



Yeast Sequencing Reports

Sequencing of a 9.9 kb Segment on the Right Arm of Yeast Chromosome VII Reveals Four Open Reading Frames, including *PFK1*, the Gene Coding for Succinyl-CoA Synthetase (β -chain) and Two ORFs Sharing Homology with ORFs of the Yeast Chromosome VIII

PAULO GUERREIRO¹, DULCE AZEVEDO¹, TANIA BARREIROS¹ AND CLAUDINA RODRIGUES-POUSADA^{1*}

¹Laboratório de Genética Molecular, Instituto Gulbenkian de Ciência, Apartado 14, P-2781 Oeiras Codex, Portugal

Received 23 July 1996; accepted 26 August 1996

A 9.9 kb DNA fragment from the right arm of chromosome VII of *Saccharomyces cerevisiae* has been sequenced and analysed. The sequence contains four open reading frames (ORFs) longer than 100 amino acids. One gene, *PFK1*, has already been cloned and sequenced and the other one is the probable yeast gene coding for the β -subunit of the succinyl-CoA synthetase. The two remaining ORFs share homology with the deduced amino acid sequence (and their physical arrangement is similar to that) of the YHR161c and YHR162w ORFs from chromosome VIII. The sequence is in the EMBL data library under Accession Numbers Z73024, Z73025, Z73026, Z73028 and Z73029. © 1997 by John Wiley & Sons, Ltd.

Yeast 13: 275–280, 1997.

No. of Figures: 4. No. of Tables: 1. No. of Refs: 9.

KEY WORDS — *Saccharomyces cerevisiae*; chromosome VII; chromosome VIII; *PFK1*; O44 protein; β -SCS

INTRODUCTION

As a contribution to the European Union BIOTECH programme to sequence the *Saccharomyces cerevisiae* genome, we have determined the sequence of a 9898 bp DNA fragment of the right arm of chromosome VII. The sequenced region is part of a 44 kb yeast segment inserted into the pWE15 cosmid that partially overlaps, at its left

side, the insert of cosmid pEGH484 sequenced by H. Steensma's group and, at its right, the insert of cosmid pUGH1273 sequenced by M. L. Carbone's group, part of which was also sequenced by ourselves.

The fragment includes the previously sequenced *PFK1* gene, the putative succinyl-CoA synthetase β -subunit encoding gene and two open reading frames (ORFs) sharing homology with other genes from yeast and rat. The latter presents a strong homology with ORFs from chromosome VIII and they both have the same physical arrangement as in this chromosome.

*Correspondence to: Claudina Rodrigues-Pousada.
Contract grant sponsor: Commission of the European Community under the BIOTECH programme.
Contract grant sponsor: Calouste Gulbenkian Foundation.

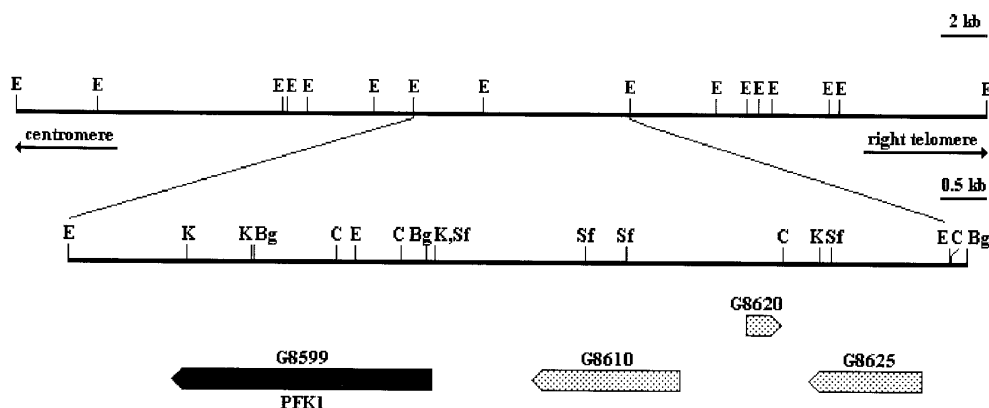


Figure 1. Restriction map of the cosmid pEGH301. The upper part presents the *EcoRI* restriction map of the insert of the cosmid. The left part of the insert overlaps the pEGH484 cosmid and the right part overlaps the pUGH1273 cosmid. The lower part shows the restriction map of the sequenced fragment here presented. The location and direction of the four ORFs are indicated by arrows, and above each arrow is the provisional name given by MIPS. The filled arrow represents the previously sequenced gene *PFKI*. Spotted arrows represent ORFs that share homologies with other genes. Bg, *Bg/II*; C, *Clal*; E, *EcoRI*; K, *KpnI*; Sf, *SfuI*.

MATERIALS AND METHODS

Strains and plasmids

The pEGH301 cosmid was provided by H. Tettelin and A. Goffeau (DNA coordinators of chromosome VII) and consists of a 44 kb DNA fragment of the chromosome VII of *S. cerevisiae* strain S288C 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. Standard techniques (Sambrook *et al.*, 1989) were used for transformation of *E. coli*, restriction mapping, plasmid preparation, DNA ligation and other manipulations.

Sequencing strategy

Random libraries were generated from pEGH301 by digestion with the *Bg/II*, *Clal*, *EcoRI*, *KpnI* and *SfuI* restriction enzymes. Overlapping clones that cover the entire DNA fragment to be sequenced were chosen to generate subclones suitable to sequencing. The sequence of the 9898 bp of cosmid pEGH301 reported herein contains: the *EcoRI-EcoRI*, E18 (3.2 kb) subfragment, the *KpnI-KpnI*, K2 (2.0 kb) subfragment, the *Bg/II-Bg/II*, Bg7 (6.2 kb) subfragment, the *SfuI-SfuI*, Sf2 (1.7 kb), Sf12 (0.4 kb) and Sf11 (2.3 kb) subfragments, and the *Clal-Clal*, C2 (1.9 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 E18, Bg7, Sf11 and C2 subfragments. Direct deletions in all subfragments using suitable restriction enzymes were used for the remaining sequence. DNA sequencing was carried out by the dideoxy chain-termination method using T7 DNA polymerase (Pharmacia) and the 'forward' and 'reverse' primers. Synthetic oligonucleotides were also used to fill a few gaps. DNA strands were sequenced in both directions and gel compressions were solved by sequencing the fragment with Deaza G/A T7 sequencing mixtures (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 with the 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 pEGH301 contains a fragment of about 44 kb located on the right arm of *S. cerevisiae* chromosome VII. Figure 1 presents the *EcoRI* restriction map of the fragment, the restriction map of the sequenced region and the location of the ORFs herein. The left extremity of the insert

Table 1. Characteristics of the ORFs.

Name	Position	Strand ^a	No. of amino acids	CAI ^b	FASTA opt. score ^c	Best homology with	Notes ^d
G8599	4049–1089	C	987	0.47	100% identical	PFK1, phosphofructokinase 1	—
G8610	6896–5193	C	568	0.16	1137	YHR161c	1 TM span
G8620	7651–8088	W	146	0.10	509	YHR162w	—
					269	O44 protein (rat)	—
G8625	9634–8354	C	427	0.19	1080 ^e	Succinyl-CoA synthetase β -subunit (<i>N. frontalis</i>)	5 ESTs

^aW, Watson strand; C, Crick strand.

^bCAI, codon adaptation index following Sharp and Li (1987).

^cFASTA optimal scores, database search following Pearson and Lipman (1988).

^dTM, transmembrane span, search following Klein *et al.* (1985); EST, cDNA tags.

^eThe best score obtained with succinyl-CoA synthetases from eight different organisms.

G8610	MSSLYTKLVKGGATKIKMAPPKQKYVDPILSGTSSARGLQEIETHALDIRLSDTAWTIVYKA	60
YHR161c	MTT.F.....S.....L....L...NEEDFY..VKG..S.IN.....S	59
G8610	LIVLHLMIQQGEKDVTLRHYSHNLDVFLRKRISHTTKWSSNDMRALQRYDEYLKTRCEEY	120
YHR161c	.L.V....RE.S...A..Y..R..EF.DIEN.RGSSNGSA.G.....D...N...V..R.F	119
G8610	GRLGMDHLRDNYSSLKLGSKN-----QLSMDELDHVESLEIQINALIRNKYSVSDLE	173
YHR161c	.KIKK.YV..G.RT...N.G.YGSSRNK.H.INIA.....V..Q...K...TQY..S	179
G8610	NHLLLYAFQLLVQDLLGLYNALNEGVTLLSEFFELSIHAKRTLIDLYKDFVDMTEYVVR	233
YHR161c	.E.IIFG.K..I...A.....I.....HHN.E.....T...L..H...	239
G8610	YLKIGKAVGLKIPVIKHITTKLINSLEEHLREETKRQRGEPSEPPQDRKPTAISSTSSH	293
YHR161c	...S..TA.....VR.....I.D-----KTHN.FVPVD..Q	286
G8610	NNNSNDKNSIAQKKLEQIREQKRLLEQQQLQNLISPTV----PQDAYNFFGQQQDL	349
YHR161c	GSAGAVVA..T..ER.....I..A..K.E.VA...ALTTVTAQS....TDS---	343
G8610	NNDFSEFTEPQPMTAQVQPTA-NPFLIPQQQQALQLTSAST----MPQPSEIQITP	403
YHR161c	----.MHTNI.MAV.NQT.QI.N...VSQT.P.VMNTP.AHTEPANLNV.EYAAV.H.V	398
G8610	NLN-----NQQTGMYSNLQYTNFTGSGFGGYTT-ENNAIMTGLDPTKTSNNP	454
YHR161c	.F.PVQDAGVSA...Y.SI.NHL..T...A....SVSQDPT.ASNQQVSHSQ.....	458
G8610	FSLENIA-----REQQQNFQNSPNPFTLQQAQTPILAHSQGTGNPFQAQN	500
YHR161c	.A.H.A.TIATGNPAHENVLNNPFSRP..DEQNTNMP...-IISNPFQY..Y.QQ.F.Q	517
G8610	--VVTSPMGTYMTNPFVAGQLQYASTGAQQQPMQMQGQQTGYVMVPTAFVFINQQQQQQH	558
YHR161c	QRKPL.SINSV..T.TSM.GSMNIPQRFDMEFQAHYTONHLQQQQQQQQQ.....F	577
G8610	QQENPNLIDI*	568
YHR161c	..GYVVPATAGANPVTNITGTVPQNFPPYQQQPPQPEQSQTQQPVLGNQYANNLNLIDM*	637

Figure 2. Alignment of the predicted amino acid sequence encoded by the ORF G8610 with the yeast YHR161c protein. 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 alignment.

overlaps H. Steensma's cosmid pEGH484 and the right extremity overlaps M. L. Carbone's cosmid pUGH1273.

The DNA sequence of the 9898 bp fragment was determined and is deposited in the EMBL data library under accession numbers Z73024, Z73025, Z73026, Z73028 and Z73029. This region comprises a total of four complete ORFs longer than 100 amino acids, which are listed in Table 1. Other genetic elements, such as tRNAs, Ty elements or introns, were not found.

G8599

The G8599 ORF has already been sequenced and identified as the *PFK1* gene, which encodes the α -subunit of the enzyme phosphofructokinase (Heinisch *et al.*, 1989). Phosphofructokinase is an octameric enzyme composed of four α -subunits and four β -subunits that catalyses the conversion of fructose-6-phosphate and ATP to fructose-1,6-biphosphate and ADP in glycolysis. Seven differences have been found between the two sequences,

```

G8620          MSASAFNEA--FRFEWNETGPKTVHFWAPTLKWGLVFAGLNDI 42
YHR162w        M.T.SVR.---.---.Q.....FS.M 42
P044_rat       MAAAGARGLRATYHRLMDKVELLPPKLL.PLY.HPA..R..F....IM....C...A.M 60

G8620          KRPEVKVSGAQNLSLLATALIWRWSFVIKPKNYLLASVNFGLGCTAGYHLTRIANFRIR 102
YHR162w        .....I.....S.....I.....S..CL....Q.G...Y... 102
P044_rat       A..A..L.T..STV.M..GF..S.Y.L..I.K.WS.FA....V.SAGASQ.F..WKY--- 117

G8620          NGDSFKQVIHYIIKGETPAVAARKQTASTSMNKGVIETNPPITH* 146
YHR162w        ....IS.LCS..LS.----.DES.KEIT.GR* 129
P044_rat       .QELKSKG.Q* 127

```

Figure 3. Alignment of the amino acid sequence encoded by ORF G8620 with the yeast YHR162w protein and the rat O44 protein. 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 alignment.

probably due to the use of different strains. One is a deletion located in the 3' non-coding region (one C is missing at 932/933 position) and the remainder are located near the C-terminal end of the coding region. These differences do not cause any amino acid substitution (G to A at positions 1620 and 1716, G to T at position 1692, G to C at position 1782, T to C at position 1794, and A to G at position 1875).

G8610

The G8610 ORF encodes a putative protein of 568 amino acids that shows 51.4% identity with YHR161c (Figure 2), a putative protein of 637 amino acids from chromosome VIII (Johnston *et al.*, 1994). Indeed, the comparison between the sequence of chromosome VIII and the present DNA sequence reveals significant similarities. Interestingly enough, both the deduced amino acid sequence and the physical arrangement of G8610 and G8620 ORFs on chromosome VII are similar to those of YHR161c and YHR162w on chromosome VIII, indicating that duplications/transpositions had occurred during chromosome segregation. However, the low homology of less than 50% observed between the intergenic regions of G8610/G8620 (755 bp) and of YHR161c/YHR162w (981 bp), suggests that this may represent an old event. Studies regarding the expression of these ORFs have shown that their transcripts are not detected by Northern analysis in exponentially growing cells, even when amounts corresponding to 20 µg of total RNA were used.

G8620

The G8620 ORF encodes a protein that shows 73.6% identity with YHR162w (Johnston *et al.*, 1994) and 45.9% identity with the rat O44 protein (Figure 3; Tsou *et al.*, 1986), whose functions are

unknown. However, the rat *O44* gene was found to be transcribed into nearly 100 different mRNAs due to alternative transcription initiation sites and polyadenylation sites. A putative alternative splicing site was also detected that produces mRNAs coding for proteins with different C-termini. The *O44* gene is expressed in all tissues tested but in brain it transcribes longer mRNAs than those transcribed in liver or kidney (Tsou *et al.*, 1986).

G8625

The G8625 ORF encodes a protein of 427 amino acids that is strongly homologous to the β-subunit of the succinyl-CoA synthetase from eight organisms (Figure 4), showing a maximum of 55.7% identity with the rumen fungus *Neocallimastix frontalis* enzyme. This suggests that G7742 defines the yeast *SCS-β* gene. *SCS* is a mitochondrial heterodimeric enzyme of the citric acid cycle that catalyses the conversion of succinyl-CoA to succinate and CoA together with the formation of GTP (Bridger, 1974). Assuming that G8625 is the yeast succinyl-CoA counterpart, it will be the longest *SCS* so far detected, with an extended N-terminus and three unique residues at position 245–247. The first 30 residues of the protein share typical characteristics of presequences of imported

Figure 4. Alignment of the amino acid sequence encoded by ORF G8625 with the β-subunit of the succinyl-CoA synthetase from the rumen fungus *Neocallimastix frontalis* (PIR2: S52384), the protozoa *Trichomonas vaginalis* (PIR2: A45242), pig (PIR2: A44529) and the bacteria *Coxiella burnetii* (EMBL: X77919), *Escherichia coli* (PIR1: SYEC5B), *Haemophilus influenzae* (PIR2: C64189), *Methylobacterium extorquens* (PIR2: B55230) and *Thermus aquaticus* (PIR2: S15950). 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 alignment.

G8625	MYSRKSLSLSISKGCQLSRINAQAALQARRHLSIHEYRSQALLREYIGTPEGFPAPTPEE	60
<i>N. frontalis</i>MN..H..NVNA.K.IV.K....	22
<i>T. vaginalis</i>MLSSSF.A.NFN.L.WQ.KEICAK.NVAAGINLV.RS...	39
Pig	MAFRPPL.AARSQVVQ.TP..W.NLQ..Q.KK.MSDN.VKVQRF.V.D.AN.	52
<i>C. burnetii</i>MNL...Q.KY..NK.N.AV.ASEVV.N.DA	30
<i>E. coli</i>MNL...QAK..FAR..LPA.V.YACT..R.	30
<i>H. influenzae</i>MNL...QAK..FEH..LPVKN.AVCQSV.D	30
<i>M. extorquens</i>MDV...QAKE..ASF.VAV.K.AV..S.DQ	30
<i>T. aquaticus</i>MNL...QAKEI.AR..VPV.P.KV.Y....	30
G8625	AFEAAKLNNTNKLVIKAQALTGGRGKGF-DTGYKSGVHMIESPQAEDVAKEMLNHLI	119
<i>N. frontalis</i>	..YQ.....ED.....V.A.....S.LQG..KLCYT.E.VK.Y.SK..G.KL.	81
<i>T. vaginalis</i>	..A...FR.M.LPAA.....VYC.....WLE..F....FVK.ADE.AKI.....G.HLV	99
Pig	..L....R..AKEI.L...I.A.....V..SS.L.G...LTKD.EVVQQL..Q.IGY.LA	111
<i>C. burnetii</i>	..VD..A.IGGDRW.V..VHA....A-----G..RLVKNKEELKSAV.AL.GTRLV	82
<i>E. coli</i>	..E...S.IGAGPW.V.C.VHA....A-----G..KVVN.KEDIRAF.ENW.GKRLV	82
<i>H. influenzae</i>	..VDLVLQA.SGG.WAA.C.VHA....A-----G..KLVQDVVEE.RAF.EKW.GQRIV	82
<i>M. extorquens</i>	..VY..TE.GGSFWAV..IHA.A...A-----G..LKLCTYNEVR.A.RDL.GKRLV	82
<i>T. aquaticus</i>	..KRI.EEF-GKRV....VHV...A-----G..KLADT..E.YEK.QAT.GM.T-	80
G8625	TKQGTGIAGKPVSAVYIVKRVDTKHEAYLSIIMDRQTKKPMIASSQGGMNIEVAERTPD	179
<i>N. frontalis</i>A..RDCN...V.E.QYAAR.Y.FA..L.A.RG.VLV...E..VE....KTN..	141
<i>T. vaginalis</i>KD.LLQC..MLSDE.EV.R.L.FA..L....QS.VV...TE..VE....AHH.E	159
PigPKE.VK.NK.MVAEAL.ISR.T.A....SCNG.VLVG.P...VD....ASN.E	171
<i>C. burnetii</i>	..Y..DER.Q..NQILVEQTS.IAR.L..GAVI..ASQRIVEM..TE..VE..K...K.S.E	142
<i>E. coli</i>	..Y..DAN.Q..NQILVEAAT.IAK.L..GAVV..SSRRVEM..TE..VE..K...E..H	142
<i>H. influenzae</i>	..F..DKL.Q..NQI.FEETC.IDK.F..SAVV..TSQ.VVF...E.....VQNS.H	142
<i>M. extorquens</i>	..L...PE...QR..VETADPFER.L..GYVL..KAERVV...QR...D...I.AKE.E	142
<i>T. aquaticus</i>	-----K.LT.KK.LVAEA..IAK.Y.AGLIL..AK.RVVLML.KE..VD....AER.E	134
G8625	AIKKFSIETSKGLSPQMAKDVAKSLGFSE-DAQDEAAKAVNSLYKIFMERDATQVEINPL	238
<i>N. frontalis</i>	..LTVPFSIKT..TREV.L.T..KM..TE-KCVFQ..DTFMK....I.K..M...M	200
<i>T. vaginalis</i>	K.H..VLGVE.ITEEV..NISTK..LTG-K.Y.NGVEMQK.W.L.VGS....V...V	218
Pig	L.F.EQ.DIIE.IKDSQ.QRM.EN..LG-PL.NQ..DQIKK..NL.LKI....V..F	230
<i>C. burnetii</i>	K.L.VT.DPAT..Q.FQCRQLFFG..LQDLKQMRSPFDIVMG..RL.T...LSLL....	202
<i>E. coli</i>	L.H.VALDPLT.PM.YQGREL.FK..LEG-KLVQQT.IFVG.AT..L..LALI....	201
<i>H. influenzae</i>	LLH.VT.DPLF.GL.YQGREL.FK..L.G-TQNKQFTDIFVG.SRL.L.K.LSLL.V...V	201
<i>M. extorquens</i>	..LIQVVV.PAV..QQFQ.REI.FQ..L-NIKQVSA.V.TIM.A.RA.RDC.G.ML....	201
<i>T. aquaticus</i>	..H..W.DPH..FR.FE.REMV.RA.LEG--NLNKL.QVLVA..RAYEGV..SIA.I...192	192
G8625	SEIEHDPHKIMCTDAKFGFDDNASFRQEKIYSWRDLSDQEDPDEVKA-KKYDLNFVKLKG	297
<i>N. frontalis</i>	A.NNRG--EVV.M.....YKPKG.FAL..TT....R..AGHA.WN..Y.GME.	257
<i>T. vaginalis</i>	A.TTDG---ITV.S..N...S.HY..KQ.FGY..K.VN.F.IR.-E..G..Y.P.D.	274
Pig	G.TPEG---QVV.F...IN....E...KD.FAMD.KSENE.I.NE.-A...KYIG.D.	286
<i>C. burnetii</i>	VITGSG---ELI.L...INI..S.LY..SELREM..TT...EH.TM..QQWE..YI..D.	258
<i>E. coli</i>	VITKQG---DLI.L.G.L.A.G..L...PDLREM..Q....R.AQ.-AQWE..Y.A.D.	257
<i>H. influenzae</i>	VITPQG---NLV.L...ISV...L..HKDLLALQ..T.N.AR.AE.-E.FQ..Y.A.E.	257
<i>M. extorquens</i>	VVTKD---RVLAL...MS...L..RRN.ADMH.P..G..R.AQ.-AEHNSYIG.E.	257
<i>T. aquaticus</i>	VVTTDG---G.VAA...IVL...L..HPDLAELREVEA.H.L..E.-SN.GFAY...D.	248
G8625	NIGCLVNGAGLAMATMDVIKINGGDPANFLDCGGGATPETIKQGFELILSNKNVDAIFVN	357
<i>N. frontalis</i>I.....V.....V..S..AKQV.EA.KI.S.D.A.S..L..	317
<i>T. vaginalis</i>	..VA.....A.....L..A.SEAATE..TI.SQQSH.K..L..	334
Pig	..A.F.....C.I.F...K....L...VKESQVY.A.K.LTADPK.E..L..	346
<i>C. burnetii</i>M.....L..S.....V..S..K.RVTEA.KI.V.D...KG.L..	318
<i>E. coli</i>M.....G...IV..H..E...V...K.RVTEA.KI..DDK.K.VL..	317
<i>H. influenzae</i>	D...M.....G...IV..Y..K.....V....Q.RVAEA.KI..TDP.S.KV.L..	317
<i>M. extorquens</i>	E...I.....M..HA..E.....V...S.DRVATA.R.V..DR..K..L..	317
<i>T. aquaticus</i>	...IIG...V.Y.L.LVNRV..K.....I....KADVYNALKVV.KDPP.KGV..I.	308
G8625	IFGGIVRCDYVALGLVEAARELEVRVPIVARLQGTKEVEEGRDIINKSGVKIYSFDELDEPA	417
<i>N. frontalis</i>M...I..E.VIQ.VK..GLDI.L.V.....A.EL.KN.NLSL.AI.D..K.	377
<i>T. vaginalis</i>M..A.VIA.FKKVGLK..L.V..E..N..A.KKL.RE..LP.I.A.N.TD.	394
PigN.A.II.N.ITK.C...L.K..L.V..E..N.H.AQN.LTN..LP.T.AVD.ED.	406
<i>C. burnetii</i>L.I.D.IIS.VK.VGID..V.V..E.NNAQL.AKKLAD..MN.IAAKGFAD.	378
<i>E. coli</i>L.I.D.IIG.VA.VG.N..V.V..E.NNA.L.AKKLAD..LN.IAAKGFAD.	377
<i>H. influenzae</i>L.I.E.VIA.VN.VG...V.V..E..NNA.M.Q.LAE.DVN.LTAQS.QQ.	377
<i>M. extorquens</i>	..A..N..W..E.V.K...VKID..LIV..A..N.D..KK.LAE..LDLITA.T.TE.	377
<i>T. aquaticus</i>T.A.E..K.VIR.LE.GLLTK.V.M.VA..AE..AKKLEGGK.PM.PTS--IE.	366
G8625	AKKVVELTQN*	427
<i>N. frontalis</i>Q.ANVVGLAKENGIKIKIEN*	402
<i>T. vaginalis</i>	GE.A.KAAKGEKF*	407
Pig	...A.ASVTKK*	417
<i>C. burnetii</i>	..EQI.KQVGVI*	390
<i>E. coli</i>	..QQ..AAVEGK*	388
<i>H. influenzae</i>	..ELA.NAAKGEH*	389
<i>M. extorquens</i>	..R.A..ACHGAKH*	390
<i>T. aquaticus</i>	..AI.AMVGAA*	378

Figure 4.

mitochondrial proteins, namely seven basic residues (M1, R4, K5, K12, R18, R28 and R29), five serines (S3, S6, S8, S11 and S17) and no acidic residues (Hurt and van Loon, 1986).

ACKNOWLEDGEMENTS

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

REFERENCES

- Bridger, W. A. (1974). Succinyl CoA synthetase. In Boyer, P. D. (Ed.), *The Enzymes*. Academic Press, Inc., New York, pp. 581–606.
- Heinisch, J., Ritzel, R. G., von Borstel, R. C., Aguilera, A., Rodicio, R. and Zimmermann, F. K. (1989). The phosphofructokinase genes of yeast evolved from two duplication events. *Gene* **78**, 309–321.
- Hurt, E. C. and van Loon, A. P. G. M. (1986). How proteins find mitochondria and intramitochondrial compartments. *Trends Biochem. Sci.* **11**, 204–207.
- Johnston, M., Andrews, S., Brinkman, R., et al. (1994). Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome VIII. *Science* **265**, 2077–2082.
- Klein, P., Kanehisa, M. and Delisi, C. (1985). The detection and classification of membrane-spanning proteins. *Biochem. Biophys. Acta* **815**, 468–476.
- 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, New York.
- 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.
- Tsou, A. P., Lai, C., Danielson, P., Noonan, D. J. and Sutcliffe, J. G. (1986). Structural characterization of a heterogeneous family of rat brain mRNAs. *Mol. Cell. Biol.* **6**, 768–778.