

The Complete Sequencing of a 24.6 kb Segment of Yeast Chromosome XI Identified the Known Loci URA1, SAC1 and TRP3, and Revealed 6 New Open Reading Frames Including Homologues to the Threonine Dehydratases, Membrane Transporters, Hydantoinases and the Phospholipase A₂-Activating Protein

MARIA TZERMIA†, OURANIA HORAITIS§ AND DESPINA ALEXANDRAKI†‡*

[†]Foundation for Research and Technology-HELLAS, Institute of Molecular Biology and Biotechnology and [‡]University of Crete, Department of Biology, P.O. Box 1527, Heraklion 711 10 Crete, Greece [§]The Murdoch Institute for Research into Birth Defects, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia

Received 15 August 1993; accepted 1 December 1993

We report the entire sequence of a 26.4 kb segment of chromosome XI of *Saccharomyces cerevisiae*. Identification of the known loci *URA1*, *TRP3* and *SAC1* revealed a translocation compared to the genetic map. Additionally, six unknown open reading frames have been identified. One of them is similar to catabolic threonine dehydratases. Another one contains characteristic features of membrane transporters. A third one is homologous in half of its length to the prokaryotic hydantoinase HyuA and in the other half to hydatoinase HyuB. A fourth one is homologous to the mammalian phospholipase A_2 -activating protein. A fifth one, finally, is homologous to the hypothetical open reading frame YCR007C of chromosome III. The sequence has been deposited in the EMBL data library under Accession Number X75951.

KEY WORDS — Genome sequencing; Saccharomyces cerevisiae; chromosome XI; catabolic threonine dehydratase; membrane transporter; hydantoinase; phospholipase A_2 -activating protein.

INTRODUCTION

In the course of the European community (BRIDGE) project of sequencing of the yeast Saccharomyces cerevisiae chromosome XI, we have determined the complete sequence of 24 577 nucleotides on a DNA fragment mapped near the left telomere. This fragment includes three previously sequenced genes, URA1, TRP3, RSD1

*Author to whom correspondence should be addressed.

CCC 0749-503X/94/050663-17 © 1994 by John Wiley & Sons Ltd (SAC1) and part of the 3' non coding region of the UBA1 gene. In addition, it contains six unknown open reading frames (ORFs), the function of which will be discussed below.

MATERIALS AND METHODS

Strains and vectors

Cosmid pEKG100 was provided in *Escherichia* coli strain TG1 (Δ (lac pro), thi1, supE44, hsdD5, F'



Figure 1. (a) EcoRI restriction map of the 24 577 base pairs of cosmid pEKG100. The remaining 7556 bases 5' of the yeast sequence in that cosmid are presented separately (Alexandraki and Tzermia, 1994). The 3' end of the insert is a Sau3AI site. The numbers below the bar indicate the size of each EcoRI fragment. (b) 6-phase ORF map of the 24 577 bases. Small bars indicate initiation codons and full bars indicate stop codons. The location and the direction of nine ORFs are indicated by arrows. The number in the name of each ORF indicates its size in amino acids and the letter identifies each of the 6 possible reading frames.

(traD36, $proA^+$ B⁺lacI^QlacZ\DeltaM15)) from Agnès Thierry and Bernard Dujon (Thierry and Dujon, in preparation). It is one of the cosmids from the library of chromosome XI, derivative of pWE15 plasmid, containing a 32·1 kb partial Sau3AI yeast DNA fragment. Escherichia coli strain DH5a (supE44 Δ lac U169 (φ 80lacZ Δ M15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) and pUC18 vector were used for all subsequent subcloning and sequencing steps. Gene disruptions in yeast strains were performed according to Rothstein, 1983.

Sequencing strategy

We have used directed sequencing of ordered restriction fragments. Cosmid DNa was digested with *Eco*RI, electrophoresed and purified from low melting point agarose. Six *Eco*RI fragments were subcloned into pUC18 vector. The order of the *Eco*RI fragments is shown on the map of Figure 1a.

Double stranded template DNA was prepared by the alkaline lysis-PEG precipitation method (Ausubel, *et al.*, 1987) and sequenced using ³⁵S]dATP and the Sequenase kit (United States Biochemical Corp.) following the supplier's protocols. Sequencing of both strands of fragments subcloned in both orientations was performed by 'universal' and 'reverse' primers on nested ExoIIImung bean deletions (Ausubel, et al., 1987) of the *Eco*RI fragments. Synthetic oligonucleotides (made on an Applied Biosystems synthesizer by the Department of Microchemistry at I.M.B.B.-Crete) corresponding to internal sequences were used as primers to fill in the gaps. The junctions between the sequenced EcoRI fragments have been determined by sequencing from primers corresponding to sequences near the ends using cosmid DNA as template. Samples of sequenced DNAs were electrophoresed on 40 cm long 6% or 4% polyacrylamide gels with single or double loadings.

Sequence analysis software

Restriction and ORF mapping of the sequences were accomplished by the DNA Strider software (Marck, 1988). Comparisons of the nucleotide ECOR I 1 <u>GRATIC</u>TART TETERAGGEG GRAATGEGTA CITTERAGGAR CEPCTATIET CETATARAGE TITERCART ARCTARGECA APETRIERTA OFFICECOUTE

101	GCCITTTTCT	TTCCGACTAT	ATAZATGCAA	ATAGTCAGAA	GTTGTAACTC	СААТААСААА	CCCCTCATAC	CGTCCTCAAA	TTGGCATATC	GCTATCAGAA		
201	TGAGAGGCGA	TGGGGTTAAA	GATAGGAGTA	TAGATGIGTT	ATCCCTCAAA	CATTTCGAAT	CCC AGAAA GT	TGTTCTGCCT	CAAGACCTTT	TCATGGATAA		
301	CTTCACCTGG	ATGTTTTATG	AGTTCTTCAA	CTGCTTCACG	TITCGTACCT	GGTTGCTATT	ATTACTGCTA	ATGTOSTTCC	CAGGTTTTCT	TTCGCAGATA		
401	AAGTCCATCA	ATCGGATTTT	TCCGTTTAAA	CITIGIATI	TAGTCTCATG	CCTAGTGGGG	ATATTTTAC	CCANTATTA	TTCATTCTCT	CATAAAAGTG		
501	TGTTAACGAA	TCAACTTACT	CAGTTCTCCA	AAGAAATTGT	TGAACATGCA	CCAGGTACTG	ATACTCATGA	TTGGGAAACA	GTTGCGGCAA	ATCTAAATTC		
601	TTACTTCTAT	GAAA ATAAAG	CTTGGAATAC	TGAGTACTTC	TITTTCAATG	CTOCAGAGIG	TCAAAAAGCA	TTCAGAAAAG	TTCTTCTCGA	ACCATTCICT		107
701	GTGAAGAAAG	ATGAATCTTC	GAAAATA AAA	TCATTTOGGG	ATTCTGTCCC	CTACATCGAA	GAGGCCTTGC	AAGTCTATTC	CACAGAATTT	GACAAAAAGT	Ì	Š
801	GGAAGTTGTT	CAATACTGAG	AAAGTGTGGA	OCCCTGATAA	CTTGGAACAT	GTICAGCTCC	CCANGAAAAC	TTATOSTTAT	AAGTTIACCT	GGGTTCTCAA		
901	GAGGATTTTC	AATCTCTGGC	TTTTTCCAGC	ATTTATTCTG	TTTCTGGCTT	GTATCTATGT	GTCATGGGAT	AAGGGCCATC	TATTTOGTAT	CTTGTGTTGT		
1001	GGGGGAGGTT	TTCTOCTTAT	GGTAAGAGTT	TITCAAAATA	TGAGGCCCTT	TTOCATGCAC	ATGGAAGACA	AAATGCAGTT	CTTGTCAACG	ATTATAAATG		
1101	AGCAAGAAAG	TEGTECGAAT	GGATGGGACG	алаттосала	GAAAATGAAC	AGGTACTTGT	TTGAGAAAAA	AGTCTGGACT	AGTGAAGAGT	TTTTCTITGA		
1201	TGGGATTGAC	TGTGAATGGT	TTTTTAACCA	CITCTTTTAC	COCCTTCTAT	CTACAAAGAA	ACCIATGTTT	GATAGACCTC	TAAACGTGGA	ACTATGOCCA	IJ	1
1301	TACATTAAAG	AAGCGCAATT	GACCCGCAAA	CAGGCGCCTC	COSTOTAGGA	ACAAGCATCC	CTTTTCAATT	TATAGATCGT	GTAGTITATT	AGATGCGAAT	۲	
1401	Асссталаса	AGGTCGCTTA	TTATTATTAT	TACAACTATT	CTTTTTCCTA	TTOCCCTCTG	GTTAACAAAA	ACCAGTATAT	таасадааат	AAATAATAAT		
1501	ААСАААААА	TATGAATATA	AGAATAAATT	ссаатааата	GTAATATITA	TAAAATATAA	аладаадаа	TTTTTATTCT	TTTTATATTT	TATTATAAAA		
1601	TAGTTCAATT	ATTTATATTA	TTTAAATGTT	GIAATTIATT	ATTAATAATT	ATATTTTATC	AATATTTATA	CGATATAAAA	аттталала	ATTATAATAT		
1701	AGAAATAATA	AATAATAATA	ATATGATAAT	AAAATAATAT	аталаталаа	ATGATAAGAT	TATATTAATT	TATTIAATTT	TCAAATTATT	алатааатаа		
1801	таасстаата	CTTACTTTTC	CTAATATTAT	TAAAAATGGT	TATTGTCAAA	CCCATTAACT	GATTATATTA	TAATATTTCA	ATATTTAATT	CACTTAATTT		
1901	GTTGTTATAA	TAATTAATGA	AATAGAGAAA	CTATAATCCA	ATTGATTGTG	CAACAGTTIG	ATCTTCCAAA	TCATAAAAGA	ATATTGAGTG	TTAGTTOCTA		
2001	CTGGTCATAA	CTATCCATT	TGAACATAAC	CATGTAACTC	TTTTCCCGTC	TACTGTTAAT	салоддалса	TGTTGAAAGT	TATATCAATA	CATGTGTTGA		
2101	GTATTTCACG	CGAATTTATT	GCTACAAAAG	AGCACTAAGT	AATACATCGT	TAATTTCTGC	GGTAGAACTG	ACCTOGTGAC	CTTGCTTGAT	GCTTTATTGT		
2201	TTGTGGGATC	TGTCACGTAT	ATCGCCATAT	TTTTTTCCTA	COGATATAAT	GTAAAACACT	AATTTATCCT	TCAGOGTTGC	TCTTTAAAAC	TCAATTGTTG		
2301	AATAACTTGG	ТАТТСАТАТА	TGTGGCTGCT	CGAATATCT	TGACACATAT	TATGATTATC	алалалаадт	GAAGAACTAA	ATTTAGGAGG	ATTTCTACAG		
2401	TTTTGGTCTT	GGTTAGGTAA	тааттестта	AAATAGAGGC	TTATCCTTTC	CTGTGGACCA	TAAATATCTT	CATGCAAGAT	CTCGATTGAT	TAGTAATAAG		
2501	AGCAGACGAT	ATTGTGTGTG	GAAAAGAGAA	TTTTTTCTGC	GINTANANTA	GGACATTATG	ACAGTTACCT	TTAATTGAAG	TAGAGATATG	TTTTTTCAAC		
2601	TTAGAAGCTA	TGTCTPTACT	GGGTGTAAGC	CTAGGAACGC	TCTTCTGTTT	TAGTTCAAAA	TATOCAAAAG	аластаатат	ACCAGTAATG	CTCATAGTCA		
2701	CAGTCAGAAG	TTATATTATT	AGTTTGACAT	AAGCAGTATT	CTTCCCTCTC	ACACACTATC	ATTCTGCCAA	AAAATTTTCC	CACATAAGGG	TGTCAGTTCA		
2801	GTGAAATATA	ATACTAGTAG	AAAAATATGT	AGAGTTATCA	GAGTATAAAT	CAGCTAGAGG	CTTTTTTTGC	TGAACTCACT	TCAAAATCTC	AACCATGTTA		
2901	TTAATGTTAA	GTCAAATGAT	ATATGCCGTG	tgaaagtaga	CAACTAAAGC	TTTCCAATTC	ттсастатаа	AACTGAACTC	ааааттсааа	AGTATACACT		
3001	AATCGTTGG	GCCATCTTCC	TTCCCAGAGA	TTAAAGTAGC	ATATCTCTTC	ATGTCTACAT	TACCTCCACT	TAGTATTATG	CCTACTTTCT	TCCCAACTAG		
3101	CTCTTCCTTT	TTTAGAAGTG	CACCTGCAAA	TCCCAAACAA	GCTGTGGGCT	CANCAACCAC	CTTCATGCGT	тесссалдал	AGTGCATGCA	TTTTACTAGC		
3201	TCTTGGTCGC	TAACCGTTAA	AATATCATCG	ACATTTTCGC	GAATAATGGC	AAATGTGTAC	TCACCGAGGT	GTTGIGTTTG	AGCGCCATCT	GCGATAGTTT		
3301	TTGGCGTATT	GATATGAACA	ATGGAACCCG	ATCTGAAGGA	TIGTTGTCCA	TCGTTACCAG	CTTCAGGTTC	ААССССАААА	ATTTTGCAGC	CIGGAGAAAG		
3401	GCTTCTAGCG	GCAAGTGCAG	ATCCTGAAAG	GAGCCCACCA	COGCCCAAAG	GAACAAATAA	TGCATCAAGT	TGTCCAACCT	CTTCTAATAG	CTCTTTTGCC		26
3501	GACGTACCTT	GCCCTGCAAT	AACATCAGGA	TGATCGTAGG	GOGGAATCAA	TGCAAAACCA	TGTTCGGCTG	CTAGTTGACG	CCCAATCTGC	TCGCGATCTT		ã
3601	CAGTATACCT	GTTATACCTT	ATGATATGTG	CTCCGTAACC	GOCTGTAGCA	GCAACTTTAA	GGGCGGGGGGGG	ATCCICGGGC	ATAACAATTG	TIGCAGGIAC		
3701	ATTTAATAGT	TTTGCACTAA	GAGCAATAGC	CTGTGCATGG	TITCCTGATG	AGAAGGCAAT	TACTCCCTTA	CTTCITTTTT	CATCACTTAA	TITTGAAACA		
3801	GCATTCATTG	CTCCACGAAA	CTTGAACGCT	CCACTOGTT	GGAAATTCTC	ACCTTTAAAG	TATATTTGTG	CTCCAAGTCG	ATCATTAAGC	ATTCGTGAAG		
3901	TGAGAACCGG	TGTTTTATTT	ACATATTCTT	TAATTCIGTT	GCTGGCATCC	AAAACGTCTC	CATAAGTGGG	AACAATCATT	TTTCCAACTC	AATGAAAACT	'	
4001	AAATATCTCG	ААСТАТАТТА	ATTATTTGTA	CTTTTTAAGT	TTTTTTCAA	TACTCCTTTA	TTTGAATTTT	TGCACAATGA	AAATGTAACT	CTTTATATTC		
4101	CAAAGCTTAT	CTTTTTCTCA	CGTACCGTCC	TATAGACCCT	GATTTTATGA	TAAGAGGCTG	AATTTACTAA	TTAGICATTT	TTTTCAGTTT	TTTTGTCATA		
4201	GAAGCCACAT	TTAAAAGAAG	TCAGCCAAGT	TACCTGTAAA	TATTAAAACA	ACCTGCTTAG	AATGACTGTT	AATTTACAGA	ATTTACGATA	AAAATATAAC		
4301	TTGGCTCTAG	GGCGTGATGA	TAGTGACAAA	AAACTCGTCA	AGTAAATGTT	GGGAGCTTAC	TTGCCGTCGT	TTTCTAACGC	CAAAATACAC	GAATCATGTT		
4401	TGGGAAATAT	AAGTGATTAG	AACTTGTCCA	ACGATATIGT	CAAACAATAA	CATACTACAA	TAATCCTAGA	TCGTGTATTA	TTTGCGTATG	GATAAATATA		

A6

4501 CCTCAAGCTT TATGAGGCAT AGAAAATTCA TTAGAGTCAC TACTTACGTC GCACTCTCCA GGATTGGAAT GACGAAATCT TCTTCACGAT TTCACCATTC 4601 ACCTITICAA GOTTOCTATO TOGTTOCTTA AGTAGAATGA AAGCATAGCA GOCCAAAAGA AAAGCOTOCO TOAGCAAGAA TATTAACAAA GAACAATCA TEGACCENTT CTACTACTTA ATATCTCTAT CACAACTETC AGTECTCAATC AGTACTTAAA TEACETCAC GAAAAGATA TEATTATTTT TATCATCCCC 4701 CTAATAAAAAT AAGTAATGAA TCAATGCCAT TAATTCGGGA ACTAGTCATA TACCTAAGGA ATCAGGCTTT GATAGCTCTT TAATAATAAT ATTAGGGCAT 4801 4901 TITIGATATE STAAAAGATT AGATTAGGAA ACCCCTTITE TAGACCATCE CACTACAATA TETECAAAAA CGGTCACCTT GCATAGGCAG AACACGCTAC 5001 CTITIGAGITA ATTCACAGGG GAGAGCCCTT ITTITITCTT ACCAANAGIG CITAACAATC ITAATTCCTG GAATGITITC GGCGGATTGT IGTATAITGT TTGAATTTGT TCGTTCTAAA AAGTGAGACA CCTGCACTTT CCACGCCAAG AACACTAATA CCTTTAAACA TGCCACGCGTA CTTTAAAATG CCACGAGTCG 5101 EcoR I CCCCCGTACCA AACATGCTGG CGATTGTAAT TTTTCGCCTT TTATTGGAAT CGACAATTCG GGCACTGTCC TTTTCTTCGT CATTTGAATG TTATAATCAT 5201 TITTATCGCC AMATTITCTT CANTANANGG ICTCTTITAN ANTAIGITAC INIGITITTI ATTITICIAL TAITTITCTA IGTGATITTI CAGACGAITT 5301 5401 ATTCTTGTT CGAAGGAGAA CAGATGCAAC GGTCAAGTTT TIACTATCCA CATGAGGAGC CGTAATATTC TAAAGGCAAA AGCATGTGTG CTTGTAAAAA AAAATCAAAA AAAGTTACTG CTCCAAGCCA GCAAGCTTTC TCTTTTGCGA AATCTTTAAC ACGGTTAAAA TCAAGACAA TTACTTCATT GACAAGGTCA 5501 ATGCTAATAT TTACGETCTC CECACGAAAG ACTCTGAAAA ATGTTAAAAT TTATAGTGGC TAATECTAAT TCAAGAATAG TCTCGGETTG TTTGATCACA 5601 5701 TTATTTAAT TTTTATTGAG CACTGAATAG GACCCGTATG TICCAATGAG TCTATTTCTT GCAACTTAAT CTGGTTTGTT TTTAGGGGAG IGGCATTTTT TTTAGGGTTA TTCAMATATT CAATTCTGTA ATTTGCTTAC GACAMAAAGA AAAGGAGAGA TGAMAAAGG GCGACGGTA TAATCTATTC AGCCGCCAAT 5801 GTCCCGACGT ATGCCAAAAT GTGGTGCCCC GETAACAGTA GCATCACACA GACCATGAGA AAGCAAGAA AAGTGGAAGT CTGTATGTCA CATTTTATT 5901 GANAATGCTA TITGCTATAA GGTTCGCTAT TCAACTGATT GAACCGAGAT AGACAAAATG TTAAACTTTT CATAGAGACG TTATTGTGAT AGGGTTTACG 6001 6101 CTCAGAATCG TTGAAACATA CTTGCTAATA AACCACGTCA CGTCAARATA TCAGTTTTTC TTTTTAGTT ATTCCCTTCT CACTTTTGAC TTGAGATGAG 6201 ATCGTTACAC TTTTATTGAA TACGAACGGA TCACGATACA TCCACATATA CGAATGTTCG GCCAGACAGG GCATCAGAGT ATATACCCTT TTTCCTAAAA TEGAAGEGEC CTAGATCGTT GECATCAGEC GTTACCTAAC CGAAGTGGAT CANATCAGEG CCTATTTTTC CGTTTAGAAC GATTAGATGE CGTCAATATG 6301 6401 TATGTGCTTA CTGTATCATC CACGTAGTAA ASCTGGGGGG TAGCATCOCT AGGTTACTT CTCAATACTG GGGTOCCTGA CCAAAGATAA AACATAACCT 6501 ATTATGGCGT ACGAGTCAAA GTTAAACGCT ATTGACTCG GCCAGCATGC ATCACCCGGG GGAATGGAAA ACTTCTCATG GCTACCCCAA CGAAGCCAGA CARGATCING GAGATGICIC CACCITICCGA GCTACACTCIT CACATCICIT ACATAGACAC TGTGAAGTAA GTGATTGCAC ACAATCCAAA GAAGCACATC 6601 ANATANTATE TANATATTAC COGTACANAT GEGAAATCAT GECACTITC COGTAGTCA TOGGETOGGE COGTOCTOC ACTITOCITT TOTTCOGAI 6701 GTTAATTAAT ATATAAATGG ATCTATAATT ITCAAAAGGT ATAAACGCAC AGTATGAAAA ATATTGCAAG AATAATGGTGT TTATAAATTA TTTTTTTTGC 6801 TEGETAGEARA ATCAACTEAT TEACATETECAT TEAGAGEETA ATEGAACGET ATEGAACACT TEACACACTE TEACAACAATA COATTAGET TAAGTEGATG 6901 GACCCCCACG CTTAGTGTTC CACAGGTTTG TCCCCACTGT TTTTACATTC CACTGTACAT TTTTGCAATA GAAGGTCATT GTATGCTACC TTGGGCGCGCT 7001 AAGAATACCT GTAAAAATTT GGAGAAATTA GATTCGTAAA GAATGACTCG CAACGACTCC AATGATTCT TCTTTCACC CTTTGACGG CCGATATCCG 7101 CECCEGEATCE TEACCEEGEA ATTTACTEEA CTAGACEGEC SIGTTTETET TITTECTTT CETEGEGETTA GAGECCAAGA GETAATAGEC GACAAAGGA 7201 CTCCARARAR ARARGEAGEC ACAGGACARA CECAGGACCT GOSTCATTCA CECTGARGES GEAGGAGCA TTTTOGATCA GETECAATTA AATGAAGACT 7301 ATTCCCCCTA CCGTTCCCAG ATGGCTCCGA AAGTCACTGA TOSAGGAAGT TATTGACCCC GCGCCTTGAA ACTATTTCTC CATCTCAGAG CCGCCAACCC 7401 7501 TACCATTATT CTCCACCAGG AAGTTAGTTT GTAAGCTTCT GCACACCATC CGGACGTCCA TAATTCTTCA CTTAACGGTC TTTTGCCCCC CCTTCTACTA TAATGCATTA GAACGTTACC TGGTCATTTG GATGGAGATC TAAGTAACAC TTACTATCTC CTATGGTACT ATCCTTTACC AAAAAAAAA AAAAAAAAA 7601 AAAAAAAAAA ATCAGCAAAG TGAAGTACCC TOTTGATGTA TAAATACATT GCACATCATT GTTGAGAAAT AGTTTTGGAA GTTGTCTAGT COTTCTCCCT 7701 TAGATCTARA AGGAAGAAGA GTAACAGTTT CAAAAGTTTT TOOTCAAAGA GATTAAATAC TGCTACTGAA AATATGTCGT CGTCAATTAC AGATGAGAAA 7801 ATATCTGGTG AACAGCAACA ACCTGCTGGC AGAAAACTAT ACTATAACAC AAGTACATT GCAGAGCCTC CTCTAGTGGA CGGAGAAGGT AACCCTATAA 7901 ATTATGAGCC GEAAGTTTAC AACCCGGATC ACGAAAAGCT ATACCATAAC CCATCACTGC CTGCACAATC AATTCAGGAT ACAAGAGATG ATGAATTGCT 8001 GGAAAGAGTT TATAGCCAGG ATCAAGGTGT AGAGTATGAG GAAGATGAAG AGGATAAGCC AAAGCTAAGC GCTGGGTCCA TTAAAAGTTA TGCTTTAACG 8101 AGATTTACGT CCTTACTGCA CATCCACCAG ITTTCTTGGG AGAATGTCAA TCOCATACCC GAACTGCGCA AAATGACATG GCAGAATTGG AACTATTTT 8201 8301 TTATEGETTA TITTEESTEE TTETETEESE CITEGESETT CITTEESETT TEASTATEAS TESETEATTA GECTEAACTA TATEACAGAC CAACCAACGA CATCACCTCC GEGTTEGGAT TEGTEGTTATT TETTECTTCA GEAGETECTE TEATATTEE TTTATEGACA GATAMETETT COAGAMAETE CCCGTACATT 8401 ACATGITIGT ICTIDITIGT CATTGCACAA CICIGIACIC CATGGIGIGA CACATACGAG AAATITCIGG GCGIDAGGIG GATAACCGGI ATTGCIAIGG 8501 GAGGAATTTA CCCATCTCCT TCTCCAACAG CEATTGAAGA TCCACCTGTG AAACCACCTT CCTTCCTATC AGGTCTATTT TTTTCTCCTT ACGCTATGG 8601 GTTCATATTT GCTATCATTT TTTACAGAGC CTTTGGCTAC TITAGGGAIG ATGCTGGAA AATATTGTTT TGGTTTAGTA TTTTTCIACC AATTCTACTA 8701 ATTTTCTGGA GATTGTTATG GCCTGAAACG MAATACTTCA CCAAGGTTTT GAMAGCCCGT AAATTAATAT TGAGTGACGC AGTGAAAGCT AATGGTGGGG 8801 AGCCTCTACC AAAAGCCAAC TTTAAACAAA AGATGCTATC CATGAAGAGA ACAGTTCAAA AGTACTGGTT GTTGTTCGCA TATTTGGTTG TTTTATTGGT ¥ 8901

Figure 2. (Continued).

9001	GGGTCCAAAT	TACTTGACTC	ATGCTTCTCA	AGACTTGTTG	CCAACCATGC	TGOGTGCCCA	ATTAGGCCTA	TCCAAGGATG	CTGTCACTGT	CATTGTAGTG	1:	
9101	GTTACCAACA	TCGGTGCTAT	TTGTGGGGGGT	ATGATAITTG	GACAGTTCAT	GGAAGTTACT	GGAAGAAGAT	TAGGOCTATT	GATTGCATGC	ACAATGOGTG	li	
9201	GTTGCTTCAC	CTACCCTGCA	TTTATGTTGA	салсссаала	GOCTATATTA	GGIGCCGGIT	TCATGTTATA	TTTTIGTGTC	TTTGGTGTCT	GGGGTATCCT		
9301	GCCCATTCAC	CTTGCAGAGT	TGGCCCCTGC	TGATGCAAGG	GCTTTGGTTG	CCOGTTTATC	TTACCAGCTA	GGTAATCTAG	CTTCTGCAGC	GCTTCCACG		
9401	ATTGAGACAC	AGTTAGCTGA	TAGATACCCA	TTAGAAAGAG	ATGCCTCTGG	TGCTGTGATT	AAAGAAGATT	ATGCCAAAGT	TATGGCTATC	TTGACTOGTT		
9501	CTGTTTTCAT	CTTCACATTT	GCTTGTGTTT	TTGTTGCCA	TGAGAAATTC	CATCGTGATT	тотототос	TGTTATGAAG	алатататаа	ACCANGIGGA		
9601	AGAATACGAA	GCCGATGGTC	TTTCGATTAG	TGACATTGTT	GAACAAAAGA	CGGAATGTGC	TTCAGTGAAG	ATGATTGATT	CGAACGTCTC	AAAGACATAT		
9701	GAGGAGCATA	TTGAGACCGT	TTAATCACTT	TTCATTGCTC	TCTAGGGCGT	GTTCGCTTCT	CTATGTAACT	GCATTICACA	TATATTCATA	TTATAGOCTT	¥	
9801	TATAACATTA	CATAAAAGAC	ATGATATACG	ACATACITTA	AAGTTAGAAT	TTTTGTTGIA	TTAATTTTCT	CGAAQGGATT	AGAATGTAAG	TACATTIAAA		
9901	TAGCGCCGCA	адааасатаа	CAGCCGTTAA	CACTAGCAAG	ATGGTAGATT	ACIGIGTCIA	AGAATTTAAA	TIGTITIGCT	TECTTITECC	ACTATAAGAC		
10001	AATAATAGTA	атадалалда	AATGTTAGTC	GAACAGCAAA	TIGGTAAAAG	tggaaaaat	TATTTCGTTA	TTCATTTTAG	GCCGTTCAGA	аататассат		
10101	TITIGGTTCTT	GCACTGAATG	GGCCCTGAGG	ACAGTGTGCG	GCTGGATAGC	CATATACTIC	GACATTACTT	TTCACCTTTG	AAGGGGGATG	TATGCTTATA		
1 02 01	CACAATCAAA	GCTCGGCCTG	TGCAGGTTTC	ACTACATAAG	CTTTTTTACC	ATAGAAATCT	TGCAGAAGGA	AGTAAGTATC	AAGATATTAC	GTATCGCATC		
10301	CATAGAACGA	ТАССАСАЛАА	ATGCAGTATA	TGATAACGAT	ATGCTTCACA	GCTGCGGCCC	ATTTGGGGAA	алалссалал	GGAGCATCCT	CGATCTIACA		
10401	GATGTCGCTT	TTTCCTCTTG	AGGTCATTCC	CTAACCTTGG	GATCCAAGTC	AAICTGGTAT	CTTOCCACCC	TAAATAGTGT	ATTCCCCATG	TAATAAGATC		
10501	GGACGATAAA	CTTCGAAACA	ATTCTAGCCT	TITTTATTAG	AUATTTACTC	TACTTTAGAT	TTGGTGATTT	TACTAAATGT	TATCCTTCAG	CCACTACAGC		
10601	CTACTTTTTC	TCACTAATAT	GTTCTTGAAG	CATTTICTT	CACCGGGCTT	TGCCGAAGGT	TATTTCACGA	TCTTOGGCAT	TCGCCGAAAA	TTTTTCAAGA		
10701	TGCCCATCAC	САААААТАА	AAAAAGCTTT	ATATATGGCA	TOGCATACGC	ATCEGGAACC	GGTATTCTTC	TCTTQGATTA	CACAGTTTGG	ACTACTAGAC		
10801	GTTICAAATA	GATATACACA	аттастсааа	AAAAAATTGA	ACTTCCGTAC	CANAGATGAC	AGCCAGTTTA	ACTACCAAGT	TCTTGAACAA	TACCTATGAA	1	
10901	AACCCATTTA	TGAATGCATC	CGGTGTTCAT	TGCATGACTA	CACAAGAATT	AGATGAATTA	GCAAACTCTA	AAGCTGGCGC	ATTCATTACA	AAGAGTCCTA		
11001	CAACCTTAGA	AAGAGAAGGT	AACCCTGAAC	CACGTTACAT	TICTGTCCCT	CTAGGCAGTA	TCAACTCCAT	GGGTTTACCA	AACGAAGGTA	TCGACTACTA		
11101	TTTGTCCTAT	GTATTAAACC	бтсалалдаа	TTATCCTGAT	GCACCTGCTA	ттттсттстс	AGTIGCTGGT	ATGAGCATTG	ATGAAAATTT	AAATTTGTTG		
11201	AGGAAAATCC	Eo AAGATAGC <u>GA</u>	OR I <u>attc</u> aacggt	ATTACCGAGT	TAAACTTGTC	TTGTCCTAAT	GTGCCTGGGA	AACCACAAGT	TGCTTATGAC	TTTGACTTGA		314
11301	CAAAGGAAAC	CTTGGAAAAG	GTTTTTGCCT	TTTTCAAAAA	ACCTCTTOGT	GTCAAGTTGC	CTCCTTATTT	TGATTTTGCC	CATTTIGATA	тсатебсала		<
11401	AATATTGAAC	GAGTTCCCAT	TAGCTTATGT	CAACTCUATC	AATAGTATAG	GAAATGGTCT	TTTCATTGAT	GTGGAGAAGG	AGAGTGTAGT	AGTGAAGCCA		
11501	AAGAATGGTT	TCGGGGGTAT	TGGAGGTGAA	TATGTTAAGC	CAACCGCQCT	CGCCAATGTT	CGTOCATTTT	ACACTCGTTT	GAGACCTGÀA	ATCAAAGTTA		
11601	TCGGTACAGG	TGGAATTAAG	TCCGGTAAGG	ATGCATTTGA	AGATCTTCTA	TGTGGTGCCT	CTATECTACA	GATTOGTACA	GAATTACAAA	AAGAGGGCGT		
11701	CAAGATTTTT	GAACGTATCG	AAAAAGAATT	AAAAGACATA	AIGGAAGCTA	AGOGTTATAC	ATCOATAGAT	CAGTTCCGTG	GGAAGTTGAA	CAGCATTEAA	¥	
11801	TTATCGGAAT	TTGAAGAACT	ATGATOGGAC	AAGTTATGGA	AGAGTGTTAA	GATTCGTATG	AATATTTTAT	TATTTGAAA	TTGCATAAAT	CAAGGADAGT		
11901	TGGTAAATGC	TAATTCCTTA	ATTTTTCGGT	GAGTGCATTC	GCAATTGOCA	TCATACATTA	GGGTTCAATG	TGCTTTAATA	GATGATGATA	CTAAATACTT		
12001	ATACTTTGAA	ATAACTTAAT	CAATGTAAAT	DATTGTCTTG	ATTCATTIAT	ACTTAAATTT	TTATATCTCA	ACATATGATA	TATCCTTACA	TTAAGACTTA		
12101	TTCAAGAAAC	ATTTTTCAAC	GATAATTATT	AGTAGGCTAA	AAATGGCATA	GGAACCAGAA	ANTIMITGTIC	AACAMATGAA	GCCATCIGTT	TAATATTTGG		
12201	GGAAAAAAGT	GAAGAAATTA	ACGAATATGT	CATATATAT	ATATATATAT	ATATATATAT	ATATTGCACA	GGTGATTTTG	ATGTAGAATG	TAGCCTTTTC		
12301	GAAAGAGTAA	TTTGGTTTTA	TCTACAGAGT	ICCGAAACAT	TGTTAAGTGC	CTGTTCTTAT	атааттааас	TGTACTTTCT	GTAAAAGTCA	ATTTGTQGCT		
12401	GGCCCATCCA	CGGTATACCA	ATGGTICCGT	GTAATTOGAT	ATTGACTAAG	аассалалас	алатссттта	ATCCTTGTAT	TGTCCAAATC	OCCACOACC	A	
12501	AGCGGTTTTA	ATGATGAATC	GATCACCGGG	TTGAGCGTAA	ATCGTATTTT	TACCGCCGAC	ATTGATTAAT	GCGCCTGTGC	TATGTCTTAC	CATAAATTC		
12601	TCACCTCTAC	TACCGTCTTG	TCCTCCCTTT	ATACCATGAG	GCCAATGAC	ACGACGTTCT	GATAATATAG	ATGCIGTTAC	AGCCTICCGA	AATTGAACAT		
12701	CCCTTACTAC	GCCATTTCCG	CCAGTATATT	DACCTTICCC	TOCTGAACCT	CTICTGATCG	AGAACTCTTT	CAACAGAACG	GGGTATCTCC	TTTCAAATAC		Ś
12801	CTCAGTATCT	GICATICTIG	TATTGGTCAT	ATTCGTGTGT	ACAGCGTCTG	AGCCATTCCA	ACCACTTCCT	CTCCATGAAT	CTGCACCAGC	CCCCGAGCCG		280
12901	CCACATATGG	TTTCATAATA	TCCAAAGCCT	TIGATCIGCT	TATCGGTITT	GTTACCAGAA	TTTOCTCCAG	TTCCAAATGT	AAAGTTGTTA	CAGTCACCTT		L
13001	GAGAATCTCC	CATGACGTTA	AATGTTTTTA	AGATTACATC	TGTCACTCTT	TGGGAAGTTA	ATACATTGCC	TCCGACAACA	GCAGCACCGG	ATCGGGGGCT		
13101	CAAGAGGGAA	CCAGCAGGAA	TTTTTATTGT	DAAGGGTTTT	AAACAACCTT	GATTTAAAOG	AATATCTTCG	CCTACTAGAC	ACCGTAGACA	ATAAAGAATC		
13201	GCAGAATTAC	TAATAGCCTC	TGGTGCATTT	AAATTACCAT	AAACTTGGGG	AGAAGTCCCA	TCGAAATTAA	AAATATACTC	CTCTTTTCA	GGTCTTATCA		
1 3 3 0 1	TALOPPOOR		Chackan	CTABCOCATO		ABOTTOCTO	TCCCAABATC	TTCASCCATT	TETCOM		i	
13301	TAACITGTAG	I I IAATTAAG	CI ICCATCAT	CIAAGCGA1C	TICOCCAGAA	AAGLIGGTIG	IGCOMMANIG	LIGHNCGATT	TTTGCAGGCA	TTTTTTT	•	

667

Figure 2. (Continued).

E203

13401 GGATTCAGAT GCATTTGTCT GAATTGCAGC CATGTATTTC AATATTGTAG CCAGATCATA TTCCTTAGTT AATGACCCCAA TTAACTGTAT CCCTTTGTGT 13501 TRACTOCCA CITEGECTIT TARATOCCIT AUGUTATCAC TRAATCTICT TEAACCIGAE CASOCIGET ATTICCIES ATCTICIACE ARAACCIGI 13601 ARAITRATTC ITCCIGGARA AITCCCICIT IGACARCCAR ITCGGARTAR AITCCGGITC CITCITCGIA CARTICITIG GARTITEGAG GCACIGARCO 13701 GGGTAGAATA CCTCCAATAT CTGCATGATG GECTCTTGAA GCCACATAAA ATATCAATTE GCCAGTTGAA GAGAAAGAAG GAGTTATTAC TGTTATGTCT GETAAATGAG TACCACCAAT ATCCGGATGG TTTGTGATTA AAACATCACC AGGCTTTAAT TTGCCTTCCC ATAACTTCGC CTGTCCTGAT ATGCAGGTT 13801 ACATEGATCE TANGEGAACA GETACATEGAG GACCETTAGE GACTANGETT COTTEAGAGE COARTAGEGE ACATEGAAAAA TOTAATETEE CETTEAGET 13901 14001 AGTIGAAACA GAAGTITICC TCAATIGAGT ACCCATCIGG AGGCGATAT CCATAAACCT ATGACIGAAA ATAGACAATA AGATIGGATC TATATCAAGT 14101 TCATATCCCC TTTTAGATAA TETCTTCCCA CCCTTTTGGT TCATCTTCAT AAATATATCA GAGTTTAAAA TTGTTGCTTG GCTGTTTGGT AAAATAATAT 14201 TEGTTGGGT ACCEPCAGEC ANTATCECTE GECCTOCAT ANTESTICES SCARGAAGAT CATCAATTIT GAAGACCECA STETEGACCC ATTISTICT 14301 ANAATATGET TTETGGETAA AACTAGEATE ETTEGACAEG TEGGETTTET TEANATEGEN AATTTEANTE AATTGETEAT CAACEGETTT TTETTTTET/ 14401 ACCOUTGATT TOCCANTTGE TOTTATCOTT ATATCATCAA TINTAATTGE ITTETCATCE AATCAGAATE CEAACTOTTT TTTATCAGOT TOAGAGAACC 14501 ATTOCCTARA ATTOCATTGE TOATCAPATT ITTGEBATAT CATCARACTE GTITCAGEAC CITCATATCI GAGATITARA TATOTITICOA AGACTAIGIU 14601 CTCCCCAGAA AACGACTGAG ATAAAAGGTT ITTGATTGAA TITTTAGAAA GTTCTAAGAA ICTTITTTTC ACTITCAAGA IGGTCTCTGG CTCACCCAAG ATANANGAGE AAGGETETTE STITTETEA ATAACATETE EINAGAAAAT GEGATAAGGE GACAAGATAG AAGAATATET ATGAATAAA ACAGTGTEAA 14701 TECCCAATER ATCTECAACE SCTATESEGT SITESCOCEC ASCCCEGECA AAGEAAACTA ACCEATECTE AGATACTACA TETECTITTE CITETERA 14801 ECOR I GGCCCTCACA GGTCTTGCCA TGCATTCGTT AGCTACTTTG ATAAATCCAT AGGCAACCTC TTCCATGGTT AGATTT<u>CAAT</u>ACATCCTT ATTGATGACA 14901 TETETTAATT CECEGAATTE CAAAGTAGTE GETECEAAGE CAAGTGATEC ATCCCCATE GECCCAAAAA TETETCEGGAA GAACTCAGGA ACCAAACETC 15001 CTANANAGAG GITGGCGTCT GINATGGTCA MAGGCCGGCC CITTCIGIAG GCTCCAGGGC CTGGATCAGC AGCTGCTGAA TCAGGGCCTA CTCTGAATAA 15101 15201 CCCGTTTTTC CATGACAAAA TAGAACTTCC TCCAGCAGCA ACAGTATGEA TGTCAAGTTE AGGACATTET ATGATGATCC CTGCGGTAAC AGTTTCAAAA 15301 ACATGITCIA ACCICCAIC GEOGRACIA CIGACAICAG IAGAIGITEC ICOCAIGICA AAGOCIAITA AGGGIAIGII AITAITITIA ICAIAGOAIG TACTAGAATA ACCTATTACG CCACCTGCAG GACCAGATAA AATAGATTTC AAACCGGAAA ATTTTCCCCC ATCCACTAAA CCACCATCAG ACTGCATGA/ 15401 TTGAATGTGG GTATCTTCTG CGTGGCTCAA ACCTGCAGAA ATGCTATTCA AGTACTTTT AATTACCGGA GTAAGGTAGG CATCTGCGAC AGAACTGTGJ 15501 15601 CCCTCCCCTA AAAATTTTAT CATTCACAGAG ACTTCACATG ATAATGAAAC ATCAGAAAAC CCAATCTCCC TGGCAATATT ACCAACAATT CGCTCATGG 15701 CTGGAAAAGT GTATGAGTGT AAAAATGCTA TCGCAATAGA CTTGATTCCG CTAGCATAAA GCACTTTCAG AATAGATCTT ACGCTTGATT CATCTGGCT 15801 CTITATGACC CITACCATIT CCCCACIAIT ACCITCCAGG ATACCITCIT GITCATIAGG AGACATITG STAAAATAAG GGTCITCCGA GAAATCITCI 15901 ARAGTARCTC TCTCATCTAT TTCARCTACA STGTCATATA ATGGCACTAC CTTTTTTATA TTTARGTTGA AGATATCTGG CCTGGTTTGG TCTCCAATCI 16001 COMMEGANCE CTEAMAGECT TECHTAGEAR TEAMEGEACA AGETERICA TELETICATE GAGGEGAGET TETAGETAGE GETERECEA TECTEAMACE TOTOTACATTA GAGATGTOTA GAGGAATACC GEGAGGAATT GICTIGIGIT CAMCACCIC CAACAGAEGT CITATGCCTT CAAGGGGAGC ATCTGGATAC 16101 TITTIGGGAT CCACCEACAA GAGCTIGATE ACAGTATEAT GITCITEGITT ACCEGITACEA ATAITCCCCCA CACAATCAGT GAAAGTICCA CCETTATEGA 16201 16301 TEGECANTTET TATETTEET TETECATES INTACCANGA ASTTATATT ASSITTTEAN TECTANANT TATTACTETT TANGTEENST TETTASTTETT 16401 TEGCATTETT TETTEAACAE CAEGTAACAE TEGATETTAT ETTEETTTE AGUEGTAATE TEATEGEGAT CETECTAAT CEGAAAAGCA CETEAGECAA ACAMAACAAT TETTEMAGTA TEATECAAA TAATACTEAA TACAECEGAG TACECITETT CITMATTCCA GAGATTEETA AAGTETTAT CATTECTAC 16501 16601 CTCTAGAAAT AAAAATTAAT TATTTAGGTA AAAAGTAGTA CTGAGGTTTC ATATCTGTAG ACTGATGGAG TAAAAAAACA TTGAAAAGAC TATAAAGAAA TTATAAGTAG ATACACAATT GCTTAACCTT TCATGTAAGC ATCTAATTCA GCGTCAAGAG CTTGACAGA AGTAGGAGTG TTATTCTTCT CACGGCCCTT 16701 GTTCTTTTTG GAGACTCTGT TAGGCTTTTG @CGATAATGG CICCCTGGCC TG@CTCTTCC GTGTTCTTGA GCCGICTGTT TGCGAQGGTT GTGGGCATTC 16801 16901 ATTCTAGAAT COTTICTCTG TGGTTGGTAA ATCTCCACTT CAATTTTTGC ATTATGTAAC TCATGTCCAT TATAACGCTC GACAATCTTT TCCAAAACTT 17001 CTOGGETCTIC ANALTEGTAN ATACATGTAC GATECTETIT ATTEFCAMA ATTITEGAM ATAITGEGA GECAMACTET TIANTGATAT CATEGAGEGT 17101 STASTCAGAA ACGTCTAGTG GGATTTTGGA TATTCTAATT CTCTTAGGCA AAGGAGGCTC TTCCTCTCT CTATAAAGGC GATCTTTCGA CCTTGATGCA 17201 SCATCCGAAG SCSCAAATCC CASTCTAGGT CCTAATTCAT TECTTAAATC CCTTCTGCGS TATCTCGTTA CCTTATGATT AGATGASCTG TCTGTGTGAGA 17301 TATTICCAAT AATTICGTCA AATGCTTTGT CAATGCAAA ATGCAGTATT GAATCGTCTA ATGGAAAAGG TACTAGACGC TTATCGTTTC ACTTTTTAG 17401 TETTTECAT CTATTATCA GECEGACTE ANAATAACCE TATEGEAGTE AACCETTTEC ATTERCGAT GATAAFTEET TGECGACTEA GACAGAAAAT 17501 GACCETGATA GTITETCATA ATTITCAATA GEGCAGAAAG AATTITTAAAG ATTATTECT ATCIAGACAT TATETETTTT ATATGATTEC TETAAAAGTA 17601 TTAGGAGAGA TOGTOGAAAA TATCOTTAAA COTTGGCACG TTOCCATAGO TOCTTTTGAT GTTTGCAAGC CAAGTTACGG AGTTGCCAAA TTGTCTTAAT 17701 GTAGGTTCAA CAGTAGCCAA ATTTCOGTAA GCTACGGTTA GCCTATATGC GGCTTCTTCG CATTCTTGAT ACTCTTCCAA TGGCCCATAC TTGGTGTTGA 17801 TEGOSTCASC TACGATAGGC AAAAGTTCTA AATCAGAGTT COCTTCGTG ACGAGAGGTG AATACTTGAA AATTACTGTG GAAACTGCAA TAGCCAAATT

Figure 2. (Continued).

17901	CTGTGACTGT	TTAGCAGATG	CCTGAGAAAA	TTCAGTATCT	ATAGTCTCAA	AAATTGATTT	ATAAACTTGA	TTTGATTCCA	GTAATTTCAC	ACCCCAGITT		
18001	TCATTATTGA	AGCAATTTAC	CANANTACGC	ACAGTTAACA	TOGTTAGAGT	GATATICITA	TTGCCAAGC	CTTCTTCGAT	GTAATCTTTT	ATATCTGAAG		
18101	AATAAGGTAA	TTTTTTACA	ATCAATCGTA	CTATGTOGTA	TOCAGGAGIC	TTAATTTCCC	AGTTAGATCT	TATAGIGITA	gcaaatgaaa	GTAGCAACTC		
18201	CCAGCTTTCA	TCTATGTCAT	GTAGTOCGCC	ACCAATTIGC	GOCAGAATTT	CATCATCAAA	GGTCTTCTCA	TTTGAATTTA	TTTTTACTAT	TCCATTGAAA		
18301	ATAGTATCTG	GGTTATAATT	TTCCATGATA	AGATACIGCT	TGACAGGAAG	AACCTTCATT	ACGCTCGTTT	TTG AA GGTGA	TACTGCAGAA	GATGATCCAT		
18401	TATCGTTGGG	CTGGTCTAGC	GATATGCCAT	TAGTGTTCTT	CAGTATGAAC	TGTACTACTT	GATOCCTATA	ACTCATTGGC	AGTTCATAAC	GAGCCAAAAA		
18501	ATTGTCGGCT	GCTGTATATG	GGTTGTCGCT	GACATTAATC	GCAATTICA	AGGCAGGTTT	TCCATCTTCA	ATATOCACAT	CAAATACGTA	GTCATAOGTT		
18601	TTACCCTOGA	OR I <u>ATTC</u> TATTTT	TITGICCTIT	CAGTAGCAC	CTGCACCGAC	AACATCACCA	ACCITTTTCC	ATGAAGAATT	tgaaaattga	TGAGCTTCTA		15
18701	TAGTACCTTG	AGGTGATTTG	ACTACCACAA	TTTGACCCTC	TITGCGCCCC	GGCGATTGTA	алаттсата	AGGTGATAGT	TTAGACTCAT	CANATTCTAT		E
18801	TGTCTTGGAA	CTTATAGTGG	ATTTTTCCAC	TTGCGTAGAT	AGITCATITA	TTICATCCTC	GGAAGCCCAT	CTTGACTTCT	CTTGCGAAAA	TATTCTCACA		
18901	AGATTATCAC	TACTTCCCAC	AATAATATCA	CATTAGACA	TACAATCAAC	GGACCAAATA	GAGATGGCAG	GAAGTGTAAT	TACCTOCTTT	AACGAACCAT		
19001	TTTCCTTTGA	CCATATACGG	ACTGTTCGGT	CTTCGCCGCA	gcttacaata	TCTCCATTOG	GCAGCAGTTT	аатссаатаа	ACAAAGCTTT	CATGGCCTTC		
19101	ATATGTTCTT	AGTACGTCAC	CTGTATGCAT	ATCAACCAAC	TTGATGAGGC	CATCGTTTGA	ACAACTGATA	ANATOSCOAT	CATCCACCAC	AGCTARATGT		
19201	CTTACTACAT	CATTGTGAAT	CCCACTGAAA	GTTTTGATCA	CTTTATCGTT	TTOCCACAGT	TTAATCGTCT	TATCOCCGA	AGCTGTCAAG	AATTTATTT		
19301	CTGAGAAGGA	TACGACTTTG	GCATCCCATA	CAGAGGCATT	GIGTGCTIGC	AAATTGTAAA	CCAACGATCC	TTCTTTCCAA	ACCTTTGCGG	TITTGTCCCA		
19401	ACTACCGCTT	ATCACAACAC	CATCTTGGAA	ggataaacta	CAAACGTTCC	CTIGGTGGCC	AATCAAGGTA	TACAAGGGAT	CCTCTCCTGA	GGTGGCAAAC		
19501	AAGGGCACAC	CGTTGATCAT	AGTATCCTTA	остесалала	GCAACAACTC	CTICTCGGAA	TCATRACACA	састатттаа	AAATCCTTGT	CCTGTATAAA		
19601	CTACCGTACC	GAGCCATTGG	TCATCTTTAG	ACCACAAGCG	AACTGTTCCA	TCOCTCGAAA	CACTAGCAAC	CTTTGAATCA	TCCACAGCTA	CACATOCT		
19701	GACGTCCTGA	TCGTGCCCTT	TAAGTGTTGC	ACTCAATTGA	TATCCCATAC	TCCAAATCTG	CTGCTCTACA	CCTTACTATC	асасатбаат	ATATATATAT	ł	
19801	AAAATAAGCC	AAGACAGTGG	CCTTCCCTTA	TTATCAGCGT	ACTAAAATCT	CATATGATT	ATTITTCGTG	GTCCTGAACG	AGTGTGAAAA	ATTTTGAAAA		
19901	ACTOGCAAAG	GAAATGCCAA	GGTCAGCAAC	TTTCTCGACG	TOCAAGCOTC	GCICCTATIG	TTTTAGTAAC	CTGTAGACCT	TCAATATTCA	AGAGCTTGCA		
20001	CACGCACGCA	TTGCTGCACT	ACTGCTTACC	CACACTACAA	ATTTAATTCC	AGCITATIGA	AGTIGAAAAG	GCAAQGGAAA	AATACCACAG	GTTTAGATAA		
20101	ggaaatagga	GAAAQGATTA	GAAACCATAT	CCTATAACA	GTAACGATAA	TATTTATATA	CACGTATATT	ттстоятста	GATATGACAG	GTCCAATAGT		
20201	GTACGTTCAA	AATGCGGACG	GTATCTTCTT	CAAGCTTGCT	GAGGGCAAAG	GAACTAACGA	TGCIGTTATT	CACTTGGCCA	ATCAAGATCA	AGGTGTTCGG		
20301	GTCCTTGGAG	CAGAGGAATT	TCCTGTGCAA	getcaagtag	TANAGATIGC	GTCCTTGATG	GGGTTCATTA	AGCTAAAGTT	GAACAGGTAT	GCCATTATCG		
20401	CAAATACTGT	GGAAGAGACC	GGTAGATTCA	ATGGCCACGT	TTTCTATAGA	GTGTTGCAAC	ATTCTATCGT	ATCTACCAAG	TTTAACTCGA	GAATCGATTC		
20501	TGAAGAAGCC	GAATATATCA	AGCTACTGGA	GTTGCATTTG	AAAAATTCCA	CCTTTTATTT	TTCATACACA	TATGATTTAA	CAAATTCCTT	ACAAAGAAAT		
20601	GAAAAGGTTG	GTCCTGCAGC	CTCCTGGAAA	ACCGCTGATG	AACGATTCTT	TTOGAACCAT	TACITAACTG	AAGATTTGAG	AAACTTTGCT	CATCAAGATC		
20701	CTAGAATTGA	CTCCTTTATA	CAACCTGTTA	ICTATGOGTA	TOCCAAGACA	GTOGACGCOG	TTTTGAATGC	CACCOCTATC	GTTCTTGGTT	TGATTACCAG		
20801	ACGTAGTATA	TTTAGGCGG	GCACAAGATA	CTTCCGTCGT	GGTGTTGACA	AAGACGGTAA	CGTTGGCAAT	TTCAATGAAA	CTGAGCAAAT	TTTACTOGCT		
20901	GAGAATCCAG	AGAGTGAAAA	AATACACGTT	TTCTCCTTCT	TACAGACAAG	AGGATCTGTG	CCAATATACT	GGGCTGAAAT	CAACAACTTG	AAGTACAAGC		
21001	CAAATCTTGT	TCTTGGAGAA	AACTCATTAG	атессалсала	AAAGCATTTT	GACCAGCAAA	AGGAGTTATA	TGGCGACAAC	TACTTGGTTA	ACCTAGICAA		53
21101	CCAAAAGGGC	CACGAACTAC	CCGTGAAAGA	OGGCTATGAA	TCAGTCGTGC	ACCCGCTAAA	CGATCCGAAG	ATTCACTACG	TGTATTITGA	CTTCCACCAT		B62
21201	GAATGTCGTA	AGATGCAATG	GCATAGAGTG	AAATTGTTAA	TIGATCACCT	gg a gaaatta	GGTTTATCTA	ACGAAGATTT	CTTCCACAAG	GTCATAGACT		
21301	CTAATGGTAA	CACCGTTGAA	ATTGTTAATG	AGCAACATTC	CGTTGTAAGA	ACAAACTGTA	TGGATTGTTT	GGACAGAACA	AATGTOGTTC	AATCTGTTTT		
21401	AGCCCAGTGG	GTTTTGCAAA	AGGAGTTTGA	AAGTGCCGAT	GTCGTTGCTA	CTOGAAGCAC	TTGGGAAGAC	AACGCTCCAT	TGTTAACTTC	TTACCAAAAC		
21501	TTATGGGCTG	ATAATGCAGA	TGCAGTTAGT	GTGGCATATT	CGGGCACTGG	AGCTTTGAAG	ACCONTTCA	CAAGAACCGG	TAAGCGTACA	CGTCTAGGTG		
21601	CATTCAACGA	EcoR TTTTTT <u>GAAT</u>	I I <u>TC</u> AGCATCAC	GTTATTACCA	GAACAATTGG	ACTGATGGTC	CAAGACAGGA	TTCATACGAT	TTATTOCTTG	GTGGATTTAG		
21701	ACCACATACC	GCTTCTATCA	AGTCGCCATT	CCCGACAGA	AGACCAGTGT	ATATTCAGCT	GATCCCAATG	ATCATTTGCG	CAGCCTTGAC	CGTTTTAGGT		
21801	GCTACGATAT	TTTTTCCTAA	AGATAGATTT	ACCAGCAGTA	AGAATTTOCT	GTATTTTGCA	GGTQCGTCGA	TTGTCTTGGC	GCTTTCAACC	AAATTCATGT		
21901	TTAAGAACGG	TATTCAGTTT	GTCAACTGGC	CTAAGTTAGT	AGACGTTOGG	TTCTTGGTCG	TTCATCAAAC	GCATGATAAA	GAACAACAAT	TCAAAGGTTT		
22001	GAAATATGCA	CAGAGTCCTA	AATTTTCCAA	OCCGGATCCT	TTAAAAAGAG	ATTAATGAGT	TCTATATGTG	ATAAAATGAT	GATTATIGTA	AATCCAAAAC	¥	1
22101	GTAAAACCGA	AAATATGATA	алаассадаа	AAAAAGTAAG	ттааааата	 AAAAAATAAA	AAAATGAACG	CCGGCCCAG	AGGGAAGAAC	AAGCTATTTA		
22201	TTTTCTTTTT	TACTITATCT	TTCATGGTCA	TATATATATA	TATATATATA	TATATATATC	TTCIATAGGC	TACTATACAT	GAAGGOGTTC	GCCCTTACAA		



Figure 2. (Continued).

D484

22301	CGTTAAGTTT	TTTTCCCTAT	TGTACCGTGT	AAGTTTCAAT	TATTCGCATA	ATTCATGAAT	GAACTTCTTC	ACATOGGTTG	ATTTCATTAG	GCTTCACCC	1
22401	ACTAAAAATC	CATGGACACC	TTCTTTTTTG	TATTTTCAG	CATCGTCCCT	GGTGGTAATT	CCCCATAGAG	CAATTAGAAG	AACATCCTTT	GGAATAGATT	
22501	CTACCAAGTT	ACTGGTGGTA	TTTAGGTCTA	ogttgaatga	ATGCAGGTCC	CTATTATTGA	CACCTACAAC	TTTAGCACCA	ATTTCTAGAG	CCCTTTGTAA	
22601	TTCCTCTTTG	GAGTTCACCT	CAACGAGAGG	TTCCATGTTC	AAATCTTIAC	TGTAGCTGTA	CAGITCCTTC	AATAAGGGTT	GAGATAGCAT	CTTGACTATA	
22701	AGAAGGACAG	TGTCAGCTCC	AGCTAATCTT	GCTTCTAGTA	TTTGATACTT	GCTGAAAATA	AATICTTTTC	тсаааасаса	AGGCCTCTCC	TTGGGAGGAA	
22801	ATTICAAATC	TAGGATTTTC	CTCACATTTA	CTAAATOCTG	TAACGAACCG	TGAAACCAAT	GACCTTCGGT	CAATACGGAA	ATTGCGGATG	CACCAGOCTC	
22901	TGCGTATTTG	AGAGCCTGTT	CAGCAGCAAC	AGCTITIAAA	CAAATGGGTC	CCTTCGATGG	AGAGGCACGC	TTGACTTCAG	CAAGAACAAC	GGCTCTTTTA	
23001	TGGGATGATG	ACAACACCGT	GTAGAAATCC	IGTAACGGTG	GOGCAAGACC	TAAATCATAG	TTAGATTGTA	AGTCTTGAAA	GGTGAAACCT	GGGATTTTAG	
23101	ACTGCTCATT	GACGTCTATT	TTACGCCGAG	CATAGATACG	GTCCAAAATA	GAATTTGAAG	GAGATGATTT	GTTTTCCTCC	CAAGTOCCTC	CACTGACGTT	
23201	TAAAATGTTC	CTGAICATCA	GATGACCTTC	CTCGGTCAAG	ATGGATTCCG	GAIGAAATIG	CACACCTTCT	ACAGIGTACT	TCTTGTGTCT	TACACCCATG	
23301	ATAATTCCAT	TTTCOGTACT	CGCAGTAACC	TTCAAGCAGG	ATCGTAGCCA	CGATTCTGTC	CCGGCCAATG	AATGGTATCT	TGTCACAGCA	ATACCTIGCG	
23401	GCACGTTCTT	GAAAATTCCA	CAGTTGTCGT	GAGAGATIGG	GGACGTTTTA	CCGTGGACAA	TCTCACCAGC	GTAGGCAACT	TCACCACCAA	ATACGTCAAA	İ
23501	CATGCATTGC	TGGCCCATAC	AGATTCCAAA	TACAGGAATT	TTCCCAGTAA	AGTACCGGAT	ACAGICICTT	GAAATGCCAG	AATCTGFCTT	TGGGTGGCCT	
23601	GGTCCAGGCG	Адатаадсаа	TGTGTCGGGA	TTCAAGGCGG	CAATTTCTGG	AACIGTAATT	GCATCGTTAC	GGTAGACGCT	CACTTTGGQG	CCTCCIGGC	
23701	ACAAGTACTC	GTAAACGTTC	CAGGTAAAGG	AATCGTAGTT	GTCAATTAGA	ACCACATGCT	TATIGATIGG	GTTTGTTGCA	GCGTGCACAG	ACATTTTTTT	
23801	CTTACTTTTA	TGGCGTTCTA	TGTAGAACCA	AACAGATITG	GTAATGTGTT	AATACTGACC	AGCTCCTTGT	CGAACITIT	GCGTACCTTA	АТАТСТОСТА	
23901	TAGATTAGAA	TGAGTCAACG	AGCCATTGAT	GATCATICAA	TGATTACTAC	AAGAGGCATT	TTCOSTCGTC	AGTGATATGA	ATAGCACGGT	GAGGTCCAGA	
24001	AGAGATAGTA	GGAAGTTTAA	TACTGGCTCA	AAGGCATACT	TATATCGACA	AGCTAAGCGC	АЛАДАСТАСТ	TTGTOGTACA	TCATCAGCAT	TGCTTTOGAG	
24101	AAAACAGCAT	TTAAAATCAG	CCATTIGCTG	OGCTTTCTTT	AAGCCCTAAC	AGAAAATAGT	TCAMAAAGGC	CTATAAGAAG	CAAAAAAACT	GAAAAATACG	
24201	алалаатаст	CCAGOGGAGG	AGTATAAAAG	GGAGGAAAAT	ACAAAATGTA	GTTGTGTTGT	TACOCATCTC	аладаадаа	GAAGAAGAAG	AAGAAGAGGA	
24301	CGTCGCACTT	GCAAGCAAAA	ACAGCGACTG	OCGCGAGAAC	AACAAACACT	AATGCATGTC	TETGTCCAGE	AGGAAAAGCG	AGCCTTAGTG	TTTCTTATTT	
24401	TCTOGTTCGC	CTCACTAGTT	TGTAGTAGGG	TGGGAGAAAC	GATGGAGTGG	ATIGTATGAT	GAGGAACCAG	ATAGOGACAA	CAGATICTCA	AGTGACGAGG	
24501	AACATCTTTT	AAAGCCCAGT	TTTTAGTAGA	GCTTAGGGCG	CCTTTACTGA	CTATCCACTC	CTTCTTAAGC	Sau3A I TAT <u>GATC</u>			

Figure 2. (Continued).

Figure 2. Complete sequence of the 24 577 bases of chromosome XI. The sequence reads 5' to 3' from the left telomere to the centromere. *Eco*RI sites are underlined. ORFs are boxed. The direction of each ORF is shown by arrow.

and the amino acid sequences were performed to the GenBank, EMBL, SWISS-Prot and NBRF libraries using the GCG package software by us at the I.M.B.B. MicroVAX and by the staff at MIPS.

RESULTS AND DISCUSSION

Sequence determination

The 24.6 kb sequence was determined from overlapping ExoIII-produced deletions and from internal priming to fill in the gaps. An average length of 280 nucleotides was read manually from each sequencing reaction. Readings up to 400 bases were achieved on 4% polyacrylamide gels. (Selected sequences were determined using an A.L.F. sequencer (Pharmacia)). Compressions seen at several specific positions were solved by repeating the sequencing reactions using dITP (6 different instances). Occasional base ambiguities were resolved by new preparations of DNA templates and from opposite strand readings. Sequence assembly was performed manually according to the restriction map and to the sequences obtained from oligonucleotide primers connecting the restriction fragments. Verifications were performed manually by careful re-reading of original sequences and deciding between differences found on the two strands.

Sequence analysis

Six phase ORF map analysis of the sequence included within the 24 577 bases by the DNA Strider program revealed nine ORFs >100 codons (Figure 1b). Their sizes range from 203 to 1286 codons and constitute 60.2% of the sequence (14 792 bases). This percentage agrees with the organization found on chromosome III (Oliver *et al.*, 1992).

ORF	Homologous or Identical protein	Optimized score	Highest score	Reference
A407	S. cerevisiae Hyptothetical protein YCR007C (239aa)	152	2300	Aigle <i>et al.</i> , 1992 EMBL: X59720
D326	20.0% identity in 140aa <i>E. coli (tdc)</i> Threonine dehydratase (329aa) 28.0% identity in 214aa	578	1517	Data <i>et al.</i> , 1987 GB: X14430
A616	<i>E. coli (proP</i> <fv;1) proline/betaine transporter (500aa)</fv;1) 	235	3263	Culham <i>et al.</i> , 1993 EMBL: M83089
A314	22.0% identity in 354aa S. cerevisiae (URA1) Dihydroorotate oxidase	1519	1519	Roy, 1992 GB: M83295
F1286	S. cerevisiae Hypothetical protein (URA1 3' region) (283aa)	1386	6147	Roy, 1992 EMBL: X59371
	Pseudomonas sp. Hydantoinases (hyuA-hyuB) (690+592aa) 24.4% identity in 1190aa	783	6147	Watabe <i>et al.</i> , 1992 GB: D10494
E203 F715	no homology found Mouse PLAP: Phospholipase A ₂ -activating protein (325aa)	509	973 3427	Clark <i>et al.</i> , 1991 GB: M57958
B623	S. cerevisiae SACI (RSDI) protein 100.0% identity in 623aa	3109	3109	Cleves et al., 1989
D484	S. cerevisiae Anthranilate synthase (TRP3) 99.8% identity in 484aa	2336	2338	Zalkin <i>et al.</i> , 1984 EMBL: K01386

Table 1. Best optimized FastA scores obtained by the comparison of the putative translation product of each ORF with the protein databases.

The complete sequence of the 24 577 bases of cosmid pEKG100 is given in Figure 2. FastA analysis (Pearson and Lipman, 1988) of this sequence revealed that three fragments were previously sequenced including the genes URA1 (2884 bases), RSD1 (2406 bases), TRP3 (2815 bases), and part of the 3' non coding region of the UBA1 gene (359/4795 bases). The database files of the last three genes are partially overlapping. Our sequence data are in complete agreement with the URA1 published sequence. There are minor differences in the non-coding regions of the other sequences and one nucleotide substitution in the coding region of TRP3 changing the arginine residue 130 to lysine, which is a conservative

change. This region of our sequence determination has been verified independently by another group participating in the sequencing of chromosome XI. Therefore the discrepancies found with the previously published data could be due to strain polymorphisms.

Our sequence analysis also showed differences with the published genetic map as well as with the physical map of chromosome XI (Mortimer *et al.*, 1989). Genes URA1, SAC1 (RSD1) and TRP3 were placed 105–115 kb from the left telomere and in reverse order. This distance from the telomere according to the sequences of cosmid pEKG100 and pUKG040 is 25–38 kb (Alexandraki and Tzermia, 1994). Leaving aside

A407a	50	VLPQDLFMDNFTWMFYEFFKCFTFRTWLLLLLLMWLPGFLSQIKSINRIFPFKLC
А407Ь	244	QLPKKTYRYKFTWVLKRIFNLWLFPAFILFLACIYVSWDKGHLFRILC
YCR007C	32	TLPEDTFKSYMTYLLYEMAHYKPMIFS-FLALSVSILIVVIFHNVKACDVVFGFS
		** . * *
A407a		ILVSCLVGIFLPNIYSFSHKSVLTNQLT~QFSKEIVEHAPGTDTHDWETVAANLNSYF
A407b		C-GGGFLLMVRVFQNMRPFSMHMEDKMQFLSTII-NEQESGANGWDEIAKKMNRYL
YCR007C		IFVTSILFLSTLIPFNVYISDEGFRIKLLLEVITHRPAVKGKEWRAITDNMNOYL
		** . * ***.
A407a		YENKAWNTEYFFFNAAECQKAFRKVLLEPFSVKK 195
А407Ь		FEKKVWTSEEFFFDGIDCEWFFNHFFYRLLSTKK 389
YCR007C		LDNGLWSTRYYFYSSERCYKFFR-FLVKEKPPGV 173
		** . * *



Figure 3. Analysis of the A407 ORF product. (a) Alignment of the indicated amino acid regions in A407 and YCR007C proteins using the CLUSTAL program. Asterisks indicate identities and dots indicate conservative substitutions. (b) Hydrophobicity profile (Kyte and Doolittle, 1982) of A407 protein derived using the DNA Strider program.

human errors, these discrepancies could reflect stain variabilities or alternatively, a variation in the frequency of recombination across this region.

Analysis of the ORF products

The putative translation products of the identified ORFs have been compared to protein databases using FastA (Table 1). Scores higher than 200 have been considered as significant, although in some instances lower scores due to homologies in restricted areas of the protein sequence indicated conservation of specific domains. For better evaluation of each score's significance we have also included the highest FastA score, obtained by the comparison of each ORF with itself. Similarities were found for all but one of the ORFs contained in the sequenced 24 577 bases, either with known yeast proteins or with proteins from other organisms. Protein patterns (motifs) have been identified by the ProSite program (Bairoch, 1991) of the GCG package.

All ORFs correspond to expressed genes, evidenced by DNA hybridization analysis, using polyadenylated RNA and radioactively labelled single stranded oligonucleotide probes designed according to the sequence and the direction of transcription of each hypothetical gene (data not shown). We have additionally performed gene disruption/deletion analyses for two of the identified ORFs, the D326 and the F1286. Below we describe some interesting findings on the ORF sequences.

Dot matrix analysis of the A407 product revealed an internal region of about 145 residues which has been duplicated and diverged (data not shown). This duplicated area shares sequence similarities with one region in two other yeast

IlvA ILV1 D326	MADSQPLS-GAPE MSATLLKQPLC-TVVRQGKQSKVSGLNLLRLKAHLHRQHLSPSLIKLHSELKLDELQTDN MIVPTYGDVLDASNRIKEYVNK
Tue	*
Il vA ILV1 D326 Tdc	GAEYLRAVLRAPVYEAAQVTPLQKMEKLSSRLDNVILVKREDRQPVHSFKLRGAYAMMAG TPDYVRLVLRSSVYDVINESPISQGVGLSSRLNTNVILKREDLLPVFSFKLRGAYNMIAK TPVLTSRMLNDRLGAQIYFKGENFQRVGAFKFRGAMNAVSK TGMPRSNYFSERCKGEIFLKFENMQRTGSFKIRGAFNKLSS
IlvA ILV1 D326 Tdc	LTEEQKAHGVITASAGNHAQGVAFSSARLGVKALIVMPTATADIKVDAVRGFGGEVLLHG LDDSQRNQGVIACSAGNHAQGVAFAAKHLKIPATIVMPVCTPSIKYQNVSRLGSQVVLYG LSDEKRSKGVIAFSSGNHAQAIALSAKLLNVPATIVMPEDAPALKVAATAGYGAHIIRYN LTDAEKRKGVVACSAGNHAQGVSLSCAMLGIDGKVVMPKGAPKSKVAATCDYSAEVVLHG * ** *
IlvA ILV1 D326 Tdc	ANFDEAKAKAIELSQQQGFTWVPPFDHPMVIAGQGTLALELLQQDAHLDRVFVPVGG NDFDEAKAECAKLAEERGLTNIPPFDHPYVIAGQGTVAMEILRQVRTANKIGAVFVPVGG RYTEDREQIGRQLAAEHGFALIPPYDHPDVIAGQGTSAKELLEEVGQLDALFVPLGG DNFNDTIAKVSEIVEMEGRIFIPPYDDPKVIAGQGTIGLEIMEDLYDVDNVIVPIGG * .**.** ******* * . *
IlvA ILV1 D326 Tdc	GGLAAGVAVLIKQLMPQIKVIAVEAEDSACLKAALDAGHPVDLPRVGLFAEGVAVKRIGD GGLIAGIGAYLKRVAPHIKTIGVETYDAATLHNSLQRNQRTPLPVVGTFADGTSVRMIGE GGLLSGSALAARSLSPGCKIFGVEPEAGNDGQQSFRSGSIVHINTPKTIADGAQTQHLGE GGLIAGIAVAIKSINPTIRVIGVQSENVHGMAASFHSGEITTHRTTGTLADGCDVSRPGN *** .* *
IlvA ILV1 D326 Tdc	ETFRLCQEYLDDIITVDSDAICAAMKDLFEDVRAVAEPSGALALAGMKKYIALHNIRGER ETFRVAQQVVDEVVLVNTDEICAAVKDIFEDTRSIVEPSGALSVAGMKKYISTVHPEIDH YTFAIIRENVDDILTVSDQELVKCMHFLAERMKVVVEPTACLGFAGALLKKEEL LTYEIVRELVDDIVLVSEDEIRNSMIALIQRNKVVTEGAGALACAALLSGKLDQYI ******
IlvA ILV1 D326 Tdc	LAHILSGANVNFHGLRYVSERCELGEQREALLAVTIPEEKGSFLKFCQLLGGRSVT TKNTYVPILSGANMNFDRLRFVSERAVLGEGKEVFMLVTLPDVPGAFKKMQKIIHPRSVT VGKKVGIILSGGNVDMKRYATLISGKEDGPTI QNRKTVSIISGGNIDLSR-VSQITGFVDA
IlvA ILV1 D326 Tdc	EFNYRFADAKNACIFVGVRLSRGLEERKEILQMLNDGGYSVVDLSDDEMAKL EFSYRYNEHRHESSSEVPKAYIYTSFSVVDREKEIKQVMQQLNALGFEAVDISDNELAKS
IlvA ILV1 D326 Tdc	HVRYMVGGRPSHPLQERLYSFEFPESPGALLRFLNTLGTYWNISLFHYRSHGTDYGRVLA HGRYLVGGASKVP-NERIISFEFPERPGALTRFLGGLSDSWNLTLFHYRNHGADIGKVLA
IlvA ILV1 D326 Tdc	AFELGDHEP-DFETRLNELGYDCHDETNNPAFRFFLAG GISVPPRENLTFQKFLEDLGYTYHDETDNTVYQKFLKY

Figure 4. CLUSTAL alignment of the entire sequences of the anabolic threonine dehydratases ILV1 and IlvA with those of the catabolic threonine dehydratase Tdc and the D326 protein.

Table 2. Pairwise similarity scores of threonine dehydratase sequences from yeast (D326 and ILV1) and *E. coli* (Tdc and IlvA) using the PileUp and FastA (GCG) programs

Compared protein sequences	PileUP Scores	FastA Scores
ILV1 × IlvA	0.88	1251
$D326 \times Tdc$	0.78	578
$IlvA \times Tdc$	0.76	558
ILV1 × Tdc	0.71	519
$D326 \times IlvA$	0.68	501
$D326 \times ILV1$	0.67	468

hypothetical proteins YCR007C and YCR048W, of unknown function, found on chromosome III.

A multiple alignment of both homologous A407 regions and of the similar area in YCR007C protein is shown in Figure 3a. The YCR048W hypothetical protein (Grivell *et al.*, and Bolotin-Fukuhara *et al.*, 1992, EMBL: X59720) of 610 amino acids showed a lower degree of similarity to A407 (FastA score: 128). The hydrophobicity profile of A407 ORF showed that the duplicated area consists of a stretch of hydrophobic amino acids followed by a hydrophilic domain (Figure 3b). Its conservation in other proteins implies some specific structural or functional property possibly with a dual role in A407 protein.

The gene encoding for the D326 protein is not essential for viability, based on our gene disruption-deletion analysis. D326 ORF product showed extensive similarities to all known prokaryotic and eukaryotic threonine dehydratases



Figure 5. Hydrophobicity profiles of the A616 ORF product and of the ProP protein.

COMPLETE SEQUENCING OF A 24.6 Kb SEGMENT OF CHROMOSOME XI

F1286 HVUA-HVUB	MQKGNIRIAIDKGGTFTDCVGNIGTGKQEHDTVIKLLSVDPKNYPDA <u>PLEGIRRLL</u> EVLE
nyun nyub	** ****** * *
	↑HyuA
F1286	HKTIPRGIPLDISNVRSLRMGTTLATNCALERNGERCAFITTKGFKDSLLIGDQTRPDIF
НуиА-НуиВ	RQYIDRTAIDHVFHGTTIATNAILEYDGAKTGMITTEGYRDIIHIGRHQRPQNY
	* * ***.*** ** .* *** ** . ** . ** .
F1286	NLNIKKVVPLYD-TVVEIDERVTLEDFSEDPYFTKSSPNEQEGILEGNSGEMVRVIKKPD
HyuA-HyuB	SIMQEIPWQDRPLVQRRHRLAIAERMGPVKGQVITPVQ
	.** * .** *. *. *. *. *.
F1286	ESSVRSILKVLYASGIKSIAIAFLHSYTFPDHE-RIVGNIAREIGFSHVSLSSEVSPMIK
НуиА-НуиВ	EDQVRGAVATLKERGVDSIIVNFLFSYTNPEHEQRVKEIIEEEYPEAFVTISSEVSPQFR
	* ** * *. ** . ** *** *. * * * . * .
F1286	FLPRAHSSVADAYLTPVIKKYLNSISAGLSHAE-DTHIOFMOSDGGLVDGGKFSGLKS
HyuA-HyuB	EFERFTTASINGFVGPKVKNYIQNLEQSLKDSGISAELHIMCSNGGVATPKTVSEKPVNT
	· * ·· ···. * ·* <i>·</i> *··· ·* ·· ···• *.**. * ···
F1286	ILSGPAGGVIGYSSTCYDKNNNIPLIGFDMGGTSTDVSRYGDGRLEHVFETVTAG
HyuA-HyuB	LLSGPAAGILGGAWAG-ELTNRQKLITFDVGGTSADIGIITDSGYGESSARDTWIAG
	.****.** * ** **.***.* ** **
F1286	IIIQSPQLDIHTVAAGGSSILSWKNG-LFRVGPDSAAADPGPAAYRKGG-PLTITDANLF
HyuA-HyuB	YPVMVPMIDIHTIGAGGGSIAHIDEGGAFKVGPRSAGSRPGPACYGHGGLKPTVSDANVV
	. * .*******.** .* *.*** ** **** * ** *
F1286	LGRLVPEFFPKIFGPNEDESLDLETTTLKFRELTDVINKDLNSNLTMEEVAYGFIKVA
НуиА-НуиВ	LGRIDEHNFLGGEMKIYTNAAYKVIDELAGQLDLSRERTAEGVLQIM
	**** •****
F1286	NECMARPVRAITEAKGHVVSQHRLVSFGGAGGQHAIAVADSLGIDTVLIHRYSSILSAYG
HyuA-HyuB	NNNMANAIREKTIQKGEDPREFSLVAFGGAGPLHAVEVAQILNIPEVIIPLYPGINSATG
	*. *** * ** **.**** **. **. * * *.* *.* *.*
F1286	IFLADVIEENQEPCSFILGEPETILKVKKRFLELSKNSIKNLLSQSFSREDIVLERYLNL
НуиА-НуиВ	LLTTDLKYDVIKT-EFMMSTNMDFSGLNEDLAGLETQLINQLKEDGVSKQDIRILRSADC
	* * * . ** *** .
F1286	RYEGTETSLMI~~LQKYDDQWNFREWFSEAHKKEFGFSFDDKRIIIDDIRIRAIG-K
HyuA-HyuB	RYAGQGYELRVDLPDVFLDEETIVDALNNFHESHKAEYGHNFTDSPIEFVNIRVTGTGYM
	** * * * * * *.** *.* * * * *** *
F1286	SGVRKEKTVDEQLIEISHFKKADVSKDASFTQKAYFDNKWVDTAVFKIDDLPAGTIIEGP
HyuA-HyuB	PKIEKQAIHHEYQLEDALLKTGDATFNIDGSLVKVEINFYQREKIPVGAEFNGP
	· · *· ·* ·* · * · * * * · · · * · · · · * *· · · *
F1286	AILADGTQTNIILPNSQATILN-SHIFIKINQKAAKTLSKSGYELD-IDPILLSIFSHRF
НуиА-НуиВ	CIVLQKDTTTVIPPNCTAYIEEYGNMIIKVGVMSKIHTDLKKIDPITVQVVLGSL
	* * .* ** * ***** .* ****
	HyuB
F1286	MDIALQMGTQLRKTSVSTNVKERLDFSCALFDSKGNLVANAPH-VPVHLGSMSTCISAQA
НуиА-НуиВ	ENVAVEMGHKLARMSYSSIIRESEDFGCALVDVRGQQLCESSHSTPLQSGPIPGYIKGIR
	** .* . * ** **.*** * .** * * .
F1286	KLWEGKLKPGDVLITNHPDIGGTHLPDITVITPSFSSTGELIFYVASRAHHADIGGI
НуиА-НуиВ	EIMEDRNDTFNQGDVIMHNSPYHGASHGPDVGFCIPVFYK-DELIGFSVTTAHHLDIGSS
	· *·· ·· ***·· * * *··* * * * · *** · · ******

Figure 6.

F1286	LPGSVP-PNSKELYEEGTAIYSELVVKEGIFQEELIYKLFVEDPGKYPGCSGSRRFSDNI
НуиА-НуиВ	TPGSCGIVDAVDAYAEGLQFKAIKVYDQGV-KNRYVWDILKDNIRAPKLVV
F1286	SDLKAQVAANTKGIQLIGSLTKEYDLATILKYMAAIQTNASESIKKMLAKMVE-HFGTTK
НуиА-НуиВ	GDMEAQIAAARIGAQRYIEIIEKYGLDTVQAASEELMNYSEKMMRDAIKKLPDGEYTAEG
	.*. **.** * * * . *.* * *
F1286	FSGEDRLDDGSLIKLQVIIRPEKEEYIFNFDGTSPQVYGN-LNAPE-AITNSAILY
HyuA~HyuB	FL-DGYLDSDDPAKKDLRINVTVKVDGSDLTVDLTGTSPQVTDKPINMPLLGTVDIAIYL
	* *** ******* * **
F1286	CLRCLVGEDIPLNQGCLKPLTIKIPAGSLLSPRSGAAVVGGNVLTSQRVTDVILK
HyuA-HyuB	TLRSILLDSTVYGNFPQNSGLIRPIKIVAPKGTLCNPIFPAPTIA-RFNSGNAVADTLMK
	** * * * . * * *. * *
F1286	TFNVMADSQGDCNNFTFGTGGNSGNKTDKQIKGFGYYETICGGSGAGADSWRGSGWNGSD
НуиА~НуиВ	ALAQVVPHQVSAGVGNLQVVAFSGQSNENYWVYMDIMEGSYGGRYGKDGMD
	*
F1286	AVHTNMTNTRMTDTEVFERRYPVLLKEFSIRRGSGGKGKYTGGNGVVRDVQFRKAVTASI
НуиА-НуиВ	AVDTLYANTRNNPIEDIESHYPLRVNRYELRDNDSAPGKWRGGIGSIREVSFLADGSFSV
	.* .* * .* .*** ** ** * .*.* * . *.
F1286	LSERRVIGPHGIKGGQDGSRGENLWVRHS-TGALINVGGKNTINTI
HyuA-HyuB	EADGHKYAPWGFDDGQDGYVG-SLSIRDNETNELVQLPSKLPNRHAQSGSTIQLVGPCGG
	··· · * *. ·**** * ·* ·* · * * ··· * * ·* * ·* *
F1286	-YAQPGDRKD
HyuA-HyuB	GYGNPLEREPEKVLSDYLDGFITKEKALVEYGVTITDSEEIDYEKTNELRKV
	**.* ** ** •
	Figure 6. (Continued).

Figure 6. CLUSTAL alignment of the F1286 ORF sequence with the two hydantoinases, HyuA and HyuB. The mitochondrial energy transfer protein motif is underlined.

as well as some similarities to serine dehydratases. Multiple sequence alignment analysis revealed that it is most probably the yeast biodegrative threonine dehydratase (Figure 4 and data not shown). Our conclusion was based on the following observations summarized in Table 2. D326 was more similar to the E. coli tdc gene product, which catalyzes the catabolic dehydration of L-threonine to α -ketobutyrate and ammonia, than to the ILV1 yeast threonine dehydratase, which catalyzes the first step in the isoleucine biosynthetic pathway (Kielland-Brandt et al., 1984, PIR1: DWBYT, 36.2% identity in 287 overlapping amino acids). ILV1, on the other hand, appeared more homologous to the E. coli biosynthetic IlvA threonine dehydratase (Lawther et al., 1987, PIR1: DWECTS, 47.8% identity in 517 overlapping amino acids). The corresponding similarity of the D326 product with the IlvA threonine dehydratase is 35.5% identity in 318 overlapping amino acids. In addition to their homologies, the two catabolic enzymes are similar in size (326 and 329 amino acids respectively) and quite different from the two anabolic enzymes (576 and 514 amino acids). Finally, the *CHA1* gene product reported to be responsible for the catabolism of both L-serine and L-threonine (Bornaes *et al.*, 1992) was very clearly grouped with the serine dehydratases in our multiple alignment analysis (not shown).

The product of ORF A616 is quite possibly a membrane metabolite transporter. It is significantly similar to the prokaryotic ProP osmoregulatory proline/betaine transporter and less similar to a number of proteins from various species, as permeases and drug resistance proteins (FastA scores: 100–154). The region of homology, residues 180 to 520 of A616 and 70 to 415 of ProP, coincides with the region of ProP which is homologous to the citrate and α -ketoglutarate transporters (Culham *et al.*, 1993). Finally, a comparison of the hydrophobicity profiles of the two proteins indicated extensive topological similarities (Figure 5). They both contain the characteristic twelve potentially membrane spanning domains and both

A F715 PLAP	1 MGYQLSATLK <u>GH</u> DQDVRDVVAVDDSKVASVSRDGTVRLWSKDDQWLGTVVYTGQGFLNSV 1 MHYMSGHSNFVSYVCIIPSSDIYP
F715 PLAP	CYDSEKELLLFGGKDTMINGVPLFATSGEDPLYTLI <u>GHQ</u> GNVCSLSFQD-GVVISGSW <u>D</u> K HGLIATGGNDHNICIFSLDSPMPLYILK <u>GH</u> KDTVCSLSSGKFGTLLSGSW <u>D</u> T *. **.* · · · ** * **. ***** * · · *****
F715 PLAP	TAKV <u>W</u> KEGSLVYNLO <u>AH</u> NASVWDAKVVSFSENKFLTASADKTIKLWQNDKVIKTFSGIHN TAKVWLNDKCMMTL <u>OGH</u> TAAVWAVKILP-EQGLMLTGSADKTIKLWKAGRCERTFLG-HE ***** **.* *.** * **.********
F715 PLAP	DVVRHLAVVDDGHFISCSNDGLIKLVDMHTGDVLRTYE <u>GH</u> ESFVYCIKLLPNG-DIVSCG DCVRGLAILSETEFLSCAN <u>D</u> ASIRR <u>WQ</u> I-TGECLEVYF <u>GH</u> TNYIYSISVFPNSKDFVTTA * ** ***.*.**.**. ***. * * *** ***. ***.
F715 PLAP	EDRTVRIWSKENGSLKQVITLPAISIWSVDCMSNGDIIVGSSDNLVRIFSQEKSRWASED EDRSLRIWKHGECAQTIRLPAQSIWCCCVLENGDIVVGASDGIIRVFTESEERTASAE ****** .* * * *** *** . ****.***.*** ** .
F715 PLAP	EINELSTQVEKSTISSKT 316 EIKASLSRES 261 ** **

B

F715 A33928	1 1	MGYQLSATLK <u>GH</u> DQDVRDVVAVDDSKVASVSR <u>D</u> GTVRL <u>W</u> SKDDQWLGTVVYTGQGF MTEQMTLRGTLK <u>GH</u> NGWVTQIATTPQFPDMILSASR <u>D</u> KTIIM <u>W</u> KLT * * * * ****** * · · · · · * *** * · *
F715 A33928		LNSVCYDSEKELLLFGGKDTMINGVPLFATSGEDPLYTLI <u>GHO</u> GNVCSLSFQDGVVISGS RDETNYGIPQRALR <u>GH</u> SHFVSDVVISSDGQF <u>ALSGS</u> *******
F715 A33928		WDKTAKVWKEGSLVYNLQAHNASVWDAKVVSFSENKFLTASADKTIKLWQNDKVIK WDGTLRLWDLTTGTTTRRFVGHTKDVLSVAFSSDNR <u>OIVSGSRDKTIKLWNT</u> LGVCK ** ** ** * ** *.***. **
F715 A33928		TFSGIHNDVVRHLAVVDDGHFISCSNDGLIKLVDMHTGDVLRTYE <u>GH</u> ESFVYCIKLLPNG -YTVQDVQDE <u>SH</u> SEWVSCVRFSPNS *.* *.* *.*.
F715 A33928		DIVSCGEDRTVRIWSKENGSLKQVITLPAISIWSVDCMSNGDIIVGSSDNLVRIFSQ SNPI <u>IVSCGWDKLVKVWNL</u> ANCKLK ***** *. **. * **
F715 A33928		EKSRWASEDEINELSTQVEKSTISSKTIEFDESKLSPYEILQSPGRKEGQIVVVKS 345 TNHI <u>GH</u> TGYLNTVTVSPDGSL <u>CASGCKDGOAMLWDL</u> 220

Figure 7. (a) Alignment of the F715 ORF sequence with the mammalian phospholipase A_2 -activating protein (PLAP). The underlined residues indicate the **GH** (19–23N)**D**(5N)**W** repeat. (b) Alignment of the F715 ORF sequence with the chicken GTP binding protein β chain homologue (A33928). The repeated β transducin motif for the β subunit of G proteins is underlined.

have two extended hydrophilic domains, one loop at the centre of the molecule and one at the carboxyl terminus where it is predicted to form an α -helical coiled coil (Secondary Structure

Prediction Suite (PREDICT) of the CCP4 package).

Part of the gene sequence of the ORF F1286 was previously known as neighbouring the 3' region of

the URA1 gene. It is a gene not essential for life based on our gene disruption/deletion analysis. The F1286 product showed a significant similarity to hydantoinases. The hydantoinases HyuA and HyuB are involved in the conversion of D- and L-5-substituted hydantoins to the corresponding N-carbamyl-D- and N-carbamyl-L-amino acids respectively. The hyuA and hyuB genes have been isolated from a native plasmid of Pseudomonas sp. strain NS671 along with three more enzymes all of which are responsible for the asymmetric production of L-amino acids from the corresponding racemic 5-substituted hydantoins (Watabe et al., 1992). Both HyuA- and HyuB-like proteins, appear to be represented in yeast in a single ORF, as HyuA is similar to the amino end half of F1286 (29.1% identity in 619 overlapping amino acids, FastA score: 559) and HyuB to the remaining carboxy end half (24.9% identity in 566 overlapping amino acids, FastA score: 383) (Figure 6). Therefore the F1286 product may be a bifunctional enzyme, which is not unprecedented in yeast (Donahue et al., 1982). The resemblance of the veast and bacterial molecules was also clearly seen by examining their hydrophobicity profiles and the distribution of acidic and basic amino acids (DNA Strider program, data not shown). The F1286 ORF contains a rare motif starting on residue 48, not present in HyuA sequence, which characterizes mitochondrial energy transfer proteins (P-x-[DE]-x-[LIVAT]-[RK]-x-[LR]-[LIVMFY]). We are currently testing its significance.

No homologous sequences or motifs were found for the product of ORF E203. Its hydrophobicity profile indicated a very hydrophilic protein which probably exists in cells since we have detected the corresponding RNA by blot-hybridization analysis (data not shown).

The F715 ORF product showed a significant similarity to the mouse protein PLAP. This protein activates phospolipase A_2 in specific inflammatory disease processes and results in the release of active oxygenated eicosanoids. The observed homology involved the entire length of PLAP spanning only to about 300 residues of the amino terminus of F715. (We have not found any potential frame-shifts in either F715 or PLAP DNA sequences.) This difference may indicate a multiple role for the F715 protein in yeast (Figure 7a). F715 product also showed regional similarity to the chicken GTP binding protein β chain homologue (Guillemot *et al.*, 1989, PIR2: A33928) (FastA score: 148) as well as to a number of β chain homologous sequences

from various species including yeast (FastA scores: 100-140). This similarity is localized at the same amino terminal area as that with the PLAP protein and it is mainly at positions which contain a non perfect β -transducin motif, also called Trp-Asp motif (Duronio et al., 1992). (Consensus pattern: [LIVMSAC]-[LIVMFYWSTAGC]-[LIMSTAG]-[LIVMSTAGC]-x(2)-[DN]-x(2)-[LIVMWSTAC]x-[LIVMFSTAG]-W-[DEN]-[LIVMFSTAGC]) (Figure 7b). The sequence similarity is also extended to the GH dipeptide that precedes the central D residue by 19-22 residues as recently described by Peitsch et al. (1993). This motif exists in several copies in a number of proteins not all of which are associated to the plasma membrane but they could potentially participate in the transmission of signals. The protein F715 may be similarly involved in a signal transduction pathway.

ACKNOWLEDGEMENTS

We thank Bernard Dujon for the excellent coordination of the chromosome XI sequencing project. We thank Martina Haasemann, Irmi Becker and all MIPS staff for help with the sequence analysis. We also thank Yannis Papanikolaou for help with the computer analyses at IMBB, and Alekos Athanasiadis and Manolis Pittarokilis for introducing us to the inhouse computer system. We thank Georgia Houlaki for help with the preparation of the figures. We thank Morten Kielland-Brandt for communicating information on the ILVI gene. Finally, we thank the co-contractor on this project George Thireos for fruitful discussions. This work was supported by the Commission of the European Communities under the BRIDGE program of the Division of Biotechnology and by the Greek Ministry of Industry, Energy and Technology.

REFERENCES

- Alexandraki, D. and Tzermia, M. (1994). Sequencing of a 13.2 kb segment next to the left telomere of yeast chromosome XI revealed five open reading frames and recent recombination events with the right arms of chromosomes III and V. Yeast 10, S81–S92.
- Asubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (Eds) (1987). Current Protocols in Molecular Biology. Greene Publishing Associates and Wiley-Interscience.
- Bairoch, A. (1991). A dictionary of sites and patterns in proteins. Nucl. Acid Res. 16, 2241-2245.

COMPLETE SEQUENCING OF A 24.6 Kb SEGMENT OF CHROMOSOME XI

- Bornaes, C., Petersen, J. G. L. and Holmberg, S. (1992). Serine and threonine catabolism in *Saccharomyces cerevisiae*: The CHA1 polypeptide is homologous with other serine and threonine dehydratases. *Genetics* 131, 531–539.
- Clark, M. A., Özgür, L. E., Conway, T. M., Dispoto, J., Crooke, S. T. and Bomalaski, J. S. (1991). Cloning of a phospholipase A₂-activating protein. *Proc. Natl. Acad. Sci. USA* 88, 5418–5422.
- Cleves, A. E., Novick, P. J. and Bankaitis, V. A. (1989). Mutations in the SAC1 gene suppress defects in yeast Golgi and yeast actin function. J. Cell. Biol. 109, 2939–2950.
- Culham, D. E., Lasby, B., Marangoni, A. G., Milner, J. L., Steer, B. A., van Nues, R. W. and Wood, J. M. (1993). Isolation and sequencing of *Escherchia coli* gene *proP*<fv;1 reveals unusual structural features of the osmoregulatory proline/betaine transporter. ProP. J. Mol. Biol. 229, 268-276.
- Datta, P., Goss, T. J., Omnaas, J. R. and Patil, R. V. (1987). Covalent structure of biodegrative threonine dehydratase of *Escherchia coli*: homology with other dehydratases. *Proc. Natl. Acad. Sci. USA* 84, 393–397.
- Donahue, T. F., Farabaugh, P. J. and Fink, G. R. (1982). The nucleotide sequence of the *HIS4* region of yeast. *Gene* 18, 47–59.
- Duronio, R. J., Gordon, J. and Boguski, M. S. (1992). Comparative analysis of the β transducin family with identification of several new members including *PWP1*, a nonessential gene of *Saccharomyces cerevi*siae that is divergently transcribed from *NMT1*. *Proteins* 13, 41–56.
- Guillemot, F., Billault, A. and Auffray, C. (1989). Physical linkage of a guanine nucleotide-binding protein-related gene to the chicken major histocompatibility complex. *Proc. Natl. Acad. Sci. USA* 86, 4594–4598.
- Higgins, D. G. and Sharp, P. M. (1988). Clustal: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**, 237–244.
- Kielland-Brandt, M. C., Holmberg, S., Petersen, J. G. L. and Nilssdon-Tillgren, T. (1984). Nucleotide sequence of the gene for threonine deaminase (*ILV1*) of Saccharomyces cerevisiase. Carlsberg Res. Commun. 49, 567-575.

- Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydrophobic character of a protein. J. Mol. Biol. 157, 105-132.
- Lawther, R. P., Wek, R. C., Lopes, J. M., Pereira, R., Taillon, B. E. and Hatfield, G. W. (1987). The complete nucleotide sequence of the *ilvGMEDA* operon of *Escherchia coli*: K-12. *Nucl. Acid Res.* 15, 2137– 2155.
- MacGrath, J. P., Jentsch, S. and Varshavsky, A. (1991). UBA1: an essential yeast gene encoding ubituitinactivating enzyme. Embo J. 10, 227–236.
- Marck, C. (1988). 'DNA-Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computer. *Nucl. Acid Res.* 16, 829–1836.
- Mortimer, R. K., Schild, D., Contopoulou, C. R. and Kans, J. A. (1989). Genetic map of *Saccharomyces cerevisiae*, Edition 10. *Yeast* **5**, 321-403.
- Oliver, S. G. et al. (1992). The complete DNA sequence of yeast chromosome III. Nature 357, 38–46.
- Pearson, V. R. and Lipman, D. J. (1988). Improved tools for biological sequence analysis. Proc. Natl. Acad. Sci. USA 85, 2444–2448.
- Peitsch, M. C., Borner, C. and Tschopp, J. (1993). Sequence similarity of Phospholipase A2 activating protein and the G protein β-subunits: a new concept of effector protein activation in signal transduction? *Trends Biochem.* 18, 292–293.
- Rothstein, R. J. (1983). One-step gene disruption in yeast. *Methods in Enzymology* vol. 101, 202-211.
- Roy, A. (1992). Nucleotide sequence of the URA1 gene of Saccharomyces cerevisiae. Gene 118, 149–150.
- Watabe, K., Ishikawa, T., Mukohara, Y. and Nakamura, H. (1992). Cloning and sequencing of the genes involved in the conversion of 5-substituted hydantoins to the corresponding L-amino acids from the native plasmid of *Pseudomonas sp.* strain NS671. J. Bacteriol. 174, 962–969.
- Zalkin, H., Paluh, J. L., van Cleeput, M., Moye, W. A. and Yanofsky, C. (1984). Nucleotide sequence of *Saccharomyces cerevisiae* genes *TRP2* and *TRP3* encoding bifunctional anthranilate synthase: indole-3-glycerolphosphate synthase. *J. Biol. Chem.* **259**, 3985–3992.