
Structure of yeast regulatory gene *LEU3* and evidence that *LEU3* itself is under general amino acid control

Kemin Zhou, Paula R.G.Brisco, Ari E.Hinkkanen¹ and Gunter B.Kohlhaw

Department of Biochemistry, Purdue University, West Lafayette, IN 47907, USA

Received May 8, 1987; Accepted May 29, 1987

Accession no. Y00360

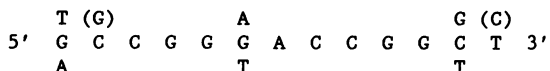
ABSTRACT

Determination of the nucleotide sequence of a DNA region from *Saccharomyces cerevisiae* previously shown to contain the *LEU3* gene revealed one long open reading frame (ORF) whose 887 codons predict the existence of a protein with a molecular mass of 100,162 daltons. The codon bias index of 0.02 suggests that *LEU3* encodes a low-abundance protein. The predicted amino acid sequence contains a stretch of 31 residues near the N-terminus that is rich in cysteines and basic amino acids and shows strong homology to similar regions in five other regulatory proteins of lower eukaryotes. Additional regions with a predominance of basic amino acids are present adjacent to the cysteine-rich region. A stretch of 20 residues, 19 of which are glu or asp, is found in the carboxy terminal quarter of the protein. The 5' flanking region of *LEU3* contains a TATA box 111 bp upstream from the beginning of the long ORF and two transcription initiation elements (5'TCAA3') 58 and 48 bp upstream from the ORF. The 3' flanking region shows a tripartite potential termination-polyadenylation signal. The predicted 5' and 3' ends of the transcript are in very good agreement with the previously determined size of the *LEU3* message. Analysis of a *LEU3'*-*lacZ* translational fusion suggests that the *LEU3* gene, whose product is involved in the specific regulation of the leucine and possibly the isoleucine-valine pathways, is itself under general amino acid control. Consistent with this observation is the finding that the 5' flanking region of *LEU3* contains two perfect copies of the general control target sequence 5'TGACTG3'.

INTRODUCTION

At present, *LEU3* is the only specific genetic element in leucine biosynthesis in yeast known to exert a regulatory function. Strains carrying a spontaneous mutation or partial deletion in *LEU3* grow very sluggishly in the absence of leucine because of greatly diminished expression of the *LEU1* and *LEU2* genes which encode isopropylmalate (IPM) isomerase and β -IPM dehydrogenase, respectively (1,2). The fact that *leu3* mutations are uninducible and are recessive in diploids suggests that the *LEU3* product acts in a positive fashion. The 5' flanking region of *LEU2* has been shown to contain a G+C-rich, palindromic sequence important for the expression of that

gene (3). Similar "leucine-specific" upstream activating sites (UAS_L) with the consensus sequence:



are also present in the 5' flanking regions of *LEU1* (4), *LEU4* (5), *ILV2* (6), and *ILV5* (7). It is possible that these sites represent target sequences for the *LEU3* product, although other possibilities cannot be ruled out at this time.

The *LEU3* gene has recently been cloned and shown to specify an mRNA of 2.9-3.0 kilonucleotides (2). Here we report that *LEU3* is capable of elaborating a 100 kDa protein with features typical of DNA binding proteins. Analysis of an in-frame *LEU3*'-'*lacZ* fusion shows that *LEU3* expression increases when general amino acid control signals are given.

MATERIALS AND METHODS

Strains, Plasmids, and Special Materials

The following strains were used as hosts in transformations: *S. cerevisiae* CG219 (*MAT α ura3-52*) (2), *E. coli* MC1000 (Δ [*lacIPOZYA*] X74, *galU*, *galK*, *rpsL*, Δ [*ara, leu*] (8), and *E. coli* JM101 and JM103 (9). Plasmids pSEY101 and pSEY102 (10) (gifts from S. Emr of CalTech) contain the yeast *URA3* gene, a polylinker region, and either yeast 2 μ m DNA (pSEY101) or yeast *ARS1-CEN4* DNA (pSEY102). DNA to be sequenced was derived from plasmids pGB4 and pTSC36 (2). Cloning vectors M13mpl8 and M13mpl9, universal single-stranded M13 primer (17 bases), and mung bean nuclease were obtained from P-L Biochemicals. DNA polymerase I (Klenow enzyme), T4 DNA ligase, and exonuclease III from *E. coli* B were from Boehringer Mannheim. All other biochemicals, including restriction endonucleases, were purchased from various national suppliers. [α ³²P]dCTP (800 Ci/mmol) and [γ -³²P]ATP (5000 Ci/mmol) were from Amersham Corp.

Growth Conditions

The growth conditions and media were as described elsewhere (2), except that yeast cells were harvested at an OD₅₈₀ of about 0.8. 3-Amino-1,2,4-triazole was added to exponentially growing cells for the final doubling prior to harvest.

DNA Preparation and Treatment

Plasmid DNA isolation, restriction enzyme digestions and fragment isolation procedures were described previously (2). Mung bean nuclease and Klenow enzyme treatments were performed as described (5). Ligations were performed

at 22°C for >12 hours using 1-2 units of T4 DNA ligase per μg DNA. Ligation mixtures for blunt end ligations also contained T4 RNA ligase at a concentration of 1 μg per μg DNA. Synthetic linkers were phosphorylated essentially as described by Maniatis, Fritsch and Sambrook (11) using unlabeled ATP.

Transformation Procedures

Yeast transformations (12) and bacterial transformations (11) were performed essentially as described.

Nucleotide Sequencing

All sequencing was performed by the dideoxy chain termination method (13). Fragments of plasmids pGB4 or pTSC36, generated by restriction enzyme digestion, were inserted into the replicative form of two M13 vectors (9). When fragments were too long, exonuclease III was used to create a series of controlled unidirectional deletions (14). In addition, synthetic oligonucleotides corresponding to known *LEU3* sequences were sometimes used in place of the universal primer.

Assay For β -Galactosidase

Yeast cell-free extracts were prepared as described by Baichwal *et al.* (1), except that the extraction buffer contained 0.1 M TRIS-HCl, pH 8.0, 20% (v/v) glycerol, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, and 0.05% Triton X-100. β -Galactosidase activity was measured by the procedure of Miller (15). Protein was determined by the method of Bradford (16) using bovine serum albumin as a standard. For plate assays, bacteria were put on minimal plates containing ampicillin (50 $\mu\text{g}/\text{ml}$) and X-Gal (5-bromo-4-chloro-3-

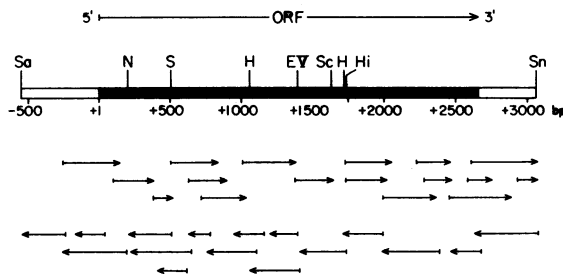


Figure 1. Sequencing strategy and position of long open reading frame.

The long open reading frame is indicated by the top arrow. The short arrows indicate the direction and the extent of sequencing. Arrows pointing to the right, noncoding strand; arrows pointing to the left, coding strand. For more than 95% of the DNA between positions -250 and +3060, sequence was obtained from both strands. The restriction sites shown are those whose presence was confirmed by digestion with the appropriate enzymes. EV, *EcoRV*; H, *HpaI*; Hi, *HindIII*; S, *SalI*; Sa, *Sau3A* (only the most upstream site is shown); Sc, *ScaI*; Sn, *SnaBI*.

Nucleic Acids Research

ATCTCTCTCTG GCGTTCGTCG TCGTCTCCCG ACTGACACTG GAGCAGATTC TAGCGATATA GCTAGATATA TCGCTGTGTC TGTCTTCTCT GGTGACTACT CTCTCTCTCC GCGTCTCTG -442
CGGTCGCTCT GGTAGACAGC CATTTCTTCT ACACGCTATA ACACAGACAA GCGATGATAC TCGCAGACC AGTAGACAA GCGCGCGCGCT GTAGACTCTC ACATATATA TACTCTCTGT -382
AGTTATGATA GTATGACTTT ATTTCTTTGT ATTTCTTACT TATTCTAAT CATCTGATG TATATAGAG TTATGACAA TGTCTCTCTC CTATAGACT CTCTCTCTCT GAGGTGCGC -302
CGTATTTTAT CCGTACAGC TCGCTTTTT CAGATTTTT CAGAGGAAA ATTAGAAGT ATTAGACAA ACATAGAGG ATATGATGTA TATATAGST CATAGACTA GGTTCGAGG -82
ACTATGCGA ATTTAGBAAA CCGTCTCT GCGTACGTA AAGATGCTT CATACGATA ATACAGATTC TTTTCTCAT T
GTA GAA GAA ABA TCA TTT TTT GCG ACT TCA CAG TCC GAA AGT GAA ATG AGC CAT AGT GAA ACT AGG ANT ABA ACT GAG ATG ANT GCT AGG AAA AGG 99
Met Gly Gly Arg Ser Asp Phe Val Ala Thr Ser Gly Ser Gly Ser Gly Met Ser His Ser Gly Thr Arg Arg Arg Thr Gly Met Asn Ala Arg Lys Arg 33
AAA TTC GCG TGT GTS GAA TGT CAT CAG CAG AAG TCG AAA TGT GAT GCT CAC GAA ABA GCA CCG GAG CCA TGC ACT AAG TGT GCT AAA AAG ANT GTC CCA 130
Lys Phe Ala Cys Val Gly Cys Arg Gly Gly Lys Ser Lys Cys Asn Ala His Gly Arg Ala Pro Gly Pro Cys Thr Lys Cys Ala Lys Lys Asn Val Pro 66
TGC ATT TTA AAA CAA GAT TTC ABA AAA ACT TAT AAA ABA GCA AAG AGC GAA GCG ATT GAA AAA ABA TTC AAG GAA CTC ACC ABA ACT TTB ACA ANT TTA 297
Cys Ile Leu Lys Arg Asp Phe Arg Thr Tyr Lys Arg Ala Arg Asn Gly Ala Ile Gly Lys Arg Phe Lys Gly Leu Thr Arg Thr Leu Thr Asn Leu 59
ACT TCG GAT GAA AAT TTA AAG AAA ATT GAA GAG GAA CAG GAA GAT GTT TTB GAT AGC AGT ANT TTC ACA AAA AAA ABA GTA AAA CAA CTC AAG AAG AGT 395
Thr Ser Asp Gly Ile Leu Lys Ala Ile Gly Ile Gly Ile Gly Ile Val Leu Asp Asn Ser Asn Phe Thr Lys Gly Leu Lys Val Lys Gly Asn Ala Arg Lys Arg 132
GCT TTT GAG ACG AAA GAA AYA CAG AAA TCA TAC AAA CTA CTT CAA GAA GAA GCT ATT TET TAC AGT ACC AAC ABA ABA CAT ACG GAT TET TCT CCG 454
Ala Phe Gly Thr Thr Gly Ile Gly Pro Arg Ser Tyr Lys Thr Leu Arg Gly Ala Pro Ile Ser Tyr Ser Thr Asn Arg Arg His Thr Asp Ser Ser Pro 163
TTA ACG CTC TTA ACG TCA TCG ACA AAG TTC GAC CCT GTT CAC TCA ACA AAG GAT TTA ACA GAT GAT CAA CTT AAG TGC TTB CCA AAA AGC CTG GAC GAC 594
Thr Val Asp Gly Ile Leu Lys Ser Ser Asn Phe Thr Asn Phe Asp Pro Val His Ser Thr Asn Val Met Thr Asp Asp Gly Met Lys Lys Cys Leu Pro Lys Ser Leu Val 198
GTA TAT TTB TCA ACG GAT ATT GCT GAG CTG TTT CAA GAA TTT GCG ACA AAA TAT CAT CAA TTT TTA GCG CTC GTT GAC CTT TCG AAA GAA GCA GAG 693
Val Tyr Leu Ser 231
CAA ATT TAT CAC TTA TCT CCT TCC TTA TCT TCG GTC ATC CTG CTT ATT GAT TTA AAG CAG AAA TTT GAG ACC AAG CAA TTA ATC ACT GYA TCA GTA 736
Arg Ile Tyr His Leu Ser Pro Cys Leu Phe Trp Val Ile Leu Leu Ile Gly Leu Arg Arg Lys Phe Gly Ala Thr Asp Leu Met Thr Arg Leu Ser Val 264
CTA GTA AAG TCA GTT TTA TCA GAA ACT ACA TAT TCT CCA ATA ATT CAA TAT ACT CCA TCA GAT AAG AGC GAA CCC GTT CTA ANT GTA GCA TCT GTA TAT 891
Thr Val Asp Val Ser Gly Thr Val Ala Ser Ser Phe Gly Phe Pro Ile His Arg Tyr Thr Val Pro Ser Asp Val Lys Val Lys Val Asn Val Ala Ser Val Lys 297
TCC GTS CAA CAA TTT CTT TTA TAC AGC TTC TCG CCG CCG TTA ACC TET TCA TTA ACG GCG GAC ACT TCG TGG ANT ACC ATA ABA GCA GCG ATG TTC CAA 990
Ser Val Gly Ala Phe Leu Lys Thr Tyr Pro Leu Thr Ser Pro Leu Thr Ser Thr Ser Thr Ser Thr Ser Thr Ser Thr Ser Thr Ser Thr Ser Thr Ser Thr Ser 330
GCG CTT CAG GTA GAA CTA ANT TGT CAA GAT TTT TCA AAA GAG TAT GCT TCG GCA ANT TCA GAA TTA GTT ACC GAG CAA ATA CAA ACT TGG ATT TBC TBC 1089
Ala Leu Arg Val Gly Leu Ser Ser Ser Ser Ala Gly Phe Lys Gly Tyr Ala Ser Ala Asn Ser Gly Leu Val Lys Val Ser Thr Ser Thr Ser Thr Ser Thr Ser Thr Ser Thr Ser Thr Ser 1429
AAT GTT GTA TCT CAA ACA GTT ACA TCA TCA TTT GAT TCC CCA GCT TAT GTT TCA TTT GAT TTT TTA GTA ATC ACG TCT ATT GAA GTA CCA ANT TCA AAA 1188
Asn Val Val Ser Gly Thr Val Ala Ser Ser Phe Gly Phe Pro Ala Tyr Val Ser Phe Asp Tyr Leu Val Ile Ser Ser Ile Arg Val Pro Asn Ser Lys 395
AGC CAA GTA ANT ATA CCG ANT GAA TCA ABA CAA ATG GCT CAA ATT GCT ABA TTT GAG AAG CAA GTT GTA ACG ACA ATG AAG TCC ACC CCG GCG AGT GTT 1287
Ser Gln Val Asp Phe Pro Asn Gly Ser 1459
ACT GAG ATG GTA AGT CAG GAA AAG AAG CAG CCG TTB TTA CAC GCT CTT ANT CAA CTA AGT CAA TTB GAG AGT AGT CTT GAA GAA ANT AAG CTA GAT 1386
Thr Gly Met Val Ser 462
GAT ACT CAA AAA TTT TTA TTA CTA GTS GCG AAG GTT CAC TTA TCA ACC TAT TCA TTC ACT GAT GTT ACC TCC CCA GAT GCT GAA TCA ANT GAT ANT 1465
Arg Ile Arg Lys Phe Leu Leu Val Ala Lys Val His Leu Leu Thr Tyr Tyr Phe Thr Asp Val Thr Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser 495
ATT TAT GAG GCA TCA TAT TCC ATT ATG GAA CTC GAT ACA AGT TTT GAA ACG GAT GAA TTB GTS AAA GTT TAT ANT GCG GCT GTA ACA TTT CTT ATA 1584
Ile Tyr Gly Gly Ser Thr 528
CAT GCG ANT AGT ATG TGG GAA CAT GAT CCG ACC ATT ATT AAG TAC TTT CCG AGT TTB TTT GTC TTB ANT ATA TGG CAA TCT GCG TST GAT ATT AGT AAA 1643
His Ala Asn Ser Thr Thr Gly His Asp Pro Thr Thr Ile Lys Tyr Phe Pro Thr Gly Leu Phe Val Thr Gly Ser Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr 561
CTC ATA CAT TCA TCA CTT CAT ATG CTA GAT GTT AAG TCA GCG AAA AAA GCT TAT AAG AAG ACA ANT TCA TTB ACG TTT ANT GCG TCA GTT TTA AAA 1782
Leu Ile His Ser Ser Leu His Ser Ser Met Leu Asp Val Asn Ser Gly Lys Phe Ala Tyr Asn Asn Ala Ile Ser Leu Thr Phe Asn Ala Ser Thr Asn Leu Thr Lys 594
TAT GAT ATG GCG TAC GAA TCA TCC GAA ATA ATG CAA AGC ATA TGG AGT TTA TTT GCT ANT ATG TAT GAT GCG TGG AAA AAG CAA AAG GAA GAT GAA 1881
Tyr Asp Met Ala Tyr Arg Ser Ser Gly Ile Met Arg Ser Ile Trp Ser Leu Phe Ala Asn Met Tyr Asp Ala Trp Lys Asn Asp Ile Lys Gly Leu Lys Cys 660
GAT ABA CTA ANT GAT TTC ANT TTA GCG ATC ACC ATA AAA TCT AAG ATG TCA STA AAG GAT TTT TTT GAG TGC TTA TAT ATT CTA AAA GAG AAA TST 1980
Gly Arg Leu Asn Asn Asp Phe Asn Leu Gly Ile Thr Ile Lys Ser Arg Met Ser Val Asn Val Phe Phe Asp Cys Leu Tyr Ile Leu Lys Gly Lys Cys 660
GAT ATG GCG AAA TTB GAG ABA GAG ACC AAG GTT TCT ACA GCT TAC ANT GTT GAT GAA GAG GAA GAG GAA GAT GAA GAT GAG GAG GAA GAA GAA GAA 2079
Gly Met Ala Lys Leu Gly Arg Gly Thr Lys Val Ser Thr Ala Tyr Asn Val Asp Gly Gly Gly Gly Gly Asp Gly Asp Gly Gly Gly Gly Gly Gly Gly 693
GAA GAA GAA CTA ANT ANT ABA GTT CCA GAA ANT ATG GAT AAG CAG CAA CTA AAG ACA AAG AAA TTC ACC ANT GTA AAG CAT CCA GAA AAG AAA GCA 2178
Gly 725
AAG AAA ATA ATT GAA ACA ATT CCG CTA ACG CCA ANT CCA ATA ANT GCA GCG TCT ACC AAG ANT GAA AAG TCA TTA ACG ACC CCA ANT ANT CAA GTA GCG 2277
Arg Lys Ile Ile Gly Thr Ile Pro Leu Asp Thr Asn Pro Ile Asn Ala Gly Ser Thr Ser Ser Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser 759
AGC ACT ATA TCA TAT ABA GAA ATC ATC ANT AAA ATG TCA CCG AAG GAA CAA CTA ANT CAT GCA ANT TTA ANT TCC AGT GTT TCT ACA AGC ATC AAG AAG 2376
Asn Thr Ile Ser Tyr Arg Gly Ile Leu Asn Lys Met Ser Pro Arg Gly Ile Asn Leu Asn His Ala Phe Leu Asp Asp Ser Ser Val Ser Ser Thr Ile Lys Asp 792
ACT GAA GCT GTC ANT GAA GCT CTG CCA ATA GAG AAG ANT GCT GAA CAT CCG GCA ANT CAA CCG CCG CTT CTA TTA ACA CAA ATG CAA AAA AAG ACA CTA 2475
Gly Ala Val Asn GCG ACC ANT GCG Pro Thr Gly Arg Asn Ala His Pro Asn Ala Gly Ser Thr Ser Ser Pro Leu Thr Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser 825
CGT GCG ACA CAA GCG TCT TCT CTA TTA AAA GCG TAT CCG ANT GTT CAA TCA AAG CCG GTT ACC ACT ACA ATC ABA AAA TCA CCG ANT TCC ATC ATG 2574
Pro Ala Thr Gly Ala Asn Ser Ser Ser Leu Leu Gly Thr Tyr Pro Ile Val Gly Ser Asn Pro Val Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr 858
ACA GAT TGG GAT AAG TGG GAA TCT GAT ATG GTT TGG AAG GAT GTT GAT ATT TTA ATG ANT GAA TTT GCG TCC ANT CCG AAG GTT TAA 2661
Ala Gly Tyr Asp Asn Trp Gly Ser Asp Met Val Trp Arg Asp Ile Leu Met Asn Gly Ile Phe Leu Asp Asn Ser Val Ser Thr Thr Thr Thr Thr Thr Thr Thr Thr 886
AGTCTTTCT TTTTCTTCTG TATTTCTC TTTCTCTCA AAGATTTCTA TATATGACT GCGTCTCTC ABAAGATTA ATATTTTCT TATATGCTT TATATTTCT GATTTCTC 2781
AGCTATTTT TTTCTCTCA ATATTTTCTA TATATGACT TATATGCTT GCGTCTCTC ABAAGATTA ATATTTTCT TATATGCTT TATATTTCT GATTTCTC 2901
AAGCAGCAGC CCGCTCTC CCGAGATTA ATATTTTCTA TATATGACT TATATGCTT GCGTCTCTC ABAAGATTA ATATTTTCT TATATGCTT TATATTTCT GATTTCTC 3021
TATATGCTT AAGCAGC CCGCTCTC TATATGACT TA 3063

indolyl- β -D-galactoside; Research Organics Inc., Cleveland, OH) (40 μ g/ml). Yeast cells were spotted onto minimal (SD) plates containing X-Gal (17).

RESULTS AND DISCUSSION

Nucleotide and Deduced Amino Acid Sequence of LEU3

Figure 1 shows a partial restriction map of the *LEU3* region and the sequencing strategy. The nucleotide sequence between restriction sites *Sau3A* (left end of clone) and *SnaBI* is shown in Figure 2. A computer-aided search of all six reading frames revealed only one extended open reading frame (ORF). Its 886 amino acid codons are indicated by the shaded region in Figure 1. The next-longest ORF contains 90 codons. Identification of the long ORF as that of *LEU3* is based on the following arguments: (i) Earlier subcloning experiments showed that of three DNA fragments, all of which extended beyond the *SnaBI* site, only the one that included the *Sau3A* site at position -561 was able to complement *leu3* mutations; two others, starting at the *NsiI* and the *SalI* site, respectively, were unable to do so (2). (ii) Deletion of the 0.66 kb *HpaI-HpaI* fragment from the *LEU3* gene destroyed *LEU3* function (2). (iii) The size of the long ORF (2.658 kb) agrees well with the size of the *LEU3* message, previously determined to be 2.9-3.0 kilonucleotides (2).

The deduced amino acid sequence of the *LEU3* protein yields a calculated molecular weight of 100,162. Codon usage is shown in Table 1. The codon bias index is 0.02, implying that the *LEU3* product is a low-abundance protein (18).

Special Features of the Predicted LEU3 Protein

Inspection of the deduced amino acid sequence reveals several prominent features. One of these extends from residue 37 to residue 67. This segment of 31 amino acids contains six cysteines, is rich in basic residues, and shows a remarkable homology to five other lower-eukaryotic regulatory proteins (Figure 3). In addition to the six cysteines, one arginine, one lysine, and one proline residue are conserved in all six proteins; two more lysine residues are present in comparable positions in five of the six proteins; and

Figure 2. Nucleotide sequence of the LEU3 gene and flanking regions (noncoding strand).

The sequence covers the region shown in Fig. 1. Nucleotides upstream from the proposed translation start at +1 carry negative numbers, those downstream carry positive numbers. The deduced amino acid sequence of the long ORF is shown below the nucleotide sequence. The following features have been highlighted by wavy lines: Two general amino acid control boxes, a potential Goldberg-Hogness box, two sequences potentially signalling transcription starts (all in the 5' noncoding region), and a potential tripartite transcription termination signal (3' noncoding region). A potential DNA binding motif and an acidic amino acid cluster have been underlined (coding region). See text for further details.

Table 1. Codon usage in *LEU3*.

TTT Phe	21 (2.4)	TCT Ser	17 (1.9)	TAT Tyr	19 (2.1)	TGT Cys	7 (0.8)
TTC Phe	13 (1.5)	TCC Ser	8 (0.9)	TAC Tyr	7 (0.8)	TGC Cys	7 (0.8)
TTA Leu	29 (3.3)	TCA Ser	32 (3.6)	TAA End	1 (0.1)	TGA End	0 (0.0)
TTG Leu	12 (1.4)	TCG Ser	7 (0.8)	TAG End	0 (0.0)	TGG Trp	11 (1.2)
CTT Leu	10 (1.1)	CCT Pro	11 (1.2)	GAT His	10 (1.1)	GGT Arg	2 (0.2)
CTC Leu	7 (0.8)	CCC Pro	9 (1.0)	CAC His	5 (0.6)	CGC Arg	0 (0.0)
CTA Leu	15 (1.7)	CCA Pro	13 (1.5)	CAA Gln	26 (2.9)	GGA Arg	8 (0.9)
CTG Leu	5 (0.6)	CCG Pro	6 (0.7)	CAG Gln	7 (0.8)	CGG Arg	2 (0.2)
ATT Ile	26 (2.9)	ACT Thr	16 (1.8)	AAT Asn	39 (4.4)	AGT Ser	19 (2.1)
ATC Ile	10 (1.1)	ACC Thr	13 (1.5)	AAC Asn	21 (2.4)	AGC Ser	12 (1.4)
ATA Ile	17 (1.9)	ACA Thr	23 (2.6)	AAA Lys	37 (4.2)	AGA Ser	19 (2.1)
ATG Met	23 (2.6)	ACG Thr	8 (0.9)	AAG Lys	17 (1.9)	AGG Arg	13 (1.5)
GTT Val	27 (3.0)	GCT Ala	16 (1.8)	GAT Asp	30 (3.4)	GGT Gly	9 (1.0)
GTC Val	5 (0.6)	GCC Ala	12 (1.4)	GAC Asp	11 (1.2)	GGC Gly	4 (0.5)
GTA Val	16 (1.8)	GCA Ala	14 (1.6)	GAA Glu	52 (5.9)	GGA Gly	14 (1.6)
GTG Val	6 (0.7)	GCG Ala	9 (1.0)	GAG Glu	18 (2.0)	GGG Gly	4 (0.5)

The numbers in parentheses are percentages.

one aspartate and one asparagine residue each is conserved in four of the six proteins. The cysteine/basic amino acid-rich regions of *GAL4* and *qa-1F* have been shown to be part of the DNA binding domain (19,20); it is likely that the corresponding regions of the other four proteins have the same function.

It has been pointed out that the cysteine-containing region of the *GAL4* and *PPRI* products resembles the putative DNA binding regions in TFIIIA of *Xenopus* as well as proteins encoded by the *Krüppel* and *Serendipity* genes of *Drosophila*, and the yeast regulatory protein encoded by *ADRI*, and might conform to the zinc-binding "finger" motif found in those proteins (21). How-

<i>LEU3</i> (S.c.)	37	C	V	E	C	R	Q	Q	R	S	K	C	D	A	H	E	R	A	P	E	P	C	T	K	C	A	K	K	N	V	P	C
<i>PPRI</i> (S.c.)	34	C	K	R	C	R	L	K	K	I	K	C	D	Q	-	E	F	-	P	-	S	C	K	R	C	A	K	L	E	V	P	C
<i>ARGRII</i> (S.c.)	21	C	N	T	C	R	G	R	K	V	K	C	D	L	-	R	H	-	P	-	H	G	Q	R	C	E	K	S	N	L	P	C
<i>GAL4</i> (S.c.)	11	C	D	I	C	R	L	K	K	L	K	C	S	K	-	E	K	-	P	-	K	C	A	K	C	L	K	N	N	M	E	C
<i>LAC9</i> (K.L.)	95	C	D	A	C	R	K	K	K	W	K	C	S	K	-	T	V	-	P	-	T	C	T	M	C	L	K	Y	N	L	D	C
<i>qa-1F</i> (N.c.)	76	C	D	Q	C	R	A	A	R	E	K	C	D	G	-	I	Q	-	P	-	A	C	F	P	C	V	S	Q	G	R	S	C

S.c. = *SACCHAROMYCES CEREVISIAE*; K.L. = *KLUYVEROMYCES LACTIS*; N.c. = *NEUROSPORA CRASSA*

Figure 3. Comparison of potential DNA binding regions in six lower-eukaryotic regulatory proteins.

The numbers preceding the sequences indicate the distance (in residues) from the N-terminus. The hyphens designate spaces introduced for maximum homology. Sequence data are from the following references: *PPRI* (40), *ARGRII* (41), *GAL4* (19), *LAC9* (42), *qa-1F* (20).

ever, this interpretation ignores the potential significance of the highly conserved third and sixth cysteines of the region shown in Figure 3, as well as other differences. It is more likely that the sequences listed represent a variation of the finger motif. Metal binding has yet to be demonstrated for any of these proteins. The sequence present in *LEU3* is unique even among the lower eukaryotic examples, since its middle region contains three additional amino acids, including one histidine. Participation of this histidine in the motif would create a very regular Cys-X₂-Cys-X₆-Cys-X₂-His-X₆-Cys-X₂-Cys-X₆-Cys structure. The importance of cysteine/histidine-rich domains in nucleic acid binding proteins is emphasized by the fact that at least six classes of such proteins are now known, including *gag* encoded proteins of retroviruses, amino acyl tRNA synthetases, and steroid hormone and vitamin D receptors (22,23).

Another notable feature of the *LEU3* protein is the uneven distribution of charged amino acids. Thus, an accumulation of lysine and arginine residues is found between amino acid positions 31 and 94 and again between positions 123 and 148. In both of these regions, the basic amino acids constitute about one third of the total number of amino acids present, as opposed to 8.4% for the remainder of the protein. An extraordinary accumulation of acidic amino acids occurs between positions 678 and 697, in the carboxy terminal quarter of the protein. In this stretch of 20 amino acid residues, one glycine is surrounded by 16 glutamates and 3 aspartates. While similar sequences exist in a number of other proteins, there appears to be no common denominator with respect to their function. Among the proteins containing a cluster of acidic amino acids are yeast ubiquinol-cytochrome c reductase (24), a pig neurofilament protein (25), the major capsid protein of adenovirus 2 (26), homeotic proteins (27), bovine non-histone nucleosomal proteins HMG1 and HMG2 (28), and frog nucleoplasmin, a histone binding protein (29). The last three examples are of obvious interest with respect to potential functions of the *LEU3* product.

Features of the 5' and 3' Flanking Regions

The promoter region of *LEU3* contains both near-upstream "selector" and potential far-upstream "modulator" elements. A good example of a Goldberg-Hogness (TATA) box is present at position -111. This is followed by two 5'TCAA3' sequences approximately 50 and 60 bp further downstream (at positions -58 and -48, respectively). The TCAA element was recently recognized as one of two preferred sequences in yeast that usually signal transcription initiation when present 50-120 bp downstream from the TATA box (30). The first ATG downstream from the putative transcription initiation sites is located at position +1. This is followed by two additional in-frame ATG's at

positions +49 and +82. A decision as to whether any of the latter two ATG's could be utilized as a translation start will have to await experimental determination of the transcription start point(s).

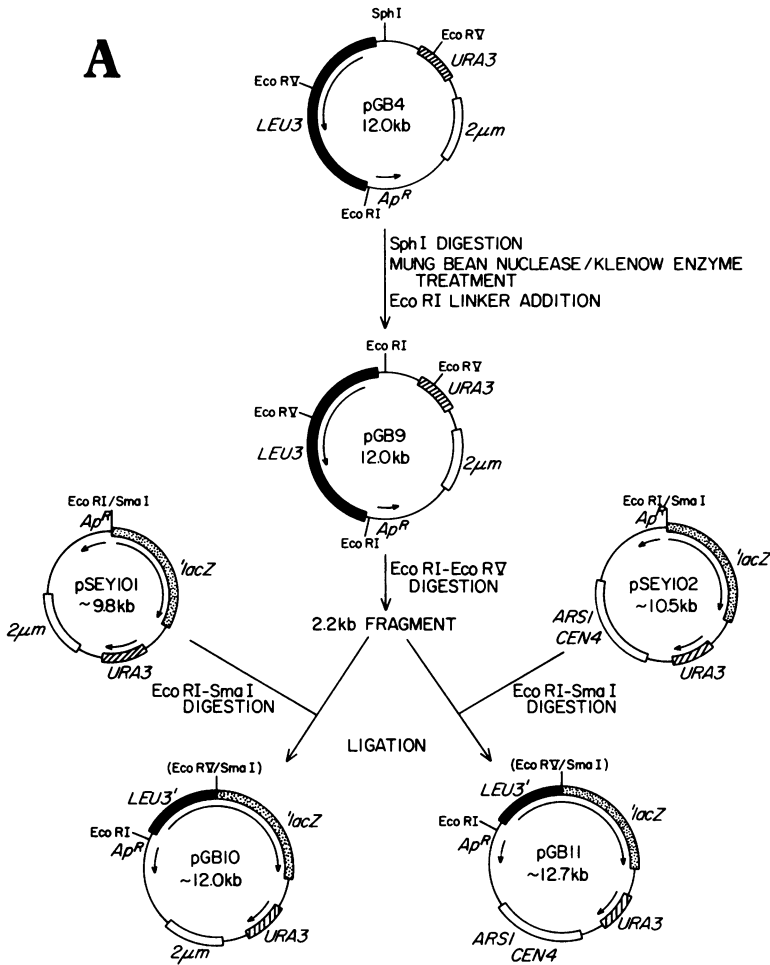
Two hexanucleotide sequences with a perfect homology to the general amino acid control box 5'TGACTC3' are present at positions -417 and -241. The sequence TGACTC appears in at least one copy in the 5' flanking region of all genes that are subject to the cross-pathway regulation termed general control of amino acid biosynthesis (reviewed in 31). It is part of the target sequence for positive control by the *GCN4*-encoded protein. *GCN4* itself is the last, most direct-acting element in a hierarchy of regulatory genes that respond to amino acid starvation. It has been shown that binding of the *GCN4* protein extends to bases on either side of TGACTC (32,33). The environment of the *LEU3* general control boxes would predict intermediate affinity for the *GCN4* protein (32).

A sequence somewhat akin to the UAS_L consensus sequence (see INTRODUCTION) is present between positions -442 and -429. However, this sequence contains two additional bases in a region (5'ACCGG3') that is perfectly conserved in the five known examples of UAS_L sequences, and its significance is therefore unclear.

The likely translational stop codon of the *LEU3* message (UAA at position 2659) is followed within the next 90 nucleotides by five additional stop codons in all three reading frames. A tripartite termination-polyadenylation signal of the kind proposed by Zaret and Sherman (34) is found between positions 2840 and 2867 (TAG...TATGT..[A+T rich]..TCTT). The proposed positions of transcription start and transcription termination agree very well with the approximate length of the *LEU3* encoded message (2.9-3.0 kilonucleotides) determined earlier (2). In combination with the absence of the sequence 5'TACTAAC3' (35,36), these results also suggest that introns are not present within the *LEU3* gene.

Construction of *LEU3*'-'*lacZ* Fusions

Utilizing the unique *EcoRV* restriction site between nucleotide positions 1387 and 1392, a translational fusion to the *E. coli lacZ* gene was constructed for the dual purpose of proving the existence of a long ORF and of being able to study regulation of *LEU3* expression at the protein level. Plasmid pGB4 (2) which contains the *LEU3* region (including the *LEU3* promoter) served as starting material (Figure 4A). Its unique *SphI* site was changed to an *EcoRI* site by *SphI* digestion, mung bean nuclease and Klenow enzyme treatment, and the addition of *EcoRI* linkers. The plasmid was subsequently digested with



B

		from			
LEU3		polylinker		lacZ	
... GAT GAT	GGG	GAT	CCC	GTC	GTT ...
Asp Asp	Gly	Asp	Pro	Val	Val
462 463				9	10

Figure 4. Construction of LEU3'-lacZ translational fusion plasmids. **A**, details of construction of in-frame fusion that contains 561 bp of the LEU3 5' flanking region, the first 463 LEU3 codons, and a truncated lacZ gene lacking its promoter and the first eight codons. Blocks represent yeast DNA, thin lines, bacterial DNA. Arrows within the plasmid circles indicate direction of transcription. See text for additional information. **B**, details of the fusion region.

Table 2. β -Galactosidase activities of a *LEU3'*-'*lacZ* fusion protein in cell-free extracts.

Strain/plasmid ¹⁾	Additions to minimal media	Specific activity (nmoles/min x mg)	-fold change ⁴⁾
CG219	Uracil, 0.2 mM	N.D. ²⁾	
CG219/pSEY101	None	N.D.	
CG219/pSEY102	None	N.D.	
CG219/pGB10	None	3.03 \pm 0.06 ³⁾	30.3
CG219/pGB11	None	0.10 \pm 0.02	1.0
	3-AT ²⁾ , 10 mM	0.31 \pm 0.02	3.1
	Leucine, 2 mM	0.21 \pm 0.03	2.1
	Leucine, 2 mM, plus valine, isoleucine, 1 mM	0.08 \pm 0.02	0.8

¹⁾ See MATERIALS AND METHODS for description.

²⁾ N.D. = not detectable; 3-AT = 3-amino-1,2,4-triazole.

³⁾ Experimental error is given as standard deviation ($n \geq 6$).

⁴⁾ Relative to the β -galactosidase level in CG219/pGB11 cells grown with no addition.

EcoRI and *EcoRV*. Among the fragments thus generated was one, 2.2 kb in length, that contained the *LEU3* promoter and *LEU3* ORF sequence up to the *EcoRV* site. This fragment was ligated to plasmids pSEY101 and pSEY102 (described in MATERIALS AND METHODS), both of which had been digested with *EcoRI* and *SmaI*. This strategy forced the insertion of the 2.2 kb fragment to occur in the proper orientation. The resulting new plasmids were designated pGB10 and pGB11. That the fusion had occurred at the desired point (Figure 4B) was confirmed by diagnostic digestion with restriction enzymes *HpaI* and *SaII* (one *HpaI* and one *SaII* site each are carried in by the *LEU3* fragment), and by nucleotide sequencing.

LEU3'-'*lacZ* Expression is Under General Control of Amino Acid Biosynthesis

β -Galactosidase activity was absent from yeast strains containing the parent plasmids pSEY101 or pSEY102, but was present in strains containing the *LEU3'*-'*lacZ* fusions (Table 2). This result demonstrates that a long ORF indeed exists in *LEU3*, at least to the *EcoRV* site. The strain carrying the 2 μ m DNA-containing plasmid pGB10 showed about 30 times as much β -galactosidase activity as the strain carrying the *CEN4*-containing plasmid pGB11. This probably reflects copy number differences. For the study of *LEU3* regulation, we concentrated on a strain carrying plasmid pGB11. Centromere-containing

plasmids are known to be stably maintained in a copy number that is very close to one (37, and S. Emr, personal communication). Extracts from strain CG219/pGB11 grown in minimal medium showed low but statistically significant fusion enzyme levels (Table 2). When this strain was grown under two conditions known to elicit a general control derepression signal, i.e., in the presence of 3-amino-1,2,4-triazole (3-AT) or in the presence of excess leucine, the β -galactosidase levels increased 3.1-fold and 2.1-fold, respectively. Growing strain CG219/pGB11 in the presence of leucine, isoleucine and valine resulted in essentially unchanged β -galactosidase levels, compared to cells grown with no additions. The histidine analog 3-AT acts by causing a histidine deficiency (38). Leucine is one of several amino acids which, when present by themselves at elevated concentrations, cause amino acid imbalance, resulting in "derepression factors" of 1.2-2.0 (39). The leucine effect disappears when isoleucine and valine are also present (39). It is evident from the results shown in Table 2 that expression of the *LEU3*'-'*lacZ* fusion is increased by general control stimuli to an extent typical for this system. Qualitatively similar results were seen with a strain harboring the multicopy plasmid pGB10 (data not shown). The conclusion that *LEU3* is under general amino acid control is corroborated by the presence of two 5'TGACTG3' boxes in the promoter region of this gene. To our knowledge, this is the first known example of general control being exerted on a regulatory, as opposed to a structural gene, and thus of an intertwining of general and specific controls. Stimulation of the production of a positive regulator by the general control system might reinforce the overall upward trend of amino acid biosynthesis in situations of amino acid imbalance. However, it is not clear at present exactly where and how an increase in *LEU3* expression would manifest itself. It was observed previously that cells harboring multicopy *LEU3*-containing plasmids exhibited elevated levels of *LEU3* message, but did not show significant changes in the levels of IPM isomerase or β -IPM dehydrogenase (2). It is possible that other potential targets of *LEU3* action (*LEU4*, *ILV2*, *ILV5*) are more sensitive to changes in *LEU3* expression. Also, the existence of additional controls of *LEU3* expression cannot be excluded at this point. Experiments to study these questions are underway.

ACKNOWLEDGEMENTS

This work was supported by research grant GM15102 from the National Institutes of Health. This is Journal Paper No. 11151 of the Agricultural Experiment Station, Purdue University.

¹Present address: Max-Planck-Institut für Immunbiologie, Stübeweg 51, 7800 Freiburg-Zähringen, FRG

REFERENCES

1. Baichwal, V.R., Cunningham, T.S., Gatzek, P.R., and Kohlhaw, G.B. (1983) *Curr. Genet.* 7, 369-377.
2. Brisco, P.R.G., Cunningham, T.S., and Kohlhaw, G.B. (1987) *Genetics* 115, 91-99.
3. Martinez-Arias, A., Yost, H.J., and Casadaban, M.J. (1984) *Nature* 307, 740-742.
4. Hsu, Y.-P. and Schimmel, P.R. (1984) *J. Biol. Chem.* 259, 3714-3719.
5. Beltzer, J.P., Chang, L.L., Hinkkanen, A.E., and Kohlhaw, G.B. (1986) *J. Biol. Chem.* 261, 5160-5167.
6. Falco, S.C., Dumas, K.S., and Livak, K.J. (1985) *Nucl. Acids Res.* 13, 4011-4027.
7. Litske Petersen, J.G. and Holmberg, S. (1986) *Nucl. Acids Res.* 14, 9631-9651.
8. Casadaban, M.J., Martinez-Arias, A., Shapira, S.K., and Chou, J. (1983) *Methods Enzymol.* 100, 293-308.
9. Messing, J. (1983) *Methods Enzymol.* 101, 20-78.
10. Emr, S.D., Vassarotti, A., Garrett, J., Geller, B.L., Takeda, M., and Douglas, M.G. (1986) *J. Cell Biol.* 102, 523-533.
11. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1983) In: *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
12. Beggs, J.D. (1978) *Nature* 275, 104-108.
13. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
14. Henikoff, S. (1984) *Gene* 28, 351-359.
15. Miller, J.H. (1972) In: *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. Bradford, M.M. (1976) *Anal. Biochem.* 72, 248-254.
17. Rose, M., Casadaban, M.J., and Botstein, D. (1981) *Proc. Natl. Acad. Sci. USA* 78, 2460-2464.
18. Bennetzen, J.L., and Hall, B.D. (1982) *J. Biol. Chem.* 257, 3026-3031.
19. Keegan, L., Gill, G., and Ptashne, M. (1986) *Science* 231, 699-704.
20. Baum, J.A., Geever, R., and Giles, N.H. (1987) *Mol. Cell. Biol.* 7, 1256-1266.
21. Hartshorne, T.A., Blumberg, H., and Young, E.T. (1986) *Nature* 320, 283-287.
22. Berg, J.M. (1986) *Science* 232, 485-487.
23. McDonnell, D.P., Mangelsdorf, D.J., Pike, J.W., Haussler, M.R., and O'Malley, B.W. (1987) *Science* 235, 1214-1217.
24. VanLoon, A., DeGroot, R.J., DeHaan, M., Dekker, A., and Grivell, L.A. (1984) *EMBO J.* 3, 1039-1043.
25. Geislor, N., Fischer, S., VandeKerckhove, J., Plessmann, U., and Weber, K. (1984) *EMBO J.* 3, 2701-2706.
26. Jörnvall, H., Akusjärvi, G., Aleström, P., Bahr-Lindström, H., Pettersson, U., Appella, E., Fowler, A., and Philipson, L. (1981) *J. Biol. Chem.* 256, 6181-6186.
27. Colberg-Poley, A.M., Voss, S.D., Chowdhry, K., and Gruss, P. (1985) *Nature* 314, 713-718.
28. Walker, J.M., Gooderham, K., Hastings, J., Mayes, E., and Johns, E.W. (1980) *FEBS Lett.* 122, 264-270.
29. Dingwall, C., Dilworth, S.M., Black, S.J., Kearsley, S.E., Cox, L.S., and Laskey, R.A. (1987) *EMBO J.* 6, 69-74.

30. Hahn, S., Hoar, E.T., and Guarente, L. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8562-8566.
31. Hinnebusch, A.G. (1986) *CRC Crit. Rev. Biochem.* **21**, 277-317.
32. Arndt, K. and Fink, G.R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8516-8520.
33. Hope, I.A. and Struhl, K. (1985) *Cell* **43**, 177-188.
34. Zaret, K.S. and Sherman, F. (1982) *Cell* **28**, 563-573.
35. Pikielny, C.W., Teem, J.L., and Rosbash, M. (1983) *Cell* **34**, 395-403.
36. Langford, C.J. and Gallwitz, D. (1983) *Cell* **33**, 519-527.
37. Gunge, N. (1983) *Ann. Rev. Microbiol.* **37**, 253-276.
38. Wolfner, M., Yep, D., Messenguy, F., and Fink, G.R. (1975) *J. Mol. Biol.* **96**, 273-290.
39. Niederberger, P., Miozzari, G., and Hütter, R. (1981) *Mol. Cell. Biol.* **1**, 584-593.
40. Kammerer, B., Guyonvarch, A., and Hubert, J.C. (1984) *J. Mol. Biol.* **180**, 239-250.
41. Messenguy, F., Dubois, E., and Descamps, F. (1986) *Eur. J. Biochem.* **157**, 77-81.
42. Wray, L.V., Witte, M.M., Dickson, R.C., and Riley, M.I. (1987) *Mol. Cell. Biol.* **7**, 1111-1121.