

Characterization and crystallization of the helicase domain of bacteriophage T7 gene 4 protein

Louise E. Bird, Kjell Håkansson, Hu Pan and Dale B. Wigley*

Laboratory of Molecular Biophysics, Rex Richards Building, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Received March 12, 1997; Revised and Accepted May 9, 1997

ABSTRACT

Limited proteolysis of bacteriophage T7 primase/helicase with endoproteinase Glu-C produces several proteolytic fragments. One of these fragments, which is derived from the C-terminal region of the protein, was prepared and shown to retain helicase activity. This result supports a model in which the gene 4 proteins consist of functionally separable domains. Crystals of this C-terminal fragment of the protein have been obtained that are suitable for X-ray diffraction studies.

INTRODUCTION

The replication of DNA requires a number of proteins in addition to DNA polymerase (1). Amongst these are DNA primase, which initiates DNA replication by synthesis of RNA primers, and at least one DNA helicase. DNA helicases utilize the energy of NTP hydrolysis to unwind the DNA duplex (2). Helicases have been characterized biochemically according to their nucleic acid substrate specificity. There are two broad classes of enzyme based on the requirement for either a 3' or a 5' flanking region of single-stranded nucleic acid adjacent to the duplex to be unwound. The proteins are known as 3'→5' or 5'→3' helicases respectively. Helicases have been classified further into at least five families based on sequence alignment (3).

The bacteriophage T7 chromosome provides a simple *in vitro* model system for DNA replication, requiring just five proteins: the phage encoded T7 DNA polymerase (gene 5) and T7 primase/helicase (gene 4, gene products 4A and 4B), single-stranded DNA binding protein (gene 2.5) and *Escherichia coli* thioredoxin (4–6).

Gene 4 of bacteriophage T7 encodes two polypeptides of 56 and 63 kDa. The 56 kDa product (termed 4B) is translated from an internal initiation codon (Met64) in-frame within the coding sequence for the 63 kDa product (termed 4A; 7). Both proteins are 5'→3' helicases, but 4A also has primase activity (8,9). The difference between the two proteins is the presence of a Cys4 zinc binding motif in 4A, which is proposed to function in recognition of priming sites (8). Mutation of Met64 to Leu in 4A (termed 4A') demonstrated that 4A alone is sufficient to support bacteriophage T7 DNA replication (10). In contrast, in *E.coli* the primase and helicase activities are two separate proteins (DnaG and DnaB respectively).

The primase is 64 kDa and the replicative helicase is 52 kDa (1). However, there is an association of the two proteins (11,12).

Alignment of the protein sequence of 4A with those of *E.coli* DnaB and DnaG is suggestive of two functional domains in 4A (13). The N-terminus of 4A shows homology with DnaG and the C-terminal region with DnaB. Moreover, mutation of the 'Walker A' nucleotide binding site in 4A destroys helicase activity but does not completely abolish primase activity (14), suggesting a degree of separation of helicase and primase activities. Low resolution structural data, obtained using electron microscopy, have recently shown that there are two structural domains for each monomer in a protein complex that consists of a hexameric ring (15). DnaB, which is a member of the same family of helicases as the gene 4 proteins, is also a hexameric ring consisting of monomers with a bilobal structure (16). Limited proteolysis of DnaB with trypsin released a C-terminal fragment (17). Interestingly, all the motifs suggested as being necessary for helicase activity are in this fragment (3,18). This suggests that there may be a common structural domain for helicases of the DnaB family.

In order to understand the differences between the structures and mechanisms of both classes of DNA helicase we have undertaken high resolution structural studies. We have recently solved the structure of a 3'→5' helicase, PcrA, from *Bacillus stearothermophilus* (19). In this paper we have used limited proteolysis to investigate whether there are functional domains in the bacteriophage T7 gene 4A protein and thus define a minimal 5'→3' helicase for high resolution structural studies. We describe the subsequent cloning and expression of fragments of 4A that are helicases and crystallization of one of these proteins. The results obtained support a model where the 4A protein does indeed consist of functionally discrete domains of defined biochemical function.

MATERIALS AND METHODS

Bacterial strains

Escherichia coli MC1061 F^- , *araD139*, $\Delta(\text{ara-leu})7696$, *galH*, *E15*, *galK16*, $\Delta(\text{lac})X74$, *rpsL*, (*Str*^r), *hsdR2* (r_K^- , m_K^+), *mcrA*, *mcr B1* (20) was used for cloning. *Escherichia coli* HMS174(DE3) F^- , *recA*, *hsdR*, (r_{K12}^- , m_{K12}^+), Rif^R (21) and B834(DE3) F^- , *ompT*, *hsdS_B*, (r_B^- , m_B^-), *gal*, *met*, *dcm* (22) were used for protein production.

* To whom correspondence should be addressed. Tel: +44 1865 275381; Fax: +44 1865 275182; Email: wigley@biop.ox.ac.uk

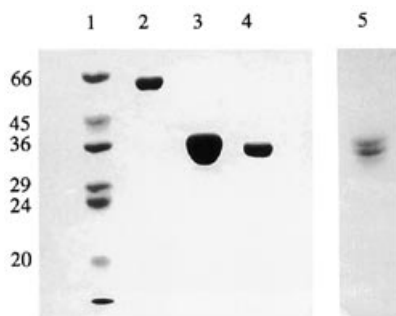


Figure 1. Purified proteins. Lane 1, Dalton VII low molecular weight markers; lane 2, 4A'; lane 3, 4C purified from HMS174(DE3)[pLysS]; lane 4, 4D purified from B834(DE3)[pLysS]; lane 5, 4C purified from B834(DE3)[pLysS].

DNA, nucleotides and enzymes

Bacteriophage T7 genomic DNA and endoproteinase Glu-C were purchased from Sigma. Single-stranded M13mp18 was purified as described (23). Restriction enzymes were purchased from New England Biolabs and Gibco BRL. *Taq* polymerase was purchased from Promega. Oligonucleotides were synthesized using an Applied Biosystems 381A DNA synthesizer. Radiolabelled nucleotides were purchased from Amersham.

Chemicals and column resins

All chemicals were obtained from Sigma unless stated otherwise. Protein purification resins were obtained from Pharmacia.

Expression and purification of 4A'

HMS174(DE3)pAR5018 (10) was grown, induced and harvested as described previously (24). The cell pellet was resuspended in buffer A [50 mM Tris-HCl, pH 7.5, 2 mM EDTA, 1 mM dithiothreitol (DTT)] + 10% w/v sucrose and frozen at -80°C until required. The cells were lysed by sonication and the supernatant was clarified by centrifugation at 20 000 *g* for 10 min. 4A' was precipitated from the supernatant using polyethyleneimine and ammonium sulphate as described previously (24). The pellet from the ammonium sulphate precipitation was resuspended in buffer A to bring the conductance of the solution to that of buffer A + 200 mM NaCl, prior to loading onto a Mono-Q 10/10 column. The column was washed with buffer A + 300 mM NaCl and the protein eluted with a 160 ml 300–600 mM NaCl gradient in buffer A. The fractions containing the protein were pooled and concentrated using a Centricon 30 (Amicon) to 10 mg/ml. The protein was applied to a Superdex 200HR 10/30 gel filtration column pre-equilibrated with buffer A + 100 mM NaCl. The protein elutes as a single peak and is >99% pure as estimated by Coomassie stained SDS-PAGE (Fig. 1).

Limited proteolysis

Proteolytic cleavage of purified 4A' (1 mg/ml) was carried out at room temperature (22°C) for 15 min. A number of proteases were tried, but the best results were obtained using endoproteinase Glu-C (V8 protease) with a buffer containing 50 mM sodium phosphate, pH 7.8, + 100 mM NaCl. The reactions were terminated by adding SDS-PAGE gel loading buffer and heating

to 95°C for 5 min. The samples were loaded immediately onto a 15% SDS-polyacrylamide gel for analysis.

Amino acid sequence analysis

N-Terminal sequencing of the proteolytic fragments was carried out using an Applied Biosystems model 470A protein sequenator connected to an on-line 120A high pressure liquid chromatograph.

Cloning and expression of 4C and 4D

To enable overexpression of 4C, PCR was used to amplify the coding sequence. The DNA primers were based on the published T7 genomic sequence (7; GenBank accession no. G15572) and on the amino acid sequence of the proteolytic fragments of 4A'. The 5' primer had a *Nco*I site and the 3' primer had a *Hind*III site after the stop codon. This allowed cloning into T7 promoter-based pET expression vectors (Novagen). The 5' primer was 5'-GATATACC ATG GCT GCA CAG GTT CTA CCT G-3' and the 3' primer was 5'-GCA TCC AAGCT TCA GAA GTC AGT GTC GTT G-3'. PCR was performed in 50 μl using these primers and T7 genomic DNA. The PCR reaction was electrophoresed in 1.0% agarose (Gibco BRL) and the product excised from the gel and purified using a QiaexII DNA extraction kit (Qiagen). The product was digested with *Nco*I and *Hind*III and further gel purified. The PCR products were ligated to dephosphorylated, *Nco*I/*Hind*III-cleaved pET21d, transformed into *E. coli* MC1061 and plated onto LB agar plates containing 100 $\mu\text{g}/\text{ml}$ ampicillin.

Positive clones were identified by restriction digestion. Clones containing the correct insert were screened for expression by transforming into B834(DE3)[pLysS], growing selected colonies in 5 ml Luria broth (LB) containing ampicillin and chloramphenicol at 37°C until the A_{600} reached 0.6, induction with 1 mM isopropyl-1-thio- β -D-galactopyranoside (IPTG) and growing for a further 3 h. The level of induction was monitored using SDS-PAGE. This clone was designated pET214C. Cloning and expression of 4D was carried out similarly to 4C; the 5' primer was 5'-GATATA CC ATG GCT GCA CAG GTT CTA CCT G-3', the 3' primer was as for 4C. The expression vector used was pET28a. This clone was designated pET284D.

Purification of 4C and 4D

One litre cultures of LB containing 100 $\mu\text{g}/\text{ml}$ ampicillin and 34 $\mu\text{g}/\text{ml}$ chloramphenicol or 30 $\mu\text{g}/\text{ml}$ kanamycin and 34 $\mu\text{g}/\text{ml}$ chloramphenicol were inoculated with a 5 ml culture of HMS174(DE3)[pLysS pET214C] or B834(DE3)[pLysS pET284D]. The cultures were grown with shaking at 37°C until the A_{600} reached 0.6–0.7. The cultures were induced by addition of 1 mM IPTG. Growth was continued for 3 h, before the cells were harvested by centrifugation at 5000 *g*. The cell pellets were resuspended in 20 ml buffer A + 10% sucrose and frozen at -80°C until required. The cells were lysed by sonication and the supernatant was clarified by centrifugation at 20 000 *g*, prior to precipitation by addition of an equal volume of saturated ammonium sulphate. The precipitate was harvested by centrifugation at 20 000 *g*. The pellet was resuspended in buffer A in a volume such that the conductivity of the solution was equal to or less than the conductivity of buffer A + 100 mM NaCl. This was applied to a 40 ml Q-Sepharose column equilibrated with buffer A + 100 mM NaCl. The column was washed with buffer A + 300 mM NaCl. The protein was eluted with a gradient of

300–800 mM NaCl in buffer A. The peak fractions were pooled and diluted to buffer A + 200 mM NaCl and loaded onto a Mono-Q 10/10 column. The column was washed with buffer A + 300 mM NaCl and the protein eluted with a 160 ml 300–700 mM NaCl gradient in buffer A. Purity of the sample was monitored using SDS–PAGE (Fig. 1).

SDS–PAGE

Protein samples were analysed by SDS–PAGE in 10, 12 and 15% gels with 4% stacking gels (25). Gels were stained with Coomassie Brilliant Blue and destained in 10% acetic acid and 25% methanol.

Gel filtration

Gel filtration was performed at 22°C in the absence of nucleotide. The protein samples were concentrated using a Centricon 30 (Amicon). Gel filtration was performed using a Superdex 200 HR 10/30 column. The column was equilibrated with buffer A + 100 mM NaCl and calibrated with protein standards. The protein sample was applied to the column in 100 µl. The apparent molecular weights (M_{app}) were calculated from an interpolation of a semi-log plot of partition coefficient (K_{av}) of the protein standards versus molecular weights.

Nucleotide hydrolysis assay

dTTP hydrolysis was monitored by following the production of inorganic phosphate using acidic ammonium molybdate with malachite green (26). The reaction mixtures (50 µl) contained 50 mM Tris–acetate, pH 7.5, 50 mM sodium acetate, 0.1 mM EDTA, 10 mM magnesium chloride, 0.1 mg/ml BSA, 1.5 µg single-stranded M13mp18 DNA, dTTP and enzyme. The reactions were started by addition of enzyme and were incubated at 37°C for 10 min. Acidic malachite green/ammonium molybdate solution (0.8 ml) was used to stop the reaction, followed by addition of 0.1 ml 34% (w/v) sodium citrate 1 min later. The colour was left to develop for 20 min and the absorbance was read at 660 nm using a reaction minus enzyme for each concentration of substrate as a blank. All assays were performed in duplicate. The amount of phosphate produced was quantified using a standard curve.

Helicase assay

A 68mer oligonucleotide with 22 bases homologous to M13mp18 and 5' and 3' tails of single-stranded DNA (27) was 5'-end-labelled using T4 polynucleotide kinase (New England Biolabs). The oligonucleotide was annealed to single-stranded M13mp18 DNA. The unannealed oligonucleotide and free label were removed by centrifugation through a 1 ml Sepharose CL-6B spin column.

The reactions (10 µl) contained 20 mM Tris, pH 7.5, 50 mM NaCl, 3 mM MgCl₂, 5 mM DTT, 10% glycerol, 9.5 nM substrate, enzyme and 1 mM dTTP. The reactions were started by addition of enzyme and incubated at 37°C for 30 min. One fifth volumes of 2% SDS, 200 mM EDTA and 50% glycerol were used to stop the reaction. The mixture was electrophoresed through a non-denaturing 6% polyacrylamide gel containing TBE (23). After electrophoresis, the gel was fixed and dried under vacuum. Autoradiography was carried out overnight at –80°C using Fuji RX X-ray film. Alternatively, quantification of the degree of displacement of the oligonucleotide was carried out using a

phosphorimager and ImageQuant software (Molecular Dynamics). The percentage displacement was calculated from the amount of radioactivity (adjusted for background) in the annealed and unannealed bands and corrected for the background of unannealed product in the protein minus control.

Crystallization of 4D and X-ray data collection

The protein was concentrated to 7 mg/ml using a Centricon 30 (Amicon). The protein was exchanged into 10 mM Tris, pH 7.5, 50 mM NaCl, 1 mM DTT during this process. Crystals of 4D were obtained by vapour diffusion at 20°C in hanging drops (equal volumes of protein and well solution) over a 1 ml well solution of 50 mM CHES–NaOH, pH 9.5, 0.8 M sodium acetate, 1 mM sodium azide, polyethylene glycol (PEG) of molecular weight 4000 at a concentration of 20–30% (w/v). The crystals were either mounted in capillaries at room temperature or flash frozen, in the same solution as the well solution with the addition of 15% glycerol (v/v), at 100 K. Data were collected using a Rigaku rotating anode X-ray generator and a MAR image plate detector and processed with DENZO and SCALEPACK (28).

RESULTS

Proteolytic analysis of 4A'

4A' was digested under limiting conditions with a range of proteases, including trypsin, chymotrypsin, clostripain, elastase, endoproteinase Glu-C and thermolysin. Despite the differences in specificity, all enzymes produced a major product of between 33 and 36 kDa (data not shown). However, endoproteinase Glu-C gave rise to discrete fragments that were resistant to further proteolysis (Fig. 2a). Fragments 1 and 2 were produced rapidly and were cleaved further. N-Terminal sequencing was used to identify the cleavage sites (summarized schematically in Fig. 2b). Fragment 1 has the same N-terminal sequence as full-length 4A', showing that cleavage occurs near the C-terminus, resulting in the loss of ~3 kDa. We did not carry out C-terminal sequencing and so were unable to determine the cleavage site at the C-terminus. However, we have indicated an approximate position (Fig. 2b) based on the apparent molecular weight of fragment 1 as determined by SDS–PAGE (Fig. 2a). Following the initial cleavage, the 51 N-terminal amino acids were also removed by proteolysis (fragment 2). The 33 kDa fragment 3 was the result of proteolysis between residues 218 and 219. The position of this cleavage site is in the same region of the protein as the trypsin-sensitive site at amino acid 275 reported recently by Washington *et al.* (29). This suggests that there may indeed be a helicase structural domain with a flexible region of protein linking it to the N-terminus. Sequence comparison of 4A' from amino acid 219 (Fig. 3) with a similar tryptic fragment of DnaB (17) showed that the C-terminal fragment has all the motifs identified in the DnaB family of helicases (13). The fragment derived from 4A' was also similar both in size and sequence composition to the 30 kDa hexameric 5'→3' helicase from the incompatibility group Q plasmid RSF1010 (30,31), suggesting that the C-terminal fragment derived from 4A' may be a helicase.

In order to establish the relationship between activity and structure for the C-terminal region of the bacteriophage T7 gene 4 proteins we decided to make a protein based on the endoprotease Glu-C-sensitive site at amino acid 218. Since we do not know which amino acid is the C-terminus of fragment 3, we

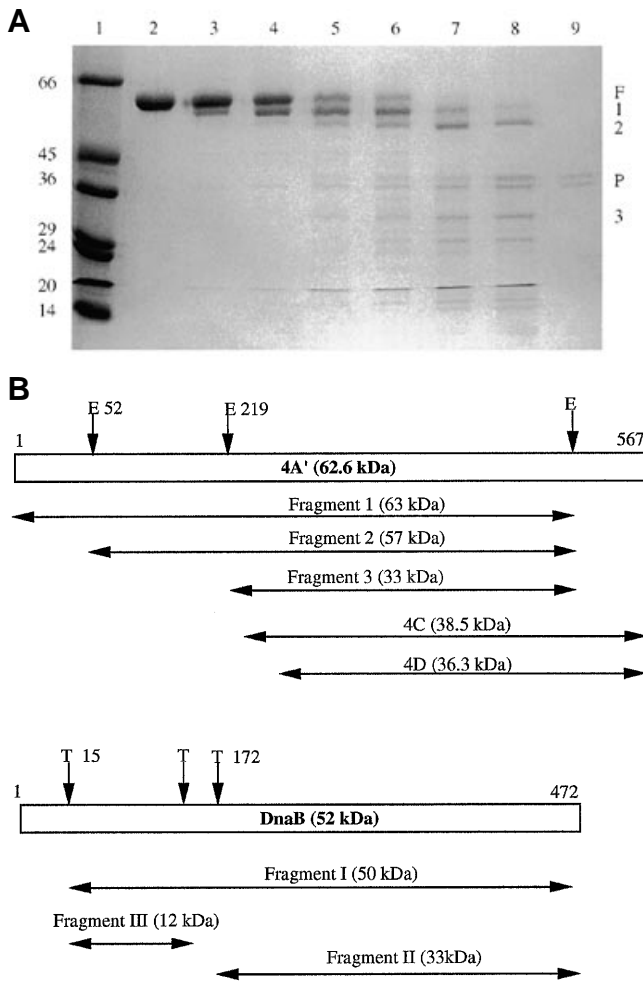


Figure 2. Proteolysis of 4A' with endoproteinase Glu-C. (a) 4A' was proteolyzed with increasing amounts of protease. From left to right: lane 1, Dalton VII low molecular weight markers; lanes 2–8, 5 µg 4A' and 0, 0.05, 0.1, 0.25, 0.5, 1.0 and 2.5 U endoproteinase Glu-C; lane 9, control containing 0.25 U endoproteinase Glu-C but no 4A'. The full-length protein (F), fragments (1–3) and protease (P) are indicated. (b) Schematic diagram of proteolysis of 4A' with endoproteinase Glu-C compared with a tryptic digest of DnaB (17). The protease-sensitive sites are designated E and T respectively; the cleavage positions are indicated. The amino acid residue numbers relative to the full-length proteins of all proteins are shown; unknown residue numbers are indicated by a question mark. The molecular weights of the proteins as determined by SDS-PAGE are also indicated. The cloned fragments used in this study are also shown for comparison.

decided to make a protein that included the 3 kDa C-terminal peptide. The recombinant protein from amino acid 219 of 4A' to the C-terminus (amino acid 567) was cloned by PCR and over-expressed in vector pET21d. This protein was designated 4C (Fig. 2b) and the vector pET214C.

Purified 4C is a mixture of two proteins: 4C and 4D

Purified 4C, expressed in B834(DE3)[pLysS], was visualized using SDS-PAGE and was observed to be a mixture of two proteins, with the majority having a lower molecular weight than predicted (Fig. 1, lane 5). 4C purified from HMS174-(DE3)[pLysS] was >99% pure (Fig. 5, lane 3). However, when lower concentrations of 4C were analysed by SDS-PAGE (data

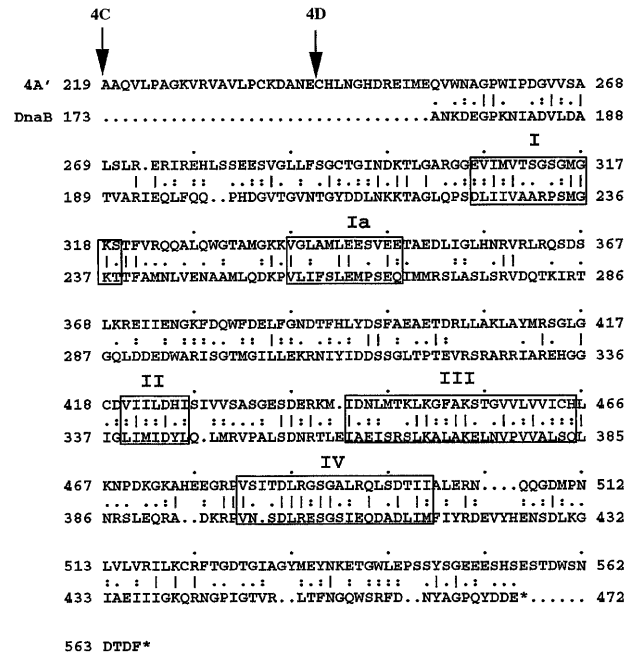


Figure 3. Sequence comparison of the C-terminal fragment of 4A' with the 33 kDa tryptic fragment of DnaB (17). The sequence alignment was compiled using the GCG Gap program (43). The conserved motifs identified by Ilyina *et al.* (13) are boxed. The identity (|) and degree of similarity (: and .) of residues are indicated.

not shown) it can be seen that, although the majority of the protein is equivalent to the expected molecular weight, a low level of the lower molecular weight product seen previously was still observed. Consequently, HMS174[pLysS] was used for production of 4C, since the purified protein is a more homogeneous population.

A sample of 4C purified from B834(DE3)[pLysS] was subjected to electrospray mass spectrometry. Two peaks were observed (data not shown). Comparisons of the observed molecular weights (38 329 ± 4 and 35 994 ± 3 Da, a difference of 2335 Da) with calculated molecular weights for truncations of either the N- or C-terminus suggested that the larger protein is the 4C protein lacking its N-terminal methionine (calculated molecular weight 38 318 Da) and that the smaller protein lacks the 24 N-terminal amino acids (calculated molecular weight 35 983 Da, 2335 Da smaller than 4C). The gene for the smaller protein (designated 4D) was cloned by PCR. 4D was over-expressed using pET28a (designated pET284D) and purified (Fig. 1, lane 4).

The purified proteins were >99% pure as assessed by SDS-PAGE (Fig. 1).

Oligomerization of 4C and 4D

It has been shown previously that hexamerization of 4A_{M64G} (M64→G in order to produce only 4A) is necessary for single-stranded DNA-dependent NTP hydrolysis, translocation and DNA unwinding (32). 4A' (M64→L) assembles into stable hexamers in the presence of MgdTMP-PCP. However, in the absence of nucleotide, the protein oligomerizes in a concentration-dependent manner. At low protein concentrations 4A' is reported to form dimers and trimers, while at increasing protein concentrations higher

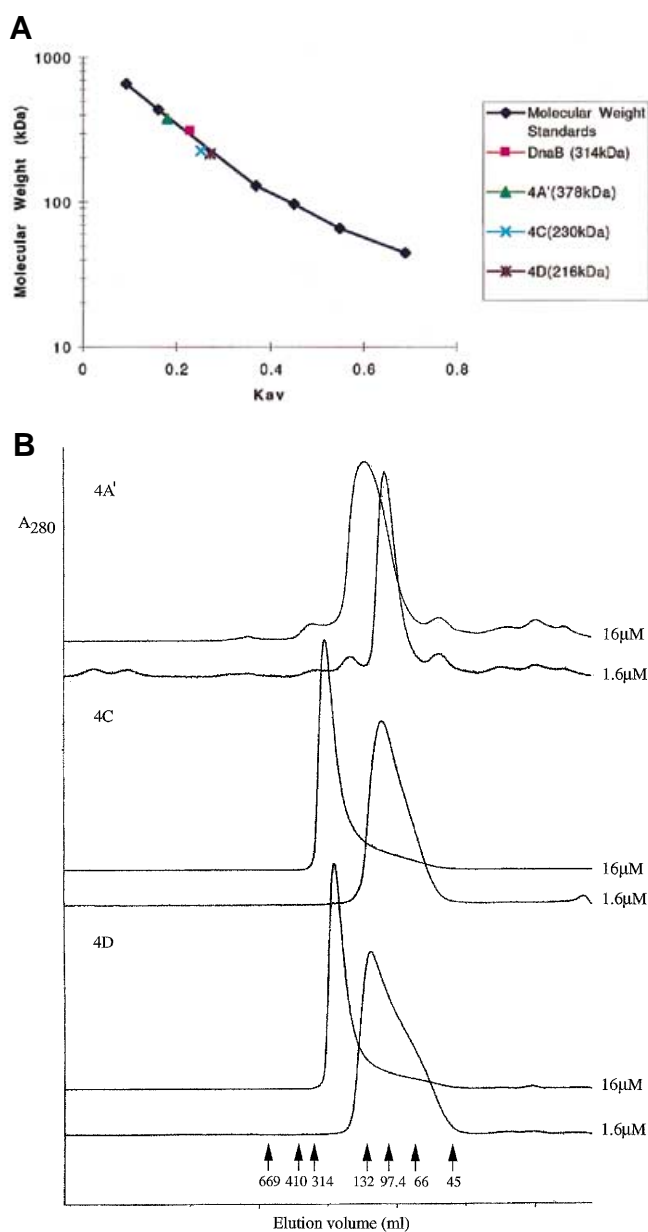


Figure 4. Gel filtration of gene 4 proteins. (a) High protein concentrations of 4A', 4C and 4D in the absence of ligands. All samples were loaded onto the column at 160 μ M. Molecular weight markers were: thyroglobulin (669 kDa), β -galactosidase (464 kDa), myosin (410 kDa), bovine serum albumin (66 kDa, 132 kDa), phosphorylase b (97 kDa) and ovalbumin (45 kDa). DnaB (314 kDa), a hexameric helicase with a toroidal structure, was also run on the column. K_{av} is plotted against the molecular weight for all proteins. (b) Low protein concentrations of 4A', 4C and 4D. All samples were loaded onto the column at both 16 and 1.6 μ M. The elution points for the molecular weight standards are indicated. The heights of the peaks are scaled to be similar sizes.

order oligomers were observed (33). In order to examine the oligomerization of 4C and 4D, their oligomeric state was compared with 4A' for both high and low protein concentrations in the absence of nucleotide using gel filtration. At high protein concentrations (160 μ M), 4A' (M_{app} 405 kDa) as well as 4C (M_{app} 264 kDa) and 4D (M_{app} 231 kDa) eluted from the column close to the position expected for a hexamer (Fig. 4a). This behaviour

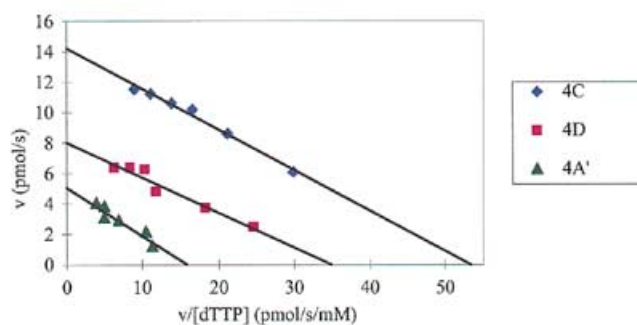


Figure 5. Eadie-Hofstee plot of dTTP hydrolysis by 4A', 4C and 4D. Hydrolysis of dTTP by 4A' (0.1 μ M), 4C (0.3 μ M) and 4D (0.8 μ M). The velocity (v) is expressed as pmol phosphate produced per second and the concentration of dTTP ($[dTTP]$) is in mM. $v/[dTTP]$ is plotted against v . The kinetic parameters determined from these data are shown in Table 1.

was similar to the behaviour of the toroidal hexameric helicase DnaB (M_{app} 300 kDa compared with a calculated molecular weight of 314 kDa). No larger or smaller order oligomers were observed for 4C and 4D, although a small amount of a lower order oligomer was observed for 4A' (data not shown). Since a small amount of a lower order oligomer was observed for 4A' but not for 4C and 4D, the extent of oligomerization of all the proteins was analysed at lower concentrations (Fig. 4b). At 16 μ M both 4C and 4D were hexameric, with small amounts of lower order oligomers. In contrast, the M_{app} of the 4A' peak (132 kDa) is consistent with the peak being a dimer. The broadening of the peak suggests an equilibrium between monomer and dimer in rapid exchange, with the dimeric form predominating. At even lower concentrations (1.6 μ M) we also observed this effect for both the 4C and 4D proteins, suggesting that there is a similar dynamic equilibrium between lower order oligomers.

Thus, under the conditions utilized for gel filtration in this study, both 4C and 4D can form stable hexamers at high protein concentrations in the absence of nucleotide, in common with 4A'. The differences in degree of oligomerization at both high and low protein concentrations suggests that 4C and 4D may form more stable hexamers, in the absence of nucleotide, than 4A'.

dTTPase activity of 4C and 4D

Nucleotide hydrolysis is essential for the activity of helicases (2). 4A is able to hydrolyse both dNTPs and rNTPs using single-stranded DNA as an effector (4). However, the preferred substrate is dTTP (34). For most of the nucleotides the concentration dependence of the NTPase activity is hyperbolic, however, the ATPase activity is sigmoidal (24). Since dTTP is the preferred substrate, the dTTPase activity of 4C and 4D, using single-stranded M13 as an effector, was compared with 4A'.

4C and 4D were able to hydrolyse dTTP. A plot of the primary data is hyperbolic (data not shown) and thus the activities obey Michaelis-Menten kinetics as reported previously for 4A' (24,34). The kinetic parameters were determined by plotting $v/[dTTP]$ against v (Fig. 5) and are shown compared with those of 4A' determined at the same time in Table 1. The values of K_m for 4C and 4D are of a similar order of magnitude to 4A' and are in good agreement with the value reported previously for 4A'

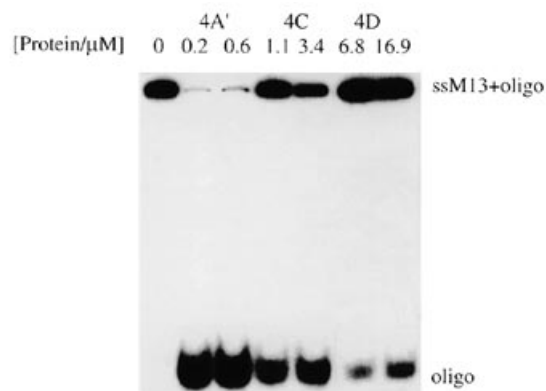


Figure 6. The helicase activity of 4A', 4C and 4D. dTTP was used as an energy source. The protein and concentration utilized is indicated above each lane. The positions of the substrate and the displaced oligonucleotide are also indicated.

(24,34). However, while 4A' and 4C have similar values of k_{cat} , the value for 4D is 5-fold lower.

Table 1. Steady-state dTTPase kinetic parameters of 4A', 4C and 4D

Protein	K_m (mM)	k_{cat} (per s) ^a	k_{cat}/K_m (per s/M)
4A'	0.36	0.9	2.5×10^3
4C	0.25	1.1	4.4×10^3
4D	0.22	0.2	0.9×10^3

^a k_{cat} is expressed per mol monomer.

Helicase activity of 4C and 4D

The ability of 4C and 4D both to hexamerize and to hydrolyse dTTP suggests that they may be helicases. The helicase activity of 4C and 4D compared with 4A' was analysed as a function of protein concentration. The displacement of an oligonucleotide substrate annealed to single-stranded M13 was used as an assay system. The substrate had a 22 bp duplex and both 5' and 3' tails of single-stranded DNA. dTTP was used as an energy source.

4C and 4D are both helicases active upon a tailed substrate (Fig. 6). However, the activities of 4C and 4D were markedly less than that of 4A'. At a concentration of 0.21 μM 4A' displaced 89% of the annealed oligonucleotide, whereas, even at 3.4 μM , 4C only displaced 40% of the oligonucleotide. 4D is an even poorer helicase than 4C; at a concentration of 16.9 μM 4D displaced only 18% of the oligonucleotide.

Crystallization of 4D

Despite the reduction in activity of the fragments, we conclude that we have isolated a minimal helicase domain and that this is likely to be a structural domain. In order to obtain high resolution structural data we have obtained crystals of 4D by the hanging drop method (Fig. 7). Preliminary characterization of these crystals reveals that they are tetragonal and belong to the space group $P4_12_12$ (or $P4_32_12$) with unit cell dimensions $a = b = 122 \text{ \AA}$, $c = 285 \text{ \AA}$. The crystals diffract to 3.5 \AA using a rotating anode source. Assuming one hexamer/asymmetric unit, the V_m is 2.5 $\text{\AA}^3/\text{Da}$ and the approximate solvent content is 50%. However,

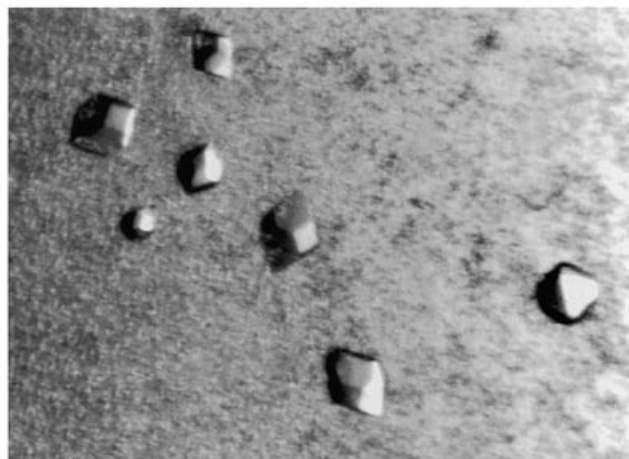


Figure 7. Crystals of 4D. The largest crystals shown are 150 μm in their largest dimension.

it is possible that there may be half a hexamer in the asymmetric unit if the hexamer is sitting on a crystallographic 2-fold axis. A heavy atom derivative search is underway.

DISCUSSION

Previous work has indicated that the gene 4 proteins have defined structural domains (15). These low resolution structural data are consistent with the idea proposed by Ilyina *et al.* (13) that gene 4 proteins evolved through fusion of a primase and helicase. It is therefore likely that there would be two major structural and functional domains. The limited proteolysis presented in this study has helped delineate the domain structure of the 4A' protein in terms of function. The susceptibility of the N-terminus of 4A', a region only found in 4A', to proteolysis is consistent with a discrete structural domain consisting of the Cys4 zinc binding motif that functions to recognize priming sites (8,9). Rapid cleavage of the C-terminus also suggests a discrete region of the protein. This region of the protein is highly acidic (8/24 residues) and may be involved in protein-protein interactions such as that observed with the T7 DNA polymerase (35). The remainder of the protein is likely to consist of the two structural domains observed by electron microscopy (15). The evidence presented in this paper suggests that these may also be functional domains, since the proteins derived from the C-terminal region of 4A' are both helicases. However, both 4C and 4D are poorer helicases than 4A'. The reduction in helicase activity is not due solely to an effect on the DNA-dependent dTTPase activity, because, while 4D has a lower k_{cat} than 4A', 4C has very similar kinetic parameters to 4A'. The rate of the helicase reaction is dependent upon enzyme concentration even at levels considerably in excess of the DNA substrate concentration (Fig. 6). This suggests that the reduction in helicase activity is likely to be due to a reduced affinity for DNA or to reduced processivity of the enzyme. With the present data we are unable to distinguish between these alternatives. The defect in helicase activity could possibly be due either to a difference in oligomerization relative to 4A or because of other sequences that are not present in the fragments. The C-terminal tryptic fragment of DnaB, which only possesses single-stranded DNA-dependent ATPase activity and no detectable helicase activity (36), is

defective in oligomerization; only stable dimers and not hexamers are formed. In contrast, both 4C and 4D seem to form more stable oligomers than 4A'. The mechanism by which this could decrease helicase activity is not clear and needs further investigation. 4C and 4D are larger than the DnaB tryptic fragment, with most of the additional amino acids being at the N-terminus of the proteins. Since all of the conserved motifs thought to be necessary for helicase activity in the DnaB family of proteins are present in both fragments, it is possible that sequences additional to the observed conserved motifs are necessary for helicase activity and could either lie just N-terminal of Motif I or within the N-terminal domain. If sequences just N-terminal of Motif I are important, this could explain why 4C is a better helicase than 4D. There is evidence for a degree of cooperativity between the two enzymatic activities of 4A, since mutation of the 'Walker A' motif, which abolishes helicase activity, also reduces primase activity (14). The reduction in primase activity is due to a reduced affinity for DNA, the helicase activity in this case acting to tether the primase to the DNA. A similar situation is seen in *E.coli*, where DnaG primase requires DnaB to stimulate its activity (37,38) but hydrolysis of ATP by the helicase and therefore duplex unwinding is not necessary (39). However, it need not follow that the helicase requires the primase domain. Serial truncations of 4A from the N-terminus and addition of the purified N-terminus to the helicase reaction will help address this question.

The high resolution structure of a 3'→5' helicase, PcrA from *B.stearothermophilus*, has recently been reported (19). Interestingly, the structure revealed two sub-domains, both with homology to RecA. RecA has the same ATP binding fold as F1-ATPase (40), a pseudo-hexameric protein which has sequence homology with the bacterial hexameric RNA helicase Rho (41). Furthermore, an alignment of RecA with a secondary structure prediction of 4A (29) suggested that the C-terminus of 4A would have the same structure as both RecA and F1-ATPase. Since 3'→5' helicases have been shown to have a RecA fold we have predicted that all helicases have a common structural core, namely a RecA domain. Comparison of the residues that comprise the nucleotide binding site of RecA with those in PcrA shows that they are conserved spatially but not conserved in primary sequence (19,42). For example, the adjacent DE residues of Motif II in PcrA are found far apart in the primary sequence in RecA (D144 and E96) but adjacent in space in the structure. The Walker A and B motifs (B contains the D of the DE found in Motif II in PcrA) found in RecA align well with Motifs I and II in gene 4 protein (29), suggesting that the structure of hexameric helicases, in terms of the relative positions of amino acids in primary sequence, may in fact be more similar to RecA than the 3'→5' helicases. In order to answer this question, we have crystallized the helicase domain of the gene 4 protein and await the structural determination with interest.

ACKNOWLEDGEMENTS

We thank A.Rosenberg for the 4A' clone and data prior to publication, V.Cooper for synthesis of oligonucleotides, T.Willis for N-terminal sequencing of proteolytic fragments, K.Marians for DnaB and R.Aplin for mass spectrometry. This work was supported by the MRC and Wellcome Trust.

REFERENCES

- Marians,K.J. (1992) *Annu. Rev. Biochem.*, **61**, 673–719.
- Lohman,T.M. (1992) *Mol. Microbiol.*, **6**, 5–14.
- Gorbalenya,A.E. and Koonin,E.V. (1993) *Curr. Opin. Struct. Biol.*, **3**, 419–429.
- Kolodner,R. and Richardson,C.C. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 1525–1529.
- Tabor,S., Huber,H.E. and Richardson,C.C. (1987) *J. Biol. Chem.*, **262**, 16212–16223.
- Kim,Y.T. and Richardson,C.C. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 10173–10177.
- Dunn,J.J. and Studier,F.W. (1983) *J. Mol. Biol.*, **166**, 477–535.
- Bernstein,J.A. and Richardson,C.C. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 396–400.
- Bernstein,J.A. and Richardson,C.C. (1989) *J. Biol. Chem.*, **264**, 13066–13073.
- Rosenberg,A.H., Patel,S.S., Johnson,K.A. and Studier,F.W. (1992) *J. Biol. Chem.*, **267**, 15005–15012.
- Wu,C.A., Zechner,E.L., Reems,J.A., McHenry,C.S. and Marians,K.J. (1992) *J. Biol. Chem.*, **267**, 4074–4083.
- Tougu,K., Peng,H. and Marians,K.J. (1994) *J. Biol. Chem.*, **269**, 4675–4682.
- Ilyina,I.V., Gorbalenya,A.E. and Koonin,E.V. (1992) *J. Mol. Evol.*, **34**, 351–357.
- Mendelman,L.V., Notarnicola,S.M. and Richardson,C.C. (1993) *J. Biol. Chem.*, **268**, 27208–27213.
- Egelman,E.H., Yu,X., Wild,R., Hingorani,M.M. and Patel,S.S. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 3869–3873.
- San Martin,M.C., Stamford,N.P.J., Dammerova,N., Dixon,N.E. and Carazo,J.M. (1995) *J. Struct. Biol.*, **114**, 167–176.
- Nakayama,N., Arai,N., Kaziro,Y. and Arai,K.-I. (1984) *J. Biol. Chem.*, **259**, 88–96.
- Hodgman,T.C. (1988) *Nature*, **333**, 22–23.
- Subramanya,H.S., Bird,L.E., Brannigan,J. and Wigley,D.B. (1996) *Nature*, **386**, 379–383.
- Casabadian,M.C. (1980) *J. Mol. Biol.*, **138**, 179–207.
- Studier,F.W. and Moffat,B.A. (1986) *J. Mol. Biol.*, **189**, 113–130.
- Wood,W.B. (1966) *J. Mol. Biol.*, **16**, 118–133.
- Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Patel,S.S., Rosenberg,A.H., Studier,F.W. and Johnson,K.A. (1992) *J. Biol. Chem.*, **267**, 15013–15021.
- Laemilli,U.K. (1970) *Nature*, **227**, 680–685.
- Lanzetta,P.A., Alvarez,L.J., Reinach,P.S. and Candia,O.A. (1979) *Anal. Biochem.*, **100**, 95–97.
- Crute,J.J., Mocarski,E.S. and Lehman,I.R. (1988) *Nucleic Acids Res.*, **16**, 6585–6596.
- Otwinowski,Z. (1993) In Sawyer,L., Isaacs,N. and Bailey,S. (eds), *Data Collection and Processing*. SERC Laboratory, Daresbury, UK, pp. 56–62.
- Washington,M.T., Rosenberg,A.H., Griffin,K., Studier,F.W. and Patel,S.S. (1996) *J. Biol. Chem.*, **271**, 26825–26834.
- Haring,V. and Scherzinger,E. (1989) In Thomas,C.M. (ed.), *Promiscuous Plasmids of Gram Negative Bacteria*. Academic Press, London, UK, pp. 95–124.
- Scherzinger,E., Haring,V., Lurz,R. and Otto,S. (1991) *Nucleic Acids Res.*, **19**, 1203–1211.
- Notarnicola,S.M., Park,K., Griffith,J.D. and Richardson,C.C. (1995) *J. Biol. Chem.*, **270**, 20215–20224.
- Patel,S.S. and Hingorani,M.M. (1993) *J. Biol. Chem.*, **268**, 10668–10675.
- Matson,S.W. and Richardson,C.C. (1983) *J. Biol. Chem.*, **258**, 14009–14016.
- Nakai,H. and Richardson,C.C. (1988) *J. Biol. Chem.*, **263**, 9818–9830.
- Biswas,S.B., Chen,P.-H. and Biswas,E.E. (1994) *Biochemistry*, **33**, 11307–11314.
- Arai,K. and Kornberg,A. (1981) *J. Biol. Chem.*, **256**, 5260–5266.
- Arai,K., Low,R.L. and Kornberg,A. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 707–711.
- Shrimankar,P., Stordal,L. and Maurer,R. (1992) *J. Bacteriol.*, **174**, 7689–7696.
- Abrahams,J.P., Leslie,A.G.W., Lutter,R. and Walker,J.E. (1994) *Nature*, **370**, 621–628.
- Dombroski,A.J. and Platt,T. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2538–2542.
- Story,R.M. and Steitz,T.A. (1992) *Nature*, **355**, 374–376.
- Devereux,J., Haerberli,P. and Smithies, O. (1984) *Nucleic Acids Res.*, **12**, 387–395.