Chem Soc Rev

Dynamic Article Links

Cite this: *Chem. Soc. Rev.*, 2011, **40**, 4332–4345

www.rsc.org/csr

TUTORIAL REVIEW

Chemical genetics†

Cornelius J. O' Connor, Luca Laraia and David R. Spring*

Received 24th February 2011
DOI: 10.1039/c1cs15053g

Chemical genetics can be defined as the study of biological systems using small molecule tools. Cell permeable and selective small molecules modulate gene product function rapidly, reversibly and can be administered conditionally in either a cellular or organismal context. The small molecule approach provides exacting temporal and quantitative control and is therefore an extremely powerful tool for dissecting biological processes. This *tutorial review* has been written to introduce the subject to a broad audience and highlights recent developments within the field in four key areas of biology: modulating protein–protein interactions, malaria research, hepatitis C virus research, and disrupting RNA interference pathways.

Importance of small molecules

Small molecules are fundamental to our understanding of life; they form critical components of signal transduction cascades and are central to most biological processes. They are used to modulate chemoreceptors in the cell membrane, enabling extracellular entities to enter the cell and they interact with transmembrane proteins to control ion concentration, and therefore pH, inside the cell. They function as secondary messengers between proteins and orchestrate a plethora of biological events in a concentration dependent manner. Steroid hormones (*e.g.* cortisol, testosterone, and progesterone) influence physiological metabolism, by binding to specific DNA sequences and by activating (or suppressing) gene transcription.[1] Other small molecules play important roles in cell–cell signalling in early development, for example, retinoic acid establishes the anterior–posterior axis in vertebrate organogenesis by controlling cell differentiation and another morphogen, thyroxine, induces metamorphosis in tadpoles.[2] A selection of important small molecule signaling molecules is shown below (Fig. 1).

Perhaps not surprisingly, small molecules are also employed by simple prokaryotes in a myriad of processes. Bacteria utilise small molecules as an intercellular signalling mechanism to coordinate gene expression in a cell population density-dependent manner.[3] Cross-kingdom cell signalling (between prokaryotes and eukaryotes) is also possible whereby hormones and

University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge, United Kingdom CB2 1EW. E-mail: spring@ch.cam.ac.uk
† Part of the themed issue on small molecules in biology.

Cornelius J. O' Connor

Cornelius (Con) O' Connor grew up in Carlow, in the south east of Ireland. He obtained his BA (Mod.) in Chemistry in 2004 from Trinity College Dublin where he also undertook his PhD under the supervision of Dr Mike Southern, studying nucleoside mimetics and developing synthetic methodology. He subsequently joined the Connon Group at Trinity, where he worked on the development of novel catalytic systems. Con then moved to the University of Cambridge where he is currently a member of the Spring Group, researching the use of small molecules to modulate protein–protein interactions, diversity-oriented synthesis and chemical genetics.

Luca Laraia

Luca Laraia was born in Rome and grew up in Vienna, Austria. He completed his MSci at Imperial College London before starting his PhD on the Cancer Research UK medicinal chemistry training programme at the University of Cambridge. Under the supervision of Dr David Spring and Prof. Ashok Venkitaraman he is looking at disrupting protein–protein interactions relevant for anti-cancer therapy.

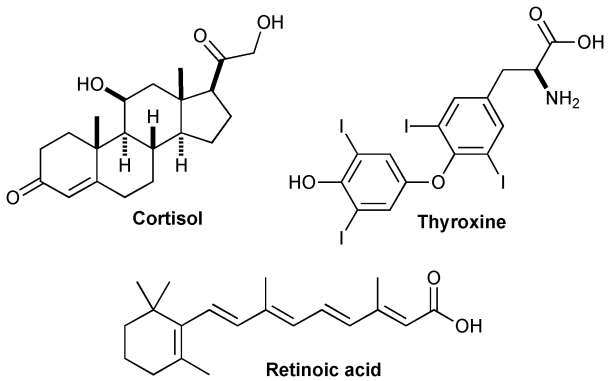


Fig. 1 Some small molecule modulators of protein function.

hormone-like small molecules are used to communicate between microorganisms and their host.[4] In addition, some of the same compounds are used in very similar roles across hugely differing life forms, which further highlights their significance, particularly when considered in an evolutionary context.

Put simply, the importance of bioactive small molecules cannot be over-emphasised. Of course, this is not a new discovery; natural products derived from plant extracts have been harnessed by mankind for centuries as both therapeutics and stimulants. This trend continues today with the majority of drugs currently approved by the FDA being natural products or derivatives thereof.[5]

Despite humanity's significant dependence on bioactive small molecules, it was not until the latter half of the twentieth century, when the field of molecular biology was born, that any detailed understanding of the cellular targets and mechanism of action of these small molecules was discovered. Ever increasing degrees of scientific sophistication have led to the systematic discovery of small molecules with specific biological activity. Chemistry has been described as the *lingua franca* of medical and biological sciences,[6] but today the 'druggable' genome (the number of macromolecules affecting disease states amenable to small molecule modulation) comprises approximately 500 genes of more than 20 000 genes, which code for $>100\,000$ proteins.[7] A significant challenge for chemical biologists is to expand the druggable genome. In order to accomplish this, the development of novel systems for interrogating biological processes is essential. The study of biological systems using small molecule (or 'chemical') intervention instead of genetic intervention has been termed '*chemical genetics*'.[8–10] This tutorial review builds on a previous report from this group,[9] which, in addition to providing a brief introduction to the field of chemical genetics, aims to provide an account of how the subject has grown in recent years and key developments will be illustrated in four important case study examples.

David R. Spring

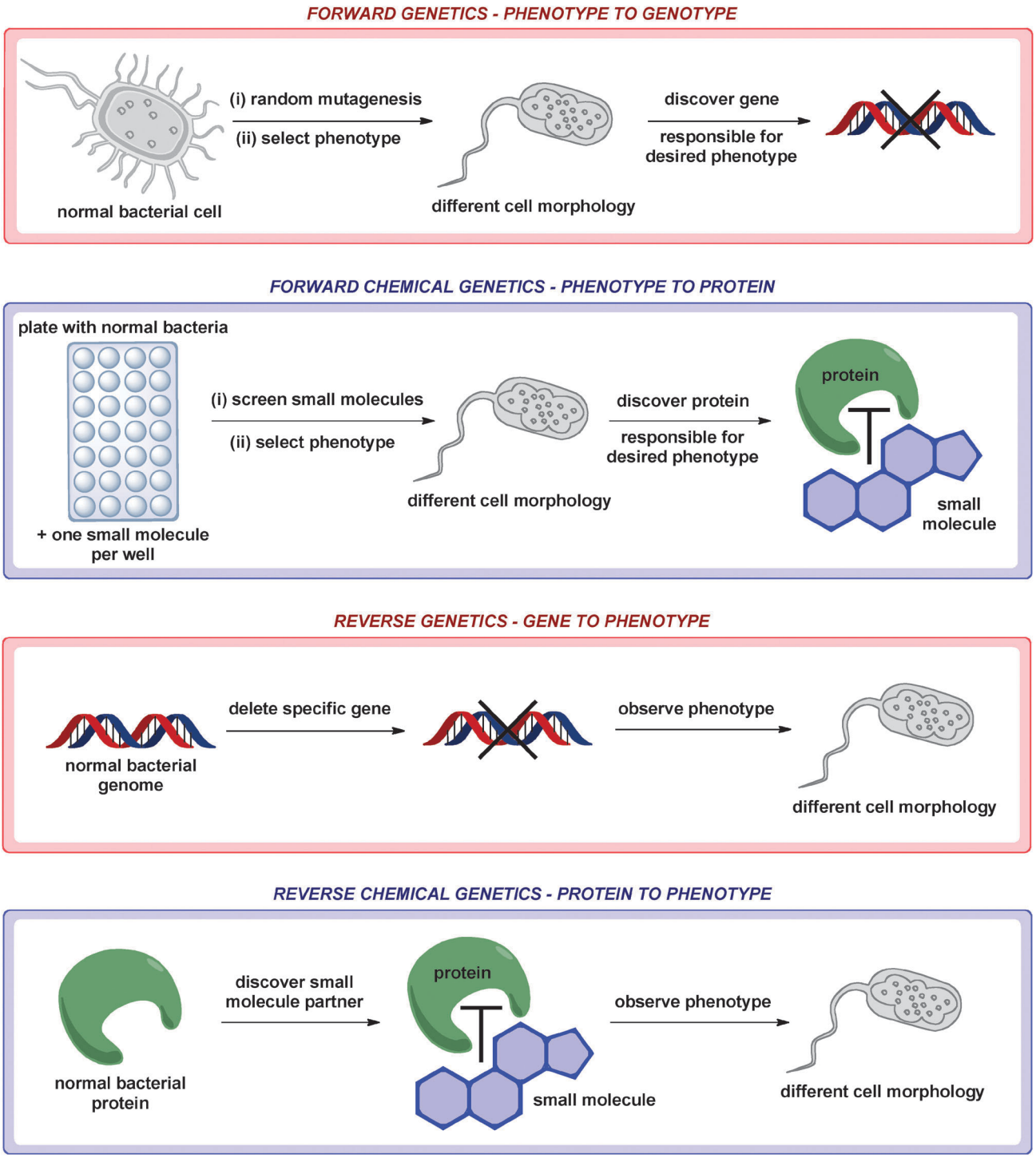
David Spring is currently a Senior Lecturer at the University of Cambridge within the Chemistry Department. He received his DPhil (1998) at Oxford University under Sir Jack Baldwin. He then worked as a Wellcome Trust Postdoctoral Fellow at Harvard University with Stuart Schreiber (1999–2001), after which he joined the faculty at the University of Cambridge. His research programme is focused on the use of chemistry to explore biology.

Chemical genetics

Chemical genetics is the study of biological systems using small molecule tools. As in classical genetics, which uses mutagenesis to investigate the relationship between genes and phenotypes (physiological effects), chemical genetics can be divided into two approaches: forward and reverse (Scheme 1).[10] Classical forward genetics aims to identify the genes responsible for a particular phenotype of interest—it operates from phenotype to genotype. Random mutagenesis of cells or organisms, introduced using DNA damaging agents or radiation, produces mutant strains which are subsequently screened for a phenotype of interest (*e.g.* different cell morphology); the gene responsible for the phenotype of interest can then be identified through mutation mapping.[9,10] Forward chemical genetics uses small molecules to modulate gene-product function; compounds that induce a phenotype of interest are selected and then the protein which they target must be identified.[11] Target identification gives an insight into the role of the gene and/or the function of the protein. Classical reverse genetics operates from genotype to phenotype. Firstly, the gene of interest must be selected; secondly, the gene must be manipulated to produce mutant progeny and finally, the phenotypic differences between the mutant and the wild-type cells or organisms must be identified to discover the function of the gene of interest.[11] Classical reverse genetics has become a very powerful tool for studying disease and the development of transgenic, knock-out, knock-in and conditional allele animal models has revolutionised our understanding of disease states. Reverse chemical genetics similarly involves a specific protein (or gene product) that is screened with libraries of small molecules to identify ligands which perturb its function. Once an appropriate protein binding partner is identified, it is introduced into a cell or organism and the resulting phenotypic changes are studied; the ligand is used to mimic the effects of a genetic mutation.

Advantages and disadvantages

The use of small molecules to study complex biological systems offers several advantages over traditional genetics; chemical genetics will have its greatest impact in the study of systems which cannot be interrogated using a classical genetic approach.[12] Despite the awesome power of classical genetics, particularly for revealing the molecular processes and pathways in lower organisms, its application in higher organisms, especially mammals, can be difficult due to their diploid genome, physical size, prolonged gestation periods and slow



Scheme 1 Comparing genetics with chemical genetics.

rate of reproduction.[11] Conversely, the chemical genetic approach can be readily applied in either cellular or organismal contexts. In addition, small molecules often induce their biological effects reversibly, due to metabolism or clearing and the phenotypic effects can be monitored quickly.[9] In order to achieve reversibility using classical techniques, conditional alleles are used, such as temperature sensitive (ts) or cold sensitive (cs) mutations.[13] This approach works well for lower organisms, however, pleiotropic effects such as the heat shock response can be problematic, making the interpretation of gene product modulation difficult. The induction of conditional alleles is rarely possible in animal models. The small molecule approach is conditional; ligands can be added or removed, and the induction of their effects is usually rapid, this allows for direct *in vivo* kinetic analysis, which cannot be accomplished using classical techniques.[10] This type of temporal control enables the characterisation of protein modulation in a time and dose dependent manner.

Chemical genetics is used to study gene products, and thereby the function of genes. Multi-functional proteins can possess several binding partners; therefore, selective ligand binding can modulate individual functions. Classical genetic knockouts delete the protein from the organism and therefore remove the possibility of studying distinct effects. Also, when a gene codes for more than one protein, the small molecule approach can provide specific protein binders, so the effects of modulating a single protein are observed rather than shutting down an entire pathway or removing several proteins from the system. In addition, knockout of essential genes results in lethality, whereas sub-lethal dosing of exogenous ligands provides a partial knockout phenotype.[14] By administering varying concentrations of the small molecule binder, changes in the phenotype can be graded and dose response data can be obtained. Chemical knock-in alleles are also possible whereby exogenous ligands stimulate rather than suppress protein function.[14]

Antisense oligonucleotides, interfering RNA (RNAi) and intracellular ribozymes can also be used as alternatives to conditional alleles, however, they do not provide the exacting temporal control that the small molecule approach affords, and in some instances, particularly in mammalian cells, their application can result in non-specific gene expression and off-target effects.[15] One of the main difficulties associated with the advancement of the RNAi approach in drug discovery programmes is the development of safe, specific, systemic delivery systems, although considerable progress in this area has been made in recent years.[16]

The main disadvantage of the 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 for the gene product of interest. Although there are examples where the specificity of chemical ligands approaches that of a classical gene knockout, finding highly specific chemical entities, which do not produce off-target effects, thus enabling, the definition of specific protein functions remains a significant challenge.[8,9,12,14]

Currently, only a fraction of proteins have known ligand partners. In order to increase the discovery of specific protein–ligand partners, greater exploration of biologically relevant chemical space is required. Traditional combinatorial library screening has had a disappointing hit return ratio, and it is now widely accepted that this poor rate of return is due to a lack of structural diversity within the libraries. Pharmaceutical compound collections have, in the past, rigidly conformed to predefined criteria, such as Lipinski's 'rule of five', as an indication of oral bioavailability, and are fundamentally biased towards known bioactive chemical space.[17] Despite the vast size of these libraries, lack of diversity has resulted in a dramatic decline in the discovery of new drugs. In order to make the chemical genetic approach as generally applicable as classical genetics, the current rate of protein–ligand partnership discovery must be increased dramatically; improvements in high-throughput screening coupled with the design and synthesis of more structurally diverse compound collections (diversity-oriented synthesis) could fulfil its promise and achieve the goal of finding 'a small molecule partner for every gene product'.[8]

Relationship to other fields

In this, the 'omic' era, chemical genetics on a genome-wide scale is referred to as chemical genomics, which essentially is the systematic search for small molecule modulators for each function of every gene product.[9] This is an enormous challenge, requiring tremendous advancements in screening techniques, target identification, and current approaches to small molecule compound assembly; but as highlighted in the sections to follow, significant progress has been made in these areas.

Chemical genetics is a subsection of chemical biology and as such it operates at the interface of several research fields. However, it is important to note that chemical genetics is not drug discovery. The aim of chemical genetics is to interrogate biological systems using small molecule probes. These probes are not drugs,[7] nor are they, by definition proof-of-concept compounds for drug discovery. Therefore, they do not have to conform to certain physicochemical properties, or adhere to particular health and safety requirements. Many toxicity issues can be ignored, as these compounds are not intended for human use *per se*. This implies that certain chemical motifs which are usually avoided in drug discovery programmes, such as Michael acceptors, or potential alkylating agents can be included in chemical genetic screens; however, caution must be exercised when interpreting assay results as with increased reactivity there comes an increased propensity for off-target effects.

Despite this fundamental difference, chemical genetics is very closely linked with drug discovery. The tools used by the pharmaceutical industry for drug development, *e.g.* high throughput screening, protein binding assays, phenotypic assays, biophysical analysis *etc.*, are all used in chemical genetics. However, chemical genetics is not solely the pursuit of big pharmaceutical companies; it can be very successfully researched in an academic setting.[18] In fact, it is immensely important that chemical genetics remains an active area of research in academia. The spiralling cost of drug discovery has reshaped the pipelines of almost all pharmaceutical companies, with an emphasis now being placed on discovering therapeutics for highly prevalent diseases, with a market place large enough to gain a return on investment. Rather alarmingly, a significant number of pharmaceutical companies have moved away from antibiotic research, because the cost of finding new antibiotics with novel modes of action is seen as too great, especially considering the disturbing rate of resistance development. It is therefore vital that academic research continues to challenge what is considered druggable.[19] Dissecting biological systems, with small molecule probes, will provide greater insight into their complexities at a molecular level, and invariably lead to the discovery of new drug targets.

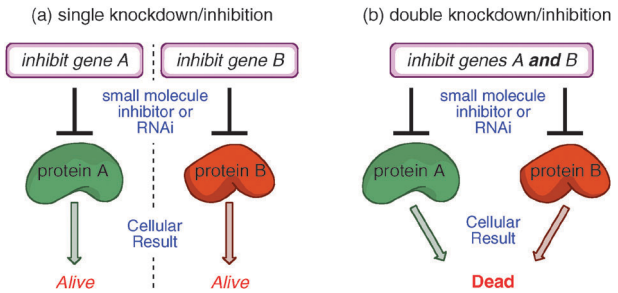
In the last decade, the availability of small molecule compound collections has dramatically increased, *via* the accessibility of commercial libraries and public repositories. Advancements in screening approaches and synthetic techniques mean that it is now possible to synthesise and screen literally tens of thousands of compounds in an academic setting.[20] What once was solely an industrial process, the practice of accruing large compound collections for screening

is now commonplace in large academic institutions. Several schemes have been set up to facilitate academic research and to accelerate breakthroughs in drug development, the most notable of which are the National Institutes of Health (NIH) Chemical Genomics Center and the National Cancer Institute's (NCI) Initiative for Chemical Genetics, both US initiatives.

Combination chemical genetics

In addition to supplying compounds for drug discovery efforts and identifying chemical probes for basic biology research, chemical genetics can be used in combination with traditional genetics to discover novel targets for combination therapies. This is particularly efficacious in the context of cancer.[21] In classical genetics two genes are said to be synthetically lethal if individual mutation corresponds with viability but simultaneous mutation results in lethality.[21] However, the effects of a selective small molecule modulator on gene product function are generally concentration dependent, with the transition from inactivity to activity being dependent on the level of dosage.[22] If a second perturber is introduced to the system its effect can be synergistic or antagonistic. Therefore, once a specific small molecule modulator of a cancer target has been discovered, it can be used in a synthetic lethality screen with small interfering RNAs (siRNAs) of other cancer associated genes. This is referred to as a combination chemical genetic synthetic lethality screen (see Scheme 2); it uses small molecules at a sub-lethal concentration, and observes which gene product elimination by siRNAs restores the lethal effect. It can then be envisaged that a small molecule targeting the new protein can be used in conjunction with the original hit for a more powerful response. This approach identified synthetic lethality between the Breast Cancer 2 susceptibility protein (BRCA2) and Poly (ADP–ribose) polymerase (PARP), two proteins involved in different forms of DNA repair. Tumours which are BRCA2 deficient are highly susceptible to PARP inhibitors, highlighting the strength of this approach.[23] Such screens can also be conducted in a purely chemical genetic setting, where combinations of small molecules (rather than small molecules and siRNAs) are screened.[24] These approaches have been successful in identifying compounds suitable for administration in combination regimens with significant therapeutic benefit to cancer patients.

With the ever-increasing impact of interdisciplinary research, it is beyond doubt that chemical genetics will coalesce with other fields of research and add a significant value to the work already being conducted. The field of chemical biology is expanding rapidly; continued collaborative efforts between chemists and biologists will result in greater understanding of the intricacies of biological complexity at a molecular level; this will inevitably result in better understanding of disease states and improved drug treatment programmes for patients.



Scheme 2 Synthetic lethality: (a) no desired phenotype with single knockdown/inhibition; (b) desired phenotype with double knockdown/inhibition

Screening tools for chemical genetics

Both forward and reverse chemical genetics require robust and reliable assays to detect the effect of a small molecule on a biological system. However there are substantial differences between the assay systems in both approaches. Reverse chemical genetic screens aim to identify small molecules that interact specifically with a desired protein, RNA or DNA. Such assays are usually conducted in an *in vitro* setting, where a direct binding or displacement event can be detected. One of the main requirements for these screens is the availability of a purified protein or RNA/DNA. Obtaining pure protein can be a great challenge in itself, and many research labs invest a significant amount of time and money in this endeavour.[25] Once this challenge has been surmounted, an appropriate screen must be devised. A plethora of screening techniques and assays have been devised, and the technique of choice will be dependent on library size, resource availability, expertise and the type of target being investigated.

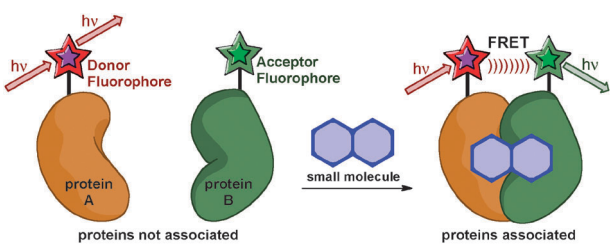
In recent years, library sizes in both industry and academia have increased dramatically, though now the emphasis is set on the quality of the library as well as the quantity. It has been recognised that different target types require different compound selection. For example, compounds that will target kinases may not be appropriate for proteases. With the expansion of the druggable genome more diverse libraries appropriate for different sets of targets are required. Diversity oriented synthesis and biologically oriented synthesis were borne out of this requirement.[17] In addition to synthetic small molecule libraries, natural products are a great source of diversity. Whilst a disproportionately large percentage of drugs originate from natural products, their complexity often hinders their development. However many remain excellent tools in chemical genetics and natural product inspired libraries which combine complexity with relative ease of synthesis are becoming increasingly popular.[26]

If the resources for assembling such a large and diverse library are missing but appropriate structural information of the target is available, a more cost effective screening method such as virtual screening could be used. In this approach, large commercial libraries can be screened for predicted interaction with the target (docking), or for similarity with a known ligand of the target (ligand based virtual screening). Alternatively, models based on the surface properties of the target can be developed, and ligands that contain complementary properties will be selected (pharmacophore screening). Once commercial hits have been identified based on the ranking, a percentage of the highest scoring molecules can be purchased. The ease of use and widespread availability of these programmes means that almost any researcher can attempt this approach; however,

one should be warned that without appropriate knowledge and experience, the results can be difficult to interpret. The best results are achieved when several crystal structures of the same target are available, and when a range of techniques are applied. Computational approaches are also commonly used to filter compound libraries for predicted physiochemical properties, selecting compounds to be screened in primary assays.[27]

Once compounds have been selected, primary and secondary *in vitro* assays are typically devised. Currently, the most widely selected primary assay types are those that produce a quantifiable response with a fluorescent read-out as a result of the small molecule disrupting an interaction or enhancing/reducing a reaction. An example of this is fluorescence polarisation. Here a fluorescently labelled substrate (either a peptide or a molecule such as ADP/ATP) differentially reflects polarised light depending on its state. For example, polarised light will remain such when reflected off a fluorescently labelled peptide bound to a protein. However, if the peptide is unbound (displaced by a small molecule), it will reflect light in an unpolarised fashion. This type of assay has become especially favoured when screening for inhibitors of protein–protein interactions, due to low reagent requirement and simplicity of the read-out. Other fluorescence based screening tools include Förster resonance energy transfer (FRET). Two fluorophores (a donor and an acceptor) are appropriately placed either on the same macromolecule or on different molecules, the main requirement being that an active small molecule brings the fluorophores close enough to allow energy transfer or increases their distance to reduce energy transfer upon excitation. The utility of this assay has been demonstrated in the measurement of molecules that can stabilise heterodimeric proteins (Scheme 3). When the two proteins are sufficiently close, FRET is observed. An excellent review on available screening assays for different protein classes is available for more detailed information.[28] Once hit compounds are selected, these must be validated by further primary assays (similar to those described), but also more complex secondary assays.

Secondary assays in reverse chemical genetics should be orthogonal to the primary assay (not measure the same property) and are typically lower throughput, and more protein expensive. The most popular techniques include biophysical measurements that detect the binding event between a small molecule and its protein partner.[29] These can be used to extract binding constants (K_i) and thermodynamic parameters (Gibbs Free Energy, enthalpy, and entropy of the interaction) as well serving as an additional ranking tool for compound development. An example of this is isothermal titration calorimetry (ITC), which measures the heat change required to keep a solution at a constant temperature when a small molecule is titrated against its protein partner. An additional example is nuclear magnetic resonance (NMR). A series of different NMR experiments can be used to determine thermodynamic parameters and as the technique is very sensitive, it can be effectively used to assess weakly binding molecules.



Scheme 3 Förster Resonance Energy Transfer (FRET) for a small molecule stabiliser of a dimeric protein complex.

Forward chemical genetic screens are usually conducted in mammalian cell lines or whole organisms such as yeast, worms (*C. elegans*), plants or zebrafish. A lot of the same principles apply for forward screens as for reverse screens. However libraries tend to be selected based on expected cell permeability rather than target specificity, as the target is unknown. The read-out of the primary screens can also vary from simple organism viability (is the compound lethal) to more complex signalling pathway effects. These screens have increased in their throughput, just like reverse chemical genetic screens, due to the advent of miniaturisation in plate size and reagent requirement for an observable effect. Additionally, technological advances in automated microscopy have also improved phenotypic screens in terms of quality and breadth of read-out.

Stockwell and co-workers adopted a forward chemical genetics approach to identify small molecule inhibitors of apoptosis (programmed cell death) in cells harboring a mutant huntingtin protein.[30] Mutations in this protein are common in patients with Alzheimer's and Huntington's disease and have been shown to cause protein misfolding. This leads cells to apoptose by a previously unknown mechanism. Small molecules were screened for their ability to rescue cell viability in a cell line over-expressing mutant huntingtin by using Alamar blue fluorescence (a compound reduced by a functional mitochondrion). The target of a small molecule hit was identified as protein disulfide isomerase (PDI), which was shown to have a role in apoptosis by causing permeabilisation of the mitochondrial membrane. An exciting example of a forward chemical genetic approach in a whole organism was the discovery of neuroactive small molecules in zebrafish using a behavioural read-out.[31] The response of the zebrafish to a light stimulus in the presence of known neuroactive compounds was compared to the response in the presence of novel small molecule probes to predict likely target areas. For example, STRs (Fig. 2) were identified as novel acetylcholinesterase inhibitors, the latter being a promising therapy for Alzheimer's and dementia patients. Well characterised plants have also been used for whole organism screens. Traditionally, general pesticides were screened for; however, more specific modulators of function are now being sought. For example, a chemical genetic screen was undertaken in *Arabidopsis thaliana* to identify molecules that stopped strigolactone accumulation.[32] The strigolactones are a family of plant hormones involved in inhibiting the growth of lateral buds (meristems), which regulate branching patterns. Whilst this is a normal process required in several plants, strigolactones also stimulate germination of several parasitic plant species which are damaging to crops. CTLs (Fig. 2) were identified as stopping germination, and pathways that they, as well as strigolactones, regulate

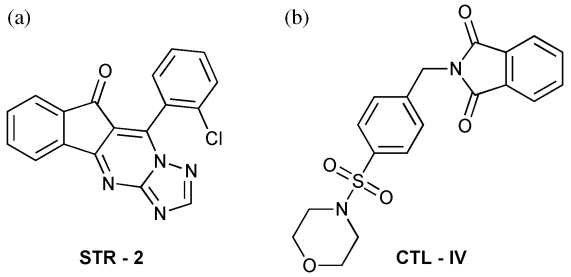


Fig. 2 (a) STR-2, a probable acetylcholinesterase inhibitor, (b) CTL-IV, a modulator of strigolactone levels.

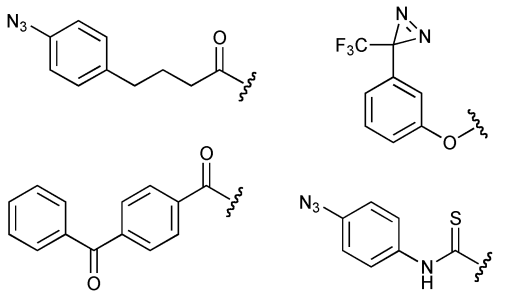


Fig. 3 Reactive groups for differential photoimmobilisation.

were also identified. These tools are proving increasingly important in commercial agriculture to increase crop yields and target pests more effectively.

These examples outline just a few of the many possible forward chemical genetic screens. The possibilities are almost endless and provided that a suitable read-out can be identified, screens for almost any phenotype can be designed. The clear advantage of forward chemical genetic screens is the ability to obtain a cell permeable molecule with the desired phenotype. However a formidable challenge for these systems lies in identifying the specific target(s) of a small molecule, a process which will be discussed in the following section.

Target identification[33]

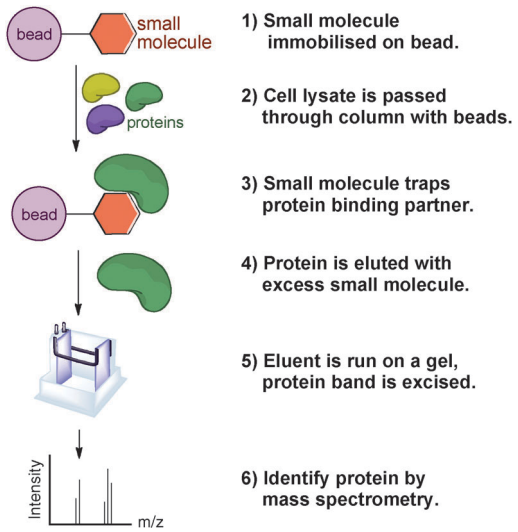
A key aspect of forward chemical genetics is the target identification (ID) stage. This is crucial in determining the mode of action of small molecules identified from a phenotypic screen. Several different techniques have been developed to achieve this, and although no single systematic solution exists, a few key approaches have been developed. The two most advanced and widely used include affinity-based and genetic based approaches.

Affinity based techniques are often the first port of call for target ID, as they have been extensively studied and many variations exist. In general, immobilisation of the small molecule is required, which has been successfully carried out on glass slides, slides for microarrays, magnetic beads, and beads for affinity chromatography. Small molecules can also be tagged with fluorous tags or biotin which bind strongly to fluorous or streptavidin coated beads, respectively, for non-covalent immobilisation.[34] This can all be achieved by a wide variety of different chemistries, the discussion of which is beyond the scope of this review.[35] Immobilisation can in itself be problematic, because it is important that it does not affect the small molecule binding to its molecular target. Photochemical immobilisation is one of the most promising recent developments to address this issue. Whilst the technique itself is not new, a recent application has utilised a wide variety of different photochemical labels to immobilise small molecules to a solid support (Fig. 3).[36] Dilly *et al.*[36] were able to show that by using molecules with different photochemical reactivities they could increase the probability of finding an immobilisation position that did not interfere with the binding of abscisic acid to its target.

Once the small molecule has been immobilised, the most widely exploited target ID strategy is affinity chromatography. This involves passing cell lysates or protein extracts through a column of beads with immobilised small molecules. Proteins that bind very tightly to the small molecule will be retained, whilst those that do not will be washed away. A different buffer system then allows the bound proteins to be removed, separated by SDS-PAGE and characterised by mass spectrometry completing the target ID process. The main disadvantages are the inability to isolate proteins of low binding affinity, due to the stringent conditions of the final washing steps, and a lack of sensitivity in the MS read-out. These can be circumvented by eluting with an excess of the non-immobilised small molecules or alternatively by covalently cross-linking the protein to the small molecule, and by using stable isotope labelling by amino acids in cell culture (SILAC) to improve the MS read-out. The latter involves culturing cells with ^{13}C and ^{15}N labelled lysine and arginine, which are incorporated in proteins. Identified proteins that show an enrichment in the 'heavy' state are classed as *bona-fide* targets, whereas those where the ratio is approximately 1 are not.[37] The approach was tested on a small amount of kinase inhibitors and identified a target for immunosuppressant small molecules. Affinity chromatography based techniques have also been successful in identifying the targets of rapamycin, as well as identifying additional targets of clinical kinase inhibitors.[38] Many variations of affinity-based target ID strategies exist, and two comprehensive reviews highlight a wide range of examples in more detail.[34,39] A general outline of the affinity based approach is shown in Scheme 4.

Alternative strategies for target ID which also require immobilisation steps (whether of small molecules or proteins) have been devised over the years, including phage display, yeast three hybrid assays and protein microarrays.[22] However, these have been less utilised, as chemists and biologists have sought out new strategies that do not require modification of the bioactive small molecule and have the key advantage of not interfering with any section of the molecule that may be responsible for its bioactivity. Additionally, this would be more time-efficient as identifying a successful immobilisation strategy would not be required. Ironically, *genetic* techniques can play an important role in this, as the boundaries between the latter and *chemical* genetics continue to blur.[22]

Yeast-based systems offer new possibilities for genetic based target ID. Recently, Butcher *et al.*[40] described a novel protocol involving yeast strains over-expressing a library of target proteins. The resistance of the yeast to a small molecule for yeast strains with different overexpression plasmids was used as a simple measure of whether the particular target was being



Scheme 4 Affinity chromatography.

modulated by a small molecule.[22] Conversely, as diploid organisms have two copies of each gene, the removal of one allele and reduction in the dosage of a single gene from two copies to one copy result in the production of a heterozygote that is sensitised to any small molecule that acts on the product of that gene. The identification of small molecule targets using this approach is known as haploinsufficiency profiling (HIP).[41] The overexpression and HIP approaches are complementary and have been combined to offer an increasingly accurate mode of target ID.[40] Yeast-based approaches have become increasingly popular and reliable, and the growing number of reviews on the subject highlights its importance.[33,42] Despite these advantages, there are still problems with characterising a compound destined to study human pathways using yeast. Certain targets may not be conserved between humans and yeast or may be distinctly different. Therefore genetic approaches have also been extended to mammalian cell lines. This field is still in its infancy and is not yet able to provide definitive identification of a compound's target; however, it has been very successful in mapping small molecules into clusters of similar phenotypic effects, suggesting common pathways of action.

As well as the affinity-based and genetic techniques which have been developed and refined over the last decade, a recent paper by Lomenick *et al*. describes a completely novel immobilisation-free target ID strategy.[43] Drug affinity responsive target stability (DARTS) involves incubation of a small molecule with a cell lysate followed by proteolysis and separation by gel electrophoresis. The technique exploits the propensity of a protein bound to a small molecule to be *less* susceptible to proteolysis. Therefore the protein target(s) of the small molecule should be the only one not hydrolysed. This has been successfully employed in a proof-of-concept study and to identify a novel potential target for resveratrol, an active constituent of grapes suggested to increase longevity. Although target ID still remains a significant challenge in forward chemical genetics, the examples outlined give us a positive outlook. With the increasing sophistication of such experiments, target ID will hopefully be less challenging in the future.

Case studies

The use of small molecules is becoming increasingly popular as a powerful approach to dissect their role in individual molecular interactions within biological systems. Below, four important areas of biology are highlighted to represent the merits and to illustrate the potential of the chemical genetic approach. Rather than present an exhaustive list of examples, only a few important examples, illustrating the awesome power of the chemical genetic concept, have been focused upon in detail.

Case study 1: small molecule modulation of protein–protein interactions (PPIs)[44]

Most synthetic small molecule modulators of biological systems have been designed to act on single proteins with a well-defined binding pocket or active site. However in recent years there has been increasing interest in disrupting interactions between two proteins to expand the druggable genome beyond more traditional targets such as kinases and G-Protein coupled receptors. Importantly, a small molecule can be used to inhibit a very specific function of a protein through its interaction with another protein in contrast to siRNA knockdown, which abolishes all protein activity. Therefore using small molecule modulators of PPIs as a more specific approach will provide invaluable information for dissecting and manipulating biological pathways. However despite the clear advantages of modulating PPIs with small molecules, they have classically been described as undruggable targets, due to the relatively 'flat' and large surface area at the protein interface, covering up to 3000 Å.[2] Designing small molecules to cover this area may appear to be a daunting task, requiring relatively large 'small' molecules with few possibilities of making strong interactions with the protein of interest.

Fortunately, over the last decade this notion has almost completely been dispelled, due to the identification of 'hot-spots' on the interaction surfaces of proteins.[45] These, often very hydrophobic regions, contribute a large degree of binding affinity between the proteins without covering the whole surface area of the interaction. Several examples have shown that small molecules can be designed to target these hot-spots without the need to increase the molecule size too far above what is considered drug-like.[46] The identification of these sites is crucial for identifying small molecule modulators of PPIs, and can be achieved through a series of mutational techniques including alanine-scanning.[47] This technique sequentially replaces all residues involved in the interaction with alanines and changes in binding energy can be observed using biophysical techniques such as isothermal titration calorimetry (ITC).

To date, several PPIs have been successfully targeted by small molecules; two of the earliest and most studied examples will be described here. The first is the interaction between the tumour suppressor p53 and human double minute 2 (HDM2), which is crucial for the regulation of pro-apoptotic signals. As HDM2 inhibits p53 function and is overexpressed in a large percentage of cancer patients, inhibiting this interaction has become of great importance for clinical application. Since the initial discovery of the nutlins by Roche[48] several distinct small molecule inhibitors of this interaction have been

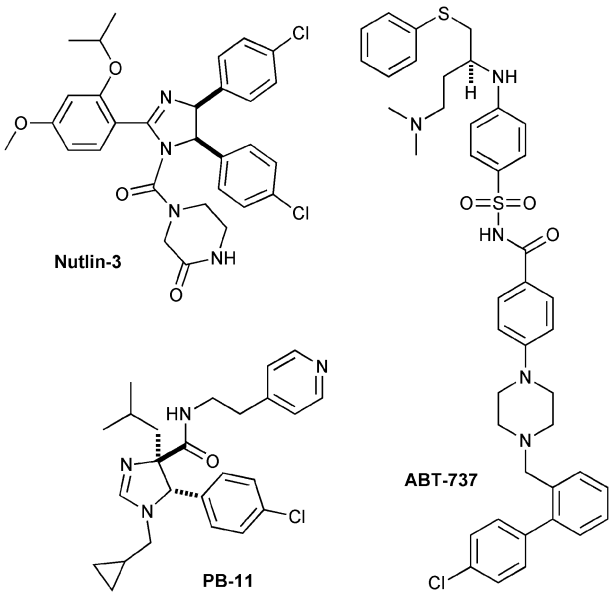


Fig. 4 Small molecule modulators of PPIs.

discovered (Fig. 4). These have been identified by HTS,[49] *in silico* screening,[50] as well as peptidomimetic approaches.[51] Following initial identification of hits, several *in vitro* and biophysical experiments were undertaken to confirm the molecules, such as PB-11, as *bona-fide* hits. One of the key factors to the success of these projects was the availability of high resolution X-ray crystal structures of bound inhibitors. This allowed the exact binding site and mode to be determined as well as constructive SAR to be formed. All of the inhibitors were shown to bind to the predicted 'hot-spot' on the target protein, highlighting the importance of identifying these sites.

In a second example, one of the successful screening approaches used was to start from very small molecules known as fragments.[52] These are typically <300 Da in molecular weight and display a high ligand efficiency. Fragment hits can either be grown, *i.e.* expanded to fit a binding pocket, or linked to contain two differentially binding fragments. This has been used successfully to find activators of the Bcl-x_L and Bcl-2 PPI. These structurally similar pro-apoptotic proteins are activated by binding to an α-helical segment of the Bid and Bad proteins, termed the BH3 helix.

Two hit fragments were combined by parallel synthesis affording ABT-737. Despite its relatively large size for a small molecule (893 Da), this compound still displays good oral bioavailability and reasonable ligand efficiency, particularly when compared to the original BH3-helix binding partner (Fig. 4).

Case study 2: *Plasmodium falciparum*

Malaria is a devastating disease estimated to infect 300–500 million people and cause almost 1 million deaths each year.[53] The most deadly form of the disease is caused by the pathogen *Plasmodium falciparum*. To date, several programmes have been established to prevent or reduce the spread of malaria by eliminating the disease vector and by treatment with chemotherapy. The artemisinins (sesquiterpene lactone endoperoxides) represent the only remaining effective antimalarial. Vaccine development has not yet been successful and with the increasing emergence of drug resistant *P. falciparum* strains, clearly, new therapies with novel mechanisms of action or unique resistance profiles are seriously needed.

P. falciparum has a complex life cycle involving both mosquito and human hosts.[54] A bite from the female *Anopheline* mosquito inoculates the human host with infective sporozoites. (These are spores formed after sexual reproduction of the protozoan parasite within the mosquito gut.) The sporozoites subsequently invade liver cells (hepatocytes) and each sporozoite can develop into thousands of merozoites within the hepatocyte. Merozoites, when released from the liver, infect red blood cells (erythrocytes) where asexual reproduction occurs and it is at this point that the disease becomes symptomatic.[54,55] The parasite differentiates and multiplies inside a membrane bound vacuole known as a parasitophorous vacuole, which is established upon invasion of the erythrocyte (Fig. 5). Over 48 hours approximately 20 daughter merozoites are formed per mature parasite, which must rupture their host cell, a process known as *egress*, in order to infect new erythrocytes. The blood stage of malarial infection is when most associated pathology is experienced including: fever, hypoglycaemia and renal failure.[55] In recent years considerable progress has been made to elucidate how the parasite invades the human host red blood cells[56] but in comparison relatively little is known of the processes by which the parasite is released after replication has occurred.

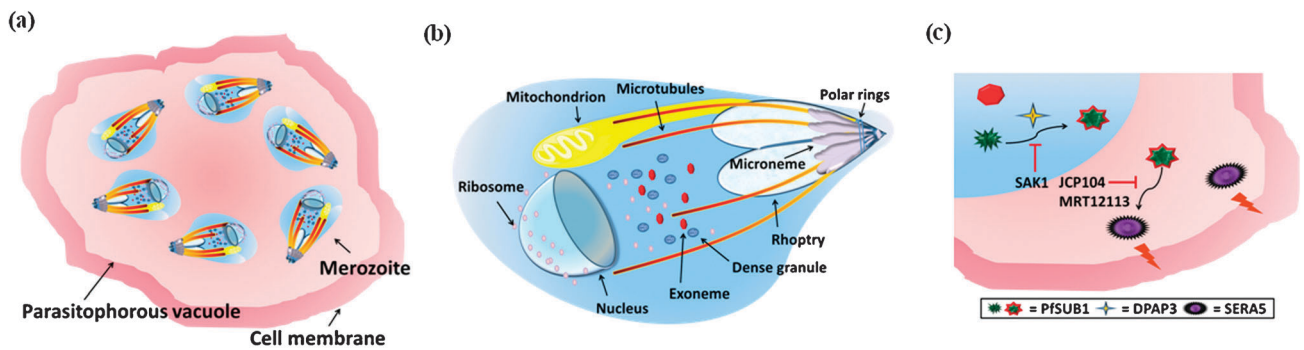


Fig. 5 Parasite egress from host red blood cell. (a) Merozoite accumulation in erythrocyte; (b) structure of merozoite; (c) protease pathways leading to degradation of the parasitophorous vacuole and host cell including chemical perturbagens.

In 2007, Yeoh *et al.*[57] reported that immediately prior to egress, an essential serine protease called PfSUB1 (a member of the subtilisin-family of proteases) is discharged from a previously unrecognised parasite organelle called exonemes into the parasitophorous vacuole.[58] PfSUB1 interacts with the papain-like serine repeat antigens (SERAs) which results in rupture of the host cell and the release of the infective merozoites. Using transfection based genetic approaches Yeoh *et al.*[57] demonstrated the essential nature of the proteolytic activity of PfSUB1. The authors developed a multiwell fluorescence assay of PfSUB1 activity using a fluorogenic rhodamine labelled peptide substrate and carried out a high throughput screen of over 170 000 small molecules obtained from a number of commercial and proprietary sources, from which they isolated a selective inhibitor of PfSUB1, termed MRT12113 (Fig. 6).[57] The small molecule protease inhibitor was subsequently employed in chemical knockdown studies. Addition of MRT12113 to cultured parasites resulted in partial inhibition of schizont rupture and reinvasion by those merozoites which were released was considerably reduced. Blackman and co-workers later confirmed that in addition to the role of PfSUB1 in host cell rupture it also plays a role in merozoite reinvasion of healthy erythrocytes.[59]

A related but independent study was carried out by Arastu-Kapur *et al.*[60] who adopted a forward chemical genetics approach and screened a library of over 1200 serine and cysteine covalent protease inhibitors to discover compounds which block host cell rupture by *P. falciparum*. Arastu-Kapur *et al.*[60] identified one serine protease inhibitor (JCP104) and by lead optimisation of two non-specific cysteine protease inhibitors, the authors developed a specific dose-dependent cysteine protease inhibitor (SAK1) of parasite egress.[58] The covalent binding mode of the small molecules used simplified target identification for these compounds. PfSUB1 was identified as the target of JCP104 and the cysteine protease inhibitors targeted members of the dipeptide peptidase (DPAP) family of proteins (DPAP1 and DPAP3). Structure–activity relationship studies produced SAK1 a specific DPAP3 inhibitor. Yeoh *et al.*[57] discovered that inhibition of PfSUB1 was associated with a build-up of a member of the serine repeat antigens (SERA) family, SERA5, which implied that PfSUB1 is involved in SERA5 processing which in turn leads to parasite egress.[58] This prompted Arastu-Kapur *et al.*[60] to monitor the effects of JCP104 and SAK1 on SERA5 levels within the parasitophorous vacuole and they too observed decreased SERA5 processing. Cessation of the proteolytic processing of SERA5 directly correlated to a block in parasite release from the erythrocyte. The findings of Arastu-Kapur *et al.*[60] strongly support the theory that PfSUB1 processing of SERA5 triggers pathways leading to host cell rupture (Fig. 5c).

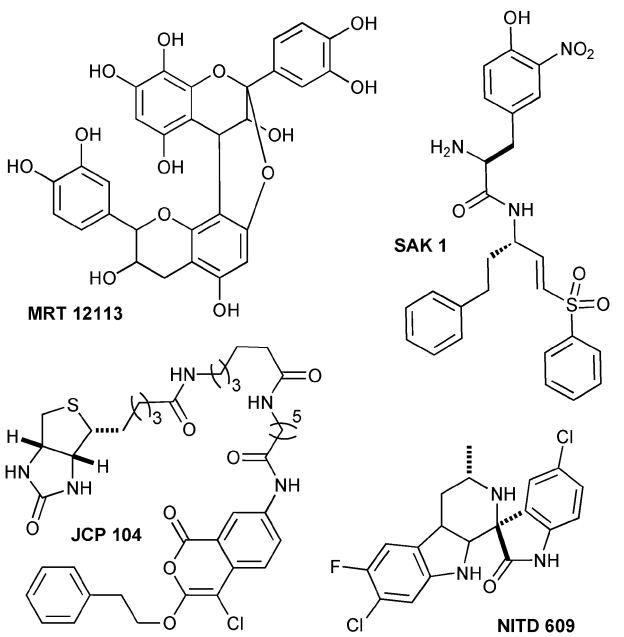


Fig. 6 Inhibitors of *P. falciparum* merozoite egress.

The confirmation that PfSUB1 acts on SERA5 which subsequently leads to egress, combined with Blackman and colleagues'[59] findings that PfSUB1 primes proteins on the surface of merozoites for reinvasion prior to egress, has identified PfSUB1 as a possible drug target.

The high mutation rate of *Plasmodium* has resulted in reduced efficacy of the current repertoire of antimalarials. PfSUB1 has multiple protein substrates each of which is processed in a substrate specific manner, Blackman and co-workers therefore hypothesise that anti-PfSUB1 drugs are less likely to be rendered ineffective by parasite resistance as the simultaneous co-evolution of both protease and substrate protein would be required.[59]

Over the last few years, three reports are of particular note which demonstrate the power and elegance of chemical genetics in malaria research and drug development; all three projects involved large forward chemical genetics screening.[61–63] A team of researchers from GlaxoSmithKline[61] invigorated the global antimalarial research effort, when they created and screened a library consisting of nearly 2 million compounds, of which over 8000 compounds displayed inhibitory activity against the multi-drug resistant Dd2 strain. The antimalarial activity and associated structures of these compounds have now been made public in an effort to encourage further research and kick-start anti-malarial drug development campaigns.

In a similar study, Guiguemde *et al.*[62] screened almost 310 000 compounds in phenotypic assays of *P. falciparum*; 172 compounds were subjected to detailed profiling and a reverse chemical genetic study. Nineteen new inhibitors of validated malaria drug targets were discovered. In addition, a panel of *P. falciparum* strains with acquired resistance to known antimalarial chemotherapeutics were tested for sensitivity to cross-validated hit compounds from the reverse chemical genetic screen; similar potencies ($EC_{50} \leq$ 3-fold) were observed for approximately 34% of the compounds, possibly indicating novel modes of action. The authors carried out preliminary pharmacokinetic and pharmacodynamics studies on a focused group of lead compounds and found them to be suitable for further development. The *in vivo* efficacy of one compound was determined in a murine mouse model for malaria, though it proved 25-fold less potent than chloroquine in the same model.

Rottmann *et al.*[63] using similar *Plasmodium* whole cell assays to those mentioned above screened a focused library of about 12 000 compounds of natural and synthetic origin.

Primary hits were then screened for activity against multi-drug resistant strains and any compounds with cytotoxic effects against mammalian cells were discarded. Pharmacokinetic and physical studies were carried out on 17 compounds of which one class of spiroindolone compounds stood out as appropriate candidates for further development. Optimisation studies produced NITD609 (Fig. 6) which rapidly suppressed protein processing by inhibiting the gene that encodes the P-type cation transporter ATPase-4 (PfATP4). Developing mutant *Plasmodium* strains with resistance to NITD609 proved difficult and no cross resistance to other antimalarials including artmenisinin and mefloquine was observed. The mechanism of action of NITD609 is distinct from that of existing antimalarials and it is currently undergoing pre-clinical trials with Novartis.

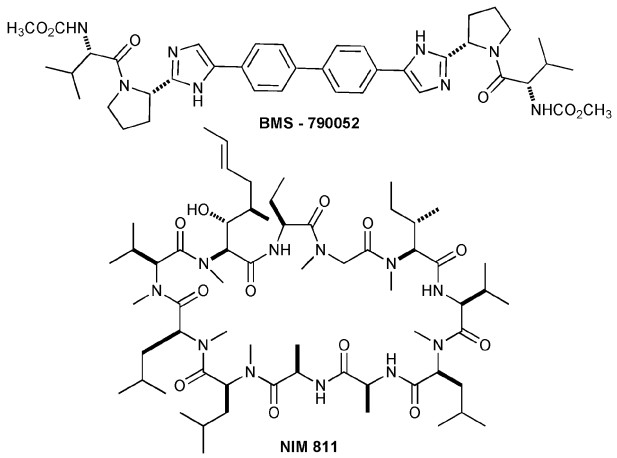


Fig. 7 Structures of HCV inhibitors.

Case study 3: hepatitis C virus

In 2007, 170 million people were estimated to be chronically infected with hepatitis C virus (HCV) and currently 3 to 4 million people are newly infected with HCV each year.[42] More than 80% of people infected with HCV progress to a chronic form of the disease,[64] and the incidence of HCV related hepatocellular carcinoma is increasing.[65] The current standard of care treatment for chronic HCV infections is a combination therapy consisting of pegylated interferon alpha (pegIFN-α) and ribavirin. After a 48-week course of treatment approximately 50% of those individuals chronically infected with genotype-1 HCV experience a sustained virological response.[66] Treatment is often accompanied by debilitating side effects including depression and flu-like symptoms. However, increased knowledge of the viral life cycle is providing new opportunities for therapeutic intervention. In recent years, several small molecule inhibitors which target specific viral proteins have entered and are now showing promise in clinical trials. So far research has focused on developing inhibitors of the NS3-4A protease and the RNA-dependent RNA polymerase NS5B.[67] The high mutation rate of the HCV virus is a major cause for concern as the number of drug resistant strains continues to rise. It is likely that future treatments for HCV will predominantly be comprised of combination therapies as targeting multiple viral proteins helps offset the development of resistance.

In an attempt to discover new chemical entities with distinct modes of action Gao *et al.*[67] screened over 1 million compounds, from the Bristol-Myerrs Squibb proprietary collection, for the selective inhibition of HCV replication using HCV replicon cells.[68] The library consisted of compounds of both natural and synthetic origin. Only compounds which did not display inhibitory activity against NS3 protease, NS3 helicase and NS5B polymerase in biochemical assays were selected as preliminary hit compounds. Extensive optimisation and chemical refinement using a traditional medicinal chemistry approach produced the most potent inhibitor of HCV replication reported to date, BMS-790052 (Fig. 7), with mean half maximum effective concentration (EC_{50}) values of 50 and 9 pM for genotype 1a and 1b replicons.

In order to determine the drug target, BMS-790052 was used to select for resistance on genotype 1a and 1b replicons. Resistance could be mapped to a single protein NS5A. NS5A is a key phosphoprotein involved in viral RNA replication and modulation of cellular processes, with no known enzymatic function.[69,70] Gao *et al.*[71] also performed pull-down experiments with a biotin tagged derivative providing further evidence that NS5A is indeed the target of BMS-790052. No binding to NS3 and NS5B was observed suggesting selective binding to NS5A.

The mode of action of NS5A in regulating viral replication and cellular processes remains unclear. X-Ray diffraction studies have revealed that sections of NS5A form dimeric structures,[69,72] which may oligomerise to form extended two-dimensional assemblies through which RNA substrates are transported.[73] Gao *et al.*[71] speculate that the mechanism of action of BMS-790052 may be based on the disruption of NS5A dimerisation, with a subsequent effect on oligomerisation which could explain the extreme potency of the compound. BMS-790052 performed very well in phase 1 clinical trials where it was found to be safe and well tolerated, and the highest dose administered produced a $3.3\log_{10}$ reduction in mean plasma viral RNA levels after 24 hours.[71] BMS-790052 is currently in phase 2 clinical trials as part of a combination therapy with pegylated interferon alpha and ribavirin.

In 2003 Watashi *et al.*[74] performed a forward chemical genetics screen in search of compounds which affect the rate of replication of HCV in HCV replicon cells. The immunosuppressant cyclosporin A (CsA) was found to be a strong inhibitor of HCV replication. Shimotohno and colleagues[75] subsequently used CsA as a bioprobe, to investigate the cellular processes involved in HCV replication and discovered that the interaction of CsA with a member of the cyclophilin family of proteins, cyclophilin B (CypB), was suppressing viral genome replication. The cyclophilins are a family of peptidyl–prolyl *cis–trans* isomerases (PPIase) involved in numerous cellular processes such as protein folding and trafficking.[75,76] Further investigation revealed that CypB directly promoted the RNA binding activity of NS5B and identified the NS5B–CypB protein–protein interaction as a potential drug target for the development of HCV therapeutics.[75,77]

Despite the impressive activity of CsA as an inhibitor of HCV RNA replication, its strong immunosuppressive activity

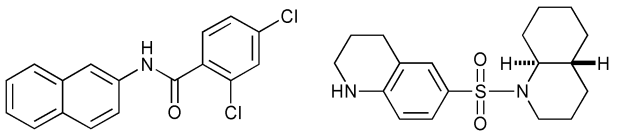


Fig. 8 Inhibitors of the miRNA-122 pathway.

makes it an undesirable drug candidate. However, non-immunosuppressive analogues of cyclosporin A (NIM811, alisporivir and SCY-635) have since been developed which are currently in performing well clinical trials (Fig. 7).[76]

Case study 4: RNA interference pathways[78]

An exciting new area of research where chemical genetics can successfully be applied is the understanding of RNAi pathways, which have been shown to regulate up to 30% of all genes. MicroRNAs (miRNAs) are considered one of the most important regulators in these pathways. These 25–30 nt long RNA segments regulate genes at the post-transcriptional level by binding to a complementary sequence of mRNA. Although a large proportion of regulating miRNAs are known, the factors that regulate these RNAs themselves are still not well understood. To overcome this, assays that can detect the changing levels of miRNAs in response to small molecule modulators have been developed and successfully applied. Two landmark papers by the Deiters group have identified small molecule inhibitors of the miR21 pathway, which is upregulated in many forms of cancer, and the miR122 pathway, which is downregulated in certain hepatocarcinomas and upregulated in hepatitis C virus (HCV).[79,80] So far, these are the only miRNA specific small molecule modulators to have been identified (see Fig. 8); however, other molecules that target the general miRNA biogenesis pathway have been reported. The current limitation in all the pathway screens mentioned previously is the difficulty in identifying the targets. This is also the case for small molecule miRNA modulators; however, the field is just in its infancy, and improving target ID strategies should help uncover these pathways further.

Conclusions

The last few years have been a remarkable period for chemical genetics. Tremendous advancements in screening methods and target identification strategies have dramatically increased the accessibility and generality of the chemical genetic approach. Chemical genetics uses cell-permeable and selective small molecules to perturb gene-product function rapidly, reversibly and conditionally with temporal and quantitative control. Chemical genetics can be readily applied in either a cellular or organismal context and as such offers several advantages over the classical approach. However, rather than being considered an alternative to traditional genetics used by chemists, chemical genetics will have a greater impact on contemporary science when used as a complimentary tool.

The application of chemical genetics in drug discovery programmes is becoming increasingly popular, as it holds the promise of discovering new drug targets outside of the 'druggable' genome and can provide 'hit' compounds with novel modes of action for further development. This is particularly evident in the search for new antimalarials. Both academic and industrial research groups have adopted (often in collaboration) forward chemical genetic phenotypic screening as the starting point in chemical biology and drug discovery initiatives. The whole parasite screening approach can be viewed as being more expedient in lead candidate identification, particularly in drug discovery, when compared with traditional rational design approaches.[81] Once a hit compound is discovered, a combination of classical and chemical genetic approaches can be used to decipher its mode of action, and once a molecular target is identified, target based approaches can be used to develop appropriate leads for drug development.

In order to generalise the chemical genetic approach, the speed at which large collections of chemically diverse compounds can be synthesised will have to dramatically increase. The pace at which complex biological processes, such as protein synthesis or genetic manipulation of organisms, can be achieved is more efficient in many cases than small molecule synthesis. Diversity oriented synthesis can provide structurally complex small molecules which explore a greater degree of chemical space than traditional combinatorial synthesis, but greater efficiency is still required. The evolution of a systematic chemical genetic approach will expand the druggable genome, providing novel targets for medical intervention and inevitably this will lead to better chemotherapeutic treatments.

Notes and references

1 B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts and J. D. Watson, *Molecular Biology of the Cell*, Garland, New York, 3rd edn, 1994.
2 L. Stryer, *Biochemistry*, Freeman, New York, 4th edn, 1995.
3 W. R. J. D. Galloway, J. T. Hodgkinson, M. Welch and D. R. Spring, *Chem. Biol.*, 2009, **16**, 913–914.
4 D. T. Hughes and V. Sperandio, *Nat. Rev. Microbiol.*, 2008, **6**, 111–120.
5 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2007, **70**, 461–477.
6 A. Kornberg, *Chem. Biol.*, 1996, **3**, 3–5.
7 K.-H. Altmann, J. Buchner, H. Kessler, F. Diederich, B. Kräutler, S. Lippard, R. Liskamp, K. Müller, E. M. Nolan, B. Samori, G. Schneider, S. L. Schreiber, H. Schwalbe, C. Toniolo, C. A. van Boeckel, H. Waldmann and C. T. Walsh, *ChemBioChem*, 2009, **10**, 16–29.
8 S. L. Schreiber, *Bioorg. Med. Chem.*, 1998, **6**, 1127–1152.
9 D. R. Spring, *Chem. Soc. Rev.*, 2005, **34**, 472–482.
10 B. R. Stockwell, *Nat. Rev. Genet.*, 2000, **1**, 116–125.
11 B. R. Stockwell, *Trends Biotechnol.*, 2000, **18**, 449–455.
12 J.-R. J. Yeh and C. M. Crews, *Dev. Cell*, 2003, **5**, 11–19.
13 X. F. S. Zheng and T.-F. Chan, *Drug Discovery Today*, 2002, **7**, 197–205.
14 D. P. Walsh and Y.-T. Chang, *Chem. Rev.*, 2006, **106**, 2476–2530.
15 Y. Dorsett and T. Tuschl, *Nat. Rev. Drug Discovery*, 2004, **3**, 318–329.
16 N. Blow, *Nature*, 2007, **450**, 1117.
17 W. R. J. D. Galloway, A. Isidro-Llobet and D. R. Spring, *Nat. Commun.*, 2010, **1**, 80.
18 M. Kawasumi and P. Nghiem, *J. Invest. Dermatol.*, 2007, **127**, 1577–1584.
19 K.-H. Altmann, *et al.*, *ChemBioChem*, 2009, **10**, 16–29.
20 O. Kwon, S. B. Park and S. L. Schreiber, *J. Am. Chem. Soc.*, 2002, **124**, 13402–13404.
21 W. G. Kaelin, *Nat. Rev. Cancer*, 2005, **5**, 689–698.
22 J. Lehár, B. R. Stockwell, G. Giaever and C. Nislow, *Nat. Chem. Biol.*, 2008, **4**, 674–681.
23 R. Brougha, J. R. Frankuma, S. Costa-Cabrala, C. J. Lorda and A. Ashworth, *Curr. Opin. Genet. Dev.*, 2011, **21**, 34–41.

24 A. A. Borisy, P. J. Elliott, N. W. Hurst, M. S. Lee, J. Lehár, E. R. Price, G. Serbedzija, G. R. Zimmermann, M. A. Foley, B. R. Stockwell and C. T. Keith, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 7977–7982.
25 R. K. Scopes, *Protein Purification*, Springer, 3 edn, 1994.
26 R. A. Bauer, J. M. Wurst and D. S. Tan, *Curr. Opin. Chem. Biol.*, 2010, **14**, 308–314.
27 D. J. Huggins, A. R. Venkitaraman and D. R. Spring, *ACS Chem. Biol.*, 2011, **6**, 208–217.
28 J. Inglese, R. L. Johnson, A. Simeonov, M. Xia, W. Zheng, C. P. Austin and D. S. Auld, *Nat. Chem. Biol.*, 2007, **3**, 466–479.
29 J.-P. Renaud and M.-A. Delsuc, *Curr. Opin. Pharmacol.*, 2009, **9**, 622–628.
30 B. G. Hoffstrom, A. Kaplan, R. Letso, R. S. Schmid, G. J. Turmel, D. C. Lo and B. R. Stockwell, *Nat. Chem. Biol.*, 2010, **6**, 900–906.
31 D. Kokel, J. Bryan, C. Laggner, R. White, C. Y. J. Cheung, R. Mateus, D. Healey, S. Kim, A. A. Werdich, S. J. Haggarty, C. A. MacRae, B. Shoichet and R. T. Peterson, *Nat. Chem. Biol.*, 2010, **6**, 231–237.
32 Y. Tsuchiya, D. Vidaurre, S. Toh, A. Hanada, E. Nambara, Y. Kamiya, S. Yamaguchi and P. McCourt, *Nat. Chem. Biol.*, 2010, **6**, 741–749.
33 J. N. Y. Chan, C. Nislow and A. Emili, *Trends Pharmacol. Sci.*, 2010, **31**, 82–88.
34 B. J. Leslie and P. J. Hergenrother, *Chem. Soc. Rev.*, 2008, **37**, 1347–1360.
35 M. Uttamchandani, D. P. Walsh, S. Q. Yao and Y.-T. Chang, *Curr. Opin. Chem. Biol.*, 2005, **9**, 4–13.
36 S. J. Dilly, M. J. Bell, A. J. Clark, A. Marsh, R. M. Napier, M. J. Sergeant, A. J. Thompson and P. C. Taylor, *Chem. Commun.*, 2007, 2808–2810.
37 S.-E. Ong, M. Schenone, A. A. Margolin, X. Li, K. Do, M. K. Doud, D. R. Mani, L. Kuai, X. Wang, J. L. Wood, N. J. Tolliday, A. N. Koehler, L. A. Marcaurelle, T. R. Golub, R. J. Gould, S. L. Schreiber and S. A. Carr, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 4617–4622.
38 J. Li, U. Rix, B. Fang, Y. Bai, A. Edwards, J. Colinge, K. L. Bennett, J. Gao, L. Song, S. Eschrich, G. Superti-Furga, J. Koomen and E. B. Haura, *Nat. Chem. Biol.*, 2010, **6**, 291–299.
39 U. Rix and G. Superti-Furga, *Nat. Chem. Biol.*, 2009, **5**, 616–624.
40 R. A. Butcher, B. S. Bhullar, E. O. Perlstein, G. Marsischky, J. LaBaer and S. L. Schreiber, *Nat. Chem. Biol.*, 2006, **2**, 103–109.
41 G. Giaever, D. D. Shoemaker, T. W. Jones, H. Liang, E. A. Winzeler, A. Astromoff and R. W. Davis, *Nat. Genet.*, 1999, **21**, 278–283.
42 A. M. Smith, R. Ammar, C. Nislow and G. Giaever, *Pharmacol. Ther.*, 2010, **127**, 156–164.
43 B. Lomenick, R. Hao, N. Jonai, R. M. Chin, M. Aghajan, S. Warburton, J. Wang, R. P. Wu, F. Gomez, J. A. Loo, J. A. Wohlschlegel, T. M. Vondriska, J. Pelletier, H. R. Herschman, J. Clardy, C. F. Clarke and J. Huang, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 21984–21989.
44 J. A. Wells and C. L. McClendon, *Nature*, 2007, **450**, 1001–1009.
45 O. Keskin, A. Gursoy, B. Ma and R. Nussinov, *Chem. Rev.*, 2008, **108**, 1225–1244.
46 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Delivery Rev.*, 1997, **23**, 3–25.
47 M. Sattler, H. Liang, D. Nettesheim, R. P. Meadows, J. E. Harlan, M. Eberstadt, H. S. Yoon, S. B. Shuker, B. S. Chang, A. J. Minn, C. B. Thompson and S. W. Fesik, *Science*, 1997, **275**, 983–986.
48 L. T. Vassilev, B. T. Vu, B. Graves, D. Carvajal, F. Podlaski, Z. Filipovic, N. Kong, U. Kammlott, C. Lukacs, C. Klein, N. Fotouhi and E. A. Liu, *Science*, 2004, **303**, 844–848.
49 B. L. Grasberger, T. Lu, C. Schubert, D. J. Parks, T. E. Carver, H. K. Koblish, M. D. Cummings, L. V. LaFrance, K. L. Milkiewicz, R. R. Calvo, D. Maguire, J. Lattanze, C. F. Franks, S. Zhao, K. Ramachandren, G. R. Bylebyl, M. Zhang, C. L. Manthey, E. C. Petrella, M. W. Pantoliano, I. C. Deckman, J. C. Spurlino, A. C. Maroney, B. E. Tomczuk, C. J. Molloy and R. F. Bone, *J. Med. Chem.*, 2005, **48**, 909–912.
50 A. Czarna, B. Beck, S. Srivastava, G. Popowicz, S. Wolf, Y. Huang, M. Bista, T. Holak and A. Dömling, *Angew. Chem.*, 2010, **122**, 5480–5484.
51 M. Liu, C. Li, M. Pazgier, C. Li, Y. Mao, Y. Lv, B. Gu, G. Wei, W. Yuan, C. Zhan, W.-Y. Lu and W. Lu, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 14321–14326.
52 T. Oltersdorf, S. W. Elmore, A. R. Shoemaker, R. C. Armstrong, D. J. Augeri, B. A. Belli, M. Bruncko, T. L. Deckwerth, J. Dinges, P. J. Hajduk, M. K. Joseph, S. Kitada, S. J. Korsmeyer, A. R. Kunzer, A. Letai, C. Li, M. J. Mitten, D. G. Nettesheim, S. Ng, P. M. Nimmer, J. M. O'Connor, A. Oleksijew, A. M. Petros, J. C. Reed, W. Shen, S. K. Tahir, C. B. Thompson, K. J. Tomaselli, B. Wang, M. D. Wendt, H. Zhang, S. W. Fesik and S. H. Rosenberg, *Nature*, 2005, **435**, 677–681.
53 R. W. Snow, C. A. Guerra, A. M. Noor, H. Y. Myint and S. I. Hay, *Nature*, 2005, **434**, 214–217.
54 L. H. Miller, D. I. Baruch, K. Marsh and O. K. Doumbo, *Nature*, 2002, **415**, 673–679.
55 D. C. Greenbaum, *Trends Pharmacol. Sci.*, 2008, **29**, 51–56.
56 J. Baum, T.-W. Gilberger, F. Frischknecht and M. Meissner, *Trends Parasitol.*, 2008, **24**, 557–563.
57 S. Yeoh, R. A. O'Donnell, K. Koussis, A. R. Dluzewski, K. H. Ansell, S. A. Osborne, F. Hackett, C. Withers-Martinez, G. H. Mitchell, L. H. Bannister, J. S. Bryans, C. A. Kettleborough and M. J. Blackman, *Cell (Cambridge, Mass.)*, 2007, **131**, 1072–1083.
58 D. A. Fidock, *Nature*, 2010, **465**, 297–298.
59 K. Koussis, C. Withers-Martinez, S. C. Yeoh, M. Child, F. Hackett, E. Knuepfer, L. Juliano, U. Woehlbier, H. Bujard and M. J. Blackman, *EMBO J.*, 2009, **28**, 725–735.
60 S. Arastu-Kapur, E. L. Ponder, U. P. Fonovi, S. Yeoh, F. Yuan, M. Fonovi, M. Grainger, C. I. Phillips, J. C. Powers and B. Matthew, *Nat. Chem. Biol.*, 2008, **4**, 203–213.
61 F.-J. Gamo, L. M. Sanz, J. Vidal, C. de Cozar, E. Alvarez, J.-L. Lavandera, D. E. Vanderwall, D. V. S. Green, V. Kumar, S. Hasan, J. R. Brown, C. E. Peishoff, L. R. Cardon and J. F. Garcia-Bustos, *Nature*, 2010, **465**, 305–310.
62 W. A. Guiguemde, A. A. Shelat, D. Bouck, S. Duffy, G. J. Crowther, P. H. Davis, D. C. Smithson, M. Connelly, J. Clark, F. Zhu, M. B. Jiménez-Díaz, M. S. Martinez, E. B. Wilson, A. K. Tripathi, J. Gut, E. R. Sharlow, I. Bathurst, F. El Mazouni, J. W. Fowble, I. Forquer, P. L. McGinley, S. Castro, I. Angulo-Barturen, S. Ferrer, P. J. Rosenthal, J. L. DeRisi, D. J. Sullivan, J. S. Lazo, D. S. Roos, M. K. Riscoe, M. A. Phillips, P. K. Rathod, W. C. Van Voorhis, V. M. Avery and R. K. Guy, *Nature*, 2010, **465**, 311–315.
63 M. Rottmann, C. McNamara, B. K. S. Yeung, M. C. S. Lee, B. Zou, B. Russell, P. Seitz, D. M. Plouffe, N. V. Dharia, J. Tan, S. B. Cohen, K. R. Spencer, G. E. González-Páez, S. B. Lakshminarayana, A. Goh, R. Suwanarusk, T. Jegla, E. K. Schmitt, H.-P. Beck, R. Brun, F. Nosten, L. Renia, V. Dartois, T. H. Keller, D. A. Fidock, E. A. Winzeler and T. T. Diagana, *Science*, 2010, **329**, 1175–1180.
64 P. Marcellin, *J. Hepatol.*, 1999, **31**, 1–2.
65 A.-C. Cardoso, R. Moucari, C. Figueiredo-Mendes, M.-P. Ripault, N. Giuily, C. Castelnau, N. Boyer, T. Asselah, M. Martinot-Peignoux, S. Maylin, R. J. Carvalho-Filho, D. Valla, P. Bedossa and P. Marcellin, *J. Hepatol.*, 2010, **52**, 652–657.
66 E. De Clercq, *Med. Res. Rev.*, 2009, **29**, 571–610.
67 M. Gao, R. E. Nettles, M. Belema, L. B. Snyder, V. N. Nguyen, R. A. Fridell, M. H. Serrano-Wu, D. R. Langley, J.-H. Sun, D. R. O'Boyle II, J. A. Lemm, C. Wang, J. O. Knipe, C. Chien, R. J. Colonno, D. M. Grasela, N. A. Meanwell and L. G. Hamann, *Nature*, 2010, **465**, 96–100.
68 Research into HCV genomic replication has until very recently been hampered by the lack of a reliable *in vitro* cell culture system. However, in 1999, Lohmann *et al.* (V. Lohmann, *et al.*, *Science*, 1999, **285**, 110–113) established a subgenomic HCV replicon system in which subgenomic HCV RNA autonomously replicates in Huh-7 cells. This replicon cell system enables the investigation of HCV replication in cell culture.
69 T. L. Tellinghuisen, J. Marcotrigiano and C. M. Rice, *Nature*, 2005, **435**, 374–379.
70 A. Macdonald and M. Harris, *J. Gen. Virol.*, 2004, **85**, 2485–2502.
71 M. Gao, *et al.*, *Nature*, 2010, **465**, 96–100.
72 R. A. Love, O. Brodsky, M. J. Hickey, P. A. Wells and C. N. Cronin, *J. Virol.*, 2009, **83**, 4395–4403.

73 N. Appel, T. Schaller, F. Penin and R. Bartenschlager, *J. Biol. Chem.*, 2006, **281**, 9833–9836.
74 K. Watashi, M. Hijikata, M. Hosaka, M. Yamaji and K. Shimotohno, *Hepatology*, 2003, **38**, 1282–1288.
75 K. Watashi, N. Ishii, M. Hijikata, D. Inoue, T. Murata, Y. Miyanari and K. Shimotohno, *Mol. Cell*, 2005, **19**, 111–122.
76 K. Lin, *Virol. Sin.*, 2010, **25**, 246–266.
77 K. Watashi and K. Shimotohno, *Rev. Med. Virol.*, 2007, **17**, 245–252.
78 Y. Li, C. He and P. Jin, *Chem. Biol.*, 2010, **17**, 584–589.
79 K. Gumireddy, D. D. Young, X. Xiong, J. B. Hogenesch, Q. Huang and A. Deiters, *Angew. Chem., Int. Ed.*, 2008, **47**, 7482–7484.
80 D. D. Young, C. M. Connelly, C. Grohmann and A. Deiters, *J. Am. Chem. Soc.*, 2010, **132**, 7976–7981.
81 T. N. C. Wells, *Science*, 2010, **329**, 1153–1154.