

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.

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 cerevisiae* (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
<i>Drosophila</i>	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

¹From 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 bioinformatics, 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 cerevisiae* (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.

