Computer-aided target selection – prioritizing targets for antifungal drug discovery

Frank Spaltmann, Martin Blunck and Karl Ziegelbauer

The entire DNA sequence of the *Saccharomyces cerevisiae* genome was completed in 1996 and represents the first entirely decoded eukaryotic genome. Because major human pathogenic fungi such as *Candida albicans* are closely related to *S. cerevisiae* on a molecular level, the question arises as to how this new information can be used to identify and prioritize those genes that are most suitable as targets for antimycotic drug discovery. To tackle this challenge, a software tool called CATS (computer-aided target selection) was developed. The authors describe how it allows an automated and periodically updated assessment of all *S. cerevisiae* genes to be carried out with regard to their suitability as antifungal targets.

Recently, many complete microbial genome sequences have been released and analysed, such as those of *Saccharomyces cerevisiae*, *Escherichia coli*, *Haemophilus influenzae*, *Mycoplasma genitalium*, *Methanococcus janaschii* and *Mycoplasma pneumoniae*[1–6]. Furthermore, projects have been established to sequence the genomes of *Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Candida albicans* and *Arabidopsis thaliana*[7–11]. Most of the data generated by genome projects are publicly available and offered via the fast communication tools provided by the Internet. This entirely new quality of information combined with publication via the Internet offers exciting new possibilities to the scientific community. Genomics and bioinformatics have a high impact on many disciplines in biological science and will revolutionize drug discovery. For example, by using data from genome projects many disease-related genes have been discovered, such as the hereditary non-polyposis colon cancer gene (*MSH1*, *MLH1* in yeast[12]), the neurofibromatosis type 1 gene (*IRA2* in yeast[13]) and the Werner's syndrome gene (*SGS1* in yeast[14]). In the field of anti-infective research, scientists hope that microbial genomics will help identify novel, pathogen-specific genes as targets for novel antibiotics. A drug designed by this strategy might be more selective and, therefore, have fewer side-effects.

Among the infectious diseases, serious systemic infections caused by fungi are an increasing problem in intensive care units, in recipients of transplanted organs and in other immunosuppressive conditions[15]. The most important pathogens, besides *Aspergillus fumigatus*, are *Candida* species. Treatments of systemic fungal infections are limited to a few agents such as amphotericin B, 5-fluorocytosine, fluconazole and itraconazole. Moreover, these drugs have severe disadvantages because of toxicity, the emergence of drug resistance and their therapeutic spectra[16–20]. Therefore, novel antifungal agents with new mechanisms of action and pharmacodynamic characteristics suitable for the treatment of systemic fungal infections are urgently needed.

In addition to traditional whole-cell screening, target-oriented biochemical assays were introduced for the identification of novel antifungal compounds[21,22]. However, until now, only a few proteins were pursued as antifungal targets and no attempt to classify systematically all fungal proteins with regard to their suitability as antifungal targets has been reported. Since *S. cerevisiae* is closely related to

Frank Spaltmann, Institute for Antiinfective Research, Bayer AG, D-42096 Wuppertal, Germany; **Martin Blunck**, Scientific Information and Documentation, Bayer AG, D-42096 Wuppertal, Germany; and **Karl Ziegelbauer***, Bayer Yakuhin Ltd, Research Center Kyoto, 6-5-1-3 Kunimidai, Kizu-cho, Soraku-gun, Kyoto, 619-02, Japan. *tel: +81 774 75 2459, fax: +81 774 75 2506, e-mail: karl.ziegelbauer.kz@bayer-ag.de

 1359-6446/99/$ – see front matter PII: S1359-6446(98)01278-1

C. albicans – the major human pathogenic fungi – on a molecular level[23], the question arises as to what extent the genomic data of *S. cerevisiae* can be used to identify novel antifungal targets. Because of their huge number, the assessment of all yeast genes for their suitability as antifungal targets can no longer be achieved manually. Therefore, we developed a computational tool, called CATS (computer-aided target selection), to reduce the number of targets to be assessed individually from about 6200 to about 50–100 using uniform criteria.

Computer-aided analysis of the yeast genome

The entire DNA sequence of the genome of *S. cerevisiae* was released in April 1996 and is publicly available[1,24,25]. In addition, several locations on the Internet offer an automated and up-to-date analysis of the yeast genome[26]. Among them the Munich Information Center for Protein Sequences (MIPS; http://speedy.mips.biochem.mpg.de/) offers a very sophisticated and complete analysis[27,28]. The genome sequence of *S. cerevisiae* is approximately 12 Mb in size. The yeast genome is noteworthy because of its compactness. About 72% of the sequence is covered by coding regions. The predicted number of *S. cerevisiae* proteins is about 6200 (Ref. 24). A typical *S. cerevisiae* gene is 2000 bp long, including an open reading frame (ORF) extending over 1450 bp corresponding to 483 amino acids[24]. The analysis of the sequence data of the *S. cerevisiae* genome revealed two groups of genes: two-thirds with known functions and one-third for which the function of the gene is unknown. The latter are unknown sequences or sequences that are homologous to ORFs with unknown functions[29]. As part of the analysis and annotation of the yeast genes, Mewes and colleagues at MIPS compiled table-formatted lists of all yeast ORFs ordered by their corresponding chromosome numbers (e.g. see http://speedy.mips.biochem.de/htbin/get_chromosome_table/?chr=1). In these chromosome tables many properties of the registered genes are described. In addition, the analysis performed by MIPS includes a homology search – using the FastA program – for every ORF against the protein database PIR International[28]. The results of these FastA searches are offered via a hypertext link. Another hypertext-linked table includes information on the function of the ORFs. MIPS classified the ORFs according to their function into a defined set of functional classes (http://muntjac.mips.biochem.de/ycd/ funcat/index.html/).

Parameters for target assessment

In order to assess the value of a gene as an antifungal target, many properties have to be considered. For example, how important is the respective gene for the survival of *S. cerevisiae*, as hitting an essential gene will probably lead to cell death? And, what is the distribution of the gene among pathogenic fungi to be covered by the antimycotic drug? Furthermore, if a gene of interest is also present in man, this may result in unwanted side-effects from a drug targeting that particular gene. Finally, any biochemical and functional information that can be used to develop a high-throughput (HT) screen is of interest. To cover all these different aspects of what is important for an antifungal drug target, we established four different parameters or criteria and used them to assess each gene. The parameters were named 'Quality', 'Occurrence', 'Specificity' and 'Assay Development'. The parameter Quality assesses whether or not the gene product is essential for the survival of *S. cerevisiae* (Box 1). The parameter Occurrence assesses the type and the number of organisms in which the analysed protein is characterized (Box 2). Specificity quantifies the degree of similarity to homologous proteins from fungal species and higher organisms, respectively (Box 3). The parameter Assay Development assesses the practicability of an HT screening (Box 4). The parameters were chosen in such a way that they could be subjected to automated analysis based on the yeast genome data and, in addition, that they could be quantified by scores. To score the degree of homology to other genes, which is important for the parameters Occurrence and Specificity, a 'FastA quotient' was defined (Box 5). The scores for the four parameters were optimized by performing several test runs using 30 genes and then applied to all genes.

Generation of data and analysis of output

As a first step to realizing CATS, a database including all information on all yeast genes necessary for target assessment was generated. To create this database, we used the results of the sequence analysis of the *S. cerevisiae* genome

Box 1. Scoring parameter 'Quality'

The parameter Quality assessed the importance of a gene for the organism. A gene essential for survival of a pathogenic microorganism is a more suitable target than a nonessential gene. As a data source for the parameter Quality we used the column 'gene disruption' of the chromosome tables (URL; see main text). After test scoring with ~30 randomly chosen yeast genes, we set a score of 30 points for the register 'gene disruption lethal' and 0 points for the property 'gene disruption viable' or no register in this column.

Box 2. Scoring parameter 'Occurrence'

The score Occurrence assessed the type and the number of organisms in which the analysed protein is characterized. It is favorable for an antifungal target to be characterized not only in *S. cerevisiae* but also in other pathogenic fungi, especially *C. albicans*. If this is not the case, this has to be proven during target evaluation, which requires time and resources. However, a yeast protein that is also present in higher eukaryotes is scored lower because it is more difficult to find a specific antimycotic drug. Therefore, we scored presence in non-fungal, higher eukaryotic organisms negatively. We used the results of the FastA searches performed for every *S. cerevisiae* gene to analyse the number and the degree of homology of characterized homologs of a given gene. We used a 'FastA quotient' (Box 5) as a measure for the relationship of a given yeast gene with its potential homologs. All proteins possessing a FastA quotient above a given threshold we scored positively for fungal homologs and negatively for higher eukaryotic homologs. However, every species was scored only once. After iterative optimization, we used a FastA quotient of 0.3 as the lower threshold. According to their clinical relevance we assigned different scores to the different fungal species. After test runs with ~30 randomly chosen genes, we assigned the scores shown in brackets for a homolog in the following species: *C. albicans* (10), other *Candida* species (10), *Aspergillus* species (10), other fungi (5). Higher eukaryotes such as human, mouse, rat, pig, bovine, dog, rabbit, cat, chicken and others were scored at −5. Since many protein families are characterized mainly in higher eukaryotes and others mainly in fungi, we observed a distortion by calculating the Occurrence score simply by the sum of all scores for every register. Therefore, we applied an iterative process to avoid a biased view of this parameter. Briefly, a score from a higher eukaryote was registered only if a fungal score above the threshold was found. Thus, for every positive score from fungi only one negative score from higher eukaryotic organisms was allowed.

Box 3. Scoring parameter 'Specificity'

The score Specificity quantified the degree of similarity to homologs from fungal species and higher organisms, respectively. A very high similarity to genes from other fungi led to a high score, while high similarity to an organism listed under higher organisms resulted in a low score. The value Specificity was, therefore, the sum of the two scores Specificity 'Fungi' and Specificity 'Higher Organisms'.

The value Specificity Fungi was determined as follows: first, the gene with the highest homology to the query gene among all fungi was identified; then, the FastA quotient of this gene was used to award the score. The scores shown in parentheses were defined for FastA quotients ranging from no homolog among fungi (0), FastA quotient 0–0.3 (0), FastA quotient 0.31–0.4 (4), FastA quotient 0.41–0.6 (12), FastA quotient 0.61–0.8 (16), FastA quotient 0.81–1 (20).

To determine the score Specificity Higher Organisms we used a similar approach. We looked for the protein with the highest similarity among higher organisms and calculated the FastA quotient. The scores were defined as follows: no homolog among higher eukaryotes (20), FastA quotient 0–0.2 (20), FastA quotient 0.21–0.4 (15), FastA quotient 0.41–0.6 (10), FastA quotient 0.61–0.8 (4), FastA quotient 0.8–1 (0). A limitation of the score Specificity was that, apart from *S. cerevisiae*, no other fungal or eukaryotic genome has yet been completely determined. Thus, we might have missed potential homologs. We tried to overcome this problem by using sequence data of many species.

provided by Mewes and coworkers[25] via the Internet (http://speedy.mips.biochem.mpg.de/) or CD-ROM. Using TRIP software (TRIP Version 3.1-2; a full-text database management system from Fulcrum Technologies, Ottawa, Canada), all data were combined and calculations derived. The TRIP data were then exported into the RS/1 data analysis system (RS/1 Version 5.2; a comprehensive data analysis package from BBN Software Products, Cambridge, MA, USA) to perform all analysis and scoring steps.

Then, on the basis of the parameters described above, the CATS program calculated the scores according to the procedure described in detail in Boxes 1–4. In addition, a 'Total Score' was calculated as the sum of the scores Quality, Occurrence, Specificity and Assay Development. As the sum of all individual scores, the Total Score is a final measure of how suitable a gene is as an antifungal target. The result of this calculation, the CATS output table, was a hierarchical database containing a list of all *S. cerevisiae* ORFs that could be sorted according to different criteria. For example, by request, scores of single parameters or scores including only three out of four parameters could be received.

To analyse the results, first the distribution of various populations of genes in the output table was determined (Figs 1–3). Figure 1 (red bars) shows the Total Score plotted versus the number of genes. The distribution pattern of the genes suggested that CATS was able to select a reasonable amount of antifungal targets out of the ~6200 yeast genes; 54 genes had a Total Score higher than 90 and only 20 ORFs scored higher than 95. In-depth analysis of such a number of genes can be performed manually. The blue bars in Fig. 1 show the overall distribution of the Total Scores of genes with unknown function. As expected, such genes

Box 4. Scoring parameter 'Assay Development'

The parameter Assay Development assessed the resources necessary to set up a high-throughput (HT) screen for the respective gene function. As we also wanted to assess this parameter automatically, on the basis of known data, we had to introduce some simplifications. Our basic assumptions were:

- A cell-based *S. cerevisiae* assay requires fewer resources to be established and run compared with a biochemical assay.
- Transmembrane proteins are generally more difficult to handle and might require additional resources compared with soluble proteins.
- A well-characterized protein will be more suitable for developing an assay than an unknown protein.
- Proteins with different functions (i.e. proteins of different 'functional classes') are more or less suitable for HT assays. In addition, different protein classes differ in their lead probability. A survey in 1995 revealed that, out of the top 100 pharmaceutical drugs, 18 target G-protein-coupled receptors, 10 target nuclear receptors, 16 target ion channels and the rest inhibited enzymes[30].

After careful evaluation of the available data, we used the following four properties of a certain gene:

- 'Gene Disruption'. For an essential gene it is more difficult to develop a target-specific cell-based assay using growth as the read-out. Therefore, the property 'gene disruption lethal' was scored with 0 points. The property 'gene disruption viable' was assigned 10 points and 'no information', 5 points. The property 'gene disruption lethal' was scored both negatively and positively, which seems to be contradictory. However, this allowed us to perform more sophisticated analysis, such as including or excluding certain parameters.
- 'Transmembrane Domains'. Transmembrane proteins are difficult to handle with regard to protein purification or native reconstitution, for example. Therefore, a protein without a transmembrane domain is more suitable in terms of developing a biochemical assay. In the column 'tm' of the chromosome table (URL; see main text) the sum of possible transmembrane domains was given. If a number >0 appeared here, this property was scored with 0 points. If 0 appeared, CATS assigned 10 points.
- 'Status of Characterization'. A well-characterized protein is more suitable for developing an assay than an unknown protein. To assess for the degree of characterization we used the parameter 'class' in the chromosome table (URL; see main text). Here, all proteins are classified according to their degree of characterization. Each protein class was assessed for its suitability to develop an HT assay by the scores shown in parentheses: 1.1 – known protein, completely characterized (10); 1.2 – known protein, not completely characterized (9); 2.1 – homologous to protein of known biochemical and physiological function (8); 2.2 – homologous to protein of known biochemical function (7); 3 – weak homology to known protein (4); 4 – homologous to unknown protein (0); 5 – no similarity (0); 6 – questionable ORF (0).
- 'Functional Class'. The function of a protein also influences the feasibility of an HT assay. One can imagine that it is more difficult to establish an assay, for example, on the basis of a structural protein than to develop a screening procedure for an enzyme. In an attempt to assess the properties of different functional classes, we assigned different scores (shown in parentheses) to different functional classes used in the classification of all ORFs (open reading frames) performed by Mewes and coworkers (see http://muntjac.mips.biochem.mpg.de/ycd/funcat/index.html/): nitrogen and sulphur metabolism (10), nucleotide metabolism (8), amino acid metabolism (10), phosphate metabolism (10), carbohydrate metabolism (7), phospholipid/sterol/sphingolipid/fatty acid metabolism (7), biosynthesis of vitamins/cofactor/prosthetic groups (10), energy (5), cell growth/cell division/DNA synthesis (7), transcription (7), protein synthesis (10), protein destination (6), transport facilitators (3), intracellular traffic (6), cell structure and cellular organization (2), signal transduction (8), cell rescue (8), classification not yet clear-cut (0), unclassified proteins (0), retrotransposons/plasmid proteins (0).

Finally, the score for Assay Development was calculated as the sum of the scores for Gene Disruption, Transmembrane Domains, Status of Characterization and Functional Class.

ranked in the middle and lower part of the output table. The middle part of the CATS output table also contained many nonessential proteins and proteins that are highly conserved among organisms, such as ribosomal proteins. Figure 2 represents the positioning of essential genes (yellow bars) in front of the overall distribution of all genes. Essential genes were found at middle and high Total Scores.

The distribution of the closest homologous proteins among other fungal species and higher organisms was investigated. Figure 3 graphically represents the distribution of genes with specific ranges of FastA quotients towards other fungi and higher organisms. The distribution suggested that:

- Using the FastA quotient, homologous genes over the entire range could be identified except for the range close to a FastA quotient of 1;

Box 5. Measurement of homology

As part of the analysis of the yeast genome, for every gene a search for homologous proteins using the FastA program was performed[28]. As a measure of homology the FastA program provides a so-called 'optimized FastA score'. Because the 'optimized FastA score' is dependent on the length of the query gene or protein, it is not suitable for comparative homology studies. Therefore, we defined a so-called 'FastA quotient' as the quotient of the 'optimized FastA score' of a hit in the FastA result table divided by the 'optimized FastA score' of the yeast protein used as query sequence.

As an example, the calculation of the FastA quotient for the hits of the FastA search using pyruvate kinase as query is given. The optimized score of pyruvate kinase to itself is 2316 (100% identity). The optimized score to the *S. cerevisiae* ORF (open reading frame) YOR347c is 1806. Therefore, the FastA quotient was calculated as: ('optimized FastA score' YOR347c)/('optimized FastA score' pyruvate kinase) = 1806/2316 = 0.78. The same calculation was performed for every FastA result for every yeast gene.

- For a significant part of all genes, no homologs could be found (Fig. 3, left corner, behind);
- Many genes exist for which a close homolog in higher organisms but no counterpart in other fungi could be identified (Fig. 3, right corner, behind). This underlined the necessity to calculate the Occurrence score by a more balanced procedure (see Box 2) rather than by simply adding the scores for each organism;
- Some genes were identified that had no close homolog in higher organisms but a high degree of homology to a gene from other fungi (Fig. 3, left corner, front).

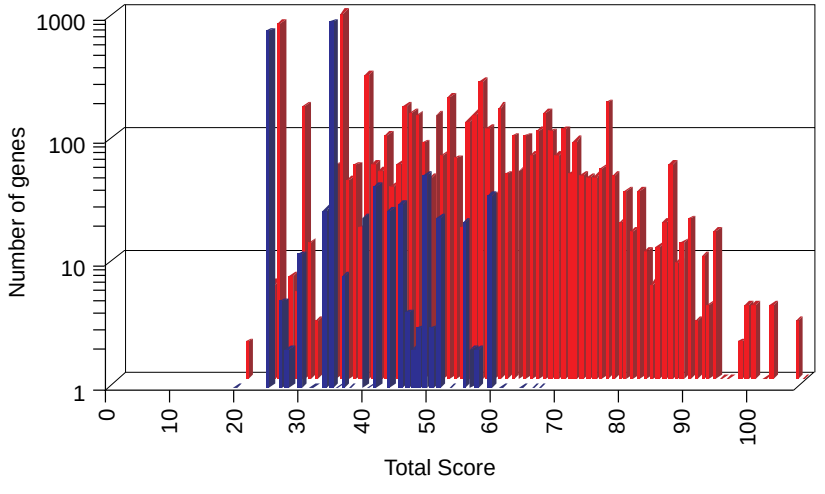

Figure 1. *Distribution of Total Scores I. The number of genes is plotted versus the Total Score. Red bars, all yeast genes; blue bars, genes with unknown function.*

The part of the output table containing the 25 genes with the highest scores ordered by their Total Score is shown in Table 1. The top 25 targets included only a few genes previously considered as good antifungal targets – such as the elongation factor 3 (Refs 31,32) and the proton pumping P-type ATPase (Ref. 33) – but many genes that had not previously been described as antifungal targets.

Assessment of known targets

To test the scoring parameters, the ranking of both antimycotic targets that are widely accepted in the antifungal research community[34] and targets of antifungal drugs in clinical development or on the market were determined (Table 2).

Elongation factor 3

Elongation factor 3 (EF-3) is an essential translation elongation factor that is present in all pathogenic fungi but absent in higher eukaryotes[31,32]. Thus, it is considered a promising antifungal target. EF-3 was ranked, together with two other proteins, at position 10. As expected, EF-3 acquired highest scores regarding the parameters Quality, Occurrence and Specificity, but a lower score for Assay Development. This is in agreement with the difficulty of establishing a target-specific, functional HT assay for a translation factor such as EF-3.

Topoisomerase I

Topoisomerase I scored at position 624. The Total Score calculated by CATS was 74 (Table 2), which is low in comparison with the maximum score of 108. The relatively low scoring resulted from a lack of points in the parameter Quality. Topoisomerase I is not essential for fungal growth. Nevertheless, topoisomerase I-targeting drugs such as camptothecin can lead to cell death by converting topoisomerase I into a cell poison by stabilizing an intermediate, covalent DNA–enzyme complex. Targets with properties like topoisomerase I were not prioritized

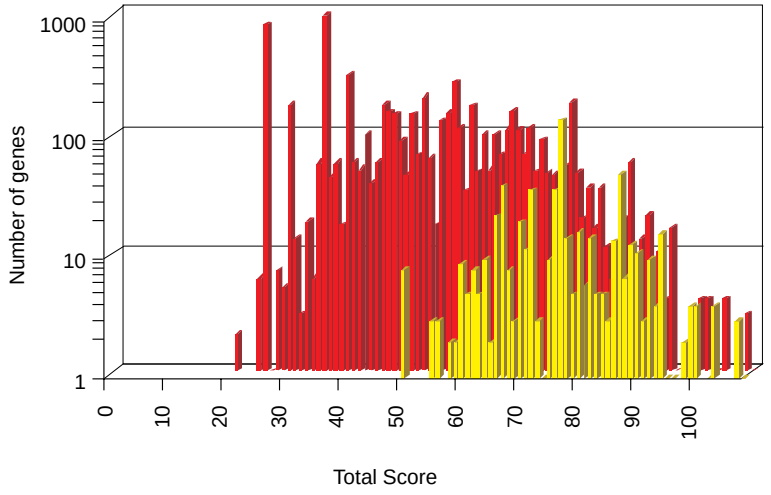

Figure 2. *Distribution of Total Scores II. The number of essential genes (yellow bars) at a certain Total Score is plotted in front of the Total Score of all yeast genes (red bars).*

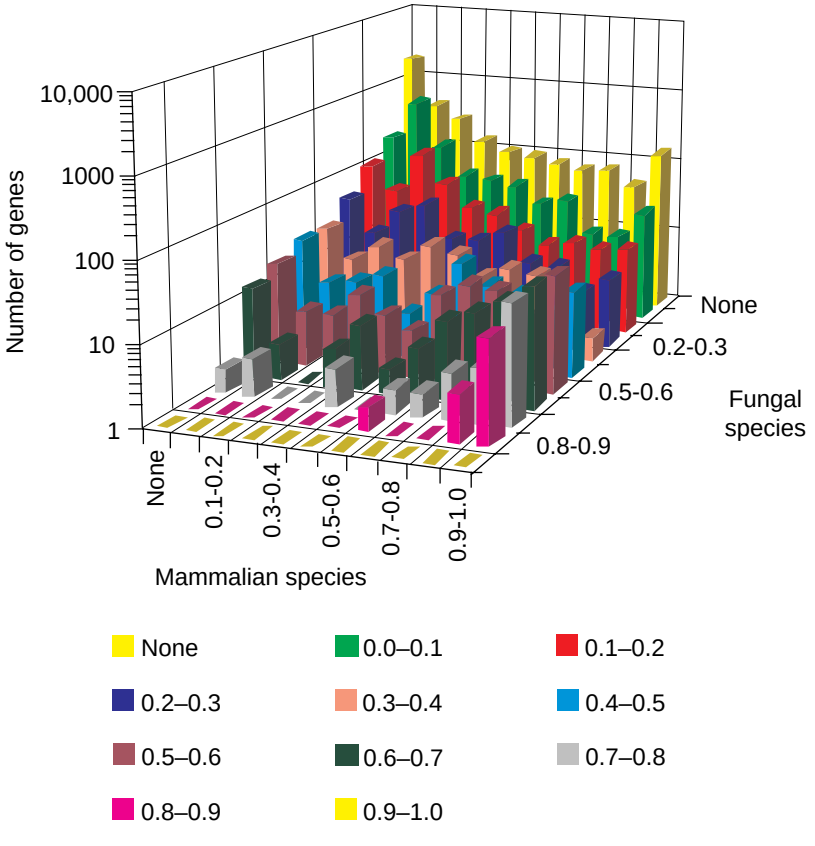

Figure 3. *Distribution of the closest homologs. The number of genes with homologous genes among other fungi and higher organisms within specific ranges of FastA quotients are shown (Version 1.0).*

by CATS. However, the extraordinary cytotoxic mechanism of topoisomerase I-targeting drugs is an exception, not the rule[35].

Fungal cell wall proteins

Other widely discussed targets are the fungal cell wall proteins and the enzymes responsible for the synthesis of the cell wall constituents such as 1→3 and 1→6 branched β-glucans, chitin and mannoproteins[36–38]. Because there is no equivalent of the fungal cell wall in humans, it is considered an ideal target and several compounds have reached clinical trials. In the *S. cerevisiae* genome at least five glucan synthase proteins were identified, in part as subunits of larger protein complexes. For all these genes, deletion mutants are viable and all possess at least one transmembrane domain. Both features received a low score by CATS. According to the Total Score, the two subunits of 1→6 β-glucan synthase, KRE6 and SKN1, were ranked at position 668. In the same range, many proteins that make up the cell wall or that are responsible for cell wall synthesis were found. Our low scoring suggests that HT screens with individual target proteins may not be promising and that other approaches, such as whole-cell screening, are more appropriate. Indeed, most, if not all, compounds targeting the fungal cell wall were discovered by microbial rather than target-oriented screening.

Assessment of the target proteins of antifungal drugs

Azoles

Another feasibility test was the positioning of the target proteins of known antifungals (Table 2). To date, the most successful systemic antifungal class of compounds are the azoles. As the molecular target of these drugs, the enzyme cytochrome P450 lanosterol 14α-demethylase (ERG11) was identified. This enzyme catalyses an important step in ergosterol biosynthesis. Ergosterol is a major sterol in membranes of fungi, but is absent in membranes of higher eukaryotes.

Table 1. CATS output table – the best 25 genes sorted by Total Score

Name	ID	Gene	Quality	Occurrence	Specificity	Assay Development	Total Score
α,α-Trehalose-phosphate synthase, 56 kDa subunit	YBR126c	TPS1	30	10	40	28	108
Protein-tyrosine phosphatase	YFR028c	CDC14	30	10	40	27	107
Phosphatidylinositol/phosphatidylcholine (PI/PC) transfer protein	YMR079w	SEC14	30	10	40	27	107
Polyadenylated RNA-binding protein	YPL190c	NAB3	30	10	40	27	107
Nucleolar rRNA processing protein	YHR089c	GAR1	30	10	36	27	103
Nonsense-mediated mRNA decay protein	YHR170w	NMD3	30	10	36	27	103
Myo-inositol-1-phosphate synthase	YJL153c	INO1	30	10	36	27	103
Profilin	YOR122c	PFY1	30	10	36	27	103
Telomere TTAGGG repeat-binding factor 1	YPL128c	TBF1	30	10	40	22	102
Translation elongation factor EF3	YLR249w	YEF3	30	10	40	20	100
Ubiquitin–protein ligase	YER125w	RSP5	30	10	32	28	100
Adenylate cyclase	YJL005w	CYR1	30	10	32	28	100
Inorganic pyrophosphatase, cytoplasmic	YBR011c	IPP1	30	10	30	30	100
Regulatory subunit for the mitotic function of type I protein phosphatase	YKL193c	SDS22	30	10	32	27	99
Suppressor of toxicity of Gal4-IKB	YLR197w	SIK1	30	10	32	27	99
Ran binding protein	YOL021c	DIS3	30	10	32	27	99
DNA-directed RNA polymerase I, 190 kDa alpha subunit	YOR341w	RPA190	30	10	32	27	99
Vesicular-fusion protein, functional homolog of NSF	YBR080c	SEC18	30	10	32	26	98
Mannose-6-phosphate isomerase	YER003c	PMI40	30	10	31	27	98
ATP-dependent RNA helicase	YOR204w	DED1	30	10	26	30	96
H^+-transporting P-type ATPase	YGL008c	PMA1	30	10	40	15	95
Phosphatidylinositol 3-kinase	YKL203c	TOR2	30	10	36	18	94
Aluminium resistance protein	YOL130w	ALR1	30	10	36	18	94
Member of the Mcm2p, Mcm3p, Cdc46p family	YBL023c	MCM2	30	10	27	27	94
Cyclin-dependent kinases regulatory subunit	YBR135w	CKS1	30	10	27	27	94

Table 2. Position of known antifungal targets

Name	ID	Gene	Quality	Occurrence	Specificity	Assay Development	Total Score	Position
Translation elongation factor EF-3	YLR249w	YEF3	30	10	40	20	100	10
H^+-transporting P-type ATPase	YGL008c	PMA1	30	10	40	15	95	21
Cytochrome P450 lanosterol 14 α-demethylase	YHR007c	ERG11	30	10	31	18	89	76
DNA topoisomerase I	YOL006c	TOP1	0	10	27	37	74	624
Glucan synthase subunit	YPR159w	KRE6	0	10	36	27	73	668
Glucan synthase subunit	YGR143w	SKN1	0	10	36	27	73	668
Cytosine permease	YER056c	FCY2	0	5	36	23	64	1446
Cytosine deaminase	YPR062w	FCY1	0	0	20	38	58	1926

CATS scored ERG11 on position 76, together with several other proteins. Thus, ERG11 is considered as good antifungal target, in agreement with the successful application of azoles in clinical practice. However, azoles have limitations in terms of spectrum and resistance development[39–42]. Furthermore, azoles cause side-effects that result from the inhibition of

mammalian enzymes of the cytochrome P450 superfamily involved in processes such as xenobiotic inactivation and biosynthetic events. Our scoring confirmed this drawback. The homologies between ERG11 from *S. cerevisiae* and cytochrome P450 enzymes, including the 14α-demethylase of *C. albicans*, had FastA quotients in the 0.7 range. At the same time, CATS identified cytochrome P450 proteins in higher eukaryotes with FastA quotients to ERG11 of ~0.3. This resulted in a Specificity score for ERG11 of 31 points, which is significantly lower than the highest possible score of 40 points. In addition, the score for Assay Development was rather low. Here, the lethal phenotype of the deletion mutant and the membrane association of ERG11 scored negatively.

5-Fluorocytosine

After uptake by a cytosine permease (FCY2), the antifungal drug 5-fluorocytosine (5-FC) is converted by a fungal-specific deaminase (FCY1) to 5-fluorouracil (5-FU). 5-FU is subsequently converted by the enzymes of the pyrimidine salvage pathway into toxic nucleoside triphosphates[43]. Both target proteins of 5-FC were found at positions lower than 1000 (Table 2). The low positions were mostly due to viable phenotypes of the deletion mutants, which are considered to be responsible for the increased resistance development observed for 5-FC. In addition, the cytosine deaminase had comparatively low scores for Occurrence and Specificity. Overall, the low ranking of the target proteins of 5-FC is in agreement with its limitations in clinical use due to limited spectrum and resistance development[34].

Identification of new targets

Among the top-scoring ORFs several genes that represent interesting novel targets were identified (Table 1).

Trehalose phosphate synthase

The top-scoring gene was α,α-trehalose phosphate synthase (TPS1), which has only recently been proposed as an antifungal target[44]. This protein received highest scores in all four parameters. *S. cerevisiae* mutants defective in TPS1 are unable to grow in the presence of physiologically relevant concentrations of glucose. On the other hand, trehalose has not yet been found at a significant level in mammals. The *C. albicans* TPS1 (CaTPS1) gene was cloned by functional complementation in *S. cerevisiae*[45]. The CaTPS1 protein showed high homology to the TPS1 proteins of *S. cerevisiae* and *Klyveromyces lactis*. A *C. albicans TPS1* deletion mutant showed impaired yeast-to-hypha transition, a phenomenon involved in the infection process. Furthermore, the TPS1 deletion mutant showed dramatically reduced pathogenicity in a mouse model of *C. albicans* infection[45]. All these results support the top scoring of TPS1.

SEC14

Another new target with a high score was phosphatidylinositol/phosphatidylcholine (PI/PC) transfer protein, also known as SEC14. SEC14 is an essential protein in *S. cerevisiae*. It is involved in the production of secretory vesicles from the yeast Golgi apparatus[46] and is therefore necessary for the export of proteins. It exhibits PI/PC exchange activity, but the exact mode of action is not known[47]. In many other fungi including *C. albicans* a SEC14 homolog was cloned, which also seems to be essential[48]. The *C. albicans* SEC14 showed >60% identity on the amino acid level in comparison with the *S. cerevisiae* counterpart. A putative human homolog that exhibits a similar structure and lipid-binding properties was also cloned. This protein showed only 37% similarity and much lower identity to SEC14p on the amino acid level[49]. However, since the human genome data are not yet complete, the existence of a homolog to SEC14p in man cannot be ruled out. Yeast strains harboring compensatory mutations for the lethal SEC14 deletion have been described[50]. With these strains it should be possible to generate differential screening models to identify SEC14-specific inhibitors. SEC18, a vesicular fusion protein of the secretory apparatus, was also scored high by CATS (Ref. 51). Thus, assays targeting proteins of the fungal secretion apparatus are worth considering.

Inorganic pyrophosphatase

Another top-scoring protein was inorganic pyrophosphatase (iPPase). This essential enzyme is required for the hydrolysis of pyrophosphate, which is a by-product of many biosynthetic and metabolic processes, involving DNA and RNA synthesis, activation of amino acids to build aminoacyl-tRNAs and fatty acid activation. The iPPases are present in every organism; however, identity at the amino acid level among fungal iPPases is ~90%, while the identity between the fungal and, for example, the bovine enzyme is only ~50% (Ref. 52).

Other new targets

Other potential novel targets were CDC14 (a dual-specificity protein phosphatase that functions in late mitosis), profilin (an essential protein involved in the regulation of actin polymerization), the mRNA-binding protein NAB3 and *myo*-inositol 1-phosphate synthase. The latter enzyme catalyzes the conversion of D-glucose 6-phosphate to *myo*-inositol 1-phosphate, the first committed step in the production of all inositol-containing compounds.

Relevant criteria for target evaluation – a never-ending discussion

The main goal of CATS was to provide a uniform assessment of all genes of the *S. cerevisiae* genome as targets for new antifungal drugs based on an automated analysis. The scoring and weight assigned to the different parameters to be considered represent our point of view and many colleagues may disagree with our assessment. For example, a strong emphasis was put on the parameter Quality, which depends on the knock-out phenotype. A lethal knock-out event leads to a high score, because hitting an essential gene will more likely lead to cell death or growth arrest. Cytotoxic or fungicidal activity of new compounds was considered important because most patients suffering from fungal infections are immunocompromised. Furthermore, fungicidal activity is an important differentiation factor compared with azoles, which are fungistatic. This justified, from our point of view, the high scores for essential genes. However, many genes, if deleted, result in viable phenotypes under laboratory conditions, but contribute to fitness under less optimal growth conditions[53]. In addition, by analyzing the phenotype of gene disruptions manually, knock-out strains were often found that were limited in growth but nevertheless viable. This growth retardation might be sufficient to reduce pathogenicity; however, it remains to be seen whether a low molecular weight drug against such a target would show superior activity compared with azoles.

Another point of discussion is the score Assay Development. This parameter evaluated the feasibility of developing an HT assay on the basis of the protein under consideration. Therefore, we preferred nonessential genes because cell-based assays might be possible, ORFs without membrane domains because soluble proteins are much easier to handle, and well-characterized proteins because these proteins do not require extensive characterization prior to assay development.

The highly flexible design of CATS makes it possible for scores to be changed easily and individual parameters such as Quality to be included or excluded. This allows the selection of a target portfolio based on different criteria and thus maximizes the likelihood of success. Even expertise in a certain area, such as cell wall biochemistry, can be included in the assessment by increasing the score for genes belonging to this group.

CATS in the context of drug discovery

The CATS procedure was designed to reduce the number of targets to be assessed manually to a manageable size and to help choose the novel and most suitable targets. The targets selected will then be subjected to a more detailed theoretical and experimental analysis. One of the first steps will be target validation in a pathogenic fungus such as *C. albicans*. Recent improvements in gene knock-out techniques and the possibility to test mutants in animal models of human fungal infection make genetic analysis the best choice for target validation[54]. For targets that are either lethal or significantly reduce virulence, assay development and HT screening for inhibitors will follow.

Update of CATS and further improvements

The output of CATS and the quality of the data do not only depend on the completed *S. cerevisiae* genome sequence, but also on data of ongoing projects:

- The data of the human genome required for the identification of fungal-specific genes;
- The genome sequences of pathogenic fungi, especially *C. albicans*. Although some companies may already have access to a fairly complete version of the *C. albicans* genome sequence, publicly available data are still incomplete (see http://alces.med.umn.edu/Candida.html);
- The systematic functional analysis of all yeast genes and the phenotypic characterization of *S. cerevisiae* mutants. Efforts are ongoing both in Europe[55] and in the USA (Refs 56,57) (see also http://sequence-www.stanford.edu/group/yeast_deletion_project/deletions3.html or http://sgag-www.stanford.edu:8080/footprint/chrV/prechrV.html).

Thus, an important property of the CATS software is the possibility of periodically doing an automated update. We have already completed the first two updates (Versions 1.0–1.2). Not surprisingly, because of the continuous addition of new data, within only five months, six new genes entered the 25 top-scoring genes.

Conclusion

The computer-aided system for antifungal target prioritization was developed for automated analysis of the complete yeast genome data. The program exploits basic molecular information on every gene of *S. cerevisiae* and has the advantage that all genes are assessed with the same criteria. As the CATS scoring program is a user-friendly RS/1 scoring procedure, parameters and scores can be easily changed to include new information or the investigator's personal point of view as to what might be a suitable target. Several novel targets have been identified by CATS – for example, proteins of the yeast secretion apparatus and inorganic pyrophosphatase. To our knowledge no antifungal compound identified by target-oriented screening has

reached clinical trails, so our prioritization remains speculative. Nevertheless, we consider that CATS will be useful, not only to select antifungal targets, but also for assessing bacterial and human genes that are important for other diseases.

ACKNOWLEDGEMENTS

We thank H. Labischinski and W. Schönfeld for discussion and H.W. Mewes and his colleagues at MIPS for their excellent and extensive analysis of the yeast genome data, which were the basis for this study.

REFERENCES

1 Goffeau, A. *et al.* (1996) *Science* 274, 546–567
2 Blattner, F.R. *et al.* (1997) *Science* 277, 1453–1462
3 Fleischmann, R.D. *et al.* (1995) *Science* 269, 496–512
4 Fraser, C.M. *et al.* (1995) *Science* 270, 397–403
5 Bult, C.J. *et al.* (1996) *Science* 273, 1058–1073
6 Himmelreich, R. *et al.* (1997) *Nucleic Acids Res.* 25, 701–712
7 Pennisi, E. (1998) *Science* 281, 148–149
8 Rubin, G.M. (1996) *Genome Res.* 6, 71–79
9 Tait, E. *et al.* (1997) *Fungal Genet. Biol.* 21, 308–314
10 Berks, M. (1995) *Genome Res.* 5, 99–104
11 Bevan, M. *et al.* (1998) *Nature* 391, 485–488
12 Strand, M. *et al.* (1993) *Nature* 365, 274–276
13 Gutmann, D.H. *et al.* (1993) *Oncogene* 8, 761–769
14 Sinclair, D.A., Mills, K. and Guarente, L. (1997) *Science* 277, 1313–1316
15 Hay, R.J. (1991) *J. Antimicrob. Chemother.* 28 (Suppl. B), 17–25
16 Fan-Havard, P. *et al.* (1991) *Antimicrob. Agents Chemother.* 35, 2302–2305
17 Hay, R.J. (1991) *J. Antimicrob. Chemother.* 28 (Suppl. B), 36–46
18 Law, D. *et al.* (1994) *J. Antimicrob. Chemother.* 34, 659–668
19 Lyman, C.A. and Walsh, T.J. (1992) *Drugs* 44, 9–35
20 Odds, F. (1996) *Int. J. Antimicrob. Agents* 6, 141–144
21 Ryley, J.F. and Barrett-Bee, K. (1992) in *Emerging Targets in Antifungal and Antibacterial Chemotherapy* (Sutcliffe, J.A. and Georgopapadakou, N.H., eds), pp. 546–567, Chapman and Hall
22 Mayor, J.S. (1995) in *Antifungal Agents, Discovery and Mode of Action* (Dixon, G.K., Coppong, C.G. and Hollomon, D.W., eds), pp. 9–16, BIOS Scientific Publishers
23 Sudbery, P.E. (1994) *Yeast* 10, 1707–1726
24 Dujon, B. (1996) *Trends Genet.* 12, 263–270
25 Mewes, H.W. *et al.* (1997) *Nature* 387 (Suppl. 6632), 7–65
26 Walsh, S. and Barrell, B. (1996) *Trends Genet.* 12, 276–277
27 Mewes, H.W. *et al.* (1997) *Nucleic Acids Res.* 25, 28–30
28 Mewes, H.W. *et al.* (1998) *Methods Microbiol.* 26, 33–51
29 Casari, G. *et al.* (1996) *Trends Genet.* 12, 244–245
30 Lehmann, J. *et al.* (1996) *Nature* 384 (Suppl. 6604), 1–5
31 Belfield, G.P. and Tuite, M.F. (1993) *Mol. Microbiol.* 9, 411–418
32 Kanath, A. and Chakrabarty, K. (1989) *J. Biol. Chem.* 264, 15423–15428
33 SetoYoung, D. *et al.* (1997) *Biochim. Biophys. Acta* 1326, 249–256
34 Georgopapadakou, N.H. and Walsh, T.J. (1994) *Science* 264, 371–373
35 Shen, L.L. *et al.* (1992) *Antimicrob. Agents Chemother.* 36, 2778–2784
36 Gozalbo, D. *et al.* (1993) *Pharmacol. Ther.* 60, 337–345
37 Debono, M. and Gordee, R.S. (1994) *Annu. Rev. Microbiol.* 48, 471–497
38 Tanner, W. *et al.* (1995) *Acta Biochim. Pol.* 42, 505–508
39 Marichal, P. and Van den Bosche, H. (1995) *Acta Biochim. Pol.* 42, 509–516
40 Horne, T.J. and Hollomon, D.W. (1997) *FEMS Microbiol. Lett.* 149, 141–149
41 Hoesley, C. and Dismukes, W.E. (1997) *Semin. Respir. Crit. Care Med.* 18, 301–309
42 Denning, D.W., Baily, G.G. and Mood, S.V. (1997) *Eur. J. Clin. Microbiol. Infect. Dis.* 16, 261–280
43 Polak, A. (1977) *Contrib. Microbiol. Immunol.* 4, 158–167
44 Eck, R. *et al.* (1997) *Microbiology* 143, 3747–3756
45 Zaragoza, O., Blazquez, M.A. and Gancedo, C. (1998) *J. Bacteriol.* 180, 3809–3815
46 Novick, P., Field, C. and Schekman, R. (1980) *Cell* 21, 205–215
47 Bankaitis, V.A. *et al.* (1990) *Nature* 347, 561–562
48 Monteoliva, L. *et al.* (1996) *Yeast* 12, 1097–1105
49 Arita, M. *et al.* (1995) *Biochem. J.* 306, 437–443
50 Kearns, B.G. *et al.* (1997) *Nature* 387, 101–105
51 Graham, T.R. and Emr, S.D. (1991) *J. Cell Biol.* 114, 207–218
52 Yang, Z. and Wensel, T.G. (1992) *J. Biol. Chem.* 267, 24641–24647
53 Thatcher, J.W. *et al.* (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 253–257
54 Fonzi, W.A. and Irvine, M.Y. (1993) *Genetics* 134, 717–728
55 Oliver, S.G. *et al.* (1998) *Trends Biotechnol.* 16, 373–378
56 Shoemaker, D.D. *et al.* (1996) *Nat. Genet.* 14, 450–456
57 Smith, V. *et al.* (1996) *Science* 274, 2069–2074