Model Organisms in Drug Discovery

Edited by

Pamela M. Carroll and **Kevin Fitzgerald**

Applied Genomics
Pharmaceutical Research Institute
Bristol-Myers Squibb
Princeton, New Jersey
USA

John Wiley & Sons, Ltd

Model Organisms in Drug Discovery

Model Organisms in Drug Discovery

Edited by

Pamela M. Carroll and **Kevin Fitzgerald**

Applied Genomics
Pharmaceutical Research Institute
Bristol-Myers Squibb
Princeton, New Jersey
USA

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

E-mail (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Cover photograph by Charlotte Raymond of Bristol-Myers Squibb, taken at Exelixis, Inc

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Model organisms in drug discovery / edited by Pamela M. Carroll and Kevin Fitzgerald.
p. cm.
Includes bibliographical references and index.
ISBN 0-470-84893-6 (Cloth : alk. paper)
1. Pharmacogenomics. 2. Pharmacogenetics -- Animal models. 3. Drugs -- Research. I. Carroll, Pamela M. II. Fitzgerald, Kevin J., 1967-
RM301.3.G45M635 2003
615′.19 -- dc21

2003013114

British Library Cataloguing in Publication Data

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

ISBN 0 470 84893 6

Typeset by Dobbie Typesetting Ltd, Tavistock, Devon
Printed and bound in Great Britain by TJ International, Padstow, Cornwall
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production..

In loving memory of
Constance Fitzgerald
and
James J. Carroll

Contents

Model Organisms in Drug Discovery. Edited by Pamela M. Carroll and Kevin Fitzgerald.
 ISBN 0 470 84893 6

List of Contributors

Hector Beltrandelrio Lexicon Genetics Incorporated, 8800 Technology Forest Place, The Woodlands, TX 77381, USA

Thierry Bogaert Devgen NV, Technologiepark 9, B-9052 Ghent-Zwijnaarde, Belgium

Lynn Butler Devgen NV, Technologiepark 9, B-9052, Ghent-Zwijnaarde, Belgium

Pamela M. Carroll Department of Applied Genomics, Pharmaceutical Research Institute, Bristol-Myers Squibb Company, PO Box 5400, Princeton, NJ 08543, USA

Neil Dear Ingenium Pharmaceuticals AG, Fraunhoferstrasse 13, D-82152 Martinsried, Germany

Steven A. Farber Department of Microbiology & Immunology, Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA, USA

Kevin Fitzgerald Department of Applied Genomics, Pharmaceutical Research Institute, Bristol-Myers Squibb Company, PO Box 5400, Princeton, NJ 08543, USA

Barbara Froesch The Genetics Company, Inc., Wagistrasse 27, CH-8952 Zürich-Schlieren, Switzerland

Dan Garza Department of Functional Genomics, Novartis Pharmaceuticals Corporation, 556 Morrive Avenue, SEF2026, Summit, NJ 07901, USA

Johannes Grosse Ingenium Pharmaceuticals AG, Fraunhoferstrasse 13, D-82152 Martinsried, Germany

Ernst Hafen The Genetics Company, Inc., Wagistrasse 27, CH-8952 Zürich-Schlieren, Switzerland

Shiu-Ying Ho Department of Microbiology & Immunology, Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA, USA

Model Organisms in Drug Discovery. Edited by Pamela M. Carroll and Kevin Fitzgerald.
 ISBN 0 470 84893 6

Titus Kaletta Devgen NV, Technologiepark 9, B-9052 Ghent-Zwijnaarde, Belgium

Francis Kern Lexicon Genetics Incorporated, 8800 Technology Forest Place, The Woodlands, TX 77381, USA

Rachel Kindt Exelixis, Inc., 170 Harbor Way, South San Francisco, CA 94083-0511, USA

Thomas Lanthorn Lexicon Genetics Incorporated, 8800 Technology Forest Place, The Woodlands, TX 77381, USA

Hao Li Department of Functional Genomics, Novartis Pharmaceuticals Corporation, 556 Morrive Avenue, SEF2026, Summit, NJ 07901, USA

Geert Mudde Ingenium Pharmaceuticals AG, Fraunhoferstrasse 13, D-82152 Martinsried, Germany

Michael Nehls Ingenium Pharmaceuticals AG, Fraunhoferstrasse 13, D-82152 Martinsried, Germany

Tamas Oravecz Lexicon Genetics Incorporated, 8800 Technology Forest Place, The Woodlands, TX 77381, USA

Michael Pack Departments of Medicine and Cell and Developmental Biology, University of Pennsylvania School of Medicine, Room 1212 BRB 2/3, 421 Curie Blvd., Philadelphia, PA 19104, USA

James Piggott Lexicon Genetics Incorporated, 8800 Technology Forest Place, The Woodlands, TX 77381, USA

David Powell Lexicon Genetics Incorporated, 8800 Technology Forest Place, The Woodlands, TX 77381, USA

Ramiro Ramirez-Solis Lexicon Genetics Incorporated, 8800 Technology Forest Place, The Woodlands, TX 77381, USA

Petra Ross-Macdonald Department of Applied Genomics, Pharmaceutical Research Institute, Bristol-Myers Squibb Company, PO Box 5400, Princeton, NJ 08543, USA

Andreas Russ Ingenium Pharmaceuticals AG, Fraunhoferstrasse 13, D-82152 Martinsried, Germany

Arthur T. Sands Lexicon Genetics Incorporated, 8800 Technology Forest Place, The Woodlands, TX 77381, USA

Andreas Schröder Ingenium Pharmaceuticals AG, Fraunhoferstrasse 13, D-82152 Martinsried, Germany

Stefan Schulte-Merker Exelixis, Deutschland GmbH, Spemannstrasse 35, D-72076, Tübingen, Germany

Corina Schütt The Genetics Company, Inc., Wagistrasse 27, CH-8952 Zürich-Schlieren, Switzerland

Reinhard Sedlmeier Ingenium Pharmaceuticals AG, Fraunhoferstrasse 13, D-82152 Martinsried, Germany

Gabriele Stumm Ingenium Pharmaceuticals AG, Fraunhoferstrasse 13, D-82152 Martinsried, Germany

Sigrid Wattler Ingenium Pharmaceuticals AG, Fraunhoferstrasse 13, D-82152 Martinsried, Germany

Brian Zambrowicz Lexicon Genetics Incorporated, 8800 Technology Forest Place, The Woodlands, TX 77381, USA

Acknowledgments

It has been an exciting time to translate model systems and the new age of genomics into relevant technologies for drug discovery. We thank all of the excellent authors for their creative contributions. We also thank our colleagues at Bristol-Myers Squibb for their continued support of our work and for allowing us to pursue model organism approaches. In particular, we would like to acknowledge the visionary leadership of Drs Elliott Sigal and Mark Cockett at Bristol-Myers Squibb and Drs Geoff Duyk and Greg Plowman at Exelixis. We thank our excellent editor Joan Marsh and editorial assistant Layla Paggetti at John Wiley & Sons, Ltd for their guidance, suggestions, expertise and especially their patience.

Model Organisms in Drug Discovery. Edited by Pamela M. Carroll and Kevin Fitzgerald.
 ISBN 0 470 84893 6

1

Introduction to Model Systems in Drug Discovery

Kevin Fitzgerald and **Pamela M. Carroll**

A major challenge in the 'post-genomic' world is to rapidly uncover the proteins that may become the high-quality therapeutic targets of the future. This book will focus on the utility of model organisms as a systematic approach to a broad array of disease-based questions. The recent publication of the human genome revealed the most complete set of human genes to date, yet most of these genes have not been assigned a biological function and an even smaller number have been linked to a human disease process. Comparative genomic analysis of simple model systems with that of the human has revealed the evolutionary conservation of gene and protein structure as well as 'gene networks'. This evolutionary conservation is now being exploited with model systems as critical 'functional genomics' linchpins, in associating conserved genes with therapeutic utilities. Genes of unknown function can now be studied in the more tractable model systems and inferences can be drawn about their roles in complex biological processes.

1.1 Integrating model organism research with drug discovery

Pharmaceutical drugs in the modern era are something we all take for granted. We swallow a pill if we have a headache and magically the pain abates. Infections that in the past caused limb amputations, paralysis, lung damage or death are treated by antibiotic tablets and the infection and symptoms abate.

Model Organisms in Drug Discovery. Edited by Pamela M. Carroll and Kevin Fitzgerald.
 ISBN 0 470 84893 6

Diseases such as diabetes, AIDS, high blood pressure and cholesterol that often resulted in a host of serious and medical issues are now controlled with medications. Life expectancy has increased and the quality of life in old age continues to improve. Drug discovery and development have a remarkable history of success considering that the quest for new pharmaceuticals traditionally has encompassed searching for a needle in a chaotic and disorganized haystack of complex human biology and disease. It was not until the release of a complete draft of the human genome sequence in 2001 that scientists were provided with a list of all possible drug targets for pharmaceutical intervention. The current and future challenges are to identify those genes implicated in disease and to leverage the genome information into an understanding of complex biological systems, efficiently paving the way for drug discovery.

The genome information provides the rudimentary gene list for all possible drug targets but still leaves scientific research a great distance from understanding the role of each of these protein targets in normal biology and disease processes. Years from now the sequencing of not only the human genome but the genomes of *Saccharomyces cerevisae* (yeast), *Caenorhabditis elegans* (nematode), *Drosophila melanogaster* (fruit fly), *Danio rerio* (zebrafish) and *Mus musculus* (mouse), as well as a large number of unpleasant pathogenic bacteria and viruses, will be looked upon as watershed events in the development of novel medicines. Parallel to the sequencing of the genome are advances in chemistry, engineering, microscopy and genetics that are having a major impact on the drug discovery process. The purpose of this book is to update and forecast how these technological advances are being combined with model organisms in biology to have an impact on modern drug discovery.

A useful analogy of model organism studies is the hobby of constructing 'model' cars or planes. Such model kits arrive with a parts list, a large number of pieces and an assembly manual that describes the function of each part and how the various parts fit together into a three-dimensional working object. Models can be manipulated by removing a part and determining the overall structure and function of the model without that part. The same is true of model organisms in drug discovery. The genome sequences of 'model' systems described in this book are the list of parts. Of course, we are not handed the assembly manual (therein lies both the challenge and the promise) but biologists are arduously writing this very complex manual in small bits at a time. Organisms arrive whole and functioning, and scientists strive to deconstruct the functioning end product into its various parts and then hypothesize about the functions of individual parts and the connections between them. This is actually more akin to someone handing you a functioning F-16 fighter jet along with a parts list and requiring you, without any instruction manual, to assemble a new fighter jet or, in an analogy to a

Table 1.1 Genome comparisons of model organisms

Organism	Transcriptome size	% Genes[1] similar to a human gene	Cellular complexity	Generation time
Yeast	6200 genes	46%	1 cell	2 h
Nematode	18 300 genes	43%	~959 cells	3 days
Drosophila	14 400 genes	61%	$>10^6$ cells	10 days
Zebrafish	30 000–80 000 genes	>80%	$>10^8$	6–8 weeks
Mouse	30 000–80 000 genes	95–97%	$>10^9$ cells	6 weeks

[1]From Lander, E. S., *et al.* (2001) *Nature* **409**, 860–921.

human disease state, to diagnose and fix a malfunctioning jet. The progress in genetic and molecular tools has allowed us to begin the process of deconstructing normal and disease biology, but the process remains daunting and in reality will most likely take decades to complete. Because we cannot dismantle the human organism, we rely upon the fact that biology has evolved in a similar fashion from the single cell yeast to the system complexity of the mouse. We utilize organisms such as *C. elegans* and *Drosophila* because scientists have the tools to deconstruct these organisms and ask questions about the functions of every gene. Scientists can leverage the fact that evolution, for the most part, did not reinvent the same processes many times. For instance, the process by which one cell divides to make a second cell is a conserved function and biological pathway in yeast and humans. Throughout this book you should begin to gain an appreciation for how few biological differences there are between animal models and humans, and how to exploit this similarity to uncover the causes of and find new treatments for human disease.

This book will review the technical and innovative advantages that are specific for each model organism, as well as provide detailed accounts of 'disease models' in simple organisms that have had an impact on the understanding of human biology. The model organisms of focus are yeast, nematodes, fruitflies, zebrafish and mice. Many of these organisms have the advantage of a complete genome sequence and recent sophisticated advances in 'forward' (going from a phenotype *in vivo* to the causative gene mutation) and 'reverse' (going from a gene to the phenotype of a mutation in that gene *in vivo*) genetic tools that allow for genome-wide functional discoveries.

Table 1.1 offers a glance at comparisons of the systems in terms of the number of genes, similarity to humans and life cycle length (personal communication with Ethan Bier). When embarking on research projects it is not always clear which organism to choose for human relevance and speed of discovery. With increasing biological complexity comes greater similarities to humans; therefore, the mouse would be the clear system of choice if it were not

for its long generation time and cumbersome technologies. For example, when carrying out mutation studies, embryonic lethal mutations are often more easily characterized in the zebrafish than the mouse. In the last decade, we have seen experimental models such as *Xenopus laevis* (the frog) lose favor. In the case of *X. laevis* this is due to a large and polyploid genome making genomics and genetic undertakings unreasonable. On the horizon are new model systems that have not entered the subject of this book but may soon be on all our research radar screens. Sometimes a new system needs the commitment of powerful scientists to lead the research community. Would zebrafish have seen the massive worldwide undertaking of genetic screens and technologies without the commitment of *Drosophila* geneticist and Nobel Laureate Christian Nusslein-Volhard? Will Sydney Brenner, the founding father of *C. elegans* as a model organism and Nobel Laureate, leverage his interest in the Japanese pufferfish (*Fugu*) and its complete genome into an important experimental model?

Specific model organisms were chosen as this book's focus because they are widely accepted as valuable experimental models in genomics and genetics. Many biotechnology and pharmaceutical companies have programs centered on model organisms for an array of drug discovery and development platforms. Applications covered herein range from target identification, target validation, compound discovery and toxicology screening. Important models in drug development, such as rat and monkey, were not included largely due to less developed genetic tools. Each model system has a set of unique advantages and disadvantages offered by that particular genetic model. The biological problems that are chosen for study in each system depend on how likely a model system is to yield insights into human biology. For example, zebrafish offers an unparalleled visualization of a multi-organ vertebrate system and many of the organ systems (such as the circulatory system) are good models for human organs, but the technologies available for forward and reverse genetics are still relatively costly and time-consuming. Conversely, yeast offers rapid, efficient genetic approaches, but only about 50% of the gene networks are functionally conserved with humans and they lack the complex nature of human organ tissue systems. *Drosophila* in many cases represents a good 'happy medium' in that they integrate multiple complex organ systems yet have the rapid genetic tools used to deconvolute complex biology.

The chapters of this book are ordered along increases in evolutionary complexity towards humans, starting with yeast, nematodes and fruitflies and then proceeding into chapters centered around zebrafish and mice. One could also view this as a progression of technology development with an abundance of powerful genetic tools available in yeast, fruitflies and nematodes and the quest of zebrafish and mice researchers to develop similar technologies. The book will detail the incorporation of advances in the application of bio-informatics, proteomics, genomics, biochemical and automation technologies

to simple organisms and how these advances constitute an integrated drug discovery platform. Detailed accounts of the application of model organism technology to specific therapeutic areas will be covered. The authors include leading experts in each field who will examine state-of-the-art applications of individual model systems, describe real-life applications of these systems and speculate on the impact of model organisms in the future. The first of these authors will delve into the relatively simple model organism, yeast.

Chapter 2 by Ross-Macdonald of Bristol-Myers Squibb describes the history of *Saccharomyces cerevisae* (yeast) research in drug discovery and how this simple eukaryote historically has been utilized mainly as a production vehicle due to its ability to produce compounds and proteins but also as a valuable tool in understanding biology. Yeast researchers have an unparalleled breadth of reagents to probe the genome, making it a natural choice for studying conserved targets and mechanisms of basic biological processes. With the sequencing of the yeast genome and the advent of such tools as transcriptional profiling, protein–protein interaction assays and genetic tools such as deficiency, overexpression and haploinsufficiency strain sets, yeast is now a workhorse in uncovering hidden links among genes and defining cell signaling circuits. Many of the genomics tools that are being applied to the other model systems were developed in yeast and the yeast model system continues to be an invaluable source of innovation and technology development. For this review, Ross-Macdonald has chosen to highlight the contributions of biotechnology and pharmaceutical researchers in order to focus this broad field.

Caenorhabditis elegans is a tiny worm composed of just around 900 cells and a life cycle of about three days, yet it contains many of the cell types and genes found in humans. It was the first multicellular organism to have its complete genome sequenced. It is in *C. elegans* where we begin to see the development of rudimentary tissues, organs and the beginnings of a more sophisticated nervous system. The level of complexity (complex but not so complex as to have little chance of ever understanding all of the various neuronal connections) is one of the attributes of *C. elegans* that first attracted Sydney Brenner to *C. elegans* as a model system. Research into *C. elegans* has played an essential role in our general understanding of more complex human diseases such as cancer (i.e. Ras oncogene), depression (i.e. neuronal signaling and drug mechanism of action), Alzheimer's disease (i.e. presenilin genes) and cell death. In Chapter 3, Kaletta, Butler and Bogaert from DevGen review the short but impactful career of *C. elegans* in drug discovery. They also take us through the detailed process of applying *C. elegans* technologies of 'high-throughput' target identification and compound screening. Clearly, there is a great future for *C. elegans* in drug discovery.

For nearly 100 years *Drosophila* genetics has been a central contributor of research on inheritance, genome organization and the development of an organism. *Drosophila* represents a 'happy medium' in that terrific genetic tools are available and yet there is a level of complexity to the organism that more closely resembles vertebrates. In *Drosophila* there is the emergence of a complex nervous system and visual and digestive organs. Chapter 4, authored by Li and Garza from Novartis, describes the *Drosophila* technologies that have evolved over this long history, and in Chapter 5 Ernst Hafen and colleagues at the Genetics Company and the University of Zurich show how these technologies have been implemented to decipher several important disease pathways. For example, recent genetic studies have revealed the *Drosophila* insulin-mediated signaling pathway and its astounding similarity to mammals, suggesting that *Drosophila* research deserves a place in the studies of metabolic diseases such as diabetes. Any discussion of drug discovery would be incomplete without a clear discussion of compounds that lie at the very heart of and are the ultimate goal of the process. It is clear that one of the emerging areas of model systems will be 'chemical genetics'.

Chemical genetics consists of combining the genetic tools of model organisms with novel compounds in order to get a better understanding of their mode of action. It also encompasses screening for compounds that interfere with biological processes and then using those compounds as tools, which, when combined with genetics, allow you to unravel pathways of gene interaction. Every chapter of the book touches upon this new emerging field and Chapter 6, authored by the editors and Rachel Kindt at Exelixis, is dedicated to this concept. Perhaps the most striking revelation contained in these pages is that compounds work on conserved targets across species and, although ultimately the compound affinities may differ, the mechanisms of action are similar. Chapter 6 highlights the utility and benefits of having multiple genetic systems to unravel a problem. Examples of relevance in understanding the mode of action of gamma secretase inhibitors in Alzheimer's disease and natural products in inflammation are discussed, and these examples explore the integration of compounds with genetics.

The emerging power of the zebrafish system is captured in Chapter 7 by Schulte-Merker at Exelixis and in Chapter 8 by Ho, Farber and Pack at Thomas Jefferson University and the University of Pennsylvania. Zebrafish are a vertebrate model that develop externally and transparently; thus the formation of many structures and biological processes can be easily monitored. The progress of genome mapping, mutagenesis screens and new 'knock-out' and overexpression technologies will provide significant insights into these biological processes (Chapter 7). Chapter 8 discusses a specific model where zebrafish are being utilized to study lipid metabolism with strong parallels to those found in humans.

Finally, Chapters 9 and 10 explore the advances in one of the workhorses of modern drug discovery, the mouse. Mice have been involved in drug discovery for some time as models of human disease but the adaptation of higher throughput technologies is just beginning to have an impact on the search for novel targets. In addition, the mouse model is coming into its own as a tool to 'de-orphan' the biology of novel targets and allow compounds to be tested in mouse models lacking any gene. In some areas such as neuroscience, a phenotype in a mouse model is the gold standard (besides active compounds or human genetics) that associates a given gene with a disease. The mouse-focused chapters are divided into forward genetic approaches contributed by Ingenium AG (Chapter 9) and the reverse genetics approaches based on work at Lexicon Genetics (Chapter 10). In forward genetics a phenotype is identified first and then the molecular basis of a given trait is identified. Historically, the process of phenotype to mutation has been laborious and time-consuming, but new genomics technology is rendering the process more robust. Chapter 9 reveals new approaches for novel, rapid, chemical genetic screens and mutation identification that allow for *in vivo* target discovery in unprecedented ways. Conversely, Lexicon Genetics (Chapter 10) describes its undertaking of systematic large-scale gene knock-outs of the 'druggable genome' in mice and the process in place to associate a gene's functions with disease. Because most drugs act as antagonists, knock-out phenotypes should mimic drug action.

An exciting paradigm for drug discovery is evolving. The current processes by which drugs are discovered are long and expensive. Many compounds still fall out of the discovery pipeline due to lack of efficacy and mechanism-based toxicity. Central to these reasons is a failure to understand properly all of the biological roles of potential drug targets in normal and disease processes (also referred to as 'target validation'). This knowledge failure results in ignorance of the many potential unpleasant consequences that could be rendered by compound modulation of the target's activity *in vivo*. The integration of model systems into the drug discovery process, the speed of the tools and the amount of *in vivo* validation data that these models can provide will clearly help to define better the disease biology and thereby result in better validated targets. Better targets will lead to high efficacy and less toxic therapeutic compounds. The future will see a merging of the genetics of model systems with proteomics, bioinformatics, structural biology and compound screening, creating the exciting new framework of drug discovery for the 21st century.

2

Growing Yeast for Fun and Profit: Use of *Saccharomyces cerevisiae* as a Model System in Drug Discovery

Petra Ross-Macdonald

Yeast has great utility as a surrogate system to study aspects of mammalian biology. This utility extends to the drug discovery process, where yeast has been used to reveal the mechanism of action of compounds, to discover and characterize components of signaling pathways and to dissect protein function. These applications of yeast are illustrated by examples of research published by major pharmaceutical companies.

2.1 Introduction

This chapter is intended to illustrate the use of yeast (*Saccharomyces cerevisiae*) as a model organism in drug discovery research. Yeast has had a long utility as the workhorse of pharmaceutical discovery research, whether as a representative of its pathogenic cousins or as a living eukaryotic vessel for bringing together reagents such as the two-hybrid system components or carrying reporter constructs for screening. However, I will confine this review to applications where yeast has been used as a true 'model' for vertebrate biology in the area of disease. To demonstrate the value of yeast in applied

Model Organisms in Drug Discovery. Edited by Pamela M. Carroll and Kevin Fitzgerald.
 ISBN 0 470 84893 6

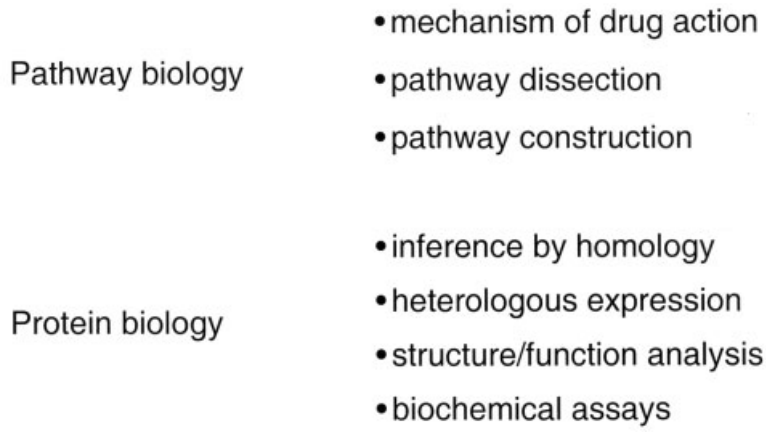

Figure 2.1 Outline of the areas in which yeast has been used as a model system for the biology of higher eukaryotes. Pharmaceutical research in these areas is described in the text

pharmaceutical research, my examples and citations are almost completely culled from publications by research scientists at major pharmaceutical companies (i.e. roughly the top 20 companies based on the market share). This approach results in the omission of many fine pieces of academic work that may have had publication priority, but the aim of this chapter is to demonstrate the type of yeast research that drug discovery organizations historically have regarded as worthwhile, informative and likely to affect their bottom line. Unfortunately this approach also unwittingly leads to the omission of much excellent biomedical research using the fission yeast *Schizosaccharomyces pombe*, because few examples of its application have been published by pharmaceutical companies. The uses of yeast described in this chapter are laid out in Figure 2.1; they include sections on the use of yeast in elucidating pathways and their components, including pathways that are not native to yeast and pathways involved in the mechanism of action of compounds. I will also describe more targeted experiments to characterize the functions of specific proteins. Finally, I will review the 'post-genome' tools, technologies and information resource advances that now enable yeast research.

2.2 *Saccharomyces cerevisiae* and its genome: a brief primer

Commonly known as baker's, brewer's or budding yeast, *S. cerevisiae* has been a standard laboratory microorganism since the 1950s. It has many endearing attributes, including the ability to fill a laboratory with a pleasant 'warm-bread' odor, yet also to survive years of abandonment in a fridge or

freezer or even on a desiccated piece of agar in a forgotten petri dish. (Almost every yeast biologist has had the need to test this last assertion.) It is cheap to feed, non-pathogenic and divides every 2 h. It can grow either aerobically or anaerobically, depending on the nutrients provided, and in solid or liquid media. It can exist stably as a haploid or a diploid, and haploids can be mated and put through meiosis to recover haploid progeny in a matter of days. Although a unicellular organism, it can on occasion display such group characteristics as pseudohyphal growth, intercellular signaling and programmed cell death. Finally, a highly versatile transformation (transfection) system has been available for several decades. You can choose a vector that is linear, circular or integrating, high or low copy number, with a positive or negative selection system, and you can express your favorite gene from several types of regulated promoters. In addition, homologous recombination occurs with high efficiency, allowing the integration of transformed DNA into chromosomes at precise locations, replacing and deleting host DNA as desired.

The *S. cerevisiae* genome sequence was completed and almost entirely annotated for genes in 1996 (Goffeau *et al.*, 1996) but it has not remained static (Kumar *et al.*, 2002c). By comparison to most eukaryotes, coding regions are enviably simple to identify in yeast: about 70% of the genome encodes protein, and only about 4% of yeast genes contain introns, usually as a small insertion very near the 5′ end of the coding region. For expedience, the primary annotaters of the genome set the ability to encode a 100-amino-acid protein as the cutoff for a gene (unless other evidence existed). Each resulting open reading frame (ORF) was given a unique and informative seven-character identifier, e.g. YOR107w. This name immediately tells a yeast biologist that the gene lies on the Watson strand of the right arm of chromosome XV, 107 genes distal from the centromere. Unlike the Dewey decimal system, this left no room for additions; fortunately there have been relatively few subsequent modifications of genes because these have had to be dealt with by inelegant suffixes (A, -B, etc.; inconsistencies in their syntax are a common source of error in data handling). This systematic name complements and conforms to the yeast genetic nomenclature adopted by consensus in the 1960s, in which the upper case notation informs us that a wild-type gene is being discussed, a lower case notation would indicate a mutant and Yor107w is the name of the encoded protein. All yeast genes thus have a systematic ORF name; about half of them also have one or more traditional three-letter gene names that are intended to reflect some property of interest, e.g. *RAD1* to identify the first gene identified from a mutant screen for radiation sensitivity. Yeast biologists have concluded that clarity in the literature is more important than their egos and nowadays they commonly agree on a single, rational primary gene name maintained in a central registry. These names are certainly duller than those for *Drosophila* – yeast never had *ether-a-gogo* but for over a

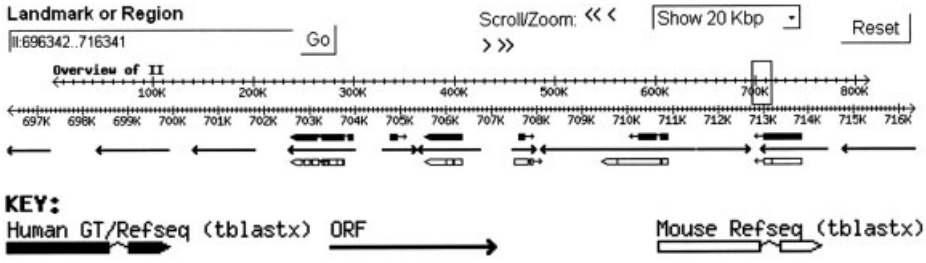

Figure 2.2 A graphical view of a 20-kilobase region of yeast chromosome II, showing 11 open reading frames (ORFs) encoding proteins. The NCBI Reference Sequence project (RefSeq) clones from human and mouse that show significant homology at the protein level are overlaid on their yeast homologs. The view was generated using the browser created by the Generic Model Organism Database Project (www.gmod.org) following customization by Dr N. Siemers

decade it did have *WHI1* (whiskey1, named in a pub in Scotland) until the title's overturn by the more prosaic name *CLN3* (cyclin3).

The number of recognized genes in yeast hovers just above 6000, remaining in flux due to continued research on which of these are spurious and what additions should be made (see Kumar *et al*., 2002c). Figure 2.2 provides a visual snapshot of a region of the yeast genome and illustrates the significant homology between some of the proteins coded therein and proteins from the mouse and human genomes. Unfortunately, no quantitative cross-comparison between yeast and human genomes has been published since 'completion' of the human genome sequence. An analysis performed in 1997 found that about one-third of yeast proteins had significant homology to a mammalian GenBank sequence (Botstein *et al*., 1997); by 1997 the results from Bassett *et al*. (1996) had been updated to suggest the existence of yeast homologs for 34% of the 84 disease-related human genes that were positionally cloned at the time. In 1998 a very stringent comparison between yeast and the newly finished *Caenorhabditis elegans* genome (Chervitz *et al*., 1998) predicted that about 40% of yeast proteins were orthologous to about 20% of those encoded in worm. Many of the remaining 80% of worm proteins contained domains also present in yeast, but their arrangement within proteins was not identical. Because 80% of *C. elegans* proteins apparently lack a close relative in yeast, it might seem that there is a low probability of a given gene from a multicellular organism having a yeast homolog that can be studied productively. However, these numbers are skewed by the 'bulking out' of the *C. elegans* proteome by gene duplication events that lead to huge multigene families such as that for the nuclear hormone receptors. Within core metabolic and structural functions there is virtually complete conservation across eukaryotes. A recent comparison between the predicted proteins of the *S. pombe* and *S. cerevisiae*

genomes and 289 human disease proteins found 182 *S. cerevisiae* proteins with significant similarity with about 50 probable orthologs (Wood *et al.*, 2002). The shared proteins covered a range of human disease areas from neurological to metabolic, the largest group being those implicated in cancer. Also, in many situations where a more intensive analysis has been brought to bear, proteins previously cited as absent from yeast have been found. A recent example is the identification of a caspase-type protein in yeast (Uren *et al.*, 2000) and demonstration of its orthology to metazoan caspases (Madeo *et al.*, 2002).

2.3 Yeast in pathway and mechanism elucidation

Selection of appropriate targets remains a major hurdle in drug discovery. When a biological pathway is of interest for therapeutic intervention, a broad understanding of its components is essential to allow the design of assays that can address both desired and undesired effects of that intervention. Knowledge of pathway biology is at its most advanced in yeast, owing to the ground cleared by decades of academic yeast research. Observations of cell cycle mutants of *S. pombe* and *S. cerevisiae* in the 1970s led directly to identification of the same pathway in humans and to the first generation of cyclin-dependent kinase (CDK) inhibitors currently in the clinic (Senderowicz, 2000). Yet examples of the use of yeast by pharmaceutical companies in further dissection of this pathway are rare, although Novartis has reported a yeast system to screen for Cdk4-specific antagonists (Moorthamer *et al.*, 1998). It seems that translation of observations in yeast to the relevance in mammalian systems and into pharmaceutical application continues to be underutilized. Mammalian biologists often feel that yeast is too simple to be of relevance to the process they study, or they point to incongruities in data to insist that yeast 'does it differently'.

Such reservations are partly justified: there are many examples of mammalian target proteins or drug effector mechanisms that are simply not present in yeast. For example, components of the cholesterol biosynthesis pathway, including the target for basic biochemical inhibitory action of the statin drugs, are largely conserved from yeast to humans. Yeast was used extensively by companies such as Bristol-Myers Squibb (Robinson *et al.*, 1993) and Zeneca (Summers *et al.*, 1993) in the identification and characterization of targets within this pathway. However, statins exert the majority of their cholesterol-lowering effect in humans by a feedback mechanism that leads to upregulation of the hepatic low-density lipoprotein (LDL) receptor, and this protein is not conserved in yeast (although feedback mechanisms responding to lowered sterol level do exist). Yeast also has no nuclear hormone receptors and thus lacks a form of regulation that overlays many conserved metabolic pathways in higher eukaryotes. Conversely,

examples also exist of cases where yeast has proved to contain the target for a drug, even though that drug has its therapeutic effect in a process such as immunity, which has no apparent parallel in yeast. There are also cases where a very clear conservation exists and yet the published work is almost exclusively academic, e.g. the use of yeast in the determination of the mechanism of action of the topoisomerase inhibitors (reviewed by Bjornsti *et al.*, 1994). A search of the literature on camptothecin produces only one example of the use of yeast by industry: Takeda laboratories used *S. pombe* to demonstrate that the mechanism of a novel topoisomerase I inhibitor differs from that of camptothecin (Horiguchi and Tanida, 1995).

2.4 An example of mechanism elucidation: immunosuppressive agents

Three sterling examples of how yeast can contribute to the identification of a drug target and characterization of the responding pathway are provided by the immunosuppressive agents cyclosporin A, FK506 and rapamycin. The story of this research is also the story of what would have been an overwhelmingly difficult mechanism of action study without yeast, because it is a case where compounds interact with structurally unrelated binding partners to affect the same target and, conversely, compounds interact with the same binding partner to affect different targets (see Figure 2.3). The mechanism runs contrary to established wisdom on the feasibility of modulating protein–protein interactions. Finally, the binding partners are not the therapeutic target but, to throw in a couple of red herrings, they do have a common enzymatic activity that is inhibited by the compound! Without academic and industry groups striving neck and neck for the answer, and without yeast to identify additional components and provide genetic dissection and stringent hypothesis-testing, determination of their mechanisms within a decade of research is extremely unlikely to have occurred. Ironically, the ultimate targets are a kinase and a phosphatase, and today no right-thinking pharmaceutical company would put any money into a compound that took such a convoluted path to reach these targets. But these compounds were clinical successes before their mechanisms were established, and their efficacy has yet to be matched by small molecules from a rational development process. Cyclosporin A was identified in the 1970s at Sandoz (now Novartis) and approved for use as a transplant rejection therapeutic in 1983. As an interesting footnote, Novartis's own web page states that the initial observations on the natural product indicated a very weak compound that was regarded as being of little practical value. Fortunately an intellectual curiosity prevailed and allowed work to continue until Dr Jean Francois