

FOR THE RECORD

New protein functions in yeast chromosome VIII

CHRISTOS OUZOUNIS,¹ PEER BORK,² GEORG CASARI, AND CHRIS SANDER

European Molecular Biology Laboratory–Heidelberg, D-69012 Heidelberg, Germany

(RECEIVED May 4, 1995; ACCEPTED September 6, 1995)

Abstract: The analysis of the 269 open reading frames of yeast chromosome VIII by computational methods has yielded 24 new significant sequence similarities to proteins of known function. The resulting predicted functions include three particularly interesting cases of translation-associated proteins: peptidyl-tRNA hydrolase, a ribosome recycling factor homologue, and a protein similar to cytochrome *b* translational activator CBS2. The methodological limits of the meaningful transfer of functional information between distant homologues are discussed.

Keywords: computational sequence analysis; function prediction; genome evolution; translation factors; yeast genome

The yeast genome sequencing project is making rapid progress, and the completion of the entire sequence is expected in the winter of 1995–1996 (Oliver et al., 1992; Dujon et al., 1994; Feldmann et al., 1994; Johnston et al., 1994; Bussey et al., 1995; Dietrich et al., 1995; Murakami et al., 1995; B.G. Barrell, pers. comm.; K. Kleine, P. Mordant, & A. Goffeau, pers. comm.). The original study of chromosome VIII (Johnston et al., 1994) reported the function of 46% of the probable 269 gene products, either as experimentally known or as predicted by exploiting evolutionary relationships to proteins of known function.

As of the submission date of this paper, we have been able to add functional information to another 9% of the putative ORFs, by searching more recent versions of sequence databases and using improved similarity search tools (Bork et al., 1992; Koonin et al., 1994; Scharf et al., 1994). As a result, the level of tentative functional assignment for chromosome VIII is now above 50% of all ORFs (Fig. 1).

Results and discussion: The newly identified tentative functions range from enzymes (e.g., chelate synthase, glutaminase-asparaginase, gamma-butyrobetaine hydroxylase, 2-deoxyglucose-

6-phosphate phosphatase, NH(3)-dependent NAD(+) synthetase) to regulatory proteins (e.g., hexose transport activator protein, putative BTF3-like transcription factor). For three particularly interesting cases, we include alignments (Fig. 2) and a brief description.

YHR189w, a putative peptidyl-tRNA hydrolase: PTH (EC 3.1.1.29) is a well-characterized cytoplasmic enzyme that cleaves peptidyl-tRNA or *N*-acyl-aminoacyl-tRNA to yield free peptides of *N*-acyl-amino acid and tRNA; its natural substrates probably are peptidyl-tRNAs that drop off the ribosome during pro-

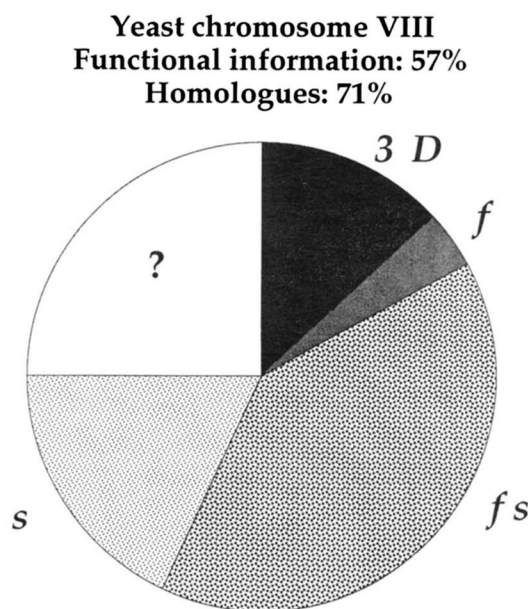


Fig. 1. Information clock for the function assignment and homology detection for yeast chromosome VIII. Information decreases clockwise. 3D, homology implies known 3D structure, and function is known (mostly from homology); *f*, function known from experimentation; *fs*, function predicted based on significant sequence similarity to a protein in the database; *s*, significant sequence similarity to a protein of undetermined function; *?*, no significant sequence similarity to any database protein. Note the large proportion of structural and functional homologues.

¹ Present address: AI Center, SRI International, Menlo Park, California 94025.

² Present address: Max-Delbrueck Center for Molecular Medicine, D-13112 Berlin, Germany.

Abbreviations: RRF, ribosome releasing factor (ribosome recycling factor); PTH, peptidyl-tRNA hydrolase; CBS2, cytochrome *b* translational activator; ApbA, a protein in a genetic locus involved in thiamine biosynthesis; ORF, open reading frame.

tein synthesis (Garcia-Villegas et al., 1991; Meinel et al., 1993, and references therein). PTH is a rare enzyme with about 25 molecules per cell (Dutka et al., 1993). There are several hypotheses regarding the role of PTH in cell growth.

YHR189w and *Escherichia coli* PTH have about 200 amino acids each and are 27% identical over 183 residues (Blastp score 85, probability $P(N) 7.9 \times 10^{-13}$). The significance of this similarity is strongly supported by the conservation pattern in a multiple alignment of several other prokaryotic proteins related to PTH (Fig. 2A). The sequences cover a wide species spectrum within bacteria and have diverged considerably (less than 36% identical residues between different prokaryotic PTHs). We predict that all of the proteins in Figure 2A have peptidyl-tRNA hydrolase function.

YHR038w, a ribosome recycling factor homologue: Ribosome releasing factor is responsible for the dissociation of ribosomes from mRNA after translation termination (Ichikawa & Kaji, 1989) and is also called ribosome recycling factor (Janosi et al., 1994). RRF appears to be essential for growth in *E. coli* (Janosi et al., 1994). Searching databases with YHR038w, the nuclear protein D2 from the plant *Daucus carota* (S. Schrader, R. Kaldenhoff, & G. Richter, 1993, unpubl.) is identified as clearly similar (Blastp score 77, probability $P(N) 6 \times 10^{-6}$), whereas its well-characterized bacterial homologue RRF is less similar, at 23% sequence identity (Blastp score 50, probability $P(N) 3 \times 10^{-4}$). Both similarities were confirmed by sensitive motif searches (Tatusov et al., 1994) with conserved regions.

The implication of these similarities is that RRF also occurs in plants (Fig. 2B), and we therefore predict that this final step of translation is ubiquitous in eubacteria and various eukaryotic phyla. The evolutionary distances do not allow, however, an unambiguous functional prediction, as the plant sequence is much more closely related (45% identity) to *E. coli* than to the yeast homologue (23% identity).

YHR063c, a mitochondrial translation activator? The protein sequence of YHR063c is similar (25% identity) to ApbA protein from *Salmonella typhimurium* LT2, which is involved in thiamine biosynthesis (Downs & Petersen, 1994), and similar to CBS2, a nuclear-encoded mitochondrial protein on chromosome IV of *Saccharomyces cerevisiae* (Michaelis et al., 1988) involved in the translational activation of mitochondrial cytochrome *b* mRNA (Michaelis & Rödel, 1990). The similarity of YHR063c with ApbA (Blastp score 72, probability $P(N) 8 \times 10^{-8}$; Fig. 2C) is stronger than that with CBS2 (Blastp score 69, probability $P(N) 3 \times 10^{-7}$); a motif search in the sequence database (Tatusov et al., 1994) exclusively matches the three related proteins with a strict cutoff value of 0.005.

Despite the incomplete functional characterization of ApbA, we predict, based on the data from CBS2, that these proteins represent a family of factors that are involved in (mitochondrial) translation. Organelle genomes have been significantly reduced during evolution (Gray, 1993), and their translation mechanisms mostly resemble prokaryotic systems. Thus, CBS2 or YHR063c may even be nuclear-encoded mitochondrial proteins of pro-

Table 1. Novel functions in yeast chromosome VIII^a

Query	Access	Identifier	Protein family and/or predicted function
YHL021c	P80193	BODG_PSESK	Gamma-butyrobetaine hydroxylase
YHL003c	P28496	YKA8_YEAST	TRAM/UOG1 family
YHR003c	P36101	YKC7_YEAST	THIF/HESA/MOEB family
YHR032w	D24172	OSR14662A	DNA damage inducible Dinf-like
YHR038w	P16174	RRF_ECOLI	• Ribosome releasing factor
YHR043c	U00062	YSCH8179_15	* 2-Deoxyglucose-6-phosphate phosphatase
YHR044c	U00062	YSCH8179_14	* 2-Deoxyglucose-6-phosphate phosphatase
YHR058c	P13511	CZCA_ALCEU	Cation efflux system proteins
YHR063c	P37402	APBA_SALTY	• APBA cytochrome <i>b</i> translational activator
YHR074w	Q03638	NADE_RHOCA	NH(3)-dependent NAD(+) synthetase
YHR075c	Q03565	YKD7_CAEEL	Lipase/hydroxylase family; chelate synthase
YHR090c	U13948	HSU13948_1	Zn-binding PHD-finger
YHR093w	U00060	None	* Hexose transport activator protein
YHR099w	L34075	HUMFRAPX_1	FKBP-rapamycin associated protein
YHR115c	Q06003	GOLI_DROME	RING finger
YHR137w	X78503	RMMOCCABR_6	rocR gene product
YHR138c	P01095	IPB2_YEAST	* Protease B inhibitor 2
YHR139c	S33203	S33203	Glutaminase-asparaginase
YHR146w	Z14127	SCSPM1	Glucose repression protein GAL83
YHR154w	X62676	SPRAD4_1	RAD4 gene product (<i>S. pombe</i>)
YHR160c	P08468	PT11_YEAST	PET111 protein precursor
YHR161c	S27867	S27867	Clathrin assembly phosphoprotein
YHR189w	P23932	PTH_ECOLI	• * Peptidyl-tRNA hydrolase
YHR193c	Z28479	HSB20A092	Putative transcription factor, BTF3-like

^a Query: identifier of the yeast chromosome VIII ORF. Access, Identifier: database accession number and identifier of the closest homologue. Protein family and/or predicted function: family or function as deduced from the detected homology. Bullets indicate the three cases presented in detail. Asterisks indicate the cases that are detected automatically by standard software.

A	Pth_Ecoli	65
	Sp5c_Bacsu	62
	Pth/Bb	62
	Pth/Mg	0
	Pth/Ct	38
	Yhr189w	70
	Consensus	70
	Pth_Ecoli	134
	Sp5c_Bacsu	130
	Pth/Bb	130
B	Nip_Dauca	0
	Rrf_Ecoli	14
	Yhr038w	60
	Consensus	60
	Nip_Dauca	55
	Rrf_Ecoli	73
	Yhr038w	116
	Consensus	120
	Nip_Dauca	107
	Rrf_Ecoli	125
	Yhr038w	176
	Consensus	180
	Nip_Dauca	167
	Rrf_Ecoli	185
	Yhr038w	230
	Consensus	240

Fig. 2. See caption on facing page.

C	Apba_Salty	59
	Yhr063c	78
	Cbs2_Yeast	79
C	Consensus	82
	Apba_Salty	117
	Yhr063c	152
	Cbs2_Yeast	161
C	Consensus	164
	Apba_Salty	166
	Yhr063c	218
	Cbs2_Yeast	243
C	Consensus	246
	Apba_Salty	222
	Yhr063c	296
	Cbs2_Yeast	312
C	Consensus	328
	Apba_Salty	281
	Yhr063c	378
	Cbs2_Yeast	389
C	Consensus	410
	Apba_Salty	
	Yhr063c	
	Cbs2_Yeast	

Fig. 2. Alignments for three translation-associated proteins. **A:** Alignment of YHR189w with the *E. coli* peptidyl-tRNA hydrolase and other bacterial homologues. Sequence positions are given at the right margin. The consensus line is calculated on a similarity basis. Boxed residues represent similarity with the consensus. Four boxes of conserved and polar residues are candidate functional regions. Figure generated by PrettyPlot (by P. Rice). Identifiers and names (database: accession number): Pth_Ecoli, *E. coli* peptidyl-tRNA hydrolase (Swiss-Prot: P23932); Sp3c_Bacsu, *Bacillus subtilis* stage V sporulation protein C (Swiss-Prot: P37470); Pth/Bb, *Borrelia burgdorferi* PTH homologue (EMBL: L32144); Pth/Mg, *Mycoplasma genitalium* fragment (EMBL: U02185); and Pth/Ci, *Clamydia trachomatis* hypothetical protein fragment (PIR: C37840). **B:** Alignment of YHR038w with the *E. coli* ribosome releasing factor (Rrf_Ecoli, Swiss-Prot: P16174) and the *Daucus carota* protein D2 (Nlp_Dauca, Swiss-Prot: P37706). **C:** Alignment of YHR063c with Apba from *S. typhimurium* (Apba_Salty, Swiss-Prot: P37402) and Cbs2 from yeast (Cbs2_Yeast, Swiss-Prot: P14905).

karyotic origin transferred from the endosymbiotic mitochondrial genome to the nuclear genome (Thorsness & Fox, 1990; Nugent & Palmer, 1991), where they have further duplicated.

Conclusions: For newly sequenced yeast proteins, the rate of homology detection in current sequence databases has surpassed the 70% level (Bork et al., 1994; unpubl. results). The level of identification of probable function from database similarity searches is more difficult to estimate. The proposed functions (Table 1) are, for the most part, the functions of related database proteins. However, depending on the species context and evolutionary distance, the similarity of function between related proteins varies substantially. The assigned functions should, therefore, be treated as plausible hypotheses, ranging from certain to approximate. With this caveat in mind, we conclude that the level of plausible functional identification by bioinformatics methods will surpass 60% as the yeast genome sequence is completed. For some classes of proteins, however, only experimental work can ultimately elucidate function.

Data and methods: The 269 sequences of putative proteins from yeast chromosome VIII (Johnston et al., 1994) were obtained from the *Saccharomyces* Genomic Information Resource at Stanford University School of Medicine via anonymous ftp. Each single protein sequence was subjected to a variety of homology search tools that retrieve information from numerous sequence databases (for details, see Bork et al., 1992; Koonin et al., 1994) using GeneQuiz (Scharf et al., 1994). For initial searches, programs of the Blast series (Altschul et al., 1990, 1994) were used with default parameters (including masking of compositionally biased regions; Altschul et al., 1994). Using ClustalW (Thompson et al., 1994), multiple alignments were constructed and evolutionary trees calculated. Weak sequence similarities were inspected and distant homologies verified by pattern search procedures (Tatusov et al., 1994).

After all ORFs had been analyzed, entries were annotated (by assigning features of interest, such as phylum information or functional classes) using a relational database system and expert modules incorporated into GeneQuiz (Scharf et al., 1994). A complete list of functions assigned to proteins from yeast chromosome VIII will be accessible via Internet using the URL <http://www.sander.embl-heidelberg.de/genequiz/yeast/chromosome8>. The searches reported here are valid as of November 10, 1994.

References

- Altschul SF, Boguski MS, Gish W, Wootton JC. 1994. Issues in searching molecular sequence databases. *Nature Genet* 6:119–163.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.
- Bork P, Ouzounis C, Sander C. 1994. From genome sequences to protein function. *Curr Opin Struct Biol* 4:393–403.
- Bork P, Ouzounis C, Sander C, Scharf M, Schneider R, Sonnhammer E. 1992. Comprehensive sequence analysis of the 182 ORFs of yeast chromosome III. *Protein Sci* 1:1677–1690.
- Bussey H, Kaback DB, Zhong W, Vo DT, Clark MW, Fortin N, Hall J, Ouellette BF, Keng T, Barton AB, et al. 1995. The nucleotide sequence of chromosome I from *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 92:3809–3813.
- Dietrich FS, Mulligan J, Hennesy K, Allen E, Araujo R, Aviles E, Berno A, Brennan T, Carpenter J, Chen E, Cherry JM, Chung E, Duncan M, Guzman E, Hartzell G, Hunnicke-Smith S, Hyman R, Kayser A, Komp C, Lashkari D, Lew H, Lin D, Mosedale D, Nakahara K, Namath A, Norgren R, Oefner P, Oh C, Petel FX, Roberts D, Sehl P, Schramm S, Shogren T, Smith V, Taylor P, Wei Y, Yelton M, Botstein D, Davis R. 1995. The complete sequence of *Saccharomyces cerevisiae* chromosome V. Forthcoming.
- Downs DM, Petersen L. 1994. *apbA*, a new genetic locus involved in thiamine biosynthesis in *Salmonella typhimurium*. *J Bacteriol* 176:4858–4864.
- Dujon B, Alexandraki D, et al. 1994. Complete sequence of yeast chromosome XI. *Nature* 369:371–378.
- Dutka S, Meinnel T, Lazennec C, Mechulam Y, Blanquet S. 1993. Role of the 1–72 base pair in tRNAs for the activity of *E. coli* peptidyl-tRNA hydrolase. *Nucleic Acids Res* 21:4025–4030.
- Feldmann H, Aigle M, et al. 1994. Complete DNA sequence of yeast chromosome II. *EMBO J* 13:5795–5809.
- Garcia-Villegas MR, De La Vega FM, Galindo JM, Segura M, Buckingham RH, Guarneros G. 1991. Peptidyl-tRNA hydrolase is involved in λ inhibition of host protein synthesis. *EMBO J* 10:3549–3555.
- Gray MW. 1993. Origin and evolution of organelle genomes. *Curr Opin Genet Dev* 3:884–890.
- Ichikawa S, Kaji A. 1989. Molecular cloning and expression of ribosome releasing factor. *J Biol Chem* 264:20054–20059.
- Janosi L, Shimizu I, Kaja A. 1994. Ribosome recycling factor (ribosome releasing factor) is essential for bacterial growth. *Proc Natl Acad Sci USA* 91:4249–4253.
- Johnston M, Andrews S, et al. 1994. Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome VIII. *Science* 265:2077–2082.
- Koonin EV, Bork P, Sander C. 1994. Yeast chromosome III: New gene functions. *EMBO J* 13:493–504.
- Meinnel T, Mechulam Y, Blanquet S. 1993. Methionine as translation start signal: A review of the enzymes of the pathway in *E. coli*. *Biochimie* 75:1061–1075.
- Michaelis U, Rödel G. 1990. Identification of CBS2 as a mitochondrial protein in *Saccharomyces cerevisiae*. *Mol Gen Genet* 223:394–400.
- Michaelis U, Schlapp T, Rödel G. 1988. Yeast nuclear gene CBS2, required for translational activation of cytochrome *b*, encodes a basic protein of 45 kDa. *Mol Gen Genet* 214:263–270.
- Murakami Y, Naitou M, Hagiwara H, Shibata T, Ozawa M, Sasanuma SI, Sasanuma M, Tsuchiya Y, Soeda E, Yokoyama K, Yamazaki M, Tashiro H, Eki T. 1995. Analysis of the nucleotide sequence of chromosome VI from *Saccharomyces cerevisiae*. *Nature Genet* 10:261–268.
- Nugent JM, Palmer JD. 1991. RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. *Cell* 66:473–481.
- Oliver SG, van der Aart QJM, et al. 1992. The complete DNA sequence of yeast chromosome III. *Nature* 357:38–46.
- Scharf M, Schneider R, Casari G, Bork P, Valencia A, Ouzounis C, Sander C. 1994. GeneQuiz: A workbench for sequence analysis. In: Altman R, Brutlag D, Karp P, Lathrop R, Searls D, eds. *Proc 2nd Intl Conf Intell Syst Mol Biol*. Menlo Park, California: AAAI Press. pp 348–353.
- Tatusov RL, Altschul SF, Koonin EV. 1994. Detection of conserved segments in proteins: Iterative scanning of sequence databases with alignment blocks. *Proc Natl Acad Sci USA* 91:12091–12095.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
- Thorsness PE, Fox TD. 1990. Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. *Nature* 346:376–379.