Volume 9 Number 18 1981 Nucleic Acids Research

Nucleotide sequence of the asnA gene coding for asparagine synthetase of E.coli K-12

Masataka Nakamura, Masao Yamada, Yukinori Hirota, Kazunori Sugimoto*, Atsuhiro Oka* and Mituru Takanami*

National Institute of Genetics, Mishima, Shizuoka 411, and *Institute for Chemical Research of Kyoto University, Uji-shi, Kyoto 611, Japan

Received 7 July 1981

ABSTRACT

We have subcloned the *asnA* gene of *E. coli* K-12, a gene coding for asparagine synthetase, from a previously cloned 6 mega-dalton segment of *E. coli* chromosome containing the DNA replication origin, *ori*, and *asnA*. The complete nucleotide sequence of the *asnA* gene was determined: the region of the structural gene extends 990 base-pairs at nucleotide positions 1434-2423 (see Fig. 3), which codes for a polypeptide of 330 amino-acid residues with a molecular weight of 36,688 daltons. The nucleotide sequences of the promoter and the ribosome-binding site of the gene are also assigned. We discuss the properties of its polypeptide.

INTRODUCTION

Since Yasuda and Hirota (1) isolated a chromosomal fragment carrying the replication origin, *ori*, of *E. coli*, extensive studies on the structure and function of *ori* have been carried out (2-7) and the *asnA* gene was found on the cloned 6 mega-dalton (Mdal) DNA as an adjacent gene of *ori* (3, 5, 7, 8).

The *asnA* gene encodes an enzyme protein, an asparagine synthetase which catalyzes the reaction transfering ammonium to aspartic acid to form asparagine in the presence of ATP. The molecular weight of the enzyme was determined to be 80,000 daltons by measuring the enzyme activity in the eluates after chromatography of extracts of *E. coli* K-12 on Sephadex G-100 (9). Recently, another asparagine synthetase gene, *asnB*, has been found at 15 min on the chromosome (10, 11). These two enzymes have similar catalytic activities but differ in substrate specificity and temperature stability. Either enzyme alone is sufficient for the growth of the bacteria under the conditions of asparagine limitation. The molecular weight of 80,000 daltons has been observed for asparagine synthetase from *asnA* (11). On the contrary, von Meyenburg and Hansen reported the molecular weight of the gene product of *asnA* to be 39,000 daltons (7). Thus, the studies of biochemical characteristics of this enzyme do not seem satisfactory. We approached this problem

by the use of technologies for gene-cloning and nucleotide sequencing of *asnA*.

In the present paper, we describe the complete nucleotide sequence of *asnA* gene including the promoter and the ribosomal binding site. The amino-acid sequence of the asparagine synthetase was deduced from the nucleotide sequence and the properties of its polypeptide were discussed.

MATERIALS AND METHODS

All strains used were derivatives of *E. coli* K-12: JE6279 (F⁻ *asnA asnB thi thy str recA*) and GM31 (*thr leu dcm his thi ara lac galK galT xyl mtl str tonA tsx supE*). Restriction endonucleases were purchased from Takara Co., Ltd. or Bethesda Research Laboratories, or prepared by published procedures (12). Plasmid DNA was isolated as previously described (2). Transformation using the plasmid DNA was done as previously described (13). DNA sequence determination was performed by the procedure of Maxam and Gilbert (14).

RESULTS

Cloning of *asnA* gene

From previous studies (3, 7), the 6 Mdal *Eco*RI fragment containing the replication origin of *E. coli* was shown to carry the *asnA* gene. The details of the restriction map of the 6 Mdal fragment were constructed by measuring of molecular weight of the fragments which were generated by clevage of λ*gt·ori* (3) DNA with several restriction enzymes. Only the sites of *Bam*HI, *Pst*I and *Hind*III are shown in Fig. 1A. The *asnA* gene was extended into a region of *Bam*HI B and D fragments (3). In order to determine a more precise location of *asnA* gene, the following plasmids carrying subfragments of the 6 Mdal fragment were constructed (Fig. 1B). pMY102 and pMY123 carried *Hind*III-*Eco*RI A and *Bam*HI-*Eco*RI D fragments on pBR322, respectively. pMY114 and pMY111 carried *Pst*I C fragment on pBR322 at the *Pst*I site. The orientation of the *Pst*I C fragments of both plasmids was opposite. pTSO103 possessed the *Bam*HI B fragment on pBR322 (3).

In order to examine whether the above plasmids carry the *asnA* gene or not, the JE6279 strain (F⁻ *asnA asnB recA*) was transformed with each plasmid. pMY102, pMY111 and pMY114 could transform the Asn⁻ strain into Asn⁺. However, pMY123 and pTSO103 failed to transform the Asn⁻ bacteria into Asn⁺. It is concluded that the complete genetic information on the production of asparagine synthetase is in *Pst*I C fragment and the *Bam*HI site in *Pst*I C fragment

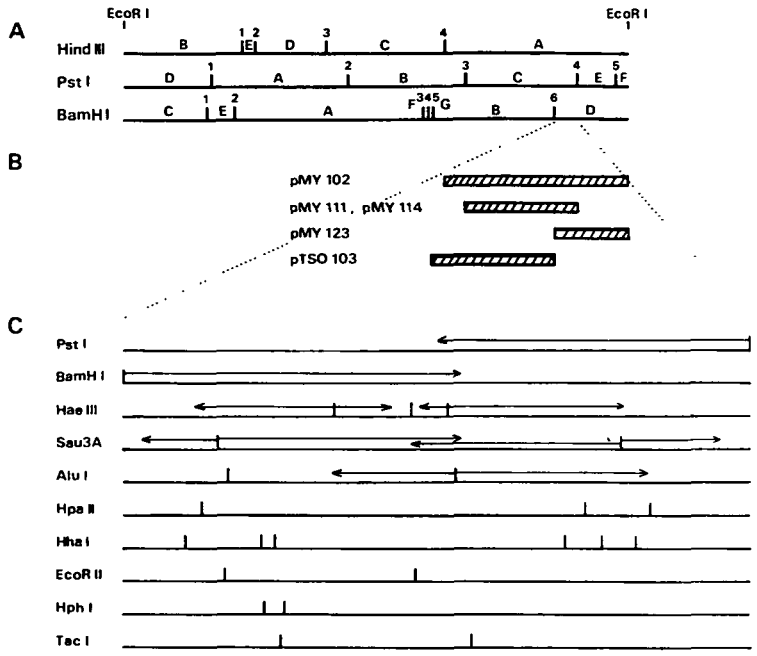

Figure 1. A: Restriction map of the 6 Mdal chromosome fragment. Numbers at each vertical bar indicate restriction enconuclease recognition sites and the capital letters indicate restriction fragments, respectively. B: Shaded lines indicate cloned fragments of pMY102, pMY111, pMY114, pMY123 and pTSO103, respectively. C: Fine physical map and sequencing strategy of *Bam*HI-*Pst*I region.

cleaves *asnA* gene. A similar result has been obtained by the analysis of deletion mutants isolated from λ*gt·ori* (4).

Nucleotide sequence

The nucleotide sequence of the *Bam*HI F, G and B regions has been established and a part of an open reading frame region corresponding to a part of the *asnA* gene has been proposed in previous studies (3, 4). In this study we determined the nucleotide sequence of the region from the *Bam*HI site 6 to the *Pst*I site 4 (see Fig. 1A). The entire *asnA* gene is shown to be located within the *Pst*I C fragment as shown in a above section. A fine restriction map of the *Bam*HI-*Pst*I region of pMY111 was constructed by using a series of eight different restriction enzymes (Fig. 1C). The strategy used to sequence the *Bam*HI-*Pst*I region is presented in Fig. 1C. Over 95% of the region was sequenced for both strands. We determined the sequence of the *Bam*HI-*Pst*I

fragment containing 379 base-pairs. Thus the addition of the sequence to the previously determined sequence (2291 base-pairs) completed the entire sequence (2670 base-pairs) of the fragment from the *Bam*HI site 4, to *Pst*I site 4 containing the *ori* and *asnA* genes. We adopted the numbering system of nucleotide positions used in previous papers (2-4).

Coding region of *asnA* and the gene product

We sought the reading frame of the *asnA* on the nucleotide sequence of the *Pst*I C fragment: initiation (ATG) and termination (TAA, TAG and TGA) codons for protein synthesis were plotted on each of the six reading frames, three phase frames on both orientations, ranging from nucleotide positions 489 to 2570 (Fig. 2). The cloning experiments have shown that the complete genetic information coding for the production of the asparagine synthetase is encoded within the *Pst*I C fragment and the *Bam*HI site 6 is located in the essential region of *asnA*. The apparent molecular weight of the *asnA* gene product has been reported to be 39,000 daltons (7) or 80,000 daltons (9, 11). Only one open reading frame which can code for a peptide having an approxi-

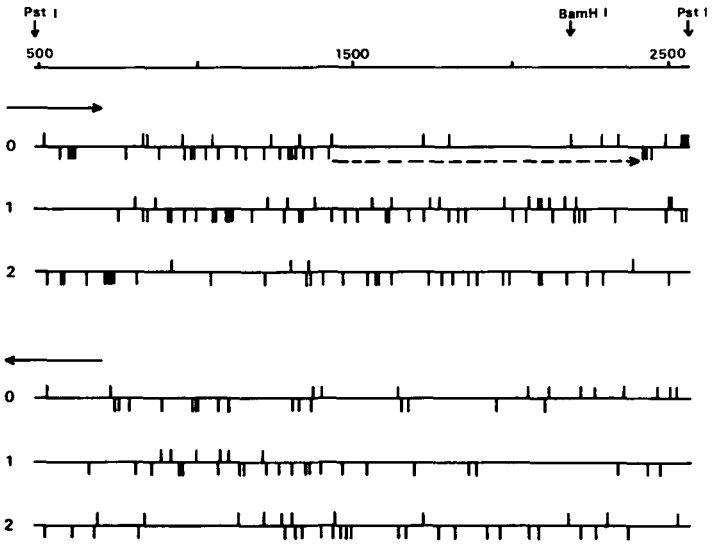

Figure 2. Initiation and termination codons on each of the three phase frames (0, 1, 2) on both directions within the nucleotide positions of 489 through 2570. Vertical bars above the lines represent initiation codons and bars under the lines indicate termination codons. Solid arrows indicate the direction of transcription. An open reading frame coding for asparagine synthetase and the direction of transcription are indicated with a broken line and an arrow.

mate molecular weight of 39,000 daltons was found. We concluded that the region coding for the *asnA* was initiated with ATG at the nucleotide positions 1434-1436 and terminated with a doubled terminator, TAATAA, at nucleotide positions 2424-2429 (Fig. 2 and 3).

Based on the assumption that the *asnA* gene product is not processed after translation *in vivo*, the polypeptide of the asparagine synthetase consists of 330 amino-acid residues having a molecular weight of 36,688 daltons.

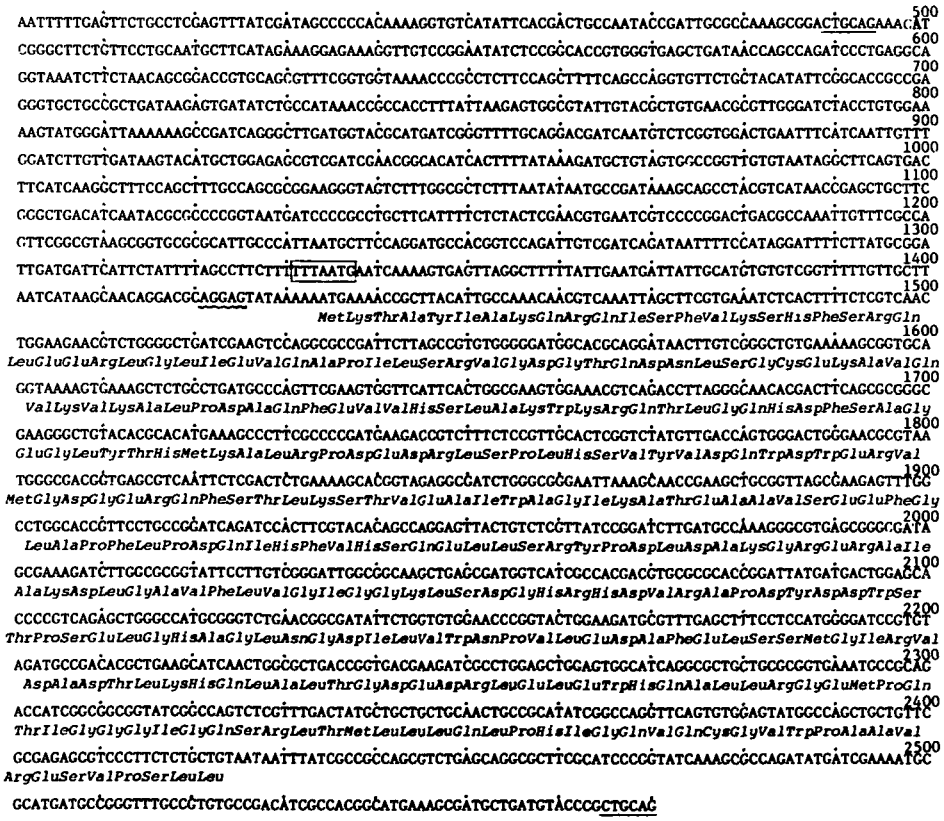

Figure 3. Nucleotide sequence of *Pst*I C fragment and amino-acid sequence of *asnA*. Bases are ordered in the *Pst*I site 3 to 4 direction. (▭) a potential Pribnow sequence of *asnA*; (∿∿) a potential ribosome binding site of *asnA*; (—) *Pst*I recognition sequence. A part of other open reading frame which initiates from TAC at positions 734 through 732 from right-to-left direction was also found. We named the reading frame *proX* gene which codes for a hypothetical X-protein. Presence of *proX* gene was described elsewhere (4).

DISCUSSION

An enormous amount of information on nucleotide sequences of the structural as well as regulatory regions of DNAs of many organisims has been accumulating. We report here additional information on the complete nucleotide sequence of *asnA* of *E. coli* K-12. Our results indicate that the structural gene of *asnA* consists of 990 nucleotide base-pairs (nucleotide positions 1434-2423) which code for 330 amino-acid residues and that a Pribnow-box-like sequence TTTAATG (nucleotide positions 1331-1337) is found upstream from the structural gene of *asnA*. A possible ribosome-binding site is also found at positions 1421-1425 (AGGAG). The region encompassing the regulatory region of the *asnA* gene including the Pribnow-box-like sequence and ribosome-binding sequence is extremely A-T rich (4). It should be noted that there are numerous inverted repeats in the start sites of transcription and translation. Those occurring at positions 1305-1313 and 1334-1342 and at positions 1393-1401 and 1406-1414 are particularly significant. It is possible that these repeats may participate in regulation at the level of transcription and translation of the *asnA* gene. At positions 2466-2513, there is an inverted repeat which continued to a run of three T residues at its 3'-end. It is possible that the sequence is to serve as the termination site of RNA polymerase of *asnA* (15).

The molecular weight of the *asnA* product is estimated to be 36,688 daltons from the amino-acid sequence deduced from the nucleotide sequence. Its molecular weight was consistent with that reported by von Meyenburg and Hansen (7). They estimated the apparent molecular weight at 39,000 daltons from the relative mobility of the polypeptide on SDS-polyacylamide gel. On the other hand, the apparent molecular weight of asparagine synthetase determined by other groups was reported as 80,000 daltons (9, 11), twice as large as that of ours. They measured the enzyme activity of the asparagine synthetase in the eluates after chromatography of cell lysates on a column of Sephadex G-100. We found no open reading frame corresponding to such an 80,000 dalton polypeptide within the sequenced region. An explanation to reconcile these conflicting results might be as follows: a single polypeptide having a molecular weight of 36,688 daltons is a monomeric subunit of asparagine synthetase and the apparent molecular weight of 80,000 daltons (9, 11) corresponds to a dimer form which may have enzyme activity of asparagine synthetase.

Nichols and Yanofsky found no tryptophan residue in the α-subunit of the tryptophan synthetase of *E. coli* (16). Similarly, the asparagine synthe-

Table 1. Codon usage in *asnA*

Phe	UUU	3	Ser	UCU	6	Tyr	UAU	3	Cys	UGU	2
	UUC	7		UCC	1		UAC	2		UGC	0
Leu	UUA	2		UCA	3	Term	UAA	1	Term	UGA	0
	UUG	3		UCG	3		UAG	0	Trp	UGG	8
Leu	CUU	7	Pro	CCU	2	His	CAU	6	Arg	CGU	10
	CUC	0		CCC	1		CAC	7		CGC	7
	CUA	0		CCA	1	Gln	CAA	7		CGA	0
	CUG	28		CCG	10		CAG	13		CGC	2
Ile	AUU	6	Thr	ACU	2	Asn	AAU	0	Ser	AGU	0
	AUC	7		ACC	6		AAC	3		AGC	9
	AUA	1		ACA	0	Lys	AAA	12	Arg	AGA	0
Met	AUG	6		ACG	4		AAG	3		AGG	0
Val	GUU	5	Ala	GCU	5	Asp	GAU	15	Gly	GGU	6
	GUC	4		GCC	6		GAC	9		GGC	14
	GUA	8		GCA	3	Glu	GAA	13		GGA	2
	GUG	7		GCG	13		GAG	10		GGG	7

tase contained only two asparagine residues in the amino-acid sequence. Thus, these two amino-acid synthetases lack or have very few amino-acid residues of the respective reaction product. Two cysteine residues were found at the positions of the 51st and 275th residues in the amino-acid sequence. These residues of the enzyme may contribute to the formation of the structure(s) responsible for the enzyme activity. This idea is supported by the observation that the enzyme activity is stabilized in the presence of 2-mercaptoethanol and inactivated by agents which react with sulfhydryl groups (9). The content of acidic and basic amino-acids of the polypeptide deduced from the nucleotide sequence was 14.3% and 10.3%, respectively. This indicates that its polypeptide is slightly acidic. This is supported by the finding of Cedar and Schwartz who measured its isoelectric point as 5.5 (9).

Three hydrophobic regions of its polypeptide were found in the C-terminal area of the polypeptide. Values of hydrophobic index (17) at the three areas are 2.18, 1.00 and 1.91, respectively. These hydrophobic indexes are consistent with the fact that the asparagine synthetase is a soluble protein.

The frequency of codon usage was also investigated. As shown in Table 1, it is not random (18) but the tendency is more random than that of codon choices in ribosomal protein gene (19). The net preference for a codon responding to the major tRNA species apparent in the *asnA* was as follows: CUG for leucine (20), CCG for proline (21), GCG for alanine (22), AAA for lysine (23), CGY (Y=pyrimidine) for arginine (24) and AUY for isoleucine (25).

ACKNOWLEGEMENTS

This work was supported by Research Grant 56122005 from the Ministry of Education of Japan and a grant from the Volkswagen foundation.

REFERENCES

1. Yasuda, S. and Hirota, Y. (1977) Proc. Natl. Acad. Sci. USA *74*, 5458-5462.
2. Sugimoto, K., Oka, A., Sugisaki, H., Takanami, M., Nishimura, A., Yasuda, Y. and Hirota, Y. (1979) Proc. Natl. Acad. Sci. USA *76*, 575-579.
3. Hirota, Y., Yasuda, S., Yamada, M., Nishimura, A., Sugimoto, K., Sugisaki, H., Oka, A. and Takanami, M. (1979) Cold Spring Harbor Symp. Quant. Biol. *43*, 129-138.
4. Hirota, Y., Yamada, M., Nishimura, A., Oka, A., Sugimoto, K., Asada, K. and Takanami, M. (1981) in Progress in Nucleic Acid Research and Molecular Biology, Cohn, W.E. Ed., Vol 26, pp.33-48, Academic Press, New York.
5. Miki, T., Hiraga, S., Nagata, T. and Yura, T. (1979) Proc. Natl. Acad. Sci. USA *75*, 5099-5103.
6. Meijer, M., Beck, E., Hansen, F.G., Bergmans, H.E.N., Messer, W., von Meyenburg, K. and Schaller, H. (1979) Proc. Natl. Acad. Sci. USA *76*, 580-584.
7. von Meyenburg, K. and Hansen, F.G. (1980) in ICN-UCLA Symposia on Molecular and Cellular Biology, Alberts, B. Ed., Vol XIX, pp.137-159, Academic Press, New York.
8. Bachmann, B.J. and Low, K.B. (1980) Microbiol. Rev. *44*, 1-56.
9. Cedar, H. and Schwartz, J.H. (1969) J. Biol. Chem. *244*, 4112-4121.
10. Felton, J., Michaelis, S. and Wright, A. (1980) J. Bacteriol. *142*, 221-228.
11. Humbert, R. and Simoni, R.D. (1980) J. Bacteriol. *142*, 212-220.
12. Takanami, M. (1973) FEBS Letters *34*, 318-322.
13. Lederberg, E.M. and Cohen, S.N. (1974) J. Bacteriol. *119*, 1072-1074.
14. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Sci. USA *74*, 560-564.
15. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. *13*, 319-353.
16. Nichols, B.P. and Yanofsky, C. (1979) Proc. Natl. Acad. Sci. USA *76*, 5244-5248.
17. Segrest, J.P. and Feldmann, R.J. (1974) J. Mol. Biol. *87*, 853-858.
18. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Nucleic Acids Res. *9*, r43-r74.
19. Post, L.E., Strycharz, G.D., Nomura, M., Lewis, H. and Dennis, P.P. (1979) Proc. Natl. Acad. Sci. USA *76*, 1697-1701.
20. Blank, H.U. and Söll, D. (1971) J. Biol. Chem. *246*, 4947-4950.
21. Söll, D., Cherayil, J.D. and Bock, R.M. (1967) J. Mol. Biol. *29*, 97-112.
22. Lund, E. and Dahlberg, J.E. (1977) Cell *11*, 247-262.
23. Chakraburtty, K., Steinschneider, A., Case, R.V. and Mehler, A.H. (1975) Nucleic Acids Res. *2*, 2069-2075.
24. Celis, T.F.R. and Maas, W.K. (1971) J. Mol. Biol. *62*, 179-188
25. Yarus, M. and Barrell, B.G. (1971) Biochem. Biophys. Res. Commun. *43*, 729-734.