Published in final edited form as:
Nat Rev Genet. 2000 November ; 1(2): 116–125. doi:10.1038/35038557.

CHEMICAL GENETICS: LIGAND-BASED DISCOVERY OF GENE FUNCTION

Brent R. Stockwell
Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

Brent R. Stockwell: stockwell@wi.mit.edu

Abstract

Chemical genetics is the study of gene-product function in a cellular or organismal context using exogenous ligands. In this approach, small molecules that bind directly to proteins are used to alter protein function, enabling a kinetic analysis of the *in vivo* consequences of these changes. Recent advances have strongly enhanced the power of exogenous ligands such that they can resemble genetic mutations in terms of their general applicability and target specificity. The growing sophistication of this approach raises the possibility of its application to any biological process.

Classic FORWARD-GENETIC screens in model organisms have revealed the molecular basis for diverse biological processes, including cell division in the yeast *Saccharomyces cerevisiae*[1], programmed cell-death in the nematode *Caenorhabditis elegans*[2], and embryonic pattern formation in the fly *Drosophila melanogaster*[3]. Such genetic screens use a three-step procedure (FIG. 1a) that entails: first, random mutagenesis of a large number of organisms; second, screening the resulting mutants to identify those with a defect in the process of interest (for example, cell division, apoptosis or embryonic patterning); and last, identifying the mutations in specific genes that underlie the mutant phenotypes of interest. This is a powerful method for identifying genes that regulate biological processes in lower organisms.

It is difficult, however, to use a genetic approach in mammals because of their slow rate of reproduction, large physical size, and large, diploid genome[4]. But recent attempts at doing large-scale genetic screens in vertebrates, such as mice and zebrafish, suggest that such efforts are warranted[4–8].

A second limitation of the genetic approach is that most mutations are not conditional — they cannot be turned on or off at will. Even when conditional mutations can be found (for example, temperature-sensitive mutations or mutant alleles under the control of an inducible promoter), expressing the mutant allele usually entails causing an environmental stress to the organism, such as a temperature shift, which itself can have marked consequences[9] and may confound the interpretation of results. This drawback prevents the functions of essential genes from being investigated because organisms with mutations in these genes are not viable. In addition, organisms that carry constitutive mutations may have time to compensate for their effects by upregulating related genes, which can obscure the initial effects of a mutation.

DATABASE INFORMATION | p53 | MEK1 | somatostatin | somatostatin receptor subtype-2 | somatostatin receptor subtype-5 | TGF-β | plasminogen activator inhibitor type 1 | interleukin-2 | interleukin-6
FURTHER INFORMATION | The Gram stain | Paul Ehrlich | Animations of reporter-gene assays and antibody-based cellular immunoassays | *Saccharomyces* genome deletion project | Field lab page | Stockwell lab page

Despite these limitations, the power of the genetic approach stems from the ability of mutations to alter the function of a single gene product within the context of a complex cellular environment[10,11]. In a chemical-genetic approach, exogenous ligands are likewise used to alter the function of a single gene product within this complex cellular context. Chemical-genetic screens use a three-step procedure (FIG. 1b) that entails: first, assembling a set of mutation equivalents (ligands capable of altering protein function); second, doing a high-throughput screen for ligands that affect a biological process of interest (such as cell division, programmed cell death or embryonic pattern formation); and last, identifying protein targets of these active ligands.

The chemical-genetic approach is conditional and readily applicable to cells derived from complex organisms such as vertebrates. This is because a ligand can be added or removed at any time, enabling a kinetic analysis of the consequences of changes in protein function *in vivo*. This permits an analysis of the cellular or physiological events that occur immediately after altering the activity of a specific protein. In higher organisms, such as mice and humans, the approach can be used for 'genetic-like' screens in cells and tissues of these organisms, complementing genetic-screening methods. Although this review focuses on the use of protein-binding reagents, other exogenous ligands can be similarly used, such as RNA-binding antisense reagents and DNA-binding reagents[12,13].

This review summarizes the evolution of the conceptual framework and technologies necessary for implementing chemical-genetic analyses. The current techniques in generating chemical diversity, high-throughput screening and TARGET IDENTIFICATION are surveyed, and recent successes with these technologies are highlighted.

Early history: pure substances and receptors

The two theoretical underpinnings of chemical genetics are: first, that pure biologically active substances can be obtained; and second, that such substances act by binding to specific molecular targets within an organism. The origin of these pivotal ideas can be traced back to the eighteenth century, when many researchers began to suggest that plant extracts that affect animals each contain a single active ingredient that acts on a discrete part of the animal[14]. Subsequently, Frederick Serturner succeeded in isolating the first active compound (morphine) (FIG. 2e) from a medicinal plant (opium)[15]. The first conceptual pillar of chemical genetics was therefore solidified — that biological activity resides within pure substances.

The second premise of chemical genetics — that low-molecular-weight compounds act by binding to specific protein receptors — took another hundred years to develop. In the early nineteenth century, François Magendie and Claude Bernard proposed that small molecules act within a specific body structure, and Rudolf Buchheim formulated the idea that drug action could be explained on the basis of a physicochemical reaction between cell constituents and a particular drug[16]. However, the crucial breakthrough came at the beginning of the twentieth century, when Paul Ehrlich developed both the term for, and the concept of, a 'receptor' — the single, specific protein target of a small molecule[17,18]. (Note that a small-molecule receptor may or may not function as a cell-surface receptor in the traditional biological sense.) Today, Ehrlich's breakthrough has been extended with the vision that we will eventually identify a small-molecule partner for every protein[10].

Finally, the usefulness of SMALL ORGANIC MOLECULEs for studying biological systems was discovered at the same time that the theoretical bases of chemical genetics were developed. Before the nineteenth century, uncharacterized, naturally occurring plant extracts were used to perturb biological systems. Then Paul Ehrlich showed, somewhat surprisingly,

that small organic molecules derived from an industrial waste product, coal tar, could selectively interact with cells and tissues[19]. Ehrlich discovered that methylene blue (FIG. 2a) could stain nerve cells selectively, and he promoted the idea that low-molecular-weight organic compounds could be of value for studying their receptors in biological systems[14,18]. Furthermore, in 1884, Hans Christian Gram found that a particular dye, crystal violet (FIG. 2b), could selectively stain certain types of microorganism in a process now referred to as 'the Gram stain'[20]. Several useful small molecules, such as aspirin (FIG. 2c) and saccharin (FIG. 2d), were also discovered by studying coal-tar extracts, showing that even very simple organic molecules could have a range of interesting and useful biological effects[21,22]. Before long it was accepted that simple organic molecules could profoundly affect cellular and organismal systems, paving the way for a comprehensive chemical-genetic approach to understanding biological systems.

So, by the early twentieth century it was substantiated, first, that pure substances have biological activity, and second, that these substances interact with specific proteins within an organism. Furthermore, small molecules were discovered to be generally useful tools for probing biological systems because of their ability to interact selectively with different cells, tissues and organisms. Nonetheless, the development of these principles into chemical genetics, a general method for determining protein function, took almost another century. It is only now that we are beginning to witness the full power and extent of these simple ideas.

Chemical-genetic tools: chemical libraries

In the first step of the chemical-genetic process, a collection of diverse ligands, capable of altering protein function, is assembled. These ligands can be small organic molecules (FIG. 2) or PEPTIDE APTAMERs[23,24]. Both types of ligand act as mutation equivalents because they alter the function of their target protein. In either case, a LIBRARY of such ligands must be obtained.

The origin of chemical libraries

Modern pharmaceutical companies have chemical libraries of small molecules that collectively number in the millions. These enormous collections have been assembled one-by-one during the course of synthesizing many variants of each drug candidate over the past century. Furthermore, many tens of thousands of natural products, derived from marine sponges, fungi, bacteria and plants, have gradually been added to these libraries. These collections have become a valuable resource for screens related to therapeutic drug discovery and development. However, this screening approach was largely unavailable to academic researchers until recently, when small collections of organic molecules became commercially available and, more importantly, rapid syntheses of small molecule libraries were invented[25,26]. The synthesis of organic-compound or peptide libraries has developed into a discipline in itself — the field of chemical diversity.

Early efforts to generate chemical diversity focused on chemically synthesized peptides because they are composed of the same building blocks as proteins and were therefore thought to be good candidates for ligands that bind to proteins. These syntheses followed the pioneering work of Merrifield, who developed highly efficient SOLID-PHASE SYNTHESES of peptides[27]. Chemical syntheses of such libraries have been done using numerous methods: in solution using conventional chemistry[28], on small plastic beads using solid-phase peptide synthesis[29,30], and on silicon wafers using light-directed synthesis[31]. Genetically encoded peptide-aptamer libraries have also been developed, in which short peptides are displayed *in vitro* on phage particles[32], or *in vivo* on scaffold proteins[33]. Peptide libraries have yielded ligands for many different protein targets (TABLE 1).

Over the past decade, methods for the synthesis of chemical libraries of non-peptidic, low-molecular-weight organic compounds have been developed[25,26]. Organic compounds have the advantage of often being able to permeate the plasma membrane of target cells[34]. Although it is theoretically possible that completely unbiased libraries of small molecules might lack any biological activity, this approach has, in fact, been quite successful. Synthetic methods have been developed for producing a wide range of chemical structures using COMBINATORIAL CHEMISTRY (TABLE 2). These libraries have yielded ligands for many proteins (TABLE 1), as well as compounds that are active in cell-based assays[35–38].

However, despite the apparent success of combinatorial chemistry, the field is still in its infancy. It remains to be seen which types of combinatorial library will yield small molecules with exquisite protein-binding specificity, which is required if such compounds are to be used as mutation equivalents. Most of the large combinatorial libraries synthesized so far contain simple compounds with few or no stereocentres[26], in contrast to the stereochemical complexity that has been found in many natural products (FIG. 2)[39,40]. Proteins are stereochemically complex because they are composed of many α-amino acids, most of which have at least one stereocentre. It is likely that small molecules containing stereocentres will show greater specificity in their protein-binding properties than small molecules lacking stereocentres. But despite promising preliminary research[41], compounds rivalling the structural complexity, high potency and target specificity of natural products are not available using existing combinatorial synthetic methods.

Surprisingly, the optimal types of chemical structure for perturbing biological systems are not known. Most synthetic efforts in the pharmaceutical industry have focused on flat aromatic compounds, similar to the dyes synthesized in the nineteenth century (FIG. 2). The common view is that such compounds will be more readily developed into human therapeutic agents[42], although it is just as likely that many human therapeutic agents have been developed using these compounds simply because they were the only compounds available for testing. So despite the progress in developing new synthetic reactions during this century, there has been a lack of creativity in the selection of target chemical structures until recently. We are beginning to witness an explosion in creatively selected, novel, small organic molecules that will function as powerful tools for perturbing biological systems[43–45]. In time, perhaps, we will learn the fundamental principles that govern the relationship between chemical structure and both potency and selectivity. These will help guide the selection of target molecules for combinatorial syntheses.

Chemical-genetic tools: screens

To test the effects of thousands or millions of small molecules, rapid methods need to be devised to conduct many assays in parallel. Such methods are referred to as high-throughput screens because of their ability to put numerous candidate compounds through the screen in question. Testing small molecules in a high-throughput screen requires that each small molecule be spatially segregated to prevent reagent mixing. This feature has stimulated efforts to minimize the volume of each assay and so maximize the number of assays done in a given time[46].

As the number of compounds available for screening has increased, the throughput of assay technology has kept pace. Assay volumes were reduced to 100–300 μl in the 96-well plate format in the late 1970s, and to 20–80 μl in the 384-well plate format by the 1990s. The past few years have witnessed a proliferation of ‘nanowell’ plates, containing thousands of small wells that permit assay volumes of several hundred nanolitres[47–49]. This fulfils two purposes: first, decreasing the length of time that is required per assay (and therefore increasing the number of assays that can be completed in a given time), and second,

decreasing the cost of each assay. The combined effect has enabled the testing of hundreds of thousands of compounds each day in a fully automated high-throughput screening laboratory[46,49].

So now the rate-limiting step in applying this approach to a biological question is the development of a robust high-throughput screen for the process of interest. Towards this end, two types of screen are possible — target-based and phenotype-based screens.

Target-based screens

These screens are used to identify ligands for a specific protein of interest. Subsequently, such ligands can be used to inhibit or to activate the function of this protein *in vivo*. This method is conceptually analogous to REVERSE-GENETIC methods, such as gene targeting in mice, in that both methods search for the phenotypic consequences of altering the activity of a single protein (FIG. 1).

Target-based screens are done by purifying the protein target of interest and developing binding *in vitro* or enzymatic assays. In either case, the assay must be miniaturized to 2–100 μl to allow screening in 96-, 384- or 1,536-well plates. Typically, between 10,000 and 1,000,000 compounds are tested in a target-based screen, yielding 10 to 100 candidate ligands. These candidates are re-tested several times at various concentrations and only confirmed hits are taken forward for tests of specificity and *in vivo* functionality.

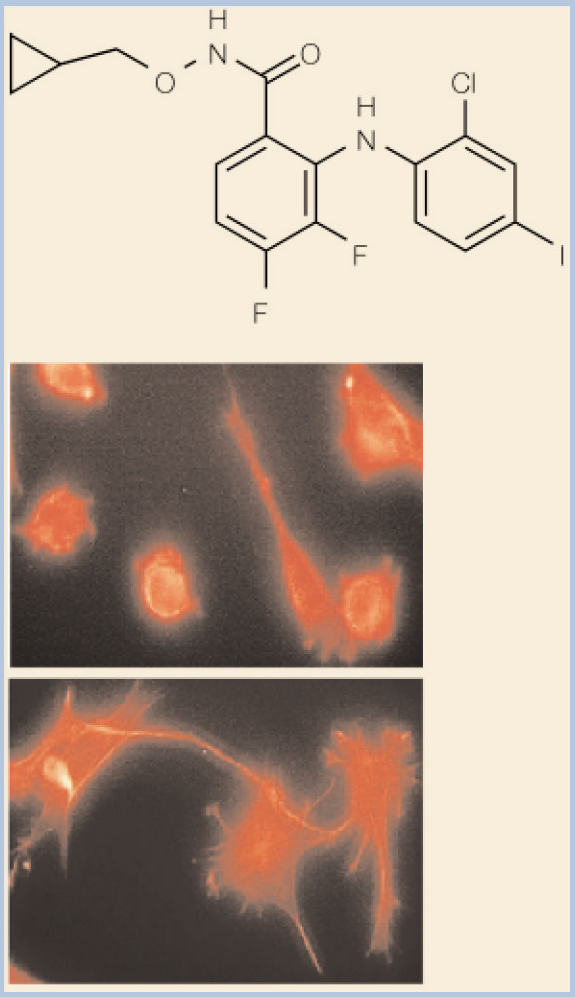
Ligands to a specific protein can be used to elucidate the phenotypic consequences of inhibiting or otherwise altering this protein[50]. For example, a small-molecule inhibitor of p53 was identified in a target-based screen and was used to show that loss of p53 function reduced the side effects of anti-tumour therapeutic regimens[51]. In addition, a small-molecule inhibitor of the mitogen-activated protein kinase kinase, MEK1**,** has been used to determine the function of this kinase in tumour cell growth *in vivo* (BOX 1)[52]. Target-based screens have yielded small-molecule or peptide ligands for many proteins, including kinases[33], phosphatases[53], proteases[54], cell-surface receptors[55,56], SH3 (Src homology region 3) domains[57], E3 ubiquitin ligases[58] and steroid receptors[59]. In general, small-molecule inhibitors of specific proteins can be enormously valuable for rapidly assessing whether the protein is involved in a particular biological process.

Box 1

Reverse chemical-genetics: a case study

The reverse chemical-genetic approach has recently been used to study the biology of MEK1[52]. An *in vitro* high-throughput kinase assay was used to select PD 184352 (see the figure), a small organic molecule that potently inhibits the kinase activity of MEK1 *in vitro*. It was found to be a specific inhibitor of MEK1 *in vitro* because it could not inhibit the kinase activities of several other kinases that were tested. To define the function of MEK1 in cell-cycle progression, cell growth and cell morphology, the effect of PD 184352 on these processes was tested in colon tumour cells. These studies found that MEK1 activity is required for: the progression of cells from the G1 to S phase of the cell cycle, anchorage-independent growth, cell scattering and the conversion of cells from a flattened to a round morphology (as shown in the figure, PD 184352 causes the round morphology of colon tumour cells (upper panel) to revert to a flattened morphology (lower panel))[52]. Treating colon tumour-bearing mice with PD 184352 caused tumour size to shrink, indicating that MEK1 function is also required for *in vivo* colon tumour cell growth. (PD 184352 had no effect in mice with leukaemia.) So in this study, a small-molecule MEK1 inhibitor was discovered and used to define the function of MEK1 in tumour processes. It is interesting to note that in the analogous reverse-genetic approach,

Mek1-deficient mouse embryos died in early embryogenesis[94] and, because it is an essential gene, its function could not be studied in tumour processes from adult mice by reverse-genetic methods. In such cases, reverse chemical-genetic methods complement and extend reverse-genetic methods for studying specific gene product functions *in vivo*.



(Adapted with permission from *Nature Med.* (REF. 52) © (1999) Macmillan Magazines Ltd)

The availability of whole-genome DNA sequences offers the possibility of doing target-based screens against all members of a protein family. Where whole-genome sequences are available, it is possible to screen all test compounds against each member and so identify specific inhibitors or activators of each family member. An example of this strategy was recently reported in which a COMBINATORIAL LIBRARY was screened against five subtypes of somatostatin receptors[60]. Specific agonists were identified for each receptor subtype, and these agonists were used to identify downstream signalling events activated by each receptor[60]. For example, somatostatin receptor subtype-2 activation led to inhibition of glucagon release from mouse pancreatic α-cells, whereas somatostatin receptor subtype-5 activation altered insulin secretion from pancreatic β-cells[60]. This study also showed the feasibility of using small molecules to activate, rather than inactivate, specific members of a protein family. Such functional manipulation is difficult to achieve using conventional reverse-genetic methods. However, the difficulty of identifying specific agonists for each

member of a receptor family increases as the number of proteins in the family increases. For example, the kinase family, with hundreds of members, represents a formidable challenge. Identifying selective inhibitors of each kinase will probably require larger and more complex libraries of exogenous ligands than have been synthesized so far.

Phenotype-based screens

These screens test the ability of each member of a library to induce a specific phenotypic outcome in a cell or organism (FIG. 3). Such screens are analogous to classic forward-genetic screens in model organisms (FIG. 1)[11].

Phenotype-based chemical-genetic screens have been done in single-cell organisms, such as bacteria[25,26] and fungi[61], and in cells from multicellular organisms, such as mice and humans[62,63]. No phenotype-based, chemical-genetic screen has been done in whole animals because of the difficulty of maintaining, injecting and observing thousands, or tens of thousands, of animals. However, recent progress in vertebrate phenotype-based genetic screening[4–6,64,65] indicates that such an approach may ultimately be feasible.

Detection methods for phenotype-based screens include the use of markers, functional assays and microscopic imaging of cells. Assays using marker genes or proteins (for example, reporter-gene assays and antibody-based cellular immunoassays) measure the abundance of a specific gene or protein epitope as a marker of a more broad cellular phenotype. For example, transforming growth factor-β (TGF-β) induces numerous transcriptional and post-translational changes in target cells, but the expression level of a single gene (plasminogen activator inhibitor type I) is often used as a marker for active TGF-β signalling[66]. Functional assays directly measure cellular activities such as cell division, metabolism, apoptosis, chemotaxis or adhesion. Finally, automated microscope-based imaging systems can detect changes in cellular morphology or changes in subcellular localization of marker proteins[67].

Small molecules or peptides discovered in phenotype-based screens can be used to identify proteins regulating a specific phenotype. The discovery of calcineurin as the target of the immunosuppressant FK506 (FIG. 2f) in 1987 represents the first successful application of the three steps of the chemical-genetic process. Before the discovery of this small molecule, the structurally unrelated cyclic peptide cyclosporin A (CsA) (FIG. 2g) was used to treat solid-organ transplant rejection. CsA was known to inhibit the production of T-cell-derived cytokines such as interleukin-2 (IL-2). Using a phenotype-based screen, Kino *et al.* tested fermentation broth extracts from *Streptomyces* for their ability to phenocopy CsA by blocking IL-2 production[68,69]. This screen resulted in the discovery of FK506, and subsequent studies placed the molecular target of FK506, calcineurin, in the T-cell receptor-initiated signalling pathway. These results led to the elucidation of the molecular events that govern this membrane-to-nucleus signalling pathway. More recent phenotype-based screens have yielded reagents that block mitotic progression[61,62,70], induce or suppress cell-cycle arrest[11,66,71], prevent mating-pheromone-induced cell-cycle arrest[23,72], prevent interleukin-6 (IL-6) secretion[73], induce apoptosis[74] and prevent endothelial cell activation[75].

Target identification

Since Paul Ehrlich's invention of the protein receptor concept, it has been understood that small molecules exert their effects on biological systems by interacting with specific protein targets[76]. Despite this early insight, identifying the protein target of a specific small molecule was not feasible until methods were invented for purifying and sequencing proteins and cloning their cognate genes. Perhaps the first example of protein target identification was the covalent labelling[77], peptide sequencing[78] and subsequent cloning[79]

of the acetylcholine (ACh) receptor. A similar strategy was used to identify the intracellular receptors for various steroid hormones[80]. In these studies, ACh and steroid hormones were used to label their respective protein targets, allowing the proteins to be purified from cellular extracts. This process became a powerful method for identifying protein targets of small molecules, and was used successfully to identify the targets of many natural products, including FK506 (REF. 81), cyclosporin[81], rapamycin[82], trapoxin[83], fumagillin[84], depudecin[85] and lactacystin[86].

Although biochemical purification of protein targets has been successfully used for many natural products, the method is laborious, difficult and often unreliable. As a result, new methods for rapid and efficient protein-target identification are now being developed[87]. An exciting new tool that can be applied towards this goal is DNA microarray technology. In one approach, the changes in gene expression patterns that occur in response to treatment of cells with a small molecule are determined. This pattern of gene expression may reveal specific gene expression changes that might reflect or explain the activity of a small molecule. For example, in a recent study, the induction of metal-ion-responsive genes by a small molecule revealed that the direct targets of the small molecule were specific metal ions[66].

In a related approach, the pattern of gene expression changes observed are treated as a molecular signature or fingerprint of the compound. Pattern-matching algorithms are then used to determine whether a small molecule has the same molecular signature as a gene deletion. For example, a recent study found that the pattern of gene expression changes that occurred in response to treatment of yeast cells with the anaesthetic dyclonine was the same as that which occurred in response to deletion of the *ERG2* gene. So it was revealed that the protein target of dyclonine was Erg2 (REF. 88).

DNA arrays can also be used to determine those yeast strains that grow under selective pressure, such as in the presence of a small molecule in a culture containing many different yeast strains. This is a particularly powerful approach if the pool of yeast strains consists of a large number of strains, each with a different gene deletion. A yeast strain bearing a deletion in the gene that encodes the protein target of a small molecule will not respond to it. Furthermore, although many genes in yeast are haplosufficient, the gene encoding the protein target of a small molecule may effectively be rendered haploinsuffficient by the presence of the small molecule. Therefore, in a recent study, the known targets of six small molecules were verified using DNA microarrays, by determining those deletion strains that showed compound-induced haploinsufficiency[89].

Finally, DNA arrays can be used for rapid ALLELIC SCANNING and for mapping loci of interest. In this forward-genetic approach, genetic loci responsible for a specific phenotypic trait are identified. If the phenotypic trait is susceptibility or resistance to an exogenous ligand, such as a small molecule, then this genetic-mapping strategy will identify the locus that encodes the target of the ligand. This strategy was recently applied in *S. cerevisiae* to map a drug resistance locus for cycloheximide by comparing a sensitive strain with a resistant strain[90].

An alternative method of identifying ligand-interacting proteins is the yeast two-hybrid system[91] — a yeast transcription-based system that measures the interaction of a test protein with many other proteins. In the ‘three-hybrid’ variant of this system, a DNA-binding domain is fused to the protein target of a known ligand, which is then covalently linked to a small molecule of interest. The small molecule of interest is displayed against a library of proteins fused to a transcriptional activation domain. If the small molecule is capable of binding to one such protein, the activation domain will be recruited to the small molecule

and the DNA-binding domain, resulting in activation of a reporter gene[47,92]. The simpler two-hybrid approach has been quite useful in determining interacting proteins for peptide aptamers. In one case, peptide aptamers that blocked mating pheromone-induced arrest in *S. cerevisiae* were used in a series of two-hybrid screens to identify interacting proteins[23]. Numerous targets were identified that had not previously been linked to pheromone-induced cell-cycle arrest, including cell wall biosynthesis kinase 1 (Cbk1)[23]. So in this study, a chemical-genetic approach was successfully used to identify proteins involved in the pheromone responsive signalling pathway.

Validation and specificity

Once a small molecule or peptide is selected in either a forward or reverse chemical-genetic screen, the specificity of this ligand for its protein target must be assessed and its ability to selectively perturb the protein target validated. Specificity can be assessed initially by testing whether the ligand binds to or inhibits the function of other related proteins *in vitro* (BOX 1)[60]. With emerging genomic technologies, it will probably become possible to test the ability of a ligand to bind to every protein in an organism. However, the fact that a ligand does not bind to proteins other than a candidate protein, even in such global surveys, cannot be used to conclusively prove its specificity because a ligand might bind to a protein–protein interface or a conformational epitope *in vivo* that is not found *in vitro*.

The ultimate test of the specificity of a ligand must occur in a cellular context, where all possible target molecules are present in their native conformations. Recent advances in EXPRESSION PROFILING offer new avenues for determining the specificity of a ligand in a cellular context. One test for specificity is done using reverse-genetic methods to delete the gene for a candidate target of a ligand. If a ligand has an effect in cells that do not express a candidate target protein, there must be a secondary target for this ligand. Norman *et al.*[61] observed such a result for a peptide ligand discovered in a forward chemical-genetic screen that aimed at identifying suppressors of the spindle assembly checkpoint (BOX 2).

Box 2

Forward chemical-genetics: a case study

In an analysis of the spindle checkpoint[61], a library of peptides containing 16 random amino acids was displayed on a scaffold protein (a staphylococcal nuclease) and was expressed in yeast cells. In a screen for suppressors of the spindle checkpoint, three peptides that suppressed the Mps1 (monopolar spindle 1)-induced activation of this checkpoint were identified from 6.5 million screened peptides. One peptide also suppressed spindle checkpoint arrest induced by depolymerization of the mitotic spindle and by the introduction of short linear minichromosomes, indicating that it was a general inhibitor of the spindle checkpoint. Subsequently, the protein target of this peptide was determined. Using a two-hybrid analysis, the peptide suppressor of the spindle checkpoint was found to interact with the product of *YDR517W*, a yeast open reading frame of unknown function. Subsequent deletion of *YDR517W* revealed that this gene is required for a functional spindle checkpoint, validating this forward chemical-genetic screen.

Moreover, the global pattern of changes in gene expression occurring in response to a ligand can be used to create an expression profile for the ligand[93]. By comparing such an expression profile with the expression profiles obtained after deleting a candidate gene, it is possible to determine whether a given ligand is targeting one or more proteins. For example, Marton *et al.* measured the expression profile of 3-aminotriazole, an inhibitor of His3, in wild-type yeast cells and found more than 1,000 changes in gene expression[93]. The

expression profile of *HIS3*-deleted cells had a nearly identical pattern of changes in gene expression, indicating that His3 is probably the only protein target of 3-aminotriazole[93]. Furthermore, it is possible to measure the expression profile of a small molecule in cells deleted for the primary target to determine whether there are transcriptional changes caused as a result of binding to other proteins. Using this logic, Marton *et al.* measured the expression profile of 3-aminotriazole in *HIS3*-deleted cells and found no changes in gene expression, indicating that 3-aminotriazole is a specific inhibitor of His3 in yeast cells[93].

Conclusion

The power of classic genetic screens stems from their ability to select, in an unbiased manner, the few genes out of an entire genome that regulate a particular biological process. Such screens have been enormously powerful in dissecting the molecular events governing the behaviour of lower organisms. Much effort is now being devoted to developing technologies that allow genetic-like screens to be done in mammalian systems.

The utility of mutations in a classical genetic approach derives from their specificity and their broad generality. In other words, a mutation can specifically alter the function of a single gene within a complex genome, but because mutations can be made in any gene, mutations are generally applicable and can be used to study nearly any biological process. The use of exogenous ligands in the chemical-genetic approach in place of mutations requires that such ligands be likewise both general and specific. The former criterion seems likely to be met with small-molecule and peptide ligands, as such ligands can be found for nearly any protein. The latter criterion might also be met, because in some cases even very simple small molecules can interact specifically with a single protein target[93]. Therefore, it is possible that exogenous ligands can be both generally applicable and specific in their mechanism of action.

It is, however, difficult to screen exogenous ligands in whole animals, such as mice, because of the difficulty and expense of testing numerous ligands in animals. Furthermore, determining the target specificity of an exogenous ligand is laborious, although essential to the chemical-genetic process. Future advances in this area will probably involve improvements in the ease and speed of target identification and specificity assessment. In addition, future studies may reveal fundamental relationships between chemical structure and target specificity. Knowledge of such relationships will greatly aid the design of future chemical libraries.

Finally, the chemical-genetic approach requires three critical technologies: first, diverse chemical or peptide libraries; second, high-throughput screening; and last, protein-target identification. Early efforts using these technologies focused on drug development, rather than on the dissection of biological systems. However, these technologies have become sufficiently robust and widespread that we can now conceive of using them to study virtually any biological process. Such an approach promises to provide a generally useful, rapid and conditional method of altering protein function in mammalian systems. It is possible that chemical genetics will prove to be as revealing in the century to come as genetics has been in the previous century.

Acknowledgments

The author thanks J. M. Scharf and M. S. Stockwell for critical reading of the manuscript.

Glossary

FORWARD GENETICS	Genetic analysis that proceeds from phenotype to genotype through positional cloning or a candidate-gene approach
TARGET IDENTIFICATION	The process of finding cellular macromolecules that bind to a small molecule or peptide
SMALL ORGANIC MOLECULE	(also called small molecule) Carbon-containing compounds that usually have a molecular weight of less than 2,000 g mol^{-1}
STEREOCENTRE	A carbon atom that bears four distinct functional groups
PEPTIDE APTAMER	A short oligomer of amino acids, often fused to a protein scaffold
LIBRARY	A large collection of peptides, small molecules or other reagents
SOLID-PHASE SYNTHESIS	Chemical synthesis method using a solid support, such as a plastic bead
COMBINATORIAL CHEMISTRY	The synthesis of numerous organic compounds by combining variations of each of the building blocks that comprise the compounds
REVERSE GENETICS	Genetic analysis that proceeds from genotype to phenotype through gene-manipulation techniques
COMBINATORIAL LIBRARY	A collection of small molecules synthesized by combinatorial chemistry
ALLELIC SCANNING	The rapid detection of DNA sequence variations between two strains by analysing total genomic DNA on high-density DNA microarrays
EXPRESSION PROFILING	The use of DNA microarrays to determine the expression level of thousands of genes simultaneously

References

1. Hartwell LH. Twenty-five years of cell cycle genetics. Genetics. 1991; 4:975–980. [PubMed: 1783298]
2. Metzstein MM, Stanfield GM, Horvitz HR. Genetics of programmed cell death in *C. elegans*: past, present and future. Trends Genet. 1998; 14:410–416. [PubMed: 9820030]
3. Nusslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in *Drosophila.* Nature. 1980; 287:795–801. [PubMed: 6776413]
4. Justice MJ, Noveroske JK, Weber JS, Zheng B, Bradley A. Mouse ENU mutagenesis. Hum Mol Genet. 1999; 8:1955–1963. [PubMed: 10469849]
5. Mullins MC, Hammerschmidt M, Haffter P, Nusslein-Volhard C. Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. Curr Biol. 1994; 4:189–201. [PubMed: 7922324]
6. Vitaterna MH, et al. Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. Science. 1994; 264:719–725. This study was the first to successfully use random mutagenesis, phenotype-based screening and positional cloning (that is, a forward-genetic approach) in mice. [PubMed: 8171325]
7. Reppert SM, Weaver DR. Forward genetic approach strikes gold: cloning of a mammalian *Clock* gene. Cell. 1997; 89:487–490. [PubMed: 9160739]

8. Munroe RJ, et al. Mouse mutants from chemically mutagenized embryonic stem cells. Nature Genet. 2000; 24:318–321. [PubMed: 10700192]
9. Schena M, et al. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA. 1996; 93:10614–10619. [PubMed: 8855227]
10. Schreiber SL. Chemical genetics resulting from a passion for synthetic organic chemistry. Bioorg Med Chem. 1998; 6:1127–1152. [PubMed: 9784856]
11. Stockwell BR, Haggarty SJ, Schreiber SL. High-throughput screening of small molecules in miniaturized mammalian cell-based assays involving post-translational modifications. Chem Biol. 1999; 6:71–83. [PubMed: 10021420]
12. Agrawal S, Zhao Q. Antisense therapeutics. Curr Opin Chem Biol. 1998; 2:519–528. [PubMed: 9736926]
13. Dervan PB, Burli RW. Sequence-specific DNA recognition by polyamides. Curr Opin Chem Biol. 1999; 3:688–693. [PubMed: 10600731]
14. Parascandola J. The theoretical basis of Paul Ehrlich's chemotherapy. J Hist Med Allied Sci. 1981; 36:19–43. [PubMed: 7009720]
15. Schmitz R. Friedrich Wilhelm Serturner and the discovery of morphine. Pharm Hist. 1985; 27:61–74. [PubMed: 11611724]
16. Parascandola J. Reflections on the history of pharmacology: the 1980 Kremers award address. Pharm Hist. 1980; 22:131–140. [PubMed: 11610767]
17. Ehrlich P. Chemotherapeutics: Scientific principles, methods and results. Lancet. 1913; ii:445–451.
18. Travis AS. Science as a receptor of technology: Paul Ehrlich and the synthetic dyestuffs industry. Sci Context. 1989; 3:383–408. [PubMed: 11622453]
19. Travis AS. Perkin's mauve: ancestor of the organic chemical industry. Technol Cult. 1990; 31:51–82.
20. Popescu A, Doyle RJ. The Gram stain after more than a century. Biotech Histochem. 1996; 71:145–151. [PubMed: 8724440]
21. McTavish JR. Aspirin in Germany: the pharmaceutical industry and the pharmaceutical profession. Pharm Hist. 1987; 29:103–115. [PubMed: 11612139]
22. Kaufman GB, Priebe PM. The discovery of saccharin: a centennial retrospect. Ambix. 1978; 25:191–207. [PubMed: 11615708]
23. Geyer CR, Colman-Lerner A, Brent R. 'Mutagenesis' by peptide aptamers identifies genetic network members and pathway connections. Proc Natl Acad Sci USA. 1999; 96:8567–8572. This is the first successful use of peptide aptamers in a forward chemical-genetic screen. [PubMed: 10411916]
24. Colas P. Combinatorial protein reagents to manipulate protein function. Curr Opin Chem Biol. 2000; 4:54–59. [PubMed: 10679377]
25. Dolle RE. Comprehensive survey of chemical libraries yielding enzyme inhibitors, receptor agonists and antagonists, and other biologically active agents: 1992 through 1997. Mol Diversity. 1998; 3:199–233.
26. Dolle RE, Nelson KH Jr. Comprehensive survey of combinatorial library synthesis: 1998. J Comb Chem. 1999; 1:235–282. [PubMed: 10748736]
27. Merrifield B. Concept and early development of solid-phase peptide synthesis. Methods Enzymol. 1997; 289:3–13. [PubMed: 9353714]
28. Schneider CH, Rolli H, Blaser K. Liquid–liquid extraction in peptide synthesis. Int J Pept Protein Res. 1980; 15:411–419. [PubMed: 7440052]
29. Lam KS, et al. A new type of synthetic peptide library for identifying ligand-binding activity. Nature. 1991; 354:82–84. [PubMed: 1944576]
30. Furka A, Sebestyen F, Asgedom M, Dido G. General method for rapid synthesis of multicomponent peptide mixtures. Int J Pept Protein Res. 1991; 37:487–493. References 29 and 30 describe split-pool synthesis, a revolutionary method for synthesizing very large chemical libraries. [PubMed: 1917305]
31. Fodor SP, et al. Light-directed, spatially addressable parallel chemical synthesis. Science. 1991; 251:767–773. [PubMed: 1990438]

32. Scott JK, Smith GP. Searching for peptide ligands with an epitope library. Science. 1990; 249:386–390. [PubMed: 1696028]
33. Colas P, et al. Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. Nature. 1996; 380:548–550. [PubMed: 8606778]
34. Lien EJ, Wang PH. Lipophilicity, molecular weight, and drug action: reexamination of parabolic and bilinear models. J Pharmacol Sci. 1980; 69:648–650.
35. Tan DS, Foley MA, Stockwell BR, Shair MD, Schreiber SL. Synthesis and preliminary evaluation of a library of polycyclic small molecules for use in chemical genetic assays. J Am Chem Soc. 1999; 121:9073–9087.
36. Sullivan RW, et al. 2-Chloro-4-(trifluoromethyl)pyrimidine-5-*N*-(3′,5′-bis(trifluoromethyl)phenyl)-carboxamide: a potent inhibitor of NF-κB- and AP-1-mediated gene expression identified using solution-phase combinatorial chemistry. J Med Chem. 1998; 41:413–419. [PubMed: 9484492]
37. An H, Haly BD, Cook PD. Discovery of novel pyridinopolyamines with potent antimicrobial activity: deconvolution of mixtures synthesized by solution-phase combinatorial chemistry. J Med Chem. 1998; 41:706–716. [PubMed: 9513599]
38. Boger DL, Chai W, Jin Q. Multistep convergent solution-phase combinatorial synthesis and deletion synthesis deconvolution. J Am Chem Soc. 1998; 120:7220–7225.
39. Hung DT, Jamison TF, Schreiber SL. Understanding and controlling the cell cycle with natural products. Chem Biol. 1996; 3:623–640. [PubMed: 8807895]
40. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol. 1998; 5:245–249. The authors describe a powerful method for discovering new natural products using bacterial genetics.
41. Tan DS, Foley MA, Shair MD, Schreiber SL. Stereoselective synthesis of over two million compounds having structural features both reminiscent of natural products and compatible with miniaturized cell-based assays. J Am Chem Soc. 1998; 120:8565–8566 . This is the first demonstrated use of combinatorial chemistry to synthesize a large library of structurally complex small molecules.
42. Wang J, Ramnarayan K. Toward designing drug-like libraries: a novel computational approach for prediction of drug feasibility of compounds. J Comb Chem. 1999; 1:524–533. [PubMed: 10748729]
43. Gellman SH. Foldamers: a manifesto. Acc Chem Res. 1998; 31:173–180.
44. Schreiber SL. Target-oriented and diversity-oriented organic synthesis in drug discovery. Science. 2000; 287:1964–1969. [PubMed: 10720315]
45. Vaino AR, Janda KD. Euclidean shape-encoded combinatorial chemical libraries. Proc Natl Acad Sci USA. 2000; 97:7692–7696. [PubMed: 10884401]
46. Sundberg SA. High-throughput and ultra-high-throughput screening: solution- and cell-based approaches. Curr Opin Biotechnol. 2000; 11:47–53. [PubMed: 10679349]
47. Borchardt A, Liberles SD, Biggar SR, Crabtree GR, Schreiber SL. Small molecule-dependent genetic selection in stochastic nanodroplets as a means of detecting protein–ligand interactions on a large scale. Chem Biol. 1997; 4:961–968. [PubMed: 9427663]
48. You AJ, Jackman RJ, Whitesides GM, Schreiber SL. A miniaturized arrayed assay format for detecting small molecule–protein interactions in cells. Chem Biol. 1997; 4:969–975. [PubMed: 9427662]
49. Mere L, et al. Miniaturized FRET assays and microfluidics: key components for ultra-high-throughput screening. Drug Discov Today. 1999; 4:363–369. [PubMed: 10431146]
50. Wetterau JR, et al. An MTP inhibitor that normalizes atherogenic lipoprotein levels in WHHL rabbits. Science. 1998; 282:751–754. [PubMed: 9784135]
51. Komarov PG, et al. A chemical inhibitor of p53 that protects mice from the side effects of cancer therapy. Science. 1999; 285:1733–1737. [PubMed: 10481009]
52. Sebolt-Leopold JS, et al. Blockade of the MAP kinase pathway suppresses growth of colon tumors *in vivo*. Nature Med. 1999; 5:810–816. [PubMed: 10395327]
53. Aramburu J, et al. Affinity-driven peptide selection of an NFAT inhibitor more selective than cyclosporin A. Science. 1999; 285:2129–2133. [PubMed: 10497131]

54. Apletalina E, Appel J, Lamango NS, Houghten RA, Lindberg I. Identification of inhibitors of prohormone convertases 1 and 2 using a peptide combinatorial library. J Biol Chem. 1998; 273:26589–26595. [PubMed: 9756897]
55. Tian SS, et al. A small, nonpeptidyl mimic of granulocyte-colony-stimulating factor. Science. 1998; 281:257–259. [PubMed: 9657720]
56. Zhang B, et al. Discovery of a small molecule insulin mimetic with antidiabetic activity in mice. Science. 1999; 284:974–977. [PubMed: 10320380]
57. Chen JK, Lane WS, Brauer AW, Tanaka A, Schreiber SL. Biased combinatorial libraries: Novel ligands for the SH3 domain of phosphatidylinositol 3-kinase. J Am Chem Soc. 1993; 115:12591–12592.
58. Bottger V, et al. Identification of novel mdm2 binding peptides by phage display. Oncogene. 1996; 13:2141–2147. [PubMed: 8950981]
59. Norris JD, et al. Peptide antagonists of the human estrogen receptor. Science. 1999; 285:744–746. [PubMed: 10426998]
60. Rohrer SP, et al. Rapid identification of subtype-selective agonists of the somatostatin receptor through combinatorial chemistry. Science. 1998; 282:737–740. These authors were the first to apply reverse chemical-genetics to a protein family. They identify specific small molecule partners of each member of a family of somatostatin cell-surface receptors. [PubMed: 9784130]
61. Norman TC, et al. Genetic selection of peptide inhibitors of biological pathways. Science. 1999; 285:591–595. [PubMed: 10417390]
62. Mayer TU, et al. Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. Science. 1999; 286:971–974. [PubMed: 10542155]
63. Rosania GR, et al. Myoseverin, a microtubule-binding molecule with novel cellular effects. Nature Biotechnol. 2000; 18:304–308. [PubMed: 10700146]
64. Hrabe de Angelis MH, et al. Genome-wide, large-scale production of mutant mice by ENU mutagenesis. Nature Genet. 2000; 25:444–447. [PubMed: 10932192]
65. Nolan PM, et al. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. Nature Genet. 2000; 25:440–443. [PubMed: 10932191]
66. Stockwell BR, Hardwick JS, Tong JK, Schreiber SL. Chemical genetic and genomic approaches reveal a role for copper in specific gene activation. J Am Chem Soc. 1999; 121:10662–10663.
67. Boland MV, Markey MK, Murphy RF. Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. Cytometry. 1998; 33:366–375. [PubMed: 9822349]
68. Kino T, et al. FK-506, a novel immunosuppressant isolated from a *Streptomyces* I. Fermentation, isolation, and physico-chemical and biological characteristics. J Antibiot (Tokyo). 1987; XL: 1249–1255. [PubMed: 2445721]
69. Kino T, et al. FK-506, A novel immunosupressant isolated from a *Streptomyces* II. Immunosuppressive effect of FK-506 *in vitro*. J Antibiot (Tokyo). XL:1256–1265.1987; [PubMed: 2445722]
70. Haggarty SJ, et al. Dissecting cellular processes using small molecules: identification of colchicine-like, taxol-like and other small molecules that perturb mitosis. Chem Biol. 2000; 7:275–286. [PubMed: 10780927]
71. Roberge M, et al. High-throughput assay for G2 checkpoint inhibitors and identification of the structurally novel compound isogranulatimide. Cancer Res. 1998; 58:5701–5706. [PubMed: 9865726]
72. Caponigro G, et al. Transdominant genetic analysis of a growth control pathway. Proc Natl Acad Sci USA. 1998; 95:7508–7513. [PubMed: 9636180]
73. Yamaguchi K, et al. 4-Phenylthiazole derivatives inhibit IL-6 secretion in osteoblastic cells and suppress bone weight loss in ovariectomized mice. Bioorg Med Chem Lett. 1999; 9:957–960. [PubMed: 10230619]
74. Lu XP, et al. Novel retinoid-related molecules as apoptosis inducers and effective inhibitors of human lung cancer cells *in vivo*. Nature Med. 1997; 3:686–690. [PubMed: 9176499]
75. Rice JW, Davis JE, Crowl RM, Johnston PA. Development of a high volume screen to identify inhibitors of endothelial cell activation. Anal Biochem. 1996; 241:254–259. [PubMed: 8921195]

76. Goldstein A. The interactions of drugs and plasma proteins. Pharmacol Rev. 1949; 1:102–163.
77. Karlin A, Winnik M. Reduction and specific alkylation of the receptor for acetycholine. Proc Natl Acad Sci USA. 1968; 60:668–674. [PubMed: 5248823]
78. Raftery MA, Hunkapiller MW, Strader CD, Hood LE. Acetylcholine receptor: complex of homologous subunits. Science. 1980; 208:1454–1457. [PubMed: 7384786]
79. Noda M, et al. Primary structure of α-subunit precursor of Torpedo californica acetylcholine receptor deduced from cDNA sequence. Nature. 1982; 299:793–797. [PubMed: 6182472]
80. Kornel L. On the effects and the mechanism of action of corticosteroids in normal and neoplastic target tissues: findings and hypotheses. Acta Endocrinol Suppl (Copenh). 1973; 178:1–45. [PubMed: 4361334]
81. Liu J, et al. Calcineurin is a common target of cyclophilin–cyclosporin A and FKBP–FK506 complexes. Cell. 1991; 66:807–815. [PubMed: 1715244]
82. Brown EJ, Schreiber SL, et al. A mammalian protein targeted by G1-arresting rapamycin–receptor complex. Nature. 1994; 369:756–758. [PubMed: 8008069]
83. Taunton J, Hassig CA, Schreiber SL. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science. 1996; 272:408–411. The authors demonstrate the power of affinity chromatography and biochemical purification to identify the protein target of the natural product trapoxin. [PubMed: 8602529]
84. Sin N, et al. The anti-angiogenic agent fumagillin covalently binds and inhibits the methionine aminopeptidase MetAP-2. Proc Natl Acad Sci USA. 1997; 94:6099–6103. [PubMed: 9177176]
85. Kwon HJ, Owa T, Hassig CA, Shimada J, Schreiber SL. Depudecin induces morphological reversion of transformed fibroblasts via the inhibition of histone deacetylase. Proc Natl Acad Sci USA. 1998; 95:3356–3361. [PubMed: 9520369]
86. Fenteany G, et al. Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. Science. 1995; 268:726–731. [PubMed: 7732382]
87. Sche PP, McKenzie KM, White JD, Austin DJ. Display cloning: functional identification of natural product receptors using cDNA-phage display. Chem Biol. 1999; 6:707–716. [PubMed: 10508685]
88. Hughes TR, et al. Functional discovery via a compendium of expression profiles. Cell. 2000; 102:109–126. [PubMed: 10929718]
89. Giaever G, et al. Genomic profiling of drug sensitivities via induced haploinsufficiency. Nature Genet. 1999; 21:278–283. [PubMed: 10080179]
90. Winzeler EA, et al. Direct allelic variation scanning of the yeast genome. Science. 1998; 281:1194–1197. [PubMed: 9712584]
91. Colas P, Brent R. The impact of two-hybrid and related methods on biotechnology. Trends Biotechnol. 1998; 16:355–363. [PubMed: 9720324]
92. Licitra EJ, Liu JO. A three-hybrid system for detecting small ligand–protein receptor interactions. Proc Natl Acad Sci USA. 1996; 93:12817–12821. [PubMed: 8917502]
93. Marton MJ, et al. Drug target validation and identification of secondary drug target effects using DNA microarrays. Nature Med. 1998; 4:1293–1301. The authors demonstrate the utility of expression profiling for identifying the target(s) of a small molecule. [PubMed: 9809554]
94. Giroux S, et al. Embryonic death of Mek1–deficient mice reveals a role for this kinase in angiogenesis in the labyrinthine region of the placenta. Curr Biol. 1999; 9:369–372. [PubMed: 10209122]
95. Blum JH, Dove SL, Hochschild A, Mekalanos JJ. Isolation of peptide aptamers that inhibit intracellular processes. Proc Natl Acad Sci USA. 2000; 97:2241–2246. [PubMed: 10688899]
96. Adey NB, Kay BK. Identification of calmodulin-binding peptide consensus sequences from a phage-displayed random peptide library. Gene. 1996; 169:133–134. [PubMed: 8635738]
97. Wrighton NC, et al. Small peptides as potent mimetics of the protein hormone erythropoietin. Science. 1996; 273:458–463. [PubMed: 8662529]
98. Wrighton NC, Barrett RW, Dower WJ. Peptide agonist of the thrombopoietin receptor as potent as the natural cytokine. Science. 1997; 276:1696–1699. [PubMed: 9180079]

99. Yayon A, et al. Isolation of peptides that inhibit binding of basic fibroblast growth factor to its receptor from a random phage-epitope library. Proc Natl Acad Sci USA. 1993; 90:10643–10647. [PubMed: 7504274]

100. Yanofsky SD, et al. High affinity type I interleukin 1 receptor antagonists discovered by screening recombinant peptide libraries. Proc Natl Acad Sci USA. 1996; 93:7381–7386. [PubMed: 8693002]

101. Maly DJ, Choong IC, Ellman JA. Combinatorial target-guided ligand assembly: identification of potent subtype-selective c-Src inhibitors. Proc Natl Acad Sci USA. 2000; 97:2419–2924. [PubMed: 10716979]

102. Richter LS, Moos WH. Discovery of nanomolar ligands for 7-transmembrane G-protein-coupled receptors from a diverse *N*-(substituted)glycine peptoid library. J Med Chem. 1994; 37:2678–2685. [PubMed: 8064796]

103. Liang R, et al. Parallel synthesis and screening of a solid phase carbohydrate library. Science. 1996; 274:1520–1522. [PubMed: 8929411]

104. Szardenings AK, et al. Identification of highly selective inhibitors of collagenase-1 from combinatorial libraries of diketopiperazines. J Med Chem. 1999; 42:1348–1357. [PubMed: 10212120]

105. Bunin BA, Ellman JA. A general and expedient method for the solid phase synthesis of 1,4-benzodiazepine derivatives. J Am Chem Soc. 1992; 114:10997–10998.

106. Peisach E, et al. Interaction of a peptidomimetic aminimide inhibitor with elastase. Science. 1995; 269:66–69. [PubMed: 7604279]

107. Bailey N, et al. A convenient procedure for the solution phase preparation of 2- aminothiazole combinatorial libraries. Bioorg Med Chem Lett. 1996; 6:1409–1414.

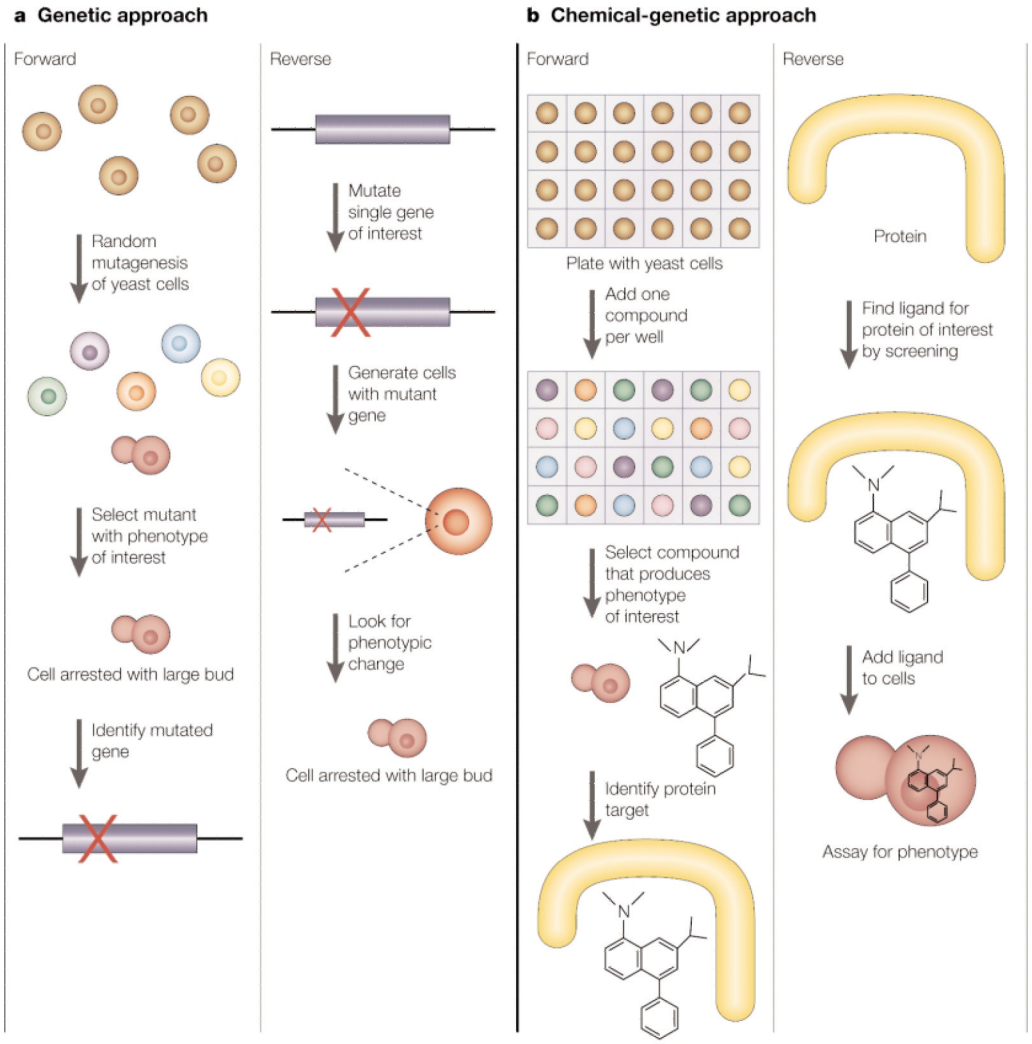


Figure 1. Genetic and chemical-genetic approaches identify genes and proteins, respectively, that regulate biological processes

a | Forward genetics entails introducing random mutations into cells, screening mutant cells for a phenotype of interest and identifying mutated genes in affected cells. In the example shown, yeast cells are randomly mutated, cells showing a large-bud phenotype are selected, and genes mutated in these cells are identified. Reverse genetics entails introducing a mutation into a specific gene of interest and studying the phenotypic consequences of the mutation in a cellular or organismal context. In the example shown, a single mutated gene is introduced into yeast cells and a large-bud phenotype is observed. **b** | Forward chemical-genetics entails screening exogenous ligands in cells, selecting a ligand that induces a phenotype of interest, and identifying the protein target of this ligand. In the example shown, one compound that induces a large-bud phenotype is selected and the protein target of this ligand is subsequently identified. Reverse chemical-genetics entails overexpressing a protein of interest, screening for a ligand for the protein, and using the ligand to determine the phenotypic consequences of altering the function of this protein in a cellular context. In the example shown, a ligand for a specific protein is found to induce a large-bud phenotype.

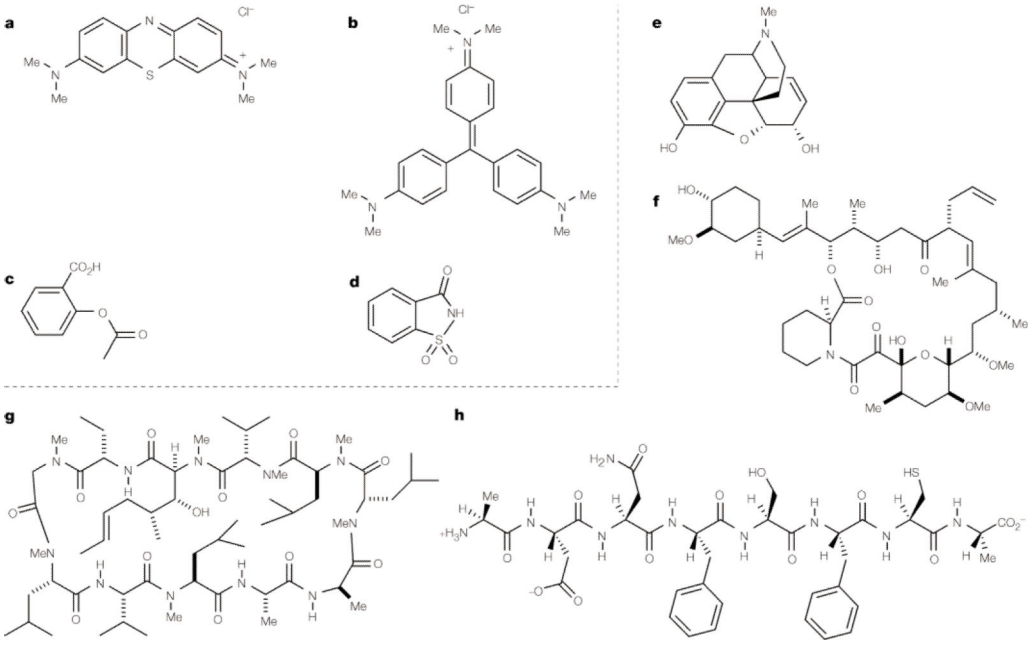


Figure 2. Small molecules have a large range of structural complexity

Simple small molecules, such as **a** | methylene blue, **b** | crystal violet, **c** | aspirin and **d** | saccharin, lack STEREOCENTRES, whereas complex small molecules, such as **e** | morphine, **f** | FK506, **g** | cyclosporin and **h** | peptides, have one or more stereocentres.

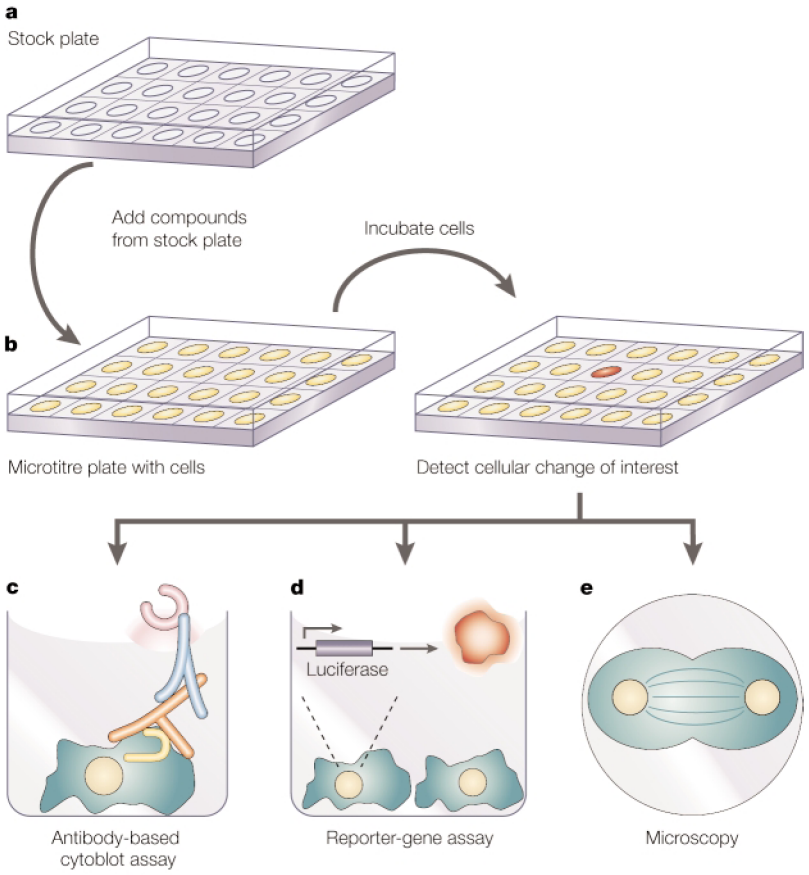


Figure 3. High-throughput cell-based assays

These assays measure the effect of a collection of exogenous ligands on a specific biological process. **a** | To screen a large number of small molecules, a microtitre plate is prepared with the cells of interest. **b** | A small volume of each compound to be tested is added to a well of the microtitre plate. **c–e** | One of several possible methods is used to detect the effect of each compound on the cells. **c** | An antibody detects post-translational or biosynthetic changes within the cell. **d** | A reporter-gene assay detects changes in gene expression at a specific promoter. **e** | Morphological or other changes in the cells are visualised by microscopy. In each case, those reagents that cause the desired cellular change are selected for further study (shown as the red well in **b**).

Table 1

Target-based ligand discovery

Ligand type	Target	Reference
Peptide	Thymidylate synthase	95
Peptide	Calmodulin	96
Peptide	Cyclin-dependent kinase	33
Peptide	Erythropoietin	97
Peptide	Thrombopoietin	98
Peptide	Fibroblast growth factor	99
Peptide	Interleukin-1	100
Small organic molecule	SH3 domains	101
Small organic molecule	Kinases	102
Small organic molecule	G-protein-coupled receptors	103
Small organic molecule	Cytokine receptors	60
Small organic molecule	Lectins	104
Small organic molecule	Proteases	105

Table 2

Combinatorial libraries

Small-molecule class	Reference
Benzodiazepines	105
Peptidomimetic organic compounds	106
Non-peptidic amino-acid polymers	57
Aminothiazoles	107
Shikimic acid derivatives	41
Carbohydrates	103