Nucleotide sequence of the *Escherichia coli* replication gene *dnaZX*

Kuo-Chang Yin, Aleksandra Blinkowa and James R. Walker

Department of Microbiology, University of Texas, Austin, TX 78712-1095, USA

Received 16 June 1986; Accepted 17 July 1986

ABSTRACT

The *Escherichia coli* 2.2 kilobase *dnaZX* region contains one 1929 nucleotide reading frame which directs the synthesis of two protein products involved in DNA polymerization. The larger consists of 643 amino acids in a deduced 71,114 dalton chain which could be the τ subunit of DNA polymerase III. The smaller, the DNA polymerase III γ subunit, is encoded by the same reading frame as the larger. The *dnaZX* sequence contains a region homologous to ATP binding sites, suggesting that these replication factors are adenine nucleotide binding proteins.

INTRODUCTION

The *Escherichia coli* *dnaZ* and *dnaX* genes were originally thought to be separate but closely linked on the chromosome (1). However, Mullin et al. (2) and Kodaira et al. (3) cloned *Z* and *X* complementing activities on a 2.2 kilobase pair (kb) fragment, which has a maximum coding capacity of 80,000 daltons of protein in one reading frame, and showed that this region directs the synthesis of two proteins, one of about 75,000 daltons, the other of 56,500. There are two clues that the larger protein, the *X* product, might be cleaved to generate the smaller, the *Z* protein. First, peptide mapping after limited proteolysis showed that the smaller is part of the larger (3). Second, a *dnaZ*⁺*X*⁺ plasmid directs the synthesis of both in minicells but only the larger in an *in vitro* coupled transcription-translation system, suggesting that a processing enzyme is lacking in the *in vitro* system (2).

The 75,000 dalton *dnaX* protein migrates during electrophoresis with purified DNA polymerase III holoenzyme component τ and it is possible, therefore, that the *dnaX* protein is the τ subunit (2,3). The 56,500 dalton *dnaZ* protein was identified as holoenzyme subunit γ by demonstrations that extracts of a temperature-sensitive (Ts) *dnaZ* mutant could be complemented by purified holoenzyme component γ (4,5) and that cloned *dnaZ* fragments stimulated overproduction of γ activity (4,6,7).

To determine how the *dnaZX* region directs the synthesis of both *Z* and *X* proteins, the 2.2 kb region has been sequenced. A single open reading frame encodes a (deduced) polypeptide of 71,114 daltons. This result rules out the possibility that different reading frames direct the synthesis of the two products and supports models in which the smaller *dnaZ* protein is identical to a portion of the larger *dnaX* protein.

METHODS

Phage and Plasmid

M13mp9 phage (8) was obtained from New England Biolabs. The pJH16 plasmid (2) consists of a 3.0 kb dnaZ⁺X⁺ EcoRI-PstI fragment cloned in pBR322.

DNA Purification

Plasmids and M13 replicative form DNA were extracted by the rapid alkaline procedure (9). M13 single strand DNA was prepared by phenol extraction (10).

Enzymes

Restriction enzymes, T4 DNA ligase, Bal31 nuclease, and the Klenow fragment of DNA Polymerase I were obtained from New England Biolabs or Bethesda Research Labs and used in accordance with the supplier's instructions, except that ligations were conducted overnight at 10°C.

Nucleotide Sequencing

Fragments cloned in M13mp9 were sequenced by the chain terminator method (11). The strategy for sequencing the dnaZX region was based on the generation of overlapping fragments by Bal31 nuclease digestion (12). DNA of the dnaZ⁺X⁺ plasmid pJH16 (2) (Fig. 1) was opened with HindIII, digested with Bal31 nuclease for various intervals, and treated with PstI to create a set of fragments with one blunt end and one PstI end. For sequencing, these fragments were cloned into phage M13mp9 DNA which had been cut with HincII (blunt end) and PstI. The sequence of the opposite strand was determined by opening plasmid pJH16 with PstI, Bal31 digestion, cleavage with HindIII, and cloning the resulting fragments into M13mp9 cut with HincII and HindIII.

RESULTS

The SmaI-PvuII fragment which contains both dnaZ⁺ and X⁺ complementing activities consists of 2218 base pairs which contain only one long open reading frame (Fig. 2). A potential initiating ATG at position 142 is followed by an open reading frame, which would encode 71,114 daltons of protein, and a TGA terminator. The other five possible reading frames all contain multiple stop codons; the next longest potential polypeptide could contain 167 amino acids with a molecular weight of 18,400. Presumably 71,114 daltons is the correct size of the dnaX polypeptide previously reported as 75,000 (2) to 83,000 daltons (13). Thus, both the Z and X products are generated from one reading frame. The amino acid composition of the deduced dnaX protein reflects a slightly acidic protein of pI 6.72 (Table 1). The amino acid composition of a purified preparation of γ (the Z product) has been reported (4); however the reported composition differs considerably from the deduced composition of Z product, irrespective of whether Z overlaps the N-, as reported (3), or C-terminal ends of the X product.

Analysis of the deduced amino acid sequence revealed a region likely to be an adenine nucleotide binding site. Walker et al. (14) reported that proteins which bind adenine nucleotides contain sequence A which is Gly-X-X-

Fig. 1. Strategy for sequencing the dnaZX region. The heavy line represents chromosomal DNA present in plasmid pJH16 (2). pJH16 was opened with HindIII, treated with Bal31 nuclease, and cut with PstI. The resulting fragments were cloned into the PstI-HincII sites of M13mp9 for sequencing the regions indicated by the rightward arrows. For sequencing the complementary strand, pJH16 was opened with PstI, digested with Bal31 nuclease, and cut with HindIII. The resulting fragments were cloned into HindIII-HincII cut M13mp9 for sequencing the regions indicated by the leftward arrows. E, EcoRI; H, HindIII; Sm, SmaI; N, NdeI; B, BstEII; Pv, PvuII; Ps, PstI; Hc, HincII; B, BamHI; H, HindIII.

Gly-X-Gly-Lys-Thr-X-X-X-X-X-Ile/Val, in most cases preceded five or six residues upstream by a basic amino acid. Residues 45 to 59 of the deduced dnaX product are Gly-X-X-Gly-X-Gly-Lys-Thr-X-X-X-X-X-Ala. In addition, a histidine is located six residues upstream. Although most such sequences contain isoleucine or valine in the last position, the dnaX protein has an alanine in position 59. Another exception to the consensus sequences occurs in the rho protein adenine nucleotide binding site, which also has an alanine, rather than isoleucine or valine, as the last amino acid (15). Inasmuch as the X protein probably originates from the N-terminal portion of the ZX gene (3), both the X and Z products contain nucleotide binding sites. It has previously been reported (16) that the τ and γ subunits of holoenzyme bind ATP and dATP. Adenine nucleotide binding proteins sometimes contain a sequence B (14) to which there was no homology in the dnaX sequence.

The proposed dnaZX reading frame is preceded by a potential promoter consisting of a -35 sequence separated from the -10 sequence by 17 base pairs (Fig. 2). A potential Shine Dalgarno AGAG sequence is located seven base pairs upstream of the ATG initiating codon.

The codon usage predicts that dnaZX is a weakly expressed region. As reviewed by Grosjean and Fiers (17), strongly expressed genes have codon distributions different from those of weakly expressed genes. The codons NNC are favored over NNU for phenylalanine, tyrosine, asparagine, and isoleucine in genes for abundant proteins. In weakly expressed genes, NNC codons are favored over NNU for proline, arginine, glycine, and alanine (Table 2) (17). The dnaZX open reading frame distribution of NNC and NNU codons is typical of weakly expressed genes except for the asparagine codons (Table 2).

Another characteristic of genes with high expression efficiencies is the

tendency not to contain codons which correspond to certain minor species of tRNAs (17,18). The codons AUA (isoleucine), CGA/CGG/AGA/AGG (arginine), CUA (leucine) and GGA (glycine) are not used at all or are very rarely used in genes for abundant proteins, but are more often used in weakly expressed genes (17). The *dnaZX* gene contains all of these except AGG (Table 2).

Table 1. Amino acid composition of the deduced dnaX protein.

<u>Amino Acids</u>	<u>Number</u>
Nonpolar	
Ala	86
Val	43
Leu	79
Ile	27
Pro	41
Met	15
Phe	11
Trp	6
Polar	
Gly	24
Ser	30
Thr	37
Cys	6
Tyr	9
Asn	18
Gln	40
Acidic	
Asp	27
Glu	53
Basic	
Lys	28
Arg	47
His	16
Total	643

DISCUSSION

The 2.2 kb dnaZX region consists of one long open reading frame which, if translation begins at ATG, would encode a protein of 71,114 daltons. The ATG codon is preceded seven base pairs upstream by a potential AGAG Shine-Dalgarno sequence. A potential promoter consists of a TCGCCG -35 region separated by 17 base pairs from a TAGCAT -10 sequence. The deduced product of 71,114 daltons is likely to be the dnaX product, previously reported as 75,000 to 83,000 daltons (2,3,13). This open reading frame directs the synthesis also of the Z protein, reported as 52,000 to 56,500 daltons (2,3).

One model to explain the appearance of two products is based on the possibility that the larger protein is proteolytically cleaved to generate the smaller. A second model proposes that the dnaX mRNA is processed, or its synthesis is terminated prematurely, to generate a shorter messenger, the

Table 2. Codon Distribution of the dnaZX Gene^a.

UUU Phe 7 1.1%	UCU Ser 2 .3%	UAU Tyr 6 .9%	UGT Cys 3 .5%
UUC Phe 4 .6%	UCC Ser 4 .6%	UAC Tyr 3 .5%	UGC Cys 3 .5%
UUA Leu 7 1.1%	UCA Ser 4 .6%	UAA ---	UGA ---
UUG Leu 11 1.7%	UCG Ser 8 1.2%	UAG ---	UGG Trp 6 .9%
CUU Leu 9 1.4%	GCU Pro 2 .3%	CAU His 12 1.9%	CGU Arg 14 2.2%
CUC Leu 4 .6%	CCC Pro 3 .5%	CAC His 4 .6%	CGC Arg 23 3.6%
>CUA Leu 3 .5%	CCA Pro 9 1.4%	CAA Gln 10 1.6%	>CGA Arg 4 .6%
>CUG Leu 45 7.0%	CCG Pro 27 4.2%	CAG Gln 30 4.7%	>CGG Arg 5 .8%
AUU Ile 15 2.3%	ACU Thr 3 .5%	AAU Asn 6 .9%	AGU Ser 5 .8%
AUC Ile 10 1.6%	ACC Thr 21 3.3%	AAC Asn 12 1.9%	AGC Ser 7 1.1%
>AUA Ile 2 .3%	ACA Thr 0 0%	AAA Lys 21 3.3%	>AGA Arg 1 .2%
AUG Met 15 2.3%	ACG Thr 13 2.0%	AAG Lys 7 1.1%	>AGG Arg 0 0%
GUU Val 9 1.4%	GCU Ala 11 1.7%	GAU Asp 15 2.3%	GGU Gly 7 1.1%
GUC Val 13 2.0%	GCC Ala 17 2.6%	GAC Asp 12 1.9%	GGC Gly 12 1.9%
GUA Val 5 .8%	GCA Ala 15 2.3%	GAA Glu 35 5.4%	>GGA Gly 2 .3%
GUG Val 16 2.5%	GCG Ala 43 6.7%	GAG Glu 18 2.8%	>GGG Gly 3 .5%

^aEach square contains four columns for codons, amino acids, number of times the codon appears in the dnaZX open reading frame, and the percentage contribution of the codon to the total number of codons. Rectangles with bold lines indicate NNC codons which are preferred over NNU codons in strongly expressed genes. Rectangles with broken lines indicate NNC codons which are preferred over NNU codons in weakly expressed genes. Arrows indicate codons underrepresented in strongly expressed genes but often present in weakly expressed genes.

translation of which synthesizes the Z protein. A third model assumes that two different promoters direct the synthesis of different length messengers. If this model is correct, it must be assumed that the promoter for the Z transcript was inactive in the coupled in vitro transcription-translation system which generated the larger, but not the smaller protein (2) or that the shorter messenger was synthesized but not translated. A fourth model assumes that an internal methionine codon, such as the ATG at 529-531 (Fig. 2) serves to initiate translation of the Z protein. Two possibilities which

can be ruled out by the sequence analysis include a two reading-frame model and read-through of a nonsense codon (which terminates Z protein) to generate a longer X protein.

The possibility that this one gene directs the synthesis of two DNA Polymerase III subunits is very interesting. There is strong evidence that the dnaZ protein is the γ subunit. Wickner and Hurwitz (5) isolated the dnaZ protein assaying wild-type extract activity in a heated dnaZ(Ts) extract. The native molecular weight of purified protein was 125,000 daltons. Huebscher and Kornberg (4) also purified the dnaZ product using a complementation assay and demonstrated that the denatured protein has a molecular weight of 52,000 and that the native form is a dimer. The evidence that the protein isolated by these approaches was the product of the Z gene was the observations that plasmids and transducing phages which carried the dnaZ gene caused the overproduction of dnaZ complementing activity (4,6,7). Huebscher and Kornberg (4) also showed that purified dnaZ protein and the γ subunit of DNA polymerase III had similar molecular weights and activities.

The other product, the X protein, could be the τ subunit (2,3). τ complexes with the α , ϵ , and θ subunits to form DNA Polymerase III', a form intermediate in complexity and processivity between the core and holoenzymes (13,19). The X protein and τ migrate similarly in two-dimensional gel electrophoresis (3). However, there is some doubt about the identity of X and τ because the defect of a dnaX(Ts) mutant extract was complemented in vitro by the 32,000 dalton δ subunit of DNA polymerase III (20). Moreover, there is no functional assay for τ , which prohibits the testing of mutant extracts for τ thermolability. In addition, there are other proteins known to interact with DNA and to be approximately the same size as X but not yet identified with genes. These include an 81,000 molecular weight DNA-dependent ATPase (ATPase IV) which also migrates with τ (21) and a 74,000 dalton topoisomerase III (22,23).

The adenine nucleotide binding site deduced for both the X and Z proteins tend to corroborate the report (16) that both the τ and γ subunits bind ATP and dATP.

The apt gene immediately upstream of dnaZX encodes adenine phosphoribosyltransferase, an enzyme involved in adenine salvage (24,25; Hershey and Taylor, personal communication). Hershey and Taylor (personal communication) have sequenced the apt gene and the 3' region which extends toward dnaZX. The sequence overlaps nucleotides 1 through 113 of Fig. 2 (with perfect agreement). The (deduced) apt protein terminates with a histidine (nucleotides 8-10 of Fig. 2) meaning that both are transcribed from left to right (in Fig. 2). Only 131 nucleotides separate the apt and dnaZX reading frames. Hershey and Taylor (personal communication) have determined that apt transcripts in vivo end at positions 100, 105, and 117 of Fig. 2, although it is unknown whether these represent sites of transcription termination or sites of mRNA cleavage. Hershey and Taylor also have pointed out two regions of dyad symmetry between apt and dnaZX (nucleotides 24-56 and 65-89 of Fig. 2). The function of these symmetrical regions is unknown,

although they are thought not to be involved in rho-independent termination (Hershey and Taylor, personal communication). It is interesting that the second stem and loop region overlaps the potential promoter of dnaZX (Fig. 2).

ACKNOWLEDGEMENTS

We thank Dr. Howard V. Hershey for communication of results before publication. This work was supported by American Cancer Society Grant NPL69 and, in part, by PHS Grant GM34471 and Welch Foundation Grant F949.

REFERENCES

1. Henson, J. M., Chu, H., Irwin, C. A. and Walker, J. R. (1979) *Genetics* **92**, 1041-1059.
2. Mullin, D. A., Woldringh, C. L., Henson, J. M. and Walker, J. R. (1983) *Mol. Gen. Genet.* **192**, 73-79.
3. Kodaira, M., Biswas, S. B. and Kornberg, A. (1983) *Mol. Gen. Genet.* **192**, 80-86.
4. Huebscher, U. and Kornberg, A. (1980) *J. Biol. Chem.* **255**, 11698-11703.
5. Wickner, S. and Hurwitz, J. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 1053-1057.
6. Wickner, S. H., Wickner, R. B. and Raetz, C.R.H. (1976) *Biochem. Biophys. Res. Commun.* **70**, 389-396.
7. Yasuda, S. and Takagi, T. (1983) *J. Bacteriol.* **154**, 1153-1161.
8. Messing, J. and Vieira, J. (1982) *Gene* **19**, 269-276.
9. Birnboim, H. C. and Doly, J. (1979) *Nucleic Acids Res* **7**, 1513-1523.
10. Schreier, P. H. and Cortese, R. (1979) *J. Mol. Biol.* **129**, 169-172.
11. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467.
12. Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E. and Surrey, S. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4298-4302.
13. McHenry, C. S. (1982) *J. Biol. Chem.* **257**, 2657-2663.
14. Walker, J. E., Saraste, M., Runswick, M. J. and Gay, N. J. (1982) *EMBO J.* **1**, 945-951.
15. Pinkham, J. L. and Platt, T. (1983) *Nucl. Acids Res.* **11**, 3531-3545.
16. Biswas, S. B. and Kornberg, A. (1984) *J. Biol. Chem.* **259**, 7990-7993.
17. Ikemura, T. (1981) *J. Mol. Biol.* **151**, 389-409.
18. Fay, P. J., Johanson, K. O., McHenry, C. S. and Bambara, R. A. (1982) *J. Biol. Chem.* **257**, 5692-5699.
19. Huebscher, U. and Kornberg, A. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 6284-6288.
20. Meyer, R. R., Brown, C. L. and Rein, D. C. (1984) *J. Biol. Chem.* **259**, 5093-5099.
21. Dean, F., Krasnow, M. A., Otter, R., Matzuk, M. M., Spengler, S. J. and Cozzarelli, N. R. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 769-777.
22. Srivenugopal, K. S., Lockshon, D. and Morris, D. R. (1984) *Biochemistry* **23**, 1899-1906.
23. Kocharyan, S. M., Livshits, V. A. and Sukhodolets, V. V. (1975) *Genetika* **11**, 1417-1425.
24. Taylor, M. W., Simon, A. E. and Kothari, R. M. (1985) *The APRT system in Molecular Cell Genetics.* ed Michael Gottesman. John Wiley, p. 311-332.