Functional Protein Microarrays in Drug Discovery
Drug Discovery Series

Series Editor

Andrew Carmen

Johnson & Johnson PRD, LLC
San Diego, California, U.S.A.

1. Virtual Screening in Drug Discovery, edited by Juan Alvarez and Brian Shoichet
2. Industrialization of Drug Discovery: From Target Selection Through Lead Optimization, edited by Jeffrey S. Handen, Ph.D.
3. Phage Display in Biotechnology and Drug Discovery, edited by Sachdev S. Sidhu
6. In Silico Technologies in Drug Target Identification and Validation, edited by Darryl León and Scott Markel
7. Biochips as Pathways to Drug Discovery, edited by Andrew Carmen and Gary Hardiman
Table of Contents

Section 1
Functional Protein Content for Microarrays ....................................................... 1

Chapter 1 High-Throughput Gene Cloning Using the Gateway® Technology .......................................................... 3
Scott N. Peterson, Patrick Burr, Getahun Tsegaye, and Pratap Venepally

Chapter 2 Protein Expression for MicroArrays.................................................. 23
Harry H. Yim, Thomas G. Chappell, and Steven H. Harwood

Chapter 3 Emerging Trend: Cell-Free Protein Expression ................................ 39
Federico Katzen and Wieslaw Kudlicki

Section 2
Fabrication of Functional Protein Microarrays................................................. 51

Chapter 4 The Critical Role of Surface Chemistry in Protein Microarrays ......................... 53
Athena Guo and X.-Y. Zhu

Chapter 5 Fabrication of Sol-Gel-Derived Protein Microarrays for Diagnostics and Screening ........................................................... 73
Nicholas Rupcich and John D. Brennan

Chapter 6 Printing and QC of Functional Protein Microarrays.......................... 99
Dee Shen, Fang X. Zhou, and Barry Schweitzer

Chapter 7 Protein Engineering for Surface Attachment.................................. 115
Aparna Girish, Grace Y. J. Chen, and Shao Q. Yao

Chapter 8 Protein In Situ Arrays through Cell-Free Protein Synthesis .......... 133
Mingyue He, Farid Khan, Elizabeth Palmer, Mingwei Wang, and Michael J. Taussig
Section 3
Detection Methods for Protein Microarrays .................................................... 145

Chapter 9  Fluorescent Detection Methods for Protein Microarrays......... 147
Steven Roman and Scott Clarke

Chapter 10 Functional Analysis of Protein Interactions Using Surface
Plasmon Resonance-Based Microarrays......................................................... 181
Alan McWhirter and Stefan Löfås

Chapter 11 Leaving the Surface Behind: At the Intersection of Protein
Microarrays and Mass Spectrometry .............................................................. 199
Darrell P. Chandler, Daniel S. Schabacker, Sergei Bavykin,
and Igor M. Gavin

Chapter 12 High-Resolution Label-Free Detection Applied to
Protein Microarray Research ........................................................................ 217
Lance G. Laing and Brian Cunningham

Section 4
Applications of Functional Protein Microarrays ............................................. 237

Chapter 13 Studying Protein–Protein Interactions with Protein Microarrays:
Rapid Identification of 14-3-3 Protein Binding Partners .................. 239
Jun-ichi Satoh

Chapter 14 A Combined Force of Chemical Genetics
and Protein Microarrays................................................................................. 261
Heng Zhu and Jing Huang

Chapter 15 Antibody Profiling for Protein Drug Development
and Clinical Development............................................................................. 275
Steve H. Herrmann

Chapter 16 Humoral Response Profiling Using Protein Microarrays ............ 301
Arun Sreekumar, Barry S. Taylor, Xiaoju Wang, David Lubman,
and Arul M. Chinnaiyan
Chapter 17  DNA Interactions with Arrayed Proteins ...........................................313
Marina Snapyan and Vehary Sakanyan

Chapter 18  G Protein–Coupled Receptor Microarrays for Drug Discovery ....333
John Salon, Michael Johnson, Brian Rasnow, Gloria Biddlecome, Yulong Hong,
Brian Webb, Ye Fang, and Joydeep Lahiri

Chapter 19  Kinase Substrate Identification Using Yeast
Protein Microarrays..................................................................................................351
Geeta Devgan and Michael Snyder

Section 5
Bioinformatics & Data Analysis...........................................................................361

Chapter 20  Protein Microarray Image Analysis.................................................363
Minzi Ruan

Chapter 21  The Analysis of Protein Arrays ..........................................................381
Brad Love

Chapter 22  Evaluating Precision and Recall in Functional
Protein Arrays ........................................................................................................403
Keith Robison

Chapter 23  Visualization of Protein Microarray Data........................................415
Kevin Clancy

Index..........................................................................................................................443
Preface

Since their introduction in the 1990s, microarray-based technologies have had a tremendous impact on the biological sciences. One of the most exciting recent developments in this field is functional protein microarrays: microarrays with large numbers of correctly folded and functional proteins. Initially considered an impractical if not impossible goal, high-content functional protein microarrays have now proven their utility in a multitude of applications. While the “field’s” early successes have set the stage for the rapid growth now being witnessed, it is not without its challenges. Indeed, challenges are to be expected in a fast-moving interdisciplinary endeavor such as this, where molecular biology, protein chemistry, bioinformatics, engineering, and physical sciences all intersect.

Currently no book has addressed all aspects of functional protein microarrays in a coherent and integrated fashion. This book is intended to provide the first comprehensive reference for the field, addressing basic principles, methods, and applications. While intended primarily as a reference for industrial, academic, and government scientists, it is also suitable as a graduate-level supplementary text. The book is divided into five main sections, each addressing critical aspects of the field. The first focuses on the generation of functional protein content, which is the first and perhaps most challenging aspect of protein microarrays. The second section describes both “standard” and state-of-the-art fabrication methods, focusing on issues of particular significance to functional protein microarrays. Similarly, the third section reviews current and next-generation approaches to assay detection, which hold one key to the future of the field. The fourth and largest section is dedicated to applications. This section spans the breadth of published applications, from biomolecular interaction discovery and characterization (proteins, antibodies, DNA, small molecules) to humoral response biomarker profiling, enzyme substrate identification, and drug discovery. The final section addresses fundamental computational issues including image and data analysis as well as data visualization.

The intent of this book is to provide the first integrated reference for functional protein microarrays. In doing so, I have aspired to create a volume worthy of the promise of functional protein microarrays, a practical resource capable of conveying the excitement and enabling the development of this field. This book would not have been possible, however, without the hard work of its many authors and Kathie McCoy, to whom I am truly grateful.
Contributors List

Sergei Bavykin  
Argonne National Laboratory  
Argonne, Illinois

Gloria Biddlecome  
Amgen Incorporated  
Thousand Oaks, California

John D. Brennan  
Department of Chemistry  
McMaster University  
Hamilton, Ontario

Patrick Burr  
The Institute for Genomic Research  
The Pathogen Functional Genomics Resource Center  
Rockville, Maryland

Darrell P. Chandler  
Argonne National Laboratory  
Argonne, Illinois  
and  
Akonni Biosystems, Incorporated  
Frederick, Maryland

Thomas G. Chappell  
Invitrogen Corporation  
Carlsbad, California

Grace Y. J. Chen  
Departments of Chemistry and Biological Sciences  
Medicinal Chemistry Program of the Office of Life Sciences  
National University of Singapore  
Republic of Singapore

Arul M. Chinnaiyan  
Department of Pathology  
and  
Department of Urology  
and  
Bioinformatics Program  
and  
Comprehensive Cancer Center  
University of Michigan Medical School  
Ann Arbor, Michigan

Kevin Clancy  
Informax  
Frederick, Maryland

Scott Clarke  
Molecular Probes  
Eugene, Oregon

Brian Cunningham  
Department of Electrical and Computer Engineering  
Micro and Nanotechnology Laboratory  
University of Illinois at Urbana-Champaign  
Urbana, Illinois

Geeta Devgan  
Department of Molecular, Cellular and Developmental Biology  
Yale University  
New Haven, Connecticut

Ye Fang  
Biochemical Technologies, Science and Technology Division  
Corning Incorporated  
Corning, New York
Igor M. Gavin  
Argonne National Laboratory  
Argonne, Illinois  
and  
University of Illinois at Chicago  
Chicago, Illinois  

Aparna Girish  
Departments of Chemistry and  
Biological Sciences  
Medicinal Chemistry Program of  
The Office of Life Sciences  
National University of Singapore  
Republic of Singapore  

Athena Guo  
MicroSurfaces, Inc.  
Minneapolis, Minnesota  

Steven H. Harwood  
Eugene, Oregon  

Mingyue He  
Technology Research Group  
Protein Technologies Laboratory  
The Babraham Institute  
Cambridge, U.K.  

Steve H. Herrmann  
Biological Technologies  
Wyeth Research  
Cambridge, Massachusetts  

Yulong Hong  
Biochemical Technologies, Science  
and Technology Division  
Corning Incorporated  
Corning, New York  

Jing Huang  
Department of Molecular and Medical  
Pharmacology  
David Geffen School of Medicine  
University of California  
Los Angeles, California  

Michael Johnson  
Amgen Incorporated  
Thousand Oaks, California  

Federico Katzen  
Invitrogen Corporation  
Carlsbad, California  

Farid Khan  
Technology Research Group  
Protein Technologies Laboratory  
The Babraham Institute  
Cambridge, U.K.  

Wieslaw Kudlicki  
Invitrogen Corporation  
Carlsbad, California  

Joydeep Lahiri  
Biochemical Technologies, Science  
and Technology Division  
Corning Incorporated  
Corning, New York  

Lance G. Laing  
SRU Biosystems  
Woburn, Massachusetts  

Stefan Löfås  
Biacore AB, Rapsgatan  
Uppsala, Sweden  

Brad Love  
Corporate Research Laboratory  
Invitrogen Corporation  
Carlsbad, California  

David Lubman  
Department of Chemistry  
and  
Department of Surgery  
University of Michigan Medical School  
Ann Arbor, Michigan  

Alan McWhirter  
Biacore AB, Rapsgatan  
Uppsala, Sweden  

Elizabeth Palmer  
Technology Research Group  
Protein Technologies Laboratory  
The Babraham Institute  
Cambridge, U.K.
Scott N. Peterson
The Institute for Genomic Research
The Pathogen Functional Genomics Resource Center
Rockville, Maryland

Brian Rasnow
Amgen Incorporated
Thousand Oaks, California

Keith Robison
Computational Biology
Millennium Pharmaceuticals Incorporated
Cambridge, Massachusetts

Steven Roman
Invitrogen Corporation
Carlsbad, California

Minzi Ruan
VigeneTech Incorporated
Boston, Massachusetts

Nicholas Rupcich
Department of Chemistry
McMaster University
Hamilton, Ontario

Vehary Sakanyan
Unité Biotechnologie, Biocatalyse, Biorégulation
Université de Nantes and ProtNeteomix
Nantes, France

John Salon
Amgen Incorporated
Thousand Oaks, California

Jun-ichi Satoh
Department of Immunology
National Institute of Neuroscience
Tokyo, Japan

Daniel S. Schabacker
Argonne National Laboratory
Argonne, Illinois

Barry Schweitzer
Protein Array Center
Invitrogen Corporation
Branford, Connecticut

Dee Shen
Protein Array Center
Invitrogen Corporation
Branford, Connecticut

Marina Snapyan
Centre de Recherché Université Laval
Robert-Giffard (CRULRG)
Quebec, Quebec

Michael Snyder
Department of Molecular, Cellular and Developmental Biology
Yale University
New Haven, Connecticut

Arun Sreekumar
Department of Pathology and
Comprehensive Cancer Center
University of Michigan Medical School
Ann Arbor, Michigan

Michael J. Taussig
Technology Research Group
Protein Technologies Laboratory,
The Babraham Institute
Cambridge, U.K.

Barry S. Taylor
Department of Pathology and
Bioinformatics Program
University of Michigan Medical School
Ann Arbor, Michigan
Getahun Tsegaye
The Institute for Genomic Research
The Pathogen Functional Genomics Resource Center
Rockville, Maryland

Pratap Venepally
The Institute for Genomic Research
The Pathogen Functional Genomics Resource Center
Rockville, Maryland

Mingwei Wang
National Center for Drug Screening
Shanghai Institute of Materia Medica
Shanghai, China

Xiaoju Wang
Department of Pathology
University of Michigan Medical School
Ann Arbor, Michigan

Brian Webb
Biochemical Technologies, Science and Technology Division
Corning Incorporated
Corning, New York

Shao Q. Yao
Departments of Chemistry and Biological Sciences
Medicinal Chemistry Program of the Office of Life Sciences
National University of Singapore
Republic of Singapore

Harry H. Yim
Invitrogen Corporation
Carlsbad, California

Fang X. Zhou
Protein Array Center
Invitrogen Corporation
Branford, Connecticut

Heng Zhu
Department of Pharmacology and HiT Center
The Johns Hopkins University School of Medicine
Baltimore, Maryland

X.-Y. Zhu
Department of Chemistry
University of Minnesota
Minneapolis, Minnesota
Introduction

As central actors in most biological responses, proteins are the subject of intense study for both basic and drug research. This, in turn, has driven the development of increasingly sophisticated approaches for the study of proteins, which, in recent years has extended to proteomic level methodologies. Despite this need, however, microarray technologies for proteins have lagged behind those for nucleic acids. This has been particularly evident in the case of functional protein microarrays, where formidable technical challenges must be surmounted. However, while technical challenges still remain, the past few years have witnessed movement of the field from basic proof of concept\textsuperscript{1,2} to the use of microarrays for important scientific work, landmark discoveries\textsuperscript{3} to proteomic characterizations.\textsuperscript{4} At the same time, the number of publications in the field has increased exponentially. The purpose of this book is to provide the reader with an up-to-date overview of the field, as well as the background required to actually design and develop arrays or perform and analyze array experiments. The five sections of this book reflect five key considerations in the field: protein content, array fabrication, assay detection, applications and data analysis.

FUNCTIONAL PROTEIN CONTENT

The development of functional protein content is one of the most challenging, and often rate-limiting, aspects of protein microarray experimentation. These challenges can be largely eliminated in cases where protein content or even protein arrays can be acquired commercially. However, in many cases protein content must be generated by the investigator. This content is most typically generated from DNA clones using recombinant expression technology. High-throughput methods for expression clone generation have been developed at The Institute for Genomic Research, and are described in detail in Chapter 1. Important considerations such as information management, automation, quality control and clone validation are addressed. The second chapter addresses expression and purification of proteins in heterologous host systems (\textit{E. coli}, yeast and insect cells), and provides guidance for selecting an appropriate system based on a variety of parameters such as yield, functionality, post-translational modifications, throughput and cost. The final chapter of this section reviews cell-free protein expression systems, and discusses specific considerations for protein microarrays. Together, these chapters provide a thorough overview of the basic considerations for protein content generation.
FABRICATION

The functional and structural heterogeneity of proteins makes arraying and functional surface attachment a considerable challenge. Chapter 4 provides a thorough examination of the challenges of surface chemistry for protein microarrays, which include minimizing nonspecific interactions and maximizing the presentation of conformationally correct proteins. A completely different approach is described in Chapter 5 with the entrapment of proteins in a three-dimensional sol-gel. The various strategies are illustrated in Figure 1.

Critical aspects of array manufacture are addressed in Chapter 6. These include a brief review of commercially available printing technologies, the myriad challenges presented by protein microarrays, and quality control in manufacturing.

The final chapters of this section describe novel strategies for generating protein arrays. Chapter 7 focuses on oriented immobilization strategies based on protein engineering and chemistry, while Chapter 8 addresses the *in situ* generation of proteins. Both chapters describe methods that can “compress” the steps involved in making an array, by combining purification (Chapter 7) or expression and purification (Chapter 8) into the array printing process. These simplified techniques promise to make protein microarray technology more accessible to “average” labs, although at the potential cost of less well controlled array content.

DETECTION

The varied applications of functional protein microarrays all require sensitive assay detection technologies. The most common detection method, fluorescence, is described in chapter 9. This chapter provides a detailed discussion of the basics: fluorescent dyes, fluorescent proteins, time-resolved fluorescence, fluorescent quantum dotes, signal amplification, labeling methods and instrumentation. It concludes

---

**FIGURE 1** Protein immobilization strategies. (a) Proteins are directly attached to the surface based on one (or few) site-specific interactions. This approach has the advantage of a relatively homogeneous presentation of protein to the solution, but regions of the protein may be systematically “hidden.” (b) Proteins are attached in a nonspecific orientation. This approach has the advantage of (collectively) displaying a large fraction of the protein surface, but some protein molecules may be functionally blocked. (c) Proteins are not attached but “caged” in an aqueous environment. This approach has the advantage of displaying proteins in a more “native” manner, but can support only a limited range of applications.
by describing a variety of examples of applications enabled by fluorescent detection technology. As useful as fluorescent detection has proven, however, there is a clear need for “label-free” detection methods. This is especially true of small molecule assays, where the addition of a fluorescent group can significantly alter the chemical and biological properties of the compound under investigation. The most common label-free detection technology, surface plasmon resonance (SPR) is described in Chapter 10. This chapter reviews the basic physics behind the SPR phenomenon, and discusses special considerations for the adaptation of SPR to arrays. Chapter 11 describes recent advances in the application of MALDI (matrix assisted laser desorption ionization) mass spectrometry to protein microarrays. In addition to detection, mass spectrometry can be used for molecular identification, potentially enabling highly multiplexed experiments. Chapter 12 describes a recently commercialized alternative to SPR based on photonic crystal biosensors.

APPLICATIONS

Functional protein microarrays have been adapted for a variety of applications in both basic research and drug discovery. Two basic classes of experiments can be performed with functional protein microarrays: interaction assays and activity assays. Interaction assays profile the ability of molecules (or even cells) to bind to proteins on the array surface. Activity assays profile the activity of proteins either in solution or on the arrays themselves (see Figure 2). The breadth of applications generated through these types of experiments is summarized in Table 1, and discussed in more detail in Chapters 13 to 19.

Chapter 13 describes the use of functional protein microarrays for profiling protein-protein interactions, with a focus on 14-3-3 proteins. The use of protein

![Interaction assays and Activity assays diagram](image)

**FIGURE 2** Basic types of assays. Interaction assays monitor the ability of a molecule/complex (B) to bind proteins on the array (A). Activity assays monitor the activity of proteins in solution, such as an enzyme (E) modifying (m) a substrate (S) on the array. Alternately, such assays can monitor the activity of proteins on the array. (curved arrow represents a biochemical reaction). (Reprinted with permission, Invitrogen)
TABLE 1
Applications of Functional Protein Microarrays. A Summary of Many of the
Basic and Drug Research Applications of Functional Microarray Experiments

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Basic Research Application</th>
<th>Drug Research Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein–protein interaction profiling</td>
<td>Pathway mapping</td>
<td>Target discovery</td>
</tr>
<tr>
<td>Protein–DNA interaction profiling</td>
<td>Protein interaction mapping</td>
<td>Early target validation</td>
</tr>
<tr>
<td>Protein–lipid interaction profiling</td>
<td>Protein function determination</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_d$ estimation</td>
<td></td>
</tr>
<tr>
<td>Substrate assays</td>
<td>Pathway mapping</td>
<td>Target discovery</td>
</tr>
<tr>
<td></td>
<td>Substrate identification</td>
<td>Early target validation</td>
</tr>
<tr>
<td>Enzyme activity profiling</td>
<td>Pathway mapping</td>
<td>Target discovery</td>
</tr>
<tr>
<td></td>
<td>Enzyme activity discovery</td>
<td>Early target validation</td>
</tr>
<tr>
<td>Protein–small molecule interaction profiling</td>
<td>Pathway mapping</td>
<td>Target/mechanism determination</td>
</tr>
<tr>
<td></td>
<td>Metabolomics</td>
<td>Drug rescue</td>
</tr>
<tr>
<td></td>
<td>Chemical genomics</td>
<td>Alternate target identification</td>
</tr>
<tr>
<td>Antibody specificity profiling</td>
<td>Antibody characterization</td>
<td>Biotherapeutic development and optimization</td>
</tr>
<tr>
<td>Immune response profiling</td>
<td>Biomarker discovery</td>
<td>Diagnostic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Biomarkers for efficacy and safety</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vaccine design</td>
</tr>
<tr>
<td>Enzyme inhibitor profiling</td>
<td>Enzyme characterization</td>
<td>Specificity profiling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lead selection and optimization</td>
</tr>
<tr>
<td>Enzyme activity assay</td>
<td>Enzyme kinetics</td>
<td>Specificity profiling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IC_{50}$ determination</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lead selection and optimization</td>
</tr>
</tbody>
</table>


microarrays for small molecule target identification is described in chapter 14, with an emphasis on the author’s pioneering use of protein arrays to study chemical genetics with the compound SMIR4. Chapter 15 discusses the possible uses of functional protein microarrays for biotherapeutic drug development, with a particular
focus on using arrays to identify cross-reactive therapeutic antibodies, as well as monitoring for autoimmune side effects. Chapter 16 describes the use of protein arrays to discover antibody immune response biomarkers, an application which also has implications for vaccine design and testing. The use of protein arrays to study DNA binding is described in Chapter 17, including a discussion of the clinical significance of such investigations. One of the most important and challenging classes of proteins, multi-transmembrane spanning G protein-coupled receptors (GPCRs), is addressed in Chapter 18. This chapter provides a thorough description of this application, from surface chemistry to binding assay protocols and assay validation. Finally, Chapter 19 describes the use of protein arrays for the identification of kinase substrates, focusing on the application of this technique to the yeast proteome.

While exhaustive coverage of all applications is not possible, this section describes all of the major uses of protein microarrays currently under investigation. No doubt, as the field evolves, new applications will be developed. The basics described in this section, though, should provide a good foundation for understanding these future developments.

DATA ANALYSIS

Data analysis is one of the most important, but often underappreciated, aspects of the use of protein microarrays. This section starts with a thorough discussion of image analysis in Chapter 20. Numerous considerations, from spot boundary assignment and contaminant removal to statistical analysis and visualization, are described. Chapter 21 takes off from there, describing approaches to analyzing the numerical data generated directly from the images. Although focusing on biomarker discovery, many of the approaches described in Chapter 21 are directly applicable to the other applications described in this book. Chapter 22 uses computer simulations to help evaluate the potential of protein microarrays for kinase substrate identification. Like the previous chapter, however, the basic approach is applicable to many other applications. The final chapter examines the software requirements for visualizing, sharing and integrating the results of experimentation. It is only with this ability, after all, that the full potential of this technology will be realized.

REFERENCES

Section 1

Functional Protein Content for Microarrays
1 High-Throughput Gene Cloning Using the Gateway® Technology

Scott N. Peterson, Patrick Burr, Getahun Tsegaye, and Pratap Venepally

CONTENTS

Introduction................................................................................................................3
Gateway Recombinational Cloning...........................................................................6
  PCR Amplification of ORFs..............................................................................7
    PCR Product Verification and Quantitation..................................................8
  BP Clonase Reaction.........................................................................................9
  E. coli Transformation......................................................................................9
Gateway Clone Resource Validation Procedure....................................................10
  Sequence Assembly.......................................................................................13
  Validation and Reports....................................................................................15
Destination Vector Cloning..................................................................................16
Technology and Robotics.......................................................................................17
Methods and Materials..........................................................................................19
  PCR Amplification of ORFs............................................................................19
  PCR Product Verification and Quantitation...................................................19
  BP Clonase Reactions......................................................................................20
  DH10B-T1 E. coli Transformation....................................................................20
  Clone Sequence Validation.............................................................................20
  Plasmid Extraction...........................................................................................20
  Sequencing Template Production by TempliPhi..............................................21
References.............................................................................................................21

INTRODUCTION

The genomic era has produced an ever-increasing number of complete genome sequences from a wide variety of organisms. The large number of annotated gene sequences being produced has driven the advancement of numerous complementary technologies that enable research scientists to exploit the availability of genome
sequence data in new and powerful ways. One such technology is the Gateway®
cloning system made available by Invitrogen Inc.

The introduction of this technology was particularly well timed in relation to
genomic sequencing, since the Gateway platform provided a vehicle for the cloning
and expression of complete open reading frames (ORFs). Prior to the introduction
of the Gateway cloning technology, classical cloning strategies using restriction
enzymes and DNA ligase were fully entrenched. The primary limitation of traditional
cloning procedures was the difficulty in implementing high-throughput approaches.
The displacement of the traditional methods would require a clear and substantial
improvement in efficiency, ease of use and automation potential. The Gateway
technology delivered these requirements. The increased use of this cloning system
is in turn driving the development of a novel series of technologies that these
expression clones feed directly. Foremost are those technologies associated with in vivo and in vitro protein expression and purification for functional and structural
analysis of proteins. The improved efficiency and ease of generating the raw mate-
rials (DNA expression clones, purified recombinant proteins) are supporting a vig-
orous growth in the use of immobilized proteins on glass surfaces. These advances
hold promise for accelerating the discovery of functional roles of genes and provide
new strategies for identifying drug targets and therapeutics.

In response to the challenge put forward by the Nation Institute for Allergy and
Infectious Disease (NIAID) to generate and distribute cloned ORFs to the scientific
community, to enable functional genomics of microbial pathogens, viruses and
parasites, the Pathogen Functional Genomics Resource Center (PFGRC) at The
Institute for Genomic Research (TIGR) was motivated to identify a cost-effective
and efficient cloning technology. It was important to select a strategy that not only
provided the necessary efficiency for high-throughput cloning, but also one that
would be widely recognized and well-accepted by the diverse scientific end-user.
The widespread adoption of the Gateway cloning platform was fortuitous since major
cloning efforts performed in other laboratories are now commonly using the Gateway
platform and have therefore enabled the collaboration and clone sharing among
scientists with diverse scientific interests (see appendix for other users of the tech-
nology). While individual applications may vary, the primary purpose of the Gateway
platform is the generation of cloned ORFs in one or more expression vectors
(destination vector). This is accomplished in two steps that together mimic the
recombination reaction that occurs between the genome of E. coli and that of phage
lambda. The lambda phage genome contains an attP site that undergoes recombi-
nation, with the aid of lambda phage and E. coli–encoded proteins with the 25 bp
attB site in the E. coli genome. Upon recombination, the attP site divides in two
halves, attL and attR, that flank the lambda genome. During lytic phase, lambda
undergoes a second recombination between attL and attR thus reconstituting the
attP site while leaving behind the original attB site. In the context of gene cloning,
a PCR product (ORF) is generated that is flanked by two nonidentical, primer
encoded attB sites (attB1 and attB2). The PCR product undergoes recombination
with a vector containing two nonidentical attP sites (attP1 and attP2). The recom-
bination reaction is efficient and directional. The clones derived from this recombi-
nation reaction are referred to as entry clones. Entry clones contain inserts that are
High-Throughput Gene Cloning Using the Gateway® Technology

flanked by two nonidentical attL sites. The entry clone has no direct function other than to serve as the substrate for the transfer of the ORF into an expression vector via an LR reaction, named because the attL sites in the entry clone recombine directionally with attR sites in the destination vector. The entry clone is considered to be a useful resource since the cloned insert can be readily shuttled into any number of commercially available or user-designed expression vectors (destination vectors) without a priori knowledge of the intended downstream application (Figure 1.1).

The generation of entry clones in a high-throughput manner is a multistep process that, taken together, results in a high overall cloning efficiency. This efficiency can be attributed to two features. First, recombination of att sites in both BP and LR reactions is nearly stoichiometric and requires the input of a purified PCR product and vector DNA into a proprietary mixture of enzymes that catalyze the recombination of the PCR product into the cloning vector. Second is the system’s use of both negative and positive selection in the subsequent transformation of E. coli. The successful recombination between vector and PCR product displaces the resident “stuffer fragment” that consists of the markers ccdB and Cmr. The ccdB gene product interferes with gyrA activity and is therefore toxic to E. coli. Nonrecombinant vector will retain the ccdB and therefore not be frequently recovered following E. coli transformation. The vector backbones used contain standard antibiotic resistance genes for positive selection of transformants. The vast majority (>99%) of colonies that form are recombinant clones. The simplicity of the cloning reaction allows full automation of the steps leading up to and including the cloning reaction itself.
We have developed a nearly fully automated pipeline for the cloning and sequence validation of ORFs using Gateway. The automation not only provides the potential to generate large numbers of recombinant clones but also the implementation of a process for the tracking of materials through a Laboratory Information Management System (LIMS). The development of a functional LIMS serves to reduce sources of human error and reagent waste. The ability to automate the process is very important to our pipeline since, for several steps, second and even third attempts are made on a small number of failed cases requiring “cherry-picking” and subsequent reintegration with the complete clone set. The creation of a fully functional and automated clone validation sequence analysis process has led to increased throughput and efficiency of clone production. The Gateway clone production pipeline (Figure 1.2) illustrates the integration of this multistep process.

GATEWAY RECOMBINATIONAL CLONING

Primer Design: Each of the unique open reading frames (ORFs) identified and annotated in a genome are potential targets for forward and reverse primer design. Recently duplicated genes displaying a high degree of sequence identity may be difficult, if not impossible, to amplify in pure form. Each forward primer contains a 5′, 25 nt attB1 sequence (see Materials and Methods) appended to each gene specific sequence representing the start codon and the neighboring 3′ nucleotides required to achieve a predefined $T_m = 60–65°C$. The reverse primer has a 5′, attB2 sequence appended to gene specific sequence beginning at the nucleotide just upstream of the stop codon. This design feature allows the subsequent flexibility to
create COOH-terminal fusion proteins, wherein a stop codon is conferred by the cloning vector, just downstream of the cloned ORF in each of the three possible reading frames. Some investigators prefer to include a stop codon in the PCR primer. This is accommodated by altering the primer design to include either an endogenous stop codon or a standard (uniform) stop codon. Each primer contains four G residues at the 5’ end. These residues are important for recombination efficiency and serve to internalize the attB sequences so that they do appear at the very end of PCR products. We sort our primer pairs with respect to the anticipated PCR product size from smallest to largest. By restricting the T_m of each primer used within a small range, we can define efficient cycling conditions based on the single variable of extension time. For whole genome applications the range of size of ORFs arranged by size in any 384 grouping is relatively small allowing us to define extension times that are nearly optimal for all targets. We have had very good success using oligonucleotides obtained from Illumina Inc. The forward and reverse primers are synthesized in identical well locations in paired 96-well plates, facilitating manual or automated robotic setup of PCR reactions.

**PCR Amplification of ORFs**

The production pipeline developed in the PFGRC is quite generalized and its overall efficiency is not strongly influenced by the specific ORFs to be cloned. One exception is the prior optimization of PCR conditions for the genomic DNAs of interest. The most pronounced variable to account for is the G+C content of the genome. We have found that species by species optimization of the strategy used for amplification of ORFs using the PCR is critical, especially when large numbers of reactions are to be performed. We have identified four proof-reading polymerases that when applied to particular genomes, perform well (Table 1.1). This list is by no means exhaustive but provides a guideline for robust polymerases for use in a high-throughput cloning process.

Once PCR reaction optimization is complete, high-throughput reaction setup and cycling is ready to begin. We perform PCR in a 35 µl reaction volume in 384-well format. The scale of the reaction provides sufficient yield of product for subsequent cloning reactions. Primer dimers containing both attB sites represent clonable products and therefore behave as active competitors with the ORF in BP cloning reactions. An alternative process that we have not investigated thoroughly is the use of a two-step PCR reaction. The primers used differ from those described above and include only the 3’ half of the attB sequence. After a limited number of PCR cycles, the products are cleaned-up to remove the initial primers and a second set of universal primers containing a complete attB site are used in all reactions. Since the second primer pair is used in all reactions, the cost of primer synthesis can be driven down. After cycling, the PCR products are transferred to 384-well filtration plates (Millipore) using a Beckman Coulter Biomek-FX 96 probe liquid handling robot. For lower throughput applications, a multichannel pipette is a useful alternative. PCR products are purified according to the manufacturer’s suggested procedure and products are eluted in 50 µl of H_2O and finally transferred to a clean, 384-well MJ Research hardshell plate using a Beckman Coulter Biomek-FX 96 probe liquid handling robot.
The purified PCR product yield is determined using a Caliper ASM90 SE capillary electrophoresis instrument (Caliper LifeSciences). The Caliper System uses a “sipper” mounted on a robotic arm to remove ~1 µl from each well. Each PCR product is electrophoresed through the single capillary, where its mobility is compared to a set of size standards. The quantity and relative purity (single band) of each PCR fragment is determined in a matter of 30 seconds. The size estimates in our experience are accurate to ±5%. PCR products deviating by more than 10% from an expected size are flagged. In less than 1% of the cases we observe size estimates outside this range. Interestingly, a significant proportion of these are ultimately determined to

TABLE 1.1
PCR Kits for High-Throughput ORF Amplification

<table>
<thead>
<tr>
<th>Product</th>
<th>Manufacturer</th>
<th>Catalog</th>
<th>Description</th>
<th>Target Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phusion High-Fidelity DNA Polymerase</td>
<td>Finnzymes New England</td>
<td>F-530-L</td>
<td>Proprietary Pyrococcus-like enzyme with processivity enhancing domain and proofreading capacity</td>
<td>H. pylori, B. anthracis, S. agalactiae, S. typhimurium, S. pneumoniae, V. cholerae, Y. pestis</td>
</tr>
<tr>
<td>Platinum PCR SuperMix High Fidelity</td>
<td>Invitrogen</td>
<td>12532–016</td>
<td>Complex of recombinant Taq polymerase and Pyrococcus species GB-D with proofreading capacity</td>
<td>S. aureus COL</td>
</tr>
<tr>
<td>Takara LA PCR kit</td>
<td>Takara</td>
<td>RR013</td>
<td>Proprietary modified Taq polymerase with proofreading capacity</td>
<td>M. tuberculosis</td>
</tr>
<tr>
<td>Advantage –HF PCR kit</td>
<td>BD Biosciences</td>
<td>K1909-1</td>
<td>Proprietary mix of a modified Taq polymerase with a Pfu-like proofreading polymerase</td>
<td>M. tuberculosis</td>
</tr>
</tbody>
</table>

PCR Product Verification and Quantitation

The purified PCR product yield is determined using a Caliper ASM90 SE capillary electrophoresis instrument (Caliper LifeSciences). The Caliper System uses a “sipper” mounted on a robotic arm to remove ~1 µl from each well. Each PCR product is electrophoresed through the single capillary, where its mobility is compared to a set of size standards. The quantity and relative purity (single band) of each PCR fragment is determined in a matter of 30 seconds. The size estimates in our experience are accurate to ±5%. PCR products deviating by more than 10% from an expected size are flagged. In less than 1% of the cases we observe size estimates outside this range. Interestingly, a significant proportion of these are ultimately determined to
be the expected ORF with no structural rearrangements. The DNA yield, size, and purity are stored directly in the instrument’s online computer where it is then classified as passing or failing. Common reasons for scoring a reaction as a failure include low or no yield (<10 ng/µl) or the formation of two or more PCR products (poor primer specificity). Failed reactions (~10%) are identified automatically in a report form that is used to “cherry-pick” appropriate primer pairs for a second pass attempt. A second attempt to amplify failed PCR reactions, using either the identical reaction conditions or those of another kit, generally results in an additional 1 to 8% increase in overall success achieved (~95%). In our experience 0.5 to 1% of PCR failure is attributable to oligonucleotide synthesis. Resynthesis of oligonucleotides for failed reactions imposes additional economic burden on a project with limited returns. A summary of two recent whole genome microbial ORF cloning projects are shown in (Table 1.2). Successful PCR products are then merged together before proceeding to the BP reaction.

### TABLE 1.2
Summary of ORFeome Projects

<table>
<thead>
<tr>
<th>Project</th>
<th>PCR 1st Round</th>
<th>PCR Follow-up</th>
<th>Transformation Heat Shock</th>
<th>Transformation Electroporation</th>
<th>Validation Success Colony 1</th>
<th>Validation Success Colony 2</th>
<th>Overall Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. aureus COL</td>
<td>93.1%</td>
<td>94.4%</td>
<td>98.0%</td>
<td>99.9%</td>
<td>74.3%</td>
<td>14.3%</td>
<td>88.6%</td>
</tr>
<tr>
<td>F. tularensis SHU S4</td>
<td>87.3%</td>
<td>94.6%</td>
<td>96.4%</td>
<td>99.3%</td>
<td>65.8%</td>
<td>18.8%</td>
<td>84.6%</td>
</tr>
</tbody>
</table>

BP Clonase Reaction

The output file generated by the Caliper contains the concentration of each PCR product that is then converted to a molar concentration. This information is fed directly into the Biomek-FX SPAN-8 liquid handling robot for automated set up of BP cloning reactions. We have adopted the use of a scaled down version of the BP reaction using 50 fmol of target vector and PCR product insert. A master mix containing all components other than the PCR product insert is prepared and aliquoted into individual wells. Each PCR product is diluted to a concentration of 25 fmol/µl, and subsequently 2 µl of each are added to the master mix. The BP cloning reaction efficiency is inversely proportional to the size of the PCR product to be cloned. This relationship is only strongly limiting for very large genes (~5 Kb). In an earlier version of our pipeline we set up BP reactions at 2 separate scales 25 fmol for PCR products <2.5 Kb and 100 fmol for PCR products >2.5 Kb. More recently we have used 50 fmol scale reactions for all PCR products.

E. coli Transformation

We have not yet identified a reliable 96-well device for electroporation of electrocompetent E. coli cells. The efficiency of most cloning efficiency chemically prepared competent cells is more than adequate for recovery of recombinant clones.
Functional Protein Microarrays in Drug Discovery

_E. coli_ transformations are conducted in 96-well trays and are set up robotically. After heat shock and recovery in nonselective media for 1 hour, cells are robotically plated onto 20 × 20 cm dishes that are sub-divided into 48 grids. Each grid is preseeded with 6 to 8 glass beads (3 mm) and when all 48 transformation reactions are distributed, the plate is shaken gently until all liquid is absorbed into the solid media. The glass beads are removed by inverting the plate into an appropriate receptacle. Any grid yielding one or fewer colonies is considered a failure. A list of failed transformations is compiled and those BP reactions are used a second time to manually transform electro-competent _E. coli_ cells. An interesting feature we observe is that virtually all of the PCR reactions scored as failures lead to colony generation, following this two step procedure. Slightly more than half of these cases result in valid full-length recombinant clones.

**GATEWAY CLONE RESOURCE VALIDATION PROCEDURE**

The validation of Gateway clones is an important aspect of clone production. It is also the most challenging to conduct. Since Gateway clones are most commonly used for expression, it is important to validate the clones for sequence and length integrity. We have not observed substantial DNA rearrangements of cloned inserts; however, nucleotide substitutions introduced during PCR are frequent enough that further consideration is warranted. Thus far, no standards exist for clone validation. In an attempt to initiate such a standard we have adopted a standard set by The Harvard Institute for Proteomics (HIP). The HIP sequence validation standard is stringent but reasonable and involves rejecting any clones containing more than 2 nonsilent substitutions. Clones containing indels (frame shift), nonsense, and/or mutations in the att sites are rejected. These validation criteria have driven our validation strategy to include a two-tier process that begins with the sequence validation of a single colony. If that DNA insert fails the validation criteria, a second colony is then analyzed. Mutations introduced early in PCR cycling will be present in a large fraction of the resulting colonies; however, in practice we find a relatively high degree of utility in going to a second colony in instances where the first colony was deemed unacceptable. Data in Table 1.2 illustrate the utility derived from the analysis of a second colony and our future interest to determine the point of diminished returns in terms of the number of colonies to select for sequence validation. This option must be weighed against the alternative that is to begin the process again from the start.

Initial attempts to sequence validate Gateway entry clones resulted in unacceptably low sequencing success frequencies. This was particularly evident for clones containing small inserts, <600 bp, but also negatively affected end reads obtained for larger inserts. It was suggested that the _attL_ sites flanking the cloned inserts have significant potential to form secondary structure that polymerases have difficulty traversing. Specific blocking primers were designed to inhibit the secondary structure formation, thus partially alleviating the barrier to polymerase processivity. We have verified the utility of the blocking primers and observed discrete improvement to our overall entry clone sequencing efficiency. Despite the improvement afforded by the use of blocking primers, our sequencing success was
High-Throughput Gene Cloning Using the Gateway® Technology

still below that routinely obtained for other cloning vectors (~90%). We have developed further improvements to the sequence validation of entry clones using phi29 polymerase (Amersham, Inc.) on crude lysates prepared directly from colonies. Templates prepared in this manner are advantageous, although for reasons that are not completely understood. The use of sequencing templates generated through the random priming of plasmid DNAs by rolling circle amplification alleviate the observed sequencing failure for clones containing inserts <600 bp. This strategy has allowed us to generate sequencing results consistent with expected success frequencies and quality. For cloned inserts 600 bp and larger we have had comparable high frequency sequencing success using either templates derived from templiPhi or double-stranded plasmid DNAs. A remaining dilemma in sequence validation of clones involves confirmation of the correctness of the attB site itself. Sequence traces from failed validation attempts are often due to abrupt termination in signal strength as the polymerase reaches the att site. Yet another possible solution to sequence validation is to direct efforts on the destination clone which is flanked only by attB sites.

The generation of DNA templates for sequence validation using templiPhi is simple, inexpensive, and easily automated. A sample of 10 µl from an overnight culture is used to prepare lysates by brief heat treatment, 93°C for 3 minutes. A small volume from the cleared lysate is used as template for templiPhi reactions. The reaction products are diluted to a final volume of 40 µl with H2O. We typically obtain yields between 20 and 40 ng/µl, which is sufficient for approximately 20 sequencing reactions.

The number of sequencing reactions performed to validate a cloned ORF is based on its length. For ORFs 500 bp or less, only end-reads are performed. For ORFs larger than 500 bp internal walking primers are designed to generate reads in both directions. The optimal density of walking primers is dependent on average read length and overall sequencing success. We have compared the outcomes of applying walking primers at a regular spacing of 250 bp and 500 bp (Figure 1.3). Given that an average read length from automated sequencing instruments is now in excess of 800 bp, it may be surprising that walking primers at such high density are required for high-throughput sequencing. The failure of some sequencing primers and reactions is a given and if the spacing of sequencing primers is too great, these failures will result in an inability of neighboring primers to fill the gap. The first-pass sequencing attempt (end reads and walking primers) results in a number of outcomes ranging from perfectly validated clones (2X coverage, no mutations) to assemblies with only partial coverage. The classification of sequence validated clones is described below (Table 1.3). We can see that 500 bp spacing among walking primers compares unfavorably to 250 bp spacing in terms of the frequency at which validated clones are identified. From our perspective the choice between 250 bp and 500 bp walking primers spacing is a matter of decision drivers like economics and throughput. The average gene requires seven walking primers for validation and therefore represents a substantial cost. Reducing these costs by nearly 50% is potentially attractive but does carry the consequence of increasing the amount of second-pass sequencing attempts required to fully validate a clone. These additional attempts also carry a cost.
Depending on the nature and number of remaining clones, directed efforts are applied to close remaining gaps and confirm sequence ambiguities in the assembled sequence. After applying brute force and manual efforts to elusive clones, a final sequence validation report is generated that directs the representation of acceptable clones for distribution to the scientific community. The list of acceptable ORFs is used to direct the robotic compression of the two freezer copies (colony 1 and 2), into a final set that is replicated into several glycerol stock copies.

**FIGURE 1.3** The effect of walking primer spacing on sequence validation. (A). Primer design schema (left). Walking primer pairs (forward and reverse arrows) at an interval of 250 (a) or 500 bases (b) or single alternating forward and reverse primers at 250 base intervals (c, forward first primer; d, reverse first primer) are used for sequencing. The circles (•) demarcate 250 base intervals. The table on the right lists number of primer pairs designed for various ORF length intervals. (B) Primer intervals vs. sequence coverage. Number of clones in each single-contig sequence validation class A, B, C, and D (see text and Table 1.3) obtained with different walking primer intervals, illustrated in Figure 3A, are plotted. The inset table shows the percentage of full-length A and B wild-type class clones and A, B and C class (full-length clones with mutations) clones seen with different walking primer intervals relative to 250 base interval primer pairs (set to 100). The average length of sequences considered in the validation of clones in each class is shown at the bottom.

Depending on the nature and number of remaining clones, directed efforts are applied to close remaining gaps and confirm sequence ambiguities in the assembled sequence. After applying brute force and manual efforts to elusive clones, a final sequence validation report is generated that directs the representation of acceptable clones for distribution to the scientific community. The list of acceptable ORFs is used to direct the robotic compression of the two freezer copies (colony 1 and 2), into a final set that is replicated into several glycerol stock copies.
SEQUENCE ASSEMBLY

The sequences obtained from Gateway entry clones are validated using a novel high-throughput assembly pipeline, called CLASP (CLone validation ASsembly Pipeline) developed by the PFGRC bioinformatics group at TIGR. This software can be accessed upon request (www.pfgrc.tigr.org). Initially, the algorithm performs the assembly of individual sequences, generated from the cloned insert, to form a consensus sequence. Subsequently, the consensus is compared to the reference ORF and a validation report detailing the quality of the cloned sequence is generated.

The clone assembly validation pipeline exploits the fact that the sequence of the insert in the vector is already known. As shown in the Figure 1.4, it uses two separate assembler programs to optimize the assembly of the sequence reads and achieve maximum accuracy in the consensus sequence. The first and the most important of these is AMOScmp (Figure 1.4A) which places a premium on the agreement between the reads and the reference — rather than on the phred quality scores generated from the trace files. This guides the selection of which sequences to use for the final contig and thus the consensus. The AMOScmp assembly allows joining of even short overlapping sequences, resulting in a high recovery of single,
Functional Protein Microarrays in Drug Discovery

full-length contigs (classes A, B and C, Table 1.3) which otherwise might remain as a single (class D) or multiple (class E) partial-length contigs. For sequences that align very poorly with the reference sequence and result in either partial-length contigs or no contigs after applying AMOScmp, a second assembler, called Minimus (Figure 1.4B), is used. In Minimus the amount of sequence used from any read in the assembly are determined by the phred quality scores instead of how well they align with the reference. For that reason, the consensus sequence(s) obtained from the Minimus assembly have less identity with respect to the reference sequence than those generated by the AMOScmp assembler (classes M, and N; Table 1.3). However, the Minimus results are indispensable for (a) ‘catching’ clones which are good but mislabeled in any of the steps along the cloning process (class W) and also in the (b) identification of clones that do not meet the acceptance criteria for valid clones and thus require repeat efforts at cloning and validation. After the assembly with AMOScmp or Minimus, the base calls in the consensus sequence(s) obtained for each clone is verified for accuracy against chromatograms using autoEditor*. In the case of multiple contigs with gaps, autoJoiner* is used to join the neighboring reads.

* Sequencing closure software developed by TIGR.
by relaxing (extending) the clear ranges if they align well with each other above a
set threshold value (Figure 1.4B). The final contig(s) for the clones, which are
processed by both assemblers, is chosen based on the best coverage in length and
the identity shown vis-à-vis the reference.

VALIDATION AND REPORTS

The consensus sequence generated in the assembly process is analyzed not only
for the integrity of the insert but also of the flanking attL sites up to BsrGI sites
(5’ TGTACA 3’ sites at 651 on the forward strand and at 2903 on the reverse strand —
Figure 1.5.) Following sequence validation by BLASTN and BLASTX analysis
against reference nucleotide and protein sequences, respectively, the clones are
classified into valid and invalid categories as defined in Table 1.3. The details of the
final validation data are presented in two reports. One of them, clone_distribution_ report.html (Figure 1.6, partly shown), shows the details of validation for each clone
including the sequences for the cloned insert and the reference ORF — via hyperlinks
shown in the last column. The Class and Mutations fields in each case are hyperlinked.
Functional Protein Microarrays in Drug Discovery

FIGURE 1.7 Clone alignment and summary of mutations. (A) Nucleotide and protein alignments (top and bottom panels) and a summary of mutation(s) (middle panel) are shown. See text for the details. (B) A partial screenshot of clone_distribution_report_C_class_mutations.html. Various categories of C class clones, grouped based on the number and types of mutations at the protein level, are shown (links at the top and the details of a link at the bottom are shown as an example). In addition, clones categorized on the basis of whether the mutations occur in the CDS or the flanking "att" sites, or both at the nucleotide level are shown in the middle (only links shown).

to the files showing nucleotide and protein alignments with the reference sequence (Figure 1.7A, top and bottom panels, respectively) and the summary of mutations, if any (Figure 1.7A, middle panel). Positions of mutations and the associated consensus base call quality values, calculated using the procedure described by Churchill and Waterman, are displayed below each nucleotide alignment. In addition, a quality class is assigned to each mutation suggesting the level of confidence in the base call (Figure 1.7B, middle panel). A second report (clone_distribution_report_C_class_mutations.html) shows the further sub-division of C class clones based on whether the mutations occur within the coding DNA sequence (CDS) or the flanking 'att' sequences and whether they represent silent or missense or nonsense mutations at the protein level.

The first pass attempt results in a number of outcomes ranging from perfectly validated clones (classes A and B, see above) to assemblies with only partial coverage. In cases where the first pass validation attempts indicate that a clone has more than 2 nucleotide substitutions relative to the reference sequence, or no cloned insert, the second colony, held as a glycerol, stock is used for template production and sequence validation. Upon generating comparable sequence data from the second colony and the validation process, additional clones which pass the acceptable
criteria are identified and consolidated with those selected from the first colony. At this juncture, depending on the nature and number of remaining invalid clones, further sequencing on both first and the second colony templates are performed in an effort to close remaining gaps and resolve any sequence ambiguities in the assembled sequence.

Following the final consolidation of the validated clone set, reports with various details on the validation are generated. The list of acceptable ORFs from these reports is used to direct the robotic compression of the two freezer copies (colony 1 and 2), into a final set that is replicated into several glycerol stock copies (5 to 10). Since accurate storage and retrieval of samples is essential to a facility managing the distribution of thousands of clones, PFGRC has acquired and installed a Biophile Storage and Individual Vial Retrieval System (Biophile, TekCel) for this purpose. The system consists of five −80°C storage units (BSU) and a −40°C individual vial retrieval unit (IVR). The system is integrated into a relational database, utilizing bar code information to identify each clone and its location in storage. The IVR system automatically records, stores, and retrieves each requested vial based on a prepared worksheet. The “rearraying” of individual vials into new sets allows accurate retrieval of clones or clone sets.

DESTINATION VECTOR CLONING

The ability to automate a large portion of the Gateway Clone Resource production pipeline accounts for the high-throughput capabilities that the PFGRC now offers. The ease of use of the Gateway technology makes possible the construction and validation of 10,000 or more expression clones annually. The scale up of the procedure is largely dependent on the acquisition of additional robots. The ability to transfer cloned and validated entry clone inserts into Gateway expression vectors is straightforward. Purified entry clones and destination vectors are mixed and recombination occurs faithfully via an LR clonase reaction. The screening of recombinant clones and validation of their quality can be performed in a streamlined manner as well, since there is no need to repeat the sequence validation, the only important check being to establish that the complete ORF is present in the expression vector. In general, a single colony can be selected for insert validation. Direct PCR from selected colonies using forward and reverse Gateway® primers allow a rapid and cost-effective method for qualifying the expression clones.

TECHNOLOGY AND ROBOTICS

The PFGRC utilizes two versions of the Beckman, BioMek FX platform to automate most steps in the process, including PCR reaction setup, PCR product purification, cloning reaction setup, transformation of chemically competent cells, plating of transformed cells, plasmid isolation, setup of sequencing plate format, and replication of clone stock copies for distribution. The FX-96 platform transfers equal volumes of
96 samples in parallel, and is used for processes that have pre-equalized concentrations of all reagents, such as PCR reaction setup, and clone stock replication. The FX-Span8 platform transfers samples individually with one of eight pipetting tips with independent volume and sample well location control. Additionally the range of movement and software flexibility allows the deposition of transformation reactions onto oversized 48-well, sectored agar culture plates for colony isolation. Together these two instruments have provided the necessary flexibility and accuracy to make high-throughput Gateway cloning efficient and reliable. One adaptation to reaction setup necessitated by the use of robotics is to ensure that each individual pipetting step delivers 2 µl or more, since overall pipetting accuracy below these volumes is lower.

Accurate qualitative and quantitative assessment of PCR amplification products is an essential part of the high-throughput application of Gateway technology. PCR products must be screened for the presence of multiple bands, incorrect size products, and failed reactions. The precise size (bp) and concentration of successfully amplified products must be known to ensure that optimal recombinational cloning efficiency is achieved. The PFGRC utilizes the Caliper ASM 90 SE Capillary Electrophoresis platform for this purpose. Other similar technologies offered by Agilent Inc and others, perform in a similar way. The Caliper instrument has the ability to accomplish these tasks in parallel with a high degree of accuracy and walk away automation. The system performs electrophoresis in a single gel filled micro capillary channel with high resolution and processing speed (100 to 5000 bp, 30 seconds per sample), allowing the characterization of a 384-well microtiter plate in approximately three hours. The PCR products are detected using a fluorescent dye that provides high sensitivity detection of secondary products and smears that could go unnoticed using traditional agarose or polyacrylamide slab gels. Sizing of the detected PCR products is automated and accurate within ±5%. The concentration of any bands detected is also calculated automatically based on a standardization sample. The area under the peak is calculated for product bands and compared to the standard. This method provides results that are more accurate than traditional absorbance readings taken at 260 nm as it relies on an intercalating fluorescent dye rather than absorbance that can be skewed by multiple factors. The output from the system is then transferred to liquid handling robots for subsequent automated reaction set up.

To achieve a high level of success in a high-throughput endeavor such as the Gateway clone validation pipeline, tracking various laboratory and data processing steps in a systematic way is very critical. A software system like LIMS (Laboratory Information Management System) or a similar resource will be very helpful in that effort and can aid in the creation of high-quality Gateway clones in the following ways: (a) by capturing measurement data, a LIMS can ensure that the correct values are used for PCR evaluation and calculations, (b) by automatically generating robot rearray scripts, a LIMS can prevent plate-to-plate transfer errors, as well as speed up lab processing, and (c) by providing analysis and reporting tools, it can provide valuable metrics that allow lab personnel to evaluate the quality of their techniques over time to improve them.
METHODS AND MATERIALS

PCR AMPLIFICATION OF ORFS

Forward and reverse primers are designed to amplify each ORF from a reference genome sequence. Each oligonucleotide sequence is then appended 3’ of the attB1 (forward) and attB2 (reverse) sequence.

Forward Primer:
5‘ GGGG ACAAGTTTGTACAAAAAAGCAGGCTTC (N18-25) Gene Specific Seq 3’

Reverse Primer:
5‘ GGGG ACCACTTTGTACAAAAAGCTGGGTC (N18-25) Gene Specific Seq 3’

Forward and reverse primer pairs (Illumina/Invitrogen, Carlsbad, CA) at a concentration of 25 µM are combined into a master mix containing 0.15 µM of each dNTP, reaction buffer and 40 ng of genomic DNA with total reaction volume of 35 µl. Typical cycling conditions after a 1 minute initial denaturation at 98°C are as follows: 98°C for 10 seconds, 55°C for 30 seconds and 72°C for 1 minute per kb intended product size. Reactions are cycled through these temperatures 25 times followed by a 72°C final extension of 10 minutes. After cycling, the PCR products are transferred to 384-well Millipore filter plates (Millipore, Billerica, MA) using a Beckman Coulter (Beckman Coulter, Fullerton, CA) Biomek FX 96 probe liquid handling robot. Filter plates are then subjected to a vacuum of 10 inches of Hg for approximately 10 minutes. Then, 50 µl of Milli-Q water is added to each filter plate, PCR products are eluted by aspiration and then transferred to a clean, 384-well MJ Research hardshell plate (Bio-Rad Waltham, MA) using a Beckman Coulter Biomek FX 96 probe liquid handling robot.

PCR PRODUCT VERIFICATION AND QUANTITATION

Each PCR product is analyzed by capillary electrophoresis on the Caliper Life-Sciences (Hopkinton, MA) AMS 90 SE Instrument using LabChip HT 2.4.1 software. A 384-well or 96-well plate containing PCR sample volumes no less than 25 µl is placed in the instrument. Two trays containing ladder and buffer respectively are equipped alongside a “Caliper Chip” which must be cleaned, primed, and prepared with gel dye and marker. Before beginning a run, an input file, containing only ORF IDs and ORF lengths are loaded into the computer along with user input of the plate type and allowed percent deviation from actual ORF length. This information will be used to later calculate the concentration of the fragments found. A 96-well plate will take about an hour to resolve; consequently a 384-well plate will resolve within 4 hours.

After finishing the run, two files are generated and saved. One output file contains the actual electronic gel images and the second output file (of entirely text format) contains the summary data obtained from every well including concentration, size, and if the original fragment was found. This second output file is converted by virtue...
of a simple script into a .csv file and is directly imported into the Biomek FX Span-8 liquid handling robot for automated set up of BP cloning reactions using equimolar quantities (50 fmol) of target vector and PCR product insert.

**BP Clonase Reactions**

BP cloning reactions are performed in 96-well, MJ Research (Bio-Rad Waltham, MA) plates and are conducted in a 15 µl total volume containing: 50 fmol entry clone vector, pDONR221, 50 fmol PCR product, 2 µl of proprietary BP clonase enzyme, (Invitrogen, Carlsbad, CA) 3 µl of 5x BP clonase buffer (Invitrogen, Carlsbad, CA) and brought to volume with 1 × TE. Reaction plates are then incubated for 16 hours at 25°C in a thermal cycler, followed by 4°C hold until recovered. BP clonase reactions are terminated through the addition of 2 µg of proteinase K (Invitrogen, Carlsbad, CA) for 20 minutes at 37°C.

**DH10B-T1 E. coli Transformation**

Chemically competent, DH10B-T1 E. coli cells in 96-well format (Invitrogen, Carlsbad, CA) are thawed on ice and 2 µl of the BP cloning reaction are added using the Biomek FX 96 probe instrument. The plates are sealed with sterile covers and incubated on ice for 30 min. The plates are transferred to a thermal cycler, prewarmed to 45°C. The plates are held for 30 seconds and immediately transferred to ice for 2 min. Cells are allowed to recover by adding 40 µl of SOC media and incubating at 37°C without shaking for 1 hour. Qtray bioassay trays (Genetix Limited, U.K.) with 48 divided areas containing LB media supplemented with 50 µg/ml kanamycin and 2% agar are warmed to room temperature. Several 3 mm glass beads are added to each well and 30 µl of cells are then pipetted onto the agar surface. The Qtrays are shaken gently until all visible liquid has been absorbed into the plates. The beads are discarded and the plates are incubated at 37°C for 16 to 18 hours. Transformation efficiencies are scored by colony count estimations (1–10, 10–50, >50) for each transformation. The Qtrays are held at 4°C.

**Clone Sequence Validation**

Colonies are picked with sterile toothpicks into 1250 µl of 2x YT media, supplemented with kanamycin 50 µg/ml, in 96 deep well blocks. The blocks are sealed with an airpore tape pad strip and incubated at 37°C for 17 hours (11 hours static and 6 hours shaking at 800 rpm). These inoculations are performed in duplicate, one being specified for stock generation. Freezer copies of clone sets are generated in 96-well Matrix Track Mate 2-D bar-coded vials (Matrix Technologies Hudson, NH) by combining 50 µl of overnight culture to an equal volume of 75% glycerol using the Biomek FX 96 probe liquid handling robot.

**Plasmid Extraction**

Plasmid DNA is purified using what is essentially the Qiagen (Qiagen, Valencia, CA) R.E.A.L preparation method using Qiagen’s QIAfilterTM and appropriate buffers.
High-Throughput Gene Cloning Using the Gateway® Technology

E. coli overnight cultures are collected by centrifugation of deep-well blocks at 3200 rpm for 15 minutes at 4°C. The growth medium is decanted and the pellets are resuspended in 300 µl of R1 buffer (50 mM Tris pH 8) containing a final RNase concentration of 100 ng/ml. The cells are lysed by the addition of 300 µl of R2 buffer (1% [w/v] SDS, 200 mM NaOH) with gentle mixing. The lysates are incubated at room temperature for 5 min. The lysates are neutralized by the addition of 300 µl of R3 buffer (3 M KOAc, pH 5.5) followed by mixing and incubation on ice for 10 min. The Biomek FX 96 probe instrument is used to transfer lysates into QIAfilter TM filter plates. The lysates are cleared by vacuum filtration. The plasmid DNA is precipitated through the addition of 625 µl of isopropanol. After mixing, the plates are spun at 3200 rpm for 30 min. at 4°C. The supernatants are decanted and the pellets are washed with 300 µl 70% (v/v) ethanol (–20°C). The plates are spun at 3200 rpm for 15 minutes. The supernatants are decanted and allowed to air dry to completion. The plasmids are resuspended in 50 µl of Blue Tris dye (1 mM Tris pH 8.0, bromophenol blue 1.25 mg/ml) by shaking for 30 minutes on a platform shaker.

SEQUENCING TEMPLATE PRODUCTION BY TEMPLIPHI

Sequencing templates generated by TempliPhi (Amersham Biosciences, U.K.) uses the Phi29 DNA polymerase and rolling cycle amplification to generate linear concatenated copies of plasmid templates. The Biomek 96 probe instrument is used to transfer 10 µl of overnight culture into 384-well plates. The cells are collected by centrifugation at 3200 rpm for 5 minutes. The media is decanted by inverting the plates on absorbent material followed by low-speed centrifugation at 500 rpm for 2 to 3 sec. The Biomek Span-8 is used to add 2 µl of lysis buffer to cell pellets/well. The plates are then placed in a thermal cycler and incubated at 93°C for 3 min. The plates are returned to the Biomek and 34 µl of Milli-Q H₂O is added to each well. The plates are then sealed and spun at 3200 rpm for 5 min. Two microliters of the supernatant are transferred to a clean 384-well MJ Research plate (Bio-Rad Waltham, MA). The enzyme (4 µl) is added to the supernatants and the plates are sealed and incubated at 30°C for 16 hours. The reactions are stopped by heat treatment at 96°C for 5 min. The Biomek FX 96 probe liquid handling robot then adds 34 µl of Milli-Q H₂O Blue Tris dye to each well.

REFERENCES

References

Introduction


Chapter 2. Protein Expression for MicroArrays


17. Kapust, R.B. and Waugh, D.S., Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused, Protein Sci., 8, 1668, 1999.


23. Cornvik, T. et al., Colony filtration blot: A new screening method for soluble protein expression in


36. St. John, T.P. and Davis, R.W., The organization and transcription of the galactose gene cluster of


49. Long, S.B., Campbell, E.B., and Mackinnon, R., Crystal
structure of a mammalian voltage-dependent Shaker family \( K^+ \) channel, Science, 309, 897, 2005.


61. Williams, G.V. and Faulkner, P., Cytological changes


71. Hollister, J.R., Shaper, J.H., and Jarvis, D.L., Stable expression of mammalian β-1,4galactosyltransferase extends the N-glycosylation pathway in insect cells, Glycobiology,


Chapter 3. Emerging Trend: Cell-Free Protein Expression


12. Shimizu, Y., Kanamori, T., and Ueda, T., Protein synthesis by pure translation systems, Methods, 36, 299,


24. Kang, S.H. et al., Cell-free production of aggregation-prone proteins in soluble and active forms,


44. Tan, Z. et al., De novo genetic codes and pure translation display, Methods, 36, 279, 2005.


Chapter 4. The Critical Role of Surface Chemistry in Protein Microarrays


Chapter 5. Fabrication of Sol-Gel-Derived Protein Microarrays for Diagnostics and Screening


25. Braun, S. et al., Biochemically active sol-gel-glasses:


49. Yamanaka, S.A. et al., Nicotinamide adenine dinucleotide phosphate fluorescence and absorption monitoring of enzymic activity in silicate sol-gels for...


60. Lee, M.-Y. et al., Metabolizing enzyme toxicology assay


6 Chapter 6. Printing and QC of Functional Protein Microarrays


Chapter 7. Protein Engineering for Surface Attachment


10. Choi, Y.S., Pack, S.P., and Yoo, Y.J., Development of a


20. (a) Tolbert, T. and Wong, C.H., Intein mediated synthesis of proteins containing carbohydrates and other molecular probes, J. Am. Chem. Soc., 122, 5421, 2000; (b) Yeo, S.Y.D. et al., Cell-permeable small molecule probes


30. Weng, S. et al., Generating addressable protein microarrays with PROfusion covalent mRNA-protein fusion


9 Chapter 9. Fluorescent Detection
Methods for Protein Microarrays

1. Dittrich, P.S. and Manz, A., Single-molecule
fluorescence detection in microfluidic channels — The holy

2. Schena, M. and Davis, R.W., Genes, genomes, and chips,
in DNA Microarrays: A Practical Approach, Schena, M., Ed.,

3. Schermer, M.J., Confocal scanning microscopy in
microarray detection, in DNA Microarrays: A Practical

4. Cutler, P., Protein arrays: The current

5. Haab, B.B., Methods and applications of antibody

chip development in terms of fabrication and application,

7. Sun, C. et al., Advances in the study of luminescence

8. Ekins, R.P. and Chu, F.W., Multianalyte microspot
immunoassay — microanalytical “compact disk” of the future,

9. Zhu, H. et al., Global analysis of protein activities

10. Telford, W.G., Hawley, T.S., and Hawley, R.G., Analysis
of violet-excited fluorochromes by flow cytometry using a

11. Gosalia, D.N. et al., Profiling serine protease
substrate specificity with solution phase fluorogenic
peptide microarrays, Proteomics, 5, 1292, 2005.

12. Berlier, J.E. et al., Quantitative comparison of
long-wavelength Alexa Fluor dyes to Cy dyes: Fluorescence
of the dyes and their bioconjugates, J. Histochem.


49. Waugh, D.S., Making the most of affinity tags, Trends


62. Martin, K. et al., Strategies and solid-phase formats for the analysis of protein and peptide phosphorylation


74. MacBeath, G. and Schreiber, S., Printing proteins as microarrays for high-throughput function determination,


97. Hueber, W. et al., Antigen microarray profiling of


Chapter 10. Functional Analysis of Protein Interactions Using Surface Plasmon Resonance-Based Microarrays


Chapter 11. Leaving the Surface Behind: At the Intersection of Protein Microarrays and Mass Spectrometry


25. Ogorzalek-Loo, R. et al., Mass spectrometry of proteins

26. Gross, J. and Strupat, K., Matrix assisted laser
desorption/ionization-mass spectrometry applied to
biological macromolecules, Trends Anal. Chem, 17, 470,
1998.

27. Katayama, H., Nagasu, T., and Oda, Y., Improvement of
in-gel digestion protocol for peptide mass fingerprinting
by matrix assisted laser desorption/ionization
time-of-flight mass spectrometry, Rapid Comm. Mass

28. Shevchenko, A. et al., Charting the proteomes of
organisms with unsequenced genomes by
MALDI-quadrupole-time-of-flight mass spectrometry and BLAST

29. Bienvenut, W.V. et al., Toward a clinical molecular
scanner for proteome research: Parallel protein chemical
processing before and during western blot, Anal. Chem., 71,
4800, 1999.

in surface-enhanced laser desorption/ionization-time of

glass slides for matrix assisted laser
desorption/ionization analysis: Application to

32. Bundy, J.L. and Fenselau, C., Lectin and carbohydrate
affinity capture surfaces for mass spectrometric analysis

33. Fung, E.T. et al., Protein chips for differential

34. Nedelkov, D. and Nelson, R. W., Analysis of native
proteins from biological fluids by biomolecular interaction
analysis/mass spectrometry (BIA/MS): Exploring the limit of
detection, identification of nonspecific binding and
detection of multiprotein complexes, Biosensors &

35. Lundquist, M. et al., Descriimation of Francisella
tularensis subspecies using surface enhanced laser
desorption ionization mass spectrometry and multivariate


Chapter 12. High-Resolution Label-Free Detection Applied to Protein Microarray Research


Chapter 13. Studying Protein–Protein Interactions with Protein Microarrays: Rapid Identification of 14-3-3 Protein Binding Partners


12. Kawamoto, Y. et al., 14-3-3 proteins in Lewy bodies in


18. Toyo-oka, K. et al., 14-3-3ɛ is important for neuronal migration by binding to NUDEL: A molecular explanation for Miller-Dieker syndrome, Nature Genet., 34, 274, 2003.


23. Tzivion, G. and Avruch, J., 14-3-3 proteins: Active


47. Michaud, G.A. et al., Analyzing antibody specificity with whole proteome microarrays, Nature Biotechnol., 21,


57. Ichimura, T. et al., 14-3-3 proteins modulate the expression of epithelial Na+ channels by phosphorylation-dependent interaction with Nedd4-2 ubiquitin ligase, J. Biol. Chem., 280, 13187, 2005.


63. Lurquin, C. et al., Two members of the human MAGEB gene family located in Xp21.3 are expressed in tumors of various histological origins, Genomics, 46, 397, 1997.


80. Chan, T.A. et al., 14-3-3σ is required to prevent mitotic catastrophe after DNA damage, Nature, 401, 616,


93. Zheng, G. et al., Multiplexed electrical detection of
14 Chapter 14. A Combined Force of Chemical Genetics and Protein Microarrays


15 Chapter 15. Antibody Profiling for Protein Drug Development and Clinical Development


24. Wadman, M., London's disastrous drug trial has serious


49. Ghetie, V. and Ward, E.S., Multiple roles for the major


60. Ferry, H. et al., B-cell tolerance, Transplantation, 81, 300, 2006.


75. Spellerberg, M.B. et al., Dual recognition of lipid A and DNA by human antibodies encoded by the VH4-21 gene: A
possible link between infection and lupus, Hum. Antibodies Hybridomas, 6, 52, 1995.


86. de Bruijn, I.A. et al., Clinical experience with inactivated, virosomal influenza vaccine, Vaccine, 23 (Suppl. 1), S39, 2005.

87. De Serres, G. et al., Oculo-respiratory syndrome after


89. Chiavaroli, C. and Moore, A., An hypothesis to link the opposing immunological effects induced by the bacterial lysate OM-89 in urinary tract infection and rheumatoid arthritis, BioDrugs, 20, 141, 2006.


93. Perosa, F. et al., Generation of biologically active linear and cyclic peptides has revealed a unique fine specificity of rituximab and its possible cross-reactivity with acid sphingomyelinase-like phosphodiesterase 3b precursor, Blood, 107, 1070, 2006.


95. Polyak, M.J. and Deans, J.P., Alanine-170 and proline-172 are critical determinants for extracellular CD20 epitopes; heterogeneity in the fine specificity of CD20 monoclonal antibodies is defined by additional requirements imposed by both amino acid sequence and quaternary structure, Blood, 99, 3256, 2002.


110. Caspi, D. et al., Synovial fluid levels of anti-cyclic


120. Ganor, Y. et al., Antibodies to glutamate receptor subtype 3 (GluR3) are found in some patients suffering from epilepsy as the main disease, but not in patients whose epilepsy accompanies antiphospholipid syndrome or Sneddon's syndrome, Autoimmunity, 38, 417, 2005.

121. Ganor, Y. et al., Autoimmune epilepsy: Distinct
subpopulations of epilepsy patients harbor serum autoantibodies to either glutamate/AMPA receptor GluR3, glutamate/NMDA receptor subunit NR2A or double-stranded DNA, Epilepsy Res., 65, 11, 2005.


143. Li, N. et al., Complete FcRn dependence for intravenous Ig therapy in autoimmune skin blistering


152. Huizinga, T.W. et al., Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins, Arthritis Rheum., 52, 3433, 2005.


13. de Wildt, R.M. et al., Antibody arrays for


Chapter 17. DNA Interactions with Arrayed Proteins


10. Xuan, Z. et al., Genome-wide promoter extraction and analysis in human, mouse, and rat, Genome Biol., 6, R72, 2005.


27. Bulyk, M.L. et al., Quantifying DNA-protein


44. Morin, A. et al., Hyperthermophilic Thermotoga arginine repressor binding to fulllength cognate and heterologous arginine operators and to half-site targets, J. Mol. Biol., 332, 537, 2003.


50. Lebon, G. et al., unpublished data.


26. Hovis, J.S. and Boxer, S.G., Patterning barriers to lateral diffusion in supported lipid bilayer membranes by


19 Chapter 19. Kinase Substrate Identification Using Yeast Protein Microarrays


Chapter 21. The Analysis of Protein Arrays


10. Yang, Y. et al., Development of a toxicogenomics in vitro assay for the efficient characterization of compounds, Pharmacogenomics, 7, 177, 2006.


12. Berkvens, D. et al., Estimating disease prevalence in a bayesian framework using probabilistic constraints,


23 Chapter 23. Visualization of Protein Microarray Data


