

Nucleotide sequence of the *Bacillus subtilis* IMP dehydrogenase gene

N.Kanzaki and K.Miyagawa

Integrated Technology Laboratories, Takeda Chemical Industries, Ltd, Osaka 532, Japan

Submitted October 11, 1990

EMBL accession no. X55669

We have cloned the IMP dehydrogenase gene from a *Bacillus subtilis* guanosine producing strain¹⁾.

Here, we report the nucleotide sequence of the XbaI-HincII fragment (1.85 kb) on which the gene is located²⁾. The fragment contains an open reading frame of 1,539 bp encoding 513 amino acid residues. The deduced amino acid sequence of the gene is 52% homologous to that of the *Escherichia coli* guaB encoding IMP dehydrogenase³⁾.

A good Shine-Dalgarno sequence is located 10 bp upstream from the putative start codon. The sequence TTGACA at

nucleotide 123-128 coincides with the consensus sequence of *B. subtilis* -35 region. Proposed -35 and -10 regions are underlined in the figure.

REFERENCES

1. Miyagawa, K., Kimura, H., Nakahama, K., Kikuchi, M., Doi, M., Akiyama, S. and Nakao, Y. (1986) *Bio/Technology* 4, 225-228.
2. Miyagawa, K., Kanzaki, N., Kimura, H., Sumino, Y., Akiyama, S. and Nakao, Y. (1989) *Bio/Technology* 7, 821-824.
3. Tiedemann, A. A. and Smith, J. M. (1985) *Nucl. Acids Res.* 13, 1303-1316.

```

TCTAGAAAATTTCTGGGTTGATTCAACGGAGTAAAAATAGCGCTAAGTCTTTCCACTCATGATATGTCATGTATATTTCTTAATCCTTCCGGTTATCTAAAT      100
ATTTCACACTCTTCCCGCTTCCCTTGACATGCTCTTGGCTAGTGTGATAAATCTACATATAATATTTGCCGAAAAAGAGGGGATTACTAATGTGGGAAAAGT      200
          -35          -10          SD          M W E S
          N          - - - - -
AAATTTTCAAAGAAGCGTTAACGTTGACGATGTGCTGCTTGTGCCAGCAAAGCTGAGGTACTTCCGCATGATGTGGATTTATCTGTAGAACTTACAA      300
K F S K E G L T F D D V L L V P A K S E V L P H D V D L S V E L T
- - * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
AAACGTAAAGTAAATATTCCTGTCAATCAGCGCAGGATGGACACTGTAACAGAAATCAGCAATGGCAATGGCAATGGCAAGACAGGGCGGCTTGGGCAT      400
K T L K L N I P V I S A G M D T V T E S A M A I A M A R E G G L G I
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
CATTACAAAAATATGTCATTGAACAGCAGGCTGAACAAAGTGAATAAAGCGTTCTGAGCGCGCGTATCACAAATCCCTCTTTTTAACTCCT      500
I H K N M S I E Q Q A E Q V D K V K R S E R G V I T N P F F L T P
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
GATCAACAGTATTTGATCGGGAGCATTGATGGGAAATACAGAATTTCCGGTGTTCGGATTGTAATAACGAAGAAGACCAGAAGCTTGTGGAAATTA      600
D H Q V F D A E H L M G K Y R I S G V P I V N N E E D Q K L V G I
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
TTACAACCGTGCCTTCTGTTATTTCTGACTACTCAATGAAAAATCAGCGACGTCATGACGAAAGAAGAGCTAGTTACTGCATCTGTAGAACTACTCT      700
I T N R D L R F I S D Y S M K I S D V M T K E E L V T A S V G T T L
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
GGATGAAGCTGAAAAGATTTGCCAAAACATAAAAATGAAAAGCTTCCCTCTCGTAGATGACCAGAATAAATAAAAGGTCCTTACAAATAAAGACATT      800
D E A E K I L Q K H K I E K L P L V D D Q N K L K G L I T I K E
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
GAAAAAGCTCATTGATTCGGAACTCACTAAGACATTCACGGCGCCGTGATGTTGGCGGCGCAGTTGGTGAATGAACTGCGGATACAATGACTCGCGTCA      900
E K V I E F P N S S K D I H G R L I V G A A V G V T G D T M T R V
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
AAAAGCTTGTGAAGCCATGTTGATGTGATTTGATCGATACAGCTCAGGACCTCTCAAGCGTTTTAAACAGAGTTACAAAAATCCGTGAAACGTA      1000
K K L V E A N V D V I V I D T A H G H S Q G V L N T V T K I R E T Y
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
TCCGAAATAACATTTGCTGGAAACGTGGCAACAGCTGAAGCGACAAGACCGCTTATCGAAGCTGGAGCAGAGTTGTCAAAAGTTGGAATAGGGCCT      1100
F E L N I I A G N V A T A E A T R A L I E A G A D V V K V G I G P
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
GGTCAATTTGTACTACAGCTGTGTAGCCGGGTGGGTTCCGCAAATACAGCAATTTATGATTTGCGACTGAAGCAAGAAAACACGGCAAGAA      1200
G S I C T T R V V A G V G V P Q I T A I Y D C A T E A R K H G K T
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
TCATCCCGCAGCGTGGGATTAAATCTCTGGCGATATCACTAAAAGCATTGGCAGCGCGGACATGCTGTTATGCTCGGAAAGCTTGGCTGCAGGCACATC      1300
I I A D G G I K F S G D I T K A L A A G G H A V H L G S L L A G T S
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
AGAAAGCCCTGGTGAAGCTGAAATCTACCAAGGCAGAAAGATTAAAGTATACCGCGCATGGGATCAGTTGCTGCAATGGAAAAAGGAAAGTAAGACCGT      1400
E S P G E T E I Y Q G R R F K V Y R G M G S V A A H E K G S K D R
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
TACTTCAAGAAGAAAAAAATTTGCTCGTGAAGAAATTGAAGACGACACCTTACAAGGCGCAAGTTGAAGAAACCGTTATACAGCTAGTGGGAG      1500
Y F Q E E N K K F V P E G I E G R T P Y K G P V E E T V Y Q L V G
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
GCCTTCGTTCTGGTATGGGGTATGGCGGTCCAAAAGATCTCGGTGCGCTAAGAGAAGAAGCTCAGTTTCATCGCATGACTGGCGCAGGACTTCCGGAAG      1600
G L R S G M G Y C G S K D L R A L R E E A Q F I R M T G A G L R E S
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
CCATCCGCATGACGTACAGATTACAGTGCATCGTAATAAGCGCTTCCCTGGTCTATTTGGTCTCATCAGAAAAAACAGGATTGTTGATGATGATGT      1700
H P H D V Q I T V H R N K A L P G L F G S H Q K K T G F V Y D E C
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
TGTCATCCGGCTTTTTTCATCGATTGATGGTGGAAAGCGTACAGCCCGATAAGCAAATACTATTATCAGCATGGGGAATACCGCGCGATATCGCAT      1800
C Q S G F F S S D
CTATCTCTCCTGTATAAACAGTAGTT      1829

```

* indicates amino acid residue identical with that of *E. coli* IMP dehydrogenase.
- indicates presence of one extra amino acid residue in *E. coli* IMP dehydrogenase.
N and C indicate N- and C-terminal amino acid residues of *E. coli* IMP dehydrogenase, respectively.