Complete nucleotide sequences of two phosphoglucoisomerase isozymes from *Bacillus stearothermophilus*

Wufan Tao, Lili Wang, Renquan Shen and Zujia Sheng

Institute of Genetics, Fudan University, Shanghai, China

Submitted October 30, 1989 EMBL accession nos X16639, X16640

Two phosphoglucoisomerase isozyme genes from Bacillus stearothermophilus have been cloned in E.coli (1) and sequenced with the Sanger method (2). The *Eco*RI-*Pst*I fragment containing the *pgiB* gene is 1866 bp in length. An open reading frame, starting from a Met codon at position 453 preceeded by a ribosome binding site, AGAAAGGAG (G 19.4 kcal/mol) (3), codes for a protein of 445 amino acids (MW 50120). The *Eco*RI-*Pvu*II-*Pvu*II fragment containing the *pgiA* gene is 1822 bp in length. An open reading frame starting from a Met codon at position 95 preceeded by a ribosome binding site, GGAGG (G 14.4 kcal/mol) codes for a protein of 449 amino acids (MW 50337). The spacing between the -35 region and the -10 region, between the ribosome binding sites and the initiation codon and the G of ribosome binding site conform to previous reports (4, 5, 6). The G+C% of *pgiA* and *pgiB* are 43.2% and 40.7%, and the G+C% of the third base of the codons of *pgiA* and *pgiB* are 40.7% and 35.4% respectively. This finding together with the report on the *bgaB* gene (4) seems to be not consistent with the speculation of Barstow, et al. (7).

The homology between the nucleotide sequences of the two isomerase genes is 66% and the homology between the deduced amino acid sequences of the two isomerases is 69%. In the following figure the A and B before the nucleotide sequences represent *pgi*A or *pgi*B and * represents the amino acid of PGIB homologous to that of PGIA.

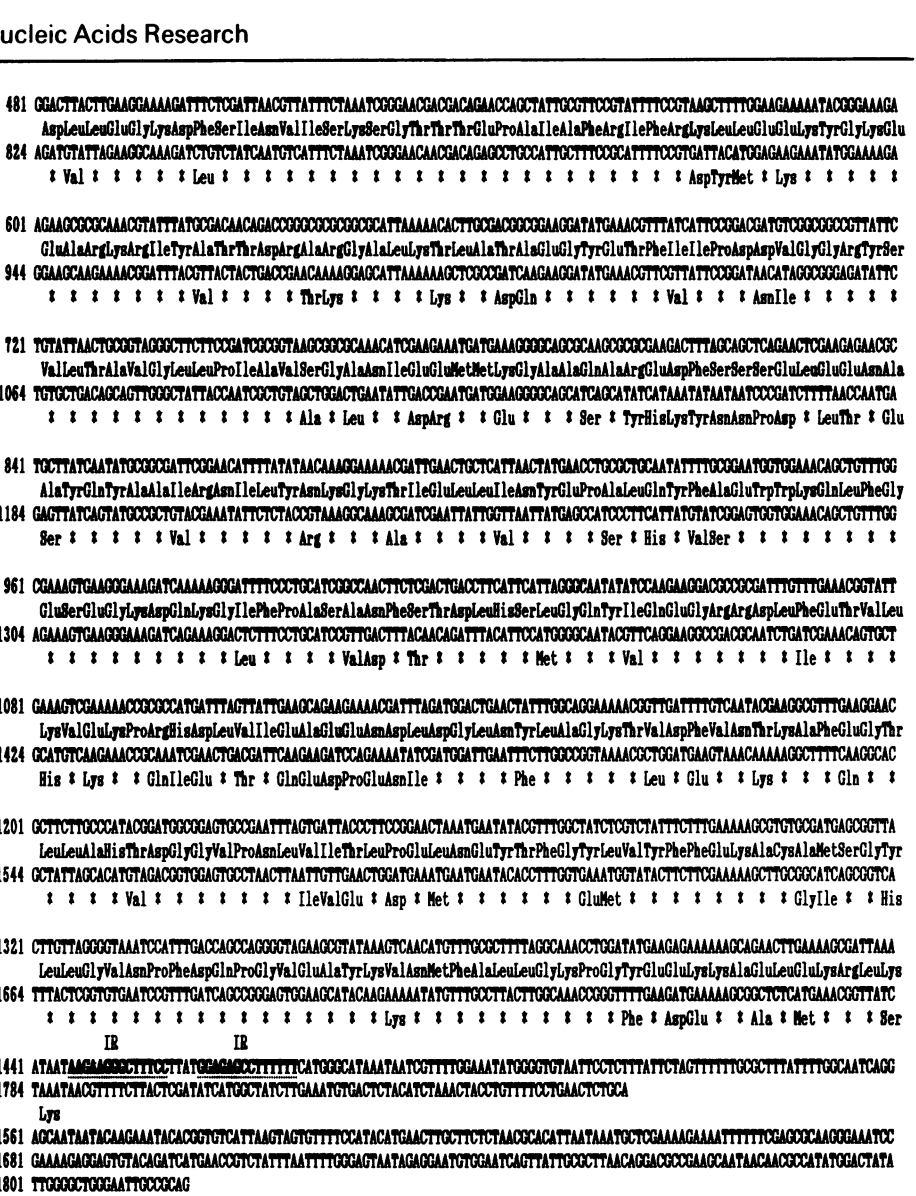
```
B   1 TTAAGATTATGACTATGAAAAAGTTGCAAGAGTGTGGAATAATGGTTCGGTTATTCGTTCATGGCTCATGGAATTAACAGAACGCGCATTTTCAAAAGATGCAAAATTAGAGGAAATCAA
B 121 AGGAATTATGTATTCTTCCGGTGAAGGGAAATGGACAGTCGAAACGGCTTTAGATCTACAAACTGCCACTCCTGTCATCGCCATGTCCTTATTGATGCGGTATCGTTCATTAGAAAACGA
B 241 TACATTTACAGGCAAAGTGGTAGCAGCTCTACGCAACGAATTTGCCGGACATGCTGTAGAAAAAAATGAGAAGAAATAAACCGTTATAAGCATTTGAAGCATCTCCAGAAAAAACTTTTA
                                                                    AT rich region                  -35 region        -10 region

A   1 AATTCGTTTACACCGATAGGGCGCGACTTAATTCCTTCTGATTGTTTCGTTAAAGTGTAAAAAGGGAAAACATTTTTGTTTGGAGGATCGAGCGTATGACACATATTCGTTTCGACTATTC
                                          -35 region        -10 region                SD         METThrHisIleArgPheAspTyrSer
B 361   AGGAAATAACATAAAATCCCCTTTCAGACGTAAGGTCCAAAATGATCAATGGACATGCACAGCAAAATTGGAAGAGAAAGGAGAAGTGTTATATGGCT    ATCTCTTTTGATTACTC
                                                                                        SD         METAla   * Ser *  *  *  *

A 121 AAAAGCATTATCGTTCTTTGGCGAACATGAGCTTACATATTTGCGCGATGCGGTGAAAGTTGCTCACCATTCCCTTCATGAAAAAACCGGAGTGGGCAATGATTTTTTAGGCTGGCTTGA
        LysAlaLeuSerPhePheGlyGluHisGluLeuThrTyrLeuArgAspAlaValLysValAlaHisHisSerLeuHisGluLysThrGlyValGlyAsnAspPheLeuGlyTrpLeuAsp
B 476 TAATGCATTACCATTTATGCAAGAGAATGAACTAGATTATTTAAGCGAATTTGTAAAAGCAGCTCATCATATGCTTCATGAAAGGAAAGGTCCAGGATCGGATTTTCTCGGTTGGGTTGA
        Asn *  * Pro * MetGln * Asn *  * Asp *  * SerGluPhe *  * Ala *  *  * Met *  *  * ArgLys * Pro * Ser *  *  *  *  * Val *

A 241 TTTGCCGGTGAATTATGACAAAGAAGAATTTGCCCGCATTCAAAAAGCGGCCGCAAAAATTCAAGCGGATTCCGACGTATTATTAGTCATCGGAATCGGCGGTTCGTATTTGGGAGCGCG
        LeuProValAsnTyrAspLysGluGluPheAlaArgIleGlnLysAlaAlaAlaLysIleGlnAlaAspSerAspValLeuLeuValIleGlyIleGlyGlySerTyrLeuGlyAlaArg
B 596 TTGGCCGATCCGCTATGATAAAAACGAATTTTCGCGAATCAAACAAGCTGCTGAAAGAATCCGGAATCATTCAGACGCTCTTGTGGTGATCGGCATTGGCGGATCTTATTTGGGTGCAAG
        Trp * IleArg *  *  * Asn *  * Ser *  * LysGln *  * GluArg * ArgAsnHis *  * Ala * Val *  *  *  *  *  *  *  *  *  *  *  *

A 361 TGCAGCGATCGAAATGCTTCATCATTCGTTTTATAATGCGCTTCCGAAAGAAAAACGAAATACGCCGCAAATCATTTTCGTCGGCAACAACATTAGTTCGACATACATGAAAGAAGTAAT
        AlaAlaIleGluMetLeuHisHisSerPheTyrAsnAlaLeuProLysGluLysArgAsnThrProGlnIleIlePheValGlyAsnAsnIleSerSerThrTyrMetLysGluValMet
B 616 GGCGGCTATTGAAGCATTATCCCATACCTTTCATAACCAAATG   AACGAT     ACGACA   CAAATTTATTTTGCTGGTCAGAATATCAGTTCTACTTATATCTCTCATTTATT
         *  *  *  * Ala * Ser * Thr * His * GlnMet   AsnAsp     Thr *     *  * Tyr * Ala * Gln *  *  *  *  *  * IleSerHisLeuLeu
```



REFERENCE

1) Tao, W.F., Wang, L.L., Shen, R.Q. and Sheng, Z.J. Acta Genetica Sinica (in press). 2) Sanger, F., Nicklen, S and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467. 3) Tinoco, Jr. I., Borer, P.N., Dengler, B., Levine, M., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41. 4) Hirata, H., Fukazawa, T., Negoro, S., and Okada, H. (1986) J. Bateriol. 166, 722-727. 5) Ihara, H., Sasaki, T., Tsoboi, A., Yamagata, H., Tsukagoshi, S and Usaka, S. (1985) J. Biochem. 98, 95-103. 6) Takagi, M., Imanaka, T and Aiba, S. (1985) J. Bateriol. 163, 824-831. 7) Barstow, D.A., Sharman, A.F., Atkison, T. and Minton, N.P. (1986) Gene 46, 37-45.