Volume 13 Number 11 1985 Nucleic Acids Research

Nucleotide sequence of the yeast *ILV2* gene which encodes acetolactate synthase

S.Carl Falco*, Kathy S.Dumas and Kenneth J.Livak

Central Research and Development Department, E. I. du Pont de Nemours and Company, Wilmington, DE 19898, USA

Received 27 February 1985; Revised and Accepted 7 May 1985

ABSTRACT

We have determined the nucleotide sequence of the yeast ILV2 gene which codes for the amino acid biosynthetic enzyme acetolactate synthase (ALS). ALS has recently been shown to be the target in bacteria, yeast and plants, of the potent new herbicide sulfometuron methyl. The coding sequence for the ILV2 polypeptide contains 2061 base pairs. Comparison of deduced amino acid sequences indicates considerable conservation between the yeast protein and the large subunits of the E. coli ALS II and ALS III isozymes. A major distinction between the three proteins is the presence of an additional 90 amino acids at the amino terminal of the yeast protein. The amino acid sequence in this region shows similarities to yeast mitochondrial transit sequences and may function as such, since yeast ALS is localized in the mitochondria. Consensus sequences for initiation and termination of transcription that are consistent with the ends of the ILV2 mRNA, as well as general amino acid control regulatory sequences have been identified.

INTRODUCTION

The first common step in the biosynthesis of the branched chain amino acids is catalyzed by acetolactate synthase (also known as acetohydroxy acid synthase, EC 4.1.3.18). Detailed studies on the genetics, biochemistry and regulation of this enzyme have been carried out in bacteria (for a review see ref.1). There are several genes, some cryptic, that encode acetolactate synthase (ALS) isozymes in E. coli and Salmonella typhimurium. Biochemical analysis has shown that ALS isozymes I and II are composed of two different polypeptides of approximately 60 kD and 10 kD (2,3); genetic and DNA sequence analyses suggest that isozyme III is also composed of two subunits (4,5). Isozymes I and III, but not II are feedback inhibited by valine (1). Two of the genes, ilvG and ilvI, which encode the large and catalytically essential subunits of isozymes

© IRL Press Limited, Oxford, England.

II (inactive in wild-type E. coli K-12 due to a frameshift mutation) and III, respectively, have been isolated and sequenced (5,6). Considerable sequence conservation at both the nucleotide and amino acid levels was observed (5). Immediately downstream from each of these genes and downstream from the ilvB gene, which encodes the large subunit of isozyme I, are open reading frames which encode polypeptides of the expected size for the small subunits (5,6,7). The predicted amino acid sequence from the open reading frame downstream of ilvG in E. coli (designated ilvM) matches the amino terminal sequence of the small subunit of isozyme II from Salmonella typhimurium (3). The predicted amino acid sequence of the open reading frame downstream from ilvI, designated ilvH, shows some homology with that of ilvM (5). Genetic and biochemical evidence indicates that the ilvH gene product confers valine sensitivity on the ALS III isozyme, but is not required for catalytic activity (4).

In the yeast Saccharomyces cerevisiae a single gene essential for ALS activity has been identified genetically and designated ILV2 (8,9,10). The ILV2 gene has been mapped to the right arm of chromosome XIII (9,10), molecularly cloned (10,11) and shown to be unique in the yeast genome (10). Yeast ALS has been shown to be feedback inhibited by valine (12) and localized in the mitochondria with the other isoleucine and valine biosynthetic enzymes (13). It is not known whether yeast ALS is composed of multiple subunits.

New interest in ALS has been generated recently by the discovery that it is the molecular target of two structurally unrelated classes of herbicides, the sulfonylurea herbicides and the imidazolinone herbicides (10,14,15,16,17,18). The sulfonylurea herbicide sulfometuron methyl is a potent inhibitor of ALS from bacteria, yeast and higher plants with an apparent Ki in the 10-100 nM range (10,14,15,16). Among ALS enzymes involved in amino acid biosynthesis the only naturally occuring sulfometuron methyl resistant isozyme thus far identified is ALS I of Salmonella and E. coli (19). However, sulfometuron methyl resistant mutants which produce resistant forms of ALS have been isolated in Salmonella, yeast and tobacco (10,14,16). In Salmonella and yeast the mutations have been mapped to ilvG

and ILV2, respectively (10,14). In yeast at least twelve phenotypically distinct resistance mutations in the ILV2 gene have been identified and several have been cloned (10).

In this paper we present the first nucleotide sequence of a gene encoding acetolactate synthase from a eucaryote, the ILV2 gene of Saccharomyces cerevisiae. We compare the deduced amino acid sequence of the ILV2 polypeptide to that of the large subunits of the E. coli ALS isozymes II and III. We also identify potential regulatory sequences, as well as sequences likely to play a role in the mitochondrial localization of yeast ALS.

MATERIALS AND METHODS

Strains, phages and media

E. coli strain JM103 [Δ(lac pro), thi, rpsL, endA, sbcB15, hsdR, supE/F'(traD36, proA, proB, $lacI^Q$, lacZ M15)] was used as a host for growth of phages M13mp8 and M13mp9 (20) and their derivatives. WB373 is E. coli strain MM294 [endA1, thi, hsdR17, supE44] carrying the plasmid pWB373 which confers kan^r, amp^r and $M13^s$; it was used as a host for growth of M13WB2348 phage (21) and its derivatives. JM103 and WB373 were grown in LB media (22). Ampicillin was added to growth media at a concentration of 50 μg/ml when required.

Cloning DNA fragments into M13

The double-stranded replicative form (RF) of the M13 vectors mp8 and mp9 were purchased from New England Biolabs (NEB). The RF from M13WB2348 and the plasmids pCP2-4 and pKD1-2 that carry the ILV2 were prepared by CsCl equilibrium density gradient centrifugation (22). DNA was digested with restriction endonucleases purchased from either NEB or Bethesda Research Laboratories (BRL) according to the suppliers protocols. DNA fragments were size-separated by agarose gel electrophoresis in low melting temperature agarose and purified using a NACS PrePack column (BRL). Ligation, transfection and purification methods for recombinant M13 phages were as described in the instruction manual provided with the NEB M13 dideoxynucleotide sequencing system or as published (21). Relative orientations of cloned DNA fragments in M13 were determined as follows. Single-stranded

phage DNA was electrophoresed in agarose gels and transferred to nitrocellulose (23). A ^{32}P-labeled hybridization probe complementary to any one cloned DNA fragment was prepared as described (24). The nitrocellulose filter was pre-hybridized for one hour at 42°C in a solution of 50% formamide, 5 X SSPE (22), 1% sarkosyl and hybridized to the labeled probe in the above solution at 42°C overnight. (1 X SSPE is 0.18M NaCl, 10mM NaH_2PO_4, pH 7.4, 1mM EDTA.) The filter was then washed with 0.1 X SSPE, 0.1% SDS at 50°C and exposed to Kodak XRP film with a DuPont Cronex intensifying screen overnight. Cloned fragments in the opposite orientation from the probe form hybrids.

DNA sequencing

DNA sequencing reactions were performed by the dideoxy chain termination method (25). Synthesis was primed with the universal 15 nucleotide primer (NEB) complementary to a region of the M13 phages immediately upstream of the cluster of restriction endonuclease cloning sites. Polyacrylamide gel electrophoresis and autoradiography was as described (26).

S1 nuclease mapping of ILV2 mRNA

For S1 protection studies, the following fragments were inserted in the M13 vector mWB2348 in both orientations: pKD1-2 Δ4 BamHI/HindIII-2747 bp; pKD1-2 BglII-820 bp; and pKD3-4 Sau3A-1220 bp (see Figure 1). The BamH1 site in pKD1-2Δ4 was introduced as a linker in a Bal31 generated deletion mutation (10). The end of the deletion was determined by DNA sequencing. Yeast RNA was isolated from strain FY138 carrying plasmid pCP2-4 (10) as described(27). Poly A^+ RNA was selected on poly[U]-sepharose. A typical hybridization reaction contained 100 ng yeast polyA^+ RNA and 100 ng single-stranded M13 DNA with an inserted DNA fragment complementary to the ILV2 mRNA in 20 1 5 X SSPE, 50% (v/v)formamide. This reaction was incubated at 95°C for 1 min and then incubated at 47°C for 18 h. For the RNA^- control, the yeast RNA was replaced with H_2O. For the DNA-DNA control, the yeast RNA was replaced with 100 ng single-stranded M13 DNA with the inserted DNA fragment of the same polarity as the mRNA and the hybridization temperature was reduced from 47°C to 42°C. Following hybridization, each sample was mixed with 20 µl 0.1M Na-acetate, pH 4, 0.3M NaCl, 6mM $ZnSO_4$; 300 units of S1

nuclease was added and the reaction were incubated at 37°C for 30 min. Protected fragments were precipitated by the addition of 40 µl 5M NH_4-acetate plus 200 µl ethanol and stored at -20°C. Samples were electrophoresed on 1.0% alkaline agarose gels (25), blotted onto nitrocellulose, and hybridized with ^{32}P-labeled ALS-specific probe prepared as described above from an M13 phage which carried the 2747 bp BamHI/Hind III DNA fragment. Hybridization conditions and autoradiography were as described above.

RESULTS AND DISSCUSSION

Nucleotide sequence of the ILV2 gene

The molecular cloning of the yeast ILV2 gene on a 5.6 kb pair DNA fragment in the vector YEp24 has been described previously(10). Several methods were used to localize the gene on this fragment including functional analysis of in vitro constructed deletion mutations (10) and in vivo constructed Tn5 insertion mutations (T. Van Dyk, personal communication), as well as S1 nuclease mapping of the ILV2 mRNA (see below). A restriction endonuclease map of the ILV2 gene and flanking sequences is shown in Figure 1.

Using the bacteriophage M13 and the dideoxy chain

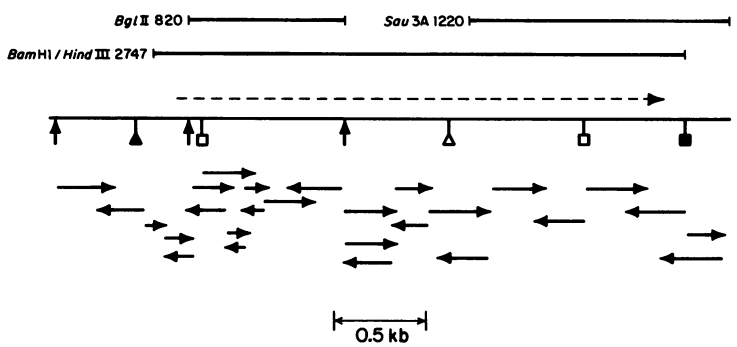

Figure 1. DNA sequencing and S1 mapping strategy. A restriction endonuclease map of the ILV2 DNA fragment is shown: (↑)BglII; (▲)KpnI; (□)EcoRI; (△)PvuII; (■)HindIII. The DNA fragments shown above the map were used in S1 nuclease protection experiments. The dashed arrow indicates the extent of the ILV2 mRNA and the direction of transcription. The arrows below the map show the DNA sequencing strategy.

termination method, a total of 3523 nucleotides including the entire ILV2 gene and its 5' and 3' flanking regions was sequenced employing the strategy shown in Figure 1. The sequence was confirmed using nine custom synthesized oligonucleotide primers, spaced 250 nucleotides apart, to re-sequence the gene from the non-coding strand. More than 80 percent of the sequence was determined from both DNA strands. The regions not sequenced from both strands were sequenced at least twice from different M13 clones using different primers with no ambiguity in the sequence.

The nucleotide sequence includes a continuous open reading frame of 2061 bases that starts with an ATG designated nucleotide +1 in Figure 2 and ends with a TGA. Northern blot analysis of yeast mRNA demonstrated the presence of an approximately 2.5kb mRNA that hybridizes to the cloned ILV2 gene (not shown). The entire functional ILV2 gene is contained on the 2747bp BamHl/HindIII DNA fragment of plasmid pkDl-2Δ4(10) which extends from nucleotide -339 to +2408 (Figure 2). From this 2747bp fragment an approximately 2.4kb DNA fragment is protected from Sl nuclease digestion by Poly A^+ RNA (Figure 3A). This result indicates that there are no intervening sequences within the ILV2 coding region which is consistent with the DNA sequence information. Further Sl nuclease mapping indicates that the 5' end of the ILV2 mRNA is upstream of the Bgl II site at -190, since the entire 820bp Bgl II fragment (from -190 to +630) is protected from Sl nuclease digestion by the ILV2 mRNA (Figure 3B). Thus transcription of ILV2 must begin in the 150 base pair interval from -190 to -339. The sequence TATA, which has been shown to be important for initiation of transcription of eucaryotic mRNA (28), is present in this interval (underlined in Figure 2), suggesting that transcription begins between -190 and -267. The ATG codon at nucleotide +1 starts the long open reading frame encoding ALS. This is the most 5' ATG shown to be present on the ILV2 mRNA by the Sl protection experiment and is probably the first ATG on the ILV2 mRNA. Three other ATG codons are upstream in the -190 to -339 interval and therefore could be present on the ILV2 mRNA. However, two of these are upstream of the TATA box, the third is immediately adjacent to it, and all three are out of frame with the ILV2 coding sequence and would

```
                                                                                                                    Bgl II
-897                                                                                                                AGATCT TCCATAGGAT CACAGGACAG TTCCGATGTT GAGGACGTTA AGGAAGGGAG

-840 ATTACCCGCA GAAGTAGAAA TACCAAAGAA TGTTGACATC TCTAACATGT CGCAAGGTGA GTTTTTAAGA CTTTACGAAA GTTTGAGGAG GGGGGAACCG ACAATAAAGT AAATAGATAA ATGGAGGAGA TATTCATAGT

-700 GTCTTTTCTT GTGTTGTCTA TTTATTCTAA CATGATTATA TATTTATTAT GCTTTTCAGT GTTTCCATTA GTTTTATATG ATAGCATCCT TATGTCTTTT TTCATACCTA TATAAGCAAT AAGAAGTTAT GTAAACAGAA
                                                                                                                                                  Kpn I
-560 CTTTGCCACT AATACCATAA TTACTGTCTT TTAGTATCAT TGTCTGTCAG TCGGCACGGA TGAAAGGTGA CAAACGCCTA GCCGCCGGAG CCTGCCGGTA CCGGCTTGGC TTCAGTTGCT GATCTCGGCG CGGAAAAATC

-420 AGCGCCCACG CCAAAAGGTT CGTATTTTTT CTTTTTTTTT CTAATCTTCC ATCTATTCGG TAGCGATGAT TCATTTCTCG GAAAAAAAAA AAAAAAAAAA AAAATGAAAA AGAATATTTT TTTGATGAAC TTGTATTTCT
                                                                                                                                  Bgl II
-280 CTTATCTGGT TGATATATAT GCTATCATTT ATTTTCTTAT CAAGTTTCCA AATTTCTAAT CCTTTCTCCA CCATCCCTAA TTAATAATTC AGATCTACGT CACACCGTAA TTTGTATTGT TTTTTTCCTT CATTGTCTAA
          Eco RI
-140 ACCGAAGAAT TCATCAGCCA CAGTTACTAG TTCATTTGAA GCGAAATTAC ACACATTTTC CCTGTTACAA TAGAAAGTAT TTTACAAAAT CTAAACCCTT TGAGCTAAGA GGAGATAAAT ACAACAGAAT CAATTTTCAA

+  1 ATG ATC AGA CAA TCT ACG CTA AAA AAC TTC GCT ATT AAG CGT TGC TTT CAA CAT ATA GCA TAC CGC AAC ACA CCT GCC ATG AGA TCA GTA GCT CTC GCG CAG CGC TTT
     MET Ile Arg Gln Ser Thr Leu Lys Asn Phe Ala Ile Lys Arg Cys Phe Gln His Ile Ala Tyr Arg Asn Thr Pro Ala MET Arg Ser Val Ala Leu Ala Gln Arg Phe   36

 108 TAT AGT TCG TCT TCC CGT TAT TAC AGT GCG TCT CCA TTA CCA GCC TCT AAA AGG CCA GAG CCT GCT CCA AGT TTC AAT GTT GAT CCA TTA GAA CAG CCC GCT GAA CCT
     Tyr Ser Ser Ser Ser Arg Tyr Tyr Ser Ala Ser Pro Leu Pro Ala Ser Lys Arg Pro Glu Pro Ala Pro Ser Phe Asn Val Asp Pro Leu Glu Gln Pro Ala Glu Pro   72

 216 TCA AAA TTG GCT AAG AAA CTA CGC GCT GAG CCT GAC ATG GAT ACC TCT TTC GTC GGT TTA ACT GGT GGT CAA ATA TTT AAC GAA ATG ATG TCC AGA CAA AAC GTT GAT
     Ser Lys Leu Ala Lys Lys Leu Arg Ala Glu Pro Asp MET Asp Thr Ser Phe Val Gly Leu Thr Gly Gly Gln Ile Phe Asn Glu MET MET Ser Arg Gln Asn Val Asp  108

 324 ACT GTA TTT GGT TAT CCA GGT GGT GCT ATC CTA CCT GTT TAC GAT GCC ATT CAT AAC AGT GAT AAA TTC AAC TTC GTT CTT CCA AAA CAC GAA CAA GGT GCC GGT CAC
     Thr Val Phe Gly Tyr Pro Gly Gly Ala Ile Leu Pro Val Tyr Asp Ala Ile His Asn Ser Asp Lys Phe Asn Phe Val Leu Pro Lys His Glu Gln Gly Ala Gly His  144

 432 ATG GCA GAA GGC TAC GCC AGA GCT TCT GGT AAA CCA GGT GTT GTC TTG GTT ACT TCT GGG CCA GGT GCC ACC AAT GTC GTT ACT CCA ATG GCA GAT GCC TTT GCA GAC
     MET Ala Glu Gly Tyr Ala Arg Ala Ser Gly Lys Pro Gly Val Val Leu Val Thr Ser Gly Pro Gly Ala Thr Asn Val Val Thr Pro MET Ala Asp Ala Phe Ala Asp  180
                                                                                                                                  Bgl II
 540 GGG ATT CCA ATG GTT GTC TTT ACA GGG CAA GTC CCA ACT AGT GCT ATC GGT ACT GAT GCT TTC CAA GAG GCT GAC GTC GTT GGT ATT TCT AGA TCT TGT ACG AAA TGG
     Gly Ile Pro MET Val Val Phe Thr Gly Gln Val Pro Thr Ser Ala Ile Gly Thr Asp Ala Phe Gln Glu Ala Asp Val Val Gly Ile Ser Arg Ser Cys Thr Lys Trp  216

 648 AAT GTC ATG GTC AAG TCC GTG GAA GAA TTG CCA TTG CGT ATT AAC GAG GCT TTT GAA ATT GCC ACG AGC GGT AGA CCG GGA CCA GTC TTG GTC GAT TTA CCA AAG GAT
     Asn Val MET Val Lys Ser Val Glu Glu Leu Pro Leu Arg Ile Asn Glu Ala Phe Glu Ile Ala Thr Ser Gly Arg Pro Gly Pro Val Leu Val Asp Leu Pro Lys Asp  252

 756 GTT ACA GCA GCT ATC TTA AGA AAT CCA ATT CCA ACA AAA ACA ACT CTT CCA TCA AAC GCA CTA AAC CAA TTA ACC AGT CGC GCA CAA GAT GAA TTT GTC ATG CAA AGT
     Val Thr Ala Ala Ile Leu Arg Asn Pro Ile Pro Thr Lys Thr Thr Leu Pro Ser Asn Ala Leu Asn Gln Leu Thr Ser Arg Ala Gln Asp Glu Phe Val MET Gln Ser  288

 864 ATC AAT AAA GCA GCA GAT TTG ATC AAC TTG GCA AAG AAA CCT GTC TTA TAC GTC GGT GCT GGT ATT TTA AAC CAT GCA GAT GGT CCA AGA TTA CTA AAA GAA TTA AGT
     Ile Asn Lys Ala Ala Asp Leu Ile Asn Leu Ala Lys Lys Pro Val Leu Tyr Val Gly Ala Gly Ile Leu Asn His Ala Asp Gly Pro Arg Leu Leu Lys Glu Leu Ser  324

 972 GAC CGT GCT CAA ATA CCT GTC ACC ACT ACT TTA CAA GGT TTA GGT TCA TTC GAC CAA GAA GAT CCA AAA TCA TTG GAT ATG CTT GGT ATG CAC GGT TGT GCT ACT GCC
     Asp Arg Ala Gln Ile Pro Val Thr Thr Thr Leu Gln Gly Leu Gly Ser Phe Asp Gln Glu Asp Pro Lys Ser Leu Asp MET Leu Gly MET His Gly Cys Ala Thr Ala  360
```

```
                                                                                                                 Pvu II
1080 AAC CTG GCA GTG CAA AAT GCC GAC TTG ATA ATT GCA GTT GGT GCT AGA TTC GAC GAC CGT GTC ACT GGT AAT ATT TCT AAA TTC GCT CCA GAA GCT CGT CGT GCA GCT
     Asn Leu Ala Val Gln Asn Ala Asp Leu Ile Ile Ala Val Gly Ala Arg Phe Asp Asp Arg Val Thr Gly Asn Ile Ser Lys Phe Ala Pro Glu Ala Arg Arg Ala Ala  396

1188 GCC GAG GGT AGA GGT GGT ATT ATT CAT TTC GAG GTT AGT CCA AAA AAC ATA AAC AAG GTT GTT CAA ACT CAA ATA GCA GTG GAA GGT GAT GCT ACG ACC AAT CTG GGC
     Ala Glu Gly Arg Gly Gly Ile Ile His Phe Glu Val Ser Pro Lys Asn Ile Asn Lys Val Val Gln Thr Gln Ile Ala Val Glu Gly Asp Ala Thr Thr Asn Leu Gly  432

1296 AAA ATG ATG TCA AAG ATT TTC CCA GTT AAG GAG AGG TCT GAA TGG TTT GCT CAA ATA AAT AAA TGG AAG AAG GAA TAC CCA TAC GCT TAT ATG GAG GAG ACT CCA GGA
     Lys MET MET Ser Lys Ile Phe Pro Val Lys Glu Arg Ser Glu Trp Phe Ala Gln Ile Asn Lys Trp Lys Lys Glu Tyr Pro Tyr Ala Tyr MET Glu Glu Thr Pro Gly  468

1404 TCT AAA ATT AAA CCA CAG ACG GTT ATA AAG AAA CTA TCC AAG GTT GCC AAC GAC ACA GGA AGA CAT GTC ATT GTT ACA ACG GGT GTG GGG CAA CAT CAA ATG TGG GCT
     Ser Lys Ile Lys Pro Gln Thr Val Ile Lys Lys Leu Ser Lys Val Ala Asn Asp Thr Gly Arg His Val Ile Val Thr Thr Gly Val Gly Gln His Gln MET Trp Ala  504

1512 GCT CAA CAC TGG ACA TGG AGA AAT CCA CAT ACT TTC ATC ACA TCA GGT GGT TTA GGT ACG ATG GGT TAC GGT CTC CCT GCC GCC ATC GGT GCT CAA GTT GCA AAG CCA
     Ala Gln His Trp Thr Trp Arg Asn Pro His Thr Phe Ile Thr Ser Gly Gly Leu Gly Thr MET Gly Tyr Gly Leu Pro Ala Ala Ile Gly Ala Gln Val Ala Lys Pro  540

1620 GAA TCT TTG GTT ATT GAC ATT GAT GGT GAC GCA TCC TTT AAC ATG ACT CTA ACG GAA TTG AGT TCT GCC GTT CAA GCT GGT ACT CCA GTG AAG ATT TTG ATT TTG AAC
     Glu Ser Leu Val Ile Asp Ile Asp Gly Asp Ala Ser Phe Asn MET Thr Leu Thr Glu Leu Ser Ser Ala Val Gln Ala Gly Thr Pro Val Lys Ile Leu Ile Leu Asn  576

1728 AAT GAA GAG CAA GGT ATG GTT ACT CAA TGG CAA TCC CTG TTC TAC GAA CAT CGT TAT TCC CAC ACA CAT CAA TTG AAC CCT GAT TTC ATA AAA CTA GCG GAG GCT ATG
     Asn Glu Glu Gln Gly MET Val Thr Gln Trp Gln Ser Leu Phe Tyr Glu His Arg Tyr Ser His Thr His Gln Leu Asn Pro Asp Phe Ile Lys Leu Ala Glu Ala MET  612
                                                                      Eco RI
1836 GGT TTA AAA GGT TTA AGA GTC AAG AAG CAA GAG GAA TTG GAC GCT AAG TTG AAA GAA TTC GTT TCT ACC AAG GGC CCA GTT TTG CTT GAA GTG GAA GTT GAT AAA AAA
     Gly Leu Lys Gly Leu Arg Val Lys Lys Gln Glu Glu Leu Asp Ala Lys Leu Lys Glu Phe Val Ser Thr Lys Gly Pro Val Leu Leu Glu Val Glu Val Asp Lys Lys  648

1944 GTT CCT GTT TTG CCA ATG GTG GCA GGT GGT AGC GGT CTA GAC GAG TTC ATA AAT TTT GAC CCA GAA GTT GAA AGA CAA CAG ACT GAA TTA CGT CAT AAG CGT ACA GGC
     Val Pro Val Leu Pro MET Val Ala Gly Gly Ser Gly Leu Asp Glu Phe Ile Asn Phe Asp Pro Glu Val Glu Arg Gln Gln Thr Glu Leu Arg His Lys Arg Thr Gly  684

2052 GGT AAG CAC TGA ATTTCAAAAA CATTTATTTC AAAAGCATTT TCCAGTAAAA AATGCAGACT TTATTATTAT TTAATCGTGC TTCTTATATA TGACATTCTA CCAAATCGGT AGTCATGTAT ATTTTTTTCG
     Gly Lys His  .                                                                                                                  1     1       1

2184 TATATACTTT ATATATTTTT TTCTAAAAAA CTAATGACGG CTAAAATTAA GTCATAGATG AATAATAAGT TCAATTCAAG TGAGTTGGTA GTATTTGATA AATCTAAAGT GGATACGTAG CATATGTATT CAAATGGTGT
                                   2               2   3        3                          2,3                              4     4
                                                                                     Hind III
2324 GTTTTAACTC GTCGGACACG ACTTCTTTCT CGTTTTGTCA ATCATCAAAA ATATTTCTCG AAAAGGGGCC GTCAAAAATT GCAAGCTTGG AAAATACAAA CTGGACTATG TATTTAAAGG TTATAGTTTA TAATAAAGCA
      4

2464 TAAGAACAGT GGCCAATACG AATTTAACCG CTTTATAGAA ATGGCTATCT TAAAAAGAGG AGCTAGAAAA AAGGTACATC AGGAGGCCAG CTAAAGCGTC TGCGAATATC AAGAAAGCTA CTTTTGATTC CTCGAAGAAG

2604 AAAGAAGTTG GTGTGTCTGA TC
```

thus require an RNA splice in order to serve as the translation initiation codon.

Upstream of the TATA box the sequence TGATTC appears starting at nucleotide -354 (Figure 2). This is similar to the TGACTC sequence implicated in transcriptional regulation of amino acid biosynthetic genes by the general control system of yeast (29). Since the nucleotide at the fourth position of these sequences is variable (30) and since expression of the ILV2 gene is regulated by the general control system (unpublished results), it is likely that the TGATTC sequence is important for this regulation.

The localization of the 5' end of the ILV2 transcript to the interval -190 to -339 together with the protection of a 2.4 kb DNA fragment from the 2747 bp (-339 to +2408) fragment that includes the entire functional gene positions the 3' end of the mRNA in the interval +2060 to 2300. The 3' end must be downstream of the +2061 because that is the position of the stop codon for the ILV2 open reading frame. The 3' end of the message has been further localized by S1 nuclease mapping. The ILV2 mRNA protects an approximately 820 bp fragment from the 1220 bp Sau3A fragment that extends from nucleotide 1402 to 2622 (Figure 3B), suggesting that the end of the mRNA is in the vicinity of nucleotide +2220. There are several potential yeast trancription termination signals (31) in this region (underlined in Figure 2).

The S1 protection experiments do not eliminate the possibility of intervening sequences in the 5' or 3' untranslated regions. Such introns and their corresponding exons, if they exist, would have to be very small because the entire functional ILV2 gene is present on a 2750 bp DNA fragment, approximately 2400 contiguous base pairs of that fragment are protected from S1

Figure 2. DNA sequence of ILV2 and the deduced amino acid sequence of the ILV2 polypeptide. The A of the presumed translation initiation codon is numbered +1; the preceding A is numbered -1. Sequence elements outside of the coding region of possible significance are indicated as follows: the TATA sequence at -268 is underlined; the TGACTC-like sequence (implicated in general control of amino acid biosynthesis) at -354 is marked by a wavy line; potential transcription termination signals beginning at +2163 are underlined and grouped as indicated by numbers 1 to 4.

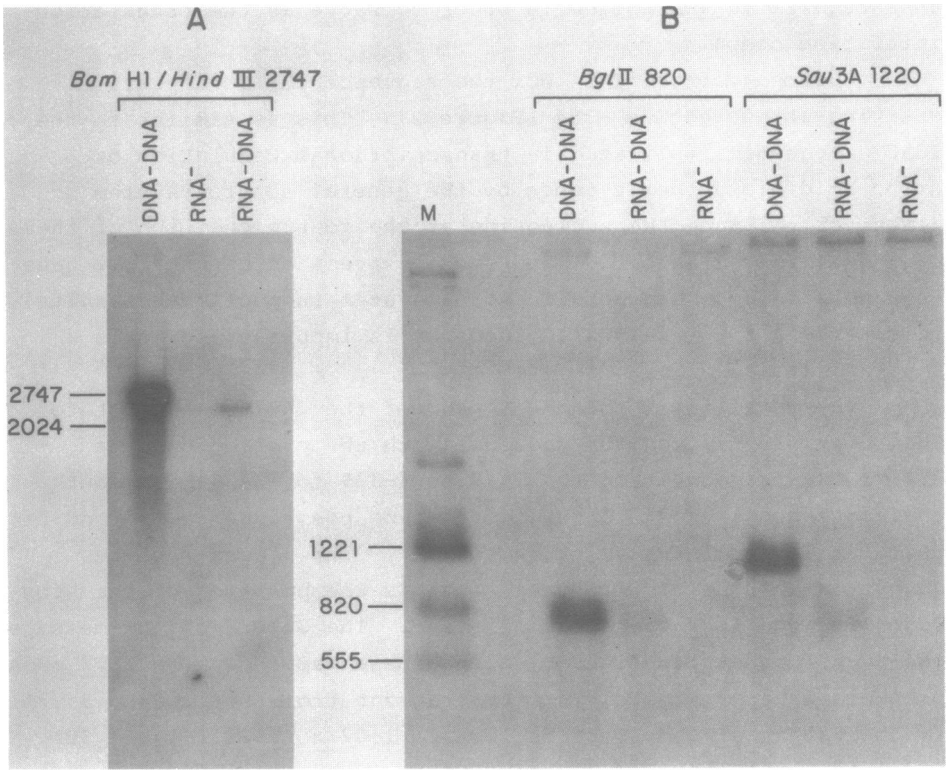

Figure 3. S1 nuclease mapping of the ILV2 mRNA. A)The BamH1/HindIII 2747 bp DNA fragment from -339 to +2408 includes the entire ILV2 coding region (see Figure 1). B)The BglII-820 bp DNA fragment is at the 5' end of the gene from -190 to +630. The Sau3A-1220 bp DNA fragment is at the 3' end of the gene from +1402 to +2622 (see Figure 1). Lane M contains DNA size markers.

digestion by the ILV2 mRNA and the size from Northern blot analysis of the ILV2 mRNA, including the poly A sequence, is about 2.5 kb.

Codon usage within the ILV2 gene does not show the strong bias observed in yeast genes that are highly expressed (32). Fifty-nine of the possible sixty-one codons are utilized in ILV2 (Table 1). Based on the method of Ikemura (33), the frequency of optimal codon usage in ILV2 is 0.59, a value similar to those calculated for TRP5 (34) and CPA2 (35) both of which are moderately expressed genes in yeast. The ILV2 codon usage is

Table 1. Codon frequencies in yeast ILV2

TTT-Phe	11	(1.6)	TCT-Ser	15	(2.2)	TAT-Tyr	5	(.7)	TGT-Cys	2	(.3)
TTC-Phe	16	(2.3)	TCC-Ser	7	(1.0)	TAC-Tyr	9	(1.3)	TGC-Cys	1	(.1)
TTA-Leu	16	(2.3)	TCA-Ser	7	(1.0)	TAA- .	0	(.0)	TGA- .	1	(.1)
TTG-Leu	18	(2.6)	TCG-Ser	1	(.1)	TAG- .	0	(.0)	TGG-Trp	7	(1.0)
CTT-Leu	4	(.6)	CCT-Pro	10	(1.5)	CAT-His	10	(1.5)	CGT-Arg	10	(1.5)
CTC-Leu	2	(.3)	CCC-Pro	1	(.1)	CAC-His	6	(.9)	CGC-Arg	4	(.6)
CTA-Leu	9	(1.3)	CCA-Pro	32	(4.7)	CAA-Gln	28	(4.1)	CGA-Arg	0	(.0)
CTG-Leu	3	(.4)	CCG-Pro	1	(.1)	CAG-Gln	4	(.6)	CGG-Arg	0	(.0)
ATT-Ile	19	(2.8)	ACT-Thr	18	(2.6)	AAT-Asn	12	(1.7)	AGT-Ser	10	(1.5)
ATC-Ile	8	(1.2)	ACC-Thr	6	(.9)	AAC-Asn	18	(2.6)	AGC-Ser	2	(.3)
ATA-Ile	10	(1.5)	ACA-Thr	11	(1.6)	AAA-Lys	25	(3.6)	AGA-Arg	14	(2.0)
ATG-MET	21	(3.1)	ACG-Thr	8	(1.2)	AAG-Lys	20	(2.9)	AGG-Arg	2	(.3)
GTT-Val	28	(4.1)	GCT-Ala	29	(4.2)	GAT-Asp	18	(2.6)	GGT-Gly	43	(6.3)
GTC-Val	17	(2.5)	GCC-Ala	15	(2.2)	GAC-Asp	14	(2.0)	GGC-Gly	4	(.6)
GTA-Val	2	(.3)	GCA-Ala	18	(2.6)	GAA-Glu	26	(3.8)	GGA-Gly	3	(.4)
GTG-Val	7	(1.0)	GCG-Ala	3	(.4)	GAG-Glu	13	(1.9)	GGG-Gly	4	(.6)

consistent with the moderate levels of mRNA and ALS activity observed.

Amino acid sequence of the ILV2 polypeptide

The deduced amino acid sequence of the ILV2 polypeptide is shown in Figure 2. Initiation of translation is indicated at the first ATG codon demonstrated to be present on the ILV2 mRNA (see above). While the first ATG in the mRNA is usually the start codon for translation of eucaryotic mRNAs (36,37), the ATG at nucleotide +1 has a pyrimidine (C) at the -3 position which is unusual in yeast. Generally a purine, most often A, is found at this position (37). The second in-frame ATG codon, which appears 78 nucleotides downstream, has a G at the -3 position; in addition, the second ATG is preceded, 11 nucleotides upstream, by the sequence ACACAC. The sequence PuCACACA has been noted at this approximate position upstream from a number of yeast translation start codons (38), but its significance remains to be determined. It is interesting that a very similar sequence arrangement is present in the ILV1 gene of yeast which encodes threonine deaminase(39). In ILV1 there are also two in-frame ATG codons, separated by 117 nucleotides. The first ATG is preceded by a C in the -3 position; the second has a G at -3 and ACACA begining at -14. However, in ILV1 there is a TAA stop codon between the two ATG codons indicating that the second ATG serves as the translation start for threonine deaminase. Which of the

Table 2. Amino acid composition of three acetolactate synthases

Amino Acid	Percentage by Weight Yeast ALS	E. coli ALS Isozyme II	E. coli ALS Isozyme III
A - Alanine	9.4	11.5	9.0
C - Cysteine	0.4	1.6	1.8
D - Aspartic Acid	4.7	6.0	4.8
E - Glutamic Acid	5.7	4.6	3.9
F - Phenylalanine	3.9	3.3	2.1
G - Glycine	7.8	8.6	8.6
H - Histidine	2.3	3.3	3.2
I - Isoleucine	5.4	3.8	5.3
K - Lysine	6.5	3.1	4.1
L - Leucine	7.6	10.9	8.8
M - Methionine	3.1	3.6	4.1
N - Asparagine	4.4	3.1	3.0
P - Proline	6.4	5.6	4.9
Q - Glutamine	4.7	6.2	5.5
R - Arginine	4.4	3.6	4.6
S - Serine	6.1	4.0	5.6
T - Threonine	6.3	5.5	5.5
V - Valine	7.8	7.8	9.9
W - Tryptophan	1.0	1.3	2.1
Y - Tyrosine	2.0	2.4	3.2

ILV2 ATG codons is used for initiation of translation (or whether both are used) remains to be determined.

Yeast ALS is found sequestered in the mitochondria along with the other isoleucine and valine biosynthetic enzymes (13). The amino-terminal region of the ILV2 polypeptide (starting from either the first or the second ATG codon) shows some of the characteristics noted previously for proteins destined for transport into the mitochondria (40,41,42). The amino-terminal sequence is basic; there are no acidic residues among the first fifty-five amino acids from the first ATG, but there are seven arginines, three lysines and one histidine. From the second ATG no acidic residues appear among the first twenty-nine amino acids, which include four arginines and one lysine. Starting from either ATG there is a stretch of four consecutive serine residues, as seen in the mitochondrial transit sequence of yeast cytochrome c peroxidase (40). There is also a stretch of ten apolar amino acids (Tyr 43 through Ser 52); this hydrophobic stretch is considerably shorter than that seen in cytochrome c

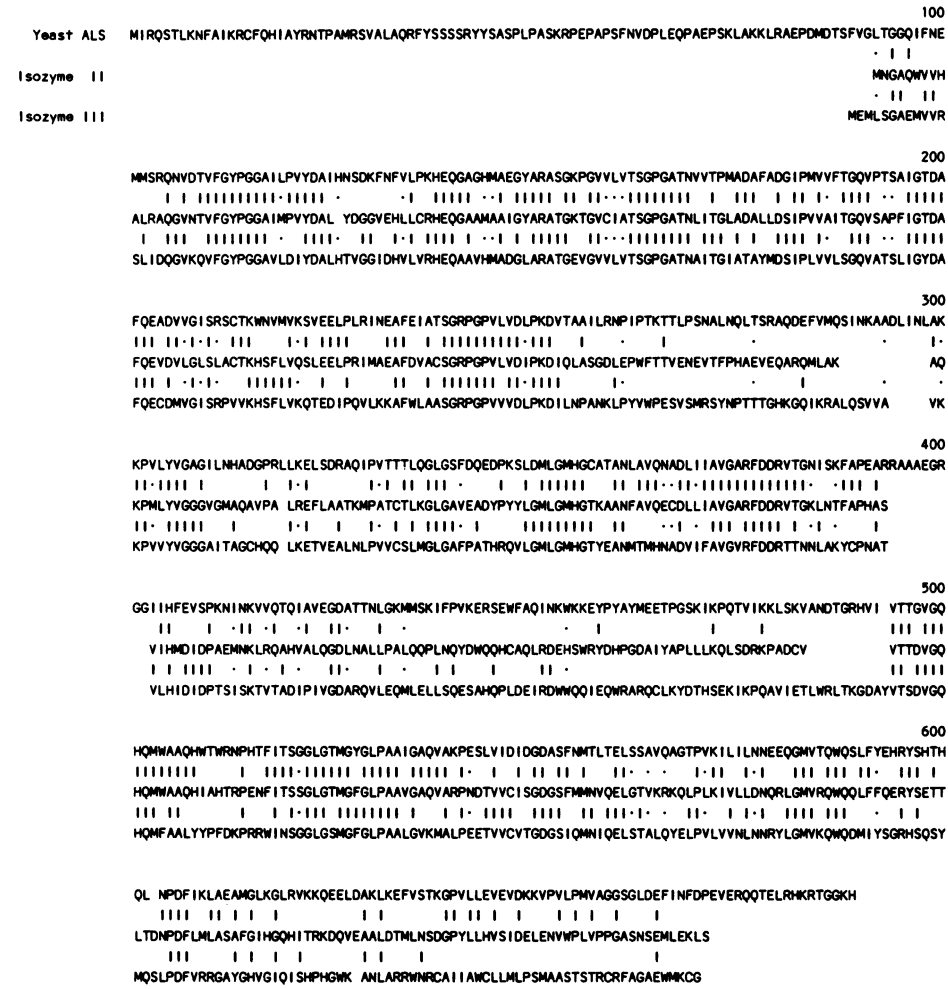

Figure 4. Comparison of deduced amino acid sequences of yeast and E. coli acetolactate synthases. The (|) indicates conserved amino acid residues between adjacent proteins; the (•) indicates conserved residues between yeast ALS and the large subunit of isozyme III only.

peroxidase (23 amino acids, 40) and in the 70kd mitochondrial outer membrane protein (28 amino acids, 41), but similar to that seen in Δ^1-pyrroline-5-carboxylate dehydrogenase, a mitochondrially located enzyme involved in proline degradation (8 amino acids, 43).

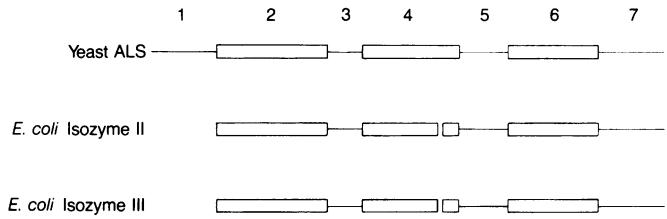

Figure 5. Schematic representation depicting similar structural organization of yeast and bacterial acetolactate synthases. The extent of the proteins is given by the lines; the open bars indicate homologous regions (see Table 3).

Structural homology between yeast and bacterial ALS

The deduced amino acid composition of the ILV2 protein (assuming translation initiation at the first ATG codon) is compared to that of the large subunits of the ALS II and ALS III isozymes of E. coli in Table 2. The compositions of the three proteins are very similar. The only major difference is the much lower cysteine content of the yeast protein. There are three cysteines in the yeast protein compared to nine in the large subunit of ALS II and ten in that of ALS III. Most of the cysteine residues are unconserved; none are present at the same position in all three proteins. One of the yeast cysteines is in the amino terminal portion of the protein (if translation starts at the first ATG) and may be lost if this region is cleaved during transport into the mitochondria. The lack of conservation of cysteine residues makes it unlikely that they play any unique role in the structure or function of ALS.

The deduced amino acid sequence of the ILV2 polypeptide is compared to that of the large subunits of E. coli isozymes ALS II and ALS III in Figure 4. It should be noted that in this analysis translation of the ALS II large subunit is started at the ATG codon at bp 271-273 (see ref.6), as suggested originally by Lawther and Hatfield (44), rather than at bp 352-354, as later suggested by Lawther et al. (6). That the ATG at bp 271-273 is the initiation codon is based on the amino-terminal sequence of the purified Salmonella ALS II isozyme (3). Considerable amino acid sequence conservation among the three proteins is evident; approximately 40% of the amino acid residues are identical. The

Table 3. Amino acid sequence conservation of yeast and bacterial acetolactate synthases

Region[a]	Number of Amino Acids[b]	Percent Conserved Amino Acids: Yeast ALS and Isozyme II	Yeast ALS and Isozyme III	Isozyme II and Isozyme III
1	91	-	-	-
2	162	57	52	59
3	45	4	7	4
4	129	46	40	40
5	65	6	5	5
6	113	58	49	52
7	82	22	7	7
Total	687	41	35	36

a. The regions of the proteins are shown schematically in Figure 5.
b. Number of amino acids is taken from yeast acetolactate synthase.

conserved residues are not uniformly dispersed throughout the proteins, however. Three regions of highly conserved amino acids separated by regions with no discernable similarity were previously noted in the two bacterial proteins (5). The same structural organization is evident in the ILV2 polypeptide. The percentage of conserved amino acids in the seven regions of the proteins shown diagrammatically in Figure 5 are presented in Table 3. It is remarkable that at the amino acid sequence level the yeast protein is as similar to either of the two bacterial proteins as the bacterial proteins are to each other.

Amino acid sequence homology between the yeast and bacterial proteins begins near the amino end of the bacterial proteins, within the first twenty residues. The ILV2 polypeptide possesses about 90 amino acids at its amino terminal which are absent from the bacterial proteins. (If translation of the ILV2 mRNA begins at the second ATG, rather than the first, the yeast protein would possess 64 amino acids absent from the bacterial protein.) As described above, it is likely that the amino end of the ILV2 polypeptide plays a role in transport of the protein into the mitochondria. Whether amino terminal residues are removed in the process of transport, as is the case for other proteins (40,42), remains to be determined.

The nucleotide sequence of the ILV2 gene provides the foundation for further studies on the regulation and

intracellular localization of yeast ALS. In addition, sequence comparison of the many ILV2 mutations which encode variants of ALS resistant to inhibition by the herbicide sulfometuron methyl (10) will provide insight into small molecule-protein interactions. Since many of the mutations affect ALS activity and/or allosteric interaction with the feedback inhibitor valine (10, unpublished results), the amino acid sequence differences will be helpful in understanding the function and regulation of ALS. Finally, the sequence information is essential for the construction of chimaeric genes for the introduction of sulfonylurea herbicide resistance into plants by transformation.

ACKNOWLEDGEMENTS

We thank Eric Whitehorn for help with DNA sequencing and Northern blot analysis. We thank John Schloss, Anthony Rafalski, and Barbara Mazur for critical reading of the manuscript.

*To whom correspondence should be addressed.

REFERENCES

1. DeFelice, M., Lago, C.T., Squires, C.H. & Calvo, J.M. (1982) Ann. Microbiol. (Paris) 133A:251-256.
2. Eoyang, L., & Silverman, P.M. (1984) J. Bacteriol. 157:184-189.
3. Schloss, J.V., Van Dyk, D.E., Vasta, J.F. & Kutny, R.M. (1985) submitted to Biochemistry.
4. DeFelice, M., Guardiola, J., Esposito, B. & Iaccarino,M. (1974) J. Bacteriol. 120:1068-1077.
5. Squires, C.H., DeFelice, M., Devereux, J & Calvo, J.M. (1983) Nucleic Acids Res. 11:5299-5313.
6. Lawther, R.P., Calhoun, D.H., Adams, C.W., Hauser, C.A., Gray, J. & Hatfield, G.W. (1981) Proc. Natl. Acad. Sci. USA 78:922-925.
7. Friden, P., Newman, T. & Freundlich, M. (1982) Proc. Natl. Acad. Sci. USA 79:6156-6160.
8. Magee, P.T. & de Robichon-Szulmajster, H. (1968a) Eur. J. Biochem. 3:502-506.
9. Petersen, J.G.L., Kielland-Brandt, M.C., Holmberg, S. & Nillson-Tillgren, T. (1983) Carlsberg Res. Commun. 48:21-34.
10. Falco, S.C. & Dumas, K.S. (1985) Genetics 109:21-35.
11. Polaina, J. (1984) Carlsberg Res. Commun. 49:577-584.
12. Magee, P.T. & de Robichon-Szulmajster, H. (1968b) Eur. J. Biochem. 3:507-511.
13. Ryan, E.D. & Kohlhaw, G.B. (1974) J. Bacteriol. 120:631-637.
14. LaRossa R.A. & Schloss, J.V. (1984) J. Biol. Chem. 259:8753-8757.
15. Ray, T.B. (1984) Plant Physiol. 75:827-831.
16. Chaleff, R.S. & Mauvais, J. (1984) Science 224:1443-1445.

17. Shaner, D.L., Anderson, P.C. & Stidham, M.A. (1984a) Plant Physiol. 75S:50.
18. Shaner, D.L., Anderson, P.C. & Stidham, M.A. (1984b) Plant Physiol. 76:545-546.
19. LaRossa, R.A. & Smulski, D.R. (1984) J. Bacteriol. 160:391-394.
20. Messing, J. (1981) Third Cleveland Symposium on Macromolecules: Recombinant DNA, A. Walton, ed., Elsevier, Amsterdam, 143-153.
21. Barnes, W.M., Bevan, M. & Son, P.H. (1983) Methods in Enzymology 101:98-122.
22. Davis, R.W., Botstein, D. & Roth, J.R. (1980) Advanced Bacterial Genetics: A Manual for Genetic Engineering, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
23. Southern, E.M. (1975) J. Mol. Biol. 98:503-517.
24. Hu, N. & Messing, J. (1982) Gene 17:271-277.
25. Sanger, F., Nicklen, S. & Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74:5463-5467.
26. Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. Carlson, M. & Botstein, D. (1982) Cell 28:145-154.
28. Breathnach, R. & Chambon, P.A. (1981) A. Rev. Biochem. 50: 349-383.
29. Donahue, T.F., Daves, R.S., Lucchini, G. & Fink, G.R. (1983) Cell 32:89-98.
30. Hinnebush, A.G. & Fink, G.R. (1983) J. Biol. Chem. 258: 5238-5247.
31. Zaret, K.S. & Sherman, F. (1982) Cell 28:563-573.
32. Bennetzen, J.L. & Hall, B.D. (1982) J. Biol. Chem. 257: 3026-3031.
33. Ikemura, T. (1982) J. Mol. Biol. 158:573-597.
34. Zalkin, H. & Yanofsky, C. (1982) J. Biol. Chem. 257:1491-1500.
35. Lusty, C.J., Widgren, E.E., Broglie, K.E. & Nyunoya, H. (1983) J. Biol. Chem. 258:14466-14472.
36. Sherman, F. & Stewart, J.W. (1982) The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression, Strather,J.N., Jones, E.W. & Broach, J.R., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
37. Kozak, M. (1983) Microbiol. Rev. 47:1-45.
38. Stiles, J.I., Szostak, J.W., Young, A.T., Wu, R., Consaul, S. & Sherman, F. (1981) Cell 25:277-284.
39. Kielland-Brandt, M.C., Holmberg, S., Petersen, J.G.L. & Nillson-Tillgren, T. (1984) Carlsberg Res. Commun. 49:567-575.
40. Kaput, J., Goltz, S. & Blobel, G. (1982) J. Biol. Chem. 257: 15054-15058.
41. Hase, T., Riezman, H., Suda, K. & Schatz, G. (1983) EMBO J. 2:2169-2172.
42. Viebrock, A., Perz, A. & Sebald, W. (1982) EMBO J. 1: 565-571.
43. Krzywicki, K.A. & Brandriss, M.C. (1984) Mol. Cell. Biol. 4: 2837-2842.
44. Lawther, R.P. & Hatfield, G.W. (1980) Proc. Natl. Acad. Sci. USA 77:1862-1866.