Identification, molecular cloning and characterization of the gene encoding the \( \chi \) subunit of DNA polymerase III holoenzyme of Escherichia coli

Jeffrey R. Carter, Mary Ann Franden, Julie A. Lippincott, Charles S. McHenry

The University of Colorado Health Sciences Center, Department of Biochemistry, Biophysics and Genetics, B-121, 4200 E. Ninth Avenue, Denver, CO 80262, USA

Received: 25 January 1993 / Accepted: 14 May 1993

Abstract. We have identified a previously reported open reading frame (ORF13) that maps between \( pepA \) and \( valS \) at 96.6 centisomes of the \textit{Escherichia coli} genome as the structural gene for the \( \chi \) subunit of DNA polymerase III holoenzyme. This conclusion is supported by a perfect match of the amino-terminal 24 residues of \( \chi \) with the DNA sequence of ORF13 and a demonstration that ORF13 directs expression of a protein that co-migrates with authentic \( \chi \) on SDS-polyacrylamide gels. ORF13, designated \( hocl \), was isolated from the \textit{E. coli} chromosome and inserted into a \textit{tac} promoter-based expression plasmid to direct production of the \( \chi \) subunit to 5–7% of the total soluble protein. The 3' end of \( hocl \) was sequenced to resolve discrepancies between two published versions.

Key words: DNA polymerase III holoenzyme – \( \chi \) subunit – \( hocl \) gene – Protein overexpression – Codon usage

Introduction

DNA polymerase III holoenzyme is the major replicative enzyme of \textit{Escherichia coli}. The special properties that distinguish holoenzyme from other polymerases include its very high processivity (Wu et al. 1992b), resistance to physiological concentrations of salt (Burgers and Kornberg 1982; Crut et al. 1983; Gries and McHenry 1989) and spermidine (Fay et al. 1981), and a requirement for single-strand DNA binding protein (Geider and Kornberg 1974; Zechel et al. 1975). Additionally, holoenzyme has the ability to interact with primosome proteins at the replication fork (Wu et al. 1992b; Zecher et al. 1992) and to form an asymmetric, dimeric complex that can coordinate leading- and lagging-strand DNA synthesis (Johanson and McHenry 1984; McHenry and Johanson 1984; Maki et al. 1988; Wu et al. 1992a).

Holoenzyme contains at least ten distinct subunits. The core polymerase (pol III; McHenry and Crow 1979) is composed of the \( \alpha \) catalytic subunit (\( dnaE \); Welch and McHenry 1982), \( \varepsilon \) (\( dnaQ \); DiFrancesco et al. 1984; Scheuermann et al. 1983), the 3' to 5' proofreading subunit, and \( \theta \) (\( holE \); Carter et al. 1993; P. Studwell-Vaughan and M. O’Donnell, submitted). The major processivity subunit is \( \beta \) (\( dnaN \); Burgess et al. 1981; Crut et al. 1983), which forms a sliding clamp on primed template DNA (Kong et al. 1992; Stukenberg et al. 1991). The \( \beta \) sliding clamp is loaded onto primed template by the \( \gamma \) complex in a reaction that requires ATP hydrolysis (Wickner 1976; Wickner and Kornberg 1973; O’Donnell 1987). The \( \gamma \) complex contains five subunits (Maki and Kornberg 1988; McHenry et al. 1986; McHenry and Kornberg 1977; O’Donnell 1987): \( \gamma \) (\( dnaX \); Kodaira et al. 1983; Mullin et al. 1983; Flower and McHenry 1986; Lee et al. 1987), \( \delta \) (\( holA \); Carter et al. 1992; Z. Dong, R. Onrust and M. O’Donnell, submitted), \( \delta' \) (\( holB \); Carter et al. 1993; Z. Dong, R. Onrust and M. O’Donnell submitted), \( \psi \) (\( holD \); Carter et al. 1993; R. Crombie, H. Xiao, Z. Dong, R. Onrust, and M. O’Donnell, submitted) and \( \chi \). The \( \tau \) subunit (\( dnaX \); Kodaira et al. 1983; Mullin et al. 1983), dimerizes pol III core (McHenry 1982; Studwell-Vaughan and O’Donnell 1991), specifically stimulates lagging-strand synthesis and prevents formation of abnormally short Okazaki fragments in the presence of physiological glutamate concentrations (Wu et al. 1992a).

Five purified holoenzyme subunits, \( \alpha \), \( \varepsilon \) and \( \beta \) plus \( \gamma \), \( \delta \) or \( \delta' \), reconstitute a polymerase capable of processive synthesis of phage RF II molecules (O’Donnell and Studwell 1990). The \( \chi \) and \( \psi \) subunits are not required for reconstitution of this minimal processive complex in vitro. Evidence that \( \chi \) and \( \psi \) are integral components of holoenzyme is provided by: (1) co-purification with holoenzyme (McHenry et al. 1986), pol III* (holoenzyme lacking the \( \beta \) subunit) (McHenry 1985; Maki et al. 1988) and the \( \gamma \) complex (McHenry et al. 1986; Maki and Kornberg 1988); (2) stimulation of the \( \gamma \) complex DNA-dependent ATPase (Onrust et al. 1991); and (3) stabiliza-
tion of holoenzyme to salt (O’Donnell and Studwell 1990; Crombie and O’Donnell 1992). Further characterization of the contribution of χ to the holoenzyme reaction requires identification of the structural gene for χ and construction of an overproducing strain. We report here our accomplishment of these goals plus determination of the sequence of the structural gene for χ.

**Materials and methods**

**Chemicals.** SDS, acrylamide, N,N’-methylene-bis-acrylamide and Coomassie brilliant blue R-250 were purchased from Bio-Rad. TRIS-HCl, polyvinylpyrrolidone, dextran sulfate, bovine serum albumin, and Ficoll were purchased from Sigma. Urea was from Fisher. γ-[32P]ATP was purchased from ICN. Low-molecular-weight protein standard was from Pharmacia, bacteriophage χ DNA digested with HindIII, restriction enzymes, T4 DNA ligase and T4 polynucleotide kinase were purchased from Promega, and SeaKem LE agarose was obtained from FMC BioProducts. All other chemicals were reagent grade.

**Oligonucleotides.** The 50mer complementary to the structural gene for χ was 5'-ATGAAAAACCGAGCTTC-TACCTCTGGAAATGGACACCCGCTGATGG-3'. The two oligonucleotides used in construction of the χ overproducing plasmid to reconstruct the 5' end of the gene were 5'-GATCTGAGGATTAAATAATTGA-AAATGCTACCTTTTATCTGTTGGATAACGATA-CTACTGTTGACGCTTAAGATACCTGGCGCGG-CTTGCA-3' and 5'-GGGCGGCGCAAGTTATCCATGAGCGTCAAACGATGATGCTTGTTAATCCAAACAGAT-AAAAGGTACATTTTTCATATTATATACCC-3'.

**Bacterial strains, plasmids and media.** XL1Blue (Stratagene) was the host strain for routine plasmid transformation experiments. MAF102 (Carter et al. 1992), a lexA3, uraD (Washburn and Kushner 1991) derivative of the wild-type strain MG1655 (Guyer et al. 1980), was the source of *E. coli* chromosomal DNA. Holoenzyme was purified from MGC100, a derivative of RS320 [Δlac- (1P^-O^-Z^-Y^-A^-) U169 Δon araD139 strA supF; gift of R. Sclafani, University of Colorado Health Sciences Center; Carter et al. 1992] that is resistant to a bacteriophage that often contaminates our fermentor. HB101 [supE44 hsdS20 (rpsB-m') recA13 ara-14 proA2 lacY1 galK2 rpsL20 xyl-5 mtl-l; Boyer and Roulland-Dussoix 1969] was the host strain for IPTG-induced overexpression of the χ subunit. pBlueScript II SK+ (Stratagene) was the primary cloning vector. Plasmid pJCl, used to overexpress the HIV nucleocapsid protein (You and McHenry, submitted) was used as a positive control plasmid in the overexpression experiment. pBBMD11 (Fürste et al. 1986; McHenry et al. 1990) was the overexpression vector into which the gene encoding χ was cloned to create the χ overproducer. L broth (Miller 1972) was used for routine bacterial growth. Medium for overexpression of χ contained 1.4% yeast extract, 0.8% peptone, and 1.2% potassium phosphate pH 7.5. When required, ampicillin and tetracycline were used at 50 and 10 μg/ml, respectively.

**Plasmid purification.** Large-scale plasmid purification was performed using the alkaline-SDS lysis method (Birnboim and Doly 1979). Further purification was achieved by processing plasmid DNA through two CsCl-ethidium bromide equilibrium density gradients in a Sorvall T127.0 rotor at 15°C for 20 h at 188 000 × g or 40 h at 132 000 × g.

**Amino-terminal sequencing of χ.** Preparation of holoenzyme subunits for amino-terminal sequencing has been described (Carter et al. 1992). Holoenzyme was purified (Oberfelder and McHenry 1987) and concentrated 35-fold in a collodion bag apparatus (Schleicher and Schuell) to yield about 500 μg of holoenzyme in 50 μl. Individual holoenzyme subunits were separated in a 7.5-17.5% SDS-polyacrylamide gel (Laemmli 1970), transferred onto a ProBlott PVDF membrane (Applied Biosystems) and stained with Coomassie brilliant blue R-250. Densitometric quantitation of the blot was performed as described (Carter et al. 1992) using a Molecular Dynamics Computing Densitometer, Model 3000, running the software ImageQuant version 3.0. About 10 μg of χ subunit had been transferred to the blot. The χ subunit was excised from the blot and sequenced at the University of California Cancer Center Protein Microsequencing Core Facility on an Applied Biosystem 477A Protein Sequencer.

**Computer searches and analysis of DNA/protein structure.** Protein sequences contained in GenPept (version 64.3) and Swiss-Prot (version 18.0) were searched for similarity to the amino-terminal sequence of χ using the FASTA program, which uses the Pearson and Lipman (1988) search method. Protein translation, molecular mass determination, and codon usage analysis were performed using the protein and DNA analysis software PC/GENE version 6.5 (IntelliGenetics).

**Preparation of 5' end-labeled oligonucleotide.** Fifty picomoles of the 50mer oligonucleotide complementary to the structural gene for χ was 5' end-labeled by incubation with 160 μCi γ-[32P]ATP (600 Ci/mmol, 160 μCi/μl), and 10 units of T4 polynucleotide kinase in 50 μl of kinase buffer (50 mM TRIS-HCl pH 7.5, 10 mM MgCl2, 10 mM DTT) at 37°C for 30 min. Unincorporated nucleotide was removed using a G-25 Sephadex (Pharmacia) gel-filtration spin column equilibrated and developed with 20 mM TRIS-HCl pH 8.0, 1 mM EDTA, and 50 mM NaCl. Purified oligonucleotide was stored at −20°C.

**Agarose gel electrophoresis.** For routine analysis of plasmid DNA, gels containing 0.7-1.0% agarose in TRIS-borate buffer (89 mM TRIS base, 2.75 mM EDTA, 89 mM boric acid) were electrophoresed at 2-10 V/cm in TRIS-borate buffer in a Hoefer submarine gel electrophoresis apparatus at room temperature for 5-1 h. Gels were stained with ethidium bromide and illuminated with ultraviolet light.
were stained in a 0.8 µg/ml solution of ethidium bromide, illuminated with 254-nm light from a Fotodyne transilluminator and photographed using Polaroid type 667 film. Size-fractionation of chromosomal DNA for purification or for Southern transfer was performed as described above, except that the gels were electrophoresed at 4°C and 0.5–2 V/cm for 16–20 h. Excision of restriction fragments without UV irradiation was performed as described (Carter et al. 1992). DNA was eluted from the gel using an Elutrap apparatus (Schleicher and Schuell) and then ethanol-precipitated.

For purification of plasmid DNA restriction fragments, gels containing 0.6–0.8% agarose in TRIS-acetate buffer (40 mM TRIS-HCl pH 8.0, 20 mM acetic acid, 10 mM EDTA) were electrophoresed for 2–5 h at room temperature and at 3–5 V/cm in TRIS-acetate buffer. DNA-containing gel slices were excised without UV irradiation and DNA was purified from the gel using the GeneClean II DNA purification kit (Bio 101).

Identification of chromosomal restriction fragments complementary to the 50mer oligonucleotide. Southern blotting (Southern 1975) and hybridization were performed as described (Carter et al. 1992). Chromosomal DNA digested with either BamHI, BglII, HindIII, EcoRI, EcoRV, PstI or with all possible pairs of these enzymes was separated by agarose gel electrophoresis and transferred to GeneScreen nylon membrane (New England Nuclear). The 50mer was 5' end-labeled and hybridized to the immobilized DNA following instructions provided with the GeneScreen membrane. The blot was autoradiographed using a Molecular Dynamics phosphorimager screen, which was scanned and analyzed on a Molecular Dynamics Phosphorimager, Model 400E using the ImageQuant version 3 program.

Colony hybridization. Screening for a clone containing the structural gene for the χ subunit was performed by colony hybridization according to the method of Sambrook et al. (1989), using 5' end-labeled 50mer oligonucleotide as the hybridization probe.

DNA sequencing. DNA sequencing was performed by dideoxy chain-termination method (Sanger et al. 1977). The template DNA was plasmid pJRC110, the original clone of chromosomal DNA containing the gene encoding χ. Both strands of the DNA were sequenced.

Determination of the amino-terminal sequence of χ and tentative identification of its structural gene

To obtain the amino-terminal sequence of the χ subunit, holoenzyme was concentrated to 10 mg/ml as described (Carter et al. 1992), loaded into one well of a 7.5–17.5% SDS-polyacrylamide gel and subjected to electrophoresis. The separated subunits were transferred to a PVDF membrane. The portion of the membrane containing the χ subunit was excised from the blot and the amino-terminal sequence of χ was determined. The first 24 residues were determined to be MetLysAsnAlaTyrPheTyrLeuLeuAspAsnAspTyrTyrTyrValAspGlyLeuSerAlaValGluGlnLeu.

A search of two protein databanks, SwissProt (version 18.0) and GenPept (version 64.3), revealed a previously uncharacterized open reading frame (ORF13; Colloms 1990) located between pepA (aminopeptidase A) and valS (valyl-tRNA-synthetase) that encoded a protein with the same amino-terminus as χ. The complete DNA sequence for ORF13 was derived by combining overlapping DNA sequences reported by different groups. The first 306 bp of the open reading frame were reported by Colloms (1990) and the remainder of the open reading frame was reported as sequence upstream of the valS gene by Härtlein et al. (1987) and Heck and Hatfield (1988a). Discrepancies between the sequences reported by the latter two groups resulted in disagreement as to the location of the stop codon of the open reading frame. However, translation of ORF13 terminated with either predicted stop codon resulted in proteins approximately the size of χ, about 16 kDa as measured by SDS-polyacrylamide gel.
electrophoresis. Based on the perfect 24 amino acid match, we tentatively identified ORF13 as the structural gene for \( \chi \). Confirmation of this required resolving the discrepancy in the position of the stop codon, demonstration that the gene actually expressed a protein, and demonstration that the protein expressed co-migrated with authentic \( \chi \) on SDS-polyacrylamide gels.

**Cloning the structural gene for \( \chi \)**

The wild-type ORF13 was cloned by synthesizing an oligonucleotide identical to the unambiguously determined sequence of the 5' end of ORF13 which agrees completely with the experimentally determined amino-terminal sequence of \( \chi \). This oligonucleotide, 5'-ATG-AAAAACGCGACGTCTTACTATTGCAATTGA-CACCACCGTGGATGG-3', was 5' end-labeled and used as a hybridization probe in a Southern blot of MAF102 chromosomal DNA digested with the restriction enzymes \( \text{PstI}, \text{EcoRV}, \text{HindIII}, \text{EcoRI}, \text{BamHI} \) and \( \text{BglII} \) used singly and in all two-enzyme combinations. A unique restriction fragment was identified by autoradiography in each digest (Table 1), indicating a single locus of DNA complementary to the probe. The sizes of these restriction fragments allowed determination of a restriction map of the region of the chromosome containing the structural gene for \( \chi \) (Fig. 1) that confirmed the map assembled from the \( E. \ coli \) chromosomal restriction map (Kohara et al. 1987; Rudd et al. 1990; Rudd 1992) and DNA sequence (Härtlein et al. 1987; Heck and Hatfield 1988a, b) and gave us confidence that a 2.2-kb \( \text{BglII} \) fragment contained the entire sequence of ORF13.

In 1982, a chromosomal DNA digest was prepared by digesting with \( \text{BamHI} \) and \( \text{BglII} \). A PCR amplification was performed on the \( \text{BglII} \) fragment to isolate the entire sequence of ORF13.

**Table 1. Sizes* of MAF102 chromosomal restriction fragments complementary to ORF13**

<table>
<thead>
<tr>
<th>Restriction Enzyme</th>
<th>( \text{PstI} )</th>
<th>( \text{EcoRV} )</th>
<th>( \text{HindIII} )</th>
<th>( \text{EcoRI} )</th>
<th>( \text{BglII} )</th>
<th>( \text{BamHI} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{PstI} )</td>
<td>20.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{EcoRV} )</td>
<td>0.8</td>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{HindIII} )</td>
<td>2.0 b</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{EcoRI} )</td>
<td>3.1</td>
<td>0.8</td>
<td>1.7</td>
<td>23.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{BglII} )</td>
<td>2.2</td>
<td>2.2</td>
<td>1.4</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>( \text{BamHI} )</td>
<td>3.3</td>
<td>2.8</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>3.3</td>
</tr>
</tbody>
</table>

*Sizes expressed in kb

b Indicates that the fragment was not detected. The reported size of this fragment is 456 bp (Colloms 1990), which would have run off the gel used to determine the fragment sizes.

**Fig. 1. Restriction map of the region of the chromosome that contains the putative \( \chi \) structural gene. The restriction map was constructed from the information presented in Table 1. Restriction enzyme abbreviations are: B, \( \text{BamHI} \); G, \( \text{BglII} \); E, \( \text{EcoRI} \); V, \( \text{EcoRV} \); H, \( \text{HindIII} \); P, \( \text{PstI} \). The gene encoding \( \chi \) is depicted by an arrow and the truncation of \( \text{valS} \) is depicted by a pair of small vertical lines. The \( \text{EcoRV} \) sites in parentheses indicate that the 0.8-kb \( \text{EcoRV} \) fragment was unambiguously located only after sequencing the DNA. The regions containing the \( \text{valS} \) and \( \text{pepA} \) genes were taken from the literature.

**Sequencing the structural gene for \( \chi \)**

The sequence of the 3' one-third of the putative gene encoding \( \chi \) has been reported by two groups who sequenced \( \text{valS} \). The sequence reported by Härtlein et al. (1987) predicts that the gene encoding \( \chi \) extends 14 bases into the \( \text{valS} \) coding region, whereas the sequence reported by Heck and Hatfield (1988a) predicts that the gene ends 73 bases before the \( \text{valS} \) initiation codon. To resolve this discrepancy and define the termination codon of the gene encoding \( \chi \), the 3' one-half of the gene was sequenced (Fig. 2A) and compared with the reported sequences (Fig. 2B). Both of the previous reports contained errors. Our sequence of the 3' region of ORF13 was combined with the reported 5' sequence of ORF13 (Colloms 1990) to form a corrected sequence of ORF13, the putative gene encoding \( \chi \) (Fig. 3).

The putative \( \chi \) structural gene contains 441 bases of coding sequence and is predicted to encode a 147-amino acid protein of 16 599 Da. The termination codon of the gene overlaps the initiation codon of \( \text{valS} \), suggesting that translational coupling between these two genes is partially responsible for \( \text{valS} \) expression (Gold and Stormo 1987). Within the 159 bp between the \( \text{pepA} \) termination codon and the initiation codon of the putative gene encoding \( \chi \) is a potential promoter that contains two elements which are similar to the consensus —35 and —10 sequences which are required for translational coupling, the ribosome-binding site (Fig. 3) and the 2 G/C box from the sequence pairing at the putative site (Stormo 1987).
1...TTG CTG GCA CAG TTC CTG TTA AAC CCG GCT GGG TTT AAC

stop pepA

GOCC GAA CGG TAA TGGCGTCAAGCAAGGCGCTTTATCAGcGGAAGC

-10

GCG GAA CAG TAA TGGCGTCAAGCAAGGCGCTTTATCAGcGGAAGC

+2

GCG GAA CAG TAA TGGCGTCAAGCAAGGCGCTTTATCAGcGGAAGC

138

CACGACACGCCACCTGCGCCGGGCGCTGGATACGATGCGATCTTT

205

TCCGCCACAAAAGGCCCCCA TTT ATG AAA AAC GCG ACG TTC TAC

MET Lys Ala Thr Phe Tyr

232

CTT CTG GAC AAT GAC ACC ACC GTT GAT GCG TTA ACG GCC

Leu Leu Asp Asn Thr Thr Asp Gly Leu Ser Ala

271

GTT GAG CAA CTG GTG TTT GAA ATT GCC GCA GAA GCT TGG

Val Glu Gin Leu Val Cys Glu Ile Ala Ala Glu Arg Trp

310

CGG AGC GGT AAG CGC GTG ATC GTT GAA GAT GAA

Arg Ser Gly Lys Arg Val Leu Ile Ala Cys Gly Asp

349

AAG CAG GCT TAC CGG CTG GAT GAA GCC CGT TGG CGG CGT

Lys Gin Ala Tyr Leu Asp Arg Val Pro Ala Ala Leu Arg

388

CCG GCA GAA AGC TTT GGT CCG CAT AAT TTA GCG GAA GAA

Pro Ala Glu Ser Phe Val Pro His Asn Leu Ala Gly Gly

427

GGA CCG CGC GGC GGT GCA GCG GTG GAG ATG TTC TGG CCG

Gly Pro Arg Gly Gly Ala Pro Val Glu Ile Ala Leu Trp

466

CAA AAG CAG ACC AGC CGG CGC GAT ATA TTT ATT ACG GAA

Gln Lys Arg Ser Ser Ser Arg Arg Asp Ile Leu Ile Ser

505

CTG CGA ACA ACC TTT GCA GAT TTT GCC ACC GCT TTC ACA

Leu Arg Thr Ser Phe Ala Asp Phe Ala Thr Phe Thr

544

GAA GTG GTA GAC TTC GTT CCT TAT GAA GAT TCT CTG AAA

Glu Val Val Asp Phe Val Pro Tyr Glu Asp Ser Leu Lys

583

CAA CTG GCG CGC GAA CGG TAT AAA GCC TAC CGG GTG GCT

Glu Leu Ala Arg Glu Arg Tyr Lys Ala Tyr Arg Val Ala

622

GTT TTC AAC CTG AAT ACG GCA ACC TGG AAA TAT TGG AAA

start valS

Gly Phe Asn Leu Asn Thr Ala Thr Trp Lys Thr

661

AGA CAT ATA ACC CAC AA

start ORF13. The sequence from

The 159 bp between pepA and the putative χ structural gene contains two repetitive extragenic palindromic (REP) elements (Higgins et al. 1982; reviewed in Gilson et al. 1984; Lupski and Weinstock 1992) separated by 23 bp. The REP element is a consensus of about 31 bp (Gilson et al. 1984) and is estimated to occur several

The 159 bp between pepA and the putative χ structural gene contains two repetitive extragenic palindromic (REP) elements (Higgins et al. 1982; reviewed in Gilson et al. 1984; Lupski and Weinstock 1992) separated by 23 bp. The REP element is a consensus of about 31 bp (Gilson et al. 1984) and is estimated to occur several
hundred times in the extragenic regions of the *E. coli* and *Salmonella typhimurium* chromosomes, possibly comprising up to 1% of the chromosome (Stern et al. 1984). These elements are reported to increase mRNA stability (Newbury et al. 1987) and could contribute to transcription termination (reviewed in Gilson et al. 1984). The consensus REP sequence contains inverted repeats separated by at most 25 bp, but more typically 5 or fewer base pairs (Gilson et al. 1984); the inverted repeats of the two REP elements upstream of the putative gene encoding $\chi$ are separated by 2 and 4 bp, respectively. (Fig. 3; Stirling et al. 1989; D. Sherratt, Genbank accession number X15130). The potential promoter identified for the putative gene encoding $\chi$ is almost entirely within the first REP element.

### Codon usage

A set of eight are codons (AUA, UCG, CCU, CCC, ACG, CAA, AAU, and AGG) was identified following a survey of 25 highly expressed genes of *E. coli* (Konigsberg and Godson 1983). It has been proposed that rare codon usage contributes to the low expression of regulatory proteins (Konigsberg and Godson 1983) that exist in very low cellular concentrations. There are approximately 10–20 copies of holoenzymes per cell. The coding region of the putative gene encoding $\chi$ was analyzed to address the possibility that rare codon usage in the gene contributes to its weak expression. Rare codons are used at a frequency of 6.8% in the putative gene encoding $\chi$, comparable to the values for three other holoenzyme genes, *dnaX*, *holA* and *dnaQ*, and higher than the values for the holoenzyme genes *dnaE* and *dnaN* (Table 2). Likewise, the value for the putative gene encoding $\chi$ is 1.5-fold higher than the average value (4.2% rare codons) for the 25 highly expressed genes surveyed by Konigsberg and Godson (1983), and 4-fold higher than the value for the subset of ten very highly expressed ribosome protein genes (1.7% rare codons). It is therefore possible that rare codon usage contributes to the low cellular abundance of $\chi$.

#### Construction of a $\chi$ overproducing plasmid

To determine whether ORF13 expresses a protein and whether it is the size of $\chi$ required construction of a plasmid that overexpresses ORF13 (Fig. 4). The expression vector chosen was pBBMD11 (Fürste et al. 1986; McHenry et al. 1990), a *dnaX* overexpression plasmid.

![Diagram of plasmid construction](image)

Table 2. Percentage of rare codons* in the three possible reading frames of nine *Escherichia coli* genes

<table>
<thead>
<tr>
<th>Genes</th>
<th>Frame 1</th>
<th>Frame 2</th>
<th>Frame 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF13</td>
<td>6.8</td>
<td>12.2</td>
<td>6.8</td>
</tr>
<tr>
<td>pepA</td>
<td>5.4</td>
<td>12.5</td>
<td>9.3</td>
</tr>
<tr>
<td>valS</td>
<td>2.3</td>
<td>13.9</td>
<td>11.9</td>
</tr>
<tr>
<td>dnaE</td>
<td>4.9</td>
<td>13.4</td>
<td>10.8</td>
</tr>
<tr>
<td>dnaN</td>
<td>4.1</td>
<td>12.5</td>
<td>11.4</td>
</tr>
<tr>
<td>dnaX</td>
<td>6.8</td>
<td>15.6</td>
<td>6.7</td>
</tr>
<tr>
<td>dnaQ</td>
<td>7.0</td>
<td>13.2</td>
<td>11.5</td>
</tr>
<tr>
<td>holA</td>
<td>8.7</td>
<td>11.4</td>
<td>8.7</td>
</tr>
<tr>
<td>dnaG</td>
<td>11.3</td>
<td>12.4</td>
<td>12.9</td>
</tr>
</tbody>
</table>

* Rare codons include AUE (Ile), UCG (Ser), CCU and CCC (Pro), ACG (Thr), CAA (Gln), AAU (Asn), and AGG (Arg) (Konigsberg and Godson 1983)


* Coding frame

Fig. 4. Construction of a $\chi$ overexpression plasmid. Plasmid pBBMD11 was digested with *Bgl*II and *Pst*I to remove all of the *dnaX* gene. A synthetic, double-stranded oligonucleotide was inserted into this plasmid to create pMAF50. The oligonucleotide contains *Bgl*II and *Pst*I half sites, and a consensus ribosome-binding site 9 A/T bp upstream, of the ATG start codon that begins a sequence of codons encoding the first 17 amino acid residues of $\chi$. The 17 codons are followed by *Afl*I and *Not*I sites. A 1.140 bp *Afl*I-Not*I restriction fragment containing all but the 5' 17 codons of the wild-type gene encoding $\chi$ was removed from pJRC110, the original clone of the gene, and cloned into pMAF50 to generate pMAF51, the $\chi$ overexpression plasmid. Abbreviations: are: *Pac* tac promoter; *Afl*I; *Bam*HI; *G, Bgl*II; *N, Not*I; *P, Pst*I. The *dnaX* gene is represented by stippled fill, the oligonucleotide by solid fill, the 2.2 kb *Bgl*II fragment containing the gene encoding $\chi$ by no fill, the tac promoter-based vector by horizontal fill, pBlueScript SK+ by vertical fill, and the gene encoding $\chi$ by the curved arrow.

### Overexpression

The overexpression of the gene is induced with IPTG (isopropyl-thio-

Fig. 5. Overexpression of the gene is induced with IPTG (isopropyl-thio-

massee bromide) and protein expression is induced by a protein expression pe-

Fig. 1. Overexpression of the gene is induced with IPTG (isopropyl-thio-

pacing plasmid, pBBMD11, was purified homogeneously.

94–

68–

43–

20–

14,4–
mid. ORF13 was inserted in place of the dnaX coding sequence to create the plasmid pMAF51 (Fig. 4). Transcription from the expression vector is driven by the strong tac promoter, ensuring that any protein produced by ORF13 would be detectable.

A key step in the overproduction strategy was the use of a double-stranded oligonucleotide to replace the 5' end of the gene. The oligonucleotide included 17 codons that encode the first 17 amino acid residues of the χ subunit. These codons contain 14 deviations from the wild-type sequence, that result in codons synonymous to the wild-type codons. The alternate codons were included to produce high-usage codons and to change the DNA sequence, possibly reducing negative regulation of χ expression.

Overexpression of χ

The overexpression plasmid pMAF51 was transformed into HB101 to produce MAF151, which was used in an IPTG-induction experiment to test for production of a protein with the same electrophoretic mobility as the χ subunit of holoenzyme. Two control strains, HB101 and HB101 containing pJC1, which expresses the HIV nucleocapsid protein from the same tac promoter-based expression plasmid as pMAF51, were included in this experiment. MAF151 produces a protein with the same electrophoretic mobility as χ obtained from purified DNA polymerase III holoenzyme; production of this protein is dependent on IPTG induction (Fig. 5, compare lanes 5 and 6). HB101, the negative control strain, does not produce this protein, even when induced with IPTG (Fig. 5, compare lanes 1 and 2). The positive control strain, HB101 with pJC1, demonstrates IPTG-dependent production of the 6380-Da HIV nucleocapsid protein (Fig. 5, compare lanes 3 and 4), but does not produce a protein with the mobility of χ. Lanes 5 and 6 of the gel in Fig. 5 were analyzed by densitometry (data not shown). It was calculated that χ is expressed to a level equivalent to 5–7% of total soluble protein.

Discussion

The goal of this work was to isolate the structural gene for the χ subunit of DNA polymerase III holoenzyme. We used a reverse genetic approach in which the amino-terminal 24 amino acid residues of χ were determined and used to identify a DNA sequence for this segment of the structural gene. A perfect match to the 24-amino acid sequence was identified as being the amino-terminus of a protein predicted to be encoded by an open reading frame (ORF13; Colloms 1990) downstream of pepA, that had been mapped to 96.5 min of the chromosome. ORF13 was isolated from the chromosome and subcloned into a tac promoter-based expression vector to determine if it expressed a protein and if that protein was the same size as authentic χ found in purified holoenzyme. The resulting expression plasmid produced a protein to about 5–7% of total soluble protein that comigrates with χ from holoenzyme. Taken together, these data indicate that we have isolated the structural gene for the χ subunit. This gene has also been independently isolated by the O’Donnell laboratory (Crombie and O’Donnell 1992; R. Crombie, H. Xiao, Z. Dong, R. Onrust, M. O’Donnell, submitted). As suggested by Dr. K. Marians (Sloan Kettering), both the O’Donnell laboratory and our laboratory have agreed to call the gene holC.

To resolve the discrepancy between the two reported sequences (Härtilt et al. 1987; Heck and Hatfield 1988a) of DNA upstream of valS, we sequenced the 3' one-half of holC. We found that neither of the reported sequences proved to be entirely correct, and that the holC termination codon overlaps the valS initiation codon. The errors found in the two reported sequences were in DNA upstream of valS, the gene of interest to the authors. Typically, sequences outside of actual coding regions are not treated with the same rigor as coding DNA, but are nevertheless reported to aid researchers who are interested in identifying additional genes.

The codon usage of holC was analyzed and a codon bias, consistent with the theory that rare codons contribute to low gene expression (Konigsberg and Godson 1983), was found. However, as discussed by Anderson and Kurland (1990), a high percentage of rare codons does not necessitate a low level of expression. Nevertheless, the possibility of codon bias negatively influencing gene expression prompted us to eliminate several rare codons from the beginning of the gene during construction of the holC overexpression plasmid.

Analysis of the DNA sequence in the vicinity of holC revealed several features that could provide genetic regula-
tion of holC, pepA and valS. We have identified two sequences, with similarity to the consensus — 35 and 10 promoter elements, that could form a holC promoter beginning 108 bp upstream of the holC initiation codon, and a ribosome-binding site 7 bases upstream of the holC initiation codon. The putative holC promoter is almost entirely within the first of two REP elements found in the pepA-holC intergenic region. REP elements are sequences of about 35 bp that contain an inverted repeat and occur with great frequency in intergenic regions of the E. coli chromosome (Stern et al. 1984; Lupski and Weinstock 1992). One role of REP elements is to protect mRNA from attack by 3'→5' exonucleases (Newbury et al. 1987), perhaps following endonucleolytic processing of non-translated mRNA.

At the 3' end of holC, we find that the holC termination codon overlaps the initiation codon of valS. Analysis of the DNA sequence revealed no evidence for a rho-independent transcription terminator following the holC termination codon. Thus, transcription of holC is likely to proceed through valS. Given the overlap of initiation and termination codons, the possibility exists that translation of valS mRNA occurs through translational reinitiation following translation of holC mRNA (Gold and Stormo 1987), which could supplement valS expression directed from the two active valS promoters (Heck and Hatfield 1988a).

The products of the pepA, holC and valS genes are involved in seemingly unrelated activities in the cell. valS produces vanyl-tRNA synthetase, pepA is the gene for aminopeptidase A, which contributes to protein degradation (reviewed in Miller 1987) and which is required for cer-mediated site-specific recombination in ColEl (Summers and Srborn 1984; Strirling et al. 1989), and holC produces the χ subunit of DNA polymerase III holoenzyme. We find this theme to be very common for genes encoding holoenzyme subunits. For example, dnaE (a polymerase subunit of holoenzyme) is in an operon that includes cdxA (CDP-diglyceride synthetase), pX (UDP-N-acetylglucosamine acetyltransferase) and pXb (lipid A disaccharide synthetase) (H. Tomaszewicz and C. McHenry, manuscript in preparation), dnaX (γ and ω subunits of holoenzyme) is in an operon with recR (a recombination protein) and apt (adenine phosphoribosyltransferase) (Mahdi and Lloyd 1989), and holA (δ subunit of holoenzyme) seems to be collocated with rlpB (a rare lipoprotein) and perhaps with several other proteins involved in cell envelope biogenesis (Carter et al. 1992).

At present, the role of χ in DNA polymerase III holoenzyme is not well understood. In combination with the ω subunit, χ moderately stimulates the DNA-dependent ATPase activity of χ (Ornust et al. 1991), and stabilizes reconstituted DNA polymerase III holoenzyme to salt (Crombie and O’Donnell 1992). We are presently purifying large quantities of the χ subunit from the overproducing strain described in this work. Having large quantities of the χ subunit available will facilitate defining its contribution to the replicative reaction of χ polymerase III holoenzyme.

Acknowledgements. We thank Dr. David Sherratt for providing helpful information on the sequence downstream of pepA. Robin Gutell for performing the initial GenPept and Swiss-Prot database searches, Joline Mills for oligonucleotide synthesis, and Dr. Michael O’Donnell and colleagues for exchanging DNA sequence of holC with us prior to publication. This work was supported by NIH grant GM36255 and American Cancer Society grant NP-819 to C.S.M.

References


Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513–1523


Flower AM, McHenry CS (1986) The adjacent dnaZ and dnaX genes of Escherichia coli are contained within one continuous open reading frame. Nucleic Acids Res 14:8091–8101


Gold L, Storer LK, Low KD, Escherichia coli, a molecular biology manual, D, N. Griepe MA, inhibition of. 264:14


Härtelein N, Escherichia coli, a physical and chemical overview. 1982, J Biol Chem 257:11474–11478


Stirling CJ, Colloms SD, Collins JF, Szatmari G, Sherratt DJ (1989) xerB, an Escherichia coli gene required for plasmid ColE1 site-specific recombination, is identical to pepA, encoding aminopeptidase A, a protein with substantial similarity to bovine lens leucine aminopeptidase. EMBO J 8: 1623–1627
Wickner W, Kornberg A (1973) DNA polymerase III star requires ATP to start synthesis on a primed DNA. Proc Natl Acad Sci USA 70: 3679–3683

Deletion in Chromatium matuticium

Evelyn M. S. Hoeffler

Institut Our Lady of the Lake

Received 26 June 1989

Abstract

Controlled deletions in the genome of Chromatium matuticum (the lower phototrophic bacterium with the largest genome, 3.8 kb) were obtained by means of insertional inactivation followed by integrated selection. The sequencing of the deletion clones indicates that, other than the 200-bp deletion in the mutD gene, no other mutagenic effectors were found to lead to deletions in the genome. These deletions are transmissible in vivo and, when combined with a mat + mutation in crosses of C. matuticum, the deletion results in loss of fertility. No functional compensation was observed at the mutD sites.

Key word(s): Chromatium matuticum, Ectoplasma

Introduction

For several years, we have been exploring the opportunities of using the microorganism Chromatium matuticum as a model for the study of the genetics of the microorganisms of the purple bacteria. In particular, we are interested in the effects of radiation on the microorganisms, and we have been using a number of different approaches to that end. While the results of our studies are not yet complete, we believe that the microorganisms of this group are well suited for such experiments, and that the data obtained will be of great importance in understanding the effects of radiation on the microorganisms of the purple bacteria. Furthermore, we believe that the results of these studies will be of great importance in understanding the effects of radiation on the microorganisms of the purple bacteria, and that the data obtained will be of great importance in understanding the effects of radiation on the microorganisms of the purple bacteria.