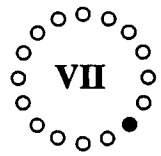
YEAST VOL. 12: 273–280 (1996)



Yeast Sequencing Reports

Sequencing of a 17·6 kb Segment on the Right Arm of Yeast Chromosome VII Reveals 12 ORFs, Including *CCT*, *ADE3* and *TR-I* Genes, Homologues of the Yeast *PMT* and *EF1G* Genes, of the Human and Bacterial Electron-Transferring Flavoproteins (β-chain) and of the *Escherichia coli* Phosphoserine Phosphohydrolase, and Five New ORFs

PAULO GUERREIRO, TANIA BARREIROS, HELENA SOARES, LUISA CYRNE‡, ALEXANDRA MAIA E SILVA AND CLAUDINA RODRIGUES-POUSADA*

Laboratório de Genética Molecular, Instituto Gulbenkian de Ciência, Apartado 14, P-2781 Oeiras Codex, Portugal
‡Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade de Lisboa, P-1700 Lisboa, Portugal

Received 4 May 1995; accepted 28 August 1995

A 17·6 kb DNA fragment from the right arm of chromosome VII of *Saccharomyces cerevisiae* has been sequenced and analysed. The sequence contains twelve open reading frames (ORFs) longer than 100 amino acids. Three genes had already been cloned and sequenced: *CCT*, *ADE3* and *TR-I*. Two ORFs are similar to other yeast genes: *G7722* with the *YAL023* (*PMT2*) and *PMT1* genes, encoding two integral membrane proteins, and *G7727* with the first half of the genes encoding elongation factors 1γ, *TEF3* and *TEF4*. Two other ORFs, *G7742* and *G7744*, are most probably yeast orthologues of the human and *Paracoccus denitrificans* electron-transferring flavoproteins (β chain) and of the *Escherichia coli* phosphoserine phosphohydrolase. The five remaining identified ORFs do not show detectable homology with other protein sequences deposited in data banks. The sequence has been deposited in the EMBL data library under Accession Number Z49133.

KEY WORDS — *Saccharomyces cerevisiae*; chromosome sequencing; chromosome VII; *CCT*; *ADE3*; *TR-1*; *PMT*; *EF19*

INTRODUCTION

As a part of the BRIDGE/BIOTECH program of the European Union to sequence the *Saccharomyces cerevisiae* genome, we have determined the sequence of a 17,640 bp DNA fragment of the right arm of chromosome VII. The sequenced region is a part of a 26,229 bp yeast segment inserted into the pWE15 cosmid that partially overlaps the insert of cosmid pEGH059 sequenced by C. Nombela's group in Spain (Guerreiro *et al.*, 1995).

*Corresponding author.

CCC 0749–503X/96/030273-08

The fragment includes the *ADE3* locus and has revealed the presence of two other previously sequenced genes (*CCT* and *TR-I*), four sharing homology with other genes from yeast or from other organisms, and five unknown genes, of a total of 12 complete open reading frames (ORFs) detected.

MATERIALS AND METHODS

Strains and plasmids

The pEGH620 cosmid was provided by H. Tettelin and A. Goffeau (DNA coordinators of

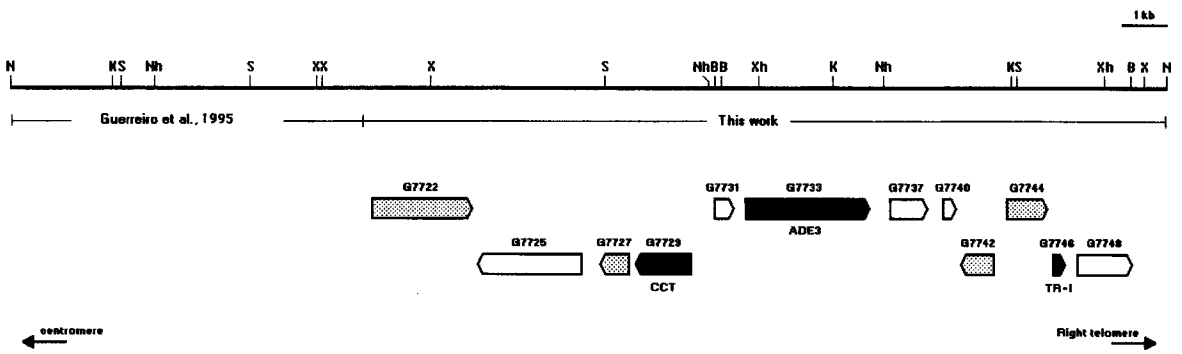


Figure 1. Restriction map of the 26·2 kb insert of the cosmid pEGK620. The left part of the insert overlaps Dr Nombela's pEGH059 cosmid and is presented separately (Guerreiro *et al.*, 1995). The location and the direction of the 12 ORFs are indicated by arrows, and above each arrow is the provisional name for each ORF given by MIPS. Filled arrows represent previously sequenced genes, with their name presented below. Spotted arrows represent ORFs that present homologies to other genes, and open arrows represent ORFs that do not show detectable homologies with other genes. B, *Bam*HI; K, *Kpn*I; Nh, *Nhe*I; N, *Not*I; S, *Spe*I; X, *Xba*I; Xh, *Xho*I.

chromosome VII) and consists of a 26·2 kb DNA fragment of the chromosome VII of *S. cerevisiae* strain S228C inserted in the pWE15 cosmid.

Plasmid pUC19 was used for subcloning experiments. *Escherichia coli* strain XL1-Blue was the host used for transformation and amplification of plasmids. Sequencing was also performed in M13mp18 and M13mp19 vectors using *E. coli* strains XL1-Blue and TG1 as hosts for transformation and growth of M13 phages. Standard techniques (Sambrook *et al.*, 1989) were used for transformation of *E. coli*, restriction mapping, plasmid preparation, DNA ligation and other manipulations.

Sequencing strategy

The entire DNA insert of the cosmid pEGH620 was subcloned into a total of 18 subclones. Twelve of these subclones were used in sequencing the 18,230 bp reported here: the *Hin*dIII-*Hin*dIII H9 (4·5 kb), H39 (2·5 kb), H44 (1·6 kb), H4 (2·3 kb) and H2 (1·4 kb) subfragments, the *Eco*RI-*Eco*RI E11 (0·9 kb), E4·1 (1·0 kb), E37 (3·2 kb), E40 (1·6 kb), E31 (0·3 kb) and E39 (4·5 kb) subfragments, and the *Kpn*I-*Kpn*I K2 (4·0 kb) subfragment. Clones suitable for sequencing were generated in two ways. Nested deletions were made using the Pharmacia Exonuclease III nested deletion kit in order to sequence one of the strands of the *Hin*dIII-*Hin*dIII H9, H39, H44 and H2 subfragments, and of the *Eco*RI-*Eco*RI E37, E40 and E39 subfragments. Direct cloning in M13 and pUC19 vectors using suitable enzymes was used for the remaining sequence. DNA sequencing was carried out manually by the dideoxy chain-termination method using T7 DNA polymerase (Pharmacia) and the M13 forward ('−40') and 'reverse' primers. Synthetic oligonucleotides were also used to fill a few gaps. Both strands were sequenced and gel compressions were solved by subcloning the fragment in M13 vectors and/or by sequencing it with Deaza G/A T7 sequencing mixes (Pharmacia).

Sequence analysis software

Sequences from the individual reactions were assembled into longer contigs with the aid of DNASIS 5·0 analysis software (LKB-Hitachi). Comparison of DNA sequences and deduced amino acid sequences to the GenBank and Swiss-Prot databases was done using MIPS facilities and the GCG package.

RESULTS AND DISCUSSION

The cosmid pEGH620 contains a fragment of 26·2 kb located on the right arm of *S. cerevisiae* chromosome VII. Figure 1 presents the restriction map of the fragment and the location of the ORFs in the sequenced region. The left extremity of the insert overlaps C. Nombela's pEGH059 cosmid, and has already been analysed (Guerreiro *et al.*, 1995).

The DNA sequence of the 17,640 bp fragment was determined and deposited in the EMBL data library under accession number Z49133. This region comprises a total of 12 complete ORFs longer than 100 amino acids, which are listed in Table 1.

Table 1. Characteristics of the ORFs.

Name	Position	Strand[a]	No. of amino acids	CAI[b]	FASTA opt. score[c]	Best homology with	Motifs[d]
G7722	213–2489	W	759	0·13	2013 682	YAL023 (=PMT2) PMT1, dolichyl protein mannosyltransferase (*S. cerevisiae*)	6 TM spans
G7725	4981–2618	C	788	0·18	—	No significant homology	—
G7727	5908–5234	C	225	0·13	574 398	TEF3, elongation factor 1γ (*S. cerevisiae*) cEF1γ homolog TEF4	—
G7729	7454–6188	C	424	0·16	100% identical	CCT, cholinephosphate cytidylyltransferase	—
G7731	7948–8391	W	148	0·06	—	No significant homology	—
G7733	8645–11 482	W	946	0·28	100% identical	ADE3, C1-tetrahydrofolate synthase	—
G7737	11 924–12 793	W	290	0·17	—	No significant homology	ATP/GTP-binding site
G7740	13 143–13 445	W	101	0·15	—	No significant homology	—
G7742	14 339–13 557	C	261	0·18	631 604	Electron-transferring flavoprotein β chain (*P. dentifricans; H. sapiens*)	—
G7744	14 594–15 520	W	309	0·21	409	Phosphoserine phosphohydrolase (*E. coli*)	—
G7746	15 938–15 627	C	104	0·42	100% identical	TR-I, thioredoxin I	—
G7748	17 449–16 217	C	411	0·13	—	No significant homology	1 TM span ATP/GTP-binding site

[a]W, Watson strand; C, Crick strand.
[b]CAI, codon adaptation index following Sharp and Li (1987).
[c]FASTA optimal scores, database search following Pearson and Lipman (1988).
[d]TM, transmembrane, search following Klein *et al.* (1985).

Table 2. Characteristics of the genetic elements.

Feature	Name	Sequence	Position	Strand[a]	Observations
Intron	5′ site	GTTAGG	12 739–12 744	W	Consensus: GTATGT
	branchpoint	TACTAAC	12 879–12 885		Distance to 3′ site: 229 nucleotides
	3′ site	ATAG	13 114–13 117		
Regulatory motif	GCN4	GAAAGAGTCAAT	14 435–14 446	W	Positioned at −148 of *G7744* (W strand) Positioned at −96 of *G7742* (C strand)

[a]W, Watson strand; C, Crick strand.

Other genetic elements, such as tRNAs, Ty elements or introns, were also searched, and the results are presented in Table 2.

G7722

The centromere-proximal ORF *G7722* encodes a putative protein of 759 amino acids that shows 44·8% identity with YAL023 and 32·8% identity with the *S. cerevisiae* dolichyl protein mannosyltransferase, PMT1 (Figure 2). *YAL023* is a functional gene encoding a 70 kDa integral membrane protein (Ouellette *et al.*, 1993) that was recently identified as the second member of the PMT family (Lussier *et al.*, 1995). PMTs are integral endoplasmic reticulum membrane proteins that catalyse the initial reaction of protein O-glycosylation in *S. cerevisiae* (Strahl-Bolsinger *et al.*, 1993), and are encoded for by at least five genes (Gentzsch *et al.*, 1995). The hydropathy profile of the G7722 protein is quite similar to those of YAL023 and PMT1, and also predicts an integral membrane protein with six potential transmembrane spans, whereby the lipophilic parts of the protein are concentrated within the N- and C-terminal regions, whereas the central part corresponds to a hydrophilic domain (Figure 3). The peptide sequence also predicts four potential N-glycosylation sites (underlined in Figure 2), with one of them, 403-NGT-405, localized at comparable positions in the three proteins. This suggests that *G7722* is most probably the sixth member of the PMT family, *PMT6*.

G7725

The *G7725* ORF encodes a putative protein of 788 amino acids that does not show any significant homology to other known genes.

G7727

The *G7727* ORF encodes a 225 amino acid protein which exhibits a high degree of homology with the elongation factor EF-1γ. This factor is encoded in *S. cerevisiae* by multiple genes, with two members have already been sequenced (Kinzy

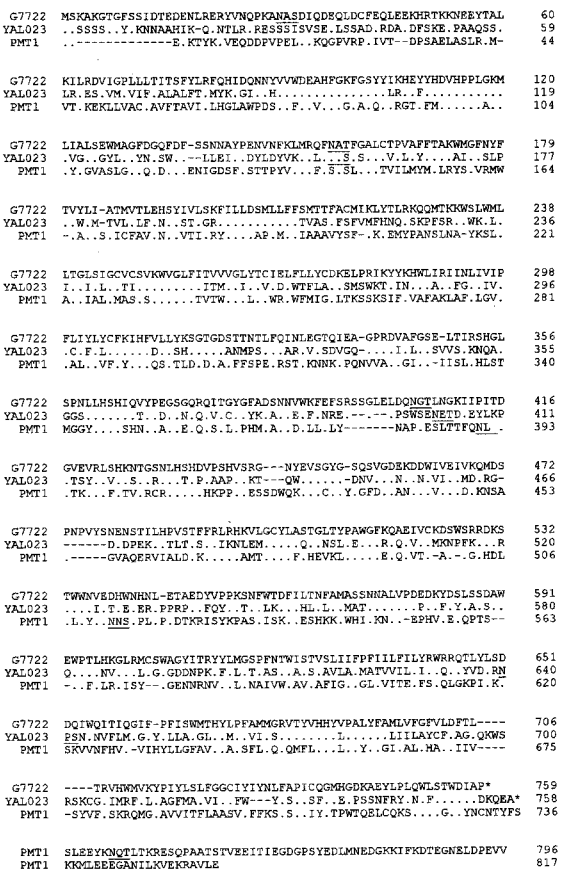


Figure 2. Alignment of the amino acid sequence encoded by ORF *G7722* with the dolichyl protein mannosyl transferases, YAL023 (PMT2) and PMT1, from yeast. Only the amino acids differing from the first sequence are shown; dots indicate identical amino acids and bars indicate gaps in the sequence that were introduced to improve the alignment. Underlined sequences represent potential N-glycosylation sites.

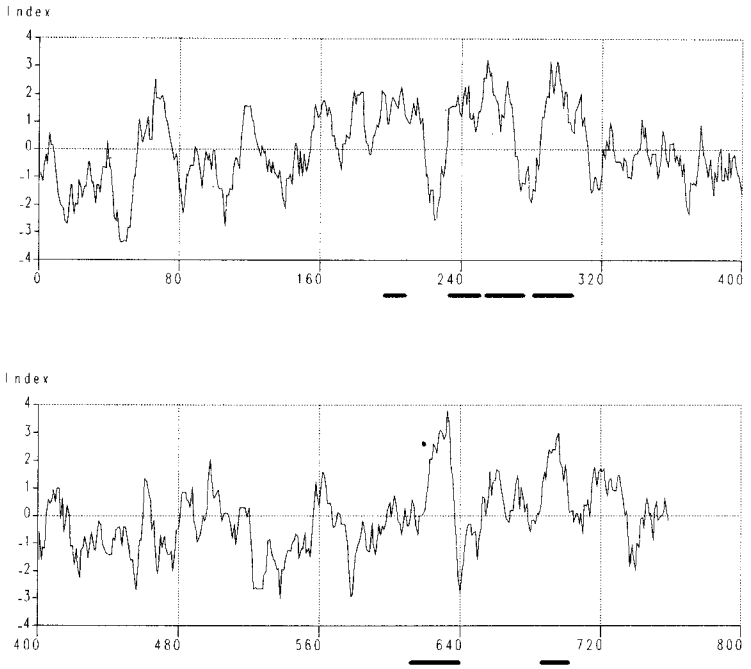


Figure 3. Hydropathy profile of the amino acid sequence predicted by ORF G7722. The hydropathy profile was determined according to the method of Kyte and Doolittle (1982). Horizontal bars below the profile represent the predicted transmembrane spans according to the method of Klein *et al.* (1985).

et al., 1994). However, although G7727 presents 44·4% identity with EF1-γ (TEF3) and 30·7% with YKL081 (TEF4), it is significantly shorter than these two proteins, both having more than 410 amino acids (Figure 4). The EF-1γ is one of the three subunits of the elongation factor EF-1, which functions to facilitate binding of aminoacyl-tRNA to the ribosomal A site (Moldave, 1985). The α subunit binds aminoacyl-tRNA in a GTP-dependent manner, the ternary complex binds the ribosome, and after the aminoacyl-tRNA binds the A site of the ribosome, GTP is hydrolysed leaving GDP bound to EF-1α. The β subunit stimulates nucleotide exchange to regenerate EF-1α-GTP but the function of EF-1γ is unknown, although it seems that it can stimulate the nucleotide exchange activity of the β subunit (Janssen and Moller, 1988). However, EF-1γ may play several roles in the cell since *TEF3* and *TEF4* seem to have different properties. *TEF3*, but not *TEF4*, is a weak suppressor of *drs2*, a cold-sensitive mutant deficient in the assembly of 40S ribosomal subunits, and the double mutant *tef3 tef4* does not present defects in polyribosomes or growth (Kinzy *et al.*, 1994).

```
G7727   MSDGTLFTDLKERKLIRTIVPRGLVRSLKLDVKLADPSDAQQLYEREFPLRKYPTFVGPH    60
TEF3    ..Q...YANFR----...W.....KA.....VVT.DA.AEQFA.D...K.V.A....-    55
TEF4    ..Q...YINRS----P.NYASEA.ISYF.....IV.LEQSSE-FASL...KQA.A.L..-    54

G7727   DEWTLTEAMAIDYYLIHLSSDKEAVRQLLGPEGDFKTRADILRWESLSNSDFLNEVCEVF   120
TEF3    KGYK.....N...VK..Q.DKMKT....ADD.LNAQ.Q.I..Q..A...LCIQIANTI   115
TEF4    KGLK....L..QF..ANQVA.EKERAR...S--.VIEKSQ....A..A...VMSNIARP.   112

G7727   FPLIGVKPYNATEFKAARENVDTIVSLYEKRLKKQQYLVCDDHETLADLISAAAFSLGFI   180
TEF3    V..K.GA...KKSVDS.MDA..K..DIF.N...NYT..ATEN-IS....VA.SI.TRY.E   174
TEF4    LSFK.LI...KKDVD.CFVKI.NLAAVFDA..RDYTFVATEN-IS.G..HA.GSWAF.LA   171

G7727   SFFDETWRSKHPEVTRWFNRVIKSRFFEGEFESFKMCETEM---QPIK*              225
TEF3    .L.GTE..AQ..AIV....T.RA...LKD.YKD..FADKPL--SPPQKKKEKKAPAAAPA   232
TEF4    TILGPE..A...HLM....T.AA.PIVKTP.AEV.LA.KALTTYTPPKKQKAEKPKAEKSK   231

TEF3    ASKKKEEAKPAATETETSSKKPKHPLELLGKSTFVLDDWKRKYSNEDTRPVALPWFWEHY   292
TEF4    AEKKKDEAKPA--DDAAPAKKPKHPLEALGKSTFVLDDWKRKYSNDDTRPVALPWFWEHY   289

TEF3    NPEEYSLWKVTYKYNDELTLTFMSNNLVGGFFNRLSASTKYMFGCLVVYGENNNNGIVGA   352
TEF4    NPEEYSIWKVGYKYNDELTLTFMSNNLVGGFFNRLSASTKYMFGCLVVYGENNNNGIVGA   349

TEF3    VMVRGQDYVPAFDVAPDWESYDYAKLDPTNDDDKEFINNMWAWDKPVSVNGEPKEIVDGKVLK*  415
TEF4    VMVRGQDFAPAFDVAPDWESYEYTKLDPTKEEDKEFVNNMWAWDKPVVVNGEDKEIVDGKVLK*  412
```

Figure 4. Alignment of the amino acid sequence encoded by ORF *G7727* with the elongation factors 1γ TEF3 (EF-1γ) and TEF4 (YKL081) from yeast. Only the amino acids differing from the first sequence are shown; dots indicate identical amino acids and bars indicate gaps in the sequence that were introduced to improve the alignment.

G7729

The *G7729* ORF has already been sequenced and identified as the *CCT* gene, which encodes for the enzyme cholinephosphate cytidylyltransferase

(Tsukagoshi *et al.*, 1987). Three differences have been found between the two sequences probably due to the use of different strains. Two of them are insertions located in the 5′ non-coding region (two G at positions 7736 and 7842) and the third one is a mismatch in the coding region that promotes an amino acid substitution (G to A at position 6886, changing His to Tyr at amino acid position 192).

G7731

The *G7731* ORF encodes a putative protein of 148 amino acids that does not exhibit any significant homology to previously sequenced genes. However, it had already been partially sequenced since it is included in the DNA fragment that contains the *ADE3* gene (Staben and Rabinowitz, 1986; see below). Four insertions were found between the two sequences (GC at position 8196/7, two G at positions 8237 and 8398) probably made by misreading of the original sequence due to gel compressions observed in these regions, which were solved in the present work by sequencing with deaza nucleotides. This gene presents the lowest codon adaptation index (CAI; Sharp and Li, 1987) of the 12 ORFs found (Table 1), suggesting a low level of expression.

G7733

The *G7733* ORF is identical to *ADE3*, a gene that encodes the multifunctional protein C_1-tetrahydrofolate synthase. This enzyme catalyses the three sequential reactions of the interconversion of tetrahydrofolate in 5,10-methylenetetrahydrofolate (Staben and Rabinowitz, 1986). No differences were found between the two sequences.

G7737

The *G7737* ORF encodes a putative protein of 290 amino acids that does not show any significant homology to other known genes. However, the *G7737* gene has already been sequenced as part of the DNA fragment that contains *ADE3* (Staben and Rabinowitz, 1986). One deletion was found between the two sequences (one missing G at 12,806/12,807), probably made by misreading the original sequence due to gel compressions observed in this region. The G7727 protein presents, as its principal feature, a ATP/GTP-binding site near its N-terminus, 38-**GPQGSGKS**-45, that fits the 'P-loop' motif [AG]-N4-GK-[ST] (Saraste *et al.*, 1990). Another unusual feature is a long stretch of 24 T residues immediately following the stop codon and whose function is unknown. Staben and Rabinowitz (1986) did not detect any expression of the *G7737* gene by Northern analysis when they studied the expression of the *ADE3* gene, suggesting that the transcript, if it exists, is present at much lower levels than the *ADE3* mRNA. The difference observed between the CAI of the two genes (0·17 against 0·28; Table 1) also suggests that the putative protein G7737 is less expressed than ADE3.

G7740

The *G7740* ORF encodes a putative protein of 101 amino acids that does not exhibit any significant homology to previously sequenced genes. It is also possible that G7740 is an exon of G7737. The intergenic region between these two genes presents a canonical branchpoint at positions 12,879/12,885, and it is possible to find in-frame a 5′ site and a 3′ site of an intron (Table 2). There are, however, two facts which argue against the presence of an intron: first, the 5′ splice site, GTTAGG, is similar but not identical to the GTATGT standard pattern; second, the distance from the branchpoint TACTAAC to the 3′ splice site is longer than usual (180 nucleotides; Fondrat and Kalogeropoulos, 1994). A search in GenBank also detected a short cDNA read (accession number Z13182) that overlaps both ends of ORFs G7740 and G7742 (from position 13,313 to position 13,598). Since the GenBank sequence is a (−) strand and complementary to ORF G7740, it should correspond to a cDNA of this gene, suggesting that it is functional and expressed.

G7742

The *G7742* ORF shows 48·3% and 47·4% identity with the β subunit of *Paracoccus denitrificans* and human electron-transfer flavoproteins (ETF), respectively (Figure 5) (Bedzyk *et al.*, 1993; Finocchiaro *et al.*, 1993). This suggests that the *G7742* gene is the yeast *ETF*-β gene. ETF is a mitochondrial heterodimeric flavoenzyme that functions as an electron acceptor of several primary flavoprotein dehydrogenases. The β subunit is one of the few mitochondrial proteins that does not have a leader peptide directing the subunit from cytosol to the mitochondria. The import is energy-dependent and seems to be targeted by the first dozen amino acid residues, a conserved region (Figure 5) that also shows similarity with the N-terminal region of another mitochondrial

```
G7742          MSAKQQLRILVPVKRVVDFQIKPRVNKTLTGIETSGIKFSINPFDDIAVEEAIRIKEKNK   60
P.denitrificans      MKV......LI.YNV.A..KSDGS.VDLANV.M.M....E........L...GQ   54
human              MAE..V..A....I.YAV.I..KPDR..VV.D.V.H.M...CE......V.L...-.   56

G7742          SLVESTHAVSIGSAKAQDILRNCLAKGIDTCSLID--SVGKENIEPLAIAKILKAVVEKK   118
P.denitrificans A--.EII.....VKQ.AET..TA..M.A.RAILVVAADDVQQD.....V....A..ARAE   112
human          -..KEVI...C.P.QC.ETI.TA..M.A.RGIHVEVPPAEA.RLG..QV.RV.AKLA..E   115

G7742          GSNLVLMGKQAIDDDCNNTGQMLAGLLNWPQATNAAKVEFLDNGRVQVTREIDDGEEVIE   178
P.denitrificans .TE.IIA......N.M.A......AI.G.A...F.S...-IEGAKAK....V.G.LQT.A   171
human          KVD...L..........Q....T..F.D...G.F.SQ.T-.EGDKLK.E....G.L.TLR   174

G7742          ASLPMVITTDLRLNTPRYVGLPKLMKAKKKPIEKLDIAKDFPEINIEPQLKIVSMEEPKT   238
P.denitrificans V...A.V.A.....E...AS..NI.......LDE-KT.A.Y-GVDVA.R.EV..VR..EG   229
human          LK..A.V.A.....E...AT..NI......K..V-IKPG.L-GVDLTSK.SVI.V.D.PQ   232

G7742          KSPGVKLNSVDELIEKLKEVKAI*                                      261
P.denitrificans RKA.I.VG.....VG....AGV.*                                     252
human          RTA...VETTED.VA....IGR.*                                      255
```

Figure 5. Alignment of the amino acid sequence encoded by ORF *G7742* with the β subunit of the electron-transfer flavoproteins of *Paracoccus denitrificans* and human. Only the amino acids differing from the first sequence are shown; dots indicate identical amino acids and bars indicate gaps in the sequence that were introduced to improve the alignment.

```
G7744                                  MSKFVITCIAHGENLPKETIDQIAKEITESSAKDV   35
PSP (serβ)     MPNITWCDLPEDVSLWPGLPLSLSGDEVMPLDYH GRSGWLLYGRGLDKQRL.QYQS.LG   60

G7744          SINGTKKLSARATDIFIEVAGSIVQKDLKNKLTNVIDSHNDVDVIVSVDNEYRQAKKLFV   95
PSP (serβ)     AAMVIVAAWCVEDYQV.RL...LTARATR------LAHEAQL..APLGKIPHLRTPG.L.   114

G7744          FDMDSTLIYQEVIELIAAYAGVEEQVHEITERAMNNELDFKESLRERVKLLQGLQVDTLY   155
PSP (serβ)     M.....A.QI.C.DE..KL..TG.M.A.V.....RG....TA...S..AT.K.ADANI..   173

G7744          DEIKQKLEVTKGVPELCKFLHKKNCKLAVLSGGFIQFAGFIKDQLGLDFCKANLLEVDTD   215
PSP (serβ)     QQVREN.PLMP.LTQ.VLK.ETLGW.V.IA....TF..EYLR.K.R.TAVV..E..I-M.   232

G7744          GKLTGKTLGPIVDGQCKSETLLQLCNDYNVPVEASCMVGDGGNDLPAMATAGFGIAWNAK   275
PSP (serβ)     ..F..NVI.D...A.Y.AK..TR.AQE.EI.LAQTVAI...A....MIKA..L...YH..   292

G7744          PKVQKAAPCKLNTKSMTDILYILGYTDDEIYNRQ*                           309
PSP (serβ)     ...NEK.EVTIRHADLMGVFC..SGSLNQK*                               322
```

Figure 6. Alignment of the amino acid sequence encoded by ORF *G7744* with the *Escherichia coli* O-phosphoserine phosphohydrolase encoded by the *ser*β gene. Only the amino acids differing from the first sequence are shown; dots indicate identical amino acids and bars indicate gaps in the sequence that were introduced to improve the alignment.

enzyme devoid of a transient presequence, the rat 3-oxo-acyl-CoA thiolase (Arakawa *et al.*, 1987; Finocchiaro *et al.*, 1993).

G7744

The *G7744* ORF encodes a protein of 309 amino acids that is 24·9% identical to the O-phosphoserine phosphohydrolase (PSP) of *E. coli* (Neuwald and Stauffer, 1985), suggesting it as the yeast *PSP* gene. The alignment presented in Figure 6 shows that the two proteins have a highly homologous core, with an identity of 46·5% between amino acids Leu93 and Val278, and divergent N- and C-terminal regions. It is possible that this core contains the active region of the enzyme. PSP is the enzyme that catalyses the hydrolysis of the phosphate group from phosphoserine to yield serine, which is the final step in serine biosynthesis in both prokaryotes and eukaryotes. However, eukaryote PSP-encoding genes have never been cloned before, the yeast *PSP* gene being the first one sequenced. A GCN4 motif was found at position − 148 of the initiation codon (Table 2) using the EXPLORA v.1.1 program (Fondrat and Kalogeropoulos, 1994), suggesting that the yeast *PSP* gene may be controlled by the GCN4 transcription factor.

G7746

The *G7746* ORF is identical to *TR-I*, a gene that encodes one of the two yeast thioredoxins (Gan, 1991). Thioredoxins are a group of small proteins containing a dithiol active site sequence CGPC that are general thiol-disulfide oxidoreductases (Holmgren, 1985). No differences were found between the two sequences.

G7748

The *G7748* ORF encodes a putative protein of 411 amino acids that does not show any significant homology to other known genes. The G7727 protein presents two features: a putative transmembrane span and an ATP/GTP-binding site at N-terminal, 11-**GKPSSGKS**-18, that fits the 'P-loop' motif [AG]-N4-GK-[ST] (Saraste *et al.*, 1990).

ACKNOWLEDGEMENTS

We thank H. Tettelin, coordinator of sequencing of chromosome VII, for providing us with the recombinant cosmid. We wish to acknowledge K. Kleine at MIPS for his help in computer analysis. We also thank Dr A. Kalogeropoulos for help with the search of possible upstream activating sequences. This work was supported, in part, by the Commission of the European Community under the Biotechnology program for sequencing the yeast genome and by grants of the Calouste Gulbenkian Foundation given to C.R.-P.

REFERENCES

Arakawa, H., Takiguchi, M., Amaya, Y., Nagata, S., Hayashi, H. and Mori, M. (1987). cDNA-derived amino acid sequence of rat mitochondrial 3-oxoacyl-CoA thiolase with no transient presequence: structural relationship with peroxisomal isozyme. *EMBO J.* **6**, 1361–1366.

Bedzyk, L. A., Escudero, K. W., Gill, R. E., Griffin, K. J. and Frerman, F. E. (1993). Cloning, sequencing, and expression of the genes encoding subunits of *Paracoccus denitrificans* electron transfer flavoprotein. *J. Biol. Chem.* **268**, 20211–20217.

Finocchiaro, G., Colombo, I., Garavaglia, B., *et al.* (1993). cDNA cloning and mitochondrial import of the β-subunit of the human electron-transfer flavoprotein. *Eur. J. Biochem.* **213**, 1003–1008.

Fondrat, C. and Kalogeropoulos, A. (1994). Approaching the function of new genes by detection of their potential upstream activation sequences in *Saccharomyces cerevisiae*: application to chromosome III. *Curr. Genet.* **25**, 396–406.

Gan, Z.-R. (1991). Yeast thioredoxin genes. *J. Biol. Chem.* **266**, 1692–1696.

Gentzsch, M., Immervoll, T., Strahl-Bolsinger, S. and Tanner, W. (1995). O-Glycosylation in *Saccharomyces cerevisiae*. *Yeast* **11**, S433.

Guerreiro, P., Maia e Silva, A., Barreiros, T., *et al.* (1995). The complete sequence of a 9000 bp fragment of the right arm of *Saccharomyces cerevisiae* chromosome VII contains four previously unknown open reading frames. *Yeast* **11**, 1087–1091.

Holmgren, A. (1985). Thioredoxin. *Ann. Rev. Biochem.* **54**, 237–271.

Janssen, G. M. C. and Moller, W. (1988). Elongation factor 1βγ from *Artemia*. Purification and properties of its subunits. *Eur. J. Biochem.* **171**, 119–129.

Kinzy, T. G., Ripmaster, T. L. and Woolford Jr, J. L. (1994). Multiple genes encode the translation elongation factor EF-1γ in *Saccharomyces cerevisiae*. *Nucl. Acids Res.* **22**, 2703–2707.

Klein, P., Kanehisa, M. and Delisi, C. (1985). The detection and classification of membrane-spanning proteins. *Biochem. Biophys. Acta* **815**, 468–476.

Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132.

Lussier, M., Gentzsch, M., Sdicu, A.-M., Bussey, H. and Tanner, W. (1995). Protein O-glycosylation in yeast. *J. Biol. Chem.* **270**, 2770–2775.

Moldave, K. (1985). Eukaryotic protein synthesis. *Ann. Rev. Biochem.* **54**, 1109–1149.

Neuwald, A. F. and Stauffer, G. V. (1985). DNA sequence and characterization of the *Escherichia coli serB* gene. *Nucl. Acids Res.* **13**, 7025–7039.

Oullette, B. F. F., Clark, M. W., Keng, T., *et al.* (1993). Sequencing of chromosome I from *Saccharomyces cerevisiae*: analysis of a 32kb region between the *LTE1* and *SPO7* genes. *Genome* **36**, 32–42.

Pearson, W. R. and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Saraste, M., Sibbald, P. R. and Wittinghofer, A. (1990). The P-loop — a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430–434.

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.

Staben, C. and Rabinowitz, J. C. (1986). Nucleotide sequence of the *Saccharomyces cerevisiae ADE3* gene encoding C_1-tetrahydrofolate synthase. *J. Biol. Chem.* **261**, 4629–4637.

Strahl-Bolsinger, S., Immervoll, T., Deutzmann, R. and Tanner, W. (1993). *PMT1*, the gene for a key enzyme of protein O-glycosylation in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **90**, 8164–8168.

Tsukagoshi, Y., Nakawa, J. I. and Yamashita, S. (1987). Molecular cloning and characterization of the gene encoding cholinephosphate cytidylyltransferase in *Saccharomyces*. *Eur. J. Biochem.* **169**, 477–486.