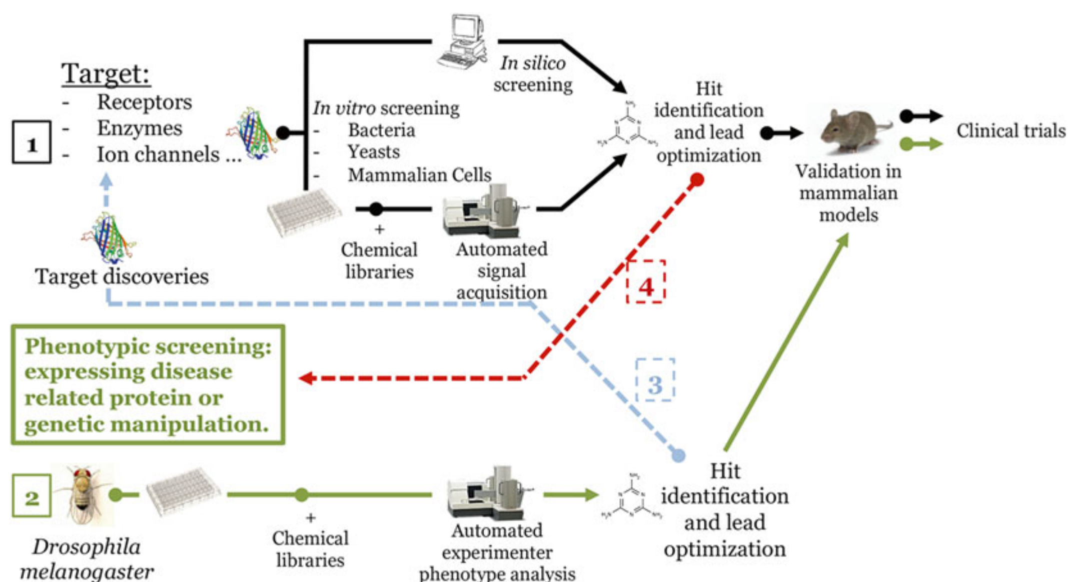


Fig. 1 *Drosophila melanogaster* in high-throughput screening (HTS) of chemical libraries. (*1, black solid lines*) An overview of the different steps in a target-based drug discovery process, including traditional HTS. Without prior knowledge of the therapeutic targets, this type of screening cannot be performed. (*2, green solid line*) Another approach is the phenotypic screening of chemicals with small model animals like *D. melanogaster*. (*3, blue dash line*) Identification of hits in *D. melanogaster* may facilitate the discovery of new therapeutic targets and the elucidation of the molecular mechanisms of diseases. (*4, red dash line*) Hits from target-based HTS can be quickly analyzed in *D. melanogaster* models, before further validation in mammalian models

drug screening will reduce the cost, accelerate lead discovery, and facilitate the identification of new therapeutic targets for human diseases.

Drosophila has been used to validate post-HTS hits. In an in vitro study, a chemical library containing 2,800 small molecules was screened against the aggregation of polyglutamine proteins in cultured cells [14]. The researchers identified 740 small molecules that could inhibit the protein aggregation. One of the best hits (Y-27632) has been further analyzed in *Drosophila* and it displayed protective effects against eye degeneration in a spinal and bulbar muscular atrophy fly model. Applying this approach, additional small molecules have been identified for further therapeutic development [15].

Drosophila could also be directly used to identify small molecules that could modify the phenotypes associated with human diseases. The first example was the small-molecule screening against the fruit fly model of fragile X syndrome [16]. The discovery that *dFmr1*-deficient *Drosophila* will die when they are reared on food containing increased levels of glutamate is consistent with the theory that loss of fragile X mental retardation protein (FMRP) disrupts the regulation of glutamate signaling. Two thousand compounds were screened against this lethal phenotype, and nine compounds were identified for their abilities to rescue the lethality,

including three that implicate the GABAergic inhibitory pathway. These discoveries led to the development of a GABA antagonist for the therapeutic intervention of fragile X syndrome.

The fragile X mental retardation 1 (*FMR1*) gene contains a highly polymorphic CGG repeat in the 5′-untranslated region [17]. Whereas normal individuals generally possess 5–54 repeats, individuals with more than 200 CGG repeats, referred as the full mutation, develop fragile X syndrome (FXS) [18]. Premutation alleles, defined as 55–200 CGG repeats of the *FMR1* gene, are known to contribute to the FXS phenotype through genetic instability, as the copy number of CGG repeats in these alleles can expand to the full mutation during germline transmission [19]. Within the last decade, fragile X-associated tremor/ataxia syndrome (FXTAS), a late-onset neurodegenerative disorder distinct from the neurodevelopmental disorder, FXS, has been recognized mainly among many male premutation carriers in or beyond their fifth decade of life [20]. The most common clinical symptom of FXTAS is a progressive tremor with ataxia. So far there is no effective therapeutic intervention for FXTAS. Previously we developed a model of FXTAS using *Drosophila* expressing the *FMR1* untranslated-CGG repeats 5′ to the EGFP coding sequence and demonstrated that premutation-length CGG RNA (rCGG) repeats are toxic by themselves and sufficient to cause neurodegeneration [21].

We discovered that flies with modest expression of fragile X premutation $rCGG_{90}$ repeats exclusively in the neurons do not reach adulthood. Lethality occurs primarily during embryonic development before larval formation. Taking advantage of this lethality, we designed a high-throughput method to screen chemical libraries for small-molecule modulators of FXTAS. Here we describe a generic protocol based on our experience in FXTAS *Drosophila* model. This chapter describes initial screening of a small-molecule library in a life span or death rescue assay to identify initial hits. Initial hits are then confirmed through additional validation assays such as locomotion behavioral assay. Similar screenings have substantial potential for the development of therapeutic agents for other human neurological disorders.

2 Materials

Prepare all solutions, media, and materials 1 day before use. Prepare and store all reagents at room temperature (unless indicated otherwise). Diligently follow all waste disposal regulations when disposing waste materials and flies.

1. Jazz-Mix *Drosophila* food: In a bucket, mix 56.7 g of Jazz-Mix with 300 mL of tap water. Mix thoroughly with a blender. Autoclave the Jazz-Mix food for 18 min using the liquid cycle with no drying and a total cycle time of 45 min.

Wait for 20 min before taking the food out of the autoclave. Allow it to cool down at room temp with occasional mixing with the blender every 30 min for about 1.5 h until the temperature reaches 50 °C. This volume represents enough food to prepare 100 vials.

2. Green food dye (Fast Green FCF: Green No. 3, Sigma Aldrich): In a 1.5 mL Eppendorf tube, mix 0.5 g of Green food dye with 1 mL of water thoroughly.
3. Small-molecule library: Dilute a fraction of the initial drug library in 96-well plates to 1 mM in DMSO. For the specific study described below, the Spectrum Collection from MicroSource Discovery Systems was used.
4. Sterile plastic culture tubes.
5. Fly strain mimicking human neurological disease (*see* **Note 1**).
6. Water bath.
7. Pyrex glass tubes.
8. Plastic activity tube caps (Trikinetics Inc.).
9. Fly pad (Flystuff).
10. CO_2.
11. Fly brush (Flystuff).
12. Cotton balls.
13. *Drosophila* activity monitor (DAM) (Trikinetics Inc.).
14. Environmental incubator with temperature and diurnal control (Tritech Research, Inc.).
15. DAM System data collection and scan software (Trikinetics Inc.).
16. Microsoft Excel.

3 Methods

Culture all flies on standard medium at 25 °C in 50–70 % humidity and maintain with chromosome balancer which carries a heat shock-inducible apoptotic gene such as hs-*hid* (*see* **Note 2**). It is necessary to amplify fly stocks before screening. Twenty to thirty vials flipped out every 3 days provide enough flies to screen 300 drugs per week. It is noteworthy that different types of screenings can be performed on many phenotypes (Table 1) (*see* **Note 3**). Here we describe a lethality rescue screening which is the most frequently used when diseases reduce life span or induce death.

3.1 Primary Lethality Screening

1. Prepare 100 vials with drugs during the autoclaving and cooling steps. Drop off 40 μL of 50 % (w/v) green dye (final concentration is 1 %). It is used as an indicator of uniform drug dispersion in the food (*see* **Note 4**).

Table 1
Screening throughput in fruit fly

Stage	High throughput	Medium throughput	Low throughput
Embryo			Axon wiring Cell fate determination Neuronal development Organogenesis Pattern formation
Larva	Body size Necrotic spots Lethality	Olfactory Weight	Body contraction Locomotor defect Muscle contraction Tissue development
Adult	Anesthesia response Body dimension Flight capacity Lethality Stress assay	Aggression Fecundity Sleep, awakening, and stand-off behavior Weight Wing expansion behavior	Climbing assay Courtship and mating behavior Electrophysiology Feeding behavior Learning and memory behavior Life span Pain reaction Phototaxis Retinal degeneration Seizure and tremor behavior Visual discrimination

2. Add 80 μL of drugs at 1 mM to provide a final concentration of 40 μM (*see* **Note 5**).
3. Add 1.9 mL of cool Jazz-Mix to each vial, vortex immediately, and make sure that the green color is homogenously dispersed.
4. While the media hardens, cover the tubes for 2 h with coverage tissues to prevent the buildup of condensation and to avoid contamination by "free" flies (*see* **Note 6**).
5. Drop off five males and five females (3 days old) in each vial supplemented with drugs.
6. Raise the crosses in an incubator for 2 days at 25 °C with food supplemented with 40 μM of individual compound.
7. At day 2, heat shock the crosses at 37 °C in water bath for 1 h to eliminate progenies with chromosome balancer.
8. Keep the vials at 25 °C for 10–15 days and count viable flies that are otherwise lethal.
9. Compare results to the vial containing food but no drug. Estimate the efficacy of the drugs using Table 2.

Table 2
Primary lethality screening cutoff

Number of flies	Results	Interpretation
0	No rescue	No hit
0–5	Very weak rescue	Potential hit
		Needs to be confirmed with drug concentration modulation (lower and higher dosages) in this assay
5–10	Weak rescue	Potential hit
		Needs to be confirmed with drug concentration modulation (lower and higher dosages) in this assay
10–15	Good rescue	Hit
		Should be confirmed in another validation assay
>20	Very good rescue	Hit
		Should be confirmed in another validation assay

3.2 Validation Assays

To confirm the hits from the primary screen, it is critical to develop secondary or validation assays. The assay that we used for this protocol is behavior-based circadian assay that measures the locomotor activity of *Drosophila*. Locomotor movement of *Drosophila* can be measured with special apparatuses. These devices are housed in environmentally controlled incubators located in a darkroom and use a beam of infrared light to record the locomotor activity of individual flies contained inside small tubes. When measured over a week, *Drosophila* exhibit daily cycles of activity and inactivity, a behavioral rhythm that is governed by the animal's endogenous circadian system.

Here, to confirm the hits identified in the initial lethality rescue screening in Subheading 3.1, the same fly lines used in Subheading 3.1 are subject to additional viability assays such as the locomotion behavioral assay (*see* **Note 3**) that will be performed on a larger scale using 40 μM of each compound (*see* **Note 7**). It is crucial to use the animals that are reared in the same environmental conditions and of the same age. In general, adult male flies that are reared in 25 °C and between 1 and 5 days old are used for the locomotor activity assays (*see* **Note 8**).

1. For each drug, prepare 16 small glass vials with food supplemented with the compound (*see* **Note 9**). In a 50 mL beaker, drop off 80 μL of 50 % (w/v) green dye with a final concentration of 1 %.
2. Add 160 μL of 1 mM drugs to the food in the activity tubes, which will provide a final concentration of 40 μM. Make sure to prepare tubes without drug for controls.

3. Add 20 mL of melted Jazz-Mix food precooled to 50 °C to the 50 mL beaker.
4. Mix immediately and make sure that the green color is homogenously dispersed.
5. Plunge 16 activity Pyrex tubes tightened with a rubber band into the mixture, covering approximately 1 cm of the activity tube.
6. Let the medium harden for 1 h.
7. Recover the activity tubes and cap them on the food side with plastic caps. Let the tubes dry for 1–2 h.
8. Load 1 adult male fly used in Subheading 3.1 per vial for all 16 glass vials.
9. Anesthetize the flies with carbon dioxide.
10. Use a fine paintbrush to gently transfer a fly into an activity tube.
11. Insert a small piece of cotton into the nonfood end of the activity tube to plug the opening and prevent the fly from escaping (*see* **Note 10**).
12. Insert the tubes into the activity monitors and hold the tubes in place with rubber bands to ensure that the infrared beam passes through the center of the tube.
13. Adjust the incubator temperature to 25 °C and set a 12-h light/12-h dark regime using the DAM System light controller.
14. Put the activity monitors into the incubator and connect them to the data collection system.
15. Use the DAM System collection software to make sure that all the monitors are connected properly and data is being collected from each of them (*see* **Note 11**).
16. Synchronize and entrain flies to the 12-h light/12-h dark regime for 2 full days. No data recording is needed.
17. On day 3 record 2 full days on the 12-h light/12-h dark regime.
18. Set the DAM System light controller to constant dark and record data for 4 full days.
19. Recover raw binary data collected using the DAM System software onto a portable data storage device.
20. Process the raw binary data using DAM Scan software and summed into 0.5 and 1-h bins.
21. Export the processed data to Excel files for the 2 days in light-dark and the 4 days in dark-dark cycle.
22. Calculate the activities, defined as the average number of movements recorded in 0.5 or 1 h, and plot graphs for each condition, drugs, and genotype (*see* **Note 12**).

23. Identify significant differences between genotypes/drug treatments using ANOVA, post hoc *t*-tests (two samples assuming equal variances).
24. Repeat the experiment once if the results are inconclusive.

4 Notes

1. Many *Drosophila* models have been developed for human neurological diseases and disorders. Genetic manipulation such as mutation, overexpression, and knockdown (RNAi) can be applied in the whole organism or in specific tissues to mimic the disease in human. In addition to our studies, similar screenings could be conducted using fly models for other neurological disorders. Several online resources are available. These websites provide important information about available whole-genome sequences, mutant alleles, RNAi knockdown lines, and human disease homologs in *D. melanogaster*. The most convenient and comprehensive Internet-based resource for the researcher community is Flybase. This database provides an overview of data obtained from published scientific information, sequence archives, and extensive data providers of fly material resources (i.e., mutant stocks or cell lines). In addition, Flybase contains valuable information on genes, annotation, sequences and transgene constructs, and references to the literature. Moreover, Flybase provides direct links to many other important *D. melanogaster*-related websites, making it easier for researchers to reach the available resources from one place. In Flybase, the links to these additional Web resources are arranged in an alphabetical order.

   Flybase and Other Websites

   - Flybase: flybase.org.
   - DrosDel *Drosophila* Isogenic Deficiency Kit (an isogenic collection of *Drosophila* strains comprising deletions overlapping most of the region genome): www.drosdel.org.uk.
   - *Drosophila* Genetic Resource Center (Kyoto) (information database and stock center in Japan): www.dgrc.kit.ac.jp/en/index.html.
   - *Drosophila* Genomics Resource Center (source center for cell lines, cDNA clones and vectors): dgrc.cgb.indiana.edu.
   - FlyPNS (a specialized database on embryonic and larval peripheral nervous system): www.normalesup.org/vorgogoz/FlyPNS/page1.html.
   - Gene Disruption Project P-Screen Database (database of gene disruption strains): flypush.imgen.bcm.tmc.edu/pscreen.

- Homophila (database providing relation of human diseases to *D. melanogaster* genes): superfly.ucsd.edu/homophila.

Public Stock Centers

The public stock centers of *D. melanogaster* are a precious source for acquiring a diversity of tools for research (balancers, mutants, RNAi strains, and deficiency kits). One of the most universally used stock centers by fly community is the Bloomington *Drosophila* Stock Center at Indiana University.

- The Bloomington *Drosophila* Stock Center at Indiana University: flystocks.bio.indiana.edu.
- *Drosophila* RNAi Screening Center: flyrnai.org/RNAi_index.html.
- *Drosophila* TILLING Project: tilling.fhcrc.org/fly.
- Duke University Model System Genomics: www.biology.duke.edu/model-system/FlyShop/index.html.
- Exelixis *Drosophila* Stock Collection at Harvard Medical School: Drosophila.med.harvard.edu.
- Fly stocks of National Institute of Genetics (NIGFLY): www.shigen.nig.ac.jp/fly/nigfly/index.jsp.
- Gene Disruption Project Database: flypush.imgen.bcm.tmc.edu/pscreen/.
- Vienna *Drosophila* Resource Center (RNAi strain collection): www.vdrc.org.

2. All flies should be kept with a chromosome balancer that carries a heat shock-inducible apoptotic gene. This strategy avoids extensive work to remove or determine genotype of the progeny and only flies with the right genotype can survive (homozygous mutant or driver/transgene).
3. The choice of readout is essential for the speed of the screening. There are many different assays available in *Drosophila* (Table 1). However, visible phenotypes or markers are almost always preferred. A cell sorting method, equivalent to flow cytometry, was previously used on *C. elegans* embryos [22]. The same method can also be used to seed fly embryo into 96- or 384-well plates. Afterward, an appropriate phenotype (such as variation in fluorescent markers) compatible with automated HTS could be used as the readout. The throughput of behavioral assays can vary from low to medium. The simple assessment of both locomotor and circadian activities can be performed in medium-throughput screenings, such as DAM System described above. Two to three thousand compounds can be screened per month using DAMS. Learning and memory assays are included in semiautomatic methods with low throughput. These assays require conditioned stimulus training where only a small population of flies (8–32) can be trained [23, 24].

Usually 25–50 drugs can be screened by one person per week using a 32-channel trainer. Courtship behavior and aggression are the most common phenotypes used in social interaction assays. The recording of these social interaction phenotypes is difficult to automate and therefore requires a human observer. However, some important parts have been automated by video tracking software. Thus, locomotor activity, aggression, and courtship behavior can be analyzed simultaneously on 96 video channels [25, 26].

4. Uniform drug dispersion in the fly food is critical for the success of small-molecule screenings. Inclusion of food color or other indicator is critical.

5. For many drugs, 40 μM is the concentration that has been determined to be effective without significant toxicity. However, some drugs can be lethal to the fly at 40 μM. For these drugs, tests should be performed at a lower concentration (e.g., 5–10 μM). Obtaining small-molecule libraries with good coverage in chemical space is important. Different sets of compound libraries are listed below. The advantage of using the Spectrum Collection in the initial screening is that all the compounds in this library are FDA-approved drugs, making the translation of screening discovery into human clinical trials easier.

   - MicroSource Inc. (offers FDA-approved drugs, natural products, and synthetic compounds in 96-well plate): http://www.msdiscovery.com/home.html.
   - ChemBridge Corporation (provides over 700,000 different compounds and chemical design): http://www.chembridge.com/products.html.
   - Comgenex (offers drug libraries of 200,000 compounds, and technical expertise): http://www.rdchemicals.com/targetedcompound-libraries/comgenex.html.
   - TimTech (provides over 1,000,000 of natural compounds and synthetic organic drugs, and diverse customizing services): http://www.timtec.net/.

6. Water condensation should be removed. If not, flies will stick to the tube wall and die.

7. To further validate the initial hits, it is important to test different concentrations of small molecules in a secondary screen to identify the most potent ones for further studies.

8. Male flies (instead of female flies) are traditionally used because egg-laying activity will affect true measurement of locomotor activity. Because of sexual dimorphism, sometimes assaying female flies might be informative.

9. Generally, a sample size of at least 16 flies is necessary to obtain reliable results for a particular genotype.
10. Make sure that the tubes are laid on their sides until the fly awakens, or else there is the risk of the fly getting stuck to the food.
11. The monitor emits an infrared light beam across the center of each glass activity tube. The locomotor activity of the flies is recorded as raw binary data where "one" is recorded each time the infrared beam is broken or a "zero" is recorded in which the infrared beam is not broken.
12. Routinely we use Microsoft Excel to calculate activity. However, FaasX software (M. Boudinot and F. Rouyer, Centre National de la Recherche Scientifique, Gif-sur-Yvette Cedex, France) and Insomniac (Matlab-based program; Leslie Ashmore, University of Pittsburgh, PA) can be used to examine circadian (e.g., period and power) and sleep/rest (e.g., percentage sleep, mean rest bout length) parameters, respectively.

Acknowledgements

This work was supported by NIH grants (R01 NS051630 and R21 NS067461 to P.J.).

References

1. Andretic R et al (2008) Drosophila D1 dopamine receptor mediates caffeine-induced arousal. Proc Natl Acad Sci U S A 105:20392–20397
2. Bainton RJ et al (2000) Dopamine modulates acute responses to cocaine, nicotine and ethanol in Drosophila. Curr Biol 10:187–194
3. McClung C, Hirsh J (1998) Stereotypic behavioral responses to free-base cocaine and the development of behavioral sensitization in Drosophila. Curr Biol 8:109–112
4. Moore MS et al (1998) Ethanol intoxication in Drosophila: genetic and pharmacological evidence for regulation by the cAMP signaling pathway. Cell 93:997–1007
5. Nichols CD et al (2002) Hallucinogens and Drosophila: linking serotonin receptor activation to behavior. Neuroscience 115:979–984
6. Rothenfluh A, Heberlein U (2002) Drugs, flies, and videotape: the effects of ethanol and cocaine on Drosophila locomotion. Curr Opin Neurobiol 12:639–645
7. Satta R, Dimitrijevic N, Manev H (2003) Drosophila metabolize 1,4-butanediol into gamma-hydroxybutyric acid in vivo. Eur J Pharmacol 473:149–152
8. Wolf FW, Heberlein U (2003) Invertebrate models of drug abuse. J Neurobiol 54: 161–178
9. Rubin GM, Spradling AC (1982) Genetic transformation of Drosophila with transposable element vectors. Science 218:348–353
10. Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118:401–415
11. McGuire SE, Roman G, Davis RL (2004) Gene expression systems in Drosophila: a synthesis of time and space. Trends Genet 20: 384–391
12. Roman G et al (2001) P[Switch], a system for spatial and temporal control of gene expression in Drosophila melanogaster. Proc Natl Acad Sci U S A 98:12602–12607
13. Dietzl G et al (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448: 151–156
14. Pollitt SK et al (2003) A rapid cellular FRET assay of polyglutamine aggregation identifies a novel inhibitor. Neuron 40:685–694

15. Desai UA et al (2006) Biologically active molecules that reduce polyglutamine aggregation and toxicity. Hum Mol Genet 15: 2114–2124
16. Chang S et al (2008) Identification of small molecules rescuing fragile X syndrome phenotypes in Drosophila. Nat Chem Biol 4:256–263
17. Warren ST, Sherman SL (2001) The fragile X syndrome. In: Scriver CR, Beaudet al, Valle D, Childs B, Kinzler KW, Vogelstein B (eds) The metabolic & molecular bases of inherited disease. McGraw-Hill Companies, New York, pp 1257–1290
18. Sherman S (2002) Epidemiology. In: Hagerman RJ (ed) Fragile X syndrome: diagnosis, treatment and research. The Johns Hopkins University Press, Baltimore, MD, pp 136–168
19. Hagerman RJ, Hagerman PJ (2002) The fragile X premutation: into the phenotypic fold. Curr Opin Genet Dev 12:278–283
20. Hagerman PJ, Hagerman RJ (2004) The fragile-X premutation: a maturing perspective. Am J Hum Genet 74:805–816
21. Jin P et al (2003) RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. Neuron 39:739–747
22. Pulak R (2006) Techniques for analysis, sorting, and dispensing of *C. elegans* on the COPAS flow-sorting system. Methods Mol Biol 351: 275–286
23. Ali YO et al (2011) Assaying locomotor, learning, and memory deficits in Drosophila models of neurodegeneration. J Vis Exp 49:2504
24. Scott R et al (2002) CREB and the discovery of cognitive enhancers. J Mol Neurosci 19: 171–177
25. Branson K et al (2009) High-throughput ethomics in large groups of Drosophila. Nat Methods 6:451–457
26. Dankert H et al (2009) Automated monitoring and analysis of social behavior in Drosophila. Nat Methods 6:297–303

Chapter 11

High-Throughput Small-Molecule Screening in *Caenorhabditis elegans*

Sunitha Rangaraju, Gregory M. Solis, and Michael Petrascheck

Abstract

Chemical compounds, which modulate enzymatic activities or those which induce specific phenotypes of interest, are valuable probes to study biological phenomena, as they allow modulation of enzymatic activities and temporal control of protein action. Here, we describe the methodology to conduct large-scale screens for chemical compounds that induce a desired phenotype in the roundworm *Caenorhabditis elegans* (*C. elegans*) using 96- or 384-well microtiter plates.

Key words *Caenorhabditis elegans*, High-throughput, Small molecules, Drug screen, Chemical libraries

1 Introduction

One strategy to study biological phenomena in *C. elegans* is to screen for chemical compounds that induce a phenotype of interest. Such screens are often referred to as forward pharmacological screens, in analogy to forward genetic screens, since both screen for phenotypic changes rather than activation or inhibition of a specific target protein. Elucidating the mechanisms by which such chemical probes perturb a specific process and cause phenotypes of interest reveals insights into the regulation of the process of interest. Chemical probes are valuable tools that provide multiple technical advantages. Chemical probes allow modulation of severity of the phenotype through use of different probe concentrations. They also allow transient modulation of a process and are easily combined with other compounds or strains that carry mutations, thus making them very flexible experimental tools. Furthermore, chemical probes have distinct advantages to study phenotypes such as sterility, larval arrest, or lethality, which oftentimes result in strains that are difficult and laborious to maintain.

Identification of chemical probes that induce a phenotype of interest requires screening of large compound libraries for compounds

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_11, © Springer Science+Business Media New York 2015

that induce the phenotype of interest. To date, *C. elegans* is the best characterized metazoan that, as a whole animal, is amenable to liquid handling technologies and can be cultured in 96-well or even 384-well microtiter plates.

Here, we outline the general methodology of culturing *C. elegans* in 96- as well as 384-well plates and the materials required to culture worms in such a format. We further lay out the considerations necessary to plan and implement such phenotype-based screens and conclude with some general notes. In Subheading 2, we describe the materials necessary and in Subheading 3, we outline the general protocol. In Subheading 4, we delineate the general considerations that have to be taken into consideration while planning a chemical screen using *C. elegans*. The protocol outlined in Subheading 3 assumes that phenotypes will be scored using an inverted microscope and that the compounds will be added to day 1 adults. Dependent on the type of phenotype analyzed, this may change.

1.1 Planning and Considerations

In this section, we outline what considerations have to be taken when planning to screen for compounds that induce phenotypic changes of interest in *C. elegans*. While the details are highly dependent on the size of the screen, the phenotype of interest, and the assay conditions, there are some general considerations that are common to all screens. In addition to the recommendations that are provided below, a good place to find information on how to develop an in vivo screen is the "Assay Guidance Manual" that can be found in the NIH bookshelf [1].

1.1.1 Choice of Screening Library

Forward pharmacological screens are generally more difficult, laborious, and an order of magnitude smaller than in vivo screens against a known target. Choosing the right library to screen is therefore one of the most important aspects in the preparation for such a screen. If phenotype assessment is difficult and laborious, it may be advisable to screen a smaller library consisting of compounds with known biological activity at the price of novelty.

Many universities today have repositories or facilities with various libraries available. For the purpose of this protocol, we distinguish four classes of screening libraries that differ in various aspects such as size, biological activities, novelty, and characterization of the different compounds. The four main types of libraries are (a) collection of pharmacological compounds; (b) targeted libraries; (c) natural product libraries; and (d) compound libraries (Table 1).

(a) *Collection of pharmacological compounds* is an attractive starting point. These collections/libraries consist of well-characterized compounds with known pharmacology or FDA-approved drugs. The library ranges in size from 500 to 5,000 compounds. Hit compounds will be commercially available from various

Table 1
Compound screening libraries

Library type	Size	Advantages	Disadvantages	Providers
Collection of pharmacological compounds	500–5,000	Commercially available, low toxicity, known pharmacology, high biological activity, known targets	Not novel, biased	Tocris, Sigma, Prestwick, NIH, Microsource
Targeted library	100–10,000	Target identification	Biased	Asinex, Chemdiv
Natural product library	Varies	Biological activity, wide range of different pharmacophores	Difficult to synthesize, difficult to obtain pure compound, target identification is challenging	Prestwick, Enzo Lifesciences, Microsource
Compound library	500,000 or more	Industry standard, unbiased, novelty	Expensive, laborious	Chembridge, Chemdiv, Maybridge, Asinex

Note: The list of providers is certainly not complete and we apologize for any inadvertent omissions

sources in larger quantities. Toxicity of these compounds is generally low as these compounds are already extensively optimized and all compounds are known to be biologically active. Knowledge about the pharmacology can guide the post-screening experiments and help elucidate the underlying mechanisms.

(b) *Targeted libraries* consist of compounds that have been designed to target specific classes of proteins (e.g., kinases). They are generally constructed around pharmacophores that, in the past, have provided successful starting points for compounds targeting a specific class of proteins. If the phenotype of interest is known to be regulated by proteins belonging to this particular class, targeted libraries are advantageous as they reduce the size of the screen combined with the possibility of identifying novel compounds.

(c) *Natural product libraries*, or phytochemical libraries, are another potential library for forward chemical screens. These libraries vary in size and generally have a good hit rate since all compounds can be expected to be biologically active. While the biological activity is an attractive feature of these compounds, commercial availability, purity, as well as structure-activity studies are the challenges associated with phytochemicals.

(d) *Compound libraries*, a collection of compounds designed to cover a large pharmacophore space, are the most widely used screening libraries. To date, the sizes of these libraries range in the several hundreds of thousands of compounds and reach the million compound mark. Such libraries certainly have a low bias and screening these libraries can lead to truly novel compounds. However, in vivo screens of more than 10,000 compounds are strenuous endeavors in terms of time, cost, and effort and will require sophisticated, high-throughput equipment. Since *C. elegans* feeds on bacteria and most of the high-throughput facilities specialize in and are designed for in vivo or cell culture screens, the facilities will need to take special precautions to prevent contamination arising from bacteria.

1.1.2 Automated Liquid Handling

Depending on the size of the screen, use of robotic liquid handling should be considered for distributing *C. elegans* into plates as well as to add the compounds to the culture plates. Robotic pipetting reduces variability and errors. There are two main difficulties associated with pipetting *C. elegans*. The first difficulty is to maintain a uniform worm concentration over the extended period of time it takes for distributing the worms into plates. Worms have a tendency to sink to the bottom of a container within less than a minute. If possible, worms should be pipetted as L1 larvae. L1 larvae are highly active and have lower mass than adults, which makes it easier to keep them in suspension. However, even for L1 larvae, a stir bar or a similar way to constantly stir the liquid is required. The second problem is that *C. elegans* has a high tendency to stick to plastic. This problem is most easily solved by pipetting the worms in combination with the feeding bacteria. The feeding bacteria will saturate any potential “sticky” sites. Another possibility is to use non-disposable Teflon-coated tips.

Considerations for pipetting the library are mainly dependent on the solvent used for the construction of the library. Most, if not all, compound libraries are dissolved in DMSO at concentrations of around 10 mM. We have found that *C. elegans* tolerates up to 0.55 % of DMSO with only minimal effects to their biology. Therefore, the maximum volume of DMSO-dissolved library compound that can be pipetted into each well of a 150 μL liquid culture (96-well plate) is 0.8 μL, resulting in ~50 μM final concentration of a given compound in the library. For 384-well plates, the corresponding maximum volume of compound dissolved in DMSO that can be added is 0.3 μL. Most mass-produced disposable tips for liquid handling today have a minimum limit of 0.5 μL. These problems can be either solved by using a pin tool, or by diluting compounds first in a buffer before adding them to the *C. elegans* culture.

1.1.3 Choice of Microtiter Plates

The choice of microtiter plate will be contingent upon the phenotype, the length of the screening procedure, and the overall size of the screen. There are some general points to consider. For the biology of the worm, the most important consideration is the distance from the bottom of the well to the surface of the liquid. This distance determines the oxygen levels that the animal is exposed to. Oxygen diffuses in liquid very poorly. As a result, at the same culture volume, worms in plates with larger well area have more oxygen. For a 96-well plate using a final volume of 150 μL per well, the well bottom area should be 0.29 cm^2 or greater. The plates should be sterile and if possible the bottom should not be treated or coated (functionalized). To avoid problems with evaporation, plates also have to be sealed with tape sealers that are not gas permeable. Gas-permeable sealers will not prevent evaporation. Evaporation across plates is very irregular and most pronounced at the border wells, affecting salt and compound concentrations. Because the plates have to be sealed, it is important not to fill the well all the way to the rim and to leave space as a reservoir for oxygen. Oxygen can be replenished by exchanging the sealers periodically, typically once every week.

384-Well plates allow much higher throughput but screens in 384-well plates are more challenging. The density of animals in a well of a 384-well plate will be necessarily higher (5–10 animals) than for 96-well plates (5–15 animals). The main problem is to provide the animals with sufficient OP50 bacteria to sustain them without causing them to go into starvation. However, too much bacteria is problematic, as it will deplete oxygen. Bacterial concentrations of ~7 mg/mL will sustain 5–10 worms in a well of a 384-well plate throughout development (up to 192 h) but not much longer and refeeding will be necessary.

1.1.4 Dead Versus Live Bacteria

Another aspect of designing a *C. elegans* screen is whether to feed the worms with live or dead bacteria. Worms prefer feeding on living bacteria. We have verified that the bacteria in the below described *C. elegans* liquid culture conditions do not grow, as there are no nutrients for bacterial growth in the media. However, it is likely that bacteria will metabolize some of the library compounds. Furthermore, some library compounds could affect the bacteria, causing potential indirect effects on the worms [2]. The most effective way to prepare dead bacteria is by gamma irradiation with 1,000 gray. In contrast to most other methods we have tried, 1,000 gray gamma irradiation can kill large batches of bacteria to less than 1 in 100,000 bacteria alive. We have tested various alternatives to gamma irradiation (heat, sodium azide, UV irradiation, and combinations of antibiotics), but all of them are less effective, less reliable, and often result in feeding bacteria that do not allow *C. elegans* to develop into adults.

1.1.5 Statistical Considerations

Before any large-scale screens are conducted, it would be necessary to conduct a small pilot screen to determine the distribution of the measured phenotype. It is widely assumed that phenotypic measurements follow a normal distribution. Surprisingly, we have observed that this is often not the case. For measurements that are not normally distributed, standard measurements such as the Z'-factor, which are used to determine the feasibility of a screen, may not be adequate [3]. It is highly advisable to involve a trained statistician well ahead of time during the early planning stages of the screen, before the screen is conducted.

2 Materials

All solutions are prepared using ultrapure deionized water (Millipore) and analytical grade reagents. All solutions can be stored at room temperature unless otherwise indicated.

2.1 Feeding Bacteria

1. Luria Broth (LB): Dissolve 10 g of Bacto Tryptone, 5 g of yeast extract, and 10 g of NaCl into 800 mL of deionized water. Adjust the pH to 7.0 with a concentrated NaOH solution and dilute to 1,000 mL, and then autoclave.
2. Solution of 0.17 M KH_2PO_4/0.72 M K_2HPO_4: Dissolve 23.1 g of KH_2PO_4 and 125.4 g of K_2HPO_4 in 900 mL of deionized water. After the salts have dissolved, adjust the volume of the solution to 1,000 mL with deionized water and sterilize by autoclaving.
3. Terrific Broth (TB): Dissolve 12 g of Bacto Tryptone, 24 g of yeast extract, and 4 mL of glycerol into 900 mL of deionized water, and then autoclave the solution. Allow to cool down to 55 °C and add 100 mL of the 0.17 M KH_2PO_4/0.72 M K_2HPO_4 solution.
4. 100 mg/mL Ampicillin: Dissolve 1 g of ampicillin in 10 mL of sterile deionized water. Sterile filtrate, aliquot, and store at −20 °C. Use at a final concentration of 50–100 μg/mL.
5. 250 μg/mL Amphotericin B (Fungizone): Dissolve 1 mg of amphotericin B in 4 mL of ethanol. Aliquot and store at −20 °C. Use at a final concentration of 0.1 μg/mL (*see* **Note 1**).
6. 1 M Potassium phosphate buffer, pH 6.0: Dissolve 136 g of KH_2PO_4 in 900 mL of sterile deionized water. Adjust the pH to 6.0 with 5 M KOH and then add deionized water to 1,000 mL and autoclave the solution.
7. S-basal medium: Weigh 5.9 g of NaCl and add 50 mL of 1 M potassium phosphate, pH 6.0. Add 1,000 mL deionized water and then autoclave. Let the solution cool to 55 °C and then add 1 mL of 5 mg/mL cholesterol (dissolved in ethanol).

8. 1 M Potassium citrate, pH 6.0: Dissolve 268.8 g of tripotassium citrate ($C_6H_5K_3O_7$) and 26.3 g of citric acid monohydrate ($C_6H_8O_7 \cdot H_2O$) in 900 mL of deionized water. Adjust the pH to 6.0 with 5 M KOH. Add deionized water to 1,000 mL and then autoclave.
9. Trace metal solution: Dissolve 1.86 g of Na_2EDTA, 0.69 g of $FeSO_4 \cdot 7H_2O$, 0.20 g of $MnCl_2 \cdot 4H_2O$, 0.29 g of $ZnSO_4 \cdot 7H_2O$, and 0.016 g of $CuSO_4$ in 1,000 mL deionized water and autoclave. Store this solution in the dark or cover with aluminum foil.
10. S-complete medium: Combine 974 mL of S-basal, 10 mL of 1 M potassium citrate pH 6.0 (sterile), 10 mL of trace metal solution (sterile), 3 mL of 1 M $CaCl_2$ (sterile), and 3 mL of 1 M $MgSO_4$ (sterile).

2.2 *C. elegans* Culture

1. Nematode growth medium (NGM) agar plates: Dissolve 3.0 g of NaCl, 2.5 g of peptone (from casein, pancreatic digest), and 17 g of agar in 975 mL of deionized water with a stirring bar. Autoclave the solution, allow to cool down to 55 °C, and then add 0.5 mL of 1 M $CaCl_2$ (sterile), 1 mL of 5 mg/mL cholesterol in ethanol, 1 mL of 1 M $MgSO_4$ (sterile), and 25 mL of potassium phosphate buffer, pH 6.0 (sterile). Once everything is dissolved after stirring with the stir bar, pour the still hot NGM into Petri dishes using sterile technique. Leave the plates at room temperature for 2 days to cool off and for moisture to evaporate.
2. M9 buffer: Dissolve 6 g of Na_2HPO_4, 3 g of KH_2PO_4, 5 g of NaCl, and 0.25 g of $MgSO_4 \cdot 7H_2O$ in 1,000 mL of deionized water and autoclave.
3. Stock solution of 0.6 mM 5-fluoro 2′-deoxyuridine (FUDR): Dissolve 100 mg of 5-fluoro 2′-deoxyuridine in 670 mL of sterile S-complete. Make 10 or 45 mL aliquots and store at −20 °C. Use at a final concentration of 100–120 μM.
4. 100 mg/mL Carbenicillin: Dissolve 1 g of carbenicillin in 10 mL of sterile deionized water. Sterile filtrate, aliquot, and store at −20 °C. Use at a final concentration of 50 μg/mL.
5. Bleach/NaOH solution (prepared fresh before synchronization): Mix 7.0 mL of sterile deionized water, 2.5 mL of household bleach, and 0.5 mL of 10 M NaOH.

3 Methods

This is a step-by-step protocol on how to culture *C. elegans* in 96-well plates [4]. The method has been developed for screening purposes and is not suited for general strain maintenance.

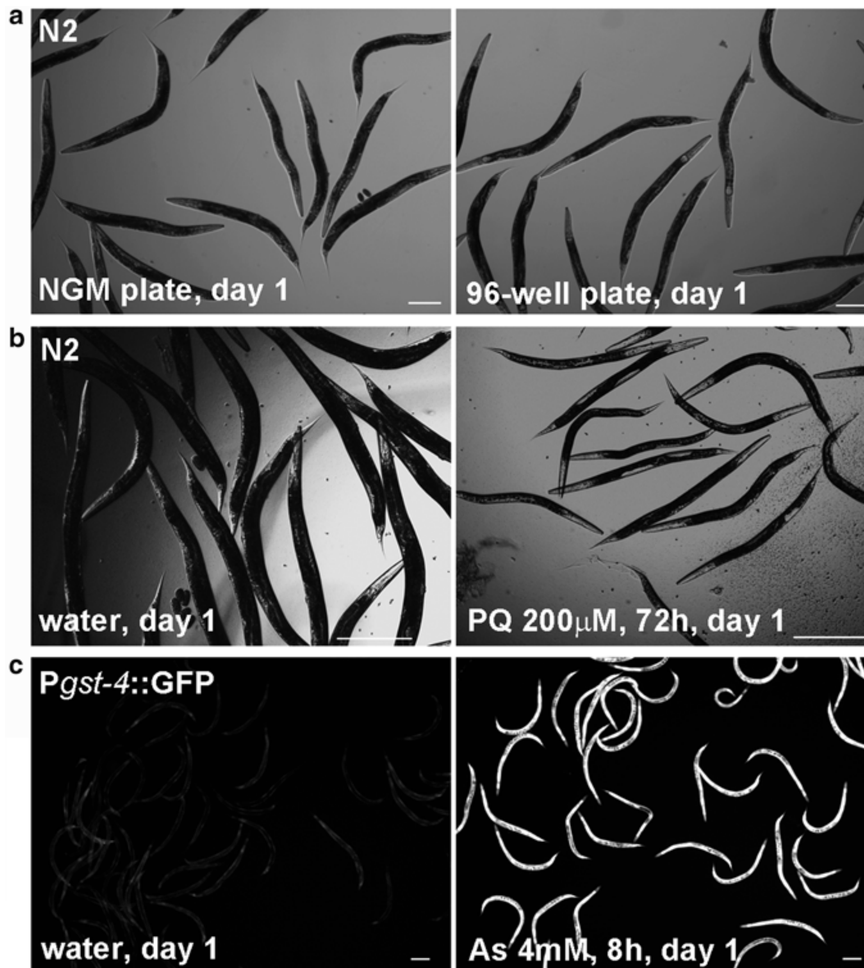


Fig. 1 Examples of *C. elegans* phenotypes that can be imaged in 96-well plates. (**a**) Synchronized day 1 adult N2 worms grown on either solid NGM plates (*left panel*) or using the 96-well (*right panel*) protocol described in this chapter. Worms grown by these two methods are morphologically indistinguishable. (**b**) Paraquat treatment delays development. N2 worms were seeded in 96-well plates using the protocol (Subheading 3.2). L1 larvae were treated with either water or 200 μM paraquat (PQ) and imaged 72 h later. While water-treated control worms developed into adults (*left panel*), PQ-treated worms did not (*right panel*). (**c**) Arsenite induces gst-4 expression. Wild-type transgenic worms expressing GFP driven by the glutathione S-transferase-4 promoter (P *gst-4*::GFP) [10] were cultured in 96-well plates and imaged after 8-h water or arsenite (4 mM) treatment. All scale bars are 100 μm long

Strains should be maintained on NGM plates as described in "The genetics of *Caenorhabditis elegans*" [5]. *C. elegans* grown in NGM plate and 96-well plates look identical morphologically and phenotypically (Fig. 1a). This protocol described in this chapter has been used to screen for compounds that extend life span and stress resistance and could be adapted for other assays [6–8]. For 384-well plates, the changes in volume are indicated at the end of each paragraph. We have split the methodology sections

into the following subsections: Preparation of the feeding bacteria (Subheading 3.1), Preparation of the 96-/384-well *C. elegans* culture (Subheading 3.2), and Scoring phenotypes (Subheading 3.3). To aid in experimenter's protocol planning, week, day, and critical times are provided for each step and are organized in additional subsections. The entire protocol uses the L4 stage, the last larval and molting stage before adulthood, as a day [0] reference point.

3.1 Preparation of the Feeding Bacteria

In this protocol, *C. elegans* are fed an *E. coli* strain, OP50, which is resistant to carbenicillin as well as ampicillin (available from the *Caenorhabditis* Genetics Center (CGC)). In our protocol, we make use of the ampicillin resistance while preparing OP50 (Subheading 3.1) and carbenicillin resistance during the setup of 96-/384-well plates with *C. elegans* (Subheading 3.2). Antibiotic resistance allows selective growth in the presence of these antibiotics to prevent cross contamination with other bacteria. Preparation of OP50 using sterile conditions should occur 4–5 days before it is added to the worm culture.

3.1.1 Day [−8]: Wednesday (Week 1): Prepare OP50 Pre-inoculum

1. In the morning, pick a single colony of OP50 and add to 5 mL of LB medium. Add ampicillin to a final concentration of 100 μg/mL and amphotericin B (Fungizone) to a final concentration of 0.1 μg/mL (*see* **Note 1**). Incubate this pre-inoculum for 7–10 h in a bacterial shaker at 37 °C. In the evening, transfer ~1 mL of the pre-inoculum to a flask of 250 mL of TB containing 50 μg/mL ampicillin and continue incubating in the shaker at 37 °C for 8–12 h overnight until the bacterial culture reaches saturation. Do not allow the culture to grow longer than 16 h.

3.1.2 Day [−7]: Thursday (Week 1): Prepare OP50 Solution

1. Transfer the overnight OP50 culture into six 50 mL centrifugation tubes, being sure to pre-weigh two of the empty centrifugation tubes. Centrifuge the tubes for 10 min at 2,200 × *g* in a tabletop or knee-well centrifuge to pellet the OP50 (*see* **Note 2**).
2. Discard the supernatant and use 25 mL of autoclaved, deionized water to resuspend the OP50 pellet in each tube. Spin down and wash once more.
3. After washing, add 10 mL of autoclaved, deionized water to each tube, resuspend each pellet, and then combine everything into the two pre-weighed 50 mL centrifugation tubes. Spin down both tubes once more.
4. After the last spin, carefully aspirate the remaining water, being sure none is left in the tubes. Weigh the centrifugation tubes containing the OP50 pellets and subtract the pre-weighed values of the empty tubes to determine the wet weight of the OP50.

5. Resuspend the pellet in S-complete to a concentration of 100 mg/mL, making sure that no clumps are present (*see* **Note 3**). Combine both OP50 solutions into one tube if the volume allows it.
6. To verify if the correct concentration of OP50 was prepared, use a spectrophotometer to measure the optical density and determine the number of bacteria per mL. The relationship between bacterial counts and optical density at 600 nm (OD_{600}) has to be established for each spectrophotometer individually. One mg of OP50 (pellet at high speed, remove liquid, and measure wet weight) suspended in 1 mL corresponds to roughly 2×10^8 bacteria/mL. Adjust the concentration with S-complete as needed.
7. The OP50 solution should be stored at 4 °C until it will be used to prepare the worm seeding solution in Subheading 3.2.3, **steps 2** and **3** (*see* **Note 4**).

3.2 Preparation of the 96-/384-Well Plate *C. elegans* Culture

The aim of this section is to prepare an age-synchronous population of worms that can be distributed evenly in 96- or 384-well plates. It is important that all steps after the bleach treatment in this section (Subheading 3.2.2, **step 3**) are done in a sterile manner (*see* **Note 5**). The timing in this protocol has been worked out for worms cultured at 20 °C. If other worm culture temperatures are used, the timing has to be adjusted accordingly.

3.2.1 Day [−6]: Friday (Week 1), 4:00 pm: Transfer Worms to a Fresh NGM Plate

1. From a 5–10-day-old NGM plate with mostly starved L1 larvae, wash off a section of the plate (500–5,000 animals) with sterile water onto a new 10 cm plate with freshly seeded OP50 (*see* **Note 6**).
2. Incubate the plate for approximately 65 h at 20 °C until a majority of the animals are gravid adults containing many eggs. This incubation time may vary from strain to strain (*see* **Note 7**).

3.2.2 Day [−3]: Monday (Week 2), 10:00 am: Synchronize Worms

1. Wash worms off the 10 cm plate from Subheading 3.2.1, **step 2**, with 10 mL of sterile water and transfer to a 15 mL conical tube.
2. Wash the worms with water once by centrifuging at $280 \times g$ and aspirating the supernatant.
3. Prepare a fresh bleach/NaOH solution. Add 5 mL of this solution to the 15 mL conical tube with worms. Carefully monitor the tube under a microscope while vortexing the worms every minute (min) until the adult worms break open, releasing the eggs. The worms should break open within 3–5 min after addition of the bleach solution.
4. Immediately when most of the adult worms break open and release their eggs, add 5 mL of M9 buffer to the egg solution to

neutralize the bleach and stop the reaction. This step is critical: adding M9 too early will result in under-bleaching, leaving the adults intact, while adding M9 too late will overbleach the eggs and reduce viability (*see* **Note 8**).

5. Wash the eggs one to two times (*see* **Note 9**) with 10 mL of M9 buffer, centrifuging for 2 min at 1,100 × *g*.
6. After the M9 wash, wash once with S-complete for 2 min at 1,100 × *g*.
7. Aspirate the supernatant and add 10 mL of S-complete. Place the tube on a nutator or similar device, gently rotating overnight at room temperature.

3.2.3 Day [−2]: Tuesday (Week 2), 12:00 Noon: Seed Worms into Microtiter Plates

1. Check the conical tube from Subheading 3.2.2, **step 7**, under a dissecting microscope to ensure that worms hatched during the night. To determine the concentration of worms, pipette ten individual drops of 10 μL each taken from the worm solution and place onto a clear surface for inspection under the dissecting microscope (*see* **Note 10**). Count the number of worms per drop, averaging over the ten drops, and use this to calculate the worm concentration in worms/mL (*see* **Note 11**).
2. Start to prepare the worm seeding solution that will be distributed into the microtiter plates (96- or 384-wells). A single 96-well plate requires ~12.5 mL of worm seeding solution and is best prepared in 15–50 mL Falcon tubes. If preparing larger quantities of worms ($n > 100{,}000$), use a 600 mL Nunclon flask. Dilute the worms from the previous step (Subheading 3.2.3, **step 1**) in S-complete to a final concentration of 70–100 worms/mL. Add carbenicillin (100 mg/mL stock) and amphotericin B (stock 250 μg/mL in ethanol) to final concentrations of 50 μg/mL and 0.1 μg/mL, respectively. Rotate on a nutator or similar device at room temperature to mix until the OP50 is added in the next step. At this point, the OP50 prepared in Subheading 3.1.2 should be warmed up to room temperature by placing it on a nutator (for 384-well plates: make a seeding solution of 120 worms/mL).
3. At 2:30 pm add the OP50 to the worm seeding solution to a final concentration of 6 mg/mL. Let the solution rotate on a nutator for 10 min to thoroughly mix the worm seeding solution before seeding it into 96-well plates. Without OP50 feeding bacteria the worms are unable to develop. Addition of OP50 will start their developmental cycle (for 384-well plates: use 7 mg/mL OP50).
4. Seed 120 μL of the worm seeding solution into each well of a 96-well plate, making sure that the worms stay suspended while pipetting (for 384-well plates: seed 50 μL of the worm seeding solution into each well).

5. Seal the plate using tape sealer to prevent contamination and evaporation (*see* **Note 12**). Shake the plate for 2 min on a microtiter plate shaker set at medium speed level, and then place in a 20 °C incubator for 2 days. After 2 days, the worms should have reached the L4 stage.

3.2.4 Day [0]: Thursday (Week 2), Before Noon: Add 5-Fluoro 2′-Deoxyuridine (FUDR) to Sterilize Animals

1. For some screens, it will be necessary to prevent the worms from producing offspring (*see* **Note 13**). This can be achieved by adding 30 μL of 0.6 mM FUDR stock solution to each well (final concentration 0.12 mM) using a multichannel pipette (*see* **Note 14**). This addition increases the final volume in each well to 150 μL (*see* **Note 15**) and decreases the final concentration of OP50 from 6 to 5 mg/mL (1×10^9 bacteria/mL). Seal the plate with tape sealers and shake the plate for 2–3 min on a microtiter plate shaker to distribute the FUDR evenly. It is important to add FUDR within 45 h of adding OP50 to the L1 larvae (Subheading 3.2.3, **step 3**). For instance, if the OP50 was added at 2:30 pm on day [−2], it is essential to add the FUDR before noon of day [0] to prevent the animals from making functional sperm. Store the plate at 20 °C (for 384-well plates: add 12 μL of 0.6 mM FUDR stock solution to each well).

3.2.5 Day [1]: Friday (Week 2): Add Compound Library to C. elegans Culture

1. On day [1], by 9:00 am, the animals seeded in Subheading 3.2.3 should have developed into adults with several eggs visible inside (*see* **Note 16**). The eggs should be visible irrespective of whether the animals were FUDR treated or not, as FUDR prevents fertilization of eggs, but not their production. For our experiments, we generally add the compound library of interest on day 1 of adulthood, but this time point will differ from screen to screen (*see* **Notes 17** and **18**). After adding compounds, reseal the plates with tape sealer and shake for 2–3 min on a microtiter plate shaker to distribute the compounds evenly in the culture. Return the plates to the 20 °C incubator.
2. Occasionally, the compound addition step could lead to dead worms, especially when using a solvent other than water. Using an inverted microscope, check the plate for dead worms 6 h after adding the compounds to the culture. Animals that have died due to compound addition artifacts should be excluded from the analysis.

3.2.6 Day [4]: Monday (Week 3), Before Noon: Change Sealers to Replenish Oxygen

1. Some screens will last for several days into adulthood. To ensure fresh oxygen supply, remove the sealer, wait for 1 min, and reseal the plate with a fresh tape sealer. Keep the lid on the 96-well plates during this 1 min to prevent contamination of the culture during the aeration. After resealing the plate, shake it for 2–3 min on a microtiter plate shaker. Repeat this step every week until the experiment is complete (*see* **Notes 19**, **20**, and **22**).

3.2.7 Day [5]: Tuesday (Week 3): Refeed Culture with OP50 to Prevent Starvation

1. The amount of feeding bacteria (OP50) supplied during the initial seeding step (Subheading 3.2.3, **step 3**) allows up to 15 animals to develop into adults and to reach day 5 of adulthood. If the culture is kept longer than that additional feeding bacteria have to be added to prevent starvation. Add 5 μL of the 100 mg/mL OP50 solution prepared in Subheading 3.1.2 (*see* **Note 4**) to each well (*see* **Notes 20** and **21**) (for 384-well plates: add 2 μL of 100 mg/mL OP50 per well).

3.3 Scoring of Phenotypes

Most *C. elegans* screens allow for phenotype scoring using an inverted microscope in clear, flat-bottom plates. Examples of such phenotypes include development (Fig. 1b), activation of fluorescent markers (Fig. 1c), sterility, movement, life span, resistance against toxic compounds, and neuronal degeneration [9]. For less obvious phenotypes or higher throughput, specialized plate readers or high-content imaging systems may be required. Ideally, a phenotype should be quantifiable and easy to assess. Before starting the screen, it will be necessary to determine mean and standard deviation of phenotypic measurements, unless the phenotypes are very obvious. For visually scoring 96- or 384-well plates for obvious phenotypes, we suggest using 2× to 4× objectives, which allow for the rapid screening of many plates by eye without sophisticated instrumentation, as these objectives allow visualizing the contents of the entire well.

3.4 Post-screening Considerations

After the identification of primary hits, it will be important to verify the identity of the compound that causes the desired phenotype. Even slight impurities such as synthesis intermediates could also be responsible for the observed phenotypes. To confirm the identity of a compound, it would be necessary to test a highly pure sample of the compound of interest.

4 Notes

1. Amphotericin B does not dissolve well in ethanol and precipitation is common. Be sure to vortex and resuspend before each use.
2. For the preparation of the feeding bacteria OP50, we recommend centrifuging at 4 °C, as a lower temperature facilitates pelleting.
3. OP50 clumps are best broken up by continuously pipetting up and down or by vortexing.
4. The sterile OP50 prepared on day [−7] should be stored in a clean place at 4 °C until used for seeding on day [−2], and for refeeding on day [5].
5. Always use sterile techniques to avoid contamination of the *C. elegans* 96-/384-well culture. Prior to work, wipe down the

working bench, pipettes, and gloves with 75 % ethanol or 20 % bleach and work next to a flame from a Bunsen burner.

6. Plate seeding is a standard procedure to maintain *C. elegans* [5]. In an NGM plate (*see* Subheading 2.2, **item 1**), 100 μL of diluted (1:50–100) pre-inoculum culture of OP50 (Subheading 3.1.1) droplet is placed and spread around by swirling the plate gently, such that the droplet is dispersed into a circular lawn in the center. Incubate this plate overnight at 37 °C. The bacteria grow as a lawn in the center of the plate. Leave the plate outside at room temperature for 2 days and allow the lawn to get thicker. Use this plate for maintaining the *C. elegans* strains. The bacterial lawn serves as the food source.
7. The timing described in this protocol has been determined for N2 animals. Many mutant strains develop considerably slower than N2, and time adjustments will be necessary for transferring (Subheading 3.2.1), synchronizing (Subheading 3.2.3), and adding FUDR (Subheading 3.2.4). These adjustments in timing are best made based on the morphological characteristics of the different larval stages.
8. Different strains of *C. elegans* have distinct sensitivity to bleach. While the 3–5-min bleaching works well for N2 worms, some strains are inherently more or less sensitive to bleach and may require lesser or longer time for disruption with bleach solution. Under-bleaching, a failure to disrupt the worms to release the eggs, will result in carcasses with multiple eggs inside. Any of these carcasses inadvertently included in a well of a 96-well plate will dramatically increase the number of worms in that well. Overbleaching will dramatically reduce the viability of the eggs. To avoid under- and overbleaching, monitor the worms every minute after adding bleach solution, and neutralize with M9 buffer immediately after most of the worm bodies have disrupted in two halves. We have also noted that as the household bleach becomes older, it becomes less effective.
9. Too many M9 washes can lead to loss of eggs, as the eggs stick to the side of the tube and pipettes. If a preparation has a lot of worm debris after bleach disruption and high egg counts, wash twice. If the preparation has low egg counts, wash once.
10. Worm drops can be placed on a clear surface such as the lid to a transparent 96-well plate or a Petri dish.
11. On day [–2] (Subheading 3.2.3, **step 1**), when the synchronized larvae are counted to determine the worm concentration, unhatched eggs may be observed. The viability of eggs can vary from strain to strain, and without prior knowledge, it may be difficult to decide whether these eggs will hatch and

contribute to the worm concentration or not. As a rule of thumb, we recommend to assume that 50 % of the eggs will hatch and contribute to the final worm concentration. This adjusted measure avoids unexpected increases or decreases in the number of worms per well.

12. Tape sealers with perforations for gas exchange will not work for this step. The wells are sealed to (a) prevent the liquid from drying out during the assay and (b) to prevent contamination, both essential for screens that run for several days.
13. FUDR should be added to L4-stage animals. Adding FUDR too early can cause developmental defects. The precise developmental timing varies dependent on strain and incubation temperature. FUDR intercalates into sperm DNA, thereby impairing the sperm to fertilize the eggs.
14. While adding FUDR to the wells using a multichannel pipette, make sure that the entire droplet of FUDR stock has been dispensed into the worm culture. If the final concentration of 0.12 mM FUDR is not achieved due to pipetting error, the eggs laid by the worms would become viable, leading to hatching of offspring, which would then render the well useless because of the increase in total number of worms by the offspring of a lesser age.
15. For a long-term assay such as life span, depriving the animals of oxygen is prevented by keeping the liquid volume to half the maximum well volume. For example, in a 96-well plate with a maximum liquid volume capacity of 300 μL per well, seed 150 μL or less final culture volume.
16. Similar to the point addressed in **Note** 7, depending on the development of the strain, time adjustments may be made for compound adding, if a strain develops considerably slower than N2. The exact stages of *C. elegans* development are well defined and can be distinguished morphologically (*see* www.wormbook.org [or] www.wormatlas.org). These morphological characteristics provide the ability to distinguish slow-developing *C. elegans* strains from strains that develop at wild-type rates, but are smaller or thinner as adults.
17. The timing of when the compound library is added to the culture depends very much on the phenotype of interest. Adding the compounds at the stage outlined in the protocol serves merely as an example.
18. Most compound libraries are dissolved in DMSO. DMSO concentrations cannot exceed 0.55 % v/v (0.8 μL DMSO per 150 μL culture) without long-term effects on the worms. DMSO at or below final concentrations of 0.33 % v/v has minimal effects on development and life span of *C. elegans.*

DMSO of low purity (<99.9 %) can substantially blunt the effect of a compound, and we recommend using DMSO of ≥99.9 % purity for compound preparation.

19. To prevent deprivation of oxygen to the animals, sealers should be changed once a week until the assay is completed.
20. Carbenicillin and amphotericin B added in the culture should prevent bacterial and fungal contamination, respectively. The highest chance of contamination occurs during refeeding the worms on day [5], Subheading 3.2.7, and during weekly changing of sealers. If contamination is observed, add 50–100 μL of mineral oil to the well to seal it off. Ideally, the oil is added through a small incision made in the sealer before the sealer is removed and replaced to prevent spreading of the contamination.
21. Certain phenotypes may be influenced by the number of worms per well. The amount of feeding bacteria (6 mg/mL) that is seeded into the plate is generally enough for up to 15 animals to reach the post-reproductive stage. If more than 15 animals are seeded and/or the animals are kept for extended periods, they will run out of feeding bacteria and starve, which can affect the phenotype of interest. It is therefore advisable to exclude wells with more than 15 worms/well for the analysis, unless it has been specifically tested that high worm numbers do not affect the phenotype of interest.
22. During routine handling of assay plates, it is possible to accidentally tilt the plates, causing liquid to touch the tape sealer. If this happens, plates should be gently spun down in a 96-well plate centrifuge, at $180 \times g$ for 15 s, at room temperature, to aid the liquid to gravitate back into the well. Liquid on the sealers is something to avoid and is important to be quickly scanned for and taken care of before changing sealers and refeeding plates. Liquid spraying from opening the sealers leads to changes in concentration and loss of worms, and increases the chance of contamination.

Acknowledgements

This work was supported by grant funding to M.P., from the NIH (DP2 OD008398), a grant from the Ellison Medical foundation (AG-NS-0928-12), an MDA Development Grant for S.R., and an NSF GRFP Fellowship for G.M.S. Strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440).

References

1. Haas J et al (2004) In vivo assay guidelines. In: Sittampalam GS et al (eds) Assay guidance manual. Eli Lilly & Company, Bethesda, MD
2. Cabreiro F et al (2013) Metformin retards aging in *C. elegans* by altering microbial folate and methionine metabolism. Cell 153: 228–239
3. Zhang JH, Chung TD, Oldenburg KR (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen 4:67–73
4. Solis GM, Petrascheck M (2011) Measuring *Caenorhabditis elegans* life span in 96 well microtiter plates. J Vis Exp 49
5. Brenner S (1974) The genetics of *Caenorhabditis elegans*. Genetics 77:71–94
6. Petrascheck M, Ye X, Buck LB (2007) An antidepressant that extends lifespan in adult *Caenorhabditis elegans*. Nature 450:553–556
7. Petrascheck M, Ye X, Buck LB (2009) A high-throughput screen for chemicals that increase the lifespan of *Caenorhabditis elegans*. Ann N Y Acad Sci 1170:698–701
8. Ye X et al (2014) A pharmacological network for lifespan extension in *Caenorhabditis elegans*. Aging Cell 13:206–215
9. Kwok TC et al (2006) A small-molecule screen in *C. elegans* yields a new calcium channel antagonist. Nature 441:91–95
10. Leiers B et al (2003) A stress-responsive glutathione S-transferase confers resistance to oxidative stress in *Caenorhabditis elegans*. Free Radic Biol Med 34:1405–1415

Chapter 12

Whole-Organism Screening for Modulators of Fasting Metabolism Using Transgenic Zebrafish

Philipp Gut and Didier Y.R. Stainier

Abstract

Organismal energy homeostasis is maintained by complex interorgan communications making the discovery of novel drugs against metabolic diseases challenging using traditional high-throughput approaches in vitro. Here, we describe a method that rapidly identifies small molecules with an impact on organismal energy balance in vivo. Specifically, we developed a whole-organism screen for modulators of fasting metabolism using transgenic bioluminescence-reporter zebrafish for the gluconeogenic gene *phosphoenolpyruvate-carboxykinase 1 (pck1)*.

Key words Zebrafish, Gluconeogenesis, Diabetes, Metabolic disease, Fasting metabolism, Small-molecule screening, Bioluminescence, Phosphoenolpyruvate-carboxykinase, *pck1*, Transgenesis

1 Introduction

Small molecules that mimic physiological responses toward fasting or calorie restriction are believed to exert beneficial effects on aging-associated pathologies, including cardiovascular disease, cancer, and neurodegenerative conditions [1–4]. The identification of novel drugs with these properties is difficult due to homeostatic feedback loops across organs that regulate whole-body energy balance [5]. In addition, most drugs that were discovered in target-based in vitro approaches fail when they are subjected to the xenobiotic defense mechanisms that occur in living animals. An emerging alternative approach is to design phenotype-based screens in vivo using whole organisms, including C. *elegans* and zebrafish [6–9].

We previously developed transgenic zebrafish that monitor the transcriptional activation of the fasting-inducible gluconeogenic gene *pck1* when larvae transition from a high-energy state to calorie deficit after the yolk is consumed [7]. This "yolk feeding-to-fasting transition" occurs between 4 and 6 days postfertilization (dpf), i.e., at a stage when larvae are small enough to be kept in 96-well plates. We chose gluconeogenesis as a readout for fasting

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_12, © Springer Science+Business Media New York 2015

because it is dynamically regulated by feeding and fasting via hormonal signals [10]. In addition, chronically elevated gluconeogenesis is a hallmark of diabetes, and drugs that lower hepatic glucose production are among the most important therapeutic options to control glycemia in diabetic individuals [11].

Bioluminescence reporter proteins are characterized by a high signal-to-noise ratio and can be rapidly quantified in 96-well format using a microplate reader. For our screen we engineered firefly luciferase under the control of the 2.8 kb promoter upstream of the start codon of *pck1*. The stable transgenic reporter zebrafish *Tg(pck1:Luc2,cryaa:mCherry)*^{S952} (hereafter named *Tg(pck1:Luc2)*) expresses a fluorescent eye marker cassette in addition to the *pck1:Luc2* reporter for easy selection of transgenic carriers.

The protocol described here spans over eight experimental days. First, transgenic embryos are generated by crossing wild-type and transgenic "*pck1* reporter" zebrafish. At 4 dpf, the transgenic larvae are distributed in 96-well plates and treated with small molecules. After a 48-h incubation time, the bioluminescence signal is measured and serves as a readout for *pck1* promoter activity.

This strategy allows one to rapidly probe the effect of approximately 800 chemicals as potential modulators of gluconeogenesis on 2,400 zebrafish larvae per round of screening (*see* Fig. 1 and Table 1). Below we describe a detailed outline for this procedure. The technology we present can be modified as a rapid and sensitive readout for any gene of interest with a dynamic upstream promoter, opening avenues for novel drug discovery strategies beyond the application described here.

2 Materials

2.1 Husbandry and Transgenic Zebrafish

1. Egg water for embryos: Add 60 mg of Instant Ocean Sea Salt (Instant Ocean Blacksburg, VA, USA) to 1 L of distilled H_2O. Supplement with 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) for the egg water used in the 96-well screening plate (*see* **Note 1**).
2. Wild-type strain AB or TL and zebrafish.
3. 100 mm × 20 mm cell culture petri dishes.
4. Zebrafish anesthesia: Prepare a stock solution of 400 mg of tricaine (ethyl 3-aminobenzoate methanesulfonate) in 97.5 mL of egg water.
5. Microscopy: Fluorescence stereoscope, for example ZEISS Discovery V8 with mCherry or dsRED fluorescence filter set (Carl Zeiss AG, Oberkochen, Germany).

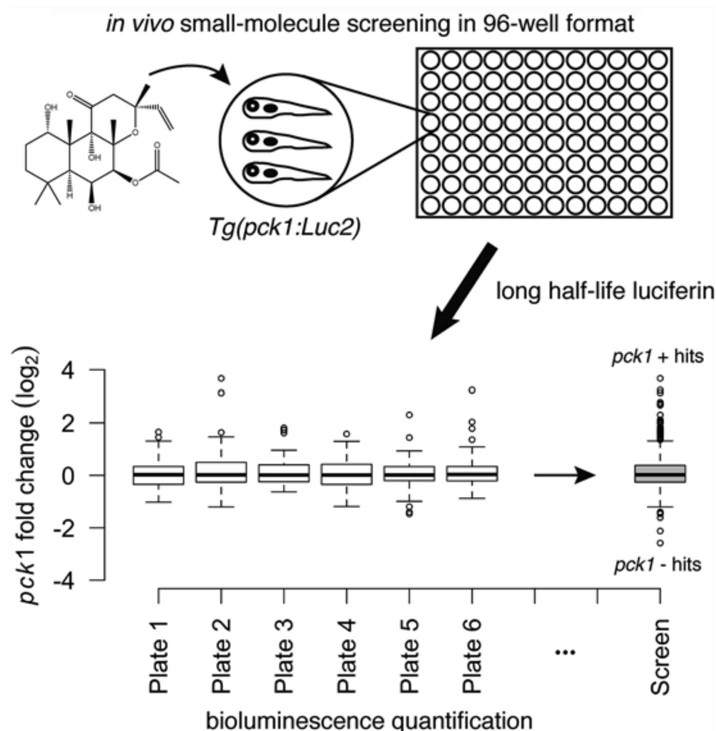


Fig. 1 In vivo small-molecule screening for modulators of fasting metabolism. Three transgenic *Tg(pck1:Luc2)* zebrafish larvae are distributed at 4 dpf into a 96-well plate. Chemicals from a small-molecule collection are added to each well at a final concentration of 10 μM. A long half-life luciferin is added after 48 h of incubation to dissolve the fish and to release a bioluminescent signal that can be detected using a luminometer. This semi-automated approach can rapidly identify drugs that modulate the promoter activity of the fasting-inducible gene *pck1*

Table 1
Workflow of small-molecule screen

Experimental day	Zebrafish stage	Task	Time effort
1		• Set up 10–30 crosses to obtain 1,000–3,000 embryos	20–40 min
2	0 dpf	• Collect and clean embryos	30–120 min
6	4 dpf	• Verify eye marker in 100 % of the progeny • Sort for healthy embryos • Distribute in 96-well plates • Treat with small molecules	60–180 min (~15 min to load 1 plate with 240 larvae and to add compounds)
8	6 dpf	• Visually check wells for toxicity • Add long half-life luciferin + incubate for 1 h • Read *pck1* promoter activation with luminometer	60–180 min

2.2 Small-Molecule Libraries and Drug Treatments

1. Small-molecule libraries: Bioactive small-molecule libraries Tocriscreen Mini screening library (1,120 compounds, Tocris Bioscience, Bristol, England), Sigma Lopac (1,280 compounds, Sigma Chemical Company, St. Louis, MO, USA) (*see* **Note 2**). Aliquot in 1 mM stock plates using dimethyl sulfoxide (DMSO) as solvent and keep at −20 °C.
2. Screening plates: Opaque-walled 96-well microplates with clear bottom (*see* **Note 3**) and lid (Costar/Thermo Fisher Scientific, Fremont, CA, USA).
3. Small-molecule storage plates: 96-Well deep well plate, polypropylene, 1.1 mL, round bottom with lids (Cole-Parmer, Vernon Hills, IL, USA).
4. Disposable polystyrene pipetting reservoir.

2.3 In Vivo Bioluminescence Screening

1. SteadyGlo Luciferase assay system (long half-life luciferin) (Promega, Madison, WI, USA). Store in 5 mL aliquots at −80 °C.
2. Microseal "F" Film PCR Sealer (Bio-Rad Laboratories, Hercules, CA, USA).
3. Bioluminescence 96-well microplate reader (GloMax 96-well Luminometer, Promega, Madison, WI, USA).

3 Methods

3.1 Generation of Transgenic Reporter Zebrafish (Experimental Days 1 and 2)

1. Set up 10–30 crosses of homozygous *Tg(pck1:Luc2)* with wild-type zebrafish to generate 1,000–3,000 embryos that are hemizygous for the *pck1* reporter transgene (*see* **Note 4**). For each cross use one male and one to two females.
2. Collect embryos into 100 mm petri dishes and place in an incubator at 28.5 °C until the afternoon of 0 dpf.
3. Transfer the embryos into clean petri dishes with a density of 80–100 per petri dish and cover with ~40–50 mL of egg water.
4. Incubate the embryos at 28.5 °C until 4 dpf. Exchange the egg water of each plate once between 2 and 3 dpf to remove chorion debris (optional).

3.2 Administration of Small Molecules (Experimental Day 6)

1. At 4 dpf verify expression of the red fluorescence marker in the eye that indicates transgenic *Tg(pck1:Luc2)* reporter zebrafish using a fluorescence stereoscope.
2. Pour the transgenic embryos through a fine strainer and rinse thoroughly with clean egg water. Collect embryos in a clean petri dish.
3. Use 2 mL of tricaine solution for ~40 mL of egg water to anesthetize animals.

4. Transfer healthy larvae into a pipetting reservoir (*see* **Note 5**). Exchange the medium with HEPES-buffered egg water several times to eliminate the tricaine (*see* **Note 6**).
5. Distribute three larvae per well into a 96-well plate using a P200 pipette adjusted to a 200 μL volume. Cutting a small piece off the end of the tip helps to capture the embryos more easily. Load the centered 80 wells with larvae for drug treatments (*see* **Note 7**). The left and right columns can be loaded with larvae to include positive and negative controls. Alternatively these columns can be kept empty if the bioluminescence value is normalized to the plate median to retrieve fold changes (*see* **Note 8**).
6. Transfer 2 μL of small molecules from the 1 mM concentrated stock plate into the screening plate using a multi-pipette (1:100 dilution for a final screening concentration of 10 μM in 1 % DMSO). A beta-adrenergic agonist, for example isoprenaline, or DMSO can be used as positive and negative control, respectively (optional).
7. Cover with a lid and wrap the 96-well plate in aluminum foil to protect it from light. Place the plate for several minutes on a horizontal shaker.
8. Incubate for approximately 48 h until 6 dpf at 28.5 °C.

3.3 Analysis (Experimental Day 8)

1. Thaw the SteadyGlo luciferin reagent. Keep it on ice once thawed and protect from light.
2. Carefully remove 100 μL of medium from the small-molecule-treated plate using a multi-pipette. Make sure not to remove any larvae at this point.
3. Visually examine the larvae in each well for signs of drug-induced toxicity (*see* **Note 9**). Note wells that contain larvae with apparent signs of toxicity and exclude from the analysis. Note also deviation from three larvae per well to adjust accordingly the luminescence value obtained from the respective well (*see* **Note 10**).
4. Cover the clear bottom of the 96-well plate with adhesive foil using Microseal "F" Films. Add 50 μL of SteadyGlo luciferin to each well using a multi-pipette. Place the plate on a horizontal shaker for 1 h at room temperature (*see* **Note 11**).
5. Measure the bioluminescence signal of each well using standard settings of the luminometer.
6. Data analysis: Normalize the reads per well and assign hit compounds (*see* **Note 8**) (*see* Table 2 for representative data).

Table 2
Representative data: Hit compounds of plate 8 of the Tocriscreen Mini bioactive screening library

Compound name	Annotated target	Plate	Well	Raw signal run 1/ plate median	Raw signal run 2/ plate median	Fold change (log2)	Result
Isoproterenol hydrochloride	β-Adrenergic agonist	8	C3	128701/15771	82202/15441	2.753	*pck1* + hit compound
Ibudilast	Nonselective phosphodiesterase inhibitor	8	A10	65878/15771	69167/15441	2.114	*pck1* + hit compound
YM 90709	Interleukin-5 receptor antagonist	8	A4	26127/15771	99037/15441	2.013	*pck1* + hit compound
Raclopride	D2/D3 receptor antagonist	8	E4	31624/15771	64074/15441	1.622	*pck1* + hit compound
NSC 663284	Cdc25 phosphatase inhibitor	8	F8	4718/15771	2723/15441	–2.072	*pck1* – hit compound

Two independent runs were performed

4 Notes

1. We found that using HEPES increased the robustness of the screen as low pH values can affect the health of the larvae. In addition, variations in pH likely affect the pK_a value of a chemical and therefore its uptake.
2. A large variety of small-molecule libraries are commercially available for phenotypic screens in zebrafish. Whole-organism screens are commonly of lower throughput than conventional screens. Using bioactive libraries or collections of clinically approved drugs increases the likelihood of identifying a hit compound in a low- to medium-scale screen (500–5,000 compounds). For bioactive collections, a defined target is commonly annotated for each chemical, which can help to find the target associated with a phenotype. Aliquots of the original plates can be kept at −20 °C for several independent screens. Avoid multiple freeze-thaw cycles. Spin-down the plates before use to avoid spilling drops of chemicals that adhere to the lid. Furthermore, we recommend deep-well 96-well plates to avoid cross-contamination during freeze-thaw cycles.
3. Clear-bottom 96-well plates allow one to visually assess the health of larvae after the incubation with small molecules and before proceeding to the bioluminescence assay.
4. When setting up a large number of crosses it is advantageous to keep populations of one genotype in a gender and/or background strain different than that of the wild-type animals. The morphological differences (male-female or AB-TL strains) help to separate the wild-type from the transgenic zebrafish rapidly when returning them into their tanks. For example, all homozygous *Tg(pck1:Luc2)* zebrafish can be used as males in a TL background strain and crossed into females of the AB strain. This strategy diminishes the risk of contamination of one population with the other and guarantees a clean population of hemizygous reporter animals for downstream steps over the course of the screen.
5. Sorting larvae at late 4 dpf by the characteristic of an inflated swim bladder and healthy appearance ensures a cohort of healthy animals, thereby decreasing false-positive mortality events and the variance of the bioluminescence signal.
6. To remove the anesthetic and to replace the water with HEPES-buffered medium, exchange as much of the volume as possible several times without discarding animals.
7. Commercial small-molecule 96-well storage plates commonly contain 80 chemicals leaving columns 1 and 12 empty. Loading matching 80 wells with larvae makes the transfer of compounds from the stock plate to the screening plate fast and easy using

a multi-pipette. It takes about 10 min for a trained person to distribute three larvae in each well of a 96-well plate making this task one of the limiting steps of the screen.

8. Each luciferase value should be normalized against control values to obtain fold changes associated with a given chemical treatment. The normalization can be achieved by loading columns 1 and 12 with animals that serve as vehicle controls (1 % DMSO). Alternatively, we recommend the following normalization procedure to save time and larvae: Rank-sort all 80 light-unit values and exclude the ten highest and the ten lowest numbers. Compute the median of the remaining 60 values and use this number to normalize all reads of a plate. We recommend at least one repeat for each chemical screened to reduce the number of false positives before considering drugs for secondary assays and further validation (*see* Table 2 for representative results).
9. Toxic compounds should be excluded from the analysis. Typical signs of drug-induced toxicity are a deflated swim bladder, pericardial edema, or morphological malformations.
10. When analyzing the plate for toxicity, verify that three larvae have been placed correctly in each well. Note if more or less animals have been placed so that the luminometer value can be adjusted to the actual number of larvae per well. Plates that have been manually loaded frequently contain a few wells with the incorrect number of larvae.
11. SteadyGlo luciferin has a half-life of several hours (5 h according to the manufacturer). It contains a strong detergent, which readily dissolves the larvae. We recommend an incubation time between 45 min and 2 h to obtain robust values.

Acknowledgements

This study was supported in part by a postdoctoral fellowship DFG GU 1082/101 from the German Research Foundation to P.G., NIH grant RO1 DK60322, a pilot and feasibility award from the University of California—San Francisco diabetes center funded by NIH U01 DK089541 and the Packard Foundation to D.Y.R.S.

References

1. Baur JA et al (2006) Resveratrol improves health and survival of mice on a high-calorie diet. Nature 444:337–342
2. Le Couteur DG et al (2012) Aging biology and novel targets for drug discovery. J Gerontol A Biol Sci Med Sci 67:168–174
3. Guarente LL (2006) Sirtuins as potential targets for metabolic syndrome. Nature 444: 868–874
4. Anson RM et al (2003) Intermittent fasting dissociates beneficial effects of dietary restriction on glucose metabolism and neuronal resistance

to injury from calorie intake. Proc Natl Acad Sci U S A 100:6216–6220

5. Moller DE (2012) Metabolic disease drug discovery- "hitting the target" is easier said than done. Cell Metab 15:19–24
6. Lemieux GA et al (2011) A whole-organism screen identifies new regulators of fat storage. Nat Chem Biol 7:206–213
7. Gut P et al (2013) Whole-organism screening for gluconeogenesis identifies activators of fasting metabolism. Nat Chem Biol 9: 97–104
8. Andersson O et al (2012) Adenosine signaling promotes regeneration of pancreatic β cells in vivo. Cell Metab 15:885–894
9. Zon LI, Peterson RT (2005) In vivo drug discovery in the zebrafish. Nat Rev Drug Discov 4:35–44
10. Lin HV, Accili D (2011) Hormonal regulation of hepatic glucose production in health and disease. Cell Metab 14:9–19
11. Rena G, Pearson ER, Sakamoto K (2013) Molecular mechanism of action of metformin: old or new insights? Diabetologia 56:1898–1906

Chapter 13

High Content Screening for Modulators of Cardiovascular or Global Developmental Pathways in Zebrafish

Charles H. Williams and Charles C. Hong

Abstract

Major developmental pathways play critical roles in wide array of human pathologies. Chemical genomic screening allows for the discovery of novel tools not only to target known pathway interactors but also to discover new, chemically tractable targets for known pathways. The zebrafish has emerged as a useful model for developmental biology and has been well characterized. The zebrafish represents a hardy conglomerate of totipotent cells that are massively and simultaneously assessing all significant pathways in parallel in an endogenous context. This represents a gold standard for biological assays, chemically targeting select pathways without causing pleiotropic effects. Here, we describe methods used to develop high content screening assays implementing transgenic zebrafish to assess phenotypic changes that have identified several classes of novel compounds that effect developmental pathways.

Key words Zebrafish, Small molecules, Phenotypic screening, High content screening, Cardiovascular development

1 Introduction

Since zebrafish first emerged onto the scene of academic research it has had a close relationship with development and chemicals. In the late 1950s the zebrafish was used in teratological studies investigating the effect of chemicals such as actetylaminofluorene and zinc sulfate on development [1, 2]. Genome-wide mutagenesis screens have generated a library of mutants that have been published, providing novel insights into various aspects of development including melanocyte development, neural development and cardiovascular development [3–10]. This strategy calls for random mutagenesis of the organism's genome in F0, and subsequently breeding the heterozygous F1 generation to have a population of homozygous mutants in F2. This mutagenesis strategy has been successfully employed in a variety of other organisms including but not limited to *C. elegans*, *Drosophila* and *Medaka*. However, this approach has a few limitations: the most prominent would be the

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_13, © Springer Science+Business Media New York 2015

lack of dominant effectors where putatively embryonically lethal or sterilizing dominant mutations cannot be bred. Secondly, there are logistical difficulties in screening for maternally contributed components. Finally, functional redundancies of homologues such as those created by genome duplication during evolution can make investigation of a gene's function difficult.

Utilizing chemical entities to perturb a protein's function offers multiple advantages to that of more traditional mutagenesis screening methods. The first of these advantages include a screening at F0, which means that the effect is seen immediately with no need for breeding the mutation to homozygosity. This allows the study of critical proteins that where loss of homozygosity could be lethal or sterilizing. Furthermore, since chemicals target functionality irrespective of the origin, they target maternally deposited proteins and transcript products with the same ease as zygotic products, allowing for easier perturbation of maternal components. Secondly, small molecules are scalable both temporally and in concentration. Unlike mutations that are immutable and persistent, the scalability of small molecules gives one the ability to transiently study critical protein functions at different time points and simple dose dependent interrogation of protein activity. Finally, novel pharmacophores can serve as a starting point for the analogues that have therapeutic potential.

Although this phenotypic screening approach allows an unbiased screening platform at medium to high-throughput, it is limited by target identification. Our lab uses a chemical library enriched for chemical structures with no known molecular target but predicted biological activity. We treat embryos arrayed in a 96-well plate with 3–5 embryos per well at an initial concentration of 10 μM at 5 h post fertilization (hpf) (Table 1). This screen has

Table 1
Experimental workflow

Experimental timing	Zebrafish stage	Task	Time effort
–1		Set up 10–30 crosses to obtain 1,000–3,000 embryos	20–40 min
0	0 dpf	Collect and clean embryos Array embryos at 3 hpf Treat embryos at 6 hpf	90–160 min (~20 min to load 1 plate with embryos.)
1	24 hpf	Observe live/dead Remove dead embryos Document deviants found in BF, Green and red channels	30–120 min
2	48 hpf	Document deviants found in BF, Green and red channels	30–120 min

allowed us to identify numerous classes of targets across multiple pathways ranging from kinase inhibitors such as Dorsomorphin to the histone acetyl transferase inhibitor Windorphen [11–13]. Others have successfully utilized similar methodologies to identify cardiogenic compounds [14], and modulators of pigment cell formation [15]. Here we provide a method for chemical genomics applied to zebrafish to find novel modulators of either global development or cardiovascular development (*see* **Note 1**).

2 Materials

1. Minimum of 20 pairs of adult Tg(cmlc2:mcherry; fli:egfp). These double transgenic fish have a red fluorescent protein driven by a heart specific cmlc2 (cardiac myosin light chain 2) promoter, and an enhanced green fluorescent protein driven by the endothelial promoter for fli1a (friend leukemia integration 1a).
2. Fish nets.
3. Breeding tanks, with removable inner container and dividers.
4. Petri dishes (10 cm).
5. Plastic tea strainer.
6. Wash bottle containing embryo water.
7. Disposable polyethylene transfer pipettes.
8. Polystyrene 96-well round-bottom assay plates.
9. Glass Pasteur pipette.
10. Manual pipette pump, 10 mL.
11. 60× E3 embryo medium: Add 34.8 g of NaCl, 1.6 g of KCl, 5.8 g of $CaCl_2 \cdot 2H_2O$, and 9.78 g of $MgCl_2 \cdot 6H_2O$ to 2 L of H_2O. Adjust the pH to 7.2 with HCl and autoclave.
12. 10× 1-Phenyl-2-Thiourea (PTU): Dissolve 0.3 g of PTU in 1 L of 1× E3 embryo media. Solutions containing PTU should be protected from light by covering with aluminum foil.
13. Embryo Screening Medium: 5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$, 0.003 % 1-Phenyl-2-Thiourea (PTU). Combine 16.5 mL of the 60× E3 embryo medium, 100 mL of 10× PTU, 0.5 g of Kanamycin and fill to 1 L with H_2O.
14. 12-Channel pipettes, 2–20 μL and 12-channel pipettes, 3–300 μL.
15. Disposable polystyrene pipette basin, 50 mL.
16. Small molecule library of structurally diverse compounds arrayed in a 96-well format as 10 mM stocks in DMSO. Aliquot each master plate into 96-well polypropylene storage plates, and store at −80 °C until use.

17. Aluminum sealing tape for 96-well plates (Nunc, Rochester, NY).
18. DMSO.
19. Desiccation chamber containing Drierite (W.A. Hammond Drierite Co., Xenia, OH).
20. Incubator at 28.5 °C.
21. Stereomicroscope with transmitted light base, and fluorescence attachment (Leica Microsystems, Bannockburn, IL).

3 Methods

3.1 Zebrafish Egg Collection

1. On the afternoon prior to the day of the chemical screen, set up 10–20 zebrafish breeding tanks. Fill each tank with water from the aquaculture system.
2. Using a fish net, transfer one adult male and one to two adult females to the inner container in each breeding tank. Separate the male and female fish from each other with a divider. Label the cages and put a lid over them.
3. On the morning of the screen, remove the dividers from the breeding tanks and allow zebrafish to mate. Over the course of the next 1 h, allow fertilized eggs to fall through grid at the bottom of each inner container.
4. After 1 h, return the adult zebrafish back to permanent storage tanks, remove the inner container and collect the eggs by straining the water in each breeding tank through a plastic tea strainer.
5. Invert the strainer over a Petri dish and rinse the strainer gently to flush the eggs into the Petri dish by using a wash bottle containing the 1× Embryo screening medium.
6. All unfertilized eggs, which appear opaque, should be removed using a disposable plastic pipette. Each mating cross should yield approximately 200 embryos.
7. Place embryos into a 28.5 °C incubator until the embryos reach 1,000 cell stage, approximately 3 h post fertilization (hpf).

3.2 Treating Embryos in 96-Well Plates

1. Using a P200 pipette set to 100 μL with the tip clipped to widen the bore, transfer 3–5 embryos in embryo screening medium into each well of a 96-well plate.
2. Put the 96-well plates into a 28.5 °C incubator until the embryos reach 50 % epiboly stage, approximately 5 h post fertilization (hpf) (*see* **Note 2**).
3. Small molecule libraries are typically supplied in a 96-well format, with each compound stored in DMSO as a 10 mM stock. About 60 min before the embryos reach the stage when the

compounds are to be added, thaw a desired number of 96-well plates containing aliquots of small molecules (source plate). Take note of the serial or other identification number of the source plates. To minimize condensation on the plates, thawing can occur in a desiccation chamber.

4. Briefly spin down the compound source plates at 300 × *g* for 2 min in a tabletop centrifuge equipped with multi-well plate adaptor.
5. Remove the aluminum sealing tape from source plate. Using a 12-channel pipette, dilute the compounds in the source plate to the concentration of 1 mM (for example, if starting with 250 nL aliquots of 10 mM stock, add 2.25 μL of DMSO to each well).
6. When the embryos in the 96-well plate (recipient plate) reach the desired stage for treatment, 6 hpf for this protocol, use a 12-channel pipette (0.1–2.5 μL) to transfer 1.0 μL of compounds (1 mM) from the source plates into the recipient plates containing the embryos (*see* **Note 3**).
7. Record the identification number of the source plates on the recipient embryo plates. Cover the recipient plates now containing the embryos and compounds with lids, gently mix the plates by swirling, and place them in a 28.5 °C incubator.
8. Cover each source plate containing unused small molecules (1 mM) with aluminum sealing tape and place them in a −80 °C freezer for long-term storage.

3.3 Screening for Effects of Small Molecules by Visual Inspection of Phenotypes

1. Prior to performing the screen, formulate a specific criterion for what will constitute a "hit." The criteria described herein are a subset of possibilities using the transgenic reporter and morphological differences (Table 2).
2. At 24 hpf, remove the 96-well plates containing compound-treated embryos from the incubator and examine each well under a stereomicroscope using transmitted light. This time point is good for initial toxicity and live/dead assessment. Gross morphological perturbations, such as potent dorsalization or ventralization can be seen at this point (*see* **Notes 4** and **5**).
3. Switch to fluorescent lighting and investigate formation of vasculature on the green channel. Primary axial vessels should be present and support circulation.
4. Quickly scan the 96-well plates for any well in which at least 2 of 3/3 of 4/3 of 5 embryos exhibit the prescribed "hit" phenotype. Record the identity of the plate and the well location of each potential hit.
5. Switch to the red channel and investigate form and function of the heart tube.

Table 2
Definition of subset of possible "hit" phenotypes observable in this screen

	Bright field	Red channel cmlc2:mcherry	Green channel Fli:EGFP
5 hpf (50% Epiboly)	Treatment/General health	N/A	N/A
24 hpf	Circulation Axis length Eye formation	Heart tube size Contractility	Axial vessel formation
48 hpf	Pigmentation Otic placode size Otolith formation Hemorrhage	Heart looping Heart size Tachycardia Bradycardia	Formation of ISV Formation of DLAV Vascularization of retina

6. Quickly scan the 96-well plates for any well in which at least 2 of 3/3 of 4/3 of 5 embryos exhibit the prescribed "hit" phenotype. Record the identity of the plate and the well location of each potential hit and place plates back in incubator.
7. At 48 hpf, remove the 96-well plates containing compound-treated embryos from the incubator and examine each well under a stereomicroscope using transmitted light. This time point is good for initial motility assessment as embryos twitch randomly at this age.
8. Switch to fluorescent lighting and investigate formation of vasculature on the green channel. Secondary vessels, ISV (intersegmental vessels), and DLAV (dorsal longitudinal anastomotic vessel) should be formed at this stage.
9. Switch to the red channel and investigate form and function of the two chamber heart.
10. Quickly scan the 96-well plates for any well in which at least three out of five embryos exhibit the prescribed "hit" phenotype. Record the identity of the plate and the well location of each potential hit and place plates back in incubator (*see* **Note 6**).
11. Reconfirm a potential hit by retesting the effects of the compound at several doses (1, 5, 10, and 50 μM). For each dose, ten embryos are tested in 0.5 mL of E3 medium in a 48-well plate format. The timing of compound addition for retesting should be identical to that of the original screening. A hit is confirmed when the elicited phenotype is reproduced on retesting of the compound.
12. Identify the hit compound from the database of small molecules in the chemical library (*see* **Note 7**).

4 Notes

1. This is a simple protocol that is easily adjusted for other organ systems. There are numerous transgenic lines with tissue-specific promoters that can help visualize specific anatomical features. For more information about obtaining lines of transgenic zebrafish please contact ZIRC (Zebrafish International Resource Center; Website: http://zebrafish.org).
2. Successive inbreeding of laboratory zebrafish can give rise to embryos that are abnormally patterned. Occasionally poor nutrition will also have effects on progeny. Therefore, it is always best to leave a small handful of embryos (10–15) untreated to observe any baseline morphological defects in the sample population.
3. Treating zebrafish embryos with DMSO concentrations above 1 % v/v can cause nonspecific effects that mirror the desired phenotype. When treating zebrafish embryos, in this and subsequent studies, keep DMSO at no more than 1 %. It is also important to ensure that DMSO is handled properly, as contaminants in DMSO caused by poor handling can increase rates of false positives.
4. Initial observation 24 h after treatment gives significant time for development to progress. The timing of treatment and subsequent observation can be adjusted according to what part of the zebrafish the screen is interested in observing. However, screening done on ages above 6 days post fertilization (dpf) becomes difficult as the embryos start needing to switch completely off of their yolk for nutrients and need to be fed.
5. If embryos are all dead in a well after 24 h of treatment the compound in the well could be toxic due to a nonspecific mechanism. However, if counter screening on embryos that are older results in a normal nontoxic phenotype, one can assume that the embryos died during earlier development because of a specific interaction.
6. This screening protocol uses qualitative measures as a read out. However, many high content screening (HCS) tools have been developed which are capable of imaging zebrafish in 96-well formats and could result in more quantitative measures for the phenotype screened.
7. General workflow after identifying the compound is reasonably standard. First, purchase the compound from the supplier to confirm a phenocopy and to confirm that the compound is not mis-annotated. It is also important to independently synthesize the compound personally or through a third party to be

sure that the product is indeed not contaminated with impurities from the synthesis and purification processes. Subsequently, investigations should move onto target identification, if the target is un-annotated, and mechanistic studies.

Acknowledgements

C.C.H. is supported by the U.S. NIH/NHLBI (R01HL104040), Veterans Administration Merit Award (101BX000771), and the Cali Family Foundation.

References

1. Hisaoka KK (1958) The effects of 2-acetylaminofluorene on the embryonic development of the zebrafish II. Histochemical studies. Cancer Res 18:664–667
2. Skidmore JF (1965) Resistance to zinc sulphate of the zebrafish (Brachydanio rerio Hamilton-Buchanan) at different phases of its life history. Ann Appl Biol 56:47–53
3. Choi T-Y et al (2007) Zebrafish as a new model for phenotype-based screening of melanogenic regulatory compounds. Pigment Cell Res 20:120–127
4. George EL et al (1993) Defects in mesoderm, neural tube and vascular development in mouse embryos lacking fibronectin. Development 119:1079–1091
5. Haffter P et al (1996) The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. Development 123:1–36
6. Chen JN et al (1996) Mutations affecting the cardiovascular system and other internal organs in zebrafish. Development 123: 293–302
7. Driever W et al (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. Development 123:37–46
8. Kelsh RN et al (1996) Zebrafish pigmentation mutations and the processes of neural crest. Development 123:369–389
9. Malicki J et al (1996) Mutations affecting development of the zebrafish retina. Development 123:263–273
10. Schier AF et al (1996) Mutations affecting the development of the embryonic zebrafish brain. Development 123:165–178
11. Hao J et al (2008) Dorsomorphin, a selective small molecule inhibitor of BMP signaling, promotes cardiomyogenesis in embryonic stem cells. PLoS One 3:e2904
12. Hao J et al (2010) In vivo structure-activity relationship study of dorsomorphin analogues identifies selective VEGF and BMP inhibitors. ACS Chem Biol 5:245–253
13. Hao J et al (2013) Selective small molecule targeting β-catenin function discovered by in vivo chemical genetic screen. Cell Rep 4:898–904
14. Ni TT et al (2011) Discovering small molecules that promote cardiomyocyte generation by modulating Wnt signaling. Chem Biol 18:1658–1668
15. Colanesi S et al (2012) Small molecule screening identifies targetable zebrafish pigmentation pathways. Pigment Cell Melanoma Res 25:131–143

Part III

Small Molecules in Chemical Biology

Chapter 14

Extraction Methods of Natural Products from Traditional Chinese Medicines

Jingyu Zhang, Jianying Han, Ayokunmi Oyeleye, Miaomiao Liu, Xueting Liu, and Lixin Zhang

Abstract

In recent years, many research activities have focused on Natural Products (NPs) derived from Traditional Chinese medicines (TCMs), thus making a renaissance in the drug discovery process of TCMs. Maximizing the diversity of extracts from those plants is the key for the chemical biology process. Methods for the preparation and pretreatment of plant extracts are very important for further purification and discovery of active compounds present in minor quantities. In this chapter, two methods of extraction, including one of the most broadly applicable method (solvent extraction) and one newly developed technique (supercritical fluid extraction), have been described in detail.

Key words Traditional Chinese medicines, Extraction, Solvent extraction, Supercritical fluid extraction

1 Introduction

The use of Traditional Chinese medicines (TCMs) has a long history in China [1]. Because they are enriched with active compounds, TCMs are used to treat many kinds of diseases, not only in China but also in the USA, Australia, and European countries [2–6]. TCMs contain structurally diversified bioactive components, and as such could be considered as a natural combinatorial chemical library. Such bioactive constituents can be isolated and characterized from various plant parts including leaves, stems, flowers, and fruits. Therefore, TCMs represent a potential resource from which new bioactive compounds can be derived [7–10].

Natural products (NPs) from TCMs serve as a valuable reservoir for drug discovery [7]. More drug leads have been identified by newly developed screening methods: for instance, berberine has been highlighted as a combination agent for the treatment of fungal infections by high-throughput synergy screening developed by Zhang et al. [9]. Recently developed techniques for the isolation of

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_14, © Springer Science+Business Media New York 2015

NPs have significantly accelerated the discovery of novel substances from natural resources and this has subsequently enhanced the application of NPs as useful probes and molecular tools for the investigation of biological targets in cells. In addition, the short supply of clinically important molecules such as Taxol and Artemisinin from medicinal plants has challenged researchers to investigate their biosynthesis by metabolic engineering of microorganisms [11–14]. Hence, developing new and rapid isolation and identification techniques for active NPs would provide crucial and efficient approaches for the discovery of natural drug leads.

Extraction of crude extracts is the first and the most crucial step in the procedure for the isolation of active compounds from TCMs. A pipeline for the extraction process is composed of the following steps:

(a) Drying and milling of plant material (e.g., roots and barks) or uniformizing fresh material (e.g., leaves and flowers).
(b) Choosing an appropriate solvent for the extraction (e.g., polar solvent: water, methanol/ethanol; medium polar: acetone/ethyl acetate/chloroform/methylene chloride/diethyl ether; non-polar: hexane/petroleum ether).
(c) Choosing an appropriate extraction method: in this chapter, two methods of extraction have been described in detail, i.e., solvent extraction and supercritical fluid extraction.

1.1 Solvent Extraction

Solvent extraction is widely used in the process of natural products extraction due to its simplicity, inexpensive extraction apparatus, and adequately high extraction rates; however, high solvent consumption and thermal oxidative degradation of unstable compounds limit its development. Solvent extraction is a traditional method for the extraction of natural products from TCMs when compared with other extraction methods due to its simplicity and robustness in the laboratory [15–20]. Aqueous extraction and organic-solvent extraction could be applied based on the compound of interest. In general, a 60–90 % mixture of methanol/ethanol and water is the most economical and efficient solvent for the extraction. If the components have a more polar nature, water or a mixture of lower proportion of methanol/ethanol in water is selected as the extracting solvent to obtain the compounds of interest. While it is not suitable for constituents with non-polar nature, organic-solvent extraction is preferential instead. An efficient solvent for extraction should be of low toxicity, must vaporize easily at low temperatures, and should not form aggregates or decompose in the process [19]. The most commonly used solvents are ethanol, methanol, acetone, chloroform, methylene chloride, diethyl ether, or a mixture of some of them. For moderately polar to non-polar components, chloroform, methylene chloride, and diethyl ether could be chosen as the appropriate solvents; however, ethanol, methanol and acetone are normally used for the extraction

of the preparation of compounds with more polar properties. Besides, based on the extraction temperature, solvent extraction involves soaking at room temperature and heat reflux extraction; the latter temperature depends on the boiling point of the solvent chosen. Therefore, if the temperature is higher, the risk of degrading the compounds of interest also becomes greater.

1.2 Supercritical Fluid Extraction

Supercritical fluid extraction (SFE), a newly developed technique, is used for laboratorial and industrial purposes because it presents a series of advantages compared to the conventional extraction processes, especially for the extraction of thermolabile components [21, 22]. It was first presented as a patent for decaffeination of coffee [23]. Since then, SFE has been used for many years as an alternative extraction method, which causes less pollution to the environment. The concept of the critical point was defined in 1822 as the highest pressure and temperature at which a pure substance could exist in vapor-liquid equilibrium. Above this point, supercritical fluid (SCF) is formed. These qualities make SCFs have higher diffusivities and less degradation of solutes than ordinary solvents to extract active components.

Carbon dioxide is the most commonly used solvent because it is nontoxic, readily available, inexpensive, and especially easy to eliminate from extracted products. SFE with carbon dioxide is operated at ca. 40 °C, avoiding degradation of NPs when exposed to high temperatures and atmospheric oxygen. Other solvents also used include ethane, butane, pentane, ammonia, and water. The extracted efficiency of SCFs can be increased by adding modifier (entrainer or cosolvent) in the solvents. The modifier, such as 1–10 % methanol, can enhance the polarity of SC-CO_2 to extract moderately polar compounds. A diagram of typical SFE instrumentation is given in Fig. 1 [24].

2 Materials

Prepare all analytical grade reagents, chromatographic grade reagents, and solutions using ultrapure water (produced by a Milli-Q system, 18.2 MΩ, Millipore, Bedford, MA, USA). Prepare and store all reagents at room temperature (unless indicated otherwise).

1. Dried plant materials (5–200 g that can be determined by the size of extraction container).
2. Grinder (motor power: 220 V, 1,300 W).
3. Extraction solvents: For polar components, water or a mixture of lower proportion of methanol/ethanol in water, ethanol, methanol, and acetone. For medium polar and non-polar components, chloroform, methylene chloride, and diethyl ether.
4. Sieve (60 mesh).

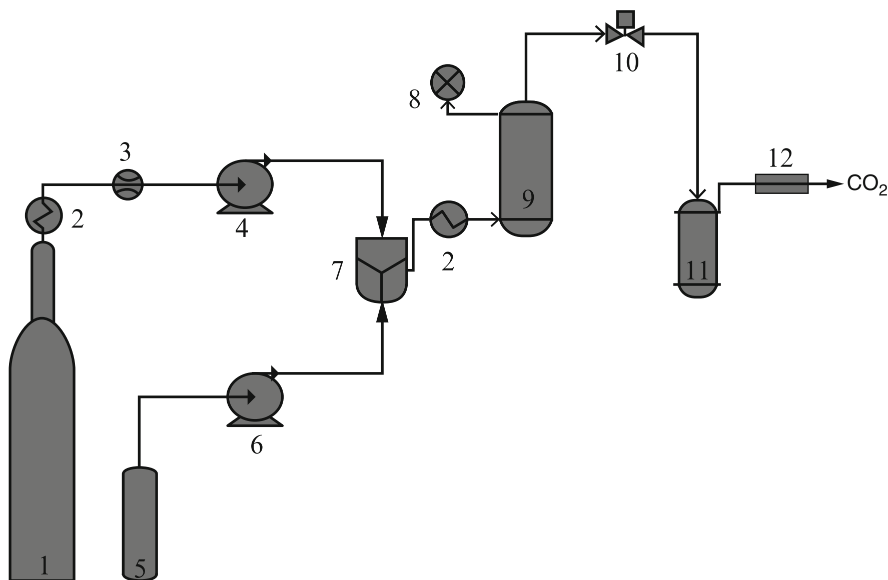


Fig. 1 The typical diagram of SFE apparatus. (1) CO_2 Tank; (2) Heat Exchanger; (3) Flow Meter; (4) CO_2 Pump; (5) Cosolvent vessel; (6) Cosolvent Pump; (7) Mixer; (8) Gauge; (9) Extraction Cell; (10) Control Valve; (11) Yield Trap; (12) Wet/Dry Gas Meter (reproduced from ref. [24] with permission from Elsevier)

5. Rotary evaporator.
6. Vacuum pump.
7. Syringe filter (40 and 0.20 μm).
8. HPLC apparatus (two LC-20A micropumps (Shimadzu, Japan)) equipped with a UV–Vis detector.
9. Agilent ZORBAX SB-Aq (150 mm × 4.5 mm, 5 μm) column (Agilent, USA).
10. A commercial SFE system (Spe-ed SFE Laboratory System, 7071, Applied Separations, Allentown, PA, USA).
11. Cosolvent: add the cosolvent (methanol) using modifier pump (Spe-ed Max for the Helix, Applied Separations, Allentown, PA, USA).
12. Carbon dioxide.
13. HPLC apparatus (Agilent 1100 LC system including degasser, binary gradient pump, autosampler, column thermostat and diode array detector, Agilent Technologies, Waldbronn, Germany).
14. Kinetex C18 (100 mm × 2.1 mm, 2.6 μm, pore size: 100 Å) column (Phenomenex, Torrance, CA, USA).

3 Methods

3.1 Solvent Extraction of Crude Extracts

Here we take heat reflux extraction as an example for preparation of crude extracts.

1. Dry plant materials under ambient temperature for 48 h or longer to constant mass (*see* **Note 1**).
2. Mill the plant materials to powders by a grinder, and then filter with a 60-mesh sieve (*see* **Note 2**).
3. Weigh out a 10 g sample of milled plant material.
4. Put the powdered sample in a 250 mL flask containing 100 mL methanol (*see* **Note 3**).
5. Put a condenser (with running cold water) attached to the flask.
6. Reflux the sample using a water bath at 60 °C for 240 min.
7. Cool the solvent to room temperature.
8. Filter the solvent with a 40 μm filter to remove any solid matter.
9. Concentrate the extract solvent in vacuo to afford an extract (*see* **Note 4**).
10. Store the crude extract in the refrigerator at 4 °C until it is used to further analysis or separate.
11. Dissolve the extract in a certain amount of methanol and make the final concentration of the sample solution to be 10 mg/mL.
12. Filter the sample solution by a 0.20 μm Millipore filter and follow by HPLC analysis (Fig. 2).

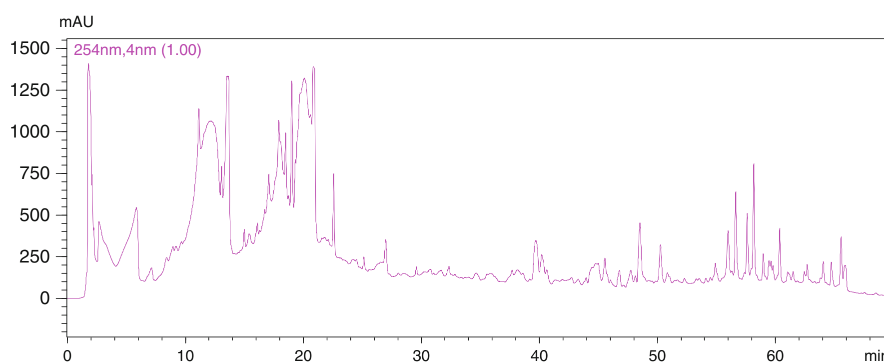


Fig. 2 HPLC chromatography of the extract (*Ardisia japonica* (Thunb.) Blume) from solvent extraction (Flow rate: 1.0 mL/min; 0 min: 0 % (v/v) B, 5 min: 0 % (v/v) B, 60 min: 100 % (v/v) B, 65 min: 100 % (v/v) B, 66 min: 0 % (v/v) B, 70 min: 0 % (v/v) B; Injection volume: 20 μL)

3.2 Supercritical Fluid Extraction

1. Dry the solid samples at 105 °C for 8 h or longer to constant mass.
2. Crush the dry sample to powder with mean diameters between 0.25 and 2.0 mm (*see* **Notes 5** and **6**).
3. Weigh 5 g of the raw materials.
4. Load the samples into a 50 mL extraction vessel with the aid of a nylon sleeve presenting approximately the same diameter of the vessel.
5. Pack glass wool at both ends of the vessel (*see* **Note 7**).
6. Design an orthogonal array to investigate the effects of the pressure, temperature, cosolvent level (methanol), and extraction time on the yield of raw material. The orthogonal matrix contains four factors, and each factor includes three levels. As a rule, the flow rate of SCFs is 25 kg/h (*see* **Notes 8–11**) (Table 1). Keep other independent variables (e.g., sample size and solvent flow rate) constant during the above experiments.
7. Investigate the influence of the chosen parameters upon the efficiency of the extraction.
8. Choose the optimized extraction variables to scale up using a 1 L vessel.
9. Weigh 150 g of ground materials.
10. Load the samples into a 1 L extraction vessel with the aid of a nylon sleeve presenting approximately the same diameter of the vessel.
11. Pack glass wool at both ends of the vessel.
12. Perform the scaling-up SFE under the optimized extraction condition.
13. Collect the extracts using a 100 cm^3 amber glass bottle placed in an ice bath during the extraction (*see* **Note 12**).
14. Remove the methanol from the extracts using a rotary evaporator in vacuo.

Table 1
Four considered variables for the SFE experiments

Variable	Low	Medium	High
Pressure	100 bar	200 bar	300 bar
Temperature	40 °C	50 °C	60 °C
Cosolvent level	0 %	5 %	10 %
Extraction time	120 min	150 min	180 min
	(30 min static + 90 min dynamic)	(30 min static + 120 min dynamic)	(30 min static + 150 min dynamic)

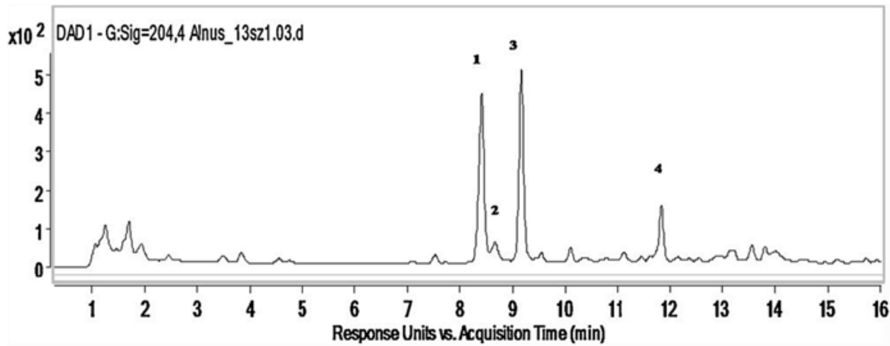


Fig. 3 LC-DAD chromatogram of *Alnus glutinosa* L. (Gaertn.) supercritical extract. Numbering of peaks are the following: 1 = betulinic acid, 2 = ursolic acid, 3 = betulin, 4 = β-amyrin. (Flow rate: 0.2 mL/min; 0 min: 65 % (v/v) B, 3 min: 80 % (v/v) B, 10 min: 100 % (v/v) B, 12 min: 100 % (v/v) B, 13 min: 65 % (v/v) B, 16 min: 65 % (v/v) B; Injection volume: 2 μL) (reproduced from ref. [25] with permission from Elsevier)

15. Store the extract in a refrigerator at 4 °C until HPLC analysis.
16. Dissolve the extract in methanol at a concentration of 10 mg/mL.
17. Filter the sample solution with a 0.20 μm Millipore filter and follow by HPLC analysis (Fig. 3) [25].

4 Notes

1. For plant materials (e.g., roots, barks), if there is no obvious boundary between the inner and outer after snapping them, it suggests that the plant materials are dried. For plant parts (e.g., flowers, leaves, and herbs), if it is easily broken up or powdered, it demonstrates that the parts are dried. For seeds and fruits of plant, if it is definitely dried, it is hard to bite or snap.
2. The aim of powdering the plant materials is more efficient for extreme extraction. The plant parts should not be excessively shattered to avoid the loss of samples in the process of grinding.
3. The choice of appropriate solvent used should have lower boiling point to avoid NPs degradation. In the process of extraction, extraction temperature should be equal to or lesser than the boiling point of chosen solvent. Therefore, heat reflux extraction cannot be used to extract thermolabile components.
4. The temperature of the rotary evaporator should be controlled under 60 °C especially for the extract containing volatile or thermolabile component.
5. Sample size plays a decisive role in the process of SFE. A smaller size can increase the efficiency of diffusion of solvent. However, channeling might appear inside the extraction bed if the sample size is too small.

6. The extracted sample must have fine solubility in SCFs. SCFs (CO_2) is more suitable for fat-soluble chemical constituents. The solubility of samples in SCFs can be modified by adding cosolvent (such as 1–10 % methanol), which can enhance the polarity of SCFs (CO_2) to extract moderately polar compounds.
7. The aim of packing the ends of extractor is to stop entrainment of the sample.
8. The efficacy of SFE is affected by a variety of variables, including extraction time, the pressure, temperature, sample size, flow rate of SCFs, and modifier.
9. If the flow rate of SFE is too high, the yield of extraction will be decreased due to the time reduction in the extractor.
10. The cosolvent, typically 1–10 % methanol, can enhance the polarity of SC-CO_2 to extract moderately polar compounds.
11. Static extraction means no liquid flow through the extractor, while dynamic extraction is to subject the sample by flowing SCFs at a set rate.
12. The bottle is placed in an ice bath to improve the collection efficiency.

Acknowledgement

This work was supported in part by the National Program on Key Basic Research Project (973 program, 2013CB734000), the National Natural Science Foundation of China (81102369, 30901849, 30911120483, 81102356, 30973665, 30911120484), the CAS Pillar Program (XDA04074000), the Ministry of Science and Technology of China (2013ZX10005004, 2011ZX11102-011-11, and 2007DFB31620), Genzyme, a Sanofi company, the Institute for Molecular Bioscience, the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 312184. L. Z. is an Awardee of the National Distinguished Young Scholar Program in China.

References

1. Drašar P, Moravcova J (2004) Recent advances in analysis of Chinese medical plants and traditional medicines. J Chromatogr B 812: 3–21
2. Ernst E, White A (2000) The BBC survey of complementary medicine use in the UK. Complement Ther Med 4:32–36
3. Chen FP et al (2007) Use frequency of traditional Chinese medicine in Taiwan. BMC Health Serv Res 7:26
4. Cassidy CM (1998) Chinese medicine users in the United States part I: utilization, satisfaction, medical plurality. J Altern Complement Med 4:17–27
5. MacLennan AH, Wilson DH, Taylor AW (2002) The escalating cost and prevalence of alternative medicine. Prev Med 35:166–173
6. Molassiotis A et al (2005) Use of complementary and alternative medicine in cancer patients: a European survey. Ann Oncol 16:655–663

7. Koehn FE, Carter GT (2005) The evolving role of natural products in drug discovery. Nat Rev Drug Discov 4:206–220
8. Balunas MJ, Kinghorn AD (2005) Drug discovery from medicinal plants. Life Sci 78: 431–441
9. Zhang L et al (2007) High-throughput synergy screening identifies microbial metabolites as combination agents for the treatment of fungal infections. Proc Natl Acad Sci U S A 104:4606–4611
10. Bradbury J (2005) From Chinese medicine to anticancer drugs. Drug Discov Today 10: 1131–1132
11. Paddon C et al (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496:528–532
12. DeJong JM et al (2006) Genetic engineering of taxol biosynthetic genes in Saccharomyces cerevisiae. Biotechnol Bioeng 93:212–224
13. Engels B, Dahm P, Jennewein S (2008) Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. Metab Eng 10: 201–206
14. Ajikumar PK et al (2010) Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. Science 330:70–74
15. Giannoccaro E, Wang YJ, Chen P (2006) Effects of solvent, temperature, time, solvent-to-sample ratio, sample size, and defatting on the extraction of soluble sugars in soybean. J Food Sci 71:59–64
16. Weidner S et al (2012) Extracts of phenolic compounds from seeds of three wild grapevines—comparison of their antioxidant activities and the content of phenolic compounds. Int J Mol Sci 13:3444–3457
17. Castro-Vargas HI et al (2010) Extraction of phenolic fraction from guava seeds (*Psidium guajava* L.) using supercritical carbon dioxide and co-solvents. J Supercrit Fluids 51: 319–324
18. Xiao W, Han L, Shi B (2008) Microwave-assisted extraction of flavonoids from *Radix Astragali*. Sep Purif Technol 62:614–618
19. Ncube N, Afolayan A, Okoh A (2008) Assessment techniques of antimicrobial properties of natural compounds of plant origin: current methods and future trends. Afr J Biotechnol 7:1797–1806
20. Zhang JY et al (2013) Caesanines A–D, new cassane diterpenes with unprecedented N bridge from *Caesalpinia sappan*. Org Lett 15:4726–4729
21. Smith RM, Burford MD (1992) Supercritical fluid extraction and gas chromatographic determination of the sesquiterpene lactone parthenolide in the medicinal herb feverfew (*Tanacetum parthenium*). J Chromatogr A 627:255–261
22. Hedrick JL, Mulcahey LJ, Taylor LT (1992) Supercritical fluid extraction. Microchim Acta 108:115–132
23. Zosel K (1964) Method for separation of mixtures. German Patent 1493:190
24. Azmir J et al (2013) Techniques for extraction of bioactive compounds from plant materials: a review. J Food Eng 117:426–436
25. Anikó FG et al (2012) Supercritical fluid extraction of *Alnus glutinosa* (L.) Gaertn. J Supercritical Fluids 61:55–61

Chapter 15

Bioassay-Guided Identification of Bioactive Molecules from Traditional Chinese Medicines

Jianying Han, Jingyu Zhang, Wenni He, Pei Huang, Ayokunmi Oyeleye, Xueting Liu, and Lixin Zhang

Abstract

Traditional Chinese medicines (TCMs) serve as a major source of a variety of drug lead compounds. In the process of natural products development, bioassay-guided isolation is a rapid and validated method for isolation of compounds with bioactivities. This chapter describes bioassay-guided separation and purification of compounds from the crude extracts of TCMs. Two approaches including size-exclusion chromatography (SEC) and high performance liquid chromatography (HPLC) are described in detail.

Key words Traditional Chinese medicines, Biological activity assays, High-speed counter current chromatography, Structure characterization

1 Introduction

Traditional Chinese medicines (TCMs) have been used for thousands years and have provided an important source of drug discovery leads. In China, there are more than 31,000 species of plants; however, only a small number of them have been chemically investigated. The importance of TCMs' biodiversity is increasingly prominent in the process of drug discovery. Bioassay-guided separation and characterization that can directly link the activity of the crude extract to its components are a crucial process to obtain the active compounds after the extraction from TCMs [1–4]. Analytical HPLC fractionation, combined with a 96-well fluorescent bioassay screen, including different bioassay models, has been developed and used for the separation and screening for active fractions or compounds, due to the high efficiency and sensitivity of 96-well screening techniques for identifying the active components [5–7].

As soon as the crude extract is obtained via the appropriate extraction approaches, a bioassay-guided fractionation technique can be applied for identifying the principal active components [1, 2].

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_15, © Springer Science+Business Media New York 2015

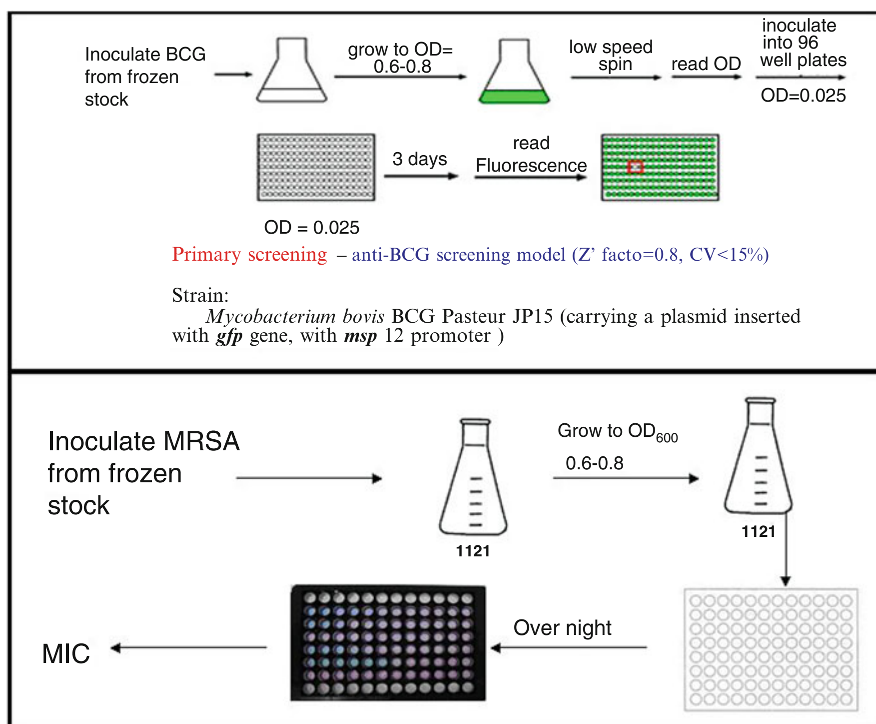


Fig. 1 Flowchart of the biological activity assays

It is the most common method for natural products discovery and has been successfully applied for obtaining active compounds from TCMs [1, 8–11]. Wang et al. [12] developed an efficient assay to identify inhibitors in an aerobic, logarithmic growth screen of Bacillus Calmette-Guérin (BCG) (Fig. 1). The BCG used in the following biological activity assays was a *Mycobacterium bovis* BCG 1173P2 strain transformed with a green fluorescent protein (GFP) constitutive expression plasmid pUV3583c, with direct fluorescence readout as a measure of bacterial growth [13]. However, isolation and purification methods, such as column chromatography (CC), size-exclusion chromatography (SEC), medium pressure liquid chromatography (MPLC), and high performance liquid chromatography (HPLC) could be applied for further identification of active compounds [14].

SEC, known as gel-filtration chromatography, is a chromatographic method in which molecules are separated according to their sizes and molecular weights [15] and can be used as a primary fractionation method for large amounts of crude extract. It involves the use of porous gel molecules of agarose, cross-linked dextran, or polymers of acrylamide that allow the separation of compounds based on their molecular sizes. There is no sample loss because generally solutes used do not interact with the stationary phase, an

advantage which is of great importance for the small quantity of compounds [16]. In the experimental process, the large molecules incapable of penetrating the porous column particles are eluted first. However, small molecules diffusing into the pores are eluted at a later time [17]. Sephadex LH-20 is the widely used medium that is designed for molecular sizing of natural products such as steroids, terpenoids, lipids, and low molecular weight peptides. Due to the unique physicochemical properties of this medium, it can be used either during initial purification prior to polishing by high performance ion exchange or reverse phase chromatography or as the final polishing step, e.g., during the separation of diastereomers.

HPLC is a technique in analytic chemistry used to separate the components in a mixture, to identify each component, and to quantify each component. HPLC has been used for medical, legal, research, and manufacturing purposes. HPLC has been widely used for both separation and quantification of natural products (NPs) from the active fractions of TCMs [18–22]. The efficiency of HPLC analysis and preparation might be affected by many factors, including different support materials in the column, mobile phase, and others [23]. This chapter describes purification of bioactive molecules from TCMs based on bioassay-guided isolation integrated with SEC and HPLC preparation approaches.

2 Materials

2.1 Biological Activity Assays

1. *Mycobacterium bovis* BCG 1173P2 strain.
2. *Staphylococcus aureus* (SA, ATCC 6538).
3. *Bacillus subtilis* (BS, ATCC 6633).
4. Methicillin-resistant *S. aureus* (MRSA).
5. *Pseudomonas aeruginosa* (PAO1).
6. EnVision 2103 Multi-label Plate Reader (Perkin-Elmer Life Sciences).
7. Oleic albumin dextrose catalase (OADC): Add 0.85 g of NaCl, 5 g of Bovine Serum Albumin Fraction V, 2 g of dextrose, 0.05 g of oleate acid, and 0.004 g of catalase to 100 mL of water and filter through a 0.22 μm membrane. Store at 4 °C.
8. 7H9 Medium broth: Add 1.88 g of 7H9 broth, 0.8 mL of glycerol, and 0.2 mL of Tween-80 to 360 mL of water and mix. Sterilize at 121 °C for 10 min. Add 40 mL of filtered OADC to sterilized 7H9 broth and store at 4 °C.
9. Mueller-Hinton Broth (MHB): Add 2.4 g of MHB powder to 100 mL of water, and sterilize at 121 °C for 15 min.
10. Luria–Bertani (LB) agar plate: Add 0.5 g of yeast extract, 1 g of NaCl, 1 g of tryptone, and 2 g of agar to 100 mL of water.

Sterilize at 121 °C for 20 min. Add about 20 mL sterilized LB broth to a 90 mm dimensional sterilized flat.

11. 96-Well clear flat-bottom microtiter plate.
12. Constant-temperature incubator.
13. Inoculating loop.
14. Dimethyl sulfoxide (DMSO).
15. Isoniazid, 0.032 mg/mL: Dilute 0.032 mg of isoniazid with 1 mL of DMSO. Store at −20 °C.
16. Vancomycin, 0.32 mg/mL: Dilute 0.32 mg of vancomycin with 1 mL of DMSO. Store at −20 °C.
17. Ciprofloxacin, 0.32 mg/mL: Dilute 0.32 mg of ciprofloxacin with 1 mL of DMSO. Store at −20 °C.

2.2 Size-Exclusion Chromatography (SEC)

1. Sephadex LH-20 powder.
2. Chromatographic column.
3. Methanol.
4. Dichloromethane.
5. Isopropanol.
6. Fraction collector.

2.3 High Performance Liquid Chromatography (HPLC)

1. Acetonitrile.
2. Agilent 1200 Series HPLC or equivalent instrument.
3. Diode array and multiple wavelength detectors.
4. Agilent ZORBAX SB-C18 column, 5 μM, 4.6 × 150 mm and Agilent ZORBAX SB-C18 column, 5 μm, 9.4 × 250 mm.

3 Methods

In this section, two biological activity assays, anti-BCG assay and general antimicrobial assay, together with two of the most applicable methods of fractionation, SEC and HPLC, have been described in detail.

3.1 Biological Activity Assays

3.1.1 Anti-BCG Assay

1. Grow BCG at 37 °C to mid log phase in 7H9 medium broth. Adjust the suspension to an OD_{600} of 0.025 with culture medium as a bacterial suspension (Fig. 1) (*see* **Note 1**).
2. Dissolve an appropriate amount of the crude TCM extract with DMSO to a concentration of 4 mg/mL.
3. Perform a twofold serial dilution of the TCM extract DMSO solution in a 96-well microplate or tubes using DMSO to achieve eight concentrations of 4,000, 2,000, 1,000, 500, 250, 125, 62.5, and 31.25 μg/mL (*see* **Note 2**).

4. Perform a twofold serial dilution of the positive control isoniazid DMSO solution in a 96-well microplate or tubes using DMSO to achieve eight concentrations of 32, 16, 8, 4, 2, 1, 0.5, and 0.25 μg/mL (*see* **Note 3**).
5. Perform a twofold serial dilution of the negative control DMSO solvent in a 96-well microplate or tubes by adding 7H9 medium broth to achieve eight concentrations accordingly.
6. Add 2 μL of the crude TCM extract dilution series and controls solutions to the appropriate wells of a 96-well microplate.
7. Add 78-μL aliquots of the bacterial suspension to each well (*see* **Note 4**).
8. Incubate the plate at 37 °C for 3 days (*see* **Note 5**).
9. Measure GFP fluorescence with a Multi-label Plate Reader using the appropriate read mode, with excitation at 485 nm and emission at 535 nm (*see* **Note 6**).
10. Define minimum inhibitory concentration (MIC) as the minimum concentration of crude extract that inhibits more than 90 % of bacterial growth reflected by fluorescence value.

3.1.2 General Antimicrobial Assays

Antimicrobial assays should be performed according to the Antimicrobial Susceptibility Testing Standards outlined by the Clinical and Laboratory Standards Institute (CLSI) using the bacteria *S. aureus* (ATCC 6538), *B. subtilis* (ATCC 6633), and methicillin-resistant *S. aureus* (MRSA) (Fig. 1) (*see* **Note 7**).

1. Streak a loopful of each organism from a glycerol stock onto an LB-agar plate respectively and incubate overnight at 37 °C (*see* **Note 1**).
2. Pick a single bacterial colony and suspend the colony in MHB medium to approximately 1×10^4 CFU/mL as a bacterial suspension (*see* **Note 8**).
3. Dilute the crude TCM extract with DMSO to a concentration of 4 mg/mL.
4. Perform a twofold serial dilution of the TCM extract solution with DMSO in a 96-well microplate or tubes to achieve eight concentrations of 4,000, 2,000, 1,000, 500, 250, 125, 62.5, and 31.25 μg/mL (*see* **Note 2**).
5. Perform a twofold serial dilution of the positive control vancomycin and ciprofloxacin DMSO solution in a 96-well microplate or tubes using DMSO to achieve eight concentrations of 320, 160, 80, 40, 20, 10, 5, and 2.5 μg/mL.
6. Perform a twofold serial dilution of the negative control DMSO solvent in a 96-well microplate or tubes by adding MHB medium broth to achieve eight concentrations accordingly.

7. Add 2 μL of the crude TCM extract dilution series and controls solutions to the appropriate wells of a 96-well microplate.
8. Add 78 μL of bacterial suspension to each well to give final crude extract concentrations of 100–0.78 μg/mL in 2.5 % DMSO.
9. Incubate the plate at 37 °C aerobically for 16 h.
10. Measure the optical density (OD) of each well at 600 nm with a Multi-label Plate Reader.
11. Define MIC values as the minimum concentration of crude extract that inhibits visible bacterial growth (*see* **Note 9**).
12. Repeat the experiment to verify the reliability of the results.

3.2 Size-Exclusion Chromatography (SEC)

1. Weigh 100 g of Sephadex LH-20 powder into 500 mL of an appropriate solvent according to the physicochemical properties and solubility of the sample and swell the medium for at least 3 h. Prepare a medium slurry in a ratio of 75 % settled medium to 25 % solvent (*see* **Note 10**).
2. Pack the Sephadex LH-20 slurry into a glass column in one continuous motion. Open the bottom outlet of column until the media bed reaches a constant height.
3. Before applying sample, equilibrate the column with eluent to be used in the separation until the baseline becomes stable (at least two bed volumes) (*see* **Note 11**).
4. Dissolve the crude extract in an appropriate solvent, of which methanol and dichloromethane are most widely used.
5. Filter the sample solution through a 0.45-μm solvent-resistant filter or centrifuge before use (*see* **Note 12**).
6. Load the sample solution onto the Sephadex LH-20 gel bead surface carefully.
7. Elute the column using the designed solvent system (*see* **Note 13**).
8. Continuously monitor the UV absorption of extract eluted from the outlet of the column and collect the solvent into test tubes with a fraction collector (Fig. 2) [24].
9. Dry the organic solvent under vacuum to obtain the fraction extracts, which are ready for further biological activity testing (*see* **Note 14**).

3.3 High Performance Liquid Chromatography (HPLC)

In this section, the samples can be either those obtained out of the SEC fractions or crude extracts. The analytical columns could be used for analyzing the samples in order to get an overview of the components their properties. The semi-preparative column is needed for further purification of pure compounds.

1. For sample preparation, dissolve the crude TCM sample in an appropriate solvent based on its solubility to a concentration of 1 mg/mL and 10 mg/mL for pure compounds and fractions,

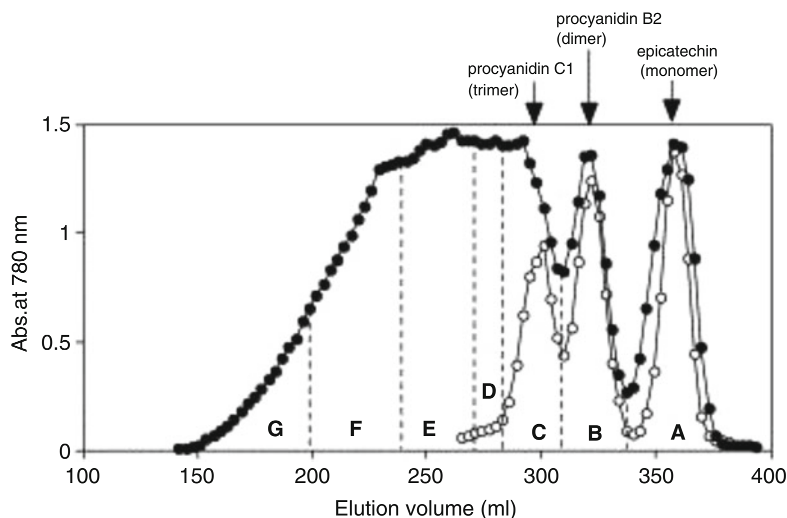


Fig. 2 SEC elution profile of apple procyanidins using a Toyopearl HW-40 F column. Experimental conditions: the column size was 950 × 25 mm inner diameter; mobile phase solvent was acetone–8 M urea (pH 2) (6:4); flow rate was 1.0 mL/min; the eluent was fractionated into 3 mL fractions each; the total phenolics content (i.e., the value corresponding to the absorbance at 760 nm) in each fraction was estimated by modified Folin–Ciocalteu assay; *filled circle* = chromatogram of apple procyanidins (10 mg/0.5 mL of mobile phase); *open circle* = chromatogram of a mixture of standard oligomers from monomer to trimer (each 2 mg/0.5 mL of mobile phase) (reproduced from ref. [24] with permission from Elsevier)

respectively, of which methanol and acetonitrile are the two most commonly used solvents.

2. Filter the sample solution through a 0.45 μm pore size membrane (*see* **Note 15**).
3. Add the prepared sample (at least 500 μL) to a HPLC vial.
4. Equilibrate the UV detector for 30 min.
5. Equilibrate the columns by running 20 times the column volume full of solvent. Column volumes are approximately 11.6 and 182 mL for the small and large columns (*see* **Note 16**).
6. Choose an appropriate analytical column according to the properties of the sample and set up the solvent system accordingly. A common solvent gradient program such as 5–100 % acetonitrile in water for 30 min can be applied for HPLC analysis of crude extracts on a reverse phase column (*see* **Note 17**).
7. Optimize the method for further application on the semi-preparative column to separate the peaks.
8. Collect the peaks based on the retention time.
9. Dry the organic solvent under vacuum to obtain pure compounds.

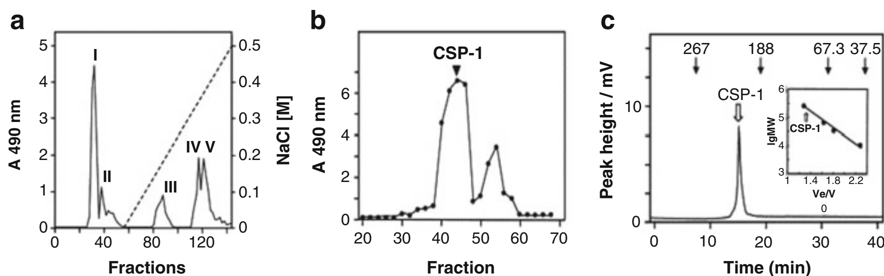


Fig. 3 Purification of the isolated polysaccharide CSP-1 from cultured Cordyceps. (**a**): Partial purification of polysaccharides from cultured Cordyceps mycelia by DEAE-cellulose chromatography. Extract was loaded onto the DEAE column (3.5 × 30 cm), and eluted with 0 to 0.5 M NaCl, as indicated by *dotted line*, in 10 mM Tris–HCl pH 7.4 having a flow rate of 30 mL/h. Five milliliter fractions were collected. (**b**): The polysaccharide-enriched fractions were applied onto a Sephacryl S-300 column equilibrated with 0.2 M NaCl in 10 mM Tris–HCl buffer pH 8.0. The first-peak fractions (CSP-1) were collected. (**c**): HPLC profile of CSP-1 by using TSKgel G3000 SWXL (7.8 mm ID × 30 cm) column. Molecular markers are indicated by *arrows* in kDa. The polysaccharide profile was detected by refraction index detector. The insert shows the molecular weight determination of CSP-1 (reproduced from ref. [25] with permission from Elsevier)

10. Analyze the purity of isolated compounds by using an analytical column (Fig. 3) [25] (*see* **Note 18**).
11. The biological activity of isolated pure compounds will be determined by the biological assays described in Subheading 3.1 for their further chemical and biological profiling.

4 Notes

1. In view of the potential risk, operations involving infectious strains must be careful and the experiments must be operated in class II biohazard safety cabinet.
2. The recommended starting concentration of test sample is 4 mg/mL for serial twofold dilution, which could be changed according to the specific requirement as well.
3. Isoniazid is one of the first-line drugs for treatment of tuberculosis. It is recommended to choose isoniazid as a positive control for the anti-BCG biological activity assay.
4. A 96-well plate can be used for testing up to 10 crude extracts together with positive and negative controls under normal circumstances.
5. Bacterial growth inhibition can be observed after 3 days of cultivation. If not, the incubation time needs to be extended to 4 days.
6. The top and bottom read mode could be chosen according to the corresponding function of the Multi-label Plate Reader.

Before GFP fluorescence measurement, the instrument should be on for a half an hour in advance.

7. Other strains of bacteria can also be used with this assay protocol following appropriate optimization of the growth conditions.
8. Thoroughly mix by vortexing to make sure that the bacterial colonies are evenly distributed.
9. Since the difference of OD_{600} between the concentrations above and below MIC is very obvious, the MIC values that are determined based on the OD_{600} is reliable.
10. Sephadex LH-20 is supplied as a dry powder and must be swollen before use. During swelling excessive stirring should be avoided as it may break the beads. Do not use magnetic stirrers.
11. Dichloromethane can be chosen as the eluting solvent. Make sure that there are no air bubbles during column packing.
12. Make sure the concentration is appropriate in order to avoid a viscous or over-diluted sample. The sample volume should be in the range of 1–2 % of the total bed volume.
13. Keep adequate eluting solvent in the column to avoid drying column. The recommended flow rate range is dependent on the application. Flow rates of 1–10 cm/h are recommended. Generally, the lower the flow rate, the better the resolution.
14. Do this step according to the biological activity assays described above.
15. The sample should be centrifuged at $6{,}000 \times g$ for at least 5 min or should be filtered with a 0.22 μm filter.
16. Never run the column dry. Make sure there is enough solvent in the reservoir.
17. Before injecting the sample, wash the column for at least 20 min and equilibrate with the initial solvent for at least 10 min.
18. If there are impurities in the analyzed compounds, perform Subheading 3.2, **steps 5** and **6** again.

Acknowledgement

This work was supported in part by the National Program on Key Basic Research Project (973 program, 2013CB734000), the National Natural Science Foundation of China (81102369, 30901849, 30911120483, 81102356, 30973665, 30911120484), the CAS Pillar Program (XDA04074000), the Ministry of Science and Technology of China (2013ZX10005004, 2011ZX11102-011-11, and 2007DFB31620), Genzyme, a Sanofi company, the Institute for

Molecular Bioscience, the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 312184. L.Z. is an Awardee of the National Distinguished Young Scholar Program in China.

References

1. Si W et al (2006) Bioassay-guided purification and identification of antimicrobial components in Chinese green tea extract. J Chromatogr A 1125:204–210
2. Ejele AE et al. (2014) Bioassay-guided isolation, purification and charaterization of antimicrobial compound from acidic metabolite of piper umbellatum seed extract. Int J Chem 6:61–70
3. Koehn FE, Carter GT (2005) The evolving role of natural products in drug discovery. Nat Rev Drug Discov 4:206–220
4. Lam KS (2007) New aspects of natural products in drug discovery. Trends Microbiol 15: 279–289
5. Fitch RW et al (2003) Bioassay-guided isolation of epiquinamide, a novel quinolizidine alkaloid and nicotinic agonist from an Ecuadoran poison frog, epipedobates tricolor. J Nat Prod 66:1345–1350
6. Wang Q et al (2013) Abyssomicins from the South China Sea deep sea sediment *Verrucosispora sp.* : natural thioether Michael addition adducts as antitubercular prodrugs. Angew Chem-Int Ed 52:1231–1234
7. Liu M et al (2014) Endophytic *Streptomyces sp.* Y3111 from traditional Chinese medicine produced antitubercular pluramycins. Appl Microbiol Biotechnol 98(3):1077–85
8. Han QB et al (2009) Bioassay guided discovery of apoptosis inducers from gamboge by high-speed counter-current chromatography and high-pressure liquid chromatography/electrospray ionization quadrupole time-of-flight mass spectrometry. J Chromatogr B 877: 401–407
9. Ho CC, Kumaran A, Hwang LS (2009) Bioassay guided isolation and identification of anti-Alzheimer active compounds from the root of *Angelica sinensis*. Food Chem 114:246–252
10. Lin LZ et al (1998) Liquid chromatographic–electrospray mass spectrometric study of the phthalides of *Angelica sinensis* and chemical changes of *Z*-ligustilide. J Chromatogr A 810: 71–79
11. Koleva II, Niederlander HA, van Beek TA (2000) An on-line HPLC method for detection of radical scavenging compounds in complex mixtures. Anal Chem 72:2323–2328
12. Wang JF et al (2010) Antituberculosis agents and an inhibitor of the para-aminobenzoic acid biosynthetic pathway from hydnocarpus anthelminthica seeds. Chem Biodivers 7:2046–2053
13. Stanley SA et al (2012) Identification of novel inhibitors of M. tuberculosis growth using whole cell based high-throughput screening. ACS Chem Biol 7:1377–1384
14. Harborne JB (1984) Phytochemical methods. 278: Springer
15. Paul-Dauphin S et al (2007) Probing size exclusion mechanisms of complex hydrocarbon mixtures: the effect of altering eluent compositions. Energy Fuels 21:3484–3489
16. Skoog DA (2006) Principles of instrumental analysis, 6th edn. Thompson Brooks, Belmont, CA, Chapter 28
17. Mori S, Barth HG (1999) Size exclusion chromatography. Springer
18. Stalikas CD (2007) Extraction, separation, and detection methods for phenolic acids and flavonoids. J Sep Sci 30:3268–3295
19. Naczk M, Shahidi F (2004) Extraction and analysis of phenolics in food. J Chromatogr A 1054:95–111
20. Robbins RJ (2003) Phenolic acids in foods: an overview of analytical methodology. J Agric Food Chem 51:2866–2887
21. Kalili KM, de Villiers A (2011) Recent developments in the HPLC separation of phenolic compounds. J Sep Sci 34:854–876
22. Flamini R (2003) Mass spectrometry in grape and wine chemistry. Part I: polyphenols. Mass Spectrom Rev 22:218–250
23. Merken HM, Beecher GR (2000) Measurement of food flavonoids by high-performance liquid chromatography: a review. J Agric Food Chem 48:577–599
24. Yanagida A, Shoji T, Shibusawa Y (2003) Separation of proanthocyanidins by degree of polymerization by means of size-exclusion chromatography and related techniques. J Biochem Biophys Methods 56:311–322
25. Li SP et al (2003) A polysaccharide isolated from *Cordyceps sinensis*, a traditional Chinese medicine, protects PC12 cells against hydrogen peroxide-induced injury. Life Sci 73:2503–2513

Chapter 16

NMR Screening in Fragment-Based Drug Design: A Practical Guide

Hai-Young Kim and Daniel F. Wyss

Abstract

Fragment-based drug design (FBDD) comprises both fragment-based screening (FBS) to find hits and elaboration of these hits to lead compounds. Typical fragment hits have lower molecular weight (<300–350 Da) and lower initial potency but higher ligand efficiency when compared to those from high-throughput screening. NMR spectroscopy has been widely used for FBDD since it identifies and localizes the binding site of weakly interacting hits on the target protein. Here we describe ligand-based NMR methods for hit identification from fragment libraries and for functional cross-validation of primary hits.

Key words Fragment-based drug design, Ligand-based NMR, Hit identification, STD, STDD, NMR-based functional assay

1 Introduction

Fragment-based drug design (FBDD) has emerged as a field in which smaller numbers of compounds with lower molecular weight (MW) are screened as compared to traditional high-throughput screening (HTS) campaigns. Advantages of fragment-based screening (FBS) over HTS are (1) more efficient sampling due to the smaller chemical space of fragment-sized compounds and (2) a higher probability of fragments possessing good complementarity with the target resulting in good ligand efficiency [1]. Due to their low MW, fragment-based hits are typically weak inhibitors and/or binders with half maximal inhibitory concentration (IC_{50}) and/or the equilibrium dissociation constant (K_d) in micromolar to millimolar range. As a consequence, they need to be screened at high concentrations using sensitive techniques that can reliably detect the weak signals generated by subtle modulation of the drug target and/or interacting compounds, e.g., nuclear magnetic resonance (NMR), surface plasmon resonance (SPR), high-concentration functional screening (HCS), or X-ray crystallography [2]. And more effort has to be spent on optimization to obtain lead

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_16, © Springer Science+Business Media New York 2015

compounds with an acceptable affinity. X-ray crystal structures may play a crucial role in accomplishing this goal efficiently [3].

Typical fragment libraries are designed following the “rule of three”—molecular weight <300 Da, the calculated octanol-water partition coefficient (cLogP) < 3, and number of hydrogen bond donors and acceptors < 3 [4]. These criteria may be considered as guidelines rather than strict rules. For example one might vary the limits in the above criteria to design a more project-specific fragment library. The size of fragment libraries may vary from several hundred to several thousand of compounds; however, it might be desirable to include a sufficient number of close analogues with variations in substituents in order to avoid missing hits with a potentially good core structure [5]. Determination of the so-called SPAM (Solubility, Purity and Aggregation of the Molecule) is another important step in the design of a high-quality fragment library [3, 6, 7]. However, care should be given for each screening campaign to confirm the true positive hits through orthogonal validation with multiple biophysical techniques.

NMR-based screening of protein targets has become a well-established part of the drug discovery process especially with respect to fragments [2, 3, 8–12]. A variety of NMR approaches exist depending on whether signals from the drug target or the ligand are detected to characterize the intermolecular interaction which can be broadly categorized into target based and ligand based. Each of these methods has advantages and limitations, and can provide information about the ligand-target interaction at various levels of detail, including determination of ligand affinities and potencies, their binding site, and binding mode on the drug target. Typical criteria for selection of specific NMR experiments are the target size and type, and isotope-labeling protein resources. Therefore, different NMR screening and follow-up strategies can be selected for different FBDD campaigns.

Target-based detection methods, mainly using heteronuclear single-quantum correlation (HSQC) spectra of ^{13}C-^{1}H or ^{15}N-^{1}H pairs, have proven to be very robust and information rich. It can reveal structural information about the ligand-binding site and ligand-binding mode with the drug target. HSQC experiments can detect site-specific ligand binding over a virtually unlimited affinity range, and can be used to derive ligand affinities for weak fragment hits that are in fast exchange on the NMR timescale ($K_d > \sim 10\ \mu M$) or for submicromolar affinity hits when combined in a competition format [8, 9, 11]. But, target size limitation is the major impediment to NMR-based screening using target-detected techniques; in addition, large amounts (>50 mg) of isotope-labeling protein is required, necessitating high-expression yields (>1 mg/L), and cost-effective isotope labeling (i.e., *Escherichia coli*). Also knowledge of the 3D structure of the drug target and NMR assignments of the active site residues are required to localize active site binders.

Therefore, target-detected NMR approaches are typically limited to a subset of drug targets (MW < ~60 kDa) that give quality NMR spectra and do not aggregate at relatively high concentrations (~25–80 μM) in an aqueous NMR buffer.

Ligand-based NMR methods are often applied much more broadly than target-detected fragment screens. Typically, they require about 1–10 % the amount of target protein, do not require isotope labeling, and have no upper MW size limitation (in fact, they work better on large proteins) [13]. Although some details about the ligand-binding epitope can be obtained, ligand-based NMR methods do not reveal the binding site on the target protein. Ligand-based screens rely on monitoring changes in the ligand upon its binding to the protein. This includes techniques that are based on magnetization transfer like saturation transfer difference (STD) [14], saturation transfer double difference (STDD) [15], water-LOGSY [16], diffusion-edited spectral change [9], heteronuclear (^{19}F, ^{31}P) detection [17, 18], and paramagnetic effect from the target (SLAPSTIC) [10], or the more specialized target-immobilized NMR screening (TINS) [19]. One of the most useful of these NMR methods is saturation transfer difference (STD) NMR spectroscopy, and its variant, competition-STD (c-STD) NMR spectroscopy [20]. If spins anywhere in the protein are selectively saturated, the saturation will quickly spread throughout the protein by spin diffusion, and will be transferred to a ligand if it has a long-enough residence time in the binding site. If the ligand has a fast-enough off-rate, the bound-state saturation will be observed on the free state of the ligand. In practice, the STD experiment works well for K_d range of 0.1 μM to 10 mM, with protein concentrations of 0.2–5.0 μM and the ligand present in 50- to 500-fold molar excess. However, as with other ligand-based NMR methods, it suffers from inherent limitations. Nonspecific binding of molecules to a protein target can confound identification of true binders. If there is a known ligand with known binding site (competitor), competition-STD (c-STD) may be used to determine site-specific binding and localize the binding site of a screened compound. In this way, c-STD is a two-step experiment. First, the STD spectrum of the reference compound is obtained. Then the competitor is added, and the STD spectrum of the ternary mixture (reference compound, competitor, and protein target) is obtained (Fig. 1). If both compounds are competing for the same binding site, the STD signal of the reference compound will decrease. The magnitude of the decrease can be used to estimate the affinity of the competitor [11] when the affinity of the reference (K_d) is known and the two compounds are strictly competitive with each other for the same binding site. Since c-STD can help rule out nonspecific binding, it is a highly valuable addition if well-characterized competitor compounds are known for the target.

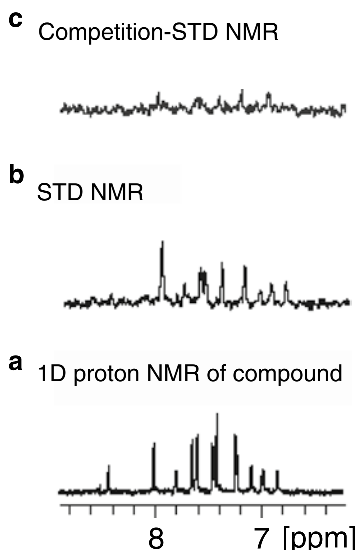


Fig. 1 Ligand-based NMR methods for fragment library screening. (**a**) 1D ¹H NMR control spectrum of reference compound. (**b**) STD NMR spectrum of reference compound in the presence of the target protein. Only resonances of atoms which contact the protein are present in the STD spectrum; (**c**) c-STD of reference compound and known competitor in the presence of the target protein. The STD signal of reference compound decreases because the known competitor with higher affinity displaces it from the binding site

Recently, several reports describing the use of NMR-based functional assays have appeared in the literature [17, 21]. An enzymatic reaction can be followed by monitoring changes in the NMR signals of substrate(s) and product(s) in the presence of submicromolar enzyme concentrations. The extent of completion of the reaction can then be compared between different inhibitors when present at a fixed concentration. In our experience (unpublished), a simple 1D proton NMR-based functional assay can be used to derive the percentage of inhibition by measuring changes of substrate (or product) concentrations against an internal standard (i.e., TSP) of known concentration. A 1D proton NMR-based functional assay might also reveal valuable details about the mode of action of modulators, since all reaction components can directly be monitored in real time during the enzymatic reaction. We can directly quantitate the depletion of native substrates, the formation of products, and the concentration of a ligand; even buffer components can be detected, thus minimizing errors in sample handling. However, finding an appropriate substrate for the assay can be challenging. Therefore, key to the success of a robust NMR-based functional assay is the selection of a suitable substrate. The chemical shifts of NMR signals between substrate(s)

and product(s) need to be distinguishable and observable in the NMR spectrum; and the concentration may be too low for tightly binding substrates to be detectable by NMR [21]. On the other hand, NMR-based functional assays are particularly attractive for substrates with higher K_m values and to study fragment hits that often are only weak inhibitors with IC_{50}s in the micromolar to millimolar range.

In this chapter we present the detailed setup of ligand-based NMR experiments and an NMR-based functional assay using a 30 kDa enzyme protein that makes bacteria resistant to a broad range of beta-lactam antibiotics. STD, c-STD, and STDD were used for fragment screening and an NMR-based functional assay that monitors the hydrolysis of ampicillin was used for functional validation of fragment hits with weak affinities/activities.

2 Materials

1. NMR spectrometer: Bruker 500 or 600 MHz Avance III spectrometer equipped with a 5 mm TCI CryoProbe with an ATM accessory (automatic tuning and matching) and a SampleJet automated sample changer (Bruker) (*see* **Note 1**).
2. Liquid handler: Gilson 215 liquid handler.
3. 96-Well plate, polypropylene 1.2 mL.
4. NMR tubes: 5 mm OD NMR tubes (Wilmad-LabGlass) in 96-tube rack.
5. Recombinant protein: High purity (>95 %) of target protein (*see* **Note 2**).
6. NMR buffer: Adjust the buffer conditions to optimize for target protein stability. Use of deuterated buffer components such as HEPES-d_{18}, Tris-d_{11}, and Glycerol-d_8 in D_2O solvent (Cambridge Isotope Laboratories, Inc) is beneficial (*see* **Note 3**). For the sample described here, use 25 mM Tris(d_{11})-DCl (pD 7.4), 150 mM NaCl, 100 μM $ZnSO_4$, and 2 mM DTT-d_{10} in D_2O.
7. Functional assay reaction buffer: 100 mM phosphate (pD 7.0), 10 μM $ZnSO_4$ in D_2O.
8. 1 mM stock of 2,2,3,3-tetradeutero-3-trimethylsilylpropionic acid (TSP) in D_2O. Store the stock at room temperature.
9. Fragment library: Merck Fragment library consisted of largely “rule of three” compliant compounds which are stored as 100 mM concentration stocks in DMSO-d_6 at ambient temperature in 96-well plates placed in a desiccator; fragments should have good solubility in aqueous buffer, at least 200 μM (*see* **Note 4**).

10. Competitor compounds: A lead compound for competition-STD NMR or a known inhibitor/activator to validate site-specific binding of a screening hit to the target or an enzyme substrate/ATP for kinase.
11. Substrate: Ampicillin sodium salt crystalline powder. The substrate was dissolved in D_2O with the final concentration 25 mM. Store the stock at −20 °C.
12. NMR software: Topspin 3.2 and ICON NMR (Bruker AG, Karlsruhe, Germany).

3 Methods

3.1 Hit Identification, Confirmation, and Qualification

3.1.1 Quality Control (QC) Samples for Fragment Compounds

1. Prepare a sample for each fragment at 200 μM in the NMR buffer which contains 25 μM TSP as the internal integration and spectral reference standard.
2. Load the NMR samples into a SampleJet Automated Sample Changer.
3. After setting up the sample temperature at a desired value (i.e., 298 K), follow the routine setup procedure for every sample changed: lock, tune and match, shim, and perform ^1H 90° pulse calibration.
4. Record a regular 1D proton NMR spectrum of each fragment with the optimized parameters, typically using relaxation delays = 8 s, transmitter frequency offset = solvent ^1H frequency, time domain = 8 K, and spectral width = 15 ppm.
5. Filter and flag insoluble (low solubility) fragments on the plate maps (i.e., < ~10–25 μM) by visual inspection of the 1D proton NMR spectra (*see* **Note 5**).

3.1.2 STD and Competition-STD NMR

1. Prepare samples of target protein in the NMR buffer at 3 μM concentration.
2. Combine groups of 5–10 compounds from fragment library into clusters: Add and mix 1 μL of each fragment from 100 mM DMSO-d_6 stock into a 5 mm NMR tube containing 500 μL NMR buffer with the target protein. Final concentration of each fragment is 200 μM.
3. Incubate samples for 30 min at room temperature.
4. Obtain STD NMR spectra of clusters of 5–10 fragment compounds: Apply selective saturation of the protein with switching the on- and off-resonance saturation frequency after each scan. For the current protein target, we used a train of Gaussian shape pulses with 50 ms pulse length (corresponding to an excitation width of 100 Hz) separated by a delay of 1 ms, with the total length of the selective saturation set to 3 s, and the

on- and off-resonance saturation frequencies set to 50 Hz and 20,000 Hz, respectively (*see* **Note 6**).

5. If there is any binding, a positive STD NMR signal will come up. The real binder is identified by comparison of the STD NMR signal and the 1D proton NMR reference data (Fig. 1).
6. Categorize cluster hits into three groups (1 = "no hit," 2 = "medium hit," 3 = "strong hit") based on increased STD NMR signal intensities (*see* **Note 7**).
7. Deconvolute the high-score clusters (mainly "3") by running STD for each compound individually as singletons.
8. After the hit identification, add 1 μL of known competitor from the 100 mM DMSO-d_6 compound stock and acquire the STD experiment again, to validate the site-specific binding of the fragment hit. If there is good competition, the STD signal for the fragment should decrease with the addition of a competitor (Fig. 1c) (*see* **Note 8**).

3.1.3 STDD and Competition-STD NMR

In case residual STD signals from target protein or additives in the sample hampers hit identification, STDD NMR can be applied to reduce such artifacts [15].

1. Prepare samples as described in **steps 1–3** of Subheading 3.1.2 except that each cluster is added twice into 500 μL NMR buffer solution either in the presence or absence of target protein.
2. Obtain STD NMR spectra of clusters of 5–10 fragment compounds in the presence and absence of target protein.
3. Subtract the reference cluster spectrum without protein from that with protein to obtain STDD NMR spectrum (Fig. 2).
4. Compare the STDD NMR cluster spectrum with the 1D proton NMR reference spectra of each individual compound in that cluster to identify potential hits (*see* **Note 9**).
5. Confirm potential hits by repeating **steps 1–4** of Subheading 3.1.3 with each potential hit being tested as a singleton.
6. Add a known competitor (active site) to those samples that contain a confirmed STDD NMR hit.
7. Qualify those confirmed hits as "active-site"-specific hits whose STD NMR signals are reduced upon competitor addition.

3.2 1D Proton NMR-Based Functional Assay

The conversion of the antibiotic substrate into its hydrolyzed product can easily be followed by simple 1D proton NMR spectra (Fig. 3). Changes in chemical shifts of ampicillin occur as the amide bond in the β-lactam ring is hydrolyzed by β-lactamase. Resonances from the substrate at ~1.4 ppm [22] lose intensity, whereas resonances from the product at ~1.1 ppm gain intensity. This allowed

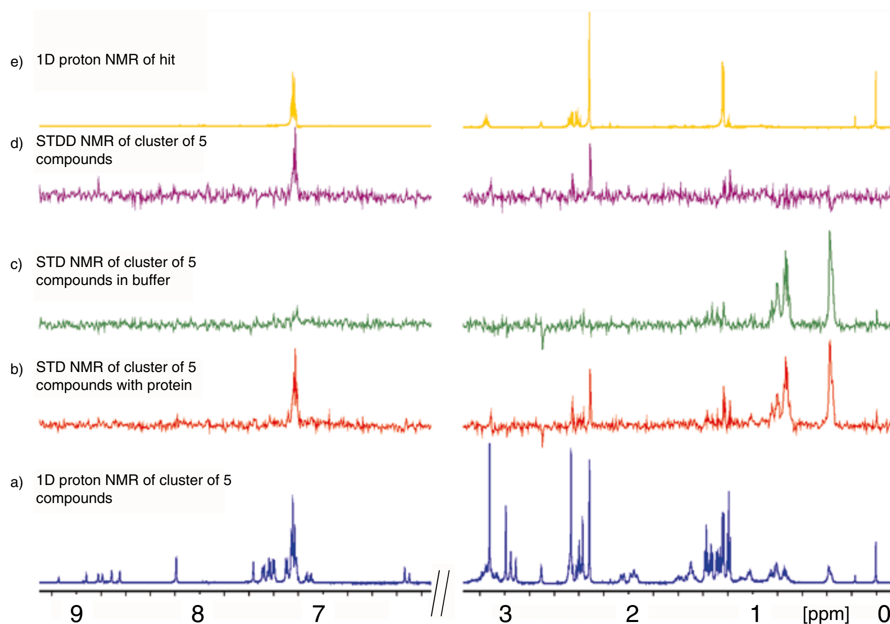


Fig. 2 Ligand-based NMR methods for fragment library screening illustrating STDD can "clean up" artifacts in STD NMR spectra. (**a**) 1D ^{1}H NMR control spectrum of a cluster of five compounds which contains a compound with a cyclopropyl signal near the irradiation frequency at 0.33 ppm. (**b**) STD NMR spectrum of that same cluster of five compounds in the presence of the target protein; (**c**) STD NMR control spectrum of that cluster of five compounds in the same buffer without target protein revealing large artifactual signals between 0 and 1 ppm; (**d**) STDD NMR spectrum reveals only STD NMR signals of the interacting compound; (**e**) 1D ^{1}H NMR control spectrum of the identified hit

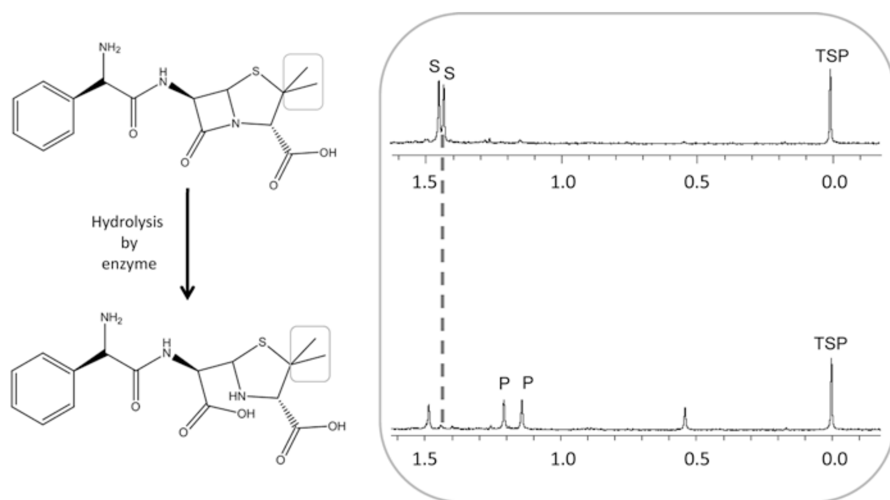


Fig. 3 The enzyme-catalyzed substrate (ampicillin) hydrolysis reaction was carried out at room temperature in the reaction buffer. Both depletion of substrate **S** (ampicillin) and formation of its hydrolyzed product **P** were monitored by 1D proton NMR spectra

us to use an unmodified β-lactam antibiotic as the substrate to monitor the enzyme's activity directly which is desirable for developing a robust, high-quality enzymatic assay to cross-validate primary screening hits.

1. Optimize the buffer conditions for 1D proton NMR-based functional assay through variations of pH, salt concentration, and cofactor (Zn) concentration (*see* **Note 10**).
2. Prepare the enzyme in the functional assay reaction buffer at 2 nM concentration.
3. Aliquot 500 μL of the enzyme solution into each well of a 1.2 mL 96-well plate.
4. For each individual fragment add 1 μL from a 100 mM DMSO-d_6 stock into a well containing 500 μL of the enzyme solution to get a final compound concentration of 200 μM.
5. Add 1 μL of substrate from the 25 mM ampicillin stock to each individual well to get a final substrate concentration of 50 μM.
6. Incubate for a proper reaction time (i.e., 25 min), and then quench by heating in a 95 °C water bath for 5 min to prevent further reaction during the NMR acquisition.
7. Transfer the enzyme solution from the 96-well plate to a 96 NMR tube rack using a liquid handler.
8. Load NMR samples into a SampleJet Automated sample changer and acquire a 1D proton NMR at the desired temperature (298 K). Each experiment takes 3 min with 128 scans.
9. Use a desired water suppression pulse program to obtain a flat baseline. In automation, the time duration for changing the sample, locking, shimming, and data collection can be within 3 min per sample. Script-based automation can be applied for baseline and phase correction.
10. Define three interesting resonance regions (i.e., substrate, product, and TSP).
11. Scale the spectral intensities to the TSP as the internal standard. Individual intensities of substrate and product can be automatically integrated based on the volume of TSP (*see* **Note 11**).
12. Export the integrations to data analysis program for the calculation of percentage (%) inhibition.
13. The percentage of inhibition is calculated according to the following equation [21]:

$$\%\,\text{inhibition} = 100 \times \left(1 - \left[\frac{P_{\text{w}}}{P_{\text{w/o}}}\right]\right)$$

where [P_w] and [$P_{w/o}$] are given by the integrals of the product signal in the presence and absence of the inhibitor, respectively.

14. The threshold for hit selection was set ≥40 % inhibition in this assay.

4 Notes

1. NMR screening work flow and the sample preparation can be in automated fashion with a programmable platform such as Hamilton (Tecan), Liquid handler (Bruker BioSpin), and SampleJet™ (Bruker BioSpin). The SampleJet™ is compatible with the Bruker BioSpin spectrometer control software such as TopSpin as well as IconNMR. This allows for higher throughput and savings in spectrometer time, while the sample is always freshly prepared prior to data acquisition. Automation workflow details can be found on the following Bruker webpage: http://www.bruker.com/products/mr/nmr/automation/automation/samplejet/overview.html.
2. Target protein needs to be stable and active in the presence of DMSO (1–5 %) for at least the duration of sample preparation and data collection, ~24 h for practical reasons.
3. Avoid non-deuterated buffers or ensure that the solvent signals do not obscure important signals in the spectrum. This can be assessed with the reference compound prior to the analysis of the spectra.
4. Fragments in mixtures can sometimes precipitate due to the high total fragment concentration in solution, which could be up to 5 mM. In most cases where precipitation is observed, we have noted that the other fragments in the mixture are still soluble and give good NMR signals. Thus the mixture is still usable.
5. NMR signal intensity of each fragment sample is compared to internal TSP signal to identify solubilized amount of the fragment.
6. The selective saturation of protein resonances can be obtained by irradiating regions of the proton NMR spectra (typically, the aliphatic region of the spectrum, between −1 and 1 ppm) that are usually well populated by methyl groups of the protein but devoid of resonances from fragment compounds.
7. Relative STD or STDD signal intensities are estimated based on an overall comparison between samples. Experimental conditions should be kept constant from sample to sample (i.e., concentrations of protein and compounds, temperature, and buffer composition).

8. Competition-STD NMR experiments can be performed within the same NMR samples for the deconvolution. The competitor is just added to the solution in the NMR tube and mixed well.
9. Fragments in the mixtures that bind to the protein target can easily be identified by comparing the STDD NMR spectrum of the hit with the 1D proton NMR spectrum of the individual fragments. Prefer to have 1D "proton chemical shift-encoded" fragment library for straightforward hit identification, from which each fragment mixture can be prepared easily such that each mixture contains only compounds with nonoverlapping proton NMR signals.
10. Possible problems to solve are that the chemical shifts between the NMR signals of substrate and product may either not be different enough and/or sensitive enough to be measurable in the NMR spectrum, that their intensity might be too low, or that they may be obstructed by other signals like buffer components.
11. Use specialized programs in order to compare multiple resonances in multiple spectra. For example, the multi-integration tool in Amix (Bruker license required): it provides a useful tool to define integration areas, perform peak integration, and carry out data analysis.

Acknowledgements

We would like to thank Yan Hou and Mark McCoy for developing the 1D proton NMR-based functional assay and Hugh Eaton and Corey Strickland for valuable discussions.

References

1. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov 6:211–219
2. Harner MJ et al (2013) Fragment-based drug discovery using NMR spectroscopy. J Biomol NMR 56:65–75
3. Wyss DF et al (2012) Combining NMR and X-ray crystallography in fragment-based drug discovery: discovery of highly potent and selective BACE-1 inhibitors. Top Curr Chem 317:83–114
4. Congreve M et al (2003) A 'rule of three' for fragment-based lead discovery? Drug Discov Today 8:876–877
5. Lau WF et al (2011) Design of a multi-purpose fragment screening library using molecular complexity and orthogonal diversity metrics. J Comput Aided Mol Des 25:621–636
6. Eaton HL, Wyss DF (2011) Effective progression of nuclear magnetic resonance-detected fragment hits. Methods Enzymol 493:447–468
7. Dalvit C et al (2006) NMR-based quality control approach for the identification of false positives and false negatives in high throughput screening. Curr Drug Discov Technol 3:115–124
8. Campos-Olivas R (2011) NMR screening and hit validation in fragment based drug discovery. Curr Top Med Chem 11:43–67

9. Hajduk PJ et al (1999) NMR-based screening in drug discovery. Q Rev Biophys 32: 211–240
10. Jahnke W et al (2001) Spin label enhanced NMR screening. J Am Chem Soc 123:3149–3150
11. Pellecchia M et al (2008) Perspectives on NMR in drug discovery: a technique comes of age. Nat Rev Drug Discov 7:738–745
12. Fernandez C, Jahnke W (2004) New approaches for NMR screening in drug discovery. Drug Discov Today Technol 1:277–283
13. Ludwig C, Guenther UL (2009) Ligand based NMR methods for drug discovery. Front Biosci (Landmark Ed) 14:4565–4574
14. Mayer M, Meyer B (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. Angew Chem Int Ed Engl 38:1784–1788
15. Claasen B et al (2005) Direct observation of ligand binding to membrane proteins in living cells by a saturation transfer double difference (STDD) NMR spectroscopy method shows a significantly higher affinity of integrin alpha(IIb)beta3 in native platelets than in liposomes. J Am Chem Soc 127:916–919
16. Dalvit C et al (2000) Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. J Biomol NMR 18:65–68
17. Dalvit C et al (2003) A general NMR method for rapid, efficient, and reliable biochemical screening. J Am Chem Soc 125: 14620–14625
18. Manzenrieder F et al (2008) Phosphorus NMR spectroscopy as a versatile tool for compound library screening. Angew Chem Int Ed Engl 47:2608–2611
19. Vanwetswinkel S et al (2005) TINS, target immobilized NMR screening: an efficient and sensitive method for ligand discovery. Chem Biol 12:207–216
20. McCoy MA et al (2005) Screening of protein kinases by ATP-STD NMR spectroscopy. J Am Chem Soc 127:7978–7979
21. Dalvit C et al (2004) Reliable high-throughput functional screening with 3-FABS. Drug Discov Today 9:595–602
22. Tung JC et al (2000) On the 1H NMR chemical shift assignments for ampicillin. Magn Reson Chem 38:126–128

Chapter 17

Practical Strategies for Small-Molecule Probe Development in Chemical Biology

Jonathan E. Hempel and Charles C. Hong

Abstract

The effective identification, selection, and implementation of small molecules for the interrogation of biological systems require an intricate understanding of the chemical principles underlying their cellular activities. While much has been published regarding the use of screening techniques in forward chemical genetics platforms and on small-molecule target identification, less emphasis has been placed on detailed strategies for evaluating, selecting, and optimizing screening hits. This chapter provides practical tools for identifying and developing promising screening hit compounds into effective tools for biological discovery.

Key words Small molecules, Probe development, Synthetic chemistry, Chemical biology, Structure-activity relationship, Physicochemical properties

1 Introduction

Chemical biology seeks to leverage the tools of chemistry to probe complex biological systems, and small molecules continue to prove their value toward this end [1, 2]. Thus, as universities and research institutes have embraced the value of high-throughput screening platforms, access to small molecules with diverse biological activities has proliferated outside of the pharmaceutical and biotechnology industries [3–5]. Chemical probes can be defined as molecular tools with which to perturb a biological system in a controlled manner resulting in a biological response, and probes can take on many forms based on the screening assay from which they are discovered [6]. Specifically, biochemical screens against a discrete protein yield biased probes for investigating a targeted hypothesis, and this approach has predominated as a method for drug discovery in recent history [7]. Conversely, genetic reporter constructs that monitor transcriptional activity can identify unique modulators of signaling pathways and networks and whole-organism phenotypic perturbations [8–13]. Cellular and organismal morphological

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_17, © Springer Science+Business Media New York 2015

outputs such as stem cell differentiation or developmental patterning also constitute platforms for unbiased probe discovery screening. Thus, from these platforms, myriad small-molecule probes have been discovered with unique characteristics, and ongoing probe development efforts will continue to facilitate fundamental biological discoveries as the quality of newly characterized probes constantly improves.

One particular manifestation of the proliferation of high-throughput screening is the field of chemical genetics, which represents a chemical biology-based extension of targeted (reverse) and random (forward) mutagenesis [14]. However, small molecules have distinct advantages over genetic studies such as temporal and dose control, and forward chemical genetics holds the potential to discover novel cellular signaling nodes that can inform future drug discovery efforts [15]. Forward chemical genetics employs chemical libraries to identify modulators of a desired cellular or organismal phenotype in an unbiased fashion, and hit selection generally involves identifying of the most effective and reproducible inducers of the specified output. While these initial screening hits can in some cases constitute bona fide biological probes, more often various limitations associated with inherent chemical screening library weaknesses and other compound characteristics preclude their widespread use for the selective and reproducible dissection of cellular signaling processes [16]. Therefore, guidelines have been proposed to aid in the development of high-quality probes. This chapter serves to overview the generally accepted characteristics of a high-quality chemical probe, to describe practical strategies for developing forward chemical genetic screening hits into validated probes, and to profile a case study of small-molecule probes of hedgehog signaling.

2 Probe Molecule Characteristics

The intended use of a chemical probe will necessarily dictate its desirable characteristics. Probes of cellular processes used for epistatic perturbations will be subject to some of the same limitations of an in vivo probe of zebrafish developmental patterning; however, drug metabolism and pharmacokinetic (DMPK) parameters essential for the action of an in vivo probe used in mouse or rat will not necessarily apply to the function of a cellular probe. Indeed, these differences will alter priorities for the beneficial characteristics of an in vitro probe when compared to an in vivo probe. For example, Lipinski's "rule of five" remains a central tenet of medicinal chemistry in drug discovery which serves as a useful limit of molecular characteristics for the development of orally bioavailable drugs, and while these guidelines generally describe preferable probe molecule properties, the "rule of five" may be

unnecessarily restrictive and unproductive for optimizing chemical genetic screening hits for in vitro studies [17]. Importantly, Lipinski has acknowledged that probe and drug discovery exist on a continuum, and the specific set of guidelines employed will be unique to the intended use of the compound developed [18].

Although instituting overly stringent rules for probe molecule characteristics may unnecessarily limit the development of potentially useful probes, a general consensus among chemical biologists has emerged stressing the importance of applying at very least a basic set of criteria to probe development [16]. In light of examples such as the broad-spectrum kinase inhibitor staurosporine and the protein kinase A inhibitor H89, which have both been called into question as lacking predictive power due to off-target effects, a general set of probe development guidelines seeks to avoid pollution of the literature with weak or incorrect conclusions stemming from low-quality probes [19, 20].

Therefore, multiple overlapping sets of guidelines have been proposed to foster the contribution of well-characterized and reliable probes to the chemical biology toolkit [16, 19, 21]. Frye and Workman agree that generally desirable characteristics include:

- Dose-dependent in vitro biochemical potency of <100 nM and cellular potency of <1–10 μM.
- Target selectivity of >10- to 100-fold over related and unrelated target family members.
- Aqueous solubility of >100 μM and validated membrane permeability.
- Known structure freely available for use and a structurally related inactive analog.
- Biological hypothesis answered based on successful target modulation of a cellular output.

These characteristics aim to guide the development of robust, reproducible, and easily implementable probe compounds. Sufficient potency supports necessary target selectivity, and selectivity enables confident associations to be made between target modulation and phenotypic output for hypothesis testing. Membrane permeability, while often implied by the presence of a phenotypic response in a cellular or organism-based screen, also envelops passive and active transport as well as potential active reverse cellular efflux.

As forward chemical genetic screens are commonly performed in the context of a human disease hypothesis, a blurred line between probe development and drug discovery can potentially create incentives to limit the use of discovered compounds through patent protection. While the intent of chemical biology is not to stifle therapeutic development but conversely to enable future

medical advances, a distinction should be made early in the project between the intended use of probe development for biological tool discovery and direct therapeutic development; however, these two directions are not mutually exclusive as long as the final probe molecule is readily available for the research community. Importantly, unrestricted access to and use of probe compounds enables cross-laboratory validation and fosters the mission of chemical biology to support fundamental biological discovery through application of chemical tools.

With these guidelines in mind, specific approaches to probe development can be implemented to transition forward chemical genetic screening hits into validated probes of biological processes, and these strategies are highlighted in Subheading 3.

3 Screening Hit Selection and Optimization

No single prescription for probe development strategies will appropriately encompass the specific issues associated with each campaign. Therefore, the order in which validation and optimization studies occur will vary, as will the priority rank of each step. Nevertheless, it is essential to view the overall process as one not just of narrowing hits but also improving selected hits for optimization toward the principles outlined in Subheading 2 through chemical synthesis [22] (Fig. 1). Therefore, close consultation with a synthetic and/or medicinal chemist is highly recommended.

3.1 Hit Selection

Traditional forward chemical genetic screens operate on a low- to medium-throughput scale involving testing of hundreds to thousands of compounds. From the point of screening, the obvious first step involves eliminating compounds that fail to induce the desired phenotypic output. For a well-designed assay and an appropriately sized and composed compound library, a reasonable percentage of compounds should elicit a positive response, the definition of which will vary based on screen design (single concentration vs. multiple concentrations). Although tempting to select only compounds with the strongest response, setting a hit threshold too high can result in unintended potency optimization at the expense of other equally valuable probe properties. Ultimately, chemical scaffolds are more likely to have pliable potency constraints than physical characteristics; therefore, early selection for potency may unnecessarily hamstring future improvement efforts.

Additional narrowing parameters will depend on the nature of the desired probe and its intended system, but the ultimate goal of hit selection should be to identify two or more structurally unrelated hit compounds for validation, profiling, and synthetic optimization [23]. Ideally, each structurally unique hit moved into further

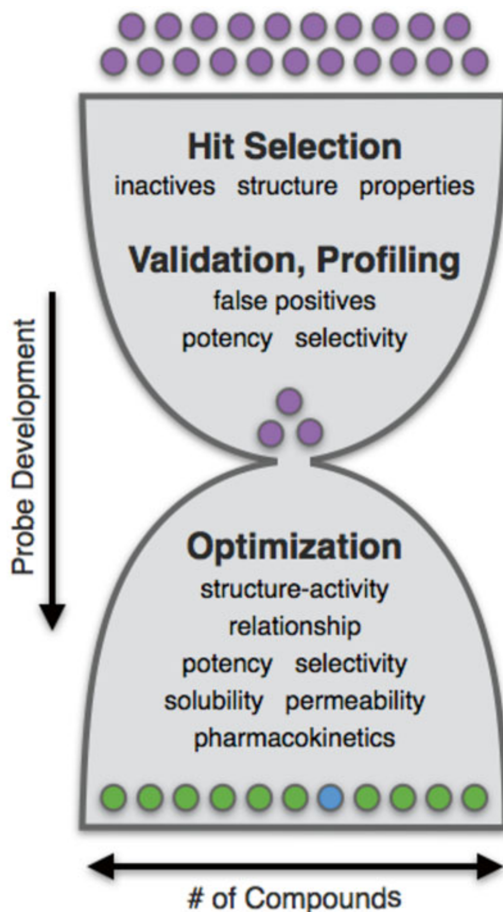


Fig. 1 The hourglass of probe development. Screening hit molecules are narrowed to a small number of structurally unique lead hits through the process of selection, validation, and profiling. Profiling provides benchmarks for essential probe properties which are optimized through structure-activity relationship (SAR) studies and analog synthesis to arrive at one or more bona fide biological probes

characterization studies should be representative of multiple active hits within its structural class, where structural similarity determinations should focus on hits with a conserved core structure and varying appended functional groups. This strategy is focused twofold on providing redundancy for potentially intractable SAR studies on an individual compound series while affording initial validation for the feasibility of structural modification with activity retention.

Hits containing certain chemical functional groups should be evaluated with caution: at very least, controversy remains over the presence of reactive groups such as aldehydes, epoxides, and α,β-unsaturated carbonyls in probe and drug molecules [24].

Exceptions certainly exist in the realm of clinical therapeutics such as the proteasome inhibitor carfilzomib and in the emerging chemical biology field of activity-based protein profiling (ABPP); however, without a direct hypothesis relating to electrophilic modification, early elimination of these hit compounds may prove fortuitous [25, 26].

Final considerations should include ease of synthetic assembly and structural variation within the proposed synthetic route as well as inherent physicochemical properties such as aqueous solubility and membrane permeability, where preferred physical properties of a lead hit molecule will improve the likelihood that optimized synthetic analogs will also possess these important properties [27]. Additionally, consideration of the calculated log partition coefficient (cLogP) of potential lead hits using in silico methods (i.e., ChemBioDraw) can provide insight into initial aqueous solubility issues and guide hit selection. For synthetic route strategies, consultation with a chemistry collaborator can illuminate concerns prior to final selections. The number of viable hit compounds will ultimately impact the number of criteria that can be applied to hit selection; however, consideration of selection parameters in addition to phenotypic response will significantly improve the success of follow-up studies and increase the potential for the development of a high-quality probe.

3.2 Hit Validation and Profiling

Screening campaigns invariably produce false-positive and nonreproducible results due to the scale of both the compound library and the screen itself [28]. Mis-annotation of compounds in either the screening collection or during the process of compound transfer is not uncommon, although the widespread use of robotic automation in the process minimizes this issue, and long-term stock solution stability issues can lead to compound degradation. Additionally, synthetic impurities in compound stocks or even bioactive contaminants in laboratory plastic containers can either reduce the effective concentration of the screening compound in the source plate stock or directly lead to nonspecific assay interference causing a false-positive readout [29]. Therefore, detailed hit validation should be undertaken prior to synthetic optimization studies.

Initially, hits should be retested within the screening assay from the original compound stock to confirm reproducible phenotype modulation. If the screen was performed at a single concentration, the half-maximal effective phenotypic response concentration (EC_{50}) should be determined to show dose-dependent output associated with the hit of interest. To associate a compound structure to a true assay response, the compound should be resynthesized in collaboration with a synthetic chemist or, if not feasible, purchased from an alternate vendor: it is important to keep in mind that even repurchased library compound stocks have been shown to contain high concentrations of inorganic metal impurities that can lead to

false-positive assay readouts [30]. Thus, retesting and resynthesis are critical for addressing the potential for mis-annotation addressed above as well as the possibility of vendor-specific impurities and screening collection compound stock degradation. Finally, the potential for false-positive hits requires an investigation into assay interference mechanisms such as compound fluorescence in the case of a fluorescent screening assay readout, or in promoter-driven reporter expression assays utilizing firefly luciferase and others [31, 32].

Once the chosen hits have been confirmed as true positives, ideally at least two structurally unique compounds will remain. With these compounds, parameters listed in Subheading 2 that describe a high-quality probe should be profiled as a starting point for optimization studies. Functional target knowledge at this early stage will dictate the immediate assay focus: either the screen design enriched for activity at a specific signaling node, on a particular signaling pathway, or was performed in a wholly unbiased manner such as for some in vivo screens. And it is widely accepted that target identification is the bottleneck of phenotypic screening: as the complexity of the biological system screened increases, this task will exponentially increase in difficulty [33]. While a confirmed functional target is not required at this stage, this knowledge will improve the number and quality of assays available for hit profiling. Nevertheless, appropriate context-specific profiling should illuminate compound limitations to be addressed by synthetic optimization.

In the case that the functional cellular target of the lead hits is known and validated, determine the in vitro biochemical half-maximal inhibitory concentrations (IC_{50}) with an accepted assay method. If the capacity for generating this data does not exist institutionally, many fee-for-service resources are available. Without a known target but knowledge of the signaling pathway modulated by the lead hits, a cellular reporter assay can be used in lieu of a biochemical assay; however, it is important to keep in mind that compensatory and multimodal effects can confound SAR efforts in cellular systems. These on-target biochemical and cellular assay profiles will serve as potency benchmarks for later SAR and optimization studies.

Off-target profiling will provide valuable information regarding selectivity concerns. Specifically, when the protein target is known, IC_{50} values should be generated for closely related family members such as genetic and splice isoforms, and broad profiling should also be undertaken. For kinase inhibitors, due to the highly conserved ATP-binding pocket, selectivity is critical, and lead hits should be at minimum tested at single concentrations of 10- to 100-fold the on-target IC_{50} against >100 kinases [21]. However, with the availability of fee-for-service resources such as the KINOMEscan®, which can test a compound against over 400 kinases, selectivity profiling requirements are ever increasing. In both cases

of known and unknown target, off-target signaling pathway profiling should also be performed. These data will ultimately guide areas of focus, such as minimizing off-target activity against an unrelated signaling pathway or a functionally related protein, for future synthetic optimization.

Finally, physicochemical properties of importance such as aqueous solubility should be benchmarked, and for in vivo probes, absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties should be subjected to preliminary evaluation. While the luxury of removing hits with undesirable properties during the selection phase may present an opportunity to de-emphasize these factors during optimization studies, issues cannot always be avoided and, with continuous profiling of hits and analogs, can be actively solved. Aqueous solubility assays can determine the remaining concentration of a compound when mixed from an organic stock solution over time (kinetic) or the ability of a pure compound to dissolve in aqueous medium (thermodynamic), and factors such as the salt composition and pH of the aqueous medium can affect the nature of the information derived from the assay; therefore, careful consideration of possible assay formats is required [34]. And although ADMET properties are essential during the in vivo probe development process, the details of optimization are outside of the scope of this chapter and should involve close consultation with a laboratory with expertise in drug metabolism and pharmacokinetics (DMPK) [27].

The principles of hit compound validation and profiling outlined here in Subheading 3.2 are aimed at completing screening hit selection to identify multiple structural series of chemical probes of a biological process of interest. As depicted in Fig. 1, the original screening library is narrowed to a limited number of lead hits, and Subheading 3.3 overviews the process of expanding these structural series through analog synthesis to arrive at an optimized chemical probe.

3.3 Hit Optimization

Some chemical genetic screening probe discovery campaigns incorrectly end at the stage of hit narrowing: it is common that screening library hits will not fulfill the guidelines laid out in Subheading 2. Therefore, structure-activity relationship (SAR) studies can illuminate synthetic modifications that can lead to improvement of potency, selectivity, solubility, and other properties of interest. The general concept of SAR investigation involves the methodical synthetic substitution of chemical functional groups in discrete locations on the hit molecule and comparison of the resultant change in biological assay activity. In this way, if for example the hit molecule has a terminal phenyl moiety, logical SAR substitutions would include addition of electron-withdrawing groups (EWG) (path **1**) such as a bromine at the 4-position as shown in Fig. 2 as well as electron-donating groups (EDG) (path **2**) at the 4-position. However,

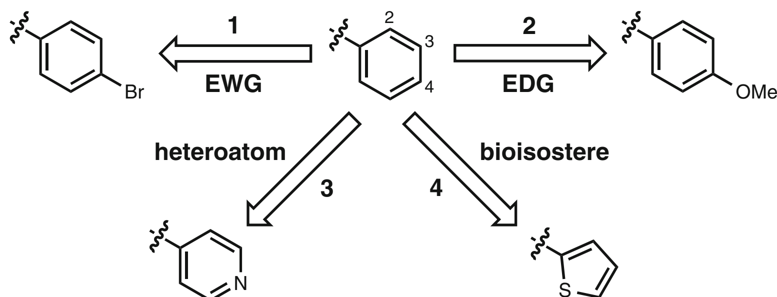


Fig. 2 The basics of structure-activity relationship modifications. From a hit compound containing a terminal phenyl moiety (*center*), SAR involves addition of electron-withdrawing groups (EWG) (path **1**) and electron-donating groups (EDG) (path **2**) at varying positions around the ring system. Heteroatom substitution can alter the binding and physical properties of the molecule (path **3**), and bioisosteric replacements (path **4**) can facilitate significant structural changes without loss of on-target potency

substitution of these functionalities at the 2- and 3-positions of the phenyl ring should also be performed. Each of these synthetic analogs should be evaluated in biochemical or cellular reporter assays for potency, and future analog synthesis should incorporate chemically similar functionalities to those that increase potency in the first round of analog synthesis.

Paths **1** and **2** of Fig. 2 represent simplistic but descriptive manifestations of the SAR logic. However, the pyridine for phenyl substitution example (path **3**) can optimize properties such as aqueous solubility due to the incorporation of a polar atom, in this case nitrogen. And finally, path **4** represents the important SAR concept of bioisosterism in which substitutions can be rationally made based on electronic structure and atomic size of functional groups and have literature precedent for activity enhancement: for example, the thiophene for phenyl substitution in path **4** [35]. Another valuable example of bioisosterism is the fluorine for hydrogen substitution which has successfully aided numerous probe and drug development programs [36].

While a comprehensive review of SAR and analog synthesis in probe and drug discovery commands its own textbook, certain general concepts remain pertinent to this chapter. Specifically, although chemical appendages such as ring systems at the molecular termini are relatively easily modified during SAR studies and can be targeted first for initial optimization, more difficult synthetically is the modification of the core scaffold structure which brings together the appendages. But placing value in its own SAR can enable biological activity improvements that may otherwise be unattainable with the hit scaffold, and this process is neatly described by the term scaffold hopping that has been previously

reviewed and continues to evolve [37]. Overall, the process of SAR and probe optimization can be viewed as a closed loop in which the first round of analog synthesis informs the modifications made for the second round and so on [22].

With the general strategies outlined for the logic of SAR studies, similar logic should be applied to the first-line biological assays employed to monitor and evaluate newly synthesized analogs. Assays of importance will be driven by the properties needing optimization, for example aqueous solubility, for which all compounds should be evaluated in parallel for biochemical potency as well as an optimized solubility assay. However, potency in either biochemical or cellular assays should always be evaluated regardless of the starting hit's potency, as chemical modifications can render synthetic analogs less potent. And as previously mentioned in Subheadings 3.1 and 3.2, moving multiple structurally unique hit series into SAR studies can stem potential roadblocks associated with the intractability of a single scaffold to SAR improvement.

As prioritized properties are improved through iterative SAR optimization, secondary biological assays for off-target selectivity and ADMET properties should fill in final profiling of the putative chemical probes. If characteristics of a final probe molecule are still lacking, fine-tuning through additional synthesis can address these issues as well. Ultimately, each optimization effort will take its own direction based on the desired final properties and the starting characteristics, but the general structure of iterative SAR workflow should largely remain the same.

4 Case Study: Small-Molecule Probes of Hedgehog Signaling

Hedgehog (Hh) signaling has a rich history in chemical biology with a wide variety of small-molecule probes acting on numerous signaling nodes [38]. Hh inhibition in development leads to cyclopia in vertebrates, and its dysregulation in adult humans causes basal cell carcinoma (BCC), medulloblastoma, pancreatic cancer, and others. Thus, the available tools for its perturbation can serve as a useful case study for the development and use of probe molecules in biology.

Many details of Hh signaling are well understood, and numerous Hh proteins have been targeted by probe development programs; however, significant knowledge gaps remain (Fig. 3). In the absence of extracellular sonic hedgehog (Shh) ligand, the twelve-pass transmembrane receptor Patched (Ptc) exerts inhibitory influence on the ciliary translocation and activity of the seven-pass transmembrane G-protein -coupled receptor (GPCR) Smoothened (Smo). Without Smo activity, Suppressor of Fused (Sufu) guides Gli transcription factors for proteolytic processing and degradation leading to repression of transcription. When Shh is present and binds to

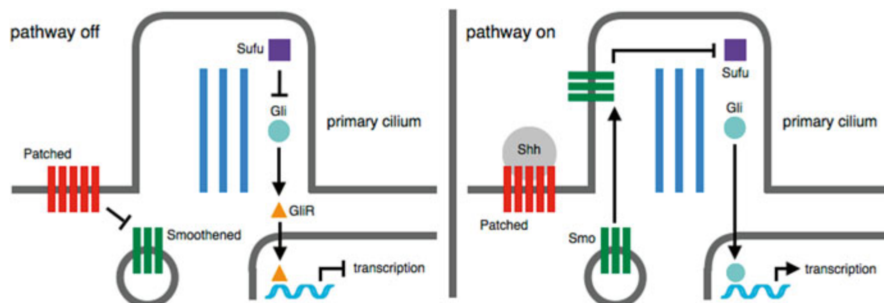


Fig. 3 Hedgehog signal transduction. In the absence of Shh ligand, Patched exerts inhibition on Smoothened (Smo), and Suppressor of Fused (Sufu) directs processing of Gli transcription factors into their repressor forms leading to transcription inactivation. When Patched binds Shh, Smo translocates to the primary cilium to inhibit the activity of Sufu, and active Gli promotes Hh target gene transcription

Table 1
Small-molecule probes of hedgehog signaling

Hh probe	Hh target	Mode of action	Discovery	Hh EC_{50} (nM)	cLogP
Robotnikinin	Shh	Inhibitor	SM microarray to ShhN	3,100	3.22
Cyclopamine	Smo	Antagonist	Teratogenic natural product	300	5.02
GDC-0449	Smo	Antagonist	Shh-induced Gli-luc HTS	13	2.74
SANT-1	Smo	Antagonist	ShhN-induced Gli-luc HTS	20	4.42
Purmorphamine	Smo	Agonist	MEF differentiation HTS	1,000	7.17
SAG	Smo	Agonist	Shh-induced Gli-luc HTS	3	5.29
Ciliobrevin A	dynein	Inhibitor	SAG-induced Gli-luc HTS	7,000	3.66
GANT-61	Gli1/2	Inhibitor	Gli1-induced Gli-luc HTS	5,000	4.91

Numerous probes of hedgehog (Hh) signaling target varying levels of signal transduction with a wide range of effective concentrations. Physical properties of these probes, including their calculated aqueous solubility, affect the ease of their implementation as a tool for Hh signaling interrogation

Ptc, Ptc no longer inhibits Smo ciliary translocation. Smo activity leads to Sufu dissociation with Gli transcription factors which translocate to the nucleus and activate Hh target gene transcription.

A retrospective study of Hh probes, as well as chemical biological probes in general, will inevitably highlight shortcomings in the context of guidelines described in Subheading 2; general probe development guidelines were only proposed around 2010 in response to mounting concerns regarding reporting of low-quality probes. Additionally, detailed off-target selectivity profiling has only recently become a common step in probe development. However, the compounds listed in Table 1 are generally representative of successful Hh signaling probes, and acknowledgement of their limitations can inform future efforts in all areas of chemical genetics.

Historically, cyclopamine was the first compound to known to inhibit Hh signaling, identified as a teratogen leading to the birth of cyclopic lambs in Idaho [38]. It has successfully illuminated the cholesterol dependence of Shh maturation, and its fluorescent BODIPY analog is commonly used in the determination of Hh inhibitor binding to Smo [39]. Cyclopamine's limited aqueous solubility demonstrated by a cLogP of >5 necessitated the synthesis of more highly cell-permeable and soluble analogs such as KAAD-cyclopamine which also exhibits increased Hh potency. And the orally available cyclopamine analog IPI-926 (saridegib) has been tested through phase 2 clinical trials for pancreatic cancer.

Smo is the most commonly targeted node of the Hh pathway with many more antagonists discovered in recent years than those listed in Table 1. However, in addition to cyclopamine, SANT-1 and GDC-0449 represent differing probes for Smo antagonism [40, 41]. Both compounds were identified from high-throughput cellular reporter screens for inhibitors of Shh-induced Gli transcription and inhibit Hh signaling with similar Hh EC_{50} values; however, GDC-0449 importantly is the US Food and Drug Administration (FDA) approved for the treatment of advanced basal cell carcinoma (BCC) and is more commonly used in efforts to identify Hh pathway inhibitors as a positive control.

Smo agonists have played an integral role in the identification of downstream of Smo Hh inhibitors, which has become increasingly important as gain-of-function Smo inhibitor resistance has been encountered in the clinic [42]. Purmorphamine was identified in a screen for inducers of osteogenic differentiation in the mouse embryonic mesenchymal mesoderm fibroblast C3H10T1/2 cell line and was determined to function through Smo agonism [43, 44]. Due to its limited aqueous solubility and physical properties, many recent cellular Hh studies have relied on the Smo agonist SAG which activates Gli transcription at a concentration of 3 nM [40].

Although most small molecules target Hh at the level of Smo, increasing numbers of probes both upstream and downstream are being identified. Upstream of Smo, robotnikinin has been shown to inhibit the N-terminally processed form of Shh (ShhN), and the Gli1/2 inhibitor GANT-61 represents the most well-characterized modulator of transcription factor-level Hh inhibition [45, 46]. And the recently reported dynein inhibitor ciliobrevin A is an example of the emerging utility of phenotypic screening in identifying novel biological targets for regulation of downstream transcriptional activity [47]. However, the limited Hh signaling inhibition potency of these probes may hamper their widespread chemical biological applicability.

Overall, small-molecule probes of Hh signaling encompass their own case study of phenotypic screening for probe development, and their further use in modulating the activity of Hh

signaling will continue to aid with identifying novel Hh probes and improving fundamental understanding of the intricacies of cellular signaling.

5 Conclusions

Small-molecule probes of cellular processes are invaluable tools for answering fundamental biological questions, and phenotypic screening can support the identification of novel probes. However, to constitute a useful in vitro or in vivo probe, certain characteristics including on-target potency, off-target selectivity, physical properties, and availability should be fulfilled. As screening hits often fail to meet preferred probe criteria, a practical approach to identifying lead probe compounds and optimizing their desired properties has been outlined within this chapter. With these strategies, it is hoped that more high-quality probes will be developed for use within the greater chemical biology community toward improved biological understanding.

Acknowledgements

We thank Gary Sulikowski for his helpful guidance and input. J.E.H. is supported by an American Heart Association Postdoctoral Fellowship (14POST19550002). C.C.H. is supported by the US NIH NHLBI (1R01HL104040), the US Veterans Administration (101BX000771), and the Cali Family Foundation.

References

1. Bucci M, Goodman C, Sheppard TL (2010) A decade of chemical biology. Nat Chem Biol 6:847–854
2. Alaimo PJ, Shogren-Knaak MA, Shokat KM (2001) Chemical genetic approaches for the elucidation of signaling pathways. Curr Opin Chem Biol 5:360–367
3. Frearson JA, Collie IT (2009) HTS and hit finding in academia–from chemical genomics to drug discovery. Drug Discov Today 14: 1150–1158
4. Hasson SA, Inglese J (2013) Innovation in academic chemical screening: filling the gaps in chemical biology. Curr Opin Chem Biol 17:329–338
5. Huryn DM (2013) Drug discovery in an academic setting: playing to the strengths. ACS Med Chem Lett 4:313–315
6. Garcia-Serna R, Mestres J (2011) Chemical probes for biological systems. Drug Discov Today 16:99–106
7. Macarron R et al (2011) Impact of high-throughput screening in biomedical research. Nat Rev Drug Discov 10:188–195
8. Castoreno AB, Eggert US (2011) Small molecule probes of cellular pathways and networks. ACS Chem Biol 6:86–94
9. Prior M et al (2014) Back to the future with phenotypic screening. ACS Chem Neurosci 5:503–513
10. Lee JA, Berg EL (2013) Neoclassic drug discovery: the case for lead generation using phenotypic and functional approaches. J Biomol Screen 18:1143–1155
11. Eggert US (2013) The why and how of phenotypic small-molecule screens. Nat Chem Biol 9:206–209
12. Tamplin OJ et al (2012) Small molecule screening in zebrafish: swimming in potential drug therapies. Wiley Interdiscip Rev Dev Biol 1:459–468

13. Gonzalez C (2013) Drosophila melanogaster: a model and a tool to investigate malignancy and identify new therapeutics. Nat Rev Cancer 13:172–183
14. O'Connor CJ, Laraia L, Spring DR (2011) Chemical genetics. Chem Soc Rev 40: 4332–4345
15. Spring DR (2005) Chemical genetics to chemical genomics: small molecules offer big insights. Chem Soc Rev 34:472–482
16. Workman P, Collins I (2010) Probing the probes: fitness factors for small molecule tools. Chem Biol 17:561–577
17. Lipinski CA et al (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3–25
18. Lipinski C, Hopkins A (2004) Navigating chemical space for biology and medicine. Nature 432:855–861
19. Frye SV (2010) The art of the chemical probe. Nat Chem Biol 6(3):159–161
20. Murray AJ (2008) Pharmacological PKA inhibition: all may not be what it seems. Sci Signal 1:re4
21. Cohen P (2010) Guidelines for the effective use of chemical inhibitors of protein function to understand their roles in cell regulation. Biochem J 425:53–54
22. Lipinski CA (2010) Overview of hit to lead: the medicinal chemist's role from HTS retest to lead optimization hand off. In: Hayward MM (ed) Lead-seeking approaches: topics in medicinal chemistry, 5th edn. Springer, Berlin, pp 125–140
23. Wagner BK, Clemons PA (2009) Connecting synthetic chemistry decisions to cell and genome biology using small-molecule phenotypic profiling. Curr Opin Chem Biol 13:539–548
24. Rishton GM (2003) Nonleadlikeness and leadlikeness in biochemical screening. Drug Discov Today 8:86–96
25. Singh J et al (2011) The resurgence of covalent drugs. Nat Rev Drug Discov 10:307–317
26. Heal WP, Dang THT, Tate EW (2011) Activity-based probes: discovering new biology and new drug targets. Chem Soc Rev 40:246–257
27. Gleeson MP et al (2011) Probing the links between in vitro potency, ADMET and physicochemical parameters. Nat Rev Drug Discov 10:197–208
28. Matson SL et al (2009) Best practices in compound management for preserving compound integrity and accurately providing samples for assays. J Biomol Screen 14:476–484
29. McDonald GR et al (2008) Bioactive contaminants leach from disposable laboratory plasticware. Science 322:917
30. Hermann JC et al (2013) Metal impurities cause false positives in high-throughput screening campaigns. ACS Med Chem Lett 4:197–200
31. Thorne N, Auld DS, Inglese J (2010) Apparent activity in high-throughput screening: origins of compound-dependent assay interference. Curr Opin Chem Biol 14:315–324
32. Thorne N et al (2012) Firefly luciferase in chemical biology: a compendium of inhibitors, mechanistic evaluation of chemotypes, and suggested use as a reporter. Chem Biol 19:1060–1072
33. Schenone M et al (2013) Target identification and mechanism of action in chemical biology and drug discovery. Nat Chem Biol 9:232–240
34. Williams HD et al (2012) Strategies to address low drug solubility in discovery and development. Pharmacol Rev 65:315–499
35. Meanwell NA (2011) Synopsis of some recent tactical application of bioisosteres in drug design. J Med Chem 54:2529–2591
36. Böhm H-J et al (2004) Fluorine in medicinal chemistry. ChemBioChem 5:637–643
37. Sun H, Tawa G, Wallqvist A (2012) Classification of scaffold-hopping approaches. Drug Discov Today 17:310–324
38. Heretsch P, Tzagkaroulaki L, Giannis A (2010) Modulators of the hedgehog signaling pathway. Bioorg Med Chem 18:6613–6624
39. Heretsch P, Tzagkaroulaki L, Giannis A (2010) Cyclopamine and hedgehog signaling: chemistry, biology, medical perspectives. Angew Chem Int Ed 49:3418–3427
40. Chen JK et al (2002) Small molecule modulation of smoothened activity. Proc Natl Acad Sci 99:14071–14076
41. Robarge KD et al (2009) GDC-0449—a potent inhibitor of the hedgehog pathway. Bioorg Med Chem Lett 19:5576–5581
42. Yauch RL et al (2009) Smoothened mutation confers resistance to a Hedgehog pathway inhibitor in medulloblastoma. Science 326:572–574
43. Wu X et al (2002) A small molecule with osteogenesis-inducing activity in multipotent mesenchymal progenitor cells. J Am Chem Soc 124:14520–14521
44. Sinha S, Chen JK (2005) Purmorphamine activates the Hedgehog pathway by targeting Smoothened. Nat Chem Biol 2:29–30

45. Stanton BZ et al (2009) A small molecule that binds Hedgehog and blocks its signaling in human cells. Nat Chem Biol 5:154–156
46. Hyman JM et al (2009) Small-molecule inhibitors reveal multiple strategies for Hedgehog pathway blockade. Proc Natl Acad Sci U S A 106:14132–14137
47. Firestone AJ et al (2012) Small-molecule inhibitors of the AAA+ ATPase motor cytoplasmic dynein. Nature 484:125–129

Chapter 18

Principal Component Analysis as a Tool for Library Design: A Case Study Investigating Natural Products, Brand-Name Drugs, Natural Product-Like Libraries, and Drug-Like Libraries

Todd A. Wenderski, Christopher F. Stratton, Renato A. Bauer, Felix Kopp, and Derek S. Tan

Abstract

Principal component analysis (PCA) is a useful tool in the design and planning of chemical libraries. PCA can be used to reveal differences in structural and physicochemical parameters between various classes of compounds by displaying them in a convenient graphical format. Herein, we demonstrate the use of PCA to gain insight into structural features that differentiate natural products, synthetic drugs, natural product-like libraries, and drug-like libraries, and show how the results can be used to guide library design.

Key words Principal component analysis (PCA), Medium rings, Macrocycles, Ring expansion, Natural products, Drugs, Libraries, Diversity-oriented synthesis

1 Introduction

Principal component analysis (PCA) is a mathematical method for dimensionality reduction that allows for multidimensional datasets to be visualized using two- or three-dimensional plots with minimal loss of information [1, 2]. When applied in the context of diversity-oriented synthesis, PCA is primarily used to visualize similarities and differences within collections of compounds based on structural and physicochemical parameters, and can be leveraged in library design [3]. Molecular weight, stereocenters, rotatable bonds, hydrophobicity, and aqueous solubility are a few examples of parameters commonly included in such analyses. Herein, we selected 20 structural and physicochemical parameters for analysis based on previously identified correlations of these parameters with oral bioavailability, cell permeability, solubility, and binding selectivity, as well as their ability to distinguish synthetic drugs

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_18, © Springer Science+Business Media New York 2015

from natural products (*vide infra*). Each compound in our analysis is represented as a 20-dimensional vector defined by the structural and physicochemical parameters. PCA rotates these vectors onto a new set of orthogonal axes called principal components, in which the variance retained from the original data is maximized on each successive principal component. As such, the three-dimensional plot we show in this example retains 75 % of the variance from the full 20-dimensional dataset.

PCA can also be used to guide the design of chemical libraries. This is important in drug discovery because current drugs are limited in both structure and function. For example, current small-molecule drugs address only about 1 % of the protein targets encoded in the human genome [4], and half of those target only four protein classes: rhodopsin-like G-protein receptors, nuclear receptors, and voltage- and ligand-gated ion channels. In contrast, natural products are known to target a broader range of protein classes and have led to the majority of antibacterials (65 %) and anticancer drugs (75 %) [5]. Therefore, novel libraries of compounds that share the structural features of natural products are attractive for the discovery of lead compounds to evaluate new therapeutic targets.

Along these lines, many macrocycle and medium-ring-containing natural products have compelling biological activities. This key cyclic framework presents functional groups to biological targets in appropriate pharmacophoric conformations [6–8]. Compared to their corresponding linear congeners, macrocycles can provide increased binding affinity [9], improved bioavailability [10], and, in some cases, enhanced cell permeability [11], which are desirable pharmacological properties in the development of new drugs.

However, despite these attractive features of macrocycles and medium rings, they remain severely underexploited in current drug and probe discovery efforts [12, 13], due to challenges associated with their synthesis. To address the underrepresentation of these compounds, we have sought to circumvent the inherent limitations of classical cyclization-based strategies for macrocycle and medium-ring synthesis by developing alternative ring-expansion approaches that are tolerant of a broad range of substitution patterns and functional groups. We recently developed two such methods (Fig. 1) [12, 13], both of which can be employed on gram scale, provide products bearing handles for further diversification, and are transferable to parallel synthesis platforms.

Herein, we describe the use of PCA to assess how libraries of compounds produced using these synthetic routes compare to natural products and what structural and physicochemical parameters distinguish them from synthetic drugs and drug-like libraries. The information harnessed from PCA can also direct downstream

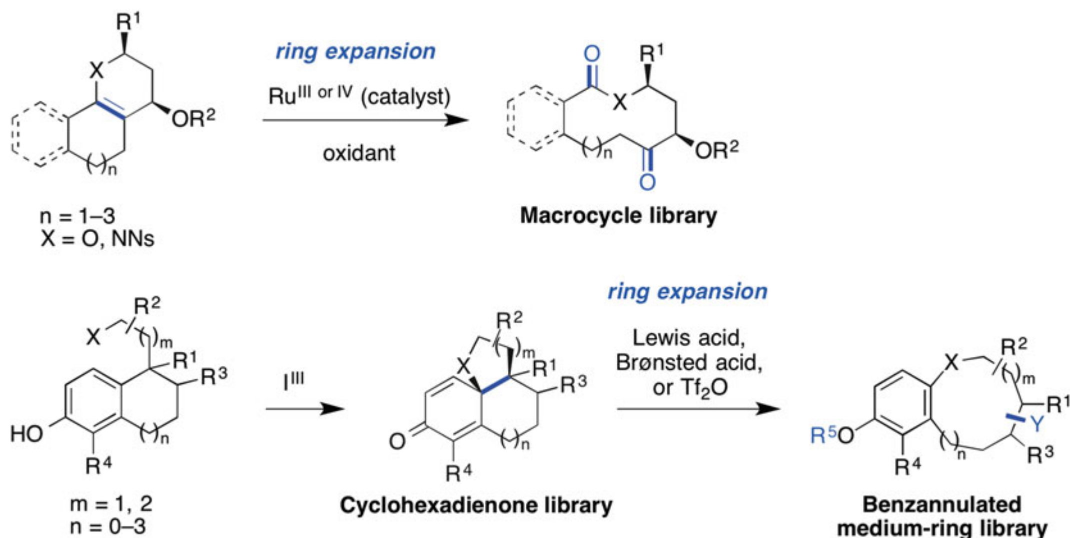


Fig. 1 Recently developed routes to natural product-like macrocycle and medium-ring libraries using ring expansion strategies

modifications of a scaffold to obtain molecules that are more characteristic of a targeted class, such as natural products.

In this example, we show that compounds appearing in the proximity of the drug-like region of the PCA plot can be modified to have greater natural product-like properties by addressing several influential structural and physicochemical parameters. The relative contributions of structural and physicochemical parameters to each principal component (PC) axis are obtained from the loading data and loading plots produced from PCA. In the analysis presented herein, the number of oxygen atoms, hydrogen bond donors, and hydrogen bond acceptors are among the most influential parameters for PC1. Stereochemical density (the number of stereocenters normalized to molecular weight) and the fraction of sp^3-hybridized carbons are large contributors to PC2. We further demonstrate that these structural and physicochemical parameters can be addressed by chemical modifications of our library members to increase their natural product-like character. Subsequent analysis of these modified compounds in PCA demonstrates their increased penetration into natural product-like regions of the plot. This work illustrates how insights gleaned from PCA can be used in the planning of chemical libraries to probe targeted areas of chemical space.

2 Materials

This analysis requires the use of several software packages that are either commonly available in chemistry labs or freely available for download. The following software and versions were used for this protocol:

1. Mac OS 10.5.8 or Windows 7 (procedure described for Mac OS with specific changes for Windows 7 users indicated).
2. CS ChemBioDraw Ultra 12.0.3 (CambridgeSoft).
3. Microsoft Office 2008.
4. Instant JChem 5.3.8 (ChemAxon, free Academic License available).
5. Virtual Computational Chemistry (VCC) Laboratory: http://www.vcclab.org/lab/alogps/start.html (requires a JAVA-enabled browser).
6. R 2.9.2 (open-source R Project for Statistical Computing, available from http://www.r-project.org).

3 Methods

3.1 Calculation of Physicochemical Parameters

1. Obtain SMILES codes for all of the compounds to be included in the PCA (*see* **Notes 1** and **2**). SMILES codes for known compounds can be obtained from PubChem (http://pubchem.ncbi.nlm.nih.gov/) or other online resources. For new compounds, SMILES codes can be generated using ChemBioDraw (*see* **Note 3**).
2. Create a new MS Excel file containing one column for compound names (Column A) and one column for SMILES codes (Column B) (*see* **Note 4**). Do not include a header row. Group the compounds by compound class (such as Drugs, Natural Products). Save the MS Excel file as a Text (tab delimited) (.txt) file that will be used in Subheading 3.1, **step 4**. Delete the compound names column and save an additional .txt file that contains only SMILES codes, which will be used later will be used later for batch processing (Subheading 3.1, **step 7**).
3. Using Instant JChem, open a new project (File > New Project), and click on "Next" (*see* **Note 5**). Enter a project name and then click on "Finish".
4. From Instant JChem, import the MS Excel file containing the compound names and SMILES codes by selecting File > Import File, and then click on "Next" (*see* **Note 6**). Click on the folder icon next to the "File to import" field, and then navigate to and select the .txt file containing compound names and SMILES codes. Under "File Format" choose "Delineated text files (*.csv, *.tab, *.txt)", and then click on "Open". After Instant JChem has finished scanning the file and indicated the number of fields found, click on "Next". The "Field details" panel gives a summary of the fields to be imported from the text file. The structure, molecular weight, molecular formula, and compound names of each entry are displayed by default

Field 1	Structure	Formula	Molecular	N Atom count	O Atom count	HBD	HBA	Rotatable	nStereo	TPSA
Abilify		C23H27Cl2N3O2	448.38	3	2	1	4	7	0	44.81
Aciphex		C18H21N3O3S	359.44	3	3	1	5	8	1	77.1
Actos		C19H20N2O3S	356.44	2	3	1	4	7	1	68.29
Adderall		C9H13N	135.21	1	0	1	1	2	1	26.02

Fig. 2 Instant JChem table showing selected chemical terms (physicochemical parameters) used for PCA

(*see* **Note 7**). The "Monitor import" window will give a summary of the imported data (*see* **Note 8**). Once fully processed, click on "Finish".

5. Use Instant JChem to determine values for physicochemical properties by selecting Data > New Chemical Term Field. Use the "Expression" drop-down menu to choose preset chemical terms (*see* **Notes 9** and **10**). Enter an appropriate Name for the column and then click on "Finish" (Fig. 2). For this example, we selected the following 16 terms (Instant JChem input syntax follows each description) (*see* **Notes 11** and **12**):

Molecular weight (MW): mass()

N atom count (N): atomCount("7")

O atom count (O): atomCount("8")

H-bond donor count (HBD): donorCount()

H-bond acceptor (HBA): acceptorCount()

Rotatable bond count (RotB): rotatableBondCount()

Stereocenter count (nStereo): chiralCenterCount()

Topological polar surface area (tPSA): PSA()

Number of rings (Rings): ringCount()

Aromatic ring count (RngAr): aromaticRingCount()

Ring system count (RngSys): ringSystemCount()

Size of largest ring (RngLg): largestRingSize()

Fraction of sp^3-hybridized carbons (Fsp3): count(filter('atno()==6 && connections()==4'))/atomCount("6")

n-Octanol/water partition coefficient at pH = 7.4 (LogD): logd('7.4')

van der Waals surface area (VWSA): vanDerWaalsSurfaceArea()

Relative polar surface area (relPSA): PSA()/vanDerWaals SurfaceArea()

6. Export the table of physicochemical parameters calculated in Instant JChem to an MS Excel file (.xls) by selecting File > Export to File. In the "Specify details" window, click on the purple folder to the right of the "File" field. Name the file and define the file format as "Microsoft Office Excel Workbook (*xls)". The following window gives the user an option to remove or rearrange columns in the exported file (*see* **Note 13**). The "Monitor progress" window summarizes the export process. When complete, click on "Finish". Open the .xls file containing these physicochemical parameters in MS Excel; additional physicochemical parameters will be added later (Subheading 3.1, **steps 8** and **9**).
7. The following two physicochemical values are calculated using the VCC Lab Website (http://www.vcclab.org/lab/alogps/start.html) (*see* **Notes 12** and **14**):

   n-Octanol/water partition coefficient alt (ALOGPs)

   Tetko's logS aqueous solubility (ALOGpS)

   To calculate ALOGPs and ALOGpS from the website's window, choose "Upload file" and select "Smiles—SMILES file—default" from the drop-down menu and click on "Proceed with file uploading" (Fig. 3). Click on "Choose file" and select the .txt file consisting of SMILES codes only that was created in Subheading 3.1, **step 2**. Click on "Upload file" and a new pop-up window should be displayed that states "Your file "yourfile.txt" was uploaded successfully." Close the pop-up window displaying this message and a new window will open that is entitled "results.txt". Copy the text from this results window and paste it into a new MS Excel file.
8. In the MS Excel sheet/tab that contains the physicochemical parameters calculated by Instant JChem (Subheading 3.1, **step 6**), add two new columns labeled "ALOGPs" and "ALOGpS." Copy the "logP" and "logS" data columns calculated by the VCC Lab website, and paste them into the "ALOGPs" and "ALOGpS" columns, respectively. Close the MS Excel file that contains only the VCC Lab data.

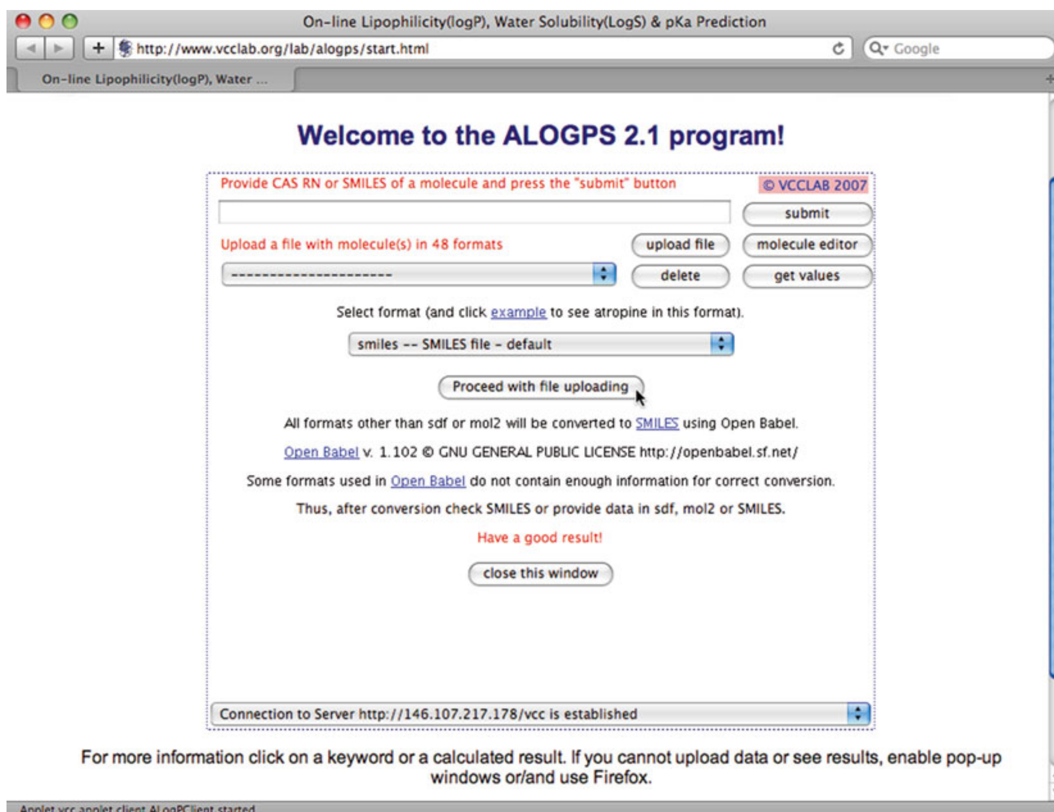


Fig. 3 Uploading SMILES codes to calculate ALOGpS and ALOGPs at the VCC Lab Website

9. The remaining two physicochemical parameters are calculated in MS Excel (*see* **Note 12**):

   nStereo ÷ *MW*, stereochemical density (nStMW)

   Rings ÷ *RingSys*, ring complexity (RRSys)

   Create two new columns in the MS Excel file that contains the other physicochemical values. For the nStMW and RRSys columns, set the column's formula according to its respective equation (Fig. 4):

   nStMW = [column for *nStereo*]/[column for *MW*]

   RRSys = IF([column for *RingsSys*] = 0,0,[column for *Rings*]/[column for *RingSys*]) (*see* **Note 15**).

   All of the physicochemical parameters needed for this PCA are now in the MS Excel file.

3.2 Principal Component Analysis

1. In the MS Excel file containing the compound names and physicochemical parameters that was created in Subheading 3.1, verify that the compounds are grouped by compound class (such as Drugs, Natural Products), and insert a new row below

PCA physicochemical parameters.xls

	C	M	N	O	P	Q	R	S	T
1	Field 1	Rings	RngAr	RngSys	RngLg	ALOGPs	ALOGpS	nStMW	RRSys
2	Abilify	4	2	3	6	5.21	-4.76	0.0000000	=IF(O2=0,0,M2/O2)
3	Aciphex	3	3	2	6	2.04	-3.03	0.0027821	1.50
4	Actos	3	2	3	6	3.17	-4.91	0.0028055	1.00
5	Adderall	1	1	1	6	1.85	-1.89	0.0073961	1.00
6	Ambien	3	3	2	6	3.15	-3.99	0.0000000	1.50
7	Avandia	3	2	3	6	2.95	-3.97	0.0027978	1.00
8	Benazepril	3	2	2	7	1.14	-4.61	0.0047115	1.50
9	Celebrex	3	3	3	6	3.99	-4.88	0.0000000	1.00
10	Concerta	2	1	2	6	1.47	-3.11	0.0085724	1.00
11	Coreg	4	4	2	6	3.05	-4.96	0.0024602	2.00
12	Crestor	2	2	2	6	1.47	-3.74	0.0041534	1.00
13	Cymbalta	3	3	2	6	4.72	-5	0.0033623	1.50
14	Diovan	3	3	3	6	3.68	-4.27	0.0022961	1.00
15	Effexor	2	1	2	6	2.69	-3.08	0.0036049	1.00
16	Flonase	4	0	1	6	3.69	-4.64	0.0179795	4.00
17	Fosamax	0	0	0	0	-1.34	-1.17	0.0000000	0.00
18	Imitrex	2	2	1	6	1.17	-3.37	0.0000000	2.00
19	Lamictal	2	2	2	6	1.87	-2.72	0.0000000	1.00
20	Levaquin	4	2	2	6	-0.02	-2.4	0.0027673	2.00

PCA Physicochemical parameters

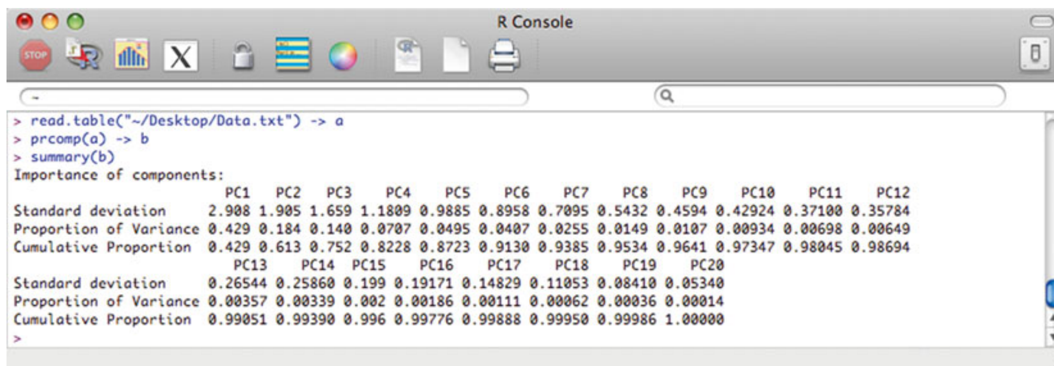
Fig. 4 Final physicochemical parameters calculated in MS Excel, with the RRsys calculation in cell T2 and the referenced cells M2 and O2 highlighted as an example

each group. Each new row will represent the average compound for a given class. Accordingly, name each new row based on the category it will represent, for example "AVG Drug". For these new rows, fill the cells associated with structural and physicochemical parameter values using the "AVERAGE" function to the left of the cell formula field and select the appropriate cells. Similarly, add two new rows below the last compound and calculate the mean ("AVERAGE") of each physicochemical parameter for the entire dataset (all compounds), as well as the standard deviation of each parameter using the "STDEV" function. Name the MS Excel sheet/tab "Raw".

2. Create a new sheet/tab within the same MS Excel file, naming it "Norm" to contain mean normalized values of the physicochemical parameters. Copy the compound names from the row header and physicochemical descriptors from the column header of the "Raw" sheet/tab and paste them into this "Norm" sheet/tab. Fill each standardized physicochemical parameter cell with the Normval value calculated using the equation

   Normval = ([val] − [column mean])/[column standard deviation]

   where "val" is the value from the corresponding cell in the "Raw" sheet/tab. The number format (Format > Cells…) should be set to four decimal places. Save the MS Excel file. Now, save only the "Norm" sheet/tab as "Data.txt" (Text-Tab Delimited) on the Desktop (Mac). Close the MS Excel file and discard the changes to the file format.



```
> read.table("~/Desktop/Data.txt") -> a
> prcomp(a) -> b
> summary(b)
Importance of components:
                         PC1   PC2   PC3    PC4    PC5    PC6    PC7    PC8    PC9    PC10    PC11    PC12
Standard deviation     2.908 1.905 1.659 1.1809 0.9885 0.8958 0.7095 0.5432 0.4594 0.42924 0.37100 0.35784
Proportion of Variance 0.429 0.184 0.140 0.0707 0.0495 0.0407 0.0255 0.0149 0.0107 0.00934 0.00698 0.00649
Cumulative Proportion  0.429 0.613 0.752 0.8228 0.8723 0.9130 0.9385 0.9534 0.9641 0.97347 0.98045 0.98694
                          PC13    PC14  PC15    PC16    PC17    PC18    PC19    PC20
Standard deviation     0.26544 0.25860 0.199 0.19171 0.14829 0.11053 0.08410 0.05340
Proportion of Variance 0.00357 0.00339 0.002 0.00186 0.00111 0.00062 0.00036 0.00014
Cumulative Proportion  0.99051 0.99390 0.996 0.99776 0.99888 0.99950 0.99986 1.00000
>
```

Fig. 5 "R" output showing the variance retained in each successive principal component of the PCA

3. Launch the "R" open-source computing package and run the following PCA commands (at the command prompt "R>") (*see* **Note 16**):

   R > read.table("~/Desktop/Data.txt") - > a

   R > prcomp(a) - > b

   R > summary(b)

   This command gives a table showing the distribution of variance from the full dataset on each principal component (Fig. 5). PCA generates as many principal components as there are parameters, but importantly, the majority of variance is represented in the first few components (*see* refs. 1, 2 for further discussion on PCA and variance). In this example, the first three principal components (PC1–PC3) retain 75 % of the variance from the 20-dimensional dataset. As such, PC1–PC3 can be used to construct a set of two-dimensional plots that will allow the visualization of the data in a more intuitive manner while still retaining the majority of the information from the full dataset. We will therefore focus our remaining analysis on PC1–PC3.

4. To obtain loading information for each structural and physicochemical parameter, continue PCA with the following command in "R":

   R > b

   The resulting table can be copied into MS Excel for convenient access (*see* **Note 17**) (Fig. 6). The loading data will become useful for directing future library design and will be discussed in Subheading 3.3.

	PC1	PC2	PC3
MW	-0.3207	-0.1482	0.1008
N	-0.2080	-0.1055	-0.3331
O	-0.3180	0.0179	0.1159
HBD	-0.3087	0.0677	-0.1180
HBA	-0.3346	-0.0057	-0.0256
RotB	-0.2587	-0.1209	-0.0540
tPSA	-0.3336	0.0109	-0.0818
nStereo	-0.2789	0.0936	0.2882
nStMW	-0.1612	0.2880	0.2964
Rings	-0.1669	-0.2384	0.0698
RngAr	-0.0475	-0.3660	-0.3227
RngSys	-0.1761	-0.2315	-0.1506
RngLg	-0.1459	-0.0128	0.2521
RRSys	-0.0015	-0.0362	0.1920
ALOGPs	0.0859	-0.4127	0.2677
ALOGpS	0.0378	0.4319	-0.2060
Fsp3	-0.0910	0.3464	0.3224
LogD	0.1850	-0.2569	0.3114
relPSA	-0.1797	0.2262	-0.3112
VWSA	-0.3122	-0.1253	0.1626

Fig. 6 Coefficients of each parameter used for PC1–PC3, highlighting those of the greatest magnitude

5. Obtain loading plots in "R" using the following command:

   R> biplot(b, choices = c(1, 2), col = c("gray", "red"))

   The plot produced illustrates loading data of structural and physicochemical parameters for PC1 vs. PC2 (*see* **Notes 18** and **19**) (Fig. 7). Save the loading plot and name it appropriately (*see* **Note 20**).

6. Generate loading plots of PC1 vs. PC3 and PC3 vs. PC2, using the following commands (and saving each loading plot):

   R> biplot(b, choices = c(1, 3), col = c("gray", "red"))

   R> biplot(b, choices = c(3, 2), col = c("gray", "red"))

   These loading plots give insight into the structural and physicochemical parameters that most distinguish a set of compounds by visually displaying each physicochemical parameter's influence on where a compound will appear in the final PCA plots. This information is also valuable to the planning of future libraries and is discussed in Subheading 3.3.

7. In "R", use the following commands to obtain the rotated PCA data (scores) and save the output as a text file (*see* **Note 21**):

   R> b$x -> c
   R> write.table(c, "~/Desktop/scores.txt", sep = "\t")

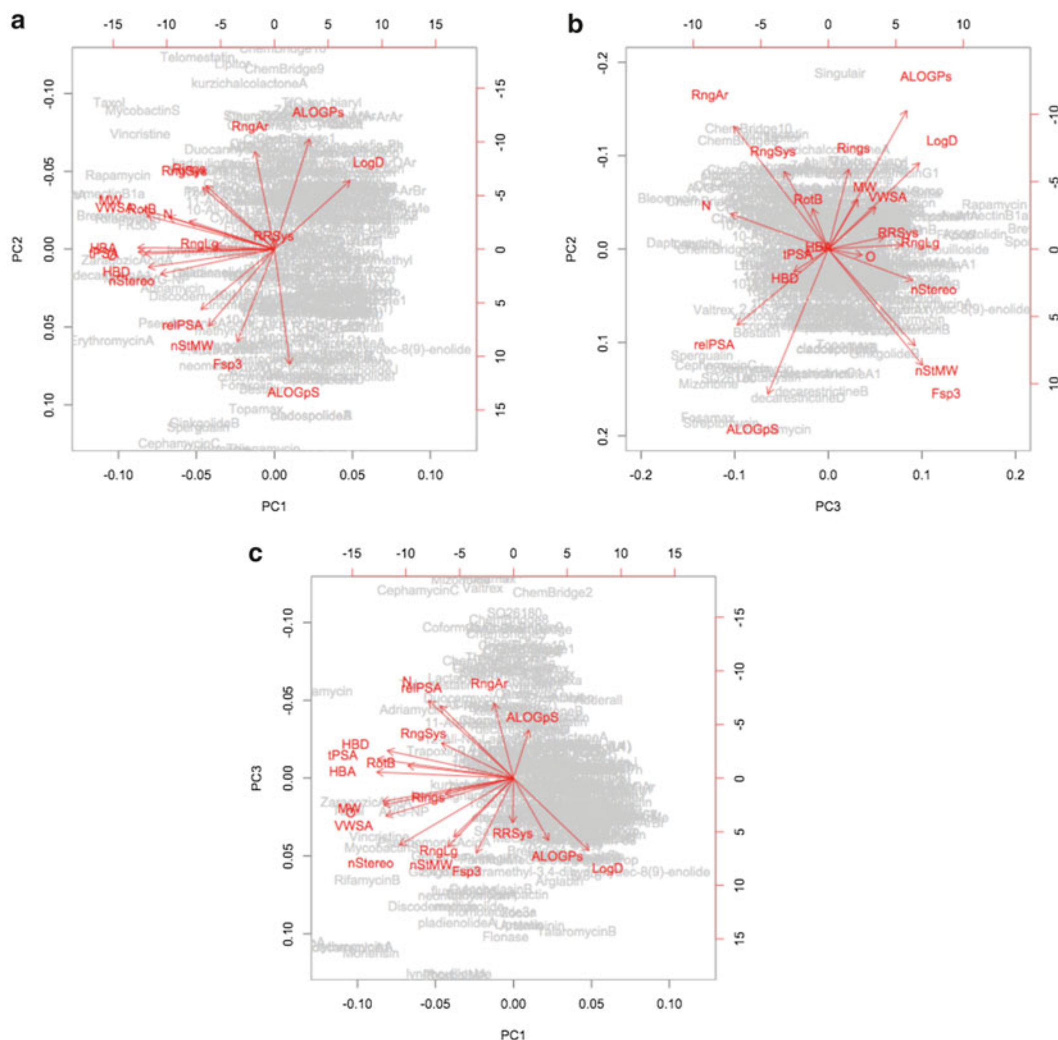
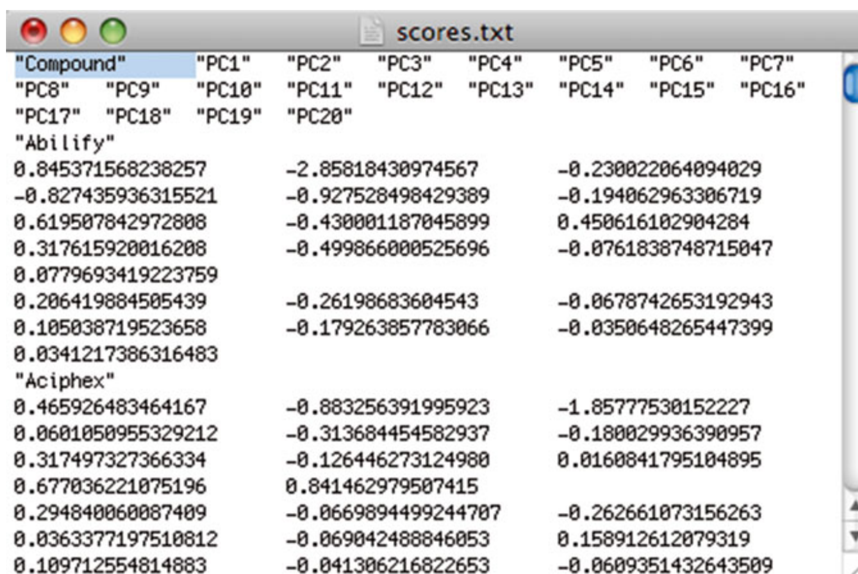


Fig. 7 Loading plots from PCA: (**a**) PC1 vs. PC2, (**b**) PC2 vs. PC3, (**c**) PC1 vs. PC 3

8. Open the MS Excel file that contains the "Raw" and "Norm" sheets/tabs (Subheading 3.2, **step 2**), and create a new sheet/tab within that MS Excel file, naming it "PCA". To transfer the scores data obtained from "R" to MS Excel, first open the scores.txt file in a text editor. At the beginning of the document, add a column header such as "Compound" followed by a tab (Fig. 8). Next, copy all of the text in the file and paste it into the MS Excel sheet/tab named "PCA". Change the number format (Format > Cells…) of the PCA cells to three decimal places.
9. In MS Excel, plot PC1 vs. PC2 from the "PCA" sheet/tab by selecting the columns and clicking on the "Chart Wizard" icon in the Standard Toolbar (View > Toolbars > Standard).



```
scores.txt
"Compound"  "PC1"  "PC2"  "PC3"  "PC4"  "PC5"  "PC6"  "PC7"
"PC8"  "PC9"  "PC10"  "PC11"  "PC12"  "PC13"  "PC14"  "PC15"  "PC16"
"PC17"  "PC18"  "PC19"  "PC20"
"Abilify"
0.845371568238257    -2.85818430974567     -0.230022064094029
-0.827435936315521   -0.927528498429389    -0.194062963306719
0.619507842972808    -0.430001187045899    0.450616102904284
0.317615920016208    -0.499866000525696    -0.0761838748715047
0.0779693419223759
0.206419884505439    -0.26198683604543     -0.0678742653192943
0.105038719523658    -0.179263857783066    -0.0350648265447399
0.0341217386316483
"Aciphex"
0.465926483464167    -0.883256391995923    -1.85777530152227
0.0601050955329212   -0.313684454582937    -0.180029936390957
0.317497327366334    -0.126446273124980    0.0160841795104895
0.677036221075196    0.841462979507415
0.294840060087409    -0.0669894499244707   -0.262661073156263
0.0363377197510812   -0.069042488846053    0.158912612079319
0.109712554814883    -0.041306216822653    -0.0609351432643509
```

Fig. 8 Modified PCA scores.txt file including a header for compound names

Under "Standard Types" choose "Scatter XY" and click on "Next". Enter series information (e.g., Drugs, Natural Products) under the "Series" tab and fill the *X*- and *Y*-values data fields with corresponding range for each series (for example PC1 data for *X*-values and PC2 data for *Y*-values). When done entering the series information, click on "Next". Enter a title for the plot and labels for the axes and then click on "Next." Select "As object in PCA" and click on "Finish" (*see* **Note 22**).

10. Follow a procedure similar to that described in Subheading 3.2, **step 9**, to generate plots for PC1 vs. PC3 and PC3 vs. PC2 using the data from the appropriate columns.

11. Copy the plots produced in MS Excel and use MS PowerPoint or graphical editing software to add colored ovals that encompass the majority of data points for a given class of compounds (Fig. 9).

3.3 Using PCA to Guide Library Design

1. Examine the PCA plots together with the loading plots to identify structural and physicochemical parameters that determine where a particular compound or a collection of compounds appears on the PCA plot. For example, many natural products appear to the left (negative PC1) of the library members in the PC1 vs. PC2 plot (Fig. 9). The corresponding loading plot (Fig. 7) indicates that HBA, tPSA, and O are all

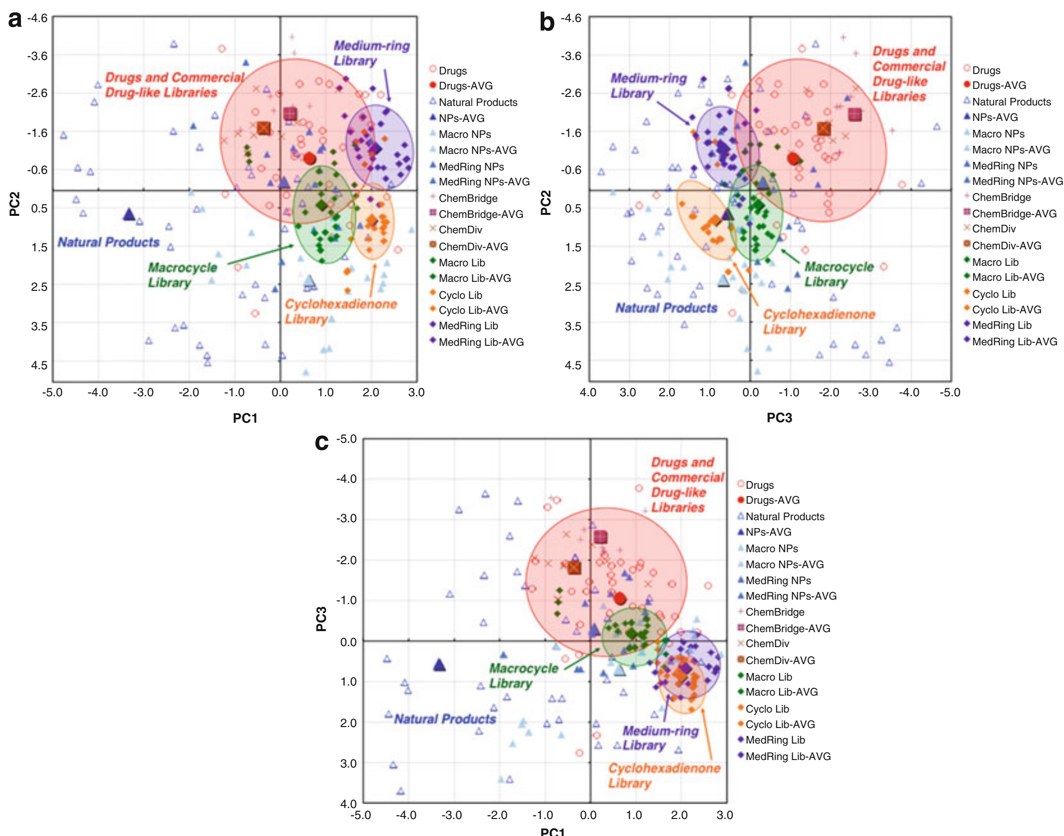


Fig. 9 Completed PCA plots: (**a**) PC1 vs. PC2, (**b**) PC2 vs. PC3, (**c**) PC1 vs. PC 3

major components of PC1. Recalling that each PC axis is a linear combination of structural and physicochemical parameters, note that the coefficients for each parameter used for a given PC axis were also obtained in Subheading 3.2, **step 4** (Fig. 6). This table provides more quantitative information regarding the parameters that have the greatest impact on the location of a compound with respect to each PC axis.

2. Consider the structural and physicochemical parameters that are the most important in differentiating a collection of compounds from the targeted region of chemical space. In this example, O, HBD, HBA, tPSA, nStereo, aqueous solubility, and Fsp^3 are among the most influential in distinguishing our library compounds from natural products. The introduction of additional oxygen atoms and additional stereocenters to our library compounds would likely lead to more natural product-like

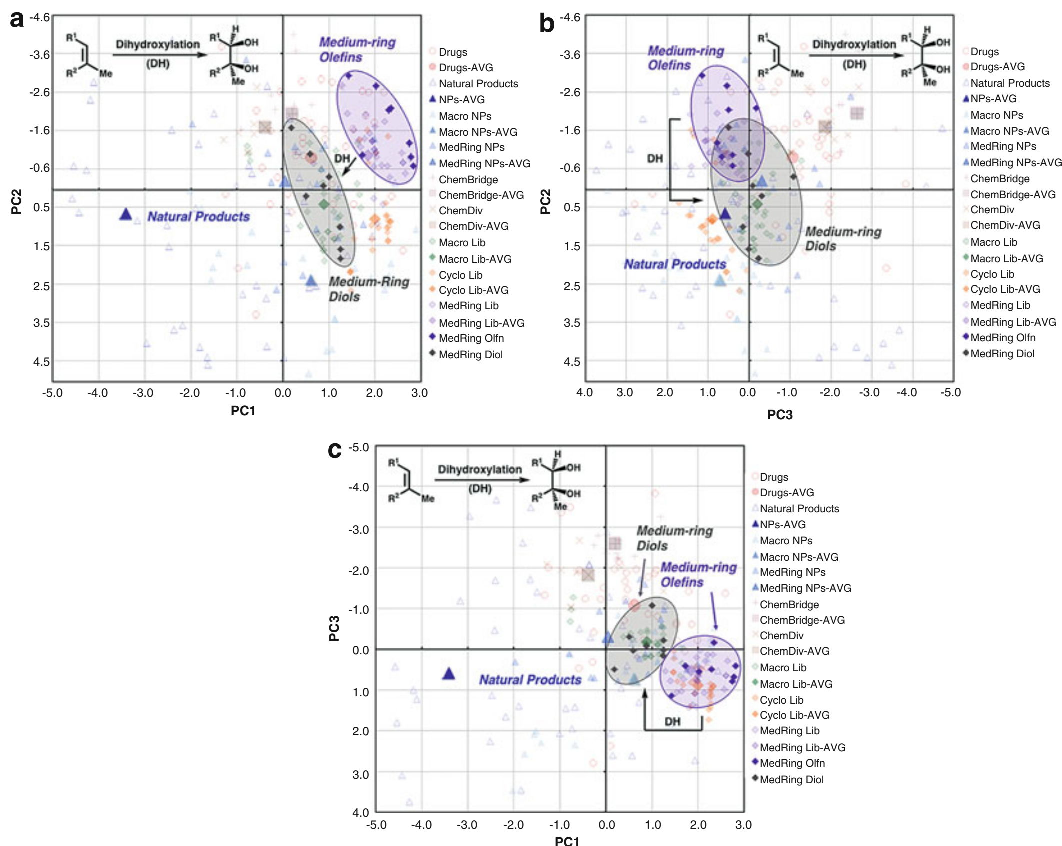


Fig. 10 Completed PCA plots: (**a**) PC1 vs. PC2, (**b**) PC2 vs. PC3, (**c**) PC1 vs. PC 3. DH = dihydroxylation

compounds. A dihydroxylation of the olefins contained in our medium-ring library members would address all of the parameters mentioned above and should result in compounds that are shifted towards natural products in our PCA plots. Leveraging this information, we proceed with the dihydroxylation of multiple medium-ring compounds to produce a collection of PCA-directed derivatives of our initial medium-ring library.

3. Include the collection of modified compounds in a new analysis to evaluate the effectiveness of the PCA-directed modifications in targeting the desired region of the plot (Fig. 10). In this example, our diol products have structural and physicochemical parameters that are more consistent with natural products compared to their parent olefins, and as a result, the compounds are shifted towards natural products in all of the PCA plots. Reiterate this process as necessary to provide a library with the desired structural and physicochemical properties.

4 Notes

1. In this analysis, we used 40 top-selling drugs from [20], 60 diverse natural products, 20 drug-like library members, 23 macrocycle natural products, 32 synthetic macrocycles, 20 medium-ring natural products, 38 synthetic medium rings, and 25 cyclohexadienone precursors to those medium rings. In the analysis described in Subheading 3.3, an additional eight synthetic medium-ring diols were included.
2. Much of the raw data used in this analysis is available from the Supplementary Information for refs. [12, 13].
3. To obtain the SMILES codes in ChemBioDraw, select the chemical structures and choose Edit > Copy As > SMILES. Paste the SMILES codes into an MS Word document. The compounds in the string are separated by a period ("."), and can be converted to a table format in MS Excel by saving the MS Word document as a text file (.txt) and importing the data in MS Excel using Data > Get External Data. Select the text file, choose the "Delimited" option in Step 1 of the Text Import Wizard, and in Step 2 of the Wizard specify the delimiters as a period (".") in the "Other" field. The imported SMILES codes can be transposed (flipped from row to column format) by copying them, then selecting Edit > Paste Special, and clicking on the "Transpose" option.
4. Some software does not handle spaces and punctuation in compound names, but underscores can be used instead. For Windows 7 users, spaces are allowed.
5. Make sure that "IJC Project (with local database)" is highlighted in the "Projects" panel.
6. Make sure that "localdb [as admin]" is highlighted in the "Projects and schemas" panel.
7. Remove any undesired fields using the "< Remove" button. Click on "Next" when finished.
8. Any error messages are saved to a file that can be opened with a text editor for review.
9. Each chemical term must be added individually.
10. User-defined chemical terms can be saved to the "Expression" menu by clicking on the yellow star to the right of the menu and naming the new term for convenient use in the future.
11. The order of the columns can be changed by clicking and dragging on the header row of a column.
12. We selected the 20 physicochemical parameters used in this example based on several criteria. First, Lipinski parameters ($MW \leq 500$, $\log P \leq 5$, $HBA \leq 10$, $HBD \leq 5$) [14], and Veber

parameters (RotB ≤ 10, tPSA ≤ 140 Å^2) [10], have been correlated with oral bioavailability. These parameters partially correlate to cell permeability [11], which is relevant to the utility of new chemical probes discovered from library screening. Second, Tetko's calculated logS solubility (ALOGpS) [15] was included because compound solubility is critical in screening and is often problematic for commercial drug-like libraries. ALOGPs was included as an alternative method for estimating solubility, and logD was included to approximate aqueous solubility at a physiological relevant pH (7.4). Third, several stereochemical parameters (nStereo, nStMW, Fsp^3) were included as approximations of three-dimensional complexity, which have been shown to be a distinguishing factor between synthetic drugs and natural products [16, 17], and also impact protein binding selectivity and frequency [18]. Fourth, relative polar surface area (relPSA) was included because it has been shown to be a distinguishing factor between Gram-positive and Gram-negative antibiotics [19]. Finally, additional physicochemical parameters (N, O, Rings, RngAr, RngSys, RngLg, RRSys) were included because they have been found to differentiate synthetic drugs from natural products [16].

13. The Structure column may be removed for faster processing, consistency of row height, and a cleaner appearance in MS Excel.
14. The browser's pop-up blocker must be turned off.
15. This formula avoids errors caused by a zero in the denominator.
16. For Windows users, the read.table command is slightly different:

    R > read.table("*file path*\\Data.txt", header = T, sep="\t", row.names = 1) - > a

    where *file path* is the entire file path beginning with the drive (usually C:\\). Users can obtain the file path by dragging and dropping the text file directly into R, which returns an error message but reports the file location including the drive.
17. To transfer the "R" output to MS Excel, copy the first section of the table without the column headers (for example the PC1–PC8 data before the first section break) and paste it into an MS Word document. Change the font to Courier and the size to 5 pt such that the text in the document resembles a table. Save the file as a Text-only (.txt) file. From MS Excel, import the data using Data > Get External Data. Select the file, then choose the "Fixed width" option in Step 1 of the Text Import Wizard, click on "Next", verify column divisions, and then click on "Finish".

18. Several PC axes were inverted in this example to maintain resemblance to our previous PCA plots [12, 13] by adding the "ylim" and "xlim" axis limit options to the "biplot" command in "R":

    R > biplot(b, choices = c(1, 2), col = c("gray", "red"), ylim = c(0.12, −0.12), xlim = c(−0.12, 0.12))

    This does not impact data interpretation because the signs of all PC axes are arbitrary.

19. If desired, loading plots can also be produced where the scores (compound names) are hidden. To do this, replace "gray" with "white" in the biplot command.
20. The plot must be saved before an additional plot command is entered.
21. For Windows users, the file path information in the write.table command will be different and needs to include the drive (such as C:\\) (cf. **Note 16**).
22. If desired, change the appearance of the plot by right-clicking on the object you wish to modify, and then select the Format option.

Acknowledgments

We thank Tony D. Davis (MSKCC) for suggesting inclusion of the logD, van der Waals surface area, and relative polar surface area parameters, and for providing modifications of this protocol for Windows users. Instant JChem was generously provided by ChemAxon. Financial support from the NIH (P41 GM076267 to D.S.T., P41 GM076267-03S1 to R.A.B., T32 CA062948-Gudas to T.A.W.), Starr Foundation, Tri-Institutional Stem Cell Initiative, Alfred P. Sloan Foundation (Research Fellowship to D.S.T.), Deutscher Akademischer Austauschdienst (DAAD, postdoctoral fellowship to F.K.), William H. Goodwin and Alice Goodwin and the Commonwealth Foundation for Cancer Research, and the MSKCC Experimental Therapeutics Center is gratefully acknowledged.

References

1. Jolliffe IT (2002) Principal component analysis. Springer, New York, NY
2. Jackson JE (2003) A user's guide to principal components. Wiley, Hoboken, NJ
3. Akella LB, DeCaprio D (2010) Cheminformatics approaches to analyze diversity in compound screening libraries. Curr Opin Chem Biol 14:325–330
4. Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? Nat Rev Drug Discov 5:993–996
5. Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod 75:311–335
6. Sánchez-Pedregal VM et al (2006) The tubulin-bound conformation of discodermolide derived by NMR studies in solution supports a common pharmacophore model for epothilone and discodermolide. Angew Chem Int Ed 45:7388–7394
7. Canales A et al (2008) The bound conformation of microtubule-stabilizing agents: NMR

insights into the bioactive 3D structure of discodermolide and dictyostatin. Chem Eur J 14:7557–7569

8. Knust J, Hoffmann RW (2003) Synthesis and conformational analysis of macrocyclic dilactones mimicking the pharmacophore of aplysiatoxin. Helv Chim Acta 86:1871–1893
9. Khan AR et al (1998) Lowering the entropic barrier for binding conformationally flexible inhibitors to enzymes. Biochemistry 37: 16839–16845
10. Veber DF et al (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45:2615–2623
11. Rezai T et al (2007) Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers. J Am Chem Soc 128:2510–2511
12. Kopp F et al (2012) A diversity-oriented synthesis approach to macrocycles via oxidative ring expansion. Nat Chem Biol 8:358–365
13. Bauer RA, Wenderski TA, Tan DS (2013) Biomimetic diversity-oriented synthesis of benzannulated medium rings via ring expansion. Nat Chem Biol 9:21–29
14. Lipinski CA et al (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3–25
15. Tetko IV et al (2001) Estimation of aqueous solubility of chemical compounds using E-state indices. J Chem Inf Comput Sci 41: 1488–1493
16. Feher M, Schmidt JM (2003) Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. J Chem Inf Comput Sci 43: 218–227
17. Lovering F, Bikker J, Humblet C (2009) Escaping from flatland: increasing saturations as an approach to improving clinical success. J Med Chem 52:6752–6756
18. Clemons PA et al (2010) Small molecules of different origins have distinct distributions of structural complexity that correlate with protein-binding profiles. Proc Natl Acad Sci U S A 107:18787–18792
19. O'Shea R, Moser HM (2008) Physicochemical properties of antibacterial compounds: implications for drug discovery. J Med Chem 51: 2871–2878
20. McGrath NA, Brichacek M, Njardarson JT (2010) A graphical journey of innovative organic architectures that have improved our lives. J Chem Educ 87:1348–1349, Also see also: Njardarson Group—Top top-selling drugs Pharmaceuticals poster; http://cbc.arizona.edu/njardarson/group/top-pharmaceuticals-poster

Chapter 19

Small-Molecule Library Screening by Docking with PyRx

Sargis Dallakyan and Arthur J. Olson

Abstract

Virtual molecular screening is used to dock small-molecule libraries to a macromolecule in order to find lead compounds with desired biological function. This in silico method is well known for its application in computer-aided drug design. This chapter describes how to perform small-molecule virtual screening by docking with PyRx, which is open-source software with an intuitive user interface that runs on all major operating systems (Linux, Windows, and Mac OS). Specific steps for using PyRx, as well as considerations for data preparation, docking, and data analysis, are also described.

Key words Virtual molecular screening, Computer-aided drug design, Molecular docking, PubChem, AutoDock, Vina, Open Babel

1 Introduction

Drug discovery is an attractive research area that enables application of cutting-edge biomedical research to improve health of many people [1]. In the past, medicines were derived from natural products, mostly from plant sources. While natural products continue to be used and researched for medicine, it is now possible to synthesize a large number of chemical compounds that are not readily available in nature. The increased number of possible chemical compounds presents both a challenge and opportunity for the pharmaceutical industry. Testing different drug candidates in human clinical trials is a long and expensive process, which is why phenotypic or target-based screening is so important in the earlier stages of drug discovery [2].

In phenotypic screening, different compounds are tested in cells or organisms to see which compound makes intended changes in the phenotype. When molecular causes of the disease are unknown, phenotypic screening is, in many cases, the only available option for finding life-saving drugs. For diseases that are well studied and understood at the molecular level, altering a single macromolecule can lead to a desired outcome. An example of such

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_19, © Springer Science+Business Media New York 2015

a macromolecular target for the common flu virus is discussed shortly. In target-based screening, compounds are tested with purified macromolecules (usually a protein) to find lead compounds that make intended macromolecular changes. For a lead compound to become a drug, it needs to be able to reach a site of action in the body, bind to its target macromolecule, and elicit the desired biological effect.

Compared to large biological molecule therapeutics, such as insulin or antibodies, which are administered through injection, small molecules can be taken orally and are better at reaching different sites in the body. This is why the majority of approved and experimental drugs are small molecules. Small molecules are also better suited for virtual molecular screening, which is the main subject of this chapter. With virtual screening, different compounds are docked from a small-molecule library to a target macromolecule (usually a protein) to find compounds with the best binding affinity [3]. Note that virtual screening is not limited to drug targets and it can be used to screen against herbicides, pesticides, or any other target of interest [4]. In all cases, finding the right target is very important for virtual screening campaign to succeed. When the three-dimensional (3D) structure of a target is available, through X-ray crystallography, NMR spectroscopy, or any other means, docking algorithms can be applied to search for the best binding mode between target macromolecule and ligand.

In this chapter, methods for performing virtual screening experiments with PyRx open-source software are outlined. The 3D structure of the influenza virus neuraminidase [5] is used as an example to show how to prepare an input file for the target macromolecule. Influenza virus neuraminidase cleaves sialic acid from the infected cell surface to release newly created viruses. Neuraminidase inhibitors bind to neuraminidase and prevent them from binding to sialic acid. This leaves the influenza virus stuck on the surface of infected cells, so that the influenza virus cannot infect nearby healthy cells [6]. Here, steps to prepare input structures for zanamivir (a neuraminidase inhibitor), sialic acid, and sucrose (table sugar) are described. These small molecules are then used to run virtual screening against influenza virus neuraminidase.

2 Materials

PyRx is written in Python programming language and it can run on nearly any modern computer, from PC (personal computer) to supercomputer. Below details are provided of the Windows PC used in Subheading 3, although similar methods also work on Linux and Mac OS as well.

2.1 Hardware and Software

1. Dell Studio 540S with Intel Core 2 Duo CPU at 2.53 GHz, 4 GB memory (RAM), ATI Radeon HD 3400 series graphics card, and 32-bit Windows Vista operating system.
2. Binary distribution of PyRx version 0.8 for Windows available free from http://pyrx.sourceforge.net.

2.2 Input Files

To start with structure-based virtual screening, structures of the target macromolecule and small molecules are needed as input files. There are a number of publicly available websites where users can download these input files. Used in this chapter are DrugBank [7] to get the structure of zanamivir, PubChem [8] for 3D structure of sucrose, and Protein Data Bank [9] to get 3D structures of influenza virus neuraminidase and sialic acid.

1. Open a preferred web browser and visit http://www.drugbank.ca/drugs/DB00558, click on SDF link next to Download, and save that page as DB00558.sdf.
2. Go to http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=5988, click SDF icon on top right, and select 3D SDF: Save.
3. Visit http://www.pdb.org/pdb/explore/explore.do?structureId=2BAT, click on Download Files, and select PDB File (Text).

The reason for choosing these particular molecules is that they are familiar to most of the readers and computations can be run relatively quickly on a PC. To apply the protocol described in Subheading 3 to other binding target and ligands, users would need to obtain input files corresponding to their specific binding target and ligands. Selection of the binding target depends on the biological problem of interest, and it is assumed that the 3D structure of the target is available in PDB format through Protein Data Bank (http://pdb.org) or other sources (*see* **Note 1**). Selection of ligands depends on whether virtual screening is used for lead discovery or lead optimization. For lead discovery, it is advised to include as many ligands with diverse shapes, sizes, and composition as possible. Since individual docking computations are independent from each other, users are practically only limited by computational power available at their disposal. For lead optimization, on the other hand, ligands are selected to closely match the lead compound [10]. One of the advantages of virtual screening is that it is not limited to commercially available compounds; a ligand file for a novel compound not found in any of the databases can also be used (*see* **Note 2**).

3 Methods

3.1 Prepare Input Files for Docking

Before input files can be used for virtual screening, they must be converted to the PDBQT file format suitable for docking with AutoDock Vina [11].

1. Start by double-clicking on PyRx icon on the Desktop.
2. Select Open Babel tab under Controls panel and click on the first icon on its toolbar with plus (+) sign on it. Navigate to the Downloads folder and select CID_5988.sdf (sucrose from PubChem).
3. Click on the first icon on the Open Babel toolbar again, and locate and open DB00558.sdf (zanamivir from DrugBank). 558 is the accession number of zanamivir in DrugBank and it is listed under the Title column in the Open Babel table. If other molecules are to be included in virtual screening, the Open Babel widget can be used to convert them to PDBQT file format (*see* **Note 3**).
4. Select the row corresponding to zanamivir with Title 558, and right-click and use the Minimize Selected option. Click OK and wait for energy minimization to complete. Notice that the title of this molecule has changed to 558_uff_E = 197.68. The _uff part corresponds to the force field used for energy minimization, which, by default, is the Universal Force Field [12] as implemented in Open Babel software package [13]. The _E = 197.68 part corresponds to the energy of the minimized molecule. The precise value for this energy is not important here. However, this notation is helpful to capture changes made to this molecule before conversion to the AutoDock ligand file in the next step.
5. Right-click on any of the rows in Open Babel table and use Convert All to AutoDock Ligand (pdbqt). This will create two pdbqt files corresponding to sucrose and zanamivir molecules under the Ligands folder.
6. Select Documents tab under the View panel, click on the Open icon (second from the left), and open the 2BAT.pdb file. 2BAT is the PDB ID for the structure of the complex between influenza virus neuraminidase and sialic acid [5]. The following steps are specific to this structure. To apply this method to targets which have no ligand attached, please go directly to **step 10** and replace 2BAT with the name of the desired target macromolecule.
7. Next select lines corresponding to sialic acid from 2BAT.pdb. Scroll down, use Ctrl-F or the Find icon on the toolbar to search for SIA residues, and select lines with HETATM 3216–3236. Use Ctrl-C or right-click Copy, click on the New icon, and paste these lines (Ctrl-V or right-click Paste) in a new file. Save this file as SialicAcid.pdb using Save icon (third from the left) on the Documents panel. If working with another target that contains a ligand that is desired for re-docking, the

Documents panel in PyRx can be used or any other text editor (such as Notepad or WordPad) to extract HETATM records corresponding to the ligand of interest. The web page for 2BAT (http://www.rcsb.org/pdb/explore.do?structureId=2BAT) also lists different ligands bound to neuraminidase, including sialic acid, which is listed under Ligand Identifier column as SIA. This web page also offers the possibility to download ligand SDF file for sialic acid.

8. Click on 2BAT.pdb tab under the Documents panel, scroll up, and left-click at the beginning of the line starting with TER 3023. The TER record indicates the end of a list of ATOM records for a chain according to PDB file format specification. In this case, it is desired to keep neuraminidase atom records only and delete all other records that correspond to different ligand and water molecules cocrystallized with this structure of neuraminidase. With the left mouse button pressed, scroll down to the end of the file and click Delete. Save this modified 2BAT.pdb file using the Save icon again.
9. From the menu bar, use File → Load Molecule menu and open SialicAcid.pdb. Right-click on SialicAcid under Molecules panel and select AutoDock → Make Ligand.
10. Use File → Load Molecule menu again and open 2BAT.pdb. Right-click on 2BAT under Molecules panel and select AutoDock → Make Macromolecule.

3.2 Run Virtual Screening Using Vina Wizard

1. Select Vina Wizard tab under the Controls panel and click on the Start button.
2. Select 558_uff_E = 197.68.pdbqt, 5988.pdbqt and SialicAcid.pdbqt under the Ligands folder (use the Shift key for selecting multiple ligands).
3. Select 2BAT under the Macromolecules folder and click on the Forward button on Vina Wizard.
4. Click on the Maximize button under Vina Search Space and then click on the Forward button. This starts AutoDock Vina and docks each ligand, one by one, to neuraminidase (2BAT). It takes less than 20 min to complete this virtual screening on a PC mentioned in Subheading 2.1 (*see* **Note 4**) (Fig. 1).
5. After virtual screening is completed, PyRx automatically advances to Analyze Results page, where results of virtual screening computation can be viewed. AutoDock Vina, by default, outputs the ten best binding modes for each docking run (*see* **Note 5**). Left-click on Binding Affinity (kcal/mol) table header cell under Analyze Results tab to sort this table by predicted binding affinity (*see* **Note 6**).

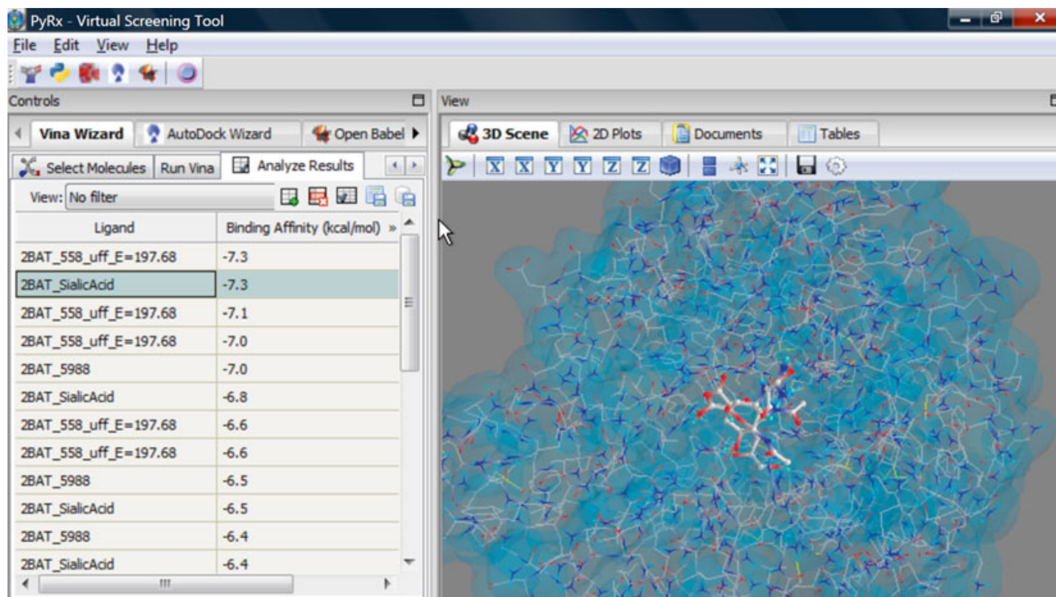


Fig. 1 A screenshot of the PyRx virtual screening tool. The table on the *left* lists predicted binding affinity of zanamivir (2BAT_558_uff_E = 197.68), sialic acid (2BAT_SalicAcid), and sucrose (2BAT_5988) for influenza virus neuraminidase (2BAT). The 3D scene on the *right* shows line drawing and transparent molecular surface of neuraminidase. Ball-and-stick models for zanamivir and sialic acid are also shown on this 3D scene

4 Notes

1. During docking runs, the 3D structure of the target is fixed while the ligand is moved and rotated to find the best binding modes. While it is possible to make some of the side chains flexible during the docking, incorporating full flexibility of the target is still a subject of active research [14].
2. There are a variety of desktop or Web-based molecular editors available that can be used to generate a ligand file for a novel compound not found in any of the databases. The Web-based molecular editors allow users to sketch molecules in 2D, while desktop tools such as Avogadro [15] can draw molecules in 3D.
3. SDF (Structure-Data File) format is commonly used to store multiple structures in a single file. It allows storing arbitrary data together with coordinates and atom types. Oftentimes, small molecules stored in SDF are flat (2D) and energy minimization is performed to get 3D structures with proper bond length between different atoms.
4. The main results from virtual screening runs are the best predicted binding modes and corresponding binding affinity. The negative values for binding affinity (or binding free energy) indicate that the ligand is predicted to bind to a target macromolecule. The more negative the numerical values for the

binding affinity, the better is the predicted binding between a ligand and a macromolecule. In this particular case of screening neuraminidase with zanamivir, sialic acid, and sucrose, Fig. 1 shows that zanamivir (2BAT_558_uff_E = 197.68) and sialic acid (2BAT_SalicAcid) are both predicted to have the best binding affinity of −7.3 kcal/mol, whereas the best binding mode for sucrose (2BAT_5988) is predicted to have binding affinity of −7.0 kcal/mol. In other words, zanamivir and sialic acid are predicted to have better binding affinity to neuraminidase than sucrose. The fact that both zanamivir and sialic acid have the same predicted binding affinity indicates that zanamivir can bind to neuraminidase and inhibit it from binding to sialic acid.

5. PyRx users can also export virtual screening results as CSV (Comma-Separated Values) or SDF files. This is useful for further analysis, filtering, or re-ranking of virtual screening results with third-party packages.

6. There are a number of approximations used to model protein-ligand interactions [16] and there are a number of unknowns when it comes to comparing virtual screening results with experiments [17], not the least of which is that a single protein is being docked with a single ligand. In practice, even with purified samples, it is hard to predict if proteins or small-molecule ligands would aggregate and whether idealistic prediction of binding affinity with single protein-ligand docking applies to diluted samples. Nevertheless, small-molecule virtual screening by docking is a very valuable in silico method that can rank small molecules according to their predicted binding affinity to a target macromolecule. The cost of running virtual screening experiments is minuscule compared to real screening experiments. Virtual screening is also a very good tool for hypothesis generation with which to test modified versions of existing compounds or custom compounds that are not commercially available. With advances in computer software and hardware, and with the increasing number of publicly available bioassay data, virtual screening will continue to remain a vibrant research field.

Acknowledgements

This work was supported by National Institute of Health (NIH) grant R01GM069832. The authors are very grateful to all those who contributed software used in this chapter. We also would like to thank Alex L. Perryman, Oleg Trott, Ruth Huey, and all members of our laboratories for helpful comments and discussions. This is manuscript # 25035 from The Scripps Research Institute.

References

1. Ng R (2008) Drugs: from discovery to approval, 2nd edn. Wiley-Blackwell, Hoboken, NJ
2. Swinney DC, Anthony J (2011) How were new medicines discovered? Nat Rev Drug Discov 10:507–519
3. Jacob RB, Andersen T, McDougal OM (2012) Accessible high-throughput virtual screening molecular docking software for students and educators. PLoS Comput Biol 8:e1002499
4. Lindell SD, Pattenden LC, Shannon J (2009) Combinatorial chemistry in the agrosciences. Bioorg Med Chem 17:4035–4046
5. Varghese JN et al (1992) The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. Proteins 14:327–332
6. McKimm-Breschkin JL (2013) Influenza neuraminidase inhibitors: antiviral action and mechanisms of resistance. Influenza Other Respir Viruses 7:25–36
7. Knox C et al (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic Acids Res 39:D1035–D1041
8. Li Q et al (2010) PubChem as a public resource for drug discovery. Drug Discov Today 15:1052–1057
9. Berman HM et al (2002) The protein data bank. Acta Crystallogr D Biol Crystallogr 58:899–907
10. Scior T et al (2012) Recognizing pitfalls in virtual screening: a critical review. J Chem Inf Model 52:867–881
11. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:455–461
12. Rappe AK et al (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. J Am Chem Soc 114:10024–10035
13. O'Boyle NM et al (2011) Open Babel: an open chemical toolbox. J Cheminform 3:33
14. Perryman AL et al (2010) A dynamic model of HIV integrase inhibition and drug resistance. J Mol Biol 397:600–615
15. Hanwell MD et al (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 4:17
16. Muchmore SW et al (2010) Cheminformatic tools for medicinal chemists. J Med Chem 53:4830–4841
17. Nicholls A (2008) What do we know and when do we know it? J Comput Aided Mol Des 22:239–255

Part IV

Small Molecule Target Identification

Chapter 20

Fluorous Photoaffinity Labeling to Probe Protein-Small Molecule Interactions

Weigang Huang and Qisheng Zhang

Abstract

Identifying cellular targets of bioactive small molecules is essential for their applications as chemical probes or drug candidates. Of equal importance is to determine their "off-target" interactions, which usually account for unwanted properties including toxicity. Among strategies to profile small molecule-interacting proteins, photoaffinity labeling has been widely used because of its distinct advantages such as sensitivity. When combined with mass spectrometry, this approach can provide additional structural and mechanistic information, such as drug-target stoichiometry and exact interacting amino acid residues. We have described a novel fluorous photoaffinity labeling approach, in which a fluorous tag is incorporated into the photoaffinity labeling reagent to enable the enrichment of the labeled species from complex mixtures for analysis. This new feature likely makes the fluorous photoaffinity labeling approach suitable to identify transient interactions, and low-abundant, low-affinity interacting proteins in a cellular environment.

Key words Fluorous tag, Photoaffinity labeling, Cross-linking, Interacting protein, Fluorous solid-phase extraction, Isotope-coded

1 Introduction

Photoaffinity labeling (PAL) was first described in 1962 [1], and has since been extensively used in conjunction with mass spectrometry (MS) to study ligand-receptor, such as oligonucleotide-protein, drug-protein, and protein-protein interactions [2, 3]. A typical PAL probe consists of three major elements: a recognition unit that directs the probe to the targeted receptor, a photoactive group that can form a covalent bond with the receptor upon light illumination, and a reporter group or tag that allows visualization and/or specific enrichment of the labeled species. Three types of photoactive groups, aryl azide, diazirine, and benzophenone, have been widely used. Fluorescent groups and stable isotopes are typically utilized for detection, while biotin is the most popular tag to selectively isolate the labeled species, but with potential problems. For example, it is difficult to fully recover the enriched products [4]

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_20, © Springer Science+Business Media New York 2015

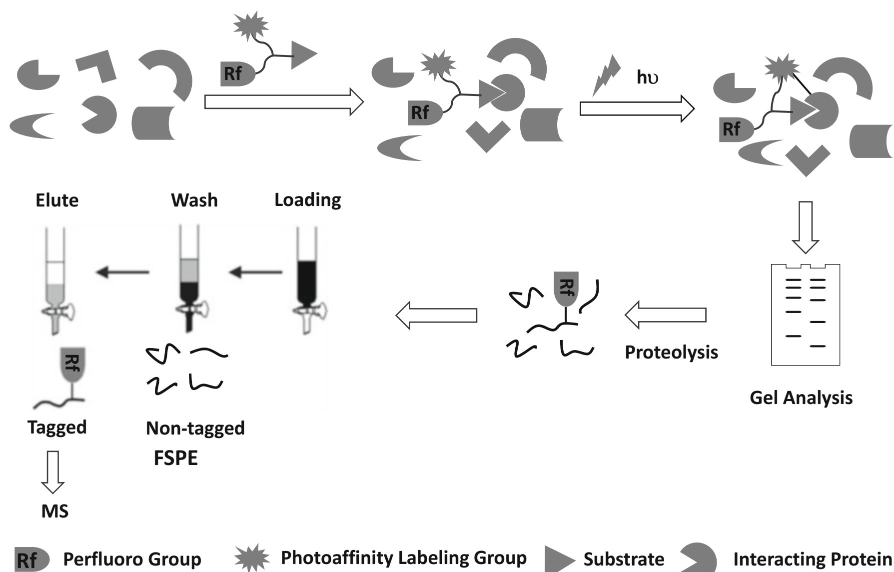


Fig. 1 Fluorous photoaffinity labeling approach to identify small molecule-interacting proteins

due to the strong binding between biotin and streptavidin, and the biotin tag often undergoes fragmentation in tandem mass spectrometry (MS/MS) [5] which complicates data analysis.

Molecules containing highly fluorinated (fluorous) units, particularly perfluoroalkyl groups, have unique features to be immiscible with either organic or aqueous phase while solubilize well in fluorous solvents. Teflon is one such example which has found a wide range of applications [6]. By incorporating a fluorous tag into small molecules, chemists have achieved selective separation of the tagged molecules from reaction or assay mixture through a simple fluorous solid-phase extraction (FSPE) [7]. This strategy has been employed to recycle catalyst [8, 9], remove reaction intermediates [10–12], and synthesize natural products using fluorous mixture synthesis [13–15]. Recently, fluorous tags have also been used in techniques such as fluorous microarray [16–18], fluorous enzymatic synthesis [19, 20], fluorous proteomics [21], and fluorous tag-facilitated nanostructure-initiator mass spectrometry (NIMS) [22], to address various biological problems. Compared to traditional biological reagents, the fluorous tags have attractive advantages. They are typically cost effective, orthogonal to biological systems, and stable in tandem mass (MS/MS) [21] analysis.

Inspired by these innovative applications of fluorous tags, we have developed a fluorous PAL technique to probe protein-small molecule interactions. In this approach (Fig. 1), a fluorous tag is incorporated into the PAL reagents to allow specific enrichment of labeled components from the complex assay mixture [23–25] for MS analysis. In a typical PAL experiment, as illustrated in Fig. 1,

the PAL probe is first incubated with its receptor to form a complex through non-covalent interaction. Subsequent light illumination activates the photoactive group to form a covalent bond with the receptor, which makes the analysis of the complex easier since harsher conditions and analytic methods, such as high-performance liquid chromatography (HPLC), can be used. The fluorous tag enables not only specific separation of labeled species but also efficient recovery and compatibility with MS analysis. We demonstrated this fluorous PAL approach with two examples. First, we incorporated the fluorous diazirine group into phosphatidylinositol 4,5-bisphosphate (PIP_2), which is a key phospholipid substrate and regulator for a number of important signaling proteins such as phosphoinositide 3-kinase (PI3K), phospholipase C (PLC), and ion channels. We showed that the fluorous PIP_2 efficiently cross-linked with its interacting protein ADP-ribosylation factor 1 (ARF1) [25]. Although we did not test the use of fluorous PIP_2 in live cells due to its poor cell permeability, the fluorous PAL approach can be used to identify cellular interacting proteins when the substrate is cell permeable. Secondly, we have developed a pair of isotope-coded fluorous diazirine reagents [24], with which we made a model peptide and demonstrated that the labeled peptides can be selectively isolated from complex protein digestion mixture via FSPE. The isotopic signature enabled a rapid differentiation of specific interactions from nonspecific binding or contaminants.

2 Materials

Prepare all solutions using deionized water and analytical grade reagents. Store all reagents at room temperature unless indicated otherwise. Strictly follow waste disposal regulations whenever applicable.

1. Probe precursor with a terminal amino functional group (for example: PIP_2 precursor [25]).
2. Fluorous diazirine NHS ester reagents [23].
3. Triethylammonium bicarbonate (TEAB) buffer: Add 95 mL of water to 5 mL of 1 M TEAB (pH 8.5) to make a 50 mM TEAB solution. For a 0.5 M solution, add 0.5 mL of water to 0.5 mL of 1 M TEAB (pH 8.5).
4. Fluorous reversed-phase silica gel (FluoroFlash® Silica Gel, ~40 μm particle size, Sigma-Aldrich).
5. 6× Tris–HCl buffer (0.3 M, pH 8.0).
6. Recombinant protein: For example, ADP-ribosylation factor 1 (ARF1) was expressed and purified as described in the literature [26]. Dilute recombinant protein to a concentration of 200 μM with cold water before use (*see* **Note 1**).

7. Black Ray® XX-20BLB UV Bench Lamp (365 nm, 20-W).
8. Tris-glycine SDS sample buffer (2×) with 50 mM DTT: Add 25 μL of 1 M DTT solution to 475 μL SDS sample buffer (2×).
9. BenchMark™ Pre-stained Protein Ladder.
10. 4-20 % Tris-glycine gel.
11. Coomassie Blue Staining solution.
12. Tris-glycine SDS running buffer (10×).
13. Antibody for probe: For example, mouse anti-PtdIns(4,5)P_2.
14. Antibody for protein of interest: For example, Anti-Arf1.
15. HRP-conjugated anti-rabbit IgG.
16. Model peptides: For example, 5-H4 and 5-D4 [24].
17. 10 mM ammonium formate.

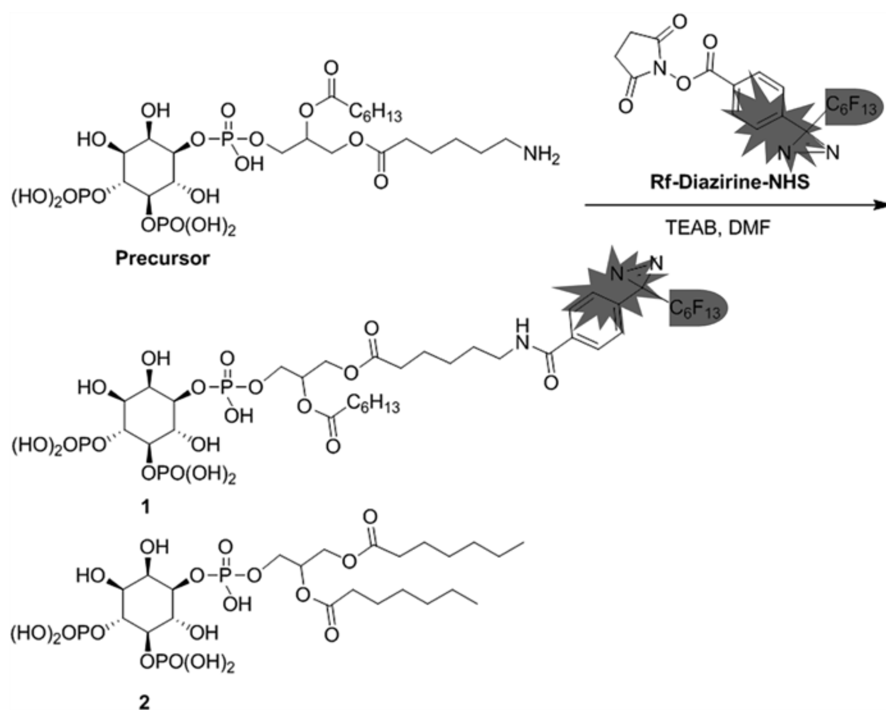
3 Methods

Perform all experiments at room temperature unless otherwise specified.

3.1 Preparation of Fluorous PAL Probe

The PAL probe can be prepared efficiently by coupling a small molecule of interest containing a terminal amine group with an activated fluorous diazirine reagent [23, 25] such as Rf-diazirine-NHS (Scheme 1). A general procedure is described below when PIP_2 is the small molecule of interest.

1. Add the solution of fluorous diazirine *N*-hydroxysuccinimide (NHS) ester (5.1 mg, 10.2 μmol) in DMF (600 μL) to the solution of PIP_2 precursor (4.5 mg, 6.3 μmol) in TEAB buffer (0.5 M, 600 μL) (*see* **Note 2**). The reaction mixture is stirred in the dark (*see* **Note 3**), and progress of the coupling reaction is monitored by MALDI-MS (*see* **Note 4**). Once the reaction is complete, remove the solvents under vacuum.
2. Suspend the fluorous reversed-phase silica gel in methanol (*see* **Note 5**), and then transfer to a glass pipette with a small amount of glass wool packed at the end. Gentle pressure can be applied to remove air bubbles. Wash the column subsequently with 100, 80, 60, 30, and 10 % methanol in water (*see* **Note 6**).
3. Treat the reaction product from Subheading 3.1, **step 1**, with 0.5 M TEAB (pH 8.5) and load it onto the column. First wash the column with 10 % and then 30 % MeOH in 50 mM TEAB (10 column volumes for each) (*see* **Note 7**).
4. Elute the fluorous-tagged probe **1** using 60 and 80 % MeOH in 50 mM TEAB (10 column volumes for each) (*see* **Note 8**).



Scheme 1 Synthesis of fluorous diazirine-PIP_2 **1** and structure of non-fluorous PIP_2 analog **2**

3.2 Cross-Linking of Fluorous PAL Probe with Protein Targets

As an extension of the traditional PAL approach, fluorous PAL probes should be useful for target identification. Limited by the lack of cell permeability of the PIP_2 probe (prepared in Subheading 3.1), only recombinant protein was tested in this demonstration experiment. However, live cells or cell lysate can also be explored similarly if the small molecule of interest is cell permeable.

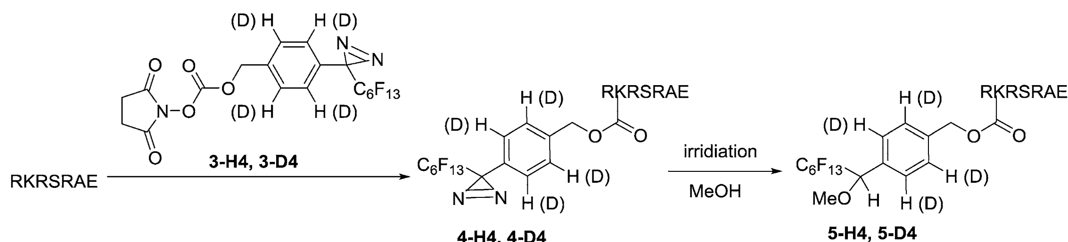
1. Incubate the solution of ARF1 (20 μM) in 1× Tris buffer (50 mM, pH 8.0, 60 μL) with the fluorous diazirine PIP_2 **1** (40 μM) for 30 min at 4 °C (*see* **Note 9**). In the competition experiment, 4 mM non-tagged PIP_2 derivative **2** is added to the assay mixture.
2. Illuminate the reaction mixture with a Black Ray® XX-20BLB UV Bench Lamp (365 nm, 20-W) at 4 °C for 10 min.
3. Add Tris-glycine SDS sample buffer (60 μL) containing 50 mM of DTT to the assay sample.
4. Heat the samples at 95 °C for 5 min, and then load them (30 μL each) onto a 4–20 % Tris-glycine gel. Benchmark prestained molecular weight standards also need to be included for each gel.
5. Following electrophoresis, rinse the gel with water and transfer it carefully to a container with Coomassie Blue Staining solution.

6. Stain the gel at room temperature overnight (with shaking) and destain by changing water once every 1 h until the background is gone.
7. Transfer the proteins on the SDS-PAGE gel to a polyvinylidene fluoride (PVDF) membrane. The Western blot is obtained by using mouse anti-PtdIns(4,5)P_2 (1:2,000) and horseradish peroxidase (HRP)-conjugated anti-mouse IgG (1:5,000) or anti-ARF1 (1:2,000) and HRP-conjugated anti-rabbit IgG (1:5,000) (*see* **Notes 10** and **11**).

3.3 Isotope-Coded Fluorous PAL Reagents and Specific Enrichment of Labeled Peptide

Photoaffinity labeling has the tendency to cross-link with nonspecific proteins. Using a pair of isotope-coded fluorous PAL reagents could minimize the nonspecific interactions as the photolabeled protein will show isotopic fingerprints of M^+ and $(M+4)^+$ ions in LC-MS/MS analysis, and thereby enhance the chances to identify cellular targets of small molecules. The isotope-coded fluorous PAL probes can be easily prepared by coupling the small molecule of interest with isotope-coded fluorous labeling reagents (**3-H4**, **3-D4**) [24] as described in Subheading 3.1. Isotope-coded, fluorous-labeled model peptides (**5-H4**, **5-D4**) were prepared and mixed with trypsin-digested bovine serum albumin (BSA) to form a relatively complex mixture (*see* **Note 12**), which was used to demonstrate that the labeled peptides from the digestion of cross-linked protein could be selectively separated from a complex assay mixture by FSPE (Scheme 2).

1. Dissolve the peptide RKRSRAE (0.25 mg) in 35 μL of water and mix with 15 μL of 1.0 M TEAB (pH 8.5). Add the isotope-coded fluorous labeling reagents (**3-H4**, **3-D4**) (0.9 equiv.) in 50 μL of DMF, and incubate the resulting mixture at room temperature with shaking for 40 min (*see* **Note 13**).
2. Remove the solvents by evaporation under vacuum.
3. Add methanol (100 μL) to the resulting residue.
4. Irradiate the methanol solution with Black Ray® XX-20BLB UV Bench Lamp (365 nm, 20-W) at 4 °C for 10 min.
5. Dry the photo-irradiated solution under vacuum, and resuspend the residue in 50 μL of 60 % methanol.



Scheme 2 Preparation of model peptide labeled with H_4/D_4 fluorous diazirine reagents

6. Mix the labeled peptide (7.5 nmol) with 9 μg of trypsin-digested BSA.
7. Load the above mixture onto a capillary column packed with fluorous silica. First wash the column with 60 % of methanol in 10 mM ammonium formate to remove the non-fluorous-tagged peptides, and then elute the fluorous-tagged peptides with 100 % methanol.
8. Remove the solvents by evaporation under vacuum, and the residue is ready for MS analysis (*see* **Note 14**). The ratio of the intensity of M^+ to that of $(M+4)^+$ should be similar to the ratio of **3-H4** to **3-D4** used in Subheading 3.3, **step 1**.

4 Notes

1. Due to the poor cell permeability of PIP_2, we did not test the probe in cells and used purified protein instead.
2. Depending on the probe, additives may be necessary to stabilize the products. For example, we added TEAB buffer to stabilize the PIP_2 probe.
3. Always keep the photoactive reagents in the dark since they are light sensitive.
4. MALDI-MS was used in this example because it is difficult to analyze the PIP_2 reaction. For most of small molecules/substrates, the reaction progression can easily be monitored by LCMS or TLC.
5. Sonication or vortex helps the fluorous silica gel get wet in methanol. The insoluble solid on the surface of solvent should be removed with a pipette.
6. We often observed unidentified white solid contaminations from the fluorous silica gel. This problem could be solved by prewashing the column with both fluorophilic and fluorophobic solutions. Wash the column with fluorophobic solution in the last step for column equilibration and sample loading.
7. In a typical FSPE, non-fluorous-tagged species are first washed off with fluorophobic solution (wet organic solvent such as 60 % methanol in H_2O), while the fluorous-tagged species will retain on the fluorous column and eluted with fluorophilic solution (such as 100 % methanol). For very hydrophilic substrates, a lower percentage of organic solvents in H_2O is required. To obtain optimal purification, stepwise or gradient washing and elution are required. At least 10 column volumes of solvents are needed for each step.
8. The majority of PIP_2 products were eluted in the 60 % MeOH fraction with high purity. The 80 % MeOH fraction also contains

small amount of the product and unreacted fluorous diazirine reagents remaining in the solution.

9. A typical photoaffinity labeling reaction mixture contains 10 μL of 6× Tris–HCl (300 mM, pH 8.0) and 6 μL of ARF1 (200 μM), 12 μL of fluorous diazirine PIP_2 (200 μM in H_2O), and 32 μL of water. For the competition experiment, 32 μL of Di-C6-PIP2 (7.5 mM in water) was added instead of water.
10. Only the cross-linked bands should be detected by both antibody against the substrate and antibody against the interacting protein. The cross-linked band will disappear in the presence of large amount of non-tagged substrate, for example compound **2** we used for the PIP_2 PAL probe.
11. Western blot was preformed to confirm that the newly formed band is indeed the photoaffinity-labeled protein. However, such a step is optional and is not necessary for the fluorous PAL approach to work.
12. To assess the sensitivity and separation efficiency, the fluorous-tagged peptide was mixed with trypsin-digested BSA to form a relatively complex mixture. The isotope-coded approach was also tested in the same experiment.
13. **Steps 1–5** in Subheading 3.3 are to generate the model peptide for the demonstration experiment, and are not required for a typical fluorous PAL experiment.
14. In this particular example, the samples were analyzed by AB Sciex 4800 Plus MALDI-TOF. However, other MS spectrometers should also work.

Acknowledgement

We thank the University of North Carolina at Chapel Hill and the National Institutes of Health (GM086558 to Q.Z.) for financial support of this work. We also thank Dr. Jonathan Goldberg (Memorial Sloan-Kettering) for providing the expression vector for ARF1.

References

1. Singh A, Thorton ER, Westheimer FH (1962) The photolysis of diazoacetylchymotrypsin. J Biol Chem 237:3006–3008
2. Geurink PP et al (2012) Photoaffinity labeling in activity-based protein profiling. Top Curr Chem 324:85–114
3. Robinette D et al (2006) Photoaffinity labeling combined with mass spectrometric approaches as a tool for structural proteomics. Expert Rev Proteomics 3:399–408
4. Steen H, Mann M (2002) A new derivatization strategy for the analysis of phosphopeptides by precursor ion scanning in positive ion mode. J Am Soc Mass Spectrom 13:996–1003
5. Oda Y, Nagasu T, Chait BT (2001) Enrichment analysis of phosphorylated proteins as a tool

for probing the phosphoproteome. Nat Biotechnol 19:379–382

6. Curran DP (2008) Fluorous tags unstick messy chemical biology problems. Science 321: 1645–1646
7. Curran DP (2001) Fluorous reverse phase silica gel. A new tool for preparative separations in synthetic organic and organofluorine chemistry. Synlett 2001:1488–1496
8. Zhang Q, Luo Z, Curran DP (2000) Separation of "Light Fluorous" reagents and catalysts by fluorous solid-phase extraction: synthesis and study of a family of Triarylphosphines bearing linear and branched fluorous tags. J Org Chem 65:8866–8873
9. Matsugi M, Curran DP (2005) Synthesis, reaction, and recycle of light fluorous Grubbs-Hoveyda catalysts for alkene metathesis. J Org Chem 70:1636–1642
10. Hungerhoff B, Sonnenschein H, Theil F (2002) Combining lipase-catalyzed enantiomer-selective acylation with fluorous phase labeling: a new method for the resolution of racemic alcohols. J Org Chem 67:1781–1785
11. Dandapani S, Curran DP (2002) Fluorous Mitsunobu reagents and reactions. Tetrahedron 58:3855–3864
12. Miura T et al (2001) Fluorous oligosaccharide synthesis using a novel fluorous protective group. Org Lett 3:3947–3950
13. Luo Z et al (2001) Fluorous mixture synthesis: a fluorous-tagging strategy for the synthesis and separation of mixtures of organic compounds. Science 291:1766–1769
14. Zhang QS et al (2004) Fluorous mixture synthesis of stereoisomer libraries: total syntheses of (+)-murisolin and fifteen diastereoisomers. J Am Chem Soc 126:36–37
15. Zhang W et al (2002) Solution-phase preparation of a 560-compound library of individual pure mappicine analogues by fluorous mixture synthesis. J Am Chem Soc 124:10443–10450
16. Ko KS, Jaipuri FA, Pohl NL (2005) Fluorous-based carbohydrate microarrays. J Am Chem Soc 127:13162–13163
17. Vegas AJ et al (2007) Fluorous-based small-molecule microarrays for the discovery of histone deacetylase inhibitors. Angew Chem Int Ed 46:7960–7964
18. Jaipuri FA, Collet BY, Pohl NL (2008) Synthesis and quantitative evaluation of glycero-D-manno-heptose binding to concanavalin A by fluorous-tag assistance. Angew Chem Int Ed 47:1707–1710
19. Huang W et al (2014) Fluorous enzymatic synthesis of phosphatidylinositides. Chem Commun 50:2928–2931
20. Cai C et al (2014) Fluorous-assisted chemoenzymatic synthesis of heparan sulfate oligosaccharides. Org Lett 16:2240–2243
21. Brittain SM et al (2005) Enrichment and analysis of peptide subsets using fluorous affinity tags and mass spectrometry. Nat Biotechnol 23:463–468
22. Northen TR et al (2008) A nanostructure-initiator mass spectrometry-based enzyme activity assay. Proc Natl Acad Sci U S A 105: 3678–3683
23. Song Z, Zhang Q (2009) Fluorous aryldiazirine photoaffinity labeling reagents. Org Lett 11:4882–4885
24. Song Z, Huang W, Zhang Q (2012) Isotope-coded, fluorous photoaffinity labeling reagents. Chem Commun 48:3339–3341
25. Huang W et al (2012) Incorporation of a fluorous diazirine group into phosphatidylinositol 4,5-bisphosphate to illustrate its interaction with ADP-ribosylation factor 1. Org Biomol Chem 10:5197–5201
26. Goldberg J (1999) Structural and functional analysis of the ARF1–ARFGAP complex reveals a role for coatomer in GTP hydrolysis. Cell 96:893–902

Chapter 21

Identification of the Targets of Biologically Active Small Molecules Using Quantitative Proteomics

Glòria Vendrell-Navarro, Andreas Brockmeyer, Herbert Waldmann, Petra Janning, and Slava Ziegler

Abstract

Currently, cell-based screenings yield a multitude of small molecule modulators of diverse biological processes. The most demanding step in the course of elucidation of the mode of action of biologically active compounds is the identification of the target proteins. Although there is no generic approach available, affinity-based chemical proteomics is the most widely applied methodology. Particularly, quantitative chemical proteomics has proven very powerful in the identification of the putative targets of small molecules. Here we describe the procedure for identification of target proteins for small molecules employing affinity chromatography and the stable isotope labeling in cell culture (SILAC) for quantitative proteomics.

Key words Affinity chromatography, SILAC, Mass spectrometry, Proteomics, Pulldown

1 Introduction

Small molecules are powerful tools to investigate and/or manipulate complex biological systems (e.g., cells or organisms). One goal of Chemical Biology is to identify small molecule modulators for every single protein (or protein's function), a quite demanding ambition considering the size of organisms' proteomes and the number of successfully purified and characterized proteins. Chemical Biology does not offer a generic approach to tackle this challenge, i.e., when compared to classical genetics. The direct exploration of modulators of given proteins is limited, inter alia, by the accessibility of the purified proteins. Rather the exposure of small molecules to living systems and thus to whole proteomes bears the potential to uncover and perturb (novel) functions of proteins. Thus, cell-based screening is increasingly being conducted by academia and pharmaceutical industry and novel small molecule modulators of various biological phenomena are continuously being reported. The major bottleneck in this process is to unravel

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_21, © Springer Science+Business Media New York 2015

the molecular targets of given small molecules, i.e., the direct binding partners, the biomolecules whose function is perturbed by the compound. Although a major effort is put in developing reliable methodologies for identification of the targets of biologically active compounds, so far there is no generic approach available.

Affinity-based chemical proteomics (also called *pulldown*) appears to be the most widely used technique [1]. For this, the active small molecule is immobilized on solid support. This step usually requires information on the structure–activity relationship (SAR), which guides the attachment of a spacer (i.e., linker) with a functional group (e.g., amine group or biotin) at a site that tolerates modifications without significant loss of activity to generate the active pulldown probe (*see* **Note 1**). In parallel, a suitable control probe is prepared that should not bind to the target proteins. This probe might represent a closely related inactive derivative or the linker alone. After immobilization the probes are exposed to protein lysate to allow for the binding of target proteins. Enriched proteins are then eluted and identified upon tryptic digest by means of tandem (MS/MS) mass spectrometry. Frequently, a simple comparison of the proteins enriched with the active probe and control probe is then preformed and proteins that are identified only using the active probe are considered as hits. However, target proteins may bind nonspecifically and thus may be enriched with both active and control probes, which would lead to false-negative results.

Quantification of the relative amounts of the identified proteins is a powerful approach which overcomes the limitations of the qualitative comparison of identified proteins. For this purpose, usually stable isotope labels are metabolically or chemically incorporated in the proteins/peptides, although label-free quantification is possible as well [2]. Stable isotope labeling by amino acids in cell culture (SILAC) is a method for metabolic labeling of proteins in cells or cell cultures [3]. In general, at least two different stable isotopically labeled cell cultures are generated under the same growth conditions with the exception of the presence of normal ("light") or labeled ("heavy") amino acids (e.g., [$^{13}C_6{}^{15}N_4$]-labeled arginine and [$^{13}C_6$]-labeled lysine), respectively (Fig. 1a). In standard SILAC experiments incorporation of labeled amino acids has to be complete. Therefore cells are cultured in presence of the labeled amino acids for at least 5–6 cell doublings. Afterwards both cell populations, i.e., with proteins labeled with "heavy" or "light" amino acids, are subjected to experiments, which require quantitative comparison of proteins. For example, to assess the influence of a small molecule on protein expression levels in cells, cells that are labeled with "heavy" amino acids are treated with the compound, whereas cells that are labeled with "light" amino acids are exposed to a control. After appropriate time, cell

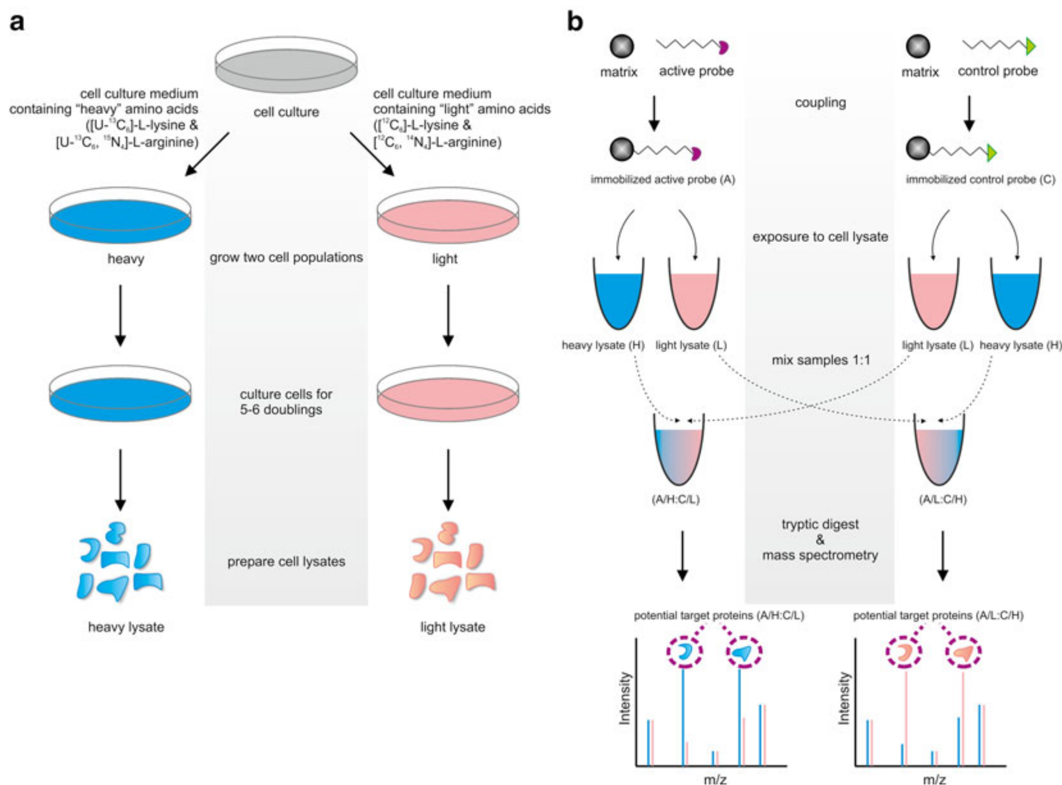


Fig. 1 SILAC-based affinity chromatography for identification of the target proteins of small molecules. (**a**) Generation of two cell populations differentially labeled with light or heavy amino acids. (**b**) Employment of the metabolically labeled cell lysates from (**a**) in affinity-based chemical proteomics

cultures are mixed in a ratio of 1:1 and lysate is prepared immediately. After tryptic digestion the samples are analyzed via nano-HPLC-MS/MS using a shotgun proteomics method. Proteins whose expression levels are not affected by the small molecule will be indicated by ratios of 1:1 regarding the precursor ion signals of corresponding "heavy" and "light" labeled peptides. Those proteins which are upregulated will give a ratio >1:1 and downregulated ones of <1:1, respectively. Labeled and unlabeled forms of the proteins are distinguished by the mass shift of the incorporated labels (e.g., $\Delta M = 10.0076$ Da for $[^{13}C_6{}^{15}N_4]$-labeled arginine and 6.0201 Da for $[^{13}C_6]$-labeled lysine). Working with ^{13}C and ^{15}N does not affect the binding properties of labeled proteins and the chromatographic behavior of tryptic peptides, which allows rather straightforward comparison of peak areas. Therefore, samples can be mixed at an early stage of the sample preparation procedure preventing random and unavoidable differences in sample

handling. This leads to a better precision of determination of SILAC ratios compared to label free approaches where samples have to be handled and measured in parallel.

For target identification, this principle can be used with slight differences [4]. In a pulldown employing the immobilized active small molecule the probe is incubated with for example the “heavy” cell lysate (Fig. 1b). The control pulldown, in which an inactive control probe is immobilized, is performed using the “light” lysate. After the pulldown procedure the samples are mixed in a 1:1 ratio and the SILAC-based quantification approach is carried out as described above. Proteins that bind nonspecifically to the probes or the matrix will be detected in a 1:1 ratio, whereas proteins that bind specifically to the active probe will be detected in a ratio >1:1 based on the intensity of the precursor ions of the corresponding peptides in both cases. As a control, the reverse experiment is carried out in parallel (pulldown with the inactive control probe using the “heavy” lysate and pulldown with the active probe using the “light” lysate). Obviously, in this case the ratio of specific binders will be <1:1. This method should enable the identification even of proteins that bind relatively weakly to small molecules since less stringent washing conditions can be applied and proteins that bind nonspecifically to the probe will result in 1:1 ratio of all intensities. A background of a rather high number of unspecific binding proteins is not a problem and offers a chance for a good normalization of protein ratios. Visual identification of potential specific binders is straightforward after normalizing the ratios to 1:1 for the median of all proteins, calculating the logarithm (log2) of normalized ratios and plotting the ratios of the forward experiment against those of the reverse one (Fig. 2).

A relatively simple method to assess the data quality is to plot the ratio distribution of all proteins quantified in one sample against the number of proteins of a distinct ratio range after calculation of the logarithm of the normalized ratios. In an ideal case, the data distribution is Gaussian-shaped and is very close to 0 for nearly all proteins with just a few outliers (Fig. 3). For more detailed data evaluation statistical tests should be used to identify outliers. The methods described in this chapter are based on the calculation of the standard deviation of the normalized and logarithmized ratios of all proteins followed by the calculation of the probability whether a ratio is a statistically significant outlier or not [5]. This test is described in two variations: first the overall standard deviation of the normalized and logarithmized ratios and second the standard deviation related to peak intensity bins (containing at least 300 ratios per bin) is used.

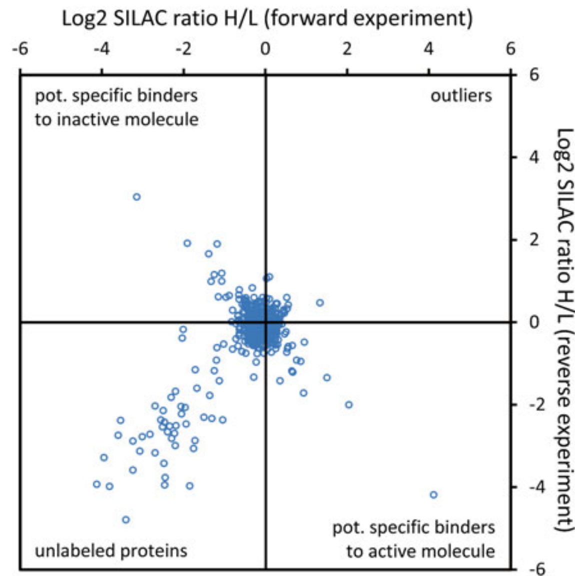


Fig. 2 Scatter plot of log2 ratios of forward and reverse replicate experiments. Plotted are the medians of each three replicates. Unspecific binding proteins will have a log2 ratio of around 0. Potential specific binders will have positive log2 values in the forward and negative log2 values in the reverse experiment. If there are potential specific binders for the inactive control probe, they will have negative log2 values in the forward and positive log2 values in the reverse experiment. Unlabeled proteins like keratins, trypsin, etc., most resulting from contaminations, will show negative log2 values in the forward and the reverse experiment. Proteins showing positive log2 values in the forward and reverse experiment are statistical outliers without any biological meaning

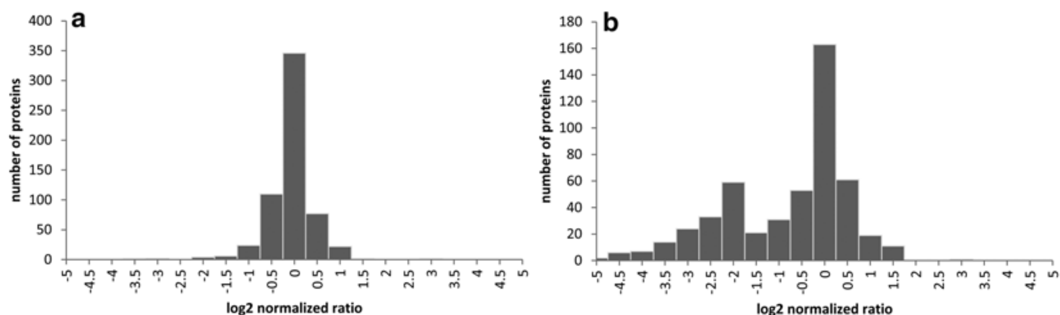


Fig. 3 Histogram of protein ratio distributions. (**a**) Nice distribution of protein ratios after an affinity pulldown experiment. (**b**) Very broad and abnormal distribution which indicates problems during the sample preparation process

2 Materials

2.1 HeLa Cell Culture

1. HeLa cells (ATCC® CCL-2™).
2. HeLa cell culture medium: High-glucose (4.5 g/L) Dulbecco's Modified Essential Medium (DMEM) supplemented with 10 % (v/v) fetal bovine serum (FBS), 4 mM L-glutamine, 1 % (v/v) nonessential amino acids, 1 mM sodium pyruvate, 100 U/mL penicillin, 0.1 mg/mL streptomycin. Add 50 mL of FBS, 10 mL of L-glutamine (200 mM, cell culture grade), 5 mL of nonessential amino acids (100×, cell culture grade), 5 mL of sodium pyruvate (100 mM, cell culture grade), and 5 mL of penicillin/streptomycin (10,000 U/mL or 10 mg/mL, cell culture grade) to 425 mL of high-glucose DMEM medium. Mix and store at 4 °C up to 1 month.
3. Unsupplemented SILAC medium: Commercial high-glucose (4.5 g/L) Dulbecco's Modified Essential Medium (DMEM) without L-Arginine, L-Glutamine, L-Lysine, sodium pyruvate, and HEPES Buffer, and with methionine and $CaCl_2$ (e.g., D-MEM-Flex, Invitrogen).
4. Light SILAC cell culturing medium (light medium): Unsupplemented SILAC medium with 10 % (v/v) of dialyzed fetal bovine serum (dFBS), 2 mM L-glutamine, 1 % (v/v) nonessential amino acids, 1 mM sodium pyruvate, 100 U/mL penicillin, 0.1 mg/mL streptomycin, 0.1 mg/mL [$^{12}C_6$]-L-lysine (MW: 146.1055 g/mol), 0.1 mg/mL [$^{12}C_6$, $^{14}N_4$]-L-arginine (MW: 174.1117 g/mol). Dissolve 50 mg of [$^{12}C_6$]-L-lysine (cell culture grade) and 50 mg of [$^{12}C_6$, $^{14}N_4$]-L-arginine (cell culture grade) each in 1 mL of unsupplemented SILAC medium. Mix well and add to 429 mL of unsupplemented SILAC medium. Add to the resultant medium 50 mL of dFBS, 5 mL of L-glutamine (200 mM, cell culture grade), 5 mL of nonessential amino acids (100×, cell culture grade), 5 mL of sodium pyruvate (100 mM, cell culture grade), and 5 mL of penicillin/streptomycin (10,000 U/mL or 10 mg/mL, cell culture grade). Mix and filter using a sterile filtering flask with 0.2 μm pore size.
5. Heavy SILAC cell culturing medium (heavy medium): Unsupplemented SILAC medium supplemented with 10 % (v/v) dialyzed fetal bovine serum (dFBS), 2 mM L-glutamine, 1 % (v/v) nonessential amino acids, 1 mM sodium pyruvate, 100 U/mL penicillin, 0.1 mg/mL streptomycin, 0.1 mg/mL [U-$^{13}C_6$]-L-lysine (MW 146.1055 g/mol), 0.1 mg/mL [U-$^{13}C_6$, $^{15}N_4$]-L-arginine (MW 174.1117 g/mol). Dissolve 50 mg of [U-$^{13}C_6$]-L-lysine (cell culture grade), and 50 mg of [U-$^{13}C_6$, $^{15}N_4$]-L-arginine (cell culture grade) each in 1 mL of unsupplemented SILAC medium. Mix well and add to 429 mL

of unsupplemented SILAC medium. Supplement the resultant medium as indicated for the light medium. Mix and filter using a 0.2 μm sterile filtering flask.

6. Trypsin–EDTA (0.05 % trypsin with 0.2 g/L EDTA in PBS) or non-enzymatic cell dissociation solution, cell culture grade.
7. Phosphate-buffered saline (PBS): 10 mM Na_2HPO_4, 137 mM NaCl, 1.8 mM KH_2PO_4, 2.7 mM KCl, pH 7.45. Add 1.44 g of Na_2HPO_4, 8 g of NaCl, 0.24 g of KH_2PO_4, and 0.2 g of KCl to ultrapure water. Adjust the pH to 7.45 with 1 N HCl and autoclave.
8. Cell culture dishes (10 cm diameter) and cell culture flasks (T75 and T175).
9. Cell scraper.
10. Trypan blue stain: 0.4 % (w/v) Trypan blue in PBS. Dissolve 40 mg of trypan blue in 10 mL of PBS buffer. Filter through a 0.2 μm filter. Store at room temperature.

2.2 Preparation of Cell Lysates

1. Lysis buffer: 50 mM PIPES (pH 7.4), 50 mM NaCl, 5 mM $MgCl_2$, 5 mM EGTA, 0.1 % (v/v) Nonidet™ P 40 alternative (4-nonylphenyl-polyethylene glycol), 0.1 % (v/v) Triton™ X-100 (*tert*-octylphenoxypolyethoxyethanol), 0.1 % (v/v) Tween® 20 (polyoxyethylenesorbitan monolaurate). Add 7.56 g of PIPES, 1.46 g of NaCl, 0.51 g of $MgCl_2 \cdot 6H_2O$, 0.93 g of EGTA to ~300 mL of ultrapure water. Add NaOH to completely dissolve the PIPES and adjust the pH to 7.4 with HCl. Bring the volume to 500 mL with ultrapure water. Add 500 μL of Nonidet™ P 40 alternative, 500 μL of Triton™ X-100, and 500 μL of Tween® 20. Filter through a 0.20 μm sterile filter and store at −20 °C. Before use, add fresh 1 mM dithiothreitol (DTT) (alternatively, add 0.1 % (v/v) 2-mercaptoethanol), protease inhibitors, and phosphatase inhibitors.
2. Commercial Bradford protein determination reagent.

2.3 Determination of the Metabolic Labeling Efficiency

1. 0.5 M Tris buffer: Add 6.06 g of Tris to ~80 mL ultrapure water. Adjust the pH to 6.8 with HCl. Add ultrapure water to obtain a 100 mL final volume.
2. 5× SDS sample buffer: 0.2 M Tris (pH 6.8), 40 % (v/v) glycerol, 100 g/L sodium dodecyl sulfate (SDS), 62 g/L dithioerythritol (DTE), 0.2 g/L bromophenol blue. Mix 20 mL of 0.5 M Tris buffer (pH 6.8), 20 mL of 40 % glycerol, 5 g of SDS, 3.1 g of DTE, and 10 mg of bromophenol blue. Adjust the volume to 50 mL with ultrapure water. Store the 5× SDS sample buffer at −20 °C in small aliquots.
3. Thermo-shaker.
4. Bovine serum albumin (BSA) solution: 50 mg/L in PBS. Store at −20 °C.

5. Pre-cast 4–20 % gradient gels for sodium dodecyl sulfate gel electrophoresis (SDS-PAGE) (e.g., Tris-glycine gels).
6. SDS-PAGE Running Buffer: 25 mM Tris, 0.2 M glycine, 0.1 % (w/v) sodium dodecyl sulfate (SDS). Prepare a 10× concentrated stock by adding 30.3 g of Tris, 144 g of glycine, 10 g of SDS in 1 L of ultrapure water. Prepare the 1× working solution by diluting 1:10 in ultrapure water. Store at room temperature.
7. Suitable protein electrophoresis system.
8. Colloidal Coomassie staining solution [6]: 0.1 % Coomassie G-250, 20 g/L H_3PO_4, 100 g/L $(NH_4)_2SO_4$, 20 % (v/v) methanol. Prepare a stock staining solution by dissolving 20 g of phosphoric acid and 100 g of ammonium sulfate in ~800 mL of ultrapure water. Add 1 g of Coomassie dye while stirring. Adjust the volume to 1 L with ultrapure water. The final staining solution is obtained by mixing 80 mL of the stock staining solution with 20 mL of methanol.

2.4 Immobilization of Pulldown Probes Bearing a Primary Amino Group

1. Pulldown probe with primary amine group and suitable control probe.
2. *N*-Hydroxysuccinimide (NHS) magnetic sepharose beads (e.g., GE Healthcare).
3. Magnetic separator and overhead rotator.
4. Protein low binding tubes and protein low bind tips.
5. Equilibration solution: 1 mM HCl. Add 0.1 mL of 1 M HCl to 99.9 mL ultrapure water. Filter through 0.2 μm chemical-resistant filter. Store at 4 °C and use as an ice-cold solution.
6. Coupling buffer: 0.15 M triethanolamine, 0.5 M NaCl, pH 8.3. Add 2.23 g of triethanolamine and 2.92 g of NaCl to ~80 mL ultrapure water. Mix and adjust the pH using NaOH. Bring up to 100 mL using ultrapure water. Aliquot and store at −20 °C.
7. Blocking buffer A: 0.5 M ethanolamine, 0.5 M NaCl, pH 8.3. Add 3.05 g of ethanolamine and 2.92 g of NaCl to ~80 mL ultrapure water. Mix and adjust the pH using NaOH. Bring up to 100 mL using ultrapure water. Aliquot and store at −20 °C.
8. Blocking buffer B: 0.1 M sodium acetate, 0.5 M NaCl, pH 4.0. Add 0.82 g of sodium acetate and 2.92 g of NaCl to ~80 mL ultrapure water. Mix and adjust the pH using HCl. Bring up to 100 mL using ultrapure water. Filter through 0.2 μm chemical resistant filter. Aliquot and store at −20 °C.

2.5 Determination of the Coupling Time and Coupling Efficiency

1. Silica TLC plates.
2. Succinimide standard: 20 mM succinimide in dichloromethane.

3. Mobile phase for developing TLC plate: dichloromethane–methanol 9:1 (v/v). Add 100 mL of methanol to 900 mL of dichloromethane and mix.
4. Permanganate stain for staining TLC plates: 10 g/L $KMnO_4$, 60 g/L K_2CO_3, and 1 g/L NaOH in water. Add 1 g of $KMnO_4$, 7 g of K_2CO_3, and 0.1 g of NaOH to 100 mL of water.

2.6 Immobilization of Pulldown Probes Bearing a Biotin Group

1. Streptavidin magnetic beads.
2. Biotinylated pulldown probe and suitable control probe.

2.7 Enrichment of Proteins That Bind to Immobilized Probes (Pulldown)

1. Lysis buffer supplemented with 75 mM $MgCl_2$: 50 mM PIPES (pH 7.4), 50 mM NaCl, 75 mM $MgCl_2$, 5 mM EGTA, 0.1 % (v/v) Nonidet™ P 40 alternative (4-nonylphenyl-polyethylene glycol), 0.1 % (v/v) Triton™ X-100 (*tert-octylphenoxypoly-ethoxyethanol*), 0.1 % (v/v) Tween® 20 (polyoxyethylenesorbitan monolaurate). Add 756 mg of PIPES, 146 mg of NaCl, 760 mg of $MgCl_2 \cdot 6H_2O$, 93 mg of EGTA to ~30 mL of ultrapure water. Add NaOH to completely dissolve the PIPES and adjust the pH to 7.4 with HCl. Bring the volume to 50 mL with ultrapure water. Add 500 μL of Nonidet™ P 40 alternative, 500 μL of Triton™ X-100, and 500 μL of Tween® 20. Filter through a 0.20 μm sterile filter and store at −20 °C. Before use, add freshly protease and phosphatase inhibitors.

2.8 Tryptic Digestion of Proteins

2.8.1 In-Gel Tryptic Digestion

1. 25 mM NH_4HCO_3: Dissolve 494.4 mg of NH_4HCO_3 in 250 mL of ultrapure water. Filter through a 0.2 μm sterile filter, aliquot and store at −20 °C.
2. Gel washing solution A: 25 mM NH_4HCO_3–acetonitrile (3:1 (v/v)). Mix 3 parts of 25 mM NH_4HCO_3 with 1 part of acetonitrile. Prepare freshly before use.
3. Gel washing solution B: 25 mM NH_4HCO_3–acetonitrile (1:1 (v/v)). Mix 1 part of 25 mM NH_4HCO_3 with 1 part of acetonitrile. Prepare freshly before use.
4. Reducing solution: 50 mM DTT in 25 mM NH_4HCO_3. Dissolve 77 mg of DTT in 10 mL of 25 mM of NH_4HCO_3. Prepare freshly before use.
5. Alkylating solution: 55 mM iodoacetamide in 25 mM NH_4HCO_3. Dissolve 101 mg of iodoacetamide in 10 mL of 25 mM NH_4HCO_3. Prepare freshly before use.
6. Trypsin stock: 0.4 μg/μL trypsin (proteomics grade) in 10 mM HCl. Aliquot and store at −20 °C.
7. Tryptic in-gel digestion solution: 10 ng/μL trypsin in 25 mM NH_4HCO_3. Add 25 μL of trypsin stock to 975 μL of 25 mM NH_4HCO_3. Prepare freshly before use.
8. Centrifugal evaporator.

2.8.2 On-Bead Tryptic Digestion and StageTip Purification

1. 50 mM NH_4HCO_3: Dissolve 494.4 mg of NH_4HCO_3 in 125 mL of ultrapure water. Filter through a 0.2 μm sterile filter, aliquot and store at −20 °C.
2. 50 mM Tris buffer, pH 7.5: Add 606 mg of Tris in ~80 mL ultrapure water. Adjust the pH to 7.5 with HCl. Add ultrapure water to give a 100 mL final volume. Autoclave and store at 4 °C.
3. Urea solution: 2 M urea in 50 mM Tris. Dissolve 1.2 g of urea in 9 mL of 50 mM Tris buffer (pH 7.5). Preserve at 4 °C and use within a day.
4. Solution E1: 1 mM DTT, 5 μg/mL trypsin in urea solution. Prepare a 1 M DTT stock by dissolving 15.4 mg of DTT in 100 μL of 50 mM NH_4HCO_3. Add 1.1 μL of DTT stock and 13.75 μL of trypsin stock (0.4 μg/μL) to 1,085.15 μL of urea solution. Prepare freshly before use.
5. Solution E2: 5 mM chloroacetamide in urea solution. Prepare a 0.5 M chloroacetamide stock by dissolving 9.35 mg of chloroacetamide in 200 μL of ultrapure water. Add 11 μL of chloroacetamide stock to 1,089 μL of urea solution. Prepare freshly before use.
6. Trifluoroacetic acid (TFA) (proteomics grade). It is important to use ampoules and to prevent storage in tubes with plastic caps.
7. Wash solution: 0.1 % (v/v) formic acid in water. Add 100 μL of formic acid to 100 mL of ultrapure water and mix.
8. Elution solution: Acetonitrile–water–formic acid (80:20:0.1 (v/v/v)). Mix 80 mL of acetonitrile, 20 mL of ultrapure water, and 100 μL of formic acid. Protect from light.
9. Extraction disks C18 (47 mm) (e.g., Supelco, #66883-U).
10. 17 or 18 G blunt ended syringe needle (StageTip syringe).
11. Microcentrifuge tip adapters. Alternatively, a drilled tube cap can be used as an adapter.
12. OPTIONAL: StageTip centrifuge (e.g., Sonation Tomy Mini Person centrifuge). Using this centrifuge, stage-tip microcolumns can be loaded directly without additional material. Otherwise, a small bench centrifuge can be used instead, but additional microcentrifuge tubes to collect the eluted buffer and microcentrifuge tip adapters are needed.

2.9 LC-MS/MS Based Analysis of Potential Target Proteins

1. Solvents, HPLC eluents: water (H_2O, LC-MS grade), acetonitrile (LC-MS grade), formic acid (LC-MS grade), trifluoroacetic acid (TFA, LC-MS grade).
2. Small instruments: ultrasonic bath, centrifuge.
3. Consumables: autosampler vials (e.g., polypropylene vials (250 μL), caps for 250 μL polypropylene vials).

4. HPLC: Nano-(U)HPLC system capable for flow-rates of 300 nL/min and pre-column desalting (e.g., UltiMate™ 3000 RSLCnano system, Dionex) equipped with a nano-HPLC column 75 μm ID C18 material (e.g., Acclaim PepMap RSLC C18, 2 μm, 100 Å, 75 μm ID × 25 cm, nanoViper, Dionex), pre-column cartridges (e.g., Acclaim PepMap100 C18, 5 μm, 100 Å, 300 μm ID × 5 mm, Dionex) and nano-spray emitter (Standard Coated SilicaTip™ Emitter, 360 μm OD, 20 μm ID, 10 μm Tip ID, 10.5 cm, New Objective, USA).
5. Mass spectrometer: High performance mass spectrometer capable for high resolution measurements of at least precursor peptide masses and MS/MS possibilities (e.g., Q Exactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer, Thermo Scientific) equipped with a nano-spray source (e.g., Nanospray Flex Ion Source, Thermo Scientific) (*see* **Note 2**).
6. Mass spectrometry peptide identification and quantification software (e.g., MaxQuant [5] or Proteome Discoverer™ Software, Thermo Scientific).

3 Methods

3.1 HeLa Cell Culture

1. Grow HeLa cells to subconfluence (80–90 %) in T75 cell culture flasks using HeLa cell culture medium.
2. Wash the cells with PBS, detach with trypsin/EDTA and resuspend the detached cells in fresh HeLa cell culturing medium (day 0).
3. Count living cells using trypan blue staining.
4. Add 10 mL of light medium to a 10 cm diameter cell culture dish. Add 10 mL of heavy medium to another 10 cm diameter cell culture dish (*see* **Note 3**).
5. Seed 4.5×10^5 HeLa cells per 10 cm diameter cell culture dish (*see* **Note 4**).
6. Incubate the cells at 37 °C and 5 % CO_2 for 5–6 cell doublings, e.g., 5–6 days for HeLa cells. Exchange the medium for the respective fresh one every 3 days. Cells are seeded in appropriate cell number, which allows maintaining them for 5–6 doublings in the same dish (i.e., without reaching confluence before harvesting).

3.2 Preparation of Cell Lysates

1. Cells grown in 10 cm diameter cell culture dishes should be nearly confluent (80–90 % confluence) for harvesting.
2. Remove the cell culture media and wash the cells with 10 mL of cold PBS.

3. Detach the cells using trypsin/EDTA or an alternative non-enzymatic dissociation solution. Collect the resuspended cells in appropriately sized centrifuge tubes by centrifugation at $500\times g$ for 5 min at 4 °C. Discard the supernatant.
4. Wash the cell pellets by resuspension in 10 mL of ice cold PBS and centrifuge at $500\times g$ for 5 min and 4 °C. Discard the supernatant.
5. Repeat the washing step twice more to ensure the removal of trace trypsin.
6. Add 0.5 mL of lysis buffer per 10 cm diameter cell culture dish to the tubes and incubate at 4 °C for 30 min. Gently vortex the samples each 5–10 min (*see* **Note 5**).
7. Homogenize the lysates by passing them five times through a 20 G needle fitted to a syringe.
8. Centrifuge the lysates for 20 min at $18{,}500\times g$ and 4 °C.
9. Collect the supernatant and determine the protein concentration using the Bradford reagent or an alternative method for determination of total protein (e.g., Lowry, Bicinchoninic Acid (BCA)).
10. Take a small aliquot to analyze the metabolic labeling efficiency, i.e., the degree of incorporation of heavy amino acids into proteins (*see* Subheading 3.3).
11. If the metabolic labeling efficiency is acceptable (*see* Subheading 3.3), upscale the cell culture (*see* **Note 3**) and prepare lysates of at least 1 mg/mL total protein. Make aliquots of 500–1,000 μg of total protein, freeze them in liquid nitrogen, and store at −80 °C (*see* **Note 6**).

3.3 Determination of the Metabolic Labeling Efficiency

This test proves that light peptides result only from light lysates and heavy peptides from heavy lysates and that protein expression levels are similar (1:1 in the 1:1 mixture). Furthermore, it determines whether arginine is converted to proline (*see* **Note 7**). The specified lysates are loaded on a SDS PAGE and a range of protein masses (e.g., 50–100 kDa) is selected for further in-gel digestion (*see* Subheading 3.8.1) and LC-MS/MS-based analysis of the resultant peptides (*see* Subheading 3.9).

1. Prepare the three following samples: light lysate, heavy lysate, and a 1:1 (v/v) mixture of the light and the heavy lysate, each one having the same total protein amount (e.g., 25–50 μg; *see* Subheading 3.2). Add an appropriate volume of 5× SDS sample buffer to the samples and heat them at 95 °C for 5 min.
2. Prepare an additional BSA control sample: add an appropriate volume of 5× SDS sample buffer to 5 pmol BSA (6.64 μL of BSA solution) and heat at 95 °C for 5 min (*see* **Note 8**).

3. Load the samples on a precast gradient gel and perform electrophoresis in SDS-PAGE Running Buffer (*see* **Note 9**).
4. Stain the gel with Colloidal Coomassie staining solution. For this, immerse the gel in Coomassie staining solution and incubate by gentle rocking at room temperature until bands become visible (usually 3–6 h or overnight). Rinse the gel 2–3 times with water and store if desired in 20–25 % (w/v) $(NH_4)_2SO_4$ in water (*see* **Note 10**).
5. Excise one or two protein bands from each lane side by side and perform tryptic in-gel digestion (*see* Subheading 3.8.1).

3.4 Immobilization of Pulldown Probes Bearing a Primary Amino Group

A general procedure for immobilization of pulldown amino probes is described below.

Prior to pulldown studies, an evaluation of the compound binding to the beads is highly recommended (*see* Subheading 3.5). Every step of liquid removal requires a rack tube with magnetic bar, whereas resuspension of solutions requires repeated inversion of the rack tube without the magnetic bar (*see* **Note 11** for the handling of the beads).

1. Add 25 μL of the resuspended NHS-activated magnetic sepharose matrix ("beads") to 1.5 mL low binding tubes. Prepare separate beads for the pulldown probe and the control (negative) probe (*see* Subheading 3.5). Prepare two tubes each with the active probe and control probe.
2. Remove the liquid and add 500 μL of ice cold equilibration solution to resuspend the beads.
3. Prepare 1 mL of the active and the control probe at a final concentration of 10 μM diluted in the coupling buffer (*see* **Note 12**). One milliliter is sufficient for two coupling reactions.
4. Immediately add 500 μL of the respective compound solution (from **step 3** of Subheading 3.4) to the beads and incubate with overhead rotation for at least 2 h at room temperature or with previously determined coupling conditions (*see* Subheading 3.5).
5. Block the residual active NHS groups on the beads by incubating the beads with 500 μL blocking buffer A for 5 min with overhead rotation. Remove the liquid.
6. Incubate the beads with 500 μL blocking buffer B for 5 min with overhead rotation. Discard the liquid.
7. Repeat Subheading 3.4, **step 5** by incubating the beads with blocking buffer A for 20 min.
8. Upon liquid removal, incubate the beads with 500 μL blocking buffer B for 5 min and remove the liquid.

9. Repeat Subheading 3.4, **steps 7** and **8** to ensure the capping of the NHS groups.
10. Wash the beads with 500 μL lysis buffer for 10 min. Remove the liquid and continue with Subheading 3.7.

3.5 Determination of the Coupling Time and Coupling Efficiency

Since the amino probe binds covalently to the NHS-magnetic matrix, it is important to verify that the reaction takes place and to estimate the time needed for optimal binding (*see* **Note 13**). For this, the released *N*-hydroxysuccinimide from the beads (as a sub-product of the reaction) is qualitatively determined over a time by means of thin layer chromatography (TLC). Different analytical methods might be needed depending on the nature of the pulldown probe (*see* **Note 14**).

1. Take 50 μL aliquots at different time points of the coupling reaction and after Subheading 3.4, **step 5**. As a starting point, incubation at room temperature for 1–4 h is suggested (e.g., 0, 30, 60, 90, 120, 240 min).
2. Dry the samples in a centrifugal evaporator.
3. Dissolve the samples in 20 μL dichloromethane–methanol (9:1 (v/v)) and spot onto a silica TLC plate. Include the succinimide standard as a reference.
4. Allow the solvent to completely evaporate off and develop the TLC using dichloromethane–methanol (9:1 (v/v)) as the mobile phase (*see* **Note 15**).
5. Dip the developed plate into the permanganate solution for staining. Dry completely using a heat gun.
6. The optimal incubation time is determined by the intensity of the succinimide spot (in the same position as the standard). An increase in intensity of the succinimide spot should be observed after time. Ideally, a small amount of succinimide compared to sample spots should be observed in the blocking step. Choose a time at which the signal is stable and no significant increase is observed afterwards.

3.6 Immobilization of Pulldown Probes Bearing a Biotin Group

Alternative to Subheadings 3.4 and 3.5, a non-covalent binding between the probe and the magnetic matrix can be used for immobilization, taking advantage from the high affinity interaction between streptavidin and biotin. For this purpose, a biotin-labeled probe is required. Streptavidin magnetic beads are handled similarly to the NHS-magnetic matrix described above (*see* Subheading 3.4) using a magnetic separator (*see* **Note 11**).

1. Dilute the biotinylated pulldown probe and the control probe to a concentration of 10 μM in PBS to a total volume of 1 mL.
2. Homogenize the streptavidin magnetic beads by gently mixing and transfer 250 μL each of the magnetic slurry to four low binding tubes (*see* **Note 11**).

3. Remove the supernatant of the beads and wash at least once with PBS.
4. Add 500 μL of the compound solution (from Subheading 3.6, **step 1**). Prepare two tubes each with the active probe and the control probe (total four samples).
5. Incubate the samples with overhead rotation at least 60 min at room temperature (or overnight at 4 °C, *see* **Note 13**).
6. Collect the beads after compound immobilization and wash at least once with PBS and once with lysis buffer. Continue with Subheading 3.7.

3.7 Enrichment of Proteins That Bind to Immobilized Probes (Pulldown)

1. Prepare samples of 1 mL of each lysate (light and heavy) with the same protein concentration (a protein concentration of at least 1 mg/mL is highly recommended) (*see* **Note 16**).
2. Add 500 μL of the light lysate from Subheading 3.7, **step 1** to beads with the immobilized active probe (active probe/light).
3. Add 500 μL of the light lysate from Subheading 3.7, **step 1** to beads with the immobilized control probe (control probe/light).
4. Add 500 μL of the heavy lysate from Subheading 3.7, **step 1** to beads with the immobilized active probe (active probe/ heavy).
5. Add 500 μL of the heavy lysate from Subheading 3.7, **step 1** to beads with the immobilized control probe (control probe/ heavy).
6. Incubate the four samples from Subheading 3.7, **steps 2–5** with overhead rotation for 2 h at 4 °C.
7. After incubation, remove the supernatant, add 500 μL of lysis buffer and combine beads as follows: active probe/light with control probe/heavy (labeled as A/L), and active probe/heavy with control probe/light (labeled as A/H). Discard the supernatants.
8. Wash the beads twice with lysis buffer at 4 °C for 10 min with overhead rotation.
9. Wash the beads with lysis buffer containing 75 mM $MgCl_2$ for 10 min at 4 °C with overhead rotation to increase the ionic strength and thus reduce nonspecific binding (*see* **Note 17**).
10. Wash the beads twice with PBS for 10 min at 4 °C with overhead rotation to remove detergents (only for on-bead tryptic digest, *see* Subheading 3.8.2). For in-gel digestion (*see* Subheading 3.8.1), lysis buffer can be used instead (*see* **Note 18**).

3.8 Tryptic Digestion of Proteins

After affinity purification of proteins using the functionalized beads, bound proteins need to be digested for further LC/MS analysis. Two different strategies are presented and their advantages are briefly discussed.

3.8.1 In-Gel Tryptic Digestion

In-gel tryptic digestion requires the separation of proteins on SDS PAGE and therefore minimizes the amount of detergents and other contaminants from previous steps that might interfere with the LC/MS analysis.

1. Boil the beads in 25–40 μL of 1× SDS sample buffer. Collect the beads and transfer the supernatant to a new tube (*see* **Note 19**).
2. Load the boiled supernatants on a pre-cast 4–20 % gradient gel. The whole samples must be loaded.
3. Load the boiled BSA control (5 pmol) as control for the mass analysis (*see* **Note 8**).
4. Perform electrophoresis in SDS-PAGE Running Buffer until the bromophenol blue front has reached the bottom of the gel.
5. Stain the gel using Colloidal Coomassie.
6. Wash the gel at least twice with water.
7. Slice each lane of the gel in 4–6 pieces with a clean scalpel. Cut the same regions for A/L and A/H. For the BSA control, only the BSA band is required. Slice an additional band of gel without sample as a control.
8. Cut the excised bands into cubes with a size of approx. 2×2 mm. Transfer gel pieces into protein low binding tubes and spin them down in a bench-top centrifuge.
9. Add at least 200 μL gel washing solution A and ensure that all gel pieces are submerged and incubate for 30 min at 37 °C with gentle mixing in a thermo-shaker by planar orbital motion at 350 rpm.
10. Remove gel washing solution A and add at least 200 μL gel washing solution B. Incubate for 15 min in a thermo-shaker at 37 °C and 350 rpm.
11. Repeat Subheading 3.8.1, **steps 9** and **10** until the gel is completely destained.
12. Incubate the gel pieces with 100 μL of reducing solution in a thermo-shaker for 45 min at 37 °C and 350 rpm.
13. Remove the supernatant and incubate the samples with 100 μL of alkylating solution in a thermo-shaker for 60 min in the dark at 37 °C and 350 rpm.
14. Wash the gel pieces twice by mixing with at least 200 μL gel washing solution B in a thermo-shaker for 15 min at 37 °C and 350 rpm.
15. Add at least 100 μL neat acetonitrile until all pieces are submerged and incubate at room temperature until gel pieces are dehydrated (i.e., gel pieces shrink). Usually this step takes 10 min.
16. Remove all liquid and let the pieces dry. Tubes should be opened only in a sterile environment, e.g., laminar flow hood.

17. Cover the dry gel pieces with at least 50 μL tryptic in-gel digestion solution and incubate in a thermo-shaker for 15 min at room temperature and 350 rpm. During this time, the solution is absorbed. After the incubation time, add more trypsin buffer to the samples, in which gel pieces remain shrunk.
18. Add 25 mM NH_4HCO_3 to yield a trypsin concentration of 3.75 ng/μL. All gel pieces should be covered. Incubate the samples in a thermo-shaker overnight at 30 °C and 350 rpm.
19. Sonicate the samples in a supersonic bath for 30 min at 4 °C. Transfer the supernatant into a new protein low binding tube.
20. Add 100 μL of neat acetonitrile to the gel pieces and incubate at room temperature until gel pieces are dehydrated. Combine this supernatant with the supernatant from Subheading 3.8.1, **step 19** and let it dry in the centrifugal evaporator at maximum 30 °C.

3.8.2 On-Bead Tryptic Digestion and StageTip Purification

On-bead tryptic digest provides several advantages over the in-gel digest. It shortens the number of steps in the protocol. It increases the reproducibility since the variability derived from slicing differences is omitted and decreases the risk of contamination (e.g., keratins). Nevertheless, pre-digest may be necessary for the identification of some target proteins (*see* **Note 20**).

1. Prepare 5 pmol BSA control (*see* **Note 8**). Take 6.64 μL of the BSA solution (50 mg/L) and treat it like the samples.
2. Add 100 μL of solution E1 to the three samples (A/L, A/H, BSA). Incubate for 1 h at room temperature with overhead rotation. Transfer the supernatants of the bead samples into new low binding tubes.
3. Add 100 μL of solution E2 to the beads and the BSA control mixture. Resuspend properly with overhead rotation for 2 min.
4. Mix the resulting supernatants of the bead samples (solution E2 from Subheading 3.8.2, **step 3**) with the supernatant from Subheading 3.8.2, **step 2** (solution E1). Incubate overnight at room temperature in a thermo-shaker at 37 °C and 350 rpm.
5. Stop the tryptic digestion by adding 2 μL of TFA.
6. Place one extraction disk C18 on top of a second one (e.g., in a petri dish) [7].
7. Press the blunt ended syringe needle into the disks to core out a piece of the filter material.
8. Insert the C18 material into the narrowest part of a 200 μL pipette tip (the end of the pipette tip) by gently pushing the cored disk pieces inside. The resultant tip is the so-called C18-StageTip microcolumn (StageTip). Prepare three StageTips, one for each sample (*see* **Note 21**).

9. Install each StageTip into the microcentrifuge tip adapter and put it on the top of a 1.5 mL microcentrifuge tube. Alternatively, when using the StageTip centrifuge, StageTips can be loaded directly into the centrifuge.
10. Activate the reversed phase by adding 100 μL methanol. The solvent will pass slowly through the StageTips (1–3 min). Repeat this step once.
11. Equilibrate the StageTips by incubation with 100 μL wash solution for 2–3 min. Centrifuge at 2,350 × *g* for 2–3 min (*see* **Note 22**) to let the wash solution flow through. Repeat the procedure once.
12. Load 100 μL of sample into the StageTips and incubate for 1 min. Centrifuge the StageTips at 2,350 × *g* for 4–5 min (*see* **Note 22**) to let the loaded liquid flow through. Load the remaining 100 μL of the sample into the StageTips following the same procedure.
13. Desalt the samples by adding 100 μL of wash solution to the StageTips and centrifuge the loaded StageTips at 2,350 × *g* for 5–10 min.
14. Install the StageTips into new protein low binding tubes to collect the eluted peptides. For the following step, the StageTip centrifuge cannot be used since the eluted liquid should be collected.
15. Add 20 μL of elution solution to the StageTip and incubate for 1 min. Centrifuge at 2,350 × *g* for 2–3 min and repeat the procedure (in the same protein low binding tube).
16. Dry the collected elution part in a centrifugal evaporator at 30 °C. Dried samples from tryptic digests can be stored for at least 1 year at −20 °C prior to analysis (*see* **Note 23**).

3.9 LC-MS/MS-Based Analysis of Potential Target Proteins

It might be useful to consult an expert in mass spectrometry or a representative of a mass spectrometry facility prior to analysis if the reader is not familiar with these methods or does not have the expertise or instrumentation to complete the protocol.

3.9.1 Sample Preparation

1. Thaw the samples from Subheading 3.8.1, **step 20** or Subheading 3.8.2, **step 16** shortly before analysis.
2. Dissolve samples in 20 μL of 0.1 % TFA in water.
3. Sonicate samples at room temperature for 15 min to dissolve the peptides as complete as possible.
4. Centrifuge samples at 15,000 × *g* for 1 min.
5. Transfer the supernatant to autosampler vials.

3.9.2 Nano-HPLC-MS/MS Analysis

Suitable nano-HPLC-MS/MS methods are typical shot-gun proteomics methods balancing the time used for full scan MS of precursor ions for subsequent relative quantification of the SILAC

ratios and the time used for MS/MS fragmentation for subsequent identification of proteins.

1. Use the following typical conditions for reversed-phase nano-LC separations: Injection of 3 μL of sample onto a pre-column cartridge using 0.1 % TFA in water as eluent with a flow rate of 30 μL/min, desalting for 5 min with eluent flow to waste followed by back-flushing of the sample from the pre-column to the nano-HPLC column during the whole analysis using 0.1 % formic acid in water as solvent A and 0.1 % formic acid in acetonitrile as solvent B performing a gradient starting at 5 % B with a final composition of 30 % B after 95 min using a flow rate of 300 nL/min. After the separation the nano-HPLC column and the pre-column should be washed, e.g., by increasing the percentage of solvent B to 60 % in 5 min and to 95 % in additional 5 min, washing the columns for further 5 min, flushing back to starting conditions and equilibration of the system for 14 min, column temperature 40 °C (*see* **Notes 24** and **25**).
2. Use the following typical conditions for MS/MS analysis using a Q-Exactive mass spectrometer: 2.1 kV spray voltage; mass range of *m/z* 300–1,650; resolution of 70,000 for full scan MS performing one micro-scan using an automated gain control (AGC) target of 3e6 and a maximum injection time (IT) of 20 ms followed by up to ten high energy collision dissociation (HCD) MS/MS scans of the most intense at least doubly charged ions (e.g., resolution 17,500, AGC target 1e5, IT 120 ms, isolation window 3.0 *m/z*, normalized collision energy (NCE) 25.0, underfill ratio 0.1 %, intensity threshold 8.3e2, dynamic exclusion 20.0 s) recording profile spectra for full scan and fragmentation spectra as well.

3.9.3 Data Evaluation and Statistical Analysis

1. Perform data evaluation using identification and quantification software. Software packages can be freely available (e.g., "MaxQuant" [5]) or commercially available like "Proteome Discoverer." "MaxQuant" uses the implemented search algorithm "Andromeda," whereas other software packages use search algorithms like "Mowse" (Matrix Science [8]) or "Sequest" [9]. Use ion chromatograms resulting from light and heavy labeled tryptic peptides for relative quantification of heavy to light ratios and fragmentation spectra for identification of the proteins. Search the experimental data against a protein database in fasta format. Use the following typical search parameters: no labeling for the light version of the peptides, [$^{13}C_6{}^{15}N_4$] labeling of arginine and [$^{13}C_6$] labeling of lysine for the heavy version of the peptides, oxidation of methionine and N-terminal acetylation of proteins as variable modifications, carbamidomethylation as fixed modification, max. number of modifications per peptide 5, max. charge 7, min. peptide length 7, false discovery rate 1 % on peptide and on protein level searching a

decoy database for discovery of false discovery rate, specific tryptic peptides allowing max. two missed cleavages, mass accuracy for full mass spectra 5 ppm, mass accuracy for MS/MS spectra 20 ppm (most of these parameters are default parameters set in MaxQuant Vers. 1.4.1.2]).

2. Perform statistical evaluation of the search results again using either freely available software like "Perseus" [10] or commercially available software like "Proteome Discoverer." Perform the following typical data processing steps: filtering off proteins which are identified with less than two peptides, filtering off quantifications with less than two ratios, logarithmizing protein ratios, looking for statistic outliers (e.g., using significance A or B implemented in "Perseus" [10]), comparing replicates and reverse replicates (ratios significantly >1 in a forward experiment should be significantly <1 in the reverse replicate for potential target proteins) (*see* **Notes 26** and **27**).

4 Notes

1. Like all pulldown procedures the described method is able to identify potential target proteins which are soluble. The expression level of the proteins has to be high enough for detection upon affinity enrichment. It is more challenging to identify insoluble proteins like membrane proteins. Very harsh solubilizing methods like using high amounts of SDS will denature the protein, so that the affinity enrichment procedure will fail. In these cases the use of trifunctional probes is recommended. For this, a photoreactive group is attached to the small molecules which enables the covalent binding to the target proteins upon irradiation. In addition, an alkyne handle is attached which allows for the covalent linkage to biotin (i.e., biotin-azide) upon binding to target proteins. The principle of this approach is employed in activity-based protein profiling [11].
2. For relative quantification of protein ratios from more or less complex samples based on ion currents of precursor ions, a high-resolution mass spectrometer is absolutely necessary. Otherwise overlaps with other peptides with a similar precursor mass will lead to wrong results.
3. For upscaling, prepare in parallel additional 10 cm diameter cell culture dishes with light or heavy medium, respectively.
4. Determine the cell number required for the metabolic labeling. For each cell line, the proper number of cells to be seeded has to be determined.
5. The composition of the lysis buffer can vary. Since usually no information on the target proteins is available, we use the mentioned lysis buffer. Other common buffers are HEPES, MOPS

and Tris. Furthermore, different detergents can be used to solubilize the membrane proteins amongst other hydrophobic proteins. Important parameters to consider are the detergent-to-protein ratio and the critical micelle concentration (cmc), which should not be exceeded. In addition, polyols can stabilize solubilized membrane proteins (typically, 5 % (v/v) glycerol). Other parameters to consider are the addition of reducing agent, salts and chelators (EGTA is a good alternative when $MgCl_2$ is also added).

6. We recommend to generate a pool of different batches of light or heavy lysates, respectively, which is sufficient to perform a series of affinity chromatography experiments. The use of pooled lysates increases the reproducibility of protein identification.
7. Some cell types are able to convert arginine to proline. In those cases partially heavy labeled proline is incorporated into the proteins in addition to completely labeled arginine. To prevent wrong quantifications of proline containing peptides different approaches, like decreasing the arginine concentration in the medium [12], addition of proline [13], using [$^{15}N_4$]-arginine in combination with normal lysine for the light medium [14], or mathematical corrections [15] are described in the literature.
8. The BSA control is used to assess the efficiency of the tryptic digest. The sequence coverage of BSA after digestion should be in a certain range depending on the instrument and the employed methods (e.g., 70–90 % on a Q-Exactive after in solution digest).
9. Self-made gels can be used as well. However, according to our experience, contamination of the samples with polyethylene glycols might occur when using self-made gels. To avoid variations, we recommend using pre-cast gels. Moreover, gradient gels will ensure proper separation of small and large proteins.
10. Due to the mixing of the samples for the active and control probe upon enrichment of the binding proteins, it is not possible to compare and thus detect differences in the appearance of the protein bands for both probes upon protein staining of the gel. The main purpose of the staining is to have reference lines and thus to cut all samples' lanes in the same way. Different protein staining methods that are compatible with mass spectrometric analysis of proteins can be used as well (e.g., some silver staining protocols, staining with fluorescent dyes).
11. Each liquid removal step is performed using the magnet bar in the tube rack. Before liquid addition, the magnet bar is removed. Then, beads are resuspended by repeated inversion of the rack (and further incubation, if needed). Further details about handling of the beads are described in the manufacturer's protocol.
12. Do not use primary amine-based buffers since they will interfere with the coupling of the pulldown probe to the NHS-activated matrix.

13. The coupling time may vary and should be assessed during determination of the coupling efficiency. Overnight incubations can be additionally tested. However the designed amino linker has to be large enough to minimize steric effects that could interfere with the kinetics of the reaction and more importantly with the pulldown of the target.
14. Alternatively, quantification by HPLC-UV can be performed. A calibration curve of the probe using an internal standard would be needed to determine the amount of coupled probe. Column and elution conditions should be chosen according to the nature of the probe. This method can be also applied to biotinylated probes.
15. Conditions to elute during TLC are optimal for the detection of the succinimide spot. However, if succinimide's and probes' spots are very close, an optimization of the eluents is required. This can be previously tested by running TLC repeatedly with gradual polarity changes (be aware of the limits, e.g., not recommendable to go below dichloromethane–methanol ratio of 8:1). Since the increase of polarity is limited, chloroform can be used alternatively (9:1, 8:2 (v/v)), or additional 0.1 % triethylamine can be added.
16. It is advisable to test different protein–compound ratios.
17. The influence of the increasing ionic strength can be analyzed using SDS-PAGE of the washing fractions. Washing steps at low ionic strength might be insufficient to remove nonspecific binders. However, washing buffers with high ionic strength might remove the target proteins from the beads.
18. Removal of detergents is essential prior to subjecting the samples to MS analysis. For the on-bead tryptic digestion, it is mandatory to perform the last washing steps with PBS to remove detergents. In the in-gel digestion approach, washing with PBS is not needed since during the SDS-PAGE detergents are usually separated from proteins.
19. Elution of enriched proteins using an excess of unmodified compound is an alternative approach and represents a second affinity step that could be explored. However, it might be very hard to displace the target protein from the beads when using unmodified compound. Moreover, it is hard to predict the optimal concentration of free compound to use without a previous knowledge of the targets and their abundance in the cells.
20. It has been described that trypsin might not be "powerful" enough to degrade some proteins under this conditions. Optional, pre-digest using Lysyl Endopeptidase (Lys C) can be performed which can decrease the occurrence of missed cleavages and increase the number of identified peptides [16].

21. The consistency of the reversed phase can be evaluated in the activation step. It should be neither overpacked (so that liquid cannot pass through it), nor feathery (so that liquid leaks without interaction with the reversed phase). A gap of several millimeters should be visible between the disk and the end of the tip.
22. The centrifugation time is adaptable. Ensure that the solvent has flowed through the StageTip. Increase centrifuge time if necessary. Do not completely dry the material.
23. If samples should not be stored after sample preparation, they can also be evaporated just nearly to dryness, diluted with 20 μL of 0.1 % TFA in water, transferred into autosampler vials and injected directly without any sonification and centrifugation.
24. If a nano-HPLC system without any on-line desalting possibilities is used, samples should be desalted before analysis (e.g., using StageTips [7]) in each case.
25. Washing of a pre-column system and therefore reduction of carryover of sample from one injection to the next can be optimized using a system that allows working with two pre-column cartridges. During the analysis of one sample the other pre-column can be washed very efficiently, e.g., by performing three very steep gradients from 0 to 95 % B using 0.1 % TFA in water as solvent A and 0.1 % TFA in acetonitrile as solvent B at a flow rate of 30 μL/min).
26. For protein identification and quantification it is necessary to decide which peptides should be taken into account, i.e., only peptides that are unique for a special protein, protein group or all peptides (in addition to unique ones also those which have identical sequences in other proteins or protein groups). A third and perhaps rather elegant method is to work with "razor + unique" peptides as performed by MaxQuant [5]. In this approach nonunique peptides are counted for the best identified protein or protein group. This method is based on the principle of Occam's razor.
27. For data evaluation of at least three replicates of a forward and three replicates of a reverse experiment it is also possible to use a *t*-test to search for proteins which behave differently in the two groups. A graphical overview of such a statistical test will be given by a Vulcano-pot.

Acknowledgement

This work was supported by the Max-Planck-Gesellschaft and the European Union (funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no. 268309).

References

1. Ziegler S et al (2013) Target identification for small bioactive molecules: finding the needle in the haystack. Angew Chem Int Ed Engl 52:2744–2792
2. Yan W, Chen SS (2005) Mass spectrometry-based quantitative proteomic profiling. Brief Funct Genomic Proteomic 4:27–38
3. Ong SE et al (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1:376–386
4. Ong S-E et al (2012) Identifiying cellular targets of small-molecule probes and drugs with biochemical enrichment and SILAC. Methods Mol Biol 803:129–140
5. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372
6. Neuhoff V et al (1988) Improved staining of proteins in polyacrylamide gels including isoelectric-focusing gels with clear background at nanogram sensitivity using Coomassie brilliant blue G-250 and R-250. Electrophoresis 9:255–262
7. Rappsilber J et al (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2:1896–1906
8. Perkins DN et al (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20:3551–3567
9. Eng JK et al (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 5: 976–989
10. Cox J, Mann M (2012) 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. BMC Bioinformatics 13(Suppl 16):S12
11. Nodwell MB, Sieber SA (2012) ABPP methodology: introduction and overview. Top Curr Chem 324:1–41
12. Blagoev B, Mann M (2006) Quantitative proteomics to study mitogen-activated protein kinases. Methods 40:243–250
13. Bendall SC et al (2008) Prevention of amino acid conversion in SILAC experiments with embryonic stem cells. Mol Cell Proteomics 7:1587–1597
14. Van Hoof D et al (2007) An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. Nat Methods 4:677–678
15. Park SK et al (2009) A computational approach to correct arginine-to-proline conversion in quantitative proteomics. Nat Methods 6: 184–185
16. Glatter T et al (2012) Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. J Proteome Res 11: 5145–5156

Chapter 22

Drug Affinity Responsive Target Stability (DARTS) for Small-Molecule Target Identification

Melody Y. Pai*, Brett Lomenick*, Heejun Hwang, Robert Schiestl, William McBride, Joseph A. Loo, and Jing Huang

Abstract

Drug affinity responsive target stability (DARTS) is a relatively quick and straightforward approach to identify potential protein targets for small molecules. It relies on the protection against proteolysis conferred on the target protein by interaction with a small molecule. The greatest advantage of this method is being able to use the native small molecule without having to immobilize or modify it (e.g., by incorporation of biotin, fluorescent, radioisotope, or photoaffinity labels). Here we describe in detail the protocol for performing unbiased DARTS with complex protein lysates to identify binding targets of small molecules and for using DARTS-Western blotting to test, screen, or validate potential small-molecule targets. Although the ideas have mainly been developed from studying molecules in areas of biology that are currently of interest to us and our collaborators, the general principles should be applicable to the analysis of all molecules in nature.

Key words Small molecules, Drugs, Target identification, Metabolites, Natural products, Proteomics, Mass spectrometry, Immunoblotting

1 Introduction

Small-molecule target identification is a critical aspect of chemical genetics, metabolomics, and drug discovery [1–4]. A variety of methods have been developed for small-molecule target identification, with affinity chromatography being the most commonly used approach [5–7]. However, affinity chromatography and related approaches are limited by the need to derivatize each small molecule, and many compounds cannot be modified without loss of binding specificity or affinity. On the other hand, genetic/genomic methods are limited to particular classes of compounds (e.g., those that affect fitness, transcription, localization, etc.) and, because they

*These authors are contributed equally.

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_22, © Springer Science+Business Media New York 2015

rely on downstream readouts, do not necessarily pinpoint the direct targets [5, 8]. These limitations have spurred the continual development of new and improved methods. Drug affinity responsive target stability (DARTS) is a paradigm-changing method developed to overcome these limitations. DARTS leverages the thermodynamic stabilization of the target protein that occurs upon small-molecule binding by detecting the binding-induced increase in resistance to proteolysis [9]. This is highly advantageous because it uses the native, unmodified small molecules and relies solely on the binding interaction but not downstream readouts to discover target proteins.

DARTS is a relatively simple technique that can easily be adopted by most labs. Unlike affinity chromatography, DARTS is not limited by the chemistry of the small molecule of interest and does not require derivatization or immobilization of the compound. Rather, DARTS is performed by simply treating aliquots of cell lysate with the compound of interest and either vehicle control or an inactive analog, followed by limited digestion of the proteins in the cell lysate with proteases. Subsequently, the samples are separated by SDS-PAGE and stained to identify protein bands that are protected from proteolysis by the small molecule. Mass spectrometry (MS) is then used to identify the proteins present in each band. This unbiased DARTS approach has been successfully utilized to identify novel protein targets for natural products and other bioactive small molecules; *see* [10, 11] for recent examples identifying a novel protein target for disulfiram, an FDA-approved drug used to treat chronic alcoholism, and a novel protein target for the metabolite α-ketoglutarate. Although this gel-based approach is the easiest to implement, more efficient gel-free proteomics approaches are also being used with DARTS to facilitate identification of the protected proteins [5, 12].

While DARTS has been successfully performed in an unbiased fashion as a discovery tool to identify unknown targets of natural products and drugs (*see* [9–11, 13, 14] for some examples), it is also powerful as a means to screen or validate binding of compounds to proteins of interest. This targeted approach has been widely used with recombinant and/or purified proteins using gel staining, endogenous proteins in lysates using Western blotting, and epitope-tagged proteins expressed in cells or in vitro and detected with epitope-specific antibodies [9–11, 15–20]. Moreover, the targeted approach could be used for high-throughput screening for compounds that bind a specific protein. Here we describe examples using DARTS to assay additional small molecule-protein interactions, including two model drug-protein pairs, methotrexate-DHFR and olaparib-PARP [21], as well as omigapil (CGP3466B)-GAPDH, which has been suggested to be protective against motor neuron apoptosis [22].

2 Materials

2.1 DARTS Materials

1. Phosphate-buffered saline (PBS).
2. Protease inhibitor cocktail (20×): Dilute one tablet of protease inhibitor cocktail (Roche) with 525 μL of ultrapure water to make 20× concentration. Mix to fully dissolve the tablet, and store at −20 °C (*see* **Note 1**).
3. Lysis buffer: For 1 mL of mammalian protein extraction lysis buffer, mix 50 μL of 20× protease inhibitor cocktail (Roche), 50 μL of 1 M sodium fluoride, 100 μL of 100 mM β-glycerophosphate, 100 μL of 50 mM sodium pyrophosphate, and 10 μL of 200 mM sodium orthovanadate with 690 μL of M-PER reagent (M-PER, Thermo Scientific) (*see* **Note 2**). Once the lysis buffer is made, keep it on ice. Make fresh lysis buffer for every DARTS experiment.
4. TNC buffer (10×): For 1 mL of 10× TNC buffer, mix 500 μL of 1 M Tris–HCl pH 8.0, 100 μL of 5 M sodium chloride, and 100 μL of 1 M calcium chloride with 300 μL of ultrapure water (*see* **Note 3**). Once 10× TNC buffer is made, keep it on ice. Store aliquots at −20 °C.
5. BCA protein concentration assay reagents (other protein concentration assays such as Bradford can be used instead). Bovine serum albumin can be used for the standard.
6. Small molecule: Dilute compounds in appropriate solvent and store accordingly in glass vials (*see* **Note 4**).
7. Pronase: Prepare a 10 mg/mL stock solution of Pronase (Roche) in ultrapure water, aliquot, and store at −20 °C.
8. Thermolysin: Prepare a 10 mg/mL stock solution of Thermolysin (Sigma) in 1× TNC buffer, aliquot, and store at −20 °C (proteases from other suppliers should also work, but they may require different amounts than we describe herein).
9. SDS-PAGE loading buffer (5×): For 50 mL of SDS-PAGE loading buffer, mix 12.5 mL of 1 M Tris–HCl (pH 6.8), 25 mL of 100 % glycerol, 5 g of sodium dodecyl sulfate, 0.25 g of bromophenol blue, and 2.5 mL of 14.3 M β-mercaptoethanol with 10 mL of ultrapure water. Aliquot and store at −20 °C.

2.2 SDS-PAGE, Visualization, and Mass Spectrometry Materials

1. SDS-PAGE gel (*see* **Note 5**).
2. Silver staining kit (must be MS compatible, such as from Sigma Aldrich), SimplyBlue stain (Invitrogen), or SYPRO Ruby stain (Invitrogen) for visualization.
3. Mass spectrometry materials (*see* **Note 6**).
4. Western blotting materials.
5. Antibody for potential small-molecule target and control protein.

3 Methods

3.1 DARTS with Complex Protein Lysate

1. Grow the cells to approximately 80–85 % confluence (*see* **Note 7**).
2. Aspirate media from the plates. Wash the cells with ice-cold phosphate-buffered saline (*see* **Note 8**).
3. Lyse the cells with an appropriate amount of lysis buffer (*see* **Note 9**).
4. Scrape cells off with a cell scraper and collect.
5. Allow lysis of the cells to occur on ice for 10 min (*see* **Note 10**).
6. Centrifuge for 10 min at 18,000 × *g* at 4 °C to pellet cellular debris and DNA.
7. Remove the supernatant (cell lysate) and transfer to a new 1.5 mL tube. Keep the tube on ice.
8. Add an appropriate volume of 10× TNC buffer to make a final concentration of 1× TNC buffer in the lysate (*see* **Note 11**).
9. Perform a BCA protein concentration assay to determine protein concentration of the cell lysate (*see* **Note 12**).
10. Create 100× stock solutions of small molecule via serial dilutions (*see* **Note 13**).
11. Split the cell lysate into identical aliquots of 99 μL (*see* **Note 14**).
12. Add 1 μL of vehicle control (solvent that the small molecule is dissolved in) and various 100× small-molecule stock solutions to each aliquot of lysate. Incubate the cell lysate with small molecule for 15–30 min at room temperature with shaking with a thermomixer (*see* **Note 15)**. Alternatively, samples can be rotated on a rotator.
13. Make protease dilutions in 1× TNC buffer (*see* **Note 16** for protease choice and solution preparation). Be sure to use a fresh aliquot of protease in every experiment.
14. After incubation with the small molecule is complete, split each sample into 20 μL samples (*see* **Note 17**).
15. Add 2 μL of the range of protease solutions (*see* **Note 16** for protease choice) prepared in Subheading 3.1, **step 13**, to achieve the appropriate final ratio of total enzyme to total substrate in each sample. Add the protease solutions at specific intervals (e.g., every 30 s) to ensure that each sample is digested for the same amount of time. Be sure to include a sample that is not digested. For the non-digested sample, add 2 μL of 1× TNC buffer instead of protease.
16. Incubate the samples at room temperature with the protease of choice for an appropriate time (*see* **Note 18**).

17. Stop each digestion reaction by adding 2 μL of 20× protease inhibitor cocktail (at the same specific time intervals as used above) and incubate on ice for 10 min.
18. Add 6 μL of 5× SDS-PAGE loading buffer to the samples to achieve a final 1× SDS-PAGE loading buffer concentration.
19. Heat the samples at 70 °C for 10 min.
20. Spin the samples down briefly with a microfuge and proceed to analysis via SDS-PAGE in Subheading 3.2. At this point, samples may be stored at −20 °C for at least several weeks if SDS-PAGE will not be performed immediately.

3.2 Identification and Validation of Potential Small-Molecule Targets

1. If the samples were frozen, thaw the samples prepared in Subheading 3.1 to room temperature.
2. For each sample, load 10–20 μg protein into each lane of a 10- or 12-well minigel (1.0 mm thickness).
3. Perform electrophoresis at room temperature using a constant voltage of 100 V until the dye front has reached the bottom of the gel (typically 1.5–2 h).
4. Carefully remove the gel from the plastic or glass plates using clean gloves and transfer it into a clean staining tray containing distilled water.
5. Wash the gel for 5 min three times with distilled water by shaking on a flat rotator.
6. Stain the gel to visualize protein samples using a mass spectrometry-compatible silver staining kit, SimplyBlue stain, or SYPRO Ruby stain (*see* **Note 19**). Follow the manufacturer's instructions for staining.
7. Look for bands that appear to be protected by incubation with the small molecule over vehicle control (*see* **Notes 20** and **21**). For example, in Fig. 1, incubation with the small molecule confers protection against proteolysis.
8. Excise gel bands corresponding to the protected proteins with a small, clean scalpel or razor blade and analyze via mass spectrometry using standard protein identification approaches. It is important to include the corresponding region from the vehicle control lane in the analysis because multiple proteins may be identified in the gel band. Quantitative mass spectrometry analysis (e.g., using spectral counting or extracted ion chromatography) is an appropriate approach to determine which protein is enriched in the drug-treated versus control sample. Once the protein(s) present in the protected band is identified, however, whether or not each is protected can be verified by immunoblotting.
9. Once potential small-molecule targets are identified via mass spectrometry, the binding of the small molecule to candidate

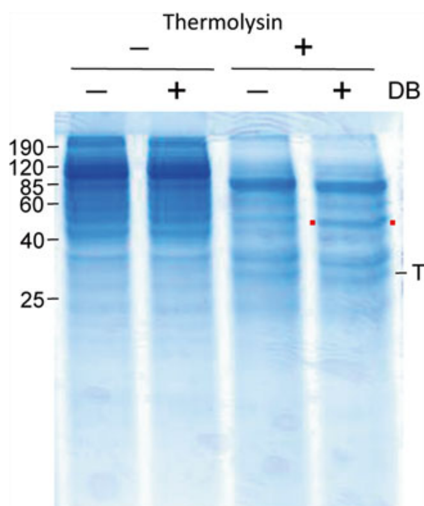


Fig. 1 Example of SimplyBlue staining visualization of unbiased DARTS with the small molecule didemnin B (DB). Red dots flank the protected band; T, thermolysin. Reprinted from [9]

targets can be immediately validated by Western blotting of samples from the unbiased DARTS experiment as well as an independent DARTS experiment using antibodies specific to the candidate target proteins (*see* **Note 22**). If no antibodies are available, then candidate proteins can be produced by in vitro translation or by expression of epitope-tagged proteins in transfected cells (see [9, 15] for examples). This DARTS-Western analysis is also extremely useful for validating any potential target proteins identified using other methods, such as omics profiling [15], pathway screening, in silico docking [20], or other computational predictions.

10. After DARTS, perform SDS-PAGE, transfer the proteins to a membrane suitable for immunoblotting, and blot the membrane with an antibody against the putative protein target as well as at least one control protein (*see* **Note 23**). For example, in Fig. 2, we used three familiar small molecule-protein target pairs to illustrate the use of DARTS for candidate target validation.

4 Notes

1. Protease inhibitor cocktails from other vendors may also work, but the concentrations for each inhibitor vary. The cocktail can also be home assembled to customize specific concentrations if necessary.

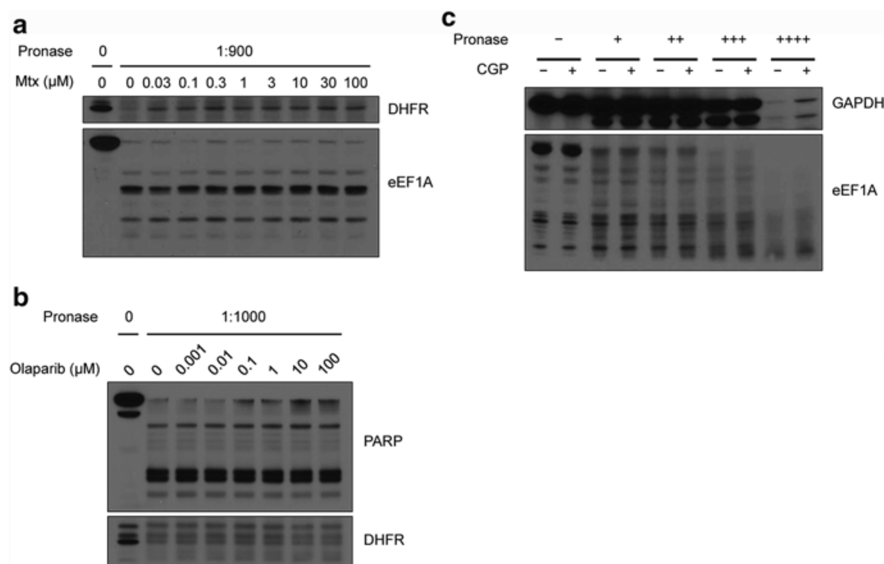


Fig. 2 (**a**) DARTS with methotrexate (Mtx) shows interaction with its known target dihydrofolate reductase (DHFR) but not eukaryotic elongation factor 1 alpha (eEF1A), which serves as a control protein. Jurkat lysates were incubated with varying concentrations of methotrexate or vehicle (in equal volume, with final 1 % DMSO), followed by digestion with 1:900 Pronase:protein ratios for 15 min. The dissociation constant for purified recombinant DHFR is ~10 nM. Its IC_{50} for cell lines varies greatly, and some cells have nM IC_{50} values corresponding to its binding affinity. We found that with ~30 nM of Mtx, there is the same level of protection of DHFR against proteolysis as with ~100 μM of Mtx. (**b**) DARTS with olaparib (O) (IC_{50} ~ 1 nM) confirms its interaction with its known target poly(ADP-ribose) polymerase (PARP), but not DHFR, which is instead the target of Mtx. Performed as in (**a**) using varying concentrations of olaparib or vehicle (in equal volume, with final 1 % DMSO). (**c**) DARTS with CGP 3466B confirms its interaction with GAPDH while eEF1A serves as a control protein. HEK293 cell lysates were incubated with 100 μM CGP 3466B or 1 % DMSO, followed by digestion with 1:1,800, 1:600, 1:400, and 1:200 Pronase:protein ratios for 15 min (*see* **Note 24**). Although CGP was reported to show strong neuroprotective effects at 1 nM [22], it is not clear that this is mediated by GAPDH

2. Other lysis buffers with various detergents (e.g., Triton X-100 or NP-40) can be used with DARTS as long as they are non-denaturing.
3. If the lysis buffer used includes any type of buffering agent (e.g., Tris or HEPES) and sodium chloride or another salt (such as potassium chloride), 10× TNC buffer is not necessary.
4. Small molecules should be stored in glass vials to avoid loss due to potential absorption by plastic tubes. This may result in a drastically lower concentration of certain compounds than intended.
5. When performing unbiased DARTS, a 4–12 % Bis–Tris gradient gel can first be used to separate the protein samples. Once potential protein targets are identified, depending on the molecular weight of those targets, a gel that best separates either small- or large-molecular-weight proteins can be used if necessary.

6. For new users, it is highly advised to collaborate with researchers who have expertise in mass spectrometry and MS-based proteomics.
7. DARTS can be performed with any cell line, tissue, or organism as the protein source, so long as the protein extraction method is non-denaturing. Usually a cell line that is sensitive to the bioactivity of the small molecule is used for target ID. The number of cells needed for each DARTS experiment will vary based on how much protein can be extracted from various cell lines. In general, the protein concentration of the lysate used is between 2.5 and 5 μg/μL. In one DARTS experiment with the natural product didemnin B (DB) where we tested Jurkat lysates at 1 and at 5 μg/μL, the protection of EF1A1 was more apparent in the 5 μg/μL DB-treated lysate. However, plenty of experiments by others have used lower concentrations, around 2–4 μg/μL, that work just as well for other compounds. We have not tested using substantially higher protein concentrations.
8. Make sure to remove all media, especially those that contain fetal bovine serum (FBS) as proteins in FBS may interfere with the protein concentration assay and downstream protease concentration calculations.
9. Use less lysis buffer for a more concentrated protein lysate. One 10 cm plate of HEK293 cells at 85–90 % confluency lysed with 600 μL of lysis buffer typically results in a protein lysate of ~2.5 μg/μL.
10. Be sure not to vortex the protein lysate as this may disrupt the native conformation of some proteins and alter or abolish their ligand-binding activity.
11. In our experience, when the 10× TNC buffer is added to M-PER lysis buffer, the lysate will become slightly cloudy. Again, if an alternate lysis buffer that includes a buffer and salt is used, 10× TNC buffer is not needed (*see* **Note 2**).
12. Any sufficiently sensitive protein concentration assay (e.g., Bradford) can be used to determine the protein concentration of the lysate. If the protein concentration of the lysate is less than 2 μg/μL, we recommend repeating the protein extraction using either more cells or less lysis buffer or concentrating the protein lysate (*see* **Note 7**). A protein concentration of 2 μg/μL in each 20 μL sample will provide sufficient protein for two Western blots if loading 20 μg per well.
13. If the small molecule is stored at 4 °C or −20 °C, make sure to allow the vials to warm up to room temperature before opening to avoid condensation and ensure that the weighing of the compound is accurate. Weigh enough of the small molecule to make a beginning stock concentration of 100 mM (or lower

depending on maximum solubility). From there, make serial dilutions from the beginning stock to create 100× stock solutions. When performing unbiased DARTS, one may begin with a higher concentration of the small molecule (5–10× the IC_{50} value for the biological/phenotypic activity of interest) to ensure optimal binding. Although this could also potentially increase the number of binding proteins identified, it may also suggest additional targets and new use of the drug. Additionally, one may begin testing concentrations near the IC_{50} of the compound to minimize identification of off-targets, and only subsequently testing higher doses if necessary. On the other hand, if the IC_{50} is unknown or there is no known specific bioactivity for the compound, then an initial high dose of up to 250 μM of the small molecule may be tested.

14. The number of aliquots of protein lysate needed depends on the number of small-molecule concentrations that are going to be tested. When performing unbiased DARTS, begin with one or two concentrations of the small molecule (*see* **Note 13** for choosing concentrations). Once a candidate target protein is determined, additional concentrations of the small molecule can be used to determine relative binding affinity. Remember to include a sample for vehicle control or inactive analog control.
15. The time required for small molecule-lysate incubation can vary. While most binding equilibria are reached in seconds, we generally incubate for at least 15–30 min to ensure optimal binding.
16. For DARTS, we recommend the proteases thermolysin and Pronase because they have been used successfully by us and others for numerous different compounds. See [5, 9, 12] for more information on choosing a protease to use. While other proteases may work equally well, we have not substantially explored alternatives as it has not been necessary. To begin, test a range of protease concentrations (e.g., from 1:100 to 1:1,000 Pronase:protein ratio) to ensure that the potential small-molecule target is neither completely digested or not digested enough. Protease concentrations can be adjusted if the proteome is over- or under-digested. To calculate protease concentrations (example):

    2.5 μg/μL protein concentration × 20 μL sample = 50 μg protein

    For a protease concentration of 1:100 protease:protein:

    50 μg ÷ 100 ÷ 2 μL = 0.25 μg/μL protease concentration needed

    2.5 μL 10 μg/μL stock protease in 97.5 μL 1× TNC buffer
17. Just prior to this, if the salt from the TNC buffer has settled to the bottom of the tube, mix by tapping the tube to ensure that the solution is homogenous.

18. Generally, begin with 20-min digestion times. This can be eventually tailored, if necessary, once potential small-molecule targets are identified. In fact, we have found that some small molecules may provide better protection with shorter digestion times (e.g., 5–10 min).
19. Many staining methods are available for gels. Silver, Simply Blue, and SYPRO Ruby stains have all been used successfully with DARTS and LC-MS/MS analysis, although other methods may also work. When performing staining, be sure that gloves and containers used are clean to prevent contamination by keratin and other environmental proteins in downstream mass spectrometry analysis.
20. If after visualization the entire lane of sample treated with the small molecule seems to be darker than the entire lane of sample treated with vehicle control, either loading is inconsistent between lanes or the small molecule has an effect on the protease used. If the latter is the case, another protease can be used.
21. Once a potential protein target is identified, the protease concentrations used can be tailored (if necessary) to the sensitivity of the potential protein target. From the DARTS experiments with various small molecules that we have tested, we observed that larger proteins (e.g., mTOR, eIF4G, dynein heavy chain) tend to be more sensitive to proteolysis, which may be explained by the increased number of flexible regions across the full length of the protein and/or the increased number of peptide bonds (protease substrates). On the other hand, many small proteins (e.g., FKBP12, eIF4E, GAPDH), especially those consisting of a single domain, are relatively resistant to proteolysis and therefore require more protease or increased digestion time. Regardless of this variability in susceptibility to proteolysis among different proteins, protection of the target protein can be seen across a range of protease concentrations in which the target is partially or fully digested in the vehicle-treated control.
22. There are two types of validation during target identification: binding vs. functional, i.e., validation of the physical binding interaction between the small molecule and the potential target and validation of the potential target as a physiological target. While quantitative mass spectrometry should be able to determine which protein identified is protected from proteolysis, we suggest validating this protection by repeating the DARTS experiment and performing Western blotting, when possible. If a hit identified to be protected against proteolysis via mass spectrometry was not detected as such when analyzed via Western blotting with a specific antibody, it may be due to different sensitivities of mass spectrometry and Western blotting

or the lack of appropriate epitope in the full-length or partially digested candidate target protein. In such a case, using a FLAG-tagged construct is recommended. Furthermore, it is important to functionally validate the protein target not only as interacting with the small molecule but also in some way modifying the protein's activity (e.g., in an in vitro biochemical assay and in vivo biological readout). The binding target identified may not necessarily be the target responsible for the bioactivity of interest of the small molecule, and functional tests must be performed to determine whether or not it is the target of interest. The functional tests used will depend on the bioactivity under study and the binding targets identified for the small molecule, a discussion of which is outside the scope of this chapter.

23. Probing for a control protein is required to show that binding is specific and that the small molecule does not have an inhibitory effect on the protease used. GAPDH, actin, and tubulin are often used as control proteins, although any protein with a similar sensitivity to proteolysis may be used. In addition, to further show that the interaction between the potential protein target and the small molecule is specific, other unrelated small molecules or inactive analogs can be used alongside the small molecule of interest when performing DARTS. If the small-molecule interaction with the protein target is truly specific to the pair, then most other small molecules should not result in protection of the protein target from proteolysis.
24. The DARTS experiments in Fig. 2 were done with both Jurkat and HEK293 cell lysates. Depending on the small molecule under study, the exact cells used for DARTS may be unimportant, as many target proteins are expressed ubiquitously [23–25]. For example, DARTS with a generally cytotoxic drug that has effects in many diverse cell types could be performed with any cell line sensitive to its effects. However, if the small molecule exhibits bioactivity in a specific cell type or under specifically induced conditions (e.g., upon starvation or radiation), we recommend using those cells because the target protein may not be expressed or active in other cell types.

Acknowledgments

Supported by the US National Institutes of Health grants R01 CA124974 (J.H.), R21 CA149774 (J.H.), U19 AI067769 (W.B.), R01 GM103479 (J.A.L.), R01 GM104610 (J.A.L.), and training grants to M.Y.P. (T32 GM007185) and B.L. (T32 CA009120).

References

1. O'Connor CJ, Laraia L, Spring DR (2011) Chemical genetics. Chem Soc Rev 40:4332–4345
2. Yang GX, Li X, Snyder M (2012) Investigating metabolite-protein interactions: an overview of available techniques. Methods 57:459–466
3. McFedries A, Schwaid A, Saghatelian A (2013) Methods for the elucidation of protein-small molecule interactions. Chem Biol 20:667–673
4. Rask-Andersen M, Masuram S, Schioth HB (2014) The druggable genome: evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. Annu Rev Pharmacol Toxicol 54:9–26
5. Lomenick B, Olsen RW, Huang J (2011) Identification of direct protein targets of small molecules. ACS Chem Biol 6:34–46
6. Ong SE et al (2012) Identifying cellular targets of small-molecule probes and drugs with biochemical enrichment and SILAC. Methods Mol Biol 803:129–140
7. Ziegler S et al (2013) Target identification for small bioactive molecules: finding the needle in the haystack. Angew Chem Int Ed Engl 52:2744–2792
8. Futamura Y, Muroi M, Osada H (2013) Target identification of small molecules based on chemical biology approaches. Mol Biosyst 9:897–914
9. Lomenick B et al (2009) Target identification using drug affinity responsive target stability (DARTS). Proc Natl Acad Sci U S A 106: 21984–21989
10. Robinson TJ et al (2013) High-throughput screen identifies disulfiram as a potential therapeutic for triple-negative breast cancer cells: interaction with IQ motif-containing factors. Cell Cycle 12:3013–3024
11. Chin RM et al (2014) The metabolite alpha-ketoglutarate extends lifespan by inhibiting ATP synthase and TOR. Nature. doi:10.1038/nature13264
12. Lomenick B et al (2011) Target identification using drug affinity responsive target stability (DARTS). Curr Protoc Chem Biol 3:163–180
13. Tohda C et al (2012) Diosgenin is an exogenous activator of 1,25D(3)-MARRS/Pdia3/ERp57 and improves Alzheimer's disease pathologies in 5XFAD mice. Sci Rep 2:535
14. Sun W et al (2014) Chemical signatures and new drug targets for gametocytocidal drug development. Sci Rep 4:3743
15. Aghajan M et al (2010) Chemical genetics screen for enhancers of rapamycin identifies a specific inhibitor of an SCF family E3 ubiquitin ligase. Nat Biotechnol 28:738–742
16. Chen T et al (2011) Chemical genetics identify eIF2alpha kinase heme-regulated inhibitor as an anticancer target. Nat Chem Biol 7:610–616
17. Gao S et al (2012) The chemistry and biology of nakiterpiosin—C-nor-D-homosteroids. Synlett 16:2298–2310
18. Xu S et al (2013) Stabilization of MDA-7/IL-24 for colon cancer therapy. Cancer Lett 335:421–430
19. Lim M et al (2014) Ligand-independent and tissue-selective androgen receptor inhibition by pyrvinium. ACS Chem Biol 9:692–702
20. Li H et al (2014) Drug design targeting protein-protein interactions (PPIs) using multiple ligand simultaneous docking (MLSD) and drug repositioning: discovery of raloxifene and bazedoxifene as novel inhibitors of IL-6/GP130 interface. J Med Chem 57:632–641
21. Molina DM et al (2013) Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. Science 341: 84–87
22. Kragten E et al (1998) Glyceraldehyde-3-phosphate dehydrogenase, the putative target of the antiapoptotic compounds CGP 3466 and R-(-)-deprenyl. J Biol Chem 273: 5821–5828
23. Lundberg E et al (2010) Defining the transcriptome and proteome in three functionally different human cell lines. Mol Syst Biol 6:450
24. Geiger T et al (2012) Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Mol Cell Proteomics 11:M111 014050
25. Geiger T et al (2013) Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse. Mol Cell Proteomics 12:1709–1722

Chapter 23

Chemical Genomic Profiling via Barcode Sequencing to Predict Compound Mode of Action

Jeff S. Piotrowski, Scott W. Simpkins, Sheena C. Li, Raamesh Deshpande, Sean J. McIlwain, Irene M. Ong, Chad L. Myers, Charlie Boone, and Raymond J. Andersen

Abstract

Chemical genomics is an unbiased, whole-cell approach to characterizing novel compounds to determine mode of action and cellular target. Our version of this technique is built upon barcoded deletion mutants of *Saccharomyces cerevisiae* and has been adapted to a high-throughput methodology using next-generation sequencing. Here we describe the steps to generate a chemical genomic profile from a compound of interest, and how to use this information to predict molecular mechanism and targets of bioactive compounds.

Key words Chemical genomics, Barcode sequencing, Functional genomics, Yeast deletion collection

1 Introduction

Chemical genomics is a powerful technique for understanding the mode of action and cellular targets for unknown compounds [1–4]. The particular strengths of chemical genomics are that it is a whole-cell assay that is not designed around a target of interest, and it provides an unbiased view of the cellular response to a compound [3]. The technique is based on exposing a large pool of defined gene deletion mutants to a compound and measuring the fitness of the individual mutants [5]. The fitness of these mutants can be measured in a number of ways (e.g., colony size, optical density). Several mutant collections have been created where the gene mutation is replaced with a specific molecular barcode, a short section of DNA with a sequence specific to the mutant [6–8]. In the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, these barcodes are 20 bp sequences flanked by common priming sites that allow amplification of the barcodes. Because each barcode is unique, the mutants can be pooled together, and the relative

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_23, © Springer Science+Business Media New York 2015

fitness of each strain can be determined by the abundance of the mutant-specific barcode, either by microarray or next-generation sequencing of the barcodes [9]. Given that most genes in *S. cerevisiae* have been functionally annotated, the fitness of the mutants in the presence of a compound can give functional insight into the chemical's mode of action. For instance, in the presence of a DNA-damaging agent like methyl methanesulfonate (MMS), mutants in genes involved in DNA damage and repair (e.g., *rad51*Δ) are significantly more sensitive to the compound compared to other mutants [1, 5]. The pool of mutants can also yield valuable knowledge for weakly bioactive compounds, as mutant performance may be dramatic even when the behavior of wild-type (WT) cells appears unaffected [4]. This "chemical genomic profile" of the mutant collection can further be paired with the yeast genetic interaction network to predict the pathways and proteins that the compounds may be directly affecting [10].

Chemical genomic profiling uses a pooled yeast deletion collection that is created by pipetting individual, liquid mutant cultures into a common pool. The pool of yeast mutants is central to all assays, thus careful construction of the pool is essential. Each time a pool is created, the starting distribution of mutants can generate a "pool signature" created by the distributions of mutants in the pool; therefore, it is best to make a large pool to cover more than the number of planned screens, so the mutant pool does not have to be remade (*see* **Note 1**). Chemical genomics can be performed using individual mutants arrayed on agar and measuring colony size to determine fitness, but the compound requirements for such assays are orders of magnitude greater than pooled assays. As novel compounds are often scarcely available, we have focused on optimizing the liquid assay to minimize compound requirements.

The major strength of chemical genomics using barcode sequencing is the ability to multiplex samples and screen many different compound conditions in a single sequencing reaction. Present sequencing technology offers a high number of reads, allowing multiple samples to be multiplexed in a single sequencing lane and later de-multiplexed via sample-specific index tags built into the PCR primers. The maximum degree of multiplexing is determined by the sequencing platform and the size of the mutant pool. When using the entire nonessential, haploid yeast knockout collection (~5,000 strains), a maximum of approximately 25 samples can be pooled and sequenced simultaneously on the Illumina MiSeq platform (using present flow cell design "V2"), while up to 96 or more samples can be sequenced on a single HiSeq 2500 lane and 192 samples in a HiSeq 2500 Rapid Run (two lanes). A single 8-lane Illumina run can accommodate nearly 1,000 chemical genomic screens and potentially much more as sequencing technology rapidly improves. Experiments should be designed with the multiplexing limitations of the sequencing platform in mind.

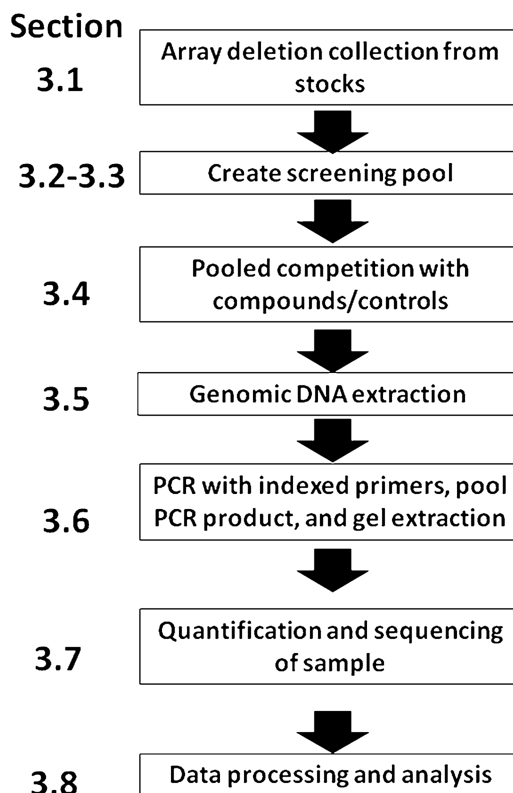


Fig. 1 Overview of the steps of chemical genomic profiling by barcode sequencing

In this chapter we describe the steps (Fig. 1) necessary to perform chemical genomic analysis using the yeast nonessential gene deletion collection. We describe creating the pool of yeast mutants and exposing these to a compound as a pooled competition. We then describe the steps to remove and amplify the molecular barcodes using multiplexed primers and then sequence these using next-generation sequencing. Finally we provide the computational tools and describe the steps to generate a chemical genomic profile and identify sensitive and resistant mutants in the presence of a chemical compound, which yields functional insight into the compound's mode of action.

2 Materials

When working with live cells, always follow proper sterile technique and use a laminar flow hood for all cell transfers. Media and reagents can be stored at 4 °C unless otherwise indicated.

2.1 Arraying the Deletion Collection to Agar

1. YPD + G418 agar: Make 2 L. For each liter of the medium, add 10 g of yeast extract, 20 g of peptone, 20 g of agar, and 950 mL of dH_2O and autoclave. To the cooled medium add 50 mL of

40 % glucose (40 g of glucose dissolved in 100 mL of H_2O and autoclaved) to make 1 L of 2 % glucose YPD. Once cooled to 50 °C, add 1 mL of 1,000× G418 (Geneticin, 200 mg/mL in dH_2O, filter sterilized) per liter.

2. 60 YPD + G418 square agar plates: Fill the plates with molten YPD + G418 agar in a sterile hood using a sterile pipette tool. Fill each plate with 35 mL of agar and let solidify.
3. Yeast nonessential deletion collection: If purchased, the collection will arrive in frozen glycerol stocks and need to be arrayed onto agar. The arrayed collection can be stored at 4 °C for 1–3 months.
4. Pipetting tool and 50 mL sterile pipettes.
5. 96-Well pin transfer tool or multichannel pipette.

2.2 Creating the Pooled Deletion Collection for Screening

1. YPD + G418 liquid medium: Make 2 L. For each liter, add 10 g of yeast extract, 20 g of peptone, and 950 mL of dH_2O and autoclave. To the cooled medium add 50 mL of 40 % glucose (40 g of glucose dissolved in 100 mL of H_2O and autoclaved) to make 1 L of 2 % glucose YPD. Once cooled to 50 °C, add 1 mL of 1,000× G418 (Geneticin, 200 mg/mL in dH_2O, filter sterilized) per liter. This medium will be used to grow the mutants prior to creating the screening pool. Medium can be stored for up to 1 month at 4 °C.
2. 96-Well, flat-bottom plates.
3. 96-Well plate shaker.
4. Sterile 100 mL reservoirs.
5. 25–50 mL pipettes.
6. 2 L sterile flask with cover.
7. Centrifuge with a capacity for vessels with a volume >50 mL.
8. Spectrophotometer.
9. Microscope and hemocytometer.
10. 30 % glycerol: Add 300 mL of glycerol to 700 mL dH_2O, autoclave.
11. Freezer tubes (1–2 mL) and boxes.

2.3 Components for Pooled Competitions

1. Screening medium: For basic screening use YPD with 2 % glucose. For 1 L, suspend 10 g of yeast extract and 20 g of peptone in 950 mL of dH_2O and autoclave. Fill to 1 L with a sterile 50 % glucose solution (*see* **Note 2**).
2. Compound(s) of interest in solution, control compounds (e.g., benomyl, MMS), and solvent for control conditions (*see* **Note 3**).

2.4 Components for Genomic Extraction, PCR, Gel Extraction, and Template Quantification

1. Genomic DNA extraction kits (*see* **Note 4**): For 1–96 samples use individual genomic extraction kits. For >96 samples, 96-well genomic extraction kits are preferred.
2. Zymolyase solution (only for 96-well genomic extractions): 1 mg/mL 100 T zymolyase in 1 M sterile sorbitol (182.17 g sorbitol in 1 L, autoclaved).
3. Taq Super Mix: Contains Taq, dNTPs, buffer.
4. TE buffer: 10 mM Tris–HCl, 1 mM EDTA, bring to pH 8.0 with HCl.
5. Indexed primer collection and common primer (*see* **Note 5**) (Table 1): Prepare indexed PCR primers, to a final concentration of 12.5 μM and the common U2 primer to a concentration of 100 μM for using TE buffer.
6. Agarose gel (2 %).
7. Agarose gel extraction kit.
8. Kapa Illumina qPCR kit.

2.5 Sequence Analysis and Target Prediction

1. Barseq counter software script package (available at www.github.com/csbio/barseq_counter).
2. Computer with Python (Version 2.7 or higher).
3. Computer with R software and Bioconductor, EdgeR, limma, and corrplot packages.

3 Methods

3.1 Arraying the Deletion Collection to Agar

Before the deletion collection can be pooled, the individual mutants must be grown in liquid culture. If purchased, the collection will arrive as frozen glycerol stocks. Rather than inoculate the liquid media directly from the glycerol, it is best to array the collection onto agar from the glycerol stocks first. This has the advantage of giving a "working collection" that can be used for single-mutant validations, or make new pooled collection without having to repeatedly thaw the deletion collection stocks and affect viability. The "working collection" can be stored at 4 °C for 1–3 months before it should be transferred. This section describes making an agar array of the deletion collection from the glycerol stocks.

1. Thaw the frozen deletion collection glycerol stocks completely.
2. In a sterile hood, use the 96-well transfer device to spot a small volume (1–2 μL) of the glycerol stocks onto the agar plates. If using a non-disposable transfer device, be sure to bleach for 5 min and flame three times between transfers to prevent contamination. Always be mindful of plate orientation (i.e., location of the A1 well).
3. Let the plates dry completely and incubate for 48 h at 30 °C.

Table. 1
Example set of 12 indexed primers and their 10 bp index tags. The entire primer contains the Illumina-specific region, the index tag, and the common priming site for barcode amplification. The common reverse primer contains an Illumina-specific region and a common priming site from the KanMX gene region of the deletion insert

Index tag	Entire indexed primer (5′–3′)
AATAGGCGCT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAATAGGCGCTGATGTCCACGAGGTCTCT
TACAGTTGCG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTACAGTTGCGGATGTCCACGAGGTCTCT
ATCCTAGCAG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTATCCTAGCAGGATGTCCACGAGGTCTCT
GATTAGCCTC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGATTAGCCTCGATGTCCACGAGGTCTCT
AATGAGCCGT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAATGAGCCGTGATGTCCACGAGGTCTCT
ACGCGGATTA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTACGCGGATTAGATGTCCACGAGGTCTCT
GCTTACGGAA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTTACGGAAGATGTCCACGAGGTCTCT
CGGTAGACTA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGTAGACTAGATGTCCACGAGGTCTCT
ATTGCCGGAA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTATTGCCGGAAGATGTCCACGAGGTCTCT
GACATGCTAG	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGACATGCTAGGATGTCCACGAGGTCTCT
TACGCTGCAT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTACGCTGCATGATGTCCACGAGGTCTCT
GTCAAGCACT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCAAGCACTGATGTCCACGAGGTCTCT
Common reverse primer	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGCACGTCAAGACTGTCAAGG

3.2 Growing the Yeast Deletion Collection to Create the Screening Pool

To create the screening pool, each member of the deletion collection is cultured in independent wells and then pooled by pipetting the cultures together into a common pool. This section describes how to culture members of the deletion collection prior to pooling.

1. Working in a sterile hood, pipette 200 μL of YPD + G418 medium into each well of 60 sterile, 96-well flat-bottomed plates. The purchased deletion collection in 96-well format is arrayed onto 57 plates. It is best to make 60 plates in case of accidents.
2. Using a 96-well transfer tool, transfer a small amount (about the size of a pinhead) of cells from each agar array to a corresponding liquid media plate (*see* **Note 6**). If not using disposable transfer tools, be sure to bleach the tips for 5 min and flame three times between transfers.
3. Shake the liquid plates gently on a plate shaker for 30 s to distribute the cells.
4. Incubate the liquid cultures for 48 h at 30 °C until the cultures reach saturation (OD ~1.5).

3.3 Pooling the Deletion Collection

After the cultures have grown they are ready to mix together to make the screening pool for chemical genomic profiling. Making the screening collection is a sensitive step, and care must be taken to ensure that the screening pool is at a high cell density with an even representation of strains. This step describes how to pool the individual cultures to ensure a homogeneous mixture, and importantly how to adjust the cell density to have adequate strain representation in downstream pooled competition assays.

1. In a sterile hood and using a multichannel pipette, remove the entire volume of the saturated liquid cultures from the 96-well plates and expel them into a reservoir. Pipette up and down or mix the plates on a plate shaker to ensure that the cells are all in suspension before removal.
2. When the reservoir is full, use a 25 (or 50) mL pipette to transfer the liquid culture into a 2 L flask with a stir bar. Do not pour. Pipette up and down with the transfer pipettes to ensure homogenization before transfer.
3. When the entire deletion collection has been pooled, cover the 2 L flask with sterile foil and allow it to mix on a stir plate for 3 min to ensure homogenization of the pool.
4. The freshly harvested screening pool will not have a high enough cell density for the competition assay, and will have to be concentrated by centrifugation at 500 × *g* before the frozen aliquots can be made. To concentrate the cell pool, use a centrifuge and large-volume vessels. Concentrate the pool to an absorbance of 70 or greater at 600 nm by centrifuging at

500 × *g* and decanting the excess media to adjust the OD. It is also advisable to confirm cells/mL using a hemocytometer. The recommended minimum density after concentration is 250 cells/strain/μL. For a 5,000 strain collection, that is 1.25×10^9 cells/mL (*see* **Note 6**).

5. Once a high enough density is confirmed, add 1:1 v/v 30 % sterile glycerol to the cell pool and gently homogenize via a stir plate. The glycerol dilutes the cell density, so it is important to have at least 250 cells/strain/μL before adding the glycerol.
6. Aliquot the pooled deletion collection into freezer tubes. Aliquot volumes of 200–500 μL is best as only a small amount will be used in each experiment, and the pool cannot be used again once thawed (*see* **Note 7**).
7. Store the aliquots at −80 °C until use.

3.4 Pooled Competition

The pooled competition is the point where the mutant pool is exposed to a compound, and this has two steps. First is determining the optimal dose of a drug for the assay. The chemical genomic assay is robust to dosing; however, signal can be maximized by dialing in the best inhibitory concentration. Using the pool as inoculum, find the final compound concentration that inhibits growth by 20–50 % compared to the solvent control after 24 h (*see* **Note 8**). Once the optimal dose is established, the next step is to perform the actual assay with replicates at the appropriate dose. It is important to include at least one control compound in the assay as well, a well-studied compound with known target. The readout of the control compound will let you know if the assay is working. Benomyl and MMS are good choices for control compounds. Benomyl targets tubulin and MMS damages DNA. A final concentration of 10–25 μg/mL of benomyl or 0.01 % MMS is an appropriate control dose. The chemical genomic profiles of these compounds can be used to assess success of the assay. In this step we describe how to determine the screening dose and then perform the pooled competition.

1. To determine the optimal dose, first thaw an aliquot of cells created in Subheading 3.3 and dilute the cells to the starting inoculum concentration (125–250 cells/strain/μL). The cells can be used directly once thawed and diluted.
2. Create cultures with 196 μL medium, 2 μL of a compound or solvent, and add 2 μL of the strain pool. Try to determine the compound and control compound dose that reduces growth of the pool by 20–50 % relative to solvent control (although low/no inhibition can still be informative depending on the performance of individual mutants). This is best accomplished using a dose curve.

3. Take a time zero (t0) measure of OD using a spectrophotometer, and then incubate for 24 h at 30 °C.
4. Measure the growth after 24 h and calculate the growth of the compound conditions relative to the solvent control. Plates can also be read continuously on an automated plate reader.
5. Find the dose of the compound that inhibits growth by 20%–50 % compared to that of the solvent, and this will be the screening dose to use in the next assay. For example, we have found a dose of 10–25 μg/mL of benomyl or 0.01 % MMS to be a good dose point for these control compounds.
6. Once dose is established, use this dose for all further experiments. Prepare wells for competition using 196 μL medium and 2 μL of compound in solvent. It is best to have at least four replicates of each test compound or control compound. Run at least four solvent control conditions.
7. Thaw another aliquot of cells created in Subheading 3.3 and dilute the cells to the starting inoculum concentration (125–250 cells/strain/μL). The cells can be used directly once thawed and diluted.
8. Add 2 μL of the pooled cells to all wells containing the growth medium, mix, and incubate at 30 °C for 48 h, recording the OD at 0, 24, and 48 h. Plates can also be read continuously on an automated plate reader.
9. Harvest the cells after 48 h of growth by centrifugation and remove the supernatant, saving the pellet. Proceed to genomic DNA extraction (Subheading 3.5) or store a cell pellet at −80 °C until extraction. Make sure to remove all cells from the wells by pipetting up and down.

3.5 Genomic DNA Extraction

After the pooled competition, the genomic DNA is extracted from the cells in preparation for amplification of the molecular barcodes. For less than 96 samples, it is preferred to perform individual genomic extractions. However, for larger scale projects, several 96-well genomic DNA kits and automated options are available, and we have found these to be comparable in quality (*see* **Note 4**).

Individual cultures: Perform genomic DNA extractions on the 200 μL cultures, scaling the kit specifications to the smaller volume. We have found that eluting the DNA with 35–50 μL of elution buffer gives a good concentration for PCR.

96-Well extractions: For 96-well extractions, both automated (e.g., Qiaextractor) and manual options exist. Many of these kits are not designed for yeast, and as such an extra cell wall digestion step is required. After pooled growth, harvest the cells by centrifugation and remove the supernatant. The cell pellets can be stored at −80 °C until needed. Before extraction, resuspend the cells in

zymolyase solution and incubate for 1 h at 37 °C to digest the yeast cell wall. After this, proceed with the extraction according to kit specifications.

3.6 PCR and Gel Extraction

After the genomic DNA has been extracted, the next step is to PCR amplify the molecular barcodes that are used to identify members of the strain pool and assess their fitness in the presence of a compound. The yeast deletion collection has two molecular barcodes for each gene deletion, an "UPTAG" and a "DOWNTAG" [6]. This method uses the "UPTAG" only.

It is at the PCR step that the index tags are added that allow multiplexed sequencing. Genomic DNA from each pooled competition will have an independent PCR reaction, each with a unique indexed primer plus the common primer. For four replicates, this means four separate PCR reactions each with a unique index primer and a common primer. These unique primers are designed with a 10 bp sequence that allows them to be pooled together (multiplexed) for sequencing and then de-multiplexed during analysis. A description of the index primer design and the resultant amplicon that will be sequenced can be found in Fig. 2a. It is very important at this step to keep detailed notes on which indexed primer is matched with each experiment. For instance, the solvent control conditions may use primers 1–4, where the compound conditions use primers 5–8. The 10 bp index tag is what is used to tell the analysis software how to de-multiplex the data into the individual experiments. We have included an example of 12 indexed primers in Table 1, and a set of 96 unique indexed primers that we have assessed for performance within the software package and supporting material (available at www.github.com/csbio/barseq_counter). This step describes how to amplify the molecular barcodes with special indexed primers, pool, and then clean up the PCR product for barcode sequencing.

1. Purchased primers should first be diluted to working concentration using TE buffer. Prepare indexed PCR primers to a final concentration of 12.5 μM for index primers and 100 μM for the common U2 primer.
2. Reaction mixture (per reaction): 20.25 μL of Taq mix, 0.25 μL of U2 primer, 2 μL of indexed primer, 2.5 μL of genomic DNA. Always add genomic DNA last to avoid any chance of contaminating the primer stock.
3. PCR conditions: An initial denaturation of 5 min at 95 °C, followed by 30 cycles of 1 min at 95 °C, 30 s at 55 °C, 45 s at 68 °C, then a final extension time of 10 min at 68 °C.
4. Pool the PCR products from the individual PCR tubes for gel extraction by combining volumes of individual reaction mixtures into a single tube. For 8–24 samples, the entire volume of each reaction mixture can be pooled. For >24 samples, pool 10 μL from each reaction.

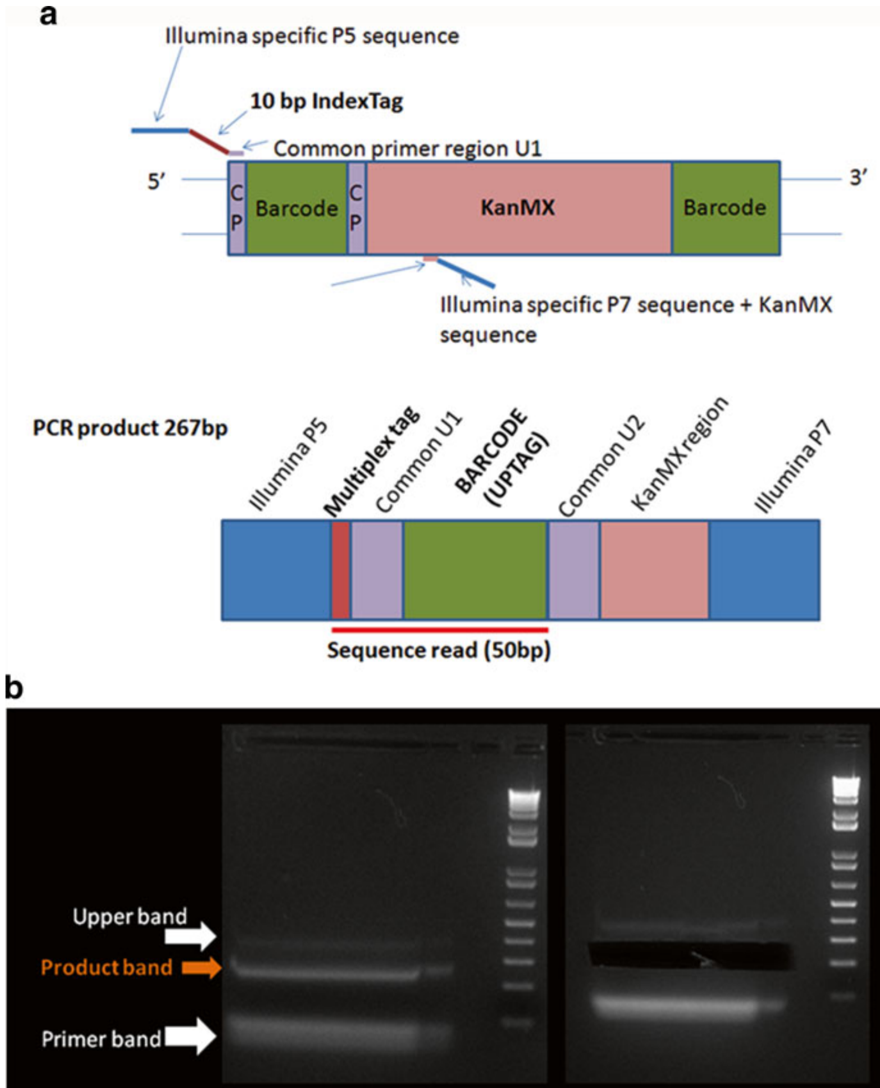


Fig. 2 PCR amplicon design and gel extraction of the amplified barcodes. The indexed primer design and components of the PCR amplicon used in sequencing (**a**). Using a 2 % agarose gel with Syber Safe of ethidium bromide, run the pooled PCR product for 45 min using a 1 kb ladder for reference and visualize under UV. Three bands will be apparent, a *lower band* of unused primers, the *middle band* at 267 bp, and an *upper band* of amplification artifacts. Excise the *middle band* for gel extraction (**b**). The other bands contain the Illumina regions of the primers and if run on the sequencing flow cell, will form clusters but will not provide usable reads (loss of read depth)

5. Prepare a 2 % agarose gel with Sybr Safe or ethidium bromide for visualization. Using tape, make an extra large well that will accommodate the entire volume of the pooled PCR products. For >96 PCR reactions multiple gel extractions may be necessary.
6. Run the gel for 30–45 min under the following conditions: 120 V, 200 mA.

7. Identify the desired 267 bp PCR product band (*see* **Note 9**) by visualizing the gel under UV. There will be three bands: a lower primer band, a center band (267 bp, the desired product), and a higher band (>450 bp) (Fig. 2b).
8. Carefully cut out the center band (267 bp) making sure not to cut any of the other bands. Try and excise a thin band, to minimize the amount of gel, and remove the slice to a tube for gel extraction.
9. Perform gel extraction on the excised gel. Elute the final product using a minimal buffer volume (25–50 μL) to ensure a high template concentration.

3.7 Quantification and Sequencing of the Samples

The amount of PCR product must be quantified prior to Illumina sequencing. It is important to load the correct amount of template DNA onto the Illumina flow cell. Too much template will cause the cluster density on the flow cell to increase to levels that severely compromise data quality. Conversely, too little template will cause low cluster densities and, consequently, insufficient reads.

There are several approaches to quantifying the library for sequencing (e.g., Kapa qPCR kits, PhiX-based qPCR, Bioanalyzer). We have found that he Kapa Illumina qPCR kits gives the most accurate estimation of the sample as it assesses only molecules amplifiable with the specific Illumina sequencing primers. A final template concentration of at least 10 nM is best to provide for sequencing. Even if a sequencing facility will quantify the samples prior to sequencing, provide quantification data when submitting the sample (*see* **Note 10**). In this step, we describe the method of sample quantification using the Kapa Illumina qPCR kit.

1. Prepare three replicates for each of the six standards.
2. Prepare three replicates of the following dilutions in 1× TE (elution buffer from the gel extraction kit) of the gel-extracted sample: 1:2,000, 1:5,000, and 1:10,000.
3. Determine the concentration of each dilution via qPCR, using machine-specific software to calculate the concentrations.
4. Calculate the sample concentration using the Kapa Kit formula.
5. Run the samples on an Illumina platform (1 × 50 cycles) at a template concentration that will yield a cluster density of 700–900 k/mm^2. We have found a template concentration of 10–18 pM to work best (*see* **Note 11**).
6. Obtain the fastq file after sequencing

3.8 Data Analysis

Following sequencing, it is finally time to get a glimpse of the data. To generate a chemical genomic profile, the raw sequence data must be processed and de-multiplexed based on the index tags, and finally a chemical genomic score assigned to each mutant in the different compound conditions. We have provided a set of open -source Python and R scripts that can be used to perform these analyses.

The script package entitled "barseq counter" is available from www.github.com/csbio/barseq_counter. Within this package are all the required instructions, scripts, and tools necessary to process, de-multiplex, count, and detect compound-mutant interactions. We have also provided a sample dataset (fastq files plus decode and control input files) with detailed instructions on using the scripts so that a first-time user can learn how to operate the script package and get a feel for what the output data will look like. The sample dataset is hosted at http://lovelace-umh.cs.umn.edu/chemical_genomics_tools/barseq_counter. "Barseq counter" requires substantial computational power, and is best run from a server with the latest versions of Python and the R-package installed. This software will preprocess the raw fastq sequence file, and then de-multiplex based on the known indexed primer list in the "decode" file that the user provides. Then, it will generate a count matrix of the data with the specific sequence counts for each mutant under each experimental condition.

The software will also provide basic quality control (QC) steps to help assess the overall success of the sequencing run. It will tell the percent of sequence reads passing the filter, the number of reads for each index tag, and the number of reads for each mutant. This QC output is important; if a particular index tag has abnormally low counts then the data from that tag may not be useful. In general, if the sequencing read has an average of at least 100 counts per strain, then it can be considered successful. Of course some strains will have either low counts or high counts based on their response to the compounds, but the average count across all mutants is what to look for to determine how well the sequencing run performed.

The final step uses runEdgeR.R, an R script wrapper around the EdgeR package, for determining the differential growth and statistical significance for each mutant relative to the control conditions. The runEdgeR.R script normalizes the data against the solvent control conditions, and allows the detection of drug-gene interactions. The use of EdgeR in barcode sequence experiments is described well in Robinson et al. 2014 [11, 12], where the list of fold changes is the chemical genomic profile, and the genes in this list give functional insight into the compound's mode of action and potential cellular target. The R script also generates box and correlation plots of the data to further assess data quality and determine the agreement between replicates. This section describes how to install, use, and interpret chemical genomic data.

1. Download and copy the folder "barseq_counter" to a computer that will run Python and the R-package for processing the data. This is a resource-intensive process, and is best run on a server running Linux. Within the "barseq_counter" package is a program called Agrep 3.14, which is used to count the barcodes and must be installed if not already present. The file folder "barcodes" contains the mapping of the strain-specific

barcodes to their ORF name, and the file "allupbarcodes.txt" contains the most up-to-date list.

2. Obtain a fastq file after sequencing and create a main folder for the experiment (e.g., Chemgen1). We have provided a test dataset here http://lovelace-umh.cs.umn.edu/chemical_genomics_tools/barseq_counter. Within this folder create a sub-folder entitled "data." Place the fastq file in the data folder. If there are multiple fastq files, these all can be placed in the "data" folder and the scripts will concatenate them during processing.

3. The first processing step is to remove the excess sequence data from the reads (e.g., the common priming region). This is accomplished with the script "preprocess_MIseq_10bp.py." In a Linux shell, navigate to the experiment folder (not the "data" folder) and enter:

   ```
   python barseq_counter/scripts/preprocess_MIseq_10bp.
   py data barseq.txt
   ```

   This command will unite them into a single processed output file "barseq.txt." The script removes the common priming region of the sequencing read, which is not used in analysis. Further, the script separates the 10 bp index tag from the gene barcode.

4. The next step is to de-multiplex the data based on the unique index tags and count the strain-specific barcodes. This step requires a user input file that tells the software which index tag corresponds to each experiment. This file is the index tag decode and is a text file with the index tag followed by the experimental condition and has the following format:

   ```
   GATTAGCCTC   DMSO
   AATGAGCCGT   DMSO
   ACGCGGATTA   DMSO
   GCTTACGGAA   MMS
   CGGTAGACTA   MMS
   ATTGCCGGAA   MMS
   GACATGCTAG   Benomyl
   TACGCTGCAT   Benomyl
   GTCAAGCACT   Benomyl
   ```

   See the example file entitled "decode.txt" as an example. Do not use special characters in the compound names (e.g., %, *), and keep a detailed file with all dose points for cross reference. Make sure that all replicates' names are the same and not indicated by a replicate number or letter in the decode file (e.g., do not use BenomylA, BenomylB, etc.), as it is necessary that they are distinguished only by the index tag and maintain the same naming scheme for downstream processing. To de-multiplex the data, have the decode file in the experiment folder and use the command

```
python barseq_counter/scripts/processbarSeq_rd.py
barseq.txt decode.txt barseq_counter/barcodes/allup
barcodes.txt barseq.processed.cerevisiae.txt
```

The output of this step is "barseq.processed.cerevisiae.txt," which contains a matrix of individual strain counts across all experiments. In this matrix, the gene names of the mutants are listed as the systematic ORF. We have provided a script that will convert the ORF name to the common yeast name. This script is called "convertORF2common.py." To run this use the following command from within the experiment folder:

```
python barseq_counter/scripts/convertORF2common.
py barseq.processed.cerevisiae.txt barseq.processed.
common.txt
```

The file barseq.processed.cerevisiae.txt is the count matrix for the experiment and will be used to determine chemical genetic interactions.

5. Before moving on to determine chemical genetic interactions, it is important to assess sequence quality, to make sure that there are enough sequence reads and spot any potential problems with the data. First make a new folder for the quality reports. While in the main project folder in the shell, use the command

```
"mkdir -p <folder name>_reports"
```

to create a folder for the read quality reports. Next use the command

```
python barseq_counter/scripts/generateReport_MIseq.
py data barseq.txt barseq.processed.cerevisiae.txt <folder
name>_reports/cerevisiae
```

This command will generate distribution plots of the index tag and barcode counts, in addition to text file summaries. Here the average sequencing counts for each index tag and barcode can be assessed. These data can be used to estimate the counts per strain for each indexed condition by dividing the total counts for an index tag by the number of strains; and ideally there will be >100 counts per strain.

6. The raw count matrix from Subheading 3.7, **step 4**, is used to determine chemical genetic interactions. This is done using the R-package "EdgeR" and a user input file that identifies the control (solvent conditions). EdgeR normalizes the count data for each experimental condition and estimates the differential growth against the control conditions to generate a fold change for each mutant in the presence of a compound. A fold change of >1 indicates increased growth, whereas <1 indicates reduced growth or sensitivity of the mutant to a compound compared to the solvent control. EdgeR also generates an adjusted *P*-value for responsive strains, which is a measure of statistical significance of the fold change, which has been corrected for multiple comparisons. The control conditions file is a text file list of index tags associated with the solvent conditions using the following format:

DMSO_GATTAGCCTC
DMSO_AATGAGCCGT
DMSO_ACGCGGATTA
To run EdgeR, use the command

runEdgeR.R--threshold 10 barseq.processed.cerevisiae.txt controls.txt

The threshold is set to make sure that EdgeR ignores conditions with very low read counts; we generally set the threshold at ten counts. In the "csv" output folder, EdgeR will provide a list of compound-responsive strains for each compound sorted based on the significance of their fitness change relative to the solvent control. The output "pdf" and "png" folders from EdgeR provide heatmap plots of the data to assess replicate agreement and global distribution of sensitive and resistant mutants as volcano plots. This is the starting point to interpret the chemical genomic data and make predictions of compound mode of action. Sort the list by fold change, and start by searching for functional enrichment among the top 10–20 sensitive or resistant strains (http://go.princeton.edu/cgi-bin/GOTermFinder). Use the control compound conditions to assess assay success. For example for benomyl, within the top sensitive strains are gene mutants in tubulin-related processes (e.g., *CIN1*, *GIM4*, *PAC2*). For MMS, there is significant enrichment for mutants involved in DNA repair (e.g., *PSY3*, *SAE2*, *RAD4*). Look for functional enrichment in the sensitive and resistant mutants of the unknown compounds to gain functional insight into what they may be targeting.

This chapter describes the basic steps for generating chemical genomic profiles using the nonessential yeast deletion collection and barcode sequencing. The profile can be used to unveil the cellular target or mode of action of novel compounds. This approach can be repeated using the heterozygous yeast deletion collection of essential genes to cover greater target space. Further, the chemical genomic profile can be correlated with the genetic interaction profiles of the yeast genetic interaction network and give further insights into mechanism [4, 10]. Chemical genomics paired with barcode sequencing can provide an unbiased, high-throughput screening method of rapidly linking compounds to their cellular targets.

4 Notes

1. If purchased, the yeast deletion collection may arrive as glycerol stocks. These will need to be thawed and pinned to YPD + G418 agar before starting the pool creation. Strain pools can be made either by scraping colonies off agar [11] or mixing liquid cultures. We have tested both methods and

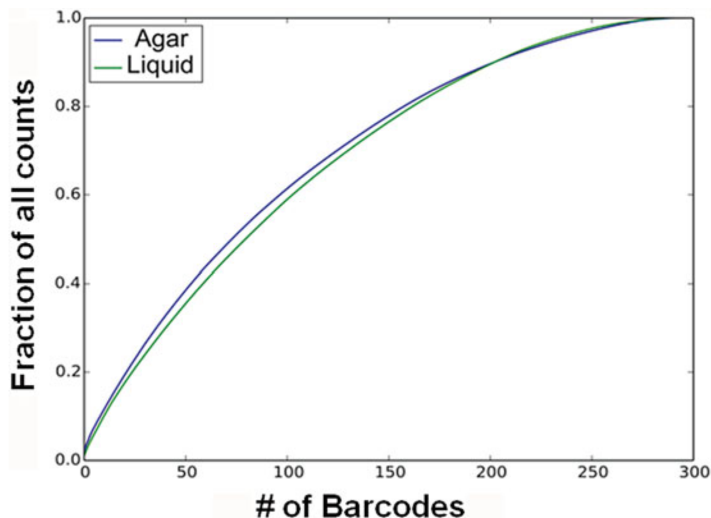


Fig. 3 Distribution of barcodes based on pooling method. We constructed two pools of 300 strains using two different pooling methods: agar scraping and mixing of liquid cultures. We then grew the pools in our chemical genomics assay. While the pools performed similarly, we found that the liquid pool had better distribution of strains, as determined by fraction of individual strains in the pool. In this figure, a *straighter line* indicates a more even strain distribution

have found that liquid cultures give a more equal distribution of strains (Fig. 3).

2. The assay described here is optimized for microcultures, which can be used in small-scale or high-throughput systems. Any growth medium can be used for the chemical genomics assays. We generally use rich media (e.g., YPD or YP galactose with 2 % sugar). We have found that using YP galactose slows cell growth and slightly sensitizes the yeast to compounds, which can yield good separation of compound-responsive mutants in the pooled competition [13].
3. Aside from water, DMSO is the preferred solvent as it has low toxicity to yeast. Stock solutions of the compound can be prepared depending on compound availability. A stock solution of 1 mg/mL is a good starting point to assess bioactivity of the compound.
4. Depending on the scale, genomic DNA extraction can be performed individually or 96 wells at a time. Successful DNA extraction with the 96-well kits requires a preincubation of the cell pellet with zymolyase, as these kits are usually not specifically designed for yeast genomic extractions. Resuspend the cell pellet in 125 μL of zymolyase solution and incubate for 1 h at 37 °C, after which the cells are ready for genomic extraction following the kit specifications.
5. Our primers are designed and optimized for the Illumina platform. We have built in 10 bp index tags so that the experiments can

be multiplexed. We have tested a range of index tag sizes, and found 10 bp to perform well.

6. The cell density of the screening pool stock, and thus the inocula used in experiments, is critical to obtaining informative chemical genomic signatures, as the ability to detect sensitive or resistant strains suffers when the number of cells per strain is either too high or too low. With too few strains there is not sufficient strain representation in the assay and functionally informative mutants may be missing; however, with too many cells, the signal of compound-mutant interactions is dampened as there is less capacity for growth. While the assay is somewhat robust to the starting cell density, we have found the optimal screening concentration to be 125–250 cells/strain. Higher densities are acceptable for the stock pool as it can be diluted prior to performing the growth assays, whereas it is more difficult to concentrate it if the density is too low.
7. The pools in glycerol can be aliquoted and stored at −80 °C. As very few cells will be used per experiment (especially if they need dilution), it is best to aliquot in small volumes (200–500 μL) so as to extend the life of the stock collection.
8. Do not exceed 1 % solvent in the cultures. If resources allow, plan on using multiple dose points so that a "dose-dependent" chemical genomic profile can be generated. Inhibition of growth by <20 % of solvent controls may still yield information given the performance of individual mutants in the pool.
9. There will be three bands: a lower band (~100 bp) that contains the unamplified primers, a middle band (267 bp) that contains the desired fragment, and a higher band (~500 bp) that results from nonspecific amplification of the Illumina regions. Cut out only the middle band. As both the lower and upper bands will contain Illumina regions, these will contaminate the flow cell if they make it to the sequencing reaction, and will result in fewer usable sequencing reads. The excised gel can be extracted following kit specifications; however, we recommend eluting the purified product with half the recommended volume to ensure a high concentration of PCR product.
10. There are several methods to quantify samples for Illumina sequencing. Sequencing facilities often have their preferred methods, and it may be best to follow their recommendations. We have found that the Kapa Illumina qPCR kits to be very reliable in estimating product, as these rely specifically on the Illumina regions to amplify, and detect the quantity of only amplifiable DNA. qPCR data can be paired with Qbit

and Bioanalyzer data to get high confidence quantification. We recommend for qPCR that all samples be run in triplicate, control curves, and dilutions of the samples. We have found that dilutions of 1:2,000, 1:5,000, and 1:10,000 are best for getting accurate quantification. 1:1,000 dilutions are often still too concentrated.

11. Our method has been optimized for Illumina sequencing; however, it could be adapted to any next-generation sequencing platform. Smith et al. [14] describes both Illumina and Solexa barcode sequencing. As these are single-end reads, we can use a higher cluster density on the Illumina flow cell, and a higher template concentration. We have found that, for experiments using the 5,000-strain pool, a running concentration of 10–18 pM as determined by the Kapa qPCR kit results in an optimal quality and quantity of sequencing reads. However, it is best to work with the sequencing facility operators when dialing in the running concentration.

For the 5,000-strain collection and the MiSeq, platform, up to 25 samples can be pooled and analyzed in one MiSeq run. For HiSeq 2500 Rapid Run, up to 192 samples can be pooled and analyzed in a HiSeq 2500 Rapid Run (two flow cells, one sample). For single-lane HiSeq 2000/2500, up to 96 samples can be pooled and analyzed in each lane of a HiSeq 2000/2500 run (768 samples per full 8-lane sequencing run). We do not recommend using PhiX in the lane with the samples, as it reduced read counts and does not help read quality.

Acknowledgements

J.P., S.M., and I.O. are funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494). SC is supported by a RIKEN Foreign Postdoctoral Researcher Award. C.M., and R.D. are supported by grants from the National Institutes of Health (1R01HG005084-01A1, 1R01GM104975-01, R01HG005853), a grant from the National Science Foundation (DBI 0953881), and the CIFAR Genetic Networks Program. S.W.S. is supported by an NIH Biotechnology training grant (5T32GM008347-22). C.B. is supported by the Canadian Institutes of Health Research (CIHR MOP-57830). R.A. is supported by the Natural Science and Engineering Research Council (NSERC) of Canada, the Canada Foundation for Innovation (CFI), and the Canadian Cancer Society Research Institute.

References

1. Parsons A et al (2006) Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. Cell 126:611–625
2. Ho CH et al (2011) Combining functional genomics and chemical biology to identify targets of bioactive compounds. Curr Opin Chem Biol 15:66–78
3. Fung S-Y et al (2013) Unbiased screening of marine sponge extracts for anti-inflammatory agents combined with chemical genomics identifies girolline as an inhibitor of protein synthesis. ACS Chem Biol 9:247–257
4. Williams DE et al (2011) Padanamides A and B, highly modified linear tetrapeptides produced in culture by a *Streptomyces* sp. isolated from a marine sediment. Org Lett 13:3936–3939
5. Giaever G et al (2004) Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. Proc Natl Acad Sci U S A 101:793–798
6. Giaever G et al (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 418:387–391
7. Kim D-U et al (2010) Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. Nat Biotechnol 28: 617–623
8. Kitagawa M et al (2006) Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. DNA Res 12:291–299
9. Smith AM et al (2009) Quantitative phenotyping via deep barcode sequencing. Genome Res 19:1836–1842
10. Costanzo M et al (2010) The genetic landscape of a cell. Science 327:425–431
11. Robinson DG, Chen W, Storey JD, Gresham D (2014) Design and analysis of bar-seq experiments. G3 (Bethesda) 4:11–18
12. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140
13. Andrusiak K (2012) Adapting *S. cerevisiae* chemical genomics for identifying the modes of action of natural compounds. Thesis. https://tspace.library.utoronto.ca/handle/1807/32456. Accessed 24 Apr 2014
14. Smith AM et al (2010) Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. Nucleic Acids Res 38:e142

Chapter 24

Image-Based Prediction of Drug Target in Yeast

Shinsuke Ohnuki, Hiroki Okada, and Yoshikazu Ohya

Abstract

Discovering the intracellular target of drugs is a fundamental challenge in biomedical research. We developed an image-based technique with which we were able to identify intracellular target of the compounds in the yeast *Saccharomyces cerevisiae*. Here, we describe the rationale of the technique, staining of yeast cells, image acquisition, data processing, and statistical analysis required for prediction of drug targets.

Key words Image processing program, CalMorph, *Saccharomyces cerevisiae*, Morphological profiling, Image-based approach, Drug target discovery, Antifungal drugs, Multivariate analysis

1 Introduction

Given that specific morphometric features of organisms are often affected by drugs, the biological activity of drugs can be investigated based on the statistical analysis of morphological changes induced by the drugs. Dose-dependent morphological changes by the drugs give much information on the detailed pharmacological response [1] as well as the characteristic features of the compounds [2].

In order to analyze morphometric features in *Saccharomyces cerevisiae*, we developed an automatic image-processing system, called CalMorph [3, 4]. CalMorph is a high-throughput, high-resolution, image processing system specialized for yeast cells that allows us to analyze and quantitate 501 cell morphology parameters from fluorescent microscopic images of triple-stained (cell wall, actin, and nuclear DNA) cells [4]. Combining the image-processing system and a comprehensive panel of the nonessential deletion mutants in yeast facilitated the high-content and large-scale phenotyping of yeast mutants [4]. It also allowed us to predict the intracellular targets of a drug after collecting data of morphological change induced by the drug [5]. The rationale here was that morphology of the drug-treated wild-type strain is supposed to have significant similarity to that of the deletion mutant of the intracellular target protein. Our morphological profiling

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7_24, © Springer Science+Business Media New York 2015

method employed an inference algorithm to estimate similarities between induced morphological changes. Using this algorithm, both known and unknown drug targets were successfully identified [6, 7]. This method can be expanded to image-based analyses in higher eukaryotes if any comparable database is available, because the method is completely based on statistics. The statistical power and utility of the method are enhanced by the multiple parameters that are extracted from high-resolution images.

Methods described in this chapter are composed of three steps. First, the appropriate concentrations of the drugs are decided. Then, dose-dependent morphometric changes are examined by CalMorph. Finally, the intracellular targets are predicted by measuring the morphological similarity between the drug-treated cells and cells of every nonessential deletion mutant strain.

2 Materials

2.1 Strain and Media

1. *Saccharomyces cerevisiae* haploid strain: *his3* derived from BY4741 (access number: Y02458: *MAT***a** *his3*::*KanMX leu2 met15 ura3*) (*see* **Note 1**).
2. Yeast-rich medium, YPD: Add 5 g of Bacto yeast extract and 10 g of Bacto peptone to 450 mL of deionized water, autoclave, and then add a sterilized 50 mL glucose solution containing 10 g of glucose in deionized water.
3. Stock solutions of compounds: Prepare stock solutions according to the manufacturer's guide (*see* **Note 2**).

2.2 Staining Reagents

1. DAPI solution: 1 μg/mL 4′,6-diamidino-2-phenylindole (DAPI) solution. Dissolve 1 mg of DAPI (Wako) in 10 mL of deionized water. Dilute 100-fold with deionized water (*see* **Note 3**).
2. Fixation solution: Mix 100 mL of 37 % formaldehyde solution and 100 mL of potassium phosphate buffer (1 M, pH 6.5) just before use.
3. P buffer: Prepare a 10 mM sodium phosphate buffer adjusted to pH 7.2. Add NaCl to a final concentration of 150 mM.
4. FITC-ConA solution: 1 mg/mL Fluorescein isothiocyanate concanavalin A (FITC-ConA) solution. Dissolve 1 mg of FITC-ConA (Sigma-Aldrich) in 1 mL of P buffer (*see* **Note 4**).
5. Mounting solution: Mix 1,995 μL of PBS and 5 μL of 0.1 N NaOH in the dark and then add 20 mg of *p*-phenylenediamine (Sigma-Aldrich). Add 18 mL of glycerol (for fluorescence microscopy use) and mix it with gentle agitation at 4 °C in the dark (*see* **Note 5**).

6. Rh-ph solution: 200 U/mL Rhodamine phalloidin (Rh-ph) solution. Dissolve 300 U of Rh-ph (Molecular Probes) in 1.5 mL of methanol (*see* **Note 6**).
7. Triton X-100 solution: 10 % (v/v) Triton X-100 in deionized water.

2.3 Equipment

1. CalMorph (ver 1.1): Image processing software CalMorph (ver 1.1) can be downloaded from the SCMD [8] site (http://scmd.gi.k.u-tokyo.ac.jp/datamine/).
2. Centrifuge with holder for 50 mL conical bottom tubes.
3. Conical bottom tubes (50 mL).
4. Coverslips (22 × 22 mm).
5. Flat-bottomed 96-well plates with lids.
6. Microscope equipment: The microscope must be outfitted with a camera, appropriate light source, appropriate filter sets, and image analysis software (*see* **Note 7**).
7. Microcentrifuge tubes (1.5 mL).
8. Microcentrifuge (Eppendorf Minispin).
9. Multiwell spectrophotometer: SpectraMax Plus384 spectrophotometer (Molecular Devices).
10. Rotator.
11. Shaking incubator.
12. Slide glass.
13. Sonicator.
14. Water bath shaker.

3 Methods

Carry out all procedures at room temperature unless otherwise specified.

3.1 Determination of the Maximum Dose of the Drug Concentration

This step is necessary to avoid side effects of the drug and to obtain sufficient information on dose-dependent morphometric changes.

1. Prepare a preculture of wild-type strain in YPD medium incubated at 25 °C. Inoculate cells into 96 wells of a microtiter plate containing fresh YPD medium and various inhibitory concentrations of the drug or the solvent alone (*see* **Note 8**).
2. Incubate the cultures at 25 °C in a shaking incubator.
3. Measure their optical densities at 600 nm at more than 10 time points over the course of 2 days using a multiwell spectrophotometer (*see* **Note 9**).

4. Calculate the average doubling time and determine the maximum concentration of the drug that gives doubling time delayed by approximately 10 % (*see* **Note 10**).

3.2 Fixation of Yeast Cells

1. Inoculate yeast cells into 100 mL flasks containing 20 mL of YPD medium with five different concentrations of a drug including no-drug control (*see* **Note 11**). Replicate experiments five times for each drug concentration (*see* **Note 12**).
2. Incubate the cultures at 25 °C for 15–17 h until concentrations of cells reach a concentration of 4×10^6–1×10^7 cells/mL.
3. Transfer the 20 mL cultures to 50 mL conical bottom tubes containing 5 mL of fixation solution.
4. After closing the caps, agitate the tubes for 30 min at 25 °C in a water bath.
5. Collect the cells by centrifugation at $1,870 \times g$ for 5 min at 25 °C.
6. Discard the supernatant. Resuspend the cells in a mixture of 2 mL of fixation solution and 8 mL of deionized water.
7. After closing the caps, agitate the tubes for 45 min at 25 °C in a water bath.
8. Collect the cells by centrifugation at $1,870 \times g$ for 5 min at 25 °C.
9. Discard the supernatant. Resuspend the cells in 1 mL of PBS (*see* **Note 13**).

3.3 Staining of Cells

1. Transfer the suspended cells into a 1.5 mL microcentrifuge tube. Keep the samples on ice until **step 11** of Subheading 3.3.
2. Collect the cells by centrifugation at $6,596 \times g$ for 30 s at room temperature.
3. Discard the supernatant. Resuspend the pellet in 600 μL of PBS. Hereafter, **steps 2** and **3** of Subheading 3.3 are referred to as "washing."
4. Collect the cells by centrifugation at $6,596 \times g$ for 30 s at room temperature.
5. Discard the supernatant. Resuspend the pellet in a mixture of 90 μL of PBS, 10 μL of Rh-ph solution, and 1 μL of Triton X-100 solution to stain actin (*see* **Note 14**). Incubate at 4 °C overnight in the dark.
6. Wash the cells with 600 μL of PBS.
7. Wash the cells with 600 μL of P buffer (*see* **Note 15**).
8. Resuspend the cells in 488 of μL P buffer and 12 of μL FITC-ConA solution to stain the cell wall (*see* **Note 16**).
9. Incubate the mixture at 4 °C for 10 min in the dark.

10. Wash the cells with 600 μL of P buffer.
11. Add 600 μL of P buffer and disperse the cells by sonication for 5 s at level 3 (TAITEC VP-5S).
12. Collect the cells by centrifugation at 6,596 × *g* for 30 s at room temperature.
13. Discard the supernatant completely (*see* **Note 17**).
14. To stain nuclear DNA, mix 7 μL of DAPI solution and 100 μL of mounting solution (*see* **Note 18**). Pipet two tiny drops (0.75 μL each) of the sample onto a slide glass.
15. Pick a small amount of cell pellet with the pipet tip and suspend it well within the drops (*see* **Note 19**).
16. Put a coverslip on the drop and leave it until the solution has spread sufficiently (*see* **Note 20**).

3.4 Image Acquisition and Data Processing

1. Acquire images of the cell wall, actin, and nuclear DNA stained with FITC-ConA (Green), Rh-ph (Red), and DAPI (Blue), respectively (Fig. 1a), in the same field of view using the "Acquire" command or its equivalent.
2. Save the images in 8-bit grayscale JPEG format (image size: 520 × 696 pixels) in the same directory (*see* **Note 21**).
3. Run CalMorph according to its manual (*see* **Note 22**) to obtain data in Excel format.

3.5 Prediction of Intracellular Drug Target

1. Summarize values of all drug concentrations and replicates for every morphological parameter into a Z score from the Jonckheere-Terpstra test (*see* **Note 23**). Each Z score represents the dose dependency of the parameter under a normal distribution.
2. Project the resultant Z scores to principal components (PCs) that are obtained with principal component analysis on null distributed data to reduce dimensions. Morphological data of wild-type replicates (*see* **Note 24**) can be used as null distributed data after the normalization by the Box-Cox power transformation (*see* **Note 25**). The cumulative contribution ratio (CCR) of the PCs should be reached above 70 %. These PC scores obtained from the Z scores represent the profile of morphologic changes that result from treatment with the drug.
3. Calculate PC scores of 4,718 nonessential deletion mutants based on our data in SCMD site (*see* **Note 24**).
4. Calculate correlation coefficient R and the associated P value based on the PC scores from the drug-treated cells and that from every deletion mutant strain. They are useful to evaluate the similarities between morphologic changes in drug-treated, wild-type cells versus mutant strains (Fig. 1b,c). High correlated R values are indicative of drug target.

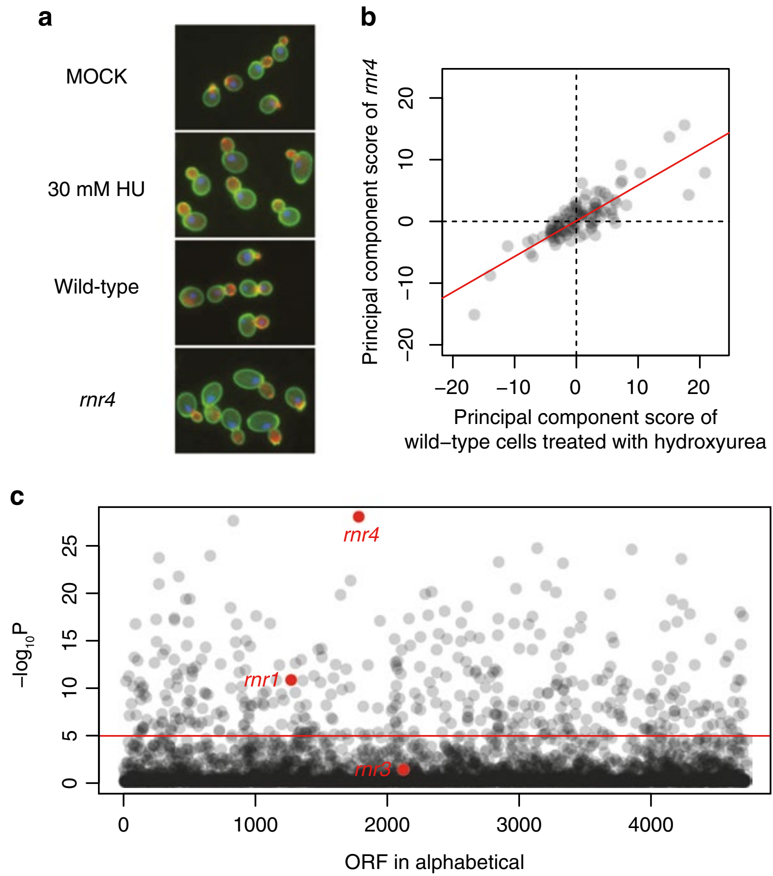


Fig. 1 Morphological profiling of hydroxyurea (HU)-treated cells. (**a**) Images of wild-type cells treated with 30 mM HU and the *rnr4* mutant. Cells were triply stained with FITC-ConA (*green*), Rh-ph (*red*), and DAPI (*blue*). Rnr4p is one of the subunits of ribonucleotide reductase (RNR) complex, an intracellular target of HU. (**b**) Similar morphological profile between HU-treated wild-type cells and the *rnr4* mutant. The scores for the 104 principal components are plotted. The *red line* indicates a linear regression line ($R = 0.836$). (**c**) Distribution of P values by one-sided t-test for correlation coefficients (R) between the mutant and HU-treated wild-type cells. The *red line* indicates the one-sided P value of 0.05 with the Bonferroni correction. *Red dots* and texts indicate nonessential gene-deletion mutants of the subunits of the RNR complex. The *rnr4* mutants showed the highest R value among the 4,718 nonessential gene-deletion mutants

4 Notes

1. The strain can be purchased from the European Saccharomyces cerevisiae Archive for Functional Analysis (EUROSCARF: http://web.uni-frankfurt.de/fb15/mikro/euroscarf/).
2. For example, 2 M hydroxyurea, 100 mM concanamycin A, 20 mg/mL lovastatin, and 2 mg/mL echinocandin B (gift

from Dr. T. Watanabe) were prepared in deionized water; dimethyl sulfoxide (DMSO); ethanolic NaOH containing 15 % (v/v) ethanol and 0.25 % (w/v) NaOH in deionized water; and DMSO, respectively.

3. Store in 1 mL aliquots at 4 °C in the dark. It keeps for about 1 year.
4. Store in 1 mL aliquots at 4 °C in the dark. Storage longer than 1 week may cause uneven staining.
5. To mix them completely, we agitate them overnight on a rotator at 4 °C. Dispense 400 μL aliquots into 1.5 mL disposable microtubes, and store at −80 °C. Keep in the dark; otherwise, the colorless mixture will turn brown. CAUTION: *p*-Phenylenediamine is carcinogenic. Avoid contact.
6. Dispense 500 μL aliquots into 1.5 mL disposable microtubes, and store at −80 °C in the dark. It keeps for several months.
7. Listed below is one example of an acceptable microscope setup:
   - Carl Zeiss Axio Imager M1 fluorescence microscope with a Carl Zeiss EC Plan-Neofluar 100×/1.30 Oil Objective.
   - Roper CoolSNAP HQ charge-coupled device (CCD) camera.
   - Carl Zeiss HBO 100 Microscope Illuminating System.
   - Carl Zeiss FITC/DAPI/Rhodamine filter sets.
   - Carl Zeiss AxioVision ver. 4.5 software with multidimensional acquisition/viewer.
8. Evaluation of the growth inhibition requires five or six different concentrations of every compound in triplicate. If the appropriate concentration cannot be determined, change the range of the concentrations.
9. Collecting the optical density data with exponentially growing cells is essential.
10. Calculate doubling time as the mean from triplicate. The maximum concentration of the drug that we used previously [5–7] gives doubling time delayed by approximately 10 %.
11. The number of inoculating cells is determined according to the growth rates in the presence and absence of the drugs. The culture volume can be reduced to 0.5 mL as described before [1].
12. Replicating the experiments at least five times is essential to analyze dose-dependent change statistically by Jonckheere-Terpstra test (*see* **Note 23**).
13. Although it is possible to store fixed cells at 4 °C for several days, immediate staining of cells is recommended.
14. To uniformly stain actin dots, do not foam during mixing the sample.

15. To uniformly stain the cell wall by FITC-ConA, reduce the amount of cell pellet as much as possible.
16. Mix cells quickly to avoid nonuniform staining. Prepare new 1.5 mL disposable microtubes and dispense 12 μL of FITC-ConA solution to make a drop on the inner surface of the microtube. Add the cell suspension (by 488 μL P-buffer) onto the bottom of the microtube and mix with the solution drop very quickly.
17. Remaining solution may cause high background fluorescence during microscopic observation. Extra centrifugation may help to remove supernatant completely.
18. The glycerol-containing solution is highly viscous. Mix gently by pipetting until it is completely dissolved. A point-cut pipet tip (cut down the point of yellow tip by scissors) is helpful to handle this procedure.
19. Too many cells results in high-density areas. In such cases, too many cells are classified as "complex," and CalMorph cannot extract quantitative data.
20. Since the cells are not fixed onto the coverslip with glue, they tend to move around. Therefore, leave the slide for at least 10 min on a flat surface in the dark before observation.
21. The file names should be:

    [folder name]-C[number].jpg for FITC-ConA images.

    [folder name]-D[number].jpg for DAPI images.

    [folder name]-A[number].jpg for Rh-ph images.
22. Java runtime version is 1.4.2 or later. Image processing software CalMorph (ver 1.1) can be downloaded from the SCMD [8] site (http://scmd.gi.k.u-tokyo.ac.jp/datamine/). Meanings of 501 morphological features are described previously [4]. A detailed manual, helpful tools, and original data can be downloaded from the SCMD site and http://www.yeast.ib.k.u-tokyo.ac.jp/CalMorph.
23. The Jonckheere-Terpstra test was performed using R (http://www.r-project.org/) to obtain the Z score.
24. The morphological data of 123 wild-type replicates and 4,718 nonessential deletion mutants can be downloaded from the SCMD site (*see* **Note 22**). Three replicates of the wild-type data had been discarded to have missing values which are indicated by a negative value as −1.
25. The Box-Cox power transformation is carried out as previously described [4]. The wild-type data x are transformed by the function $F(x)$ defined below:

$$F_{p,a}(x) > \begin{cases} \log(x+a) & \text{if } p = 0 \\ \dfrac{(x+a)^p - 1}{p} & \text{otherwise} \end{cases}$$

where p and a are transformation parameters which are chosen to minimize the Anderson-Darling static D as follows:

$$D = \max_{-\infty<r<\infty} \frac{|S(r) - N(r)|}{\sqrt{N(r)(1-N(r))}}$$

where S is the distribution function of the transformed wild-type data, and N is the distribution function of the fitted normal distribution. Then the transformed data are standardized as follows:

$$y = \frac{F_{p,a}(x) - E}{\mathrm{SD}}$$

where E is the mean of the transformed data, and SD is the standard deviation of the transformed data.

Acknowledgement

We thank Seiko Morinaga for technical assistance, and people in the Laboratory of Signal Transduction for fruitful discussions. This work was supported by grants for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan (24370002 to Y.O). S.O. and H.O. were research fellows of the Japan Society for the Promotion of Science.

References

1. Ohnuki S et al (2012) Analysis of the biological activity of a novel 24-membered macrolide JBIR-19 in *Saccharomyces cerevisiae* by the morphological imaging program CalMorph. FEMS Yeast Res 12:293–304
2. Ohnuki S et al (2007) Diversity of Ca^{2+}-induced morphology revealed by morphological phenotyping of Ca^{2+}-sensitive mutants of *Saccharomyces cerevisiae*. Eukaryot Cell 6:817–830
3. Ohtani M et al (2004) Development of image processing program for yeast cell morphology. J Bioinform Comput Biol 1:695–709
4. Okada H et al (2014) Distinct roles of cell wall biogenesis in yeast morphogenesis as revealed by multivariate analysis of high-dimensional morphometric data. Mol Biol Cell 25:222–233
5. Ohya Y et al (2005) High-dimensional and large-scale phenotyping of yeast mutants. Proc Natl Acad Sci U S A 102:19015–19020
6. Ohnuki S et al (2010) High-content, image-based screening for drug targets in yeast. PLoS One 5:e10177
7. Iwaki A et al (2013) Vanillin inhibits translation and induces messenger ribonucleoprotein (mRNP) granule formation in *Saccharomyces cerevisiae*: application and validation of high-content, image-based profiling. PLoS One 8: e61748
8. Saito TL et al (2004) SCMD: *Saccharomyces cerevisiae* morphological database. Nucleic Acids Res 32(Database issue):D319–D322

Index

A

Absorption, distribution, metabolism, excretion, and toxicity (ADMET) ... 216, 218
Acoustical spotter ... 57
Activity-based protein profiling (ABPP) ... 214, 282
Adenosine triphosphate (ATP) ... 13, 19–21, 25, 69, 202, 215
Adenylyl cyclase (AC) ... 78, 81–88
ADP-ribosylation factor 1 (ARF1) ... 255, 257, 260
Affinity chromatography ... 265, 283, 287, 288
Aging ... 93–100
Alamar Blue® ... 104
Amphotericin B ... 144, 147, 149, 151, 154
Ampicillin ... 144, 147, 201–205
Aqueous solubility ... 211, 214, 216–218, 220, 225, 230, 237, 240
Arabidopsis seedlings ... 111–124
Arabidopsis thaliana ... 111–124
Artemisinin ... 178
Assay interference ... 214, 215
AutoDock ... 245–247
Axin ... 64

B

Bacillus Calmette-Guérin (BCG) ... 188–190
Bacillus subtilis ... 189, 191
Barcode sequencing ... 299–317
Basal cell carcinoma (BCC) ... 218, 220
Benomyl ... 302, 306, 307, 312, 314
Berberine ... 177
β-catenin ... 5, 64, 66, 70, 71
β-Glucuronidase (GUS) ... 114
β-lactamase ... 203
Bioisostere ... 217
Biological activity assays ... 188–192, 194, 195
Bioluminescence ... 158, 160, 161, 163
Biotin ... 39, 253, 254, 264, 271, 276–277, 282
Bone morphogenetic protein (BMP) signaling ... 12

C

Caenorhabditis elegans (C. elegans) ... 135, 139–154, 157, 167
Calculated octanol-water partition coefficient (cLogP) ... 198, 219, 220
CalMorph ... 319–321, 323, 326
Cancer ... 4, 6, 20, 157, 218, 220
Carbenicillin ... 145, 147, 149, 154
Cardiac
- differentiation ... 43–60
- induction ... 50, 54, 56, 57

Cardiomyocyte ... 43, 44, 46, 51–55, 58–60
Cardiovascular
- biology ... 44
- development ... 167–174
- disease ... 157

Carfilzomib ... 214
Cell-free ... 63
Cell morphology ... 319
Chemical
- array ... 29–40
- biology ... 16, 209–221, 263
- genetics ... 16, 17, 111–124, 210–212, 216, 219, 287, 313
- genomic profiling ... 299–317
- genomics ... 44, 169, 299–301, 306, 310, 311, 314–316
- libraries ... 12, 20, 21, 24, 89, 128, 129, 168, 172, 177, 210, 226, 227
- probe ... 139, 209, 210, 216, 218, 240
- scaffold ... 212
- screen ... 6, 8, 11, 13, 17, 19, 20, 25, 112–114, 140, 141, 164, 170, 210

Chronological lifespan ... 93–100
Ciliobrevin A ... 219, 220
Ciprofloxacin ... 190, 191
Collagenase IV ... 45, 47, 48, 56
Competition-saturation transfer difference (c-STD) ... 199–201
Compound
- screen ... 141
- set enrichment analysis (CSEA) ... 18–21, 23, 24, 26

Jonathan E. Hempel et al. (eds.), *Chemical Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1263, DOI 10.1007/978-1-4939-2269-7,

© Springer Science+Business Media New York 2015

Critical point 179
Crosslinking 33, 36, 40, 257–258
Cyclic
  adenosine monophosphate (cAMP) 77–85, 87, 88
  guanosine monophosphate (cGMP).... 78, 79, 81, 82, 84, 85, 87, 88
  nucleotide phosphodiesterase (PDE) 78, 80–88
Cycloheximide....65, 68
Cyclopamine....219, 220
Cypridina luciferase (CL) 5, 6, 8, 13, 14
Cytochalasin D....65, 67, 68

D

Development....33, 44, 55, 59, 64, 77, 87, 94, 112, 124, 128, 129, 131, 143, 146, 149, 151, 153, 167, 169, 173, 178, 209–221, 226, 288
Diabetes....19, 20, 158
Didemnin B (DB) 292, 294
Dietary restriction 94–96, 98, 99
Digitonin.... 104, 105, 107–109
Dimethyl sulfoxide (DMSO) 7, 22, 25, 35, 36, 39, 45, 65, 69, 71, 79, 84, 86, 88, 96–98, 105, 108, 113, 118–121, 123, 130, 142, 153, 154, 160, 161, 164, 169–171, 173, 190–192, 201–203, 205, 206, 293, 315, 325
Disease allele 15–26
Disulfiram 288
Diversity-oriented synthesis 225
Docking....243–249, 292
Dorsomorphin....169
Dot blotting....4, 6, 8–12
Drosophila melanogaster.... 128, 134, 135
Drug
  affinity responsive target stability (DARTS)....287–297
  discovery....8, 103, 128, 158, 177, 178, 187, 198, 209–211, 217, 226, 243, 287
  -induced toxicity....161, 164
  repurposing....17
  screen....103–110, 128
  target discovery....128

E

Edge effect....11
Embryoid body (EB).... 47–50, 55–58
Equilibrium dissociation constant (K_d)....197
Escherichia coli (E. coli) 80, 147, 198
Ethyl methanesulfonate (EMS).... 111, 112, 114
Extraction....80, 177–184, 187, 272, 279, 289, 294, 303, 307–310, 315

F

False positive 11, 13, 39, 163, 164, 173, 214, 215
Fasting metabolism....157–164
fbp1....77
Firefly luciferase (FL)....5, 8, 13, 64, 158, 215
Fission yeast 77–90, 94
FK506. *See* Rapamycin
FKBP12 ....30, 296
5-fluoro-orotic acid (5FOA).... 77–79, 81–86, 88–90
Fluorous
  solid phase extraction (FSPE) 254, 255, 258, 259
  tag.... 254–256, 259, 260
Forward
  chemical genetics....210–212
  genetics....139
Fragile X mental retardation protein (FMRP) 128
Fragile X syndrome (FXS)....128, 129
Fragment-based drug design (FBDD) 197–207
Fragment-based screening (FBS) 197
Fruit fly.... 127, 128, 131

G

GANT-61 ....219, 220
$G\alpha_s$....78, 85
Gaussia luciferase 13
GDC-0449....219, 220
GenePattern software....21, 22, 26
Gene set enrichment analysis (GSEA)....18, 24
Genetic
  interaction.... 15–26, 300, 313, 314
  screens 111–124, 139, 210–212, 216
Gluconeogenesis....77, 157, 158
Green fluorescent protein (GFP) 56, 78, 83, 85, 86, 90, 114, 146, 169, 188, 191, 195

H

H89 211
Half-maximal effective concentration (EC_{50})....106–109, 214, 219, 220
Half-maximal inhibitory concentration (IC_{50}) 114, 197, 201, 215, 293, 295
Hedgehog signaling....210, 218–221
High content
  imaging.... 25, 52, 54, 59, 151
  screening....43–60, 167–174
High performance liquid chromatography (HPLC)....80, 181, 183, 187–190, 192–194, 255, 272, 273
High-throughput....4, 10, 20, 29, 93–100, 115, 142, 168, 220, 315, 319
High-throughput screening.... 11, 13, 19, 63–72, 77–90, 94, 106, 128, 129, 135, 139–154, 177, 197, 209, 219, 288, 314
Hit
  identification.... 202–203, 207, 296
  validation 214–216
HNF4α....19, 20

Human
diseases ... 15–26, 127, 128, 134, 135, 211
embryonic stem cells (hESC) ... 46–49, 51, 52, 54–56, 58, 59
pluripotent stem cells (hPSC) ... 43–60

I

Image-based approach ... 320
Image-based target prediction ... 319–327
Image processing program ... 319
Immunoblotting ... 291, 292
Immunostaining ... 29, 31, 46, 51–53, 57
Inhibitor of Wnt response (IWR) ... 45, 52, 56–58
Interacting protein ... 254, 255, 260
In vitro assay ... 297
Isoniazid ... 190, 191, 194
Isotope-coded ... 255, 258–260

K

Kinase inhibitor ... 16, 169, 211, 215
KNIME software ... 21, 22
Kras ... 6, 12

L

Labeling efficiency ... 269–270, 274–275
LEAFY ... 111
Libraries ... 3–6, 11–13, 20–22, 24, 29, 30, 33, 36, 39, 58, 66, 86, 89, 93, 94, 109, 115, 128–130, 136, 139–143, 150, 153, 160, 162, 163, 167–170, 172, 177, 198, 200–202, 204, 207, 210, 212, 214, 216, 225–241, 243–249, 310
Library design ... 225–241
Ligand-based NMR ... 199–201, 204
Ligand screening ... 29, 31
Locomotor activity ... 132, 136, 137
Longevity ... 94, 97, 99, 100
Luciferase assay ... 6, 8, 11, 14, 160
Lymphoblastoid cell line (LCL) ... 20–22, 25
Lymphoblasts ... 19, 20, 25

M

Macrocycles ... 226, 227, 239
Mass spectrometry (MS) ... 253, 254, 264, 273, 280, 288, 289, 291, 294, 296
Maturity onset diabetes of the young type 1 (MODY1) ... 19, 20, 24
Medium rings ... 226, 227, 238, 239
Metabolites ... 104, 288
Metabolomics ... 287
Methicillin-resistant *Staphylococcus aureus* (MRSA) ... 189, 191
Methotrexate (Mtx) ... 288, 293
Methyl methanesulfonate (MMS) ... 300, 302, 306, 307, 314
Metronidazole ... 103, 107, 109
MG132 ... 69
Microtiter plate ... 80, 82, 94, 96–100, 140, 143, 149–150, 190, 321
Miniculture ... 94
Minimum inhibitory concentration (MIC) ... 191, 192, 195
Molecular docking ... 243–249
Morphological profiling ... 319, 324
Mouse embryonic fibroblasts (MEF) ... 44, 46–50, 54, 56, 219
Multiplexing ... 11–13, 300
Mutant collection ... 299, 300
Myosin heavy chain 6 (MYH6) ... 45–47, 51–55, 57, 60

N

Nanoluc luciferase ... 13
Natural products ... 13, 24, 30, 94, 136, 140, 141, 177–184, 188, 189, 219, 225–241, 243, 254, 288, 294
Neuraminidase ... 244–249
Neuronal degeneration ... 151
NMR-based functional assay ... 200, 201, 203–206
Nodal signaling ... 58
Notch signaling ... 12
Nuclear magnetic resonance (NMR) ... 197–207, 244

O

Olaparib ... 288, 293
Omigapil ... 288
OP50 ... 143, 147–152
Open Babel ... 246
Orthogonal array ... 182
Overnutrition ... 95, 96, 98, 99

P

p53 ... 5, 6, 9, 12
p38α MAP kinase ... 30, 38
Phenotypic screening ... 3, 128, 163, 168, 215, 220, 221, 243
Phosphatidylinositol-4,5-bisphosphate (PIP_2) ... 255–257, 259, 260
Phosphodiesterase (PDE) ... 78, 80–88
phosphoenolpyruvate-carboxykinase 1 (pck1) ... 157–160, 162, 163
Phosphoinositide 3-kinase (PI3K) ... 255
Photoaffinity
labeling (PAL) ... 253–260
linker ... 30–33, 35
Photo-cross-linking ... 30
Physicochemical properties ... 189, 192, 214, 216, 229, 238

Pintool 21, 25, 57, 58
Platelet-derived growth factor receptor, alpha polypeptide (PDGFRA) 45, 47–50
Pluripotent stem cells (PSC) 43–60
Polyethylene glycol (PEG) 30–33, 35, 283
Polymerase chain reaction (PCR) 6, 12, 78, 80, 116, 120, 160, 300, 303, 307–310, 316
Principal component analysis (PCA) 225–241, 323
Probe development 209–221
Proline 30, 31, 33, 35, 78, 274, 283
Propidium iodide (PI) 104–109
Protease inhibitor cocktail 289, 291, 292
Protein degradation 71
Protein kinase A (PKA) 77–90, 94, 211
Protein turnover 64, 66–68, 71
Proteomics 115, 254, 263–285, 288, 294
Pseudomonas aeruginosa 189
PubChem 228, 245, 246
Pulldown 264, 266, 267, 270, 271, 275–277, 282–284
Purmorphamine 219, 220
Pyrabactin 112

Q

Quinabactin 112

R

Rapamycin 30
Renilla luciferase (RL) 4, 5, 8, 11, 13
Replicative lifespan 93, 94
Reporter assay 215, 217
Reporter cell line 12, 51, 52
Resazurin 104, 106
Resorufin 104, 105, 107–109
Reverse transcription quantitative polymerase chain reaction (RT-qPCR) 57, 116, 120, 122
Ring expansion 226, 227
RNA interference (RNAi) 134, 135
Robotnikinin 219, 220
R software 303
Rule
  of five 210
  of three 198, 201

S

Saccharomyces cerevisiae 93, 94, 99, 299, 300, 319, 320
SANT-1 .... 219, 220
Saridegib 220
Saturation transfer difference (STD) 199–204, 206
Saturation transfer double difference (STDD) 199, 201, 203, 204, 206, 207
SB203580 .... 30, 38
Schizosaccharomyces pombe 77, 78, 80, 81, 84–87, 93–100, 299
Screening 3–5, 10–12, 16, 17, 19, 20, 22, 25, 26, 29–31, 34, 37, 40, 43–72, 77–90, 93, 94, 103–124, 127–137, 139–154, 157–164, 167–174, 177, 187, 197–207, 209–218, 220, 221, 240, 243–249, 263, 288, 292, 302, 305–307, 314, 316
Screening hit 202, 205, 210–218, 221
Secondary screen 118–120, 136
Sialic acid 244–249
Signal-to-noise ratio (SNR) 18, 23, 24, 26, 110, 158
siRNA 4, 5, 12, 45, 51, 55, 58
Size-exclusion chromatography (SEC) 188–190, 192, 193
Small molecules 3–26, 29, 30, 44–47, 50, 55, 58, 59, 63–72, 78, 79, 84, 107, 112, 113, 128, 130, 136, 157–161, 163, 168–172, 189, 209–221, 226, 243–249, 253–260, 263–285, 287–297
Small molecule screening 26, 49, 63–72, 78, 127–137, 139–154, 159, 171, 243–249
Small RNAs 46, 55
Smoothened agonist (SAG) 219, 220
Solid medium 77–80, 111–124
Solubility, purity, and aggregation of the molecule (SPAM) 198
Solvent extraction 178–179, 181
Stable isotope labeling in cell culture (SILAC) 264–266, 268, 269, 280
Staphylococcus aureus 189, 191
Staurosporine 211
Structure-activity relationship (SAR) 213, 215–218, 264
Supercritical fluid extraction (SFE) 178–180, 182–184
Surface plasmon resonance (SPR) 197
Synthetic chemistry 212

T

Target
  identification 44, 63, 114–115, 120, 128, 141, 168, 174, 215, 257, 263–285, 287–297
  of rapamycin (TOR) 93, 96
  selectivity 211
Taxol 178
TGFβ 44, 58
Thiazovivin 45, 47
Tinidazole 103
TP53 5
Traditional Chinese medicines (TCMs) 177–184, 187–195
Transgenic reporter zebrafish 158, 160
Trichomonas vaginalis 103–110
Trichomoniasis 103
Trifluoromethylaryldiazirine (TAD) 30
Triiodothyronine (T3) 45, 52, 55
Tryptic digest 264, 265, 271–272, 277–280, 283, 284

V

Vancomycin 190, 191
Vina 245, 247–248
Virtual molecular screening 244

W

Windorphen 169
Wnt
  inhibition 44
  signaling 64

X

Xenopus egg extract 63–72
Xenopus laevis 63

Y

Y-27632 128
Yeast deletion collection 300, 305, 308, 314

Z

Zanamivir 244–246, 248, 249
Zebrafish
  development 157, 167–174
  larvae 157–159, 161, 163, 164
  screening 157–164, 167–174
Z-factor score 68–69