Nucleotide sequence of the yeast regulatory gene *GAL80*

Yasuhisa Nogi and Toshio Fukasawa

Laboratory of Molecular Genetics, Keio University School of Medicine, Shinjuku-ku, Tokyo 160, Japan

Received 26 October 1984; Accepted 20 November 1984

ABSTRACT

The *GAL80* gene in *Saccharomyces cerevisiae* encodes a negative regulatory protein for the set of inducible genes involving metabolism of galactose and melibiose. We have determined the nucleotide sequence of *GAL80* and its flanking regions and assigned the 5' end of its mRNA to the sequence. The deduced coding sequence for *GAL80* protein contains 1305 nucleotides and the calculated molecular weight of the peptide chain is 48309. The 5' end of the *GAL80* mRNA maps about 67 nucleotides upstream from the translation initiating ATG. We have also determined the nucleotide sequence of uninducible alleles *GAL80*^{S-o}, *GAL80*^{S-1} and *GAL80*^{S-2}, and found single base substitution in each of these mutant genes which would lead to alteration of amino acid in *GAL80* protein.

INTRODUCTION

In *Saccharomyces cerevisiae*, expression of the genes of *GAL1*, *GAL7*, *GAL10* and *MEL1* respectively encoding galactokinase, galactose 1-phosphate uridylyl transferase, UDP-glucose 4-epimerase and α -galactosidase is co-ordinately regulated at the transcriptional level by the functional interplay of two regulatory proteins encoded by the genes *GAL4* and *GAL80* (1-10). The current model (11) assumes that the *GAL4* protein is a positive regulator for the transcription of the structural genes whose function is hampered in the absence of inducer by the *GAL80* protein, a negative regulator. Induction would involve inactivation of *GAL80* protein, and *GAL4* protein thus liberated from the *GAL80* action would in turn promote the transcription of the target genes. A question remains to be answered: Whether does *GAL80* protein inactivate *GAL4* protein by the direct interaction or does it bind the target genes to interfere with the *GAL4* function (see ref. 10, 12, 13)?

Several groups cloned the *GAL4* gene (12, 14, 15), and Laughon and Gesteland (16) determined its nucleotides sequence. We have recently cloned the *GAL80* gene and also three types of uninducible mutant alleles, namely *GAL80*^{S-o} (originally designated *GAL80*^S in ref. 4), *GAL80*^{S-1} and *GAL80*^{S-2} (12, 13). In the present paper, we describe the complete sequence of the *GAL80*

gene together with the coding sequence of the mutant alleles. In the light of these results, we postulate probable existence of at least two functional domains in GAL80 protein.

MATERIALS AND METHODS

Plasmid and phage vectors

Plasmid pKK8005 contained the GAL80 gene in a 5.1 kilobase (kb) Sau3A fragment of yeast DNA inserted at BamHI site in YEp24 (13). Plasmid pTF8005 (17) used as a source of DNA in the sequence determination was constructed by cutting pKK8005 DNA with BamHI and SalI, and ligating the resultant 5.4 kb of fragment in between BamHI and SalI sites of pBR322. The phage vectors used for sequencing were M13mpl0 and M13mpl1 (18). The plasmid DNA and the replicative form of phage vectors were prepared by the cleared-lysate procedure (19) and purified by the ethidium bromide/CsCl equilibrium centrifugation.

Strains and media

Escherichia coli (E. coli) strain JA221 (recAL, leuB6, trp Δ E6, hsdR, lacY) harboring the plasmid pTF8005 was grown on L-broth containing 25 μ g/ml ampicillin. The E. coli strain JM105 (Δ lacpro, thi, strA, endA, sbcB15, hsdR4, F'traD36, proAB lacI^q Δ M15) used for propagation of phage vectors was grown on YT-broth. Saccharomyces cerevisiae (S. cerevisiae) strain KK4 transformed with pKK8005 was grown on YPGal medium and used as source of polyadenylated RNA (poly⁺RNA).

YT-broth contained 0.8% Bacto-trypton, 0.5% Yeast Extract and 0.5% NaCl. Other media were described previously (12).

DNA sequence determination

Sequencing was carried out by the combination of three strategies. 1. The 3.1 kb DNA fragment containing the GAL80 locus was prepared by digesting pTF8005 DNA with HindIII and eluting the desired band from low-gelling temperature agarose gel after electrophoretic separation. The isolated HindIII fragment was cleaved respectively with EcoRI, BglII, SmaI, SphI, AluI, HpaII, HinfI and TaqI. The resultant fragments were fractionated in either agarose or polyacrylamide gel electrophoresis and inserted into the polylinker region of M13mpl0 or M13mpl1. Where necessary, recessed ends were filled in with E. coli DNA polymerase I Klenow fragment in presence of dNTPs or rendered flush with S1 nuclease. Transformation of JM105 with phage DNA and preparation of single-stranded DNA templates from individual plaques were performed as described by Messig (18). The DNA sequence was determined by the dideoxynucleotide chain termination method of Sanger *et al.* (20). 2. The pTF8005 DNA was cut at the unique XhoI site locating in the vicinity of the left-hand

HindIII (Fig. 1) and digested progressively with Bal31 exonuclease. The digestion was performed as described by Guo *et al.*, (21). A series of the shortened plasmid DNA was cut with EcoRI and separated in agarose gel. The desired fragments which have EcoRI site in the GAL80 locus at one end and blunt terminus produced by the Bal31 digestion at the other end were subcloned in between EcoRI and SmaI sites of M13mp10. The DNA sequence was determined starting from the SmaI site by the dideoxy method. 3. The pTF8005 DNA was digested respectively with BglII, EcoRI, AvaI, HindIII or ClaI. The fragments were labeled either at their 5' ends with [γ - 32 P]ATP and T4 polynucleotide kinase or at their 3' ends with appropriate [α - 32 P]dNTP and *E. coli* DNA polymerase I Klenow fragment. Singly end-labeled fragments obtained by the appropriate secondary cleavage were subjected to the base-specific chemical cleavage as described by Maxam and Gilbert (22). The gel electrophoresis was performed as described previously (23) except that longer sequencing gels (80 cm in length) were used in some of the experiments.

S1 nuclease mapping of the GAL80 mRNA

The 5' terminus of the GAL80 mRNA was determined by S1 nuclease method described by Weaver and Weissmann (24). PolyA⁺RNA was prepared from *S. cerevisiae* KK4[pKK8005] grown on YPGal medium as described previously (17). The hybridization probe was 1.4 kb HindIII/BglII fragment 5' end-labeled at the BglII site. Conditions for DNA/RNA hybridization and S1 nuclease digestion were also described (17).

RESULTS

DNA sequence determination of the wild-type GAL80 gene

Shimada *et al.* (17) conducted the Northern blot analysis and also the S1 nuclease mapping to disclose the location and direction of the GAL80 mRNA on the cloned fragment. The results demonstrated that the transcription starts at about 85 bp upstream from the left BglII site, proceeds rightward and terminates at about 400 bp downstream from the SmaI site (Fig. 1). There is no indication that the GAL80 mRNA is spliced.

Thus we determined the DNA sequence of 2457 bp containing the 1.55 kb GAL80 mRNA transcribed region in the 3.1 kb HindIII/HindIII fragment. The strategy of sequencing and restriction sites used here are shown in Fig. 1. About 97% of the sequence was determined in both strands and the junction points were sequenced by the overlapping fragments. We found a long open reading frame of 1305 bp along the rightward direction within the transcribed region (Fig. 2). We assumed that this reading frame corresponds to the coding

Fig. 1 Restriction map of the GAL80 gene region and sequencing strategy. The thick line in the middle part shows a 3.1 kb HindIII fragment of yeast DNA and the bold arrow indicates the direction and localization of GAL80 transcript. The thin arrows in the upper part of the figure indicate the directions and extents of sequencing from either 5'-labeled termini or 3'-labeled termini by the method of Maxam-Gilbert. The arrows in the lower part indicate the directions and extents of sequencing by dideoxynucleotide chain termination technique.

sequence of the GAL80 gene for the following reasons: 1. Kozak suggested that translation starts at the first ATG in the 5' end in most eukaryotic mRNA (25). This holds true in most yeast mRNA (26). The ATG at +1 ~ +3 was the first ATG in the 5' end of the GAL80 mRNA as revealed by the S1 nuclease mapping described in the later section. The reading frames starting at the other ATGs in the downstream region would encode peptides which seemed to be too small to be the regulatory factor bearing the complex functions. 2. Kozak further suggested that the translation initiating ATG usually has purine (Pu) at -3 and at +4 and the surrounding sequence is ...CC(Pu)CCATGG (25). The ATG at +1 ~ +3 in GAL80 was also flanked with G at -3 and at +4. 3. Recently, Shimada *et al.* (17) have found that the GAL80 transcription is regulated by the control machinery involving GAL4 and also GAL80 itself. When we constructed a gal80-lacZ fused gene by ligating the 5' part of GAL80 cut at the BglII site (+612) to 3' coding region of 'lacZ from *E. coli* (27), the resultant fused gene exhibited β -galactosidase activity under the GAL4/GAL80 control in yeast. This could be accounted for only by assuming the longest reading frame mentioned above to represent the GAL80 coding region.

Deduced amino acid sequence of the GAL80 protein

A deduced amino acid sequence of the GAL80 protein is presented in Fig.

2. It consisted of 435 residues and the calculated molecular weight was 48309 including the NH₂-methionine. This protein would be neutral in the physiological condition since the sum of arg + lys (39 residues) was approximately equal to the sum of asp + glu (38 residues). There was however, an imbalance of charged residues in the amino-terminal portion (residues 1-70) of the GAL80 protein; the sum of arg + lys was 8 whereas the sum of asp + glu 2, indicating the net charge was +6 in this region. It was also noted that there was a heavy clustering of lys from the residue 279 to 304; 7 out of 26 residues were lys.

Mapping the 5' end of GAL80 mRNA

In the present work we extended the previous S1-nuclease mapping of 5' end (17) to the level of nucleotide. The probe employed was 1.4 kb HindIII/BglIII fragment 5' end-labeled at the BglIII site (Fig. 1). PolyA⁺RNA was prepared from the yeast strain KK4[pKK8005]. The DNA/RNA hybrids were digested with S1 nuclease and S1-resistant fragments were analyzed on the sequencing gel in parallel with the DNA sequencing ladder of the probe (Fig. 3). At least three protection bands were observed around CTGCT at the positions -67 ~ -63 in the non-coding sequence (Fig. 3, lanes 1 and 2), suggesting that the 5' end of GAL80 mRNA had a slight heterogeneity. We concluded that the major 5' end of GAL80 mRNA located at C of position -67 because the intensity of the corresponding band was the strongest among the three bands in lane 2, in which smaller amount of S1 nuclease was used than in lane 1. The possibility that the observed heterogeneity was simply due to the inherent nature of S1 nuclease (28, 29) has remained to be ruled out.

DNA sequence determination of GAL80^{S-O}, GAL80^{S-1} and GAL80^{S-2}

The dominant mutations are known in both regulatory genes GAL4 and GAL80: GAL4^C renders the yeast capable of constitutively synthesizing the GAL enzymes, while GAL80^S makes the yeast uninducible for the GAL enzymes (3, 4, 6). The latter mutations are classified into three types by the phenotype when combined with GAL4^C: The GAL enzyme synthesis is constitutive in a double mutant of GAL4^CGAL80^{S-O} (4), inducible in GAL4^CGAL80^{S-1} and uninducible in GAL4^CGAL80^{S-2} (6). We previously cloned these mutant alleles in YE24 (13). We then subcloned appropriate DNA fragments of the mutant GAL80 DNA into either M13mp10 or M13mp11 depending on the direction of sequence and determined the DNA sequence by the dideoxy chain termination method. The strategy and the extent of sequencing are shown in Fig. 4. Only the DNA sequence of the coding region encompassing 1305 bp was determined unambiguously. The results indicated that each of the mutant alleles contained the

-732

GG GTTTGTCTTC AATAGGCAGC CGTTGTCTTG

-700
CAAAAGCGCCG CCTCTGCCAT GGCAGGAAT GCTTTCCACG ACGATCATCG TAGTGCCCAA TTGGTGSCCT CATGATAGGG TAGGCGTTGG GCAAGTGCTC

-600
TTTATGTAT CTGTGAATTT ATCTCGTGGT TCTTCTGGTC TGTIAGGGCA AGGTGGCCT CTACTTACTC CATCGACAAT TCAAGATACA GAACCTCTTC

-500
CAGATGGAAT CCTTTCATA GAGAGAAGGA GCAAGCAACT GACCAATAT TGACTGCCAC TGGACCTGAA GACATGCAAC AAGTGCAAG CATAGTGGGG

-400
CCTCTTCCCA ATGCTAATCC GGTCACTGCC ACTGCTGCTA CGGAAMACCA ACCTAAGGT ATTAACTTCT TCACATAAG AAAATCACAC GAGCGCCCGG

-300
ACGATGTCTC TGTTTAAGT GCGCAAGTTT TCGCGTTTGT AATATATATT TATACCCCTT TCTTCTCTCC CTCGCAATAT AATAGTTTAA TTCTAATATT

-200
AATAATATCC TATATTTTCT TCATTATACC GCGCACTCTC GCGCGAACGA CCTCAAAATG TCTGTACAT TCATAATAAC CAAAGCTCA TAACTTTTTC

-100
TTTTGAACCT GAATATATAT ACATCATATG TCATCTGCTGG TCCTTGCCGA CCAGCGTATA CAATCTCGAT AGTTGGTTTC CCGTCTTTTC CACTCCCGTC

+1
ATG GAC TAC AAC AAG AGA TCT TCG GTC TCA ACC GTG CCT AAT GCA GCT CCC ATA AGA GTC GGA TTC GTC GGT CTC AAC GCA GCC AAA GGA
met asp tyr asn lys arg ser ser val ser thr val pro asn ala ala pro ile arg val gly phe val gly leu asn ala ala lys gly

+91
TGG GCA ATC AAG ACA CAT TAC CCC GCC ATA CTG CAA CTA TCG TCA CAA TTT CAA ACT ACT GCC TTA TAC AGT CCA AAA ATT GAG ACT TCT
trp ala ile lys thr his tyr pro ala ile leu gln leu ser ser gln phe gln ile thr ala leu tyr ser pro lys ile glu thr ser

+181
ATT GCC ACC ATC CAG CGT CTA AAA TTG AGT AAT GCC ACT GCT TTT CCC ACT TTA GAG TCA TTT GCA TCA TCT TCC ACT ATA GAT ATG ATA
ile ala thr ile gln arg leu lys leu ser asn ala thr ala phe pro thr leu glu ser phe ala ser ser ser thr ile asp met ile

+271
GTG ATA GCT ACT CAA GTG GGC AGC CAT TAT GAC GTT GTT ATG CCT CTC TTG GAA TTC TCC AAA AAT AAT CCG AAC CTC AAG TAT CTT TTC
val ile ala ile gln val ala ser his tyr asp val val met pro leu leu glu phe ser lys asn asn pro asn leu lys tyr leu phe

+361
GTA GAA TGG GCC CTT GCA TGT TCA CTA GAT CAA GCC GAA TCC ATT TAT AAG GCT GCT GCT GAA CGT GGG GTT CAA ACC ATC TCT TTA
val glu trp ala leu ala gys ser leu asp gln ala glu ser ile tyr lys ala ala glu arg gly val gln thr ile ile ser leu

+451
CAA GGT CGT AAA TCA CCA TAT ATT TTG AGA GCA AAA GAA TTA ATA TCT CAA GGC TAT ATC GGC GAC ATT AAT TCG ATC GAG ATT GCT GGA
gln gly arg lys ser pro tyr ile leu arg ala lys glu leu ile ser gln gly tyr ile gly asp ile asn ser ile glu ile ala gly

+541
AAT GGC GTG TGG TAC GGC TAC GAA AGG CCT GTT AAA TCA CCA AAA TAC ATC TAT GAA ATC GGG AAC GGT GTA GAT GTG GTA ACC ACA ACA
asn gly gly trp tyr gly tyr glu arg pro val lys ser pro lys tyr ile tyr glu ile gly asn gly val asp leu val thr thr thr

+631
TTT GGT CAC ACA ATC GAT ATT TTA CAA TAC ATG ACA AGT TCG TAC TTT TCG AGG ATA AAT GCA ATG GTT TTC AAT AAT ATT CCA GAG CAA
phe gly his thr ile asp ile leu gln tyr met thr ser ser tyr phe ser arg ile asn ala met val phe asn asn ile pro glu gln

+721
GAG CTG ATA GAT GAG CGT GGT AAC CGA TTT GGC CAG CGA GTC CCA AAG ACA GTA CCG GAT CAT CTT TTA TTC CAA GGC ACA TTG TTA AAT
glu leu ile asp glu arg gly asn arg leu gly gln arg val pro lys thr val pro asp his leu leu phe gln gly thr leu leu asn

+811
GGC AAT GTT CCA GTG TCA TGC AGT TTC AAA GGT GGC AAA CCT ACC AAA AAA TTT ACC AAA AAT TTG GTC ATT GAC ATT CAC GGT ACC AAG
gly asn val pro val ser cys ser phe lys gly gly lys pro thr lys lys phe thr lys asn leu val ile asp ile his gly thr lys

+901
GGA GAT TTG AAA CTT GAA GGC GAT GCC GGA TTG CCA GAA ATT TCA AAT GTC GTC CTT TAC TAC AGT GGA ACT AGA GCA AAC GAC TTC CCG
gly asp leu lys leu glu gly asp ala gly phe ala glu ile ser asn leu val leu tyr tyr ser gly thr arg ala asn asp phe pro

+991
CTA GCC AAT GGA CAA GCT CCT TTA GAC CCG GGG TAT GAT GCA GGT AAA GAA ATC ATG GAA GTA TAT CAT TTA CGA AAT TAT AAT GCC
leu ala asn gly gln gln ala pro leu asp pro gly tyr asp ala gly lys glu ile met glu val tyr his leu arg asn tyr asn ala

+1081
ATT GTG GGT AAC ATT CAT CGA CTG TAT CAA TCT ATC TCT GAC TTC CAC TTC AAT ACA AAG AAA ATT CCT GAA TTA CCC TCA CAA TTT GTA
ile val gly asn ile his arg leu tyr gln ser ile ser asp phe his phe asn thr lys lys ile pro glu leu pro ser gln phe val

+1171
ATG CAA GGT TTC GAT TTC GAA GGC TTT CCC ACC TTG ATG GAT GCT CTG ATA TTA CAC AGG TTA ATC GAG AGC GTT TAT AAA AGT AAC ATG
met gln gly phe asp phe glu gly phe pro thr leu met asp ala leu ile leu his arg leu ile glu ser val tyr lys ser ser asn met

+1261
ATG GGC TCC ACA TTA AAC GTT AGC AAT ATC TCG CAT TAT AGT TTA TAA
met gly ser thr leu asn val ser asn ile ser his tyr ser leu

+1331
AAGCATCTGGC CTGTGCTGG CTCCAGTGC

+1341
AATGAACGTT ATAAAAAGCA ATACTAGTA TATATCTATG TAAACAACC ATATCATTTT TTGTTCTGAA CTTTGTATAC CTAACTAGTT TTAATTTTC

+1441
CTTTTCTGTG CATCGGGTG TTCTATTAT TTAGCATACT ACATTGAAA TATCAAAATT CCTTAGTAGA AAGTGAGAG AAGGTGCACT AACACAAAA

+1541
ATAAATGCT ACGTAAACT GTCAAAACT TCGACAGCG GGCATCTTTC CATCATAGCT TCAAAACATAT TAGCGTTCT GATCTTCATA CCGCTGCTCA

+1641
AATGATCAA ACAAACTGTT ATTTGCAAGA AATAAAAGCA AGGCTGCCCT CAAAACTGA TCCATTAGAT CCTCATATCA AGCTT

single base substitution corresponding to the single amino acid alteration as listed in Table I. It should be pointed out here that the observed mutational change would always lead to an increase in the positive charge of GAL80 protein (gly to arg in both GAL80^{S-0} and GAL80^{S-1}, and glu to lys in GAL80^{S-2}) and that the mutational sites were localized within a narrow region of the gene (the amino acid positions of 301, 323 or 351 in the total number of 435 for GAL80^{S-0}, GAL80^{S-1} and GAL80^{S-2} respectively). We suggest therefore that a functional domain of the GAL80 protein resides around that region.

The other part of the DNA sequence in the respective mutant allele was perfectly identical to the wild type sequence except for two nucleotides (C at +294 and +905) both in GAL80^{S-1} and GAL80^{S-2}. These changes however would not result in the amino acid substitution because of the degeneracy of the genetic code (AGC to AGT for ser and GAC to GAT for asp). Note that the parental strain for GAL80^{S-1} and GAL80^{S-2} (6) is different from the strain from which the wild type gene was isolated (13).

The flanking sequence of the GAL80 gene

Most of the protein-coding genes in higher eukaryotes contain TATA box precisely at 30 bp upstream from the transcription initiation site (30, 31). In contrast, the typical TATA box is rarely observed at that position in yeast genes (32, 33). The 5' flanking region of GAL80 gene also seemed not to have a typical TATA box within 100 bp upstream from the transcription initiation site (Fig. 2). Instead, TATATATA octamer (-87 ~ -80) was observed about 20 bp upstream from the initiation site.

The sequence immediately downstream from the transcription initiation site contained 5'-TGCTGGTC-3' (-66 ~ -59) which was complementary to the 3' end of yeast 18 S rRNA [3'-AUUACUAG-5' (34)] at 5 out of 8 nucleotides. Zalkin and Yanofsky (35) suggested that the 5'-leader sequences of many yeast mRNAs show the complementarity to the 3' end of 18 S rRNA, which could enhance the 40 S ribosome binding to the message. A set of inverted repeats was observed (-65 ~ -58 and -52 ~ -46) within the 5'-leader sequence of GAL80 mRNA, one of which was overlapping to the 18 S rRNA complementary region.

The 5' upstream sequence of the GAL80 gene consisted of alternating AT-rich region and GC-rich region: The region from -72 to -148 and -174 to

Fig. 2 Nucleotide sequence and deduced amino acid sequence of the GAL80 and its flanking regions. The nucleotide corresponds to the non-coding strand. The numbering begins at the A in the translation initiating ATG in each direction. The amino acid sequence is translated from the nucleotide sequence. The asterisk marks the 5' ends of GAL80 mRNA. The wavy line indicates the region in which the 3' end of GAL80 mRNA should reside. An inverted repeat and a direct repeat is underlined by dotted and thin arrows respectively.

Fig. 3 Determination of the 5' end of GAL80 mRNA on the nucleotide sequence. The HindIII/BglII fragment 5' end-labeled at the BglII site was hybridized to polyA⁺RNA from KK4 with pKK8005 grown on galactose. The DNA/RNA hybrids were treated with 400 units (lane 1) or 100 units (lane 2) of S1 nuclease. The nuclease resistant fragments were electrophoresed on a 10% polyacrylamide/8.3 M urea sequencing gel along with the sequencing ladder of the same probe (lanes G, A+C, C+T, C). The partial sequence of the probe (coding strand) is shown on the right.

-273 was 74% and 76% AT-rich respectively. The region from -1 to -71, -149 to -173 and -274 to -315 was 54%, 76% and 52% GC-rich. A set of direct repeats of 5'-CGCCCGPuACGA-3' (-307 ~ -297 and -161 ~ -151) was present in two GC-rich regions.

As the signal for termination and/or polyadenylation for yeast mRNA, TAG...TATGT(or TAGT)...(AT-rich)...TTT was proposed by Zaret and Sherman (36). A related structure CAG (+1335 ~ +1337)...TATGT (+1377 ~ +1381)...(AT-rich)...TTT (+1397 ~ +1399) was observed between the termination codon and the site where the 3' terminus of GAL80 mRNA resides (Fig. 2).

Codon usage in the GAL80 gene

The codon usage frequency of the GAL80 gene is shown in Table 2. The

Fig. 4 Sequencing strategy for GAL80^S region. The thick line shows the protein-coding region of the GAL80 gene. The HaeIII site locates at -401 and the HindIII site at +1720. The coding sequence is delineated by ATG and TAA codons. The thin arrows indicate the directions and extents of sequencing by the dideoxynucleotide chain termination method.

Table 1. Nucleotide and amino acid changes in the coding region of each GAL80^s gene

Mutant allele	Position of amino acid*	Change of nucleotide**	Change of amino acid***
<u>GAL80</u> ^{s-0}	301	<u>GGA</u> → <u>AGA</u>	Gly → Arg
<u>GAL80</u> ^{s-1}	323	<u>GGA</u> → <u>AGA</u>	Gly → Arg
<u>GAL80</u> ^{s-2}	351	<u>GAA</u> → <u>AAA</u>	Glu → Lys

*The numbering of amino acid position begins at the translation initiating ATG in Fig. 2.

**Based on the DNA sequencing, the substituted positions of nucleotide in corresponding codon are underlined.

***The amino acid change is translated from the nucleotide.

GAL80 gene used 57 out of 61 possible codons, exhibiting a very weak bias in the codon utilization. Bennetzen and Hall (37) proposed that efficiently expressed yeast genes prefer only 25 codons whereas poorly expressed genes use more codons. Since the expression of the GAL80 gene is moderately efficient (17), the proposal of Bennetzen and Hall is also true in GAL80. A weak bias toward the use of codons ending with pyrimidines in the wobble position was observed in Ile Val, Ala, Thr, Gly and Pro.

Table 2. Codon usage in the GAL80 protein

	U	C	A	G
U	Phe UUU 8	Ser UCU 7	Tyr UAU 12	Cys UGU 1
	Phe UUC 12	Ser UCC 5	Tyr UAC 10	Cys UGC 1
	Leu UUA 14	Ser UCA 10	--- UAA 1	--- UGA 0
	Leu UUG 8	Ser UCG 5	--- UAG 0	Trp UGG 3
C	Leu CUU 5	Pro CCU 6	His CAU 6	Arg CGU 4
	Leu CUC 3	Pro CCC 5	His CAC 4	Arg CGC 0
	Leu CUA 4	Pro CCA 6	Gln CAA 16	Arg CGA 4
	Leu CUG 6	Pro CCG 4	Gln CAG 2	Arg CGG 0
A	Ile AUU 14	Thr ACU 6	Asn AAU 18	Ser AGU 7
	Ile AUC 15	Thr ACC 8	Asn AAC 9	Ser AGC 3
	Ile AUA 9	Thr ACA 9	Lys AAA 17	Arg AGA 4
	Met AUG 10	Thr ACG 0	Lys AAG 7	Arg AGG 3
G	Val GUU 8	Ala GCU 9	Asp GAU 11	Gly GGU 11
	Val GUC 6	Ala GCC 11	Asp GAC 7	Gly GGC 11
	Val GUA 6	Ala GCA 10	Glu GAA 13	Gly GGA 7
	Val GUG 5	Ala GCG 0	Glu GAG 7	Gly GGG 3

DISCUSSION

Extensive studies on the structure and function of the regulatory proteins of E. coli or λ phage have revealed that those proteins consist of at least two functional domains: The DNA-binding ability specified by the amino-terminal domain and the effector binding and/or oligomerization ability specified by the carboxy-terminal domain (38, 39).

The mutational sites in the dominant uninducible GAL80 alleles studied in the present work were localized within a small region consisting of 50 amino acids in the carboxy-terminal part. This may imply that the carboxy-terminal part of GAL80 protein comprises a functional domain other than that of the amino-terminal part; for instance, the domain for inducer binding. In this context, it would be worthwhile to recall that many lacI^S mutations of E. coli are mapped in the carboxy-terminal half of that gene and none in the amino-terminal part, suggesting that the carboxy-terminal but not the amino-terminal part would contribute to the inducer binding (40).

The GAL80 protein consisted of 435 amino acid residues and its overall charge was roughly neutral. However, the amino-terminal part (residues 1-70) contained more positively charged amino acids than negatively charged amino acids. This charge distribution resembles to that of lcI protein (41), which is an acidic protein but its amino-terminal part (residues 1-26) is highly basic. A similar situation is found in the lacI repressor which is neutral as a whole, but its amino-terminal part (residues 1-60) is basic (42). It is tempting therefore to speculate that the amino-terminal part of GAL80 protein also bears the similar function to the prokaryotic regulatory proteins.

On the other hand, the three-dimensional analysis of the products of lcro and lcI or the catabolite activator protein of E. coli strongly suggests that the DNA-binding domains share the common structure of α -helix-turn- α helix (43-45). Recently GAL4 protein was argued to contain the similar structure based on the nucleotide sequence of that gene, which was thus taken as an indication that GAL4 protein binds DNA (16). Such an α -helical unit could not be found in GAL80 protein from the present work, which would not support, nor rule out, the possibility of GAL80 to interact directly with its target genes. Needless to say, much more experiments are required to elucidate the mode of action of this regulatory protein. We are currently attempting to specify the hypothetical domains by the localized mutagenesis of the isolated gene on one hand and to purify the protein from yeast by an immunological assay on the other hand.

ACKNOWLEDGEMENTS

This work was supported in part by the research grant from the Ministry of Education, Science and Culture of Japan.

The authors are indebted to Dr. Mitsuru Takanami and his colleagues for valuable information on the dideoxy sequencing technique. Our thanks also due to Mr. Masahiro Tajima for help in the sequencing and to Dr. Hideo Shimada for discussions on the secondary structure of the protein.

REFERENCES

1. Douglas, H. C. and Pelroy, G. (1963) *Biochim. Biophys. Acta* 52, 155-156.
2. Douglas, H. C. and Hawthorne, D. C. (1964) *Genetics* 49, 837-844.
3. Douglas, H. C. and Hawthorne, D. C. (1966) *Genetics* 54, 911-916.
4. Douglas, H. C. and Hawthorne, D. C. (1972) *J. Bacteriol.* 109, 1139-1143.
5. Kew, O. M. and Douglas, H. C. (1976) *J. Bacteriol.* 125, 33-41.
6. Nogi, Y., Matsumoto, K., Toh-e, A. and Oshima, Y. (1977) *Mol. Gen. Genet.* 152, 137-144.
7. Perlman, D. and Hopper, J. E. (1979) *Cell* 16, 89-95.
8. Matsumoto, K., Adachi, Y., Toh-e, A. and Oshima, Y. (1980) *J. Bacteriol.* 141, 508-527.
9. St. John, T. P. and Davis, R. W. (1981) *J. Mol. Biol.* 152, 285-315.
10. Post-Beittenmiller, M. A., Hamilton, R. W. and Hopper, J. E. (1984) *Mol. Cell. Biol.* 4, 1238-1245.
11. Oshima, Y. (1982) in *The Molecular Biology of Yeast Saccharomyces: Metabolism and Gene Expression*, Strathern, J. N., Jones, E. W. and Broach, J. R. Eds., pp. 159-180, Cold Spring Harbor Lab., Cold Spring Harbor, New York.
12. Hashimoto, H., Kikuchi, Y., Nogi, Y. and Fukasawa, T. (1983) *Mol. Gen. Genet.* 191, 31-38.
13. Nogi, Y., Shimada, H., Matsuzaki, Y., Hashimoto, H. and Fukasawa T. (1984) *Mol. Gen. Genet.* 195, 29-34.
14. Laughon, A. and Gesteland, R. F. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6827-6831.
15. Johnston, S. A. and Hopper, J. E. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6971-6975.
16. Laughon, A. and Gesteland, R. F. (1984) *Mol. Cell. Biol.* 4, 260-267.
17. Shimada, H., Igarashi, M., Segawa, T. and Fukasawa, T. (1984) *J. Mol. Biol.* (submitted).
18. Messig, J. (1983) in *Methods in Enzymology*, Wu, R., Grossman, L. and Moldave, K. Eds., Vol 101, pp. 20-78, Academic Press, New York.
19. Clewell, D. B. and Helinski, D. R. (1969) *Proc. Natl. Acad. Sci. USA* 62, 1159-1166.
20. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
21. Guo, L-H., Yang, R. C. A. and Wu, R. (1983) *Nucl. Acids Res.* 11, 5521-5540.
22. Maxam, A. and Gilbert, W. (1980) in *Methods in Enzymology*, Grossman, L. and Moldave, K. Eds., Vol 65, pp. 499-560, Academic Press, New York.
23. Nogi, Y. and Fukasawa T. (1983) *Nucl. Acids Res.* 11, 8555-8568.
24. Weaver, R. F. and Weissmann, C. (1979) *Nucl. Acids Res.* 7, 1175-1193.
25. Kozak, M. (1984) *Nucl. Acids Res.* 12, 857-872.
26. Stiles, J. I., Szostak, J. W., Young, A. T., Wu, R., Consaul, S. and Sherman, F. (1981) *Cell* 25, 277-284.
27. Casadaban, M. J., Chou, J. and Cohen, S. N. (1980) *J. Bacteriol.* 143, 971-980.

28. Sollner-Webb, B. and Reeder, R. H. (1979) *Cell* 18, 485-499.
29. Hentschell, C., Irminger, J.-C., Bucher, P. and Birnstiel, M. L. (1980) *Nature* 285, 147-151.
30. Grosschedl, R. and Birnstiel, M. L. (1980) *Proc. Natl. Acad. Sci. USA* 77, 7102-7106.
31. Benoist, C. and Chambon, P. (1981) *Nature* 290, 304-310.
32. Bennetzen, J. L. and Hall, B. D. (1982) *J. Biol. Chem.* 257, 3018-3025.
33. Donahue, T. F., Farabaugh, P. J. and Fink, G. R. (1982) *Gene* 18, 47-59.
34. Rubstov, P. M., Musakhanov, M. M., Zakharyev, V. M., Krayev, A. S., Skryabin, K. G. and Bayev, A. A. (1980) *Nucl. Acids Res.* 8, 5779-5794.
35. Zalkin, H. and Yanofsky, C. (1982) *J. Biol. Chem.* 257, 1491-1500.
36. Zaret, K. S. and Sherman, F. (1982) *Cell* 28, 563-573.
37. Bennetzen, J. L. and Hall, B. D. (1982) *J. Biol. Chem.* 257, 3026-3031.
38. Sauer, R. T., Pabo, C. O., Meyer, B. J., Ptashne, M. and Backman, K. C. (1979) *Nature* 279, 396-400.
39. Matthews, B. W., Ohlendorf, D. H., Anderson, W. F. and Takeda, Y. (1982) *Proc. Natl. Acad. Sci. USA* 79, 1428-1432.
40. Miller, J. H. (1980) in *The Operon*, Miller, J. H. and Reznikoff, W. S. Eds., pp. 31-38, Cold Spring Harbor Lab., Cold Spring Harbor, New York.
41. Sauer, R. T. and Andereg, R. (1978) *Biochemistry* 17, 1092-1100.
42. Beyreuther, K. (1980) in *The Operon*, Miller, J. H. and Reznikoff, W. S. Eds., pp. 123-154, Cold Spring Harbor Lab., Cold Spring Harbor, New York.
43. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. and Matthews, B. W. (1981) *Nature* 290, 754-758.
44. Pabo, C. O. and Lewis, M. (1982) *Nature* 298, 443-447.
45. McKay, D. B. and Steitz, T. A. (1981) *Nature* 290, 744-749.