



Yeast Sequencing Reports

Sequence and Analysis of a 36.2 kb Fragment from the Right Arm of Yeast Chromosome XV Reveals 19 Open Reading Frames Including *SNF2* (5' end), *CPA1*, *SLY41*, a Putative Transport ATPase, a Putative Ribosomal Protein and an *SNF2* Homologue

RÉMY POIREY, CELINA CZIEPLUCH, EDDA TOBIASCH, AURORA PUJOL, ELISABETH KORDES AND JEAN-CLAUDE JAUNIAUX*

Angewandte Tumorstudiologie, Abteilung 0610 and Virologie Appliquée à l'Oncologie (Unité INSERM 375), Deutsches Krebsforschungszentrum, P. 101949, D-69009 Heidelberg, Germany

Received 15 August 1996; accepted 21 October 1996

The complete sequence of a 36 196 bp DNA segment located on the right arm of chromosome XV of *Saccharomyces cerevisiae* has been determined and analysed. The sequence includes the 5' coding region of the *SNF2* gene, the *CPA1* leader peptide sequence and 17 open reading frames (ORFs) of at least 100 amino acids. Two of these correspond to previously known genes (*CPA1*, *SLY41*), whereas 15 correspond to new genes. The putative translation products of three ORFs show significant similarity with known proteins: one is a putative transport ATPase, another appears to be a ribosomal protein, and the third is an Snf2p homologue. The sequence has been deposited in the EMBL databank under Accession Numbers: Z75198, Z75199, Z75200, Z75201, Z75202, Z75203, Z75204, Z75205, Z75206, Z75207, Z75209, Z75210, Z75211, Z75212, Z75213, Z75214, Z75215, Z75216. © 1997 by John Wiley & Sons, Ltd.

Yeast 13: 479–482, 1997.

No. of Figures: 0. No. of Tables: 1. No. of References: 19.

KEY WORDS — *Saccharomyces cerevisiae*; chromosome XV; *SNF2*; *CPA1*; *SLY41*; ATPase; ribosomal protein

INTRODUCTION

As participants in the European Union BIOTECH project for sequencing the genome of *Saccharomyces cerevisiae*, we have determined the nucleotide sequence of a 36 196 kb DNA fragment located on the right arm of chromosome XV. The sequence reveals four previously described genes and 15 novel open reading frames (ORFs).

*Correspondence to: Jean-Claude Jauniaux.
Contract grant sponsor: European Union (BIOTECH programme).

MATERIALS AND METHODS

Cosmid, strains and sequencing strategy

The pEOA487 cosmid was provided by B. Dujon (Institut Pasteur, Paris). It contains a 36.2 kb chromosome XV *Sau3A* DNA fragment from *S. cerevisiae* strain FY1679 (isogenic to S288C) inserted at the *Bam*H1 site of cosmid vector pWE15 (Thierry *et al.*, 1995). Cosmid pEOA487 DNA was sonicated, repaired with T4 DNA polymerase and Klenow fragment of DNA polymerase I. Fragments of 800–1200 bp were

isolated after gel electrophoresis, ligated into the pSEQ1 vector and used to transform *Escherichia coli* JM109 cells by electroporation. Cosmid pEOA487 chromosome XV-specific sub-clones were sorted out by colony hybridization using the radiolabeled 36 kb yeast insert as a probe. Double-stranded DNAs were prepared using Qiagen columns. Sequencing was performed on both strands with T7 DNA polymerase, [³⁵S]dATP, universal and reverse primers. Synthetic oligonucleotides were used as primers to complete the sequence. The 36 196 bp insert has been totally sequenced on both strands.

Computer sequence analysis

The DNASTAR software (Lasergene, Madison, Wisconsin) was used for sequence editing and assembling. Coding regions were sought with the help of the DNA Strider 1.2 software (Marck, 1988). For each ORF, the first ATG was assumed to be the translation initiation codon. Sequence analysis and search for homologies were performed by K. Kleine at MIPS (Martinsrieder Institut für ProteinSequenzen) and by us with the HUSAR software (Heidelberg Unix Sequence Analysis Resources; Senger *et al.*, 1995) based on the GCG Wisconsin package (Devereux *et al.*, 1984), which includes the FASTA software (Pearson and Lipman, 1988) with the data libraries SwissProt (release 33.0+SwNew 6/96), PIR (release 48.0), MIPSX/PATCHX (release 48.0), GenBank (release 95.0) and EMBL (release 47.0). A FASTA score below 150 was considered to be not significant. ProSite (release 13.0) was used to search for consensus patterns in the putative new proteins. Sites found at a very high frequency in almost every translated ORF, for example putative phosphorylation and/or myristoylation sites, were not reported.

RESULTS AND DISCUSSION

Sequence analysis

We have determined the sequence of the 36 196 bp insert carried by the cosmid pEOA487. Computer analysis of the sequence revealed 18 ORFs containing at least 100 contiguous sense codons and not entirely contained in a longer ORF on either DNA strand. The ORFs located at the extremities continue out of the sequenced fragment. To those ORFs we added the 25-codon ORF encoding the leader peptide of Cpa1p, which

we identified in the pEOA487 cosmid. All but one of the ORFs (YOR293w, see below) have low codon adaptation index (CAI) values (Sharp and Li, 1987), suggesting a low expression level. A search for putative introns (Kalogeropoulos, 1995) identified one in YOR293w (see below). Table 1 includes the names, positions and size of the ORFs, the CAI (Sharp and Li, 1987), the predicted transmembrane segments (Klein *et al.*, 1985), the Prosite and FASTA results and characteristics of the identified ORFs.

Previously described genes

We have identified four previously sequenced genes in our sequence: *SNF2* (*SWI2/GAM1/TYE3/HAF1*, 5'end), which encodes a component of the SWI/SNF global transactivation complex (Yoshimoto and Yamashita, 1991); YOR302w, which encodes the Cpa1p leader peptide (Werner *et al.*, 1987); *CPA1*, the carbamoyl-phosphate synthase (glutamine-hydrolysing) small-chain gene of the arginine pathway (Werner *et al.*, 1985) and *SLY41*, which encodes a protein involved in the secretory pathway (Dascher *et al.*, 1991).

ORFs with similarity to known proteins

YOR291w exhibits the characteristic features of transport ATPases. It contains a motif identical to the consensus pattern of E1-E2 ATPases phosphorylation site (amino acids (aa) 781–787) and ten transmembrane spans are predicted by the method of Klein *et al.* (1985). Moreover, it shows strong aa sequence similarity (FASTA: 1071) with a P-type cation-translocating ATPase from *Plasmodium falciparum* (Krishna *et al.*, 1993) and weaker similarities with several other transport ATPases present in the databases (Catty and Goffeau, 1996). So, YOR291w probably encodes a transport ATPase.

YOR292c encodes a putative product of 309 aa which shows a weak homology (23% over 192 aa) with the human glomerulosclerosis protein Mpv17 (Karasawa *et al.*, 1993).

YOR293w contains a putative intron of 262 bp detected by the Explora program (Kalogeropoulos, 1995). The YOR293w putative gene product is 105 aa long. It is almost identical to the YMR230w product of chromosome XVIII (S5797, 97% identity) and shows striking similarity with the superfamily of ribosomal proteins S10.2. The CAI of this ORF is very high (0.84), as expected for a ribosomal protein.

Table 1. Localization and characteristics of the ORFs identified in the 36 196 bp DNA insert of cosmid pEOA487.

ORF	Position (bp)	CAI	Size (aa)	FASTA score	Identity (%)	Genes, homologies, motifs
YOR290c	2–925		>308		100% in 308 aa	SNF2, component of SWI/SNF global transactivator complex
YOR291w	1842–6257	0.14	1472	1071	35% in 1208 aa	p-type cation translocating ATPase E1-E2 ATPases phosphorylation site Transmembrane span (10)
YOR292c	7249–6323	0.10	309	209	23% in 192 aa	Human hMpV17 protein Transmembrane span (2)
YOR293w	ORF: 7765–8516 Intron: 7816–8254	0.84	105	548	97% in 105 aa	YMR230w, chromosome XIII ORF ribosomal S10. e protein family
YOR294w	9007–9615	0.26	203	314	39% in 117 aa	<i>Homo sapiens</i> ORF of unknown function
YOR295w	9875–10558	0.11	228	475	38% in 226 aa	YMR233w, chromosome XIII ORF
YOR296w	10869–14735	0.11	1289			
YOR297c	15413–15988	0.11	192	247	33% in 165 aa	SDH4, <i>S. cerevisiae</i> succinate dehydrogenase anchor subunit Transmembrane span (1)
YOR298w	16266–17702	0.13	479			
YOR299w	19099–21336	0.11	746	1804	55% in 746 aa	YMR237w, chromosome XIII ORF
YOR300w	21239–21544	0.12	102			
YOR301w	21629–22933	0.09	435			Transmembrane span (3)
YOR302w	23429–23503	0.30	25	133	100% in 25 aa	CPA1 leader peptide
YOR303w	23563–24795	0.23	411	2000	99.7% in 411 aa	CPA1, carbamoyl-phosphate synthase (glutamine-hydrolysing) small chain
YOR304w	25178–28537	0.17	1120	2426	52% in 936 aa	SNF2, component of SWI/SNF global transactivator complex ATP/GTP-binding site motif A
YOR305c	29686–30411	0.11	242			
YOR306c	30534–32096	0.16	521	420	25% in 521 aa	JTE673, chromosome × chromosome XIII ORF Human X-linked PEST-containing transporter Transmembrane spans (10)
YOR307c	33397–34755	0.12	453	2264	99.5% in 453 aa	SLY41, involved in secretory pathway
YOR308c	35290–36195		>303			

The codon adaptation index (CAI) was calculated as described by Sharp and Li (1987); FASTA scores are given for the identical or best-related sequence; FASTA scores below 150 are not indicated; w and c indicate the direction of transcription (Watson and Crick strands); aa, amino acids. Motifs and putative transmembrane spans detected using the method of Klein *et al.* (1985) are described only for the newly identified ORFs.

YOR304w encodes a putative product which shows strong similarity (52% identity) with Snf2p, a component of the Swi/Snf global transactivator complex. It contains an ATP/GTP-binding site motif A (aa 719–726) as well.

ORFs of unpredicted function

The putative product of YOR294w shows a weak homology with a human ORF of unknown function (D25218).

The putative product of YOR295w shows significant similarities to an ORF of unknown

function, YMR233w (57600), located on chromosome XIII (38% identity over 226 aa).

YOR297w shows similarities to *SDH4*, which encodes the succinate dehydrogenase membrane anchor subunit of SDH2 (33% identity over 165 aa; Bullis and Lemire, 1994). The putative product of YOR299w shows strong similarities to an ORF of unknown function, YMR237w (S57604), located on chromosome XIII (55% identity over 746 aa). The putative product of YOR306c shows an important homology (25% identity) with the ORF of unknown function, YOL119c (S66816),

located on chromosome XV. Ten transmembrane segments are predicted by the method of Klein *et al.* (1985). Weaker similarities have been detected with the human X-linked PEST-containing transporter (Lafreniere *et al.*, 1994) and with the human monocarboxylate transporter 1 (Garcia *et al.*, 1994). Taken together, these data suggest that YOR306c encodes an integral membrane polypeptide involved in transport.

The six remaining ORFs (YOR296w, YOR298w, YOR300w, YOR301w, YOR305c and YOR308c) show no homology to other protein sequences contained in data libraries (FASTA<150).

ACKNOWLEDGEMENTS

We thank Prof. B. Dujon (Institut Pasteur, Paris) for providing cosmid pEOA487, Drs K. Kleine (MIPS) and K. Glatting (DKFZ) for their help in computer analysis, A. Grewenig for excellent technical assistance, and R. Slater and B. Volk for their assistance in sequencing and DNA preparation. We are grateful to Prof. Rommelaere for his support throughout this work. This research was supported by the European Union BIOTECH programme.

REFERENCES

- Bullis, B. and Lemire, B. (1994). Isolation and characterization of the *Saccharomyces cerevisiae* *SDH4* gene encoding a membrane anchor subunit of succinate dehydrogenase. *J. Biol. Chem.* **269**, 6543–6549.
- Catty, P. and Goffeau, A. (1996). Identification and phylogenetic classification of eleven putative P-type calcium transport ATPase genes in yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Bioscience Reports* **16**, 75–85.
- Dascher, C., Ossig, R., Gallwitz, D. and Schmitt, H. D. (1991). Identification and structure of four yeast genes (SLY) that are able to suppress the functional loss of *YPT1*, a member of the RAS superfamily. *Mol. Cell. Biol.* **11**, 872–885.
- Devereux, J., Haeblerli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387–395.
- Dujon, B., *et al.* (1994). Complete DNA sequence of yeast chromosome XI. *Nature* **369**, 371–378.
- Garcia, C. K., Li, X., Luna, J. and Francke, U. (1994). cDNA cloning of the human monocarboxylate transporter 1 and chromosomal localization of the SLC16A1 locus to 1p13.2-p12. *Genomics* **23**, 500–503.
- Kalogeropoulos, A. (1995). Automatic intron detection in nuclear DNA sequences of *Saccharomyces cerevisiae*. *Yeast* **11**, 555–565.
- Karasawa, M., Zwacka, R. M., Reuter, A., *et al.* (1993). The human homolog of the glomerulosclerosis gene *MPV17*: structure and genomic organization. *Hum. Mol. Genet.* **2**, 1829–1834.
- Klein, P., Kanehisa, M. and DeLisi, C. (1985). The detection and classification of membrane-spanning proteins. *Biochim. Biophys. Acta* **815**, 468–476.
- Krishna, S., Cowan, G., Meade, J. C., Wells, R. A., Stringer, J. R. and Robson, K. (1993). A family of cation ATPase-like molecules from *Plasmodium falciparum*. *J. Cell Biol.* **120**, 385–398.
- Lafreniere, R., Carrel, L. and Willard H. (1994). A novel transmembrane transporter encoded by the *XPCT* gene in Xq13.2. *Hum. Mol. Genet.* **3**, 1133–1140.
- Marck, C. (1988). 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucl. Acids Res.* **16**, 1829–1836.
- Pearson, W. R. and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
- Senger, M., Glatting, K. H., Ritter, O. and Suhai, S. (1995). X-HUSAR, an X-based graphical interface for the analysis of genomic sequences. *Comput. Methods Programs Biomed.* **46**, 131–141.
- Sharp, P. M. and Li, W. H. (1987). The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucl. Acids Res.* **15**, 1281–1295.
- Thierry, A., Gaillon, L., Galibert, F. and Dujon, B. (1995). Construction of a complete genomic library of *Saccharomyces cerevisiae* and physical mapping of chromosome XI at 3·7 kb resolution. *Yeast* **11**, 121–135.
- Werner, M., Feller, A., Messenguy, F. and Piérard, A. (1987). The leader peptide of yeast gene *CPA1* is essential for the translational repression of its expression. *Cell* **49**, 805–813.
- Werner, M., Feller, A. and Pierard, A. (1985). Nucleotide sequence of yeast gene *CPA1* encoding the small subunit of arginine-pathway carbamoyl-phosphate synthetase: homology of the deduced amino acid sequence to other glutamine amidotransferases. *Eur. J. Biochem.* **146**, 371–381.
- Yoshimoto, H. and Yamashita, I. (1991). The *GAM1/SNF2* gene of *Saccharomyces cerevisiae* encodes a highly charged nuclear protein required for transcription of the *STA1* gene. *Mol. Gen. Genet.* **228**, 270–280.