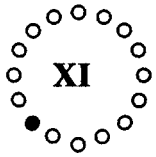
YEAST VOL. 10: 663–679 (1994)



Yeast Sequencing Reports

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 A_2-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* (*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′

*Author to whom correspondence should be addressed.

CCC 0749–503X/94/050663–17

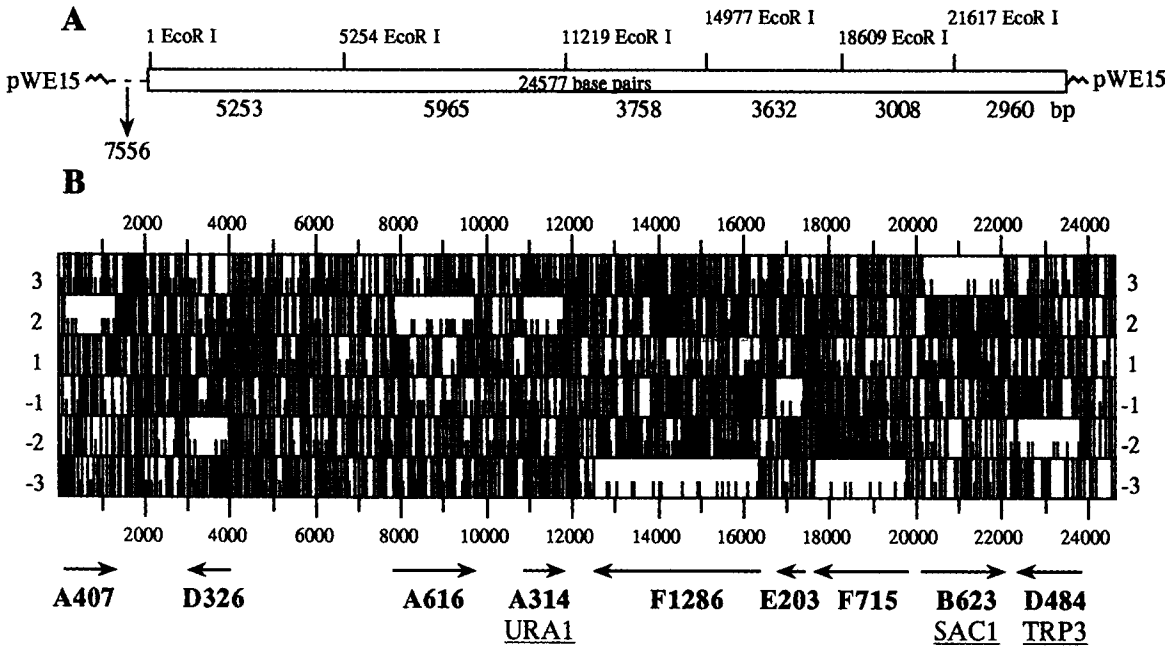


Figure 1. (a) *Eco*RI restriction map of the 24 577 base pairs of cosmid pEKG100. The remaining 7556 bases 5′ of the yeast sequence in that cosmid are presented separately (Alexandraki and Tzermia, 1994). The 3′ end of the insert is a *Sau*3AI site. The numbers below the bar indicate the size of each *Eco*RI 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*^{Q}*lacZ*ΔM15)) 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 DH5α (*supE44* Δ*lac* U169 (φ80*lacZ*ΔM15) *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 [^{35}S]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 ExoIII-mung 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 *Eco*RI 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

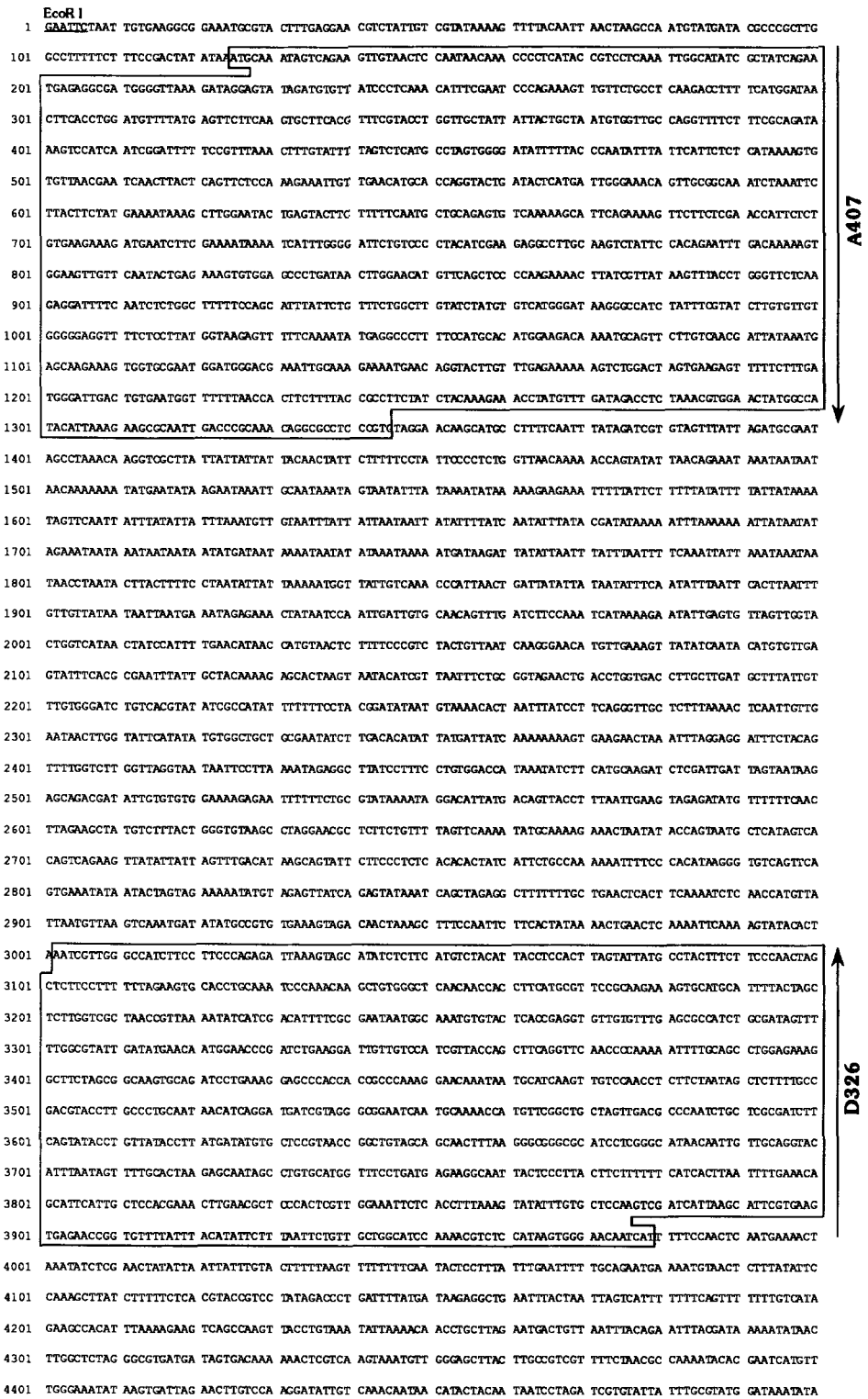


Figure 2.

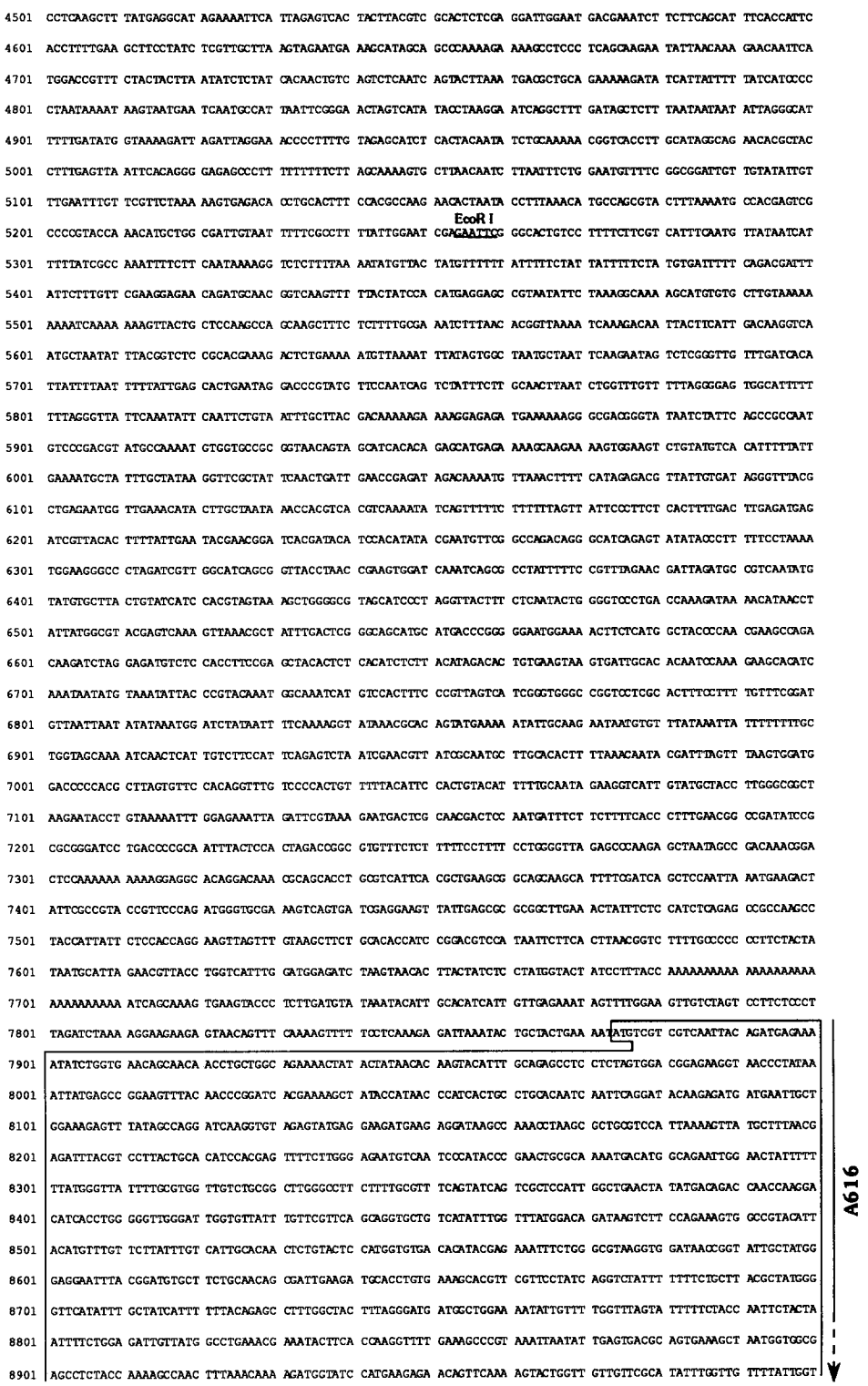


Figure 2. (*Continued*).

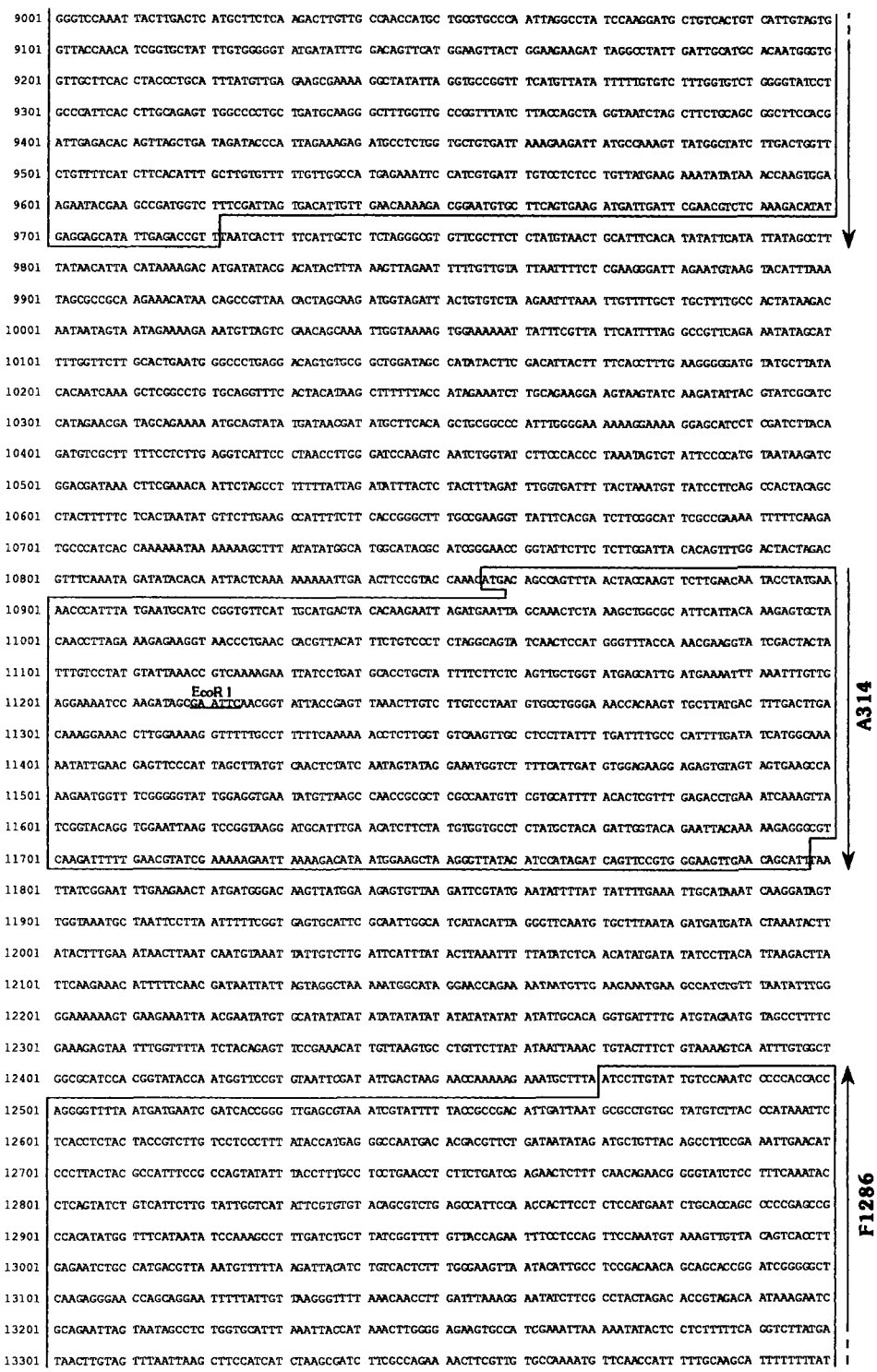


Figure 2. (*Continued*).

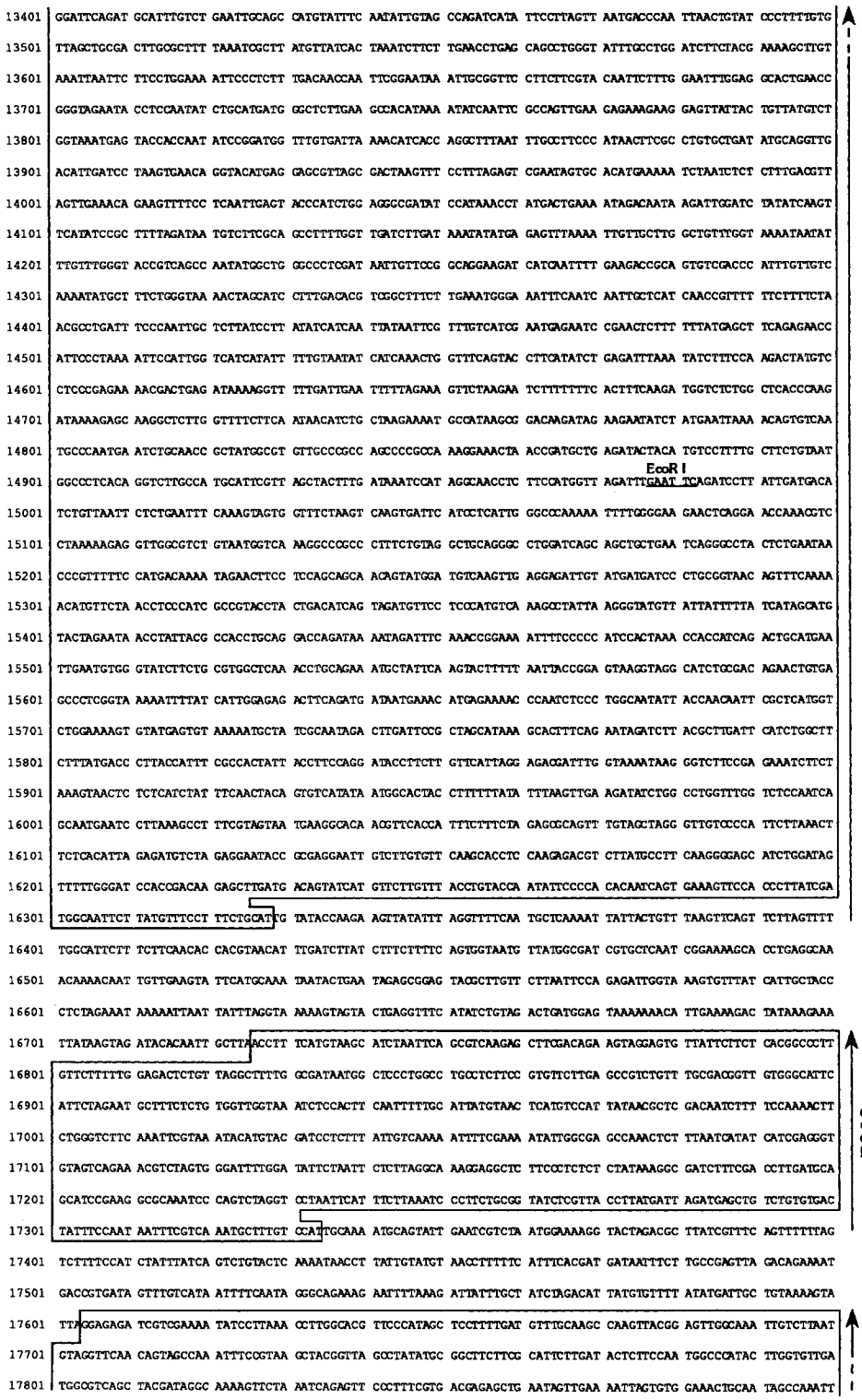


Figure 2. (*Continued*).

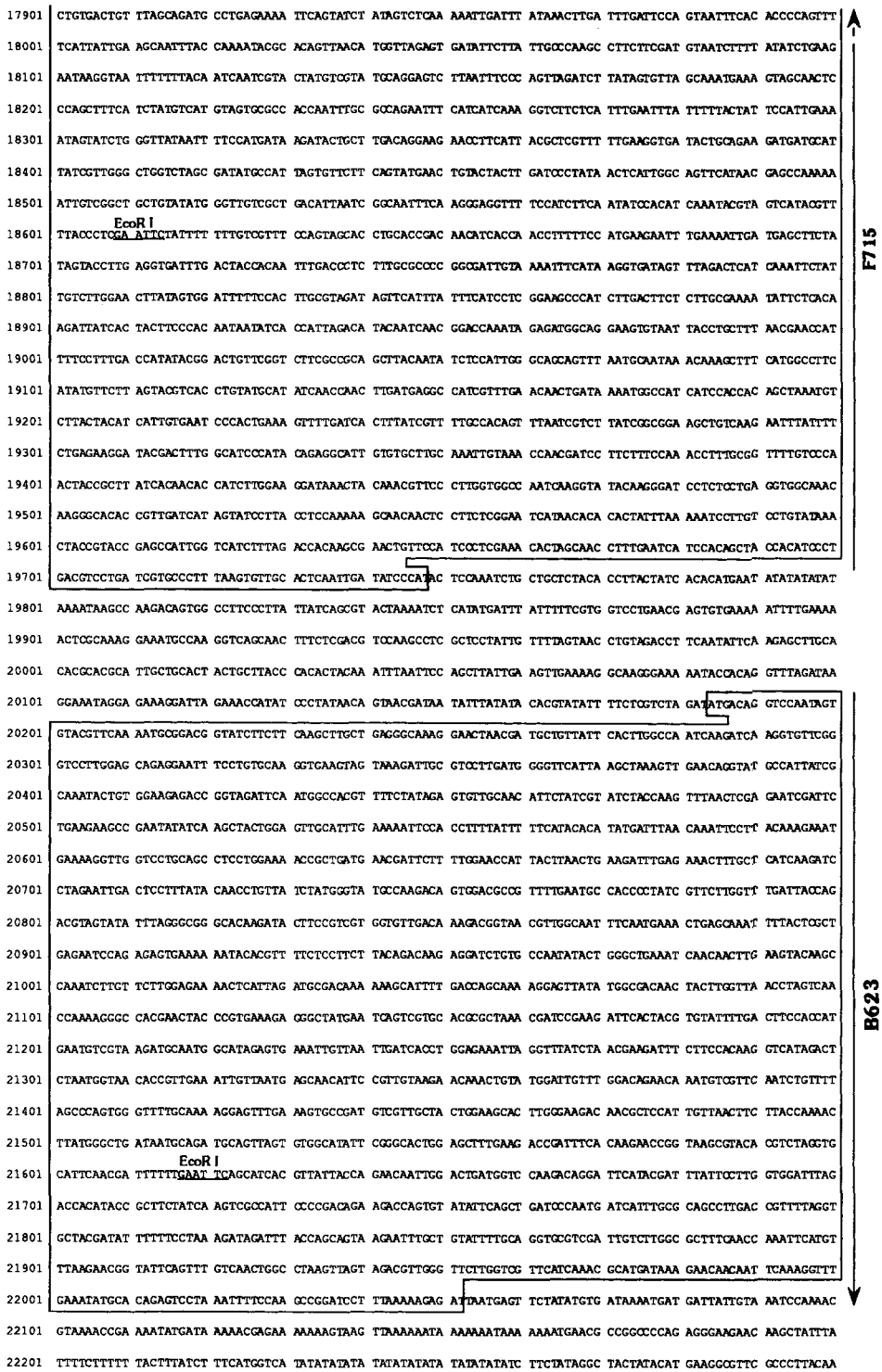


Figure 2. (*Continued*).

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).

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

ORF	Homologous or Identical protein	Optimized score	Highest score	Reference
A407	*S. cerevisiae* Hyptothetical protein YCR007C (239aa) 20·0% identity in 140aa	152	2300	Aigle *et al.*, 1992 EMBL: X59720
D326	*E. coli* (*tdc*) Threonine dehydratase (329aa) 38·9% identity in 314aa	578	1517	Data *et al.*, 1987 GB: X14430
A616	*E. coli* (*proP*<fv;1) proline/betaine transporter (500aa) 22·0% identity in 354aa	235	3263	Culham *et al.*, 1993 EMBL: M83089
A314	*S. cerevisiae* (*URA1*) Dihydroorotate oxidase 100% identity in 314aa	1519	1519	Roy, 1992 GB: M83295
F1286	*S. cerevisiae* Hypothetical protein (*URA1* 3′ region) (283aa) 100·0% identity in 283aa	1386	6147	Roy, 1992 EMBL: X59371
	Pseudomonas sp. Hydantoinases (*hyuA-hyuB*) (690+592aa) 24·4% identity in 1190aa	783	6147	Watabe *et al.*, 1992 GB: D10494
E203	no homology found		973	
F715	Mouse PLAP: Phospholipase A_2-activating protein (325aa) 40·8% identity in 250aa	509	3427	Clark *et al.*, 1991 GB: M57958
B623	*S. cerevisiae* *SAC1* (*RSD1*) 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 *et al.*, 1984 EMBL: K01386

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 VLPQDLFMDNFTWMFYEF--FKCFTFRTWLLLL---LLMWLPGFLSQIKSINRIFPFKLC
A407b      244 QLPKKTYRYKFTWVLKRI--FNLWLFPAFILFLACIYVSWDKGHLFRI----------LC
YCR007C     32 TLPEDTFKSYMTYLLYEMAHYKPMIFS-FLALSVSILIV---VIFHNVKACDVVFGF--S
                **   .   * ..    ..  *  . .    .     . .

A407a          ILVSCLVGIFLPNIYSFSHKSVLTNQLT---QFSKEIVEHAPGTDTHDWETVAANLNSYF
A407b          ----C-GGGFLLMVRVFQNMRPFSMHMEDKMQFLSTII-NEQESGANGWDEIAKKMNRYL
YCR007C        IFVTSI--LFLSTLIPFN---VYISDEGFRIKLLLEVITHRPAVKGKEWRAITDNMNQYL
                       **  .  *       .    ..  .. .       *  .. ..* *.

A407a          YENKAWNTEYFFFNAAECQKAFRKVLLEPFSVKK 195
A407b          FEKKVWTSEEFFFDGIDCEWFFNHFFYRLLSTKK 389
YCR007C        LDNGLWSTRYYFYSSERCYKFFR-FLVKEKPPGV 173
                .. * . .*. . *  *  .   .
```

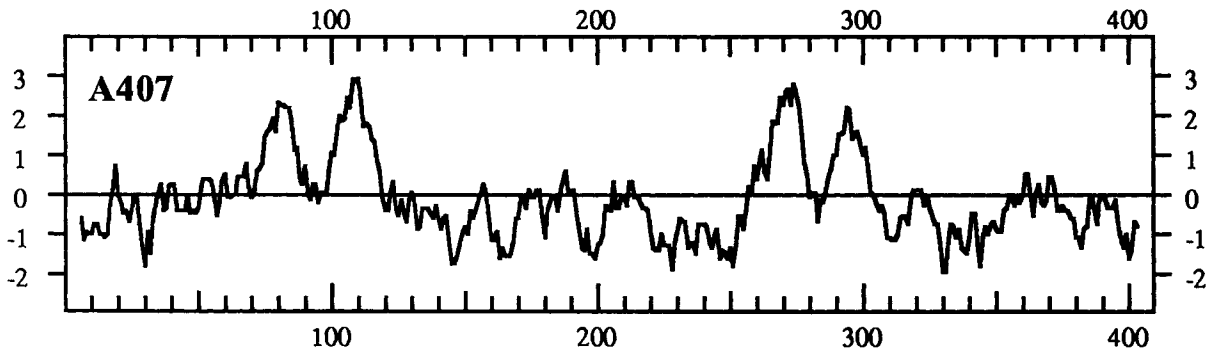


Figure 3. Analysis of the A407 ORF product. (a) Alignment of the indicated amino acid regions in A407 and YCR007C proteins using the CLUSTAL program. Asterisks indicate identities and dots indicate conservative substitutions. (b) Hydrophobicity profile (Kyte and Doolittle, 1982) of A407 protein derived 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        MADS---QPLS-G-------------------------------------------APE
ILV1        MSATLLKQPLC-TVVRQGKQSKVSGLNLLRLKAHLHRQHLSPSLIKLHSELKLDELQTDN
D326        MIVPTYG------DVLDASN---------RIKEYVNK-----------------------
Tdc         MHI-TYDLPVAIDDIIEAKQ---------RLAGRIYK-----------------------
            *

IlvA        GAEYLRAVLRAPVYEAAQVTPLQKMEKLSSRLDNVILVKREDRQPVHSFKLRGAYAMMAG
ILV1        TPDYVRLVLRSSVYDVINESPISQGVGLSSRLNTNVILKREDLLPVFSFKLRGAYNMIAK
D326        -------------------TPVLTSRMLNDRLGAQIYFKGENFQRVGAFKFRGAMNAVSK
Tdc         -------------------TGMPRSNYFSERCKGEIFLKFENMQRTGSFKIRGAFNKLSS
                              . .     .. *    .  * *.     .**.***   ..

IlvA        LTEEQKAHGVITASAGNHAQGVAFSSARLGVKALIVMPTATADIKVDAVRGFGGEVLLHG
ILV1        LDDSQRNQGVIACSAGNHAQGVAFAAKHLKIPATIVMPVCTPSIKYQNVSRLGSQVVLYG
D326        LSDEKRSKGVIAFSSGNHAQAIALSAKLLNVPATIVMPEDAPALKVAATAGYGAHIIRYN
Tdc         LTDAEKRKGVVACSAGNHAQGVSLSCAMLGIDGKVVMPKGAPKSKVAATCDYSAEVVLHG
            * .  .  **.. *.*****.....   * . . .***  ..  *       .....

IlvA        ANFDEAKAKAIELSQQQGFTWVPPFDHPMVIAGQGTLALELLQQ---DAHLDRVFVPVGG
ILV1        NDFDEAKAECAKLAEERGLTNIPPFDHPYVIAGQGTVAMEILRQVRTANKIGAVFVPVGG
D326        RYTEDREQIGRQLAAEHGFALIPPYDHPDVIAGQGTSAKELLEEV---GQLDALFVPLGG
Tdc         DNFNDTIAKVSEIVEMEGRIFIPPYDDPKVIAGQGTIGLEIMEDL---YDVDNVIVPIGG
               ..       .    *   .**.*.* ******* . *.. .      .. ..**.**

IlvA        GGLAAGVAVLIKQLMPQIKVIAVEAEDSACLKAALDAGHPVDLPRVGLFAEGVAVKRIGD
ILV1        GGLIAGIGAYLKRVAPHIKTIGVETYDAATLHNSLQRNQRTPLPVVGTFADGTSVRMIGE
D326        GGLLSGSALAARSLSPGCKIFGVEPEAGNDGQQSFRSGSIVHINTPKTIADGAQTQHLGE
Tdc         GGLIAGIAVAIKSINPTIRVIGVQSENVHGMAASFHSGEITTHRTTGTLADGCDVSRPGN
            *** .* .   . . *  . ..*.         ..             .*.*      *.

IlvA        ETFRLCQEYLDDIITVDSDAICAAMKDLFEDVRAVAEPSGALALAGMKKYIALHNIRGER
ILV1        ETFRVAQQVVDEVVLVNTDEICAAVKDIFEDTRSIVEPSGALSVAGMKKYISTVHPEIDH
D326        YTFAIIRENVDDILTVSDQELVKCMHFLAERMKVVVEPTACLGFAGA----LLKKEE--L
Tdc         LTYEIVRELVDDIVLVSEDEIRNSMIALIQRNKVVTEGAGALACAAL----LSGKLDQYI
             *. . .. .*... *  . .   .  . . .  . . * .. *. *.

IlvA        LAH----ILSGANVNFHGLRYVSERCELGEQREALLAVTIPEEKGSFLKFCQLLGGRSVT
ILV1        TKNTYVPILSGANMNFDRLRFVSERAVLGEGKEVFMLVTLPDVPGAFKKMQKIIHPRSVT
D326        VGKKVGIILSGGNVDMK--RYATLISGKEDGPTI---------------------------
Tdc         QNRKTVSIISGGNIDLS--R-VSQITGFVDA-----------------------------
                   *.**.*..    *  .      .

IlvA        EFNYRFADAKN--------ACIFVGVRLSRGLEERKEILQMLNDGGYSVVDLSDDEMAKL
ILV1        EFSYRYNEHRHESSSEVPKAYIYTSFSVVDREKEIKQVMQQLNALGFEAVDISDNELAKS
D326        ------------------------------------------------------------
Tdc         ------------------------------------------------------------


IlvA        HVRYMVGGRPSHPLQERLYSFEFPESPGALLRFLNTLGTYWNISLFHYRSHGTDYGRVLA
ILV1        HGRYLVGGASKVP-NERIISFEFPERPGALTRFLGGLSDSWNLTLFHYRNHGADIGKVLA
D326        ------------------------------------------------------------
Tdc         ------------------------------------------------------------


IlvA        AFELGDHEP-DFETRLNELGYDCHDETNNPAFRFFLAG
ILV1        GISVPPRENLTFQKFLEDLGYTYHDETDNTVYQKFLKY
D326        ---------------------------------------
Tdc         ---------------------------------------
```

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 × Tdc	0·78	578
IlvA × Tdc	0·76	558
ILV1 × Tdc	0·71	519
D326 × IlvA	0·68	501
D326 × 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

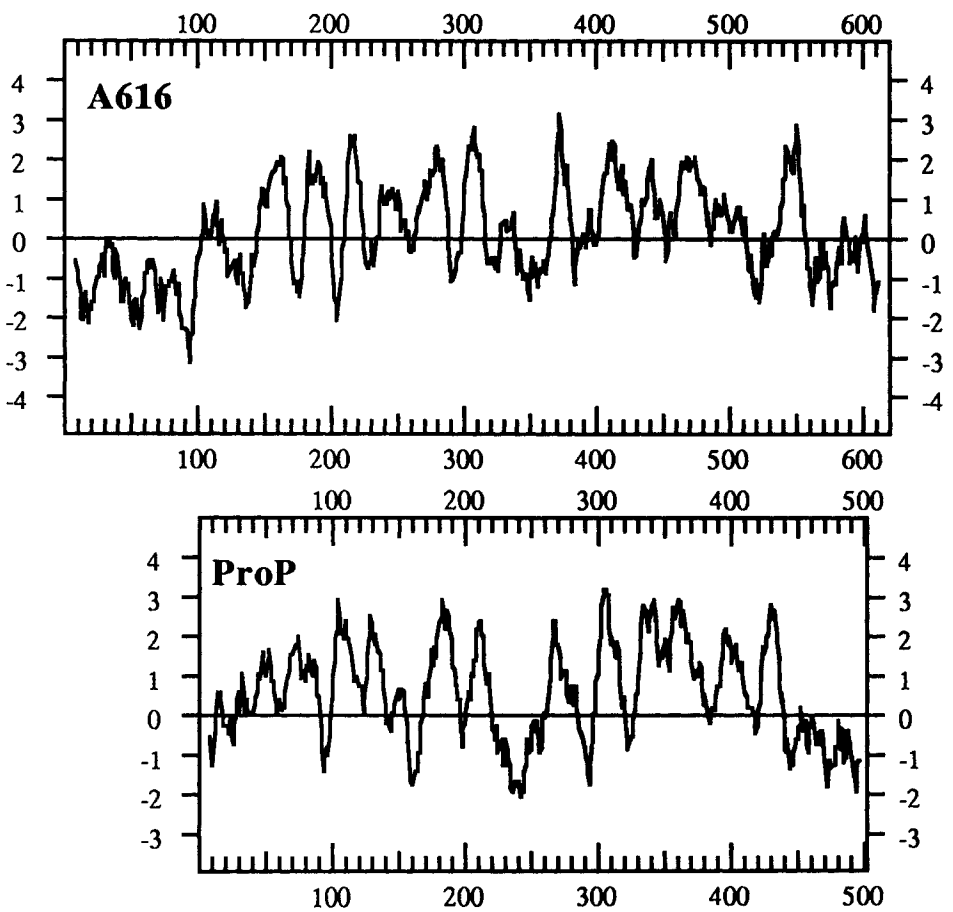


Figure 5. Hydrophobicity profiles of the A616 ORF product and of the ProP protein.

```
F1286          MQKGNIRIAIDKGGTFTDCVGNIGTGKQEHDTVIKLLSVDPKNYPDAPLEGIRRLLEVLE
HyuA-HyuB      MKL----FGVDVGGTFTDIIFS------DTETRVTAIHKVPTTLDDPSTGVVQGILELCD
               *.     ...* ****** . .      . .* .  .   *    *..   .. .**. .
               ↑HyuA

F1286          HKTIPRGIPLDISNVRSLRMGTTLATNCALERNGERCAFITTKGFKDSLLIGDQTRPDIF
HyuA-HyuB      RQYIDR------TAIDHVFHGTTIATNAILEYDGAKTGMITTEGYRDIIHIGRHQRPQNY
               .. * *      . .  .  ***.***  ** .* . . *** *..* . ** . **. .

F1286          NLNIKKVVPLYD-TVVEIDERVTLEDFSEDPYFTKSSPNEQEGILEGNSGEMVRVIKKPD
HyuA-HyuB      --SIMQEIPWQDRPLVQRRHRLAIA--------------ERMGPVKGQ------VITPVQ
                 .* . .*  *  .*.  .*...                *. * . *.      **   .

F1286          ESSVRSILKVLYASGIKSIAIAFLHSYTFPDHE-RIVGNIAREIGFSHVSLSSEVSPMIK
HyuA-HyuB      EDQVRGAVATLKERGVDSIIVNFLFSYTNPEHEQRVKEIIEEEYPEAFVTISSEVSPQFR
               *  **. .  *   *. ** . ** *** *.** *.   •  *   . *..****** ..

F1286          FLPRAHSSVADAYLTPVIKKYLNSISAGLSHAE-DTHIQFMQSDGGLVDGGKFSG--LKS
HyuA-HyuB      EFERFTTASINGFVGPKVKNYIQNLEQSLKDSGISAELHIMCSNGGVATPKTVSEKPVNT
                . *  ..  .... * .*.*....  .* ..   .....• *.**.      *   ...

F1286          ILSGPAGGVIGYSSTCYDKNNNIPLIGFDMGGTSTDVSR-----YGDGRLEHVFETVTAG
HyuA-HyuB      LLSGPAAGILGGAWAG-ELTNRQKLITFDVGGTSADIGIITDSGYGESSARDTW---IAG
               .*****.*..* . .  .  *   ** **.****.*..      **..   .      **

F1286          IIIQSPQLDIHTVAAGGSSILSWKNG-LFRVGPDSAAADPGPAAYRKGG-PLTITDANLF
HyuA-HyuB      YPVMVPMIDIHTIGAGGGSIAHIDEGGAFKVGPRSAGSRPGPACYGHGGLKPTVSDANVV
                 .  * .****..***.**    .*  *.*** **.. **** *  **   *..***.

F1286          LGRLVPEFFPKIFGPNEDESLDLETTTLKFRELTDVINKDLNSNLTM--EEVAYGFIKVA
HyuA-HyuB      LGRIDEHNF-------------LGGEMKIYTNAAYKVIDELAGQLDLSRERTAEGVLQIM
               ***.  . *            •      . . .  . .* ..* .  *  * * ...

F1286          NECMARPVRAITEAKGHVVSQHRLVSFGGAGGQHAIAVADSLGIDTVLIHRYSSILSAYG
HyuA-HyuB      NNNMANAIREKTIQKGEDPREFSLVAFGGAGPLHAVEVAQILNIPEVIIPLYPGINSATG
               *. ** ..*  *  **.   .  **.*****  **. **. * *  *.*  *..* ** *

F1286          IFLADVIEENQEPCSFILGEPETILKVKKRFLELSKNSIKNLLSQSFSREDIVLERYLNL
HyuA-HyuB      LLTTDLKYDVIKT-EFMMSTNMDFSGLNEDLAGLETQLINQLKEDGVSKQDIRILRSADC
               .. .*.  .      *...    .  ..  .  *  . *..*  .. *..** . *  .

F1286          RYEGTETSLMI--LQKYDDQWNFRE---WFSEAHKKEFGFSFDDKRIIIDDIRIRAIG-K
HyuA-HyuB      RYAGQGYELRVDLPDVFLDEETIVDALNNFHESHKAEYGHNFTDSPIEFVNIRVTGTGYM
               ** *    * .   . . *.  . .    * *.** *.* .* *  * . .**. . *

F1286          SGVRKEKTVDEQLIEISHFKKADVSKDASFTQKAYFDNKWVDTAVFKIDDLPAGTIIEGP
HyuA-HyuB      PKIEKQAIHHEYQLEDALLKTGD----ATFNIDGSLVK--VEINFYQREKIPVGAEFNGP
               . . *.   .*  .* . .* .*    *.*   . . .  *.   .. . .* *. ..**

F1286          AILADGTQTNIILPNSQATILN-SHIFIKINQKAAKTLSKSGYELD-IDPILLSIFSHRF
HyuA-HyuB      CIVLQKDTTTVIPPNCTAYIEEYGNMIIKVG-----VMSKIHTDLKKIDPITVQVVLGSL
                *. .   * .* **  * * . ....**.       .**   .*  **** . .    .
                                                    ↑HyuB

F1286          MDIALQMGTQLRKTSVSTNVKERLDFSCALFDSKGNLVANAPH-VPVHLGSMSTCISAQA
HyuA-HyuB      ENVAVEMGHKLARMSYSSIIRESEDFGCALVDVRGQQLCESSHSTPLQSGPIPGYIKGIR
                ..*..** .* . * *. ..*  **.*** * .*. . ...*  *.. *...  * .

F1286          KLWEGK---LKPGDVLITNHPDIGGTHLPDITVITPSFSSTGELIFYVASRAHHADIGGI
HyuA-HyuB      EIMEDRNDTFNQGDVIMHNSPYHGASHGPDVGFCIPVFYK-DELIGFSVTTAHHLDIGSS
                . *..   .. ***.. * *  *..* **.    * *   .*** .  . *** ***.
```

Figure 6.

```
F1286          LPGSVP-PNSKELYEEGTAIYSELVVKEGIFQEELIYKLFVEDPGKYPGCSGSRRFSDNI
HyuA-HyuB      TPGSCGIVDAVDAYAEGLQFKAIKVYDQGV-KNRYVWDILKDN---------IRAPKLVV
                ***    .. . * **  . .  *  .*. ..  .  .. ..          *     .

F1286          SDLKAQVAANTKGIQLIGSLTKEYDLATILKYMAAIQTNASESIKKMLAKMVE-HFGTTK
HyuA-HyuB      GDMEAQIAAARIGAQRYIEIIEKYGLDTVQAASEELMNYSEKMMRDAIKKLPDGEYTAEG
               .*. **.**   * *    .   *.* *.      .  .  ..  . *. . .. .

F1286          FSGEDRLDDGSL----IKLQVIIRPEKEEYIFNFDGTSPQVYGN-LNAPE-AITNSAILY
HyuA-HyuB      FL-DGYLDSDDPAKKDLRINVTVKVDGSDLTVDLTGTSPQVTDKPINMPLLGTVDIAIYL
               *  .. ** .      ....* .. . .   .. ****** .. .* *  .  . **

F1286          CLRCLVGE-----DIPLNQGCLKPLTIKIPAGSLLSPRSGAAVVGGNVLTSQRVTDVILK
HyuA-HyuB      TLRSILLDSTVYGNFPQNSGLIRPIKIVAPKGTLCNPIFPAPTIA-RFNSGNAVADTLMK
                ** .. .     ..* * * ..*. *  * *.* .*   *. ..    ... *.* ..*

F1286          TFNVMADSQGDCNNFTFGTGGNSGNKTDKQIKGFGYYETICGGSGAGADSWRGSGWNGSD
HyuA-HyuB      ALAQVVPHQVSAGVGNLQVVAFSGQSNEN----YWVYMDIMEGSYGGR-----YGKDGMD
               .. .   *       .  . **.  ..    .  *  *  ** .*       * .* *

F1286          AVHTNMTNTRMTDTEVFERRYPVLLKEFSIRRGSGGKGKYTGGNGVVRDVQFRKAVTASI
HyuA-HyuB      AVDTLYANTRNNPIEDIESHYPLRVNRYELRDNDSAPGKWRGGIGSIREVSFLADGSFSV
               **.*  .***    * .* .**. .. . .*   .. **  ** * .*.* *    . *.

F1286          LSERRVIGPHGIKGGQDGSRGENLWVRHS-TGALINVGGK---------NTI--------
HyuA-HyuB      EADGHKYAPWGFDDGQDGYVG-SLSIRDNETNELVQLPSKLPNRHAQSGSTIQLVGPCGG
                .. .  .* *. .**** * .* .*.. *  *... .*         .**

F1286          -YAQPGDR------------FIIKTPGGGGFG---------QY-------KD
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      1 MGYQLSATLKGHDQDVRDVVAVDDSKVASVSRDGTVRLWSKDDQWLGTVVYTGQGFLNSV
PLAP      1 MHY-----MSGHSNFVSYVCIIPSSDI----------------------YP--------
            * *     . ** . *  *  .  * .                        *

F715        CYDSEKELLLFGGKDTMINGVPLFATSGEDPLYTLIGHQGNVCSLSFQD-GVVISGSWDK
PLAP        -----HGLIATGGND---HNICIFSLDSPMPLYILKGHKDTVCSLSSGKFGTLLSGSWDT
                   *.  **.*   . . .*.  .  *** * **.. *****    * ..*****

F715        TAKVWKEGSLVYNLQAHNASVWDAKVVSFSENKFLTASADKTIKLWQNDKVIKTFSGIHN
PLAP        TAKVWLNDKCMMTLQGHTAAVWAVKILP-EQGLMLTGSADKTIKLWKAGRCERTFLG-HE
            ***** ..  .  **.* *.**  *...  .   **.*********. ..  .** * *.

F715        DVVRHLAVVDDGHFISCSNDGLIKLVDMHTGDVLRTYEGHESFVYCIKLLPNG-DIVSCG
PLAP        DCVRGLAILSETEFLSCANDASIRRWQI-TGECLEVYFGHTNYIYSISVFPNSKDFVTTA
            * ** **.. . .*.**.**. *.  .. **. *  * ** ...* * ..**. *.*. .

F715        EDRTVRIWSKENGSLKQVITLPAISIWSVDCMSNGDIIVGSSDNLVRIFSQEKSRWASED
PLAP        EDRSLRIWK--HGECAQTIRLPAQSIWCCCVLENGDIVVGASDGIIRVFTESEERTASAE
            ***..***   .*   * * *** ***    . ****.**.** ..*.*..   * ** .

F715        EINELSTQVEKSTISSKT 316
PLAP        EI--------KASLSRES 261
            **        *...*  .
```

B

```
F715      1 MGYQ--LSATLKGHDQDVRDVVAVDD--SKVASVSRDGTVRLWSKDDQWLGTVVYTGQGF
A33928    1 MTEQMTLRGTLKGHNGWVTQIATTPQFPDMILSASRDKTIIMWKLT--------------
            *  *  * .*****.  * .. .  .    . * *** *. .*

F715        LNSVCYDSEKELLLFGGKDTMINGVPLFATSGEDPLYTLIGHQGNVCSLSFQDGVVISGS
A33928      ---------RDETNYGIPQRALRGHSHFVSD------VVISSDGQF---------ALSGS
                    ..   .*  .  . ● . * .        .*. .*.          .***

F715        WDKTAKVWK--EGSLVYNLQAHNASVWDAKVVSFSENK--FLTASADKTIKLWQNDKVIK
A33928      WDGTLRLWDLTTGTTTRRFVGHTKDVLS---VAFSSDNRQIVSGSRDKTIKLWNTLGVCK
            ** * ..*    *.    . .*   *     *.** ..  ....* *******.   * *

F715        TFSGIHNDVVRHLAVVDDGHFISCSNDGLIKLVDMHTGDVLRTYEGHESFVYCIKLLPNG
A33928      -YT------------VQD-------------------------ESHSEWVSCVRFSPNS
             ..            *.*                        *.*   * *... **.

F715        D---IVSCGEDRTVRIWSKENGSLKQVITLPAISIWSVDCMSNGDIIVGSSDNLVRIFSQ
A33928      SNPIIVSCGWDKLVKVWNLANCKLK-----------------------------------
                ***** *. *..*.  *  **

F715        EKSRWASEDEINELSTQVEKSTISSKTIEFDESKLSPYEILQSPGRKEGQIVVVKS 345
A33928      --------------TNHIGHTGY------LNTVTVSPDGSLCASGGKDGQAMLWDL 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 yeast 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 frameshifts 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 *ILV1* 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.

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 A_2-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 cerevisiae* 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 *ilv*GMEDA 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 ubituitin-activating 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.