Using model-system genetics for drug-based target discovery

David J. Matthews and Jenny Kopczynski

The combination of medicinal chemistry and model-organism genetics is emerging as a powerful tool for the discovery and validation of drug targets. Model systems can be used to identify the cognate target for compounds that demonstrate *in vivo* efficacy but have unknown mechanisms of action. Alternatively, drugs with known cognate targets can be used to probe biochemical pathways in model organisms, revealing new targets and mechanisms within these pathways. In both cases, the availability of human genomic sequence data is opening up new opportunities for accelerating target discovery.

***David J. Matthews and Jenny Kopczynski**
Exelixis, 170 Harbor Way
South San Francisco
CA 94083 0511, USA
***tel:** +1 650 837 7000
fax: +1 650 837 8300
e-mail: matthews@exelixis.com
jennyk@exelixis.com

▼ One of the major challenges currently facing the drug discovery industry is the rapid identification and validation of new drug targets. It is estimated that current pharmaceutical agents modulate only ~400–500 proteins; by contrast, there might be 5000–10,000 novel genes in the human genome that could provide targets for therapeutic intervention[1]. Despite the imminent availability of the complete human genome sequence, the selection of genes relevant to a given disease state remains problematic. An additional challenge is the ability to assess the 'drugability' of these novel targets – in other words, to determine which targets are amenable to *in vivo* modulation by small molecules.

Conversely, there are many compounds for which a certain degree of specific biological activity has been demonstrated but for which the cognate target and/or relevant signaling pathway is unknown. The identification of these targets and pathways is particularly valuable because they are likely to be therapeutically relevant and amenable to modulation by a small molecule. This review will focus on the use of model-organism genetics to identify the cognate targets and probe the related biochemical pathways of compounds with unknown or ambiguous mechanisms, a method we refer to as the mechanism of action (MOA) approach.

Model-organism genetics – a powerful paradigm

Model-organism genetics has proven to be a valuable method for delineating protein interactions and pathways *in vivo*, as exemplified by the Ras (retrovirus-associated sequences) pathway (Fig. 1). Many years of elegant genetic analysis in the fruitfly *Drosophila melanogaster*[2] and roundworm *Caenorhabditis elegans*[3] have contributed greatly to current understanding of how Ras and associated proteins function in humans. For example, the discovery of Ksr (kinase suppressor of Ras), which was first identified in model organisms, provided a new protein target for potential therapeutic intervention[4–6].

Interfacing genetics with small organic molecules

Using similar strategies, genetic screens can be used to identify proteins that interact directly or indirectly with a drug or another small organic molecule[7]. From the geneticist's point of view, modulating target function by an agonist or antagonist is similar to modulating protein activity by a mutation that increases or decreases protein function. With this principle in mind, the MOA approach first involves administering compounds to a model organism and monitoring their physiological effects by careful examination of the resultant phenotype. If distinctive, the phenotype can provide an early indication of the process disrupted by the compound. Once such a phenotypic assay has been conducted, genes responsible for the compound-dependent phenotype can be identified by creating mutants that are resistant or hypersensitive to the compound. The mutation, or change in

1359-6446/01/$ – see front matter PII: S1359-6446(00)01612-3

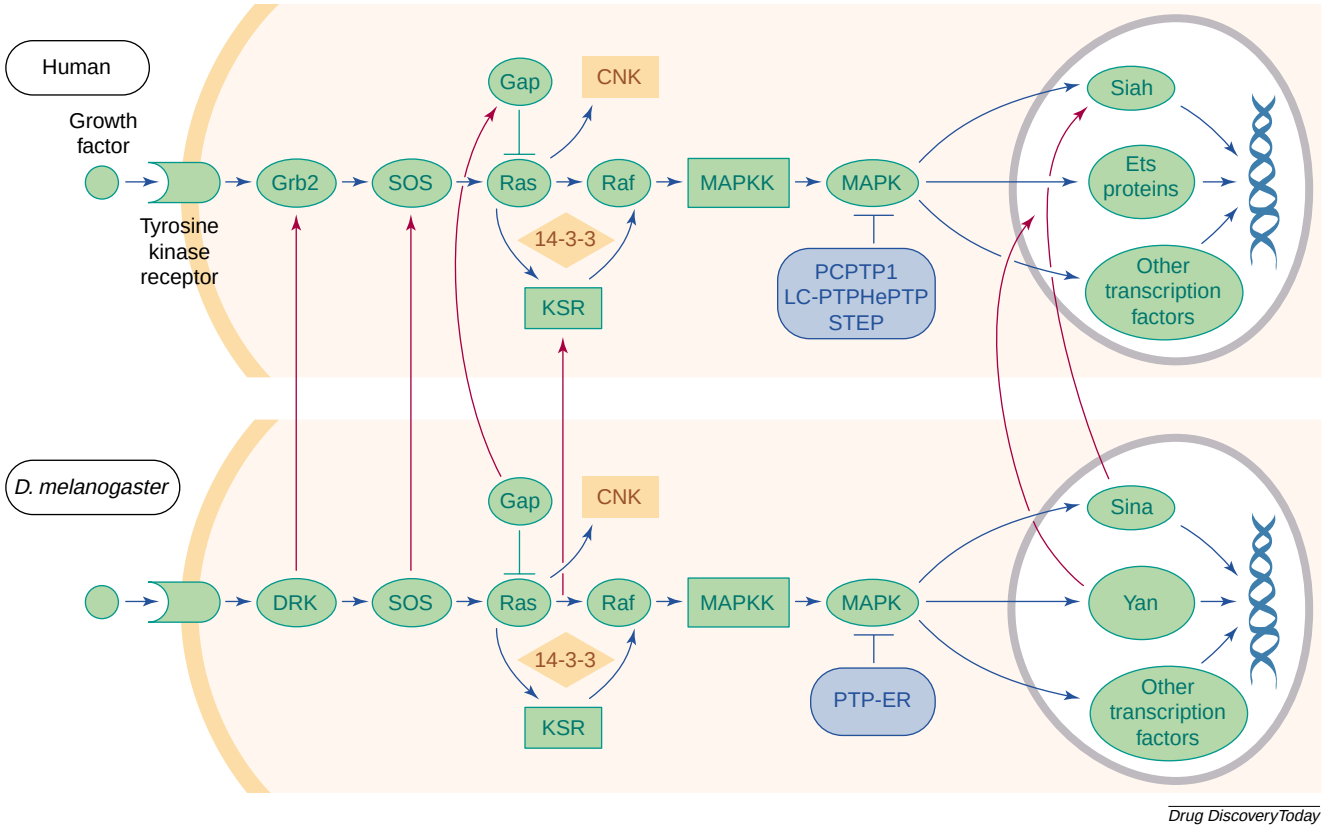

Figure 1. Comparison of the Ras pathways in *Drosophila melanogaster* and humans. Individual human and *D. melanogaster* proteins are related to each other at the level of DNA and protein sequence, as indicated by the red arrows. Moreover, the functional interactions between proteins within the pathway are highly conserved between species. Abbreviations: CNK, connector-enhancer of KSR; DRK, downstream of receptor kinase; Ets, E26/erythroblast transformation specific; Gap, GTPase activating protein; Grb2, growth factor receptor binding protein 2; KSR, kinase suppressor of ras; LC-PTPHePTP, lymphoid cell protein tyrosine phosphatase/hematopoietic cell protein tyrosine phosphatase; MAPK, mitogen activated protein kinase; MAPKK, MAPK kinase; PCPTP1, protein tyrosine phosphatase 1 from PC12 cells; PTP-ER, protein tyrosine phosphatase enhancer of ras; Raf, Ras, retrovirus associated sequences; Siah, seven in absentia human; Sina, seven in absentia; SOS, son of sevenless; STEP, striatum-enriched-protein tyrosine phosphatase.

DNA sequence, serves as a marker for subsequent gene identification. Target protein mutations that cause drug resistance have frequently been observed in microorganisms and viruses, which can replicate and generate mutations rapidly. A well known example is resistance to HIV protease inhibitors, which typically results from multiple amino acid changes in the active site of the protease[8,9]. Other cases, such as mutations that confer drug resistance upon mycobacteria[10], chlamydia[11] and malarial parasites[12], involve single amino acid substitutions.

Why use model organisms?

Model-organism genetic systems have several compelling features that favor their use for studying the MOA of organic molecules. Three of the most commonly used models – *Saccharomyces cerevisiae*, *C. elegans* and *D. melanogaster* – have short generation times and sophisticated genetic and genomic tools, which include an understanding of their complete genome sequences. Moreover, it is possible to grow large populations of these organisms inexpensively in a laboratory setting, an essential attribute that enables the identification of the often rare mis-sense mutation that disrupts the compound–target interaction[13]. The number of putative targets emerging from genetic screens is relatively small (typically 5–10 genes) – a tractable number for further evaluation and validation.

In a typical genetic screen, the model organism is treated with the compound of interest and examined closely for changes in phenotype. For example, the presence of a compound might cause developmental arrest of *C. elegans* at the larval stage. This observation can be used as the basis for a genetic screen, in which a large population of *C. elegans* is mutated (using chemical mutagens, irradiation or transposons) before treatment with the compound. Organisms that are resistant or hypersensitive to the effects of the compound are isolated and the genes responsible for

this resistance can then be mapped and cloned. A wide range of compound-induced phenotypes (representing an even wider range of underlying biochemical pathways) can be observed in model organisms and used as the basis for a screen. In *C. elegans*, these phenotypes can be morphological (e.g. body size or shape, developmental defects or lipid accumulation), behavioral (e.g. locomotion, egg laying and pharyngeal pumping) or cellular (e.g. reporter genes and immunohistochemistry). The initial characterization of compound-induced phenotypes is rapid, which allows many potentially interesting compounds to be tested in parallel. Compounds that produce a useful phenotype can subsequently be examined thoroughly before a genetic screen is initiated. Examples of compounds studied in various model organisms are presented in Table 1.

Creating a relevant phenotypic assay

The phenotype of the model organism is central to performing a genetic screen and thus must be reliable, reproducible and amenable to scoring in tens or hundreds of thousands of mutated organisms. Importantly, this approach relies on the assumption that the compound interacts with the relevant ortholog of the human target in the model organism. If the phenotypic assay reflects an interaction with a protein other than this ortholog, the relevance to human therapeutic application is questionable. However, such interactions could identify secondary targets, which might help elucidate the mechanisms of cross-reactivity and side effects. The relevance of the phenotypic assay can be addressed by using the lowest effective dose of a compound for genetic screening. Another approach is to treat the model organism with various structural derivatives of a compound, and then investigate the correlation between model-organism phenotype and vertebrate bioactivity. Alternatively, the phenotype itself can sometimes lend credence to its relationship to human physiology. If, for example, the observed vertebrate bioactivity is muscle relaxation, it would be reassuring to observe (in *C. elegans* or *D. melanogaster)* compound-induced phenotypes that are caused by the relaxation of muscle.

Model organism and human genes: a leap of faith?

The relevance of model-organism phenotypic assays to human physiology and disease is a major issue. It can be argued that the genetics of *D. melanogaster*, *C. elegans* and *S. cerevisiae* are too far removed from human genetics to provide useful information in the realm of human physiology and pathology. However, it is universally accepted that many diverse biochemical pathways show remarkable conservation between model organisms and humans, not only at the level of DNA and protein sequence, but also in the complex network of protein–protein interactions that define the signaling pathways[14] (Fig. 1). Hence, the crucial question for MOA studies is how often the compound interaction site is conserved between species. Indeed, this approach will only work if a compound-induced phenotype in the model organism is the result of a compound-induced change in the activity of an invertebrate protein, whose mammalian ortholog is the cognate drug target. Fortunately, drug molecules are typically directed against the active sites or ligand-binding domains of their target proteins. These regions are probably under significant evolutionary constraints, such that any mutations leading to changes in protein function are strongly disfavored. For example, the fungal metabolite brefeldin A is a potent inhibitor of protein secretion, and causes disruption of the

Table 1. Examples of compounds that have been studied in model organisms

Compound name	Mechanism or target	Indication	Model organism	Induced phenotype	Refs
Compactin	HMG-CoA reductase	Hypercholesterolemia	*S. cerevisiae*	Lethality	23
Fumagillin	Methionine aminopeptidase 2	Cancer	*S. cerevisiae*	Lethality on methionine aminopeptidase 2 mutants	43
Soraphen A	Acetyl-CoA carboxylase	Crop protection (fungicide)	*S. cerevisiae*	Lethality	53
Avermectin	Glutamate-gated Cl^- channel	River blindness	*C. elegans*	Lethality	54
Fluoxetine	Serotonin reuptake	Depression	*C. elegans*	Premature egg laying	32
Haloperidol	Dopamine D2 receptor	Psychoses	*C. elegans*	Movement	32
Dieldrin	GABA-gated Cl^- channel	Crop protection (insecticide)	*D. melanogaster*	Lethality	13
Cocaine	Monoamine transporter	Anesthesia	*D. melanogaster*	Behaviorial phenotypes	34

Abbreviations: *C. elegans*, *Caenorhabditis elegans*; *D. melanogaster*, *Drosophila melanogaster*; GABA, γ-aminobutyric acid; HMG-CoA, 3-hydroxy-3-methylglutaryl coenzyme A; *S. cerevisiae*, *Saccharomyces cerevisiae*.

Golgi complex in both yeast and mammalian cells. The molecular mechanism of this activity involves an exquisitely specific ternary complex between brefeldin A, a small GTP-binding protein of the ARF (ADP ribosylation factor) family and a guanine-nucleotide-exchange factor of the yeast Gea/Sec7 family[15]. Despite the profusion of GTP-binding proteins and related exchange factors in mammalian cells, sensitivity to brefeldin A is restricted to a subset of proteins that are closely related to their yeast homologs. Although the broad applicability of this approach for pharmaceutical target identification is relatively new, there are several compelling examples indicating that model-organism genetics can provide mechanistic answers that are laborious or impossible to obtain in mammalian systems.

Candidate compounds for MOA studies

What kinds of molecule are appropriate for study using this approach? One group of suitable compounds comprises marketed or late-stage clinical drugs with known pharmacological activity but unknown or ambiguous targets and/or mechanisms. Examples include the hypolipidemic agent gemfibrozil[16] and related fibrate drugs, and the antidepressant bupropion[17], which is also used as an aid to smoking cessation. These compounds, by virtue of their advancement through the clinical trials process, have a high degree of therapeutic validation in humans, and the discovery of their cognate targets (and of other components of their signaling pathways) can accelerate the discovery of improved second-generation drugs. Other compounds that are suitable for MOA studies are those that have failed at some stage during the drug discovery and optimization process. Reasons for such failure might include side effects [in which case the MOA approach might be useful for unveiling the secondary target(s)] or an inability of the *in vivo*-optimized molecule to recognize the primary target against which it was designed. Potential MOA compounds also exist in the field of natural product molecules[18]. Such molecules frequently have potent *in vivo* effects but have unknown targets and, furthermore, are difficult to synthesize. For example, phorboxazole A is a marine-sponge-derived natural product that exhibits potent anticancer activity[19]. However, its synthesis is complex[20] and its cellular target(s) remain obscure.

Identification of the cognate target using the MOA approach is of tremendous value for fueling new lead discovery efforts. An added advantage of this approach is that, once the target has been identified, the compound can be exploited as a useful reagent and a positive control for subsequent assay development and screening. Furthermore, if the target is amenable to structural elucidation, knowledge of the compound–target interaction can be used to initiate structure-based drug design efforts. Alternatively, in cases where the cognate target is known, there might still be significant value in pursuing MOA studies to delineate other potential targets within the same pathway.

Examples of MOA studies

Three examples of MOA studies in model systems, specifically chosen to highlight the strengths of each organism, will be discussed. In each case, an immense body of literature and functional annotation is available to help interpret the phenotypes that arise in the presence of the compound.

MOA in Saccharomyces cerevisiae

The first example describes the use of the yeast *S. cerevisiae*, a single-cell eukaryote with ~6200 predicted genes, ~30–40% of which have mammalian homologs[21,22]. In the mid-1980s, Rine and colleagues attempted to clone the yeast genes that encode HMG-CoA (3-hydroxy-3-methylglutaryl coenzyme A) reductase[23]. This was not a trivial task at the time because the existence of two functionally redundant HMG-CoA reductase isozymes in budding yeast made their identification by standard mutagenesis and complementation-cloning strategies unlikely[24]. Instead, a strategy was employed involving resistance to compactin, a hypocholesterolemic agent that inhibits the activity of HMG-CoA reductase, leading to fatality of the organism. A genomic library of yeast genes was transformed into yeast, treated with compactin and tested for suppression of lethality. The high copy number of these genes (~40 per cell), and the concomitant increased concentration of encoded protein, allowed yeast overexpressing the HMG-CoA reductase proteins to overcome the presence of the metabolic inhibitor and survive. The corresponding genes (HMG1 and HMG2) were consequently identified[23] and DNA sequence analysis revealed particularly high conservation between the yeast and mammalian HMG-CoA reductase genes in the catalytic domain. These proteins also provided one of the first examples of functional conservation because the mammalian gene was able to functionally replace the yeast genes[25].

One general limitation of the MOA approach is the extent to which a compound-induced phenotype can be exploited. In the above example, the drug compactin caused lethality, a severe and thus easily scored phenotype. However, the information gained from this phenotype alone is limited. Recently, the use of yeast to identify targets and uncover biochemical connections has been demonstrated using expression-profiling technology[26,27]. Hughes and coworkers[27] described the use of expression profiles to rapidly identify genes that respond to compound exposure. Here, the 'phenotype' is a compendium

of expression profiles induced by the compound. The usefulness of expression profiling in MOA studies was demonstrated by comparing the transcriptional profile of yeast treated with lovastatin (another HMG-CoA reductase inhibitor) with the transcription profiles of wild-type and mutant yeast. In this case, the expression profile of the lovastatin treated yeast was, as expected, similar to that of a yeast lacking HMG2 activity. This highlights one of the unique features of using yeast for this type of study: the ability to compare a compound with an unknown MOA, with the profile resulting from mutations in the individual yeast genes. Ideally, the profiles of both the drug and a knockout of its cognate target should be identical. Of course, the expression profiles obtained from strains that overexpress individual genes can also be collected, and one of these might be predicted to match the profile of an agonist compound. An important recent advance in this field is the development of computational-clustering algorithms[28–30], which allow the identification of co-regulated sets of genes, and grouping of functional pathways. However, frequently, such methods cannot distinguish between genes that are causally involved in the mechanism of a compound and genes whose expression is regulated as an effect of the compound. Moreover, the number of genes whose expression is regulated by a compound might be extremely large, making the identification of the actual cognate target a daunting task. Although it remains to be seen how often the genetic and drug-induced expression profiles overlap, the MOA method has the potential to make biochemical connections that were not previously known, and to test functional hypotheses.

MOA in Caenorhabditis elegans

The nematode *C. elegans* has proven to be an excellent model organism for MOA studies. The adult worm requires the coordinated activity of ~1000 cells, which include diverse cell types such as neurons, muscle, intestinal cells and germ cells. Using this complex *in vivo* system, Choy and Thomas[31] have studied the mechanism of fluoxetine, an antidepressant of the selective serotonin reuptake inhibitor class. In humans, fluoxetine causes an inhibition of serotonin reuptake, which results in an increased concentration of serotonin in neuronal synapses. The increase in serotonin levels has been proposed as the mechanism by which depression is alleviated. However, the rapid (hours) increase in serotonin levels and the long (weeks) delay in the relief from depression has suggested that fluoxetine might affect mechanisms other than serotonin reuptake. In *C. elegans*, the administration of fluoxetine causes several phenotypes, which include an increase in the egg-laying rate of the adult hermaphrodite, contraction of the muscles in the anterior portion ('nose') of the animal, and locomotory paralysis[32].

Do these phenotypes relate to human depression? At first glance, an increase in the rate of egg laying would appear to have nothing in common with psychiatric disorders. However, despite the fact that the physiological end-points are unrelated, the biochemical and cell biological processes appear to be conserved. Weinshenker and colleagues[32] demonstrated that the egg-laying phenotype is caused by an increase in serotonin levels in the neuromuscular synapses that control the contraction of the vulval muscles, presumably by antagonizing the ortholog of the human serotonin transporter. Nurrish and coworkers[33] demonstrated that mutations in a gene encoding a postsynaptically acting $G_o\alpha$ G-protein, and in a gene encoding a vesicular monoamine transporter known to act presynaptically, confer resistance to the paralysis phenotype. Taken together these results suggest that both the pre- and postsynaptic pathways impinged upon by fluoxetine are conserved in *C. elegans*.

To address the serotonin-independent mechanism of fluoxetine action, Choy and Thomas[31] chose to focus on the nose-contraction phenotype, which occurs even in worms that are unable to synthesize serotonin. They isolated mutations in genes that caused resistance to the fluoxetine-induced nose contraction, called 'Nrf' (nose resistant to fluoxetine). Of the two Nrf genes that have been cloned, ndg-4 was previously identified by resistance to nordihydroguaiaretic acid, a lipooxygenase inhibitor. Interestingly, the proteins encoded by the two Nrf genes are related to each other and are predicted to be multi-pass membrane proteins that are unlike any previously characterized transmembrane proteins. Whether these are fluoxetine transporters awaits biochemical studies and identification of relevant human homologs.

MOA in Drosophila melanogaster

The greater behavioral complexity of *D. melanogaster* relative to *C. elegans* has recently been exploited to address the molecular mechanisms that underlie drug abuse and addiction in humans. McClung and Hirsh[34] have carefully documented the behavior of *D. melanogaster* after either single or repeated exposures to cocaine and have found that, similar to their mammalian counterparts, fruit flies exhibit not only stereotypical behaviors in response to cocaine, but also behavioral sensitization in response to repeated exposure to the drug. Furthermore, the links between addiction in flies and humans probably share a common basis because sensitization to cocaine in humans is known to occur in regions of the brain that possess dopamine-containing cell bodies[35]. Hirsh and co-workers

characterized this phenomenon by identifying several mutant lines of *D. melanogaster* that respond normally to an initial dose of cocaine, but fail to sensitize to subsequent doses[36,37]. Interestingly, these results implicate a role for tyramine in the sensitization process, possibly via its interaction with monoamine transporters, and the subsequent increase in extracellular levels of dopamine. In addition, Li and colleagues[35] have recently shown that ectopic expression of an inhibitory Gα subunit in dopamine-containing neurons blocked sensitization. Similar techniques have been employed to explore mechanisms that underlie ethanol intoxication in *D. melanogaster*[38].

Identification and validation of human drug targets

Once invertebrate targets have been discovered through the MOA approach, it is necessary to identify and functionally validate the orthologous human target. The task of identification is becoming easier as the full sequence of the human genome emerges and will, in the near future, be considered more of a bioinformatics task than an exercise in cloning by hybridization. However, any gene discovered in a simple model-organism might typically have several human orthologs, and functional validation must be demonstrated empirically – particularly when the proposed cognate target for a given compound is a hitherto unknown or poorly characterized gene. Complete functional validation of a MOA-derived target will involve confirming the interaction between the compound and model organism target, and demonstrating an interaction between the compound and orthologous human protein(s).

Methods for validating compound–target interactions

In some instances, biochemical binding studies might be useful for the validation of compound–target interactions – for example, 'fishing' for targets using an immobilized variant of the compound. Such studies have been successful in several cases[39–41], although binding of the compound to a partner protein might not reveal the entire mechanism[42]. Affinity chromatography was used to isolate the cognate target, methionine aminopeptidase 2 (Met-AP2), of a derivative of the natural product fumagillin[43–45]. In this case it is interesting to note that the compound shows similar activity against the yeast ortholog of Met-AP2. However, in biochemical binding experiments the compound must be immobilized in an active conformation (which might not be obvious to determine *a priori*) and the target must be present in sufficient quantities to enable its isolation and detection. In the latter case, the availability of genomic sequence information, coupled with advances in mass spectrometry techniques[46,47], can facilitate the rapid identification of the target protein.

Following cloning of the model-organism gene and its human orthologs, the putative targets will typically need to be expressed in a heterologous system to enable further validation studies. The major challenge here is that the genes emerging from a screen could be drawn from a large number of protein families, ranging from soluble enzymes to ion channels and cell-surface receptors. Efficient *in vitro* expression systems have been described for the cells of *S. cerevisiae* and *D. melanogaster*, but not for those of *C. elegans*. Probably the most universal and rapid method for producing a candidate target protein is transient expression in mammalian cells. Although not large, the quantities of protein produced by these systems are often sufficient for confirmation of the compound–target interaction that has been implicated by direct-binding studies.

Radioligand binding offers perhaps the most direct and universal means of demonstrating a physical compound–target interaction[48]. Assuming that the compound of interest can be synthesized in a radiolabeled form, binding can be explored using whole cells, membrane and other subcellular fractions, and purified proteins, coupled with one of many standard techniques for separating bound from free ligand. Several other generally applicable, label-free systems are currently available and might have considerable use in the validation of compound–protein interactions. For example, surface plasmon resonance instrumentation [notably BIAcore[49] (BIAcore AB, Uppsala, Sweden)] has achieved sufficient sensitivity so that small molecules of <1000 Da can be detected, although it should be noted that immobilization of functional target proteins on the biosensor surface (particularly membrane proteins) can present significant technical challenges. Other emerging methods in this category include calorimetry and capillary electrophoresis.

Future directions

As described above, the MOA approach could be useful for dissecting entire biochemical pathways in addition to identifying an unknown molecular target. However, in the case of pathway identification, it is useful to be able to select *a priori* those components of the pathway (such as kinases, proteases and G-protein-coupled receptors) that are known to be amenable to modulation by a small molecule. Here again, model organisms can be used to favor selection of 'drugable' targets during the target discovery process itself. Reverse genetics (Box 1) affords the opportunity to search for specific classes of drug target that modulate the same pathway as a given compound. In this technique, detailed knowledge of the model-system genome is used to generate a systematic collection of organisms, each of which has had a specific gene knocked

Box 1. Forward and reverse genetics

Forward genetics refers to the traditional means of genetic screening, as depicted in Fig. I. In the case of mechanism of action studies, a compound-induced phenotype is used as a surrogate assay for compound function. A large population of organisms is mutated, and any individuals that are resistant (or hypersensitive) to the effects of the compound are isolated. The gene responsible for resistance is typically identified using positional cloning tools, which can be a time-consuming step.

Reverse genetics capitalizes on the fact that the genomic DNA sequence for each of the model organisms reviewed here is known. Thus, mutants can be produced in advance (using a variety of technologies) and individually tested for resistance or hypersensitivity to the compound. The advantage of this approach is that all mapping and cloning steps are eliminated, which greatly shortens the length of time between the genetic screen and target identification. However, as a result of technical limitations in generating the mutants, and the fact that a complete loss of gene function might be too severe to allow for a change in the phenotypic effect of the compound, reverse genetics is not a complete solution.

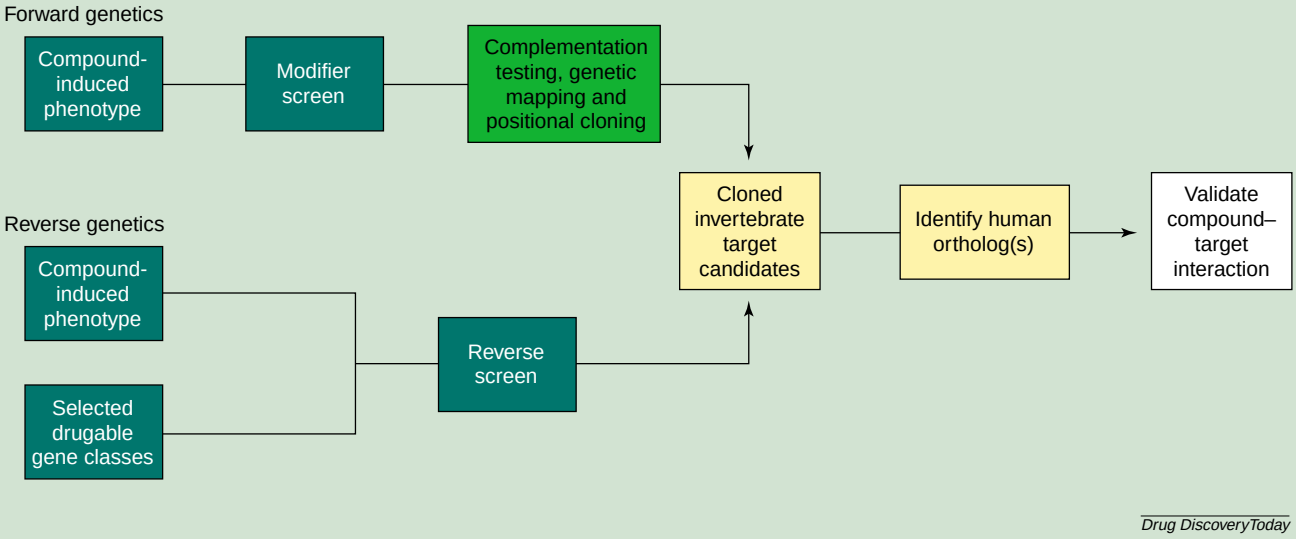

Figure I. Forward and reverse genetics.

out. This collection can subsequently be used to search entire classes of proteins (e.g. all of the kinases) that interact directly or indirectly with a small-molecule compound. One shortcoming of this reverse-genetics strategy is that it does not accommodate rare single amino acid changes that might be required to identify the cognate target. Thus, chemical mutagenesis and forward genetics are unlikely to be entirely supplanted by these more rapid tools. Another promising area of future development involves generation of sophisticated phenotypes. For example, expression profiling[50–52] (at both the RNA and protein level) could allow the identification of genes that are differentially regulated in response to certain cellular events, leading to the design of novel phenotypic readouts (such as transcriptional activation of fluorescent reporter proteins).

The industrialization of genetics

Until recently, yeast, worm and fly genetics was almost exclusively an academic pursuit. Although many key discoveries pertaining to fundamental biological processes were first made using model organisms, the drug discovery community has historically concerned itself only with the genes and proteins of higher mammals. Now, however, we are faced with an avalanche of data from the human genome sequence and need to implement new tools to extract useful information from these data. Model organism genetics is proving to be a pragmatic, rapid method for exploring the function and relevance of related human genes. Although it does not provide a complete solution (e.g. acquired immunity is not conserved in invertebrates), we can improve our chances of success by focusing on areas where proteins and pathways are well conserved, such as cancer, metabolic disease and neurological disorders. Indeed, even when genomes of higher organisms (e.g. mouse) become available and computational tools can handle such increased volumes of data, *D. melanogaster*, *C. elegans* and *S. cerevisiae* will still be used for studying well-conserved pathways by virtue of their small size (both genomic and physical), short life-cycle and experimental tractability. Coupling model-system genetics with known

bioactive compounds allows the rapid identification of targets and pathways for compounds of interest. Moreover, this drug-based target discovery approach offers a means of circumventing the crucial target validation bottleneck of the drug discovery process.

Acknowledgements

We thank our colleagues at Exelixis and Jim Thomas for helpful discussions and comments, and Jacqueline McLaughlin for assistance with graphics.

References

1 Drews, J. (2000) Drug discovery: a historical perspective. *Science* 287, 1960–1964
2 Wassarman, D.A. *et al.* (1995) The Ras signaling pathway in *Drosophila*. *Curr. Opin. Genet. Dev.* 5, 44–50
3 Sternberg, P.W. and Han, M. (1998) Genetics of RAS signaling in *C. elegans*. *Trends Genet.* 14, 466–472
4 Therrien, M. *et al.* (1995) KSR, a novel protein kinase required for Ras signal transduction. *Cell* 83, 879–888
5 Sundaram, M. and Han, M. (1995) The *C. elegans ksr-1* gene encodes a novel Raf-related kinase involved in Ras-mediated signal transduction. *Cell* 83, 889–901
6 Kornfeld, K. *et al.* (1995) The ksr-1 gene encodes a novel protein kinase involved in Ras-mediated signaling in *C. elegans*. *Cell* 83, 903–913
7 Rand, J.B. and Johnson, C.D. (1995) Genetic pharmacology: interactions between drugs and gene products in *Caenorhabditis elegans*. In *Methods in Cell Biology* (Vol. 48) (Epstein, H.F. and Shakes, D.C., eds), pp. 187–204, Academic Press
8 Condra, J.H. (1998) Resistance to HIV protease inhibitors. *Haemophilia* 4, 610–615
9 Roberts, N.A. *et al.* (1998) Resistance and cross-resistance with saquinavir and other HIV protease inhibitors: theory and practice. *AIDS* 12, 453–460
10 Lety, M.A. *et al.* (1997) A single point mutation in the embB gene is responsible for resistance to ethambutol in *Mycobacterium smegmatis*. *Antimicrob. Agents Chemother.* 41, 2629–2633
11 Wylie, J.L. *et al.* (1996) A single point mutant in CTP synthetase of *Chlamydia trachomatis* confers resistance to cyclopentenyl cytosine. *J. Biol. Chem.* 271, 15393–15400
12 Korsinczky, M. *et al.* (2000) Mutations in *Plasmodium falciparum* cytochrome b that are associated with atovaquone resistance are located at a putative drug binding site. *Antimicrob. Agents Chemother.* 44, 2100–2108
13 Ffrench-Constant, R.H. *et al.* (1993) A single-amino acid substitution in a γ-aminobutyric acid subtype A receptor locus is associated with cyclodiene insecticide resistance in *Drosophila* populations. *Proc. Natl. Acad. Sci. U. S. A.* 90, 1957–1961
14 Rubin, G.M. *et al.* (2000) Comparative genomics of the eukaryotes. *Science* 287, 2204–2215
15 Peyroche, A. *et al.* (1999) Brefeldin A acts to stabilize an abortive ARF-GDP-Sec7 domain protein complex: involvement of specific residues of the Sec7 domain. *Mol. Cell* 3, 275–285
16 Munday, M.R. and Hemingway, C.J. (1999) The regulation of acetyl-CoA carboxylase – a potential target for the action of hypolipidemic agents. *Adv. Enzyme Regul.* 39, 205–234
17 Horst, W.D. and Preskorn, S.H. (1998) Mechanisms of action and clinical characteristics of three atypical antidepressants: venlafaxine, nefazodone, bupropion. *J. Affect. Disord.* 51, 237–254
18 Harvey, A. (2000) Strategies for discovering drugs from previously unexplored natural products. *Drug Discov. Today* 5, 294–300
19 Searle, P.A. and Molinski, T.F. (1995) Phorboxazoles A and B: potent cytostatic macrolides from marine sponge *Phorbas* sp. *J. Am. Chem. Soc.* 117, 8126–8131
20 Forsyth, C.J. *et al.* (1998) Total synthesis of phorboxazole A. *J. Am. Chem. Soc.* 120, 5597–5598
21 Bussey, H. *et al.* (1997) The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XVI. *Nature* 387 (Suppl. 6632), 103–105
22 Cherry, J.M. *et al.* (1997) Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* 387 (Suppl. 6632), 67–73
23 Rine, J. *et al.* (1983) Targeted selection of recombinant clones through gene dosage effects. *Proc. Natl. Acad. Sci. U. S. A.* 80, 6750–6754
24 Basson, M.E. *et al.* (1986) *Saccharomyces cerevisiae* contains two functional genes encoding 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *Proc. Natl. Acad. Sci. U. S. A.* 83, 5563–5567
25 Basson, M.E. *et al.* (1988) Structural and functional conservation between yeast and human 3-hydroxy-3-methylglutaryl coenzyme A reductases, the rate-limiting enzyme of sterol biosynthesis. *Mol. Cell Biol.* 8, 3797–3808
26 Dimster-Denk, D. *et al.* (1999) Comprehensive evaluation of isoprenoid biosynthesis regulation in *Saccharomyces cerevisiae* utilizing the genome reporter matrix. *J. Lipid Res.* 40, 850–860
27 Hughes, T.R. *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126
28 Eisen, M.B. *et al.* (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14863–14868
29 Brown, M.P. *et al.* (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. U. S. A.* 97, 262–267
30 Sherlock, G. (2000) Analysis of large-scale gene expression data. *Curr. Opin. Immunol.* 12, 201–205
31 Choy, R.K.M. and Thomas, J.H. (1999) Fluoxetine-resistant mutants in *C. elegans* define a novel family of transmembrane proteins. *Mol. Cell.* 4, 143–152
32 Weinshenker, D. *et al.* (1995) Genetic pharmacological analysis of neurotransmitters controlling egg laying in *C. elegans*. *J. Neurosci.* 15, 6975–6985
33 Nurrish, S. *et al.* (1999) Serotonin inhibition of synaptic transmission: Gα(0) decreases the abundance of UNC-13 at release sites. *Neuron* 24, 231–242
34 McClung, C. and Hirsh, J. (1998) Stereotypic behavioural responses to free-base cocaine and the development of behavioral sensitization in *Drosophila*. *Curr. Biol.* 8, 109–112
35 Li, H. *et al.* (2000) Ectopic G-protein expression in dopamine and serotonin neurons blocks cocaine sensitization in *Drosophila melanogaster*. *Curr. Biol.* 10, 211–214
36 McClung, C. and Hirsh, J. (1999) The trace amine tyramine is essential for sensitization to cocaine in *Drosophila*. *Curr. Biol.* 9, 853–860
37 Andretic, R. *et al.* (1999) Circadian genes are required for cocaine sensitization in *Drosophila*. *Science* 285, 1066–1068
38 Moore, M.S. *et al.* (1998) Ethanol intoxication in *Drosophila*: genetic and pharmacological evidence for regulation by the cAMP signaling pathway. *Cell* 93, 997–1007
39 Chen, J.K. *et al.* (1999) The identification of myriocin-binding proteins. *Chem. Biol.* 6, 221–235
40 Crews, C.M. *et al.* (1994) GTP-dependent binding of the antiproliferative agent didemnin to elongation factor 1α. *J. Biol. Chem.* 269, 15411–15414
41 Crews, C.M. *et al.* (1996) Didemnin binds to the protein palmitoyl thioesterase responsible for infantile neuronal ceroid lipofuscinosis. *Proc. Natl. Acad. Sci. U. S. A.* 93, 4316–4319
42 Ahuja, D. *et al.* (2000) Inhibition of protein synthesis by didemnin B: how EF-1α mediates inhibition of translocation. *Biochemistry* 39, 4339–4346
43 Sin, N. *et al.* (1997) The anti-angiogenic agent fumagillin covalently binds and inhibits the methionine aminopeptidase, MetAP-2. *Proc. Natl. Acad. Sci. U. S. A.* 94, 6099–6103
44 Liu, S. *et al.* (1998) Structure of human methionine aminopeptidase-2 complexed with fumagillin. *Science* 282, 1324–1327

45 Wernert, N. *et al.* (1999) Inhibition of angiogenesis *in vivo* by *ets-1* antisense oligonucleotides – Inhibition of Ets-1 transcription factor expression by the antibiotic fumagillin. *Angew. Chem., Int. Ed. Engl.* 38, 3228–3231

46 Shevchenko, A. *et al.* (1997) Peptide sequencing by mass spectrometry for homology searches and cloning of genes. *J. Prot. Chem.* 16, 481–490

47 Gevaert, K. and Vandekerckhove, J. (2000) Protein identification methods in proteomics. *Electrophoresis* 21, 1145–1154

48 Hulme, E.C. and Birdsall, N.J.M. (1992) Strategy and tactics in receptor-binding studies. In *Receptor–ligand Interactions: A Practical Approach* (Hulme, E.C. ed.), pp. 63–176, Oxford University Press

49 Rich, R.S. and Myszka, D.G. (2000) Advances in surface plasmon resonance biosensor analysis. *Curr. Opin. Biotech.* 11, 54–61

50 Young, R.A. (2000) Biomedical discovery with DNA arrays. *Cell* 102, 9–15

51 Shimkets, R.A. *et al.* (1999) Gene expression analysis by transcript profiling coupled to a gene database query. *Nat. Biotechnol.* 17, 798–803

52 Pandey, A. and Mann, M. (2000) Proteomics to study genes and genomes. *Nature* 405, 837–846

53 Vahlensieck, H.F. *et al.* (1994) Identification of the yeast ACC1 gene product (acetyl-CoA carboxylase) as the target of the polyketide fungicide soraphen A. *Curr. Genet.* 25, 95–100

54 Cully, D.F. *et al.* (1994) Cloning of an avermectin-sensitive glutamate-gated chloride channel from *Caenorhabditis elegans*. *Nature* 371, 707–711

Elsevier Science to expand archival resources on ScienceDirect

ScienceDirect (www.sciencedirect.com) is the largest provider of scientific information online, with almost 1,200 journals allowing access to over a million articles from leading publishers. Elsevier Science has now announced its intention to supplement these resources by enabling ScienceDirect users to access backfiles that could extend to the first volume published. Tetrahedron is the longest published journal in organic chemistry and is to be made available online from its first volume, published in 1957. Titles in chemistry, mathematics and economics are the first to benefit from this move because these archives are the most frequently cited. However, Derk Haank (Chief Executive, Elsevier Science) has stated that Elsevier Science intend to make as much content available as possible on ScienceDirect. The inclusion of all chemistry journal backfiles, incorporating organic and inorganic chemistry, analytical chemistry, electrochemistry, physical and theoretical chemistry, spectroscopy and chemical engineering, will be completed by 2002. All archives will be viewable in PDF format, with abstracts, titles and references also available as HTML text. This will enable readers to benefit from using the advanced searching capabilities of ScienceDirect to reference both backfiles and recent research publications.