

Chemical genetics to chemical genomics: small molecules offer big insights

David R. Spring*

Received 20th September 2004

First published as an Advance Article on the web 28th February 2005

DOI: 10.1039/b312875j

Chemical genetics is the study of biological systems using small molecule ('chemical') intervention, instead of only genetic intervention. Cell-permeable and selective small molecules can be used to perturb protein function rapidly, reversibly and conditionally with temporal and quantitative control in any biological system. This *tutorial review* has been written to introduce this emerging field to a broad audience and focuses later on areas of biology where either it has made a significant impact, or it has the potential to do so: signalling, cytoskeleton, development, protein–protein interactions and gene transcription.

Introduction

The exploitation of small molecules in biology has been a part of scientific (and non-scientific) activity throughout human history; for example, the intake of natural products such as opium alkaloids. The twentieth century saw the introduction of pure small molecule drug treatments, such as antibiotics, including penicillin (natural product) and ciprofloxacin (synthetic), and an understanding of the biological basis for their activity. Ever greater degrees of scientific sophistication have led to the systematic discovery of small molecules with specific biological activity in order to probe biological systems. These investigations have revealed new insights into a number of biological processes, and have the potential to impact on all areas of the life sciences. The study of biological systems using small molecule ('chemical') intervention, instead of genetic intervention, has been termed '*chemical genetics*'.¹ This tutorial review will introduce the emerging field of chemical genetics and, hopefully, provide inspiration and encouragement for

further reading and involvement. The field will be described in general terms and then illustrated in detail in five important areas of biology.

Genetics has been used widely to study biology by manipulating the biological system at the level of the gene. A gene commonly is defined as the "nucleic acid sequence that is necessary for the synthesis of a functional polypeptide". Using this definition, genes encode proteins and it is the function of these gene products that we would like to understand. One way to identify the function of a protein is to perturb its function. Genetically, gene function is modulated through a mutation and then the phenotype (physiological effect) is observed. Chemical genetics studies biology by using small molecules to modulate protein function.

Genetics has been divided into forward genetics (involving random mutations followed by phenotypic screening and gene identification) and reverse genetics (involving mutation of a specific gene and phenotype characterisation). So, genetics in the 'forward' direction is from phenotype to gene; in the 'reverse' direction it is from gene to phenotype (Scheme 1). Chemical genetics can be divided similarly. Forward chemical genetics involves the use of small molecules (the 'mutations') to screen for the desired phenotypic effect on the biological system under investigation. Once a suitable small molecule has been identified, the gene product that the small molecule is modulating must be identified. Reverse chemical genetics involves the use of small molecules against a protein (gene product) of interest. Once a binding partner has been chosen it is studied to identify the phenotypic effect of adding the small molecule. So, chemical genetics in the 'forward' direction is from phenotype to protein; in 'reverse' it is from protein to phenotype. In both forward and reverse chemical genetics, the identification of a selective small molecule followed by detailed biological investigation is required.

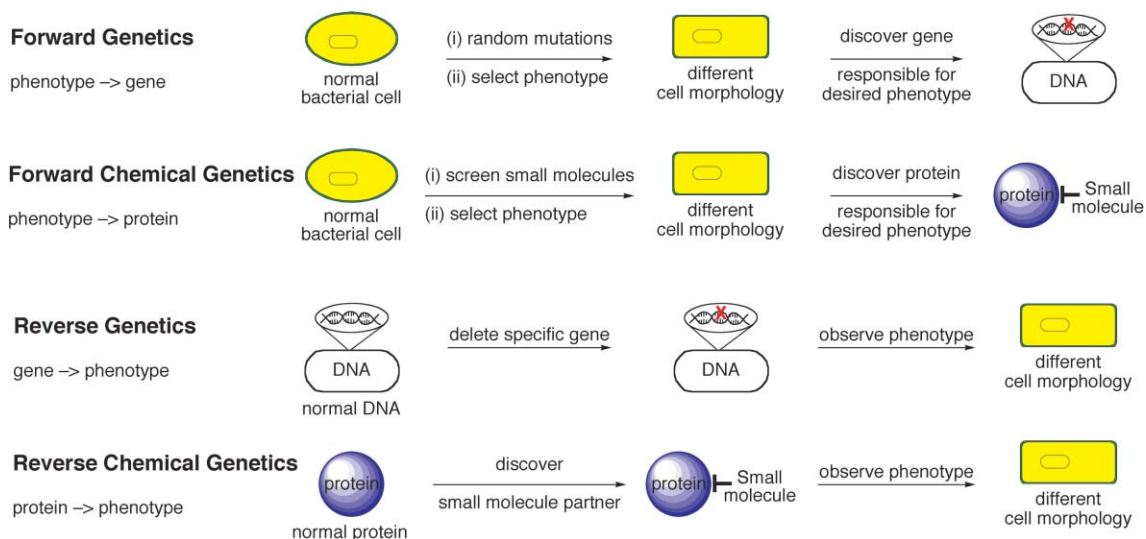
Genetics on a genome-wide scale is known as genomics. '*Chemical genomics*' has been defined loosely in the literature; for the purpose of this tutorial review, chemical genomics will be defined as the systematic search for a selective small molecule modulator for each function of all gene products, *i.e.* the extension of reverse chemical genetics to a genome-wide scale.² This is an enormous challenge! The precise number of human proteins is difficult to predict, but as an estimate if



David Spring

David Spring was born in West Bromwich and attended Oxford University for his undergraduate chemistry degree, graduating in 1995. He stayed at Oxford under the supervision of Sir Jack Baldwin and received his DPhil in 1998 for work on the proposed biosynthesis of the manzamine alkaloids. David then spent two and a half years as a Wellcome Trust postdoctoral fellow and Fulbright scholar at Harvard University with Stuart Schreiber. In 2001 he returned

to the UK as a BBSRC David Phillips Fellow and Fellow of Queens' College at Cambridge, where he is researching diversity-oriented synthesis and chemical genetics.



Scheme 1 Comparing genetics with chemical genetics.

humans have 25,000 genes, there will be at least 10 times that number of different proteins.³ This is because a single gene can give a number of different proteins through alternative splicing and RNA editing of the pre-messenger RNAs, and by post-translational modifications. Moreover, many proteins have more than one function.³ The discovery of perhaps half a million small molecules that are specific for their protein target is daunting. Nevertheless, it is a worthy ambition, and two major initiatives in the US have been set up to achieve this: National Institutes of Health's (NIH) Chemical Genomics Center and the National Cancer Institute's (NCI) Initiative for Chemical Genetics.

Advantages and disadvantages

Why bother with chemical genetics when we have genetic techniques already? There are several advantages of chemical genetics over genetics. For example, small molecules most often induce their biological effect reversibly, due to metabolism or clearing. To do this genetically a conditional allele (gene mutation) is required, such as a temperature sensitive mutation. These are difficult to identify and the pleiotropic (unrelated) effects can be problematic in identifying the effect of modulating the gene product, *e.g.* the heat shock response. In animal models induction of the conditional allele is possible only rarely. All small molecules can be used in a conditional manner; they are added at any time point in the experiment (temporal control). Moreover, the biological effect of small molecules is usually rapid (potentially diffusion-limited), allowing immediate/early effects to be characterised. With a genetic knockout, steady-state effects are observed. Many gene deletions do not lead to the expected effects because of redundancy in the system; small molecule intervention still can provide information of the immediate/early effects. Another advantage is that small molecules can be used to study critical genes at any developmental stage. If a cell line is not viable with a particular gene knockout then chemical genetics still has the potential to study the biological effect of a knockout gene product. Furthermore, multiple chemical genetic 'knockouts'

have the potential to be combined easily. Also, small molecule effects are often tuneable (quantitative control) and therefore can be used to produce 'dose-response' data, where phenotypes are graded by varying the concentration of the small molecule, and this tuning gives greater confidence in the apparent biological effect of a small molecule probe.

Other genetic technologies such as antisense oligonucleotides, short interfering RNA (siRNA) and intracellular ribozymes all work at the level of the gene. Although they can be highly specific modulators, and have their place in understanding biological systems, they have some disadvantages compared to chemical genetics, too. For example, their effects are not as rapid, delivery is a problem as they are not cell-permeable, they neglect the importance of post-translational modifications or the individual functions of protein domains and sub-domains and they do not really confirm whether a protein is a viable drug target.

The main disadvantage of a chemical genetic approach is that, at present, it cannot be applied generally. Any gene, in principle, can be manipulated by genetics; however, chemical genetics requires a selective small molecule ligand to the protein we wish to study. At this point in time only a tiny fraction of known proteins have a ligand partner identified. The current rate of protein-ligand partnership discovery must be increased dramatically if we want to make the chemical genetic approach as generally applicable as genetics, thereby making chemical genomics a reality.

The relationship of chemical genetics to other fields

The intellectual concept of chemical genomics involves small molecule probes used to study every feature of biology at the molecular level. This concept encapsulates a vast area of science. Advanced fields such as medicinal chemistry, pharmacology and the pharmaceutical industry are all concerned with the discovery and biological effects of small molecules, with the ultimate aim of making safe and effective drugs. Many techniques that are required for chemical genetics, such as high-throughput screening, protein-binding assays and

phenotypic assays have been developed and used for many years in the pharmaceutical and agrochemical industries. These fields could be considered as subsets of chemical genomics, since there is significant, but not complete, overlap. The aim of a chemical genetic study is less specific; the aim is to discover biological probes and use them to learn about a particular function of a protein and its biological context. Chemical genetics is not concerned with the extra issues involved in making drugs, such as pharmacokinetics and ADME-Tox (absorption, distribution, metabolism, excretion and toxicity). At present in the pharmaceutical industry, only around 1 in 50 drug discovery projects end up with a drug. The high rate of lead candidate attrition in drug discovery is due to these extra issues involved in preclinical and clinical development, and this is considered to be the most difficult aspect of drug discovery.

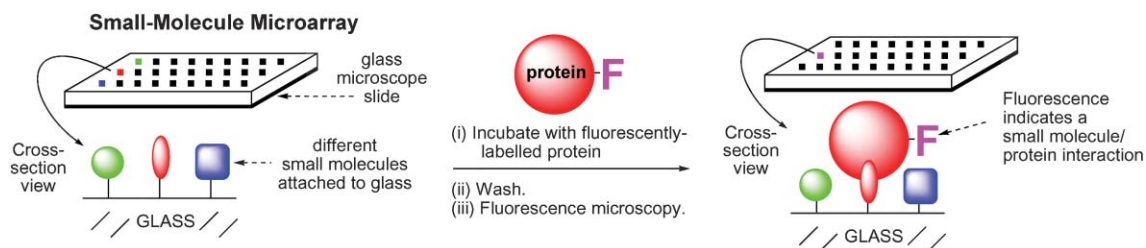
Requirements for chemical genetics

A (forward and reverse) chemical genetic study requires the involvement of at least three things: a selective small molecule, its protein partner, and biological screening. Firstly, a small molecule must be identified, but from where? In a reverse chemical genetic study the protein of interest requires a small molecule partner. The 'rational' design of protein-ligands has been successful increasingly, especially when the protein has a natural small molecule ligand, *e.g.* enzymes and receptors. To work well it requires a good understanding of the macromolecular structure, usually requiring X-ray crystallography or NMR data. The advent of structural genomics (a worldwide initiative aimed at determining a large number of protein structures; see <http://www.rcsb.org/pdb/strucgen.html>) will help this approach; however, the number of small molecules required for chemical genomics cannot be designed on a genome-wide scale. Often the protein of interest will not be an enzyme or a receptor, or even if it is, little will be known about its structure; therefore, how is a small molecule partner identified in these (common) situations? A procedure that has shown great promise is to conduct high-throughput protein-binding screens with a collection of structurally-diverse and structurally-complex small molecules. Structural complexity and diversity are necessary since ligands are required to bind selectively to any gene product. In forward chemical genetics the collection of small molecules can be used in a high-throughput screen to select the small molecule that gives the desired phenotype. Major pharmaceutical companies have proprietary compound collections consisting of around a million small molecules; unfortunately, they are not available to academic laboratories. Academics can obtain structurally diverse compound collections from several sources: commerce, nature, or diversity-oriented synthesis (DOS). Some companies have been set up to sell small compound collections, but they are typically expensive and have been considered as lacking in structural complexity.[†] Nature has provided natural products, which are undoubtedly diverse and complex structurally; however, there are disadvantages with their use. For example, natural products are not available from nature as single compounds (therefore they are screened as mixtures, making it difficult to identify the active constituent), often natural

products are isolated in low abundance, natural products are often so structurally complex that chemical derivatisation is challenging synthetically. DOS involves the efficient, simultaneous synthesis of structurally complex and structurally diverse compounds; however, synthetic strategies are not obvious and remain a significant challenge to modern synthetic chemistry.^{4,5}

Secondly, proteins are required. This requirement can be divided into two significant challenges. For forward chemical genetics, the identification of the small molecule protein partner ('target') is a longstanding challenge. This is discussed below under the heading 'target identification strategies'. If this challenge were not enough, for chemical genomics, whole proteomes, *i.e.* all proteins from an organism, or at least large proportions of proteomes, need to be expressed and purified. Expression and purification of just one correctly-folded, functional and fully-decorated protein can be difficult enough, so to isolate thousands to hundreds of thousands of such proteins is a daunting task. Nevertheless, it is a task that is being addressed by several laboratories and if achieved will be an exceptionally valuable resource for many high-throughput applications including chemical genomics.

Thirdly, biological assays are required to recognise and characterise the small molecule–protein interaction. In forward chemical genetics a collection of small molecules is screened in a phenotypic assay; for example, if a new antibiotic was sought, the assay could be looking for compounds that killed bacterial but not mammalian cells. When an interesting compound is identified, the protein target needs to be discovered, and this might involve several different assays. Of course, the target may not be a protein, which would complicate matters more. In reverse chemical genetics the protein has been pre-selected, and a small molecule binding partner is required, which could be identified through protein-binding screens. A new high-throughput method for the identification of protein-binding partners is the use of small molecule microarrays, which are also sometimes called chemical microarrays (Scheme 2).⁶ Small molecule microarrays are defined as monolithic, flat surfaces that bear a systematic arrangement (spatially-addressable) of probe sites (usually 1000 to 100,000) that each contain a small molecule (*e.g.* peptide, carbohydrate, drug-like molecule, natural product) that is immobilized either covalently or through non-specific adsorption. Instruments are commercially available now for spotting high-density arrays on glass microscope slides. Incubating the microarray probe sites with the protein of interest (labelled with a fluorescent dye, for example), washing and then detecting sites with retained protein (*e.g.* look for fluorescent sites), identifies protein–small molecule interactions. The protein–small molecule interaction is confirmed then by other techniques to verify that the interaction is reproducible and that the small molecule is not binding to the protein label. Alternatively, surface plasmon resonance (SPR) has been used by Graffinity Pharmaceuticals to detect protein–small molecule interactions in a microarray format,⁶ and in this case the protein does not need to be labelled. Once the small molecule has been identified and verified, phenotypic assays are required to characterise the effect of modulating the function of the protein.



Scheme 2 Small molecule microarrays. Different small molecules can be covalently attached to glass slides and probed with fluorescently labelled proteins requiring a small molecule partner. After the slides are washed, to remove non-specific interactions, they can be scanned for spots of fluorescence, indicating a protein–small molecule interaction.

The exploitation of small molecule microarrays for reverse chemical genetics has been utilised for the discovery of the small molecule ‘uretupamine’ (Fig. 1), which binds to the phosphoprotein Ure2p, a central repressor of genes involved in metabolism.⁷ Diversity-oriented synthesis (DOS) was used to generate around four thousand structurally-complex and -diverse small molecules that were printed onto a microscope slide to make a small molecule microarray. This microarray was challenged with fluorescently-labelled Ure2p, and uretupamine was detected. Experiments with uretupamine showed that it affects only a subset of genes controlled by Ure2p, that is to say, it only affects a subset of Ure2p function. This level of detail cannot be replicated by traditional genetics, where *URE2* would simply be deleted, and this example highlights the flexibility of chemical genetics in deciphering the individual functions of multifunctional proteins.

All this research is bound to generate a large amount of data, and will require informatics to make full use of it. PubChem (NIH Chemical Genomics Center) and ChemBank (NCI: chembank.med.harvard.edu) are initiatives to provide scientists with publicly-available databases of, and depositories for, small molecule screening data. These databases should be invaluable in assisting the design of compounds with activity in any biological system of interest.

Target identification strategies^{8,9}

Of particular importance to the success of forward chemical genetics is the target identification step, and it is usually the

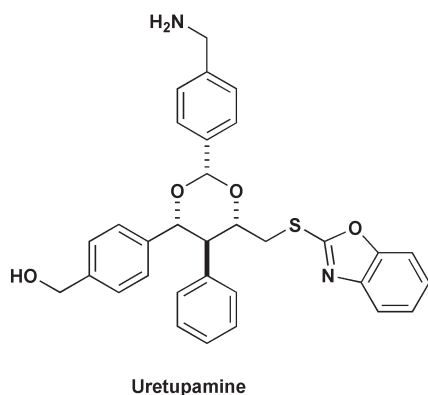
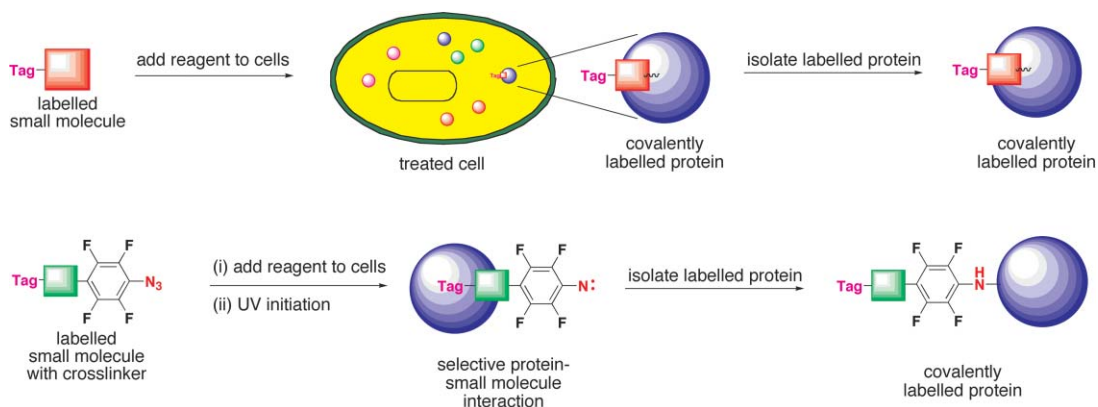


Fig. 1 Uretupamine was discovered from a small molecule microarray as a ligand to the protein Ure2p; the primary alcohol was attached covalently to the glass surface.

rate-determining step. Unfortunately, to date, there is no universal, systematic process to discover the cellular target or mechanism of action of any small molecule. The classical approach to target identification uses the small molecule as a bait in order to trap (label) the protein target. For example, if the protein becomes covalently attached to the tagged small molecule, then it can be purified by following the presence of the tag (*e.g.* radioactivity, fluorescence, biotin). Some compounds become covalently attached to their target proteins as part of their mechanism of action, *e.g.* penicillin, but most do not; therefore, photoactivatable cross-linking groups are often employed, which attach themselves covalently to nearby proteins when exposed to UV light (Scheme 3). Alternatively, affinity chromatography uses the small molecule attached to a solid phase matrix *via* a linker. Elution of protein extracts through a column of the immobilised small molecule, in some cases, retains selectively the protein target. The retained protein can be characterised by mass spectrometry microsequencing and translated back to its gene sequence *via* the genetic code and genomic data. These techniques have been successful in identifying the protein targets of acetylcholine, steroids and natural products such as cyclosporin and rapamycin. These methods require a strong affinity between the protein and small molecule partner, and are often problematic and unreliable. Therefore, newer methods for target identification have been developed, such as drug-westerns, phage display, three-hybrid assays, and protein microarrays. Inevitably, all these methods, like the earlier techniques, require the chemical derivatisation of the small molecule, which is time-consuming, since it is not clear where a tag or linker should be attached onto the small molecule.

Target identification approaches that do not require the small molecule to be derivatised are inherently more attractive. Three general strategies have this advantage: the hypothesis-based approach, profiling techniques and genetic approaches. The first of these requires the greatest degree of chemical and biological insight as it involves studying closely the chemical structure and the phenotypic effect of the small molecule, then inventing several hypotheses regarding its mechanism of action and testing each hypothesis. A notable example of this hypothesis-based approach is the identification of Eg5 as the molecular target of monstrol, which is described in more detail below.¹⁰ If the small molecule or its analogues are already known to have biological effects in other systems then this information should help to choose likely modes of action. Databases of screening data such as PubChem and Chembank



Scheme 3 Biochemical target protein identification. The target protein becomes covalently labelled (tagged) by using labelled small molecules that attach themselves covalently to their protein partner. Covalent crosslinking is involved in the normal mechanism of some small molecules such as penicillin. Other small molecules can be derivatised with chemical crosslinking reagents that have the ability to unmask reactive function groups such as a nitrene, which are capable of sigma bond metathesis with nearby bonds.

will help this approach; however, there are already databases available publicly that can be used to search the literature for chemical structures and their pharmacological activity, such as Scifinder Scholar and Beilstein Crossfire.

Eliminating known targets or unwanted modes of action is a systematic way to prioritise small molecule hits from phenotypic assays. Unlike the pharmaceutical industry, the agrochemical industry screen small molecules directly on their 'patients' (plants, insects and fungi), which usually they are trying to kill; therefore, the forward chemical genetic approach is standard. Small molecules are screened in herbicide, insecticide and fungicide assays simultaneously. Active small molecules (hits) that show relatively selective biocide activity are prioritised at an early stage of any project by screening them in a variety of mode of action assays. These assays are used to uncover small molecules that are operating *via* undesirable modes of action. For example, inhibitors of dihydrofolate reductase (DHFR), which is involved in folate biosynthesis, show up as hits in fungicide, insecticide and herbicide assays, but generally are not selective enough and probably will be toxic to mammalian cells. Although selective DHFR inhibitors are known, such as the antibacterial drug 'trimethoprim', which binds many times more tightly to bacterial DHFR compared to the human enzyme, agrochemicals have to be much more selective. Another undesirable mode of action is disruption (uncoupling) of electron transport and phosphorylation in mitochondria. Uncoupling is commonly encountered with lipophilic weak acids, such as phenols, which can carry protons across the inner mitochondrial membrane. Therefore 'uncouplers' are unlikely to be particularly potent or selective enough. Uncoupling, like DHFR inhibitors, is now widely regarded in the agrochemical industry as an undesirable mode of action; however, there are exceptions such as the commercial fungicide 'fluazinam' that works by an uncoupling mechanism. Another undesirable mode of action is biocide activity *via* the generation of reactive oxygen species, as compounds operating by this mechanism are usually unselective. All these undesirable modes of action (and others) are tested on hits discovered in phenotypic assays in the agrochemical industry. It is relatively easy to identify

small molecules that are operating by a novel mode of action and these compounds are prioritised for further investigation. The identification of the actual mode of action is not easy however, and there are some agrochemicals in use where the precise mode of action is still unknown.

Profiling techniques involve monitoring simultaneously the expression level of genes (transcriptomics) or proteins (proteomics) in an organism. These new technologies can be applied to the problem of target identification. For example, treating cells with a bioactive small molecule may result in changes in the pattern of gene expression. This pattern may reveal clues to the mechanism of action of the small molecule. Moreover, the pattern can be used as a fingerprint and matched with transcriptional profiles for specific gene deletion, if this data is available.

Perhaps ironically, genetic techniques are especially powerful approaches to target identification in forward chemical genetic screens. In one approach, mutant cell lines can be generated, if the system is amenable, that are resistant to the effects of the small molecule. Then the problem becomes identifying the mutant gene product that was responsible for resistance. These can be identified by transfecting random mutant genes into wild-type (sensitive) cells, and selecting for resistance. The transfected gene product responsible for resistance can be sequenced and therefore identified. Another recently described approach involves the integration of chemical genetic and genetic interaction data to make the link between the bioactive small molecule to its cellular target and/or pathway.¹¹ The approach uses a test known as a synthetic lethal screen, which has been used for many years in yeast genetics. The term 'synthetic' in this context has nothing to do with 'organic synthesis' but refers to the combination (synthesis) of two different genetic deletions that result in a lethal phenotype (death). In the classical yeast synthetic lethal screen combinations of non-lethal gene deletions are identified that together cause cell death. This strategy has been extended to relating this data to the treatment of the mutant cells with the small molecule (Scheme 4). In a synthetic lethal genetic interaction two individual gene deletions lead to viable mutants; however, together the double-mutant combination

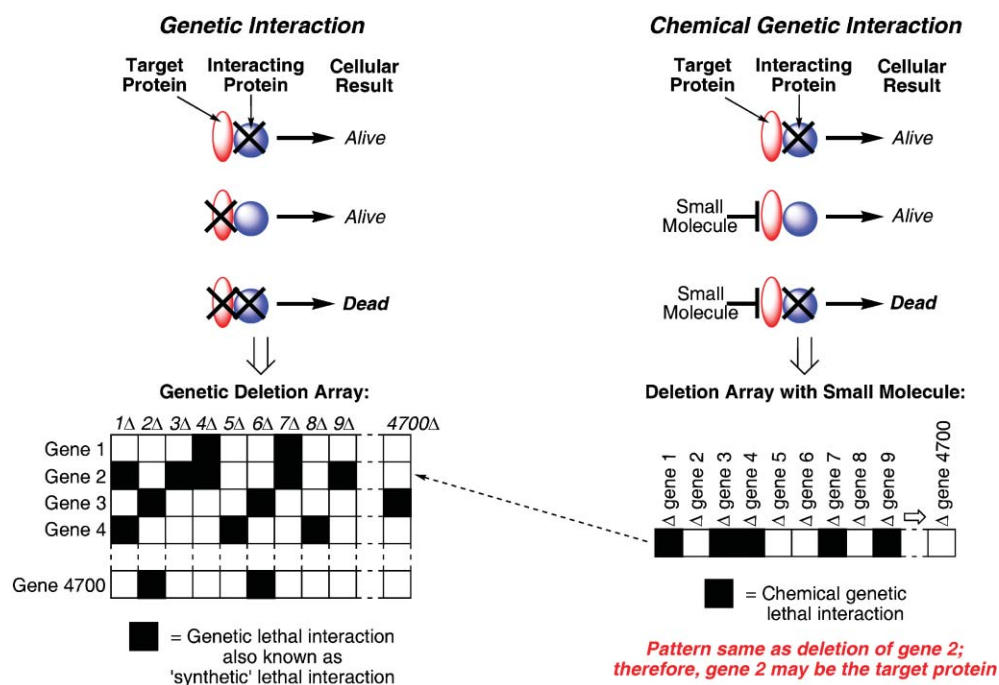
is not viable. In a chemical genetic interaction, the deletion mutant that is missing the gene product of the deleted gene represented by a black X is hypersensitive to a normally sublethal concentration of the small molecule. In this way the target gene product of the small molecule can be identified, by comparison of the small molecule profile with the compendium of synthetic lethal genetic interaction profiles. Matching the closest genetic synthetic lethal profile(s) with that of the small molecule provides the putative target(s), and a possible mechanism of action. Proof-of-principle experiments have confirmed that this technique indeed works, but not perfectly. As mentioned earlier, small molecule interactions are not identical to gene deletions. So, small molecule treatments lead to immediate/early effects of complete, or only partial, loss of gene product function, whereas only steady-state effects of gene deletions are observed. Although the profiles are not identical they are similar, and the process represents a systematic approach to target identification in yeast. Additional merits of the approach include the generation of information of the primary pathways and cellular functions affected by the small molecule treatment, which is especially useful when the compound does not inhibit a specific protein target. In order to extend this approach further it needs to be applicable to higher organisms. Large collections of defined mutants in mammalian (and other) cell lines should be possible using techniques such as RNA interference; therefore, this approach holds great promise for systematic target identification. Even so, biochemical experiments are still required to

validate that the gene product identified actually binds to the small molecule.

Target identification is a key challenge for forward chemical genetics and a systematic way to identify the protein targets of small molecule effects may be a long way off. Complications are increased when the small molecule is not entirely specific and has several protein targets. An illustration of this problem comes from one study involving the specificity of 28 reportedly-specific kinase (enzymes that add phosphate groups to proteins) inhibitors, which showed that almost all had more than one protein target.¹² Another problem is where small molecules have more than one mode of action. Nevertheless, the magnitude of the difficulties involved in target identification only serves to underline the importance of this endeavour.

Case studies

The number of reports of small molecules being used to dissect biological systems has amplified over the last decade. New basic insights into biology have been gleaned from chemical genetic studies from many laboratories worldwide. Below, five areas of biology are highlighted to represent the merits, and to illustrate the potential, of small molecule approaches in the life sciences. Rather than give an exhaustive list of examples, which can be found in other reviews,^{8,9,13–24} only a few, important examples, illustrating the chemical genetic concept, have been focused upon in detail.



Scheme 4 Synthetic lethal screens. Genetic interactions can be similar to chemical genetic interactions. In the synthetic lethal interaction (left), individual deletions of genes (represented by the black X) lead to viable mutants (alive), but double mutants are not viable (dead). In the chemical genetic interaction (right) the deletion mutant is hypersensitive to a normally sub-lethal treatment of the small molecule. A gene deletion that is lethal when cells are treated with the small molecule should also be lethal with a mutation in the compound's target gene. Therefore, comparing the matrix of synthetic lethal interactions of all non-essential genes with the profile from the small molecule treated cells should identify the pathways and targets that the small molecule is modulating.

1: Signalling^{19,20}

Signalling proteins are made up of multiple domains designed to compute and convey biological signals from different inputs. Simply to delete the gene encoding the protein cannot be used to dissect the different functions of such proteins. Protein kinases play an important role in nearly all signalling pathways and are especially difficult to study using traditional genetics, due to redundancy and high homology in their active sites, resulting in functional compensation in gene knockout experiments. The chemical genomic concept of having selective small molecules to modulate every kinase is extremely challenging due to the well-conserved active site. Until this is achieved, an innovative and potentially systematic strategy involving mutating both the kinase and a potent, but non-specific, inhibitor has been highly successful. In this approach the kinase of interest is engineered to have a functionally silent but structurally significant mutation in the active site. For example, replacement of the bulky, 'gatekeeper' residue with a glycine residue, creates a new pocket (hole) in the active site. This new pocket can be exploited by small molecule kinase inhibitor analogues that contain a sterically large substituent (bump) that could fit into the new pocket. These analogues do not inhibit the non-mutated active sites of other kinases, or at least show excellent selectivity, and therefore can be used to dissect the role of the kinase of interest. This technique of generating 'inhibitor-sensitive alleles' has been exploited to probe the function of a kinase called Cdc28 in cell-cycle regulation in yeast.²⁵ A comparison study between a traditional genetic approach and the chemical genetic approach (temperature-sensitive Cdc28 allele *vs.* inhibitor-sensitive allele) was undertaken, and discrepancies were noted. For example, the traditional genetic approach suggested that the most important function of Cdc28 was to control the G1 to S phase transition. The chemical genetic study revealed that the

inhibitor-sensitive Cdc28 mutant arrested at the G2/M transition. The inhibitor had no effect in wild-type (non-mutated) yeast; therefore, the different phenotype was not due to multiple effects of the small molecule inhibitor, but a fundamental difference in how Cdc28 function was altered. Cdc28 functions as a catalyst and as a framework for other components of the cell cycle machinery. The temperature-sensitive Cdc28 allele works by unfolding the protein at elevated temperature, and therefore results in the loss of both functions, whereas the inhibitor only blocks the catalytic activity of the enzyme, and is therefore a more specific probe of protein function.

2: Cytoskeleton^{21,22}

One of the biggest areas of impact that small molecule tools have had in biology is the study of cell morphogenesis and anatomy called the cytoskeleton. Small molecule inhibitors of the cytoskeleton are used commonly in basic cell biology research, and several have been developed into drugs, such as the natural product 'taxol', approved to treat breast and ovarian cancer. Taxol promotes the polymerisation of microtubules. It is used in cell biology experiments for purification of microtubule-associated proteins or microtubule-based motor protein studies *in vitro*, both of which require simulated microtubule polymerisation. There are many natural products that are known to modulate tubulin (*e.g.* colchicine, taxol, vinblastine, vincristine, epothilone, eleutherobin, discodermolide) (Fig. 2) and actin (*e.g.* cytochalasins, jasplakinolide, latrunculins); however, there are many regulatory and structural proteins that construct and coordinate the cytoskeleton, for which there are no known small molecule modulators. Therefore, researchers have begun to screen for compounds that modulate motor proteins, depolymerising proteins, crosslinking proteins, *etc.* In one forward chemical

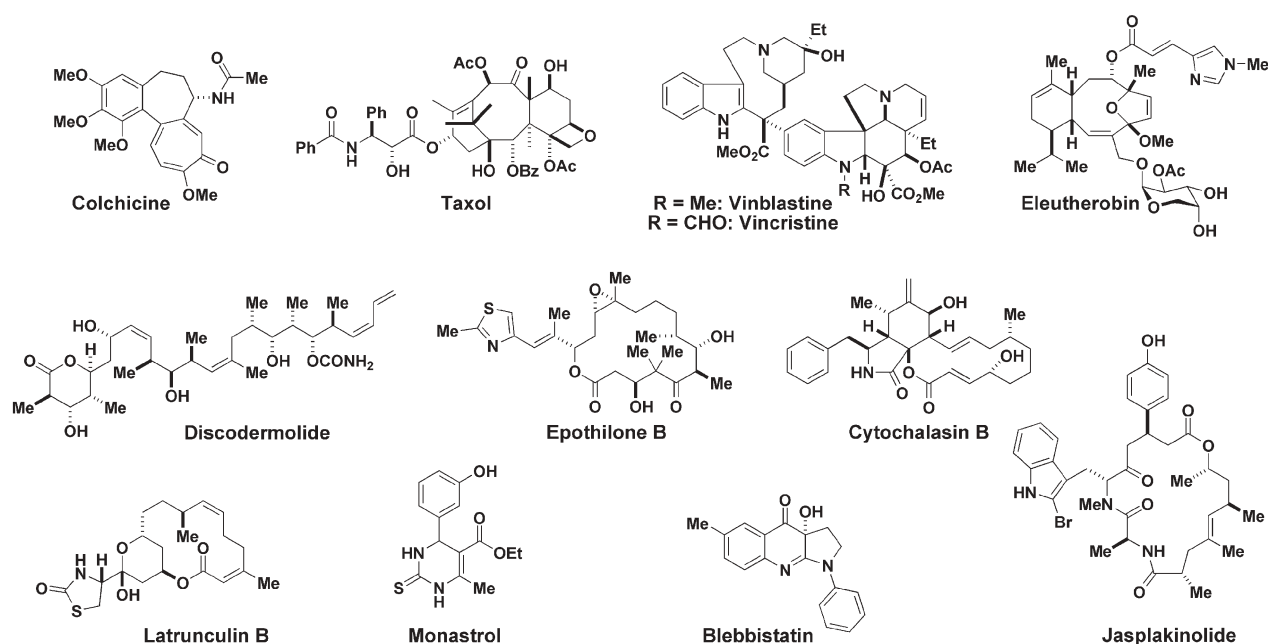


Fig. 2 Cytoskeleton-modulating small molecules.

genetic approach, small molecules were screened for the induction of mitotic arrest in tissue culture cells.⁹ This led to the discovery of the synthetic compound monastrol, which produced a remarkable reorganisation of the mitotic spindle. This monastral phenotype had been observed before on inhibition of the mitotic kinesin protein Eg5 using anti-Eg5 antibodies. Follow up biochemical experiments confirmed that monastrol inhibited the microtubule motility involving Eg5. Removing the compound from cells can reverse the effects of monastrol rapidly; so, washing out the monastrol allowed the cells to move out of mitotic arrest and complete mitosis normally. The reversibility, specificity and cell permeability of monastrol have been exploited to reveal the functions of Eg5. For example, by using monastrol the motor activity of Eg5 was shown not to be necessary for its normal spindle localisation.

Once mitosis (cell division) is complete the daughter cells are separated from one another in a process known as cytokinesis. Cytokinesis takes place on a time scale of minutes, whereby the cell cycle, cytoskeleton and membrane systems of the cell undergo a tightly coordinated series of modifications. The timing, communication, and mechanism of positioning the cleavage furrow to divide the cell into two equal parts were outstanding questions in the field of cytokinesis. In order to address these questions, a forward chemical genetic approach was taken to identify a small molecule that would arrest furrow ingression, but not perturb furrow assembly.²⁶ Non-muscle myosin II was known to provide the power for furrow ingression and therefore this activity was targeted using high-throughput screening. A compound was identified, named 'blebbistatin' because it blocked membrane swelling, known as blebbing, which blocked furrow ingression rapidly and reversibly. Then blebbistatin was used to probe the spatial organisation of the cytokinesis machinery in the absence of furrow contraction and the timing of several specific events in cytokinesis. It showed that exit from the cytokinetic phase of the cell cycle depends on ubiquitin-mediated proteolysis. The chemical genetic approach, using fast-acting and reversible-acting small molecules, is clearly invaluable to probe dynamic cellular processes.

3: Developmental biology²³

A major advantage of using chemical genetic techniques over traditional genetic techniques is that of temporal control, in other words, control over the timing of protein modulation. Since animal development is a sequence of highly regulated events requiring expression of proteins at specific times, places and concentrations, temporal control of protein function in dissecting the sequence and timing of these events is key. Another advantage is quantitative control, *i.e.* being able to add different concentrations of the small molecule in order to produce a phenotype in a dose-dependent manner. A notable example of chemical genetics in the field of developmental biology is demonstrated with zebrafish embryos.²⁷ The zebrafish is a good model vertebrate organism to study developmental biology for many reasons. They possess discrete organs, which are very similar to human organs, and they are transparent in their early stages of life, so that all organ development can be visualised in the living organism. Also,

zebrafish are small for vertebrates, so that during the first few days embryos can be grown in a single well of a 384-well microtitre plate. Moreover, a pair of adults can lay routinely hundreds of fertilised eggs every day, making it possible to screen large numbers of small molecules. Forward genetic screens on zebrafish have identified thousands of mutations that affect the development of every organ. Recently, the forward chemical genetic approach has been investigated and small molecules that modulate specifically the development of the central nervous system, the cardiovascular system, the ear and pigmentation were discovered.²⁸ For example, one small molecule, denoted 31N3 (Fig. 3), was discovered to inhibit specifically otolith development; otoliths are small, bony structures that are attached to bundles of hair cells in the ear. The otoliths move in response to gravity, allowing the zebrafish to maintain balance; 31N3-treated zebrafish swam often on their sides or upside down. Temporal and quantitative control over 31N3 administration was exploited to identify the critical time point for otolith development to be between 14 to 26 hours post-fertilisation, and to establish that each otolith developed independently.

The sonic hedgehog (Shh) pathway in mammals is implicated in many developmental processes including development of midline facial structures and limbs. A naturally occurring teratogen called cyclopamine is present in the Californian corn lily and is an antagonist of the Shh pathway, binding to a transmembrane protein receptor, called 'smoothed', involved in the signalling pathway. Lambs displayed many birth defects if their mothers had eaten this wildflower during pregnancy, the most striking defect being cyclopia, *i.e.* the fusion of two eyes into the middle of the forehead, like the Cyclops. Cyclopamine has been exploited in many developmental studies of the Shh pathway, taking advantage of its superior temporal and quantitative control over genetic methods. In genetic studies of a gene it is useful to have both 'loss-of-function' and 'gain-of-function' mutations. Similarly, in a chemical genetic study of a gene product both agonists and antagonists are valuable for dissecting pathways; therefore, a forward chemical genetic approach was applied to looking for agonists of the Shh pathway.²⁹ The potent agonist identified, labelled Hg-Ag 1.2, was found to be a direct ligand of smoothed; thus, using both cyclopamine and Hg-Ag 1.2 gave a high level of control to understand the details of the Shh signalling pathway.

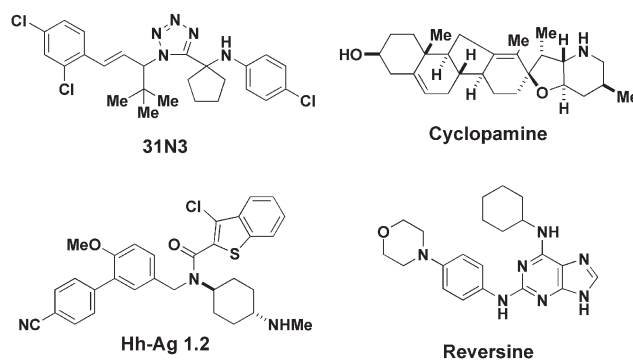


Fig. 3 Development-modulating small molecules.

Stem cells are cells from multicellular organisms that are undifferentiated (unspecialised for a particular function such as a muscle cell or neuron) and capable of giving rise to more cells of the same type indefinitely.³⁰ Given an appropriate signal, stem cells can differentiate into specialised cell types, and this is why stem cells have the potential to be used in the treatment of many human illnesses such as neurodegenerative diseases (e.g. Parkinson's disease), musculoskeletal disease (e.g. muscular dystrophy) and diabetes. Some of the hurdles that need to be overcome before this potential is realised include the controlled proliferation (growth and scale-up) and differentiation of stem cells. A forward chemical genetic approach to the discovery of small molecules that control these events has recently been undertaken. One remarkable example of the power of the chemical approach to biology is the discovery of the small molecule named reversine.³¹ The purine analogue reversine was found to reverse the differentiated state of muscle cells to stem cell-like progenitor cells, which could then re-differentiate into fat cells or bone cells under appropriate conditions. Identifying the molecular target(s) and mechanism of action of reversine will be a significant challenge, but will lead to a giant leap in our understanding of developmental biology.

4: Protein–protein interactions³²

The majority of the clinically used small molecules (drugs) target either an enzyme or a receptor protein. These are the most straightforward proteins to modulate since their function involves binding of a naturally-occurring small molecule signal or substrate. Proteins that undergo reversible conformational changes as part of their function, such as tubulin, appear to be good targets for small molecules, also; however, more rigid, flat protein structures may be less amenable to disruption by small molecules. What has become apparent increasingly with the biomolecular understanding of the cell is that the overwhelming majority of proteins exert their function as members of protein complexes, and therefore protein–protein interactions are extremely important. Since proteins have commonly more than one protein-binding surface, genetic deletion of the protein is not able to dissect the individual roles of each protein–protein interaction. The ability to modulate protein–protein interactions with small molecules offers the possibility of greater understanding of protein function and the possibility of chemotherapeutic treatments of diseases involving aberrant or inappropriate protein–protein interactions. However, there are difficulties in using small molecules to modulate protein–protein interactions. Protein–protein interfaces may lack binding sites for small molecules since they are often flat. The average recognition surface area in protein–protein complexes is 800 Å², which is much larger than the potential binding area of small molecules ($M_r < 1500$ Da) that are orally bio-available. As unrealistic as small molecule modulation may seem initially, there is hope, because only a fraction of the protein–protein interface residues account for the binding energy; these areas are known as 'hot spots'. 'Hot spot residues' tend to be congregated at the centre of a protein–protein interface and are surrounded by residues that serve probably to displace solvent water molecules rather than

strengthen protein complex binding. This phenomenon gives hope to the chemical genetic approach. Indeed, there are many important examples of small molecule modulators of protein–protein interactions, of which the anticancer natural product taxol is perhaps the most famous. As mentioned in the second case study, taxol binds to and stabilises the β -subunit of the tubulin heterodimer resulting in polymerisation to microtubules. The natural products that have proved useful in modulating protein–protein interactions (e.g. taxol, epothilone, eleutherobin, discodermolide all stabilise microtubules) are generally structurally complex, and therefore natural products have been a common source of small molecules to screen for protein–protein interaction modulators. This was the case in the recent discovery of small molecule antagonists of the oncogenic T cell factor (Tcf)/ β -catenin protein complex; disruption of the activation of Tcf-dependent genes by β -catenin represents a potential cancer therapy.³³ However, of the 7000 natural products screened, the most potent small molecule **1** was remarkably simple structurally, being a bicyclic heterocycle with no chiral centres (Fig. 4). Small molecule modulators of protein–protein interactions have most often been targeted in the pharmaceutical industry. For example, inhibition of activated leukocyte function-associated antigen-1 (LFA-1) binding to its intercellular adhesion molecule-1 (ICAM-1) has the potential to inhibit both the inflammatory and immune response. Researchers from Abbott screened drug-like compounds and identified cinnamide **2**, which again has no chiral centres, as a potent antagonist for the LFA-1/ICAM-1 interaction ($IC_{50} = 6$ nM).³⁴ Further investigation of the inhibition mechanism revealed that **2** does not directly inhibit the protein–protein interaction by binding to the protein–protein interface. The likely mode of action is by binding in a pocket of LFA-1, thereby preventing an allosteric change required for LFA-1 to adopt the conformation allowing ICAM-1 to bind. Allosteric regulation of protein–protein interactions avoids the problem of binding a small molecule to a large, flat surface, and may be a more fruitful avenue to explore the importance of protein–protein interaction with small molecules. Nevertheless, small molecule inhibitors of the LFA-1/ICAM-1 interaction have been discovered by designing rationally small molecules that replicate the LFA-1 binding epitope on ICAM-1.³⁵ After several rounds of optimisation compound **3** was discovered

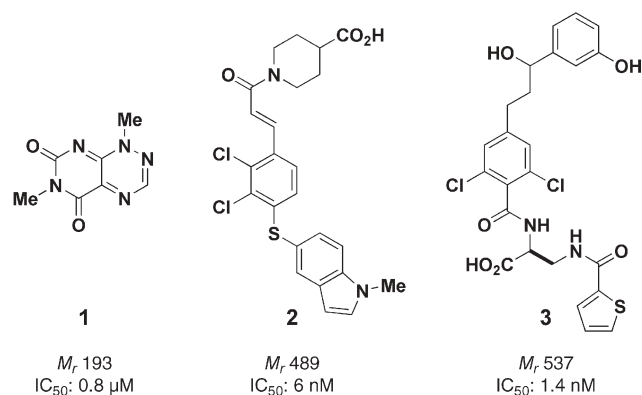


Fig. 4 Protein–protein interaction-modulating small molecules.

to be the most potent small molecule inhibitor of the LFA-1/ICAM-1 interaction known to date ($IC_{50} = 1.4$ nM).

5: Gene transcription

It is usually assumed in chemical genetics that a small molecule interacts with a protein to exert its effect; however, small molecules can interact with any molecular species within a cell. Specific interactions between small molecules and nucleic acid structures such as DNA and RNA are also useful for exploring biology. In fact, mRNA is known to bind directly to certain metabolites and use the small molecule/RNA interaction to control mRNA translation by activating ribozyme activity or changing RNA secondary structure. For example, the mRNA encoding a coenzyme B₁₂ transport protein in the bacterium *Escherichia coli* contains a domain that actually binds coenzyme B₁₂, and the B₁₂/mRNA adopts a secondary structure that is prevented from binding to the ribosome.³⁶ Such RNA domains have been termed riboswitches. To date, prokaryotic riboswitches have been identified that sense thiamine pyrophosphate (TPP), flavine mononucleotide (FMN), *S*-adenosylmethionine (SAM), guanine and adenine.³⁷ This mechanism of genetic control does not involve proteins, but still represents a target for small molecule modulation of gene expression. A different mechanism for controlling gene expression is by binding to specific DNA sequences, and antagonising expression of the gene. This approach has been exploited elegantly by the use of synthetic polyamides made up of *N*-methylimidazole and *N*-methylpyrrole amino acids.³⁸ These cell-permeable probes bind to DNA with an affinity and specificity similar to DNA-binding proteins for the same sequence.

Concluding remarks

Much of our rich knowledge of biology has been gleaned from genetics; however, there are limitations with this approach alone. The chemical genetic approach uses cell-permeable and selective small molecules to perturb protein function rapidly, reversibly and conditionally with temporal and quantitative control in any biological system. Clearly this approach is powerful, and when combined with genetic and other biochemical information gives a more complete understanding of biological processes and disease. The use of natural products such as taxol, colchicine and cyclopamine in biological investigations put this point beyond doubt, and highlights the strength of the chemical genetic approach in studying dynamic processes that are often intractable otherwise. In order to generalise the chemical genetic approach the generation of structurally diverse and complex small molecules and small molecule target identification remain significant challenges, and stress the future importance of synthetic chemistry, especially diversity-oriented synthesis, in realising the full potential of the approach towards chemical genomics. The availability of a systematic chemical genetic approach will encourage all life scientists to exploit the advantages of small molecules, resulting in a much greater appreciation and understanding of biological systems, and, inevitably, this will also lead to better chemotherapeutic treatments.

Acknowledgements

O. Loiseleur and C. Abell are thanked for their insightful comment and suggestions on this article.

David R. Spring*

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, UK CB2 1EW. E-mail: drspring@ch.cam.ac.uk

References

† Large number (> 10,000) compound collections can be bought from companies such as Asinex (asinex.com), Chembridge (chembridge.com), Chemdiv (chemdiv.com). The small molecules are usually small, flat molecules with heteroaromatic groups, *i.e.* “drug-like”. Smaller compound collections of existing drugs off patent can be bought also, for example from Prestwick (prestwickchemical.com) or Sigma (LOPAC and LIGAND-SETS libraries). Other vendors include: Akos, Biofocus, Bionet, Chemstar, Ibs, Interchim, Labotest, Maybridge, Mcl, Mdd, Microsource, Peakdale, Salor, Specs, Timtec and Tripos.

- 1 S. L. Schreiber, *Bioorg. Med. Chem.*, 1998, **6**, 1127–1152.
- 2 G. MacBeath, *Genome Biol.*, 2001, **2**, 2005.1–2005.6.
- 3 D. R. Spring, *Org. Biomol. Chem.*, 2003, **1**, 3867–3870 and references therein.
- 4 C. O'Donovan, R. Apweiler and A. Bairoch, *Trends Biotechnol.*, 2001, **19**, 178–181.
- 5 M. D. Burke and S. L. Schreiber, *Angew. Chem., Int. Ed.*, 2004, **43**, 46–58 and references therein.
- 6 G. Metz, H. Otteleben and D. Vetter, “Small Molecule Screening on Chemical Microarrays”, in *Protein-Ligand Interactions. From Molecular Recognition to Drug Design*, Ed. H.-J. Böhm and G. Schneider, 2003, Wiley-VCH, Weinheim, pp. 213–236; and references therein.
- 7 F. G. Kuruvilla, A. F. Shamji, S. M. Sternson, P. J. Hergenrother and S. L. Schreiber, *Nature*, 2002, **416**, 653–657.
- 8 B. R. Stockwell, *Nat. Rev. Genet.*, 2000, **1**, 116–125.
- 9 G. E. Ward, K. L. Carey and N. J. Westwood, *Cellul. Microbiol.*, 2002, **4**, 471–482.
- 10 T. U. Mayer, T. M. Kapoor, S. J. Haggarty, R. W. King, S. L. Schreiber and T. J. Mitchison, *Science*, 1999, **286**, 971–974.
- 11 A. B. Parsons, R. L. Brost, H. Ding, Z. Li, C. Zhang, B. Sheikh, G. W. Brown, P. M. Kane, T. R. Hughes and C. Boone, *Nat. Biotechnol.*, 2004, **22**, 62–69.
- 12 S. P. Davies, H. Reddy, M. Caivano and P. Cohen, *Biochem. J.*, 2000, **351**, 95–105.
- 13 S. L. Schreiber, *Chem. Eng. News*, 2003, **81**, 51–61.
- 14 C. M. Crews and U. Splittgerber, *Trends Biochem. Sci.*, 1999, **24**, 317–320.
- 15 B. R. Stockwell, *Trends Biotechnol.*, 2000, **18**, 449–455.
- 16 R. S. Lokey, *Curr. Opin. Chem. Biol.*, 2003, **7**, 91–96.
- 17 S. M. Khersonsky and Y.-T. Chang, *ChemBioChem*, 2004, **5**, 903–908.
- 18 K. M. Specht and K. M. Shokat, *Curr. Opin. Cell Biol.*, 2002, **14**, 155–159.
- 19 A. C. Bishop, O. Buzko and K. M. Shokat, *Trends Cell Biol.*, 2001, **11**, 167–172.
- 20 P. J. Alaimo, M. A. Shogren-Knaak and K. M. Shokat, *Curr. Opin. Chem. Biol.*, 2001, **5**, 360–367.
- 21 J. R. Peterson and T. J. Mitchison, *Chem. Biol.*, 2002, **9**, 1275–1285.
- 22 T. U. Mayer, *Trends Cell Biol.*, 2003, **13**, 270–277.
- 23 J.-R. J. Yeh and C. M. Crews, *Dev. Cell*, 2003, **5**, 11–19.
- 24 B. R. Stockwell, *Neuron*, 2002, **36**, 559–562.
- 25 A. C. Bishop, J. A. Ubersax, D. T. Petsch, D. P. Matheos, N. S. Gray, J. Blethrow, E. Shimizu, J. Z. Tsien, P. G. Schultz, M. D. Rose, J. L. Wood, D. O. Morgan and K. M. Shokat, *Nature*, 2000, **407**, 395–401.
- 26 A. F. Straight, A. Cheung, J. Limouze, I. Chen, N. J. Westwood, J. R. Sellers and T. J. Mitchison, *Science*, 2003, **299**, 1743–1747.

-
- 27 C. A. MacRae and R. T. Peterson, *Chem. Biol.*, 2003, **10**, 901–908.
- 28 R. T. Peterson, B. A. Link, J. E. Dowling and S. L. Schreiber, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 12965–12969.
- 29 M. Frank-Kamenetsky, X. M. Zhang, S. Bottega, O. Guicherit, H. Wichterle, H. Dudek, D. Bumcrot, F. Y. Wang, S. Jones, J. Shulok, L. L. Rubin and J. A. Porter, *J. Biol.*, 2002, **1**, 10.
- 30 S. Deng and P. G. Schultz, *Nat. Biotechnol.*, 2004, **22**, 833–840.
- 31 S. Chen, Q. Zhang, X. Wu, P. G. Schultz and S. Ding, *J. Am. Chem. Soc.*, 2004, **126**, 410–411.
- 32 T. Berg, *Angew. Chem., Int. Ed.*, 2003, **42**, 2462–2481 and references therein.
- 33 M. Lepourcelet, Y.-N. P. Chen, D. S. France, H. Wang, P. Crews, F. Petersen, C. Bruseo, A. W. Wood and R. A. Shivdasani, *Cancer Cell*, 2004, **5**, 91–102.
- 34 M. Winn, E. B. Reilly, G. Liu, J. R. Huth, H.-S. Jae, J. Freeman, Z. Pei, Z. Xin, J. Lynch, J. Kester, T. W. von Geldern, S. Leitz, P. DeVries, R. Dickinson, D. Mussatto and G. F. Okasinski, *J. Med. Chem.*, 2001, **44**, 4393–4403.
- 35 T. R. Gadek, D. J. Burdick, R. S. McDowell, M. S. Stanley, J. C. Marsters, Jr., K. J. Paris, D. A. Oare, M. E. Reynolds, C. Ladner, K. A. Zioncheck, W. P. Lee, P. Gribbling, M. S. Dennis, N. J. Skelton, D. B. Tumas, K. R. Clark, S. M. Keating, M. H. Beresini, J. W. Tilley, L. G. Presta and S. C. Bodary, *Science*, 2002, **295**, 1086–1089.
- 36 A. Nahvi, N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown and R. R. Breaker, *Chem. Biol.*, 2002, **9**, 1043–1049.
- 37 R. Micura, *Angew. Chem., Int. Ed.*, 2004, **43**, 4692–4694 and references therein.
- 38 J. M. Gottesfeld, L. Neely, J. W. Trauger, E. E. Baird and P. B. Dervan, *Nature*, 1997, **387**, 202–205.