

Nucleotide sequence of the *Escherichia coli* *purF* gene encoding amidophosphoribosyltransferase for *de novo* purine nucleotide synthesis

Gen-ichi Sampei and Kiyoshi Mizobuchi*

Department of Biophysics and Biochemistry, Faculty of Science, University of Tokyo, Hongo, Tokyo 113,
Japan

Submitted July 29, 1988

Accession no. X12423

The 1512 bp sequence of the *E. coli* *purF* gene encoding amidophosphoribosyltransferase (ART, EC 2.4.2.14) has been reported (1). We have also determined the nucleotide sequence of the wild type *purF* gene from *E. coli* strain W3110 and found several discrepancies with the sequence previously reported. The *purF* gene consists of 1515 bp instead of 1512 in length. Insertion of 3 bp can be seen in positions C-638, A-687 and C-689 and consequently, requires a change in the coding frame from nucleotides 637-690. Additionally, 12 other discrepancies are observed as shown in the figure. When the deduced amino acid sequences of ART from *E. coli* (E.c.), *B. subtilis* (B.s.) (2) and *S. cerevisiae* (S.c.) (3) are compared, the altered sequence gives much more similarity to the sequences of B.s. and S.c. than the reported one does. Therefore, the discrepancies observed are not probably due to the difference of *E. coli* strain used in the experiment.

ATG TCGGT ATTTCGGT ATGCCCGTTAACAGTCGATTATGATCCCTTAACCGTGCCTTACCGATCAGCGGTCAGGATGCCGCCGATCATCACCATAGATGCCAA	120
M C G I V G I A G V N P V H Q S I Y D L T V L Q H R G D A A G I I T I D A N	40
AAC T G C T T C G T T C G T T C G T T C G T G A C C G A C C G G T C G T G A C C G A T T T G A C C T G A C G C G T T C G G C A T A T G C G C A T T G C G C T T C G C C G A C G	240
E. c. N C F R L E [A] A N [G] L V S D V F E A R R H M Q R L Q G N M G I G H V R Y P T A G S	80
B.s. (53) T A H R G G Q G L (60)	
S.c. (47) C Q C [G] M (54)	
T C C A G C C G T C T C G A C C G A C C G C T T A C G T T A C T T A C C C G G T A T G C A C C A G C G T C A C G Q A G T T G C G T A A A A A C T G T T G A A G A A A A C C G	360
S S P P F Y V N S P Y G I T C P K Y [B] D L N F R K K L P F E E R K K L	120
C O C C A C A T C A C A C C A T C C G C A C T C G Q G A A T T C G C T T A A T A T C T C T C C G C A G C G T Q A H A C G T C A T T C C G C T G A A G C C G A A T A T T T C C G C T G C C G C A C A	480
R H I N T T S D S E L N H I F A S E L K Y [B] D L N F R H Y P L A D H N I P A A I A A T	160
A A C C G C T T A A T C C C C G G C G C T G T C C T G G C G T G T A T C T G C C G C A C G G T A T G T G C C C G G C A C C G G A T T C G C G C T G G C G T G A C T G Q G A A A A C C G T A T A T T G C G A G G	600
M B L I R G A Y C V V A M I I G H M V G P D R P G I R P L V L Q R R D I D E	200
A A C C G T A C A G A A T A T A T G T G C C T G C G A A V C G T A G C C G T C A G T C G C G G T T A G T T T C C G T C G T C G T G C G G C T G G C G T A C T G Q G A A A A C C G T A T A T T G C G A G G	720
E. c. M N B T E Y M V A S E S V A L D I L Q E R E V A E G E A M I Y I T E B K Q Q L F	240
B.s. (207) C A F D V V G A T Y L [B] D L N F R K K L P F E E R K K L (226)	
S.c. (218) V F R A K M Q G P E N A (234)	
A C C G C T A A T G T G C A C A C C G G T C A G C A A T C C G G C G C T G T T A G T G A T G A T A T C T T G C C G C C G G C A C T G T T A T C G A C A A A T T T C C G T T A C A C C C C C T Q T Q A A T A T G C G C	840
E. c. T B R Q C A D N P V B N P C L F E Y V V Y F A R P D B F I D K I S V Y S	280
B.s. (270) M K N L C (282) L A M G	
S.c. (282) L A M G	
A C G A A A T C G G G C G A G A A A T T G C C C G G A A T T G G A A G A T C T G G A T C G C G T G T G A T C C C G A T C C C G A A A C C T C G T G T G A T A T C C C G C T G G A A A T T T C G C G C A A C C G	960
E. c. T E G K T I A R E V E W E D L D I D V V I P I P E T S C D T I A L E I A R I L G K P	320
B.s. E M [L] A (278)	
S.c. S E [L] A (390)	
T A C C G C C A G G C T C G T T A A A A A C C G T A T G T G C C G C A C C T T T A C T A C C C G C A C G A C T G C T G C T G C T G C A A T C T G A A T C C G C A A C C C C C C G G A T T C C G G G A T	1080
Y B Q V V Y V G R F T I P N Q G Q L R B V R E L H A N H A R E F R D	360
A A A A C G T C C T G T G C A G C A C T G C A T C G C T G C T G C A C C A C T G C A G C A G T T A T C G A G A T G A T C G A C C G C A A G C C G G A A A G A T G T A C C T C C G C T T C G C G C A C C G G A A A T T	1200
E. c. K N V L L V D D S I V E G O T T S E Q I I E K A P [A] K V Y L A B A A P E I	400
B.s. (374) E S (391) S (393)	
S.c. (391) S (393)	
C O C T C C G G A A C G C T T A T G T A T T G A T A T C O C G A C C G C A C G G A A C G A T G A T C G G T G C T G A C C G G T T G A T T T T C C G G A T C T G C	1320
E. c. P R H V A K M Q G P E N A (374) E S (391) S (393)	440
A A C G A T C G T A G O A C C C G G T C G C G T G A A A A T C C G A C G A C T T G A T G C T G C T G C T A C G C G G A A G A T G T A C C T C G C G C A C C G G A A T T	1440
M D L I D V V I P I P E T S C D T I A L E I A R I L G K P	480
A C G T T A C G T A A T G A C C C G A A A C G A T G C A A C T C G A G A C G A T G O G A A A A A T C G A A A T A C A A C G A G G T A	1518
T E R R H D D A K A V Q E Q H E V E N L E N H H E G *	505

The bases and amino acid residues changed are underlined. Partial amino acid sequences of ART from B.s. and S.c. are from references (2) and (3). The identical amino acid residues between the three organisms are boxed.

*To whom correspondence should be addressed

- References:** (1) Tso, J.Y., Zalkin, H., van Cleemput, M., Yanofsky, C., and Smith, J.M. (1982) J. Biol. Chem. 257, 3525-3531. (2) Makaroff, C.A., Zalkin, H., Switzer, R.L., and Vollmer, S.J. (1983) J. Biol. Chem. 258, 10586-10593. (3) Mäntsälä, P., and Zalkin, H. (1984) J. Biol. Chem. 259, 8478-8484.