Chemical genomics in the global study of protein functions

X.F. Steven Zheng and Ting-Fung Chan

Small, cell-permeable and target-specific chemical ligands offer great therapeutic value. They can also be used to dissect diverse biological processes, such as cellular metabolism, signal transduction and intracellular protein trafficking. With cutting-edge technologies in synthetic chemistry and ligand screening and identification, chemical ligands have become more readily available for research. Chemical ligands are used increasingly in genomics approaches to understand the global functions of proteins, an emerging frontier called 'chemical genomics'. Chemical genomics should greatly accelerate discovery in biology and medicine in the near future.

***X.F. Steven Zheng**
Department of Pathology and Immunology
Ting-Fung Chan
Department of Pathology and Immunology
and
Molecular and Cellular Biology Program
*Campus Box 8069
Washington University School of Medicine
660 South Euclid Avenue
St. Louis, MO 63110, USA
*tel: +1 314 747 1884
fax: +1 314 747 2797
e-mail: zheng@pathbox.wustl.edu

▼ The turn of the century marks a dramatic breakthrough in biology and medicine. In less than a decade, we have advanced from knowing very little about the genetic makeup of living organisms to possessing databases comprised of every single genetic code. Currently, the whole genome sequences of 600 different species are publicly available, of which 170 belong to eukaryotes, according to the National Center for Biotechnology Information (NCBI) genome site (http://www.ncbi.nlm.nih.gov/Entrez/Genome/main_genomes.html). Most importantly, the entire human genome has been sequenced successfully. This complete knowledge of genomic sequences will have a tremendous impact on biological and biomedical research. New, scalable technologies have become essential to process large amounts of genomic information. Chemical ligand-based approaches are especially useful in high-throughput genomic analyses towards understanding protein functions.

Chemical genetics

Genetic analysis has been the benchmark for studying genes and proteins for nearly a century. In traditional genetics or 'forward genetics', the genome of a model organism is randomly mutagenized. Mutants that produce a change in a desirable phenotype or trait, such as growth, appearance or behavior, are used to discover the identity of genes responsible for producing the phenotype. Because of advances in molecular biology, a second type of genetic approach called 'reverse genetics' was developed. In reverse genetics, an already identified gene is mutated or deleted and the resultant phenotype is studied, providing a picture of the role of that gene in the organism. One major problem of traditional genetics is that mutations are usually constitutive. Mutations in essential proteins often lead to lethality at early life stage and make it impossible for subsequent studies. Even if mutations do not cause lethality, the mutant organisms are often able to compensate for the loss of the gene, which obscures the effect of the original mutations.

To circumvent these problems, various strategies have been developed to create so-called 'conditional alleles', in which the functions of genes become lost under certain conditions [1]. Frequently used conditional alleles are temperature-sensitive (ts) and cold-sensitive (cs) mutations. These alleles produce mutant proteins that lose functions at non-permissive temperatures as a result of misfolding and/or proteolytic degradation. This approach works well for simple organisms, such as yeast, worms, fruit flies and small plants, but it is limited for mammals. Another popular approach is the use of the inducible cre-lox P system, in which the targeted genes are flanked with lox P sites. A transgene carrying an inducible *cre* gene is also present in the same organism. Upon induction of the *cre* recombinase, the target gene is excised from the genome, thereby creating a null mutation. In addition, RNA interference (RNAi) is increasingly recognized as an effective way of producing conditional alleles by causing

1359-6446/02/$ – see front matter PII: 1359-6446(02)02118-3

degradation of messenger RNAs (mRNAs) [2]. However, these approaches are based on the blockage of steps leading to translation, which are often ineffective towards proteins with long half-lives.

Over the past two decades, target-specific chemical ligands have greatly enhanced our ability to delineate complex biological pathways and processes. Cell-permeable chemical ligands can rapidly bind to their target proteins and create loss-of-function (or gain-of-function) phenotypes. Such research, traditionally called pharmacology, originated from the need to understand mechanisms of drug action. The scope of research involving small chemical ligands is now shifting towards studying biological questions, therefore, it is now often called 'chemical biology' or 'chemical genetics' [3]. More comprehensive analysis of chemical genetics can be found in several recent reviews [4–6]. Such a new terminology genuinely reflects its current state in modern academic biological and medical research. Even in the pharmaceutical industry, it is realized that a detailed understanding of biological pathways is pivotal to successful drug development. Although chemical ligands are primarily used to target proteins, they might also bind to nucleic acids, such as DNA promoter sequences [7]. For simplicity, however, we limit our discussion here to protein target-specific chemical ligands.

In a typical chemical–biological approach, cell permeable, target-specific chemical ligands are added to the cells where they bind to protein targets and cause loss-of-function or gain-of-function phenotypes. These ligands could bind to the enzyme's catalytic site by mimicking natural substrates of the enzymes, as exemplified by the drug lovastatin (mevinolin). Lovastatin interacts with the catalytic site of hydroxymethylglutaryl (HMG)-CoA reductase and inhibits its enzymatic activity [8]. The chemical ligands could also interact with a key regulatory domain, as in the case of phorbol ester, which resembles diacylglycerol – a natural protein kinase C (PKC)-activating lipid molecule – and activates PKC activity [9]. In some cases, chemical ligands selectively interfere with certain function(s) of the target proteins but leave other functions intact. For example, the immunosuppressive drug rapamycin specifically inhibits a G1-related function of the target of rapamycin protein 2 (Tor2) but leaves intact a second essential function of Tor2, which is involved in actin depolarization [10,11]. Selective inhibition of a particular function by a drug can be a useful feature to study multifunctional proteins. Chemical ligands can penetrate into the cells and bind to their target proteins with very fast kinetics. They directly neutralize or activate the target proteins, without having to be dependent on the stability of the target proteins, like in traditional genetics. Moreover, because of their unstable nature and their rapid removal by the cellular detoxification systems, these compounds tend to stay in the cells for a short time [12]. Therefore, the effect of the chemical ligands can easily be turned on and off, thereby creating chemically conditional alleles that are dependent on their drug action. In essence, chemical ligands pose unparalleled temporal and spatial control over their protein targets.

Like traditional genetics, the early stage of chemical genetics primarily involved 'forward chemical genetics', in which natural products or a pool of synthetic compounds are screened for desired phenotypes, such as inhibition of tumor growth *in vitro*. However, during such studies it was realized that many natural products were of high specificity and potency towards certain important cellular proteins. Therefore, efforts have been made to identify the drug targets using the chemical compounds as probes. Classical examples include the proteasome inhibitor, lactacystin, and the immunosuppressant, rapamycin [13]. Both drugs are microbial natural products that inhibit cell proliferation and interfere with normal cellular functions [14,15]. Biochemical purification revealed the 20S proteasome as the sole target for lactacystin [16]; biochemical and genetic analyses led to the discovery of TOR proteins as the targets of rapamycin [17–21].

Knowledge of target proteins leads to the evolution of chemical genetics to a stage that parallels the classical reverse genetics (Fig. 1). For example, the use of lactacystin led to the discovery of the proteasome in diverse biological pathways and processes, including antigen presentation, cell-cycle control and cell-fate determination [22]. Similarly, studies using rapamycin allowed identification of a number of TOR-dependent cellular factors, including the ribosomal S6 kinases (S6K) as crucial mediators of TOR signaling to translational control in mammalian cells [23,24], and also Gln3 as a mediator of TOR signaling in nitrogen sensing and transcriptional control in yeast [25–28]. Target-specific chemical ligands have gained great popularity in recent years in research to dissect the functions of target proteins and molecular events, such as signal transduction and enzymatic reactions. For instance, there are nearly 2000 papers citing the use of the MAP kinase kinase (MKK) inhibitor PD98059 and >1500 papers reporting the use of phosphoinositide 3-kinase (PI-3K) inhibitor wortmannin [29].

Protein engineering in chemical biology

A major limitation for chemical biology is that only a small number of target-specific chemical ligands are available. Protein engineering, when combined with synthetically modified chemical ligands or substrates, can also provide

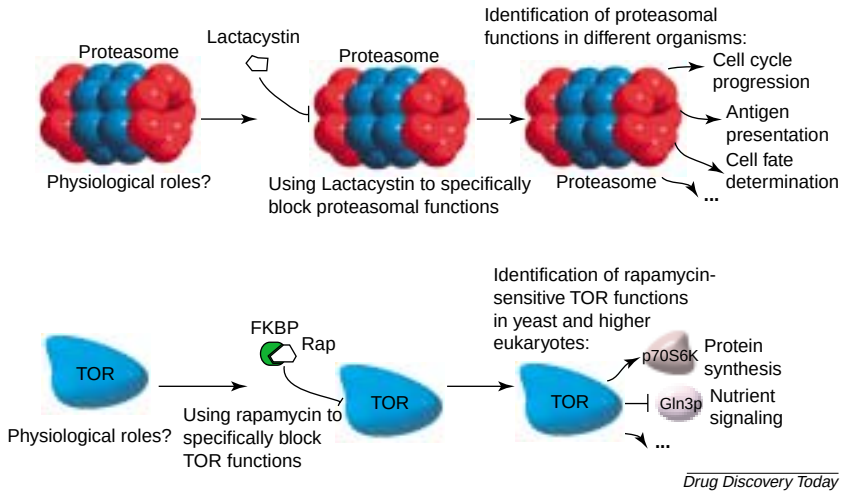

Figure 1. Dissecting protein functions using target-specific chemical ligands. The use of lactacystin unraveled important biological functions for the proteasomes, including antigen presentation, cell-cycle control and cell-fate determination. The studies with rapamycin led to the identification of the ribosomal S6 kinases (S6K) as crucial mediators of TOR signaling to translational control in mammalian cells and the GATA-type transcription factor Gln3p as a mediator of TOR signaling in nitrogen catabolite repression (NCR) in yeast.

an effective substitution for target-specific chemical ligands [30]. Protein kinases are highly conserved in their catalytic domains, making it difficult to develop highly specific inhibitors to each individual kinase. However, it is possible to engineer the catalytic site of a given kinase so that it can accommodate a modified ATP, which binds to the engineered kinase but not other kinases in the cell. This is exactly what Shokat and colleagues did with v-Src, a protein tyrosine kinase [31]. To differentiate the substrates of v-Src from that of all other kinases, they mutated the ATP-binding site of v-Src so that the engineered v-Src uniquely accepted N^6-(cyclopentyl) ATP, an ATP analog that catalyzes normal protein phosphorylation. However, N^6-(cyclopentyl) ATP is not accepted by the unmodified v-Src or other protein kinases.

Using a variation of the previous strategy, Bishop and coworkers devised a way to create allele-specific inhibitors of engineered protein kinases [32]. These inhibitors specifically inhibit the modified protein kinases, but not the wild type, endogenous kinases. Another popular approach takes advantage of the fact that many regulatory proteins require protein–protein interactions for signaling. A synthetic chemical dimerizer (FK1012), which is derived from FK506 with two FK506-binding protein 12 (FKBP12)-binding moieties, is used to mediate the interaction of proteins fused to FKBP12 [33]. When FK1012 is added to the cell, two FKBP12-fusion proteins become dimerized, thereby initiating signal transduction. Variations of this approach with a single compound – such as rapamycin, which has two distinct protein-binding surfaces – can provide higher specificity and more flexibility for such approaches [34]. These chemical ligands can be used to effectively turn on and off engineered proteins.

Why chemical genomics?

Thanks to the large-scale genome sequencing efforts, we now possess complete knowledge of the genetic codes of man and model organisms. In this genomic era, traditional genetic and biochemical approaches are no longer sufficient to process, or take advantage of, the vast amount of DNA sequence information. New, scalable technologies, such as DNA- and protein-microarrays, have become essential in keeping pace with accelerated research and discovery. Chemical compounds are particularly well suited for genome-wide, high-throughput studies. For simplicity, we call the application of target-specific chemical ligands on a genomic scale research 'chemical genomics'. It can also include the development of chemical ligands using genomic tools. Chemical genomics also comes at an opportune time when many sophisticated technologies for chemical ligand synthesis and screening and/or identification have finally become available. New synthetic methodologies, such as combinatorial chemistry [35] and structure-based chemical ligand (or drug) design [36], small chemical compound [37] and protein microarrays [38–40], are just a few examples of many innovative technologies for high-throughput chemical ligand screening and/or identification. These new technologies will inevitably power the discovery of chemical ligands for many important proteins, and accelerate research and discovery in biology and medicine. In this article, we will focus on three genomic areas where chemical ligands are expected to have a major role towards understanding the global functions of proteins.

Chemical ligands and global gene-expression profiling

Regulation of gene expression is an important way for the cell or organism to control its structure, functions and development. Therefore, studying the role of a protein in the regulation of gene expression has been at the forefront of molecular and cellular biology. Complete genome sequences enable the assembly of entire open reading frame (ORF)

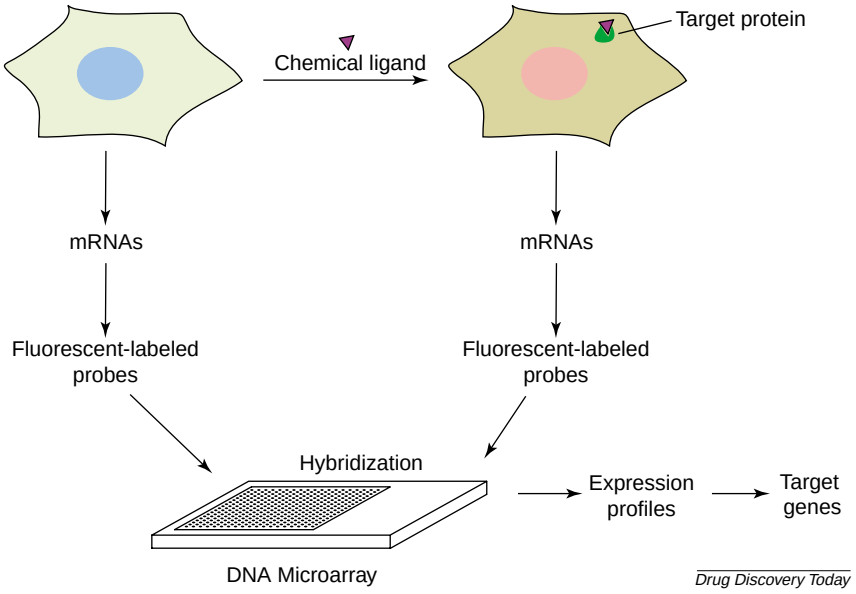

Figure 2. Chemical ligands and global gene-expression profiling. To study the role of a protein in global gene transcriptional control, a target-specific chemical ligand is used to modulate its activity. Total cellular mRNAs are extracted from the cells before and after treatment with the chemical ligand. Fluorescence-labeled cDNA are generated and used to hybridize a high-density DNA microarray. Comparing the expression profiles of the samples reveals the genes under the control of the drug target protein.

sets for individual organisms. This makes it globally possible to study gene expression by high-density oligonucleotide or complementary DNA (cDNA) microarrays (collectively called DNA microarrays) immobilized on the surfaces of glass slides [41]. In such studies, mRNAs are extracted from the sample-of-interest (cell, tissue or organism), labeled with biotin or fluorescent dyes and used to hybridize the DNA microarrays to generate complete gene-expression profiles of the sample cell, tissue or organism.

DNA microarrays have proven useful in elucidating the functions of the drug targets involved in transcriptional control (Fig. 2). In such an approach, the cell, tissue or organism is treated with chemical ligands. mRNAs are prepared from the treated and untreated cell or organism and used to produce fluorescence-labeled cDNAs to hybridize DNA microarrays and generate gene-expression profiles. By comparing the differences between the profiles before and after drug treatment, genes whose expression is modulated by the chemical ligands can then be identified. These genes are further classified into pools based on similar functions or coregulation in the cell. This information can often reveal specific transcription factors and other regulators for each gene pool, thereby allowing the assembly of potential regulatory pathways involving the drug target. A good example is provided by a recent study of yeast histone deacetylases (HDACs) [42]. There are six HDACs in budding yeast: Rpd3, Hda1, Hos1,2 and 3 and Sir2, of which Rpd3 and Hda1 are sensitive to the HDAC inhibitor trichostatin A (TSA), Sir2 and Hos1 are TSA-insensitive, Sir2 is activated by nicotinamide adenine dinucleotide (NAD), and little is known about Hos2 and 3. In this study, the effect of TSA, and the deletion of *HDA1*, *RPD3* and *SIR2* on yeast global gene expression, is compared. It demonstrated that individual HDACs have specific roles in distinct transcriptional pathways, as well as some limited overlapping functions.

Gene expression profile databases of regulatory proteins and biological pathways are rapidly expanding. It is now possible to develop and use sophisticated software to simultaneously compare the drug-induced profiles with the existing databases, thereby retrieving and identifying relevant biological pathway(s). For example, a compendium approach was used to compare the expression profile caused by a drug-induced perturbation to a large and diverse set of reference profiles, and the pathway(s) perturbed by a drug is examined by matching the drug profile with profiles of known cellular pathways [43]. DNA microarrays are also useful for drug target validation, identification of secondary drug target(s) and improvement of drug development programs. Marton and colleagues studied the effect of FK506 and cyclosporin A (CsA) on global gene-expression profiles [44]. They demonstrated that treatment of these drugs in the wild-type yeast strain pheno-copy the null mutations of *CNA1* and *CNA2*, which encode the catalytic subunits of calcineurin. A similar phenomenon was observed when 3-aminotriazole (3-AT) was used to compare gene-expression profiles to that of the *his3* mutant. In a separate experiment, the authors increased the concentration of FK506 and noticed that expression of several genes was independent of calcineurin. They subsequently found that the transcription factor Gcn4 was responsible for this new FK506 effect, placing Gcn4p as a potential secondary target of FK506.

Chemical ligands and protein profiling

Proteomics, or protein profiling, is an emerging field dedicated to the study of the complete protein complement of a cell, tissue or organism. Posttranslational modifications have crucial roles in the regulation of diverse cellular

processes, such as signal transduction, cell-cycle control, development, cytoskeleton networks and metabolism. Common posttranslational modifications include phosphorylation [29], ubiquitylation [45] and related polypeptide-dependent modifications [46], acetylation and/or methylation [47,48] and proteolytic processing [49]. In many cases, a single protein – particularly those with important regulatory functions – can be posttranslationally modified by several distinct mechanisms, depending on the regulatory needs. For example, p53, the most frequently mutated protein in human cancers, is phosphorylated at multiple sites, acetylated, ubiquitylated, oxidized and proteolytically degraded [50]. Deregulation in any of these modifications can contribute to human cancer. Therefore, it is important to study whether and how a regulatory protein controls posttranslational modification(s) of other cellular proteins. Another important aspect of protein profiling is the study of controlled assembly of protein complexes. In addition to the well-known protein complexes, such as ribosomes and proteasomes, many other regulatory proteins also function in large complexes. Dynamic changes in the components of these biological complexes often form the regulatory basis for their relevant cellular processes.

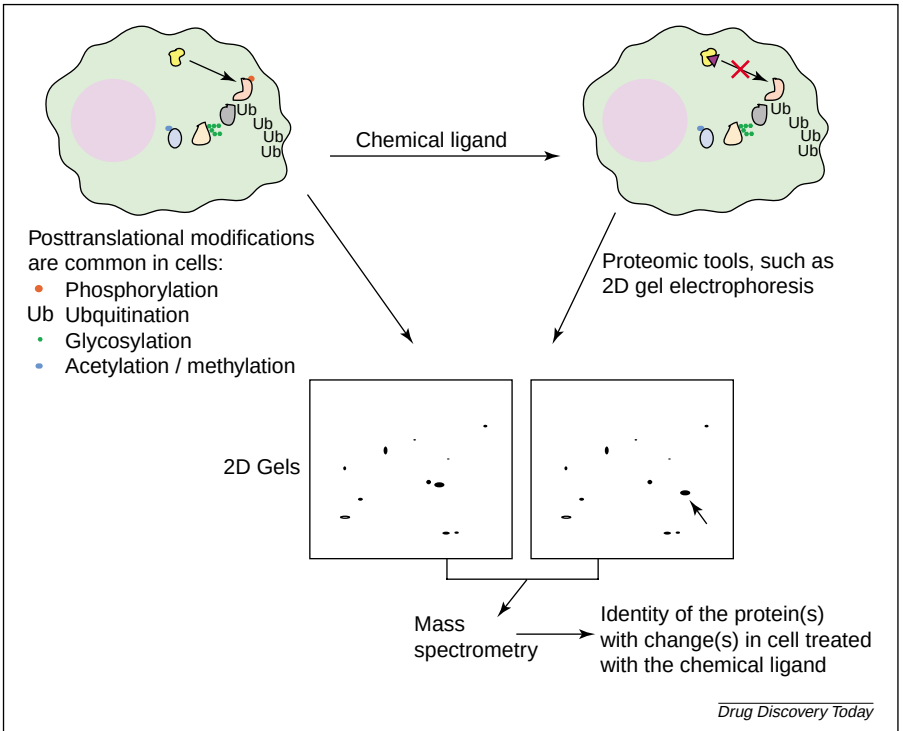

Figure 3. Chemical ligands and protein profiling. A cell-permeable, target-specific chemical ligand that perturbs posttranslational modification of a downstream target protein can be studied by protein profiling technologies, such as two-dimensional (2D) gel electrophoresis. A change in modification of a protein (as shown by an arrow) could be detected by comparison of the protein migration patterns before and after treatment with the chemical ligand. The identity of the protein can then be revealed by MS.

Small, cell-permeable chemical ligands are particularly powerful in linking a regulatory protein and posttranslational modification of its downstream targets. In a typical approach (Fig. 3), a chemical ligand is added to the cells, which leads to the inhibition (or activation) of the regulatory protein. As a result, posttranslational modification(s) of its downstream targets is changed, which can be detected traditionally by two-dimensional (2D) gels, where each polypeptide migrates at a unique position based on its molecular weight and charge [51]. When the protein is modified by ubiquitylation, for example, its mobility changes because of an addition of ubiquitin or polyubiquitin moieties. By comparison of the 2D gel patterns of samples before and after treatment with the chemical ligand, the polypeptide spots with altered mobility are localized. The identity of these polypeptides is further revealed by MS [52]. In addition, the nature of posttranslational modification for each polypeptide can be further investigated. For example, phosphorylation of a protein is easily reversed by phosphatase treatment *in vitro*. Moreover, phosphorylation-, ubiquitin-, and acetylation and/or methylation-specific antibodies can be used to detect the nature of protein modifications. Although 2D gels are useful in identifying posttranslationally modified proteins, each 2D gel can only resolve a limited number of polypeptides because of the low resolution of traditional detection methods, such as silver staining. Most low-abundant proteins are often missed as a result. Recently, highly sensitive methods have been developed. One example is to covalently conjugate proteins with fluorophores and to capture fluorescence 2D images by the charge-coupled device (CCD) camera and laser-scanner-based image acquisition devices [53]. Sophisticated software can further reveal the identity of an individual protein spot.

Several liquid-chromatography-based technologies have also been developed for advanced proteomic studies. A nonporous reverse-phase high-performance liquid chromatography (RP-HPLC) approach was used recently to efficiently separate proteins in the mass range of 5–90 kDa, which were then identified by the on-line electrospray ionization (ESI)-MS [54]. Another approach is the 2D liquid-phase

separation method, in which proteins are separated by liquid-phase isoelectric focusing (IEF) in the first dimension, and then by hydrophobicity using non-porous RP-HPLC in the second dimension [55]. Other recently described IEF fractionation techniques include the liquid-based IEF procedure of free flow electrophoresis (FFE) [56] and the multicompartment electrolyzer, which operates with IEF membranes [57]. These new technologies are not restricted by sample load and are thus more amendable to the study of low abundance proteins. In addition, new techniques, such as isotope-coded affinity tags (ICAT), greatly improved the sensitivity and quantitative analysis of protein profiling [58].

The approaches described previously are based on the analysis of total proteins, both posttranslationally modified and unmodified. Alternatively, modified proteins can be captured directly and examined for their relative abundance under different conditions. Two recent reports described smart chemistry, which allows the selective modification of phosphopeptides or phosphoproteins within complex mixtures [40,59]. Modified peptides or polypeptides are then enriched by covalent or high affinity avidin–biotin coupling to immobilized supports, thereby allowing rapid analysis of phosphorylation events in cells. In a similar vein, modifications with ubiquitin, ubiquitin-like peptides, and methyl and acetyl groups can also be used as epitope tags to capture their modified proteins. In addition, antibodies with specificity towards modified peptides can be arrayed on microchips and used to detect changes in the modification of individual proteins [60,61]. In any event, the total cellular proteins still need to be compared under different conditions to ensure accurate measurement of a change in protein modification. These new protein-profiling technologies can be integrated easily into chemical ligand-directed research.

Chemical ligands and genome-wide genetic interaction mapping

The knowledge of complete genome sequences leads to the identification of ORFs that encode for proteins. Because most of the eukaryotic ORFs are unknown or poorly studied, organized efforts are being made to generate, systematically, deletion mutants of individual genes of the entire genomes of many organisms. The *Saccharomyces* Genome Deletion Project (SGDP) is the first such initiative [62]. It has generated deletion of the entire 6217 ORFs that exist in the yeast genome in a PCR-based approach. Similar large-scale deletional projects are underway for other major model organisms, such as *Caenorhabditis elegans* (using either PCR-based deletion [63] or RNAi [64]), *Drosophila melanogaster* (using P-element insertion [65]), *Arabidopsis thaliana* and *Mus musculus* [66]. Complete sets of mutants in different organisms lay the foundation to discern protein (or gene) functions by systematically analyzing the mutant phenotypes.

Genetic interaction is a powerful way of studying the relationships among genes (or proteins). If a gene acts in the same pathway or a parallel pathway with a second gene, its mutation might affect the phenotype(s) of a mutation in the second gene. Such secondary mutations lead to modification of phenotypes of the first mutations, and are thus called modifier mutations, which enables the identification of new components in the pathway(s) of a gene of interest. In an extreme case, two normally viable mutations, when combined, can generate a lethal phenotype, which is called 'synthetic lethality'. Chemical ligands are ideal for conducting genomic screens for modifier phenotypes (Fig. 4). This is because chemical compounds can be conveniently administered into individual cells, tissues or organisms, thereby creating the same chemically induced alleles in different preexisting mutant backgrounds. The genetic interactions between the drug targets and all other genes can then be systematically measured, based on the relative sensitivity of the mutants to the drug. Even if the drug target-gene is essential, a partial loss-of-function allele can be easily produced by lowering the drug concentration. This variation is a particularly useful feature because it can be used in synthetic lethality screens. The screening process can be easily automated, either as 96-well plate-based or oligonucleotide microarray-based assays. Each yeast deletion mutant has an internal barcode to enable recognition by its unique oligonucleotide sequence [67].

Genetic screens for mutations that confer heightened sensitivity to drugs such as benomyl, have led to the discovery of important cellular events, such as the spindle checkpoint [68]. Giaever and colleagues demonstrated the feasibility of a genome-wide drug sensitivity screen by examining 233 yeast deletion strains to the drug tunicamycin [67]. The yeast genomic deletion mutants were first used in the global study of genetic interactions with TOR [69]. TOR is a highly conserved ataxia telangiectasia-related protein kinase essential for cell growth. To establish a global genetic interaction network of TOR, Chan and coworkers [69] systematically measured the sensitivity of individual yeast mutants to a low concentration of rapamycin based on the relative growth of each mutant, thereby assembling a global genetic interaction network for TOR. The genome-wide screen is obviously advantageous because it profiles every single mutant gene regardless of the severity of its phenotype. Even a moderate sensitivity could be biologically significant, which would probably be missed in a traditional

genetic screen because of a weak phenotype. Moreover, every single mutation is a complete deletion, so it avoids many complicated phenotypes by a point mutation or multi-allelic mutations in a typical traditional genetic screen. Because genes in the same biological pathway that involve the drug target should confer drug sensitivity phenotypes to various degrees, such chemical genomic screens should reveal most, if not all, of the genes in the pathway. This enables the convenient assembly of biological pathways of the drug target by simply pooling genes with similar functions.

The global study of genetic interactions is not only useful for delineating protein functions but also can unravel the genetic basis for drug sensitivity. It is well known that some cancer cells are more resistant or prone to certain anti-cancer drugs. One of the determining factors of drug sensitivity is genetic variation in genes in the same biological pathway(s) as the drug target. Some mutations can increase or decrease rapamycin sensitivity by as much as 10,000-fold [69]. Another study also revealed dramatically varied sensitivities of a collection of yeast DNA-repair and checkpoint mutants to 23 different anti-cancer compounds [70]. Drug sensitivity profiles of different types of cancers can be developed based on the relative expression of genes involved in biological pathway(s) that the drug interferes with. They should be valuable in devising customized protocols for the effective treatment of individual cancers. In addition, many hypersensitive genes could be used as secondary drug targets. Inhibition of the secondary targets can sensitize cells to the original drug, thereby significantly decreasing the dosages of the drugs to minimize side effects and achieve maximal therapeutic values. A recent study using isogenic knockout colon-cancer cells, with wild type and mutant *K-ras*, demonstrates the feasibility of such an approach in mammalian systems [71].

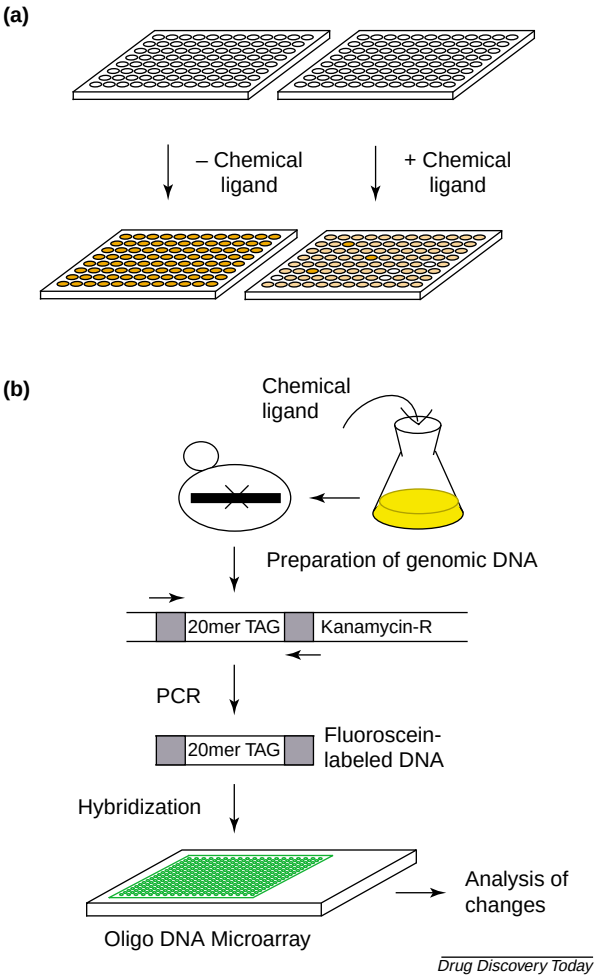

Figure 4. Chemical ligands and the global study of genetic interactions. In a typical global study of genetic interactions, the sensitivity of individual deletion mutants to a drug is systematically measured in a 96-well plate-based assay **(a)** or a bar-code oligonucleotide microarray-based assay **(b)**. Mutants that are hypersensitive or resistant to the drug can be determined by comparison of samples in the absence and presence of the drug. The genetic interaction network can be assembled by pooling genes in the same genetic or biological pathways.

Future perspectives

Chemical ligands can be integrated readily with many existing genomic tools to create assorted chemical genomics approaches, such as gene expression and protein profiling, and global analysis of genetic interactions. When combined, these approaches can produce complete biochemical and genetic profiles of the drug target-protein. Computationally assisted parallel comparison and cross examination of different profiles will enable an accurate and detailed prediction of the biological pathways involved, which form the basis for validation and further detailed study using conventional molecular and cellular approaches. We envisage, in the not-so-distant future, that chemical ligands will become available for most important proteins, which enables chemical modulation of their activities, both positively or negatively, and completely or selectively. These chemical ligands will equip us with the ultimate power to examine every molecular detail of cellular and organism physiology, as well as to deal with every disease causing condition.

Acknowledgements

There are numerous other drug-based genomic studies, particularly in the area of DNA microarray-based gene expression profiling analysis. Although these studies certainly contribute to understanding various aspects of the drug targets, they are primarily pharmacologically oriented.

Because of the space constraint, we apologize for not being able to discuss them here. We thank John Carvalho (Washington University) for helpful comments on the manuscript. This work was supported by grants from the National Cancer Institute and the American Diabetes Association.

References

1 Lewin, B. (1997) Genes are mutable units. In *Gene VI*, (Chapter 3), pp. 66, Oxford University Press
2 Sharp, P. (1999) RNAi and double-strand RNA. *Genes Dev.* 13, 139–141
3 Schreiber, S.L. (1998) Chemical genetics resulting from a passion for synthetic organic chemistry. *Bioorg. Med. Chem.* 6, 1127–1152
4 Crews, C.M. and Splittgerber, U. (1999) Chemical genetics: exploring and controlling cellular processes with chemical probes. *Trends Biochem. Sci.* 24, 317–320
5 Stockwell, B.R. (2000) Chemical genetics: ligand-based discovery of gene function. *Nature Rev. Genet.* 1, 116–125
6 Alaimo, P.J. *et al.* (2001) Chemical genetic approaches for the elucidation of signaling pathways. *Curr. Opin. Chem. Biol.* 5, 360–367
7 Dervan, P. and Burli, R. (1999) Sequence-specific DNA recognition by polyamides. *Curr. Opin. Chem. Biol.* 3, 688–693
8 Alberts, A.W. *et al.* (1980) Mevinolin: a highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme A reductase and a cholesterol-lowering agent. *Proc. Natl. Acad. Sci. U. S. A.* 77, 3957–3961
9 Newton, A.C. (1997) Regulation of protein kinase C. *Curr. Opin. Cell Biol.* 9, 161–167
10 Schmidt, A. *et al.* (1997) The yeast phosphatidylinositol kinase homolog TOR2 activates RHO1 and RHO2 via the exchange factor ROM2. *Cell* 88, 531–542
11 Zheng, X.F. *et al.* (1995) TOR kinase domains are required for two distinct functions, only one of which is inhibited by rapamycin. *Cell* 82, 121–130
12 Kuzmich, S. and Tew, K.D. (1991) Detoxification mechanisms and tumor cell resistance to anticancer drugs. *Med. Res. Rev.* 11, 185–217
13 Schreiber, S.L. (1991) Chemistry and biology of the immunophilins and their immunosuppressive ligands. *Science* 251, 283–287
14 Martel, R.R. *et al.* (1977) Inhibition of the immune response by rapamycin, a new antifungal antibiotic. *Can. J. Physiol. Pharmacol.* 55, 48–51
15 Omura, S. *et al.* (1991) Lactacystin, a novel microbial metabolite, induces neuritogenesis of neuroblastoma cells. *J. Antibiot. (Tokyo)* 44, 113–116
16 Fenteany, G. *et al.* (1995) Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. *Science* 268, 726–731
17 Kunz, J. *et al.* (1993) Target of rapamycin in yeast, TOR2, is an essential phosphatidylinositol kinase homolog required for G1 progression. *Cell* 73, 585–596
18 Cafferkey, R. *et al.* (1993) Dominant missense mutations in a novel yeast protein related to mammalian phosphatidylinositol 3-kinase and VPS34 abrogate rapamycin cytotoxicity. *Mol. Cell. Biol.* 13, 6012–6023
19 Sabatini, D.M. *et al.* (1994) RAFT1: a mammalian protein that binds to FKBP12 in a rapamycin-dependent fashion and is homologous to yeast TORs. *Cell* 78, 35–43
20 Brown, E.J. *et al.* (1994) A mammalian protein targeted by G1-arresting rapamycin-receptor complex. *Nature* 369, 756–758
21 Sabers, C.J. *et al.* (1995) Isolation of a protein target of the FKBP12-rapamycin complex in mammalian cells. *J. Biol. Chem.* 270, 815–822
22 Fenteany, G. and Schreiber, S.L. (1998) Lactacystin, proteasome function, and cell fate. *J. Biol. Chem.* 273, 8545–8548
23 Chung, J. *et al.* (1992) Rapamycin-FKBP specifically blocks growth-dependent activation of and signaling by the 70 kd S6 protein kinases. *Cell* 69, 1227–1236
24 Kuo, C.J. *et al.* (1992) Rapamycin selectively inhibits interleukin-2 activation of p70 S6 kinase. *Nature* 358, 70–73
25 Beck, T. and Hall, M.N. (1999) The TOR signaling pathway controls nuclear localization of nutrient-regulated transcriptional factors. *Nature* 402, 689–692
26 Bertram, P.G. *et al.* (2000) Tripartite regulation of Gln3p by TOR, Ure2p and phosphatases. *J. Biol. Chem.* 275, 35727–35733
27 Cardenas, M. *et al.* (1999) The TOR signaling cascade regulates gene expression in response to nutrients. *Genes Dev.* 13, 3271–3279
28 Hardwick, J.S. *et al.* (1999) Rapamycin-modulated transcription defines the subset of nutrient-sensitive signaling pathways directly controlled by the Tor proteins. *Proc. Natl. Acad. Sci. U. S. A.* 96, 14866–14870
29 Hunter, T. (2000) Signalling-2000 and beyond. *Cell* 100, 113–127
30 Bishop, A. *et al.* (2000) Unnatural ligands for engineered proteins: new tools for chemical genetics. *Annu. Rev. Biophys. Biomol. Struct.* 29, 577–606
31 Shah, K. *et al.* (1997) Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3565–3570
32 Bishop, A.C. *et al.* (1998) Design of allele-specific inhibitors to probe protein kinase signaling. *Curr. Biol.* 8, 257–266
33 Spencer, D.M. *et al.* (1993) Controlling signal transduction with synthetic ligands. *Science* 262, 1019–1024
34 Amara, J.F. *et al.* (1997) A versatile synthetic dimerizer for the regulation of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* 94, 10618–10623
35 Schreiber, S. (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* 287, 1964–1969
36 Russell, R. and Eggleston, D. (2000) New roles for structure in biology and drug discovery. *Nature Struct. Biol.* (Suppl.) 928–930
37 MacBeath, G. *et al.* (1999) Printing small molecules as microarrays and detecting protein-ligand interactions en masse. *J. Am. Chem. Soc.* 121, 7967–7968
38 Haab, B.B. *et al.* (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.* 2, RESEARCH0004.1–0004.13
39 MacBeath, G. and Schreiber, S.L. (2000) Printing proteins as microarrays for high-throughput function determination. *Science* 289, 1760–1763
40 Zhu, H. *et al.* (2000) Analysis of yeast protein kinases using protein chips. *Nature Genet.* 26, 283–289
41 Lockhart, D.J. and Winzeler, E.A. (2000) Genomics, gene expression and DNA arrays. *Nature* 405, 827–836
42 Bernstein, B.E. *et al.* (2000) Genomewide studies of histone deacetylase function in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 97, 13708–13713
43 Hughes, T.R. *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126
44 Marton, M.J. *et al.* (1998) Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nature Med.* 4, 1293–1301
45 Hershko, A. and Ciechanover, A. (1998) The ubiquitin system. *Annu. Rev. Biochem.* 67, 425–479
46 Hochstrasser, M. (2000) Evolution and function of ubiquitin-like protein-conjugation systems. *Nat. Cell Biol.* 2, E153–E157
47 Rice, J.C. and Allis, C.D. (2001) Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr. Opin. Cell Biol.* 13, 263–273
48 Sterner, D.E. and Berger, S.L. (2000) Acetylation of histones and transcription-related factors. *Microbiol. Mol. Biol. Rev.* 64, 435–459
49 Mayer, R.J. (2000) The meteoric rise of regulated intracellular proteolysis. *Nat. Rev. Mol. Cell Biol.* 1, 145–148
50 Giaccia, A. and Kastan, M. (1998) The complexity of p53 modulation: emerging patterns from divergent signals. *Genes Dev.* 12, 2973–2983
51 Moller, A. *et al.* (2001) Two-dimensional gel electrophoresis: a powerful method to elucidate cellular responses to toxic compounds. *Toxicology* 160, 129–138

52 Gygi, S.P. and Aebersold, R. (2000) Mass spectrometry and proteomics. *Curr. Opin. Chem. Biol.* 4, 489–494

53 Patton, W.F. (2000) A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. *Electrophoresis* 21, 1123–1144

54 Chong, B.E. *et al.* (2001) Differential screening and mass mapping of proteins from premalignant and cancer cell lines using nonporous reversed-phase HPLC coupled with mass spectrometric analysis. *Anal. Chem.* 73, 1219–1227

55 Wall, D.B. *et al.* (2000) Isoelectric focusing nonporous RP HPLC: a two-dimensional liquid-phase separation method for mapping of cellular proteins with identification using MALDI-TOF mass spectrometry. *Anal. Chem.* 72, 1099–1111

56 Hoffmann, P. *et al.* (2001) Continuous free-flow electrophoresis separation of cytosolic proteins from the human colon carcinoma cell line LIM 1215: a non two-dimensional gel electrophoresis-based proteome analysis strategy. *Proteomics* 1, 807–818

57 Righetti, P.G. *et al.* (2001) Prefractionation techniques in proteome analysis. *Anal. Chem.* 73, 320A–326A

58 Gygi, S.P. *et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999

59 Oda, Y. *et al.* (2001) Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* 19, 379–382

60 Huang, R.P. (2001) Detection of multiple proteins in an antibody-based protein microarray system. *J. Immunol. Methods* 255, 1–13

61 Belov, L. *et al.* (2001) Immunophenotyping of leukemias using a cluster of differentiation antibody microarray. *Cancer Res.* 61, 4483–4489

62 Winzeler, E.A. *et al.* (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285, 901–906

63 Liu, L.X. *et al.* (1999) High-throughput isolation of *Caenorhabditis elegans* deletion mutants. *Genome Res.* 9, 859–867

64 Bargmann, C.I. (2001) High-throughput reverse genetics: RNAi screens in *Caenorhabditis elegans*. *Genome Biol.* 2, REVIEWS1005

65 Spradling, A.C. *et al.* (1999) The Berkeley *Drosophila* Genome Project gene disruption project: Single P-element insertions mutating 25% of vital *Drosophila* genes. *Genetics* 153, 135–177

66 Coelho, P.S. *et al.* (2000) Genome-wide mutant collections: toolboxes for functional genomics. *Curr. Opin. Microbiol.* 3, 309–315

67 Giaever, G. *et al.* (1999) Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat. Genet.* 21, 278–283

68 Li, R. and Murray, A.W. (1991) Feedback control of mitosis in budding yeast. *Cell* 66, 519–531

69 Chan, T.F. *et al.* (2000) A chemical genomics approach toward understanding the global functions of TOR. *Proc. Natl. Acad. Sci. U. S. A.* 97, 13227–13232

70 Simon, J.A. *et al.* (2000) Differential toxicities of anticancer agents among DNA repair and checkpoint mutants of *Saccharomyces cerevisiae*. *Cancer Res.* 60, 328–333

71 Torrance, C. *et al.* (2001) Use of isogenic human cancer cells for high-throughput screening and drug discovery. *Nat. Biotechnol.* 19, 940–945