

Methods in
Molecular Biology 800

Edward D. Zanders Ed

Chemical

METHODS IN MOLECULAR BIOLOGY™

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

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

Chemical Genomics and Proteomics

Reviews and Protocols

Edited by

Edward D. Zanders

BioVillage Ltd., Cambridge, UK

 **Humana Press**

Editor

Edward D. Zanders
PO Box 133
Royston, Hertfordshire SG8 0BX
United Kingdom
ed.zanders@btinternet.com

ISSN 1064-3745 e-ISSN 1940-6029
ISBN 978-1-61779-348-6 e-ISBN 978-1-61779-349-3
DOI 10.1007/978-1-61779-349-3
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011937229

© Springer Science+Business Media, LLC 2012

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Humana Press, c/o Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

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

Preface

The discipline of chemical genomics emerged from chemical biology around the year 2000, so it is now nearing its teenage years. An earlier version of this book was published in 2005; since then, technology has improved, and chemical genomics research has delivered new biological probes and drugs. Proteins have, of course, always been at the heart of the technology, but another change over recent years has been the more explicit use of the term “chemical proteomics,” which is why the word “proteomics” is included in the title of this updated book. *Chemical Genomics and Proteomics: Reviews and Protocols* contains updated reviews of the chemistry of small molecules and their interaction with protein targets. The protocols cover different types of ligands, carbohydrates, lipids and the generation of their protein targets; methods for measuring their interactions are also covered.

The first review by Zanders provides a reminder of the relatively short history of chemical genomics (including genetics and proteomics) and gives some examples of problems of current biological interest, such as epigenetics, where chemical technologies are now playing a role. This is followed by an overview of chemical space by Yung-Sing Wong, in which he describes new chemical synthesis tools, such as the stereoselective multicomponent reactions that are being used to produce sophisticated drug-like molecules for screening. The importance of *in silico* design is not forgotten, as a review by Bernardo and Tong makes clear. There are several ways of measuring the interactions of small molecule ligands with proteins, and each has advantages and disadvantages. Surface plasmon resonance (SPR) is particularly interesting because it requires no modification of the ligand and also provides kinetic data. The throughput of SPR is improving all the time, which is why a review chapter has been devoted to this subject. Nico de Mol describes the recent introduction of array technology for SPR analysis as well as the integration with mass spectrometry to identify unknown ligands.

The protocols begin with methods for capturing ligands on affinity supports. El-Khoury and colleagues in Christopher Lowe’s group, design new ligands for immunoglobulin purification using multicomponent reactions. Kanoh’s group presents methods for covalently attaching small molecules to affinity supports using photoactivatable cross-linking. This can be achieved without prior modification of the small molecule and is independent of functional groups on the molecule. Specific types of molecular interaction are described in chapters on detecting lipoproteins (Hannoush) and finding ligands for non-ATP-binding sites on protein kinases (Simard and Rauh). Microarrays are central to genomics and proteomics of all types; chapters by Blackburn and colleagues on protein arrays, and by Cummings and colleagues on glycan arrays, provide up-to-date information on creating and using these tools. Nanotechnology is being exploited in various assay types. Nie and co-workers describe protocols for measuring low levels of small molecule analytes by binding aptamer molecules labeled with gold nanoparticles. Advances in protein expression and analysis are covered in chapters by Peleg and Unger (developments in expression systems

using *Escherichia coli* and insect cells) and Bernhard et al. (cell-free synthesis of membrane proteins). Finally, Webster and Oxley provide detailed protocols for identifying proteins using mass spectrometry.

We hope that this volume will provide inspiration to those who wish to use chemical genomics and proteomics in their work; equally, we hope that this young field develops into full maturity through the incorporation of the new biological and chemical technologies that are beginning to emerge here.

Cambridge, UK

Edward D. Zanders

Contents

| | |
|-------------------------------|-----------|
| <i>Preface</i> | <i>v</i> |
| <i>Contributors</i> | <i>ix</i> |

PART I REVIEWS

| | |
|--|----|
| 1 Overview of Chemical Genomics and Proteomics | 3 |
| <i>Edward D. Zanders</i> | |
| 2 Exploring Chemical Space: Recent Advances in Chemistry | 11 |
| <i>Yung-Sing Wong</i> | |
| 3 In Silico Design of Small Molecules | 25 |
| <i>Paul H. Bernardo and Joo Chuan Tong</i> | |
| 4 Surface Plasmon Resonance for Proteomics | 33 |
| <i>Nico J. de Mol</i> | |

PART II PROTOCOLS

| | |
|--|-----|
| 5 Biomimetic Affinity Ligands for Immunoglobulins Based on the Multicomponent Ugi Reaction | 57 |
| <i>Graziella El Khoury, Laura A. Rowe, and Christopher R. Lowe</i> | |
| 6 Preparation of Photo-Cross-Linked Small Molecule Affinity Matrices for Affinity Selection of Protein Targets for Biologically Active Small Molecules . . . | 75 |
| <i>Hiroshi Takayama, Takashi Moriya, and Naoki Kanoh</i> | |
| 7 Profiling Cellular Myristoylation and Palmitoylation Using ω -Alkynyl Fatty Acids | 85 |
| <i>Rami N. Hannoush</i> | |
| 8 Fluorescence Labels in Kinases: A High-Throughput Kinase Binding Assay for the Identification of DFG-Out Binding Ligands | 95 |
| <i>Jeffrey R. Simard and Daniel Raub</i> | |
| 9 Electrochemical Aptamer Sensor for Small Molecule Assays | 119 |
| <i>Xin Liu, Wang Li, Xiahong Xu, Jiang Zhou, and Zhou Nie</i> | |
| 10 Miniaturized, Microarray-Based Assays for Chemical Proteomic Studies of Protein Function | 133 |
| <i>Jonathan M. Blackburn, Aubrey Shoko, and Natasha Beeton-Kempen</i> | |
| 11 Glycan Microarrays | 163 |
| <i>Xuezheng Song, Jamie Heimborg-Molinaro, David F. Smith, and Richard D. Cummings</i> | |
| 12 Resolving Bottlenecks for Recombinant Protein Expression in <i>E. coli</i> | 173 |
| <i>Yoav Peleg and Tamar Unger</i> | |

| | | |
|----|--|-----|
| 13 | Recombinant Protein Expression in the Baculovirus-Infected Insect Cell System | 187 |
| | <i>Tamar Unger and Yoav Peleg</i> | |
| 14 | Systems for the Cell-Free Synthesis of Proteins | 201 |
| | <i>Lei Kai, Christian Roos, Stefan HaberstocK, Davide Proverbio, Yi Ma, Friederike Junge, Mikhail Karbyshev, Volker Dötsch, and Frank Bernhard</i> | |
| 15 | Protein Identification by MALDI-TOF Mass Spectrometry. | 227 |
| | <i>Judith Webster and David Oxley</i> | |
| | <i>Index</i> | 241 |

Contributors

- NATASHA BEETON-KEMPEN • *Division of Medical Biochemistry, Institute for Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa*
- PAUL H. BERNARDO • *Institute of Chemical and Engineering Sciences, Agency for Science Technology and Research (A*STAR), Singapore, Singapore*
- FRANK BERNHARD • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany*
- JONATHAN M. BLACKBURN • *Division of Medical Biochemistry, Institute for Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa*
- RICHARD D. CUMMINGS • *Department of Biochemistry and The Glycomics Center, Emory University School of Medicine, Atlanta, GA, USA*
- VOLKER DÖTSCH • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany*
- STEFAN HABERSTOCK • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany*
- RAMI N. HANNOUSH • *Department of Early Discovery Biochemistry, Genentech, Inc., South San Francisco, CA, USA*
- JAMIE HEIMBURG-MOLINARO • *Department of Biochemistry and The Glycomics Center, Emory University School of Medicine, Atlanta, GA, USA*
- FRIEDERIKE JUNGE • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany*
- LEI KAI • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany*
- NAOKI KANO • *Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Japan*
- MIKHAIL KARBYSHEV • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany*
- GRAZIELLA EL KHOURY • *Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK*
- WANG LI • *State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha, China*

- XIN LIU • *State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha, China*
- CHRISTOPHER R. LOWE • *Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK*
- YI MA • *School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, China*
- NICO J. DE MOL • *Department of Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands*
- TAKASHI MORIYA • *Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Japan*
- ZHOU NIE • *State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha, China*
- DAVID OXLEY • *Proteomics Research Group, Babraham Institute, Babraham, UK*
- YOAV PELEG • *Israel Structural Proteomics Center, Faculty of Biochemistry, Weizmann Institute of Science, Meyer Building, Rehovot, Israel*
- DAVIDE PROVERBIO • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany*
- DANIEL RAUH • *Fakultät Chemie, Chemische Biologie, Technische Universität Dortmund, Dortmund, Germany*
- CHRISTIAN ROOS • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany*
- LAURA A. ROWE • *Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK*
- AUBREY SHOKO • *Centre for Proteomic and Genomic Research, Cape Town, South Africa*
- JEFFREY R. SIMARD • *Chemical Genomics Centre of the Max Planck Society, Dortmund, Germany*
- DAVID F. SMITH • *Department of Biochemistry and The Glycomics Center, Emory University School of Medicine, Atlanta, GA, USA*
- XUEZHENG SONG • *Department of Biochemistry and The Glycomics Center, Emory University School of Medicine, Atlanta, GA, USA*
- HIROSHI TAKAYAMA • *Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Japan*
- JOO CHUAN TONG • *Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore*
- TAMAR UNGER • *Israel Structural Proteomics Center, Faculty of Biochemistry, Weizmann Institute of Science, Meyer Building, Rehovot, Israel*
- JUDITH WEBSTER • *Proteomics Research Group, Babraham Institute, Babraham, UK*
- YUNG-SING WONG • *Département de Pharmacochimie Moléculaire, UJF-Université Grenoble 1, Grenoble, France*
- XIAHONG XU • *State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha, China*

EDWARD D. ZANDERS • *BioVillage Ltd. Cambridge, UK*

JIANG ZHOU • *State Key Laboratory of Chemo/Biosensing and Chemometrics,
College of Chemistry and Chemical Engineering, Hunan University,
Changsha, China*

Part I

Reviews

Chapter 1

Overview of Chemical Genomics and Proteomics

Edward D. Zanders

Abstract

Chemical genetics, genomics, and proteomics have been in existence as distinct offshoots of chemical biology for about 20 years. This review provides a brief definition of each, followed by some examples of how each technology is being used to advance basic research and drug discovery.

Key words: Chemical genetics, Chemical genomics, Chemical proteomics, Connectivity map, Epigenomics, Affinity selection, Kinases

1. Introduction

Ever since genomics emerged from the more venerable disciplines of biochemistry and molecular biology in the early 1990s, a whole new world of “omics” technologies has entered the scientific lexicon. Two of these, chemical genomics and chemical proteomics, are the subjects of this review and of course the entire book. Strictly speaking, both of these subjects are included in the general field of chemical biology; the first paper to mention the latter by name was published in 1930 (1) and since then, thousands more have been published in a steadily growing literature.

It is possible to follow the development of a new field by counting the numbers of publications that explicitly mention it. Figure 1 shows a graph of the number of publications on chemical genetics, chemical genomics, and chemical proteomics listed in PubMed every year from 1990 to the end of 2010.

The first point to note is that all three subjects began to emerge at the same time around the end of the 1990s. Chemical genetics and chemical genomics grew up together, but chemical proteomics,

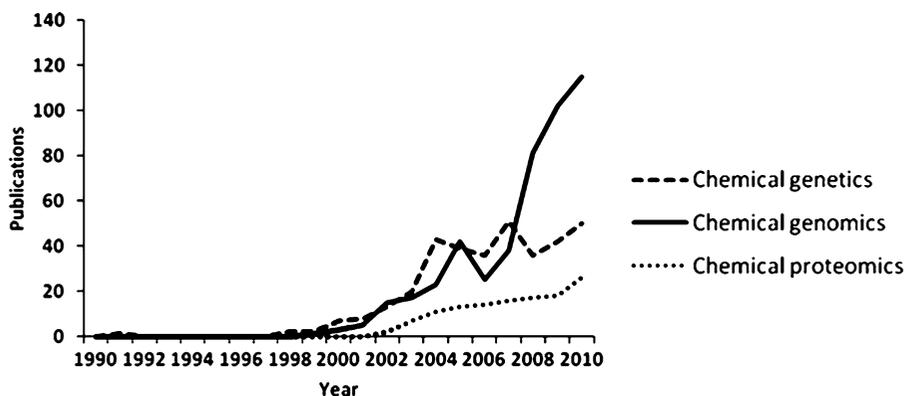


Fig. 1. Publications containing exact words listed in graph legend referenced in PubMed from 1990 to the end of 2010.

although always implicit in chemical genomics, has only recently begun to emerge as a named discipline as reflected in a more recent increase in publication rate. It is interesting to note that chemical genetics and genomics arose primarily from labs on the East Coast of the USA, in Harvard and companies such as Vertex Pharmaceuticals (2, 3). Chemical proteomics arose on the West Coast, notably in Cravatt's lab at the Skaggs Institute of Chemical Biology (4) and Bogyo's lab, then at Celera in South San Francisco (5). Of course this brief historical summary cannot do justice to the worldwide effort by scientists in academia and industry to discover more about the interactions between small molecules and the proteome. Apart from anything else, the boundaries get blurred between these very similar disciplines and publications that clearly describe chemical genomics and proteomics phenomena may not use this terminology.

The purpose of this review is to survey the state of chemical genomics and proteomics at the end of the first decade of the twenty-first century, some 20 years after the names were first coined. Some examples are given to illustrate how chemical "omic" technologies are being used to address problems of current biological and clinical interest.

2. Some Definitions

"For simplicity, we call this genome-wide genetic analysis with chemical inhibitors a chemical genomics approach." This quote from a paper by Chan et al. (6) provides a useful summary of what chemical genomics is all about. The key features of chemical genomics were laid out by Paul Caron in an earlier version of this book published in 2005 (7). A comprehensive review of chemical

genetics was published by Stockwell (8). Chemical proteomics has now entered the scene (for a review see ref. 9), so the formal list of definitions is as follows:

- **Forward chemical genetics**
The identification of targets for small molecules within whole cells; this is analogous to the creation of novel cellular phenotypes using mutagenesis. The molecular target has to be identified by deconvoluting the cellular system.
- **Reverse chemical genetics**
In this case, small molecules are identified for an isolated protein target (e.g., a kinase) and then used in cellular assays to gain information about the role of the target in vivo. This is also useful for assessing the suitability of compounds as small molecule drugs. The process is analogous to knocking out a specific gene using transgenic technology or RNA interference.
- **Chemical proteomics**
The interaction between small molecules and the proteome. More specifically, the small molecules are designed as activity-based probes which bind to the target; this means that the complex between the small and large molecule can be captured on an affinity support for further analysis. Activity-based proteomics of this type has been used to capture protein families, such as the hydrolases, and display them on SDS-PAGE gels (10). As with proteomics in general, protein identification through mass spectrometry is central to chemical proteomics. Some applications of the above technologies are described in the following sections.

3. The Connectivity Map

One of the hallmarks of chemical genomics technology is the profiling of cellular mRNA expression in response to small molecules. This transcriptomics approach generates a gene expression fingerprint which can be used to reveal subtle relationships between different compounds. Experiments of this type have been performed in individual labs for several years now, but the data sets may not be directly comparable and universally applicable. The first attempts to create a more universal database of compound-induced mRNA expression profiles used yeast cells (11) or in vivo responses in rat tissues (12). The latter were specifically designed for assessing the toxicology of xenobiotics and have proven capable of grouping compounds with similar properties according to gene expression.

In 2006, Lamb et al. published a description of the Connectivity Map (C-Map) which provides data on mRNA expression changes in human cell lines treated with 164 small molecules (13). It also provides the bioinformatic tools required to match different gene

expression fingerprints. Since many of the molecules were drugs with known targets and utility for treating particular diseases, it was possible to create a map connecting chemical genomics with physiology and pathology. One of the main problems with expression studies of this type is the variation in mRNA expression between different cell types and batches of the same cells. There is also the requirement to use a single technology platform to measure mRNA levels (the primary one used here was the Affymetrix GeneChip microarray). These problems were noticeable if standard methods of hierarchical clustering were used in the data analysis, so a different statistical process was developed and named Gene Set Enrichment Analysis (GSEA) (14). This is based on a “query signature” which is used as a reference set from which to compare different profiles. The result is a connectivity score that is either positive or negative, without any unit of measurement, meaning that the system is independent of the technology platform. The *Science* publication (13) described several applications of the C-Map, starting with validation using compounds with known mechanisms of action, such as estrogens and anti-estrogens, or HDAC inhibitors. It was also capable of assigning compounds with unknown functions to specific biological pathways, in this case the triterpenoid gedunin, which was shown to affect the HSP90 pathway in prostate cancer cells.

The C-Map data are freely available and have been used in other studies since the original publication in 2006; these studies include the search for novel antivirals for influenza (15) and the discovery of synergistic drug combination therapies for treating acute myeloid leukemia (16). The current (2011) C-Map data are based on 7,000 expression profiles representing 1,309 compounds (17).

4. Epigenomics and Small Molecules

Next generation DNA sequencing technology is rapidly expanding our understanding of genetic variation at the nucleotide level. This research will continue apace, but there is also a considerable amount of interest in the noninherited part of the genome, the epigenome. This is particularly relevant to stem cell biology (18) and diseases such as cancer (19); an understanding of different epigenetic control mechanisms could therefore lead to advances in tissue repair and new cancer treatments. The epigenome consists of methylated cytosine as well as modified histone proteins. From a chemical genomics perspective, it is these latter molecules that are of interest, because they involve both enzyme targets and protein–protein interactions. Histones are acetylated on specific lysine residues by histone acetyl transferase (HAT) leading to an opening up of the chromatin structure and greater accessibility to regulatory proteins,

resulting in enhanced gene activity (20). The opposing enzyme, histone deacetylase (HDAC) is the target for small molecules, such as trichostatin A and valproic acid, which have been shown to arrest the cell cycle and force the differentiation of tumor cells (21). There is a feeling, therefore, that small molecule inhibition of enzymes involved in chromatin structure could have therapeutic potential. The chemical genomics research group at RIKEN in Japan is currently employing forward chemical genetics to identify novel chromatin modulators. One of these, spliceostatin A, inhibits pre-mRNA splicing in the spliceosome complex giving rise to a phenotype that is similar to HDAC inhibition (21, 22). This suggests that histone modification and RNA splicing both contribute to the epigenetic regulation of chromatin structure and hence gene expression. The regulation of gene expression also involves so-called “reader proteins” that recognize epigenetic marks on histones (i.e., acetylation) caused by HAT, a “writer enzyme” and removed by HDAC, an “eraser enzyme” (23). Two laboratories have recently described small molecules that inhibit the binding of acetylated histones to members of the BET family of reader proteins (24, 25). The compound JQ1 and I-BET are remarkably similar triazolo-1,4-diazepines, the former inhibiting the proliferation of a squamous cell carcinoma and the latter inflammatory responses in macrophages. This is noteworthy for two reasons. First, it is possible to identify small molecules which target epigenetic mechanisms and yet can be selective for specific gene expression. Second, the inhibition of binding BET to acetylated histones involved breaking a protein–protein interaction, which for small molecules, seems to be the exception rather than the rule. There can be no doubt that chemical genomics (and chemical biology in general) will not only provide more tools to probe epigenetic phenomena, but also new drugs to target gene expression in specific diseases.

5. Chemical Proteomics

This final selection focuses on the use of small molecules to probe protein families containing a particular motif, such as a nucleotide-binding domain. These chemical proteomic techniques are beginning to reveal protein targets for marketed drugs that contribute to off-target effects (both beneficial and undesired). The protein kinase family is of major interest since its members are targets for current drugs and the enzymes are involved in important cell signaling pathways. The conserved ATP-binding motif in the kinase family is an obvious target for affinity selection using solid phase matrices. γ -phosphate-linked ATP-sepharose has been used to identify purine-binding proteins (26), but more selective analysis of kinases requires more specific probes. Wissing et al. (27) absorbed proteins

from cell lysates onto a series of immobilized kinase inhibitors, and then analyzed the bound proteins using SDS-PAGE and mass spectrometry. In this way, they found more than 140 different kinases and more than 200 different phosphorylation sites. Further improvements to the basic methodology have led to the kinobead system described by a group from Cellzome in which kinase binding to small molecules could be quantified using iTRAQ reagent to produce an internal reference tag for mass spectrometry (Bantscheff et al. (28) and reviewed in (29)). Seven ATP-binding ligands with broad affinity for a range of kinases and other nucleotide-binding proteins were used with cell lysates; the difference between this approach and other affinity chromatography methods lies in the quantification. This means that cells could be incubated with and without drug molecules of interest and the kinases that bound to the beads quantified using the mass spectrometry standards. Any kinase reduced in abundance relative to those from control cells was therefore bound by the drug. If the drug (in this case the Abl kinase inhibitors imatininib, dasatinib, and bosutinib) were used at different concentrations an IC_{50} curve could be produced for binding to over 500 different kinases. This powerful technology confirmed the fact that imatinib is highly selective for Abl tk and c-KIT, but also identified a new kinase target, discoidin domain receptor 1 (DDR1), as well as NQO2, an NADPH-dependent oxidoreductase. This, and similar chemical proteomic techniques are under continual development and have proven their worth in the identification of targets for clinically useful kinase inhibitors (e.g., ref. 30).

Finally, a small molecule affinity chromatography system has been used to address the long standing problem of thalidomide teratogenicity (31). The drug was prescribed to pregnant women as a sedative and treatment for morning sickness, but turned out to be teratogenic, leading to multiple defects including shortened limbs. Thalidomide is used to treat multiple myeloma and a complication of leprosy, so it would be useful to be able to identify the mechanisms of teratogenicity and produce cleaner compounds. Ito et al. used a thalidomide derivative (FR259625 containing a free carboxyl group) linked to magnetic ferrite glycidyl methacrylate beads (32). HeLa cell extracts were passed over the matrix and specifically bound proteins eluted using competition with thalidomide. Two proteins bound specifically, cereblon (CRBN) and damaged DNA-binding protein 1 (DDB1). The latter was then shown to bind indirectly to thalidomide by binding to cereblon in a functional E3 ubiquitin ligase complex. Thalidomide was shown to directly inhibit the ligase activity and produce limb defects in zebrafish. This impressive work demonstrates once again the enormous potential of chemical proteomics to produce safer drugs by designing out off-target effects such as illustrated with cereblon binding.

6. Conclusions

The fields of chemical genetics, genomics, and proteomics are in good health and continuing to advance basic research and pharmaceutical development. Confidence in the field is indicated by the presence of chemical genomics centers in different parts of the world, including at the NIH (33). This also indicates that small molecule screening is no longer the exclusive province of the pharmaceutical industry; the greater access to reagents and know-how which results from these initiatives can only be of benefit to academia and industry alike.

References

1. Leathes JB (1930) The Harveian oration on the birth of chemical biology. *Br Med J* 2:671–676
2. MacBeath G (2001) Chemical genomics: what will it take and who gets to play? *Genome Biol* doi:10.1186/gb-2001-2-6-comment 2005
3. Schreiber SL (1998) Chemical genetics resulting from a passion for synthetic organic chemistry. *Bioorg Med Chem* 6:1127–52
4. Jessani N, Liu Y, Humphrey M, Cravatt BF (2002) Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness. *Proc Natl Acad Sci USA* 99:10335–40
5. Jeffery DA, Bogyo M (2003) Chemical proteomics and its application to drug discovery. *Curr Opin Biotechnol* 14:87–95
6. Chen TF et al. (2000) A chemical genomics approach toward understanding the global functions of the target of rapamycin protein (TOR) *Proc Natl Acad Sci USA* 97:13227–32
7. Caron P (2005) Introduction to Chemical Genomics. In: Zanders ED (ed) *Chemical Genomics, Reviews and Protocols*. Humana Press, New Jersey
8. Stockwell BR (2000) Chemical genetics: ligand-based discovery of gene function. *Nat Rev Genet* 1:116–25
9. Rix U, Superti-Furga G (2009) Target profiling of small molecules by chemical proteomics. *Nat Chem Biol* 5:616–24
10. Kidd D, Liu Y, Cravatt BF (2001) Profiling serine hydrolase activities in complex proteomes. *Biochemistry* 40:4005–15
11. Hughes TR et al. (2000) Functional discovery via a compendium of expression profiles. *Cell* 102:109–26
12. Ganter B et al. (2005) Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J Biotechnol* 119:219–44
13. Lamb J (2006) The connectivity map: Using Gene-Expression Signatures to Connect Small Molecules, Genes and Disease. *Science* 313: 1929–1935
14. Subramanian A (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102:15545–15550
15. Josset L et al. (2010) Gene expression signature-based screening identifies new broadly effective influenza A antivirals. *PLoS One* doi:10.1371/journal.pone.0013169
16. Hassane DC et al. (2010) Chemical genomic screening reveals synergism between parthenolide and inhibitors of the PI-3 kinase and mTOR pathways. *Blood* 116:5983–90
17. <http://www.broadinstitute.org/cmap/> Accessed 15 February 2011
18. Barrero MJ, Izpisua Belmonte JC (2011) Regenerating the epigenome. *EMBO Rep* doi:10.1038/embor.2011.10
19. Berdasco M, Esteller M (2010) Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev Cell* 19:698–711
20. Eliseeva ED et al. (2007) Characterization of novel inhibitors of histone acetyltransferases. *Mol Cancer Ther* 6:2391–8
21. M Yoshida (2009) Chemical genomics: a key to the epigenome – an interview with Minoru Yoshida. *Int J Dev Biol* 53: 269–74
22. Kaida D et al. (2007) Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat Chem Biol* 3:533–5
23. Taverna SD Cole PA (2010) Reader's block. *Nature* 468:1050–1051

24. Fillipakopoulos P et al. (2010) Selective inhibition of BET bromodomains. *Nature* 468:1067–1073
25. Nicodeme E et al. (2010) Suppression of inflammation by a synthetic histone mimic. *Nature* 468:1119–1123
26. Graves PR et al. (2002) Discovery of novel targets of quinoline drugs in the human purine binding proteome. *Mol Pharmacol* 62:1364–72
27. Wissing J et al. (2007) Proteomics analysis of protein kinases by target class-selective prefractionation and tandem mass spectrometry. *Mol Cell Proteomics* 6:537–47
28. Bantscheff M et al. (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol* 25:1035–44
29. Peters EC, Gray NS (2007) Chemical Proteomics Identifies Unanticipated Targets of Clinical Kinase Inhibitors. *ACS Chemical Biology* 2:661–664
30. Rix U et al. (2010) A comprehensive target selectivity survey of the BCR-ABL kinase inhibitor INNO-406 by kinase profiling and chemical proteomics in chronic myeloid leukemia cells INNO-406 target profile in CML. *Leukemia* 24:44–50
31. Ito T et al. (2010) Identification of a Primary Target of Thalidomide Teratogenicity. *Science* 327:1345–1350
32. Sakamoto S et al. (2009) Development and application of high-performance affinity beads: Toward chemical biology and drug discovery. *The Chemical Record* 9:66–85
33. <http://www.ncgc.nih.gov/index.html> Accessed 16 February 2011

Chapter 2

Exploring Chemical Space: Recent Advances in Chemistry

Yung-Sing Wong

Abstract

Recent advances and concepts for exploring chemical space are highlighted in this chapter and show how the synthetic chemical world meets the demand of making large and relevant collection of new molecules for analyzing the biological world more closely.

Key words: Chemical space, Multicomponent reactions, Biology-oriented synthesis, Diversity-oriented synthesis, Divergent selectivity

1. Introduction

The terms chemical, biological, or pharmacological space have found widespread use in the last few years, more specifically in research programs involving high-throughput screening (HTS) (1–3); they encompass an even more global vision of the interconnectivity between chemistry and biology. Structural information obtained from chemical libraries in association with their performance in HTS bioassays can be qualitatively and quantitatively assessed according to descriptors, making possible data analysis and virtual screening (4). How can chemical space be defined in chemical biology or pharma arrays? Organic chemicals can be characterized by a wide range of descriptors, such as their physico-chemical properties (molecular mass, lipophilicity, etc.) or their topological features (such as molecular structure, chirality, etc.). Lipinski's "rule of five" is a good example of how a chemical space can be defined in compliance with four physicochemical descriptors which are molecular weight, hydrogen-bond donors, hydrogen-bond acceptors, and lipophilicity (5). Lipinski's analysis of the World

Drug Index has revealed that an orally active drug has no more than one violation of the following criteria: molecular weight less than 500 Da, number of hydrogen-bond donors less than 5, number of hydrogen-bond acceptors less than 10, and $\log P$ less than 5. These descriptors taken as a whole and in accordance with the latter guidelines, define a chemical space for predicting orally active drugs. Another chemical space definition is illustrated by the representation of three-dimensional (3D) molecular shapes through descriptors. Recent advances in easily and reliably capturing topological data from molecules according to 3D shape descriptors have provided routine tools for synthetic chemists (6, 7). A recent example is the use of normalized principal moments of inertia (PMI) ratios that is becoming popular for analyzing molecular shape-based diversity (8). Biological targets, like biomacromolecules, are chiral and highly complex to ensure functional specificity. Indeed, recent analysis of structural complexity of libraries shows that collections of molecules with higher numbers of carbon sp^3 centers have a greater propensity for succeeding at various stages of the drug discovery process (9, 10). As sp^3 carbon is a major source of chirality and a source of molecular complexity and diversity in synthesis, carbon saturation can be used as a measurement of structural complexity. This is defined by the fraction sp^3 (F_{sp^3}), where $F_{sp^3} = (\text{number of } sp^3\text{-hybridized carbon} / \text{total carbon in the molecule})$ (9).

In terms of making collections of molecules for biological screening, synthetic chemists have to design a way of introducing and modulating descriptor features prior to synthesis. The choice of reactions also depends on whether the synthetic chemist works for academia or the pharma world as shown in Fig. 1. However, and more importantly, both institutions share preferential criteria for reactions which must be efficient, practical, robust, functionally tolerant (without the need for protecting groups), and combinatorial (for rapidly increasing diversity). The suitability of pathways for rapid and relevant analogue synthesis is essential for HTS assays. Whereas pharma company practice more often defines a focused and restricted chemical space prior to starting library synthesis (11, 12), a more exploratory paradigm aspires to cover a much broader chemical space; this may not be covered by existing HTS programs, and it is expected that new regions in chemical space will reveal new and potent biological targets (13–15).

This brief review places emphasis on (1) the use of stereoselective multicomponent reactions to access diverse and complex pure analogues in a powerful combinatorial way and (2) to review recent synthetic pathways and concepts devoted to exploring chemical space (both defined or expanded).

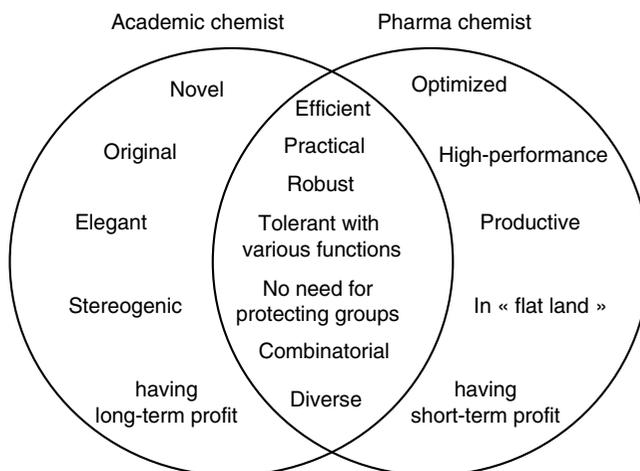


Fig. 1. Criteria required for a reaction to be selected in an array program from academic and pharma point of views.

2. Recent Advances in Stereoselective Multicomponent Reactions

Multicomponent reactions (MCRs) can be defined as three or more reactants that join together in a single synthetic step to form new products containing portions of all the components. These time and cost-effective reactions are powerful tools that are applicable to combinatorial and parallel syntheses in particular (16–20). However, MCRs for library synthesis are often selected to produce only high quantities of new compounds rather than high quality products (i.e., more diverse and chiral products in pure form). Thus, recent efforts have been made to offset this trend by building chirality into collections of new compounds with the help of efficient stereoselective MCRs (21).

2.1. Diastereoselective Multicomponent Reactions

Figure 2 represents four unique three-component reactions (3CR) (eq 1–4) in which one chiral center is set up on one reactant (symbolized by a dashed square) which can then induce a high diastereoselectivity. The first two reactions (eq 1 (22) and eq 2 (23)) are suitable for making chiral ethers **1** and **2** with *syn* configuration, whereas the last two reactions (eq 3 (24) and eq 4 (25, 26)) are designed to yield chiral amines displaying in either the *syn* **3** or *anti* **4** configuration.

2.2. Enantioselective Multicomponent Reactions

The preparation of optically pure collections of products by enantioselective MCR implies the action of efficient chiral catalysts. The Biginelli reaction that assembles aldehydes, urea/thio-urea, and enolizable carbonyls into dihydropyrimidinethiones **6** (Fig. 2, eq 5) has for a long time been selected as the MCR of

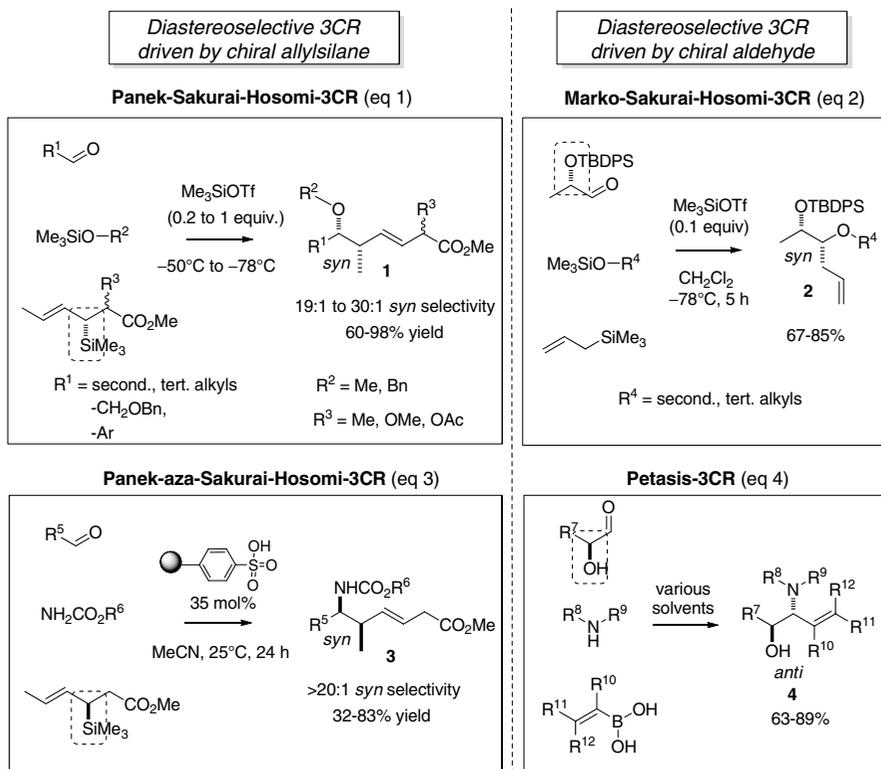


Fig. 2. Diastereoselective three-component reactions (3CR).

choice for gaining access to biologically active libraries (27). However, the control of stereochemistry in the formation of the new chiral center (symbolized by *) has been a long-standing challenge and has only recently been solved by the use of the Brønsted acid catalyst **5** having a broad tolerance of functional groups (eq 5) (28). Another milestone is the poor stereocontrol generally observed in MCR with isocyanide **7** (eq 6, see Fig. 3), a powerful functional group implicated in many MCR patterns (29). Interestingly, recent examples have shown that the enantioselective Passerini reaction can be efficiently catalyzed by chiral Lewis acids such as tridentate indan (pybox) Cu(II) **8** (eq 6) (30) or chiral salen-complex (31).

An enantioselective variant of Petasis reaction catalyzed by the chiral biphenol **10** was applied to the synthesis of optically active amines **11** (eq 7) (32). Enantioselective addition of alkyne catalyzed by the chiral copper-phosphine ligand in MCR offers a practical pathway to obtain homochiral propargylamines like **12** (eq 8) (33). The last decade has also witnessed the emergence of organocatalysts, derived from natural products like amino acids (with more emphasis

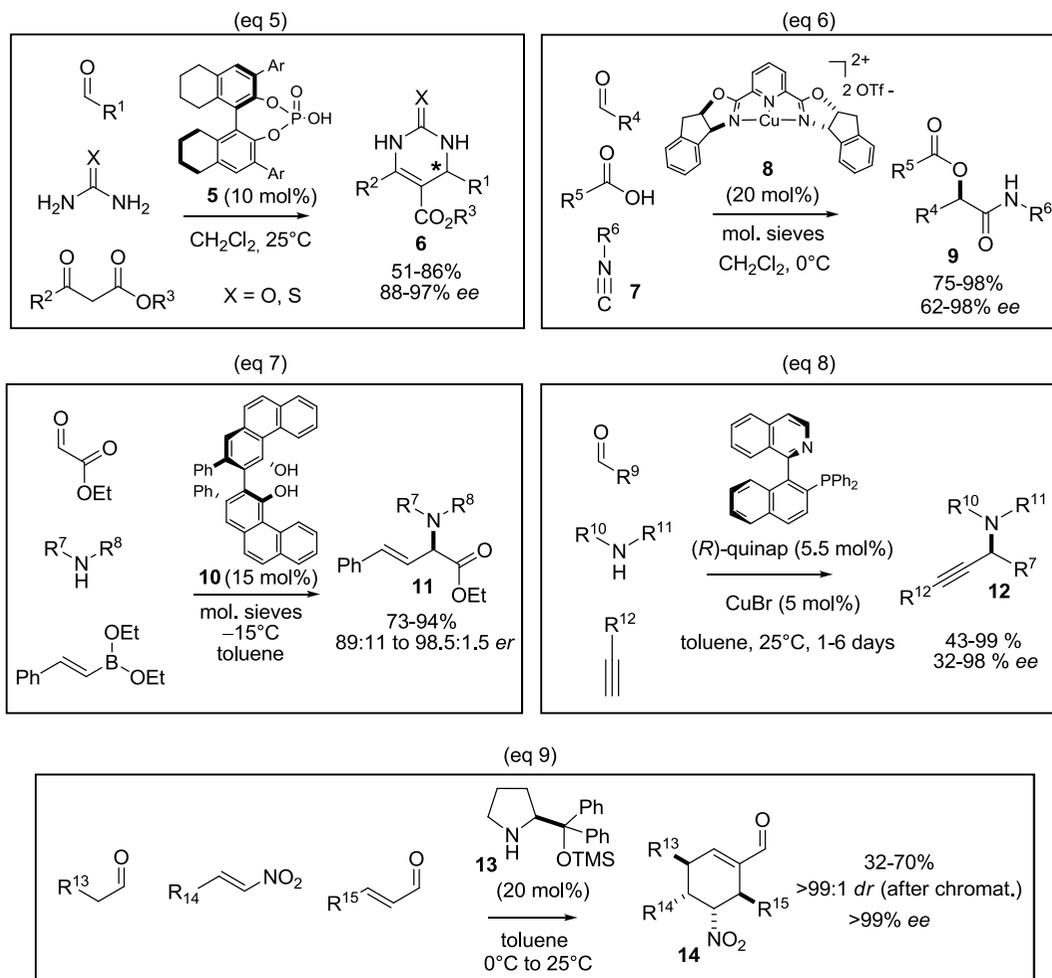


Fig. 3. Enantioselective MCRs.

on proline), as an impressive, though simple, tool for catalyzing various enantioselective MCRs (34, 35). Starting from three simple achiral reactants and in the presence of a catalytic amount of prolinol derivative **13**, a six-membered ring was created with the stereocontrol of four new chiral centers giving access to the optically pure product **14** (36).

3. Recent Concepts in the Exploration of Chemical Space

3.1. Focused Structural Diversity: Biology-Oriented Synthesis

Protein folds and natural product scaffolds are highly conserved in nature. From this observation, Waldmann proposes a concept termed biology-oriented synthesis (BIOS) (37–39). The idea rests on the hypothesis that proteins showing similar tertiary and quaternary structures (similar folding) should have the propensity

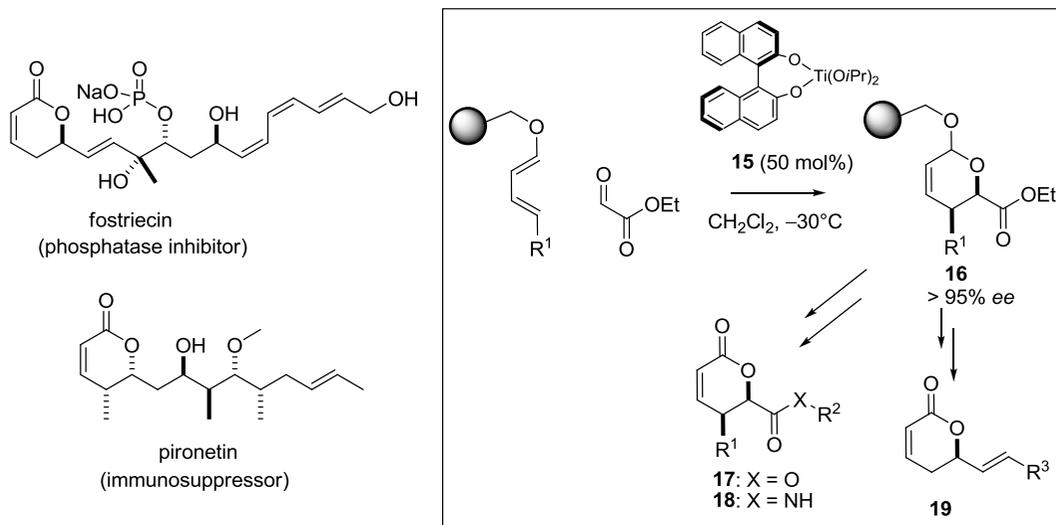


Fig. 4. Solid-phase synthesis of biologically relevant products inspired by natural α,β -unsaturated δ -lactones.

to bind natural products with similar scaffolds. Thus, synthetic chemists have to design pathways that connect natural scaffolds to each other according to a biologically prevalidated scaffold tree. In addition, natural product scaffolds are structurally complex and rich in chiral centers; they also cover chemical space that is distinct from most existing synthetic libraries (40, 41). Furthermore, nature provides efficient biomimetic reactions or pathways from which inspiration can be drawn to make compound collections (42).

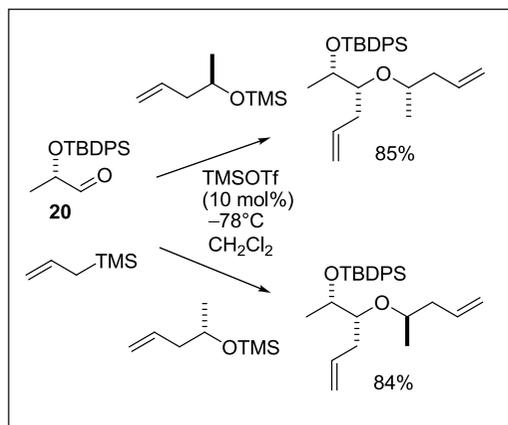
A α,β -unsaturated δ -lactone library, inspired by bioactive natural products such as fostriecin and pironetin (Fig. 4), has been synthesized on a solid support in a highly enantioselective manner using the Lewis acid Ti-(*R*)-BINOL catalyst **15** (43). After subsequent transformations, a collection of 50 analogues containing structures **17**, **18**, and **19** were subjected to biological evaluation in phenotype-based screens. A few of them have been shown to be modulators of cell cycle progression and inhibitors of viral entry into cells.

3.2. Expanded Structural Diversity: Diversity-Oriented Synthesis

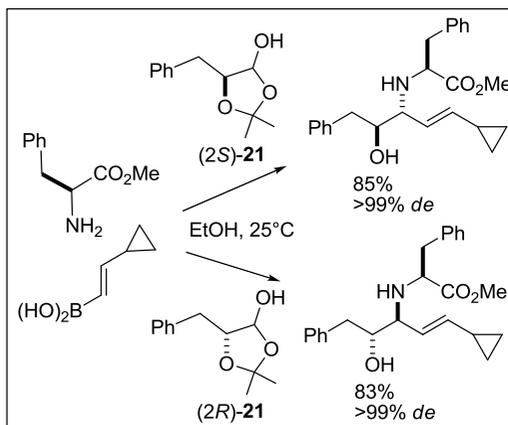
Rather than being directed toward a single and defined biological target, diversity-oriented synthesis (DOS) (13, 44–49) aims at making collections of structurally diverse molecules to gain access to wider, and often unexplored, chemical space. Thus, new strategies for library design deal with the challenge of maximizing stereochemical, appendage, and skeletal diversity within a minimal number of steps. These approaches make use of the concepts of *divergent selectivity* in either stereochemistry, reactivity, or by combination of both using short synthetic pathways.

**Divergent diastereoselective 3CR
dominated by chiral aldehyde**

Sakurai-Hosomi-3CR (eq 10)

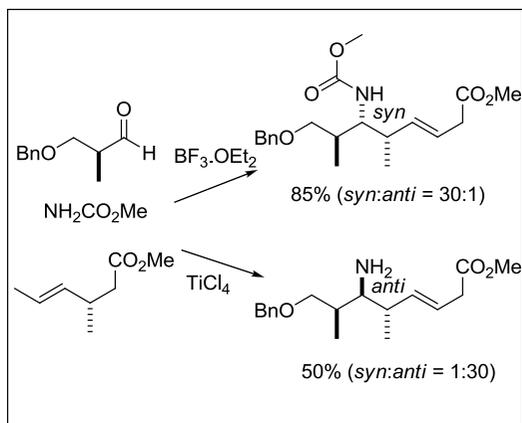


Petasis-3CR (eq 11)



**Divergent diastereoselective 3CR
driven by reagent**

Aza-Sakurai-Hosomi-3CR (eq 12)



**Divergent diastereoselective 3CR
driven by steric effect**

Roush "one-pot" reaction (eq 13)

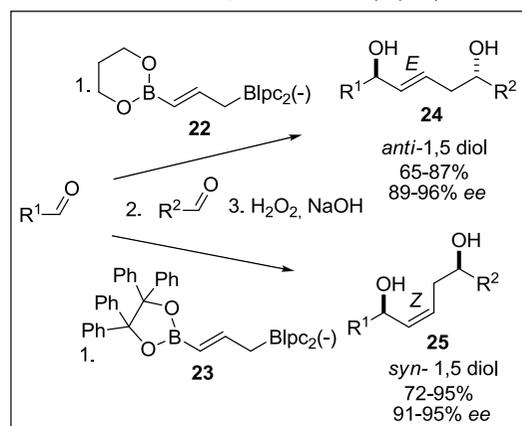


Fig. 5. Divergent diastereoselective three-components and "one-pot" reactions.

**3.2.1. Divergent
Diastereoselective
Multicomponent Reactions**

The difference between two diastereomers can be seen as the difference between the spatial orientation of their appendages. Divergent diastereoselective MCR is a straightforward method which gives rise to a unique diastereomer in each vessel. One strategy rests on the use of two chiral reactants where one chiral inductor dominates the other. Two examples are represented in Fig. 5 where the chiral aldehydes **20**, (2*S*)-**21**, and (2*R*)-**21** act as the dominant stereogenic inductor (eq 10 (**23**) and eq 11 (**50**)). Use of different reagents can also end up forming two distinct diastereoselectivities (eq 12 (**51**)). In Roush's "one-pot" double allylboration reaction

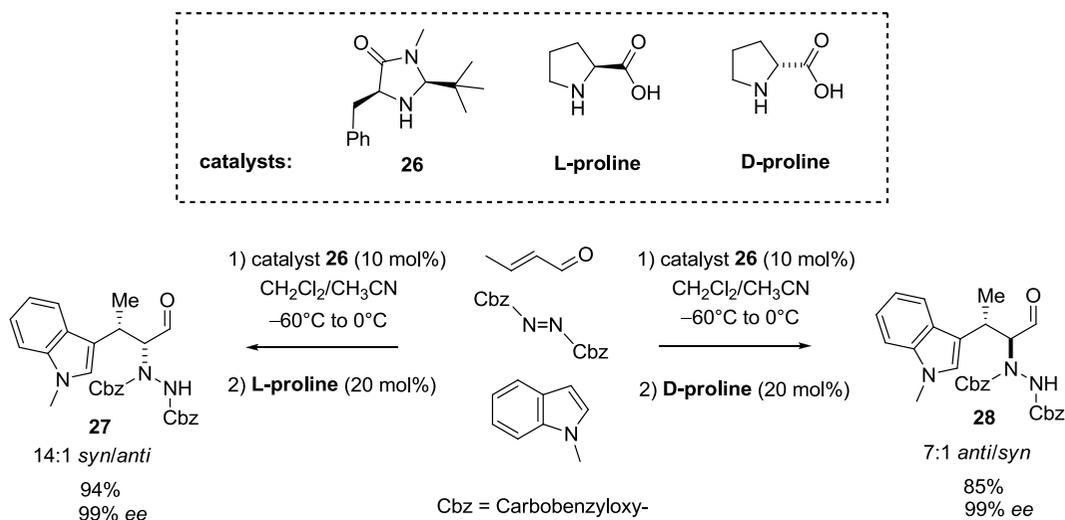


Fig. 6. Dual catalyst system in enantioselective and diastereodivergent 3CR.

sequence (**52**), the difference in steric hindrance between boronate esters **22** and **23** results in orienting the stereochemistry to form either (*E*)-1,5-*anti* diols **24** or (*Z*)-1,5-*syn* diols **25**, respectively.

3.2.2. Enantioselective Diastereodivergent Multicomponent Reactions

Starting from achiral reactants, dual and highly enantioselective catalyst system can govern the overall stereodivergent process. Ingenious examples in the field of organocatalysis have been reported (**53**). For instance, the combination of catalyst **26** and homochiral proline (Fig. 6) led to an optically pure diastereomeric product, either **27** and **28**, depending upon whether L or D proline was used (**54**).

3.2.3. The Build/Pair Strategy

The diversity-generating process can also be conducted over two steps with a Build/Pair strategy. Ideally, the first step would involve a MCR which assembles and installs (the Build) combinations of fragments having all the diversity and functionality required for performing divergent post-transformations (the Pair) (**55**). The Ugi-four-component reaction (Ugi-4CR) is a powerful build process involving an isocyanide group **29** (Fig. 7) as key reactant which forms peptide-like products such as **30**, **31**, **32**, and **33**. Subsequent conversions into novel skeletons are then possible to give diverse new scaffolds such as **34**, **35**, **36**, and **37**. Identification of lead compounds as antagonists of protein Bcl-x_L has been reported using such a Build/Pair approach (**56**).

3.2.4. The Build/Couple/Pair Strategy

After the build process the diversity potential can be enriched by adding a new fragment with functionalities which offer new pairing combinations with existing functional groups (Couple). An illustration of this Build/Couple/Pair strategy is shown in Fig. 8 (**57**).

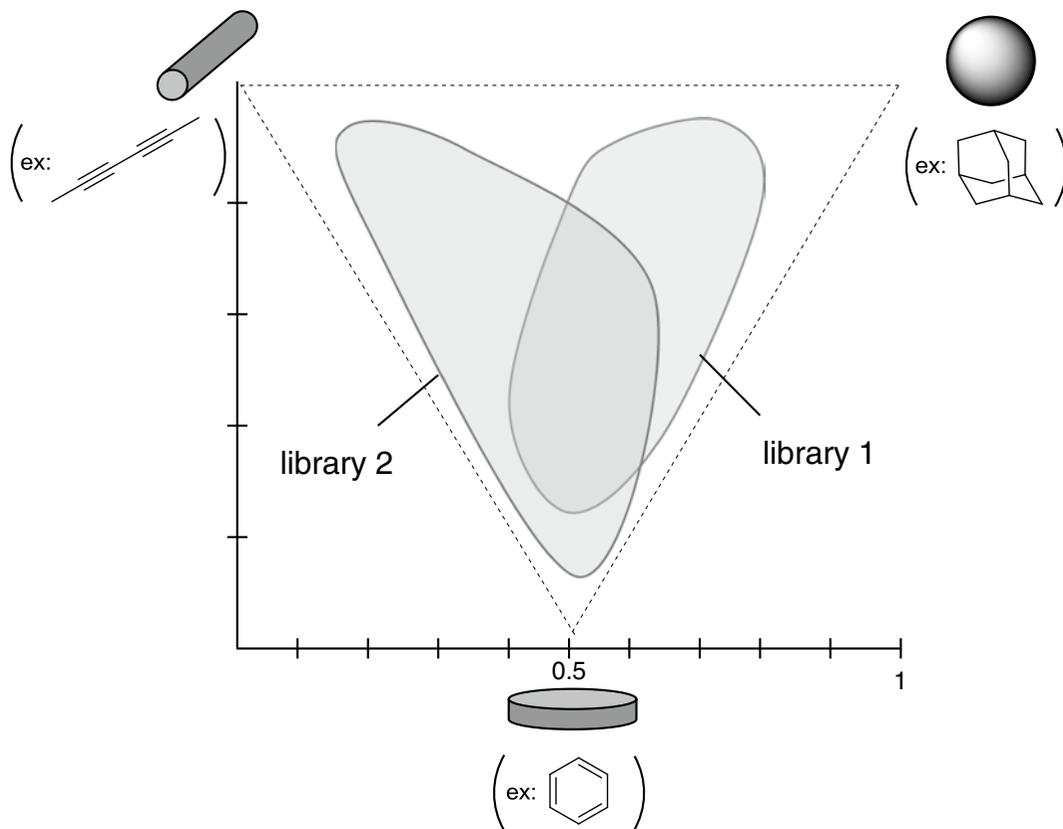


Fig. 9. Molecular shape analysis of libraries 1 and 2 by normalized principal moments of inertia (PMI) ratio plots.

Interestingly, the description of shape using normalized PMI ratios (Fig. 9) reveals that the collection of molecules issued from **38** (library 1) populates a structural region between sphere and disc shaped, whereas compounds issued from **39** (library 2) are more located in rod and disc-like regions. In this example, DOS strategy has succeeded in covering a large chemical space using no more than 73 compounds, with only five steps required for the longest synthetic pathways.

An aldol-based Build/Couple/Pair strategy has recently been applied for the discovery of new histone deacetylase inhibitors (**58**).

4. Conclusion

Targeting defined or expanded chemical space has inspired chemical and conceptual innovations that facilitate and forge close links between chemistry and biology. This cross disciplinary approach strives to reveal hidden connections between drugs (chemical space defined by small molecules) and diseases (space defined by biological targets) and will certainly provide even more promising discoveries in near future.

References

1. Dobson CM (2004) Chemical space and biology. *Nature* 432:824–828
2. Lipinski C, Hopkins A (2004) Navigating chemical space for biology and medicine. *Nature* 432:855–861
3. Paolini GV, Shapland RHB, van Hoorn WP et al (2006) Global mapping of pharmacological space. *Nat Biotechnol* 24:805–815
4. Reymond J-L, van Deursen R, Blum LC et al (2010) Chemical space as a source for new drugs. *Med Chem Comm* 1:30–38
5. Lipinski CA, Lombardo F, Dominy BW 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
6. Putta S, Beroza P (2007) Shape of things: computer modelling of molecular shape in drug discovery. *Curr Top Med Chem* 7:1514–1524
7. Nicholls A, McGaughey GB, Sheridan RP et al (2010) Molecular Shape and Medicinal Chemistry: A Perspective. *J Med Chem* 53:3862–3886
8. Sauer WHB, Schwarz MK (2003) Molecular Shape Diversity of Combinatorial Libraries: A Prerequisite for Broad Bioactivity. *J Chem Inf Comput Sci* 43:987–1003
9. Lovering F, Bikker J, Humblet C (2009) Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J Med Chem* 52:6752–6756
10. Clemons PA, Bodycombe NE, Carrinski HA et al (2010) Small molecules of different origins have distinct distributions of structural complexity that correlate with protein-binding profiles. *Proc Natl Acad Sci USA* 107:18787–18792
11. Cooper TWJ, Campbell IB, Macdonald, SJF (2010) Factors Determining the Selection of Organic Reactions by Medicinal Chemists and the Use of These Reactions in Arrays (Small Focused Libraries). *Angew Chem Int Ed* 49: 8082–8091
12. Hopkins AL, Bickerton GR (2010) Drug discovery. Know your chemical space. *Nat Chem Biol* 6:482–483
13. Burke MD, Schreiber SL (2004) A planning strategy for diversity-oriented synthesis. *Angew Chem Int Ed* 43:46–58
14. Jacoby E, Mozzarelli A (2009) Chemogenomic strategies to expand the bioactive chemical space. *Curr Med Chem* 16:4374–4381
15. Dandapani S, Marcaurelle LA (2010) Grand Challenge Commentary: Accessing new chemical space for ‘undruggable’ targets. *Nat Chem Biol* 6:861–863
16. Pulici M, Cervi G, Martina K et al (2003) Use of multicomponent, domino, and other one-pot syntheses on solid phase: Powerful tools for the generation of libraries of diverse and complex compounds. *Comb Chem High Throughput Screen* 6:693–727
17. Ulaczyk-Lesanko A, Hall DG (2005) Wanted: New multicomponent reactions for generating libraries of polycyclic natural products. *Curr Opin Chem Biol* 9:266–276
18. Sunderhaus JD, Martin SF (2009) Applications of multicomponent reactions to the synthesis of diverse heterocyclic scaffolds. *Chem Eur J* 15:1300–1308
19. Toure BB, Hall DG (2009) Natural Product Synthesis Using Multicomponent Reaction Strategies. *Chem Rev* 109:4439–4486
20. Biggs-Houck JE, Younai A, Shaw JT (2010) Recent advances in multicomponent reactions for diversity-oriented synthesis. *Curr Opin Chem Biol* 14:371–382
21. Ramon DJ, Yus M (2005) Asymmetric multicomponent reactions (AMCRs): The new frontier. *Angew Chem Int Ed* 44:1602–1634
22. Panek JS, Yang M, Xu F (1992) Diastereoselective additions of chiral (E)-crotylsilanes to in situ generated oxonium ions: a direct asymmetric synthesis of functionalized homoallylic ethers. *J Org Chem* 57:5790–5792
23. Pospisil J, Kumamoto T, Marko IE (2006) Highly diastereoselective silyl-modified sakurai multicomponent reaction. *Angew Chem Int Ed* 45:3357–3360
24. Lipomi DJ, Panek JS (2005) Three-Component, Room Temperature Crotylation Catalyzed by Solid-Supported Bronsted Acid: Enantioselective Synthesis of Homoallylic Carbamates. *Org Lett* 7:4701–4704
25. Petasis NA, Zavialov IA (1998) Highly Stereocontrolled One-Step Synthesis of anti-beta -Amino Alcohols from Organoboronic Acids, Amines, and alpha -Hydroxy Aldehydes. *J Am Chem Soc* 120:11798–11799
26. Candeias NR, Montalbano F, Cal PMSD et al (2010) Boronic Acids and Esters in the Petasis-Borono Mannich Multicomponent Reaction. *Chem Rev* 110:6169–6193
27. Kappe CO (2000) Recent Advances in the Biginelli Dihydropyrimidine Synthesis. New Tricks from an Old Dog. *Acc Chem Res* 33:879–888
28. Chen X-H, Xu X-Y, Liu H et al (2006) Highly Enantioselective Organocatalytic Biginelli Reaction. *J Am Chem Soc* 128:14802–14803

29. Doemling A (2006) Recent Developments in Isocyanide Based Multicomponent Reactions in Applied Chemistry. *Chem Rev* 106:17–89
30. Andreato PR, Liu CC, Schreiber SL (2004) Stereochemical Control of the Passerini Reaction. *Org Lett* 6:4231–4233
31. Wang S-X, Wang M-X, Wang D-X et al (2008) Catalytic enantioselective Passerini three-component reaction. *Angew Chem Int Ed* 47:388–391
32. Lou S, Schaus SE (2008) Asymmetric Pétasis Reactions Catalyzed by Chiral Biphenols. *J Am Chem Soc* 130:6922–6923
33. Gommermann N, Koradin C, Polborn K et al (2003) Enantioselective, copper(I)-catalyzed three-component reaction for the preparation of propargylamines. *Angew Chem Int Ed* 42:5763–5766
34. Carlone A, Cabrera S, Marigo M et al (2007) A new approach for an organocatalytic multi-component domino asymmetric reaction. *Angew Chem Int Ed* 46:1101–1104
35. Grondal C, Jeanty M, Enders D (2010) Organocatalytic cascade reactions as a new tool in total synthesis. *Nat Chem* 2:167–178
36. Enders D, Huettl MRM, Grondal C et al (2006) Control of four stereocenters in a triple cascade organocatalytic reaction. *Nature* 441:861–863
37. Bon RS, Waldmann H (2010) Bioactivity-Guided Navigation of Chemical Space. *Acc Chem Res* 43:1103–1114
38. Koch MA, Schuffenhauer A, Scheck M et al (2005) Charting biologically relevant chemical space: A structural classification of natural products (SCONP). *Proc Natl Acad Sci USA* 102:17272–17277
39. Schuffenhauer A, Ertl P, Roggo S et al (2007) The Scaffold Tree - Visualization of the Scaffold Universe by Hierarchical Scaffold Classification. *J Chem Inf Model* 47:47–58
40. 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
41. Reayi A, Arya P (2005) Natural product-like chemical space: search for chemical dissectors of macromolecular interactions. *Curr Opin Chem Biol* 9:240–247
42. Kumar K, Waldmann H (2009) Synthesis of natural product inspired compound collections. *Angew Chem Int Ed* 48:3224–3242
43. Lessmann T, Leuenberger MG, Menninger S et al (2007) Natural Product-Derived Modulators of Cell Cycle Progression and Viral Entry by Enantioselective Oxa Diels-Alder Reactions on the Solid Phase. *Chem Biol* 14:443–451
44. Antonchick AP, Gerding-Reimers C, Catarinella M et al (2010) Highly enantioselective synthesis and cellular evaluation of spirooxindoles inspired by natural products. *Nat Chem* 2:735–740
45. Tan DS (2005) Diversity-oriented synthesis: exploring the intersections between chemistry and biology. *Nat Chem Biol* 1:74–84
46. Thomas GL, Wyatt EE, Spring DR (2006) Enriching chemical space with diversity-oriented synthesis. *Curr Opin Drug Discov Develop* 9:700–712
47. Peuchmaur M, Wong Y-S (2008) Expanding the chemical space in practice: diversity-oriented synthesis. *Comb Chem High Throughput Screen* 11:587–601
48. Dandapani S, Marcaurelle LA (2010) Current strategies for diversity-oriented synthesis. *Curr Opin Chem Biol* 14:362–370
49. Biggs-Houck JE, Younai A, Shaw JT (2010) Recent advances in multicomponent reactions for diversity-oriented synthesis. *Curr Opin Chem Biol* 14:371–382
50. Kumagai N, Muncipinto G, Schreiber SL (2006) Short synthesis of skeletally and stereochemically diverse small molecules by coupling Pétasis condensation reactions to cyclization reactions. *Angew Chem Int Ed* 45:3635–3638
51. Schaus JV, Jain N, Panek JS (2000) Asymmetric synthesis of homoallylic amines and functionalized pyrrolidines via direct amino-crotylation of in situ generated imines. *Tetrahedron* 56:10263–10274
52. Flamme EM, Roush WR (2002) Enantioselective Synthesis of 1,5-anti- and 1,5-syn-Diols Using a Highly Diastereoselective One-Pot Double Allylboration Reaction Sequence. *J Am Chem Soc* 124:13644–13645
53. Westermann B, Ayaz M, van Berkel SS (2010) Enantiodivergent Organocascade Reactions. *Angew Chem Int Ed* 49:846–849
54. Simmons B, Walji AM, MacMillan DWC (2009) Cycle-Specific Organocascade Catalysis: Application to Olefin Hydroamination, Hydro-oxidation, and Amino-oxidation, and to Natural Product Synthesis. *Angew Chem Int Ed* 48:4349–4353
55. Sunderhaus JD, Martin SF (2009) Applications of multicomponent reactions to the synthesis of diverse heterocyclic scaffolds. *Chem Eur J* 15:1300–1308
56. Di Micco S, Vitale R, Pellicchia M et al (2009) Identification of Lead Compounds as Antagonists of Protein Bcl-xL with a Diversity-

- Oriented Multidisciplinary Approach. *J Med Chem* 52:7856–7867
57. Muncipinto G, Kaya T, Wilson JA et al (2010) Expanding Stereochemical and Skeletal Diversity Using Petasis Reactions and 1,3-Dipolar Cycloadditions. *Org Lett* 12:5230–5233
58. Marcaurelle LA, Comer E, Dandapani S et al (2010) An Aldol-Based Build/Couple/Pair Strategy for the Synthesis of Medium- and Large-Sized Rings: Discovery of Macrocyclic Histone Deacetylase Inhibitors. *J Am Chem Soc* 132:16962–16976

Chapter 3

In Silico Design of Small Molecules

Paul H. Bernardo and Joo Chuan Tong

Abstract

Computational methods now play an integral role in modern drug discovery, and include the design and management of small molecule libraries, initial hit identification through virtual screening, optimization of the affinity and selectivity of hits, and improving the physicochemical properties of the lead compounds. In this chapter, we survey the most important data sources for the discovery of new molecular entities, and discuss the key considerations and guidelines for virtual chemical library design.

Key words: Bioinformatics, Computational biology, Virtual chemical library, Virtual combinatorial library, Computer-aided drug design

1. Introduction

Computational methods are now an important component of modern drug discovery (1). They are actively employed in virtually every step of the drug discovery process, from the design and management of small molecule libraries, through initial hit identification by virtual screening, to optimization of the affinity and selectivity of hits (hit-to-lead optimization); they also improve the physicochemical properties of lead compounds (lead optimization). In silico design of small molecules in particular is a prerequisite for a drug discovery and development campaign, and plays a critical role in the success of the program. Here, we review the most important data sources for the discovery of new molecular entities, and discuss the key considerations and guidelines for virtual chemical library design.

2. Data Sources

Accessibility to quality data is an important first step in the in silico design of small molecules. A number of chemical databases have been compiled (Table 1). These include commercial databases such as the Available Chemicals Directory (ACD), World of Molecular BioActivity (WOMBAT) (2), and CSC, as well as freely available databases including the NCI Open Database, PubChem, LIGAND (3), and ZINC (4). One of the most popular and accessible ligand databases, ZINC, is maintained by Brian Shoichet's lab at the University of California, San Francisco. The ZINC database contains over 13 million purchasable compounds in ready-to-dock, three-dimensional formats, most of which have been corrected for protonation states and have the charges incorporated into the file. Other important properties such as MW, calculated LogP, rotatable bonds, and multiple conformations are annotated into these files. In addition, the database also provides information on the vendors for the compounds, facilitating the purchase of compounds after in silico screens. The Molecular Design Limited (MDL) Available Chemicals Directory (ACD) includes information on over 571,000 research-grade and bulk chemicals, while its screening compound counterpart, Screening Compounds Directory, stores over 4.5 million unique structures. LIGAND (3) includes 15,395 chemical compounds, 8,031 drugs, 10,966 carbohydrates, 5,043 enzymes, 7,826 chemical reactions, and 11,113 reactants. PubChem (<http://pubchem.ncbi.nlm.nih.gov/>), under the umbrella of National Institute of Health (NIH) Molecular Library Roadmap Initiative (<http://nihroadmap.nih.gov/>), stores information on more than 40 million small molecules and 19 million unique structures, and incorporate results from bioassays which can be useful for developing predictive models for drug discovery. DrugBank (5) contains

Table 1
Chemical databases for in silico design of small molecules

| Name | URL |
|----------|---|
| ACD | http://www.mdli.com |
| ChemDB | http://cdb.ics.uci.edu/ |
| DrugBank | http://www.drugbank.ca/ |
| PubChem | http://pubchem.ncbi.nlm.nih.gov/ |
| LIGAND | http://www.genome.jp/ligand/ |
| WOMBAT | http://www.sunsetmolecular.com/ |
| ZINC | http://zinc.docking.org/ |

information on nearly 4,800 drugs, including >1,350 FDA-approved small molecule drugs, 123 FDA-approved biotech drugs, 71 nutraceuticals, and >3,243 experimental drugs. ChemDB (6) stores ~5 million commercially available compounds. Other chemical databases exist and have been reviewed elsewhere (7).

3. Virtual Combinatorial Library

The main advantage of in silico studies is that it allows one to rapidly screen huge libraries of compounds to identify virtual hits, which can then be purchased or synthesized, and verified experimentally. Such libraries often contain large numbers of compounds with highly similar physicochemical characteristics. As such, it is desirable to minimize redundancy or to maximize the number of true leads discovered by optimizing a library's diversity or similarity to a target. Virtual combinatorial library design usually begins with the explicit enumeration of all molecular variants within the appropriate chemical space (8). Two approaches are commonly used to elaborate molecular variants: (1) Markush techniques which attach a list of alternative functional groups to variable sites on a common scaffold (9) and (2) chemical transforms which specify the part of the reacting molecules that undergoes chemical transformation and the nature of these transformations (10, 11). These libraries are then optimized for molecular diversity or similarity using molecular descriptors such as chemical composition, chemical topology, three-dimensional structures and functionality (12). They can also be optimized for drug-likeness using heuristic rules to detect ADME/Tox deficiencies (13).

4. Chemical File Formats

A variety of chemical file formats are currently available and widely adopted. Some formats, such as MOL (and MOL2), Structure Data Format (SDF) and Crystallographic Information File (CIF), are ASCII-based files which encode the chemical structure in three-dimensional space. In general, the three-dimensional structures provided are the products of a geometry optimization and thus the structures are in the lowest energy conformation. Most programs, however, can recognize freely rotatable bonds in these structures and thus produce multiple conformations on the fly. With the exception of the MOL2 format, these structures lack the requisite atomic charges required for accurate docking studies, and thus processing of the ligands is required prior to docking studies.

Other formats, such as Simplified Molecular Input Line Entry Specification (SMILES), International Chemical Identifier (InChI), and SYBYL Line Notation (SLN), are ASCII-based files which provide structures as strings based on atom connectivity. These require an interpreter to convert the information into a three-dimensional structure; however, processing string-based notation is remarkably convenient when performing substructure searches and comparisons, as well as when generating new libraries. SDF libraries may also incorporate string-based notations such as SMILES to aid in searching and sorting.

It is important to note that although ligand libraries are readily available, it is always a good idea to check the structures for any errors. This may include fixing bond lengths and angles, as well as atom types and charges. In some cases, a chemical entity is in its salt form, and thus it may be necessary to remove unwanted cations and/or anions prior to docking studies.

5. Considerations for Virtual Library Design

The discovery of new molecular entities (NMEs) is a highly combinatorial science due to the array of diverse protein targets and the huge chemical search space. The theoretical number of naturally occurring proteins is estimated to be 250,000 (14), while the number of real organic compounds with molecular weights <2,000 Da is greater than 10^{60} (14, 15). As such, several factors must be taken into account when building a virtual chemical library for a successful drug discovery campaign. First and foremost, the chemical structures must be “drug-like.” These properties are not a guideline for actual drug-like activity, rather they are simple rules (drawn from large experimental sets) that determine desirable pharmacokinetic and toxicity parameters, that is, absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox). For the most part, Lipinski’s “rule of five” (16) provides an adequate set of guidelines for selecting drug-like compounds in terms of ADME. An estimated 68.7% of compounds in the Available Chemical Directory (ACD) Screening Database (2.4 million compounds) and 55% of compounds in ACD (240,000 compounds) conform to the “rule of five” (17). This rule suggests that poor absorption and permeation are more likely to occur in molecular entities when: (1) the molecular weight is less than 500 g/mol, (2) the octanol-water partition coefficient ($\text{Clog}P$) is less than 5, (3) the number of hydrogen bond donors is less than 5, and (4) the number of hydrogen bond acceptors is less than 10. There are actually only four rules, and the name is derived from the fact that the numerical values are multiples of 5. Variants of these rules have also been proposed by other researchers (18, 19). For example, a more stringent “rule of five” was proposed by Wenlock

and colleagues (20) after analyzing 594 compounds from the Physicians' Desk Reference 1999, in favor of molecular entities with the following properties: (1) molecular weight <473 g/mol, (2) $\text{Clog}P < 5$, (3) hydrogen bond donors <4, and (4) hydrogen bond acceptors <7. A "rule of three" for lead-likeness has also been proposed by Congreve and coworkers (21), in favor of molecular entities with molecular weight <300 g/mol, hydrogen bond donors ≤ 3 and $\text{Clog}P \leq 3$. "Rule of three" compliant ligand libraries are particularly useful for fragment-based drug design. They allow one to select two or more nonoverlapping ligand hits which can be linked together to create a more potent binder which still falls under the Lipinski rules. Several companies offer both "rule of three" compliant compound libraries, as well as their corresponding virtual libraries (Table 2).

Other physicochemical properties have also been considered in the in silico design of small molecules. One such property is the topological polar surface area (TPSA), which is a measure of the entire molecular surface divided by the number of polar atoms. Drug transport across the blood brain barrier typically has a TPSA cut off at 60 \AA^2 for compounds that are completely absorbed by humans (22). Another such property is ligand charges. Partial atomic charges are often calculated for ligands in order to reflect the distribution of electron density around a particular molecule. The use of empirical and semiempirical calculations to account for ligand charges is often sufficient for molecular docking purposes. Although ab initio methods are also available, they are more time consuming and generally unsuitable for large ligand libraries. Most molecular modeling software includes packages for calculating partial charges. Examples of empirical and semiempirical force fields used to calculate ligand charges include Gasteiger-Marsili charges, AMBER, MMFF, and AM1-BCC (23–26).

Table 2
Commercially available "rule of three" compliant fragment libraries for small molecule design

| Name | URL |
|------------------------------|---|
| Vitas-M Allium Library | http://www.vitasmlab.com/ |
| ChemBridge Fragment Library | http://www.chembridge.com/ |
| Enamine Fragment Library | http://www.enamine.net/ |
| Maybridge Ro3 Library | http://www.drugbank.ca/ |
| OTAVA Fragment Library | http://www.otavachemicals.com/ |
| Prestwick Fragment Library | http://www.prestwickchemical.com/ |
| Pyxis Smart Fragment Library | http://www.pyxis-discovery.com/ |

It is also useful to remove compounds with undesirable physicochemical properties on the basis of functional groups which are known to be reactive, or known to create false positives through indiscriminate binding. Acid halides ($R-C(O)X$), for instance, are very reactive functional groups that quickly hydrolyze to carboxylic acids ($R-CO_2H$), and therefore are of little use in docking studies. Functional groups such as rhodanine (2-thioxo-4-thiazolidinone) are notorious for nonselective binding to proteins. Thus, such compounds should therefore be excluded from a ligand library prior to screening. The substructure filters for Pan Assay Interference Compounds (dubbed PAINs) by Baell and Holloway (27) is one such way for removing ligands which contain undesirable functional groups.

6. Conclusion

The availability of high quality datasets plays an ever important role in the success of computer-aided design. Here, we have reviewed the most important data sources for the discovery of new molecular entities. Combined with modern virtual screening techniques and high-throughput drug discovery programs, these advances will, for the first time, allow the realization of the full potential of lead discovery by design.

References

1. Klebe G (2006) Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today* 11:580–594
2. Olah M, Mracec M, Ostopovici L et al (2004) *WOMBAT: world of molecular bioactivity*. In: Oprea TI (ed) *Chemoinformatics in drug discovery*, Wiley-VCH, New York
3. Goto S, Okuno Y, Hattori M et al (2002) LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res* 30:402–404
4. Irwin JJ, Shoichet BK (2004) ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45:177–182
5. Wishart DS, Knox C, Guo AC et al (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 34:D668–672
6. Chen J, Swamidass SJ, Dou Y et al (2005) ChemDB: a public database of small molecules and related chemoinformatics resources. *Bioinformatics* 21:4133–4139
7. Jónsdóttir SO, Jørgensen FS, Brunak S (2005) Prediction methods and databases within chemoinformatics: emphasis on drugs and drug candidates. *Bioinformatics* 21:2145–2160
8. Agrafiotis DK, Lobanov VS, Salemme FR (2002) Combinatorial informatics in the post-genomics era. *Nat Rev Drug Discov* 1:337–346
9. Leland BA, Christie BD, Nourse JG et al (1997) Managing the combinatorial expansion. *J Chem Inf Comput Sci* 37:62–70
10. Leach AR, Bradshaw J, Green DVS et al (1999) Implementation of a system for reagent selection and library enumeration, profiling and design. *J Chem Inf Comput Sci* 39:1161–1172
11. Lobanov VS, Agrafiotis DK (2002) Scalable methods for the construction and analysis of virtual combinatorial libraries. *Combin Chem High-Throughput Screen* 5:167–178
12. Livingston DJ (2000) The characterization of molecular structures using molecular properties. A survey. *J Chem Inf Comput Sci* 40:195–209
13. Brown RD, Hassan M, Waldman M (2000) Combinatorial library design for diversity, cost

- efficiency, and drug-like character. *J Mol Graph Model* 18:427–437
14. O'Donovan C, Apweiler R, Bairoch A (2001) The human proteomics initiative (HPI). *Trends Biotechnol* 19:178–181
 15. Bohacek RS, McMartin C, Guida WC (1996) The art and practice of structure-based drug design: a molecular modelling perspective. *Med Res Rev* 16:3–50
 16. Lipinski CA (2000) Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol* 44:235–249
 17. Hou T, Wang J, Zhang W et al (2006) Recent advances in computational prediction of drug absorption and permeability in drug discovery. *Curr Med Chem* 13:2653–2667
 18. Ghose AK, Viswanadhan VN, Wendoloski JJ (1999) A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J Comb Chem* 1:55–68
 19. Oprea T (2000) Property distribution of drug-related chemical databases. *J Comput Aided Mol Des* 14:251–264
 20. Wenlock MC, Austin RP, Barton P et al (2003) A comparison of physicochemical property profiles of development and marketed oral drugs. *J Med Chem* 46:1250–1256
 21. Congreve M, Carr R, Murray C et al (2003) A “rule of three” for fragment-based lead discovery? *Drug Discov Today* 8:876–877
 22. Song CM, Tong JC (2009) Recent advances in computer-aided drug design. *Brief Bioinform* 10:579–591
 23. Gasteiger J, Marsili M (1978) A new model for calculating atomic charges in molecules. *Tetrahedron Lett* 19:3181–3184
 24. Cornell WD, Cieplak P et al (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am Chem Soc* 117: 5179–5197
 25. Halgren TA (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comp Chem* 17:490–519
 26. Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* 23:1623–1641
 27. Baell JB, Holloway GA (2010) New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J Med Chem* 53:2719–2740

Surface Plasmon Resonance for Proteomics

Nico J. de Mol

Abstract

Surface plasmon resonance (SPR) is a well-established label-free technique to detect mass changes near an SPR surface. For 20 years the benefits of SPR have been proven in biomolecular interaction analysis, including measurements of affinity and kinetics. The emergence of proteomics and a need for high throughput analysis drives the development of SPR systems capable of analyzing microarrays. The use of SPR imaging (also known as SPR microscopy) makes it possible to use multiplexed arrays to follow binding reactions. As SPR only analyzes the binding process, but not the identity of captured molecules on the SPR surface, technologies have been developed to integrate SPR with mass spectrometric (MS) analysis. Such approaches involve the recovery of analytes from the SPR surface and subsequent MALDI-TOF MS analysis, or LC-MS/MS after tryptic digestion of recovered proteins. An approach compatible with SPR arrays is on-chip MALDI-TOF MS, from arrayed spots on an SPR surface. This review describes some exciting developments in the application of SPR to proteomics, using instruments which are on the market already, or are expected to be available in the years to come.

Key words: Proteomics, Surface plasmon resonance, Mass spectrometry, SPR imaging, Microarray, On-chip MALDI-TOF MS

1. Introduction

Biomolecular interactions are at the core of virtually all processes of life that involve proteins. Proteomics is the study of the whole proteome, or large groups of proteins, which includes their structure and function. The analysis of protein expression is central in many proteomics studies, as will also appear from other chapters in this volume. However, in order to gain insight into protein function and structure on a proteomics level, it is of vital importance to be able to study molecular interactions of large groups of proteins in a high throughput fashion.

Surface plasmon resonance (SPR) has emerged over the last two decades as an excellent tool for biomolecular interaction analysis.

In the initial phase of SPR's history, it was mainly used for studies with one specific protein immobilized on the SPR sensor. This approach has been very successful, as demonstrated by the ever increasing number of papers describing the technique (1). With SPR, not only affinity of protein–protein, protein–ligand, DNA–DNA interactions, etc., can be assayed, but also kinetic information can be obtained by following the SPR signal in real time, which may help to elucidate the binding mechanism. SPR is a label-free technique, so no complications arise from, for example, the introduction of fluorescent labels or secondary antibodies. The only complication that may affect affinity and kinetics arises from the immobilization on the SPR sensor (2). SPR has high sensitivity, high dynamic ranges of molecular weight (from 300 Da), and affinity.

High throughput protein analysis is required for its application to proteomics. This can be achieved by individual SPR sensing of spots on a microarray on a suitable surface, that is, a thin gold or silver layer. Here, a specific method of SPR detection called SPR imaging (SPRi) is used, which is also referred to as SPR microscopy. Another important consideration in proteomic studies is the identity of the binding proteins. Significant effort has been directed to the problem of coupling SPR to mass spectrometry (MS). In this approach, SPR is used for affinity-capture of proteins to ligands immobilized on the SPR sensor surface. These proteins can be analyzed by MS following elution and recovery, or by direct on-chip MS analysis.

There are many applications for microarray SPR, which will generally also hold for other types of microarrays; to name just a few these are as follows: screening for ligands in drug discovery, screening for enzyme substrates, for example, kinases and proteases, screening for biomarkers in diagnostics and other clinical applications and the study of protein networks.

In this review chapter an introduction to the principles of SPR and SPR imaging is given. Secondly, we will focus on the various types of microarrays which are suitable for SPR applications; these are either in development or currently available commercially. Thirdly, an overview is given of the development of SPR/MS combinations.

2. Principles of Surface Plasmon Resonance

The following description is not intended to be a fundamental description of the physics underlying the SPR phenomenon, but rather as a means of explaining the method from a practical point of view. The following section describes some opportunities and limitations that are relevant when applying the technique. For a more basic treatment of the physics of SPR, the reader is referred to more specialized literature (3, 4).

SPR is an optical phenomenon involving the coupling of light energy to the excitation of free electrons in a metal layer. Under proper conditions, the energy carried by photons is transferred to packages of electrons on a metal surface. The electrons come in a highly delocalized state, oscillating as a free electron density against the fixed positive ions in the metal. These oscillations are known as surface plasmons. The surface plasmon resonance phenomenon has already been known from the beginning of the twentieth century, when Wood detected an anomalous diffraction pattern of light and dark bands when visible polarized light was reflected on a metal grating (5). It took until the late 1960s before the SPR phenomenon could be exploited, by the simultaneous work of Otto and Kretschmann, who independently developed configurations for SPR sensors. Many present-day SPR instruments are based on the Kretschmann configuration (see Fig. 1). Such configuration originally exists of a glass prism coated with a thin metal layer, in most cases with a 50 nm gold layer. The metal layer is exposed to a dielectric medium, usually a water phase. In many instruments a separate glass sensor with the metal coating is placed on the prism. In such settings p-polarized light falls on the sensor chip from different angles (α or β , see Fig. 1). The p-component of polarized light, which has its electric component in the plane of incidence, is used to generate the surface plasmons. Under a specific angle of incidence (e.g., β in Fig. 1) the reflection of incident light is strongly attenuated as the p-polarized light energy is transformed

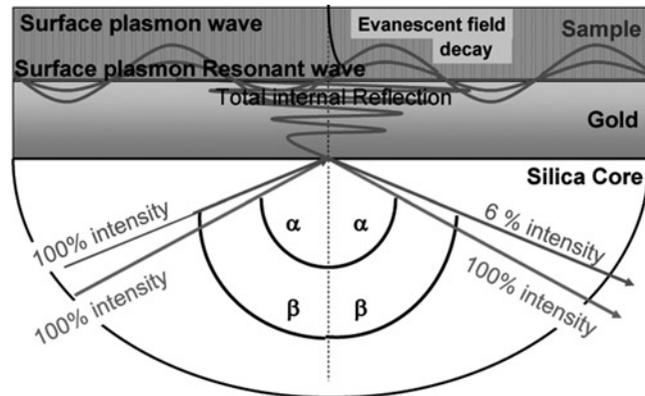


Fig. 1. An SPR sensor in the Kretschmann configuration. A hemi-cylindrical prism is shown, covered with a sensor chip with a gold layer on which a ligand can be immobilized. On top is the sample compartment. The surface is irradiated with p-polarized visible light in a range of incident angles, e.g., α and β . Under conditions of attenuated total reflection (ATR) a dip in the intensity of reflected light is observed (see Fig. 2), because the electrons in the gold layer absorb the energy of the light, resulting in a surface plasmon resonance wave. Note the limited range of the decaying evanescent field in the sample compartment (see Subheading 2). Figure by courtesy of Jan Castrop, Metrohm Autolab, Utrecht, the Netherlands.

into a surface plasmon wave (symbolized as the waves at the gold layer in Fig. 1). This condition of so-called attenuated total reflection (ATR) only occurs when the energy of the photons exactly equals the quantum energy levels of the electrons to be excited to the plasmon resonant wave. Connected to the plasmon resonant wave is an evanescent electromagnetic field pointing in the direction of the dielectric medium immediately near the gold surface, usually a water phase. This evanescent field has a limited penetration into the dielectric medium in which the binding event takes place (see also Fig. 1). The limited range of the exponentially decaying evanescent field is important for the application of the SPR technique (see Subheading 2.2).

2.1. The SPR Signal and SPR Angle

The ATR conditions can be explored by varying the incident angle of the polarized light, with concomitant detection of the intensity of the reflected light. This yields a characteristic SPR curve with a sharp minimum in reflection (approximately 6%) (see Fig. 2). The shape of the SPR curve depends on the nature of the metal, the wavelength of the incident light, and the angle of incidence (4). Most important, the angular position for minimum reflection and the shape of the curve is very sensitive to a complex dielectric constant, immediately near the metal surface, within the range of the exponentially decaying evanescent field, caused by the plasmon resonant wave (see above). This dielectric constant includes the refractive index n (4). Consequently, every change in refractive index is sensed as a change in SPR angle or a change in reflected light. The SPR angle is very sensitive to changes in refractive index, which forms the basis of SPR's use in studying binding phenomena. Absorption of material near the surface changes the refractive index which is monitored as a change in SPR signal as shown in Fig. 2. In classical non-microarray SPR analysis, the change in SPR signal is measured as the change in angle of minimum reflection (SPR angle). Under specific conditions, the change in SPR angle can be related to the absorbed mass near the gold surface. When mass is bound to a 100 nm thick dextran hydrogel layer, as in the most common types of SPR sensor surfaces (e.g., the Biacore CM5 chip), 1 ng/mm² of protein corresponds to an increase in SPR angle of 100 ± 10 millidegree (m°) (see Note 1) (6). This number is dependent on the mean distance of the bound material from the gold surface (see Subheading 2.2). For absorption of proteins, nucleic acids, or low molecular weight molecules, the same number relating absorbed mass to shift in SPR angle can be used. The relation between bound mass and shift in SPR angle is linear up to 50 ng/mm², corresponding to 5,000 m° or 50,000 RU (4). In SPR-microarray applications (SPRi), the result is not necessarily recorded as a shift in SPR angle, but in most equipment it is assayed as a change in intensity of reflected light at a fixed incident angle close to the SPR angle (see Subheading 2.3).

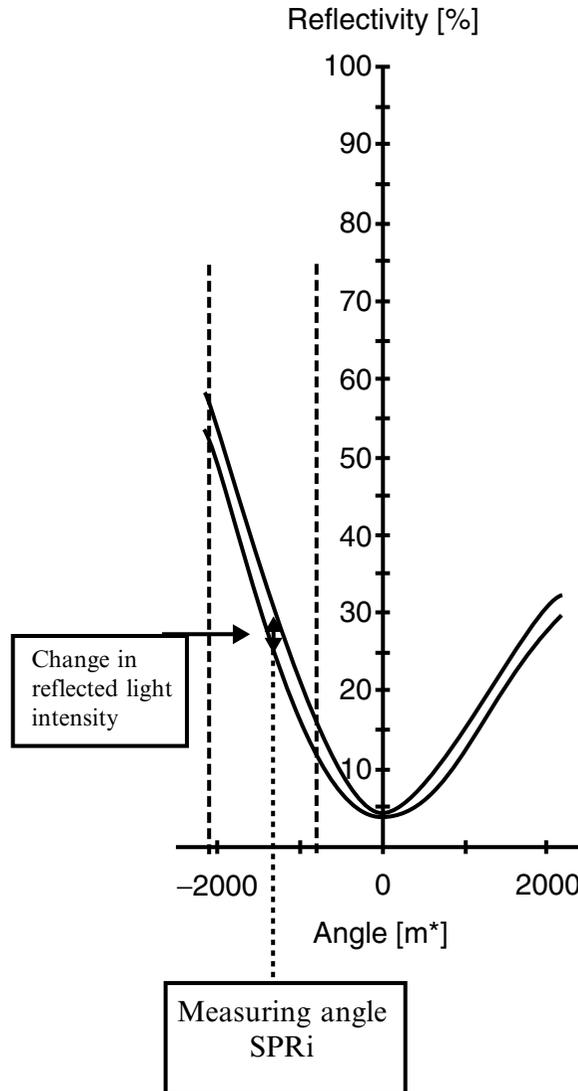


Fig. 2. SPR reflectivity curves: percentage of reflected light as function of the angle of incident light on the sensor chip. Upon binding of mass to the sensor surface the SPR curve shifts to the right. The shift can be measured as a shift in SPR angle at the minimum of reflection. For SPR imaging (SPRi), as applied for the analysis of SPR arrays, frequently the change in reflected light is measured at a fixed angle of incident light in a linear area of the SPR curve, near the SPR angle.

The refractive index is not only sensitive to bound mass, but also to small buffer changes, changes in salt concentration and temperature. The SPR signal is very sensitive to temperature, so to avoid problems due to temperature effects, the equipment must be well thermostatted. So-called bulk effects, due to changes in buffer, the presence of proteins, etc., are observed as sudden jumps in the SPR signal upon injection of the solution onto the sensor surface. Correction for bulk effects can be achieved by simultaneous exposure of solutions to a control surface on which no immobilized ligand is present.

2.2. The Strength of the SPR Signal Decays with Increasing Distance from the Gold Surface

As mentioned above, the effect of changes in the refractive index on the SPR signal can only be “sensed” within the evanescent electromagnetic field induced by the surface plasmon wave. This field penetrates the dielectric medium opposite to the side of the gold layer on which the incident light falls (see Fig. 1). The evanescent field is strongest directly near the gold surface and decays exponentially with increasing distance. At distances larger than 300 nm, the strength of the evanescent field approaches zero. The sensitivity of the SPR signal for bound mass is proportional to the strength of the evanescent wave and is strongest closest to the gold surface. SPR sensor surfaces are designed to meet this criterion of optimum sensitivity, while at the same time allowing sufficient binding capacity, which of course is larger in a thick layer. A carboxymethylated dextran hydrogel with a layer thickness of 100 nm is a good choice in many cases. But when sensitivity is more important, surfaces with shorter brush lengths are also available. For a membrane type of surface, fixation of the lipids near the gold surface will increase the sensitivity. For more detailed information, the reader is referred to the sites of suppliers of SPR sensor surfaces, for example, GE-Biacore (<http://www.biacore.com/lifesciences/products/Consumables/guide/index.html>) and Xantec (<http://www.xantec.com/new/index.php?content=7&sub=6&subsub=>).

The limited sensing range of the SPR signal of 300 nm makes SPR pre-eminently suitable for studying interactions on a molecular scale. Detection of the binding of larger objects (e.g., cells) is possible, but this hampers the quantitative interpretation of bound mass. Also, the application of layers on the gold surface larger than 300 nm (e.g., some molecular imprinted polymer films) will not be successful unless thin polymer films are used, for example, <100 nm (7).

2.3. SPR Imaging

High throughput conditions are vital for proteomics. Large number of samples can be analyzed in short time using microarrays, so new methods for analyzing microarray sensors have been developed for high throughput applications of SPR (see Subheading 3). Microarrays can be analyzed using a specific application of SPR technology, SPR imaging (SPRi), also known as SPR microscopy (8). In SPRi only the signal at a fixed angle in a linear part of the SPR curve is measured, not the complete SPR curve (see Figs. 2 and 3). Within the linear part of the SPR curve, the change in reflected light intensity is proportional to the change in refractive index due to binding of molecules to the surface. A detailed analysis of the relationship between the changes in reflected intensity in SPRi and the absolute amount of absorbed mass at the surface is given by Shumaker-Parry et al. (9). To indicate the sensitivity of SPRi, in a specific instrument setting using $200 \times 200 \mu\text{m}$ spots on the array, a detection limit of $1.2 \text{ ng}/\text{cm}^2$ corresponding to 0.5 pg per spot is observed. The dynamic range in protein concentration is $\sim 720 \text{ ng}/\text{cm}^2$ (9).

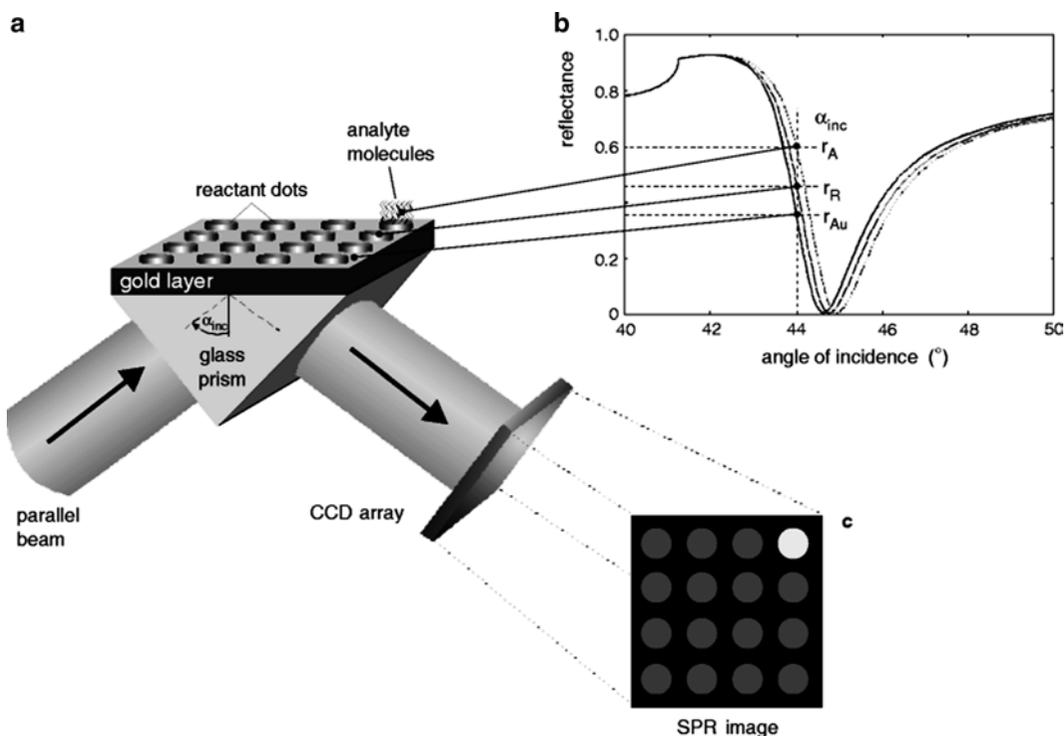


Fig. 3. Schematic diagram of SPR imaging (SPRi). (a) SPR array in the Kretschmann configuration with light source and CCD camera. (b) SPR reflectance curve for a pure gold surface (*solid line*), a reactant dot without analyte (*dashed line*) and adsorbed analyte molecules on a reactant dot (*dotted line*). (c) The contrast of the SPR image is based on the different reflectance $r_A > r_R > r_{Au}$. Reprinted from ref. 39 with permission.

Measurements are generally performed over the whole surface of the array (10) and changes in reflected intensity generally detected using a CCD (charged coupled device) camera (see Fig. 3). SPRi can also be used to assay kinetics by following the reflection pattern over time.

SPRi has a very large spatial resolution, down to $4 \mu\text{m}$ over a large sensing surface (10) which allows the positioning of a large number of spots onto a relatively small surface.

The sensitivity of SPRi is linearly dependent on light intensity (8), therefore lasers are preferred as a light source over LEDs and halogen lamps; often a He–Ne laser is chosen. However, using optical means of expanding a laser spot to the complete microarray will result in inhomogeneous illumination of the array surface. This will result in signal variation across the sensing area and in such cases, background correction is required, thus diminishing the sensitivity (8). Solutions to this problem are incorporated in various instrumental designs (see Subheading 3.2). Alternatively, collimated white light is used in combination with a narrow band pass filter, typically centered in the near-infrared region to select the excitation wavelength (11). In more sophisticated designs, the surface is scanned by the light source.

3. SPR Compatible Microarrays

Proteomics studies of many potentially interacting partners in living systems are now unthinkable without microarrays. Microarrays of large sets of ligands that are captured on a solid support in separate spots are available for research (see Fig. 3). The array can be probed with complex samples (e.g., serum, urine) containing components that recognize the captured ligands, and bind to it. Microarrays have the advantage of needing only small amounts of sample and can screen multiple interactions in a parallel manner.

Microarrays are available in various formats depending on the molecules immobilized, for example, ssDNA, ssRNA, antibody, protein, phage displayed, and peptides. Nucleic acid microarrays are applied to genomics, gene expression, and single nucleotide polymorphism genotyping (SNP). Antibody arrays are applied to the detection and profiling of multiple biomarkers as diagnostic tools in diseases, for example, cancers and cystic fibrosis and to evaluate therapeutic treatment. An alternative to antibodies are aptamers, short single-strand DNA or RNA sequences which recognize non-nucleic molecules, such as proteins, with high affinity and selectivity (12). Protein and peptide arrays are used in screening for drug development, and also to monitor enzyme activity for diagnosis in cancers, for example.

Binding to arrays can be detected in several ways; detection strategies can be classified as (13):

1. Labeled probe methods, including fluorescence, chemoluminescence, and radioactivity.
2. Label-free methods. These include, among others, atomic force microscopy (AFM), quartz-crystal microbalance analysis (QCM), and SPR.

For SPR microarrays the presence of a thin metal layer, usually gold, is essential, as follows from Subheading 2.

There are some important considerations concerning problems that may occur with microarrays in general, and also with SPR microarrays:

1. Integrity and stability of the immobilized ligand. Labile proteins in particular may be a problem. In general, peptide ligands perform better on arrays.
2. Capture of ligands should be done in an unequivocal way. A frequently used approach is immobilization through covalent binding to the surface of free amino groups which may occur in multiple sites in a protein or peptide (2). Such a strategy might lead to heterogeneity in the way ligands are coupled to the surface and might obstruct access of analyte molecules to the binding site. Alternatively, a better-defined immobilization

might occur through biotinylated or his-tagged ligands, for example. Such approaches can be successfully combined with immobilization through amine groups after initial binding to the surface through the tag (14).

3. Accessibility of the surface ligands for large protein analytes. The accessibility may be hampered due to lack of space on a closely packed surface, leading to an underestimate of the amount of bound material (2).
4. A large range of concentrations may occur in biological samples. This puts a constraint on the sensitivity and the dynamic range of the detection method.
5. Nonspecific binding of analytes in the investigated sample should be minimal. Some corrections can be made by using control spots on the surface, but such corrections should be small compared to the detected signal due to specific binding.

All these considerations are also valid for SPR microarrays. In addition to this, it is important to completely cover the gold surface, as many biomolecules will nonspecifically adsorb to the gold layer which might also hamper their biological activity (11). The general method for attaching an immobilization matrix to the gold surface involves the use of thiolated molecules as these can be irreversibly attached through a gold-thiolate bond (15). Coverage of the gold surface can be achieved using a hydrophobic layer (self-assembled monolayer, SAM) modified to immobilize the ligand, or by attaching a hydrogel (e.g., dextran-type) matrix functionalized for ligand immobilization. Another important consideration for SPR applications is the thickness of the layer, as the SPR signal decays exponentially with the distance from the gold surface (see Subheading 2.2).

Immobilization strategies for ligands, including surface chemistry, can be very diverse. Here, surface chemistries are not dealt with in detail, but general methods are reviewed (16, 17).

Ligands have to be applied on discrete spots in the μm range on the surface of microarrays. There are several methods for forming patterns on gold surfaces: UV photopatterning, microcontact printing, and robotic spotting, all of which have been used for SPRi arrays (11).

The versatility of SPR microarrays can be increased by the ability to introduce samples to specific spots of interest. This can be achieved by sophisticated microfluidics systems (Lab-on-a chip), but such approaches are still very much in development (18).

3.1. Enhancement of SPRi Sensitivity

Under “standard” conditions the sensitivity of SPRi is in the nM range (12). However, increased sensitivity is needed when studying low abundant molecules with a relatively weak affinity.

In this section, a number of methods are described for enhancing the sensitivity of microarray reading using SPRi. The Robert Corn group is the pioneer in this field and has made important contributions to it. Here, two ways of enhancing SPRi sensitivity are reviewed: (1) surface enzyme reactions and (2) functionalized nanomaterials (12).

3.1.1. Sensitivity Enhancement Using Enzyme Reactions

The use of enzymatic reactions to enhance the SPRi signal is illustrated by the amplified detection of DNA using RNase H, which specifically hydrolyzes RNA–DNA heteroduplexes on an RNA array surface (19). The principle is illustrated in Fig. 4. To an ssRNA array a mixture of ssDNA sequences and RNase H is added. The DNA sequences that are complementary to the ssRNA on the array form an RNA–DNA heteroduplex. The RNase H specifically destroys the RNA in such duplexes, and the ssDNA is released. This DNA is available for binding to the next ssRNA sequence and the RNase H reaction is repeated. In this cyclic process a small number of binding ssDNA sequences can destroy a large number of complementary RNAs, amplifying the effect of a small number of ssDNA molecules. Hydrolysis of multiple specific ssRNA sequences immobilized on a spot on the array is observed as a change in reflected light (see Subheading 2.3). Such an approach has been applied in the detection of DNA from the TSPY gene, with sensitivity in the fM range. It is estimated that in this specific experiment, 12,000 probe RNAs are hydrolyzed for every ssDNA molecule (19).

Other examples of enzymatic amplification have been published by the Corn group and others. In the rolling circle amplification (RCA) approach, antibodies provided with a DNA primer sequence

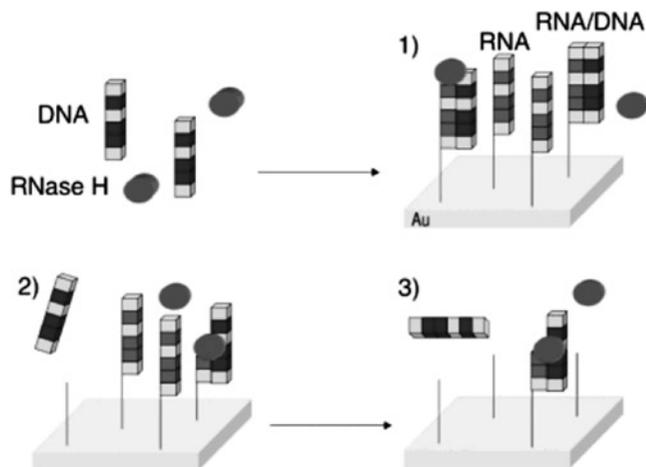


Fig. 4. Schematic representation of a surface enzymatic reaction to enhance the SPR signal by selective RNA probe removal. For details see text. Reprinted from ref. 19 with permission.

are used to detect specific proteins trapped on an array. A circular DNA with a portion complementary to the primer sequence is then hybridized to the surface. The enzyme DNA polymerase largely extends the primer by repeatedly rolling along the circular DNA-template (20). In another SPRI-based approach, a spatially resolved precipitate on the array spot (mediated by a horseradish peroxidase (HRP) reaction) is used to enhance the sensitivity (21).

3.1.2. SPRI Sensitivity Enhancement with Functionalized Nanoparticles

The amplification of the SPR signal can be very large due to absorptive coupling between the gold nanoparticles and the thin layer of gold on the array (3). However in the examples presented here, the enhancing effect is only due to changes in the refractive index as the wavelength of maximum absorption of the gold-nanoparticles (525 nm) is quite different from the wavelength in the SPRI measurement (830 nm). Under such conditions, the use of functionalized nanoparticles can lead to a greater than tenfold increase in SPR angle shift, corresponding to a more than 1,000-fold increase in sensitivity compared to nonamplified binding (22). An example is given here for the detection of single nucleotide polymorphisms (SNPs) in genomic DNA. The approach, combining enzymatic ligation and amplification with gold-nanoparticles, is illustrated in Fig. 5. Two DNA array elements are shown, one with a 3'-terminal

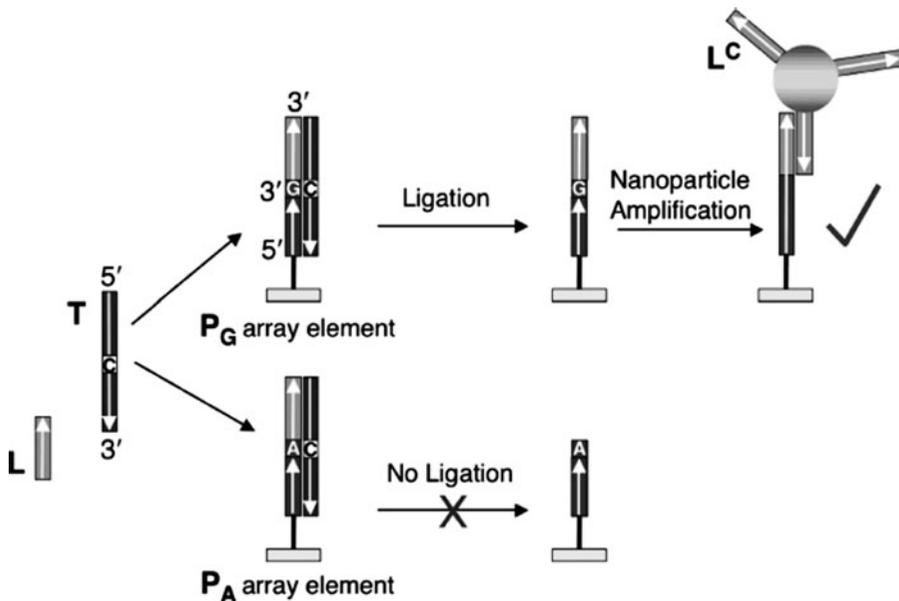


Fig. 5. Schematic representation of the SNP genotyping method based on a combination of surface ligation chemistry and nanoparticles enhanced SPRI. The two array elements (P_A and P_G) differ only by their last nucleotide at the 3' end. Target DNA (T), ligation probe (L), and *Taq* DNA ligase are simultaneously introduced. Ligation only occurs with perfect complementary duplexes on the array, leading to extension with the L sequence. The presence of L is detected by gold nanoparticles carrying oligonucleotides (L^C) complementary to L. For further details see text. Reprinted from ref. (23) with permission.

A and one with a 3' terminal G (P_A and P_G , respectively). Target DNA T, ligation DNA probe L, and *Taq* DNA ligase are introduced simultaneously. Duplexes are formed on both array elements; however, ligation of L to the array DNA is only possible if a perfect complementary duplex is formed. Upon denaturation with 8 M urea the perfectly matching P_G is extended with L, while P_A remains in the unligated state. The signal of the L sequence on the array is amplified by binding to 12–13 nm gold-nanoparticles provided with DNA sequences that are complementary to the L probe (23). With this nanoparticle-enhanced surface ligation approach, single base mismatches in ssDNA are detected down to 1 pM (23).

Other applications of nanoparticles for SPRi enhancement are the original method for DNA detection by He et al. (22), and the detection of microRNAs (miRNAs) down to a concentration of 10 fM using a surface poly(A) polymerase detection and nanoparticle amplification (24). A recent application is the detection of polymerase products in the attomolar range by template-directed polymerase extension of a surface array element with nanoparticle-enhanced detection of the reaction product (25).

3.2. Instruments for SPR Arrays

Most instruments for SPR arrays use SPR imaging at a fixed angle for analysis of the arrays (see Subheading 2.3). A number of SPR imaging instruments are now commercially available, while other potentially interesting instruments are still in development and are expected to reach the market in a few years. In this section, we will give an overview showing some differences in approach and technology, and some promising developments in this area will be covered in more detail. Differences in SPR array instrumentation exist in the optics (e.g., fixed angle or scanning angle), and the liquid handling system. Many SPRi instruments can be used in combination with dedicated equipment for spotting SPR arrays. Practically, in all SPRi systems the changes in reflectivity of the array surface are detected with a CCD camera as a microscopic view; sometimes the complete SPR curve is monitored. The most striking characteristics of each instrument are now described.

3.2.1. Biacore 4000

Biacore (part of GE Healthcare <http://www.biacore.com/lifesciences/index.html>) is the market leader in SPR analytical equipment since its introduction in 1990. Their first commercially available array-based SPR instruments were the Biacore A100 and the Flexchip, both introduced in 2005. The Biacore 4000, is now the instrument for high throughput use available from GE-Biacore.

The Biacore 4000 has 4 flow cells, each having its own needle, thus permitting parallel injections. Hydrodynamic addressing allows multiple interactants to be immobilized on different detection spots within a single flow cell. In each flow cell multiple interactants can be immobilized on 5 detection spots. This allows parallel analysis of 4 samples with 5 components, or analysis of one

sample interacting with a maximum of 20 components. With the Biacore 4000, a maximum automated throughput of up to 4,800 interactions in 24 h is possible. The Biacore 4000 is not an SPR instrument as it uses a range of angles to detect the SPR curve. Typical applications are antibody screening, screening of low molecular weight compounds and fragments for drug development and kinetic analysis, for which it is well suited.

The Flexchip instrument does not have the more common Kretschmann configuration (see Fig. 2), but is an example of a grating coupler SPR instrument. The polarized light used to create the plasmon resonance now comes from the other side and travels through the sample layer of the sensor. This has advantages and disadvantages: it is easier to apply cheaper disposable gratings; on the other hand, the flow cell has an increased height to avoid unwanted internal reflection effects (26). With the Flexchip it is possible to apply up to 400 spots on the array at one time. There is one single broad flow cell serving all these spots so only one analyte at a time can be studied. Several types of affinity chips for the Flexchip instrument are available from GE-Biacore: plain gold, streptavidin, neutravidin, and protein-A/G. The instrument is now no longer in the product range of GE-Biacore.

3.2.2. SPRImagerII, GWC Technologies

The technology (<http://www.gwctechnologies.com/gwcSPRImager.htm>) has its origin in the group of Robert Corn. A spotting robot can be used to make the array and the entire array is exposed to the same analyte in a 1-cm flow cell, allowing for simultaneous capture of data for all probes used. The change in intensity of reflected light at a fixed (adjustable) angle is recorded in the near infrared with a high sensitivity CCD camera (see Fig. 3). S-polarized light is used for calibrating the intensity of the light source. This allows conversion to absolute reflectivity values for direct comparison between experiments. The SPRImagerII is very versatile, accepting arrays containing proteins, antibodies, peptides, oligonucleotides, lectins, and aptamers, etc.

3.2.3. Plasmon Imager (Graffinity Technology)

The Graffinity technology is dedicated to drug discovery using fragment-based drug design and lead optimization (http://www.graffinity.com/t_overview.php). Libraries of low molecular weight compounds (fragments and lead-like compounds) are provided with a propriety spacer molecule which serves as an attachment point for the covalent immobilization on the array surface. These are spotted on the array to create a chemical microarray with about 10,000 compounds displayed on the top of a self-assembled monolayer (SAM) attached on a thin gold layer. The SAM serves two purposes: (1) it prevents nonspecific protein binding to the array surface, and (2) it presents anchor molecules for covalent binding of the library compounds to the array surface. In a typical experiment proteins are passed over the array.

In Graffinity's process, binding is detected as a wavelength shift that corresponds to the increase of mass concentration on the chip surface. A specific pattern or "fingerprint" emerges for each target protein, reflecting each protein's chemical-binding characteristics. Graffinity performs the screening as a service for their customers, who provide the proteins of interest.

3.2.4. *PlexArray System™* (Plexera Bioscience)

Plexera Bioscience was originally a subsidiary of Lumera, but formed a new "start up" in 2009 (<http://www.plexera.com/index.php>). The system was originally developed under the name "Proteomic Processor" and can handle large arrays of 1,000–10,000 spots. It utilizes the Kretschmann configuration for SPR-based detection and real-time monitoring of analyte binding. The instrument contains a MEMS scanning mirror from Microvision which is connected to small flexures that allow it to oscillate. The light of a diode laser is precisely directed to the spots on the array, with a scanning frequency of 60 Hz. This approach produces a homogeneous illumination over a larger area without generating interference patterns. The resolution has been measured as approximately 10 μm , which permits the resolution of 10,000 spots on a 1.4 cm^2 array surface (8).

The system is marketed by Plexera Bioscience under the name *PlexArray™* analyzer, together with specific sensor chips. *PlexArray™* Sensor Chips consist of 1" \times 3" gold-layered glass slides that are compatible with most commercially available protein microarray printers. Following printing of user-defined content, a specially designed Sensor Chip Cover is attached to create a flow chamber through which are passed sample analyte and processing reagents. Kinetic analysis of several thousand biomolecular interactions can be performed in a single experiment using K_x Array™ Technology.

3.2.5. *SPRi-Plex™ Array* *System* (GenOptics)

The SPRi array system from Genoptics (<http://www.genoptics-spr.com/>) consists of the SPRi-Plex array system, the SPRi Biochip, and the SPRi-Arrayer for printing arrays. Depending on the diameter of the spots, up to 1,000 spots (100 μm), 300 spots (200 μm), or 64 spots (500 μm) can be deposited on the 1 cm^2 sensitive surface of the biochip. Various types of surface chemistry are available: (1) Using electrochemical copolymerization, either with pyrrole or diazonium. (2) Using self-assembled mono- and multilayers, with either biotinylated biomolecules, cystamine-glutaraldehyde chemistry, or immobilization on a NHS functionalized surface to combine SPRi technology with MALDI mass spectrometry (see Subheading 4.2.1).

The SPRi-Flex uses the Kretschmann configuration with a rotating mirror for scanning the SPR angle (at the SPR reflectivity dip). A broad monochromatic polarized light bundle illuminates the whole functional array surface of the SPRi Biochip (26).

3.2.6. *ProteOn XPR36* (Bio-Rad)

The ProteOn XPR36 protein interaction array combines SPRi with a high throughput fluidics system. 6×6 spots can be applied to the array, immobilizing up to 6 different ligands in lines. Six analytes can be injected perpendicular to the lines of ligands using the fluidics system, thus allowing the detection of 36 different interactions in real time. The system measures 78 spots: 36 interaction spots and 42 reference spots at two sides of an interaction spot for signal normalization. These spots are also exposed to analyte flow and the reference signal of these spots may be subtracted from the interaction signal. The optical system generates a complete SPR curve for each interaction. ProteOn sensor chips are prepared with a modified alginate layer bound to the gold surface of the sensor prism. There are three types of sensor chips available with different binding capacities; these use activated carbonyl groups for covalent binding to primary amines. Alternatively, a chip with neutravidin is available for immobilization of biotinylated ligands.

3.2.7. *IBIS-MX96* (IBIS Technologies)

IBIS technologies (<http://www.ibis-spr.nl/>) has developed the IBIS *i*SPR instrument, which was superseded by the IBIS-MX96 in 2010. The technology consists of four elements: (1) optics, (2) fluidics, (3) sensors, and (4) a spotter. The optics component is in the Kretschmann configuration, with a scanning mirror that makes it possible to monitor the complete SPR curve of each point of interest. Scanning of the complete SPR curve is considered to yield a more sensitive signal than changes in reflection at a fixed angle (see Fig. 2), and it overcomes spot to spot sensitivity problems and linearity issues (27). Images of the total array area of 8.8×6.6 mm are monitored and in principle over 500 spots can be placed in this area. The MX96 typically has 2×48 spots. The change in SPR signal is followed in real time, thus allowing kinetic analysis of the binding. The fluidics system comprises a flow cell and two syringe pumps for generating back and forth flow during the interaction process. Long interaction times can be studied without the consumption of large amounts of often precious analyte, as occurs in a flow system. During the dissociation phase, the buffer solution is pumped slowly over the sensor area by the second syringe pump for maintaining zero analyte concentration. Captured compounds can be eluted for further analysis (e.g., by MS see Subheading 4.1). Sensor surfaces can either be pure gold, a gel-type containing preactivated esters for covalent linking of ligands, or contain streptavidin, protein A, or anti-IgG for specific immobilization.

The continuous flow microspotter (CFM) is a separate device for printing ligands onto the sensor surface in a 2×48 spot format. Ligand coupling conditions (e.g., pH) for each individual spot can be optimized for creating the highest quality spotting.

4. SPR and Mass Spectrometry for Proteomics Applications

The combination of SPR imaging with mass spectrometry (MS) is a very powerful technique for pairing SPR interaction analysis or quantification with MS for structural identification. MS, in the form of MALDI-TOF or ESI-MS, is a pivotal technique in proteomics research. The ultimate goal is online MS analysis of analytes captured on the SPR surface. Much of the early work on SPR-MS has been performed in the laboratory of Nelson and Nedelkov at Intrinsic Bioprobes, Inc. (28).

Essentially, two different methods have been developed: (1) Elution of captured analyte from the SPR surface after the binding event. (2) On-chip MALDI-TOF-MS analysis. Key examples of these approaches are presented here. It should be noted that in general, weak binding analytes with high dissociation rates do not yield sufficient material for subsequent MS analysis.

4.1. SPR/MS Analysis Using Recovered Analytes

Recovery of captured analyte from the SPR sensor chip is an obvious way to collect samples for subsequent MALDI-TOF or ESI-MS analysis. However, the yield must be sufficient for MS analysis, but there are a number of problems associated with recovery, such as loss of sample due to diffusion in the fluidic system. Nonspecific binding and carry-over of material from the injected sample to the recovered sample gives an unwanted MS response. Furthermore, recovery is not compatible with SPR arrays and high throughput analysis, unless the spots can be addressed individually by the fluidics system; the spots must also be large enough to recover sufficient amounts of material for MS analysis. On the other hand, recovery allows postprocessing of eluted analyte, such as desalting or proteolytic digestion for peptide mass fingerprinting (29).

The yield of microrecovery has been addressed in the design of the Biacore 3000 (GE Healthcare) by the introduction of air segments and reversal of the buffer flow to return the sample back through the flow cell for collection via the autosampler; this allows the recovery of a very small volume of sample (2 μ L/cycle) (30).

The recovery of hydrophobic proteins can be increased by the use of silica or C18-coated magnetic beads passed over the chip during the desorption process from the chip (31). Tests have been performed with bovine serum albumin (BSA), a protein known for nonspecific binding to many materials, including those commonly used for fluidics systems. The silica particles increased the recovery of BSA from the chip twofold. This might be a useful approach when working with hydrophobic proteins showing high nonspecific binding. Although it is often a tedious procedure, microrecovery does allow tryptic digestion for protein identification. General protocols have been described by Borch and Roepstorff for binding, washing of the fluidics system,

elution from the sensorchip, followed by tryptic digestion, and MS analysis including database searching of the peptide masses (29).

An application combining SPR and subsequent LC-MS/MS analysis after sequential elution steps improved the detection of specific binding of low-abundant proteins in complex samples (32). In this example, an IBIS iSPR instrument (see Subheading 3.2.7), equipped with a sensor surface immobilized with cGMP was used. A cell lysate was injected over the surface and fractions were recovered in subsequent elution steps using, respectively, ADP, GDP, cGMP, cAMP, and sodium hydroxide; eluted material was digested with trypsin and analyzed with LC-MS/MS. In the first elution steps with ADP and GDP, highly abundant proteins showing non-specific binding were eluted, but with cGMP elution, low abundant proteins showing specific cGMP-binding were recovered (32).

4.2. On-Chip MALDI-TOF-MS

The polysaccharide hydrogels commonly used for SPR sensor chips appear to be compatible with MALDI-TOF analysis. On-chip SPR-MS offers the intriguing possibility of studying not only individual proteins, but also in vivo assembled functional protein complexes (33). The approach of Nelson and Nedelkov has evolved on-chip MS analysis by placing the SPR surface in an appropriate MALDI target after capture of analytes. A specially developed chip cutter is used for this; this contains a circular heated cutter-head that excises the chip segment from the frequently used Biacore CM5 sensor chip in a form that fits into the MALDI mass spectrometer target (34). The MALDI matrix applicator sprays a fine mist of matrix solution over the entire surface of the cut-out chip. The matrix of choice for protein analysis is α -cyano-4-hydroxycinnamic acid (ACCA) (28). It is important that the matrix mist only moistens the surface, rather than making it completely wet. The matrix should be applied as small individual droplets, which will desorb the analytes from the capturing ligands. After rapid drying, the matrix/protein mixture will be deposited on the same area where the analytes were originally captured in the SPR analysis. More practical experimental details are given by Nedelkov (28). Upon careful matrix application, the spatial resolution is preserved between the immobilization sites in the four flow cells on the sensor surface of a Biacore 3000. This makes it possible to analyze several interactions simultaneously (35).

Grote et al. have constructed a bifunctional SPR fluid cell where optimized surfaces can be used simultaneously for binding studies and kinetics (low binding capacity) and MS (high binding capacity), with regard to the specific need of each technique (36). An IBIS instrument (IBIS Technologies BV, Hengelo, Netherlands) was used which was equipped with a cuvette cell, with the SPR sensor being at the bottom. For this specific purpose, screw-type pins, covered with a high capacity binding thick hydrogel, were mounted in the wall of the cuvette cell. The immobilization conditions are

identical for the ligand on the SPR sensor and the pin surface, both of which are located close together in the cell. The pins can be removed from the cuvette cell after the binding experiment, leaving the SPR disk in place; the pins can then be screwed into an adapted MALDI target plate. This method provides a usable MS analysis (36).

4.2.1. On-Chip MALDI-MS Detection Using an SPRi Microarray Format

The newest development in on-chip MS analysis is the analysis of SPR arrays. This will meet the strong demand for high throughput and structural identification in proteomics research. The big challenge is to retain the captured analytes on the spots during desorption by the matrix solution, which will be a problem with a high-content array of small spots and small interspot distances. There is no equipment for such an approach on the market as yet, but two promising recent developments in this area are highlighted; these are from groups cooperating with GenOptics (see Subheading 3.2.5) and from Intrinsic Bioprobes (37, 38).

Bellon et al. (37) use SPR surfaces with self-assembled monolayers consisting of short polyoxyethylene chains (POE, eight oxyethylene units) on the gold surface. These SAMs are more polar than the more frequently used alkyl chain SAMs, which are hydrophobic and readily give rise to nonspecific binding. Antigen-antibody interactions were studied as a model, with antibodies immobilized on the sensor surface in an array format of 4×3 1.5 mm diameter spots separated by a 1.5 mm interspot distance. SPR imaging (SPRi) experiments were performed using an SPRi-Plex imager (see Subheading 3.2.5). After the SPRi experiment, the chip was removed and air-dried. The captured antigens on the chip were digested by dropping trypsin onto some of the spots. The digestion was terminated by applying the MALDI matrix to the spots. The matrix consisted of HABA (2-(4-hydroxy-phenylazo) benzoic acid) or ACCA (α -cyano-4-hydroxycinnamic acid). Once air-dried, the chip was inserted into an Opti-TOF MALDI plate. HABA provided the best results for the detection of whole captured proteins, while ACCA was used for MS and MS/MS experiments on peptides. The results demonstrate that the POE surface allows for efficient and sensitive MS detection of whole captured protein antigens present at a low surface density of about 2 fmol/mm². Furthermore, a low level of nonspecific binding is observed on the nonarrayed area of the chip. The mass spectra obtained after tryptic digestion were characteristic of the peptides derived from the specifically retained proteins (37).

At Intrinsic Bioprobes (<http://www.intrinsicbio.com/>) a combination of SPR and MS detection on a high-content protein microarray has been explored (38). The gold surface of SPR chips was incubated with mercaptoundecanoic acid to form a SAM. The surface carboxyl groups were then activated with 1,1'-carbonyldiimidazole and the chips arrayed in a microarrayer, producing spots 200- μ m in diameter in a 10×10 array, with 450- μ m spacing

between each spot. SPR imaging was performed using the SPR Imager II instrument (GWC Technologies, see Subheading 3.2.2). After binding, the chip was taken out of the SPR Imager and the MALDI matrix (ACCA) applied to the surface (see Subheading 4.2). Focusing the MS laser to the spots in subsequent MALDI-TOF MS analysis, 35 laser shots were acquired for each individual spectrum from the same position within a spot on the array. The width of the laser was $\sim 100\ \mu\text{m}$ so each laser shot fits within the confines of the $200\text{-}\mu\text{m}$ diameter spot. The MALDI matrix spraying is very critical for keeping the desorbed analyte on its spot; the method developed at Intrinsic Bioprobes for this appears to be suitable for maintaining the resolution between the small spots on the array (38).

5. Outlook

The approaches and techniques described in this review have to an extent already become commercially available. This is a highly competitive field, and it is expected that new equipment and technologies will be developed over the coming years.

For high throughput and multiplex assays (analysis of multiple ligands and multiple analytes) much is expected from techniques for examining individual addressable spots on microarrays, where each spot on the array can be treated individually and has a unique functionality. Developments of Lab-on-Chip techniques, such as electrokinetic transport on arrays and microchannel fabrication to address individual spots, all point in such a direction (18). Impressive results have already been obtained by increasing sensitivity of detection on spots using enzymatic methods and functionalized nanoparticles (see Subheadings 3.1.1 and 3.1.2). More such developments for dedicated applications are expected. Further development of interfaces for on-chip MALDI-TOF MS analysis, combined with digestion of analyte protein on the spot for fingerprint identification will provide exciting opportunities for high throughput analysis combined with analyte identification (see Subheading 4.2.1).

6. Note

For some instruments, for example, from IBIS Technologies (see Subheading 3.2.7), the shift in SPR angle is expressed in millidegree (m°), for others, for example, from Biacore, the shift is expressed in response units (RU). 10 RU units correspond to $1\ \text{m}^\circ$. For a $100\ \text{nm}$ hydrogel SPR sensor, for example, the frequently used Biacore CM5 chip, $100\ \text{m}^\circ$ or 1,000 RU corresponds to a bound mass of $1\ \text{ng}/\text{mm}^2$.

References

1. Rich RL, Myszka DG (2008) Survey of the year 2007 commercial optical biosensor literature. *Journal of Molecular Recognition* 21: 355–400
2. Schuck P, Zhao H (2010) The role of mass transport limitation and surface heterogeneity in the biophysical characterization of macromolecular binding processes by SPR biosensing. *Methods Mol Biol* 627: 15–54
3. Kooyman RBH (2008) Physics of Surface Plasmon Resonance. In: Schasfoort RBM, Tudos AJ (eds) *Handbook of Surface Plasmon Resonance*, Royal Society of Chemistry, Cambridge, UK, pp 15–34
4. Salamon Z, MacLeod HA, Tollin G (1997) Surface plasmon resonance spectroscopy as a tool for investigating the biochemical and biophysical properties of membrane protein systems. I: Theoretical principles. *Biochimica et Biophysica Acta - Reviews on Biomembranes* 1331: 117–129
5. Wood RW (1902) On a remarkable case of uneven distribution of light in a diffraction grating spectrum. *Phil Mag* 4: 396–402
6. Stenberg E, Persson B, Roos H et al (1991) Quantitative determination of surface concentration of protein with surface plasmon resonance using radiolabeled proteins. *J Colloid Interface Sci* 143: 513–526
7. Li P, Huang Y, Hu J et al (2002) Surface plasmon resonance studies on molecular imprinting. *Sensors* 2: 35–40
8. Boozer C, Kim G, Cong S et al (2006) Looking towards label-free biomolecular interaction analysis in a high-throughput format: a review of new surface plasmon resonance technologies. *Curr Opin Biotechnol* 17: 400–405
9. Shumaker-Parry JS, Campbell CT (2004) Quantitative Methods for Spatially Resolved Adsorption/Desorption Measurements in Real Time by Surface Plasmon Resonance Microscopy. *Anal Chem* 76: 907–917
10. Campbell CT, Kim G (2007) SPR microscopy and its applications to high-throughput analyses of biomolecular binding events and their kinetics. *Biomaterials* 28: 2380–2392
11. Smith EA, Corn RM (2003) Surface Plasmon Resonance Imaging as a Tool to Monitor Biomolecular Interactions in an Array Based Format. *Appl Spectrosc* 57: 320A–332A
12. Lee HJ, Nedelkov D, Corn RM (2006) Surface plasmon resonance imaging measurements of antibody arrays for the multiplexed detection of low molecular weight protein biomarkers. *Anal Chem* 78: 6504–6510
13. Espina V, Woodhouse EC, Wulfkühle J et al (2004) Protein microarray detection strategies: Focus on direct detection technologies. *J Immunol Methods* 290: 121–133
14. Kimple AJ, Muller RE, Siderovski DP et al (2010) A capture coupling method for the covalent immobilization of hexahistidine tagged proteins for surface plasmon resonance. *Methods Mol Biol* 627: 91–100
15. Herne TM, Tarlov MJ (1997) Characterization of DNA probes immobilized on gold surfaces. *J Am Chem Soc* 119: 8916–8920
16. Gedig ET (2008) Surface Chemistry in SPR Technology. In: Schasfoort RBM, Tudos AJ (eds) *Handbook of Surface Plasmon Resonance*, RSC Publishing, Cambridge, UK, pp 173–220
17. Cretich M, Damin F, Pirri G et al (2006) Protein and peptide arrays: Recent trends and new directions. *Biomol Eng* 23: 77–88
18. Krishnamoorthy G, Carlen ET, Deboer HL et al (2010) Electrokinetic Lab-on-a-BioChip for Multi-ligand/Multi-analyte Biosensing. *Anal Chem* 82: 4145–4150
19. Goodrich TT, Lee HJ, Corn RM (2004) Direct Detection of Genomic DNA by Enzymatically Amplified SPR Imaging Measurements of RNA Microarrays. *J Am Chem Soc* 126: 4086–4087
20. Zhou H, Bouwman K, Schotanus M et al (2004) Two-color, rolling-circle amplification on antibody microarrays for sensitive, multiplexed serum-protein measurements. *Genome Biol* 5: R28
21. Li Y, Hye JL, Corn RM (2007) Detection of protein biomarkers using RNA aptamer microarrays and enzymatically amplified surface plasmon resonance imaging. *Anal Chem* 79: 1082–1088
22. He L, Musick MD, Nicewarner SR et al (2000) Colloidal Au-enhanced surface plasmon resonance for ultrasensitive detection of DNA hybridization. *J Am Chem Soc* 122: 9071–9077
23. Li Y, Wark AW, Lee HJ et al (2006) Single-nucleotide polymorphism genotyping by nanoparticle-enhanced surface plasmon resonance imaging measurements of surface ligation reactions. *Anal Chem* 78: 3158–3164
24. Fang S, Lee HJ, Wark AW et al (2006) Attomole microarray detection of MicroRNAs by nanoparticle-amplified SPR imaging measurements of surface polyadenylation reactions. *J Am Chem Soc* 128: 14044–14046
25. Gifford LK, Sendroui IE, Corn RM et al (2010) Attomole detection of mesophilic DNA polymerase products by nanoparticle-enhanced surface plasmon resonance imaging

- on glassified gold surfaces. *J Am Chem Soc* 132: 9265–9267
26. Schasfoort RBM, McWhirter A (2008) SPR Instrumentation. In: Schasfoort RBM, Tudos AJ (eds) *Handbook of Surface Plasmon Resonance*, The Royal Society of Chemistry, Cambridge UK, pp 35–80
 27. Beusink JB, Lokate AMC, Besselink GAJ et al (2008) Angle-scanning SPR imaging for detection of biomolecular interactions on microarrays. *Biosensors and Bioelectronics* 23: 839–844
 28. Nedelkov D (2010) Integration of SPR biosensors with mass spectrometry (SPR-MS). *Methods Mol Biol* 627: 261–268
 29. Borch J, Roepstorff P (2010) SPR/MS: recovery from sensorchips for protein identification by MALDI-TOF mass spectrometry. *Methods Mol Biol* 627: 269–281
 30. Situ C, Mooney MH, Elliott CT et al (2010) Advances in surface plasmon resonance biosensor technology towards high-throughput, food-safety analysis. *TrAC – Trends in Analytical Chemistry* 29: 1305–1315
 31. Larsericsdotter H, Jansson Ö, Zhukov A et al (2006) Optimizing the surface plasmon resonance/mass spectrometry interface for functional proteomics applications: How to avoid and utilize nonspecific adsorption. *Proteomics* 6: 2355–2364
 32. Visser NFC, Scholten A, Van Den Heuvel RHH et al (2007) Surface-plasmon-resonance-based chemical proteomics: Efficient specific extraction and semiquantitative identification of cyclic nucleotide-binding proteins from cellular lysates by using a combination of surface plasmon resonance, sequential elution and liquid chromatography-tandem mass spectrometry. *ChemBioChem* 8: 298–305
 33. Nedelkov D, Nelson RW, Kiernan UA et al (2003) Detection of bound and free IGF-1 and IGF-2 in human plasma via biomolecular interaction analysis mass spectrometry. *FEBS Lett* 536: 130–134
 34. Nedelkov D, Nelson RW (2000) Practical considerations in BIA/MS: Optimizing the biosensor-mass spectrometry interface. *Journal of Molecular Recognition* 13: 140–145
 35. Nedelkov D, Nelson RW (2003) Surface plasmon resonance mass spectrometry: Recent progress and outlooks. *Trends Biotechnol* 21: 301–305
 36. Grote J, Dankbar N, Gedig E et al (2005) Surface plasmon resonance/mass spectrometry interface. *Anal Chem* 77: 1157–1162
 37. Bellon S, Buchmann W, Gonnet F et al (2009) Hyphenation of surface plasmon resonance imaging to matrix-assisted laser desorption ionization mass spectrometry by on-chip mass spectrometry and tandem mass spectrometry analysis. *Anal Chem* 81: 7695–7702
 38. Nedelkov D (2007) Development of surface plasmon resonance mass spectrometry array platform. *Anal Chem* 79: 5987–5990
 39. Steiner G (2004) Surface plasmon resonance imaging. *Analytical and Bioanalytical Chemistry* 379: 328–331

Part II

Protocols

Biomimetic Affinity Ligands for Immunoglobulins Based on the Multicomponent Ugi Reaction

Graziella El Khoury, Laura A. Rowe, and Christopher R. Lowe

Abstract

Affinity chromatography is the method of choice for biomolecule separation and isolation with highly specific target recognition; it is ideally suited to the purification of immunotherapeutic proteins (i.e., mAbs). Conventional affinity purification protocols are based on natural immunoglobulin (Ig)-binding proteins, which are expensive to produce, labile, unstable, and exhibit lot-to-lot variability. Biological ligands are now being replaced by cost-effective, synthetic ligands, derived from the concepts of rational design and combinatorial chemistry, aided by *in silico* approaches. In this chapter, we describe a new synthetic procedure for the development of affinity ligands for immunoglobulins based on the multicomponent Ugi reaction. The lead ligand developed herein is specific for the IgG-Fab fragment and mimics Protein L (PpL), an IgG-binding protein isolated from *Peptostreptococcus magnus* strains and usually used for the purification of antibodies and their fragments.

Key words: Affinity, Biomimetic ligands, Combinatorial synthesis, Protein purification, Multi-component reaction, Rational design, Ugi reaction

1. Introduction

The field of proteomics now requires the development of cost-effective purification processes that allow the resolution of complex protein mixtures and the isolation of individual target proteins in high yield, in a relatively short period of time. This is due to the rapid increase in the number of immunotherapeutic proteins (24 mAb products have been approved by the FDA (1) and 240 mAbs are currently in clinical trials (2)). Affinity chromatography is still the only recognized purification technique that can unite the key issues of specific biomolecular target recognition (3) with suitability for large-scale production processes. A number of immunoglobulin-binding proteins (i.e., Protein A, G and L), based on

naturally occurring proteinaceous components of bacteria, have long been used for the purification of therapeutic immunoglobulins and their fragments. However, for large-scale applications, these biological ligands suffer from a wide range of limitations: They are expensive to produce and purify, may be contaminated with host DNA and viruses, show lot-to-lot variation, low scale-up potential, and instability leading to ligand leakage during conventional sterilization-in-place (SIP) and cleaning-in-place (CIP) procedures and subsequent product contamination. The conventional purification protocols are now being substituted with highly selective and sophisticated strategies based on the concept of biomimetic affinity ligands (4).

Rational design, combinatorial chemistry, and high-throughput screening have contributed to the development of synthetic biomimetic ligands for affinity chromatography. The ligand synthesis follows a defined five-part strategy which comprises: (1) Identification of a target site and design of a complementary ligand based on X-ray crystallographic studies of complexes between the natural target protein and the biological ligand; (2) solid-phase synthesis and evaluation of an intentionally biased combinatorial library of related ligands; (3) screening of the ligand library for binding to the target protein in affinity chromatography; (4) selection and characterization of the lead ligand, supported by *in silico* molecular modeling and docking of the ligand into the target protein, and (5) optimization of the adsorbent and chromatographic parameters for the purification of the target protein. This method results in a fully synthetic, chemically defined, non-toxic, and inexpensive affinity ligand which can be used for the purification of high value biopharmaceutical products (5–8).

Previously, this approach has generated a plethora of biomimetic affinity ligands specific for different biopharmaceuticals of interest, including insulin (9), elastase (10), glycoproteins (11), recombinant factor VIIa (12), and immunoglobulins (6–8, 13, 14). The ligands were synthesized on a solid-phase matrix with a triazine scaffold (15) displaying functional groups that mimic peptidal templates consisting of two or three amino acid analogues. However, this route comprised a multistep synthetic procedure requiring temperature changes between 0 and 90°C.

In this chapter, we describe the development of an alternative approach for generating synthetic affinity ligands based on the multicomponent Ugi reaction. Multicomponent reactions (MCRs) possess a number of distinct advantages over more conventional “2-component” methods previously used for synthetic ligand generation. MCRs allow for a greater diversity of ligands by incorporating three or more reactants, each of which can be varied systematically to produce a huge variety of subtle changes to the final ligand structure. The Ugi reaction is a four-component reaction in which an oxo-component (aldehyde or ketone), a primary

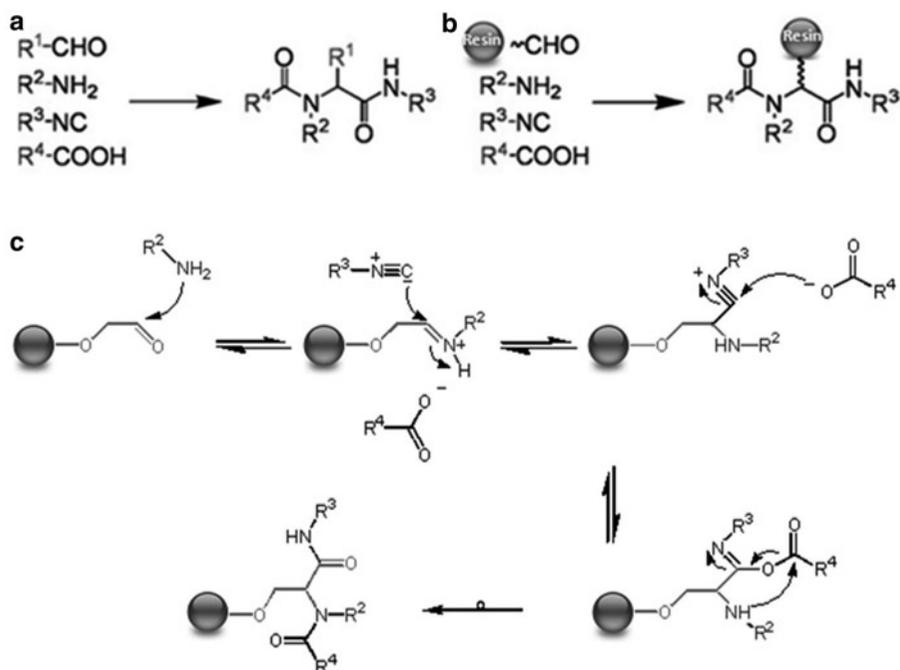


Fig.1. Ugi reaction product and scaffold formation on solid-support. (a) The Ugi reaction comprises an oxo-component (aldehyde or ketone) (R^1), a primary or secondary amine (R^2), an isonitrile group (R^3), and a carboxylic acid (R^4), which are condensed to yield a single product (1) in a “one-pot” reaction. (b) The proposed solid-phase Ugi reaction comprises an aldehyde-functionalized matrix to which the other three solution-phase components are added to yield a single ligand scaffold (2) (reproduced from (17) with permission from Elsevier). (c) Putative solid-phase Ugi reaction mechanism for affinity chromatography ligand formation on aldehyde-functionalized chromatography sorbent. The diagram shows only one ligand per bead for clarity. (Reproduced from (16) with permission from the author).

or secondary amine, a carboxylic acid and an isonitrile group are condensed, in a one-pot reaction conducted at a constant temperature ($50^\circ C$), to yield a di-amide scaffold product (Fig. 1, (16, 17)). A short survey of the number of commercially available compounds suitable for this particular multicomponent chemistry (Table 1) reveals the potential for this approach to increase scaffold diversity and expand the range of applications appropriate for these novel affinity adsorbents. The Ugi reaction also makes it possible to develop branched, cyclized, and 3D affinity ligands. Moreover, the advantage of the Ugi ligand over the triazine scaffold is the ability of the new ligand to correctly mimic a native dipeptide bond. The difference in the interatomic distances between the O^1-N-O^2 pharmacophore in the Ugi scaffold and the native dipeptide bond is less than 1.0 \AA between all three atoms.

We describe herein the design, synthesis, and evaluation of an IgG binding ligand library through the novel use of the multicomponent Ugi reaction. The rational design of the Ugi ligand library was based on previously generated, de novo designed triazine ligands: (1) Artificial Protein A (ApA) (6); (2) an optimized variant

Table 1
Current list of commercially available Ugi reaction components from the Available Chemicals Directory (ACD) as of September 2007 with a total number of 1.5×10^{15} possible Ugi product variations (corrected and reproduced from (17) with permission from Elsevier)

| Functional group | | Commercial availability |
|--------------------------|-------------------|-------------------------|
| Primary/secondary amines | R-NH ₂ | 95, 398 |
| Aldehydes | R-CHO | 10, 982 |
| Isonitrile | R-NC | 644 |
| Carboxylic acid | R-COOH | 2,158 |

There are approximately 3,000 isonitriles currently known; however, not all are commercially available (18)

of ApA known as 22/8 (8); and (3) the biomimetic Protein L 8/7 (13), which yielded putative lead ligands for whole and fragmented (Fc and Fab) immunoglobulins. The ligand library was initially screened chromatographically against whole human IgG and fragmented (Fc and Fab) molecules.

A Fab-specific ligand (A3C111) comprising an Ugi scaffold substituted with 1-amino-2-naphthol (A3), glutaric acid (C1) and isopropyl isocyanide (I1) was selected based on its ability to bind Fab differentially over Fc. Preparative chromatography of Fab and IgG from both yeast and *Escherichia coli* host cell proteins showed that 100% of Fab or IgG that loaded (100 µg/mL) was adsorbed, and subsequently eluted with a purity of 83% according to silver-stained SDS-PAGE lane densitometry. The optimized affinity adsorbent was defined with a dynamic binding capacity of 73 mg IgG/mL moist gel, and a static binding capacity of 16.1 ± 0.25 mg Fab/mL moist resin, and displayed an affinity constant of K_d : $(2.6 \pm 0.3) \times 10^{-6}$ M. The Ugi ligand was modeled in silico and docked into a human Fab fragment to suggest a putative binding interface to the constant CH₁-CL Fab terminal through six defined hydrogen bond interactions together with putative hydrophobic interactions.

This lead ligand demonstrates the potential of the Ugi scaffold for the development of future affinity ligands. Furthermore, the relative ease of conducting the Ugi reaction suggests that the scale-up of Ugi ligand synthesis may prove more cost-effective than the two-step triazine ligand synthesis route, further supporting the use of the Ugi scaffold as a credible alternative for the generation of selective affinity adsorbents for immunoproteins (18).

2. Materials

2.1. Design

1. In silico molecular modeling. Different software packages are commercially available to perform molecular modeling, such as Insight II from Accelrys, Inc. and Molegro Virtual Docker 2010, MVD v4.0.2 from Molegro ApS, which can run on a standard desktop PC. Protein X-ray and nuclear magnetic resonance (NMR) crystallographic structures are available from the Brookhaven database (<http://www.rcsb.org/pdb/>) which contains over 66,000 entries.

2.2. Synthesis of the Ugi Ligands

1. Cross-linked agarose (Sepharose™ CL-6B) (GE Healthcare, Uppsala, Sweden) can be obtained as a suspension of beads in a 20% (v/v) aqueous ethanol solution. Store at 4°C, avoiding periods of dryness. Agitation of gel suspensions, when required, should be done with an orbital shaker and not a magnetic stirrer.
2. Grade 2 sintered funnel.
3. Hybaid Maxi 14 hybridization oven (Thermo Electron, UK).
4. Captiva™ 96-well block (Varian, UK) containing a 20 μm polypropylene frit at the bottom of each well.
5. Epichlorohydrin (1-chloro-2, 3-epoxypropane). Widely available chemical, used to epoxy-activate the Sepharose™ CL-6B beads or other surfaces. A high-purity (≥99%) should be used (see Note 1).
6. Sodium hydroxide NaOH: 5 M solution in water (see Note 2).
7. Sodium periodate NaIO₄: 0.1 M solution in water (see Note 3).
8. Amines, carboxylic acids, and isonitriles. A wide range of primary/secondary amines (95,398), carboxylic acids (2,158), and isonitriles (644) are commercially available from the Available Chemical Directory (ACD) as of September 2007 (Table 1). The compounds used for the Ugi reaction on the aldehyde activated agarose vary from case to case; they must be dissolved in an appropriate buffer, either an aqueous solution or an organic solvent such as methanol. Table 2 presents the list of the different amines (A1–A8), carboxylic acids (C1–C6), and isonitrile (I1) components used for the Ugi combinatorial library prepared in this work (see Note 4).

2.3. Solution-Phase Synthesis and Ligand Characterization

1. All the reagents and solvents are of reagent grade: Sodium hydroxide, magnesium sulfate, 1-amino-2-naphthol, pentanoic acid, isopropyl isocyanide, acetaldehyde, ethyl acetate, heptane, and methanol.
2. 0.1 M Hydrochloric acid.
3. Round bottom flask (50 mL).
4. Rotary evaporator.

Table 2
Structure of the amine (A1–A8), carboxylic (C1–C6), and isonitrile (I1)
components of the IgG-binding Ugi combinatorial library

| Number | Structure | Component | Number | Structure | Component |
|--------|-----------|----------------------|--------|-----------|--------------------------------|
| A1 | | Tyramine | C1 | | Glutaric acid |
| A2 | | 4-Aminobenzamide | C2 | | 3,5-Pyridine dicarboxylic acid |
| A3 | | 1-Amino-2-naphthol | C3 | | Isophthalic acid |
| A4 | | 1-Amino-2-propanol | C4 | | Boc-glutamine |
| A5 | | 4-Aminophenol | C5 | | Benzoic acid |
| A6 | | 3-Aminophenol | C6 | | Acetic acid |
| A7 | | 4-Hydroxybenzylamine | I1 | | Isopropylisocyanide |
| A8 | | 8-Amino-naphthol | | | |

The isopropyl isocyanide (I1) was conserved for the entire combinatorial library (reproduced from (17) with permission from Elsevier)

5. Separating funnel.
6. Silica column (h: 20 cm, d: 2 cm).
7. Joel JNM Lambda LA400 FT NMR spectrometer.
8. AEI MS30 or AEI MS50 mass spectrometers.

2.4. Chromatographic Analysis

1. Binding or equilibration and elution buffers. The buffers vary from case to case, depending especially on the type of interactions between the ligand and the protein to be purified. The equilibration and elution buffers can be selected according to the standard conditions recommended for IgG-binding proteins used in affinity chromatography assays.

For example, the optimal binding and elution buffers for the lead ligand developed herein are 10 mM KH_2PO_4 , pH 6.0, and 0.1 M NaHCO_3 , 0.1% (w/v) CHAPS, pH 10, respectively.

2. Regeneration buffer: 0.1 M NaOH in 30% (v/v) isopropanol (see Note 5).
3. Disposable empty columns (4 mL) with 20 μm frits (Varian, Inc., UK) for preparative small-scale assays.
4. ÄKTA explorer™ FPLC system with a UV/Vis 920 module, operated with Unicorn™ software suite (GE Healthcare, Uppsala, Sweden).

2.5. Partition Equilibrium Studies

1. Protein solution to be tested and the immobilized ligand to be studied.
2. Unmodified Sepharose™ CL-6B beads.
3. Equilibration and regeneration buffers (Subheading 2.3).
4. Grade 2 sintered funnel.
5. Eppendorf tubes.
6. Nanodrop ND1000 spectrophotometer (Thermo Fisher Scientific, USA).

2.6. Dynamic Binding Capacity by Frontal Analysis

1. Protein solution to be tested and immobilized ligand to be studied.
2. Equilibration, elution, and regeneration buffers (Subheading 2.3).
3. Borosilicate (Pyrex) column (OMNI, Cambridge, UK).
4. ÄKTA explorer™ FPLC system.

3. Methods

3.1. Design

Rational design of biomimetic Ugi ligands. The Ugi reaction components were selected from the list of different commercially available chemicals (Table 2); they adhere to a nomenclature based on AxCyIz: “A” refers to the amine component number from the combinatorial library list, “C” the carboxylic acid, and “I” the isonitrile. The design of the combinatorial library presented in this chapter was based on a number of triazine-based biomimetic ligands previously designed for the purification of whole and fragmented IgG (i.e., Artificial Protein A (ApA), ligand 22/8, and Protein L mimetic ligand 8/7 (Fig. 2)) (6, 8, 14) together with consideration

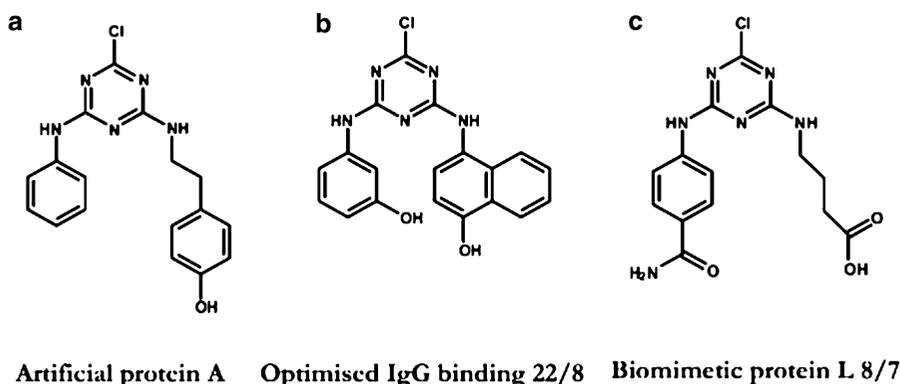


Fig. 2. Triazine-based immunoglobulin specific ligands. (a) Artificial Protein A (6); (b) optimized IgG-binding ligand 22/8 (8); (c) Protein L biomimetic ligand 8/7 (14). The ligand nomenclature refers to combinatorial triazine library components as described in the relevant references (reproduced from (17) with permission from Elsevier).

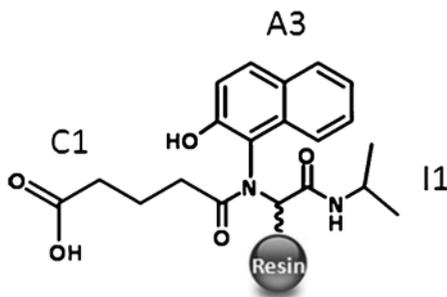


Fig. 3. Solid-phase spacer arm attachment and aldehyde group functionalization (reproduced from (16) with permission from the author).

of the literature based on the X-ray crystallographic studies of the primary binding interface between Protein L (PpL) and human Fab fragment (PDB: 1HEZ) (19, 20). These studies revealed seven key residues conserved in different Protein L domains and largely buried upon the complex formation between the protein and IgG light chains. These residues are listed below, followed by their italicized Ugi library analogues incorporated into the Ugi combinatorial library: Gln³⁵: *4-aminobenzamide* (A2) and *Boc-glutamine* (C4); Thr³⁶: *1-amino-2-propanol* (A4); Ala³⁷: *acetic acid* (C6); Glu³⁸: *glutamic acid* (C1), *2,4-pyridine dicarboxylic acid* (C2) and *isophthalic acid* (C3), Phe³⁹: *benzoic acid* (C5); Lys⁴⁰ and Tyr⁵³: *tyramine* (A1), *4-aminophenol* (A5), *3-aminophenol* (A6), and *4-hydroxybenzylamine* (A7) (see Note 6). For example, the lead ligand A3C1I1 (Fig. 3), protein L mimic, developed herein refers to an Ugi ligand comprised of amine component three (A3: 1-amino-2-naphthol, mimicking a tyrosine), carboxylic component one (C1: glutamic acid), and isonitrile component one (I1: isopropyl isocyanide). However, the aldehyde component for all the ligand libraries was restricted to the solid-phase functionality.

3.2. Solid-Phase Combinatorial Synthesis of a Ligand Library

1. Epoxyactivation of agarose beads. Wash the required amount of Sepharose CL-6B with 40 mL of distilled water/g of gel on a sintered funnel (see Note 7). Transfer the washed agarose to a 1 L conical flask and add 1 mL of distilled water/g of gel. To this moist gel, add 0.8 mL of NaOH/mL of gel and incubate the slurry for 1 h at 25°C on a rotary shaker. Raise the temperature to 34°C and add 1 mL of epichlorohydrin/mL of gel to the reaction mixture (21). Wash the epoxide-activated resin with 40 mL distilled water/g of gel on a grade 2 sintered funnel and use it directly for *cis*-diol activation (Fig. 4; Resin B). Determine the epoxy-content according to Note 7.
2. *Cis*-diol activation. Treat the washed epoxy-activated gel with 1 mL of 5 M NaOH/mL of gel and stir overnight at 34°C. This base-catalyzed procedure gradually hydrolyses the epoxide ring resulting in the formation of a *cis*-diol reaction product (Fig. 4; Resin C).

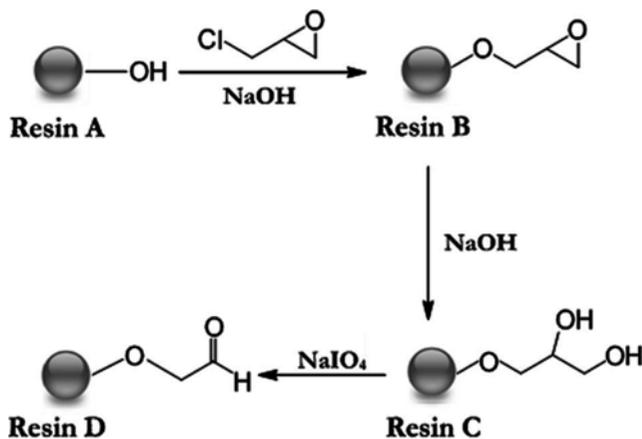


Fig. 4. Ugi lead ligand: A3C111 specific for IgG-Fab fragment (reproduced from (16) with permission from the author).

- Aldehyde functionalization. Treat the diol-activated resin with 2 mL of 0.1 M sodium periodate NaIO_4 /g of moist weight gel and leave to stir at 30°C for 3 h. This procedure causes the cleavage of the *cis*-diol, leaving a terminally functionalized aldehyde group (Fig. 4; Resin D).
- Imine (Schiff base) formation. Wash the aldehyde-functionalized resin with different methanol solutions, starting with 10% (v/v) methanol and finishing with 100% (v/v) methanol at 10% (v/v) increments (see Note 8). Slurry the methanol-saturated resin in 100% (v/v) methanol (1 mL/g of gel) and place it on a shaker with gentle shaking to prevent the resin from settling. Divide the gel into n aliquots, where n is the number of different amines used to synthesize the combinatorial library (see Note 9). Dissolve a fivefold molar excess (relative to the amount of epoxide in the gel) of each amine in methanol (0.25 mL/mL of slurry). Incubate the n aliquots with the previous mixture at 25°C in a rotary shaker (200 rpm) for 1 h. This procedure allows the amine component to become completely mixed with the supplied resin sample and promote imine formation (22).
- Addition of the carboxylic acid and the isonitrile. Dissolve the m selected carboxylic acids in 0.25 mL of methanol/mL of slurry. Each carboxylic acid is in 5 molar excess to the amount of epoxide in the gel. Finally, add a fixed aliquot (0.25 mL) of the isopropyl isocyanide (II, Table 2) component (5 molar excess, in methanol) into each of the ($n \times m$) wells (see Note 10). Incubate the reactants at 50°C in a rotary shaker (200 rpm) for 48 h. At the end of the synthesis, wash the gels with appropriate solvents. Weigh the gels, and store at $0\text{--}4^\circ\text{C}$ in 20% (v/v) ethanol (see Notes 11 and 12).

3.3. Solution-Phase Synthesis of the Ugi Ligands

Direct physicochemical analysis of synthetic affinity ligands is difficult when immobilized to the solid-phase bead (i.e., in situ). Solution-phase synthesis of lead compounds is often performed to allow characterization of the ligand by $^1\text{H-NMR}$, $^{13}\text{C-NMR}$, and mass spectrometry (MS) techniques (6, 10, 12, 14). We describe, herein, the synthesis of the solution-phase modified Ugi ligand, PpL-mimic (amine: 1-amino-2-naphthol; carboxylic acid: pentanoic acid; isonitrile: isopropyl isocyanide; aldehyde: acetaldehyde) (see Note 13).

1. Add 1.16 mL of 5 M sodium hydroxide in a 50 mL round bottom flask and 0.24 mL (0.19 g, 4.3 mmol) of acetaldehyde to 0.84 g (4.3 mmol) of 1-amino-2-naphthol in 3 mL methanol. Incubate and stir for 1 h at 21°C to ensure Schiff-base formation.
2. Add 0.47 mL (0.44 g, 4.3 mmol) of pentanoic acid in 3 mL methanol and 0.41 mL (0.30 g, 4.3 mmol) of isopropyl isocyanide in 3 mL methanol and leave to stir for 48 h at 21°C.
3. Remove the methanol with a rotary evaporator (50°C) and dissolve the product in 30 mL of ethyl acetate with 50 mL of 0.1 M hydrochloric acid. Add the solution to a separating funnel; add an additional 70 mL of ethyl acetate, mix and leave to settle.
4. Extract the lower aqueous phase and add an additional 30 mL of ethyl acetate to the separating funnel, mix, allow to settle, and extract the lower aqueous phase. Add 30 mL of distilled water to the separating funnel, mix, allow to settle, and extract the lower aqueous phase.
5. Add ~30 g of anhydrous magnesium sulfate to the remaining organic phase in order to remove any residual water, and incubate for 30 min at 21°C.
6. Filter the mixture and concentrate prior to dry loading onto a silica column (h: 20 cm, d: 2 cm) using the solvent system: Heptane: ethyl acetate (10:1).

3.4. Chromatographic Screening

The screening of the Ugi ligands is performed by affinity chromatography at room temperature.

1. Pack the synthesized adsorbents (0.4 mL resin – 50% (w/v) (prepared slurry) under gravity into 4 mL (0.8 × 6.0 cm) polypropylene columns. Wash each resin with 2 mL (20 c.v.: column volume) of regeneration buffer and with 2 mL of distilled water to bring the pH to neutral. Equilibrate the resins with 2 mL of equilibration buffer.
2. Reconstitute the protein to be purified to 0.5 mg/mL in equilibration buffer. Load 100 μL of the protein solution onto each column. Wash the columns with 2 mL of equilibration buffer and collect fractions of 200 μL (see Note 14).
3. Elute the bound protein with 2 mL of elution buffer (collect fractions of 200 μL).

4. After elution, regenerate the columns with 2 mL of the regeneration buffer. Then wash the columns with 2 mL of distilled water, equilibrate with the equilibration buffer (2 mL), and store at 0–4°C in 20% (v/v) ethanol.

3.5. Partition Equilibrium Studies

1. Wash the immobilized ligand to be studied with regeneration buffer and then equilibrate in equilibration buffer.
2. Prepare a series of Eppendorf tubes with 100 μL of standard protein solutions in equilibration buffer (0.1–10 mg/mL; confirm concentration by A_{280} measurement).
3. Add the immobilized ligand (0.01 g of moist weight gel previously dried under vacuum in a sintered funnel) to each Eppendorf tube and incubate for 24 h at room temperature and under orbital agitation.
4. Centrifuge the Eppendorf tubes (1 min; 1,430 $\times g$) to settle the matrix and take the supernatant to measure the A_{280} using a Nanodrop ND1000 spectrophotometer. The control experiment involves incubating the partitioning solute in unmodified Sepharose™ CL-6B beads.
5. Fit the data collected from these experiments into Scatchard plots to determine the association constant (K_a) and the maximum binding capacity (Q_{\max}).

3.6. Dynamic Binding Capacity by Frontal Analysis

Frontal analysis is usually the method of choice for measuring the dynamic binding capacity of an affinity support (see Note 16). The dynamic binding capacity for the lead Ugi ligand A3C111 (Fig. 4) is determined using an ÄKTA explorer™ with flow packed column.

1. Pack the affinity adsorbent into a borosilicate (Pyrex) column (c.v.: 444 μL (h: 1.3 cm, 0.66 cm i.d.)) at a flow rate of 0.17 mL/min. Wash the resin with the regeneration buffer (10 c.v.), then with distilled water (10 c.v.) and equilibrate with the equilibration buffer (10 c.v.).
2. Reconstitute the protein to be tested to 8.96 mg/mL in equilibration buffer and load the solution continuously onto the column until an A_{280} saturation plateau is achieved ($(\text{Protein})_{\text{plateau}}$, Fig. 5).
3. Wash the column with the equilibration buffer until A_{280} is re-zeroed.
4. Elute the bound protein with the elution buffer.
5. Regenerate the column with the regeneration buffer followed by distilled water, equilibrate with the equilibration buffer and store at 0–4°C in 20% (v/v) ethanol. The fractions are continuously analyzed ($A_{280\text{nm}}$) using the UVis-920 module of the ÄKTA explorer™.

Affinity chromatography is also used to test the ability of a lead ligand to purify the target protein from crude extracts (see Note 16 and Fig. 6).

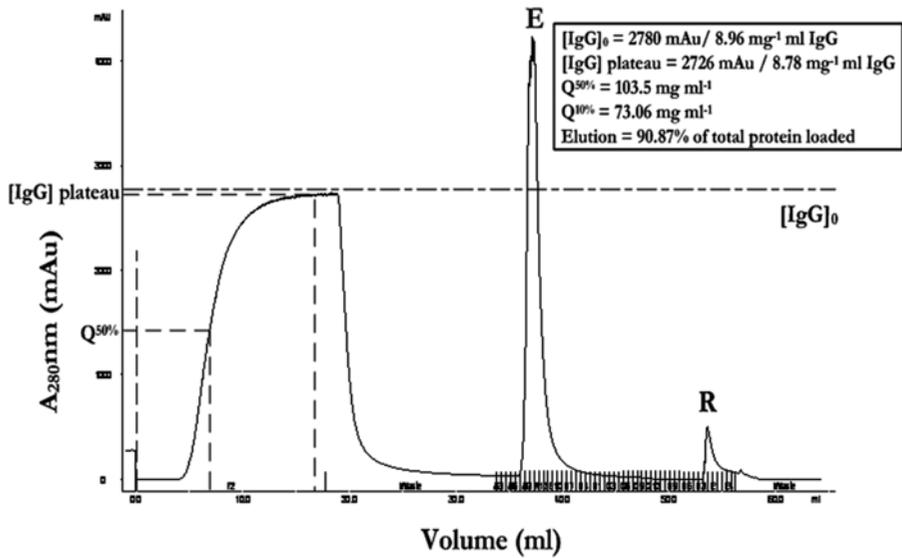


Fig. 5. Frontal analysis ÄKTA explorer™ chromatogram for A3C111 and IgG showing elution (E) and regeneration (R) peaks (reproduced from (16) with permission from the author).

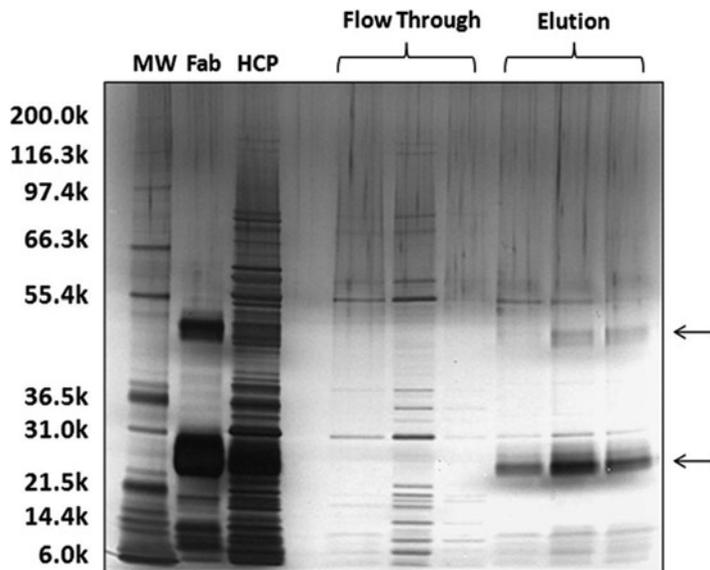


Fig. 6. SDS-PAGE silver stain analysis of the fractions generated from the chromatographic isolation of Fab fragment from *E. coli* host cell protein (HCP) extracts, with A3C111 immobilized Sepharose™. The ~25 and 50 kDa protein species (arrows) present in the elution fractions but not in the flow through fractions are of similar molecular weights to the major reduced Fab bands in the Fab standard, indicating that all the Fab was absorbed by A3C111 from the *E. coli* HCP preparation. Gel images were analyzed using Image J (V1.37V) analysis program and Fab purity (83%) was assessed for the main elution peak fraction based on gel band densitometry. (MWM: Molecular weight marker with approximate weights (kDa) on the left). (Reproduced from (16) with permission from the author).

4. Notes

1. Epichlorohydrin is a very unstable compound and must be stored in an anhydrous environment at 0–4°C. Hazards: Flammable, poison, toxic by inhalation or contact with skin, may cause an allergic skin reaction, and if swallowed may cause cancer. Toxicity data: LD50 90 mg/kg oral, rat. LC50 Inhalation – rat – 8 h – 250 ppm. Handle in a fume hood with safety glasses and gloves, and treat as a possible cancer hazard.
2. Sodium hydroxide may be harmful if inhaled, swallowed, or absorbed through skin. Hazards: Corrosive. Handle with safety glasses and gloves.
3. Sodium periodate. Harmful and toxic if swallowed, irritant to eyes, skin, and respiratory system. Toxicity data: LD50 264 mg/kg oral, rat. Handle with safety glasses and gloves.
4. Hazards and toxicity data should be considered individually for each compound according to the suppliers recommendations. Caution: The methanol is flammable and toxic and should be handled in a fume hood.
5. The regeneration buffer is used to remove any physically adsorbed protein prior to screening, and after the screening procedure to remove retained protein. Special care should be taken when using isopropanol. Hazards: Flammable, irritant to eyes, skin, and respiratory system. Toxicity data: LD50 10 g/kg oral, human. Handle with safety glasses and gloves.
6. A number of the selected library components (C1–C3, Table 2) include dicarboxylic acid groups (COOH–X–COOH) which allow incorporation into the Ugi scaffold while contributing polar functionality in the product ligand. Bifunctionality could cause bridging between two adjacent scaffolds if both acidic terminals of the molecule become incorporated in adjacent scaffold formations. At the known low epoxide (and therefore subsequent aldehyde) group densities optimum for small-synthetic ligand affinity chromatography (approximately 18–30 μmol aldehyde groups g^{-1} moist weight gel), the average distance between two adjacent aldehyde groups (100 Å, if it is assumed that the ligands are uniformly distributed throughout the matrix and positioned at the corners of a cubic lattice (1)) would be too great to allow such bridging to occur. Despite this, equimolar sodium hydroxide was added to the solution-phase dicarboxylic acid preparations (10 min before addition to the reaction mixture at 25°C) to encourage an equilibrium toward carboxylate salt formation (COOH–X–COO⁻Na⁺) and thus act as a mono-protection group for one acidic terminal. Postreaction washes of the derivatized resin promote salt removal. However, such steric constraints would not be in place for solution-phase

Ugi synthesis and di-Ugi-scaffold formation to occur, and the incorporation of glutaric acid would prove much more difficult with the potential formation of di-Ugi scaffolds. A number of carboxylic acid mono-protection systems (23) were considered, although conservation of the final ligand structure could not be ensured during the deprotection process. Therefore, a mono-functional carboxylic acid pentanoic acid was selected for the following solution phase studies comprising the same length alkyl chain component as glutaric acid.

7. Extent of epoxyactivation of the agarose beads: Sodium thio-sulfate (1.3 M, 3 mL) is added to 1 g of epoxy-activated gel and incubated at room temperature for 20 min. The mixture is neutralized with 0.1 M HCl and the amount of HCl used is recorded. The volume of 0.1 M HCl added corresponds to the number of moles of OH⁻ released (10 μmol for each 100 μL added), which equals to μmol of epoxy groups/g gel. Therefore, the extent of epoxy activation is expressed as μL HCl used/10 μmol/g gel (24). The protocol usually results in 22 μmol epoxy groups/g moist weight gel.
8. The aldehyde-functionalized resin was subjected to a series of washes of increasing methanol concentration; agarose beads may be subject to solvent-stress-related degradation and dehydration if immediately placed in 100% (v/v) methanol without gradually displacing the water absorbed by the resin.
9. In order to generate a large number of ligands simultaneously, a Captiva™ 96-well block was employed which contains a 20 μm polypropylene frit at the bottom of each well. This chemically resistant block system thereby constitutes the reaction vessel and the subsequent storage facility at the end of the final reaction.

Cut off a 1 mL Gilson™ pipette tip at approximately 2 mm from the end to allow for the easy transfer of 1 mL slurry aliquots (methanol-saturated resin in methanol) into the wells of the reaction block (8 × 12). Remove the flexible end cap-mat at this stage to allow the solvent to completely drain through and thus allow the resin to settle in the block. Replace the end cap-mat in position at the bottom of the block. After all the different amine solutions are loaded into the wells, attach the top cap-mat firmly to the block and shake (1 h, 200 rpm, 25°C).
10. The protonation by the carboxylic acid moiety to give an iminium ion followed by a nucleophilic attack of the isonitrile leads to the formation of a nitrilium ion intermediate, which is subsequently intercepted by the corresponding carboxylate anion. The resulting imino anhydride typically undergoes an irreversible transacylation (Mumm rearrangement) to give a final Ugi product (Fig. 1c). For the construction of a 2D library array, only two of the four possible components involved in the Ugi reaction are varied.

11. After the addition of the carboxylic acids and the isonitrile compounds, fix the upper cap-mat firmly to the top of the reaction block. Place the entire block in an incubation oven with a shaking platform (48 h, 200 rpm, 50°C). At the end of the reaction period, remove the lower and upper cap mats carefully and allow the wells to drain for 10 min. Then wash the wells thoroughly according to the procedure below in order to remove unreacted reagents from the resulting resin samples:
 - a. All wash steps constitute 5 mL/well with: (1) 100% (v/v) methanol; (2) 50% (v/v) dimethylformamide: 50% (v/v) methanol; (3) 50% (v/v) dimethylformamide; (4) water; (5) 0.1 M hydrochloric acid; (6) water; (7) 0.2 M sodium hydroxide in 50% (v/v) isopropanol; (8) 2 × water; and (9) 20% (v/v) ethanol.
12. In order to vary the isonitrile component, the same library can be prepared as described (Subheading 3.2), but using an alternative isonitrile component at different positions in the reaction block. In this manner, a number of secondary libraries can easily be generated with various isonitrile components, thus effectively giving rise to a 3D array of ligand structures.
13. The synthesis of A3C111 was conducted utilizing acetaldehyde as the aldehyde component of the reaction and thereby replacing the aldehyde-functionalized solid-phase component. It was noted that the use of an aromatic aldehyde would encourage imine formation, being favored by the high conjugation, but the concentration of the iminium ion in the reaction mixture would be low because of the basicity of the Schiff's base, which is often the sole isolated product in such reactions (22).
14. When performing preparative small-scale assays (4 mL polypropylene columns), total protein concentration in the collected fractions is determined by using a standard Bradford assay protocol with Coomassie Plus™ protein assay reagent. The assay is performed in 96-well standard microtiter plates.
 - a. Standard Bradford assay protocol. Prepare standard BSA (200, 100, 80, 40, and 20 µg/mL) and IgG (100, 75, 50, and 25 µg/mL) solutions and add 50 µL of each solution to a well of a 96-well microtiter plate. Distribute the samples to be quantified across the plate (50 µL/well). Prepare control solutions by adding 50 µL of equilibration buffer and 50 µL of elution buffer to control wells. Add 200 µL of Coomassie Plus™ protein assay reagent to each well. Incubate the plate for 15 min at 22°C, and then read it at 595 nm using a plate reader.

When the automatic ÄKTA explorer™ FPLC system is used, after the sample loading, wash the columns with the equilibration buffer until the absorbance of the collected fractions at 280 nm is re-zeroed (reaches a value ≤ 0.005).

15. Frontal analysis involves the continuous addition of analyte solution of known concentration ($(\text{Protein})_0$) to the affinity column until adsorbent capacity is reached ($(\text{Protein})_{\text{plateau}}$). At this point, the load “breaks through” into the mobile phase until the emerging eluate matches that of the applied solution (25). The characteristics of the breakthrough profile can be analyzed to give a dynamic binding capacity (Q) for each adsorbent at both $Q^{10\%}$ and $Q^{50\%}$ based on a function of $V_c^{10\%}$ and $V_c^{50\%}$ (the elution volume at which 10% or 50% of the total protein breakthrough profile is reached, respectively). Furthermore, the percentage difference between the concentration of the protein loaded ($(\text{Protein})_0$) and that at column saturation ($(\text{Protein})_{\text{plateau}}$) reflects unused capacity of the adsorbent due to mass transfer rate and dispersion in the mass transfer zone (i.e., protein in mobile phase versus protein in complex with the ligand) (Fig. 5) (26).
16. Example of purification of a target protein from real feedstocks. The *E. coli* bacterium is one of the most widely used hosts for the production of heterologous proteins, and an industrial fermentation system that supports high-level recombinant protein production. Dilute a crude extract of *E. coli* host cell protein to 1:4 (v/v) in equilibration buffer, filter and load 200 μL onto a pre-equilibrated column containing 0.4 mL (50% slurry of the immobilized ligand (w/v) in water). Wash the resin extensively with 2 mL of the equilibration buffer until the absorbance at 280 nm of the collected samples reaches ≤ 0.005 . Elute the bound protein with 2 mL of the elution buffer and collect samples (1 mL). Regenerate the column with the regeneration buffer. Analyze the flowthrough and elution sample under reducing SDS-PAGE conditions with silver-staining together with pure Fab and *E. coli* host cell protein (HCP) reference samples.

Prepare reduced samples by adding 13 μL sample to 5 μL NuPAGE LDS Sample Buffer (4 \times) and 2 μL Sample Reducing agent (10 \times) (Invitrogen, USA) and incubate for 15 min at 70°C. Apply samples (20 μL) to a NuPAGE 4–12% Bis–Tris gel loaded into a Novex Mini-cell system filled with 800 ml MOPS SDS Running buffer and 0.5 ml NuPAGE antioxidant (Invitrogen, USA) added to the upper gel tank chamber. Run the gels at 200 V (constant) for 50 min. Remove the gels from the precast gel case, and rinse briefly with D.I. water then silver stain, dry, scan, and analyze the bands corresponding to the target protein using the Image J (V1.37V) analysis program to determine the purity of the elution fractions based on densitometry analysis (Fig. 6).

Acknowledgment

The authors would like to thank Dr. Jonathan M. Haigh for permission to reproduce figures from his PhD thesis at the University of Cambridge (2008).

References

1. Nelson AL, Dhimolea E, Reichert JM (2010) Development trends for human monoclonal antibody therapeutics. *Nat Rev Drug Discov* 9: 767–774
2. Walsh G (2010) Biopharmaceutical benchmarks 2010. *Nat Biotech* 28: 917–926
3. Lowe CR, Dean PDG (1974) *Affinity Chromatography*. John Wiley and Sons, New York
4. Lowe CR, Lowe AR, Gupta G (2001) New developments in affinity chromatography with potential application in the production of biopharmaceuticals. *J Biochem Biophys Methods* 49: 561–574
5. Lowe CR, Burton SJ, Burton N et al (1990) New developments in affinity chromatography. *J Mol Recognit* 3: 117–122
6. Li RX, Dowd V, Stewart DJ et al (1998) Design, synthesis, and application of a Protein A mimetic. *Nat Biotechnol* 16: 190–195
7. Teng SF, Sproule K, Hussain A et al (1999) A strategy for the generation of biomimetic ligands for affinity chromatography. Combinatorial synthesis and biological evaluation of an IgG binding ligand. *J Mol Recognit* 12: 67–75
8. Teng SF, Sproule K, Hussain A et al (2000) Affinity chromatography on immobilized “biomimetic” ligands synthesis, immobilization and chromatographic assessment of an immunoglobulin G-binding ligand. *J Chrom B* 740: 1–15
9. Sproule K, Morrill P, Pearson JC et al (2000) New strategy for the design of ligands for the purification of pharmaceutical proteins by affinity chromatography. *J Chrom B* 740: 17–33
10. Filippusson H, Erlendsson LS, Lowe CR (2000) Design, synthesis and evaluation of biomimetic affinity ligands for elastases. *J Mol Recognit* 13: 370–81
11. Palanisamy UD, Winzor DJ, Lowe CR (2000) Synthesis and evaluation of affinity adsorbents for glycoproteins: an artificial lectin. *J Chrom B* 746: 265–281
12. Morrill PR, Gupta G, Sproule K et al (2002) Rational combinatorial chemistry-based selection, synthesis and evaluation of an affinity adsorbent for recombinant human clotting factor VII. *J Chrom B* 774: 1–15
13. Roque ACA, Lowe CR, Taipa AM (2005a) An artificial protein L for the purification of immunoglobulins and Fab fragments by affinity chromatography. *J Chrom A* 1064: 157–167
14. Roque ACA, Lowe CR, Taipa AM (2005b) Synthesis and screening of a rationally designed combinatorial library of affinity ligands mimicking protein L from *Peptostreptococcus magnus*. *J Mol Recognit* 18: 213–224
15. Lowe CR, Burton SJ, Pearson JC et al (1986) Design and application of bio-mimetic dyes in biotechnology. *J Chrom* 376: 121–130
16. Haigh JM (2008) Novel affinity ligands for immunoglobulins based on the multicomponent Ugi reaction. PhD thesis. University of Cambridge
17. Haigh JM, Hussain A, Mimmack MI et al (2009) Affinity ligands for immunoglobulins based on the multicomponent Ugi reaction. *J Chrom B* 877: 1440–1452
18. Veiderma M (2007) Chemistry of the isocyanides and their multicomponent reactions, including their libraries – the initiatives of Ivar Ugi. *Proc Estonian Acad Sci Chem* 56: 98–102
19. Graille M, Stura EA, Housden NG et al (2001) Complex between *Peptostreptococcus magnus* protein L and a human antibody reveals structural convergence in the interaction modes of Fab binding proteins. *Structure* 9: 679–687
20. Graille M, Harrison S, Crump MP et al (2002) Evidence for plasticity and structural mimicry at the immunoglobulin light chain-protein L interface. *J Biol Chem* 277: 47500–47506
21. Porath I, Fornstedt N (1970) Group fractionation of plasma proteins on dipolar ion exchangers. *J Chrom A* 51: 479–489
22. Marcaccini S, Torroba T (2007) The use of the Ugi four-component condensation. *Nat Protoc* 2: 632–639
23. Sundberg L, Porath J (1974) Preparation of adsorbents for affinity chromatography: I. attachment of group-containing ligands to insoluble polymers by means of bifunctional oxiranes. *J Chrom A* 90: 87–98

24. Winzor D J (2004) Determination of binding constants by affinity chromatography. *J Chrom A* 1037: 351–367
25. Kasche V (2001) Physico-chemical boundaries in the continuous and one-step discontinuous affinity-chromatographic isolation of proteins *J Biochem Biophys Methods* 49: 49–62
26. Greene TW, Wuts PGM (1991) Protective groups in organic synthesis. John Wiley and Sons, New York

Preparation of Photo-Cross-Linked Small Molecule Affinity Matrices for Affinity Selection of Protein Targets for Biologically Active Small Molecules

Hiroshi Takayama, Takashi Moriya, and Naoki Kanoh

Abstract

Small molecule-immobilized affinity matrices are indispensable tools in chemical genomics to screen and purify protein targets for biologically active small molecules. Usually, prior to immobilization, small molecules have to be derivatized at a position that does not significantly abrogate the intrinsic biological activity, or chemically synthesized to have an appropriate functional group for the immobilization chemistry. Here, we describe a photo-cross-linking technique to immobilize biologically active small molecules for protein target screening, without the need for chemical derivatization.

Key words: Photo-cross-linking, Affinity matrices, Biologically active small molecule, Immobilization, Diazirine, Photo-generated carbene

1. Introduction

Affinity purification of protein targets for biologically active small molecules using small molecule-immobilized affinity matrices has been utilized over the last several decades (1). From the late 1980s to the mid-1990s, this technology has had a tremendous impact upon the field of chemical biology, mostly through the work of Schreiber and coworkers (2, 3). These authors chemically synthesized tagged derivatives of biologically active natural products possessing a linker moiety to be attached to a solid support, and then introduced them onto affinity matrices. The prepared small molecule-immobilized affinity matrices were then incubated with cell lysates of mammalian cells, and after washing, the bound proteins were resolved on sodium dodecyl sulfate (SDS) polyacrylamide gels and analyzed by various methods. This technique has been successfully utilized to purify and

identify the cellular protein targets for bioactive natural products such as the immunosuppressant FK506 (**2**) and the histone deacetylase (HDAC) inhibitor trapoxin (**3**).

To realize the potential of small molecule-immobilized affinity matrices, the method of introducing small molecules onto the affinity matrices is a critical step. However, as stated earlier, this step usually requires substantial skills in, and knowledge of, synthetic organic chemistry; it also requires a tailor-made synthetic (or derivatization) strategy for the immobilization of the small molecules. Thus, the introduction of small molecules onto the affinity matrices constitutes a bottleneck in the process of affinity-based drug target discovery.

To circumvent the difficulties associated with the derivatization, and to facilitate the target identification of bioactive small molecules, we have developed a rapid and generic strategy in which a photo-generated carbene is used to immobilize (i.e., photo-cross-link) small molecules onto solid surfaces (**4**, **5**). In this method, aryl diazirine groups are covalently introduced to affinity matrices and transformed into highly reactive carbenes upon irradiation with UV light; these in turn bind to, or are irreversibly inserted into, a proximal small molecule in a manner that is independent of functional groups (**6**). Thus far we have developed two generations of photoactivatable affinity matrices (Fig. 1). The first one is called

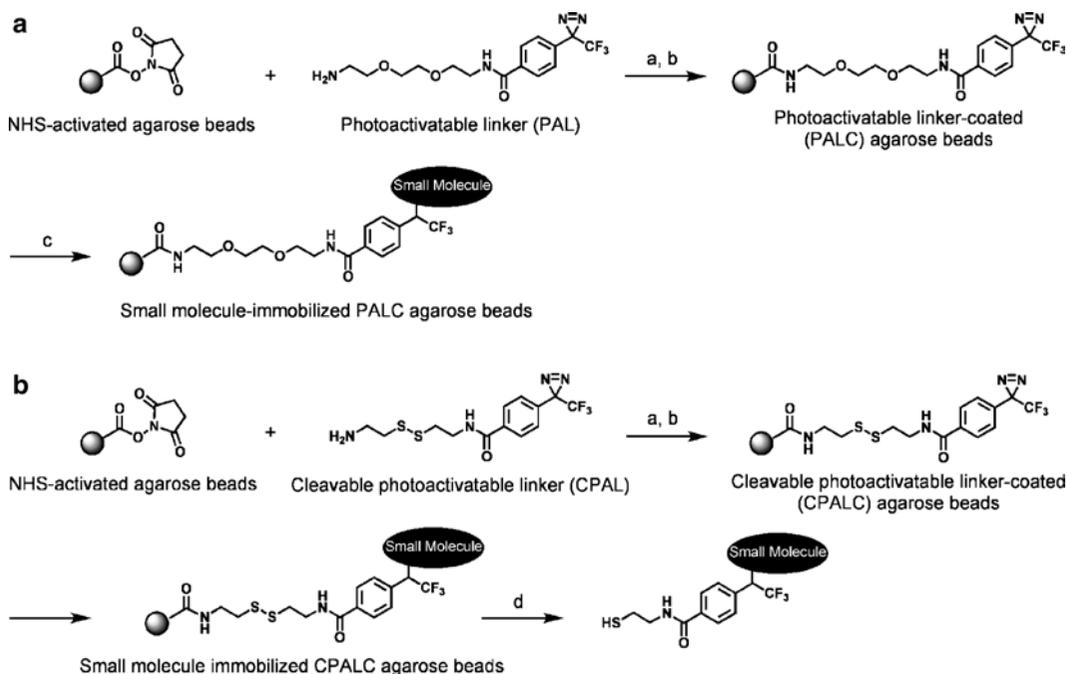


Fig. 1. Preparation of small molecule-immobilized PALC- or CPALC agarose beads (**a**, **b**) and detection of immobilized small molecules on CPALC agarose beads (**b**). Reagents and conditions: (**a**) DMF, rt, 2 h; (**b**) 1 mM ethanolamine, 100 mM Tris-HCl, pH 8.0, rt, 1 h; (**c**) small molecule, UV irradiation (365 nm, 4 J/cm²); (**d**) 100 mM dithiothreitol, EtOH, 50°C, 10 min.

the photoactivatable linker-coated (PALC) affinity matrix (5), in which an agarose bead and a diazirine group are connected through a noncleavable covalent linkage. The second generation is called the cleavable photoactivatable linker-coated (CPALC) affinity matrix (7), in which a disulfide bond is integrated between an agarose bead and a diazirine group. The selective cleavable nature of the disulfide bond not only makes it possible to verify the presence of the immobilized small molecule on the affinity matrices, but also permits the efficient detection of ligand-reacting protein targets covalently bound to the immobilized small molecule (8).

2. Materials

2.1. Preparation of Photoactivatable Linker-Coated Agarose Beads and Cleavable Photoactivatable Linker-Coated Agarose Beads

1. NHS-activated Sepharose 4 Fast Flow (GE Healthcare) (see Note 1).
2. Photoactivatable linker (synthesized in-house using the reported procedure) (see Note 2).
3. Cleavable photoactivatable linker (synthesized in-house using the reported procedure) (see Note 3).
4. 1 mM hydrochloric acid.
5. *N,N*-dimethyl formamide (reagent grade).
6. MeOH (reagent grade).
7. Blocking buffer: 1 mM ethanolamine, 100 mM Tris-HCl, pH 8.0.
8. Wash buffer: 100 mM Tris-HCl, pH 8.0.

2.2. Immobilization of Small Molecules on PALC- and CPALC Agarose Beads

1. Cyclosporin A (Wako Chemical) (see Note 4).
2. MeOH (reagent grade).
3. 2-propanol (reagent grade).
4. PBS buffer: 137 mM NaCl, 3 mM KCl, 8 mM Na₂HPO₄, 1.5 mM KH₂PO₄, pH 7.4.
5. Beads storage buffer: PBS buffer containing 1% sodium azide.

2.3. Cleavage of Immobilized Small Molecules from Small Molecule-Immobilized CPALC Agarose Beads and MS Analysis

1. Cleavage solution: 100 mM dithiothreitol in ethanol.

2.4. Equipment

1. Micro rotator RT-30mini (TAITEC).
2. Spin column (containing ca. 20 μm porous sintered polyethylene filter).

3. Rotary evaporator.
4. Oil-sealed rotary vacuum pump.
5. CL-1000 L ultraviolet cross-linker (UVP LLC, Upland, CA).
6. Dry heat block bath (FG-HB-1000, FastGene).
7. LC-MS workstation: Micromass ZQ 2000 LC mass spectrometer (Waters).

3. Methods

3.1. Preparation of Photoactivatable Linker-Coated Beads and Cleavable Photoactivatable Linker-Coated Beads

1. Add 150 μL of NHS-activated Sepharose 4 fast flow to a 1.5 mL Eppendorf tube (see Note 5).
2. Add 1 mL of 1 mM HCl to the tube.
3. Tap the tube to mix the ingredients, and centrifuge it at $2,000\times g$ for 30 s. Discard the supernatant. Repeat this washing cycle another four times immediately before use.
4. Wash the beads twice with DMF (1 mL each) as described in Subheading 3.1, step 3.
5. Add a solution of cleavable photoactivatable linker (6.3 mg, 17.3 μmol) or photoactivatable linker (6.3 mg, 17.5 μmol) in DMF (150 μL) to the tube.
6. Mix them using a micro rotator at RT for 2 h in the dark (see Note 6).
7. Centrifuge the resultant mixture at $5,000\times g$ for 3 min. Discard the coupling solution.
8. Wash the resultant beads three times with MeOH (1 mL each) as described in Subheading 3.1 step 3.
9. Add 1 mL of the blocking buffer to the resultant beads.
10. Mix them using a micro rotator at RT for 1 h in the dark.
11. Wash the beads with 1 mL of wash buffer as described in Subheading 3.1 step 3.
12. Place a spin column inside a 1.5-mL Eppendorf tube with the top cut off. Then transfer the beads into the spin column using a total of 1 mL of Milli-Q. Spindown as needed at $2,000\times g$ for 30 s to drain the water.
13. Add 400 μL of MeOH to the spin column and centrifuge it at $2,000\times g$ for 30 s. Repeat this procedure another two times, to ensure the beads are dry.
14. Transfer the beads to a 10-mL glass sample vial and use it immediately in the next step.

3.2. Photo-Cross-Linking of Cyclosporin A on PALC- and CPALC Beads

1. Add a solution of cyclosporin A (1.7 mg, 1.4 μmol) in MeOH (200 μL) (see Note 7) to the 10-mL clear glass vial containing the PALC- or CPALC agarose beads prepared in Sub-heading 3.1.
2. Concentrate the solution carefully using a rotary evaporator (see Notes 8 and 9).
3. Add 200 μL of 2-propanol to the vial. Rotavap the mixture to dryness (see Note 8), and vacuum dry it using an oil-sealed rotary vacuum pump for 1 h in the dark to ensure that the beads are completely dry (see Note 10).
4. Place the vial in a CL-1000 L ultraviolet cross-linker. Irradiate the mixture at 365 nm with 4 J/cm² (see Note 11).
5. Transfer the beads into a spin column placed in a 1.5-mL Eppendorf tube with the top cut off using water (total 5 mL) and MeOH (total 5 mL). Spindown (2,000 $\times g$) as needed and collect the filtrate (see Note 12).
6. Wash the beads successively with MeOH, DMSO, MeOH, and PBS (400 μL each).
7. Suspend the beads in 1 mL of PBS containing 1% sodium azide and store at 4°C (see Notes 13–15).

3.3. Cleavage of Immobilized Cyclosporin A from Cyclosporin A-Immobilized CPALC Agarose Beads and MS Analysis: Confirmation of Immobilization

1. Mix 10 μL of cyclosporin A-immobilized CPALC agarose beads with 200 μL of cleavage solution in a 1.5-mL Eppendorf tube (see Note 16).
2. Heat the mixture at 50°C for 10 min using a dry heat block bath.
3. Centrifuge the tube at 5,000 $\times g$ for 3 min, and collect the supernatant.
4. Perform analysis of the supernatant using LC-MS to detect cleaved products from cyclosporin A-immobilized CPALC agarose beads (see Fig. 2).

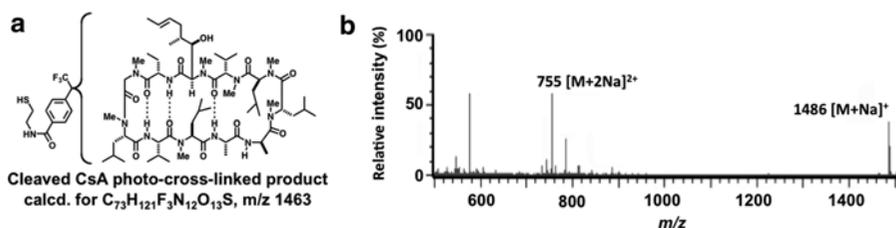


Fig. 2. Detection of cleaved photo-cross-linked products from cyclosporin A-immobilized CPALC agarose beads. (a) Expected cleaved photo-cross-linked products and their molecular weights. (b) ESI-MS spectrum of the cleaved sample. Sample injection volume: 20 μL ; Mobile phase: MeOH-H₂O (9:1, v/v) containing 0.1% formic acid; Flow rate: 1 mL/min; Ionization mode: ESI⁺; No separation column used.

4. Notes

1. According to the manufacturer's instructions, the density of the activated ester (*N*-hydroxysuccinimidyl ester) on the beads was 16–23 $\mu\text{mol/mL}$ of swollen beads.
2. The photoactivatable linker can be prepared in two steps from 4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzoic acid *N*-hydroxysuccinimide ester and *N*-(*tert*-butoxycarbonyl)-2,2'-(ethylenedioxy)diethylamine (Sigma-Aldrich) by using previously reported procedures (Fig. 3) (7). 4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzoic acid can be purchased from a commercial supplier (Tronto Research Chemicals, #T791240) or synthesized in-house according to the reported procedure (9). Synthesis should be carried out by trained chemists in facilities for chemical synthesis.
3. The cleavable photoactivatable linker can be prepared in two steps from 4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzoic acid *N*-hydroxysuccinimide ester and *N*-(*tert*-butoxycarbonyl)cystamine by using the previously reported procedures (Fig. 4). *N*-(*tert*-butoxycarbonyl)cystamine cannot currently be obtained from commercial suppliers, but can be synthesized in one step from cystamine dihydrochloride (Aldrich) by using the protocol reported in the literature (10). Synthesis should be carried out by trained chemists in facilities for chemical synthesis.
4. It should be noted that the small molecule to be immobilized must be stable under the irradiation conditions. We usually

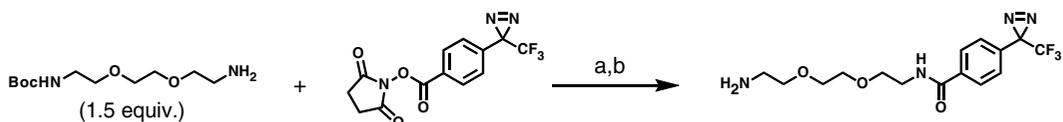


Fig. 3. Preparation of the photoactivatable linker (PAL). Reagents and conditions: (a) triethylamine, CH_3CN , rt, 1 h, 90%; (b) trifluoroacetic acid, CH_2Cl_2 , rt, 1 h, followed by purification via NH_2 -silica gel, 99%.

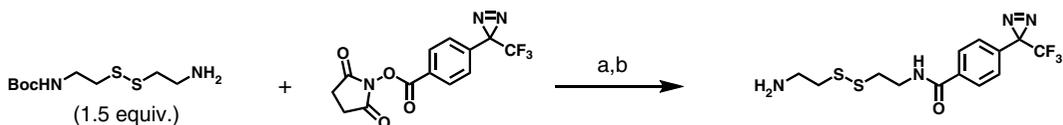


Fig. 4. Preparation of the cleavable photoactivatable linker (CPAL). Reagents and conditions: (a) triethylamine, CH_3CN , rt, 5.5 h, 94%; (b) trifluoroacetic acid, CH_2Cl_2 , rt, 3 h, followed by purification via NH_2 -silica gel, 87%.

check the stability by analyzing the HPLC profile or $^1\text{H-NMR}$ data of the sample irradiated under the same conditions.

5. Cut ca. 3 mm from the end of a 200 μL disposable pipette tip to create a wider bore and use it to dispense the beads. Draw a line at the 150 μL level on the Eppendorf tube before dispensing the beads, and then fill the tube with beads to the 150 μL line.
6. The amount of the linker introduced on the beads was estimated to be $\sim 0.92 \mu\text{mol/mL}$ of swollen beads using a spectroscopic method (11).
7. Theoretically, any volatile and neutral organic solvent can be used instead of MeOH.
8. The mixture bumps easily during the rotavaping. To avoid this, rotavap the mixture without using a warm water bath to allow the mixture to grow cold.
9. Use an adequate joint apparatus or adapter to match the joint on the rotary evaporator.
10. Release the reduced pressure inside the vial very carefully. Otherwise, the dried beads will be scattered.
11. Set the UV exposure at 1 J/cm^2 (maximum value) and repeat the irradiation four times.
12. The unimmobilized small molecule can be recovered from the filtrate.
13. The cyclosporin A-immobilized PALC- and CPALC agarose beads can be stored at 4°C for at least 6 months.
14. Prior to using the beads, wash them several times with the appropriate buffer to be used in the subsequent pulldown experiment.
15. The prepared cyclosporin A-immobilized PALC and CPALC agarose beads can equally be used to pulldown GST-tagged cyclophilin A (5) from the cell lysates of *E. coli* BL21(DE3) pLysS strain overexpressing GST-cyclophilin A (Fig. 5a). The former one can be used to pulldown endogeneous CpA from Jurkat cell lysates (Fig. 5b) (5). Other small molecule-immobilized PALC agarose beads have been used to discover or verify several ligand-protein interactions (12, 13).
16. The disulfide bond can also be cleaved with tris(carboethoxy) phosphine (TCEP) or immobilized TCEP disulfide reducing gel (Thermo Scientific). For example, it is estimated that more than 98% of the disulfide bonds are cleaved when the same amount of sample (10 μL) is treated with (a) 200 μL of 50 mM aq. NH_4HCO_3 containing 1 mM TCEP at 50°C for 10 min, or (b) 50 μL of immobilized TCEP disulfide reducing gel in 200 μL of 50 mM aq. NH_4HCO_3 .

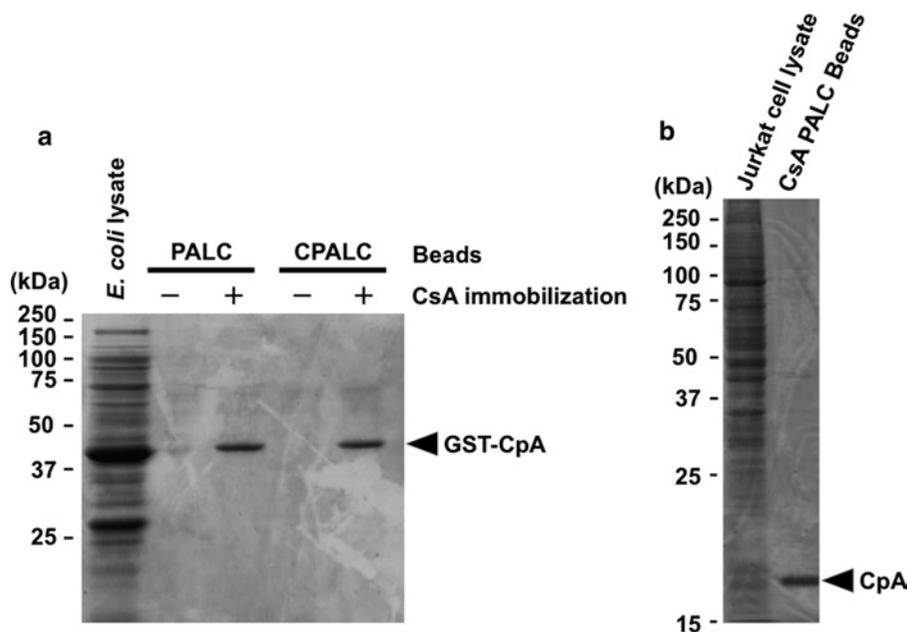


Fig. 5. Immunoprecipitation of GST-tagged cyclophilin A (GST-CpA) from GST-CpA overexpressing *E. coli* cell lysate (a) and endogenous cyclophilin A (CpA) from Jurkat cell lysate (b).

Acknowledgments

This work is supported by a Grant-in-Aid for Young Scientists (A) (No. 19681023; to NK), a Grant-in-Aid for Scientific Research (B) (No. 22310132; to NK), and a Grant-in-Aid for JAPS fellows (No. 2255053; to HT). We also thank Prof. Takayuki Doi and Dr. Masahito Yoshida for their assistance with the LC/MS analysis.

References

1. Leslie BJ, Hergenrother PJ (2008) Identification of the cellular targets of bioactive small organic molecules using affinity reagents. *Chem Soc Rev* 37: 1347–1360
2. Harding MW, Galat A, Uehling DE et al (1989) A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase. *Nature* 341: 758–760
3. Taunton J, Hassig CA, Schreiber SL (1996) A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* 272: 408–411
4. Kanoh N, Kumashiro S, Simizu S 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 Edit* 42: 5584–5587
5. Kanoh N, Honda K, Simizu S et al (2005) Photo-cross-linked small-molecule affinity matrix for facilitating forward and reverse chemical genetics. *Angew Chem Int Edit* 44: 3559–3562
6. Kanoh N, Nakamura T, Honda K et al (2008) Distribution of photo-cross-linked products from 3-aryl-3-trifluoromethyl-diazirines and alcohols. *Tetrahedron* 64: 5692–5698
7. Kanoh N, Takayama H, Honda K et al (2010) Cleavable linker for photo-cross-linked small-molecule affinity matrix. *Bioconjugate Chemistry* 21: 182–186

8. Drahl C, Cravatt BF, Sorensen EJ (2005) Protein-reactive natural products. *Angew Chem Int Edit* 44: 5788–5809
9. Hatanaka Y, Nakayama H, Kanaoka Y (1993) An improved synthesis of 4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoic acid for photoaffinity-labeling. *Heterocycles* 35: 997–1004
10. Girgenti E, Ricoux R, Mahy JP (2004) Design and synthesis of a Mn(III)-porphyrin steroid conjugate used as a new cleavable affinity label: on the road to semi-synthetic catalytic antibodies. *Tetrahedron* 60: 10049–10058
11. Gaur RK, Gupta KC (1989) A spectrophotometric method for the estimation of amino-groups on polymer supports. *Anal Biochem* 180: 253–258
12. Kawatani M, Okumura H, Honda K et al (2008) The identification of an osteoclastogenesis inhibitor through the inhibition of glyoxalase I. *P Natl Acad Sci USA* 105: 11691–11696
13. Miyazaki I, Simizu S, Okumura H et al (2010) A small-molecule inhibitor shows that pirin regulates migration of melanoma cells. *Nat Chem Biol* 6: 667–673

Profiling Cellular Myristoylation and Palmitoylation Using ω -Alkynyl Fatty Acids

Rami N. Hannoush

Abstract

Methods to detect and characterize cellular protein myristoylation and palmitoylation are invaluable in cell biology, immunology, and virology. Recently, we developed ω -alkynyl fatty acid probes for monitoring myristoylation and palmitoylation of cellular proteins. This article describes a biochemical procedure for metabolic labeling of cells with ω -alkynyl fatty acids and click chemistry.

Key words: Myristoylation, Palmitoylation, Fatty acylation, Lipids, Alkynyl fatty acids, Chemical probes, Metabolic labeling, Click chemistry

1. Introduction

Fatty acylation targets a wide range of cellular proteins which encompass kinases, GTPases, heterotrimeric G proteins, cytokines, and phosphatases (1–3). The role of fatty acylation is to regulate these proteins' physicochemical properties and spatial localization in cells. In doing so, fatty acylation controls the activation and deactivation of signaling pathways. Fatty acylation involves the enzyme-catalyzed addition of 14-carbon (myristoylation) or 16-carbon (palmitoylation) fatty acid chains to cellular proteins via amide- or thioester bonds, respectively.

Methods to monitor the status of protein myristoylation or palmitoylation are invaluable for studying cellular fatty acylation and its role in regulating protein behavior (4). Radioactivity has been the standard method for detecting myristoylation and palmitoylation of cellular proteins. Typically, (^3H)- or (^{125}I)-labeled myristic and palmitic acids are added to cells for metabolic incorporation into cellular proteins (5, 6). The signal is developed by

autoradiography and requires lengthy film exposures. Because of the hazards and costs associated with radioactivity, this method is less than ideal. Recently, the acyl-biotin exchange method was introduced for detecting protein palmitoylation (7). This method relies on blocking the free sulfhydryl groups of palmitoylated proteins followed by cleavage of the palmitate moiety and subsequent tagging of the unmasked cysteine with biotin for affinity capture and enrichment. This method is a great advance in the field, but it has limitations, primarily associated with false-positives (8), and it only detects S-palmitoylation but not myristoylation. For a detailed description of the various methods and their limitations, the reader is referred to a comprehensive review on this topic (4).

Recently, we developed a general method for the biochemical detection of myristoylated and palmitoylated proteins (9). The procedure utilizes fatty acid analogues modified at their termini with alkyne groups. We reasoned that appending an alkyne group to the terminal end of a fatty acid would not interfere with the hydrophobic nature of the fatty acid and the mechanism by which it inserts into lipid membranes. These probes are added to cultured cells and are incorporated into cellular proteins.

Proteins modified with these probes are selectively conjugated to biotin-azide or rhodamine-azide that reacts with the incorporated alkyne group via a Cu(I)-catalyzed (3+2) Huisgen cycloaddition reaction (or click reaction) (10–12). The labeled proteins are then separated by SDS-PAGE and detected by either western blotting or in-gel fluorescence (see Subheading 3 below). This method has several attractive features (4, 9). It is nonradioactive, highly sensitive, and applicable to a wide range of cellular systems. It also requires a short detection time. Furthermore, the alkynyl fatty acid reagents are portable, tunable in concentration, and conveniently stored in a freezer for immediate use. All of this makes the technique well suited for characterizing and profiling fatty acylation of cellular proteins of interest in various contexts.

To carry out the procedure outlined below, researchers first need to determine the cell type in which they are interested in studying fatty acylation. For detecting protein myristoylation, Alk-C13 or Alk-C14 probe should be used (Fig. 1). Alternatively, Alk-C16 and Alk-C18 are the appropriate probes for detecting protein palmitoylation (Fig. 1). The uptake of these probes may vary depending on the cell line used. In each case, the researcher needs to determine the optimal probe concentrations to be used and check the type of linkage via which the individual probe is incorporated into cellular proteins. This is achieved by measuring their sensitivity to hydroxylamine, a reagent which cleaves acyl chains that have been incorporated into proteins via thioester but not amide linkages. Treatment with hydroxylamine serves as a good indicator of the specificity of incorporation of the individual probe into cellular proteins (Fig. 2).

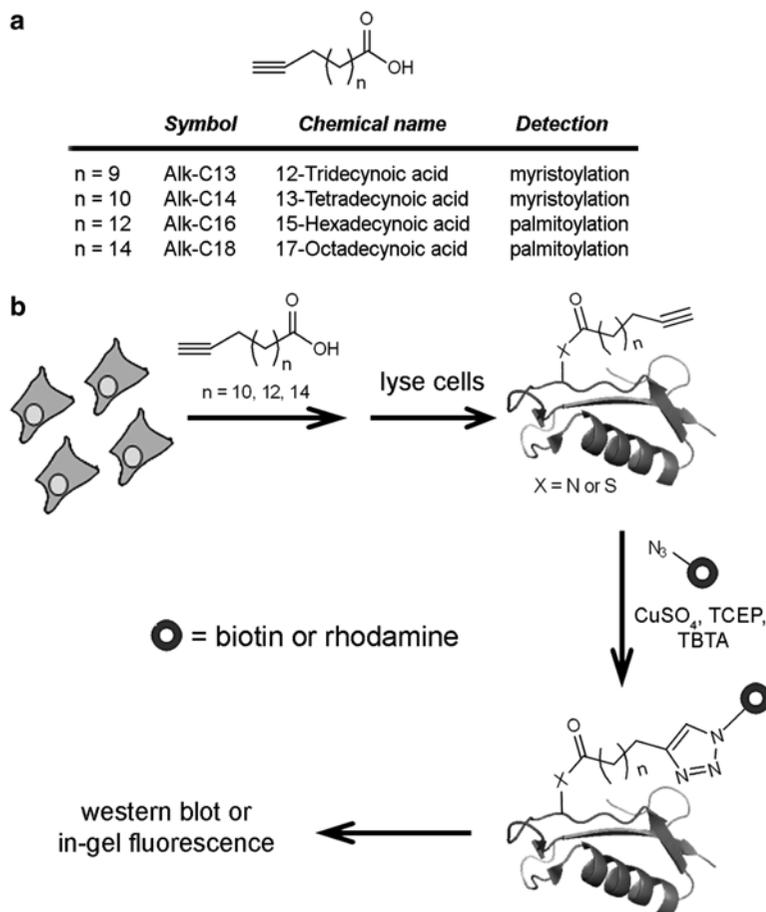


Fig. 1. Schematic of the method for monitoring myristoylated and palmitoylated cellular proteins. **(a)** ω -Alkynyl fatty acid probes for detecting protein myristoylation and palmitoylation. **(b)** ω -Alkynyl fatty acids added to growth media are metabolically incorporated into cellular fatty-acylated proteins. The cells are then lysed, Click chemistry is used to chemoselectively conjugate a biotin-azide or rhodamine-azide to fatty-acylated proteins, and the proteome is detected by western blotting or in-gel fluorescence. Alk-C13 and Alk-C14 are probes for detecting protein myristoylation, while Alk-C16 and Alk-C18 detect protein palmitoylation.

2. Materials

Use Milli-Q purified water or equivalent in all reagent preparations throughout this protocol.

2.1. Alkynyl Fatty Acid Stocks

1. Alk-C13, Alk-C14, Alk-C16, Alk-C18: 50 mM in DMSO. Aliquot solution and store at -80°C (see Note 1).

2.2. Cell Culture and Lysis

1. Tissue-culture compatible six-well plate (VWR).
2. Dulbecco's modified eagle's medium (DMEM) or any other suitable growth medium.

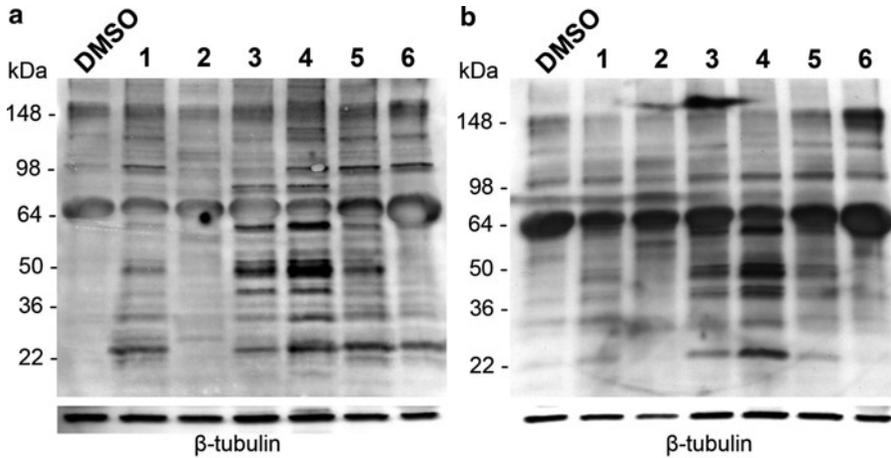


Fig. 2. Detection of myristoylated and palmitoylated proteins by metabolic labeling with ω -alkynyl fatty acids. Cellular extracts were prepared from MDCK cells treated with the different ω -alkynyl fatty acid probes (100 μ M) for 24 h. The proteome was subjected to click reaction with biotin-azide, separated by SDS-PAGE and detected by streptavidin-linked horseradish peroxidase. The blot in (b) represents an experiment done at the same time as in (a) except that it has been treated with hydroxylamine. Lanes: 1, Alk-C10 ($n=6$); 2, Alk-C11 ($n=7$); 3, Alk-C13 ($n=9$); 4, Alk-C14 ($n=10$); 5, Alk-C16 ($n=12$); 6, Alk-C18 ($n=14$). The DMSO lane serves as a background control for labeling with alkyne fatty acids.

3. Fetal bovine serum.
4. Glutamine or GlutaMax™ (Invitrogen).
5. Cells: Raw 264.7 macrophages (ATCC # CCL-2278); MDCK (ATCC # CCL-34); PC-3 cells (ATCC # CRL-1435); Mouse L-cells (ATCC # CRL-2648); HeLa cells (ATCC # CCL-2); Jurkat cells (ATCC # TIB-152); COS-7 cells (ATCC # CRL-1651). For culturing these cell types, see Note 2.
6. Fatty acid-free bovine serum albumin (BSA, Sigma-Aldrich).
7. Lysis buffer: 100 mM sodium phosphate, pH 7.5, 150 mM NaCl, 1% Nonidet P-40 (Store at 4°C up to 3 months). Add Pierce protease and phosphatase inhibitor cocktail just prior to the experiment (see Note 3).
8. Nanocep centrifugal ultrafiltration devices, MW 3 kDa (Pall Corporation).
9. BCA protein assay kit (Thermo Scientific).
10. PBS: 8 g NaCl, 0.2 g KCl, 1.13 g Na₂HPO₄ (sodium phosphate, dibasic anhydrous), 0.2 g KH₂PO₄ (potassium phosphate, monobasic), 1 L of H₂O. Adjust pH to 7.2 ± 0.1 with 6 N HCl.
11. Vortexer.
12. Thermomixer heating plate (Eppendorf).
13. Centrifuge.

2.3. Click Reaction Ingredients

1. Biotin-azide or rhodamine-azide: 5 mM in DMSO. Aliquot solution and store at -20°C . Protect from light (see Note 4).
2. Tris (2-carboxyethyl) phosphine hydrochloride (TCEP, Sigma-Aldrich): Dissolve powder in water to a concentration of 50 mM just prior to use in the click reaction. Protect from light.
3. Tris((1-benzyl-1H-1,2,3-triazol-4-yl)methyl)amine (TBTA, AnaSpec): Dissolve powder in DMSO/*t*-butanol (1:4, v:v) solution to a concentration of 10 mM. Aliquot solution and store at -20°C .
4. Copper sulfate: Dissolve powder in PBS to a concentration of 50 mM just prior to use in the click reaction.

2.4. Gel Electrophoresis and Immunoblotting

1. Novex[®] Tris Glycine SDS sample buffer (2 \times) (Invitrogen).
2. NuPAGE[®] sample reducing agent (10 \times) (Invitrogen).
3. Tris–Glycine Gels, Precast (Invitrogen).
4. PBS containing 0.1% Tween-20, v/v (PBST).
5. Non-fat dried milk (any grocery store).
6. Streptavidin-horseradish peroxidase (Zymed).
7. Enhanced Chemiluminescence Detection kit (ECL, Amersham Biosciences).
8. Restore[™] western stripping buffer (Thermo Scientific).
9. Anti- β -tubulin HRP antibody (Invitrogen).
10. Hydroxylamine, 50 wt% in water (NH_2OH , Sigma-Aldrich).
11. Amersham Hyperfilm ECL (GE Healthcare).
12. Desktop scanner.

3. Methods

3.1. Culturing and Seeding Cells

1. Grow your preferred cell type in its suitable growth medium in a 5% CO_2 humidified chamber at 37°C for at least 24 h (see Note 2).
2. Seed cells with growth medium onto six-well plates at a density of 8×10^5 cells/well.
3. Incubate for 24 h in a 5% CO_2 humidified chamber at 37°C before treating with fatty acid probes.

3.2. Preparation of Fatty Acid Probes

1. Dissolve probes in serum-free growth media supplemented with 5% BSA (fatty acid-free) at a final concentration of 100 μM just prior to the experiment. Include DMSO dissolved to 100 μM as a control (see Note 5).
2. Sonicate solutions for 15 min, and then incubate for 15 min at RT.

3.3. Treating Cells with Fatty Acid Probes

1. Remove growth media from seeded cells that have been growing for 24 h.
2. Wash cells once with PBS at RT.
3. Add 2 mL of fatty acid probe-containing medium and incubate for 24 h at 37°C/5% CO₂ (see Note 6).

3.4. Preparing Cell Lysates

1. Aspirate media and wash cells three times with cold PBS.
2. Add 400 µL lysis buffer and incubate at 4°C for 1 h with gentle shaking.
3. Centrifuge cell lysates at 16,000×*g* for 10 min at 4°C, and collect the supernatant.
4. Concentrate cell lysates by centrifugation for 15 min at 16,000×*g* at 4°C with Nanocep centrifugal ultrafiltration devices (see Note 7).
5. Measure protein concentration by BCA protein assay kit following manufacturer's protocol (see Note 8).

3.5. Labeling Cell Lysates with Biotin-Azide or Rhodamine-Azide

These volumes are based on a 25 µL reaction scale.

1. Add 23 µL of cell lysate at 2 mg/mL to a tube (see Note 9).
2. Add 0.5 µL of 5 mM biotin-N₃ to a final concentration of 0.1 mM.
3. Add 0.5 µL of 50 mM TCEP to a final concentration of 1 mM (see Note 10).
4. Add 0.5 µL of 10 mM TBTA to a final concentration of 0.2 mM (see Note 11).
5. Vortex for 5 s (high setting).
6. Add 0.5 µL of 50 mM CuSO₄ to a final concentration of 1 mM (see Note 12).
7. Vortex for 5 s (high setting).
8. Incubate for 1 h at RT in dark.

3.6. Precipitate Proteins and Prepare Samples for Gel Loading

1. Add 250 µL of ice-cold acetone.
2. Vortex and incubate for 4 h or overnight at -20°C (see Note 13).
3. Centrifuge at 16,000×*g* for 10 min at 4°C.
4. Gently aspirate the supernatant.
5. Air dry tube for 5–10 min.
6. Resuspend in 16 µL of lysis buffer.
7. Add 15.5 µL of loading buffer SDS (2×) and 3.5 µL of DTT (10×).
8. Heat sample for 5 min at 95°C.

9. Load 35 μL /lane on a ten-well gel (~30–40 μg protein/lane).

3.7. Electrophoretic Transfer and Immunoblotting

1. Resolve labeled protein lysates by SDS-PAGE using 4–20% Tris–glycine gels.
2. Transfer proteins onto a nitrocellulose membrane.
3. Block membrane with PBST containing 5% non-fat dried milk for 2 h at RT or overnight at 4°C.
4. Wash membrane three times with PBST (5 min each).
5. Incubate membrane with streptavidin-horseradish peroxidase (1:5,000 in PBST) for 1 h at RT.
6. Wash membrane with PBST three times (10 min each).
7. Develop signal using Enhanced Chemiluminescence Detection (ECL) according to manufacturer's recommendation (Amersham Biosciences). Use 1–2 min for film exposure time (see Note 14).
8. To demonstrate equal levels of protein loading, incubate streptavidin blots with Restore™ western stripping buffer for 15 min at RT.
9. Wash membrane three times with PBST (5 min each).
10. Reprobe blot with an anti- β -tubulin HRP antibody (1:10,000 in PBST) for 1 h at RT.
11. Develop signal with ECL as in step 7.

3.8. Check for Linkage-Specific Incorporation of the Probes

1. Incubate membranes from Subheading 3.7 step 2 for 65–72 h at RT with PBST and 5% NH_2OH (see Note 15).
2. Block membrane with 5% non-fat dried milk in PBST for 2 h at RT or overnight at 4°C.
3. Analyze by streptavidin blot as described in Subheading 3.7 above (steps 4–11).

3.9. Obtain Images in Film and Scanner

1. Scan films using a desktop scanner and import images into ImageJ or AdobePhotoshop.
2. The user can develop the signal by using rhodamine-azide followed by in-gel fluorescence detection. This yields minimal background and high signal-to-noise ratio compared to biotin-azide. Label cells and run gel, but replace biotin-azide with rhodamine-azide in Subheading 3.5 step 2.
3. Scan gel on a Typhoon fluorescence scanner using the Cy3/rhodamine fluorescence channel (see Note 16).
4. Export the image scan into the Typhoon image processing software (see Note 17).

4. Notes

1. Probes can be obtained from our laboratory or synthesized as described earlier (9). Alk-C13 and Alk-C18 can be obtained from Otava Ltd. and Sigma-Aldrich, respectively. The user should avoid multiple freeze–thaw cycles of the fatty acid probe stocks as this may compromise reagent quality. Dissolve all probes in a tissue culture hood (sterile environment).
2. Examples of cells lines in which this method has been validated are provided along with their growth media: Raw 264.7 macrophages (culture in high glucose DMEM supplemented with 10% FBS and 2 mM GlutaMax™); PC-3 cells (culture in F-12K Medium supplemented with 10% FBS); Jurkat (culture in RPMI supplemented with 10% FBS); MDCK, L-cells, HeLa and COS-7 (culture in DMEM supplemented with 10% FBS).
3. Adding fresh Pierce protease and phosphatase inhibitor cocktail just prior to the experiment would ensure efficacy of these inhibitors as reagent quality is compromised over time when these inhibitors are suspended in lysis buffer.
4. Probes can be obtained from our laboratory, synthesized as described earlier (9, 13–15), or purchased from commercial vendors such as Invitrogen.
5. This step should be carried out in a tissue culture hood to keep the growth medium sterile.
6. Because the optimal concentration of ω -alkynyl fatty acid needed to obtain high signal-to-noise ratio may vary depending on the particular cell type used, the user should first test the individual probe at various concentrations (1–100 μ M) and for different incubation times to determine the optimal concentration and time to be used. All treatments should be performed in a tissue culture hood to keep the media sterile.
7. This step will usually yield ~250 μ L at a final protein concentration of ~2 mg/ml. Sometimes depending on cell type and growth conditions, this concentration is not achieved and hence longer centrifugation times may be needed to achieve a protein concentration of ~1–2 mg/ml.
8. At this point, cell lysates can be snap-frozen and stored at -80°C to be thawed at a later time for subsequent steps below.
9. The total protein concentration can be ~1–2 mg/ml. The user should use 40–50 μ g of total protein per reaction.
10. TCEP is a reducing agent and should be freshly prepared just prior to use in the reaction.

11. TBTA is a ligand that stabilizes copper (I), protects it from oxidation and disproportionation and enhances its catalytic activity (12, 16).
12. Copper sulfate acts as a catalyst in this reaction. The Click reaction works best if copper sulfate has been freshly prepared just prior to the experiment. The order of addition of the reagents to the protein extracts is important for the reaction and has to be followed as described in Subheading 3.5 steps 1–6.
13. We find that 4 h incubation works for recovering the bulk of the cellular proteins. The user can also incubate overnight for convenience.
14. Sometimes, detection by streptavidin–HRP results in varying degrees of background staining which is due to streptavidin labeling of endogenously biotinylated proteins. Therefore, care should be taken to decipher nonspecifically labeled bands on the gel, and it is critical that the user include negative controls such as those lacking copper sulfate or the fatty acid probes. Another source of loss in signal is proteins that may interfere with the click reaction such as those that may chelate copper. These proteins may exist in a particular cellular system, and steps need to be taken to eliminate them if feasible. It is also possible that staining is not homogeneous throughout the membrane blot, and this could be due to uneven transfer of proteins from the gel. Furthermore, the intensity of bands on the membrane can sometimes be too high, and this can be resolved by reducing exposure times or using a more diluted stock of streptavidin–HRP.
15. This step should be done with membranes that were not previously stripped or developed with ECL. As an alternative to hydroxylamine, the user can use 0.1 N KOH in methanol for 1 h at RT.
16. Sometimes excess free rhodamine-azide may be present and it runs at the bottom of the gel. We find it useful to run gels until the rhodamine azide band elutes from the gel; this minimizes saturation of the signal when scanning the gel.
17. Once the cells are seeded in plates and all the necessary reagents are at hand, the procedure described above could be completed within 2–3 days. By combining this protocol with standard immunoprecipitation protocols, it enables researchers to analyze the myristoylation or palmitoylation status of their proteins of interest in cells that have been metabolically labeled with Alk-C13, Alk-C14, Alk-C16, or Alk-C18.

References

1. Resh MD (2006) Trafficking and signaling by fatty-acylated and prenylated proteins. *Nat Chem Biol* 2: 584–590
2. Linder ME, Deschenes RJ (2007) Palmitoylation: policing protein stability and traffic. *Nat Rev Mol Cell Biol* 8: 74–84
3. Lu JY, Hofmann SL (2006) Thematic review series: lipid posttranslational modifications. Lysosomal metabolism of lipid-modified proteins. *J Lipid Res* 47: 1352–1357
4. Hannoush RN, Sun J (2010) The chemical toolbox for monitoring protein fatty acylation and prenylation. *Nat Chem Biol* 6: 498–506
5. Schlesinger MJ, Magee AI, Schmidt MFG (1980) Fatty acid acylation of proteins in cultured cells. *J Biol Chem* 255: 10021–10024
6. Peseckis SM, Deichaite I, Resh MD (1993) Iodinated fatty acids as probes for myristate processing and function. Incorporation into pp60v-src. *J Biol Chem* 268: 5107–5114
7. Drisdell RC, Green WN (2004) Labeling and quantifying sites of protein palmitoylation. *Biotechniques* 36: 276–285
8. Roth AF, Wan J, Green WN, Yates JR, Davis NG (2006) Proteomic identification of palmitoylated proteins. *Methods* 40: 135–142
9. Hannoush RN, Arenas-Ramirez N (2009) Imaging the lipidome: omega-alkynyl fatty acids for detection and cellular visualization of lipid-modified proteins. *ACS Chem Biol* 4: 581–587
10. Rostovtsev VV, Green LG, Fokin VV, Sharpless KB (2002) A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angew Chem Int Ed Engl* 41: 2596–2599
11. Tornøe CW, Christensen C, Meldal M (2002) Peptidotriazoles on solid phase: (1,2,3)-triazoles by regioselective copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J Org Chem* 67: 3057–3064
12. Wang Q, Chan TR, Hilgraf R, Fokin VV, Sharpless KB, Finn MG (2003) Bioconjugation by copper(I)-catalyzed azide-alkyne (3+2) cycloaddition. *J Am Chem Soc* 125: 3192–3193
13. Speers AE, Cravatt BF (2004) Profiling Enzyme Activities In Vivo Using Click Chemistry Methods. *Chem Biol* 11: 535–546
14. Hsu TL, Hanson SR, Kishikawa K, Wang SK, Sawa M, Wong, CH (2007) Alkynyl sugar analogs for the labeling and visualization of glycoconjugates in cells. *Proc Natl Acad Sci USA* 104: 2614–2619
15. Lewis WG, Magallon FG, Fokin VV, Finn, MG (2004) Discovery and Characterization of Catalysts for Azide–Alkyne Cycloaddition by Fluorescence Quenching. *J Am Chem Soc* 126: 9152–9153
16. Chan TR, Hilgraf R, Sharpless KB, Fokin VV (2004) Polytriazoles as Copper(I)-Stabilizing Ligands in Catalysis. *Org Lett* 6: 2853–2855

Fluorescence Labels in Kinases: A High-Throughput Kinase Binding Assay for the Identification of DFG-Out Binding Ligands

Jeffrey R. Simard and Daniel Rauh

Abstract

Despite the hundreds of kinase inhibitors currently in discovery and pre-clinical phases, the number of kinase inhibitors which have been approved and are on the market remains low by comparison. This discrepancy reflects the challenges which accompany the development of kinase inhibitors which are relatively specific and less toxic. Targeting protein kinases with ATP-competitive inhibitors has been the classical approach to inhibiting kinase activity, but the highly conserved nature of the ATP-binding site contributes to poor inhibitor selectivity, issues which have particularly hampered the development of novel kinase inhibitors. We developed a high-throughput screening technology that can discriminate for inhibitors which stabilize the inactive “DFG-out” kinase conformation by binding within an allosteric pocket adjacent to the ATP-binding site. Here, we describe how to use this approach to measure the K_d of ligands, as well as how to kinetically characterize the binding and dissociation of ligands to the kinase. We also describe how this technology can be used to rapidly screen small molecule libraries at high throughput.

Key words: FLiK, Fluorescence, Acrylodan, DFG-out, DFG-in, Kinase inhibitors, Activation loop, High-throughput screening

1. Introduction

The highly conserved ATP-binding site of kinases is a major roadblock in the research and development of kinase inhibitors, and often significantly decreases the chance that a drug candidate will reach the market due to poor selectivity. Although inhibition of a group of kinases may have undesirable therapeutic consequences, evidence is gathering which suggests that polypharmacology may actually be the key to fighting complex diseases such as cancer (1). However, polypharmacology strategies should be developed around the concept of “controlled unselectivity,” or the targeting of a

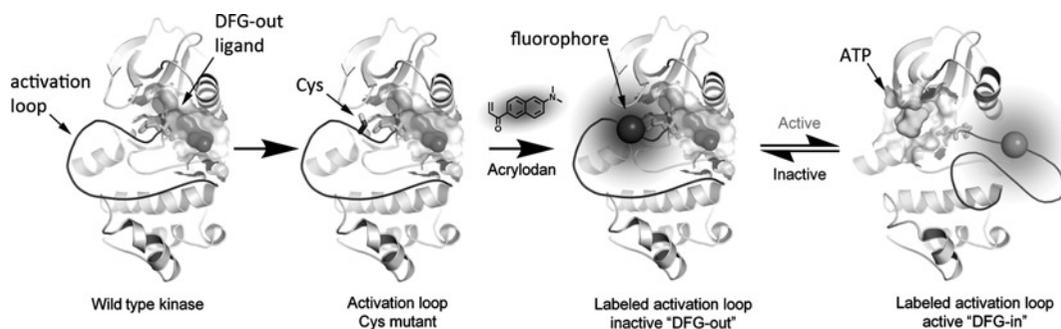


Fig. 1. The FLiK technology for the detection of allosteric Type II/III kinase inhibitors. Kinases are regulated by an activation loop which can adopt active and inactive conformations. (a) The inactive conformation (DFG-out) presents an alternative binding site in which DFG-out ligands can bind and prevent the kinase from adopting an active conformation. (b) A Cys was mutated into the activation loop and (c) used for the attachment of acrylodan. (d) The DFG motif and activation loop adopt a different conformation in inactive kinases when compared to active kinase with ATP bound, resulting in a fluorescence change. Active kinases have an activation loop which is open and extended in contrast to the inactive DFG-out conformation. Reproduced from (7) with permission from Nature Publishing Group.

desired group of kinases within one of more specific pathways in order to achieve the desired therapeutic outcome. Therefore, new strategies are needed for targeting kinases that move away from the classic ATP-competitive (Type I) inhibitors. Emerging data suggest that the issue of selectivity can be addressed by targeting the DFG-out pocket, also frequently referred to as an allosteric pocket. The DFG-out pocket is adjacent to the ATP-binding site and tends to be much less conserved with respect to amino acid sequence (2). Interestingly, the availability of this allosteric pocket requires the kinase to adopt a catalytically deficient inactive conformation in which the activation loop is in a closed conformation that disrupts the binding of ATP and substrates to the kinase (3–5).

This inactive conformation is attained by a flip of the kinase DFG motif to the DFG-out conformation in which the DFG-Phe side chain flips into the ATP-binding site, thereby inducing the rearrangement of the activation loop and opening the DFG-out pocket. This conformational equilibrium between DFG-out (inactive) and DFG-in (active) can be modulated *in vitro* or *in vivo* by the binding of various ligands, protein–protein interactions and the phosphorylation or dephosphorylation of regulatory regions of the kinase domain such as the activation loop (5). The search for Type II and Type III inhibitors that bind to and stabilize the inactive kinase conformations and the specific identification of chemical scaffolds which have affinity for the DFG-out pocket is moving to the forefront of kinase inhibitor research. However, efforts have been constrained by the lack of high-throughput assay technologies which can discriminate for ligands which stabilize the inactive conformation.

We have developed a widely applicable assay system which achieves this goal by directly tethering an environmentally sensitive fluorophore to the kinase in order to monitor the conformational changes of the activation loop upon the binding of these desired compounds (Fig. 1) (6, 7).

This approach requires the removal of solvent-exposed cysteines and the insertion of a cysteine into a desired position in the activation loop of the kinase to serve as the attachment point for the thiol-reactive fluorophore acrylodan. The kinase mutant is then expressed and purified, labeled and characterized using biochemical and biophysical methods. In a high-throughput screening scenario, in which a small library of compounds is rapidly screened at a single concentration, induced movement of the activation loop by DFG-out binding ligands results in a change in the local environment of acrylodan, causing distinct changes to the dual-emission spectrum of acrylodan. By using the ratio of emission intensities of each emission maxima, the response of acrylodan to increasing concentrations of inhibitor provides a straightforward binding assay methodology which allows direct determination of the K_d of the ligand. The assay also allows follow-up characterization of identified compounds in a manner similar to the traditional approach of using SPR (surface-plasmon resonance) by permitting the determination of k_{on} , k_{off} to better understand the contributing factors to the observed affinity ($K_d = k_{off}/k_{on}$). Thus, the Fluorescence Labels in Kinases (FLiK) technology provides a straightforward versatile means for both identifying and characterizing such compounds.

Classic ATP-competitive (Type I) inhibitors normally do not induce this type of conformational change, making them insensitive to detection when using the labeled activation loop approach described above. However, we have previously reported the detection of a few Type I inhibitors which can stabilize the DFG-out conformation of the kinase by directly interacting with residues of the activation loop rather than by directly occupying the allosteric pocket made available by this conformation (8). We have since developed other labeling strategies, including labeling of the P-loop (glycine-rich loop) to identify more Type I ligands without relying on more traditional activity-based assays which require prior knowledge of the kinase substrate (9), but these strategies are beyond the scope of this chapter.

2. Materials

2.1. Labeling Proteins with Acrylodan

1. Falcon Tubes (15 mL).
2. Aluminum foil.
3. 10 kD-MWCO Amicon Ultra Centrifugal Filter Unit with Ultracel membrane (Millipore).
4. Stock of desired protein to be labeled, stored in appropriate storage buffer at -20 to -80°C (see Note 1).
5. Dimethyl sulfoxide (DMSO).

6. 6-acryloyl-2-dimethylaminonaphthalene (acrylodan) stock (Invitrogen): 10 mM stock dissolved in DMF or DMSO (see Note 2).
7. Labeling Buffer: 50 mM Hepes, 200 mM NaCl, pH 7.0–7.5 (see Note 3).
8. FLiK Buffer: 50 mM Hepes 200 mM NaCl, pH 7.5 (see Note 4).

2.2. Fluorescence Measurements-Initial Assay Tests and K_d Determination

1. Labeled kinase (200 μ M stock, stored in desired storage buffer at -20 to -80°C).
2. FLiK Buffer.
3. Inhibitor stock solutions (10, 1, 0.1, 0.01 mM dissolved in DMSO) (see Note 5).
4. Polystyrene single-use cuvettes (see Note 6).
5. Mini cylindrical stir bars (8×3 mm).
6. Fluorescence spectrometer (JASCO FP-6500 – JASCO GmbH, Gross-Umstadt, Germany) (see Note 7).
7. Standard curve fitting software.

2.3. High-Throughput Screening

1. Labeled kinase (200 μ M stock, stored at -20°C or -80°C in desired protein storage buffer).
2. FLiK Buffer (supplemented with 0.002–0.01% v/v detergent) (see Note 8).
3. Inhibitor stock solutions (10, 1, 0.1, 0.01 mM dissolved in DMSO).
4. 384-well plates (black, small 20 μ L volume, flat bottom; black, large 120 μ L volume, flat bottom – Greiner Bio-One, Kremsmünster, Austria) (see Note 9).
5. Adhesive ThermoWell Sealing Tape for 384-well plates (aluminum, 100/stack, nonsterile – Corning).
6. Microplate Reader (Tecan Infinite).
7. For manual HTS screening tests: 8 or 16 multichannel pipettors and complementary pipetting tips.
8. For manual HTS screening tests: Dispensing troughs (Eppendorf Tip-Tub for multichannel pipettors).
9. For automated HTS screening: Automated liquid handling robot equipped with a 384-well pin tool for dispensing small volumes of compound into microtiter plates.
10. For automated HTS screening: Multidrop Combi equipped with a small tube metal tip dispensing cassette for 0.5–50 μ L volumes of kinase solution (Thermo Scientific).

3. Methods

A wide selection of fluorescence probes is commercially available which are either thiol or amine reactive and can be used for labeling proteins (<http://www.invitrogen.com>). When choosing a fluorophore for use in this approach, special consideration should be given to molecules which are highly sensitive to changes in the local microenvironment. These fluorophores will be the most sensitive detectors of changes in protein conformation and are commonly used to generate fluorescent protein conjugates, or biosensors (10). In particular, acrylodan responds with a change in overall fluorescence intensity as well as a ratiometric emission change at two emission wavelengths, thus providing an internal correction for dilution errors across several samples. For this reason, the methods described in this section will focus exclusively on the acrylodan fluorophore as an example. Further details regarding the fluorescence properties of acrylodan in the described assay are provided elsewhere (6).

Given the lower abundance of cysteine in proteins, the use of thiol-reactive fluorophores such as acrylodan allows more straightforward generation of biosensors with minimal labeling at nonspecific binding sites. To further increase labeling efficiency to a single desired site, a thorough structural analysis of tertiary protein structure should first be performed to identify additional solvent-exposed cysteines on the protein surface. This can be done using published crystal structures of the protein (Protein Data Bank; <http://www.pdb.org>) or by generating homology models (Swiss Modeller; <http://swissmodel.expasy.org>) of the protein using a published crystal structure of a highly homologous protein as a modeling template. If possible, structures or models of the kinase in both the active (DFG-in) and inactive (DFG-out) conformation should be analyzed. Further details about the selection of labeling sites are not discussed in this chapter, but further information is provided elsewhere (6, 9). Following mutation of the desired protein, it is expressed and purified according to established protocols. The final buffer of the purification should not contain dithiothreitol (DTT) or any free thiols due to likely interference with the labeling procedure (see Note 10). The labeling procedure is described below in detail. Subsequent analysis of the extent and specificity of labeling should be performed using mass spectroscopy.

3.1. Labeling Proteins with Acrylodan

1. For automated HTS screening: Automated liquid handling robot equipped with a 384-well pin tool for dispensing small volumes of compound into microtiter plates.

2. Measure the molar protein concentration of the protein stock using traditional methods.
3. Using the molecular weight (MW) and volume of the protein sample, calculate the number of moles of protein to be labeled with acrylodan.
4. Using the MW of acrylodan (MW = 225 g/mol), calculate the volume of dissolved acrylodan stock needed for a 1.5:1–2:1 mole:mole ratio of acrylodan:protein (see Note 11).
5. In a 15 mL Falcon tube, dilute the protein in Labeling Buffer such that the % v/v of DMSO/DMF stays <0.5% upon subsequent addition of acrylodan to the protein (see Note 12).
6. Wrap tube in aluminum foil, place on ice in a cold room or refrigerator (4–8°C) and allow labeling to occur overnight (see Note 13).
7. Following labeling, place the acrylodan-labeled protein into a pre-washed 10 k-MWCO centricon and centrifuge at 4°C and 3,000–3,500×g to concentrate the protein.
8. Re-suspend in the desired protein storage buffer. Wash protein by concentrating sample at least two additional times to exchange the buffer (see Note 14).
9. Concentrate the final washed sample to obtain a concentration of 100–200 μM if possible (see Note 15).

3.2. Assay Characterization

The following tests should be performed for each newly labeled kinase in order to characterize the fluorescence response induced by ligand binding. Type I and Type II (or III) inhibitors should be available as negative and positive controls. Various parameters which assess the quality of fluorescent protein biosensors have been described by de Lorimier et al. (10) and can be determined by making the measurements described in the following sections.

3.2.1. Measuring Interval Scan Emission Spectra

1. Prepare 3 mL of acrylodan-labeled kinase solution (50–100 nM) using FLiK Buffer and place into a 4.5 mL polystyrene cuvette with a mini stir bar (see Note 16).
2. Place cuvette into the sample holder of the fluorescence spectrometer and set the excitation wavelength and scan fluorescence emission over a wide range according to the fluorophore used (see Note 17).
3. Program the fluorescence spectrometer (or control manually) to measure emission spectra at regular time intervals (see Note 18).
4. Positive Control: Turn on the stir bar, inject 0.1–1 μM of a known Type II inhibitor (or any compound which is known to stabilize or bind to the inactive DFG-out kinase conformation). Immediately start the interval scan to obtain a series of changing emission spectra in response to the binding of the chosen ligand (see Note 19).

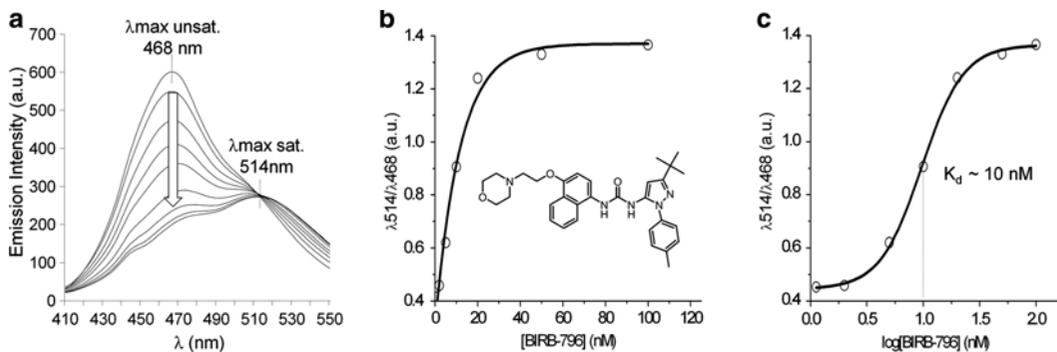


Fig. 2. K_d determination for a DFG-out binding ligand. This figure presents a typical set of endpoint measurement data obtained from the described K_d determination experiments. Shown data were obtained using acrylodan-labeled p38 α and the Type II inhibitor BIRB-796. (a) The ratiometric fluorescence values are obtained from the emission spectra obtained in the presence of increasing inhibitor concentration. (b) These data can be plotted to observe the saturation of the kinase in a particular conformation or (c) these data can be used to directly obtain the K_d for the ligand. Panels have been reprinted with permission from (6). Copyright 2008 American Chemical Society.

5. Negative Control: Repeat for a Type I ATP-competitive inhibitor which is known to bind to the active DFG-in conformation and not induce movement of the DFG motif and/or activation loop.
6. For each inhibitor type, overlay the spectra series using graphing software to visualize the changes in emission spectra that occur with ligand binding over time (Fig. 2a); identify the emission maxima of the unbound and bound states of the protein (see Note 20).
7. Select two wavelengths which are representative of each emission maximum of acrylodan and determine the ratiometric output of these two values using spectra obtained for the ligand-free labeled protein and the fully bound protein. These values will give an indication of the lowest and highest possible values, or “assay window” (see Note 21).
8. Calculate fluorescence parameters (ΔR_{\max} and ΔI_{std}) for the kinase and also Z' -factor for the assay (see Note 22).

3.2.2. Endpoint Measurements and K_d Determination

K_d values are typically obtained from endpoint measurements in which ratiometric values of emission intensities at $\lambda_{\text{max}1}$ and $\lambda_{\text{max}2}$ are monitored and plotted as a function of inhibitor concentration. These experiments can be carried out using two methods, which each have their own advantages. In the first method, a single cuvette containing the labeled protein is rapidly stirred and increasing concentrations of Type II/III inhibitor are added to a single sample while spectra are measured at each step. The advantage of this approach is that very little protein is required to obtain a single K_d curve. The disadvantage is that this approach is less accurate when adding sequential doses of a very slow binding inhibitor and if not enough time is given for each dose to reach equilibrium with

the kinase. In the second method, a series of cuvettes containing the same concentration of labeled protein are prepared and each receives a single dose of inhibitor at a specific concentration. Following a long incubation time, each cuvette is measured to examine changes in the emission spectrum associated with inhibitor binding. The advantage to this approach is that a K_d curve can be obtained at various time points and does not require constant rigorous stirring of the sample over significant time periods. This method is especially advantageous when analyzing very slow binding Type II/III inhibitors (11). The major disadvantage is that the requirement of several cuvettes requires significantly more protein. The most favorable method should be considered based on knowledge of protein availability, required binding times for DFG-out inhibitors, and stability of the kinase in buffer over time.

3.2.3. Sequential Addition of Inhibitor to a Single Sample

1. Prepare a cuvette containing the labeled kinase and place into the fluorescence spectrometer as described in steps 1–2 of Subheading 3.2.1.
2. Measure the emission spectra for the labeled protein before adding a compound which will induce the desired conformational change. Calculate and record the ratio of the two chosen emission intensities at $\lambda_{\max 1}$ and $\lambda_{\max 2}$.
3. Turn on the stir bar and inject increasing amounts of compound to the stirring suspension using the inhibitor stocks described in Subheading 2.2, item 3 (see Note 23).
4. Allow sufficient time for the binding to reach completion (see Note 24).
5. After each addition and stirring of the suspension, measure the emission spectrum. Calculate and record the ratiometric value as in step 2.
6. Plot the data using desired graphing software. A plot of inhibitor concentration (i.e., [inhibitor]) as a function of emission ratio will generate a saturable binding curve (Fig. 2b). A plot of \log [inhibitor] as a function of emission ratio values will generate a binding curve (Fig. 2c).
7. Using a sigmoidal fitting function fit the binding curve to determine the K_d from the midpoint of the curve. Repeat until at least three stable values are obtained.

3.2.4. Addition of Inhibitor to a Series of Samples

1. As an alternative method, prepare a series of cuvettes (approximately 8) containing the same concentration of labeled kinase and place as described in step 1 of Subheading 3.2.1.
2. Place the cuvettes back into their covered container (or keep dark and undisturbed). Starting at low concentrations, add the appropriate amount of inhibitor to each cuvette in the series using the prepared stocks (see Note 25). Each cuvette should receive a single dose of compound.

3. Place the box containing the prepared cuvettes on a stir-plate on the highest setting, stir for 30 s to disperse the inhibitor throughout the sample, store covered and dark until binding equilibrium is reached (see Note 26).
4. Measure the emission spectrum of each cuvette in the series and calculate the ratiometric values as in step 2 of Subheading 3.2.3. Repeat these emission scans at varying time points (see Note 27).
5. Plot the data as described in step 6 of Subheading 3.2.3 until three reproducible K_d values are obtained.

3.2.5. *Real-Time Kinetic Measurements and Determination of k_{on} and k_{off}*

Using this assay system, kinetic values are typically obtained by monitoring changes in fluorescence over time, usually at a single wavelength. The association rate constant, k_{on} , is the fastest measurable binding rate for a ligand. As the concentration of ligand is increased, the observed binding rate (k_{obs}) increases until k_{on} is achieved. This is easily observed using a plot of k_{obs} values as a function of inhibitor concentration. At low inhibitor concentrations, k_{obs} and inhibitor concentration increase linearly. A linear fit of the data will result in a line with slope equal to k_{on} . Concentrations tested should be within the linear range of this plot. At high concentrations, the plotted data level off at a maximum value for k_{obs} . At these high inhibitor concentrations, the maximum k_{obs} value is equal to the k_{on} for the ligand. Similar plots and experiments can be used to experimentally determine the maximum k_{off} value for a ligand by adding increasing amounts of unlabeled kinase to scavenge the inhibitor by shifting equilibrium away from the labeled kinase to which the inhibitor is already bound.

Kinetic rate constants such as k_{on} and k_{off} provide information and insight into the affinity of compounds since they are related to the K_d ($K_d = k_{off}/k_{on}$). Typically, these parameters are determined using various other techniques and instruments such as SPR.

Determination of k_{on}

1. Prepare several cuvettes as described in step 1 of Subheading 3.2.1 using 50 nM labeled kinase (see Note 28).
2. Place cuvette into the sample holder of the fluorescence spectrometer, set the excitation wavelength and monitor the most sensitive emission wavelength for the labeled kinase while rapidly stirring the sample (see Note 29).
3. Measure baseline fluorescence for 30 s; be sure the baseline is relatively flat (see Note 30).
4. Using a 1–10 μ L pipettor (see Note 31), add the inhibitor to the sample using the injection port above the cuvette while measuring fluorescence. Repeat for at least a total of 4 doses (see Note 32).

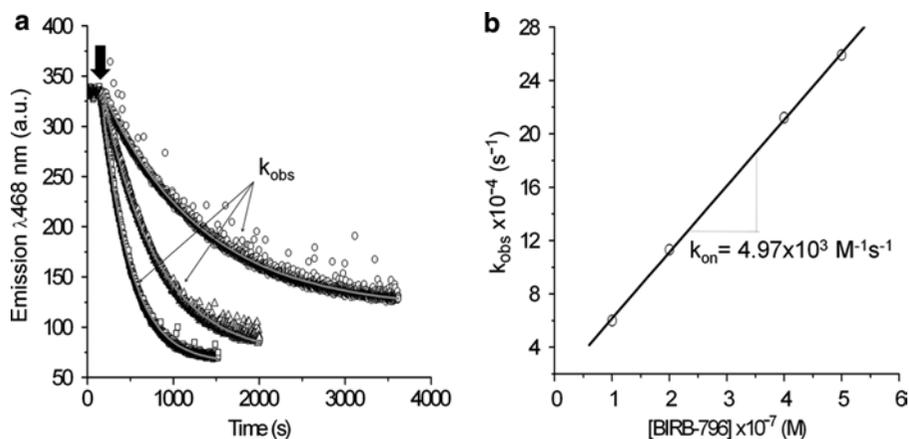


Fig. 3. Determination of k_{on} for a DFG-out binding ligand. This figure presents a typical set of kinetic binding measurements for determining the k_{on} of BIRB-796 using acrylodan-labeled p38 α . (a) The fluorescence emission is monitored at a single wavelength upon addition of increasing concentrations of inhibitor. (b) These values can be plotted as a function of inhibitor concentration to obtain a linear curve with slope = k_{on} . Panels have been reprinted with permission from (6). Copyright 2008 American Chemical Society.

5. Measure the change in emission over time; allow the fluorescence change to proceed until it linearizes.
6. Plot the data using desired graphing software. A plot of fluorescence intensity as a function of time will generate a kinetic curve (Fig. 3a). Fit the curve using a first order fitting function to determine the observed rate constant (k_{obs}) for the dose of added inhibitor (see Note 33). Repeat until at least three stable k_{obs} values are obtained for each dose.
7. Plot k_{obs} (units of s^{-1}) as a function of inhibitor concentration (units of M), fit the data using a linear fitting function (Fig. 3b). The slope of the resulting line is the true k_{on} (units of $\text{M}^{-1} \text{s}^{-1}$) of the ligand (see Note 34).

Determination of k_{off}

8. Prepare several cuvettes as described in step 1 of Sub-heading 3.2.1 using 50 nM labeled kinase.
9. Pre-incubate with an inhibitor known to bind to the DFG-out conformation of the kinase (see Note 35).
10. Monitor fluorescence emission over time as described above in step 3.
11. As described in steps 4–5, use a 1–10 μL pipettor to add a concentrated stock of unlabeled kinase to the sample while measuring the change in emission over time until it becomes linear. Repeat for at least a total of 4 doses of added unlabeled protein.
12. Plot the data using desired graphing software. A plot of fluorescence intensity as a function of time will generate a kinetic curve (Fig. 4). Fit the curve as described above in step 7 to

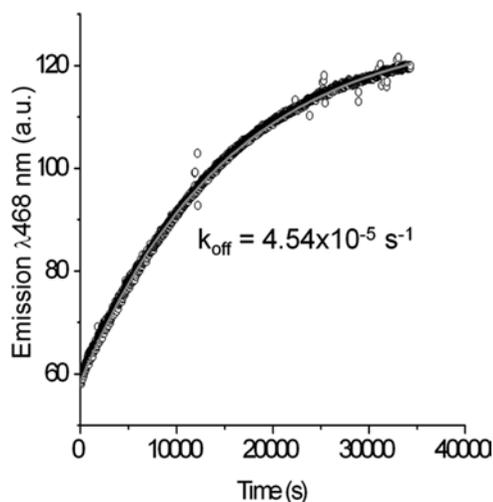


Fig. 4. Determination of k_{off} for a DFG-out binding ligand. This figure presents a typical kinetic measurement for determining the k_{off} of BIRB-796 using acrylodan-labeled p38 α . The fluorescence emission is monitored at a single wavelength upon addition of an excess of unlabeled kinase to scavenge the pre-bound ligand from the acrylodan-labeled kinase. Panel has been reprinted with permission from (6). Copyright 2008 American Chemical Society.

determine the k_{off} for the inhibitor (see Note 36). Repeat until at least three stable k_{off} values are obtained.

3.2.6. High-Throughput Screening

The FLiK assay technology can easily be adapted to several HTS formats and has been successfully employed in 96, 384, and 1,536-well microtiter plates. The protocols described in the following sections refer specifically to screening with the 384-well format and provide detailed pipetting schemes for the preparation of HTS plates either using 8 or 16-channel multichannel pipettors or liquid handling robots equipped with a 384-well pin tool. For the best data quality and reproducibility, liquid handling robots are highly preferable, if such resources are available.

In order to adapt the assay to HTS formats, the first step is to utilize the assay parameters determined using the methods described in Subheading 3.2.2 and apply them to microtiter plates. Following this initial test, various other parameters which could affect assay quality should also be investigated in order to optimize assay performance. These include %v/v DMSO, liquid volume in each well, the type and %v/v of detergent, incubation time and incubation temperature.

After the assay has been optimized, compound screening can be performed by first screening a large set of compounds at a single concentration and by analyzing the data to identify ligands which induce a significant response from the labeled kinase by stabilizing to the DFG-out conformation. The time-dependency of binding can be easily assessed by measuring plates at increasing time intervals. Type II and Type III inhibitors are known to be

slow binders to the DFG-out conformation (11), while we have previously shown that some Type I inhibitors can also stabilize the DFG-out conformation by interacting directly with amino acid side chains in the activation loop (6, 8). However, these ligands exhibit a fast binding rate. Thus, time-dependent measurements of screening plates will help to discriminate the slow binders from the fast binders. Selected binders should then be further tested in a second round of screening in which the compound is tested over a wide range of concentrations. This is done in order to determine if a binding curve can be obtained for each compound and is an easy method for identifying “false hits” which may have been detected due to auto-fluorescence.

3.2.7. Adaptation of Assay to HTS: Assay Optimization

1. Prepare the desired volume of FLiK Buffer in a Falcon tube and add 0.002–0.01% v/v Triton-X100 to the buffer (see Note 37).
2. In an empty small volume (20 μ L), black, 384-well HTS plate, prepare a dilution series of a positive control inhibitor in DMSO for subsequent use in screening plates; in a next step, pre-dilute the compound series in buffer if required (see Note 38).
3. Using a multichannel pipettor, transfer the desired amount of inhibitor to two empty 384-well HTS plates (one plate will be used for background subtraction of compound auto-fluorescence). In each plate, pipette the compound into 3–4 wells in order to obtain several replicate data points for each concentration of inhibitor tested (see Note 39).
4. Prepare a solution of 100 nM labeled kinase using the buffer prepared in step 1 and place the solution in a dispensing trough.
5. Using a multichannel pipettor, add the solution of labeled kinase to the wells of one of the HTS plates containing compound (see Note 40). Next, add the buffer (without kinase) to the other HTS plate containing compound (see Note 41).
6. The above protocol can be altered for automated HTS screens in which a liquid handling robot is used to transfer nanoliter volumes of inhibitor to screening plates (see Note 42).
7. Lightly tap the plates on the bench to gently settle the liquid in all wells.
8. Place the plate on an HTS plate shaker for 30 s to enhance mixing of kinase and compound in all wells.
9. Seal plates using aluminum adhesive seals for 384-well plates.
10. Incubate for the desired time period and temperature (see Note 43).
11. After incubation, remove the seal from each plate and place the plate inside a microtiter plate reader. For both sets of plates,

scan the fluorescence emission intensity at both desired wavelengths (see Note 44).

12. Process the acquired data by first subtracting the background fluorescence signal of the inhibitor from that obtained in the plate containing both inhibitor and kinase. Determine the ratiometric fluorescence using the corrected values for each wavelength.
13. Plot the data as described in step 5 of Subheading 3.2.3 to obtain a binding curve and a K_d value. Using data points at the bottom and top of the binding curve (i.e., no inhibitor and saturated with inhibitor), calculate the Z' -factor for the assay as described in step 8 of Subheading 3.2.1 (see Note 45).

3.2.8. Primary Screening of Compound Libraries

1. Map out the planned arrangement of a 384-well screening plate with the location of the compounds which will be added to the screening plate. Be sure to include several wells for both negative (DMSO) and positive (DFG-out inhibitor) controls (see Note 46). An example layout for a primary screen is shown in Fig. 5a (see Note 47).
2. Prepare the desired volume of FLiK Buffer in a Falcon tube and add 0.002–0.01% v/v Triton-X100 to the buffer. Using the prepared FLiK Buffer, prepare a solution of 100 nM labeled kinase (see Note 48).
3. For manually pipetted screens, transfer all compounds (prepared in 100% v/v DMSO) from highly concentrated stock plates to an empty small volume (20 μ L), black, 384-well HTS plate. According to the mapped plate scheme, place the positive (inhibitor) and negative (DMSO) controls into the desired wells of the same plate.
4. To ensure the concentration of DMSO and inhibitor will be correct in the final screening plate, pre-dilute the inhibitors using either DMSO or FLiK buffer in a similar manner to that described in Note 38.
5. Using a multichannel pipettor, transfer the desired amount of inhibitor to two empty 384-well HTS plates as described in step 3 of Subheading 3.2.7.
6. Using a multichannel pipettor, add FLiK Buffer or the prepared kinase solution each to one of the HTS plates already containing inhibitor as described in step 5 of Subheading 3.2.7.
7. Prepare and store the plates as described in steps 7–10 of Subheading 3.2.7.
8. After incubation, remove the seal from each plate and place the plate inside a microtiter plate reader to measure the fluorescence emission intensity at both desired wavelengths.
9. Process the acquired data as outlined in step 11 of Subheading 3.2.7 to determine the corrected ratiometric fluorescence.

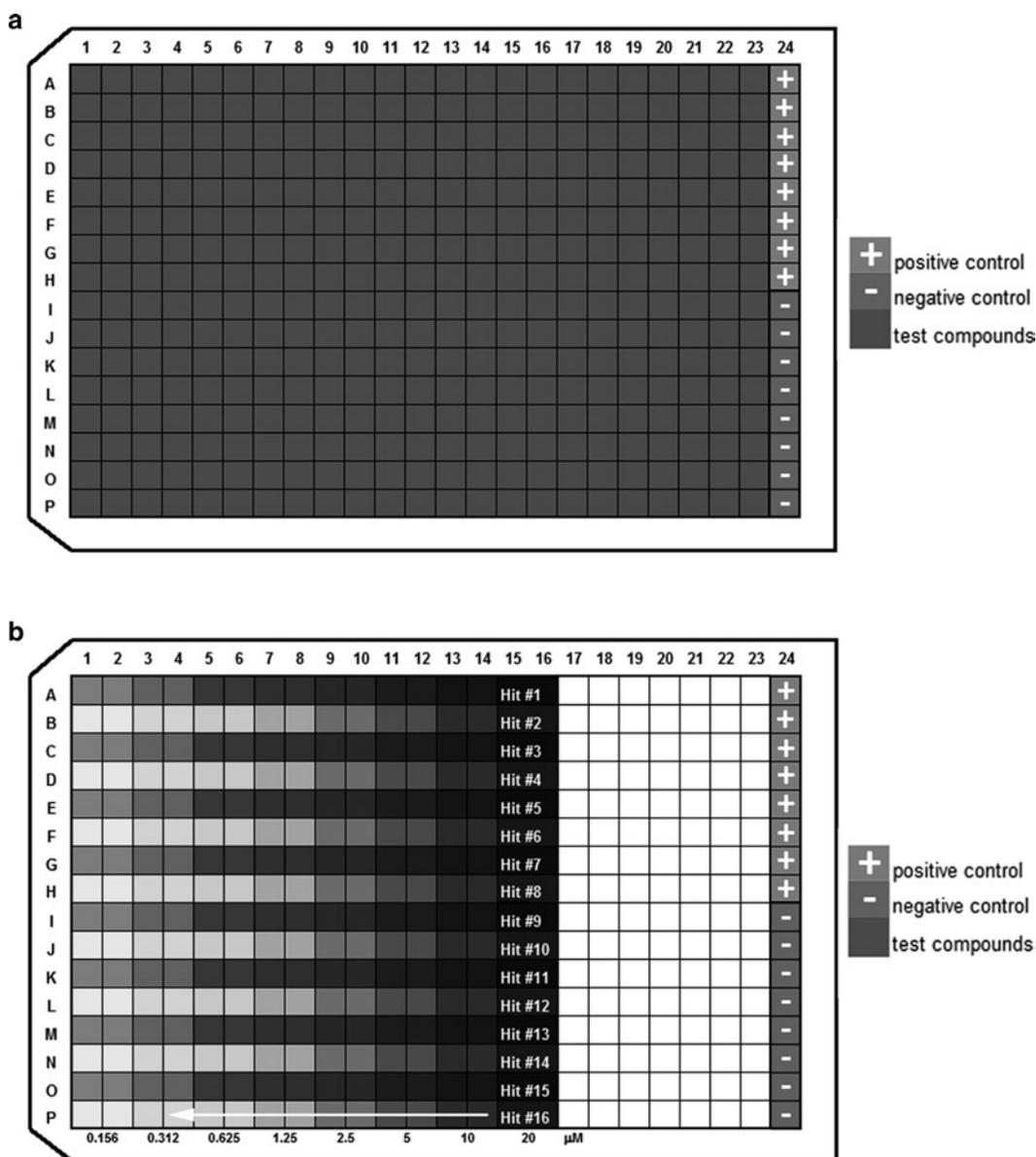


Fig. 5. Example layouts for HTS plates. (a) Primary screening plates typically consist of several wells each containing a single compound at a single concentration. These plates also contain several wells in which the positive and negative controls are placed. These controls are required to determine the % binding of hits relative to the positive control (100%). These values are calculated from the ratiometric emission change induced by ligand binding. (b) In secondary screens, chosen hit compounds are prepared and tested in a dilution series. In this example, compounds are screened in a concentration range of 20 μM to 156 nM by carrying out eight twofold dilutions. In the screening plate, each of the eight data points are pipetted side by side in duplicate.

3.2.9. Data Analysis

1. After processing the data as described in step 11 of Subheading 3.2.7, further process the data by using the ratiometric fluorescence values to determine the % binding at the concentration screened (see Note 49). The % binding is a measure of the proportion of kinase in the assay which is occupied by the ligand and locked in the DFG-out conformation.
2. Using the calculated values for % binding, a cutoff of 50% of the maximal fluorescence response (compared to 100% for the positive control) is used to select “hits” for follow-up testing (see Note 50).
3. Determine the ratiometric values and % binding for all compounds without including the subtraction of background fluorescence at λ_{max1} and λ_{max2} .
4. Compare % binding values for both the corrected and uncorrected sets of processed data as a quality control to help identify and remove likely “false hits” (see Note 51).
5. Select and follow-up with compounds which appear to be binding in both corrected and uncorrected datasets. Such compounds have the highest likelihood of being non-fluorescent binders.

3.2.10. Secondary Screening of Compound Libraries

1. Map out the planned arrangement of a 384-well screening plate with the location of a dilution series for each tested inhibitor as well as the location of several wells containing the negative (DMSO) or positive (DFG-out inhibitor) controls (see Note 46). An example layout for a secondary screen is shown in Fig. 5b (see Note 52).
2. Pre-dilution of compounds should be carried out as described in steps 2–3 of Subheading 3.2.7. In the first step, the dilution series of each compound should be prepared in DMSO prior to further pre-dilution in FLiK Buffer (for lower DMSO concentrations) or prior to transfer directly to the screening plate (for higher DMSO concentrations) (see Note 53).
3. Prepare the desired volume of FLiK Buffer in a Falcon tube and add 0.002–0.01% v/v Triton-X100 to the buffer (see Note 48). Using the prepared FLiK Buffer, prepare a solution of 100 nM labeled kinase.
4. As described in steps 3–11 of Subheading 3.2.7, dispense the pre-diluted dilution series of each compound into two sets of plates and add FLiK buffer or kinase solution into the different plates, seal the plates and incubate for an appropriate length of time before measuring each plate at two wavelengths in a microtiter plate reader.
5. Process the data into % binding and analyze as described in Subheading 3.2.9.

4. Notes

1. Concentrations of 5–10 mg/mL are preferred since this reduces the required volume of Labeling Buffer needed to dilute the sample for labeling. Dilution is required to minimize the % v/v of DMF or DMSO to $\leq 0.5\%$ when adding the appropriate amount of acrylodan stock to the protein.
2. Keep protected from light.
3. Up to 20% v/v glycerol can be added and pH can be adjusted to 7.5 if necessary. It is important not to carry out the labeling near pH 8.0 in order to avoid nonspecific labeling of lysines.
4. Depending on the target protein, up to 20% v/v glycerol can be added without effecting stability of the protein or acrylodan fluorescence; pH and NaCl concentration can be adjusted if desired. Use of different salts is also possible but the best results were obtained usually with this buffer; changes in ionic strength, salt concentration, and pH are known to effect the emission spectrum and sensitivity of acrylodan to conformational changes (12).
5. Positive inhibitor controls should be obtained in order to characterize the fluorescence response associated with the conformational switch from DFG-in to DFG-out. Common sources to identify such compounds are the Protein Data Bank or published articles which contain inhibition profiles for DFG-out binding inhibitors against a panel of kinases.
6. Quartz cuvettes can also be used without a problem. For plastic cuvettes, polystyrene seems to have optimal performance due to limited nonspecific binding of proteins and inhibitors to the surface.
7. Assay performance is not associated with this specific instrument. Any instrument should be sufficient for measuring spectral changes occurring in response to inhibitor binding. In order to perform kinetic measurements of ligand binding and dissociation, the instrument sample holder should be equipped with a magnetic stirring plate to allow rapid mixing of compounds with the protein suspension. Furthermore, if the dependency of kinetics on temperature is of interest, the instrument should be equipped with an externally regulated water bath and/or internal heating/cooling plate beneath the cuvette holder.
8. Several detergents, including Brij-35, Tween20 or Triton X-100, have been tested and did not negatively impact assay performance. Optimal buffer composition (salt concentration, ionic strength, pH, additives to stabilize proteins, detergents) and its effects on the fluorophore used to label the protein should be determined for each new labeled kinase.

9. Plates from other vendors have also been used successfully. Plates should be chosen for optimal performance in accordance with pipetting procedures, samples mixing, and the chosen instrumentation used to measure microtiter plates. However, chosen plates should be colored black for the best assay performance.
10. The reducing agent TCEP has been reported to be compatible with acrylodan labeling.
11. The protein should ideally have solvent-exposed Cys residues mutated away, leaving only the inserted Cys as the only solvent-exposed anchorpoint for the fluorophore. Under mild conditions, labeling of this Cys by the thiol-reactive fluorophore acrylodan is typically complete and specific.
12. Addition of acrylodan should be done slowly while gently stirring or mixing to avoid protein denaturation due to local high concentrations of DMSO and acrylodan in the absence of stirring; it should be noted that when 50 mM stocks of acrylodan are used, acrylodan often precipitates when added quickly to the buffered solution, resulting in poor labeling efficiency. A gentler approach is to first dissolve the appropriate amount of acrylodan and then add the concentrated protein to the buffer containing pre-dispersed acrylodan.
13. Labeling can also be carried out over shorter periods of time to minimize the chances of nonspecific labeling of proteins. The progress of the reaction can also be monitored over time by exciting acrylodan at 386 nm and monitoring the increase in intensity between 460–510 nm. Acrylodan becomes highly fluorescent upon Michael addition to Cys sulfhydryl groups of proteins. The labeling procedure should be optimized for each protein, taking into account both the length of time as well as temperature.
14. For large quantities of protein, unreacted acrylodan can also be removed by passing the protein over a size exclusion column. It should be noted that small traces of free acrylodan do not have a significant effect on assay performance. However, free acrylodan is reactive and this can have unknown consequences when compounds are introduced during screening.
15. Addition of a single acrylodan molecule increases the protein MW by 225 Da and increases the extinction coefficient by $\sim 6,000 \text{ M}^{-1} \text{ cm}^{-1}$.
16. Similar measurements may also be performed in microtiter plates if the proper instrumentation is available for reading the plates.
17. For acrylodan: excitation $\sim 385\text{--}390 \text{ nm}$; emission spectral range measured $\sim 420\text{--}550 \text{ nm}$.
18. Time intervals and compound concentration may have to be adjusted depending on the ligand and kinase used. Generally,

lower concentrations of added ligand will result in a slower response while high concentrations of ligand bind more quickly.

19. Adjust the concentration of compound and/or time interval according to the magnitude and rate of change observed in the spectra over time. Repeat until an emission spectrum with the maximum signal change is obtained.
20. Acrylodan has a primary emission maximum at 465–480 nm ($\lambda_{\max 1}$) and a secondary maximum at 505–520 nm ($\lambda_{\max 2}$). There may also be a shoulder peak visible at ~445 nm which is contributed by the FLiK Buffer. In most cases, a ratio (R) of intensities measured at $\lambda_{\max 1}$ and $\lambda_{\max 2}$ are used to discriminate between DFG-out and DFG-in binders.
21. Check several different wavelength combinations to find the ratio which gives the largest assay window and good reproducibility.
22. The Z' -factor is a statistical measure of the quality of an assay and is used to predict if the assay is useful for high-throughput screening. The Z' -factor is calculated using the equation:

$$Z \text{ factor} = 1 - \frac{3 \times (\sigma_p + \sigma_n)}{|\mu_p - \mu_n|} \quad (1)$$

where both the mean (μ) and standard deviation (σ) of both the positive (p) and negative (n) controls ($\mu_p, \sigma_p, \mu_n, \sigma_n$, respectively) are taken into account.

ΔI_{std} is the ratio of normalized intensity change compared to the average intensity of the fluorescence emission. It is one of the most important criteria for characterizing a fluorescent protein conjugate as being suitable for sensitive fluorescence spectroscopic detection of conformational changes (10). Ideally, the ΔI_{std} should have a value >0.25 and is calculated using the equation:

$$\Delta I_{\text{std}} = \left| \frac{2(I_1(\lambda_{\text{std}}) - I_2(\lambda_{\text{std}}))}{I_1(\lambda_{\text{std}}) + I_2(\lambda_{\text{std}})} \right| \quad (2)$$

where $\lambda_{\text{std}} = 1/2(\lambda_{\max, \text{unbound}} + \lambda_{\max, \text{saturated}})$ and I_1, I_2 are the fluorescence intensities at λ_{std} of each emission spectrum, respectively (i.e., if the maximum of the unbound protein is 468 nm and shifts to 514 nm in the bound state, then $\lambda_{\text{std}} = 491$ nm).

ΔR_{\max} is the maximum standard intensity change of the fluorescence emission between saturated and unsaturated kinase. It is another important criteria for characterizing a fluorescent protein conjugate as suitable for sensitive fluorescence spectroscopic detection of conformational changes (10). Ideally, the ΔR_{\max} should have a value >1.25 (although lower values are also acceptable) and is calculated using the equation:

$$\Delta R = \left| \frac{{}^0A_1}{{}^0A_1} - \frac{{}^\infty A_1}{{}^\infty A_1} \right|, \quad (3)$$

where 0A_1 , 0A_2 are two different areas in the absence of ligand and ${}^\infty A_1$, ${}^\infty A_2$ are the same areas in the presence of saturating amounts of ligand. An alternative to using the area under the curve is to measure the emission intensity at two single wavelengths to calculate ΔR_{\max} .

23. Over the titration, the %v/v of added DMSO should be kept to a minimum ($\leq 1\%$ v/v) to avoid destabilization or denaturation of the unlabeled protein as inhibitor is added in sequential doses to the cuvette. Therefore, choose the inhibitor stock concentrations that allow the sequential addition of the 0.3–1.5 μL volumes using a 0.1–10 μL pipettor (i.e., addition of 0.3, 0.6, and 1.5 μL from a 10 mM inhibitor stock will give final concentrations of 1, 2, and 5 μM , respectively, without introducing much DMSO to the cuvette).
24. This mixing time is crucial to accurate K_d determination. In the case of ligands which bind to the DFG-out conformation of a kinase, there is a time-dependence in the measured K_d since many of these ligands have a characteristically slow rate of binding (11). To estimate the time needed for a stirring sample to reach equilibrium, refer to the experiments described in Subheading 3.2.5.
25. Choose the inhibitor stock concentrations that allow the single addition of the 1.5–6 μL volumes using a 1–10 μL pipettor (i.e., addition of 1.5, 3, and 6 μL from a 10 mM inhibitor stock will give final concentrations of 5, 10, and 20 μM , respectively, without introducing $>0.2\%$ v/v of DMSO to each cuvette).
26. This incubation time will vary for each labeled kinase and must be determined; see also Note 24.
27. Measuring the series over a period of time will provide valuable information about the time-dependence of the K_d value, which is expected for slow-binding DFG-out kinase inhibitors. A slow-binding DFG-out binding inhibitor will typically produce K_d curves which shift leftward over time. Additionally, these measurements allow an easy assessment of the effect that incubation time has on the assay window. When the assay window starts to decrease, this is a sign that the protein may be denaturing after long incubation times.
28. Be sure buffer is at desired temperature since kinetics of binding and dissociation can be affected by temperature.
29. In the case of acrylodan, either $\lambda_{\max 1}$ or $\lambda_{\max 2}$ can be used. The wavelength which undergoes the largest intensity change upon ligand binding is the best choice.

30. Downward linear drift of the fluorescence signal is sometimes observed depending on the kinase. We have linked this drift to the stirring rather than to bleaching of the fluorophore by the excitation lamp. The magnitude of this drift over time is very minimal compared to magnitude of signal change induced by compound binding.
31. For the best results, make sure the pipette tips are long enough to reach far into the cuvette and ensure inhibitor is rapidly dispersed while mixing.
32. For compounds with nM – low μM affinity, a standard protocol of 4 different concentrations (spaced in equal increments between 50 and 500 nM) are used to determine k_{on} .
33. If graphing software fit provides the result as a half-time of fluorescence decay ($t_{1/2}$), the following equation should be applied to determine k_{obs} : $k_{\text{obs}} = \ln 2 / t_{1/2}$. The units of $t_{1/2}$ should be in seconds.
34. The k_{off} for the ligand can also be derived from the y -intercept of this plot.
35. Add the compound at a concentration \geq to the K_d to ensure a large enough signal change upon ligand dissociation.
36. If a nM affinity inhibitor is used in these experiments, addition of a tenfold excess of unlabeled kinase relative to labeled kinase and inhibitor will result in the fastest possible dissociation rate, or the true k_{off} of the ligand to be obtained from a first order fit of the resulting trace. If only μM affinity ligands are available, large amounts of unlabeled protein are necessary to recreate the same conditions. Therefore, in such cases, a method similar to that described in Subheading 3.2.5 for determining k_{on} should be employed. First, add increasing amounts of unlabeled kinase and determine the k_{obs} for each dose. Finally, plot the data and the slope of the linear fit should provide the value for k_{off} .
37. While concentrations of detergent in this range result in high reproducibility and superior assay results, the lowest possible concentrations of detergent should be used. In some cases, we found that detergent concentrations $\leq 0.1\%$ v/v induce a rightward shift in the binding curve of tested ligands, resulting in higher K_d values.
38. To facilitate easy transfer of compounds to screening plates, prepare the pre-dilutions of 8–10 different concentrations of compound in a separate 384-well plate using 100% v/v DMSO. The concentration of this dilution series should be 20-fold higher than the final screening concentration. If an additional pre-dilution in FLiK Buffer is required to reduce the % v/v DMSO, then the dilution series should first be prepared

in DMSO at 200-fold higher concentration than the final screening plate and then diluted 1:100 in buffer.

39. For 5% DMSO, simply transfer 1 μL of the inhibitor dilution series in DMSO to an empty small volume 384-well HTS plate. For 0.5% DMSO, first transfer 1 μL of the inhibitor dilution series in DMSO to a large volume (120 μL volume) 384-well HTS plate and then add 99 μL of FLiK Buffer. Mix thoroughly up and down at least twice. Next, transfer 10 μL of this pre-diluted compound in FLiK Buffer to an empty small volume 384-well HTS plate.
40. For screens with 5% DMSO, transfer 19 μL of kinase solution to plates already containing 1 μL of inhibitor in DMSO. For 0.5%, transfer 10 μL of 2 \times kinase solution (200 nM) to plates already containing 10 μL of inhibitor in buffer.
41. This plate will serve as the background subtraction for compound fluorescence. It is also possible to change the pipetting procedure in a way that allows compound to be pipetted into a single plate, background fluorescence is measured and then the labeled kinase solution is added, followed by a second measurement of the fluorescence. Background correction under these circumstances will also need to factor in a dilution factor as a result of the addition of kinase solution in the final pipetting step.
42. In this scenario, the compounds can be pipetted into wells containing buffer prior to measuring the background fluorescence. Kinase solution is subsequently added to the same plate, followed by a final fluorescence measurement after adequate incubation time.
43. Several factors come into play in this step and must be optimized for each kinase tested. Since this assay is designed to detect the binding of inhibitors to the inactive DFG-out conformation of kinases, a sufficient incubation time is necessary since these inhibitors bind slowly and the determined K_d value is time-dependent. Additionally, the stability of the kinase must accommodate the required incubation time under the conditions of the assay. The most straightforward way to test this is to perform the experiments described in Subheading 3.2.7 using two sets of plates. Follow the same protocol, but incubate the sealed plates at room temperature and also at 4–10°C. This procedure should be carried out in the presence of different DMSO concentrations, which can affect kinase stability over time.
44. If possible, measure the emission spectrum of the kinase in the presence and absence of the positive control inhibitor to verify that the desired maxima have not shifted in the HTS format. For some kinases, we have observed as much as a 10 nm shift

in one or both of the maxima when adapting the assay to HTS formats.

45. To assess protein stability and K_d time-dependence, re-measure the same plates over a period of time and re-plot the data. For slow-binding inhibitors of the DFG-out kinase conformation, a time-dependent leftward shift of the binding curve is often observed. Additionally, re-calculate the Z' -factor over time. A sudden decrease in assay window or data point reproducibility often indicates that the protein is no longer stable.
46. Take into consideration that positive and negative controls should be in several wells on each screening plate to permit accurate calculation of the Z' -factor.
47. A typical primary screen uses 1 well per compound (10 μM) and 100 nM labeled kinase. The kinase concentration should not be reduced to 50–75 nM unless screening inhibitors at < 10 μM . This amount of kinase and compound helps to significantly reduce the number of false auto-fluorescence hits.
48. Prepare enough of each solution according to the number of compounds and HTS plates which will be screened, taking into consideration both the number of wells and the volume in each well.
49. The % binding to the kinase is a way to quantify the percent of acrylodan-labeled kinase in the well which has adopted the DFG-out conformation in response to the binding of a compound and provides a straightforward comparison to the positive control inhibitor. The % binding is calculated using the equation:

$$\% \text{ Binding} = \left(\frac{R_{\text{hit}} - R_{\text{neg}}}{R_{\text{pos}} - R_{\text{neg}}} \right) \times 100 \quad (4)$$

50. Higher or lower cutoff values can also be used. However, using a cutoff <40% is not recommended due to the increased probability of retaining “false hits” which may have been identified in the primary screen. These hits often appear as a result of some kind of inadequately corrected auto-fluorescence effect.
51. This comparison helps to identify compounds which were picked up as “hits” in the corrected data set due to some auto-fluorescence effect which was not adequately corrected for using the described methods. This is particularly a problem with the most highly fluorescent compounds, an effect which is further enhanced by the use of two plate sets for the background subtraction, as described in Subheading 3.2.8. For highly fluorescent compounds, even the smallest differences between sets of plates will result in inadequate background subtraction. These effects can be minimized by using pintools to dispense compounds and by performing the background subtraction in a single set of plates as described in Note 41.

Compounds which have the lowest intrinsic levels of auto-fluorescence therefore will appear as “hits” regardless of the background correction, since the signal from the compounds is negligible compared to that of the labeled kinase.

52. A typical secondary screen uses at least 8 concentrations of a single compound (starting as high as 20–50 μM) and 100 nM labeled kinase.
53. Pre-dilutions are typically prepared by first preparing the highest concentration of the dilution series in a single well of a 384-well plate. The dilution series is then prepared by sequentially transferring a fixed volume of inhibitor from across a row of wells already containing DMSO. With each subsequent transfer of compound to an adjacent well containing DMSO, the contents of each well should be mixed by pipetting up and down before subsequently transferring compound to the next well. The process should be repeated until the entire dilution series is prepared.

References

1. Knight ZA, Lin H, Shokat KM (2010) Targeting the cancer kinome through polypharmacology, *Nat Rev Cancer* 10: 130–137
2. Liu Y, Gray NS (2006) Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol* 2:358–364
3. Backes AC, Zech B, Felber B, Klebl B, Müller G (2008) Small-molecule inhibitors binding to protein kinases. Part I: exceptions from the traditional pharmacophore approach of type I inhibition *Expert Opin Drug Discovery* 3:1409–1425
4. Backes AC, Zech B, Felber B, Klebl B, Müller G (2008) Small-molecule inhibitors binding to protein kinase. Part II: the novel pharmacophore approach of type II and type III inhibition. *Expert Opin Drug Discovery* 3:1427–1449
5. Rabiller M, Getlik M, Kluter S, Richters A, Tuckmantel S, Simard JR, Rauh D (2010) Proteus in the world of proteins: conformational changes in protein kinases. *Arch Pharm (Weinheim)* 343:193–206
6. Simard JR, Getlik M, Grutter C, Pawar V, Wulfert S, Rabiller M, Rauh D (2009) Development of a fluorescent-tagged kinase assay system for the detection and characterization of allosteric kinase inhibitors. *J Am Chem Soc* 131:13286–13296
7. Simard JR, Kluter S, Grutter C, Getlik M, Rabiller M, Rode HB, Rauh D (2009) A new screening assay for allosteric inhibitors of cSrc, *Nat Chem Biol* 5: 394–396
8. Simard JR, Grutter C, Pawar V et al. (2009) High-throughput screening to identify inhibitors which stabilize inactive kinase conformations in p38alpha. *J Am Chem Soc* 131:18478–18488
9. Simard JR, Getlik M, Grutter C, Schneider R, Wulfert S, Rauh D (2010) Fluorophore labeling of the glycine-rich loop as a method of identifying inhibitors that bind to active and inactive kinase conformations. *J Am Chem Soc* 132:4152–4160
10. de Lorimier RM, Smith JJ, Dwyer MA, et al. (2002) Construction of a fluorescent biosensor family. *Protein Sci* 11:2655–2675
11. Pargellis C, Tong L, Churchill L, et al. (2002) Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat Struct Biol* 9: 268–272
12. Simard JR, Kamp F, Hamilton JA (2007) Acrylodan-labeled intestinal fatty acid-binding protein to measure concentrations of unbound fatty acids. *Methods Mol Biol* 400:27–43

Electrochemical Aptamer Sensor for Small Molecule Assays

Xin Liu, Wang Li, Xiahong Xu, Jiang Zhou, and Zhou Nie

Abstract

Detection and quantification of small molecules have played essential roles in environmental analysis and clinical diagnosis. Aptamers are oligonucleic acids that bind to a specific target molecule with high specificity and affinity which are promising features for sensing small molecules. Electrochemical detection is an attractive way to exploit aptamer sensors (aptasensors) because of its high sensitivity, simple instrumentation, low cost, fast response and portability. Herein, we describe a label-free small molecular aptasensor based on a signal-amplification mechanism which uses gold nanoparticles. This aptasensor can selectively detect low nanomolar levels of ATP, the example target compound.

Key words: Aptamer, Electrochemical sensor, Gold nanoparticles, Aptasensor, Signal-amplification, Chronocoulometry, ATP

1. Introduction

The sensitive detection of small molecules is of utmost importance to environmental analysis, food and drug quality control, and clinical diagnosis (1–3). Electrochemical biosensors offer a promising platform for small molecule detection owing to its high sensitivity, fast response, low cost, simple instrumentation, and portability (4). As a commercially successful example, electrochemical glucose sensors (blood glucose meters) have been widely employed by millions of diabetics for routine tests. An electrochemical biosensor consists of a biological recognition element and a transducer (electrode) capable of converting recognition events into an electrochemical signal. The biological recognition element is the determinate factor of selectivity and sensitivity of electrochemical biosensors. Aptamers (5, 6), an emerging class of recognition elements, are single-stranded DNA or RNA sequences artificially selected through

SELEX (systematic evolution of ligands by exponential enrichment). They are able to bind a variety of targets, including small molecules, proteins, and even cells, with high affinity and specificity. Moreover, aptamers can provide a number of advantages such as simple synthesis, easy labeling, good stability and wide applicability; these features make aptamers rivals to antibodies in small molecule recognition.

Although many efforts have been made to develop feasible small molecule aptasensors, the improvement of detection limits is still a challenge because of the relatively low association constants of aptamers with small molecules. Thus, the development of effective amplification technology is essential. Several methods have been exploited to amplify the signal, such as DNAzyme (7), rolling cycle amplification (RCA) (8), and strand displacement amplification (SDA) (9), but the detection limit of these amplification methods is still not satisfactory (millimolar level). In recent years, inorganic nanoparticles (NPs) were used as signal amplification elements in various DNA or protein electrochemical assays (10–12). In these systems, the nanoparticles with a large surface area capable of loading numerous electrochemical probes can effectively amplify the biological recognition events. Furthermore, the high stability and facile, low-cost preparation of nanoparticles make the nanoparticles-based detection particularly useful in resource-limited conditions.

Here we describe two types of label-free electrochemical aptasensors for small molecule detection based on gold nanoparticle (AuNP) amplification (13, 14). These sensors are designed with either signal-off or signal-on modes, respectively (see Fig. 1).

ATP was studied as the model analyte, with an ATP-binding aptamer used as the reorganization element. Each ATP aptasensor contains three functional components: reporter DNA-functionalized gold nanoparticle (RDNA–AuNP), anchored DNA (ADNA) immobilized on electrode, and the target-responsive DNA (TRDNA). The ADNA and RDNA are thiol-modified, which can be readily immobilized on bulk or gold nanoparticle surfaces through Au–S bonds. $[\text{Ru}(\text{NH}_3)_6]^{3+}$ (RuHex), the electroactive probe, is electrostatically bound to anionic phosphates of DNA strands with an intrinsic stoichiometric ratio, thus the cumulative redox charge of RuHex measured by chronocoulometry (CC) is a direct function of the quantity of DNA strands in proximity to the electrode surface. The RDNA-modified AuNPs, loaded with about one hundred RDNA strands per particle, are used as the signal amplification element in the chronocoulometric assay through the adsorption of numerous RuHex cations onto negatively charged RDNA–AuNPs. TRDNA containing the aptamer sequence acts as the recognition element which is specific for ATP. The binding of the aptamer to the target induces a switch in the structure of TRDNA and correspondingly releases the complementary strand of TRDNA, thus triggering signal generation. In the signal-off

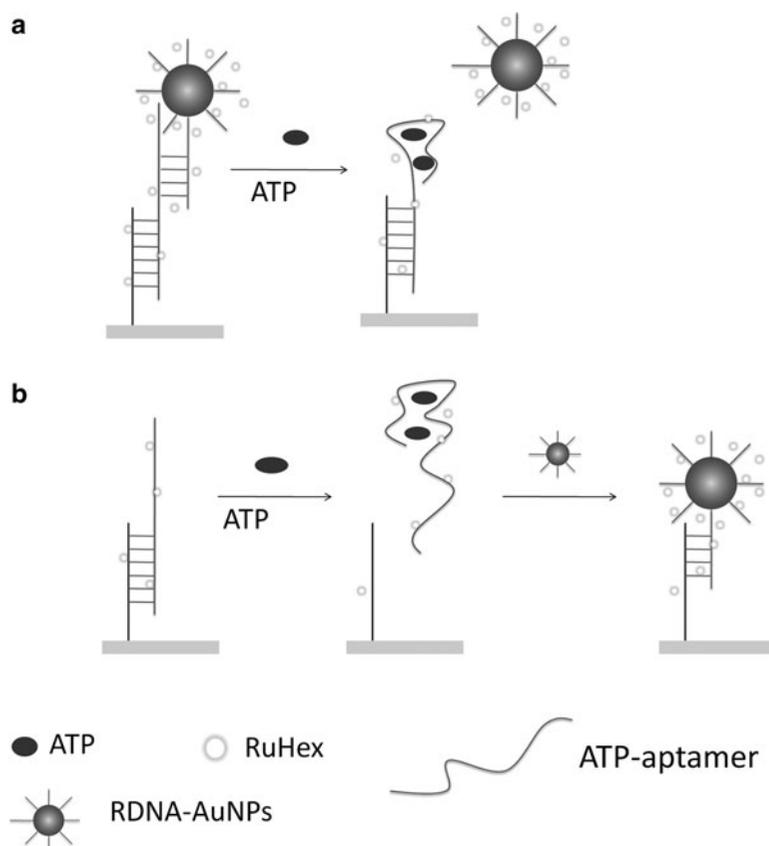


Fig. 1. Schematic representation of the aptasensors based on Au nanoparticle amplification for ATP detection: (a) Signal-off design. (b) Signal-on design.

sensor (see Fig. 1a), TRDNA1 hybridizes with ADNA1 and RDNA1–AuNPs to form a sandwich-type complex confined to the electrode. The binding of TRDNA1 with ATP causes the release of RDNA1–AuNPs, resulting in a decrease in the chronocoulometric signal. In the signal-on sensor on the other hand (see Fig. 1b), the formation of aptamer-ATP complex releases the TRDNA2, frees the ADNA2, so that the RDNA2–AuNPs can hybridize with unbound ADNA2 for signal generation. This protocol describes the detailed preparation and detection procedures of these two electrochemical aptasensors; we describe the synthesis of AuNPs and RDNA-functionalized AuNPs, preparation of the electrode, the fabrication and detection process of aptasensors, and the chronocoulometric measurement technique. Both of these aptamer-based electrochemical sensors can achieve highly sensitive and selective detection of ATP with a detection limit at the nanomolar level.

2. Materials

All samples and solutions are prepared using ultrapure water (resistivity ≥ 18.2 M Ω cm) obtained from a Millipore system. All chemicals used are of analytical grade and without further purification. All solutions are prepared and stored at room temperature (unless indicated otherwise).

2.1. DNA Sequences

Oligonucleotides: All oligonucleotides were purchased from Sangon, Inc. (Shanghai, China). The sequences of oligonucleotides are listed as follows (see Note 1):

Signal-off sensor:

ADNA1: 5'-TCA CAG ATG AGT TT-SH-3'

RDNA1: 5'-HS-CCC AGG TTC TCT-3'

TRDNA1: 5'-*ACT CAT CTG TGA* AGA GAA **CCT GGG GGA GTA TTG CGG AGG AAG GT**-3'

TRDNA1 is composed of three segments (from 5' to 3'): the ADNA-binding domain (I) (in italics), an inserted 5-nt domain as part of the RDNA-binding domain (II) (underlined) and the 27-mer aptamer sequence (III) (in bold), and ADNA1 hybridizes with segment I of TRDNA1. RDNA1 hybridizes with segment II and 7-nt part of segment III adjacent to segment II.

Signal-on sensor:

ADNA2: 5'-CCC AGG TTC TCT TT-SH-3'

RDNA2: 5'-AGA ACC TGG GTT T-SH-3'

TRDNA2: 5'-*AGA GAA* **CCT GGG GGA GTA TTG CGG AGG AAG GT**-3'

TRDNA2 is composed of two segments: the 5-nt domain as part of the ADNA-binding domain (I) (in italics) and the 27-nt aptamer sequence (in bold). ADNA2 hybridizes with segment I and 7-nt part of segment II adjacent to segment I. RDNA2 is complementary to ADNA2.

2.2. Buffer Solutions

1. Immobilization buffer (I-buffer): 10 mM Tris-HCl, 1 mM Ethylenediaminetetraacetic acid (EDTA), 0.1 M NaCl, 10 mM Tris (2-carboxyethyl) phosphine hydrochloride (TCEP), pH 7.4.
2. Electrochemistry buffer (E-buffer): 10 mM Tris-HCl, pH 7.4.
3. Washing buffer (W-buffer): 10 mM Tris-HCl, pH 7.4.
4. Hybridization buffer (H-buffer): 10 mM phosphate buffer, 0.25 M NaCl, pH 7.4.
5. MCH solution: 1 mM 6-mercapto-1-hexanol (MCH) in water.
6. RuHex solution: 10 mM hexaammineruthenium(III) chloride ($[\text{Ru}(\text{NH}_3)_6]^{3+}$, RuHex) (Aldrich, St Louis, MO, USA) in water.

7. AuNPs washing buffer: 10 mM phosphate buffer, 0.1 M NaCl, pH 7.0.
8. Salting buffer: 10 mM phosphate buffer, 2.0 M NaCl, pH 7.0.
9. HAuCl_4 solution: 25.4 mM gold (III) chloride hydrate ($\text{HAuCl}_4 \cdot 3\text{H}_2\text{O}$) in water. Dissolve 1 g of gold (III) chloride hydrate in 100 mL water (see Note 2).

2.3. Equipment

1. CHI660A electrochemical workstation (CH Instruments, Inc.).
2. Gold working electrode (2 mm in diameter).
3. A rectangular platinum foil is used as the counter electrode.
4. Saturated calomel reference electrode.
5. Cell stand (CH Instruments, Inc.). A 25-mL glass beaker is used as an electrochemical cell vial.
6. Setup of the electrochemical cell: A 25-mL clean beaker placed on the cell stand is plugged with a gold working electrode, a platinum foil counter electrode, and a saturated calomel reference electrode to form a standard three-electrode cell. For chronocoulometric measurement, E-buffer (9 mL) is added to the beaker and purged with nitrogen for 10 min through Teflon tubing. Nitrogen gas should be bubbled at a moderate flow rate to avoid splashing. Then raise the tip of the gas tubing over the top of the solution and keep the solution under nitrogen during the subsequent electrochemical measurements.
7. Alpha alumina powder, 0.3 μm (CH Instruments, Inc.).
8. Gamma alumina powder, 0.05 μm (CH Instruments, Inc.).
9. Microcloth (Buehler).
10. DU 800 UV/Visible Spectrophotometer (Beckman Coulter).

3. Methods

3.1. Preparation of Gold Nanoparticles

1. Add 4 mL of 25.4 mM HAuCl_4 solution to 96 mL water in a clean 250 mL round-bottom flask equipped with stir bar, and then attach a reflux column on the top of the flask (see Note 3). The solution in flask is heated in an oil bath to boil under stirring.
2. Add 10 mL of 38.8 mM sodium citrate to the HAuCl_4 solution in flask. Upon addition of sodium citrate, the color of solution changes from pale yellow to deep wine red, indicating the increase in size of gold nanoparticles as the result of the reduction of gold (III) by citrate ions. Keep the mixture boiling for 10 min and then cool to room temperature under continuous stirring.

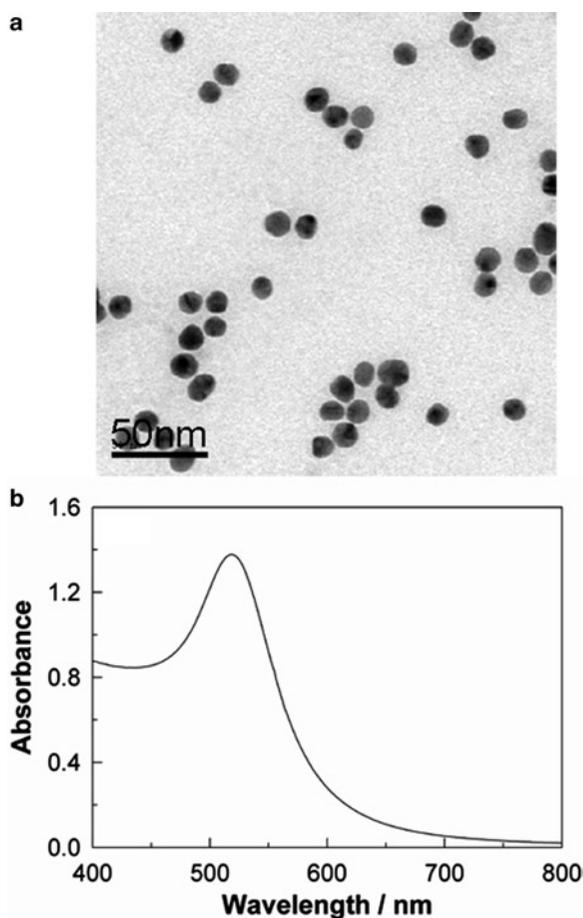


Fig. 2. (a) TEM images of the 13 nm AuNPs. The scale bar is 50 nm. (b) UV/Vis absorption spectra of 13 nm AuNPs.

3. Filter the cooled AuNPs solution through a 0.45 μm Gelman membrane filter. The resulting solution contains spherical monodisperse AuNPs of about 13 nm in diameter, which can be determined by transmission electron microscopy (TEM) and UV-Vis spectroscopy as shown in Fig. 2 (see Note 4).

3.2. Functionalization of AuNPs with Reporter DNA

1. Place 1 mL of AuNPs (13 nm, 13 nM) to a 1.5-mL centrifuge tube. Centrifuge the gold nanoparticles at $12,000 \times g$ and 4°C for 15 min. Remove supernatants and re-suspend precipitates in 1 mL of 10 mM phosphate buffer (pH 7.0). Vortex to re-disperse nanoparticles.
2. Mix 1 μL of 10 mM TCEP with 99 μL of 100 μM reporter DNA (RDNA1 or RDNA2) to activate the thiol group of RDNA. Incubate the mixture for 30 min at room temperature (see Note 5).
3. Prepare the RDNA-modified AuNPs by adding 35 μL of 100 μM RDNA (final concentration: $\sim 3 \mu\text{M}$) to 1 mL of the

resulting AuNPs solution of step 1. Incubate the mixture for 16 h at room temperature.

4. Salting step: Add 9 μL of salting buffer drop-by-drop to the solution. Repeat this salting step six times with 30 min intervals between steps. The final salt concentration is 0.1 M NaCl (see Note 6). After salting-steps, the mixture is incubated for at least 40 h at room temperature with occasional shaking.
5. Washing step: centrifuge the RDNA–AuNP conjugates at $18,600\times g$ and 4°C for 15 min. Discard the supernatant and re-suspend the precipitate in 1 mL of AuNPs washing buffer. Repeat this washing step three times.
6. Centrifuge the RDNA–AuNP conjugates at $18,600\times g$ and 4°C for 30 min. Remove supernatants and re-suspend precipitates in 1 mL of hybridization buffer (see Note 7).

3.3. Preparation of Electrodes

1. Attach a piece of microcloth to a stiff flat surface (such as a glass plate). Deposit a small amount of $0.3\ \mu\text{m}$ alumina powder on the microcloth and add distilled water to the alumina powder to form a small slurry. Polish the gold electrode under moderate pressure with the slurry while moving the electrode in a figure-of-eight pattern on the microcloth. Polish for about 5 min and then thoroughly rinse the electrodes with distilled water to remove alumina particles (see Note 8).
2. Repeat step 1 using $0.05\ \mu\text{m}$ alumina powder.
3. Sonicate the polished electrode sequentially in distilled water, ethanol, and distilled water for 5 min each to ensure complete removal of the alumina particles and other impurities.
4. Dip the sonicated electrode in piranha solution (v/v 3:1 $\text{H}_2\text{SO}_4/\text{H}_2\text{O}_2$) for 10 min (see Note 9).
5. Rinse the electrode with distilled water thoroughly and blow-dry it in a mild stream of nitrogen.
6. Immerse the electrode in 0.5 M H_2SO_4 solution. Apply a positive potential of 2 V to the electrodes for 5 s and subsequently a negative potential of $-0.35\ \text{V}$ for 10 s. Then run 5–10 cycles of cyclic voltammetry (CV) with the following parameters: potential range, -0.3 to $1.55\ \text{V}$; scan rate, $4\ \text{V/s}$.
7. Place the electrodes into a fresh 0.5 M H_2SO_4 solution. Run one CV cycle with the following parameters: potential range, -0.3 to $1.55\ \text{V}$; scan rate, $0.1\ \text{V/s}$ (see Note 10).
8. Repeat the washing process as conducted in step 5 (see Note 11).

3.4. Preparation and Detection Process of the Signal-Off Aptasensor (see Fig. 1a)

1. Dispense $4\ \mu\text{L}$ of anchored DNA (ADNA1) in I-buffer on the gold surface of an electrode (see Note 12). Cap the electrode with 1.5-mL plastic centrifuge tube and incubate for 1 h at room temperature (see Note 13).

2. Rinse the electrodes extensively with washing buffer two times and blow-dry it under a gentle stream of nitrogen.
3. Immerse the modified electrodes in 1 mM MCH solution and incubate for 2 h at room temperature (see Note 14).
4. Repeat the washing process as conducted in step 2.
5. Dispense 4 μL of 10 μM target-responsive DNA (TRDNA1) on the electrode. Cap the electrode with 1.5-mL plastic centrifuge tube and incubate for 2 h at room temperature.
6. Repeat the washing process as conducted in step 2.
7. Dispense 4 μL of RDNA1–AuNPs on the electrode. Cap the electrode with 1.5-mL plastic centrifuge tube and incubate for 2 h at room temperature.
8. Repeat the washing process as conducted in step 2 (see Note 15).
9. Dispense 4 μL of ATP sample (either test samples of unknown concentrations or the standard samples at different concentrations ranging from 1 nM to 1 mM) on the modified electrode for 1 h.
10. Rinse the electrode with a large amount of washing buffer and blow-dry electrode under a gentle stream of nitrogen before chronocoulometric measurement.

3.5. Preparation and Detection Process of the Signal-On Aptasensor (see Fig. 1b)

1. Dispense 4 μL of anchored DNA (ADNA2) in I-buffer on the gold surface of an electrode (see Note 12). Cap the electrode with a 1.5-mL plastic centrifuge tube and incubate for 1 h at room temperature (see Note 13).
2. Rinse the electrodes extensively with washing buffer two times and blow-dry under a gentle stream of nitrogen.
3. Immerse the modified electrodes in 1 mM MCH solution and incubate for 2 h at room temperature (see Note 14).
4. Repeat the washing process as conducted in step 2.
5. Dispense 4 μL of 10 μM target-responsive DNA (TRDNA2) on the electrode. Cap the electrode with a 1.5-mL plastic centrifuge tube and incubate for 2 h at room temperature.
6. Repeat the washing process as conducted in step 2.
7. Dispense 4 μL of ATP sample (either test samples of unknown concentrations or the standard samples at different concentrations ranging from 1 nM to 10 mM) on the modified electrode for 1 h.
8. Repeat the washing process as conducted in step 2.
9. Dispense 4 μL of RDNA2–AuNPs on the electrode. Cap the electrode with a 1.5-mL plastic centrifuge tube and incubate for 2 h at room temperature.

10. Rinse the electrode with a large amount of washing buffer and blow-dry electrode under a gentle stream of nitrogen before chronocoulometric measurement.

3.6. Chronocoulometric Measurement (see Note 16)

1. Add 9 mL of E-buffer to another glass cell (see Note 17).
2. Add 45 μL of 10 mM RuHex solution to the E-buffer (see Note 18). Purge the solution thoroughly with nitrogen for 10 min through Teflon tubing. Turn on the gas stream slowly and bubble the solution with a moderate flow rate (see Note 19).
3. Raise the tip of gas tubing over the top of solution and maintain the solution under nitrogen ambience during subsequent electrochemical measurements.
4. Place the resulting gold electrode (the electrode after treatment as in Subheading 3.4 for the signal-off sensor, and the electrode after treatment as in Subheading 3.5 for the signal-on sensor) in the cell, then incubate in the test buffer for 10 min.
5. Perform CC measurement to obtain a signal with the following parameters: initial potential, 0.2 V; final potential, -0.5 V; number of steps, 2; pulse width, 0.25 s; sample interval, 0.002; sensitivity (C or A/V), $1\text{e-}4$ A/V.
6. Plot charge (Q) vs. $(\text{time})^{1/2}$ ($t^{1/2}$). The Charge of the electrode can be calculated from the chronocoulometric intercept at $t=0$. (see Fig. 3) Q_{sample} stands for the charge of the electrode treated by the ATP-containing sample. Q_{control} stands for the charge of the electrode treated by the control sample without ATP. The signal is defined as the difference of the redox charge, that is, $\text{signal} = Q_{\text{control}} - Q_{\text{sample}}$ for the signal-off aptasensor; or $\text{signal} = Q_{\text{sample}} - Q_{\text{control}}$ for the signal-on aptasensor (see Note 20).

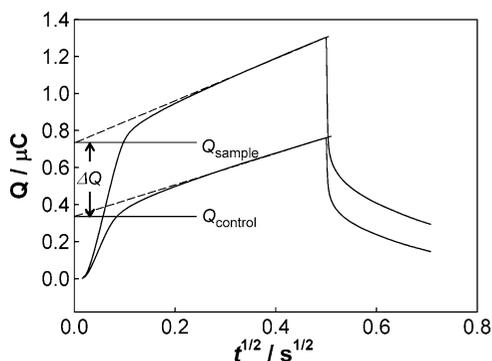


Fig. 3. Chronocoulometric measurement. Representative chronocoulometric curves of signal-on aptasensor for the control sample without ATP (Q_{control}) and the sample with ATP (Q_{sample}). Intercepts at $t=0$ in chronocoulometric curves represent redox charges of RuHex bound to DNA. Signal is defined as the difference in the redox charge of RuHex with and without ATP treatment, which is equal to the difference in the chronocoulometric intercept at $t=0$ with and without ATP treatment ($\text{signal} = Q_{\text{sample}} - Q_{\text{control}}$).

4. Notes

1. All TRDNA sequences (TRDNA1 and TRDNA2) contain the 27-mer ATP-binding aptamer (5'-ACCTGGGGGAGTATTGCGGAGGAAGGT-3'). Pay attention to the direction of each DNA sequence (from 5' to 3'); making sure not to reverse the sequence. The reverse sequence of aptamer will have no target-binding capability.
2. Gold (III) chloride hydrate is very hygroscopic. Dissolve HAuCl_4 in water as soon as possible after the original package of HAuCl_4 is unsealed.
3. The cleanness of glassware is a key factor for the preparation of good quality gold nanoparticles. All the glassware must be soaked in aqua regia for at least 30 min, rinsed with copious amount of deionized water and then ultrapure water. Prepare aqua regia by mixing concentrated HCl/HNO_3 in a 3:1 ratio of volume in a large beaker. It should be prepared in a well-ventilated fume hood with protective clothing, goggles and gloves. Be extremely careful when preparing and using aqua regia. Aqua regia should be freshly prepared and should never be stored in a closed vessel.
4. The citrate-reduced AuNPs are stable in aqueous media and can be stored at room temperature for months. Do NOT freeze, as this will induce aggregation. The final concentration of AuNPs can be estimated using UV-Vis spectrometric measurement based on an extinction coefficient of $\sim 2.7 \times 10^8 \text{ M}^{-1} \text{ cm}^{-1}$ at 520 nm for 13 nm particles.
5. TCEP is used here as a reducing agent to break disulfide bonds generated by oxidation of thiolated DNA. Because of its oxidation in air, TCEP solution is freshly prepared before use.
6. The salting buffer should be added very slowly (with frequent tapping of the tube) to avoid possible aggregation of AuNPs. The color of the solution will change from red to blue or purple if the AuNPs are aggregated.
7. Good dispersion of RDNA-AuNPs is important detection performance. The monodisperse RDNA-modified AuNPs should have the same color as the original citrate-reduced AuNPs. If the color of solution turns to blue or purple, indicating the formation of AuNPs aggregates, this batch of RDNA-AuNPs cannot be used further. Store the RDNA-AuNP conjugates away from direct sunlight at 4°C. Do NOT freeze. For consistent results, we recommend that RDNA-AuNPs should be used within 2–3 weeks of preparation.
8. To obtain a smooth planar surface, electrodes should be held vertically to the surface of the microcloth while polishing.

9. The piranha solution should be handled with great care because it can react violently with many organic substances.
10. It is important to obtain a clean gold electrode surface. A background cyclic voltammogram of the fresh 0.5 M H_2SO_4 is a good way to confirm the cleanliness of the polished working electrodes. The single sharp characteristic reduction peak located at ~ 0.9 V and multiple overlapping oxidation peaks in the range of 1.2–1.5 V are clearly visible, indicating a clean gold electrode. However, if multiple reduction peaks appear, revealing that the gold electrode is not well cleaned, run CV repetitively in a freshly prepared 0.5 M H_2SO_4 solution until a steady single sharp reduction peak appears, or re-polish the electrode if electrochemical cleaning does not work.
11. The cleaned electrodes should be used for DNA immobilization as soon as possible to prevent deposition of contaminants on the electrode surface.
12. In order to obtain a perfect sensitivity, optimization of the concentration of anchored DNA is needed. Here, we use 0.2 μM ADNA as the appropriate concentration to obtain a perfect sensitivity.
13. The electrode should be well capped to prevent solution on electrode surface from evaporation and contamination.
14. The self-assembly of MCH on the gold electrode surface is used to remove nonspecific ADNA adsorption and position the ADNA probes in a highly hybridizable orientation (15).
15. Careful and thorough rinsing is essential for removing all unhybridized AuNPs which may cause false-positive CC signals.
16. Chronocoulometry is an electrochemical technique which measures the time-dependent charge response to an applied potential step waveform. Because its key feature is that the charge due to redox process of surface-confined electroactive species can be distinguished from the charge due to redox process of electroactive species in solution, chronocoulometry is particularly useful for measurement of electroactive compounds adsorbed onto an electrode surface. Hence, chronocoulometry is employed here to measure RuHex adsorbed on surface-confined DNA strands. In an Anson plot (charge Q as a function of $t^{1/2}$; see Fig. 3), the chronocoulometric intercept at $t=0$ is the sum of the charging of double layer and the charging of surface-confined RuHex (16). Since the charging of a double layer is a constant, the difference in intercept ($Q_{\text{sample}} - Q_{\text{control}}$) represents the difference in the redox charge of RuHex after and before ATP treatment, which is defined as signal of the aptasensor.
17. The sensor relies upon the electrostatic adsorption of $[\text{Ru}(\text{NH}_3)_6]^{3+}$ on the phosphate of DNA. If high concentration

of NaCl is present in E-buffer, the sodium cations will compete with $[\text{Ru}(\text{NH}_3)_6]^{3+}$ for binding to the anionic DNA backbone and the high ionic strength will weaken the interactions between $[\text{Ru}(\text{NH}_3)_6]^{3+}$ and DNA, leading to a decrease in the amount of adsorbed $[\text{Ru}(\text{NH}_3)_6]^{3+}$ at the modified electrode (16). Thus, we use the E-buffer without NaCl for RuHex adsorption and CC measurement.

18. We find that $50 \mu\text{M}$ $[\text{Ru}(\text{NH}_3)_6]^{3+}$ is suitable for the experiments as it ensures saturation of the modified electrode with the RuHex.
19. Be careful to adjust the gas stream to the appropriate flow rate to avoid splashing the test buffer.
20. This protocol allows sensitive and selective detection of ATP. Figure 4a shows the ATP concentration-dependent chrono-coulometric curves of the signal-off aptasensor. The linear

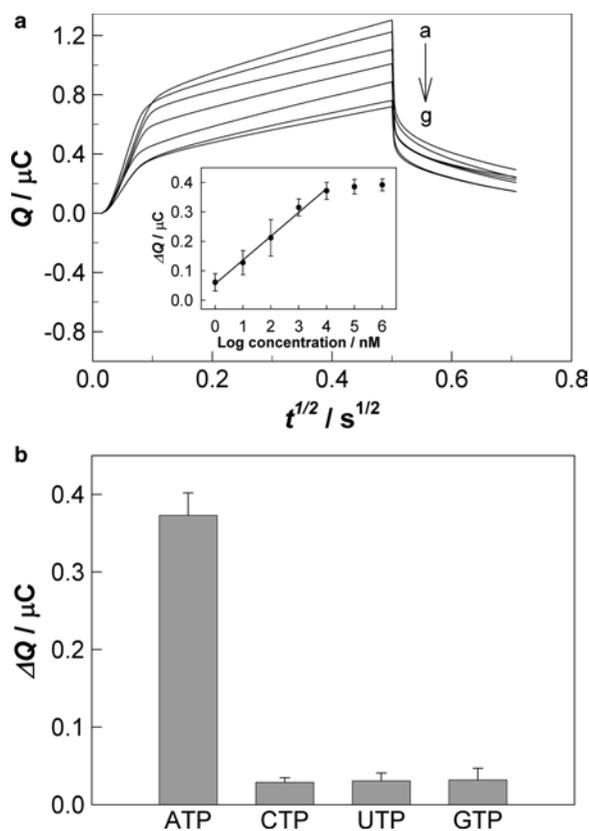


Fig. 4. (a) Chronocoulometric curves of the signal-off aptasensor for a series of concentrations of ATP (from a to g: 1 nM, 10 nM, 100 nM, 1 μM , 10 μM , 100 μM , and 1 mM). The inset is the calibration curves for the detection of ATP. (b) The selectivity of the aptasensor. The aptasensor of signal-off design was treated with 10 μM ATP, CTP, UTP, or GTP solution for 1 h. Error bars show the standard deviations of measurements taken from three independent experiments. (Adapted from (14))

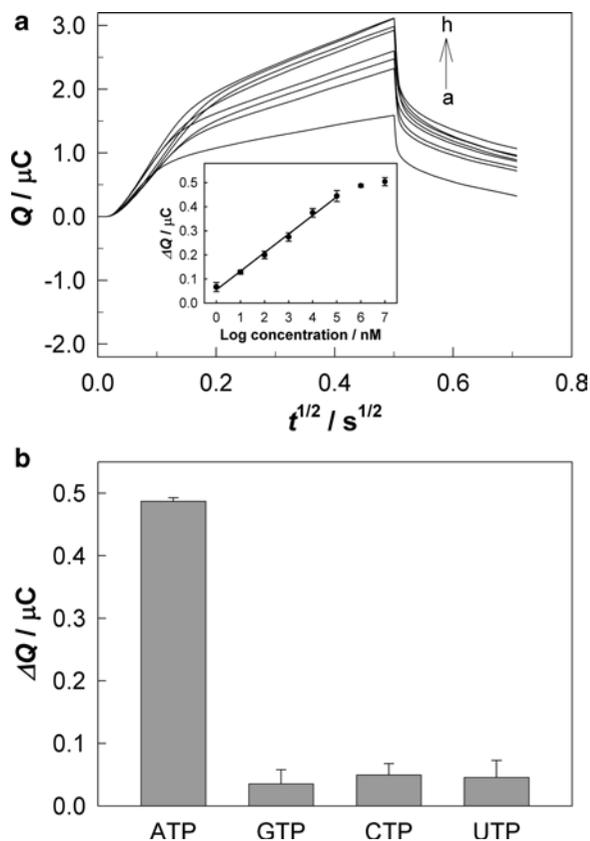


Fig. 5. (a) Chronocoulometric curves of the signal-on aptasensor for a series of concentrations of ATP (from a to h: 1 nM, 10 nM, 100 nM, 1 μM , 10 μM , 100 μM , 1 mM, and 10 mM). The inset is the calibration curves for the detection of ATP. (b) The selectivity of the aptasensor. The aptasensor of signal-on design was treated with 1 mM ATP, CTP, UTP, or GTP solution for 1 h. Error bars show the standard deviations of measurements taken from three independent experiments.

range of this signal-off aptasensor is from 1 nM to 10 μM with the minimum detectable concentration of 1 nM. This sensor can effectively differentiate ATP from other nucleoside triphosphates (see Fig. 4b). Figure 5a indicates the linear range of this signal-on aptasensor is from 1 nM to 100 μM with the minimum detectable concentration of 1 nM. Good selectivity for ATP is also demonstrated (see Fig. 5b). The detection limit of these aptasensors, which are based on nanoparticle amplification, is at least 2 orders of magnitude lower than that of aptasensors without amplification or with nucleic acid amplification (17–20). This protocol is versatile and can be employed to detect other small molecules by using a suitable aptamer.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (Nos. 20805013, 20907013), the National Basic Research Program of China (973 Program, Nos. 2009CB421601, 2011CB911002), the Natural Science Foundation of Hunan Province (Nos. 09JJ4006, 10JJ2005), and SKLCSBSC-2010-01. We acknowledge Elsevier for the original publication of Fig. 4 in *Talanta*.

References

- Zuo P, Ye BC (2006) Small Molecule Microarrays for Drug Residue Detection in Foodstuffs. *J Agric Food Chem* 54: 6978–6983
- Kerman K, Nagatani N, Chikae M et al (2006) Label-free electrochemical immunoassay for the detection of human chorionic gonadotropin hormone. *Anal Chem* 78: 5612–5616
- Liu BF, Ozaki M, Hisamoto H et al (2005) Microfluidic chip toward cellular ATP and ATP-conjugated metabolic analysis with bioluminescence detection. *Anal Chem* 77: 573–578
- Drummond TG, Hill MG, Barton JK (2003) Electrochemical DNA sensors. *Nat Biotechnol* 21: 1192–1199
- Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* 346: 818–822
- Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249: 505–510
- Li D, Shlyahovsky B, Elbaz J et al (2007) Amplified analysis of low-molecular-weight substrates or proteins by the self-assembly of DNAzyme-Aptamer conjugates. *J Am Chem Soc* 129: 5804–5805
- Cho EJ, Yang LT, Levy M (2005) Using a deoxyribozyme ligase and rolling circle amplification to detect a non-nucleic acid analyte, ATP. *J Am Chem Soc* 127: 2022–2023
- Shlyahovsky B, Li D, Weizmann Y et al (2007) Spotlighting of cocaine by an autonomous aptamer-based machine. *J Am Chem Soc* 129: 3814–3815
- He PL, Shen L, Cao YH et al (2007) Ultrasensitive electrochemical detection of proteins by amplification of aptamer-nanoparticle bio bar codes. *Anal Chem* 79: 8024–8029
- Zhang J, Song SP, Zhang LY et al (2006) Sequence-specific detection of femtomolar DNA via a chronocoulometric DNA sensor (CDS): Effects of nanoparticle-mediated amplification and nanoscale control of DNA assembly at electrodes. *J Am Chem Soc* 128: 8575–8580
- Deng CY, Chen JH, Nie Z, et al (2009) Impedimetric Aptasensor with Femtomolar Sensitivity Based on the Enlargement of Surface-Charged Gold Nanoparticles. *Anal Chem* 81: 739–745
- Deng CY, Chen JH, Nie LH et al (2009) Sensitive bifunctional aptamer-based electrochemical biosensor for small molecules and protein. *Anal Chem* 81: 9972–9978
- Li W, Nie Z, Xu XH et al (2009) A sensitive, label free electrochemical aptasensor for ATP detection. *Talanta* 78: 954–958
- Li D, Song SP, Fan CH (2010) Target-responsive structural switching for nucleic acid-based sensors. *Acc Chem Res* 43: 631–641
- Steel AB, Herne TM, Tarlov MJ (1998) Electrochemical quantitation of DNA immobilized on gold. *Anal Chem* 70: 4670–4677
- Han K, Chen L, Lin ZS et al (2009) Target induced dissociation (TID) strategy for the development of electrochemical aptamer-based biosensor. *Electrochem Commun* 11: 157–160
- Shen L, Chen Z, Li YH et al (2007) A chronocoulometric aptamer sensor for adenosine monophosphate. *Chem Commun* 2169–2171
- Yu FS, Li L, Chen F (2008) Determination of adenosine disodium triphosphate using prulifloxacin-terbium (III) as a fluorescence probe by spectrofluorimetry. *Anal Chim Acta* 610: 257–262
- Chen SJ, Huang Y F, Huang CC et al (2008) Colorimetric determination of urinary adenosine using aptamer-modified gold nanoparticles. *Biosens Bioelectron* 23: 1749–1753

Miniaturized, Microarray-Based Assays for Chemical Proteomic Studies of Protein Function

Jonathan M. Blackburn, Aubrey Shoko, and Natasha Beeton-Kempen

Abstract

Systematic analysis of protein and enzyme function typically requires scale-up of protein expression and purification prior to assay development; this can often be limiting. Miniaturization of assays provides an alternative approach, but simple, generic methods are in short supply. Here we show how custom microarrays can be adapted to this purpose. We discuss the different routes to array fabrication and describe in detail one facile approach in which the purification and immobilization procedures are combined into a single step, significantly simplifying the array fabrication process. We illustrate this approach by reference to the creation of arrays of human protein kinases and of human cytochrome P450s. We discuss methods for both ligand-binding and turnover-based assays, as well as data analysis on such arrays.

Key words: Protein microarray, Biotinylation, Proteomics, Functional analysis, Surface capture, Protein kinase, Cytochrome P450, Biotin carboxyl carrier protein, Inhibitor specificity

1. Introduction

In the post-genomic era, attention has turned towards the systematic assignment of function to proteins, including enzymes, which are encoded by genomes. Bioinformatics methods are now ubiquitously used as an essential first step in assigning predicted function to open reading frames (1); the methods are based on, for example, sequence or structural homology, or conservation of characteristic motifs, etc. However, whilst such methods give helpful insights into possible function, there remain many examples of proteins that have closely related sequences and/or structures but which prove to have quite different functions when studied experimentally (2–4). In addition, there are many other examples of proteins where the predicted function falls within a twilight zone

(e.g. *Mycobacterium tuberculosis* CYP135A1 where the sequence homology to the closest putative relative is <30%). As the number of sequenced genomes expands ever further, there is thus an ever increasing need for experimental methods that enable the determination and/or verification of protein function at high throughput. At the forefront of this monumental task, the field of proteomics can be divided into discovery- and systems-oriented proteomics (5). Discovery-oriented proteomics is mainly concerned with documenting the abundance and localisation of individual proteins, as well as building a picture of protein–protein interaction networks. This is the realm of 2-hybrid screens, 2D-gel electrophoresis and increasingly powerful, more direct, isotope-labelling-based mass spectrometry methods; these latter two methods in particular are commonly used to understand the way in which expression profiles change in response to different stimuli by comparing, for example, extracts from diseased and healthy cells. However, these discovery-oriented proteomics methods tell us little about the precise function of individual proteins or protein complexes, even when augmented by ever more sophisticated bioinformatic methods. Systems-oriented proteomics takes a different approach; rather than re-discovering each protein de novo in each new experiment, the focus is on a pre-defined target set of proteins, in principle up to an entire proteome, but in practise more typically a limited subset thereof, thus allowing the functionality of each member of that set to be dissected in great detail (6). However, obtaining quantitative and genuinely comparable functional data across large sets of proteins (particularly enzymes) with any degree of accuracy is technically difficult, requiring isolation of each individual protein in an assayable format. We and others have chosen to focus on miniaturised protein microarray-based methods and assays. The parallel, high-throughput nature of microarray experiments is attractive for a functional analysis of large numbers of proteins, the variety of assays that can in principle be carried out in this format is wide and the uniform intra-array conditions both simplify and increase accuracy (6–13). Additionally, the small volumes of ligand or reaction solution required to perform assays, typically tens to hundreds of microlitres, can provide economic advantages, for example when using expensive recombinant proteins or labelled compounds. The key element in any miniaturised protein microarray-based assay is that the arrayed immobilised proteins must retain their folded structure, such that meaningful functional interrogation can be subsequently carried out on the array surface. There are a number of approaches to this problem which differ fundamentally according to whether the proteins are immobilised through non-specific, poorly defined interactions, or through a specific set of known interactions. The former approach is attractive in its simplicity and is compatible with purified proteins derived from native or recombinant sources (14, 15) but suffers from a number of risks. Most notable

amongst these relate to the uncontrolled nature of the interactions between each protein and the surface; at best, this might give rise to a heterogeneous population of proteins in which active sites are sometimes occluded by the surface; at worst, it might destroy activity altogether due to partial or complete surface-mediated unfolding of the immobilised protein. In practice, an intermediate situation probably most often occurs, where a fraction of the immobilised proteins have either undergone conformational change as a result of the non-specific interactions, or have their binding/active sites occluded by surface attachment; these effects will effectively reduce the specific activity of the immobilised protein and therefore decrease the signal-to-noise ratio in any subsequent functional assay that is sensitive to conformation. It is therefore important to consider the possible effects of unfolding on the intended downstream assay prior to choosing an array surface: for example, an assay in which solution-phase antibodies bind to linear epitopes on the array is not likely to be affected by any unfolding of the arrayed proteins (indeed, it may even be desirable to deliberately unfold such proteins in order to expose a greater range of potential epitopes); by contrast, an assay in which an unlabeled drug candidate competes with a fluorogenic substrate for turnover by an arrayed cytochrome P450 (CYP450), may well be sensitive to disruption of the absolute and relative three-dimensional structures of CYP450 and/or cytochrome P450 reductase (POR) domains in the arrayed proteins.

The advantages of controlling the precise mode of surface attachment are that, providing the chosen point of attachment does not directly interfere with activity, the immobilised proteins will have a homogenous orientation resulting in a higher specific activity and higher signal-to-noise ratio in assays, with less interference from non-specific interactions (16). This may be of particular advantage when studying protein–small molecule interactions (including enzymatic turnover) or conformationally sensitive protein–protein interactions in an array format. The disadvantages of this approach, however, are that it is really only compatible with recombinant proteins, or with families of proteins, such as antibodies, which have a common structural element through which they can be immobilized. However, in a systems-oriented approach, the disadvantage of working with recombinant proteins is largely outweighed by the problems encountered in individually purifying large numbers of active proteins from native sources. Furthermore, experimental approaches which facilitate high throughput expression and purification of many different proteins in parallel have become generally more accessible over recent years, simplifying access to larger, defined collections of recombinant proteins. An important caveat is that it is increasingly clear that, despite its ease of use, *Escherichia coli* is not an optimal host for recombinant expression of folded, functional mammalian proteins. Furthermore,

whilst cell-free transcription/translation-based protein microarray systems have been described (17, 18), it remains unclear how reproducible such arrays are, or what proportion of mammalian proteins produced by such approaches are properly folded and therefore functional prior to immobilization.

In this chapter, we therefore describe the high throughput cloning, expression, purification, array fabrication and assay of a set of recombinant proteins in which the human proteins are expressed in insect cells and where the mode of surface attachment is tightly controlled through use of an appropriate affinity tag. Furthermore, we show how laborious pre-purification of the recombinant proteins prior to array fabrication can be avoided through the use of a suitable array surface combined with a suitable affinity tag, thus greatly simplifying array fabrication (7). We illustrate this approach to array fabrication with a set of human protein kinases, expressed in insect cells as fusions to a polypeptide tag which becomes biotinylated *in vivo* (7) (see Note 1), as well as to formation of an active, membrane-free CYP450:POR complex on an array surface. In addition, we show representative data from a number of different assays carried out on protein function arrays made in this way.

2. Materials

1. *E. coli* *Acc B* gene.
2. Human MAPK1 cDNA.
3. *Autographa californica* baculovirus “bacmid” vector pBAC10:KO₁₆₂₉.
4. *E. coli* strain HS996.
5. *Spodoptera frugiperda* SF21 cells.
6. Lipofectin (Invitrogen).
7. InsectXpress media (Lonza).
8. 6-well cell culture plates (Nunc).
9. Fetal bovine serum (FBS; Sigma).
10. 24-well deep well blocks (Nunc).
11. Biotin (Sigma).
12. Phosphate-buffered saline (PBS): 1.5 mM KH₂PO₄, 4.3 mM Na₂HPO₄, 137 mM NaCl, 3 mM KCl, adjust to pH 7.4 with HCl, prepare as 10× stock and autoclave before storage at room temperature.
13. PBST: prepare from PBS by adding 0.1% (v/v) Tween20.
14. Freezing buffer: 25 mM HEPES, 50 mM KCl, pH 7.5.

15. Lysis buffer: 25 mM HEPES pH 7.5, 20% glycerol, 50 mM KCl, 0.1% Triton X-100, 0.1% BSA, Protease inhibitor cocktail, 1 mM DTT.
16. Streptavidin–HRP conjugate (Sigma).
17. Anti-cMyc antibody (Sigma).
18. Streptavidin (Sigma).
19. 384-well V-bottomed plate (X7022; GENETIX).
20. Nexterion Slide P (Schott).
21. Lifterslips (Nunc, USA)
22. Wash buffer: 20 mM KH_2PO_4 , 0.2 mM EDTA, 5% glycerol, pH 7.4.
23. QArray II microarray robot equipped with $16 \times 300 \mu\text{m}$ solid stainless steel pins (Genetix).
24. Kinase buffer: 25 mM Tris–HCl, 5 mM beta-glycerophosphate, 2 mM dithiothreitol, 0.1 mM Na_3VO_4 , 10 mM MgCl_2 , pH 7.5.
25. Adenosine triphosphate (ATP; Sigma).
26. DNA microarray scanner (Tecan LS Reloaded).
27. Ethanolamine.
28. 150 mM Na_2HPO_4 buffer, pH 8.5: prepared by titrating 0.2 M NaH_2PO_4 into 0.2 M Na_2HPO_4 to pH 8.5 and diluting to a final concentration of 150 mM.
29. Milk/PBST: 5% (w/v) non-fat dry milk in PBST.
30. P450 Assay buffer: 100 mM KH_2PO_4 , 15% final glycerol, pH 7.4.
31. Vivid Red (Invitrogen).
32. P450 Printing buffer: 100 mM potassium phosphate, pH 7.4; 20% glycerol, 0.2 mM EDTA, 1 mM DTT.

3. Methods

The methods described below outline (1) the construction of a representative transfer vector in *E. coli*, (2) co-transfection of insect cells with this transfer vector and bacmid, (3) propagation of recombinant baculovirus, (4) induction of protein expression, (5) the extraction of the protein from insect cells, (6) the printing of a protein microarray, (7) the assay of a protein microarray for protein kinase activity, and (8) the assay of a protein microarray for cytochrome P450 turnover activity.

3.1. Construction of the Transfer Vector for Full Length Human MAPK1

3.1.1. Transfer Vector

3.1.2. Amplification and Cloning of the MAPK1 Gene as an N-Terminal Fusion to BCCP

The general baculoviral system used here is adapted from the work of Prof Ian Jones (Reading University, UK); (19). The specific *E. coli* transfer vector system used is derived from pTriEx1.1 (Novagen).

All DNA manipulations were carried out using standard recombinant DNA methods (20) to construct the transfer vector and are accordingly not described here in detail.

The gene encoding the *E. coli* biotin carboxyl carrier protein (BCCP) domain (amino acids 74–156 of the *E. coli accB* gene; Fig. 1) (21, 22) was amplified by PCR from an *E. coli* genomic DNA preparation and was cloned downstream of a viral polyhedrin promoter in an *E. coli* vector to create the transfer vector pJB1 (Fig. 2). Flanking this *polh*-BCCP expression cassette were the baculoviral 603 gene and the 1,629 genes (19) to enable subsequent homologous recombination of the construct into a replication-deficient baculoviral genome (Fig. 2).

The full length, *MAPK1* gene was amplified by PCR from a cDNA clone (obtained from the Mammalian Gene Collection) and cloned into the pJB1 transfer vector upstream of and in-frame with the BCCP tag using ligation-independent cloning methods (24), replacing the ORF region between the *Spe* I and *Nco* I sites of pJB1 in the process, to generate pJB1-MAPK1. In the course of the PCR amplification step, the stop codon of the MAPK1 gene was removed such that the resulting construct encoded an in-frame MAPK1-BCCP fusion protein; the *MAPK1* gene was sequence-verified against the RefSeq database.



Fig. 1. Structure of the *E. coli* biotin carboxyl carrier protein (BCCP) domain. Residues 77–156 are drawn (coordinate file 1bdo), showing the N- and C-termini and the single biotin moiety that is attached to lysine 122 in vivo by biotin ligase. Representation produced using SwissPDBViewer (23).

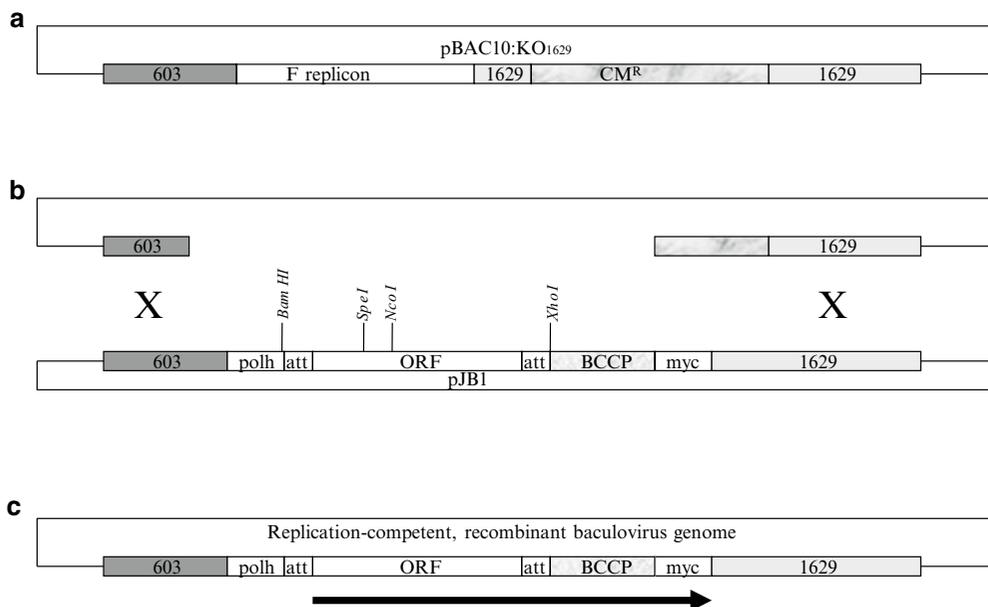


Fig. 2. Schematic of the baculoviral recombination system used in this work. (a) Disruption of the essential 1,629 gene renders the baculoviral genome replication-deficient. (b) Linearised baculoviral genomic DNA and transfer vector are co-transfected into insect cells. (c) Homologous recombination regenerates an intact 1,629 gene, enabling viral replication. Expression of the BCCP fusion protein is driven from the polh promoter.

3.2. Insertion of the MAPK1-BCCP Expression Cassette into a Baculoviral Genome

1. Bacmid pBAC10:KO₁₆₂₉ (19) was propagated in *E. coli* HS996 cells and bacmid DNA prepared according to standard procedures. pBAC10:KO₁₆₂₉ was then linearized by restriction with *Bsu*361 for 5 h at 37°C, after which *Bsu*361 was heat killed at 80°C for 15 min.
2. Linearized pBAC10:KO₁₆₂₉ was combined with undigested pJB1-MAPK1 and used to transfect SF21 cells according to standard protocols.
3. 500 ng of linearized pBAC10:KO₁₆₂₉ was combined with 500 ng undigested pJB30-MAPK1 and the total volume made up to 12 µl with water.
4. 12 µl Lipofectin (diluted 2:1 in H₂O) was then added to this DNA mix and the tube incubated at room temperature for 30 min.
5. 1 ml serum-free media (InsectXpress) was added to the lipofectin/DNA mixture.
6. A six-well plate containing 1 × 10⁶ SF21 cells/well was prepared and incubated at 27°C for 1 h to allow the cells to adhere.
7. Excess media was aspirated from the SF21 cells and replaced with the lipofectin/DNA/serum-free mix.
8. The transfected cells were incubated at 27°C overnight.

9. The media was then replaced with 2 ml InsectXpress media supplemented with 2% FBS and incubated at 27°C without agitation for a further 72 h.
10. Cells were resuspended by physical agitation and then pelleted by centrifugation at $1,000 \times g$ for 10 min.
11. The supernatant containing recombinant baculovirus was transferred to a fresh tube and stored at 4°C; this is the P₀ stock.

3.3. Amplification of Recombinant Baculovirus

Recombinant baculovirus particles were amplified according to standard procedures. Briefly:

1. A six-well plate was set up with 1×10^6 SF21 cells/well and incubated at 27°C for 1 h.
2. Excess media was removed and replaced with 500 µl of P₀ virus plus 500 µl InsectXpress media supplemented with 2% FBS and incubated at 27°C without agitation for a further 72 h.
3. P₁ virus was harvested as described above.
4. A 150 ml tissue culture flask was seeded with 20 ml of 1×10^6 SF21 cells/ml and incubated at 27°C for 1 h.
5. Excess media was removed and replaced with 500 µl of P₀ virus plus 3 ml InsectXpress media supplemented with 2% FBS and incubated at 27°C for 1 h, after which a further 25 ml InsectXpress media supplemented with 2% FBS were added and cells incubated without agitation for 72 h.
6. P₂ virus was harvested as described above.
7. The titre of the P₂ viral stock was determined by a SybrGreen-based quantitative PCR assay vs. a stock of known titre determined by plaque assay. The titre of the P₂ stock should be ca. 10^7 pfu/ml.

3.4. Protein Expression and Extraction

1. Set up a 24-well deep well plate containing 6×10^6 SF21 cells/well suspended in 3 ml InsectXpress media supplemented with 2% FBS and 50 µM biotin (see Note 3).
2. Add 200 µl of P₂ virus and incubate at 27°C for 72 h with agitation.
3. Harvest cells by centrifugation of the 24-well deep well plate prior to lysis.
4. Gently resuspend the cells in 3 ml of PBS buffer, recentrifuge the plate and discard the supernatant; repeat three times in total.
5. Gently resuspend the pellets in 350 µl of Freezing buffer ensuring thorough mixing of the cells.
6. Aliquot the cells in 50 µl volumes and store at -80°C until required for cell lysis.

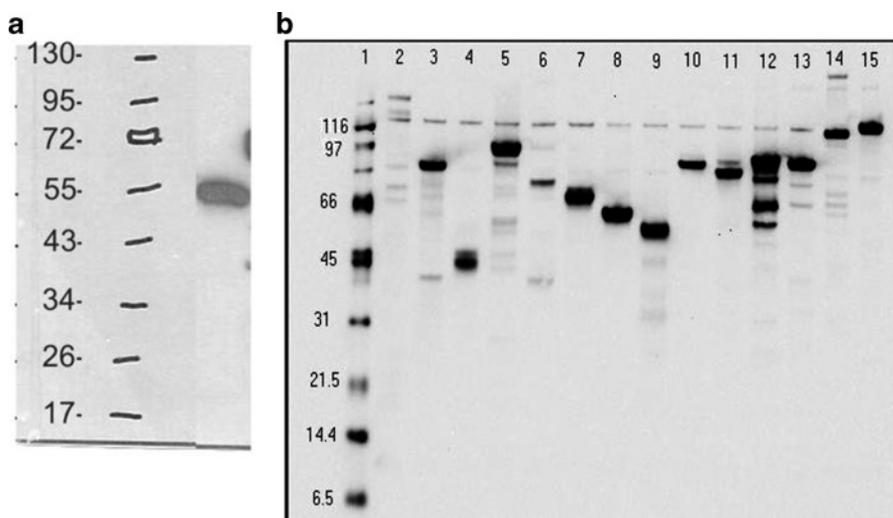


Fig. 3. Western blots of BCCP fusion proteins expressed in SF21 cells. (a) MAPK1-BCCP, encoded by pJB1-MAPK1. (b) (1) Biotinylated marker, (2) ZNF198-BCCP, (3) FUS-BCCP, (4) SDHB-BCCP, (5) STAT4-BCCP, (6) FH-BCCP, (7) MUTYH-BCCP, (8) PNTL1-BCCP, (9) GATA1-BCCP, (10) NF2-BCCP, (11) FACL6-BCCP, (12) MSF-BCCP, (13) MSN-BCCP, (14) RAB5EP-BCCP, (15) Control. All westerns were of crude insect cell lysates and were developed using a streptavidin–HRP conjugate. Good expression levels can be observed across a range of unrelated proteins, some in excess of 100 kDa.

7. Thaw a 50 μ l aliquot and add 50 μ l Lysis buffer plus 10 U Benzonsase (Pierce) and shake on ice for 30 min.
8. Remove cell debris by centrifugation, collect the supernatant and store on ice for up to 24 h before printing.
9. Determine the protein concentration of the soluble, crude protein extract using the Bradford assay (25) to confirm that effective cell lysis has occurred (see Note 4).
10. Determine the approximate expression level of soluble BCCP fusion by SDS-PAGE together with Western blot analysis (20) using a streptavidin–HRP conjugate (Fig. 3; see Note 5).
11. To determine the extent of biotinylation of the BCCP fusion protein, carry out a supershift western blot assay (with an anti-c-Myc antibody) in which equivalent crude lysate samples are pre-incubated with or without streptavidin (0.1 g/ml) (Fig. 4; see Note 6).

3.5. Multiplexed Cloning and Expression of Mammalian Proteins

The procedures described above (Subheadings 3.1–3.4) can clearly be applied to any mammalian cDNA. Moreover, we and others have found that the ligation-independent cloning methods, as well as insect cell transfection, baculovirus amplification, and protein expression and extraction are all amenable to multiplexing through use of appropriate multi-well plate formats. We have observed that by use of the approach described here, we can easily achieve a ~70% success rate from starting cDNA through to expressed, folded, biotinylated protein suitable for array fabrication. Interestingly, the ~30%

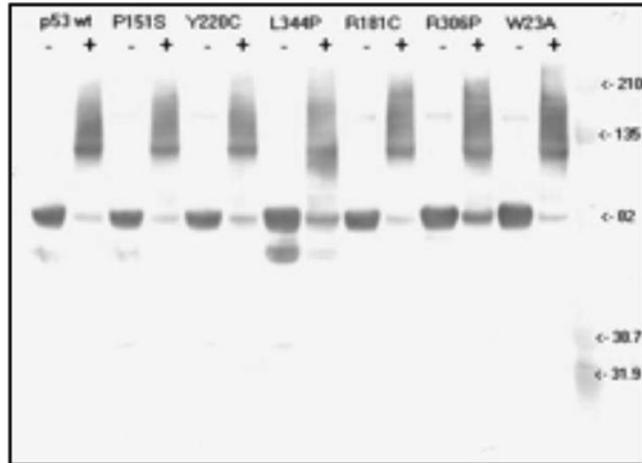


Fig. 4. Supershift assay to determine extent of biotinylation. In a supershift assay, each sample is preincubated either with (“+”) or without (“-”) streptavidin prior to separation by SDS-PAGE. The extent of biotinylation can be estimated by comparison of “+” and “-” samples. Here we show examples using wild-type and variant p53-BCCP fusion proteins.

overall drop-out rate we observe lies almost entirely in the cloning steps and we have observed a >95% success rate in progressing from sequence-verified transfer vector through to expression of a folded, biotinylated, arrayable human protein in insect cells; this compares favourably with the much lower success rates observed when attempting to express mammalian proteins in *E. coli* (26). Using these methods, we have thus been able to assemble collections of hundreds of expressed human proteins, including hundreds of human protein kinases, in a form ready for array fabrication in just a few months and at low cost. In the case of peripheral membrane proteins, such as human cytochrome P450s or human cytochrome P450 reductase, we typically remove any N-terminal hydrophobic anchors during cloning in order to facilitate expression in the soluble fraction.

3.6. Fabrication of Protein Microarrays

In the procedures described below, we do not employ a pre-purification step prior to array fabrication but instead rely on a rapid, single-step immobilization and purification procedure to create arrays of biotinylated BCCP fusion proteins (Fig. 5; see Notes 1 and 2).

3.6.1. Preparation of Source Plates for Printing

1. Transfer 40 μ l of crude protein extract for each BCCP-tagged protein to be arrayed into individual wells of a 384-well V-bottom plate and keep at 4°C. This is the source plate for the print runs.
2. Centrifuge the 384-well plate at 4,000 $\times g$ for 2 min at 4°C to pellet any cell debris that has carried over. Store plate on ice prior to print run.

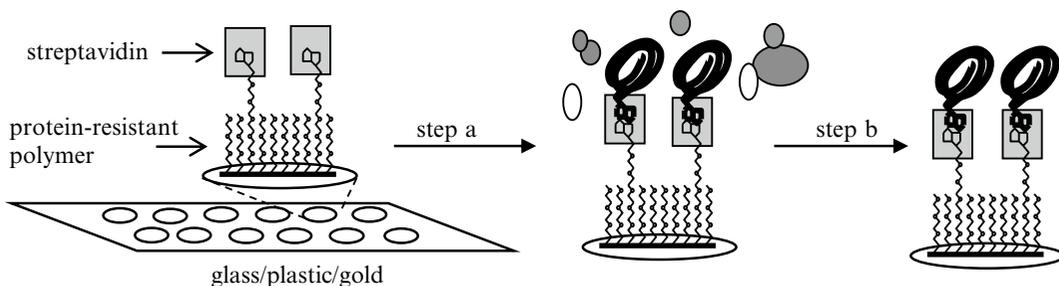


Fig. 5. Schematic of single step immobilisation/purification route to array fabrication. The array surface is intrinsically “non-stick” with respect to proteinaceous material but has a high affinity and specificity for biotinylated proteins. Crude cellular lysates containing the recombinant biotinylated proteins can then be printed onto the surface in a defined array pattern (*step a*) and all non-biotinylated proteins removed by washing (*step b*), leaving the recombinant proteins purified and specifically immobilised via the affinity tag in a single step.

3.6.2. Preparation of Streptavidin-Coated Slides for Printing

1. Equilibrate a Nexterion Slide P microarray slide to room temperature and remove from the foil package (see Note 2).
2. Make up a 1 mg/ml streptavidin solution in 150 mM Na_2HPO_4 buffer (pH 8.5).
3. Place a glass microarray “lifterslip” over the microarray surface and pipette 60 μl of the streptavidin solution along the edge of the lifterslip such that the solution is drawn under the cover slip uniformly by capillary action (see Note 2).
4. Leave for 1 h at room temperature in a humidified chamber.
5. Remove the lifterslip and wash the slide for 1 h at RT in 10 ml 150 mM Na_2HPO_4 buffer (pH 8.5) containing 50 mM ethanolamine to deactivate any remaining amine-reactive groups.
6. Wash the slide for 3 \times 5 min in 10 ml wash buffer and then for 5 min in 10 ml water.
7. Place the slide in a 50 ml Falcon tube and centrifuge at 1,000 $\times g$ for 5 min at 20°C to spin dry. Streptavidin-coated slides were placed into slide boxes, sealed in Ziploc bags and stored at -20°C.

3.6.3. Fabrication of Arrays

In general, any microarray printer could be used to print the arrays and the printing procedures described below can typically be carried out at room temperature, preferably with the atmosphere in the print chamber humidified to ca. 50% to reduce evaporation rates. If necessary, either the print bed or the entire the printing device could be cooled to 4°C; however, we have typically found that this provides little benefit in terms of protein activity on the resultant arrays. Here we describe one specific set of parameters which work well on streptavidin-coated glass microarray slides at room temperature using a Genetix QArray II robot equipped with 16 \times 300 mm tipped solid pins (Figs. 6 and 7; see Notes 7–12):

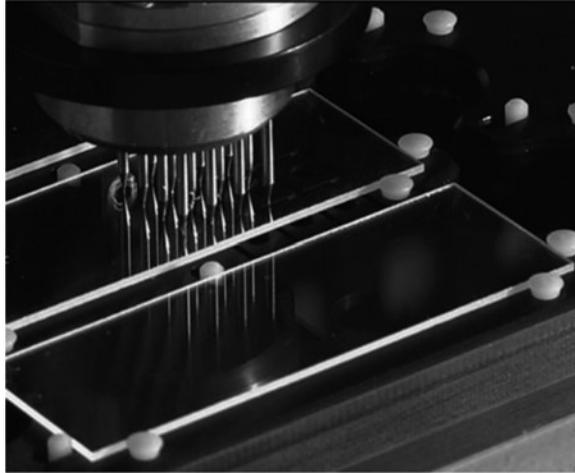


Fig. 6. Printing onto a streptavidin-coated glass microscope slide using 16 solid pins.

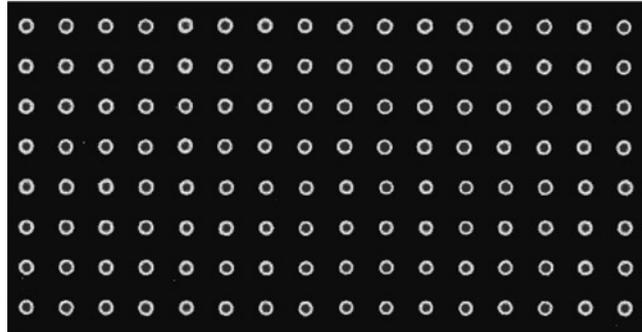


Fig. 7. Highly reproducible arrays with low intra- and inter-array variabilities are key to accurate downstream assays. Here a single biotinylated protein was printed multiple times using all pins in a 16-pin printing head. The amount of bound protein and spot morphology was determined by imaging the array after binding of a Cy3-labelled anti-His antibody to the immobilised proteins on the array.

1. Load the 384-well source plate into the QArray II.
2. Load the streptavidin-coated slides onto the print bed of the QArray II.

Print the arrays using the following key QArray II settings:

Inking Time (ms) = 500

Microarraying pattern = 7×7 , 500 μm spacing

Max stamps per ink = 1

No. stamps per spot = 2

Printing depth = 150 μm

Water washes = 60 s wash, 0 s dry

Ethanol wash = 10 s wash, 1 s dry

3.6.4. Post-printing Processing of Arrays

1. Remove slides from microarraying robot.
2. Wash the slides for 30 min in Falcon tubes containing 50 ml Wash buffer supplemented with 100 μ M biotin, 0.1 mg/ml BSA (see Note 13).
3. Store in Wash buffer at 4°C until use.

3.7. Array-Based Assays for Protein Activity

Once microarrays containing sets of folded, immobilised proteins have been physically fabricated, a wide range of different, systematic and quantitative assays can in principle be carried out on replica arrays, *inter alia*: assays of protein–protein interaction, protein–nucleic acid interaction, protein–small molecule binding, protein–lipid binding, and even enzymatic turnover reactions. Through careful experimental design, it is therefore possible to dissect the properties of diverse collections of unrelated proteins, or of specific families of proteins, in order to gain a greater understanding of, amongst other things, substrate and inhibitor selectivity. The limiting factor here remains the ability to devise an assay that is readable in a microarray format for each of the individual proteins arrayed in parallel. Here we briefly exemplify such assays by reference to arrays of protein kinases.

3.7.1. Verifying that the Individual BCCP-Tagged Proteins Have Indeed Become Immobilised

Following standard protocols for western blots, it is possible for example to probe a replica array with an anti-cMyc antibody followed by a secondary antibody–HRP conjugate, as follows:

1. Dilute a mouse anti-cMyc antibody 1:1,000 in 1 ml milk/PBST.
2. Remove the protein array from wash buffer and equilibrate in PBST at room temperature for 5 min.
3. Drain away the PBST, add 5 ml antibody solution to the array and incubate with gentle agitation at room temperature for 30 min.
4. Wash the array for 3 \times 5 min with 1 ml of PBST.
5. Dilute a goat anti-mouse antibody–HRP conjugate 1:1,000 in 1 ml milk/PBST.
6. Add the antibody solution to the array and incubate with gentle agitation at room temperature for 30 min.
7. Wash the array for 3 \times 5 min with 1 ml of PBST.
8. Add 1 ml chemiluminescent detection reagent (Pierce) to the array.
9. After 1 min, remove slide to a 50 ml Falcon tube and centrifuge for 30 s to dry.
10. In a dark room, place the array against autoradiography film for varying lengths of time before developing the film (Fig. 8).

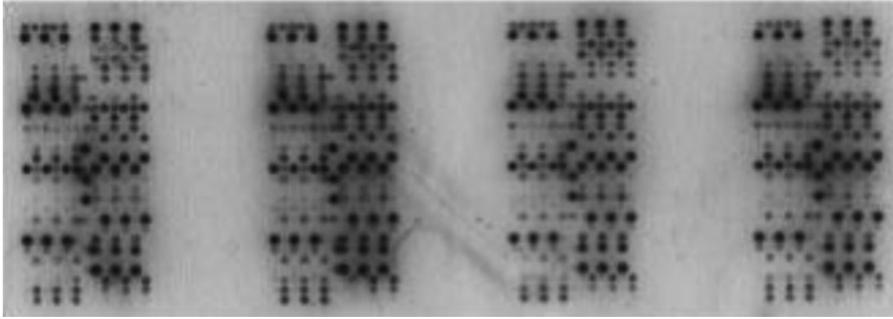


Fig. 8. Array-based qualitative analysis of immobilisation efficiencies. Here an array of 48 diverse BCCP-tagged human proteins was printed in triplicate, with 4 replica arrays per slide. Western blot-style analysis, probing with an anti-c-myc antibody, reveals that all proteins are immobilised as expected; apparent gaps on the array are due to the printing pattern used here.

3.7.2. Measuring the Extent of Autophosphorylation of Each Arrayed Protein Kinase

A subset of protein kinases will undergo autophosphorylation at elevated ATP concentrations; for those that do this, a test, based on a simple autophosphorylation assay, to demonstrate whether the arrayed kinases are active or not, can be carried out as follows:

1. Dilute a Cy3-labelled broad-specificity anti-phosphotyrosine antibody (Cell Signalling) 1:200 in 2 ml milk/PBST.
2. Remove two replica protein kinase arrays from wash buffer and equilibrate in PBST at room temperature for 5 min.
3. Drain away the PBST, add 1 ml of kinase buffer to the first array and 1 ml kinase buffer supplemented with 100 μ M ATP to the second array. Incubate both arrays at room temperature for 1 h.
4. Wash each array for 3 \times 5 min with 1 ml of PBST containing 0.1% SDS.
5. Drain away the PBST/SDS solution, add 1 ml antibody solution to each array and incubate with gentle agitation at room temperature for 30 min.
6. Wash the arrays for 3 \times 5 min with 1 ml each of PBST.
7. Remove the arrays to a 50 ml Falcon tube and centrifuge for 30 s to dry.
8. Scan the arrays at using a DNA microarray scanner (λ_{ex} = 550 nm; λ_{em} = 570 nm) and process the data using a DNA microarray data analysis software package (Fig. 9; see Note 14).

3.7.3. Phosphorylation of Arrayed Protein Kinases by Solution Phase FES Kinase

One use of protein kinase arrays is to determine the identity of the subset of the kinome that are substrates for a given solution-phase kinase. This is performed as follows:

1. Remove two replica protein kinase arrays from wash buffer and equilibrate in PBST at room temperature for 5 min.

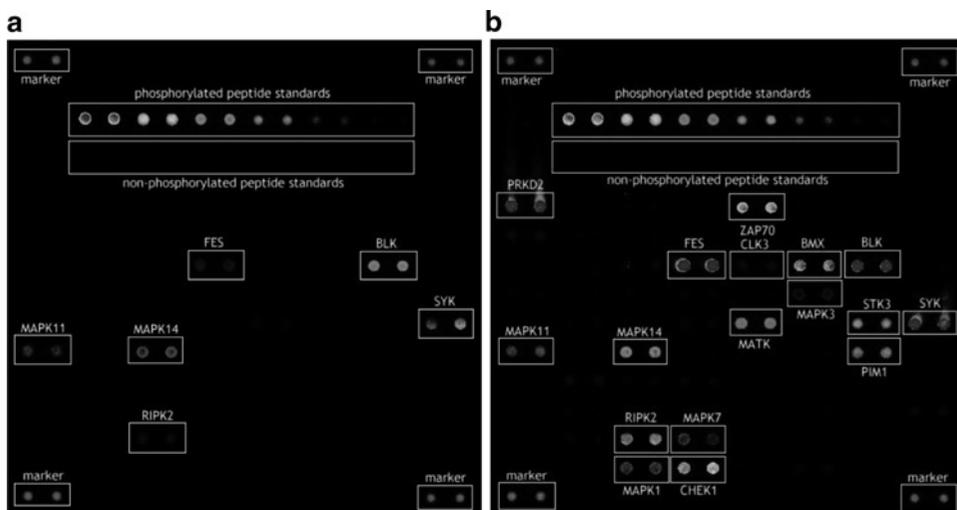


Fig. 9. Array-based autophosphorylation assays. Here an array of 80 human protein kinases was printed in duplicate and assayed for autophosphorylation activity. (a) Kinase buffer only. (b) Kinase buffer containing 100 μM ATP. The assays were developed using a fluorescently labelled anti-phosphotyrosine antibody and revealed ATP-dependent, on-array autophosphorylation for 17 human kinases, as marked.

2. Drain away the PBST, add 1 ml of kinase buffer supplemented with 100 μM ATP to each array.
3. To one of the arrays, also add 50 nM FES kinase (Upstate) and incubate both arrays at room temperature for 1 h.
4. Wash each array for 3×5 min with 1 ml of PBST containing 0.1% SDS.
5. Drain away the PBST, add 1 ml Cy3-labelled anti-phosphotyrosine antibody solution (from Subheading 3.7.2, step 1) to each array and incubate with gentle agitation at room temperature for 30 min.
6. Wash the arrays for 3×5 min with 1 ml of PBST each.
7. Remove the arrays to a 50 ml Falcon tube and centrifuge for 30 s to dry.
8. Scan the arrays using a DNA microarray scanner ($\lambda_{\text{ex}} = 550 \text{ nm}$; $\lambda_{\text{em}} = 570 \text{ nm}$) and process the data using a DNA microarray data analysis software package (Fig. 10; see Note 15).

3.7.4. Measuring the Binding of an Unlabelled Kinase Inhibitor to Each Arrayed Protein Kinase

Protein kinases are currently regarded as tractable drug targets in many different disease states. Most current drug-like kinase inhibitors target the ATP-binding pocket to a greater or lesser degree, yet data from structural biology suggests that the ATP-binding pocket of protein kinases is strongly conserved. Cross-reactivity of protein kinase inhibitors is therefore likely to be an issue, so low cost, simple, quantitative, yet high throughput approaches to assess the selectivity of such inhibitors across large panels of human

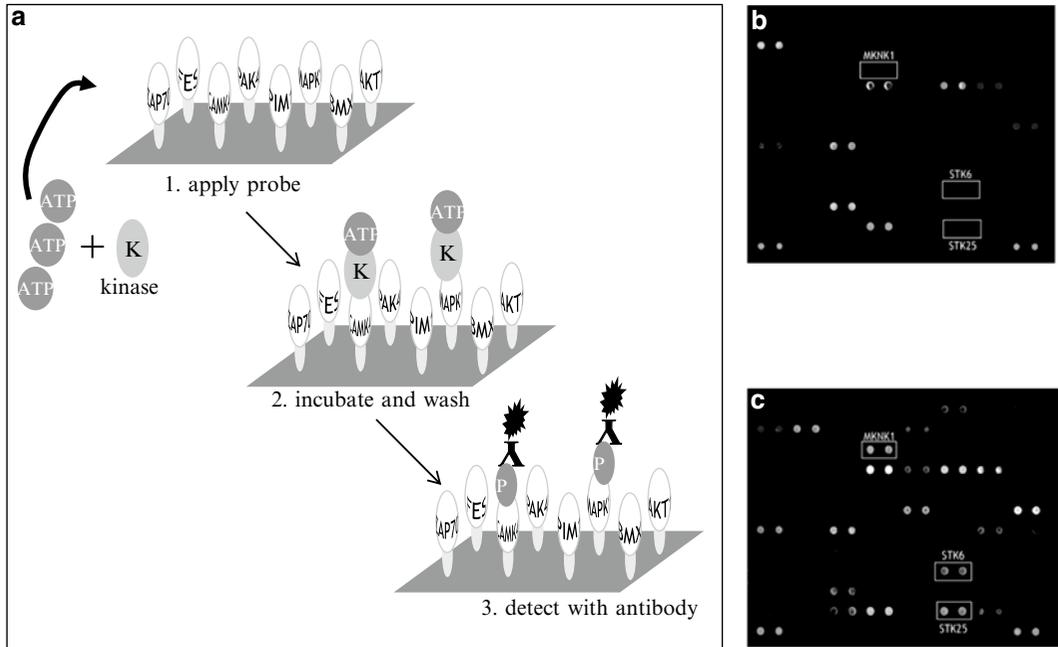


Fig. 10. Array-based phosphorylation assay using an exogenous kinase. An array of 96 human protein kinases was printed in duplicate and assayed in the presence and absence of exogenous FES kinase. (a) Schematic of assay. (b) 100 μ M ATP plus kinase buffer only. (c) 100 μ M ATP, kinase buffer plus exogenous FES kinase. The assays were developed using a fluorescently labelled anti-phosphotyrosine antibody and revealed a number of substrates for FES kinase, including MKNK1, STK6, and STK25, as marked.

kinases have considerable potential. However, an obvious limiting factor is that any chemical modification of a protein kinase inhibitor that incorporates a colorimetric or fluorescent label will potentially interfere with the structure activity relationship of the compound; on the other hand, incorporation of a radiolabel requires dedicated, de novo synthesis, so may be undesirable on cost grounds. Furthermore, a generic, activity-based assay for multiple different kinases remains elusive in an array-based format. As an alternative therefore, protein kinase arrays can be used to provide such binding affinity data on unlabelled kinase inhibitors by setting up competitive assays in which the unlabelled kinase inhibitor and a labelled ATP analog (or similar) are able to bind to the same site (typically the active site) on each arrayed kinase. Such an assay has two stages: first, the binding affinity of a labelled compound (e.g. a broad-spectrum kinase inhibitor) to each arrayed kinase is determined; second, the IC_{50} value for the displacement of the labelled compound by the unlabelled kinase inhibitor is determined, from which data the K_i for the unlabelled kinase inhibitor can be calculated. An example of such an assay is as follows.

3.7.5. Measuring the Binding of a Fluorescently Labelled Compound to Each Arrayed Protein Kinase

1. Remove eight replica protein kinase arrays from wash buffer and equilibrate in PBST at room temperature for 5 min.
2. Drain away the PBST, add 1 ml of kinase buffer to each array.
3. To each array, add an increasing concentration of a fluorescently labelled broad-spectrum kinase inhibitor [e.g. Cy3-labelled staurosporine; a fluorescent ATP analog; or fluorescently labelled polycyclic heteroaromatic compounds (27); concentration range from 0.5 to 50 nM] and incubate all arrays at room temperature for 1 h.
4. Wash each array for 3×5 min with 1 ml of PBST containing 0.1% SDS.
5. Remove the arrays to a 50 ml Falcon tube and centrifuge for 30 s to dry.
6. Scan the arrays using a DNA microarray scanner set at appropriate excitation and emission wavelengths and process the data using a DNA microarray data analysis software package to determine the K_d for the binding of the fluorescent ligand to each arrayed kinase.

3.7.6. Measuring the Binding of an Unlabelled Kinase Inhibitor to Each Arrayed Protein Kinase

1. Remove a further eight replica protein kinase arrays from wash buffer and equilibrate in PBST at room temperature for 5 min.
2. Drain away the PBST, add 1 ml of kinase buffer to each array.
3. To each array, add fluorescently labelled broad-spectrum kinase inhibitor (10 nM final concentration) plus an increasing concentration of an unlabelled kinase inhibitor (e.g. Gefitinib; concentration range from 0.5 to 50 nM) and incubate all arrays at room temperature for 1 h.
4. Wash each array for 3×5 min with 1 ml of PBST containing 0.1% SDS.
5. Remove the arrays to a 50 ml Falcon tube and centrifuge for 30 s to dry.
6. Scan the arrays at 550 nm (em) using a DNA microarray scanner and process the data using a DNA microarray data analysis software package to determine the IC_{50} for the binding of the unlabelled inhibitor to each arrayed kinase (Figs. 11 and 12).

NB. From the IC_{50} and K_d values, it is possible to derive the K_i value for the unlabelled inhibitor binding to each kinase by standard manipulations (see Subheading 3.8.2).

3.7.7. Measuring the Turnover of a Fluorogenic CYP450 Substrate by an Arrayed CYP450:POR Complex

Human cytochrome P450 (CYP450) enzymes are responsible for the primary metabolism of most prescription drugs (see Note 16) and are known to be highly polymorphic in the population (28). Single amino acid substitutions in CYP enzymes are well known to affect rates of oxidation, but can also affect substrate specificity, as

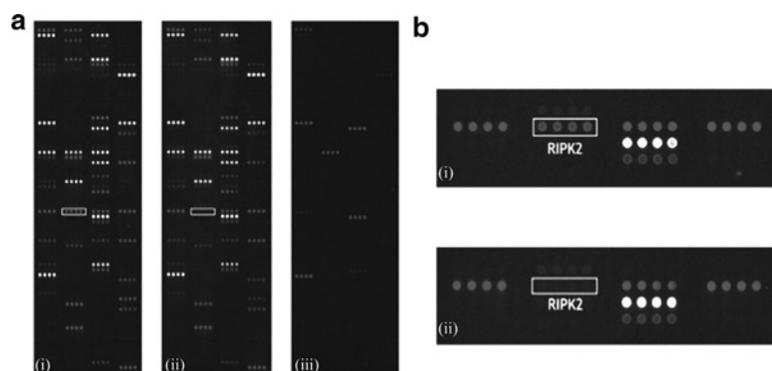


Fig. 11. Array-based inhibitor binding assays. An array of 150 human protein kinases was printed in quadruplicate, one array per slide, and assayed for binding to a universal fluorescent kinase ligand in the presence and absence of small molecule kinase inhibitors. (a) (1) Universal fluorescent ligand only; (2) Universal fluorescent ligand plus Iressa, a specific kinase inhibitor; (3) Universal fluorescent ligand plus Staurosporine, a broad-spectrum kinase inhibitor. (b) RIPK2 was identified as a target for Iressa because the drug successfully competed with the universal fluorescent ligand for binding to this kinase. Note that the specific “universal” fluorescent ligand used in this assay only bound to 48 out of 150 kinases on the array; by using combinations of such ligands, higher coverage can be achieved.

well as regio- and enantio-selectivity of oxidation (29), so can in principle lead to the formation of entirely different products. However, data are scarce regarding the effect of most polymorphisms on most drugs or drug candidates. CYP450 enzyme arrays can be used to provide such data as follows (see Note 17):

1. Transform *E. coli* DH5 α cells with expression plasmids encoding either CYP3A4[Δ 1-24]-BCCP (henceforth “CYP3A4-BCCP”) or POR[Δ 1-43]-BCCP (henceforth “POR-BCCP”) fusion proteins under the control of an IPTG-inducible promoter.
2. Grow DH5 α {CYP3A4-BCCP} and DH5 α {POR-BCCP} cells separately in Terrific Broth at 30°C, 150 rpm, supplementing the growth media in the former case with δ -aminolevulinic acid (0.5 mM) and in both cases with free biotin (50 μ M). Induce protein expression in both cases at mid-log phase with 1 mM IPTG and harvest cells by centrifugation after 18 h.
3. Prepare crude lysates of each *E. coli* culture by sonication in P450 Printing buffer and determine the relative expression levels of CYP3A4-BCCP and POR-BCCP by Western blotting.
4. Mix together the two crude lysates in proportions that create a roughly equimolar mixture of CYP3A4-BCCP and POR-BCCP proteins.
5. Fabricate CYP3A4:POR enzyme microarrays as described in Subheadings 3.6.1–3.6.4 above. Use the arrays immediately.

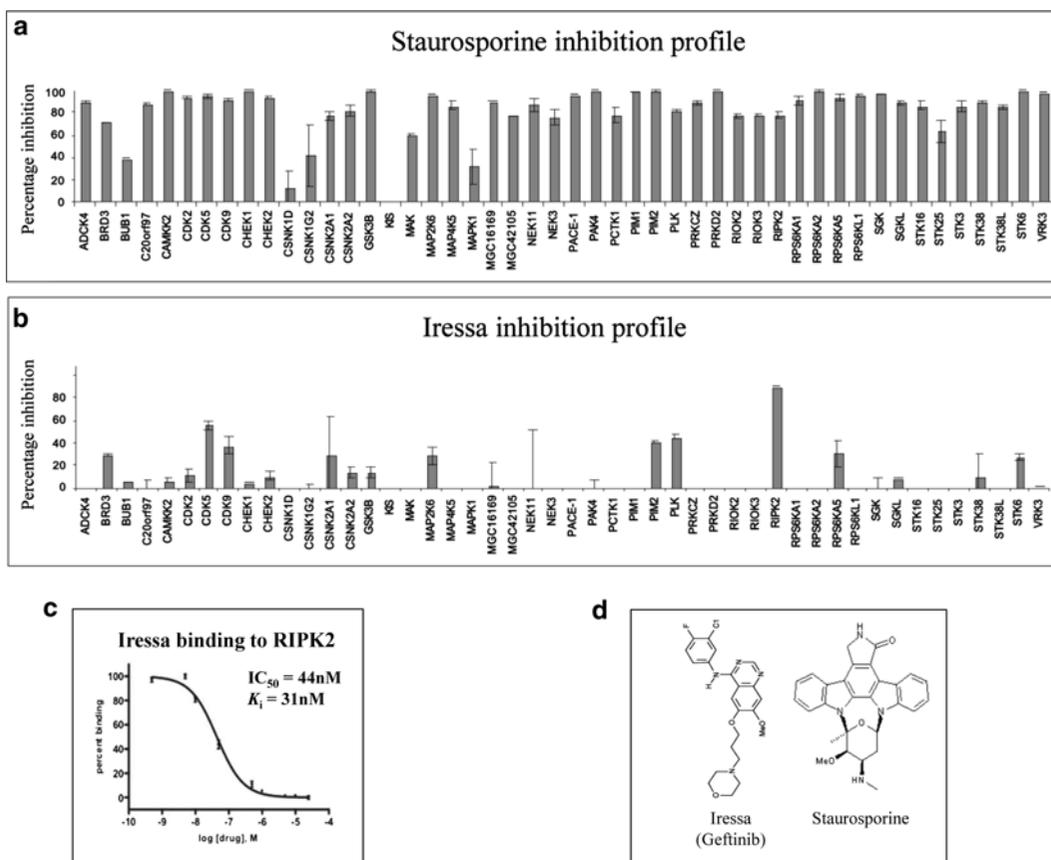


Fig. 12. Data analysis. From the primary data in Fig. 11, the ability of staurosporine and Iressa to compete with the universal fluorescent ligand for binding to each arrayed kinase can be quantified. (a) Percentage inhibition of arrayed kinases by 1 μM unlabelled staurosporine. (b) Percentage inhibition of arrayed kinases by 1 μM unlabelled Iressa. (c) Repeating the assay at different concentrations of Iressa enabled an IC_{50} value of 44 nM to be determined for the Iressa/RIPK2 interaction, in good agreement with literature values. From the IC_{50} value, we calculated that K_i (Iressa/RIPK2) = 31 nM. (d) Structures of Iressa and Staurosporine, which both bind in the ATP-binding pocket of their target kinases. As expected, at 1 μM concentration, Staurosporine competed effectively with the universal fluorescent ligand across the majority of arrayed kinases, whilst Iressa was much more selective in its binding profile.

6. In a 384-well V-bottom microtitre plate, prepare substrate & cofactor mixtures containing varying concentrations of Vivid Red (500 nM–43.25 μM) and a fixed concentration of NADPH (1 mM) in P450 Assay buffer. Also prepare a dilution series of resorufin (1–500 nM) in the same plate.
7. Place the CYP3A4:POR enzyme microarray back onto the arrayer print bed and print a single droplet of each different substrate/cofactor mixture on top of a different immobilised CYP3A4:POR enzyme complex. Also print the resorufin dilution series onto an unused area of the microarray.
8. Carefully remove the microarray from the printer and transfer to a DNA microarray scanner. Monitor product formation in

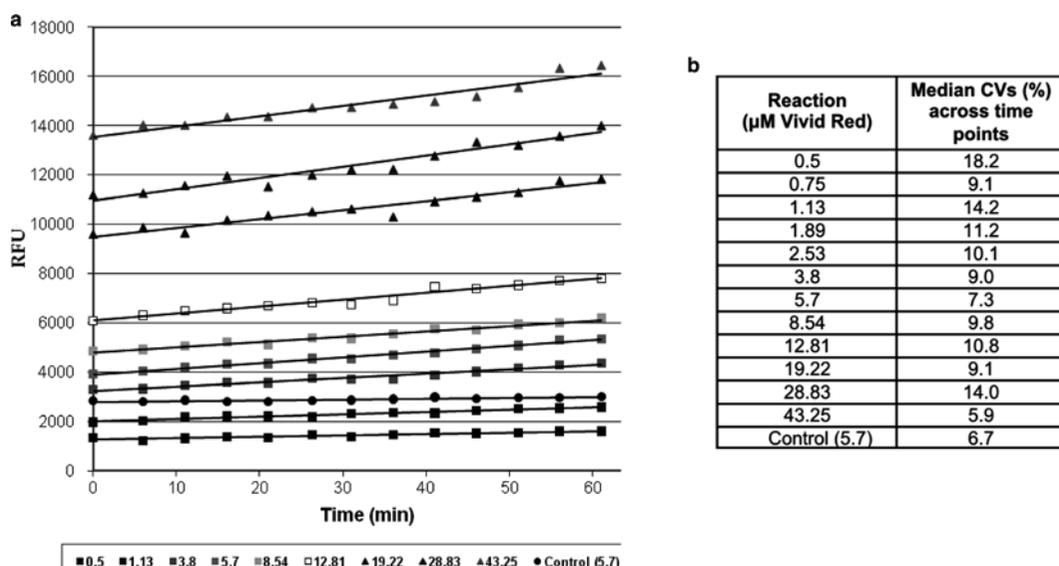


Fig. 13. Array-based cytochrome P450 turnover assays. Here an array of replica spots of a CYP3A4-BCCP:POR-BCCP complex was assayed at varying substrate concentrations, encapsulating each individual enzymatic reaction in a nano-droplet sitting above an individual immobilised protein spot on the array. Product formation was monitored in real time in a microarray scanner, acquiring data every 5 min for an hour; all reaction rates remained linear over the assay period. (a) Velocity (RFU vs. time) graphs for the turnover of Vivid Red (500 nM–43.25 μM) on a CYP3A4:POR microarray. (b) Median CV values for each of the reactions across all 13 consecutive time points. N.B. Each data point represents the average RFU of 18 replicate assays on the same microarray slide. Error bars are omitted from the graph for clarity. Data does not extrapolate to zero due to intrinsic fluorescence of the fluorogenic substrate, Vivid Red.

each droplet as a function of time by repeatedly scanning the microarray ($\lambda_{\text{ex}} = 532 \text{ nm}$; $\lambda_{\text{em}} = 585 \text{ nm}$) at 20 μm resolution every 5 min for 60 min (Fig. 13).

9. Extract the data using a DNA microarray data analysis software package and analyse the initial velocity data by non-linear regression using a biphasic multisite model in order to determine k_{cat} and K_{m} for the oxidation of Vivid Red by CYP3A4 (Fig. 14). N.B. Convert the initial velocities from RFU min^{-1} to nM (resorufin) min^{-1} using a standard curve generated from the resorufin dilution series.

3.7.8. Measuring the Binding of an Unlabelled CYP450 Inhibitor to an Arrayed CYP450:POR Complex

Having established the means to monitor the catalytic turnover of a fluorogenic substrate by a CYP450 enzyme in a microarray format, it is then relatively straightforward to set up competitive assays to determine the kinetic parameters for either unlabelled inhibitors or unlabelled substrates as follows (see Note 18):

1. Prepare a CYP3A4:POR microarray as described in Subheading 3.7.7 above.
2. In a 384-well V-bottom microtitre plate, prepare fluorogenic substrate, cofactor and test compound mixtures in P450

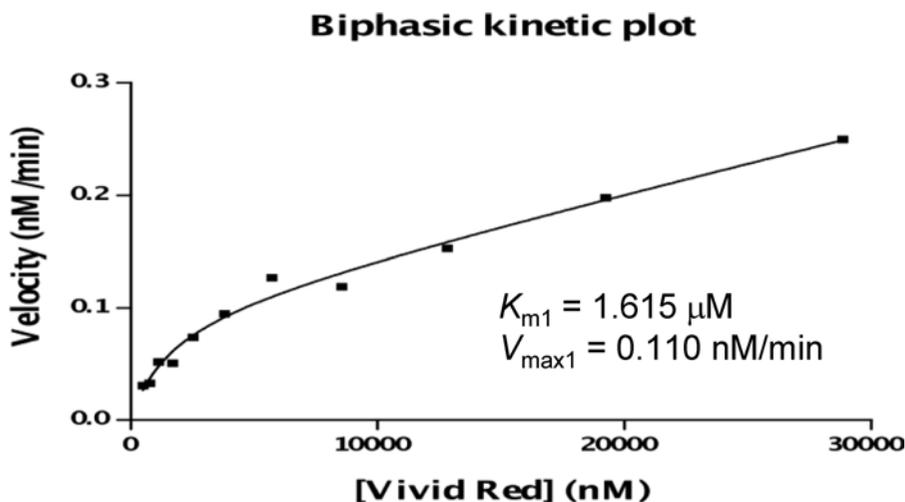


Fig. 14. A biphasic plot to determine k_{cat} and K_m for the oxidation of Vivid Red by CYP3A4 immobilised on a microarray surface. The microarray-generated kinetic data set for the turnover of Vivid Red by a CYP3A4-BCCP:POR-BCCP complex (Fig. 13) was fitted to a biphasic, multisite model using non-linear regression. The resultant catalytic parameters, $V_{\text{max}1}$ and K_{m1} , describe the enzymatic activity when only a single molecule of substrate is bound in the active site at any one time; these values conform well with literature values obtained with baculosome preparations of CYP3A4.

Assay buffer containing varying concentrations of Vivid Red (500 nM–43.25 μM), a fixed concentration of NADPH (1 mM) plus a dilution series of the unlabeled test compound (e.g. ketoconazole; concentration range from 0.1 nM to 1 mM) at each Vivid Red concentration. Also prepare a dilution series of resorufin (1–500 nM) in the same plate.

- Place the CYP3A4:POR enzyme microarray back onto the arrayer print bed and print a single droplet of each different substrate/cofactor/test compound mixture on top of a different immobilised CYP3A4:POR enzyme complex. Also print the resorufin dilution series onto an unused area of the microarray.
- Carefully remove the microarray from the printer and transfer to a DNA microarray scanner. Monitor product formation in each droplet as a function of time by repeatedly scanning the microarray ($\lambda_{\text{ex}} = 532 \text{ nm}$; $\lambda_{\text{em}} = 585 \text{ nm}$) at 20 μm resolution every 5 min for 60 min.
- Extract the data using a DNA microarray data analysis software package and analyse the initial velocity data according to standard competitive inhibitor enzyme kinetics using a biphasic multisite model in order to determine the k_i for the binding of the unlabelled test compound to CYP3A4.

N.B. Convert the initial velocities from RFU min^{-1} to $\text{nM (resorufin) min}^{-1}$ using a standard curve generated from the resorufin dilution series.

3.8. Data Analysis: General Principles

3.8.1. Raw Data Extraction

There are in principle numerous different ways in which the raw data can be pre-processed prior to analysis, depending on the precise microarray scanner software package used. We find that the most reliable method is as follows:

1. Use the highest gain setting on the microarray scanner that does not cause any signals to saturate.
2. Set a “pixel inclusion” threshold to exclude any dark pixels (i.e. pixels that strictly form part of the background) from within the analysis area.
3. Extract the raw data and determine the median foreground pixel intensity for each spot, as well as the median local background pixel intensity for each spot.
4. Subtract the median local background pixel intensity from the median foreground pixel intensity for each spot to give the net pixel intensity for each spot.
5. Using the replicate data, determine the mean of the median net pixel intensities for each arrayed protein; this is then the data to use in subsequent analyses.

3.8.2. Data Analysis

Fluorescent ligand-binding data, as well as fluorogenic substrate turnover data, can be analyzed in a number of different ways, including the use of non-linear regression algorithms. Manual analysis of ligand-binding data from arrays can be carried out by use of linear transformation Scatchard-type plots as follows.

Plot the relative amount of bound fluorescent ligand (FL) as a function of “ligand concentration in solution” for each protein in the array (7) and fit to a simple hyperbolic concentration–response curve according to:

$$R = B_{\max} [FL] / \{K_d + [FL]\}$$

where R is the response in relative fluorescence units and $[FL]$ is the ligand concentration. From this, the ligand-binding constants (K_d) and maximum ligand-binding capacities (B_{\max}) for each arrayed protein can be determined (7).

In competition binding assays of the type described above (Subheading 3.7.4), the IC_{50} values for a specific protein–ligand interaction are dependent on the intrinsic K_i of the unlabelled ligand for the protein and on the K_d of the fluorescent ligand competitor, as well as on the concentration of the fluorescent ligand in the assay. The relationship is:

$$K_i = IC_{50} / \left\{ 1 + ([FL] / K_d) \right\}$$

4. Notes

1. In order to avoid laborious pre-purification of each recombinant protein prior to array fabrication, we have developed a procedure in which we combine the immobilisation and purification stages into a single step (7). For this approach to work, the array surface itself must have a low capacity for non-specific binding of proteins, yet must have a high specificity and high binding affinity for the proteins to be arrayed. In principle, a range of different affinity tags could be used here but in practise few actually offer sufficiently high specificity interactions with the surface affinity matrix. For example, His-tags are not well suited to such an approach since the specificity of the interaction with Ni^{2+} ligands is too low (particularly in the context of expression in eukaryotic cells where there are numerous Ni^{2+} -binding proteins) and the intrinsic affinity of the interaction is also relatively low, resulting in leaching of protein from the array even in the absence of imidazole. To circumvent these problems, we have chosen to use an affinity tag that becomes biotinylated *in vivo* at a single specific residue, allowing us to make use of the very high affinity ($K_d \sim 10^{-15}$ M) and specificity of the streptavidin–biotin interaction. Array fabrication thus becomes the simple process of printing crude lysates containing the recombinant biotinylated proteins onto streptavidin-coated surfaces which have a low non-specific protein-binding capacity (see Note 2), followed by washing to remove all non-biotinylated proteins from the array surface (Fig. 5).

In the absence of any pre-purification step, it might be intuitively expected that host cell proteins which are endogenously biotinylated would compete with the biotinylated recombinant proteins for the available streptavidin-binding sites on the array surface. Insect cells have five such endogenous proteins, but under native conditions we have observed that these protein do not compete efficiently with biotinylated recombinant proteins for binding to streptavidin. This perhaps reflects low natural expression levels and the fact that in endogenous biotinylated protein, the biotin is typically less solvent accessible in the native protein (30) meaning that the biotin may not be physically accessible to bind to streptavidin.

We have also observed that using the streptavidin–biotin interaction as the basis for array fabrication confers an additional major advantage: the very high affinity of the streptavidin–biotin interaction means that we quickly start to saturate the available biotin-binding sites on the surface; this means that a crude normalisation of protein loading can be achieved without pre-adjusting the concentrations of the crude lysates to

compensate for differences in the individual expression levels of the different recombinant proteins (7).

As with many affinity tags, biotinylated tags can be positioned at either the N- or C-terminus of the protein to be arrayed, dependent on the structural and functional characteristics of the protein. In the context of oriented protein microarray fabrication, “biochemical” biotinylation seems preferable to “chemical” biotinylation since the latter offers little control over the site of biotinylation and still requires pre-purification of each protein to remove excess biotinylation reagent. There are currently three main alternatives for biochemical introduction of a biotin moiety into a recombinant protein: two involve affinity tags that can be biotinylated *in vivo* or *in vitro* whilst the third involves an intein-mediated introduction of biotinylated cysteine.

The AviTag (Avidity, Colorado, USA) is an *in vitro* synthesized 15 residue peptide that is specifically biotinylated by the *E. coli* biotin ligase (31).

Alternatively, a single biotinylated cysteine moiety can be added to the C-terminus of a recombinant protein during intein-mediated protein splicing of fusion proteins (32). However, it is not entirely clear what advantages this latter route offers over simpler use of the AviTag.

We have chosen, however, to use a compact, folded, biotinylated, ~80 residue domain derived from the *E. coli* biotin carboxyl carrier protein (BCCP; Fig. 1) (21, 22) since this affords two significant advantages over the AviTag and intein-based tag. First, the BCCP domain is cross-recognised by eukaryotic biotin ligases enabling it to be biotinylated efficiently in yeast, insect, and mammalian cells without the need to co-express the *E. coli* biotin ligase (33–35). Second, the N- and C-termini of BCCP are physically separated from the site of biotinylation by ~50 Å (Fig. 1) (21), so the BCCP domain can be thought of as a stalk which presents the recombinant proteins away from the surface, thus minimising any deleterious effects due to immobilisation.

Vectors for expressing proteins as fusions to a BCCP domain derived from *Klebsiella pneumoniae* are now available from Invitrogen (pET104 DEST Bioease). This *K. pneumoniae* domain is highly homologous to the *E. coli* BCCP protein and confers the same properties.

2. We and others have found that organic polymer coatings, such as those based on dextran or polyethylene glycol (PEG) (either in the form of long chain PEGs or of short chain PEGs supported on a self-assembled monolayer), are considerably superior to proteinaceous blocking agents such as bovine serum albumin or powdered milk in reducing the non-specific binding background

in surface-based assays (36); a number of commercially available surfaces conform to this specification. However, most, if not all, pre-coated streptavidin slides must be shipped on dry-ice in order to preserve the integrity of the streptavidin layer, so depending on location, the shipping costs can therefore be substantial. In order to circumvent this cost issue, we have found that we can custom coat amine-reactive slides (for example, the PEG-based, NHS-activated Nexterion Slide P from Schott, which ships at room temperature and therefore at significantly lower cost) with streptavidin at point of use. Importantly, we have found that our home-made slides show as good uniformity of the streptavidin coating as commercial, pre-made streptavidin surfaces (data not shown); furthermore, we found that utilizing an automated hybridization station in an effort to gain greater control over the streptavidin coating process gave no obvious advantage over the manual process described here (data not shown).

3. We observed that addition of free biotin to the growth medium increases the extent of biotinylation of the recombinant BCCP fusion protein. Importantly, we have not found it necessary to also overexpress the *E. coli* biotin ligase in any expression host when using the BCCP tag. The wash step prior to cell lysis is needed to remove free biotin in the media before array fabrication; if the protein is purified before array fabrication, this is then not necessary.
4. When expressing a large number of clones in parallel for array fabrication, the Bradford assay can conveniently be done on all clones in parallel using a microtitre plate format. However, it would be laborious to carry out SDS PAGE analysis on all clones for each and every repeat expression run, so typically we only assess a selection of clones in this way since the absolute expression level is not critical for array fabrication. For the majority of clones, we have found that a simpler dot-blot assay can form a reliable qualitative indicator that biotinylated recombinant protein has been expressed in the specific set of insect cell cultures in question (data not shown).
5. The growth temperatures and times described here have proven to be directly applicable to all BCCP fusions we have sought to express in insect cells, which number in the hundreds. Since our approach to downstream array fabrication does not require pre-purification of recombinant proteins prior to printing, we have not found it necessary to try to maximize expression levels of any one protein. On the contrary, we have found heuristically that as long as we can observe an expressed, biotinylated recombinant protein by western blotting, there will then be enough recombinant protein to array and assay.

6. The biotinylated component of the sample will be supershifted by streptavidin even under denaturing conditions, enabling a simple side-by-side comparison to be carried out. To save time, this assay need not be done on all clones since we have observed that if the recombinant BCCP proteins are expressed and folded, the BCCP domain is efficiently biotinylated.
7. We have found solid pins the easiest to clean rapidly and have observed no carry over between samples using such pins. In addition, we have found that solid pins also work well printing directly onto many different surface types. Furthermore, we have observed that smaller diameter pins also work well. It is also possible now to take advantage of nanofabricated print heads to create printed features with diameters of a few μm .
8. Each spotting event using 300 μm flat-tipped solid pins delivered ~ 10 nl liquid (estimated based on assumption of a spherical droplet of 300 μm diameter sitting on the tip of each pin). We typically use multiple stamps per spot to increase the protein loading at each position in the array and we have found that these general printing parameters work well with other surfaces, although care must be taken in calibrating the z -height and touch-down velocities on the robot when using fragile surfaces. In addition, by using the same printing parameters under the same conditions of printing (i.e. in a glycerol-containing buffer to reduce evaporation rates), pre-purified proteins can be spotted onto non-selective surfaces which bind proteins by chemical crosslinking (e.g. epoxide or aldehyde-coated glass (14, 15)) or by non-covalent adsorption (e.g. supported nitrocellulose, agarose or polyacrylamide (15)).
9. Each 50 μl aliquot of harvested expression cells provides enough recombinant protein to print 25 replica slides in 4-plex format, with each protein printed in triplicate in each sub-array. Thus, one 3 ml baculovirus culture yields enough expressed protein to fabricate 700 replica sub-arrays, or $>2,000$ replica spots of each protein, which is an important consideration when seeking to minimise array-to-array variability. It is important to note though that not all the sample volume loaded into a 96-well V-bottomed source plate can be used for printing because care needs to be taken that the uniformity of the spot size is not affected by change in volume remaining in the source plate.
10. We routinely print each protein in triplicate within each array so that we have three “technical replicates” during downstream data analysis. Furthermore, depending on the desired downstream assay format, it is simple with robotic arrayers to print either 1, 2, 4, 8, or 16 replica sub-arrays on each 7.5×2.5 cm glass slide. Obviously, the higher the number of sub-arrays per slide, the lower the number of discrete protein spots that can be accommodated in each sub-array. By way of example, using

a “4-plex” format, with $8 \times 300 \mu\text{m}$ solid pins it is possible to print eight 8×8 panels (=512 discrete protein spots) in each of the 4 replica sub-arrays with a spot-to-spot spacing of $500 \mu\text{m}$; this would enable, for example, 170 individual protein types to be printed in triplicate in each of the 4 sub-arrays.

11. Importantly, with the ready availability of reusable gasket systems today (e.g. Gentel Biosciences SIMplex™ system; Whatman FAST Frame system; etc.), it is possible now to assay each sub-array on any one slide under independent conditions, thus minimising the impact of the cost of the base slides on the “cost per assay” and “cost per data” point to much more reasonable levels. This capability also allows a number of parallel assays to be carried out on a single slide, thus reducing the possible impact of slide-to-slide variability on experimental error.
12. By printing, processing, and assaying arrays using these protocols, we have been able to achieve spot-to-spot CVs of ca. 10% without recourse to clean rooms (Fig. 7).
13. The blocking and wash steps should remove all non-biotinylated proteins from the array surface whilst the biotin in the milk powder blocks any remaining biotin-binding sites on the streptavidin surface such that any excess biotinylated proteins which have not bound within the specific spot cannot then rebind elsewhere on the array. We have found that under the simple storage conditions described here, our arrays remain viable for around 3 weeks, after which loss of activity of the arrayed proteins starts to be observed. In the absence of proteases or of elevated temperatures during storage, one possible reason for this loss of activity is that over time the arrayed proteins bury through the polyethylene glycol coating and adsorb irreversibly to the underlying glass surface; any such effect would of course be both time- and protein-dependent, but we have not characterised this in any detail.
14. We routinely observe that some human protein kinases expressed in insect cells are already phosphorylated prior to assay. However, many other kinases become newly phosphorylated during this assay. There are two obvious explanations for this: first, the autophosphorylation of these latter kinases might simply require the higher ATP concentration used in the assay (compared to in vivo concentrations), or second, that the autophosphorylation event requires dimerization of the kinases in question and that such dimerization is promoted by the surface attachment in the arrays.
15. It is also possible to carry out the same assay using radioactive ATP if desired (we have found $100 \mu\text{M}$ ATP at 60 Ci/mmol to work well, in which case the detection is by direct autoradiography). This assay format has the immediate advantage of not

being restricted by the specificity of the anti-phospho-peptide antibodies. Furthermore, in such cases, it is also then possible to mask competing autophosphorylation events by carrying out a pre-incubation with 100 μ M cold ATP.

16. During the normal catalytic cycle, human CYP450 enzymes require two electrons to be delivered sequentially by a cytochrome P450 reductase (POR) enzyme in order to reduce molecular oxygen and to generate a high energy iron (IV) oxo species (37). Ordinarily, this requires that the human CYP450 and POR enzymes be co-localised on a membrane in the correct relative three-dimensional geometry for electron transfer to occur, which presents a challenge for protein microarrays constructed using these enzymes. The so-called “peroxide shunt” pathway can be used to circumvent this requirement for the POR enzyme (38) and in principle would simplify the CYP450 assay in a microarray format, but unfortunately the CYP450 enzyme turnover numbers found using the peroxide shunt pathway are typically low, perhaps reflecting inappropriate oxidation and hence inactivation of the CYP450 enzyme itself; there is also a concern that alteration of the timing of delivery of the electrons to the CYP450 enzyme could potentially influence substrate oxidation. Thus, to date, all efforts to miniaturise CYP450 enzyme assays have focused on the use of microsomal or liposomal CYP450:POR preparations and have somehow aimed to immobilize these onto a surface via encapsulation in, for example, an agarose droplet, a sol-gel or an alginate matrix/hydrogel (39–41). In contrast, we have made use of appropriate surface chemistry to enable co-immobilisation of soluble, BCCP-tagged CYP450 and POR enzymes onto a lipid- and membrane-free surface such that POR-dependent CYP450s catalytic turnover reactions can be measured in a microarray format.
17. In seeking to measure enzymatic turnover reactions in a microarray format, an obvious additional challenge arises because, unlike simple ligand-binding assays, the product of an enzymatic reaction is free to diffuse away from the site of reaction. Classically, this problem is solved by encapsulating each assay in a vessel (e.g. a cuvette or a well of a microtitre plate) that contains a single enzyme species. In the case of an enzyme microarray containing many different arrayed enzyme species, our solution is to encapsulate each individual enzymatic reaction in a nanodroplet sitting on top of each arrayed enzyme; any product formed by a specific enzyme is thus restricted to diffusion within the volume of the relevant nanodroplet and at the end of the designated reaction period, the droplet can be dried down, thus fixing the product in the local vicinity of the enzyme responsible for its formation.

18. In principle, this miniaturized competitive assay format for unlabelled enzyme inhibitors or substrates could be used to profile many different drug-like molecules in parallel, or alternatively to determine the binding constants of an individual drug-like molecule by making kinetic measurements at varying concentrations of the test compound.

Acknowledgments

The authors thank Nashied Peton, Sarah Joyce, Colin Wheeler, Jens Koopman, Nick Workman, Steve Parham, and Mike McAndrew for their help in generating the data detailed herein. We thank Procognia Ltd (UK) for provision of human kinase arrays and also thank the Centre for Proteomic & Genomic Research, Cape Town, for access to equipment. JMB thanks the National Research Foundation (NRF; South Africa) for a Research Chair; NBK thanks the NRF for a PhD studentship. The research was supported by grants from the NRF, Procognia Ltd and Genetix PLC (UK).

References

1. Hunter S, Apweiler R et al. (2009) InterPro: the integrative protein signature database (2009). *Nucleic Acids Res.* 37: D224–228
2. Wise EY, Yew WS, Babbitt PC, Gerlt JA, Rayment I (2002) Homologous (b/a)₈-Barrel Enzymes That Catalyze Unrelated Reactions: Orotidine 5'-Monophosphate Decarboxylase and 3-Keto-L-Gulonate 6-Phosphate Decarboxylase. *Biochemistry* 41: 3861–3869
3. Schmidt, DMZ, Mundorff EC, Dojka M, Bermudez E et al (2003) Evolutionary potential of (b/a)₈-Barrels: Functional promiscuity produced by single substitutions in the enolase superfamily. *Biochemistry* 42: 8387–8393
4. The genome international sequencing consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860–921
5. MacBeath, G (2002) Protein microarrays and proteomics. *Nature Genetics* 32: 526–532
6. Wolf-Yadlin A, Sevecka M, MacBeath G (2009) Dissecting protein function and signaling using protein microarrays. *Current Op Chem Biol* 13: 398–405
7. Boutell JM, Hart DJ, Godber BLJ, Kozlowski RZ, Blackburn JM. (2004) Analysis of the effect of clinically-relevant mutations on p53 function using protein microarray technology. *Proteomics* 4: 1950–1958
8. Kodadek T (2001) Protein microarrays: prospects and problems. *Chem Biol* 8:105–115
9. Predki P (2004) Functional protein microarrays: Ripe for discovery. *Curr Opin Chem Biol* 8: 8–13
10. Zhu H, Klemic JF, Chang S, Bertone P et al. (2000) Analysis of yeast protein kinases using protein chips. *Nat Genet* 26: 283–289
11. Zhu H, Bilgin M, Bangham R, Hall D et al. (2001) Global analysis of protein activities using proteome chips. *Science* 293: 2101–2105
12. Michaud GA, Salcius M, Zhou F, Bangham R. et al. (2003) Analyzing antibody specificity with whole proteome microarrays. *Nature Biotechnology* 21: 1509–12
13. Fang Y, Lahiri J, Picard L (2003) G-Protein-coupled receptor microarrays for drug discovery. *Drug Discovery Today* 8: 755–761
14. MacBeath G, Schreiber SL (2000) Printing proteins as microarrays for high-throughput function determination. *Science* 289: 1760–1763
15. Angenendt P, Glokler J, Sobek J, Lehrach H, Cahill DJ (2003) Next generation of protein microarray support materials: Evaluation for protein and antibody microarray applications. *J Chromatogr A* 1009: 97–104

16. Koopmann J-O, Blackburn JM (2003) High Affinity Capture Surface for MALDI compatible Protein Microarrays. *Rapid Communication in Mass Spectrometry* 17: 1–8
17. He M, Taussig M (2001) Single step generation of protein arrays from DNA by cell-free expression and in situ immobilization (PISA method). *Nucleic Acids Res* 29: E73.
18. Ramachandran N, Hainsworth E et al. (2004) Self-Assembling Protein Microarrays. *Science* 305: 86–90
19. Zhao Y, Chapman DAG, Jones IM (2003) Improving baculovirus recombination. *Nucleic Acids Res.*31: e6
20. Sambrook J, MacCallum P, Russell D (2001) *Molecular Cloning, A Laboratory Manual*, Third ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
21. Athappilly FK, Hendrickson WA (1995) Structure of the biotinyl domain of acetyl-coenzymeA carboxylase determined by MAD phasing. *Structure* 3: 1407–19
22. Chapman-Smith A, Cronan JE (1999) The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity. *Trends Biochem Sci* 24: 359–363
23. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18: 2714–2723
24. Yang Y-S, Watson WJ, Tucker PW, Capra JD (1993) Construction of recombinant DNA by exonuclease recession. *Nucleic Acids Res* 21: 1889–1893
25. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248–254
26. Terwilliger TC, Stuart D, Yokoyama S (2009) Lessons from Structural Genomics. *Ann Rev Biophys* 38: 371–383
27. Brown M (2007) Novel fluorescent kinase ligands and assays employing the same. Patent application no. WO2008071937
28. Hernandez-Boussard T, Whirl-Carrillo M, Hebert J.M, et al (2008) The pharmacogenetics and pharmacogenomics database: accentuating the knowledge. *Nucleic Acids Research* 36 (Database issue), D913–D918
29. Ingelman-Sundberg M (2004) Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *TRENDS in Pharmacological Sciences* 25: 193–200
30. Choi-Rhee E, Cronan JE (2003) The biotin carboxylase-biotin carboxyl carrier protein complex of *Escherichia coli* acetyl-CoA carboxylase. *J Biol Chem* 278: 30806–30812
31. Cull MG, Schatz PJ (2000) Biotinylation of proteins in vivo and in vitro using small peptide tags. *Methods Enzymol* 326: 430–40
32. Lue RY, Chen GY, Hu Y, Zhu Q, Yao SQ (2004) Versatile protein biotinylation strategies for potential high-throughput proteomics. *J Am Chem Soc* 126: 1055–62
33. Berliner E, Mahtani HK, Karki S, Chu LF, et al. (1994) Microtubule movement by a biotinylated kinesin bound to streptavidin-coated surface. *J Biol Chem* 269: 8610–5
34. Lerner CG, Saiki AY (1996) Scintillation proximity assay for human DNA topoisomerase I using recombinant biotinyl-fusion protein produced in baculovirus-infected insect cells. *Anal Biochem* 240: 185–96
35. Parrott MB, Barry MA (2001) Metabolic biotinylation of secreted and cell surface proteins from mammalian cells. *Biochem Biophys Res Commun* 281: 993–1000
36. Zheng J, Li L, et al (2005) Strong Repulsive Forces between Protein and Oligo (ethylene glycol) Self-Assembled Monolayers: A Molecular Simulation Study. *Biophys J*. 89: 158–166
37. Makris TM, von Koenig K, Schlichting I, Sligar, SG (2006) The status of high-valent metal oxo complexes in the P450 cytochromes. *Journal of Inorganic Biochemistry* 100: 507–518
38. Pylypenko O, Schlichting I (2004) Structural aspects of ligand binding to and electron transfer in bacterial and fungal P450s. *Annual Reviews in Biochemistry* 73: 991–1018
39. Sakai-Kato K, Kato M, Homma H, Toyo'oko T, Utsunomiya-Tate N (2005) Creation of a P450 array toward high-throughput analysis. *Analytical Chemistry* 77: 7080–7083
40. Lee M, Park CB, Dordick JS, Clark DS (2005) Metabolising enzyme toxicology assay chip (MetaChip) for high-throughput microscale toxicity analyses. *Proceedings of the National Academy of Sciences USA* 102: 983–987
41. Sukumaran SM, Potsaid B, Lee M, Clark DS, Dordick JS (2009) Development of a fluorescence-based, ultra high-throughput screening platform for nanoliter-scale cytochrome P450 microarrays. *Journal of Biomolecular Screening* 14: 668–678

Chapter 11

Glycan Microarrays

Xuezheng Song, Jamie Heimburg-Molinaro, David F. Smith,
and Richard D. Cummings

Abstract

Glycan microarrays are presentations of multiple glycans or glycoconjugates printed on a single slide for screening with glycan-binding proteins (GBPs), which include lectins, antibodies, bacteria, and viruses. Glycans derivatized with functional groups can be immobilized onto appropriately activated glass slides to generate glycan microarrays where each glycan is printed at similar concentrations. Here we describe a method for fluorescently and functionally derivatizing free reducing glycans, printing microarrays, and interrogating the microarrays with GBPs.

Key words: Glycan, Microarray, Lectins, Glycan-binding proteins, Fluorescent labeling

1. Introduction

Functional Glycomics, the systematic study of structures and functions of the total glycome, has become a pivotal theme in glycobiology (1–3). The functions of glycans often are regulated through binding with relevant glycan-binding proteins (GBPs) (4), which include lectins and antibodies. While there have been many traditional methods for the study of protein–glycan interactions, they are usually not high-throughput and only deal with a small set of glycan structures. To achieve the goal of high-throughput Functional Glycomics, glycan microarrays have been developed in the past decade and have quickly become an essential tool for defining protein–glycan interactions (5–10). Glycan microarrays are prepared by immobilizing a large set of glycan structures onto a solid phase platform,

typically an activated glass slide. The immobilization can either be covalent through various chemical methods, or noncovalent through hydrophobic interactions. Since typical glycan structures are neither functionalized nor hydrophobic, they need to be synthesized or derivatized to possess such a modification. For covalent immobilization, many chemical reactions have been exploited to make glycan microarrays, such as thiol-maleimide (11, 12), azide-alkyne (13), and amino-NHS (14) or amino-epoxy (15, 16) reaction systems. For noncovalent immobilization, neoglycolipids have been synthesized and attached to nitrocellulose membranes (6, 17, 18). Once a set of functionalized glycans and corresponding solid surfaces are chosen, the actual preparation of a microarray usually involves a microarray printer, which delivers the glycans from a source plate onto microarray slides through either a contact or non-contact printing mechanism. Once a glycan microarray is prepared, it can be interrogated with GBPs or microorganisms. The GBPs or microorganisms are either directly fluorescently labeled or labeled with a tag that can be detected indirectly, such as biotin. While the concept of all glycan microarrays is similar, the success of a glycan microarray is largely determined by the size and diversity of the glycan structures on the microarray. The most widely used and successful glycan microarray is available from the Consortium of Functional Glycomics (CFG), which is comprised of more than 600 different semi-synthetic glycans. An alternative method of building a glycan library is through functional derivatizations of naturally occurring glycans, many of which are structurally complex, not accessible through synthesis, and yet more biologically relevant (19). Here we describe a method for preparing functionally and fluorescently labeled natural glycans, printing microarrays, and interrogating them with GBPs (Fig. 1).

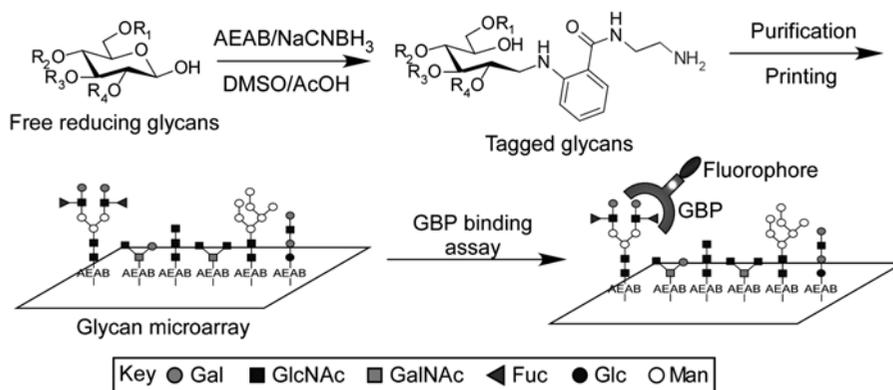


Fig. 1. A schematic of the preparation of glycan microarray and assay with GBPs.

2. Materials

We use MilliQ water (dH₂O) for all aqueous buffer preparation and HPLC solvents. All samples are stored at -20°C as soon as they are prepared unless otherwise noted. Toxic chemicals are handled in a chemical hood as much as possible and are disposed according to chemical waste procedures. Except for HPLC solvents, all chemicals are analytical grade and used without further purification.

2.1. Special Equipment

1. High performance liquid chromatography (HPLC). Use any analytical HPLC system with an appropriate detection system. We use a Shimadzu CBM-20 system coupled with SPD-20A UV/Vis dual λ detector and RF-10xl fluorescence detector. For the purification of glycan derivatives, we use a Hypercarb HPLC column (Thermo, 4.6×150 mm, $5 \mu\text{m}$) with a Javelin cartridge (4×20 mm). The column has sufficient capacity and resolving power to purify conjugates at a scale of $1 \mu\text{g}$ to 1 mg. Solvents: Acetonitrile (A), Water (C), 1% trifluoroacetic acid in water (D). Flow rate: 1 mL/min. Linear gradient: 0 min: 15% A, 75% C and 10% D; 30 min, 45% A, 45% C and 10% D; 30.1 min: 15% A, 75% C and 10% D; 40 min: stop. Fluorescence detection: 330 nm (Ex)/420 nm (Em). UV detection: 330 nm.
2. Mass spectrometry (MS) with Ultraflex II MALDI-TOF/TOF (Bruker). We use an Anchorchip target plate with anchors $200\text{--}800 \mu\text{m}$. We use 2,5-dihydroxybenzoic acid (5 mg/mL in 50% acetonitrile with 0.1% trifluoroacetic acid) as the matrix. Spot $0.5 \mu\text{L}$ of matrix solution on any clean spot, air dry, and then spot $0.5 \mu\text{L}$ sample in salt-free solution. Air dry again for MS analysis in reflective positive and/or linear negative mode for negatively charged glycan derivatives.
3. Centra-vap concentrator (Labconco). Samples are dried without heating.
4. Piezorray printer (Perkin-Elmer). Other types of microarray printers can be used according to the manufacturer's instructions. This non-contact printer has four piezo-electronic pins, a temperature-controlled source plate tray, a 5×5 microarray slide tray and two washing troughs. The printer software controls the printer. Tune the four pins so that ideal spot morphology and size distribution are reached before each printing. In a typical printing cycle, the four pins take up to four samples from the source plate, deliver 0.33 nL drops on the microarray slides, recycle the samples back to the source plate, and immerse in the wash troughs for washing before moving to the next samples. The printer generates a .GAL file incorporating all of the sample information on the microarray.

5. ProScanArray scanner (Perkin-Elmer). This scanner is installed with 4 lasers ranging from blue to red fluorescence. The images are recorded and analyzed by the Scanarray software according to the .GAL file generated by the printer. An Excel data sheet is generated and further processed in Microsoft Excel to give a histogram of binding.

2.2. Preparing Tagged Glycans

1. Free reducing glycans (commercially available from V-labs or isolated from natural sources, lyophilized).
2. DMSO/AcOH solvent: Mix dimethyl sulfoxide (ACS grade) and acetic acid (ACS grade) in a volume ratio of 7/3. This solvent is used to prepare the conjugation reagents.
3. Conjugation reagent A: Add 88 mg 2-amino-*N*-(2-aminoethyl) benzamide (AEAB) (**20**) hydrochloride salt into 1 mL DMSO/AcOH solvent. Sonicate the mixture for 5–10 min to achieve a clear solution. Reagent A can be stored at -20°C for up to 1 month.
4. Conjugation reagent B: Add 64 mg sodium cyanoborohydride (NaCNBH_3) (Sigma-Aldrich) into 1 mL DMSO/AcOH solvent. Vortex the mixture briefly to achieve clear solution. Reagent B needs to be freshly prepared.
5. Conjugation reagent C: Add 90 mg *p*-nitrophenyl anthranilate (PNPA, Fisher Scientific) into 1 mL DMSO/AcOH solvent. Vortex the mixture briefly to achieve a clear solution. Reagent B needs to be freshly prepared.
6. Acetonitrile (HPLC grade).
7. Ethylenediamine (ACS grade).
8. Acetic acid (ACS grade).

2.3. Printing Microarrays

1. Printing buffer (2 \times): sodium phosphate, 0.6 M, pH 8.5.
2. Printer wash buffer (1 \times): sodium phosphate, 0.3 M, pH 8.5 with 0.1% tween-20.
3. Heated humidity chamber: Set up a box with a loose cover inside a water bath set at 55°C .
4. Blocking buffer: 50 mM ethanolamine in 0.1 M Tris buffer, pH 9.0.
5. 384 well plate, V-shape (Biorad).
6. Centrifuge Rotor for 96- and 384-well plates.
7. NHS-activated glass slides (Schott North America).
8. Epoxy-activated glass slides (Corning).
9. Slide centrifuge (Fisher Scientific).

2.4. Assay of Glycan-Binding Proteins on the Glycan Microarrays

1. Biotinylated lectin (commercially available, ex. Vector Labs).
2. Cyanine5 (Cy5)-Streptavidin (Invitrogen).
3. Printed glycan microarray slides.

4. Slide chamber adaptors (Grace Biolabs) (see Note 1).
5. Slow speed shaker (Barnstead).
6. Humidified Slide processing chambers (Fisher Scientific, NC9091416).
7. 100 ml Coplin jars for washing slides.
8. TSM buffer: 20 mM Tris-HCl, pH 7.4 150 mM NaCl, 2 mM CaCl₂, 2 mM MgCl₂.
9. TSM Binding Buffer (TSMBB) = TSM buffer + 0.05% Tween-20 + 1% BSA.
10. TSM Wash Buffer (TSMWB) = TSM Buffer + 0.05% Tween-20.

3. Methods

3.1. Preparation of Glycan-AEAB Conjugates

1. To a lyophilized free reducing glycan sample (0.05–1 mg) in a 1.5 mL Eppendorf tube, add conjugation reagent A (10–50 μ L) and an equal volume of conjugation reagent B. A precipitate forms briefly after mixing. Heat the suspension at 65°C for 2 h in a dry heat block (see Note 2).
2. Add 10 volumes acetonitrile (200 μ L to 1 mL) to the reaction mixture. Shake briefly and store the mixture at –20°C for 30 min. Centrifuge the cold mixture at 10,000 $\times g$ for 3 min. Remove the supernatant and discard. Dissolve the pellet in 100 μ L dH₂O, centrifuge, and transfer the clear supernatant to an HPLC vial for HPLC analysis and purification (see Notes 3 and 4).
3. Inject 1 μ L of the sample for PGC-HPLC analysis. Use both fluorescence and UV detection. Use the profiles as guidance for the following purification. Inject all of the sample into the same PGC-HPLC column using UV detection only. Collect the major peak as the fluorescent glycan-AEAB conjugate.
4. Characterize the collected fraction using MALDI-TOF/TOF to confirm the expected structure.
5. Dry the collected fraction in a Centra-vap for 2 h, and lyophilize the residual solution (see Note 5). Based on its fluorescence and UV absorption, reconstitute the glycan-AEAB conjugate to 200 μ M in dH₂O. This solution can be stored at –20°C until printing.

3.2. Printing the Glycan Microarrays

1. Load 5 μ L of the 200 μ M glycan conjugate solutions described above into a 384-well source plate with V-shape wells, one well per sample. Add 5 μ L 2 \times phosphate buffer (0.6 M, pH 8.5) into each well and mix with a pipet. Centrifuge the plate at 1,000 $\times g$ for 1 min. The plate is then set securely on the printer bench top.
2. Enter the sample names, IDs, and positions into the printer software. Program the printing pattern according to the

manufacturer's instructions, including subarray number and sample replicate number. For example, a microarray can include 2–14 identical subarrays. Each subarray can incorporate a defined number of samples in replicates (3–6). The printer control software will generate a .GAL file based on the setup, which is used in the image processing step.

3. Warm NHS-activated slides from -20°C to room temperature in a vacuum desiccator before opening the package. Epoxy slides are stored at room temperature and can be opened immediately before printing. Set the desired number of slides securely onto the printer bench top.
4. Fill up all necessary liquids, including dH_2O , printing buffer, and wash buffer. Start the printing process (see Note 6).
5. After the printing is finished, put the slides into a box, loosely cover it and incubate for 1 h in a heated humidity chamber (a 55°C water bath) without water contacting the slides.
6. Wash the slides with dH_2O . Immerse the slides in blocking buffer for 1 h, wash with dH_2O , and dry the slides in a slide centrifuge. Store the printed slides at -20°C until use.

3.3. Assay of Glycan-Binding Proteins on the Glycan Microarrays

GBPs or microorganisms can be pre-labeled with any appropriate fluorescent dye or biotin, which can be further detected with fluorescently-labeled streptavidin. The following method is based on biotinylated plant lectins. The expected results are shown in Fig. 2.

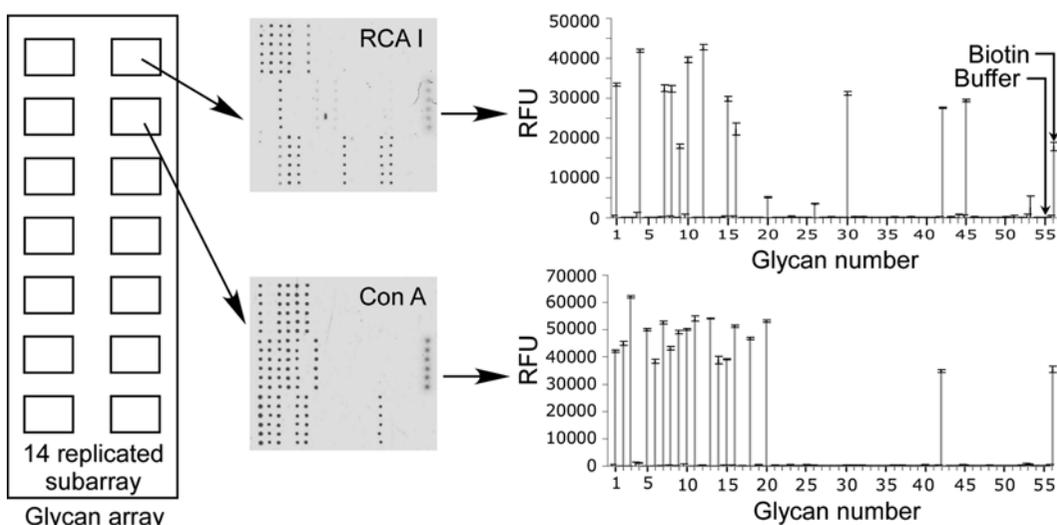


Fig. 2. An example of expected results when a glycan microarray is assayed with biotinylated plant lectins. A single microarray slide was divided into 14 identical subarrays and each subarray was assayed with a different lectin. From each subarray, a fluorescent image is generated, which can be analyzed by Scanarray software to give a histogram. All glycans bound to RCA I contain terminal β 1,4-galactose. All glycans bound to Con A contain the N-glycan core structures.

1. Prepare glycan-binding protein solutions at desired concentrations in TSMBB. The GBPs are pre-labeled with either biotin or fluorescent dyes. We generally use biotinylated lectins at 10 µg/mL or lower (see Note 7).
2. Remove printed microarray slide(s) from -20°C freezer and bring to room temperature in a vacuum desiccator.
3. Securely apply a chamber adaptor on the microarray slide to distribute the array surface into several chambers for different experiments. We use a 14-subarray format.
4. Add 100 µl wash buffer to hydrate the chambers which are to be assayed with GBPs for 5 min. Remove the liquid with vacuum suction without touching the slide surface.
5. Apply 50–100 µl GBP solution carefully to each chamber, cover the chambers with parafilm, and incubate the chambers on a slow speed shaker for 1 h.
6. Remove the parafilm cover and remove the GBP solutions by vacuum suction.
7. Wash each chamber by adding 200 µl TSMWB four times and 200 µl TSM four times, removing the buffer after each wash by vacuum suction. Add Cy5-streptavidin (0.5 µg/mL, 100 µl) in TSMBB to each chamber, cover the chambers with parafilm, and incubate the chambers on a slow speed shaker for 1 h.
8. Remove the parafilm cover and remove the Cy5-streptavidin solution by vacuum suction.
9. Wash each chamber by adding 200 µl TSMWB four times and 200 µl TSM four times, removing the buffer after each wash by vacuum suction. Remove the multichamber adaptor and rinse the slide in a Coplin jar of 100 mL dH₂O. Dry the slides using a slide centrifuge.
10. Scan the slide in the scanner using appropriate wavelengths (633 nm for Cy5). Save the images.
11. Process the images using the Scanarray software to give a datasheet, according to the .GAL file generated by the printer software. Process the datasheet in Microsoft Excel to give a histogram.

4. Notes

1. These adaptors are used to separate a single microarray slide into several chambers. Each chamber can be treated with different conditions.
2. If the reducing end ring structure is important in certain circumstances, free glycans can be derivatized to glycosyl

acrylamide. Ozone treatment of the glycosyl acrylamide generates an aldehyde group, which can be conjugated and profiled using the same condition as free reducing glycans (21).

3. An alternative method of preparing glycan–AEAB conjugates: To a free reducing glycan sample (0.05–1 mg) in a 1.5 mL Eppendorf tube, add conjugation reagent C (10–50 μ L) and an equal volume of conjugation reagent B. Heat the suspension at 65°C for 2 h in a dry heat block. Add 10 volumes acetonitrile (200 μ L to 1 mL) to the reaction mixture. Shake briefly and store the mixture at –20°C for 30 min. Centrifuge the cold mixture at 10,000 $\times g$ for 3 min. Remove the supernatant and discard. Add 20 μ L ethylenediamine to the pellet and incubate at room temperature for 5 min. Add 80 μ L dH₂O, centrifuge and transfer the clear supernatant to an HPLC vial for HPLC analysis and purification.
4. This method uses all commercially available chemicals, avoiding the synthesis of AEAB in the laboratory. The use of ethylenediamine, however, might be undesirable for certain extremely base-labile structures.
5. Extra caution is needed for handling microscale samples. Poke a single hole in the top of the Eppendorf tube before drying in Centra-vap and lyophilizing. Centrifuge the lyophilized material before opening the vial to avoid loss of material.
6. For accurate printing and material conservation, we use a Piezorray Printer to print glycan microarrays. Tune the four tips so that ideal spot morphology and size distribution are reached according to manufacturer’s instructions, which define a 0.33 nL delivery volume.
7. Although we commonly use TSM buffer, we also use other buffer systems for different GBPs as required.

References

1. Drickamer K, Taylor ME (2002) Glycan arrays for functional glycomics. *Genome Biol* doi:10.1186/gb-2002-3-12-reviews1034
2. Li J, Richards JC (2010) Functional glycomics and glycobiology: an overview. *Methods Mol Biol* 600: 1–8
3. Paulson JC, Blixt O, Collins BE (2006) Sweet spots in functional glycomics. *Nat Chem Biol* 2: 238–248
4. Varki A, et al (2009) *Essentials of Glycobiology*, Edn. 2nd. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY)
5. de Paz JL, Seeberger PH (2006) Recent advances in carbohydrate microarrays. *QSAR & Combinatorial Science* 25: 1027–1032
6. Feizi T, Chai W (2004) Oligosaccharide microarrays to decipher the glyco code. *Nat Rev Mol Cell Biol* 5: 582–588
7. Hirabayashi J (2003) Oligosaccharide microarrays for glycomics. *Trends Biotechnol* 21: 141–143; discussion 143
8. Stevens J, Blixt O, Paulson JC, Wilson IA (2006) Glycan microarray technologies: tools to survey host specificity of influenza viruses. *Nat Rev Microbiol* 4: 857–864
9. Smith DF, Song X, Cummings RD (2010) Use of glycan microarrays to explore specificity of glycan-binding proteins. *Methods Enzymol* 480: 417–444
10. Park S, Lee MR, Shin I (2008) Carbohydrate microarrays as powerful tools in studies of

- carbohydrate-mediated biological processes. *Chem Commun (Camb)* 37:4389–4399
11. Park S, Shin I (2002) Fabrication of carbohydrate chips for studying protein-carbohydrate interactions. *Angew Chem Int Ed Engl* 41: 3180–3182
 12. Ratner DM, et al (2004) Probing protein-carbohydrate interactions with microarrays of synthetic oligosaccharides. *Chembiochem* 5: 379–382
 13. Bryan MC, et al (2004) Covalent display of oligosaccharide arrays in microtiter plates. *J Am Chem Soc* 126: 8640–8641
 14. Blixt O, et al (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proceedings of the National Academy of Sciences of the United States of America* 101: 17033–17038
 15. de Boer AR, Hokke CH, Deelder AM, Wuhrer M (2007) General microarray technique for immobilization and screening of natural glycans. *Analytical Chemistry* 79: 8107–8113
 16. Song X, Xia B, Lasanajak Y, Smith DF, Cummings RD (2008) Quantifiable fluorescent glycan microarrays. *Glycoconj J* 25: 15–25
 17. Liu Y, et al (2007) Neoglycolipid probes prepared via oxime ligation for microarray analysis of oligosaccharide-protein interactions. *Chem Biol* 14: 847–859
 18. Stoll MS, (2000) Fluorescent neoglycolipids. Improved probes for oligosaccharide ligand discovery. *Eur J Biochem* 267: 1795–1804
 19. Song X, et al (2011) Shotgun glycomics: a microarray strategy for functional glycomics. *Nat Methods* 8: 85–90
 20. Song X, et al (2009) Novel fluorescent glycan microarray strategy reveals ligands for galectins. *Chem Biol* 16: 36–47
 21. Song X, Lasanajak Y, Xia B, Smith DF, Cummings RD (2009) Fluorescent glycosylamides produced by microscale derivatization of free glycans for natural glycan microarrays. *ACS Chem Biol* 4: 741–750

Chapter 12

Resolving Bottlenecks for Recombinant Protein Expression in *E. coli*

Yoav Peleg and Tamar Unger

Abstract

Escherichia coli is widely used as an expression system for production of recombinant proteins of prokaryotic and eukaryotic origin. A large body of knowledge has accumulated throughout the last few decades regarding expression of recombinant proteins in *E. coli*. However, despite this progress, protein production, primarily of eukaryotic origin, still remains a challenge. The biggest obstacle lies in obtaining large amounts of a given protein in a correctly folded form. Several strategies are being used to increase both yield and solubility. These include expression as fusion proteins, co-expression with molecular chaperones, or with a protein partner(s), and the use of multiple constructs for each protein. In this chapter, we focus on strategies for creating expression vectors, as well as on guidelines for improving recombinant protein solubility.

Key words: Protein expression, Fusion-tags, Restriction-free cloning, Protein solubility, *E. coli*

1. Introduction

The first bottleneck encountered when studying the biochemical, biophysical, or structural aspects of a given protein is the production of adequate amounts of soluble and biologically active protein. Bacterial expression systems, primarily *Escherichia coli* (*E. coli*), offer the advantage of obtaining large amounts of recombinant protein in a relatively short period and at low cost. However, despite the accumulated knowledge regarding recombinant protein expression, researchers often encounter difficulties using the *E. coli* system, primarily when the protein is of eukaryotic origin. One of the major problems is the poor solubility of the protein (i.e., formation of inclusion bodies) (1). Most researchers would prefer to avoid the time consuming re-folding process, which must be customized for each specific protein. Therefore, much effort has been devoted

to overcoming the insolubility bottleneck. Additional difficulties encountered during production of recombinant proteins in *E. coli* include low levels of expression, degradation of the protein, toxicity, and production of nonfunctional protein. The latter primarily occurs because of the lack of correct eukaryotic post-translational modification.

Since we are unable in this chapter to cover the entire spectrum of approaches for handling all these bottlenecks, we will focus only on those directly related to protein solubility. Addressing the solubility problem can be approached either at the molecular, cellular, or cell-culture level. At the molecular level, the gene of interest can be cloned downstream to a fusion tag such as Thioredoxin (Trx), maltose-binding protein (MBP), or Glutathione-S transferase (GST), which increases both solubility and level of expression (2, 3). The introduction of tags, such as GST and MBP, also provides a specific purification handle. Additionally, multiple constructs for each protein can be generated, each expressing a different domain or a truncated fragment, thus, increasing the chances of obtaining a soluble form of the protein (4). Increasing solubility at the cellular level can be approached by co-expression of the target gene with molecular chaperones (5, 6) or with specific protein partner(s) (7). A different strategy for overcoming solubility difficulties is to either alter the growth conditions, or to supplement the bacterial culture with various additives, such as betaine or sorbitol (8). From our experience, with many hundreds of recombinant proteins expressed in the *E. coli* system, one of the most promising approaches for increasing protein solubility involves lowering the culture temperature to 15–20°C during the induction stage (9). In our hands, this simple approach results in dramatic improvements in the solubility of many proteins of both eukaryotic and prokaryotic origin.

In addition to the factors described above, the cloning methodology has a major impact on the characteristics of the final protein product, as well as on the speed and costs of the entire process starting from gene to protein. The classical ligation-dependent cloning (LDC) approach, which relies on restriction enzyme digestion followed by ligation of the insert and the vector, has been used extensively in the last few decades. However, in recent years, with the establishment of structural proteomics consortia, new cloning techniques have been developed and gradually adapted, thus replacing the LDC approach (9). The new strategies are primarily ligation-independent cloning (LIC) techniques. Several of the LIC procedures rely on recombination between the insert and the destination vector (Gateway from Invitrogen® and In-Fusion™ from Clontech). A different approach uses complementary single-strand overhangs to combine the vector and the insert (10). At the Israel Structural Proteomics Center (ISPC), we have optimized the restriction free (RF) cloning procedure over the past few years;

this is an LIC approach based on whole plasmid amplification of the vector and the insert (11, 12). We have recently extended the applications of the RF cloning platform to a variety of molecular manipulations, including simultaneous multicomponent assembly, simultaneous cloning of two DNA fragments into distinct positions within an expression vector, and parallel cloning of a PCR product into a series of expression vectors (12). One of the main advantages of RF cloning is that DNA integration can be directed to any position within a circular plasmid. As a result, no extra and unnecessary sequences are introduced into the protein open reading frame, in contrast to other LDC or LIC approaches.

The protocols described herein present up-to-date guidelines for cloning and protein expression in the *E. coli* system with an emphasis on strategies for facilitating DNA cloning and increasing protein solubility.

2. Materials

2.1. Bacterial Strains

For all DNA cloning and plasmid preparation procedures, the DH5 α strain of *E. coli* (Agilent Technology, Stratagene division, Santa Clara, CA) is used. For protein expression, *E. coli* BL21(DE3) and its derivatives (EMD Chemicals, Inc., Gibbstown, NJ) are employed.

2.2. Expression Vectors

At the Israel Structural Proteomics Center (ISPC), we have constructed a series of expression vectors based on the backbone of pET- vectors (EMD Chemicals, Inc.) (12, 13). All expression vectors harbor a 6 \times His-tag followed by a tobacco etch virus (TEV) cleavage site (Fig. 1a). Besides the expression vector, pET28-TevH, which harbors only the 6 \times His-tag, the other expression vectors carry different fusion tags located at the N-terminal part of the expression cassette (Fig. 1a, b). The fusion partners employed include: GST, β 1-domain of the Streptococcal protein G (GB1), MBP, Trx, Protein Disulfide Isomerase A (DsbA), and Protein Disulfide Isomerase C (DsbC). For the DsbC and DsbA fusions, two versions were constructed, for secretion into the periplasm and for intracellular expression. For the latter expression vectors, the periplasmic secretion signal was removed. Kanamycin is used as a selection marker for all the expression vectors constructed. All expression vectors listed in Fig. 1 were engineered to have the same sequence at the multiple cloning site (MCS) to allow parallel cloning either by restriction endonucleases or integration of the same PCR product using the restriction free (RF) cloning procedure (11, 12) (see Subheading 2.3, and Note 1).

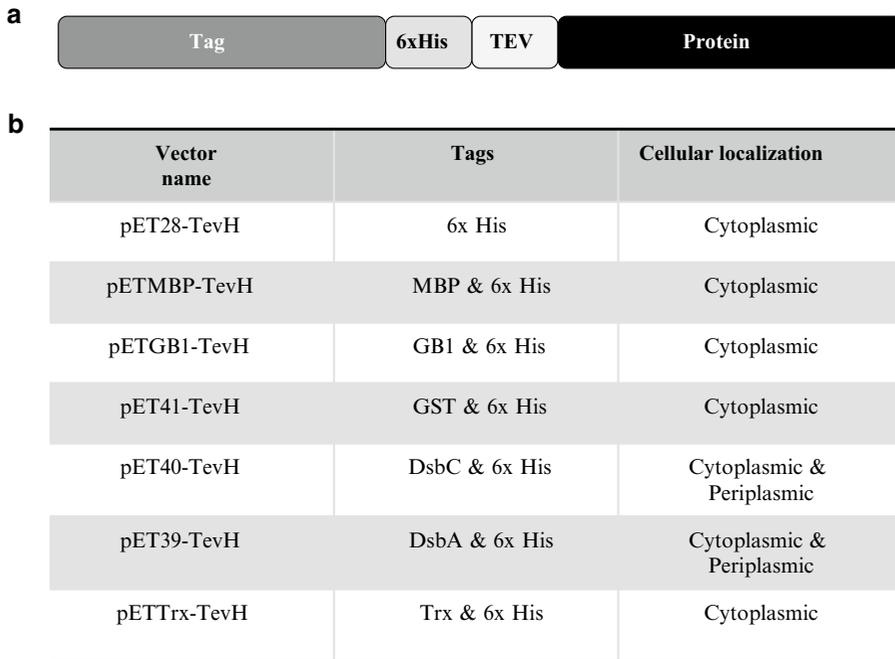


Fig. 1. pET-derived expression vectors harboring various fusion partners. **(a)** Schematic representation of the expression cassette. **(b)** List of the vectors, fusion tags, and their targeted cellular localization. *MBP* maltose-binding protein, *GB1* β 1-domain of the Streptococcal protein G, *Trx* Thioredoxin, *DsbA* Protein Disulfide Isomerase A, *DsbC* Protein Disulfide Isomerase C, *GST* Glutathione S-transferase, *TEV* Tobacco Etch Virus. For the DsbC and DsbA fusions, two versions were constructed, for secretion into the periplasm and for intracellular expression.

2.3. Establishment of Expression Vectors Using RF Cloning Reaction

1. Phusion high fidelity thermo-stable DNA polymerase (Finnzymes, Espoo, Finland).
2. 5 \times Phusion HF buffer (Finnzymes).
3. 10 mM dNTPs stock solution.
4. Synthetic primers designed for the RF cloning, and for sequencing (see Note 2).
5. *DpnI* (purchased from New England Biolabs (Ipswich, MA) or Fermentas (Ontario, Canada)).
6. PCR product purification kit (PCRquick, Intron Biotechnology, Daejeon, South Korea or equivalent).
7. Plasmid DNA purification kit (DNA-spin, Intron Biotechnology, or equivalent).
8. High efficiency *E. coli* DH5 α competent cells for heat shock transformation (see Note 3).
9. Luria-Bertani (LB) medium and agar plates supplemented with kanamycin (30 μ g/mL) and/or chloramphenicol (34 μ g/mL).

2.3.1. Agarose Gel Electrophoresis

For most applications, DNA can be analyzed on 1% (*w/v*) agarose gels in 0.5–1.0× TAE buffer.

1. Agarose gel electrophoresis apparatus.
2. Agarose.
3. Ethidium bromide (10 mg/mL stock).
4. Tris/acetic acid/EDTA (TAE) buffer, 50× stock (Bio-Rad Laboratories).
5. Microwave oven.
6. Visualization system (Gel Doc 2000, Bio-Rad or equivalent).
7. 37°C incubator.
8. Ready-mix PCR master kit (Sigma-Aldrich, St. Louis, MO or equivalent).
9. Temperature-controlled microcentrifuge (Eppendorf, or equivalent).

2.4. Protein Expression

1. Expression vectors (listed in Fig. 1) and Chaperone plasmid set from Takara (Otsu, Japan).
2. High efficiency *E. coli* BL21(DE3) competent cells for heat shock transformation (see Note 3).
3. Luria-Bertani (LB) liquid medium and plates, supplemented with the appropriate antibiotic(s); kanamycin (30 µg/mL) and/or chloramphenicol (34 µg/mL).
4. 0.2 M IPTG (isopropyl-thio-β-D-galactopyranoside) stock solution.
5. 20% (*w/v*) L-arabinose stock solution, filter sterilized.
6. 5 µg/mL Tetracycline stock solution, filter sterilized.
7. 14 and 50 mL round bottom tubes (BD Falcon, Franklin Lakes, NJ or equivalent).
8. Temperature-controlled (15–37°C) shaker.
9. Spectrophotometer.
10. Spectrophotometer cuvettes.

2.5. Cell Extraction and Protein Analysis

1. Cell-lysis buffer for sonication. Bacteria are lysed in buffer containing 50 mM Tris-HCl, pH 8, 500 mM NaCl, 20 mM imidazole, 1 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF), and 1 µL/mL protease inhibitor cocktail (Set IV, EMD Chemicals, Inc.).
2. Sonicator (vibra-cell from Sonics, Newtown, CT or equivalent) equipped with a micro-tip for 1.5 mL tubes.
3. BugBuster® extraction reagent supplemented with 1 µL/mL of Benzonase nuclease (EMD Chemicals, Inc.) and Lysozyme 6 µg/mL.

4. Temperature-adjustable microcentrifuge (Eppendorf or equivalent).
5. Rotator (Intelli-mixer, Rose Scientific, Alberta, Canada or equivalent).
6. Ni-NTA agarose beads (Qiagen or equivalent). Beads should be washed three times in Ni-binding buffer prior to use as the beads are kept in ethanol solution.
7. Wash buffer for Ni-NTA agarose beads: 50 mM phosphate buffer, pH 8.0, 300 mM NaCl and 20 mM imidazole.
8. Binding buffer Ni-NTA agarose beads: 50 mM phosphate buffer pH 8.0, 300 mM NaCl and 10 mM imidazole.
9. Elution buffer for Ni-NTA agarose beads: 50 mM phosphate buffer, pH 8.0, 300 mM NaCl and 500 mM imidazole
10. Protein gel electrophoresis system (Mini-Protean 3 Cell from Bio-Rad or equivalent).
11. Gel-staining solution (GelCode Blue Reagent, Thermo Scientific-Pierce or equivalent).
12. Protein sample (SB) buffer, both 4× concentrated (SBx4), and at final dilution (SBx1). To make a SBx4 stock solution (10 mL) mix the following components: 4.8 ml 0.5 M Tris-HCl, pH 6.8, 0.8 g SDS, 4.0 mL glycerol, 0.4 mL 14.7 M β-mercaptoethanol and 8 mg bromophenol blue. Store in aliquots at -20°C. For SBx1 dilute with 3 volumes of ddH₂O.

3. Methods

3.1. Construction of Expression Vectors

In the present protocol, the restriction free (RF) cloning procedure (11, 12) is described in detail. RF cloning can be used to introduce the same PCR product into multiple expression vectors harboring identical sequences at the sites of integration. All vectors listed in Fig. 1 are suitable for such parallel cloning.

1. Design primers for RF cloning of the target gene (see Note 4).
2. Set up PCR reactions in 0.2 mL tubes, in a final volume of 50 μL, as follows: 20 ng of template DNA, 0.4 μM of each primer (forward and reverse), 200 μM of each dNTP, 1× Phusion HF buffer, and 1.6 U of Phusion DNA polymerase. Add sterile ddH₂O to a final volume of 50 μL.
3. Perform PCR on the gene of interest: Use a single denaturation step (95°C, 1 min) followed by 25 cycles of: denaturation (95°C, 30 s), annealing (60°C, 1 min) and elongation (72°C, 1.5 min), and a final elongation step of 6 min at 72°C.

4. Analyze small aliquots (1–2 μL) from the PCR reaction mixture on an agarose gel to verify that the DNA product obtained is of the correct size.
5. Purify the PCR product with PCR purification kit according to the manufacturer's recommendations.
6. Set up RF reactions in 0.2 mL tubes, in a final volume of 50 μL , as follows: 20 ng of target DNA, 100 ng of PCR product (from the previous step), 200 μM of each dNTP, 1 \times Phusion HF buffer, and 1.6 U of Phusion DNA polymerase. Add sterile ddH₂O to a final volume of 50 μL , mix gently.
7. Perform RF reaction: Use a single denaturation step (95°C, 30 s) followed by 30 cycles of: denaturation (95°C, 30 s), annealing (60°C, 1 min) and elongation (72°C, 5 min), and a final elongation step of 6 min at 72°C (see Note 5).
8. Remove 10 μL from the RF reaction to a new tube. Add 1 μL *DpnI* and incubate for 1–2 h at 37°C.
9. Transform the *DpnI*-treated reaction into competent DH5 α *E. coli* cells. Plate on LB-kanamycin plates. Incubate plates overnight in a 37°C incubator.
10. On the following day, select 6–8 single colonies from the transformation plate, and perform colony PCR (see step 11, below) to verify that the PCR fragment was cloned (see Note 6).
11. Set reactions in 0.2 mL tubes as follows: 0.4 μM of each primer (forward and reverse, see Note 2) and 10 μL of 2 \times Master mix solution. Using a sterile toothpick, add bacteria from a selected colony and complete with sterile ddH₂O to a final volume of 20 μL , mix gently (see Note 7).
12. Perform the colony PCR reaction: Use a single denaturation step (95°C, 1 min) followed by 25 cycles of: denaturation (95°C, 30 s), annealing (60°C, 1 min), and elongation (72°C, 1.5 min), and a final elongation step of 6 min at 72°C.
13. Analyze PCR reaction products on a 1% agarose gel. Select two individual positive colonies (see Note 8).
14. Inoculate positive clones into 10 mL LB medium supplemented with kanamycin (30 $\mu\text{g}/\text{mL}$) in 50-mL conical tubes. Grow bacteria at 37°C, for 16–20 h.
15. Harvest cells and extract DNA using a plasmid DNA purification kit (see Note 9).
16. Transform two plasmids from each clone to competent BL21(DE3) *E. coli* cells (or into any other *E. coli* strain used for expression of pET vectors), spread cells onto LB agar plates supplemented with kanamycin and incubate at 37°C. On the following day, select 1–2 colonies from each plate for small-scale expression experiments (see Note 10). Inoculate single colonies

into 3 mL liquid LB medium, in 14-mL tubes, supplemented with kanamycin (30 $\mu\text{g}/\text{mL}$) and grow with shaking overnight at 37°C.

17. For co-expression experiments using Takara's chaperone plasmid set, co-transform the expression vector with each chaperone vector into BL21(DE3) *E. coli* cells (or into any other *E. coli* strain used for expression of pET vectors). Spread cells onto LB agar plates supplemented with kanamycin (30 $\mu\text{g}/\text{mL}$) and chloramphenicol (34 $\mu\text{g}/\text{mL}$) and incubate at 37°C (see Note 10). The following day, use 1–2 colonies from each plate for small-scale expression experiments. Inoculate single colonies into 3 mL liquid LB medium, in 14-mL tubes, supplemented with kanamycin (30 $\mu\text{g}/\text{mL}$) and chloramphenicol (34 $\mu\text{g}/\text{mL}$), and grow with shaking overnight at 37°C.

3.2. Protein Expression

When multiple clones are screened in parallel for protein expression, high-throughput approaches and robotics are the preferred approach for handling the process (14). However, for most laboratories this is not necessary, since only small numbers of clones must be analyzed. In the following protocol, we present a procedure for screening for protein expression on a small scale, which is suitable for wide-range of proteins expressed in the *E. coli* system. In Fig. 2 we demonstrate that co-expression of a target protein with two chaperone combinations, pGro7 and pKJE7, is necessary to induce soluble expression of a 30 kDa target protein.

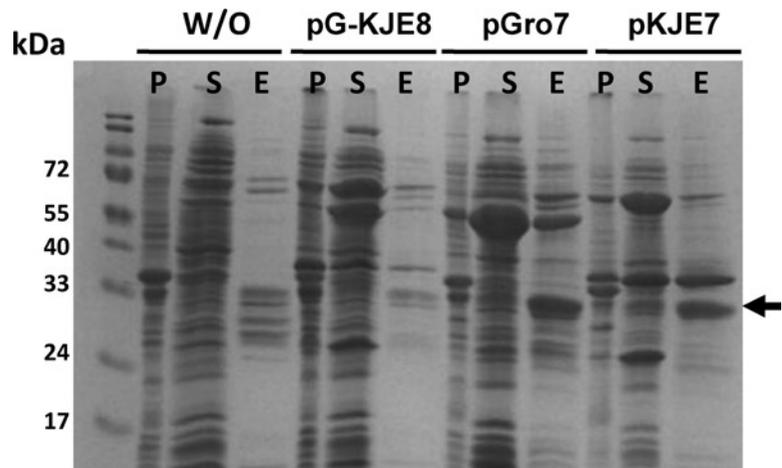


Fig. 2. Protein co-expression with molecular chaperones. Protein expression and solubility were analyzed following induction at 30°C for 3 h. Takaras' expression vectors pG-KJE8, pGro7 and pKJE7, harboring different combinations of chaperone-encoding genes, along with a target protein of 30 kDa were used for co-expression. Chaperone expression was induced with L-arabinose (for pGro7 and pKJE7), and with L-arabinose and tetracycline (for pG-KJE8). Arrow indicates position of the target protein. W/O- no chaperone-encoding gene was used. Cell extraction was performed by sonication, P Pellet fraction, S Soluble fraction, E Ni-bead elution fraction. Molecular weight marker, in kDa, is shown on the left.

1. Following growth for 16–20 h at 37°C (see Note 11), dilute the culture 1:100, into 4 mL of fresh LB medium containing the appropriate antibiotic(s), in a 14 mL tube. For each clone, two duplicate cultures should be established.
2. For co-expression experiments using Takara's chaperone plasmid set, induce chaperone expression immediately following dilution of the culture. Induce expression by addition of 0.2% L-arabinose (1:100 dilution for a 20% stock solution) and/or 5 ng/mL tetracycline (1:1,000 dilution for 5 µg/mL stock solution) (see Note 12 and Fig. 2).
3. Incubate cultures at 37°C with shaking until OD₆₀₀ reaches 0.6–0.8.
4. Induce protein expression with 200 µM IPTG (1:1,000 dilution for 0.2 M IPTG stock solution). For each clone, incubate one tube at 30°C for 3–4 h, and a second one at 15°C for 16–20 h.
5. Harvest cells by centrifugation. Store cell pellet at –20°C or proceed immediately to protein extraction, purification, and analysis.

3.3. Protein Extraction, Purification, and Analysis

Extraction of proteins from small-scale cultures of *E. coli* cells can be done in several ways, including by the sonication, freeze and thaw procedure (14), or using a detergent-based chemical lysis. Below we present protocols for protein extraction by sonication, and chemical lysis using the detergent-based reagent, BugBuster®.

3.3.1. Cell Extraction by Sonication

1. Re-suspend each cell pellet in 1 mL sonication lysis buffer (see Subheading 2.5, Note 1) and transfer to 1.5 mL micro-tubes.
2. Disrupt cells on ice, by sonication using a micro-tip. Use 10 s pulses with 30 s intervals for six pulses. If the bacterial suspension is not clear, repeat the process (see Note 13).
3. Remove cell debris by centrifugation at 4°C for 15 min at 18,000 ×g. Transfer the clear supernatant into a new 1.5 mL micro-tube.

BugBuster®, or other detergent-based extraction buffers, is a good choice for simultaneously processing multiple samples. However, we have occasionally obtained false-negative results using BugBuster® extraction (see Note 14). Figure 3 compares protein expression of the 65 kDa target protein using BugBuster® vs. sonication as the lysis procedures. The target protein was observed in the pellet, and in the soluble and elution fractions at 15 and 30°C, when sonication was used. However, protein expression was observed only in the pellet fraction when BugBuster® reagent was used for cell lysis.

1. Re-suspend cells with 800 µL of supplemented BugBuster® reagent by vortexing or pipetting and transfer to a 1.5-mL test tube.
2. Rotate extracts at room temperature for 20–30 min.

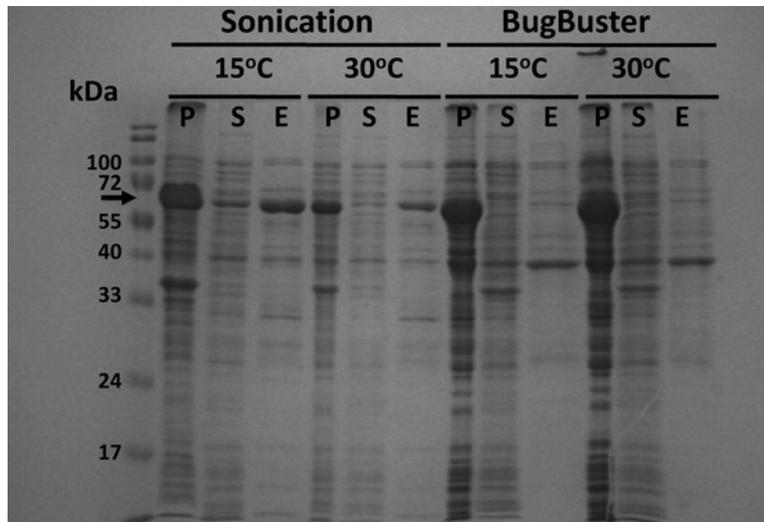


Fig. 3. Effect of extraction procedure on protein solubility. Protein expression and solubility were analyzed following induction with 200 μ M IPTG at 15°C for 20 h or at 30°C for 3 h. Cell extraction was performed in parallel by sonication or using BugBuster® reagent, to equivalent cell pellets derived from the same expression experiment. Arrow indicates position of the target protein. *P* Pellet fraction, *S* Soluble fraction, *E* Ni-bead elution fraction. Molecular weight marker, in kDa, is shown on the *left*.

3. Spin extracts for 15 min at 18,000 $\times g$ at 4°C.
4. Remove the supernatant fraction and transfer to a new 1.5-mL tube. The supernatant fraction is used for protein capture by Ni-beads.
5. Remove a 15 μ L sample from the soluble supernatant fraction and transfer to a new 1.5-mL tube. Add 45 μ L of SBx4 and heat for 5 min at 95°C.
6. Following removal of the supernatant, resuspend the final pellet in 50 μ L SBx1 and heat for 5 min at 95°C.
7. Store the pellet and soluble protein samples at -20°C for analysis by SDS-PAGE.

3.3.2. Ni-Purification and Protein Analysis

1. Add 50 μ L washed Ni-beads to the supernatant fractions, obtained following sonication or BugBuster® treatment. Incubate tubes at 4°C for 40 min to 1 h on a rotator.
2. Spin protein-bound beads at low speed (420 $\times g$ or 2,000 rpm in an Eppendorf microcentrifuge) for 1 min at room temperature. Discard supernatant, which contains unbound proteins.
3. Wash protein-bound beads two-three times with 1 mL Ni-washing buffer and spin each time at 420 $\times g$ for 1 min. Discard washing buffer.
4. Elute protein by adding 50 μ L of Ni-elution buffer. Mix gently. Incubate for 5–10 min at room temperature. Spin down at 420 $\times g$ for 1 min. Remove 45 μ L from the eluted protein and

transfer into a new 1.5-mL tube. Add 15 μ L of SBx4 and heat for 5 min at 95°C.

5. Analyze pellet, soluble and elution fractions on SDS-PAGE (Figs. 2 and 3). Load 5 μ L from the pellet fraction and 15 μ L from the soluble and Ni-elution fractions. Stain SDS-PAGE with GelCode and wash off excess dye (see Note 15).

4. Notes

1. The restriction free (RF) cloning technique has been used on a routine basis at the ISPC over the past 3 years. Other cloning techniques, including restriction enzyme-based methods or commercial systems (e.g., Gateway In-Fusion) can be employed as well. However, we found RF cloning to be highly efficient and robust. In addition, cloning using the RF approach can be performed into any vector of choice and at any position. More details describing the technique and its applications can be found elsewhere (11, 12).
2. Synthetic primers are ordered from Integrated DNA Technologies (IDT) or Sigma-Genosys. Primers up to 60-mers are ordered with only basic desalting purification. Longer primers are purified either by HPLC or SDS-PAGE. For DNA sequencing and colony PCR, the universal primers (T7 and PET REV) found in all pET vector derivatives are used for analysis of the integrity of clones. If needed, gene-specific primers may be used as well.
3. For preparation of competent *E. coli* cells, we routinely use the procedure described previously (15). In this procedure, *E. coli* cells are grown at 18°C prior to harvesting and preparation of the competent cells.
4. Design of primers for the RF cloning is described elsewhere (12). Primers used for PCR amplification, and generation of the mega-primers used for the RF cloning, are comprised of two regions: a target-specific sequence at the 3' end of the primer and 30 bp over-lapping sequences at the 5' end, which corresponds to the desired sites of integration into the destination vector. The length of the overlapping sequences can range from 20 to 40 bp. The melting temperature (T_m) of the target-specific sequence can range from 59 to 70°C, where A or T, and G or C each contribute 2 and 4°C to the T_m , respectively.
5. The RF cloning procedure can be modified if needed. One of the parameters that may influence the efficiency of the reaction is the annealing temperature. For optimization of RF cloning, we use annealing temperatures ranging from 60 to 70°C.

If necessary, dimethyl sulfoxide (DMSO), 5–10% *v/v*, can be added. If cloning into long plasmids (above 10 kb) is performed, it might be necessary to increase the elongation time from 5 min, as recommended in the standard protocol (see Subheading 3.1).

6. Colony PCR is an optional step, although highly recommended. Before selecting clones for DNA sequencing it is advisable to perform colony PCR to ensure that the selected clones harbor the target gene.
7. The colony PCR reactions are carried out using forward and reverse primers derived from sequences flanking the cloning sites in the vector (see Note 2). Alternatively, a combination of a specific primer from the insert and a primer from the expression vector can be used. It is advisable to perform a parallel PCR for the destination vector without the target gene using the same set of primers and under the same conditions. Before performing the PCR reaction, streak the selected colonies on a selective plate and grow at 37°C.
8. PCR products originating from positive clones should run at a higher molecular weight compared to that of the parental plasmid. The expected size of the PCR products should be calculated in advance.
9. It is not essential to sequence the DNA clones prior to running small-scale protein expression experiments. DNA sequencing should be performed at a later stage, e.g., prior to large-scale production. In cases in which no protein is observed at the expected molecular weight in small-scale experiments, DNA sequencing is recommended to verify the integrity of the gene.
10. For transformation into BL21(DE3) cells, we often use an alternative transformation protocol, which eliminates the extra day needed for recovery of the *E. coli* colonies harboring the expression vector. Instead of plating the transformation mix on selective agar plates, the *E. coli* cells, following the recovery stage (without antibiotic) are transferred to a new tube containing LB plus antibiotic. Cells are grown overnight and are used directly for the expression experiment. It should be noted that such a strategy is only appropriate when a defined plasmid is transformed. When an RF cloning reaction is transformed, the transformation mix must be plated on selective plates in order to first identify a correct clone.
11. It is recommended that a glycerol stock is prepared from the selected culture (final glycerol concentration 20–25% *v/v*). The stock should be stored at –80°C until used.
12. Induction of the chaperones can also be initiated about 30 min before cells reach the desired OD₆₀₀ for target protein induction. Induction with L-arabinose is required for Takaras' plasmids- pGro7, pTfl6, and pKJE7; induction with tetracycline

is used for pG-Tf2; and induction with both L-arabinose and tetracycline is necessary for pG-KJE8.

13. If proteins have to remain active following extraction, make sure that the extracts are not overheated; perform the extraction on ice, and when using a micro-tip do not exceed 40% of the maximal amplitude. When multiple samples are handled simultaneously, a multiple-tip probe is a good option (this can be purchased from Sonics).
14. BugBuster[®] is a highly efficient detergent-based extraction reagent for simultaneously processing multiple-samples. However, using BugBuster[®] can result in false-negative results, which lead to failure in detecting at least some soluble proteins (Fig. 3). When using BugBuster[®], the extract should become clear following the rotation step. However some cloudiness may appear, indicating the presence of insoluble protein. Besides BugBuster[®], other commercial bacterial detergent-based extraction reagents are also available, including those from Sigma-Aldrich (CellLytic-B) and Pierce (B-PER).
15. In order to facilitate the staining and de-staining procedures, transfer gel to a plastic container and cover with 100–150 mL water. Heat in a microwave for 1–1.5 min. Remove water and repeat the washing step for two additional times. Following the third washing step, remove the excess water and cover the gel with staining solution. Heat in a microwave for about 30 s. Remove staining solution and view gel.

Acknowledgments

We thank Prof. J. Sussman, Prof. I. Silman, Prof. G. Schreiber and Prof. Yigal Burstein for their continuous support throughout the study. This research was supported by the European Commission Sixth Framework Research and Technological Development Program “SPINE2-COMPLEXES” Project, under contract No. 031220; a grant from the Israel Ministry of Science, Culture, and Sport to the Israel Structural Proteomics Center; the Divadol Foundation; and the Neuman Foundation.

References

1. Baneyx F (1999) Recombinant protein expression in *Escherichia coli*. *Curr Opin Biotechnol* 10: 411–421
2. Malhotra A (2009) Tagging for protein expression. *Methods Enzymol* 463: 239–258
3. Esposito D, Chatterjee DK (2006) Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotechnol* 17: 353–358
4. Yumerefendi H, Tarendeau F, Mas PJ, Hart DJ (2010) ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. *J Struct Biol* 172: 66–74

5. Thomas JG, Baneyx F (1996) Protein folding in the cytoplasm of *Escherichia coli*: requirements for the DnaK-DnaJ-GrpE and GroEL-GroES molecular chaperone machines. *Mol Microbiol* 21: 1185–1196
6. Gragerov A, Nudler E, Komissarova N, Gaitanaris GA, Gottesman ME, Nikiforov V (1992) Cooperation of GroEL/GroES and DnaK/DnaJ heat shock proteins in preventing protein misfolding in *Escherichia coli*. *Proc Natl Acad Sci USA* 89: 10341–10344
7. Romier C, Ben Jelloul M, Albeck S, et al (2006) Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. *Acta Crystallogr D Biol Crystallogr* 62: 1232–1242
8. Blackwell JR, Horgan R (1991) A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Lett* 295: 10–12
9. Graslund S, Nordlund P, Weigelt J, et al (2008) Protein production and purification. *Nat Methods* 5: 135–146
10. Gileadi O, Burgess-Brown NA, Colebrook SM, et al (2008) High throughput production of recombinant human proteins for crystallography. *Methods Mol Biol* 426: 221–246
11. van den Ent F, Lowe J (2006) RF cloning: a restriction-free method for inserting target genes into plasmids. *J Biochem Biophys Methods* 67: 67–74
12. Unger, T, Jacobovitch Y, Dantes A, Bernheim R, Peleg Y. (2010) Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. *J Struct Biol* 172: 34–44
13. Peleg Y, Unger T (2008) Application of high-throughput methodologies to the expression of recombinant proteins in *E. coli*. *Methods Mol Biol* 426: 197–208
14. Berrow NS, Bussow K, Coutard B, et al (2006) Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. *Acta Crystallogr D Biol Crystallogr* 62: 1218–1226
15. Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96: 23–28

Chapter 13

Recombinant Protein Expression in the Baculovirus-Infected Insect Cell System

Tamar Unger and Yoav Peleg

Abstract

Over the last two decades, the use of eukaryotic cells for expression of recombinant proteins has become the preferred choice for many applications. This is primarily the case when posttranslational modifications and correct disulfide-bond formation are necessary for protein folding and activity. Among the eukaryotic expression systems, the baculovirus-infected insect cell platform has gained particular attention, resulting in the development and implementation of multiple strategies for protein expression. Here, we present baculovirus-infected insect cells as an efficient expression system for eukaryotic proteins. We demonstrate a simplified and a shortened procedure for recombinant virus production that is sufficient for large-scale production of proteins in insect cells.

Key words: Baculovirus, Insect cells, Protein expression and production, Secreted proteins, Intracellular proteins

1. Introduction

Expression of recombinant proteins in eukaryotic systems is a laborious, time-consuming, and costly process. However, despite this unavoidable complexity, expression in eukaryotic systems has gained increasing popularity within the scientific community over the last two decades. The main reason for this is that other expression systems, and primarily the bacterial *Escherichia coli* system, cannot properly express many proteins, primarily those requiring posttranslational modifications and/or formation of di-sulfide bonds for proper function.

One of the eukaryotic systems that attracted significant interest in recent years is the baculovirus-infected insect cell platform (1, 2). The most common species of baculoviruses used for protein expression studies is the *Autographa californica* multiple nuclear

polyhedrosis virus (AcMNPV), which relies on the lepidopteran species, *Spodoptera frugiperda*, as its host (3, 4). Two approaches were developed for expression in insect cells, secreted and extracellular (5). The choice of targeting of a given protein for expression is determined by its cellular localization in the native state. Proteins that function extracellularly are expressed as secreted proteins and are collected from the medium (6). On the other hand, proteins that function in the nucleus or in the cytoplasm are expressed intracellularly (7).

There are two main methods (both commercialized), for the production of recombinant virus for expression in insect cells (8). The Bac-to-Bac system (Invitrogen) is based on site-specific transposition of an expression cassette into a baculovirus shuttle vector, referred as the “bacmid,” propagated in *E. coli* (9). The BaculoGold system (BD Biosciences) operates on a different principle. This is based on homologous recombination between a linearized baculovirus genome containing a partial deletion of ORF1629 and an expression transfer vector containing the target gene; this restores the replicative capacity of the virus within the insect cells (10). In this chapter, we will describe the use of the BaculoGold system as a simple and a straightforward tool for producing recombinant viruses infecting *Spodoptera frugiperda* (Sf9) or *Trichopulsia ni* (High5) insect cells. We describe the procedure, starting from the cloning of a gene of interest into an expression transfer baculovirus vector, followed by virus production, protein expression, and large-scale production. Following our protocols, a high titer of viruses is produced, which is sufficient for protein production. Yet, the procedure takes about 4 weeks from cloning to protein production. We also indicate the parameters that should be taken into consideration when using the baculovirus system. Examples are given to illustrate the main principles.

2. Materials

2.1. Instruments

1. Incubator shaker, Innova 4230 floor-stackable refrigerated incubator shaker (New Brunswick Scientific).
2. Laminar flow hood.
3. Inverted Microscope.
4. Cell counter, Countess[®] automated cell counter (Invitrogen, Carlsbad, CA).
5. Refrigerated bench top centrifuge (Eppendorf 5702).
6. Liquid nitrogen tank.
7. QuixStand benchtop system for protein filtration and concentration (GE Healthcare) equipped with Hollow fiber cartridge (GE Healthcare).

2.2. Insect Cells

1. Sf9 (Invitrogen): Ovarian cell line derived from the fall armyworm, *Spodoptera frugiperda* – Recommended for transfection, amplification of viruses, and expression of intracellular recombinant proteins.
2. High5 (H5) (Invitrogen): Cell line derived from egg cells from cabbage looper, *Trichopulsia ni* – Recommended for expression of secreted recombinant proteins.

2.3. Media

1. For Sf9 –Sf900 II SFM (Invitrogen) plus 2% FBS (Gibco) supplemented with 100 µg/mL Penicillin/Streptomycin solution (Biological Industries), and 0.1% Pluronic® F-68 (Gibco, Invitrogen) to decrease membrane shearing (see Note 1).
2. For High5 (H5): Express Five SFM (Invitrogen) supplemented with 18 mM L-glutamine, 200 mM solution plus 10 µg/mL gentamycin sulfate, 50 mg/mL solution.

2.4. Vectors

1. pVL1392/1393 baculovirus transfer vector set (BD Biosciences) for intracellular protein expression. The cloned gene is under the control of the *polyh* promoter, which is responsible for high levels of expression during the very late stages of infection (see Notes 2 and 3).
2. pAcGP67B baculovirus transfer vector (BD Biosciences) for secreted protein expression. The cloned gene is under the control of the *polyh* promoter and fused to the GP67 (envelope surface glycoprotein of AcNPV) secretion signal sequence to transport the recombinant protein into the secretory pathway (see Note 3).

2.5. Flasks for Growing Cells

Adherent cells are propagated in T25, T75, or T182 tissue culture flasks. Flasks can be used from various vendors (e.g., Greiner).

Cells grown in suspension are propagated in 50 mL, 250 mL, 1 L, and 2 L flasks. Flasks can be used from various vendors (e.g., Corning) (see Note 4).

1. 12-well plate.
2. T25 flask, TC treated, sterile, plug seal cap, 50 mL, canted neck.
3. T75 flask, TC treated, sterile, plug seal cap, 250 mL, and canted neck.
4. T182 flask, TC treated, sterile, plug seal cap, 650 mL, canted neck.
5. Erlenmeyer flask, 250 mL, polycarbonate Corning® (Sigma-Aldrich).
6. Erlenmeyer flask, 500 mL, polycarbonate Corning® (Sigma-Aldrich).
7. Erlenmeyer flask, 1 L, polycarbonate Corning® (Sigma-Aldrich).
8. Erlenmeyer flask, 2 L, ml polycarbonate Corning® (Sigma-Aldrich).

2.6. Chemicals and Miscellaneous Reagents

1. Trypan blue solution.
2. Transfection buffer A and B set (BD Biosciences).
3. Ni-NTA agarose beads (Qiagen or equivalent).
4. Protease inhibitor cocktail (Set IV, EMD Chemicals, Inc).
5. Vivaspin concentrators (Sartorius) covering volumes from 100 μ L to 100 mL with MW cut off appropriate for the molecular weight of individual target protein.
6. SnakeSkin[®] Pleated Dialysis Tubing with various Molecular Weight Cut Offs (MWCO), depending on the molecular weight of the target protein (Thermo Scientific).
7. Protein gel electrophoresis system (Mini-Protean 3 Cell from Bio-Rad or equivalent).
8. Rotator (Intelli-mixer from Rose Scientific, Alberta, Canada or equivalent).
9. Microcentrifuge tubes (1.5 mL).
10. 15 mL sterile conical tubes.

2.7. Buffer Solutions

1. Dulbecco's phosphate-buffered saline without calcium chloride and magnesium chloride (1 \times PBS).
2. Insect Cell Lysis Buffer (BD Biosciences).
3. Protein dialysis buffer: 20 mM Tris-HCl pH 8.0, 200 mM NaCl.
4. Ni-binding buffer: 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 15 mM Imidazole, pH 7.0.
5. Ni-NTA agarose beads should be washed three times in Ni-binding buffer prior to use as the beads are kept in ethanol solution.
6. Ni-wash buffer: 10 mM Tris-HCl pH 7.5, 0.1% NP-40, 300 mM NaCl, 10% glycerol, 15 mM Imidazole, pH 7.0.
7. Ni-elution buffer: 10 mM Tris-HCl pH 7.5, 0.1% NP-40, 200 mM NaCl, 10% glycerol, 400 mM Imidazole pH 7.0, and 1 μ l/mL protease inhibitor cocktail.
8. Protein sample buffer concentrated four times (SB \times 4). Stock solution of 10 ml SB \times 4 contains the following: 4.8 mL 0.5 M Tris-HCl pH 6.8, 0.8 g SDS, 4.0 mL glycerol, 0.4 mL β -mecaptoethanol and 8 mg bromophenol blue. Store in aliquots at -20° C.
9. GelCode Blue Reagent (Thermo Scientific-Pierce or equivalent).

3. Methods

The pipeline for protein production in baculovirus-infected insect cells is illustrated in Fig. 1. The protein can be directed to intracellular expression or produced in a secreted form, depending on the nature of the target protein. This decision will affect the choice of vector and the type of insect cells used for infection and protein production. Proteins that normally function intracellularly are expressed internally in Sf9 cells, while proteins that are function extracellularly are expressed in High5 cells and are secreted to the medium. The vectors differ according to the mode of expression. Proteins that are secreted to the medium contain either the authentic secretion signal or a secretion signal peptide derived from the vector, i.e., the GP67 (envelope surface glycoprotein of AcNPV) secretion signal sequence found in the pAcGP67 transfer vector. Many of the recombinant proteins contain a 6× His-tag either at the N- or C terminus of the target protein, in order to facilitate the purification procedure. The His-tag can be added during the cloning procedure, as part of the primers used for amplification, or it might be incorporated into the vector sequence. Similarly, a protease cleavage site (e.g., tobacco etch virus (TEV) cleavage site) can be added in a similar way if desired.

3.1. Propagation of Adherent Sf9 and High5 Cells

Insect cells and viruses are handled inside a sterile laminar flow hood. A detailed description of growth and maintenance of insect cultures is included in the laboratory manual published by Invitrogen (11). Guidance for sterile tissue culture work with cells and viruses can be found in Methods in Molecular Biology (12). Sf9 and High5 cells can grow as adherent cells or in suspension in flasks at 27°C in a nonhumidified environment. The cells should be passaged every 2–3 days (see Note 5).

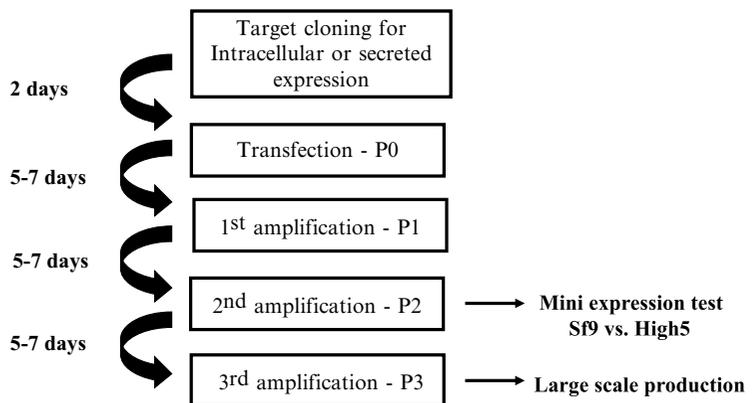


Fig. 1. Flow chart illustrating the entire process from gene cloning to protein production. Schematic presentation of the pipeline from cloning, virus production and protein production, and the time scale required for each step.

1. Seed cells from frozen stocks in T-25 or T-75 flasks containing 5 or 11 ml medium, respectively. Split cells once they reach confluence, every 2–3 days.
2. Tap on the side of the flask so that the cells become detached. Sf9 tends to attach more tightly than H5 cells.
3. Collect cells into a 15-mL conical tube.
4. Centrifuge cells at $500 \times g$ for 5 min at RT.
5. Aspirate medium. Resuspend cell pellet in 2 mL of fresh medium.
6. Count cells. Transfer 100 μ L from the cell suspension into a 1.5-mL test tube and add 900 μ L of medium (1:10 dilution). Transfer 100 μ L from the diluted cell suspension into a new 1.5-mL test tube and add 100 μ L of Trypan blue (1:1 dilution) for cell viability staining.
7. Count cells manually using a hemacytometer, or automatically using a Cell Counter (e.g., Countess[®], Invitrogen) (see Note 6).

3.2. Propagation of Suspension Cells

Cells can be propagated in suspension using different sizes of shaking flasks ranging from 50 mL to 2 L. Cells are grown in an incubator at 27°C with shaking at ~100 rpm. Sf9 cells are initially seeded at 1.8×10^6 cells/mL, while H5 cells are seeded at a density of 1×10^6 cells/mL. Cells are counted directly from the suspension culture. According to the number of cells counted, the cell culture is diluted to the desired density in the appropriate culture flask (see Note 7). The volume of the medium should not exceed 1/3 of the flask volume to allow good aeration, and should not be less than 1/5 to avoid drying out. Cells are split every 2–3 days, multiplying 2.5–4-fold.

3.3. Cloning the Gene of Interest the into Baculovirus Transfer Vector

The target gene is expressed under the control of the polyhedrin (*polh*) promoter (13). Target genes can be cloned into the multiple cloning sites (MCS) by ligation dependent cloning (LDC) using restriction enzymes, or by using a ligation independent cloning (LIC) procedure, such as restriction free (RF) cloning (14). The pVL1392 and pVL1393 (pVL1392/1393) plasmids contain the same MCS in opposite orientations. The pAcGP67 vector includes a signal peptide, which allows secretion of the recombinant protein into the medium. The pVL1392/1393 and pAcGP67 do not contain any tags, fusion proteins or protease cleavage sites. These desired elements could be engineered into the vectors by performing various molecular manipulations (15). A 6 \times His-tag and introduction of a TEV cleavage site can be included in the primer design (see Note 8).

3.4. Transfection: Virus Production – P0

For transfection, seed 0.5×10^6 Sf9 cells per well of a 12-well plate in 1 mL of Sf9 medium. Allow the cells to attach to the plate for 15 min. at RT. Confirm attachment by microscopy. Transfection

can be performed using various transfection reagents suitable for insect cells (see Note 9). When using the transfection buffers provided by BD Biosciences, the following protocol is performed.

1. Mix 1.25 μL linearized BaculoGold virus DNA (BD Biosciences) with 1 μg of baculo transfer vector containing the gene of interest (from Subheading 3.3) in a 1.5-mL tube. Incubate for 5 min at RT inside the hood (see Note 10). Replace the medium covering the cells with 250 μL of transfection buffer A (BD Biosciences).
2. Add 250 μL of transfection buffer B (BD Biosciences) to the DNA mixture and mix by pipetting.
3. Add the DNA mixture dropwise to the cells.
4. Seal plate with Parafilm to reduce evaporation of medium. Transfer the plate into a closed container containing wet paper and incubate for 4–5 h at 27°C.
5. Carefully aspirate the transfection buffers and wash cells once with 1 mL Sf9 medium.
6. Incubate the cells with 1 mL of Sf9 medium for 5–7 days at 27°C in a closed container with wet paper to prevent evaporation of the medium. Change the wet paper every 2 days.
7. Collect medium covering the cells. Spin tube at 1,000 $\times g$ for 10 min. Transfer the supernatant to a new tube and discard the pellet. The supernatant contains viruses that are referred as the P0 stock virus. Keep the virus suspension at 4°C covering the tube with aluminum foil.

3.5. Virus

Amplification: P1, P2, and P3 Stocks

1. For P1 amplification, seed 4×10^6 Sf9 cells in 5 mL of Sf9 medium into a T25 flask. Allow cells to attach to the flask for 15 min at RT inside the hood.
2. Add 100–200 μL of virus stock from P0 to the medium (see Note 11).
3. Incubate flasks at 27°C for 5–7 days. Infected cells should stop dividing, becoming larger, rounded, and shiny with granules (Fig. 2).
4. Collect medium containing viruses and centrifuge at 1,000 $\times g$ for 10 min at 4°C.
5. Keep P1 viral stock in the dark at 4°C (see Note 12).
6. For P2 amplification, seed 1.2×10^7 Sf9 cells in 12 mL of medium in a T75 flask.
7. Add 100–200 μL of virus P1 stock into the medium (see Note 11).
8. Incubate flasks at 27°C for 5–7 days.
9. Collect medium containing viruses and centrifuge at 1,000 $\times g$ for 10 min at 4°C.

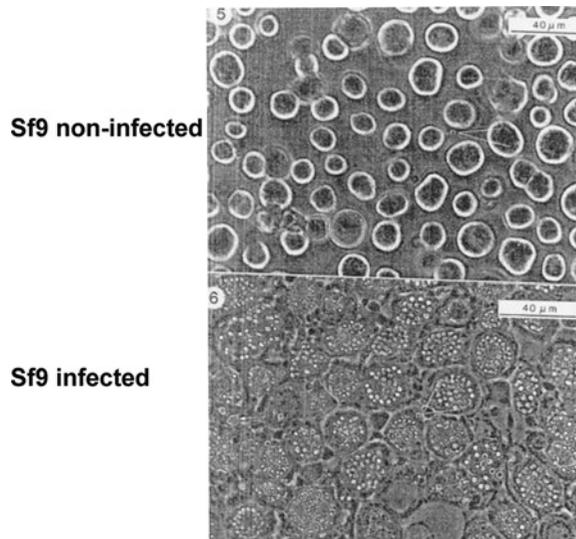


Fig. 2. Photomicrograph of uninfected Sf9 cells and Sf9 cells infected with recombinant viruses. Sf9 cells were infected with 100 μ L of recombinant virus supernatant and photographed 5 days postinfection.

10. Keep P2 viral stock in the dark at 4°C (see Note 12). P2 stock viruses can be used to check protein expression in small-scale assays.
11. For P3 amplification, seed 2.9×10^7 Sf9 cells in 25 mL of medium into T182 flask.
12. Add 100–200 μ L of viruses from P2 into the medium (see Note 11).
13. Incubate flasks at 27°C for 5–7 days.
14. Collect medium containing viruses and centrifuge at $1,000 \times g$ for 10 min at 4°C.
15. Keep P3 viral stock in dark at 4°C (see Note 12). P3 virus stocks are used for large-scale production (see Note 13).

3.6. Virus Infection for Protein Expression

3.6.1. Infection of Sf9 or H5 Cells with Recombinant Viruses

Small-scale protein expression

1. Seed 2×10^6 /mL or 1.2×10^6 /mL of Sf9 or H5 cells, respectively, in 11 mL medium into a T-75 flask. Let cells attach to the flask for 15 min at RT inside the hood.
2. Add 100 μ L of amplified recombinant P2 virus stock to the flask. For co-expression of partner proteins, 100 μ L of amplified recombinant virus encoding each protein partner is added (see Note 14).
3. Incubate flasks at 27°C for 3 days (see Note 15).

Large-scale protein expression

4. Large-scale production is usually performed in suspension culture. Seed Sf9 or H5 cells at 1.8×10^6 /ml or 1×10^6 /ml, respectively, in 2 L shaking flasks containing 700 mL medium per flask.
5. Add viruses to the flask at a 1:100 ratio (v/v) using the P3 amplified recombinant stock (see Note 14). For co-expression of protein partners, P3 amplified virus stocks of each protein partner are added. In cases where one of the protein partners is less robustly expressed compared to the other partners, a larger amount of P3 virus stock may be added from this partner.
6. Incubate flasks at 27°C for 3 days (see Note 15).

3.7. Protein Expression

3.7.1. Intracellular Expression

1. Collect floating and adherent cells into a 15-mL conical tube.
2. Centrifuge cells at $500 \times g$ for 5 min at 4°C.
3. Wash cells in cold $1 \times$ PBS without calcium chloride and magnesium.
4. Lyse cells in 1 mL of insect lysis buffer containing protease inhibitor cocktail, and transfer to 1.5-mL tubes.
5. Freeze and thaw cells twice in liquid nitrogen.
6. Centrifuge tubes at $12,000 \times g$ for 30 min at 4°C.
7. Separate supernatant from pellet. The supernatant contains the soluble proteins. Usually, over-expression of a soluble intracellular target protein can be detected by SDS-PAGE of the soluble fraction.

3.7.2. Secreted Protein Expression

1. Collect medium and spin at $1,000 \times g$ for 10 min at 4°C.
2. Concentrate medium using Vivaspin centrifugal devices (Sartorius Stedim Biotech, Goettingen, Germany) covering volumes from 100 μ l to 100 ml using the appropriate MWCO. For large volumes of medium up to 10 L, the QuixStand benchtop system is used.
3. For improved binding to the Ni column, dialysis of the medium may be performed in dialysis tubing with the appropriate MWCO against dialysis buffer for 3 h at 4°C, and a second dialysis O/N at 4°C.

Usually, over-expression of the target protein can already be detected using SDS-PAGE of the unfractionated collected medium. An example of Neuroligin 3 (NLG3) expression from the secreted medium is shown in Fig. 3.

3.8. Protein Purification by Batch Ni-Capture

1. Add 50 μ L washed Ni-beads to the supernatant.
2. Incubate protein samples with beads for 1 h at 4°C using a rotator adjusted for gentle mixing.

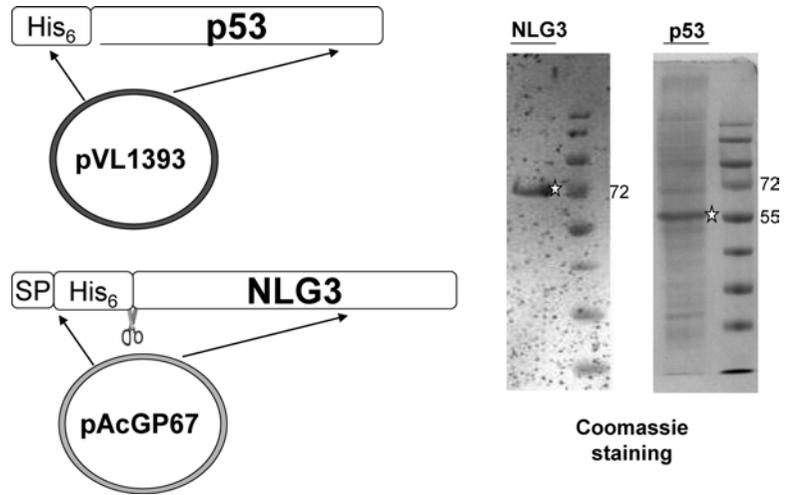


Fig. 3. Protein expression of p53 and Neuroligin 3 (NLG3). p53 and NLG3 were cloned into pVL1393 and pAcGP67 vectors, respectively. p53 is expressed intracellularly, whereas NLG3 is secreted. Both genes were designed to contain an N-terminal 6× HIS-tag followed by a TEV cleavage site incorporated into the primer sequences. The expressed proteins were purified on Ni agarose beads and analyzed using 10% SDS-PAGE. Asterisks indicate the expressed proteins.

3. Collect beads by slow centrifugation at $420 \times g$ (2,000 rpm in Eppendorf benchtop micro centrifuge) for 2 min at 4°C.
4. Discard supernatant. Wash protein-bound beads with 750 μ L washing buffer and mix gently.
5. Centrifuge at $420 \times g$ for 2 min at 4°C.
6. Repeat steps 4 and 5.
7. Elute bound proteins in 50 μ L of elution buffer. Incubate for 5 min on ice. Mix gently, and centrifuge at $420 \times g$ for 2 min at 4°C.
8. Transfer eluted protein to a new 1.5-mL tube and add 4× protein sample buffer.
9. Heat samples to 95°C for 5 min.
10. Analyze proteins on SDS-PAGE (see Notes 16 and 17).

4. Notes

1. 2% Fetal bovine serum (FBS) (v/v) is added to the medium of SF900 II SFM for Sf9 propagation, transfection, virus amplification and intracellular expression. FBS is omitted from the medium for extracellular expression of secreted protein. The presence of FBS in the medium may mask the expression of the secreted target protein and may interfere with the purification procedure.

2. pVL1393 and pVL1392 baculovirus transfer vectors contain the same MCS in opposite orientations.
3. A His₆ tag may be added at the N-terminus or C terminus of the protein to facilitate protein purification. A TEV protease recognition cleavage site (ENLYFQG) may be introduced to remove N-terminal fusion part of the target protein. Introduction of the TEV site to the C-terminal part of the target protein is not recommended. In the later case six unrelated amino acids are left at the C-terminal part of the protein.
4. Flasks can be reused several times following thorough cleaning with distilled water only, and autoclaving.
5. Keep a T75 flask for maintenance of adherent cells for both the Sf9 and H5 lines. Use the maintenance cells to expand cultures for large-scale production in suspension.
6. Cell cultures with viability below 85% should be discarded and new frozen cells should be thawed and propagated. For transfection and infection, cells should be at least 90% viable.
7. Under the conditions specified, Sf9 cells multiply 2.5–3-fold and H5 cells multiply three- to fourfold every 2–3 days.
8. Examples of primer sequences for restriction free (RF) cloning of N-terminal His or His-TEV into pAcGP67, downstream to the signal peptide, are shown in Table 1.
9. There are several commercial DNA transfection reagents suitable for insect cells. We tested the Insect Genejuice[®] (Novagen) and polyJet[™] DNA (SignaGen Laboratories) reagents, and both were efficient.
10. In case the plasmid DNA prep is from a mini prep source of low quality, it is recommended that the plasmid DNA is

Table 1
Primer design for the restriction free (RF cloning)

| Primer | Sequence |
|-----------------|--|
| pAcGP67bFHis | <u>5'-GCCTTTGCGGCGGATCTTGGATCCCATCATCATCAT</u> <u>CATCAC</u> (X ₁₉₋₂₄) |
| pAcGP67bFHisTev | <u>5'-GCCTTTGCGGCGGATCTTGGATCCCATCATCATCAT</u> <u>CATCAC TCTGAAAACCTGTACTTCCAGGGT</u> (X ₁₉₋₂₄) |
| pAcGP67bRFR | <u>5'-GATCTGCAGCGCCGCTCCGGAATTC</u> (X ₁₉₋₂₄) |

Design of three different primers are shown. pAcGP67bFHis and pAcGP67bFHisTev are forward primers, whereas, pAcGP67bRFR is a reverse primer. The two forward primers are designed to incorporate 6× His-tag at the N-terminal of the target protein (marked in bold). pAcGP67bFHisTev primer was designed to include, as well, a TEV recognition site (marked in italic letters). Gene-specific nucleotides are designated by X, ranging in size according to the Melting Temperature (T_m) of the target-specific sequence (typically size 19–24 nucleotides). In the example shown, plasmid-overlapping sequences (underlined) contain 26 and 24 bp, for the reverse and forward primers, respectively

incubated at 75°C for 20 min to denature residual proteins which could interfere with the transfection.

11. In general, 100–200 μL of virus stocks are used for virus amplifications. If there are no signs of infection (cells did not stop multiplying, and did not become enlarged, rounded, and shiny with granules), then a larger virus dose should be used, i.e., 500–750 μL . In contrast, when too high a dose of virus is used for amplification, the cells start to float 3 days postinfection.
12. Viral stocks can be kept at 4°C for many years. However, in case the virus titer drops during storage, or if more virus stock is required, additional virus can be produced from previous passages (from P0 to P1, from P1 to P2, from P2 to P3). Virus stocks can be frozen in liquid nitrogen and stored at –80°C with 20% (*v/v*) glycerol.
13. Using this procedure, the P3 stock virus roughly corresponds to a multiplicity of infection (MOI) of 5. The optimal virus dose for infection of culture to induce protein expression depends on the virus titer. However, for initial expression experiments, a ratio of virus to culture volume of 1:100 is used.
14. Since the exact titer of stock viruses is not determined in this simplified protocol, it is recommended that three different dilutions of viruses are tested (1:50, 1:100, 1:200) for insect cell infection for protein expression.
15. Since there is variation in the timing of the peak of expression for different proteins, we also recommend testing three different time points for expression: 2, 3, and 5 days.
16. In cases where protein expression is low and it is difficult to visualize the expressed protein by Coomassie staining, a Western blot is recommended, using target-specific antibodies or antibodies directed against a specific tag (i.e., His, GST, etc.).
17. The ultimate verification of the protein identity is through mass spectrometry.

Acknowledgments

We thank Prof. J. Sussman, Prof. I. Silman, Prof. G. Schreiber and Prof. Yigal Burstein for their continuous support throughout the study. This research was supported by the European Commission Sixth Framework Research and Technological Development Program ‘SPINE2-COMPLEXES’ Project, under contract No. 031220; a grant from the Israel Ministry of Science, Culture, and Sport to the Israel Structural Proteomics Center; the Divadol Foundation; and the Neuman Foundation.

References

1. Posse RD (1997) Baculoviruses as expression vectors. *Curr Opin Biotechnol* 8: 569–572
2. Luckow VA, and Summers MD (1988) Signals important for high-level expression of foreign genes in *Autographa californica* nuclear polyhedrosis virus expression vectors. *Virology* 167: 56–71
3. Fraser MJ (1986) Ultrastructural observations of virion maturation in *Autographa californica* Nuclear Polyhedrosis virus infected *Spodoptera frugiperda* cell cultures. *J Ultrastruct Mol Struct Res* 95: 189–195
4. Summers MD, Anderson DL (1972) Characterization of deoxyribonucleic acid isolated from the granulosis viruses of the cabbage looper, *Trichoplusia ni* and the fall armyworm, *Spodoptera frugiperda*. *Virology* 50: 459–471
5. Kost TA, Condreay JP, Jarvis DL (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotechnol* 23: 567–575
6. Chen X, Liu H, Shim AH, Focia PJ, He X (2008) Structural basis for synaptic adhesion mediated by neuroligin-neurexin interactions. *Nat Struct Mol Biol* 15: 50–56
7. Sun XZ, Nguyen J, Momand J (2003) Purification of recombinant p53 from Sf9 insect cells. *Methods Mol Biol* 234: 17–28
8. Jarvis DL (2009) Baculovirus-insect cell expression systems. *Methods Enzymol* 463: 191–222
9. Luckow VA, Lee SC, Barry GF, Olins PO (1993) Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. *J Virol* 67: 4566–4579
10. Vialard JE, Richardson CD (1993) The 1,629-nucleotide open reading frame located downstream of the *Autographa californica* nuclear polyhedrosis virus polyhedrin gene encodes a nucleocapsid-associated phosphoprotein. *J Virol* 67: 5859–5866
11. Invitrogen (2002) Growth and maintenance of insect cell lines. Invitrogen Life Technologies, version K, Paisley, UK
12. Shrestha B, Smeets C, Gileadi O (2008) Baculovirus expression vector system: an emerging host for high-throughput eukaryotic protein expression. *Methods Mol Biol* 439P: 269–289
13. Pennock GD, Shoemaker C, Miller LK (1984) Strong and regulated expression of *Escherichia coli* beta-galactosidase in insect cells with a baculovirus vector. *Mol Cell Biol* 4: 399–406
14. van den Ent F, Lowe J (2006) RF cloning: a restriction-free method for inserting target genes into plasmids. *J Biochem Biophys Methods* 67: 67–74
15. Unger T, Jacobovitch Y, Dantes A, Bernheim R, Peleg Y (2010) Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. *J Struct Biol* 172: 34–44

Systems for the Cell-Free Synthesis of Proteins

Lei Kai, Christian Roos, Stefan Haberstock, Davide Proverbio, Yi Ma, Friederike Junge, Mikhail Karbyshev, Volker Dötsch, and Frank Bernhard

Abstract

We describe a system for the cell-free expression of proteins based on extracts from *Escherichia coli*. Two reaction configurations, batch and continuous exchange, are discussed and analytical scale as well as preparative scale setups are documented. Guidelines for the systematic development and optimization of cell-free expression protocols are given in detail. We further provide specific protocols and parameters for the cell-free production of membrane proteins. High-throughput screening applications of CF expression systems are exemplified as new tools for genomics and proteomics studies.

Key words: Cell-free expression, Membrane proteins, Expression optimization, Detergents, Liposomes, Cell extracts

1. Introduction

Cell-free (CF) expression systems have emerged as important tools for the production of proteins at high levels (1). Besides general applicability to the production and labeling of proteins, these techniques are particularly successful for producing difficult proteins, such as integral membrane proteins (MPs), or toxic proteins. In conventional cellular systems, failure to overproduce MPs is frequently caused by overloading cellular transportation and translocation mechanisms, as well as by the toxicity of the synthesized proteins. CF expression generally eliminates these principal problems and in addition, facilitates the downstream harvesting and isolation of MPs (2, 3). Furthermore, the open nature of CF reactions provides a valuable and unique way of directly modifying

expression by introducing potentially beneficial features such as ligands, cofactors, or even artificial hydrophobic environments.

CF systems have been reported based on extracts of diverse origin, such as insect cells, protozoa, human cells, rabbit reticulocytes, wheat germ or *E. coli* cells (1, 4–7). So far however, only wheat germ and *E. coli* CF extracts have been optimized for routine preparative scale protein production in milligram amounts. This chapter focuses on *E. coli*-based systems, as protocols for preparation of extracts are highly efficient and reliable.

Two different CF expression configurations are described. The batch CF system is performed in a single reaction compartment and is ideal for automated throughput and screening applications, in particular in combination with linear DNA templates generated by PCR (8, 9). The continuous exchange cell-free (CECF) configuration requires two compartments, separated by a semipermeable dialysis membrane, and is preferred for preparative scale protein production, although high-throughput applications are also possible (10–13). The CECF reaction mixture (RM) compartment contains all the high molecular weight substances, such as enzymes and nucleic acids. The second (and larger) feeding mixture (FM) compartment contains low molecular weight precursors, such as amino acids, energy-generating compounds, and nucleotides. The FM introduces fresh precursors into the RM for a certain time period, while inhibitory breakdown products are continuously removed from the RM into the FM. Expression efficiencies and protein yields in the CECF configuration are therefore significantly increased.

CF expression offers completely new ways of producing MPs (2, 3). In the P-CF (precipitate forming CF expression) mode, no hydrophobic environments are present in the reaction and the synthesized MPs will instantly precipitate. P-CF precipitates clearly differ from cellular inclusion bodies, as functional MPs may be obtained by simple detergent solubilization procedures (14, 15). In the D-CF (detergent-based CF expression) mode, detergent micelles provide hydrophobic environments that maintain the solubility of synthesized MPs after expression. In the L-CF (lipid-based CF expression) mode, lipid bilayers are supplied for the insertion of expressed MPs. Some MPs may insert directly into artificial liposomes, while others may require additionally supplied compounds such as translocons (16, 17). The three MP production modes can only be performed by CF expression and this technique will therefore rapidly emerge as an indispensable tool for MP research.

CF expression is very successful because the complexity of MP production in natural cellular systems is reduced to the essential transcription/translation process. Initial failure to produce sufficient quantities of a target protein can thus be addressed by systematically optimizing a few crucial parameters as described in this chapter. Success rates of more than 80% expression can be obtained routinely (18). For structural and functional studies, sufficient

quantities of purified protein, even with complicated labeling schemes, can be isolated from of a few milliliters of RM volumes in less than 2 days (19, 20).

2. Materials

2.1. General Materials

1. Fermenter for 5–10 l of culture volume (B. Braun Biotech).
2. French press or other high pressure cell-disruption equipment.
3. Photometer.
4. Centrifuges and set of rotors (Sorvall or Kontron).
5. Dialysis tubes, 30 kDa MWCO (Spectrum).
6. Mini extruder (Avanti Polar Lipids).
7. Ultra sonication water bath.
8. Thermo shaker for incubation.
9. Chromatographic system (GE Healthcare).
10. Q-Sepharose column.
11. Ultrafiltration devices, 30 kDa MWCO.
12. Plasmid and PCR product purification kits (Qiagen, Machery & Nagel).
13. V-shaped 96-well microplates and 24-well microplates (Greiner).
14. Dark microplate (96 F Nunclon Delta Black Microwell SI).

2.2. Materials for CF Expression Reaction

1. 50× Complete[®] protease inhibitor cocktail (Roche Diagnostics) 1 tablet/mL of MilliQ water.
2. Amino acid mixtures: 4 mM or 8 mM of each of the 20 natural amino acids (see Note 1).
3. RCWMDE mixture: c 16.7 mM of each amino acid.
4. 1 M acetyl phosphate lithium potassium salt (AcP) (Sigma-Aldrich), adjusted to pH 7.0 with KOH.
5. 1 M phospho(enol)pyruvic acid K⁺ salt (PEP) (Sigma-Aldrich), adjusted to pH 7.0 with KOH.
6. NTP mixture: 90 mM ATP, 60 mM CTP, 60 mM GTP, and 60 mM UTP, adjusted to pH 7.0 with NaOH.
7. Pyruvate kinase (Roche Diagnostics), 10 mg/ml.
8. RiboLock[®] RNase inhibitor (Fermentas), 40 U/μl.
9. Total *E. coli* tRNA (Roche Diagnostics), 40 mg/ml.
10. Folinic acid, Ca²⁺ salt, 10 mg/ml (Sigma-Aldrich).
11. Polyethylene glycol 8000 (PEG 8000), 40% (w/v).
12. 4 M potassium acetate (KOAc).
13. 2.4 M Hepes, 20 mM EDTA, pH 8.0 adjusted with KOH.

14. 500 mM 1,4 dithiothreitol (DTT).
15. *E. coli* S30 extract, store frozen at -80°C (see Subheading 3.1).
16. T7-RNA polymerase (T7RNAP), store frozen at -80°C (see Subheading 3.2).
17. Template DNA (plasmid DNA or linear PCR products) 200–500 ng/ μL (see Subheading 3.3).
18. Reaction container: Analytical and preparative scale reaction container (see Fig. 1 and Subheading 3.4.2).
19. D-tube containers, 12–14 kDa MWCO (Merck Biosciences).
20. Slide-A-Lyzer, 10 kDa MWCO (Pierce).
21. Dialysis tubes, 12–14 kDa MWCO.
22. 10 \times Premix: 15 mM putrescine, 15 mM spermidine, 2.5 M K^+ -glutamate, 100 mM NH_4^+ -glutamate, 100 mM Mg^{2+} -glutamate, 40 mM Na^+ -oxalate, 330 mM Na^+ -pyruvate, 340 $\mu\text{g}/\text{ml}$ folinic acid, 10 mM DTT, 5.3 mM NAD^+ .
23. 30 mM CoA- Na^+ .
24. GFP assay buffer: 20 mM Tris, 150 mM NaCl, pH 7.8.

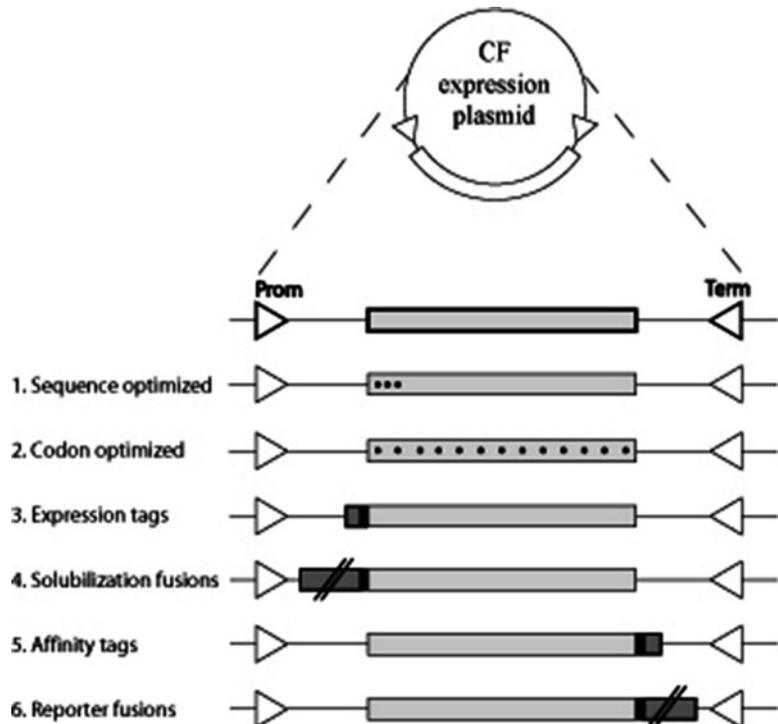


Fig. 1. DNA template design strategies for CF expression. Optional modifications are indicated. Coding region is given as *light grey bar*. *Open triangles* indicate the T7 promoter and terminator.

2.3. Materials for S30 Extract and T7RNAP Preparation

1. 40× S30-A/B buffer: 400 mM Tris–acetate pH 8.2, 560 mM Mg(OAc)₂, 2.4 M KCl. Supplement 1× S30-A buffer with 6 mM β-mercaptoethanol. Supplement 1× S30-B buffer with 1 mM DTT and 1 mM PMSF.
2. 40× S30-C buffer: 400 mM Tris–acetate pH 8.2, 560 mM Mg(OAc)₂, 2.4 M KOAc. Supplement 1× S30-C buffer with 0.5 mM DTT.
3. 2× YTPG medium: 22 mM KH₂PO₄, 40 mM K₂HPO₄, 100 mM glucose, tryptone 16 g/L, yeast extract 10 g/L, NaCl 5 g/L.
4. LB medium: Peptone 10 g/L, yeast extract 5 g/L, NaCl 10 g/L.
5. Buffer T7RNAP-A: 30 mM Tris–HCl pH 8.0, 50 mM NaCl, 10 mM EDTA, 10 mM β-mercaptoethanol, 5% glycerol.
6. Buffer T7RNAP-B: 30 mM Tris–HCl pH 8.0, 50 mM NaCl, 1 mM EDTA, 10 mM β-mercaptoethanol, 5% glycerol.
7. Buffer T7RNAP-C: 30 mM Tris–HCl pH 8.0, 1 M NaCl, 1 mM EDTA, 10 mM β-mercaptoethanol, 5% glycerol.
8. Buffer T7RNAP-D: 10 mM K₂HPO₄/KH₂PO₄ pH 8.0, 10 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 5% glycerol.
9. 20% streptomycin sulfate.
10. *E. coli* strain for extract preparation, e.g., A19 (*E. coli* Genetic Stock Center, New Haven, CT) or BL21 (Merck Biosciences).
11. BL21 (DE3) Star x pAR1219 for T7RNAP preparation (21).

2.4. Detergents and Lipids

Stock solutions of all detergents were prepared in MilliQ water (see Table 1). Stocks are stored at 4 or at –20°C for longer time. Repeated thawing/freezing cycles should be avoided.

1. *n*-dodecyl-β-D-maltoside (DDM, 10%) (AppliChem).
2. *n*-dodecylphosphocholine (Fos-12, 10%), *n*-decyl-β-D-maltoside (DM, 10%) (Anatrace).
3. 1-myristoyl-2-hydroxy-sn-glycero-3-(phospho-rac(1-glycerol)) (LMPG, 5%), 1-palmitoyl-2-hydroxy-sn-glycero-3-(phospho-rac(1-glycerol)) (LPPG, 5%), 1-myristoyl-2-hydroxy-sn-glycero-3-phosphocholine (LMPC, 5%), 1,2-dihexanoyl-sn-glycero-3-phosphocholine (DHPC, 5%) (Avanti Polar Lipids).
4. *n*-octyl-β-D-glucopyranoside (β-OG, 10%), sodium dodecyl-sulfate (SDS, 10%) (Roth).
5. Digitonin (10%), Triton X-100 (10%), polyethylene-(23)-laurylether (Brij35, 15%), polyoxyethylene-(20)-cetylether (Brij58, 15%), polyoxyethylene-(20)-stearylether (Brij78, 15%), polyoxyethylene-(20)-oleylether (Brij98, 15%) (Sigma-Aldrich).

Table 1
Detergents for CF MP expression modes

| Detergent | Mass (Da) | CMC (mM) | Working concentration (%/x CMC) |
|-------------|-----------|----------|---------------------------------|
| <i>P-CF</i> | | | |
| SDS | 288 | 7–10 | 1/3–4 |
| LMPG | 479 | 0.2 | 1/420 |
| LPPG | 507 | 0.02 | 1/986 |
| Fos-12 | 352 | 1.5 | 2/38 |
| Fos-14 | 379 | 0.12 | 2/439 |
| LMPC | 468 | 0.04 | 1/535 |
| LDAO | 229 | 2 | 2/44 |
| β -OG | 292 | 18 | 2–5/3.5–9 |
| <i>D-CF</i> | | | |
| Brij35 | 1,200 | 0.08 | 1.5/156 |
| Brij58 | 1,123 | 0.075 | 1.5/178 |
| Brij78 | 1,152 | 0.046 | 1/189 |
| Brij98 | 1,150 | 0.025 | 0.2/70 |
| Digitonin | 1,229 | 0.73 | 0.4/4.5 |
| DDM | 511 | 0.17 | 0.1/15 |
| Triton X100 | 650 | 0.3 | 0.1/7 |
| DHPC | 482 | 1.4 | 2–5/30–75 |
| DM | 483 | 1.8 | 0.2/2 |

6. *E. coli* polar lipid extract, 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC), 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) (Avanti Polar Lipids).

3. Methods

3.1. Preparation of S30 Extract

High quality cellular extracts should be prepared by cultivating cells in fermenters with good aeration. First, a growth curve should be determined and an efficient chilling technique of the fermenter broth should be established. Cells must be harvested in mid-log phase and should be chilled down to 10–14°C before centrifugation. Several strains and modifications for S30 extract preparations are described (5). We describe a basic protocol for the preparation of high quality S30 extracts for the CF expression of both soluble and membrane proteins.

3.1.1. Cell Fermentation

1. Inoculate 10 l of sterilized YTPG medium into a fermenter with 100 mL of a fresh *E. coli* overnight culture, e.g., strain A19.
2. Incubate the cells at 37°C with intensive aeration and stirring.

3. Monitor the cell growth by continuously measuring the OD_{600} .
4. Start to chill the cell broth before the cells reach mid-log phase (see Note 2) (approximately OD_{600} 3–5).
5. Quickly cool the fermenter broth to 14–10°C.
6. Harvest the cells by centrifugation at $7,000 \times g$ for 15 min at 4°C.
7. Keep the cell pellets at 4°C for all following steps. Alternatively, freeze the cell paste in thin layers wrapped in aluminum foil at –80°C for later processing.

3.1.2. Cell Extraction

1. Gently, but completely, resuspend the cell pellet in approximately 300 mL of pre-cooled S30-A buffer and centrifuge at $7,000 \times g$ for 10 min at 4°C. Discard supernatant and repeat this washing step twice. Extend the final centrifugation step to 30 min.
2. Discard the supernatant and resuspend the cell pellet in 110% (v/w) pre-cooled S30-B buffer.
3. Disrupt the cells by a high pressure cell disrupter, e.g., French press at 1,000 psi.
4. Centrifuge the lysate at $30,000 \times g$ for 30 min at 4°C. Transfer the supernatant in a fresh tube and repeat the centrifugation step.
5. Harvest the supernatant and adjust stepwise to a final concentration of 400 mM NaCl. Gently mix and incubate at 42°C for 45 min in a water bath (see Note 3).
6. Dialyze the turbid extract overnight against 100-fold excess of pre-cooled S30-C buffer using a dialysis membrane with a 12–14 kDa MWCO and with two changes of dialysis buffer.
7. Centrifuge the extract at $30,000 \times g$ for 30 min at 4°C. Harvest the supernatant and dispense suitable aliquots into plastic tubes.
8. Shock-freeze the aliquots in liquid nitrogen and store at –80°C. Aliquots are stable at –80°C for at least 1 year but should not be re-frozen.
9. The final total protein concentration of the S30 extract should be between 20 and 40 mg/mL. Each new batch of extract should be adjusted to its optimal concentration of Mg^{2+} (12–25 mM) and K^+ (250–350 mM) ions. Perform Mg^{2+} concentration screens in 2 mM steps and K^+ concentration screens in 20 mM steps while monitoring GFP expression.

3.2. T7-RNA Polymerase Preparation

T7RNAP is one of the most expensive components of CF reactions and is required in high concentrations. It can be overproduced in *E. coli* and after purification by ion exchange chromatography, approximately $0.5\text{--}1 \times 10^6$ units can be obtained out of one liter of culture.

1. Inoculate 1 L of LB medium 1:100 with a fresh overnight culture of strain BL21 (DE3) Star x pAR1219. Grow the cells on a shaker at 37°C until they reach an OD₆₀₀ of 0.6–0.8; induce T7RNAP production by the addition of 1 mM IPTG. Incubate the cells for further 5 h and harvest by centrifugation at 8,000 × *g* for 15 min at 4°C. The cell pellet can be stored at –80°C until further use.
2. Resuspend the cell pellet in 30 mL of T7RNAP-A buffer and disrupt the cells by one passage through a French press cell at 1,000 psi or by sonication. Remove cell debris by centrifugation at 20,000 × *g* for 30 min at 4°C. All subsequent purification steps should be performed at 4°C.
3. Adjust the supernatant to a final concentration of 4% streptomycin sulfate by stepwise addition of a 20% stock solution. Mix gently, incubate on ice for 5 min and centrifuge at 20,000 × *g* for 30 min at 4°C.
4. Load the supernatant on a 40 mL Q-sepharose column equilibrated with T7RNAP-B buffer and wash the column extensively with T7RNAP-B buffer.
5. Elute the T7RNAP with a gradient of 50–500 mM NaCl using T7RNAP-C buffer for 10 column volumes at a flow rate of 3–4 ml/min. Collect the fractions and analyze aliquots by SDS-PAGE (see Note 4).
6. Pool T7RNAP-containing fractions and dialyze against T7RNAP-D buffer over night. Adjust to a final concentration of 10% glycerol and concentrate the T7RNAP fraction to a total protein concentration of 3–4 mg/ml by ultrafiltration (see Note 4). Adjust to a final concentration of 50% glycerol and store aliquots at –80°C.
7. For each new batch of T7RNAP, perform concentration optimization screens in the CECF and batch configurations using, for example, green fluorescent protein (GFP) as monitor.

3.3. Design and Preparation of DNA Templates for CF Expression

As the *E. coli* CF setup is a coupled transcription/translation system, DNA is required as template. Insufficient (or no) production of a target protein is caused in most cases by problems with the central transcription/translation process. Appropriate template design and purity are thus crucial for successful CF expression. For screening and high-throughput purposes, linear templates generated by PCR may be used; the more stable circular plasmid DNA templates are recommended for preparative scale expression.

3.3.1. Strategies for Designing CF Expression Templates

Compared to conventional cellular systems, regulation of expression in *E. coli* CF systems is more straightforward; in principal, the latter requires only standard T7 regulatory sequences and requirements for the induction of expression need not to be considered.

However, a few additional options for template design are important in order to optimize expression efficiencies (see Fig. 1).

1. Regulatory sequences. Transcription of the described system is directed by the highly effective T7RNAP. Standard T7 promoters of common vectors such as the pET (Merck Biosciences) or pIVEX (RINA) series including ϵ -enhancer sequences ribosomal binding sites should be used. Transcription should be completed by a terminator region.
2. Purification tags. We recommend attaching a small purification/detection tag such as (His)₆₋₁₀-tags or a StrepII-tag to the C-terminal end of the target protein. Besides their use for purification, those tags could become valuable tools for the immunodetection of the target protein and verification of its full-length synthesis. In particular at the beginning of CF expression protocol developments, the expressed targets can often not be clearly detected with Coomassie-Blue staining (see Note 5). As target-specific antibodies are usually not available, the use of commercial antibodies directed against standard purification tags could be essential for western blot detection.
3. Expression tags. If poor or no expression of a target is observed, the improvement of the initiation of translation by modification of the 5' end of the coding region is highly recommended as a first approach to optimization. Enrichment in the overall AT content of approximately the first 10 codons by silent mutagenesis could significantly improve expression. Alternatively, the N-terminal end of the target protein could be modified by addition of small expression enhancing tags such as the T7-tag.
4. Fusion proteins. Fusions with larger reporter proteins could allow the fast monitoring of expression success and may accelerate protocol development. N-terminal fusion partners such as the maltose-binding protein could be used as combined an expression/purification tag. C-terminal fusions to GFP will allow the fast evaluation of expression success. GFP was already successfully used in cellular and in CF expression systems as an expression monitor for MPs (18, 22). Recognition sites for restriction proteases could be included in order to release fusion partners or tags after translation.
5. Codon usage. Coding sequences of target proteins should initially be analyzed for rare codons differing from the standard *E. coli* codon usage. Hot spots of rare codons could result in premature termination events which may be addressed by silent mutagenesis. If an overall high rare codon content is suspected to result in poor or incomplete expression or in misincorporation of amino acids, the increase or optimization of the supplied tRNA pool or the use of synthetic genes could be considered.

3.3.2. DNA Template Preparation

As vectors do not replicate in the CF system, plasmid compatibility and selection markers are not important. However, the supplied plasmid and the linear DNA templates must be of high quality.

1. Inoculate 100–200 mL of LB medium with the *E. coli* strain containing the desired plasmid and incubate overnight at 37°C on a shaker.
2. Plasmid templates should be purified by three step cell extraction followed by standard ion exchange chromatography columns. “Midi” or “Maxi” preparation protocols using commercial kits can be performed. However, “Mini” preparations do not provide sufficient quality.
3. Dissolve dried DNA in low volumes of MilliQ water. Template stock concentrations should be in between 0.2 and 0.5 mg/ml and can be stored at –20°C or at 4°C. Repeated freezing/thawing cycles may result in DNA precipitation and the concentration of such stocks should be checked again in case of problems.
4. Optional: The option of expressing proteins directly from linear PCR products in CF reactions is particularly useful for high-throughput applications. Different tags or fusion partners can be attached to the target protein by multistep PCR strategies. Regulatory sequences, such as the T7-promoter and T7-terminator, can be combined with any coding sequence by a two step overlap PCR approach (23). For reliable amplification of DNA templates, only proofreading polymerases should be used. The final PCR product should be purified with standard PCR purification kits. Due to decreased stability, higher molar concentrations of linear templates should be used, compared to plasmid templates, in order to maintain expression efficiency (see Fig. 2).

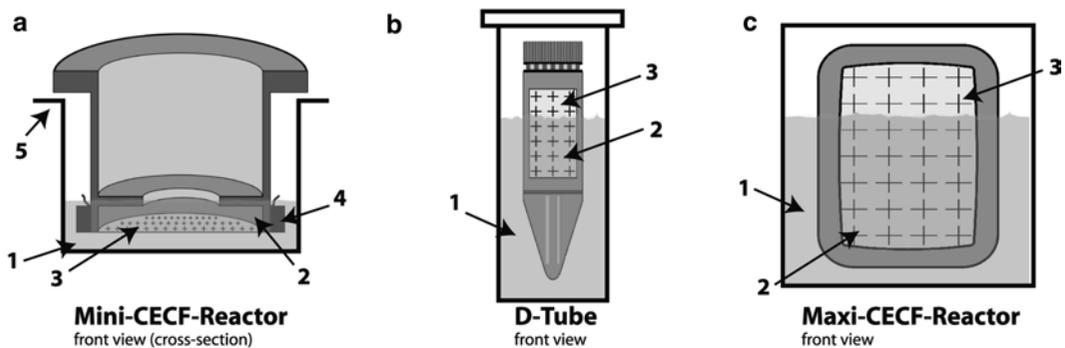


Fig. 2. The reaction container for CECF expression. (a) Mini-CECF-Reactor for RM volumes of 30–100 μ L. 24-well microplates are used as FM container. (b) Commercial D-Tube dialyzer for RM volumes of 50 μ L up to 3 mL. 2-mL Eppendorf tubes or 15–50-mL Falcon tubes can be used as FM container. (c) Maxi-CECF-Reactor. Commercial Slide-A-Lyzers are used for 1–3 mL RM volumes. Custom made Plexiglas boxes are used as FM container. 1, FM; 2, RM; 3, dialysis membrane; 4, Teflon ring; 5, 24-well microplate.

3.4. Setup of CF Configurations

Two CF reaction configurations, batch and CECF, are described. For batch reactions only a RM has to be prepared. Note that the batch RM and the CECF RM are not identical, as the two configurations have different basic compound compositions. If applicable, master mixtures of common compounds can first be prepared and subsequently aliquoted into the desired number of reaction containers. Screening compounds might then be added individually to the reactions. The volumes of master mixtures should be 110% of the calculated reaction volumes in order to compensate for potential volume losses during mixing and pipetting.

3.4.1. Batch Configuration

Standard 96-well microplates or reaction tubes can be used as reaction containers. Total volumes of $\geq 25 \mu\text{L}$ are recommended for batch screening reactions in 96-well microplates.

1. Prepare each individual compound as a stock solution (see Table 2).
2. In order to reduce pipetting time, a set of compounds can be combined in a 10 \times premix (see Table 2). The premix can be stored at -20°C (see Note 6).
3. Calculate individual compound volumes and prepare a pipetting scheme for the intended set of reactions (see Note 7).
4. Thaw the premix and other stock solutions on ice.
5. Combine all common compounds including the premix into a master mix. The premix and amino acid mix need to be resuspended before pipetting.

Table 2
Preparation of CF batch reactions

| Compound | Stock | Final |
|----------------------------|---------------------|----------------------------|
| Premix | 10 \times | 1 \times |
| Amino acid mix | 8 mM each | 2 mM each |
| PEP | 1 M | 30 mM |
| CoA-Na ⁺ | 30 mM | 0.26 mM |
| <i>E. coli</i> tRNA | 40 mg/ml | 0.17 mg/ml |
| T7RNAP | 1.4 mg/ml | 10 $\mu\text{g}/\text{ml}$ |
| 36 \times NTP-Mix | ATP: 90 mM | ATP: 2.5 mM |
| | C/G/UTP: 60 mM each | C/G/UTP: 1.7 mM each |
| Plasmid template | 0.2–0.5 mg/ml | 0.015 mg/ml |
| <i>E. coli</i> S30 extract | 100% | 24% |
| MilliQ water | X | Fill up to final volume |

6. Transfer appropriate aliquots of the master mix into the reaction containers, e.g., wells of a 96-well microplate.
7. Start the reactions by addition of the DNA template.
8. Seal the reaction container (e.g., microplates), with Parafilm. Incubate at 30°C (see Note 8) for 2–4 h with gentle shaking and then analyze the reactions.

3.4.2. CECF Configuration

Analytical scale reactions usually have a RM volume of 30–100 μL . For preparative scale reactions, we propose volumes of up to 3 mL, but higher volumes may be possible too. We further recommend a RM:FM ratio between 1:15 and 1:20 (see Note 9).

1. Calculate the individual compound volumes according to the desired number of reactions.
2. Prepare a common master mix RFM for the RM and FM (see Table 3).
3. Reconstitute RM and FM (see Table 4). The RFM master mix is vortexed briefly and an aliquot is added to the RM. The remaining volume is used for the FM.
4. RM and FM are completed by addition of MilliQ water.
5. The FM can be vortexed briefly. The RM should only be mixed by inverting or by pipetting up and down.
6. Optional: For screening reactions, common RM and FM master mixtures may be prepared and then aliquoted into the desired number of reaction containers. For compound screens comprising a series of analytical CECF reactions, master mixes are prepared with the lowest concentration of the screening compound. The RM and FM master mixes are then aliquoted according to the number of reactions and adjusted to the desired screening compound concentrations (see Note 10).
7. Add the RM and FM aliquots to the reaction containers. There are currently no commercial reaction containers available that are specifically designed for CF expression. We recommend customized containers made out of Plexiglas for analytical scale reactions (see Fig. 3). These Mini-CECF-Reactors are designed for RM volumes of 30–100 μL and can be used in combination with standard 24-well microplates with FM volumes of up to 1.5 mL. The Mini-CECF-Reactors hold the RM and are placed in the cavities of a 24-well plate holding appropriate volumes of FM. A piece of dialysis membrane is fixed to the Mini-CECF-Reactors with a Teflon ring (see Note 11). The dialysis membrane should be replaced for each new reaction while the Mini-CECF-Reactors are reusable many times (see Note 12). For preparative scale CF reactions, commercial Slide-A-Lyzer

Table 3
RFM master mix preparation for 1/16 mL CECF reaction

| Compound | Stock | Final concentration | Volume ^a (μL) |
|--|---------------|---------------------------|--------------------------|
| RCWMDE | 16.67 mM | 1 mM | 1,020 |
| Amino acid mix | 4 mM | 0.5 mM | 2,337.5 |
| Acetyl phosphate | 1 M | 20 mM | 340 |
| Phospho(enol)pyruvic acid | 1 M | 20 mM | 340 |
| 75 × NTP mix | 90 mM ATP | 1.2 mM | 226.7 |
| | 60 mM G/C/UTP | 0.8 mM | |
| 1,4 Dithiothreitol | 500 mM | 2 mM | 68 |
| Folinic acid | 10 mg/ml | 0.1 mg/ml | 170 |
| Complete [®] protease inhibitor | 50× | 1× | 340 |
| Hepes/EDTA buffer | 24× | 1× | 623.3 |
| Mg(oAc) | 1 M | 11.1 16, mM ^b | 274 |
| KOAc | 4 M | 110, 270, mM ^b | 382.5 |
| PEG 8000 | 40% | 2% | 850 |
| NaN ₃ | 10% | 0.05% | 85 |
| | | | Total: 7,057 |

^aVolumes are calculated for 1 mL RM and 16 mL FM = 17 mL RFM master mix

^bConcentration is subject to optimization and might need to be adjusted for each individual target. Volumes are calculated for final total concentrations of Mg²⁺ of 16 mM and K⁺ of 270 mM as additional amounts of 4.9 mM Mg²⁺ and 160 mM K⁺ are contributed by other compounds

devices (Pierce) that can hold up to 3 mL RM volumes may be used. We have designed Plexiglas Maxi-CECF-Reactors that perfectly combine with Slide-A-Lyzer devices to form the FM container (see Note 13). Exact blueprints of the Maxi-CECF-Reactor as well as of the Mini-CECF-Reactors have been published (24). Alternatives to the Mini- and Maxi-CECF-Reactors are the commercial D-tube dialyzers (Novagen) which are well suited for analytical and preparative scale CF reactions. The D-Tubes are available in different sizes and with different membrane MWCOs and can be used for analytical and preparative scale CF expression. The D-tube dialyzer holds the RM and need to be placed in a suitable tube holding the appropriate volume of FM. We recommend 2-mL Eppendorf tubes for the small 10–250 μL analytical scale D-tube dialyzer and 15–50 mL Falcon tubes for the larger preparative scale D-tube dialyzer

Table 4
Preparing CECF RM (1 mL) and FM (16 mL)

| Compound | Stock | Final concentration | Volume |
|----------------------------|---------------|---------------------|---------------|
| <i>FM</i> | | | |
| Master mix RFM | | | 6,642 μ L |
| S30-C puffer | | | 5,600 μ L |
| Amino acid mix | 1 \times | 0.35 \times | 2,000 μ L |
| MilliQ water | 4 mM | 0.5 mM | 1,758 μ L |
| | | | Total: 16 mL |
| <i>RM</i> | | | |
| Master mix RFM | | | 415 μ L |
| Pyruvat kinase | | | 4 μ L |
| tRNA (<i>E. coli</i>) | 10 mg/ml | 0.04 mg/ml | 12.5 μ L |
| T7RNAP | 40 mg/ml | 0.5 mg/ml | 35,7 μ L |
| Ribolock | 1.4 mg/ml | 0.05 mg/ml | 7.5 μ L |
| DNA template | 40 U/ μ l | 0.3 U/ μ l | 60 μ L |
| <i>E. coli</i> S30 extract | 0.2–0.5 mg/ml | 0.015–0.03 mg/ml | 350 μ L |
| MilliQ water | 1 \times | 0.35 \times | 115.3 μ L |
| | | | Total: 1 mL |

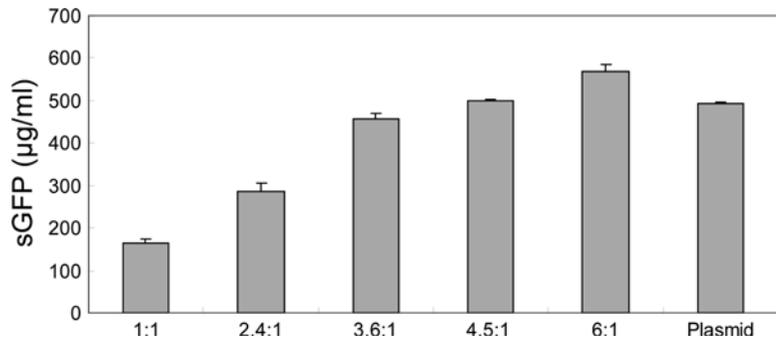


Fig. 3. Production of sGFP in the CF batch configuration with linear PCR fragments as templates. The indicated molar ratios of linear template to the plasmid template control have been analyzed. For the control with plasmid template a concentration of 15 μ g/mL was used.

(see Note 14). Small plastic boxes or beakers may further be used as the FM container for Slide-A-Lyzers. Appropriate pieces of dialysis tubing (sealed at both ends by knots), can also be used as the RM container for preparative scale reactions. The tubes can then be placed in suitable plastic vials, e.g., Falcon tubes, holding the FM.

8. CECF reactions are incubated overnight at temperatures ranging between 20 and 30°C (see Note 8). Depending on the reaction container setup, continuous agitation by shaking or rolling is necessary in order to ensure efficient substance exchange between RM and FM through the membrane. Shaking water baths or temperature-controlled cabinets with shaking plates (approximately 150–200 rpm) may be used.

3.5. CF Expression System Optimization

Each component of the CF expression reaction is important and each condition and concentration needs to be optimized. However, most the most frequent reasons for poor expression are (I) inappropriate template design and quality (see Subheading 3.3), (II) poor expression system efficiency (see Subheading 3.5.1), and (III) operation at suboptimal Mg^{2+} concentrations (see Subheading 3.5.2). Implementation of appropriate monitors of protein expression (such as GFP) at several steps of expression protocol development are therefore recommended.

3.5.1. GFP as Monitor for CF Expression Systems

GFP and its modified derivative sGFP are excellent for the rapid monitoring and quantification of CF reactions and they express at high levels. They should always be included as controls for both establishing and optimizing CF expression systems and for testing the quality of freshly prepared stock solutions and batches of extracts. Applications of GFP in CF systems are as follows:

1. GFP as quality monitor for batch and CECF reactions. Typical production kinetics of sGFP in both reaction configurations are given in Fig. 4. Production in the batch configuration stops after linear production over 2 h (see Fig. 4a), while in the CECF configuration the sGFP production continues for more than 20 h (see Fig. 4b). Yields between 400 and 600 μg sGFP per mL RM in the batch configuration and between 4 and 8 mg/ml in the CECF configuration can be obtained routinely.
2. Coexpression of GFP. Some targets might have a negative impact on the efficiency of the CF expression system, e.g., by inactivating ribosomes. If targets remain difficult to express even after optimization, coexpression with GFP could rule out potentially inhibitory side effects. In such experiments, both templates should be added in final concentrations of 20 ng/ μL each.
3. Translational fusions with GFP. GFP does not fold when it is produced as a precipitate and this approach would thus allow the direct quantification of soluble protein expression. If GFP is attached to the C-terminal end of a target protein, it may even be suitable for monitoring the folding of the protein target (22).

GFP production can be quantified by measurement of fluorescence.

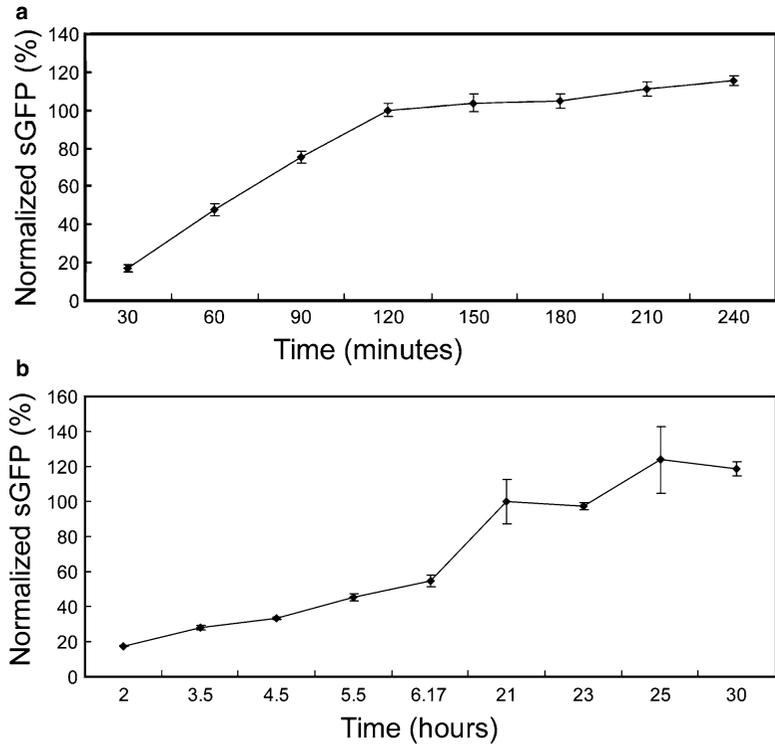


Fig. 4. Kinetics of CF sGFP production. (a) Production in the CF batch configuration. Reactions were performed in 96-well microplates with 25 μL reaction volumes, 20 ng/ μL pET22-sGFP plasmid template concentration, and 26 mM Mg^{2+} concentration. Reactions were incubated at 30°C and performed in triplicates (b) Production of sGFP in the CECF configuration. Reaction was performed in Mini-CECF-Reactors in 55 μL volume, 20 ng/ μL pET22-sGFP plasmid template and 16 mM Mg^{2+} .

1. After expression, keep GFP/sGFP samples at 4°C for 12 h to allow complete GFP folding.
2. Add 3 μL of sample into 297 μL GFP assay buffer in a 96-well dark microplate.
3. Incubate the plate with shaking for 5 min at 22°C.
4. Fluorescence measurements for sGFP are performed at an excitation wavelength of 484 nm and an emission wavelength of 510 nm. For wild-type GFP, the excitation wavelength is 395 nm with an emission wavelength of 510 nm.
5. Use a calibration curve with purified GFP/sGFP to quantify the fluorescence.

3.5.2. Optimization of Ion Concentrations

The Mg^{2+} concentration in the CF reaction is highly critical; its optimum level is normally between 12 and 18 mM for the described CECF configuration, and between 20 and 28 mM for the described batch configuration (see Fig. 5). The optima need to

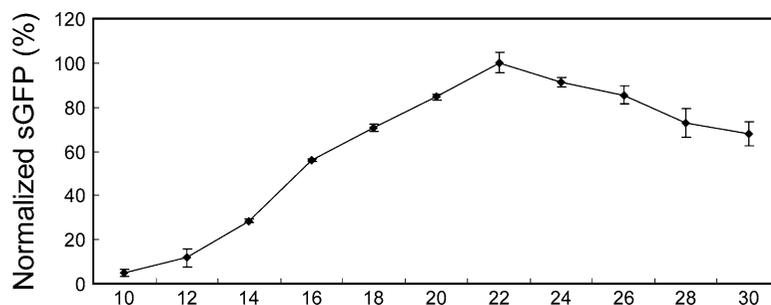


Fig. 5. Optimum of Mg²⁺ ion concentration for sGFP expression in the CF batch configuration.

be determined for each individual target, and well-defined peaks, as well as plateaus, are possible over a certain concentration range. Other ions, such as K⁺ or NH₄⁺, are less critical and optima extend over a much wider range of concentrations. A typical Mg²⁺ optimization screen from 16 mM up to 22 mM in the CECF configuration is exemplified.

1. Create the RFM master mix (see Table 3) with the lowest Mg²⁺ concentration of the screen (e.g., if the screen ranges from 16 to 22 mM, the Mg²⁺ concentration in the RFM master mix would be 16 mM).
2. Aliquot the RFM master mix and reconstitute the RM and FM of each reaction in separate vials.
3. Supplement each RM and FM with the desired additional amount of Mg²⁺ and complete the volumes with the appropriate volume of water.
4. Transfer FM and corresponding RMs to the reaction containers and incubate overnight at 30°C with gentle shaking.
5. Analyze the results, e.g., by GFP measurement or SDS-PAGE analysis.
6. Optional: For screening in the batch configuration, the master mix should also be prepared with the lowest Mg²⁺ concentration of the screen. Then aliquot the master mix (e.g., in 96-well microplate wells) and supplement each reaction with the additional amount of Mg²⁺. Finally, fill up the reactions with the appropriate volumes of MilliQ water.

3.6. CF Expression Modes for Membrane Protein Production

3.6.1. P-CF (Precipitate Generating) Mode

Depending on the nature of the hydrophobic supplements, there are three different CF expression modes for the production of MPs.

CF expression of MPs without supplemented hydrophobic compounds, such as detergents or lipids, results in their initial precipitation. Those P-CF generated MP precipitates frequently retain folded

structures which can then be solubilized with detergents, thus avoiding the extensive unfolding and refolding procedures used to solubilize inclusion bodies. Harvesting the precipitates by centrifugation may even produce relatively pure samples without further treatment. The P-CF mode is always recommended as an initial screening mode for protocol development and product yield optimization. After expression, a detergent resolubilization screen with the P-CF precipitate (see Table 1) is performed.

1. Perform analytical-scale CECF reactions in 50 μL RM (see Subheading 3.4.2).
2. After overnight incubation, the reaction should be more or less turbid and a white precipitate should be visible. The precipitate is harvested by centrifugation of the RM at $18,000\times g$ for 10 min.
3. Discard the supernatant and resuspend the pellet in S30-C buffer equal to the volume of the RM. Remove aliquots for expression yield analysis using SDS-PAGE.
4. Centrifuge the suspension at $18,000\times g$ for 10 min, discard the supernatant and resuspend in 200 μL of S30-C buffer supplemented with 100 mM NaCl and 5 mM DTT. Repeat this step at least twice (see Note 15).
5. Before the last centrifugation step, divide the suspension in appropriate aliquots for the resolubilization screen.
6. Centrifuge at $18,000\times g$ for 10 min, discard the supernatant and resuspend the pellets in S30-C buffer supplemented with the selected detergent and incubate the mixture at a temperature range between 25 and 45°C for 2 h with slight shaking (see Note 16). The volumes of the resuspension buffer must be considered for subsequent quantification. Recommended detergent concentrations are listed in Table 1 (see Note 17). Detergents usually suitable for the resolubilization of P-CF generated MP precipitates are as follows: LMPG = LPPG = LMPC > SDS > Fos12 = Fos14 > DDM > β -OG.
7. Centrifuge at $18,000\times g$ for 10 min and analyze the supernatant as well as residual pellets by SDS-PAGE.
8. Analyze the solubilization efficiencies of the selected detergents.

3.6.2. D-CF (Detergent-Based) Mode

Soluble expression of MPs can be achieved by the addition of detergents to the CF reaction. The quality of solubilized MPs can often depend upon the supplemented detergents, so a set of D-CF suitable detergents should be included in initial D-CF detergent screens (see Table 1) (see Note 18).

1. Set up analytical-scale CECF reactions with 50 μL RM volumes (see Subheading 3.4.2).

2. Detergents should be added to both reaction compartments. For initial screens the working concentration listed in Table 1 should be used (see Note 19). Avoid foam formation when pipetting the detergent solutions. The corresponding detergent volumes have to be subtracted from the final volume of water in the pipetting scheme.
3. Incubate the reactions overnight at 30°C with gentle shaking.
4. Collect the RMs and centrifuge for 10 min at 18,000 × *g*.
5. Transfer supernatants into fresh tubes. If precipitates are present, resuspend in equal volumes of water.
6. Analyze 1–3 μL of each sample by SDS-PAGE.
7. Identify those detergents giving sufficient MP solubilization. If complete MP solubilization has been achieved (i.e., if there is no residual pellet in step 6), gradually decreasing the final detergent concentration in a subsequent screen could help to achieve the best compromise between solubilization and expression yield.

3.6.3. L-CF (Lipid-Based) Mode

The addition of lipids as defined artificial liposomes, bicelles, or nanodiscs into the RM can provide alternative hydrophobic environments for the production of soluble MPs. However, the co-translational insertion of MPs into lipid bilayers may require additional compounds (such as chaperones) to support translocation and membrane targeting. The L-CF expression mode is an emerging technique and efficient membrane insertion could require intensive protocol optimizations for particular targets. Lipids tolerated at high concentrations by CF systems include DMPC, DOPC, POPC, and *E. coli* polar lipids.

1. Dissolve the selected lipid powder in chloroform at a final concentration of 10 mg/mL. The lipids should then be dried thoroughly using a rotary evaporator. Pressure and speed have to be adjusted in order to obtain uniform lipid films.
2. As chloroform can inhibit the CF reaction, it must be completely removed. This can be achieved by drying the lipid film in a desiccator overnight.
3. Resuspend the lipid films in S30-C buffer to final concentrations between 20 and 50 mg/mL. Several cycles of vortexing and sonication should be performed in order to obtain homogenous suspensions.
4. Pass the lipid solution through a 0.2 μm filter using an extruder. A minimum of 11 passages is recommended to obtain homogenous suspensions of unilamellar liposomes with size ranges between 200 and 300 nm. Extruded liposomes can be stored for few days at 4°C. For longer storage, the liposomes can be stored at –80°C and should be extruded again before usage.

5. Prepare analytical scale CECF reactions with 50 μL RM volumes (see Subheading 3.4.2).
6. Add the liposomes directly into the RM compartment of the CECF reaction. For an initial screening, final liposome concentrations of 2–4 mg/ml may be used (see Note 20).
7. Incubate the reactions overnight at 30°C with gentle shaking (see Note 8).
8. Centrifuge the RMs for 10 min at 18,000 $\times g$ (see Note 21).
9. Resuspend the pellet containing the potential proteoliposomes in appropriate volumes of buffer and directly use them for further analysis (see Note 22). Aliquots may be stored at 4°C.

3.7. High-Throughput Expression Condition Screening

The open nature and the high success rate of CF systems allow high-throughput screening of expression conditions as a new rational strategy for protein expression. The quality and yield of proteins can be optimized by extensive modification of their expression environment (see Fig. 6). In linear screens, the concentrations of one or several compounds are evaluated in one dimension (see Fig. 6). However, the concentration optima of different compounds can correlate with each other, e.g., T7RNAP and template DNA or PEG and Mg^{2+} (see Fig. 6). Such interfering compounds can be optimized in correlated screens using both dimensions of the microplate (see Fig. 6). The use of robotic liquid handling systems and calculation programs will largely facilitate such high-throughput applications. In any case, multi-pipetting should be used whenever possible in order to minimize reaction set up times and evaporation problems.

3.7.1. Linear/Correlated Compound Screening

1. Prepare stock solutions, premix and master mix for the batch configuration (see Subheading 3.4.1). Leave out the volume of water, potential screening compound(s), and template DNA.
2. Transfer suitable aliquots of the master mix to the reaction container, e.g., 96-well microplates. Reactions should be performed in triplicate in order to obtain representative results. All solutions and the microplate should be kept on ice during pipetting.
3. Add the calculated volumes of the desired screening compound(s). For new compounds, an initial evaluation of GFP expression with the tolerated concentrations is recommended.
4. Add the appropriate volume of water to achieve the final reaction volume.
5. Start the reactions by addition of template DNA.
6. Seal the microplate with Parafilm to prevent evaporation.

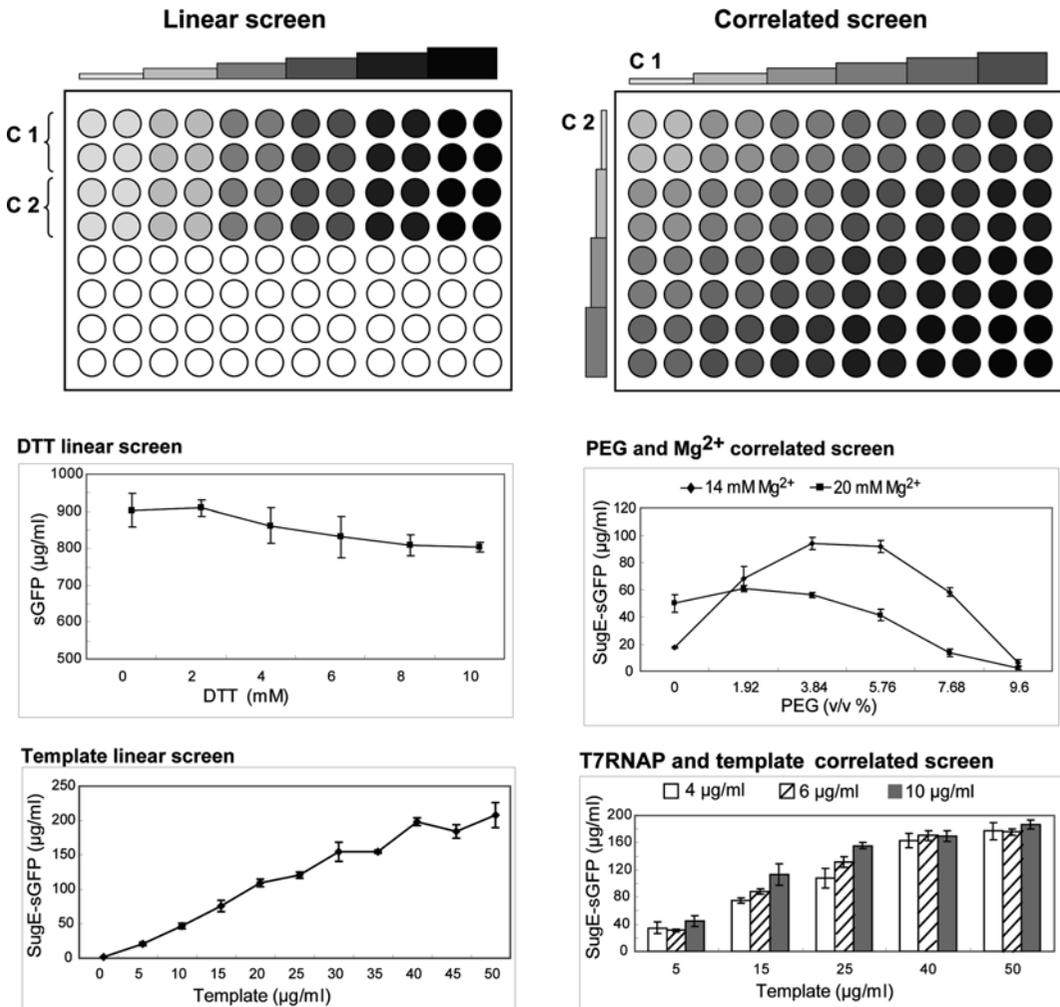


Fig. 6. CF expression protocol optimization by throughput linear and correlated compound screens. The expression of sGFP and the small multidrug transporter SugE-GFP have been analyzed. The screening compounds are indicated.

7. Incubate the microplate with shaking at 32°C for 2–4 h.
8. Analyze the reactions using suitable techniques.

3.7.2. Scaling Up Optimized Batch Reactions

Scaling up CF reactions to volumes of several mL is possible in the batch, as well as in the CECF configuration, without significant loss of efficiency (see Fig. 7). However, results obtained in the batch configuration are not directly comparable to the more productive CECF configuration. The general impact of compounds on an expressed target will be similar, but effective concentrations should be re-screened as the protocols and productivities of the two configurations are different.

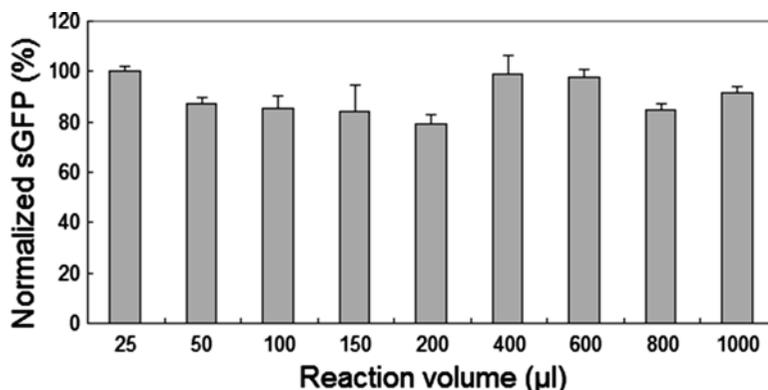


Fig. 7. Scale-up of CF batch reactions into preparative scales. Reaction conditions for sGFP expression optimized in 25 μ L volumes were stepwise scaled up to 1 mL reaction volumes. For reaction volumes ≤ 200 μ L, 96-well microplates were used. For reaction volumes ≥ 400 μ L, 24-well microplates were used.

4. Notes

1. For the 4 mM amino acid stock, first, individual stocks are prepared. Prepare a 20 mM stock solution of tyrosine and 100 mM stock solutions for the other 19 amino acids. L-tryptophan must be dissolved in 100 mM HEPES, pH 8.0. L-aspartic acid, L-cysteine, L-glutamic acid, and L-methionine must be dissolved in 100 mM HEPES, pH 7.4. All other amino acids are dissolved in MilliQ water. Sonication or heating to 60°C could improve solubility. However, some stocks (in particular W, D, E, N, C, Y) won't dissolve completely and have to be handled as suspensions. For the 4 mM amino acid mixture, take 2 mL of each the 100 mM amino acid stocks and 10 mL of the 20 mM L-tyrosine stock and make up to 50 mL with MilliQ water. For the 8 mM amino acid stock, weigh in all compounds and dissolve the powders in MilliQ water. The stock remains turbid.
2. Cells will continue to grow during cooling and must not enter the stationary phase of growth. The cells should preferably be chilled in the fermenter or quickly upon harvesting using an appropriate cooling device.
3. This step causes significant precipitation of proteins in the supernatant which is beneficial for the quality of the extract. The high salt concentration causes the ribosomes to dissociate from mRNA templates which will subsequently become degraded.
4. T7RNAP is still quite impure in the elution fractions. The protein increasingly becomes unstable at higher concentrations.

5. The apparent molecular mass of MPs by SDS-PAGE analysis often differs from their expected masses.
6. Premix composition might be variable, e.g., amino acids could be included if their final concentrations are not the subject of evaluation. The proposed premix can be re-frozen several times.
7. Calculation templates may be generated using standard programs such as Excel.
8. Temperature screens may be performed for the optimization of product quality. Lower incubation temperatures will reduce expression yields. However, this may affect folding, quality, or membrane insertion of synthesized proteins.
9. The volume ratio of the RM and FM influences the final yield of the product and should be between 1:10 and 1:50. The increase in expression efficiency is not linear with the FM volume, the proposed RM:FM ratio is therefore a good economical compromise.
10. The volume of screening compounds needs initially to be subtracted from the water volume of the RM and FM mixtures.
11. For the assembly of the Mini-CECF-Reactor, the Teflon ring is placed on a sheet of Parafilm, then a suitable piece of dialysis membrane (2×2 cm) is placed on top of the Teflon ring. Finally the container is pushed through the Teflon ring which tightly fixes the dialysis membrane between ring and container.
12. The Mini-CECF-Reactor is filled from the top by carefully touching the membrane with the pipette tip and releasing the RM. To harvest the RM after incubation, the membrane is perforated from the bottom with a pipette tip and the RM removed. After the reaction, the Mini-CECF-Reactor is disassembled and the membrane is discarded. The container and the Teflon ring are cleaned by extensive washing with deionised water and finally with MilliQ water. The container and Teflon ring should be thoroughly dried prior to next use. Microplates with the Mini-CECF-Reactors should be sealed with Parafilm to prevent evaporation during incubation.
13. Slide-A-Lyzers are filled with a syringe at one of the preformed openings. This opening should be placed upward if Maxi-CECF-Reactors are used. It must be sealed if other FM containers are used. Care must be taken not to damage the membrane upon filling the RM. Slide-A-Lyzers can be reused a few times after extensive washing with water and storage in water with 0.1% NaN_3 . We recommend reuse only for the same protein target.
14. D-Tube dialyzer may be reused a few times after extensive washing with water and storage in water with 0.1% NaN_3 . The water must be removed completely before filling with the RM.

15. Proteins from the S30 extract may co-precipitate with the MP. Those impurities can be reduced by washing the pellet with mild detergents and/or increased NaCl and DTT concentrations.
16. The optimal temperature needs to be determined and depends on the individual MP. Extended incubation times are not usually effective.
17. High detergent concentrations can be used for the resolubilization screens. Concentrations, as well as type of detergent, may be modified during subsequent affinity chromatography purification steps.
18. Besides uniform detergent micelles, environments containing hybrid micelles or even mixed micelles composed of detergents and lipids may be used for MP solubilization.
19. Most D-CF suitable detergents will still become inhibitory to the CF system above a certain concentration level. The indicated concentrations approach those maxima and should initially be used in order to obtain complete solubilization of the expressed MPs. Some detergents such as digitonin, DDM, or Triton X-100 might already have some inhibitory effects at the suggested working concentration.
20. Lipids may be mixed with detergents such as 0.5–1% Chaps in the RM. Those mixed micelles might promote MP insertion and the detergent will continuously be dialyzed out during the reaction into the FM, resulting in the subsequent formation of proteoliposomes (25).
21. The supplied liposomes may fuse during the incubation and thus precipitate. The synthesized MPs may insert into the provided liposomes but a substantial part will also precipitate.
22. Pellets from L-CF reactions usually contain mixtures of empty liposomes, proteoliposomes and associated MP precipitates. The proteoliposomes must therefore be purified (e.g., by urea washing or density gradient centrifugation) prior to further analysis.

References

1. Cell-free protein production, methods and protocols (2010) Y. Endo, K. Takai, T. Ueda, (eds.). Methods Mol. Biol. 607, Humana Press, New York.
2. Schwarz D, Junge F, Durst F, et al (2007) Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. Nat Protoc 2: 2945–2957
3. Jung, F, Schneider B, Reckel S, Schwarz D, Dötsch V, Bernhard F (2008) Large-scale production of functional membrane proteins. Cell Mol. Life Sci 65: 1729–1755
4. Katzen F, Kudlicki W (2006) Efficient generation of insect-based cell-free translation extracts active in glycosylation and signal sequence processing. J Biotechnol 125: 194–197
5. Zubay G (1973) In vitro synthesis of protein in microbial systems. Annu Rev Genet 7: 267–287
6. Mureev S, Kovtun O, Nguyen UT, Alexandrov K (2009) Species-independent translational leaders facilitate cell-free expression. Nat Biotechnol 27: 747–752
7. Vinarov DA, Loushin Newman CL, Markley J L (2006) Wheat germ cell-free platform for

- eukaryotic protein production. *FEBS J* 273: 4160–4169
8. Swartz J (2006) Developing cell-free biology for industrial applications. *J Ind Microbiol Biotechnol* 33: 476–485
 9. Kim TW, Kim DM, Choi CY (2006) Rapid production of milligram quantities of proteins in a batch cell-free protein synthesis system. *J Biotechnol* 124: 373–380
 10. Spirin AS, Baranov, VI, Ryabova LA, Ovodov SY, Alakhov YB (1988) A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* 242: 1162–1164
 11. Kigawa T, Yokoyama S (1991) A continuous cell-free protein synthesis system for coupled transcription-translation. *J Biochem* 110: 166–168
 12. Sitaraman K, Chatterjee DK (2009) High-throughput protein expression using cell-free system. *Methods Mol. Biol* 498: 229–244
 13. Aoki M, Matsuda T, Tomo Y, Miyata Y, Inoue M, Kigawa T, Yokoyama S (2009) Automated system for high-throughput protein production using the dialysis cell-free method. *Protein Expr. Purif* 68: 128–136
 14. Junge F, Luh LM, Proverbio D, et al (2010) Modulation of G-protein coupled receptor sample quality by modified cell-free expression protocols: a case study of the human endothelin A receptor. *J Struct Biol* 172: 94–106
 15. Kai L, Kaldenhoff R, Lian J, Zhu X, Dötsch V, Bernhard F, Cen P, Xu Z (2010) Preparative scale production of functional mouse aquaporin 4 using different cell-free expression modes. *PLoS ONE* doi:10.1371/journal.pone.0012972
 16. Hovijitra NT, Wu JJ, Peaker B, Swartz JR (2009) Cell-free synthesis of functional aquaporin Z in synthetic liposomes. *Biotechnol Bioeng* 104: 40–49
 17. Berrier C, Guilvout I, Bayan N, Park KH, Mesneau A, Chami M, Pugsley AP, Ghazi A (2010) Coupled cell-free synthesis and lipid vesicle insertion of a functional oligomeric channel MscL. MscL does not need the insertase YidC for insertion in vitro. *Biochim Biophys Acta* 1808: 41–46
 18. Schwarz D, Daley D, Beckhaus T, Dötsch V, Bernhard F (2010) Cell-free expression profiling of *E. coli* inner membrane proteins. *Proteomics* 10:P 1762–1779
 19. Reckel S, Sobhanifar S, Schneider B, et al (2008) Transmembrane segment enhanced labeling as a tool for the backbone assignment of α -helical membrane proteins. *PNAS* 105: 8262–8267
 20. Apponyi MA, Ozawa K, Dixon NE, Otting G (2008) Cell-free protein synthesis for analysis by NMR spectroscopy. *Methods Mol Biol* 426: 257–268
 21. Li Y, Wang E, Wang Y (1999) A modified procedure for fast purification of T7 RNA polymerase. *Protein Expr Purif* 16: 355–358
 22. Drew DE, von Heijne G, Nordlund P, de Gier JW (2001) Green fluorescent protein as an indicator to monitor membrane protein over-expression in *Escherichia coli*. *FEBS Lett* 507: 220–224
 23. Yabuki T, Motoda Y, Hanada K, et al (2007) A robust two-step PCR method of template DNA production for high-throughput cell-free protein synthesis. *J Struct Funct Genomics* 8: 173–191
 24. Schneider B, Junge F, Shirokov VA, Durst F, Schwarz D, Dötsch V, Bernhard F (2010) Membrane protein expression in cell-free systems. *Methods Mol Biol* 601: 165–186
 25. Shimono K, Goto M, Kikukawa T, Miyauchi S, Shirouzu M, Kamo N, Yokoyama S (2009) Production of functional bacteriorhodopsin by an *Escherichia coli* cell-free protein synthesis system supplemented with steroid detergent and lipid. *Protein Sci* 18: 2160–2171

Chapter 15

Protein Identification by MALDI-TOF Mass Spectrometry

Judith Webster and David Oxley

Abstract

MALDI-TOF mass spectrometers are now commonplace and their relative ease of use means that most non-specialist labs can readily access the technology for the rapid and sensitive analysis of biomolecules. One of the main uses of MALDI-TOF-MS is in the identification of proteins, by peptide mass fingerprinting (PMF). Here we describe a simple protocol that can be performed in a standard biochemistry laboratory, whereby proteins separated by 1D or 2D gel electrophoresis can be identified at femtomole levels. The procedure involves excision of the spot or band from the gel, washing and destaining, reduction and alkylation, in-gel trypsin digestion, MALDI-TOF-MS of the tryptic peptides and database searching of the PMF data. Up to 96 protein samples can easily be manually processed at one time by this method.

Key words: Proteomics, MALDI-TOF, Mass spectrometry, SDS-PAGE, 2D-gel, In-gel digestion, Peptide mass fingerprint, Protein identification, Database searching

1. Introduction

Developments in mass spectrometry technology, together with the availability of extensive DNA and protein sequence databases and software tools for data mining, has made possible rapid and sensitive mass spectrometry-based procedures for protein identification. Two basic types of mass spectrometers are commonly used for this purpose; MALDI-TOF-MS and ESI-MS. MALDI-TOF instruments are now quite common in biochemistry laboratories and are very simple to use, requiring no special training. ESI instruments, usually coupled to capillary/nanoLC systems, are more complex and require expert operators. We will therefore focus on the use of MALDI-TOF-MS, although the sample preparation is identical for both methods. The principle behind the use of MALDI-TOF-MS for protein identification is that the digestion

of a protein with a specific protease will generate a mixture of peptides unique to that protein. Measuring the molecular masses of these peptides then gives a characteristic dataset called a peptide mass fingerprint (PMF) (1). The PMF data can then be compared with theoretical peptide molecular masses that would be generated by using the same protease to digest each protein in the sequence database, to find the best match. Provided the protein being analyzed is present in the database being searched and the data are of sufficient quality, the best match should be the correct protein. In order to judge the validity of protein identification by this method, some means of scoring the quality of the match must be used.

The procedure described here involves cutting protein bands or spots from 1D or 2D PAGE gels, destaining the gel pieces, reducing and alkylating the protein, digesting with trypsin, using MALDI-TOF mass spectrometry to determine the mass of the tryptic peptides and database searching with the PMF data to identify the protein.

The sample processing steps can be performed in microfuge tubes or in 96-well plates. One person can easily process a 96-well plate in a day, but if higher throughput is required, each step can be automated, allowing the possibility of several hundred protein identifications per day. Spot cutting, sample processing, and sample plate loading robots are commercially available and routinely used in many high-throughput laboratories.

2. Materials

1. 0.5-mL microfuge tubes (or 96-well V-bottom polypropylene microtitre plates) (see Note 1).
2. One Touch Spot Picker (1.5 mm) (The Gel Company) – optional.
3. Silver destaining solution (for silver-stained gels only): dissolve potassium ferricyanide (2 mg/mL) in sodium thiosulphate solution (0.2 mg/mL). Make fresh immediately before use.
4. Sonicator bath (or vortex mixer with attachments for unattended use with 0.5-mL microfuge tubes or 96-well microtitre plates).
5. Aqueous buffer: 50 mM NH_4HCO_3 – make fresh weekly.
6. Organic buffer: 50 mM NH_4HCO_3 /acetonitrile 1:1 – make fresh weekly.
7. DTT solution: (DL-Dithiothreitol ultra pure) 10 mM DTT in aqueous buffer – make fresh immediately before use.
8. Iodoacetamide solution: (ultrapure) 50 mM iodoacetamide in aqueous buffer – make fresh immediately before use. Note iodoacetamide is toxic!

9. Trypsin solution (Modified Sequencing Grade, Promega) (see Note 2).
10. Centrifugal vacuum concentrator.
11. 10% aqueous TFA.
12. 0.1% aqueous TFA.
13. Matrix solution (see Note 3).
14. Peptide calibration mixture (see Note 3).
15. MALDI sample plate.
16. MALDI-TOF mass spectrometer.

3. Methods

Protein identification by PMF is usually performed on protein bands or spots cut from 1D or 2D PAGE gels and so this is the method described here; however, it can also be applied to proteins in solution with minor modifications. One of the biggest problems that may be encountered is contamination, particularly when low amounts of protein are analyzed. Note 1 describes some precautions to minimize the effects of contamination.

All washing steps during the processing of the gel pieces can be performed either on a vortex mixer or in a sonicator bath.

3.1. Removing Gel Spot or Band from the Gel and Destaining

Ideally, the gel should have been stained with a Coomassie stain (preferably a colloidal Coomassie if detection sensitivity is an issue, (e.g. (2, 3) or one of the commercially available stains). This method is also compatible with Sypro stains (though a UV transilluminator will be required to visualize the protein bands or spots during excision). Standard silver stains are not compatible, but certain modified silver stains (e.g. (4), or one of the commercially available “mass spectrometry compatible” silver stains) can be used; however, even these usually give inferior mass spectrometry data compared to Coomassie or Sypro stained gels.

3.1.1. Removal of Gel Piece

1. Place the stained gel in a disposable petri dish on a light box and cut out the spot/band with a manual spot picker or a clean scalpel/razor blade, without taking any excess gel.
2. If necessary, cut the gel spot/band into 1 mm pieces and transfer into a 0.5-mL microfuge tube or one well of a 96-well microtitre plate.

3.1.2. Destaining Coomassie or Sypro-Stained Gel Pieces

1. Wash with aqueous buffer (100 μ L) 1 \times 5 min.
2. Wash with organic buffer (100 μ L) 2 \times 5 min.

*3.1.3. Destaining
Silver-Stained Gel
Pieces (5)*

1. Wash with silver destaining solution (100 μL) until destained (approximately 5–30 min) – the gel will retain a pale yellow colour.
2. Wash with high purity water (100 μL) 2×5 min.
3. Wash with aqueous buffer (100 μL) 1×5 min.
4. Wash with organic buffer (100 μL) 1×5 min.

3.2. In-Gel Reduction and Alkylation of the Protein

Reduction of disulphide bonds followed by alkylation of the free cysteines to prevent reoxidation, while not essential for the digestion of most proteins, generally gives better results. This is due to increased susceptibility of the reduced/alkylated protein to tryptic digestion and the absence of any disulphide-linked peptides (which are not matched in the database search) from the PMF data (6).

1. Incubate destained gel pieces in DTT solution (100 μL) for 1 h at 50°C in an oven. If using a 96-well plate, use sealing film or a sealing lid, to exclude oxygen and prevent drying out.
2. Cool to room temperature, remove and discard DTT solution. Add iodoacetamide solution (100 μL) and incubate for 1 h with occasional mixing (vortex) in the dark at room temperature.
3. Discard supernatant and wash gel pieces with aqueous buffer (100 μL) for 5 min, then with organic buffer (100 μL) for 2×5 min.
4. Dry the gel pieces completely in a centrifugal vacuum concentrator. Caution should be exercised in handling the dried gel pieces, as they are easily lost from tubes or plates.

3.3. In-Gel Digestion of the Protein with Trypsin

1. Take an aliquot of trypsin (10 μL of $10\times$ stock solution) from the freezer, add aqueous buffer (90 μL) and mix.
2. Rehydrate gel pieces for 10 min in trypsin solution (approximately 1 μL per mm^3 gel). There should be little or no excess liquid after rehydration.
3. Add an equal volume of aqueous buffer and incubate at 37°C overnight (or for at least 3 h). Use an incubator, not a heating block. If using a 96-well plate, use sealing film or a sealing lid, to prevent drying out.
4. Add 1/10th volume of 10% TFA and sonicate/vortex for 5 min.
5. The resulting supernatant is used directly for MALDI-TOF-MS.
6. Residual peptides can be washed from the gel piece if required with a small amount of 0.1% TFA.

3.4. Mass Determination of Peptides by MALDI-TOF-MS

The quality of the MALDI-TOF spectrum that will be obtained from the sample depends crucially on the sample/matrix preparation. The basic requirements are for a uniform microcrystalline layer of matrix/sample and the removal of salts and other contaminants. There are numerous published methods, but the one described

here is robust and quite simple, requiring no pre-cleanup steps. For the best results, a high quality matrix should be used, as supplied by MALDI-MS manufacturers, or in a commercial PMF kit. Alternatively, analytical grade matrix can be recrystallized (7).

MALDI-TOF spectra must be calibrated in order to achieve sufficient accuracy for database searching. This is done by acquiring spectra on peptide standards to generate a calibration curve, which is applied to the experimental data. The calibration peptides can be analyzed separately from the experimental sample (external calibration) or they can be mixed with the experimental sample (internal calibration). Internal calibration is more accurate (typically 10–30 ppm for a strong spectrum) than external (typically 100 ppm or more), but is more difficult, as the amount of standard peptides used needs to be matched to the level of the experimental sample peptides. The addition of too much peptide standards can suppress the signal of the sample peptides and vice versa. A commonly used variation of the internal calibration method utilizes the autodigestion fragments of trypsin for calibration (842.5094 and 2211.1040 for porcine trypsin), instead of adding additional peptides (7).

3.4.1. Applying the Sample and the Calibration Mixture to the MALDI Plate

1. Apply matrix solution (0.5 μL) to a clean MALDI plate and allow to dry.
2. Apply protein digest (0.2–2 μL) to the dried matrix spot and allow to dry. If a large amount of gel was used for the digestion, the supernatant volume could be much larger than 2 μL (up to 20 μL). It is not usually necessary to use all of the supernatant unless the protein band was very weak. In this case the entire supernatant can be concentrated to 1–2 μL and used (see Note 4).
3. Apply 0.1% TFA (5–10 μL) to each sample spot, leave for 30 s, remove and discard, then repeat; this step desalts the sample.
4. If using external calibration, apply 0.2 μL of peptide calibration mixture as close as possible (but not touching) to each sample spot and allow to dry.

3.4.2. Measuring the Peptides Masses in the MALDI-TOF-MS Instrument

The precise operation of the MS is instrument-dependent, but the basics are very similar. Increasing the laser power increases the signal, but also decreases the resolution and “burns” off the sample faster. Therefore, the laser power should be set to the lowest level that gives a good signal. Set the mass range (e.g. m/z 600–3,500), the number of shots to be acquired (100–200 is reasonable) and set the laser power low. Gradually increase the power until an even distribution of noise appears across the whole mass range. Within a few shots, peaks should begin to appear above the noise. Continue to adjust the laser power until the signal level is satisfactory.

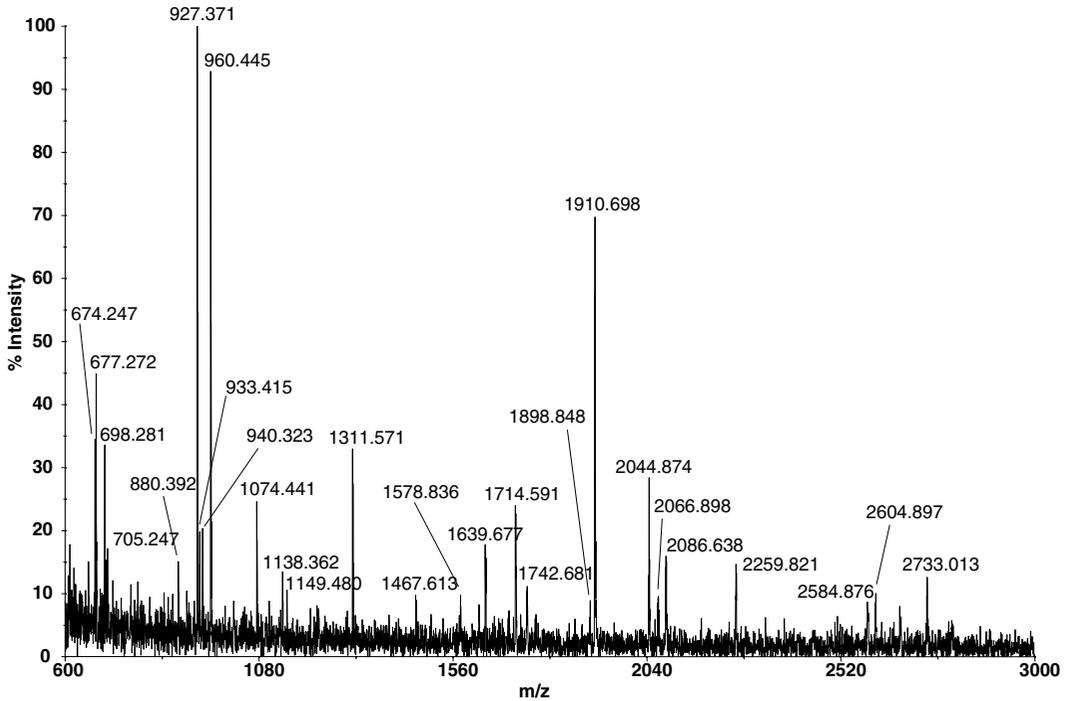


Fig. 1. MALDI-TOF spectrum of a trypsin-digested 1D gel band. Peaks are labelled with their monoisotopic masses. Note that these are not the masses of the peptides, but of the peptide (pseudo)molecular ions. In MALDI spectra, peptide molecular ions arise predominantly through the addition of a proton to the peptide, giving a mass increase of 1.007 Da. The molecular ions are usually denoted as MH^+ or $[M+H]^+$.

The heterogeneous nature of the matrix surface means that some areas of the sample spot will give better peptide signals than others. The difference can be dramatic, so it is a good idea to periodically move the laser position, while assessing the signal intensity, to find the hot-spots.

If external calibration of the spectrum is to be used, a spectrum of the calibration standards should be acquired immediately, using the same power setting. Fig. 1 shows a typical MALDI spectrum obtained from the trypsin digestion of a weak colloidal Coomassie-stained 1D gel band.

3.5. Generating PMF Data and Searching Protein Databases

The PMF data must now be extracted from the MALDI spectrum using the appropriate software associated with the MS instrument used. The baseline threshold should first be adjusted to ensure that all of the peptide signals are detected, without including any noise. If the spectrum has been obtained on a high resolution (reflectron) instrument, the peptide signals will appear as multiple peaks separated by 1 Da, due to the presence of ^{13}C isotopes in the peptides (see Fig. 2). “De-isotoping” and “centroiding” of the MS data is necessary, so that only the monoisotopic peptide masses are included in the PMF data.

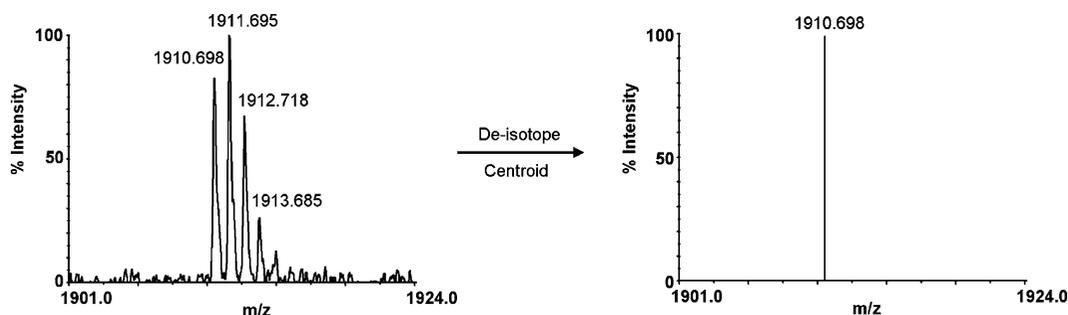


Fig. 2. Expanded view of the peptide signal at around m/z 1,910 from the MALDI-TOF spectrum in Fig. 1. De-isotoping removes the ^{13}C isotope peaks and centroiding reduces the remaining peak to a single data point.

The MS data analysis software will have an option for the data to be shown as a list of masses (peak list), which can be saved as a text file or simply copied directly into the data entry field of the database search engine. There are a number of search engines that can be used for PMF searches, some of which are freely available on the internet (see Note 5). We will use Mascot to demonstrate a search, but all have essentially the same functions. The Mascot PMF search page is shown in Fig. 3. Detailed explanations of the various terms and parameters are available from the Web site; brief descriptions are given in Note 6.

1. Select the database to be searched from the list, e.g. Swissprot.
2. Specify taxonomy information as closely as possible for the sample origin.
3. Select “trypsin” as the enzyme.
4. “Allow up to” 1 missed cleavage.
5. From the list of modifications on the right, select “Carbamidomethyl (C)” as a fixed modification.
6. Similarly, select “Oxidation (M)” as a variable modification.
7. If the MALDI spectrum was internally calibrated, the peptide tolerance can be set to 50 ppm, otherwise 100–300 ppm is a reasonable starting point.
8. Select “MH⁺” for Mass values.
9. Select “Monoisotopic” if the MALDI data were acquired in reflectron mode, or “Average” for data acquired in linear mode.
10. If the peak list was saved as a text file, browse to the file location, otherwise paste the list directly into the “Query” field.
11. Select “Auto” from “Report top” hits.
12. Start search.

Within a few seconds the search result will appear which should look similar to Fig. 4. This shows a graphical representation of the

{MATRIX}
{SCIENCE}

[HOME](#) : [WHAT'S NEW](#) : [MASCOT](#) : [HELP](#) : [PRODUCTS](#) : [SUPPORT](#) : [TRAINING](#) : [CONTACT](#)

Mascot > Peptide Mass Fingerprint

MASCOT Peptide Mass Fingerprint

| | | | |
|--|--|---|--|
| Your name | <input style="width: 95%;" type="text"/> | Email | <input style="width: 95%;" type="text"/> |
| Search title | <input style="width: 95%;" type="text"/> | | |
| Database(s) | SwissProt | Enzyme | Trypsin <input type="button" value="v"/> |
| | NCBI nr contaminants cRAP MSDB | Allow up to | 1 <input type="button" value="v"/> missed cleavages |
| Taxonomy | Homo sapiens (human) <input type="button" value="v"/> | | |
| Fixed modifications | Carbamidomethyl (C) | > < | Acetyl (K) Acetyl (N-term) Acetyl (Protein N-term) Amidated (C-term) Amidated (Protein C-term) Ammonia-loss (N-term C) Biotin (K) Biotin (N-term) Carbamyl (K) Carbamyl (N-term) Carboxymethyl (C) |
| | Display all modifications <input type="checkbox"/> | | |
| Variable modifications | Oxidation (M) | > < | |
| | | | |
| Protein mass | <input style="width: 50px;" type="text"/> kDa | Peptide tol. ± | 200 <input type="button" value="v"/> ppm <input type="button" value="v"/> |
| Mass values | <input checked="" type="radio"/> MH ⁺ <input type="radio"/> M _r <input type="radio"/> M-H ⁻ | Monoisotopic | <input checked="" type="radio"/> Average <input type="radio"/> |
| Data file | <input style="width: 95%;" type="text"/> <input type="button" value="Browse..."/> | | |
| Query NB Contents of this field are ignored if a data file is specified. | 674.247 677.272 698.281 705.247 880.392 927.371 | | |
| Decoy | <input type="checkbox"/> | | Report top AUTO <input type="button" value="v"/> hits |
| <input type="button" value="Start Search ..."/> | | <input type="button" value="Reset Form"/> | |

Copyright © 2008 Matrix Science Ltd. All Rights Reserved.

Fig. 3. Mascot PMF search page.

results and a list of significant matches. A single significant match is indicated by the graph with a score of 229, well above the significance threshold of 56. The score is related to the probability that the match is real rather than purely random, and is also expressed as an expect value equivalent to the BLAST E-value, 2.6e-19 in this case, indicating that this is almost certainly a genuine match.

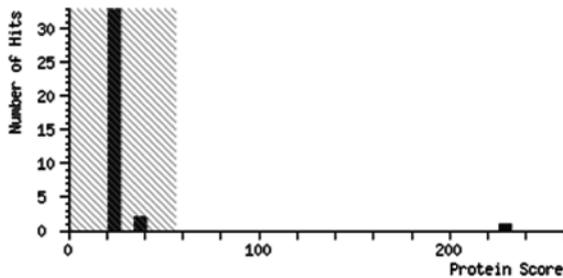
In the “Concise Protein Summary Report” shown, all of the proteins matching the same set or subset of masses are listed under a single entry. Typically, these will represent multiple database entries for the same protein as well as sequence variants, fragments and random matches. In the present example, the top hit contains

{MATRIX} Mascot Search Results *{SCIENCE}*

User :
 Email :
 Search title :
 Database : SwissProt 2010_12 (523151 sequences; 184678199 residues)
 Taxonomy : Homo sapiens (human) (20259 sequences)
 Timestamp : 10 Jan 2011 at 15:44:48 GMT
 Top Score : 229 for ALBU_HUMAN, Serum albumin OS=Homo sapiens GN=ALB PE=1 SV=2

Mascot Score Histogram

Protein score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Protein scores greater than 56 are significant ($p < 0.05$).



Concise Protein Summary Report

| | | |
|--|--|--------------------------|
| Format As | Protein Summary | Help |
| Significance threshold p< 0.05 | | Max. number of hits AUTO |
| <input type="button" value="Re-Search All"/> <input type="button" value="Search Unmatched"/> | | |
| 1. | <u>ALBU_HUMAN</u> Mass: 71317 Score: 229 Expect: 2.6e-19 Matches: 21 Serum albumin OS=Homo sapiens GN=ALB PE=1 SV=2 | |
| | <u>IDH3A_HUMAN</u> Mass: 40022 Score: 33 Expect: 11 Matches: 5 Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial OS=Homo sapiens GN=IDH3A | |
| | <u>BDNF_HUMAN</u> Mass: 28199 Score: 30 Expect: 19 Matches: 4 Brain-derived neurotrophic factor OS=Homo sapiens GN=BDNF PE=1 SV=1 | |
| | <u>TMM78_HUMAN</u> Mass: 15468 Score: 28 Expect: 31 Matches: 3 Transmembrane protein 78 OS=Homo sapiens GN=TMM78 PE=2 SV=1 | |

Fig. 4. Mascot search results summary page using PMF data from the spectrum in Fig. 1.

eleven entries, although only the top four are shown in Fig. 4. Only the first entry has a score above the significance threshold, the other ten have scores well below the threshold and are therefore almost certainly random matches.

Clicking on a protein accession number from the list brings up further information (see Fig. 5), including the matched peptides, listed and mapped onto the protein sequence, as well as the percentage of sequence coverage and any unmatched masses. There will almost inevitably be some unmatched masses; they can

{MATRIX}
{SCIENCE} **Mascot Search Results**

Protein View

Match to: ALBU_HUMAN Score: 229 Expect: 2.6e-19
Serum albumin OS=Homo sapiens GN=ALB PE=1 SV=2

Nominal mass (M_n): 71317; Calculated pI value: 5.92
NCBI BLAST search of ALBU_HUMAN against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Homo sapiens

Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 27
Number of mass values matched: 21
Sequence Coverage: 32%

Matched peptides shown in Bold Red

```

1 MKWVTFISLL FLSSAYSRG VFRDRAHKSE VAHRFKDLGE ENFKALVLIA
51 FAQLLQCCPF EDHVKLVNEV TEFAKTCVAD ESAENCCKSL HTLFGDKLCT
101 VATLRETYGE MADCCAKQEP ERNECFLQHK DDNPMLPRLV RPEVDVMCTA
151 FHDNEETFLK KYLYEIAARR PYFYAPELLF FAKRYKAAFT ECCQAADKAA
201 CLLPKLDELRL DEGKASSAKQ RLKCASLQKF GERAFAKAWAV ARLSQRFFKA
251 EFAEVSKLVT DLTKVHTECC HGDLLLECADD RADLAKYICE NQDSISSKLL
301 ECCEKPLLEK SHCIAEVEND EMPADLPSLA ADFVESKDVC KNYAEAKDVF
351 LGMFLYEPYAR RHPDYSVLL LRLAKTYETT LEKCCAAADP HECYAKVFDE
401 FKPLVEEPQN LIQNCELFE QLGEYKQNA LLVRYTKKVP QVSTFTLVEV
451 SRNLGKVGSK CCKHPEAKRM PCAEDYLSVV LNQLCVLHEK TFPVSDRVTKC
501 CTESLVNRRP CFSALEVDET YVPKEFNAET FTFHADICTL SEKERQIKKKV
551 TALVELVKHK PRATKEQLKA VMDDFAAFVE KCCKADDEK CFAEEGKKLV
601 AASQAALGL
    
```

Show predicted peptides also

Sort Peptides By Residue Number Increasing Mass Decreasing Mass

| Start - End | Observed | Mr(expt) | Mr(calc) | ppm | Miss | Sequence |
|-------------|-----------|-----------|-----------|------|------|----------------------------------|
| 25 - 34 | 1149.4800 | 1148.4727 | 1148.5686 | -84 | 1 | R.DAHKSEVAHR.F |
| 29 - 34 | 698.2810 | 697.2737 | 697.3507 | -110 | 0 | K.SEVAHR.F |
| 98 - 105 | 933.4150 | 932.4077 | 932.5113 | -111 | 0 | K.LCTVATLR.E |
| 118 - 130 | 1714.5910 | 1713.5837 | 1713.7893 | -120 | 1 | K.QEPEERNECFLQHK.D |
| 131 - 138 | 940.3230 | 939.3157 | 939.4410 | -133 | 0 | K.DDNPMLPRL.L |
| 162 - 168 | 927.3710 | 926.3637 | 926.4861 | -132 | 0 | K.YLYEIAR.R |
| 170 - 183 | 1742.6820 | 1741.6747 | 1741.8868 | -122 | 0 | R.HPYFYAPELLFFAK.R |
| 170 - 184 | 1898.8480 | 1897.8407 | 1897.9879 | -78 | 1 | R.HPYFYAPELLFFAKR.Y |
| 206 - 214 | 1074.4410 | 1073.4337 | 1073.5353 | -95 | 1 | K.LDELRLDEGK.A |
| 250 - 257 | 880.3920 | 879.3847 | 879.4338 | -56 | 0 | K.AEFAEVSK.L |
| 265 - 281 | 2086.6380 | 2085.6307 | 2085.8303 | -96 | 0 | K.VHTECCHGDLLECADDR.A |
| 265 - 286 | 2584.8760 | 2583.8687 | 2584.1105 | -94 | 1 | K.VHTECCHGDLLECADDRADLAK.Y |
| 348 - 360 | 1639.6770 | 1638.6697 | 1638.7752 | -64 | 0 | K.DVFLGMFLYEPYAR.R Oxidation (M) |
| 361 - 372 | 1467.6130 | 1466.6057 | 1466.8358 | -157 | 1 | R.RHPDYSVLLLR.L |
| 362 - 372 | 1311.5700 | 1310.5627 | 1310.7347 | -131 | 0 | R.HPDYSVLLLR.L |
| 397 - 413 | 2044.8740 | 2043.8667 | 2044.0881 | -108 | 0 | K.VFDEFKPLVEEPQNLIK.Q |
| 427 - 434 | 960.4450 | 959.4377 | 959.5552 | -122 | 0 | K.FQNALLVR.Y |
| 491 - 496 | 674.2470 | 673.2397 | 673.3395 | -148 | 0 | K.TFVSDR.V |
| 500 - 508 | 1138.3620 | 1137.3547 | 1137.4907 | -120 | 0 | K.CCTESLVNR.R |
| 509 - 524 | 1910.6990 | 1909.6917 | 1909.9244 | -122 | 0 | R.RPCFSALEVDETYVPE.E |
| 525 - 543 | 2259.8210 | 2258.8137 | 2259.0154 | -89 | 0 | K.EFNAETFTFHADICTLSEK.E |

No match to: 677.2720, 705.2470, 1578.8360, 2066.8980, 2604.8970, 2733.0130

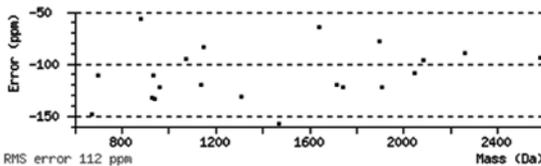


Fig. 5. Detailed information for the protein identified in the Mascot search results in Fig. 4.

be due to the presence of contaminating proteins in the digest, non-tryptic cleavages, incomplete digestion, modified peptides, errors in the database sequence, etc. (8).

The graph at the bottom of Fig. 5 plots the error for each of the matched peptide masses and is useful for assessing whether the tolerance setting used for the search was appropriate. If desired, a more appropriate setting could be chosen and the data re-searched.

3.6. Failure to Obtain a Significant Hit

Some of the most common reasons why a database search may not give a significant match are as follows:

1. Too few masses in the PMF data – depending on the size of the database and the parameters settings, a minimum of 5–7 matched peptides are typically required for a significant hit. This situation could arise simply because the spectrum is very weak and so only a few of the strongest peptide signals are visible, or because the particular protein does not yield many peptides within the useful mass range (e.g. small proteins, proteins with very high or very low numbers of potential trypsin cleavage sites, proteins which are resistant to trypsin digestion).
2. Mass spectral noise included in the PMF data – ensure that the peak detection threshold is not set too low.
3. More than one protein present in the sample – with good quality PMF data, mixtures containing 2–3 proteins can be identified, but the presence of peptide masses from the other proteins reduces the score for each individual match. With lower quality data, this can result in all of the scores dropping below the significance threshold. The presence of contaminants, e.g. trypsin or keratins has the same effect; however, known peptide masses derived from keratins and trypsin can be removed from the PMF data prior to searching (9).
4. Peptide tolerance set too low or too high.
5. Incorrectly calibrated spectrum.
6. Searching EST databases – ESTs are generally short and therefore do not usually provide sufficient numbers of matched peptides for a significant score.
7. The protein is not present in the database being searched.

4. Notes

1. Contamination is a serious concern when attempting to identify low amounts of protein by MALDI-MS. Use only the highest quality reagents, HPLC grade solvents and high quality plasticware to minimize non-protein contamination. Ideally a set of glass/plasticware and reagents/buffers should

be dedicated for MALDI-MS analysis. Laboratory dust is a major source of protein contaminants (keratins) and so every effort to exclude dust from samples, buffers, and reagents must be taken. Communal buffers, stains, destains, etc. are often contaminated with dust and should be avoided. Rinse all glassware with high quality water before making up buffers, etc. Wear nitrile gloves (latex contains protein contaminants) and a lab coat. Keep a cover over the gel during staining, etc. Never stain gels in containers that have been used for processing Western blots, as they will be contaminated with blocking proteins. Only handle the gel if absolutely necessary and then avoid touching parts of the gel that are to be cut. Gel pieces can be cut and processed in laminar flow cabinets, although this is usually not necessary if sensible precautions are taken.

2. Modified trypsin is preferred for protein digestion as it is less susceptible to autodigestion (7). The trypsin 10× stock solution is prepared by dissolving the trypsin (20 μg) in the solvent supplied with the enzyme (50 mM acetic acid; 200 μL). This is stored at -70°C in aliquots (10 μL) and is stable for at least 1 year.
3. MALDI analysis kits (e.g. Sequazyme peptide mass standards kit from AB Sciex) are a convenient way to obtain matrix, peptide standards, and solvents; however, these items can easily be purchased independently. Any peptides which span a reasonable proportion of the useful mass range can be used for calibration, provided their accurate molecular masses are known. α-cyano-4-hydroxycinnamic acid is the most common matrix used for peptide analysis. Matrix solution is prepared by dissolving 5 mg of matrix in 1 mL of 4:1 acetonitrile/water containing 0.1% TFA. A mixture of peptide calibration standards, e.g. des-Arg¹-bradykinin (monoisotopic mass 904.4681), Angiotensin 1 (monoisotopic mass 1,296.6853), ACTH (1–17 clip) (monoisotopic mass 2,093.0867), ACTH (18–39 clip) (monoisotopic mass 2,465.1989) is made to a concentration of 2 pmol/μL each in 0.1% TFA in water. This solution is then mixed 1:1 with matrix solution to give the final peptide calibration mixture.
4. Larger volumes of digests can be concentrated by vacuum centrifugation, but this can lead to significant peptide losses, particularly if the sample is concentrated to dryness. Alternatively, peptides can be concentrated and desalted by using a micro-scale pipette-tip format solid-phase extraction device (e.g. micro C18 Zip Tips from Millipore, Omix C18MB tips from Varian or STAGE tips from Proxeon). The details of their use in desalting peptide solutions for MALDI analysis are given with the product, but for the final elution step, use the matrix solution described above, to elute the peptides directly onto the MALDI plate.

5. Freely available PMF search engines include: Mascot from Matrix Science (http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF), MS-Fit from Protein Prospector (<http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msfitstandard>) and Aldente from ExPaSy (<http://www.expasy.org/tools/aldente>).
6. SwissProt is a relatively small, but highly annotated database with minimal redundancy. Uniprot, NCBIInr, and MSDB are much larger, but have multiple entries for many proteins. Other search engines may have additional database choices, and the use of an in-house copy of the search software allows custom databases to be used.

Specifying the taxonomy as closely as possible reduces the number of database entries which need to be considered in the search. This reduces search times, but more importantly, reduces the threshold score for significant matches, increasing the confidence of any protein identification.

The cleavage specificity of trypsin is C-terminal to Lys and Arg residues (except where followed by Pro). However, not every such peptide bond will be cleaved. The number of missed cleavages to consider can be specified, but increasing the number decreases the significance of any match and should normally be set to 1.

Fixed modifications are modifications to specific amino acids which are considered to be complete, i.e. every occurrence of the amino acid in the sequence is assumed to carry the modification and the unmodified amino acid is not considered. Variable modifications, on the other hand, are incomplete and therefore both the modified and unmodified amino acid is considered in the search. In the example discussed above, the reduction/alkylation should result in complete carbamidomethylation of all cysteine residues, thus “Carbamidomethyl (C)” was chosen as a fixed modification, whereas methionine oxidation, a common artifactual modification which is usually incomplete, was selected as a variable modification. The use of multiple variable modifications can greatly reduce the significance of any match and should therefore be used with caution.

Most PMF search engines allow a molecular weight for the protein to be entered; the search will then consider only database entries within a window around this value. This can be dangerous, as proteins are subject to processing/degradation which can significantly increase or decrease their molecular weight. Mascot uses a more complex method of applying this parameter, but in most cases this can be left blank.

The peptide tolerance is a window around each mass value in the peak list within which a theoretical database peptide mass must fall, in order to be matched.

Acknowledgments

The authors acknowledge the support of the Biotechnology and Biological Sciences Research Council, UK.

References

1. Pappin DJC, Hojrup P, Bleasby, AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology* 3:327–332
2. Neuhoff V, Stamm R, Eibl H (1985) Clear background and highly sensitive protein staining with Coomassie Blue dyes in polyacrylamide gels: a systematic analysis. *Electrophoresis* 6:427–448
3. Herbert B, et al (2001) Reduction and alkylation of proteins in preparation of two-dimensional map analysis: Why, when, and how? *Electrophoresis* 22:2046–2057
4. Shevchenko A, et al (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal Chem* 68:850–858
5. Gharahdaghi F, et al (1999) Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: A method for the removal of silver ions to enhance sensitivity. *Electrophoresis* 20:601–605
6. Sechi S, Chait BT (1998) Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Anal Chem* 70:5150–5158
7. Harris WA, Janecki DJ, Reilly JP (2002) Use of matrix clusters and trypsin autolysis fragments as mass calibrants in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom* 16:1714–1722
8. Karty JA, et al (2002) Artifacts and unassigned masses encountered in peptide mass mapping. *J Chrom B* 782:363–383
9. Schmidt F, et al (2003) Iterative data analysis is the key for exhaustive analysis of peptide mass fingerprints from proteins separated by two-dimensional electrophoresis. *J Am Soc Mass Spectrom* 14:943–956

INDEX

A

- AcMNPV. *See Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV)
- Acrylodan 96–101, 104, 105, 110–113, 116
- Activation loop 96, 97, 101, 106
- Adenosine triphosphate (ATP) 7, 8, 95–97, 101, 120, 121, 126–131, 137, 146–149, 151, 159, 160, 203, 211, 213
- Affinity
- capture 5, 33, 34, 40, 45, 47–50, 86
 - ligands 8, 34, 40, 41, 57–72, 77, 81, 96, 97, 103, 114, 120, 155
 - matrices 7, 8, 41, 58, 59, 67, 69, 75–82, 155
 - purification 57, 58, 63, 72, 75, 80, 136, 143, 155, 156, 224
- ÅKTA explorer™ FPLC 62, 63, 71
- ω-Alkynyl fatty acids 85–93
- Allosteric inhibitors 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117
- Analytes 33, 39–41, 45–51, 72, 120
- Aptamer 40, 45, 119–131
- Aptasensor 120, 121, 123, 125–127, 129–131
- ATP. *See* Adenosine triphosphate (ATP)
- Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) 187–188
- Autophosphorylation 146, 147, 159, 160

B

- Bacmid 136, 137, 139, 188
- Bac-to-Bac system 188
- BaculoGold system 188
- Baculovirus 136, 137, 139–141, 158, 187–198
- Baculovirus transfer vector 189, 192, 197
- BCCP. *See* Biotin carboxyl carrier protein (BCCP)
- Biology-oriented synthesis 15–16
- Biomimetic 16, 57–72
- Biomolecular interaction analysis 33
- Biotin carboxyl carrier protein (BCCP) 138, 139, 141, 142, 145–146, 150, 152, 153, 156–158, 160
- Biotinylation 133, 141, 142, 156, 157
- Build/couple/pair strategy 18–20
- Build/pair strategy 18–19

C

- Cell-free expression. *See also* Cell-free protein synthesis
- batch configuration 208, 211–212, 214–217, 220, 221
 - continuous exchange configuration (CECF) 202, 208, 210–218, 220, 221, 223
- Cell-free protein synthesis 201–224
- Cell lysates 8, 49, 75, 81, 82, 90, 92, 141
- CFG. *See* Consortium of functional glycomics (CFG)
- Chemical
- biology 3, 4, 6, 7, 11, 15, 20, 75, 147, 163
 - file formats 27–28
 - genetics 3–9
 - genomics 3–9
 - probes 5, 7, 45, 86, 145, 148
 - proteomics 3–9, 46, 133–161
 - space 11–20, 27, 28, 45
- Chirality 11–13
- Chromatographic screening 66–67
- Chronocoulometry (CC) 120, 129
- Click chemistry 87
- C-Map. *See* Connectivity map
- Colony PCR 179, 183, 184
- Combinatorial
- chemistry 12, 13, 58
 - library 12, 13, 25, 27–30, 58, 61–65
- Computer-aided drug design 30
- Connectivity map (C-map) 5–6
- Consortium of functional glycomics (CFG), 164
- Cytochrome P450 135, 137, 142, 149, 152, 160

D

- Data analysis 6, 11, 109, 133, 146, 147, 149, 151–154, 158, 233
- Diversity oriented synthesis 11, 16–20

E

- E.coli*. *See Escherichia coli*
- Electrochemical
- assays 119–131
 - sensor 119–131
- Electrospray ionization (ESI) 48, 79, 227

- Emission spectra..... 97, 100–103, 110–112, 115
Enhanced chemiluminescence..... 89, 91
Epigenomics..... 6–7
Escherichia coli
BL21(DE3)..... 81, 175, 177, 179, 180, 184, 205, 208
DH5 α 150, 175, 176, 179
ESI. *See* Electrospray ionization (ESI)
Expression vectors
pET-derived..... 175, 176, 179,
180, 183, 209
Takara pG-KJE8, pGro7 and pKJE7..... 180, 184
- F**
- Fatty acylation..... 85, 86
FLiK technology..... 96, 97, 105
Fluorescent labelling..... 34, 148, 163
Fluorescently-labelled
antibody..... 147, 148
compound..... 116, 134, 148, 149, 164, 168
Frontal analysis..... 63, 67–68, 72
Functional analysis..... 134
- G**
- GBPs. *See* Glycan binding proteins (GBPs)
Gel electrophoresis
agarose..... 177
1D..... 228, 229, 232
2D..... 134, 228, 229
sodium dodecyl sulphate polyacrylamide..... 5, 8, 60,
68, 72, 86, 88, 91, 141, 142, 157, 182, 183, 195, 196,
208, 217–219, 223 (*see also* Sodium dodecyl sulphate
polyacrylamide-PAGE)
GFP. *See* Green fluorescent protein (GFP)
Glycan binding proteins (GBPs)..... 163, 164, 166–170
Green fluorescent protein (GFP)..... 204, 207–209,
214–217, 220–222
- H**
- High5 insect cells. *See Trichopulsia ni*
High performance liquid chromatography
(HPLC)..... 81, 165–167, 170, 183, 237
High-throughput screening (HTS)..... 11, 12, 58,
97–99, 105–108, 112, 115, 116, 220
His-tag..... 41, 155, 175, 191, 192, 196
HPLC. *See* High performance liquid chromatography
(HPLC)
HTS. *See* High-throughput screening (HTS)
Huisgen cycloaddition reaction..... 86
- I**
- Immunoblotting. *See* Western blotting
Immunoglobulins (Ig)..... 47, 57–72
Inclusion bodies..... 173, 202, 218
- In-gel
fluorescence..... 86, 87, 91
reduction/alkylation..... 230
trypsin digestion..... 230
Insect cells. *See* Baculovirus
In silico design..... 25–30
- K**
- K_d determination..... 60, 97, 98, 101–103,
107, 113–116, 149, 154, 155
Kinase inhibitors
DFG-in..... 96, 99, 101, 110, 112
DFG-out..... 95–117
- L**
- LC-MS/MS..... 49
Lectin..... 45, 163, 166, 168, 169
LIC. *See* Ligation-independent cloning (LIC)
Ligation-independent cloning (LIC)..... 138, 141,
174, 175, 192
Liposomes..... 202, 219, 220, 224
- M**
- MALDI-TOF..... 33, 48–51, 165, 167,
227–239
Mascot search..... 233–236, 239
Mass spectrometry (MS)..... 5, 8, 34, 46, 48–51,
66, 134, 165, 198, 227–239
Metabolic labeling..... 88
Microarrays
glycan..... 163–170
printing..... 41, 46, 137, 142–146, 155,
157–159, 164–168, 170
scanner..... 137, 146, 147,
149, 151–154
Molecular chaperones..... 174, 177, 180, 181, 184
Multicomponent reactions
diastereoselective..... 13, 14, 17–18
enantioselective..... 13–16, 18
Ugi..... 18, 19, 57–72
Multiplexed cloning..... 141–142
- N**
- Nanoparticles (NPs)
gold..... 43, 44, 120, 121,
123–124, 128
inorganic..... 120
Natural product..... 14–16, 75, 76
Ni-capture..... 195–196
Ni-purification..... 182–183
- O**
- Oligonucleotides..... 43, 45, 122

P

Partition equilibrium 63, 67
 PCR amplification..... 138, 183
 Peptide mass fingerprint (PMF).....48, 228–237, 239
 Phosphorylation 8, 96, 146–148, 159, 160
 Photo-cross-linking.....79
 Plasmid DNA..... 176, 179, 197, 204, 208
 PMF. *See* Peptide mass fingerprint (PMF)
 Polypeptide tag 136
 Polypharmacology95
 Post-translational modification 174
 Protein
 intracellular..... 175, 176, 188, 189, 191, 195, 196
 kinases 5, 7, 8, 34, 85, 95–117, 136,
 137, 142, 145–151, 159, 203, 214
 membrane..... 142, 201, 202, 206, 209,
 217–220, 223, 224
 myristoylation 85–93
 palmitoylation..... 85–93
 secreted 188, 189, 191, 195, 196
 solubility 173–175, 180, 182, 202, 222

R

Rational design..... 58, 59, 63
 RCA. *See* Rolling cycle amplification (RCA)
 Real-time kinetic measurements 103–105
 Recombinant protein expression
 cell-free 136
 in *E.coli*..... 173–185
 in insect cells..... 136, 137, 139, 155, 157, 187–198
 Restriction-free (RF) cloning165, 174–179,
 183, 184, 192, 197
 Reverse chemical genetics..... 5
 Rhodamine-azide86, 87, 89–91, 93
 Rolling cycle amplification (RCA) 42, 120, 168

S

Screening
 primary 107–108, 116
 secondary 108, 109, 117

SDA. *See* Strand displacement amplification (SDA)
 SELEX. *See* Systematic evolution of ligands
 by exponential enrichment (SELEX)
 S30 extract 204–207,
 211, 214, 224
 Sf9. *See* *Spodoptera frugiperda* (Sf9)
 Signal amplification..... 120
 Small molecule libraries..... 25
 Sodium dodecyl sulphate polyacrylamide-PAGE
 (SDS-PAGE) 5, 8, 60, 68, 72, 86,
 88, 91, 141, 142, 157, 182, 183, 195, 196, 208,
 217–219, 223
 Spectrophotometer63, 67, 123, 177
Spodoptera frugiperda (Sf9)188, 189, 191–197
 SPR. *See* Surface plasmon resonance (SPR)
 Staining
 coomassie..... 71, 198,
 209, 229, 232
 silver60, 68, 72, 228–230
 Sypro 229
 Strand displacement amplification (SDA)..... 120
 Streptavidin
 coated slides..... 143, 144
 Cy5 166, 169
 -HRP..... 93, 137, 141
 Surface plasmon resonance (SPR)33–51, 97, 103
 Systematic evolution of ligands by exponential
 enrichment (SELEX) 120

T

Transfer vector.....137–139, 142,
 188, 189, 191–199
Trichopulsia ni.....188, 189

V

Virtual combinatorial library 27

W

Western blotting.....86, 87, 89,
 91, 150, 157