Nucleotide sequence of the 2-amino-3-ketobutyrate coenzyme A ligase (*kbl*) gene of *E. coli*

Benjamin D. Aronson, Paula D. Ravnika* and Ronald L. Somerville

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

Submitted February 9, 1988

Accession no. X06690

A 3563 bp *EcoRI* fragment that contains the genes for *E. coli* threonine dehydrogenase and 2-amino-3-ketobutyrate coenzyme A ligase (AKB ligase; EC 2.3.1.29) (1) was sequenced using the dideoxy chain termination method. These genes lie at coordinate 81.2 of the standard *E. coli* genetic map (2). The nucleotide sequence of the AKB ligase gene (designated *kbl*) and its promoter is presented. AKB ligase contains 398 amino acids. The predicted N-terminal sequence, MW, and amino acid composition are in agreement with the known properties of the purified enzyme (3). The pattern of codon usage for *kbl* is non-random and is typical of *E. coli* protein coding sequences (4). The transcriptional startpoint of the operon, ascertained by primer extension of cellular mRNA using a synthetic oligodeoxynucleotide primer, is indicated by an arrow. The -10 and -35 regions are underlined. Detailed studies on the promoter and the *tdh* structural gene will be published elsewhere. See Fig. (1).

```

GAATCTATAAATTAGTTCGGAAAGTCAAAAAAATACTAAATAATACATGTGGACGATGACCAAGGCCATTT
1
TGTGATGTGCTATACAAAGCCGCGATCTCTTCAACTAACTATCTGGGGAGAGAAAATGGTTCGATTGTTCCTGCTTCAGGAGTAACATTTAGGGGCAAAAATGACGGCC
100
TGAGAAATTTTCCTTCAAGAAAATAGGTTTTCCCTCCATTAGTGGGTAGCGTTTGTATACCAACTGAGATAGGAAAATGATGCTTTGGCATCATTTTCTTTTATATCAATGACAT
200
AAAATCAATCATCATATTATGCCGCCAAACTGTACAGTTTTTCTACTACTGTAGTGTGGAGACATCAAGTTACAGCTTATTTATCTCGAATTTGACAGAGTGTAAACGGCTTA
300
TCTGCTGCGACCATTAAGTTTGGGTAAATATGCTGCGAATTTGCCCTGTCTGGAGATGCGA
400
MET ARG GLY GLU PHE TYR GLN GLN LEU THR ASN ASP LEU GLU
ATG CGT GGA GAA TTT TAT CAG CAG TTA ACC AAC GAT CTG GAA
500
THR ALA ARG ALA GLU GLY LEU PHE LYS GLU GLU ARG ILE ILE THR SER ALA GLN GLN ALA ASP ILE THR VAL ALA ASP GLY SER GLN VAL
ACC GCA CGG GCG GAA GGG TTG TTT AAA GAA GAG CGC ATT ATT ACG TCT GCG CAG CAA GAA GAT ACT GTG GCT GAT GGA ACC GAG GTC
600
ILE ASN PRE CYS ALA ASN ASN THR LEU GLY LEU ALA ASN HIS PRO ASP LEU ILE ALA ALA LYS ALA GLY MET ASP SER HIS GLY PHE
ATT AAC TTT TGT GCG AAC AAC TAT CTC GCG CTG GCG AAT CAT CCT GAT CTG ATT GCG GCG GCA AAG GCG GAA ATG GAT TCT CAC GGT TTC
700
GLY MET ALA SER VAL ARG PHE ILE CYS GLY THR GLN ASP SER HIS LYS LEU GLU GLU GLN LYS LEU ALA PHE LEU GLY MET GLU ASP
GGC ATG GCT TCG GTG CGT TTT ATT TGC GGC ACT CAG GAC AGC CAT AAA GAG CTT GAA CAA AAA CTG GCG GCT TTC CTG GGG ATG GAA GAT
800
ALA ILE LEU TYR SER SER CYS PRE ASP ALA ASN GLY GLY LEU PHE GLU THR LEU LEU GLY ALA GLU ASP ALA ILE ILE SER ASP ALA LEU
GGC ATT CTC TAC TCT TCC TGC TTT GAT GCT AAC GGT GCG CTG TTT GAA ACG CTT CTG GGT GCG GAA GAC GGC ATT ATC TCC GAC GCA CTG
900
ASN HIS ALA SER ILE ILE ASP GLY VAL ARG LEU CYS LYS ALA LYS ARG TYR ARG TYR ALA ASN ASN ASP MET GLN LEU GLU ALA ARG
AAC CAC GCG TCT ATT ATT GAT GGT GCG GGT CTG TGC AAA GCT AAA GCG TAT CCG TAT GCG AAC AAC GAT ATG CAG GAG CTG GAA GCA CGT
1000
LEU LYS GLU ALA ARG GLU ARG GLY ALA ARG HIS VAL LEU ILE ALA THR ASP GLY LEU PHE SER MET ASP GLY VAL ILE ALA ASN LEU LYS
CTC AAA GAA GCG CGT GAA GCG GGT GCG CGT CAT GTC CTG ATC GCC ACC GAT GGG TTG TCT TCA ATG GAC GGC GTT ATT GCC AAC CTG AAG
1100
GLY VAL CYS ASP LEU ALA ASP LYS TYR ASP ALA LEU VAL MET VAL ASP SER HIS ALA VAL GLY PHE VAL GLY GLU ASN GLY ARG GLY
GGC GTT TGC GAT CTG GCA GAT AAA TAT GAT GCC CTG GTG ATG GTA GAC GAC TCC CAC GCG GGT GGT TTT GTC GGT GAA AAT GGT CGT GGT
1200
SER HIS GLU TYR CYS ASP VAL MET GLY ARG VAL ASP ILE ILE THR GLY THR LEU GLY LYS ALA LEU GLY GLY ALA SER GLY GLY TYR THR
TCC CAT GAA TAC TCG GAT GTG ATG GCG GGC GTC GAT ATT ATC ACC GGT ACG CTT GGT AAA GCA CTG GCG GGG GCT TCT GGT GGT TAT ACC
1300
ALA ALA ARG LYS GLU VAL VAL LEU TRP LEU ARG GLN ARG SER ARG PRO TYR LEU PHE SER ASN SER LEU ALA PRO ALA ILE VAL ALA ILE
GCG GCG GCG AAA GAA GAG GTG GTT GAG TGG CTG GCG CAG CGT TCT CGT CCG TAC CTG TCT TCC AAC TCG CTG GCA GCG CCG ATT GTT GCC GCG
1400
SER ILE LYS VAL LEU GLU MET VAL GLU ALA GLY SER GLU LEU ARG ASP ARG LEU TRP ALA ASN ALA ARG GLN PHE ARG GLU GLN MET SER
TCC ATC AAA GTA CTG GAG ATG GTC GAA GCG GCG AGC GAA CTG CGT GAC CGT CTG TGG GCG AAC GCG CGT CAG TTC CGT GAG CAA ATG TCG
1500
ALA ALA GLY PHE THR LEU ALA GLY ALA ASP HIS ALA ILE ILE PRO VAL MET LEU GLY ASP ALA VAL VAL ALA GLN LYS PHE ALA ARG GLU
GCG GCG GCG TTT ATT CCG TCG GCG GCG GAT CAC GCG ATT ATT CCG GTC ATG CTT GGT GAT GCG GTA GTG GCG CAG AAA TTT GCG CGT GAG
1600
LEU GLN LYS GLY ILE TYR VAL THR GLY PHE PHE THR PRO VAL VAL PRO LYS GLY GLN ALA ARG ILE ARG THR GLN MET SER ALA ALA
CTG CAA AAA GAG GCG ATT TAC GTT ACC GGT TTC TTC TAT CCG GTC GTT CCG AAA GGT CAG GCG GGT ATT CGT ACC CAG ATG TCT GCG GCG
1700
HIS THR PRO GLN ILE THR ARG ALA VAL GLU ALA PHE THR ARG ILE GLY LYS GLN LEU VAL ILE ALA 888
CAT ACC CCG GAG CAA ATT ACG GGT GCA GTA GAA GCA TTT ACG CGT ATT GGT AAA CAA CTG GCG GTT ATC GCG TGA
1800

```

Figure 1. Nucleotide and deduced amino acid sequence of *kbl* and its upstream region. Coordinate 1 is the initial G of the *EcoRI* site that defines one end of our previously described clone (1,2).

ACKNOWLEDGEMENT: This research was supported by a grant from the NIH (GM 22131). Dr. E. Dekker kindly furnished prepublication information about the amino acid composition and sequence of AKB ligase. Dr. G. Bogosian assisted in the computer analysis of the data.

*Present address: MSU/DOE Plant Research Laboratory, Michigan State University, East Lansing, MI 48824, USA

REFERENCES: (1) Ravnika, P. and Somerville, R. (1987) J. Bacteriol. 169, 4716-4721. (2) Ravnika, P., and Somerville, R. (1986) J. Bacteriol. 168, 434-436. (3) Mukherjee, J., and Dekker, E. (1987) J. Biol. Chem. 262, 14441-14447. (4) Fickett, J. W. (1982) Nucl. Acids Res. 10, 5303-5318.