Yeast genomics and proteomics in drug discovery and target validation

Ainslie B. Parsons, Ron Geyer, Timothy R. Hughes, Charles Boone

Small biologically active molecules are of interest as both research tools and therapeutic agents. In research applications, target specific inhibitory molecules are an alternative to traditional genetic methods of protein inactivation. Inhibitors designed to bind and alter the function of particular proteins can reveal information about the target with the same specificity as genetic approaches such as gene disruption or RNA interference. The transient nature of chemical ligands, however, typically allows greater temporal and spatial control over their protein targets due to their solubility, rapid diffusion and rapid removal by cellular detoxification systems. There are numerous instances where specific drugs have been used to uncover the cellular function of their targets. In one well known example, studies utilizing the microtubule depolymerizing drug benomyl led to the identification of microtubule network components and the discovery of the spindle checkpoint (1, 2). Similarly, work with rapamycin, an inhibitor of the TOR proteins, resulted in the identification of TOR function in protein synthesis mediated by the ribosomal S6 kinases (3). The development of therapeutic drugs is a second motivating factor driving the study of biologically active compounds. Recent estimates suggest that less than 2% of all human protein (~2700 at BLAST E-value <10^-10, ~1100 at BLAST E-value <10^-50; batch BLAST with default setting; ORFs) have been utilized as drug targets (5, 6). Significant opportunities exist for assigning therapeutic value to the remaining 98% of the proteome.

Despite the high value of biologically active compounds, identifying and validating the intracellular targets of bioactive molecules continues to be a significant problem in drug development. Conventional approaches to drug target identification normally involve biochemical purification of the protein target. For example, the immunophilin and cyclophilin targets of immunosuppressants FK506 and cyclosporin A (CsA) respectively were identified using drug affinity matrices and subsequent purification of their binding proteins (7). Unfortunately methods such as these are generally not amenable to high throughput analysis as they are often labour intensive and need to be tailored to individual proteins and compounds.

The budding yeast *Saccharomyces cerevisiae* offers several advantages as a biological model system for molecular studies. It is a unicellular organism with a compact genome of approximately 6000 genes (10) and a life cycle well suited to classical genetic studies. Additionally, high conservation of many cellular processes between yeast and higher organisms, especially with regards to basic cellular metabolism and cell division, makes yeast a fundamental model eukaryote (9). Just as budding yeast has proven an excellent model system for studying cell biology, it is also a very valuable organism for modelling drug action. Functional genomic and proteomic studies are now possible in *S. cerevisiae* and applications of many of these new technologies are relevant to drug mode of action studies and drug target identification.

Recent sequencing of yeast (10) and human genomes (11, 12) has revealed thousands of yeast proteins that share amino acid sequence similarity with at least one human protein (~2700 at BLAST E-value <10^-10, ~1100 at BLAST E-value <10^-50; batch BLAST with default setting; each of 6357 known or predicted *S. cerevisiae* ORFs). Several hundred of these involve proteins implicated in human disease (9). Indeed, a number of drugs (for example: FK506, CsA, rapamycin, and wortmannin) known to target human proteins also inhibit the orthologous protein in yeast (reviewed in 13).

Yeast is particularly relevant to anti-cancer drug development. Many common mutations in human cancers can be modelled in yeast (14). For example, the *SGS1* gene of yeast encodes a DNA helicase with homology to the human WRN gene, mutations in which lead to Werner’s syndrome, a disease associated with premature aging (15). Furthermore, interaction between Werner’s syndrome gene products and topoisomerases appears to be conserved (16, 17).
While *S. cerevisiae* is a largely non-pathogenic fungus, it is closely related to several common pathogenic yeast strains including *Candida albicans* and, more distantly, *Aspergillus fumigatus*. Presently, only a limited number of different types of antifungal drugs exist (18, 19) but fungal infections continue to be a serious concern in immuno-compromised persons such as AIDS and transplant patients (20). The most commonly used antifungal drugs, including the azoles, target the synthesis of ergosterol, a major component of the fungal cell membrane. The azole class of drugs work against *S. cerevisiae* as well as *C. albicans*, and *A. fumigatus*. Thus, drug target study in *S. cerevisiae* may lead to the identification of new anti-fungal compounds and targets.

A new era in cell biology research has arrived with the application of functional genomics and proteomics methodologies; yeast is currently acting as a testing ground for most of these new technologies. These tools have already contributed and continue to contribute to the understanding of yeast biology. Below, we discuss a number of new methodologies in yeast applicable to drug development including modifications of the yeast two-hybrid system for identifying ligand-protein interactions, the use of protein microarrays for direct biochemical screening, modelling biological effects of drug actions using deletion mutants, expression profiling, genome-wide synthetic lethal screens, and utilizing gene dosage effects for identifying drug targets directly.

THE YEAST TWO-HYBRID SYSTEM

The yeast two-hybrid system enables efficient detection of protein-protein interactions on a large scale (21, 22). It has also led to a number of promising technologies for identifying protein-drug interactions. The yeast two-hybrid system exploits the modular basis of eukaryotic transcription factors (23). Most eukaryotic transcription factors consist of two domains: a DNA binding domain and a transcriptional activation domain. To test for an interaction between two proteins, one protein is fused to the DNA binding domain of a transcription factor, which binds to operators upstream of a specific reporter gene. The second protein is fused to a transcriptional activation domain. When the two fusion proteins are co-expressed, the transcription factor is reconstituted if the two proteins interact. This interaction is detected through expression of a reporter gene. If all the potential ORFs identified from a complete genome sequence are incorporated into two-hybrid constructs, large-scale protein-protein interaction maps can be constructed using high through-put methodology (21, 22, 24).

VARIANTS ON THE TWO-HYBRID SYSTEM

The two-hybrid system has been adapted by Licitra and Liu (25) to study drug-protein interactions. This technique, termed the yeast "three-hybrid system", uses a synthetic heterodimer consisting of two different organic ligands to bring into proximity the DNA-binding domain fused to the receptor of one ligand and the activation domain fused to the receptor for the second ligand. The feasibility of this system was demonstrated by using as the hybrid ligand a heterodimer of covalently linked dexamethasone and FK506. The receptor for dexamethasone was fused to the LexA DNA binding domain and a Jurkat cDNA library fused to a transcriptional activation domain was screened; three overlapping clones of FKBP12, the human FK506 binding protein, were isolated (25).

A reverse two-hybrid system that can be used to select small molecules that inhibit protein-protein interactions has been discussed by Vidal et al. (26) and Huang and Schreiber (27). In the method described by Huang and Schreiber (27) expression of proteins that interact through the two-hybrid system is controlled by the GAL promoter. Following galactose induction, the two interacting proteins are synthesized and their association induces the synthesis of a toxic gene. Only cells where a small molecule inhibits the protein-protein interaction survive. Using this system, nanomolar concentrations of FK506 were shown to disrupt the association of FKBP12 with R1 of the transforming growth factor β receptor family.

YEAST BIOSENSOR

Following a strategy reminiscent of the two-hybrid system, Tucker and Fields (28) have designed a yeast biosensor that reports the binding of ligands to proteins through changes in growth of temperature sensitive yeast and may be useful for screening libraries of open-reading frames for interactions with small molecules. In this system, yeast lacking the metabolic enzyme dihydrofolate reductase (DHFR) are complemented with mouse DHFR containing a ligand-binding domain inserted in a flexible loop. Theoretically, upon binding of the appropriate ligand the yeast strains will show increased growth due to stabilization of the protein fusion. Tucker and Fields were successful in using this system to pick out FK506 from a chemical array as correctly binding to FKBP12 and estrogen as binding to estrogen receptor-α.

PEPTIDE APTAMERS

Screening for peptide aptamers is another application of the yeast two-hybrid system particularly relevant to drug target validation (29-31). Peptide aptamers are antibody-like recognition agents. They consist of peptide sequences between 10 to 20 amino acid residues long that are displayed inside cells as surface loops on a highly expressed carrier protein. Combinatorial aptamer libraries can be used for the reverse and forward analysis of biological processes. In the reverse approach, peptide aptamers that interact with a specific protein are selected from combinatorial aptamer libraries using the yeast two-hybrid system (Figure 1). Peptide aptamers are expressed in vivo where they interact with and inactivate their protein targets. This system has been successfully used to identify highly specific peptide aptamers against human cyclin-dependent kinase 2 (29), the transcription factor E2F which regulates the G1/S transition (32), and the E6 oncoprotein of human papillomaviru-
block the mating pathway in an approach has been used to generate aptamers that then used to determine the target of the aptamers. The for-hybrid system or other protein interaction methods are aptamer-induced phenotypes are isolated. The yeast two-ries are expressed within cells and those cells that display gens" that interfere with cellular processes. Aptamer libra-binatorial aptamer libraries are used as random "muta-ses (33) among others. In the forward approach, com-
binatorial aptamer libraries are used as random "muta-gens" that interfere with cellular processes. Aptamer libra-
ries are expressed within cells and those cells that display aptamer-induced phenotypes are isolated. The yeast two-
hybrid system or other protein interaction methods are
then used to determine the target of the aptamers. The for-
ward approach has been used to generate aptamers that block the mating pathway in S. cerevisiae (30, 31) and taxol-
induced cell death in HeLa cells (34).

The genetic selection of inhibitory aptamers has several possible applications to drug discovery. Aptamer study identifies new targets for drug discover-
y; a protein whose function can be inhibited with an aptamer becomes a good candidate for small molecule inhibition. Highly specific inhibitory aptamers can be used in an analogous manner to drugs to phenocopy a deletion mutant of the targeted protein. In theory, if the aptamers mimic drug action then they should also be useful for the establishment of high-throughput small molecule screens for compounds that disrupt the target-
aptamer interaction. Additionally, aptamers themselves can also be viewed as lead compounds for drug synthesis and design.

HIGH-THROUGHPUT BIOCHEMICAL SCREENING USING PROTEIN ARRAYS

Directly identifying the protein that binds a parti-
cular drug is one of the simplest methods for uncove-
ring drug targets. Martzen et al (35) recently expressed
and purified 6144 yeast ORFs as glutathione-S-transferase (GST) fusions. To identify different activities associated with specific ORFs the strains were grown in defined pools. The GST-ORFs were purified and the pools were assayed for different biochemical activities. Presumably, this set of clones could be used to identify genes encoding proteins that bind any particular ligand or drug. Work by Zhu et al (36) demonstrated the feasibility of a yeast pro-
tome microarray. After cloning, over-expressing and purifying 5800 ORFs, the proteins were printed onto glass slides and directly screened in vitro for protein and phospholipid binding; a number of new calmod-
ulin and phospholipid interacting proteins were identified. This approach will likely be valuable as a rapid and comprehensive alternative for detecting protein-drug interactions.

EXPRESSION PROFILING

Whole-genome expression profiling using DNA microarrays facilitates measurement of the transcriptional response of all genes in an organism following any change in cellular state, including genetic and chemical perturbation. Recent work in yeast has shown that the comparison of genome-wide expres-
sion profiles of cells treated with unknown drugs with strains deleted for possible drug targets has the potential to be a powerful method for linking comp-
pounds to their targets.

To generate a reference data set for the gene expres-
sion approach, a collection of expression profiles cor-
responding to almost 300 different yeast mutants was
established by Hughes et al (37). The 300 mutants studied resulted in approximately 40 discrete patterns of gene expression; by extrapolation, expression profiling of 5000 deletion mutants should yield several hundred patterns, containing on average 15-20 genes per pathway and antici-
"pated to define distinct cellular pathways.

As well as linking uncharacterized genes to their
pathways, this approach can also be used to associate drugs with the pathways they target. The patterns gene-
rated from drug treatments would be expected to mimic those from mutants in the target pathway. This was first shown formally by Marton et al (38) using calcineurin signalling as an example. Yeast cells treated with FK506
yielded a characteristic pattern of altered gene expres-
sion that was similar to the gene expression pattern obtained from a strain deleted for calcineurin, the tar-
get of FK506. Using the compendium of expression
profiles, Hughes et al (37) uncovered Erg2p as the pro-
tein target of the commonly used topical anaesthetic
dyclonine by observing that the pattern of gene expres-
sion changes that occurred upon treatment of yeast with dyclonine was similar to that which occurred in mutants defective for ergosterol biosynthesis. Interestingly, the human protein with the highest simi-
larity to Erg2p is the sigma receptor, a neurosteroid-
binding protein that positively regulates potassium flux, a process integral to nerve conductance.
This system is also useful for determining secondary drug targets. Marton et al (38) reported changes in the expression pattern from a target-less strain in the presence and absence of high concentrations of FK506 to demonstrate off-target effects of the drug. Similarly, Gray et al (39) used expression patterns from yeast to show that a kinase inhibitor (Compound 52) identified as an inhibitor of Cdc28p and Pho85p activity in vitro also has additional activities in vivo.

COMPREHENSIVE SYNTHETIC CHEMICAL STUDIES

The completion of the Saccharomyces gene deletion project (40) has added an important tool to the yeast biologist's arsenal. An international consortium of laboratories co-operated to systematically delete each of the approximately 6000 known or predicted ORFs in yeast using PCR-based homologous recombination to replace each ORF with the antibiotic resistance marker kanamycin (Figure 2A). In total, approximately 1100 genes were found to be essential for viability of the haploid organism resulting in almost 5000 viable haploid deletion mutants (41).

This collection of deletion mutants has enabled high-throughput whole-cell phenotypic studies of yeast mutants. Whole-cell assays, where the activity of a compound and often the measured output takes place on or within a cell, offer many advantages for drug target screening. First, the target is maintained in a cellular environment, favoring native conformation and association with cellular cofactors. Second, this system selects for compounds that can pass through the yeast cell and act before being inactivated or exported; indeed, most compounds that can enter yeast cells can also be expected to pass into mammalian cells because yeast pleiotropic drug resistance efflux pumps are structurally similar to the multiple drug resistance pumps in mammals. Third, as cells are self-replicating, the assay can be easily renewed and expanded.

A number of groups have screened the yeast deletion mutant collection for drug hypersensitivity to uncover the pathways and cellular functions affected by drug treatment on a global scale. One of the first of these projects (42) was aimed at studying the TOR proteins in yeast, signalling molecules implicated in translational control as well as other less well characterized functions. By streaking approximately 2200 viable yeast deletion strains onto plates containing 25 nM rapamycin, a drug that specifically inhibits TOR function, and scoring for rapamycin hyper-sensitivity and rapamycin resistance, a global profile of the rapamycin-sensitive functions of TOR was

![Figure 2. (A) PCR-based construction of deletion mutants. The first round of PCR incorporates up tags and down tags along with universal primers for each tag into the KanMX4 selectable marker. The second round of PCR incorporates homology to ends of a particular gene for integration. (B) Parallel analysis of large pools of deletion mutants. Populations of pooled mutant cells are grown in the presence or absence of a growth inhibitory drug. Genomic DNA is extracted from the pool of mutants and barcodes representing each strain are amplified by PCR using common primers labelled with the fluorescent markers Cy3 or Cy5. Drug sensitive mutants are identified by competitive hybridization of the barcode PCR products to a microarray containing oligos corresponding to each barcode.]
generated. In a similar brute-force approach, Desmoucelles et al. (43) studied approximately 5000 deletion strains for mutants affecting resistance to the immunosuppressive drug mycophenolic acid (MPA) by individually screening strains in a 96-well liquid format; approximately 100 new genes affecting MPA resistance when disrupted were identified. Bennet et al. (44) screened over 2600 homozygous diploid mutants and identified 107 new loci sensitive to ionizing radiation. The 107 strains were then screened against a number of DNA-damaging agents including UV light, bleomycin, MMS, HU, and camptothecin and nearly 90% of the mutants were found to be sensitive to at least one of the drugs.

Because many compounds are in limited supply, a major challenge is to screen the highest number of mutants in the most efficient manner possible while using the least amount of growth medium. Highly parallel analysis of large numbers of pooled deletion strains in a minimal amount of media is possible due to the 20-mer unique molecular barcodes that tag and identify each deletion strain (Figure 2B; 45, 40). In this strategy, strains are pooled and grown in parallel in liquid culture under selective conditions (for example, in the presence of drug). Genomic DNA is extracted from the pool and the barcodes from each strain in the pool are simultaneously amplified using common flanking primers fluorescently labelled with Cy3 and Cy5. This pool of PCR amplifiers is then hybridized to high density DNA microarrays containing oligonucleotides corresponding to the barcodes and the relative abundance of each strain in the pool is assessed by the strength of the resulting signals from the microarray readout compared to a mock control. Using this approach, large numbers of mutants (for example, the entire yeast deletion set of 5000 strains) can be pooled, grown competitively and quantitatively assessed in small volumes of growth medium (41).

The validity of this approach for functionally profiling the S. cerevisiae genome was shown by Giaver et al. (41). Giaver and colleagues used a parallel analysis/molecular barcoding strategy to characterize the nearly complete set of S. cerevisiae deletion strains under a number of different conditions including high salt, sorbitol, galactose, pH 8, minimal media and nystatin treatment. A similar method was used by Fleming et al. (46) to screen the yeast genome for deletion mutants affected by two proteasome inhibitors; in this manner, 52 strains were identified as hypersensitive to the drugs. Likewise, Hanway et al. (47) analysed a collection of 2827 yeast deletion strains for UV and MMS sensitivity and implicated six genes not previously known to be involved in DNA damage repair pathways in the DNA damage response.

High-density cell microarrays as described by Xu et al. (48) may be an alternative method for performing parallel phenotypic analyses on the entire yeast genome using a minimal volume of medium. Here, cell microarrays are created by arraying yeast strains onto a cellulose ester permeable membrane using a standard DNA microarrayer. After arraying the strains, membranes can be incubated on either agar or liquid medium; nutritional and chemical components of the media are accessible to the cells through the membrane. The cell microarrays respond as expected to drug treatment; when 94 different deletion strains were arrayed and incubated in the presence of rapamycin only the strain lacking FKBP1, the receptor for rapamycin, could proliferate. In this system, the distance between the centers of adjacent colonies is only 375 µm it should be possible to use this method to array the entire set of approximately 5000 gene deletion mutants on a permeable membrane within an area of approximately 6 cm².

Comprehensive synthetic chemical studies as described above can provide clues to the mechanism of drug action as well as establish an index of potential genetic and pharmacological interactions. For example, gene deletions that render cells hypersensitive to established antifungal or anticancer drugs identify potential targets for development of combination therapy to enhance drug efficacy. Presumably, this type of screening will be valuable for screening unknown compounds as well. Genome-wide screens provide a sensitivity profile of all mutants regardless of the severity of the drug-induced phenotype. Consequently, even moderate sensitivities can be identified, many of which may be relevant to drug mode of action but easily missed in traditional screens.

SYNTHETIC LETHAL INTERACTIONS AS A KEY TO INTERPRETING SYNTHETIC CHEMICAL INTERACTIONS

An ideal inhibitory molecule targets one gene product specifically. If this is the case, then a strain deleted for the drug target should provide a good model for drug activity. In particular, mutants that are synthetically lethal with the deletion mutant of a drug target should also be hypersensitive to the drug (Figure 3A). The pattern of growth defects among the approximately 5000 deletion strains exposed to a drug should theoretically mimic the pattern of synthetic lethality obtained from a mutation in the gene encoding the drug target. Thus, by comparing the synthetic chemical profiles of unknown compounds with a compendium of synthetic genetic profiles of characterised genes it should be possible to identify the pathways and targets affected by drug treatment.

We have developed a high-throughput method to study synthetic genetic interactions and specifically synthetic lethality in S. cerevisiae on a genome-wide scale (49). This system, referred to as synthetic genetic array (SGA) analysis, involves a series of robotic pinning steps to cross the ordered array of MATα deletion mutants to a MATα starting strain containing a query mutation of interest (Figure 3B). The resultant diploids are selected for and pinned onto media designed to induce sporulation, then double mutant meiotic progeny are isolated by selective growth and scored for fitness. This method of automated genetic analysis has successfully identified hundreds of synthetic genetic interactions and provides an opportunity to create a comprehensive genome-wide, genetic interaction network. Given the high degree of conservation between
genes from yeast to humans it is likely that the genetic interaction networks will be highly conserved as well.

As discussed by Hartwell et al (14), genetic approaches such as synthetic genetic array analysis may be particularly relevant to the development of anti-cancer drugs. Cancerous tumors contain specific mutations that lead to genome instability. Synthetic genetic array analysis allows for screening of mutations that display synthetic lethal interactions with yeast models of cancer causing genetic lesions and as a result can identify targets for drugs that would specifically inhibit proliferation of cells carrying a cancer-causing mutation. A drug such as this should be less toxic than generally acting anti-proliferative compounds because the target is not required for cellular viability unless the cancer-causing mutation is present.

**DRUG-INDUCED HAPLOINSUFFICIENCY ANALYSIS**

Increasing the gene copy number of a drug target increases resistance to the drug (50, 51). Likewise, decreasing the copy number of the drug target renders the cells hypersensitive to the drug. Surprisingly, lowering the gene dosage of a drug target from two copies in a diploid organism to one copy in a heterozygote deletion mutant is often sufficient to generate a hypersensitivity phenotype (Figure 4). Experiments by Giaver et al (52) established proof-of-principle on this concept by individually verifying a number of known drug targets through heterozygote strain analysis. Giaver et al (52) were able to identify the known target of tunicamycin (ALG7) and two other hypersensitive loci from a mixed culture of 233 heterozygote strains using parallel analysis and molecular barcoding. This strategy has been successfully scaled up and used to identify drug targets from pools containing 4500 individual heterozygous deletion strains (D. Shoemaker, personal communication) and offers great potential for directly identifying drug targets in yeast.

**CONCLUSION**

The value of specific drug-like inhibitors in both basic research and drug discovery continues to drive the development of new chemical genetic tools. Target-specific chemical compounds as molecular probes allow for the precise perturbation of cellular pathways; inhibitory ligands corresponding to a wide range of proteins are useful for both dissecting cellular pathways and developing therapeutic drugs. It is clear that the recent technological advances in yeast genomics

---

**Figure 3.** (A) Comparison of synthetic lethal interactions and synthetic chemical interactions. A deletion mutant sensitive to a particular drug should also be synthetically lethal with the drug target. (B) Synthetic genetic array analysis. i) A MATα strain carrying a query mutation (yfg IA) linked to a dominant selectable marker, such as the nourseothricin-resistance marker natMX, and an MFA1pr-HIS3 reporter is mated to approximately 4600 MATα deletion mutants each linked to a kanamycin resistance marker, kanMX. ii) Growth of resultant diploids is selected for on media containing nourseothricin and kanamycin. iii) The diploids are pinned onto media designed to induce spore formation. iv) MATα haploids are recovered using the MFA1pr-HIS3 reporter gene and double mutants are selected for media containing nourseothricin and kanamycin. v) Inviable double mutants are scored as synthetic lethal and slow growing double mutants are scored as synthetic sick.
and proteomics have much to offer the field of drug discovery and in particular, the challenge of drug target identification. One caveat at this point is that most the work described in this review has yet to move much beyond the “proof of concept” stage. The true test of these technologies will be when they are shown to be successful at assigning targets to biologically active compounds of unknown mechanism. However, it is now not unreasonable to imagine that there may ultimately be a chemical ligand identified for every protein in the cell. Comprehensive panels of chemical ligands would allow for an analysis of the combined effects of multiple drugs such that synthetic chemical interaction networks could be created equivalent to current genetic and protein-protein interaction networks. Synthetic chemical interaction networks have the potential to reveal functional relationships between target genes and their pathways but also to provide direct functional information about synergistic drug activities which should greatly expand drug application.

REFERENCES

Figure 4. Drug induced haploinsufficiency. Lowering the gene dosage of a drug target from two copies to one in a diploid cell results in increased sensitivity to drugs acting on the gene product. For example, Alg7 is the target of tunicamycin. At a semi-inhibitory concentration of tunicamycin, wild-type cells are viable. However, lowering the gene dosage of Alg7 from two copies in the wild-type diploid to one copy in the heterozygote deletion mutant results in hyper sensitivity to tunicamycin.