Volume 14 Number 20 1986 Nucleic Acids Research

The adjacent *dnaZ* and *dnaX* genes of *Escherichia coli* are contained within one continuous open reading frame

Ann M.Flower and Charles S.McHenry

Department of Biochemistry, Biophysics and Genetics, University of Colorado Health Sciences Center, Denver, CO 80262, USA

Received 29 July 1986; Revised and Accepted 24 September 1986

ABSTRACT

The *dnaZ* and *dnaX* loci of *Escherichia coli* have been genetically defined as separate genes, both of which are essential for DNA replication (1). The 2.1 kb region of DNA that complements mutations in both genes has a maximum coding capacity of approximately 80,000 daltons. Two protein products are produced from this region with molecular weights of 77,000 and 52,000 (2,3). We have sequenced a 2.7 kb fragment containing the *dnaZ* and *dnaX* genes and determined that it contains only one open reading frame of sufficient length to encode either of these proteins. This open reading frame may encode a protein of 71,147 daltons or of 68,451 daltons depending on which potential translational initiation codon is utilized. There are two transcriptional promoters preceding the gene as well as a ribosome binding site preceding the two potential initiation codons. Both the promoters and ribosome binding sites are predicted to be weak, perhaps contributing to the low expression of these genes.

INTRODUCTION

The holoenzyme form of DNA polymerase III is the major enzyme responsible for replicative DNA synthesis in *E. coli* (for a review, see 4). Most of the seven holoenzyme subunits have been shown to be encoded by essential *E. coli* genes: *dna*E (α structural gene) maps at 4 minutes on the *E. coli* chromosome (5), *dna*N (β) is in an operon with *dna*A at 83 minutes (6,7), *dnaZ* (γ) and *dnaX* map at 10.4 minutes (1,8,9,10) and *mut*D (*dna*Q) (ε) maps at 5 minutes (11,12). As these widely dispersed genes are not on a single large operon, the question arises as to how the synthesis of the subunits is coordinated to ensure the proper assembly of the small number of polymerase complexes present in the cell. The genes may be constitutively expressed at low levels or may be regulated, either individually or coordinately.

The *dnaZ* and *dnaX* genes were defined as separate genes by complementation studies and were shown to map very close to one another at 10.4 minutes on the *E. coli* chromosome (1). The *dnaZ* gene encodes the γ subunit of DNA polymerase III holoenzyme as shown by complementation of the inactive extract from *dnaZ*(ts) mutant cells by addition of purified γ in an *in vitro* replication assay (9) and the

overproduction of a 52,000 dalton protein by *dnaZ* cloned into a bacteriophage lambda vector (9,13). The *dnaX* gene has been tentatively designated to encode δ based on an experiment which demonstrated that the inactive extract from *dnaX*(ts) mutant cells could be complemented by addition of a purified 32,000 dalton protein (14). Knowing the two genes map in very close proximity to one another, two laboratories independently determined the minimal amount of DNA necessary to complement both *dnaX* and *dnaZ* mutations (2,3). This was found to be a 2.1 kb fragment which, when used in "Maxicell" experiments produced two protein products, with molecular weights of 77,000 and 52,000. The 52,000 dalton protein is γ; based upon its size, the 77,000 dalton protein was suggested to be τ. While it is clear that the product of the *dnaZ* gene is γ, the identity of the *dnaX* gene product is not certain.

The 2.1 kb region that produces these two polypeptides has a coding capacity, in a single open reading frame, of approximately 80,000 daltons. The production of the two proteins could be accounted for in several ways: (i) *dnaZ* and *dnaX* could utilize separate open reading frames, (ii) the two genes may be encoded on opposite strands of the DNA, (iii) there may be premature termination of translation, (iv) separate, distinct transcripts may be produced which are translated into different products, (v) the mRNA may be processed, or (vi) the 77,000 dalton protein may be proteolytically processed to produce the 52,000 dalton protein. Partial tryptic digests of the two products suggests that they share common sequences (2), indicating that the two proteins are encoded by genes which share a portion of the same open reading frame. This indicates that the two proteins are not encoded by overlapping genes either on the same or opposite strands of the DNA, but rather are produced by one of the other proposed mechanisms.

We have sequenced a 2.7 kb fragment containing the *dnaZX* region in an effort to determine how the two protein products are produced and also to begin understanding how the levels of subunits expressed from this genetic region may be regulated within the cell.

MATERIALS AND METHODS

Bacterial Strains and Plasmids

Escherichia coli K12 strain JM109 (*rec*A1, *end*A1, *gyr*A96, *thi*, *hsd*R17, *sup*E44, *rel*A1, λ^-, Δ(*lac-pro*AB), [F', *tra*D36, *pro*AB, *lac*I^{q}ZΔM15]) was used as the recipient of recombinant M13 clones and was obtained from New England Biolabs. *E. coli* K12 strain AX727 (*purE*$^+$*dnaZ*ts2016 transductant of strain S9080 [F$^-$ *thi*, *lac*, *gal*, *pur*E, *rps*L]) was obtained from J.R. Walker (1) and was used as the host strain for plasmid

pMWZ1101 which was constructed by M. Welch. Bacteriophage M13mp18 and M13mp19 (15) were obtained from New England Biolabs.

M13 Cloning

All enzymes were obtained from New England Biolabs and used according to manufacturer's directions. Ligations of *dnaZX* and M13 DNA were performed at concentrations of 0.8 pmol/ml M13 vector and 4.0 pmol/ml *dnaZX* DNA. Ligated DNA was transformed into *E. coli* JM109 and plated on media containing 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) and isopropyl-thio-galactoside (IPTG). Recombinant phage were selected by the formation of colorless plaques. Bacteriophage M13 amplification was modified by infecting at a low multiplicity of infection (2-3) and allowing phage growth for only 6.5 hours. This avoided the production of a mutant phage which could outgrow the phage containing the wild type *dnaZX* region due to possible deleterious effects of the *dnaZX* gene products. Further, the purification of phage was limited to a low number of passages.

DNA Sequencing

The *dnaZX* DNA cloned in the M13 vectors was sequenced by the Sanger method of dideoxynucleotide chain termination (16) and the incorporation of (α-^{35}S)dATP (New England Nuclear) as described by New England Biolabs with the following modifications. The preparation of template DNA was modified to minimize contaminating DNA by multiple precipitations of the phage in 2% PEG and resuspending the phage in Tris-EDTA plus 0.5% sarkosyl. DNA polymerase I (Klenow fragment) was added directly to the reaction mixture and the sequencing reactions were incubated at 37°C. Extra polymerase was added along with the second addition of deoxynucleotide triphosphates.

The sequencing gels (8% acrylamide/8M urea) were run at constant power to maintain the surface temperature at 60°C. The gels were run for 45 min after the bromophenol blue migrated to the bottom of the gel at which point the samples were reloaded in adjacent lanes. The gels were run for 10 to 15 min after the second bromophenol blue front reached the bottom. This method allowed approximately 250-300 nucleotides to be determined per gel.

Synthetic oligonucleotides were synthesized on a BioSearch oligonucleotide synthesizer and purified on 20% acrylamide/8M urea gels followed by passage over DE52 columns. The purified oligonucleotides were used as primers to allow us to sequence contiguous stretches across the gene.

Sequence analysis was performed using the Beckman "Microgenie" program. The isoelectric point of denatured proteins was calculated using a computer spreadsheet program which calculates the net charge at any given pH using the Henderson-Hasselbach equation.

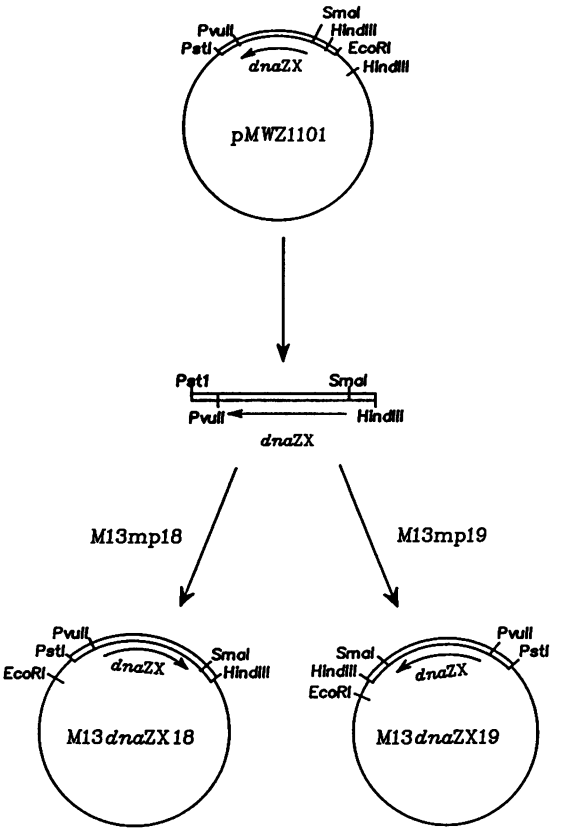

Fig. 1. Cloning of the *dnaZX* region into M13 vectors. Orientation of the *dnaZX* gene (amino to carboxy terminus) is indicated with the restriction sites that were known before sequencing. The double lines represent *E. coli* DNA.

RESULTS

To construct an appropriate vector for sequencing the *dnaZX* region, a 2.7 kb fragment of DNA was subcloned into the M13 vectors, M13mp18 and M13mp19, from plasmid pMWZ1101 which contains the entire *dnaZX* region (Fig. 1). The plasmid was digested with restriction enzymes HindIII and PstI and the purified 2.7 kb fragment was cloned into M13mp18 and M13mp19 that had been cleaved with HindIII and PstI. The vectors M13mp18 and M13mp19 differ only in the orientation of their polylinkers so that a fragment cloned into the same sites is in the opposite orientation in the two vectors, allowing the sequencing of both strands of DNA. Synthetic oligonucleotides were used as primers to sequence contiguous regions across the gene from both directions (Fig. 2).

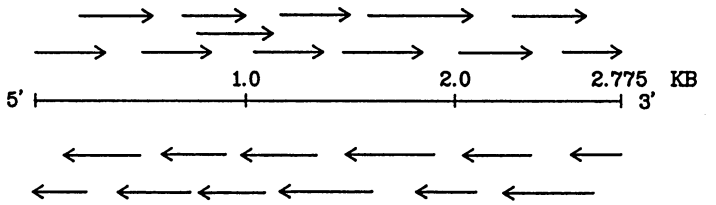

Fig. 2. Sequencing strategy for the *dna*ZX region. The arrows indicate the segment of DNA sequenced from each oligonucleotide primer (all primers were 17 nucleotides long).

The nucleotide sequence was determined for 2775 bases of the *dna*ZX region. This sequence revealed one long open reading frame in this region 1929 nucleotides long (Fig. 3). This open reading frame encodes a slightly acidic protein of 71,147 daltons. The calculated isoelectric point of this protein is 6.8. Alternatively, translation of the protein could initiate 69 bases downstream at the GUG codon. The protein produced would have a molecular weight of 68,451 and an isoelectric point of 6.9. There are no other open reading frames on either strand long enough to encode the 52,000 Dalton protein product seen in the "Maxicell" experiments (2).

There is an open reading frame extending into the region sequenced from the HindIII site and terminating 120 bases 5' to the initiation codon of the *dna*ZX region (Fig. 3). Another open reading frame begins 53 nucleotides 3' to the *dna*ZX termination codon and continues the remaining 227 bases through the PstI site (Fig. 3).

DISCUSSION

The untranslated regions of the *dna*ZX region were studied to attempt to identify possible regulatory regions. The AUG which can be predicted to be the initiator codon of the mRNA is preceded by a sequence AGAG which has some homology to the 3' end of 16S rRNA and is therefore thought to be the ribosome binding site. Based on the predictions of ribosome binding site efficiency, this would be a relatively weak site (17). This is a short Shine and Dalgarno sequence since it has only 3 bases complementary to the 3' end of the 16S rRNA. Longer sequences with more bases complementary to the 3' end of 16S rRNA have been observed to promote more efficient translation of the mRNA (18,19). For optimal translational efficiency, the region 5' to the initiating AUG (except for the Shine and Dalgarno sequence) should preferentially be composed primarily of As and Us (17); the *dna*ZX sequence does not show this composition. The base 3 nucleotides 5' to the initiator codon is especially important and should be an A in a strong ribosome binding site;

in the *dnaZX* sequence a C is found in this position. The codon following the AUG is often either GCU or AAA and in highly expressed genes the preference for GCU is very striking (17,20). The second codon is usually followed by an A in highly expressed genes. The *dnaZX* sequence does not have either the GCU or the AAA codon nor the A following the second codon. Although it is not known if there is any significant secondary structure surrounding the AUG, structures can be predicted which would inhibit the translation of the message by blocking the access of the ribosomes to the ribosome binding site.

The GUG codon at position 635 is preceded by a potential ribosome binding site that should also be weak. The Shine and Dalgarno sequence is stronger than the upstream one but the spacing between it and the initiating codon is less than optimal. Amino acid sequencing of the amino-terminus of the *dnaZX* protein will be required to determine which potential initiation site is used.

The sequence also contains two potential transcriptional promoters. Based on analysis of promoter efficiency, both promoters should be weak (21). The furthest upstream promoter (located at position 325-352) has a T to G switch in the -35 region and a T to C change in the -10 region, both of which would be expected to have severe effects on transcription. The downstream promoter (494-522) has a T to C change and an A to C alteration in the -35 region which would also be expected to greatly decrease transcription. The -10 region shows reasonably good homology to the consensus sequence with no changes which would be expected to have severe deleterious effects, and the spacing between the two sites is correct (17 bases). Regardless of the promoter used and the translational initiating site, the low expression of this gene may, in part, be due to the weak ribosome binding site and promoter.

Another potential promoter is found approximately 430 bases downstream of the proposed initiation codons. This is a strong promoter and may direct the transcription of a separate mRNA. This transcript could be translated starting at position 1358 where an AUG is found. This is also preceded by a potential ribosome binding site which, by the criteria used previously, should be weak.

Additional control on expression may be exerted upstream to the putative promoter where two sequences are found that show homology to the known *dnaA* protein binding sites (7 of 9 bases match the consensus sequence). There are four sites within the *E. coli* origin of replication for the binding of the *dnaA* protein; the *dnaA* protein is thought to induce initiation of replication by binding to these sites (22). The *dnaA* gene product also regulates its own production by binding at two sites within the two promoters for its own structural gene and inhibiting transcription (23,24). It has been proposed that when *dnaA* protein levels in the cell

```
                                                                  5'   AA

                              30                                        60
GCTTACGCTCTCAGCATCGACTTGCTGGTTGAGCGTTACAAAAATGCGGGCATTACCAAA
AlaTyrAlaLeuSerIleAspLeuLeuValGluArgTyrLysAsnAlaGlyIleThrLys

                              90                                       120
GTTGTCGGCACCGAAGCGCGTGGCTTCTTGTTTGGCGCTCCGGTAGCTCTGGGTCTGGGC
ValValGlyThrGluAlaArgGlyPheLeuPheGlyAlaProValAlaLeuGlyLeuGly

                             150                                       180
GTTGGCTTTGTACCGGTCCGTAAACCGGGCAAACTGCCGCGTGAAACCATCAGTGAAACT
ValGlyPheValProValArgLysProGlyLysLeuProArgGluThrIleSerGluThr

                             210                                       240
TACGACCTGGAATACGGCACCGATCAGCTGGAGATCCACGTTGATGCCATCAAACCGGGC
TyrAspLeuGluTyrGlyThrAspGlnLeuGluIleHisValAspAlaIleLysProGly

                             270                                       300
GACAAAGTTCTGGTGGTGGACGACCTGCTGGCAACCGGCGGCACTATCGAAGCGACCGTT
AspLysValLeuValValAspAspLeuLeuAlaThrGlyGlyThrIleGluAlaThrVal

                             330                                  ^^^^^^^^^360
AAACTGATCCGTCGTCTGGGTGGTGAAGTGGCTGACGCTGCGTTCATTATCAACCTGTTC
LysLeuIleArgArgLeuGlyGlyGluValAlaAspAlaAlaPheIleIleAsnLeuPhe

                             390                                       420
GATCTCGGCGGCGAACAGCGTCTCGAAAAACAGGGCATTACCAGCTACAGCCTTGTCCCG
AspLeuGlyGlyGluGlnArgLeuGluLysGlnGlyIleThrSerTyrSerLeuValPro

                  ^^^^^^^^^ 450                                        480
TTCCCGGGCCATTAATTATCGCCAGTCTTGTGCTGCCCACGCTACGGACAGCACAAGATG
PheProGlyHisEnd

                             510                                       540
TGCATTCAGCCTCGCCGTTCTGACGGGGCTGTGTTTAGCATTACCCCTTCGTGAATCCACC

            ****             570                                       600
TTCCAGCGTTTCAGAGCCTGCCAATGAGTTATCAGGTCTTAGCCCGAAAATGGCGCCCAC
                        MetSerTyrGlnValLeuAlaArgLysTrpArgProG

                             *****                                     660
AAACCTTTGCTGACGTCGTCGGCCAGGAACATGTGCTGACCGCACTGGCGAACGGCTTGT
lnThrPheAlaAspValValGlyGlnGluHisValLeuThrAlaLeuAlaAsnGlyLeuS

                             690                                       720
CGTTAGGGCGTATTCATCATGCTTATCTTTTTTCCGGCACCCGTGGCGTCGGAAAAACCT
erLeuGlyArgIleHisHisAlaTyrLeuPheSerGlyThrArgGlyValGlyLysThrS

                             750                                       780
CTATCGCCCGACTGCTGGCGAAGGGGCTAAACTGCGAAACCGGCATTACCGCGACGCCGT
erIleAlaArgLeuLeuAlaLysGlyLeuAsnCysGluThrGlyIleThrAlaThrProC

                             810                                       840
GCGGCGTGTGCGATAACTGTCGTGAAATCGAGCAGGGGCGCTTTGTCGATCTGATTGAAA
ysGlyValCysAspAsnCysArgGluIleGluGlnGlyArgPheValAspLeuIleGluI

                             870                                       900
TCGACGCCGCCTCGCGCACCAAAGTTGAAGATACCCGCGACCTGCTGGATAACGTCCAGT
leAspAlaAlaSerArgThrLysValGluAspThrArgAspLeuLeuAspAsnValGlnT

                             930                                       960
ACGCTCCGGCGCGTGGTCGTTTCAAAGTTTATCTGATCGACGAAGTGCATATGCTGTCGC
yrAlaProAlaArgGlyArgPheLysValTyrLeuIleAspGluValHisMetLeuSerA

                             990                                      1020
GCCACAGCTTTAACGCACTGTTAAAAACCCTTGAAGAGCCGCCGGAGCACGTTAAGTTTC
rgHisSerPheAsnAlaLeuLeuLysThrLeuGluGluProProGluHisValLysPheL
```

```
                                1050                                1080
TGCTGGCGACGACCGATCCACAGAAATTGCCGGTGACGATTTTGTCACGCTGTCTGCAAT
euLeuAlaThrThrAspProGlnLysLeuProValThrIleLeuSerArgCysLeuGlnP

                                1110                                1140
TTCATCTCAAGGCGCTGGATGTCGAGCAAATTCGCCATCAGCTTGAGCACATCCTCAACG
heHisLeuLysAlaLeuAspValGluGlnIleArgHisGlnLeuGluHisIleLeuAsnG

                                1170                                1200
AAGAACATATCGCTCACGAGCCGCGGGCGCTGCAATTGCTGGCACGCGCCGCTGAAGGCA
luGluHisIleAlaHisGluProArgAlaLeuGlnLeuLeuAlaArgAlaAlaGluGlyS

                                1230                                1260
GCCTGCGAGATGCCTTAAGTCTGACCGACCAGGCGATTGCCAGCGGTGACGGCCAGGTTT
erLeuArgAspAlaLeuSerLeuThrAspGlnAlaIleAlaSerGlyAspGlyGlnValS

                                1290                                1320
CAACCCAGGCGGTCAGTGCGATGCTGGGTACGCTTGACGACGATCAGGCGCTGTCGCTGG
erThrGlnAlaValSerAlaMetLeuGlyThrLeuAspAspAspGlnAlaLeuSerLeuV

                                1350                                1380
TTGAAGCGATGGTCGAGGCCAACGGCGAGCGCGTAATGGCGCTGATTAATGAAGCCGCTG
alGluAlaMetValGluAlaAsnGlyGluArgValMetAlaLeuIleAsnGluAlaAlaA

                                1410                                1440
CCCGTGGTATCGAGTGGGAAGCGTTGCTGGTGGAAATGCTCGGCCTGTTGCATCGTATTG
laArgGlyIleGluTrpGluAlaLeuLeuValGluMetLeuGlyLeuLeuHisArgIleA

                                1470                                1500
CGATGGTACAACTTTCGCCTGCTGCACTTGGCAACGACATGGCCGCCATCGAGCTGCGGA
laMetValGlnLeuSerProAlaAlaLeuGlyAsnAspMetAlaAlaIleGluLeuArgM

                                1530                                1560
TGCGTGAACTGGCGCGCACCATACCGCCGACGGATATTCAGCTTTACTATCAGACGCTGT
etArgGluLeuAlaArgThrIleProProThrAspIleGlnLeuTyrTyrGlnThrLeuL

                                1590                                1620
TGATTGGTCGCAAAGAATTACCGTATGCGCCGGACCGTCGCATGGGCGTTGAGATGACGC
euIleGlyArgLysGluLeuProTyrAlaProAspArgArgMetGlyValGluMetThrL

                                1650                                1680
TGCTGCGCGCGCTGGCATTCCATCCGCGTATGCCGCTGCCTGAGCCAGAAGTGCCACGAC
euLeuArgAlaLeuAlaPheHisProArgMetProLeuProGluProGluValProArgG

                                1710                                1740
AGTCCTTTGCACCCGTCGCGCCAACGGCAGTAATGACGCCAACCCAGGTGCCGCCGCAAC
lnSerPheAlaProValAlaProThrAlaValMetThrProThrGlnValProProGlnP

                                1770                                1800
CGCAATCAGCGCCGCAGCAGGCACCGACTGTACCGCTCCCGGAAACCACCAGCCAGGTGC
roGlnSerAlaProGlnGlnAlaProThrValProLeuProGluThrThrSerGlnValL

                                1830                                1860
TGGCGGCGCGCCAGCAGTTGCAGCGCGTGCAGGGAGCAACCAAAGCAAAAAAGAGTGAAC
euAlaAlaArgGlnGlnLeuGlnArgValGlnGlyAlaThrLysAlaLysLysSerGluP

                                1890                                1920
CGGCAGCCGCTACCCGCGCGCGGCCGGTGAATAACGCTGCGCTGGAAAGACTGGCTTCGG
roAlaAlaAlaThrArgAlaArgProValAsnAsnAlaAlaLeuGluArgLeuAlaSerV

                                1950                                1980
TCACCGATCGCGTTCAGGCGCGTCCGGTGCCATCGGCGCTGGAAAAAGCGCCAGCCAAAA
alThrAspArgValGlnAlaArgProValProSerAlaLeuGluLysAlaProAlaLysL

                                2010                                2040
AAGAAGCGTATCGCTGGAAGGCGACCACTCCGGTGATGCAGCAAAAAGAAGTGGTCGCCA
ysGluAlaTyrArgTrpLysAlaThrThrProValMetGlnGlnLysGluValValAlaT

                                2070                                2100
CGCCGAAGGCGCTGAAAAAGCGCTGGAACATGAAAAAACGCCGGAACTGGCGGCGAAGC
hrProLysAlaLeuLysLysAlaLeuGluHisGluLysThrProGluLeuAlaAlaLysL
```

```
                          2130                                2160
TAGCGGCAGAAGCCATTGAGCGCGACCCGTGGGCGGCACAGGTGAGCCAACTTTCGCTAC
euAlaAlaGluAlaIleGluArgAspProTrpAlaAlaGlnValSerGlnLeuSerLeuP

                          2190                                2220
CAAAACTGGTCGAACAGGTGGCGTTAAATGCCTGGAAAGAGGAGAGCGACAACGCAGTAT
roLysLeuValGluGlnValAlaLeuAsnAlaTrpLysGluGluSerAspAsnAlaValC

                          2250                                2280
GTCTGCATTTGCGCTCCTCTCAGCGGCATTTGAACAACCGCGGTGCACAGCAAAAACTGG
ysLeuHisLeuArgSerSerGlnArgHisLeuAsnAsnArgGlyAlaGlnGlnLysLeuA

                          2310                                2340
CTGAAGCGTTGAGCATGTTAAAAGGTTCAACGGTTGAACTGACTATCGTTGAAGATGATA
laGluAlaLeuSerMetLeuLysGlySerThrValGluLeuThrIleValGluAspAspA

                          2370                                2400
ATCCCGCGGTGCGTACGCCGCTGGAGTGGCGTCAGGCGATATACGAAGAAAAACTTGCGC
snProAlaValArgThrProLeuGluTrpArgGlnAlaIleTyrGluGluLysLeuAlaG

                          2430                                2460
AGGCGCGCGAGTCCATTATTGCGGATAATAATATTCAGACCCTGCGTCGGTTCTTCGATG
lnAlaArgGluSerIleIleAlaAspAsnAsnIleGlnThrLeuArgArgPhePheAspA

                          2490                                2520
CGGAGCTGGATGAAGAAAGTATCCGCCCCATTTGATCGTAAGCACAGCTTACGTTCGTCA
laGluLeuAspGluGluSerIleArgProIleEnd

                    *****  2550                                2580
TCCTTAACGTGATTGAGAGAGAAACCTATGTTTGGTAAAGGCGGTCTGGGTAACCTGATG
                             MetPheGlyLysGlyGlyLeuGlyAsnLeuMet

                          2610                                2640
AAGCAAGCCCAGCAGATGCAAGAAAAAATGCAGAAAATGCAGGAAGAGATCGCGCAGCTG
LysGlnAlaGlnGlnMetGlnGluLysMetGlnLysMetGlnGluGluIleAlaGlnLeu

                          2670                                2700
GAAGTCACCGGCGAATCTGGCGCAGGTCTGGTAAAAGTGACCATCAACGGTGCACACAAC
GluValThrGlyGluSerGlyAlaGlyLeuValLysValThrIleAsnGlyAlaHisAsn

                          2730                                2760
TGCCGTCGCGTAGAGATCGACCCGAGCCTGCTGGAAGACGACAAAGAGATGCTGGAAGAC
CysArgArgValGluIleAspProSerLeuLeuGluAspAspLysGluMetLeuGluAsp

CTGGTGGCTGCAG   3'
LeuValAlaAla
```

Fig. 3. Nucleotide sequence of the *dna*ZX region. The corresponding amino acid sequence is shown for the flanking open reading frames as well as for the *dna*ZX gene. The proposed promoters are indicated by underlining, the ribosome binding sites with asterisks, and the putative *dna*A protein binding sites by the symbol (^).

are low it binds to the origin but not to the *dna*A gene; therefore more *dna*A protein can be produced. When the levels of *dna*A protein are high, it binds to the promoter region of the gene and inhibits transcription of the gene. It is possible that one or both of the sites 5' to the *dna*ZX gene are binding sites for the *dna*A protein and when the protein is bound it may cause an inhibition of transcription of the *dna*ZX region. Synthesis of the β subunit of DNA polymerase III holoenzyme is under the control of the *dna*A protein as the structural gene for β (*dna*N) is in an operon with *dna*A (6,7). There is a potential *dna*A binding site in the region upstream to the

*dna*E gene (α) as well (H. Tomasiewicz and C. McHenry, personal communication). The possibility exists that levels of the *dna*A protein may serve to coordinate holoenzyme subunit synthesis.

As stated above, this genetic region contains only one long open reading frame. The production of the two protein products from this region must therefore be accounted for by the usage of two distinct mRNA transcripts, RNA processing, or proteolytic processing. The sequence analysis presented here eliminates the possibility of the two proteins being encoded by two genes in different open reading frames, or on opposite strands on the DNA, and also makes the possibility of premature termination of translation unlikely. This remains an intriguing question as does the possibility of an interesting mechanism existing for the regulation of expression of genes encoding holoenzyme subunits. The same situation seems to exist in *Salmonella typhimurium* where the *dna*Z and *dna*X genes (analogous to those in *E. coli*) are found to be genetically distinct by some criteria yet map within one another (25).

The open reading frame which begins 53 bases 3' to the *dna*ZX gene is preceded by a strong ribosome binding site. Further upstream (position 2330) is a strong promoter sequence. The presence of these two strong sequences indicates that this open reading frame probably is expressed. The product of this predicted gene is not yet known.

ACKNOWLEDGEMENTS

We would like to thank H. Tomasiewicz for help with DNA sequencing and analysis, P. Hagerman, W. Taylor, and J. Mills for synthesis of oligonucleotides, and G. Stormo for analyzing the ribosome binding site and promoter sequences. This work was supported by grant GM-36255 from the National Institutes of Health.

As this work was being completed an abstract appeared by A. Blinkowa, K. Yin, and J. Walker for the UCLA meeting on Mechanisms of DNA Replication and Recombination stating that they have sequenced a 2.2 kb DNA region containing the *dna*Z and *dna*X genes and also identified only one large open reading frame.

REFERENCES

1. Henson, J.M., Chu, H., Irwin, C.A. and Walker, J.R. (1979) *Genetics* **92**, 1041-1059.
2. Kodaira, M., Biswas, S.B. and Kornberg, A. (1983) *Mol. Gen. Genet.* **192**, 80-86.
3. Mullin, D.A., Woldringh, C.L., Henson, J.M. and Walker, J.R. (1983) *Mol. Gen. Genet.* **192**, 73-79.
4. McHenry, C.S. (1985) *Mol. Cell. Biochem.* **66**, 71-85.
5. Gefter, M.L., Hirota, Y., Kornberg, T., Wechsler, J.A., and Barnoux, C. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 3150-3153.
6. Burgers, P.M.J., Kornberg, A. and Sakakibara, Y. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5391-5395.

7. Sako, T. and Sakakibara, Y. (1980) *Mol. Gen. Genet.* **179**, 521-526.
8. Filip, C.C., Allen, J.S., Gustafson, R.A., Allen, R.G. and Walker, J.R. (1974) *J. Bacteriol.* **119**, 443-449.
9. Hübscher, U. and Kornberg, A. (1980) *J. Biol. Chem.* **255**, 11698-11703.
10. Wickner, S. and Hurwitz, J. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 1053-1057.
11. Echols, H., Lu, C. and Burgers, P.M.J. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2189-2192.
12. Horiuchi, T., Maki, H., Maruyama, M. and Sekiguchi, M. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3770-3774.
13. Rowen, L., Kobori, J.A., Scherer, S. (1982) *Mol. Gen. Genet.* **187**, 501-509.
14. Hübscher, U. and Kornberg, A. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 6284-6288.
15. Norrander, J., Kempe, T. and Messing, J. (1983) *Gene* **26**, 101-106.
16. Sanger, F., Nicklen, A.R. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467.
17. Stormo, G.D., Schneider, T.D. and Gold, L. (1982) *Nucl. Acids Res.* **10**, 2971-2996.
18. Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) *Nature New Biol.* **249**, 40-41.
19 Salser, W, (1977) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 985-1002.
20. Dunn, J.J. and Studier, F.W. (1983) *J. Mol. Biol.* **166**, 477-535.
21. Ackerson, J.W. and Gralla. J.D. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 473-476.
22. Fuller, R.S., Funnell, B.E. and Kornberg, A. (1984) *Cell* **38**, 889-900.
23. Atlung, T., Clausen, E.S. and Hansen, F.G. (1985) *Mol. Gen. Genet.* **200**, 442-450.
24. Braun, R.E., O'Day, K. and Wright, A. (1985) *Cell* **40**, 159-169.
25. Engstrom, J., Wong, A. and Maurer, R. (1986) *Genetics* **113**, 499-515.