CRYSTALLIZATION NOTE

Preliminary X-Ray Crystallographic and NMR Studies on the Exonuclease Domain of the $\epsilon$ Subunit of Escherichia coli DNA Polymerase III

Samir Hamdan,* Susan E. Brown,* Phillip R. Thompson,* Ji Yeon Yang,* Paul D. Carr,* David L. Ollis,* Gottfried Otting,† and Nicholas E. Dixon*

*Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia; and †Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden

Received April 6, 2000, and in revised form May 24, 2000

The structured core of the N-terminal 3'→5' exonuclease domain of $\epsilon$, the proofreading subunit of Escherichia coli DNA polymerase III, was defined by multidimensional NMR experiments with uniformly $^{15}$N-labeled protein: it comprises residues between Ile-4 and Gln-181. A 185-residue fragment, termed $\epsilon$(1–185), was crystallized by the hanging drop vapor diffusion method in the presence of thymidine-5' monophosphate, a product inhibitor, and $\text{Mn}^{2+}$ at pH 5.8. The crystals are tetragonal, with typical dimensions 0.2 mm $\times$ 0.2 mm $\times$ 1.0 mm, grow over about 2 weeks at 4°C, and diffract X-rays to 2.0 Å. The space group was determined to be $P\bar{4}_12_12_1$ ($n = 0, 1, 2, 3$), with unit cell dimensions $a = 60.8$ Å, $c = 111.4$ Å. © 2000 Academic Press

Key Words: crystallization; dnaQ; DNA replication; exonuclease; proofreading; protein NMR; structure.

DNA polymerase III (pol III) holoenzyme is the principal replicative polymerase in Escherichia coli and is composed of a core of three subunits, $\alpha$, $\epsilon$, and $\theta$, and seven other accessory subunits that together contribute to the astonishing efficiency, processivity, and fidelity of the holoenzyme (Kelman and O'Donnell, 1995). The large $\alpha$ subunit contains the polymerase active site, while $\epsilon$ (242 residues, 28 kDa), the product of the dnaQ gene (Scheuermann et al., 1983), is the 3'→5' exonuclease subunit that serves as the proofreader and thereby contributes to the high fidelity of DNA synthesis. That the polymerase and exonuclease active sites are present on separate subunits of pol III is an unusual situation, since in most DNA polymerases, determinants of both activities reside in a single polypeptide chain (Kunkel, 1988). Although the function of the small $\theta$ subunit is uncertain, it is the only one of the three core subunits whose structure has been solved, by multidimensional NMR methods (Keniry et al., 2000).

It has recently been reported that the $\epsilon$ subunit is composed of two domains: a 20.5-kDa N-terminal domain that bears the exonuclease active site and interacts with $\theta$ and, following a Q-linker sequence presumed to be flexible, a smaller C-terminal domain that interacts with $\alpha$ (Perrino et al., 1999; Taft-Benz and Schapper, 1999). This finding is confirmed by some earlier unpublished observations from this laboratory (Thompson, 1992; Yang, 1998) that are described briefly below. The interaction of $\epsilon$ and $\alpha$ stimulates the polymerase activity of $\alpha$ 2-fold and stimulates the exonuclease activity of $\epsilon$ on mismatched DNA primer templates by 10- to 80-fold (Maki and Kornberg, 1987). Alignment of the amino acid sequence of $\epsilon$ with a series of most closely related proteins thought to serve as proofreading exonucleases in other bacterial species indicates that the N-terminal exonuclease domain is well conserved, but that the C-terminal portion that interacts with $\alpha$ is not, and suggests that the domain boundary occurs between residues 170 and 180 (Fig. 1).

Amino acid alignments among polymerase-associated 3'→5' exonucleases have shown several conserved motifs (Exo I, Exo II, and Exo III-ε; Barnes et al., 1995; Blanco et al., 1992; Koonin and Deutscher, 1993; Koonin, 1997) that contain conserved acidic residues that are presumed, by analogy with DNA polymerase I (Beese and Steitz, 1991; Derbyshire et al., 1991), to interact with two divalent metal ions that participate in phosphodiester bond cleavage (cf.
By using several dnaQ mutants, Taft-Benz and Schapper (1999) showed the importance of the Exo motifs in the N-terminal domain for catalysis, and Perrino et al. (1999) showed that the isolated N-terminal domain is active alone as an exonuclease.

Initially, we constructed a new thermally induced overproducing strain for full-length dnaQ following insertion of dnaQ into the λ-promoter vector, pND201 (Elvin et al., 1990), to give pPL224 and purified the protein by a published procedure (Scheuermann and Echols, 1984) that involved its refolding from solutions in 6 M guanidinium chloride (Thompson, 1992). Attempts to crystallize the protein were not successful. A degradation product of dnaQ (dp) was obtained by chance during dialysis of one particular batch of protein, presumably as a result of proteolysis by an unidentified protease present as an impurity (Fig. 2). The protein had an apparent size of 23.5 kDa as judged by SDS–polyacrylamide gel electrophoresis (SDS–PAGE), a molecular weight of 20 600 as determined by sedimentation equilibrium experiments, and the same N-terminal amino acid sequence as determined for full-length dnaQ, STAITR-QIVL- (Thompson, 1992). Proteolysis of another batch of full-length dnaQ with chymotrypsin under native conditions cleanly generated a fragment (ch) that had the same N-terminal sequence, a molecular

**Fig. 1.** Alignment of the sequence of *Escherichia coli* (Eco) with that from *Salmonella typhimurium* (Sty; GenBank 2506367) and putative dnaQ subunits from 12 other bacterial species: *Haemophilus influenzae* (Hin), *Buchnera aphidicola* (Bap, 2833216), *Rickettsia prowazekii* (Rpr, 6685391), *Treponema pallidum* (Tpa, 6014996), *Thermatoga maritima* (Tma, 4981007), *Deinococcus radiodurans* (Dra, 6458571), *Streptomyces coelicolor* (Sco, 6117870), *Archaeglobus fulgidus* (Afu, 2649627), *Mycobacterium tuberculosis* (Mtu, 2960135), *Chlamydia trachomatis* (Ctr, 3328974), *Aquifex aeolicus* (Aae, 6014995), and *Chlamydophila pneumoniae* (Cpn, 4376954). These sequences were chosen as those that showed the highest degree of similarity to *E. coli* dnaQ in a BLAST search (Altschul et al., 1997), after excluding those that were linked in the same polypeptide chain to domains with other functions (i.e., putative polymerase or helicase domains). The alignment is based directly on the BLAST multiple-alignment output. Residues identical to those in *Eco* dnaQ are shown in black, while those that represent conservative substitutions (as defined in Altschul et al., 1997) are shown in gray. The previously defined Exo motifs (Barnes et al., 1995) are indicated. The eight residues in *Eco* dnaQ that are identical in all sequences (marked □) comprise D11, E13, D102, H161, and D166, which are presumed to be in the active site and may be involved in metal-ion coordination, as well as the residues G16, P53, and L144. The 13 residue types that are conserved (marked △) are represented in *Eco* dnaQ by T15, I29, I30, L51, I67, I72, V81, V95, H97, M106, L147, A163, and I169. Most of these are aliphatic hydrophobic residues that most likely contribute to the hydrophobic core of the molecule.
weight of \( \sim 21000 \) as determined by gel filtration, and a mobility just less than that of \( \epsilon_{\text{DP}} \) in SDS-PAGE experiments (Fig. 2). The three proteins, \( \epsilon, \epsilon_{\text{DP}}, \) and \( \epsilon_{\text{CH}}, \) had comparable activities in a proof-reading assay using oligo(dT)\(_{16}\)([\(^3\)H]dC)\(_{6}\) annealed to poly(dA)\(_{200}\) as template, and both \( \epsilon \) and \( \epsilon_{\text{CH}} \) promoted the hydrolysis of the \( p \)-nitrophenyl ester of thymidine-5'-monophosphate (pNP-TMP) with comparable turnover numbers (Thompson, 1992). This demonstrated that an N-terminal domain of \( \epsilon \) could be isolated and that it bore the exonuclease active site. Attempts at the time to determine its size more precisely by mass spectrometry (MS) were not successful, so we predicted from examination of the sequence and pattern of fragments produced by chymotrypsin treatment (Fig. 2) that \( \epsilon_{\text{CH}} \) terminated at residue Phe-186 and that \( \epsilon_{\text{DP}} \) was just slightly smaller.

To express the N-terminal domain, we chose therefore to introduce a TAA stop codon after the Ala-185 codon of dnaQ in a derivative of pPL224 by oligonucleotide-directed mutagenesis, which gave plasmid plY764 (Yang, 1998). A strain containing this plasmid produced large quantities of \((1-185)\) in insoluble form on treatment at 42°C, and the protein was purified initially following unfolding and renaturation as for the full-length protein. Its activity in pNP-TMP hydrolysis was verified, and its molecular weight (calculated to be 20 587.7) was confirmed by electrospray ionization (ESI)–MS (Yang, 1998). The protein could be concentrated to \( > 2 \) mM in 10 mM sodium phosphate, pH 6.5, 0.2 mM dithiothreitol.

![FIG. 2. Digestion of the \( \epsilon \) subunit of pol III with chymotrypsin produces a stable fragment (\( \epsilon_{\text{CH}} \)). The \( \epsilon \) subunit (45 \( \mu \)g) in 25 mM Tris · HCl, pH 7.5, 1 mM DTT, 5 mM MgCl\(_2\), 50 mM NaCl, 15% glycerol (100 \( \mu \)l) was treated at 37°C with 0.9 \( \mu \)g of chymotrypsin. At the indicated times, samples (corresponding to 7.5 \( \mu \)g of \( \epsilon \)) were withdrawn and analyzed by SDS–PAGE (15%). Proteins were visualized by staining with Coomassie blue. Samples of the untreated \( \epsilon \) subunit (\( \epsilon \)) and the isolated adventitious degradation product (\( \epsilon_{\text{DP}} \)) were included as controls. The mobilities of marker proteins (sizes in kDa) were as indicated. For preparative isolation of \( \epsilon_{\text{CH}}, \) \( \epsilon \) (0.56 mg) was treated with chymotrypsin (1 \( \mu \)g) in 1 ml of 25 mM Tris · HCl, pH 7.5, 1 mM DTT, 5 mM MgCl\(_2\), 25 mM NaCl (Buffer S) for 48 h at 4°C. Pure \( \epsilon_{\text{CH}}, \) which was eluted from a column (1 ml) of Mono-Q anion-exchange resin (Pharmacia) in a 25–200 mM gradient (10 ml) of NaCl in Buffer S, was assayed for activities and physically characterized as described in the text.](image1)

![FIG. 3. 600 MHz \(^{15}\)N HSQC spectrum of \([^{15}\text{N}]\epsilon(1–185)\) in phosphate buffer, pH 6.5, recorded at 10°C. Narrow resonances originating from mobile residues at the termini of the protein are identified with their assignment.](image2)

![FIG. 4. A crystal of \( \epsilon(1–185)\) grown in the presence of 5 mM Mn\(^{2+}\) and 5 mM TMP. Typical dimensions of crystals are 0.2 mm × 0.2 mm × 1.0 mm.](image3)
(DTT), 2.5 mM MgCl₂, 0.2 M NaCl (Buffer A), and samples were used for preliminary NMR experiments, which indicated that ε(1–185) behaved as a monomer even at these concentrations and was well-structured and that there were very few mobile residues. However, it showed a pronounced tendency to aggregate at elevated temperatures; a heavy precipitate appeared within minutes at temperatures above 35°C and within 24 h at 20°C.

Reasoning that ε(1–185) was likely to be unfolding in vivo on expression at 42°C, we transferred the mutant dnaQ gene from pJY764 into the T7-promoter vector pETMCSII (a derivative of pET3c; Studier et al., 1990) to give pSH1018. The strain BL21::DE3/pLysS/pSH1018 overproduced soluble ε(1–185) on treatment with isopropyl-β-thiogalactoside (IPTG) at 30°C, and the protein could be purified in yields of about 15 mg/liter of culture by fractionation with ammonium sulfate and two steps of anion-exchange chromatography without denaturation and refolding. Protein prepared in this way was used for crystallization. Expression at 37°C gave largely insoluble ε(1–185).

To prepare uniformly ¹⁵N-labeled ε(1–185) for NMR studies, BL21::DE3/pLysS/pSH1018 was grown at 30°C in minimal medium containing 0.5 g/liter ¹⁵NH₄Cl, essentially as described (Weigelt et al., 1999). When A₅₉₅ reached 0.5, IPTG was added to 1 mM and growth was continued for a further 4 h. Cells were harvested and [¹⁵N]ε(1–185) was purified as for the unlabeled protein. Selenomethionine (Se-Met)-substituted ε(1–185) was prepared similarly, from cells grown at 30°C in the medium described by Van Duyne et al. (1993). Essentially complete incor-

FIG. 5. X-ray diffraction image obtained from a crystal of ε(1–185). The crystal was rotated through 1.0° at a temperature of 100 K. The diffraction limit occurs at 2.0 Å.
poration of SeMet at six positions in ε(1-185) was confirmed by ESI-MS.

For NMR, [\textsuperscript{15}N]ε(1-185) was dialyzed extensively against Buffer A at 4°C and then concentrated by use of an Ultrafree-4 centrifugal filter unit (Millipore). The sample used for all measurements contained [\textsuperscript{15}N]ε(1-185) at 2.2 mM in Buffer A containing 10% D\textsubscript{2}O. NMR spectra were recorded at 10°C, under which conditions <10% of the protein precipitated over the course of 7 days of data collection with a Bruker DMX-600 NMR spectrometer. A [\textsuperscript{15}N]HSQC spectrum showed good dispersion in the amide NH region (Fig. 3), consistent with the protein being structured and containing substantial amounts of β-sheet structure. The protein contains 184 backbone and 19 side-chain amides. About 200 amide resonances were resolved in the [\textsuperscript{15}N]HSQC spectrum (Fig. 3). \textsuperscript{1}H NMR linewidths of the amide protons were about 35-45 Hz at 10°C, suggesting that [\textsuperscript{15}N]ε(1-185) is a monomeric globular domain. A series of narrow resonances (<20 Hz) was sequentially assigned using a homonuclear TOCSY experiment (Briand and Ernst, 1991) and a 3D NOESY [\textsuperscript{15}N]HSQC (τ\textsubscript{m} = 100 ms) experiment (Talluri and Wagner, 1996) to mobile residues at the N- and C-termini of ε(1-185), specifically to residues Ala-3 to Thr-5 and residues Gly-180 to Ala-185. The \textsuperscript{1}H NMR resonances of the side chains of these residues were observed at random coil chemical shifts in the TOCSY experiment, which was recorded at 25°C (cf. Miles \textit{et al.}, 1997). Nevertheless, long-range NOEs to other unidentified protons were observed in the 3D NOESY [\textsuperscript{15}N]HSQC spectrum with the side chains of Ile-4 and Gln-181, which suggests that these residues are not entirely mobile. We conclude from these experiments that the structured core of the N-terminal domain of ε comprises residues 4 to 181.

Primarily because of the necessity to work at low temperature, it will be difficult to complete NMR resonance assignments. However, we anticipate that at least half of the resonances in the [\textsuperscript{15}N]HSQC spectrum can be assigned once the crystal structure is determined. The availability of partially assigned NMR data will then be useful in determining regions of ε that interact with metal ions, nucleotides, and DNA, as well as with θ and other subunits of pol III.

For crystallization, ε(1-185) was dialyzed overnight in 50 mM Na\textcdot HEPES, pH 7.5, containing 2 mM DTT. The protein was concentrated to 11.5 mg/ml as described above. Initial screens to establish crystallization conditions were performed with three sets of samples—with the protein alone, protein with 5 mM MgCl\textsubscript{2}, and protein with 5 mM MnSO\textsubscript{4} and 5 mM TMP. Crystals suitable for diffraction experiments were finally grown by the vapor diffusion method in hanging drops at 4°C. The reservoir solution contained 20 to 21% polyethylene glycol (PEG-8K) in 0.1 M cacodylate, pH 5.8. The drop contained 3 μl of protein, 5 mM TMP, and 5 mM MnSO\textsubscript{4} mixed with 3 μl of reservoir solution. An amorphous precipitate was observed after 5 to 7 days, and crystals grew from the mixture after a further 7 to 10 days (Fig. 4). The crystal habit is that of a tetragonal prism, of variable length to 1 mm and cross section 0.2 mm × 0.2 mm. Crystals were obtained at a range of pH values between 5.8 and 6.5, with initial PEG-8K concentrations between 9 and 11%.

The crystals were transferred to 20 μl of cryobuffer (30% PEG-8K in 0.1 M cacodylate, pH 5.8) and then flash cooled in a nitrogen gas stream at 100 K using the modifications to a Rigaku Raxis-IIC detector described by Carr \textit{et al.} (1996). The X-ray generator produced CuK\textsubscript{α} radiation at a power of 5 kW (60 kV, 84 mA). Data, collected at 100 K, were processed using the program HKL (Minor, 1993; Otwinowski, 1993). Diffraction extended beyond 2 Å (Fig. 5). The data set comprised 86 992 observations of 13 606 unique reflections (Table I) in space group P4\textsubscript{2}1\textsubscript{2}1\textsubscript{2} (n = 0, 1, 2, 3). Postrefined values of the unit cell were a = 60.8Å, c = 111.4 Å. A Matthews' coefficient V\textsubscript{m} (Matthews, 1968) of 2.5 Å\textsuperscript{3}/Da was obtained with one monomer in the asymmetric unit.

Although we expect the structure of the protein to be related to those of 3′-5′ exonucleases like the proofreading domain of DNA polymerase I (Ollis \textit{et al.}, 1985), the use of known structures as models for solution of the structure of ε(1-185) by molecular replacement has so far proven unsuccessful. This is not surprising since sequence identity with any protein whose structure is solved, beyond the Exo I and Exo II motifs described above, is virtually undetectable. Although SeMet-substituted ε(1-185) could be prepared, with a view to phasing by MAD, the protein was considerably less soluble than the unsubstituted specimen, and to date our attempts to crys-
tallize it have failed. A search is under way for suitable heavy-atom derivatives.

We are grateful to Carl Braybrook and Phillip Grace for assistance with ESI-MS, Peter J effrey for analytical ultracentrifugation, and the Biomolecular Resource Facility at the Australian National University for N-terminal protein sequence determination. Part of this work was supported by a Fellowship (to S.E.B.) awarded by the Australian Research Council, by an Australian Science and Technology Travel Grant (to N.E.D.), and by grants from the Swedish Natural Science Research Council (to G.O.).

REFERENCES


Miner, W. (1993) XDISPLAYF Program, Purdue University, West Lafayette, IN.


