68–72 Nucleic Acids Research, 1998, Vol. 26, No. 1

The Yeast Protein Database (YPD): a curated proteome database for *Saccharomyces cerevisiae*

Peter E. Hodges, William E. Payne and James I. Garrels*

Proteome, Inc., 200 Cummings Center, Suite 425C, Beverly, MA 01915, USA

Received October 27, 1997; Accepted November 4, 1997

ABSTRACT

The Yeast Protein Database (YPD) is a curated database for the proteome of *Saccharomyces cerevisiae*. It consists of ~6000 Yeast Protein Reports, one for each of the known or predicted yeast proteins. Each Yeast Protein Report is a one-page presentation of protein properties, annotation lines that summarize findings from the literature, and references. In the past year, the number of annotation lines has grown from 25 000 to ~35 000, and the number of articles curated has grown from ~3500 to >5000. Recently, new data types have been included in YPD: protein–protein interactions, genetic interactions, and regulators of gene expression. Finally, a new layer of information, the YPD Protein Minireviews, has recently been introduced. The Yeast Protein Database can be found on the Web at http://www.proteome.com/YPDhome.html

INTRODUCTION

Since completion of the genome sequence of *Saccharomyces cerevisiae* in 1996 (1), yeast has been the lead organism for post-genomic analysis. Large-scale methods are being used to study the mutant phenotype of deletion of each gene (2–4), to monitor the expression level of the genes (5), to detect the protein–protein interactions (6), and to measure directly the abundance, localization and modification of the proteins (7–10). As large-scale study of proteins (proteomics) increases in importance, new challenges arise for proteome databases.

At present, the greatest body of proteome information for yeast is the research literature. The Yeast Protein Database (YPD) is a curated proteome database that seeks to compile, organize and present in a convenient format the current knowledge of yeast protein functions (11,12). The ways that proteins work together to carry out the functions of the cell are better understood for yeast than for any other cell. More than 45% of the yeast proteins have been characterized by focused investigations already (see Table 1). The pace of characterization of yeast proteins has increased rapidly in recent years, and all of the genes/proteins will soon be characterized to some degree by large-scale methods. The YPD staff gathers the new information for each yeast protein, combines it with the prior knowledge, and presents the information for each protein as a one-page Yeast Protein Report.

Table 1. Growth of YPD content by release number

YPD release	Date	Total proteins	Known function[a]	Similarity[b]	Unknown function[c]
1.2	Nov. 23, 1994	3020	1729	387	904
2.0	Dec. 8, 1994	3142	1750	450	942
3.0	Feb. 1, 1995	3512	1871	524	1117
4.0	Jun. 6, 1995	4046	1951	667	1428
4.1	Jul. 7, 1995	4305	2012	729	1564
5.0	Nov. 30, 1995	4559	2187	859	1913
6.0	Aug. 3, 1996	6021	2369	1231	2421
7.01	Jan. 21, 1997	6045	2507	1192	2346
7.42	Oct. 21, 1997	6089	2786	1084	2219

[a]Proteins characterized through genetic or biochemical experiments.
[b]Proteins that have not been characterized but have sequence similarity to characterized proteins.
[c]Proteins of unknown function.

This summary updates the status of YPD and introduces several new features. The often numerous free-text annotation lines in each Yeast Protein Report are now sorted and placed under topic headings. Pop-up windows have been added that summarize known physical and genetic interactions and the known inducers or repressors of mRNA expression for each gene. Each fact reported in YPD is associated with a reference, and the reference numbers are linked to PubMed (13) where abstracts and sometimes links to full text articles are found. Finally, this report introduces a new layer of yeast protein information, the YPD Protein Minireviews.

THE YEAST PROTEIN REPORT

The Yeast Protein Report (Fig. 1) uses a standardized format for presentation of a maximal amount of protein information on one Web page. Links can take the user to pages with a greater level of detail, such as sequence database pages (14–17), PubMed abstracts (13), or to pages with a higher degree of overview, such as YPD Protein Minireviews (see below). The four sections of each Yeast Protein Report are: (i) the one-to-two line descriptive phrase (title line), (ii) the formatted protein property section, (iii) the annotations divided by topic headings, and (iv) the reference

*To whom correspondence should be addressed. Tel: +1 978 922 1643; Fax: +1 978 922 3971; Email: jg@proteome.com

list. The title line is modified frequently as more is learned about the role of each protein. The protein property section includes alternate names, links to other sequence and model organism databases, properties calculated from the protein sequence, and experimentally determined properties. In YPD, the subcellular localizations, functional categories and post-translational modifications are based on experimental evidence or strong predictions (based on close family similarity).

The newest additions to YPD are the Interaction Reports and the Regulation Reports (Fig. 2). These new reports can be viewed in pop-up windows without leaving the Yeast Protein Report. The Interaction Report presents current information on physical interactions (protein–protein associations and multiprotein complexes) and genetic interactions, with references. The Regulation Report summarizes the current knowledge of agents (environmental conditions, small molecules, or transcription factors) that induce or repress the protein of interest at the transcriptional level.

YPD CURATION

The defining characteristics of YPD is its collection of annotations derived from the yeast literature. The annotations represent the work of curators who read the literature, extract the protein properties and restate the key biological results as free-text annotations. YPD curators are experienced PhD-level yeast researchers who are trained in the style and depth of coverage expected for YPD. Each curator is supported by a variety of tools for searching the yeast sequence and reference databases, and the curator has access to a developer's version of YPD that contains curators notes and annotations from unpublished work. Curators choose their own topics, starting with the new literature for the current week and tracing back to related papers from the older literature. Most recent papers are curated by reading the full paper; many older papers are curated from their abstract only. The YPD reference database contains ~12 000 references to papers dealing with yeast proteins. It also tracks the curation status of each paper (reviewed in full, reviewed in abstract only, or not yet done). With the current staff of eight curators, it is our goal to complete the curation of the yeast literature within the next two years.

The integrity of all data entered into YPD is ensured by an editorial step. All notes submitted by curators are reviewed by a YPD editor for proper style and clarity. Editors often put in additional cross-referencing, correct the nomenclature and style, and resolve occasional inconsistencies. The final level of review is by the YPD users, many of whom submit comments, corrections, and new data. The YPD staff welcomes Email at ypd@proteome.com.

USING THE YEAST PROTEIN DATABASE

The YPD Web pages are intended for use by yeast biologists and researchers in other fields who can benefit from the yeast model system. YPD can serve (i) to acquaint yeast researchers with newly characterized proteins, (ii) to enumerate proteins that share a particular property or function, (iii) as a guide to the yeast research literature, and (iv) as an entry point for scientists working with other organisms who seek clues to the structure, function, or interactions of an unknown protein based on the knowledge of similar proteins in yeast. Users can access a single Yeast Protein Report directly (for example, actin's page at http://www.proteome.com/YPD/ACT1.html). However, most users will be

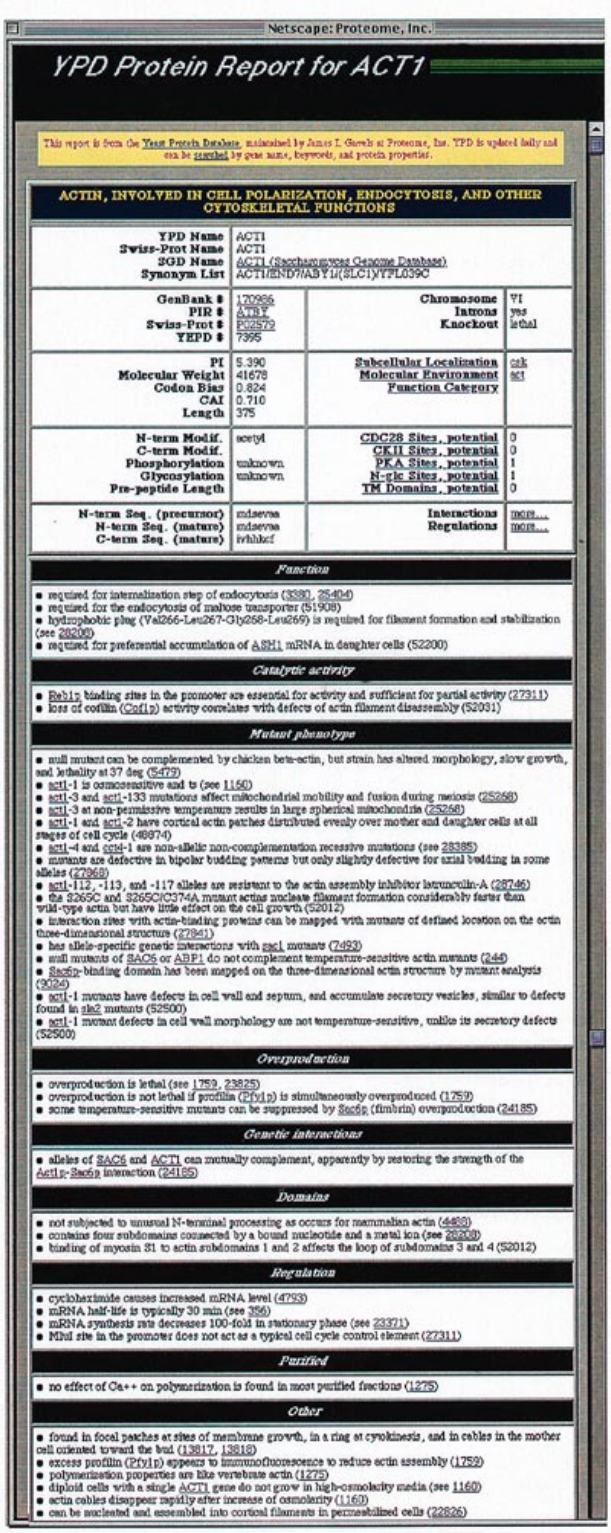


Figure 1. A representative YPD Yeast Protein Report page describing actin, encoded by the ACT1 gene. Users can access this page via the World Wide Web at http://www.proteome.com/YPD/ACT1.html . The page includes a title and description (uppermost), a table of protein properties (upper table), annotations from the literature sorted with topic headings (main body, not completely shown here) and a reference list (not shown here). All of the color-highlighted properties, gene names, protein names and reference numbers are linked to their corresponding pop-up windows (see Figs 2 and 3), Yeast Protein Report pages, PubMed reference pages, or other databases.

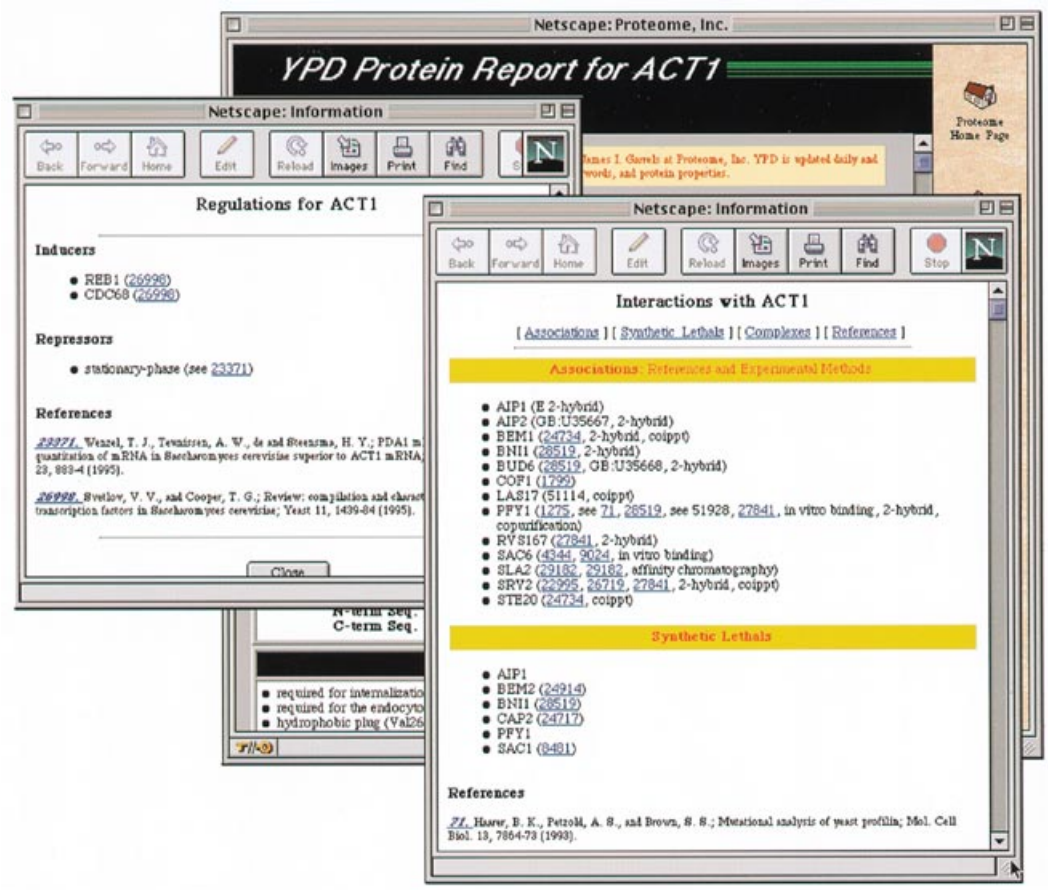


Figure 2. Representative Interactions and Regulations pop-up windows from the actin Yeast Protein Report page. The Interactions window is subdivided into Associations, Synthetic Lethals, and Complexes. The gene name corresponding to the interacting gene or protein is presented, along with references, experimental method for determining associations and the name of the multiprotein complex. The Regulations window lists factors that regulate the expression of this protein, grouped as inducers or repressors. They include regulatory proteins (listed by their gene name), environmental conditions or small molecules. References are linked and regulatory proteins are linked to their own Yeast Protein Report page.

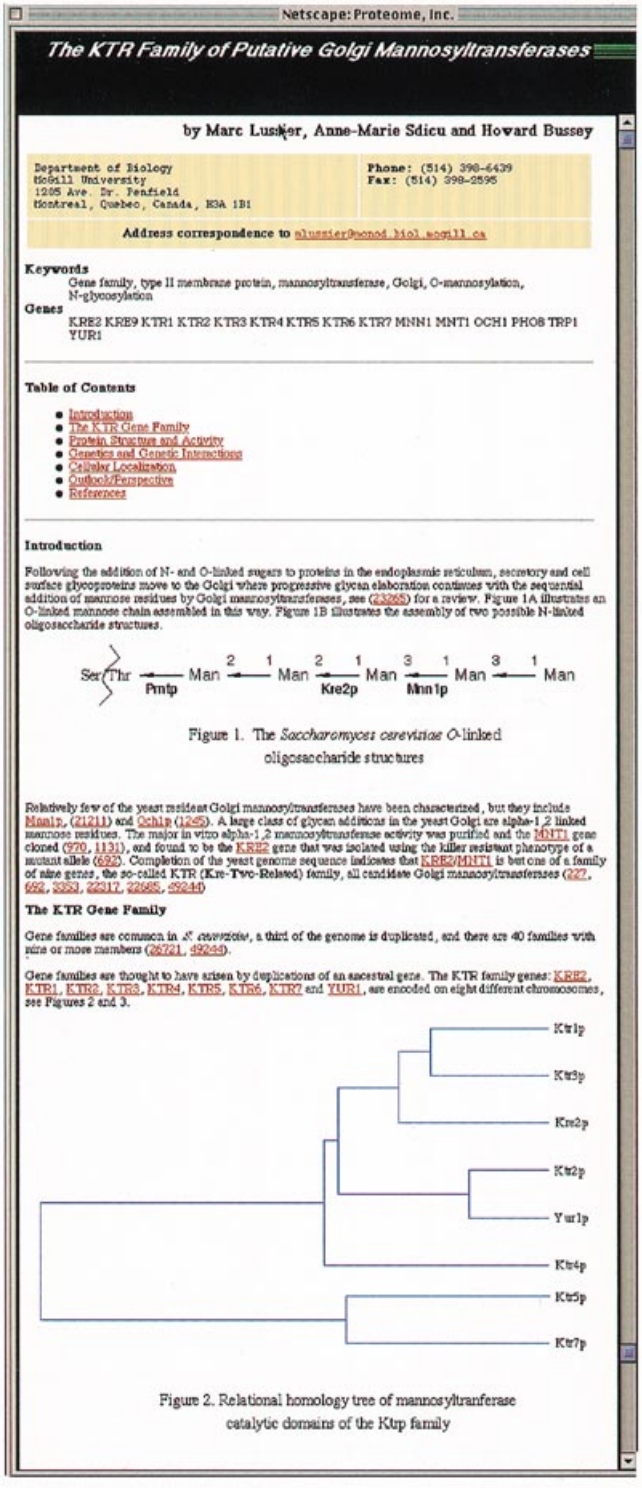
interested in accessing the Reports through the search form and summary pages available through the YPD home page (http://www.proteome.com/YPD/YPDhome.html). The YPD search form allows searches by gene names, keywords, protein properties, functional categories and amino acid sequence. Since the proteome of *S.cerevisiae* is complete, such searches take on an added power. For example, a search of protein kinases brings up every protein kinase that is encoded by the yeast genome (18).

YPD PROTEIN MINIREVIEWS: ADDING A NEW LAYER OF INFORMATION

The YPD database is built from separate Yeast Protein Reports, and within each report are long lists of annotations. To summarize these numerous, independent facts, and to present results that cover multiple related proteins, we have introduced the YPD Protein Minireviews. Each is a short summary with graphics covering a group of proteins related as part of a complex, pathway or protein family, and each is written by an expert. Each protein mentioned in a Minireview is linked to the corresponding Yeast Protein Report, and each reference citations is linked to PubMed. Keywords and the gene names of the proteins mentioned in each Minireview are indexed and searchable. Unlike a traditional review, the Minireviews are dynamic. The Yeast Protein Reports to which they link are updated daily, and each Minireview itself will be periodically updated to reflect the new information. An example shown in Figure 3, 'The KTR Family of Putative Golgi Mannosyltransferases' written by Marc Lussier, Anne-Marie Sdicu and Howard Bussey, is already available (http://www.proteome.com/YPD/YPM/mannosyltransferases.html), and many more will be added in the next year.

FUTURE DIRECTIONS FOR YPD

YPD is dedicated to completing the curation of all relevant, published data on yeast proteins, and to building a substantial collection of YPD Protein Minireviews within the next two years. We plan to continue to add new features, including a frequently updated table of related proteins from other model organisms. YPD is also expanding to encompass data from large-scale functional genomic and proteomic studies. Results of these projects will be integrated into the individual Yeast Protein Reports, with links to specialized Web pages posted by other groups. Already YPD has included the data from systematic projects to determine the phenotype of knockout mutants (2–4), two-hybrid interactions (6), the level of gene expression at the


Netscape: Proteome, Inc.

The KTR Family of Putative Golgi Mannosyltransferases

by Marc Lussier, Anne-Marie Sdicu and Howard Bussey

Department of Biology
McGill University
1205 Ave. Dr. Penfield
Montreal, Quebec, Canada, H3A 1B1

Phone: (514) 398-6439
Fax: (514) 398-2595

Address correspondence to mlussier@nonod.biol.mcgill.ca

Keywords
Gene family, type II membrane protein, mannosyltransferase, Golgi, O-mannosylation, N-glycosylation

Genes
KRE2 KRE9 KTR1 KTR2 KTR3 KTR4 KTR5 KTR6 KTR7 MNN1 MNT1 OCH1 PHO8 TRP1 YUR1

Table of Contents

- Introduction
- The KTR Gene Family
- Protein Structure and Activity
- Genetics and Genetic Interactions
- Cellular Localization
- Outlook/Perspective
- References

Introduction

Following the addition of N- and O-linked sugars to proteins in the endoplasmic reticulum, secretory and cell surface glycoproteins move to the Golgi where progressive glycan elaboration continues with the sequential addition of mannose residues by Golgi mannosyltransferases, see (23265) for a review. Figure 1A illustrates an O-linked mannose chain assembled in this way. Figure 1B illustrates the assembly of two possible N-linked oligosaccharide structures.

Ser/Thr —Man(2,1)— Man(2,1)— Man(3,1)— Man(3,1)— Man

Pmt*p* Kre2p Mnn1p

Figure 1. The *Saccharomyces cerevisiae* O-linked oligosaccharide structures

Relatively few of the yeast resident Golgi mannosyltransferases have been characterized, but they include Mnn1p, (21211) and Och1p (1245). A large class of glycan additions in the yeast Golgi are alpha-1,2 linked mannose residues. The major in vitro alpha-1,2 mannosyltransferase activity was purified and the MNT1 gene cloned (970, 1131), and found to be the KRE2 gene that was isolated using the killer resistant phenotype of a mutant allele (692). Completion of the yeast genome sequence indicates that KRE2/MNT1 is but one of a family of nine genes, the so-called KTR (Kre-Two-Related) family, all candidate Golgi mannosyltransferases (227, 692, 9353, 22317, 23685, 49244)

The KTR Gene Family

Gene families are common in *S. cerevisiae*, a third of the genome is duplicated, and there are 40 families with nine or more members (26721, 49244).

Gene families are thought to have arisen by duplications of an ancestral gene. The KTR family genes: KRE2, KTR1, KTR2, KTR3, KTR4, KTR5, KTR6, KTR7 and YUR1, are encoded on eight different chromosomes, see Figures 2 and 3.

Figure 2. Relational homology tree of mannosyltranferase catalytic domains of the Ktrp family

Figure 3. Portion of a representative YPD Protein Minireview. Users may access this Minireview at http://www.proteome.com/YPD/YPM/mannosyltransferases.html . The title and authors are given, with an Email reply directly to the authors. Keywords and a listing of the genes mentioned in the review are indexed at the beginning of the review. A table of contents lets the user skip to later sections. Each mention of a protein in the text is linked to that protein's Yeast Protein Report. References are linked to their PubMed abstract. A variety of graphics are presented to display the information.

RNA level (5), and direct protein analysis through mass spectroscopy (9) and two-dimensional gel electrophoresis (9,10). YPD has been designed to interface closely with two-dimensional gel analysis, both to help with interpretation of 2D gels (including spot identification) and as a repository of new information on protein expression and modification. A two-dimensional gel database (10) is underway that will be fully integrated with YPD.

A further role for YPD is as a model for new proteome databases. These can be generated in the near future for other microorganisms with sequenced genomes, using the same tools and methods that were developed for YPD. The development of proteome databases for other model organisms, the interconnection of these model organism databases, and the development of satellite databases for pathogenic microorganisms with links to the model organism databases will help to relate the power of genomic/proteomic approaches to biological and medical problems.

HOW TO SUBMIT PROTEIN DATA TO YPD

To supplement our curation, we appreciate feedback from our users on new data submission, additions and corrections, including personal communication of unpublished result (which will be cited as such). Any correspondence should be directed to ypd@proteome.com or by mail to the address of the authors.

CITING YPD

Authors who make use of the information provided by YPD should cite this article as a general reference for the access to and content of YPD. To cite a YPD Protein Minireview, please use the following format, or a similar format dictated by the journal of publication:

Lussier, M., Sdicu, A.-M., and Bussey, H. (1997) The KTR Family of Putative Golgi Mannosyltransferases (YPD Protein Minireviews, http://www.proteome.com/YPD/YPM/mannosyltransferases.html).

ACKNOWLEDGEMENTS

The development of the YPD has been funded by a Small Business Innovation Research grant, Phase II, (2 R44 GM54110-02) from the National Institute of General Medical Sciences. YPD relies on the expertise of a large number of scientists, and we are extremely grateful for their contributions. The quality of our annotation is due to the diligence of our curators and consultants: Maria Costanzo, Andrew McKee, Michael Cusick, David Gonda, Susan Bromberg, George Boguslawski and Rachelle Hecht. We thanks our panel of scientific experts for this contributions to the direction, scope and content of the database: Bruno André, Charles Cole, Michael Cusick, Les Grivell, Sepp Kohlwein and Jon Warner. We thank Marc Lussier, Anne-Marie Sdicu and Howard Bussey for contributing the initial YPD Protein Minireview. We thank our computer programmers Michael Benoit and Michael Tillberg, administrative assistant and reference librarian Cheryl Lengieza, and administrative assistant Shelley Lengieza. We appreciate the help and cooperative spirit of Mike Cherry, Cathy Ball, Caroline Adler, Selina Dwight, Shuai Weng and the staff at SGD, and Werner Mewes, Alfred Zollner, Jean Hani and the staff at MIPS. Most importantly, we thank our users in the scientific community. It is

their findings that make up YPD, and their comments, corrections and additions that keep it accurate.

REFERENCES

1 Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M. *et al.* (1996) *Science*, **274**, 546–567.
2 Smith, V., Chou, K.N., Lashkari, D., Botstein, D. and Brown, P.O. (1996) *Science*, **274**, 2069–2074.
3 Shoemaker, D.D., Lashkari, D.A., Morris, D., Mittmann, M. and Davis, R.W. (1996) *Nature Genet.*, **14**, 450–456.
4 http://genome.wustl.edu/gsc/yeast/chromosome8ORFs.html
5 Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Jr, Hieter, P., Vogelstein, B. and Kinzler, K.W. (1997) *Cell*, **88**, 243–251.
6 Fromont-Racine, M., Rain, J.C. and Legrain, P. (1997) *Nature Genet.*, **16**, 277–282.
7 Burns, N., Grimwade, B., Ross-Macdonald, P.B., Choi, E.Y., Finberg, K., Roeder, G.S. and Snyder, M. (1994) *Genes Dev.*, **8**, 1087–1105.
8 Niedenthal, R.K., Riles, L., Johnston, M. and Hegemann, J.H. (1996) *Yeast*, **12**, 773–786.
9 Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H. and Mann, M. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 14440–14445.
10 Garrels, J.I., McLaughlin, C.S., Warner, J.R., Futcher, B., Latter, G.I., Kobayashi, R., Schwender, B., Volpe, T., Anderson, D.S., Mesquita-Fuentes, R. and Payne, W.E. (1997) *Electrophoresis*, **18**, 1347–1360.
11 Garrels, J.I. (1996) *Nucleic Acids Res.*, **24**, 46–49.
12 Payne, W.E. and Garrels, J.I. (1997) *Nucleic Acids Res.*, **25**, 57–62.
13 http://www.ncbi.nlm.nih.gov/PubMed/
14 http://www.ncbi.nlm.nih.gov/Web/Genbank/
15 http://expasy.hcuge.ch/sprot/sprot-top.html
16 http://www.mips.biochem.mpg.de/mips/yeast/
17 http://genome-www.stanford.edu/Saccharomyces/
18 Hunter, T. and Plowman, G.D. (1997) *Trends Biochem. Sci.*, **22,** 18–22.