Identification of a Compact DNA-Binding Domain in the Gene 5 Protein of Pf1 Bacteriophage†

S. E. Plyte‡ and G. G. Kneale*

Biophysics Laboratories, School of Biological Sciences, University of Portsmouth, Portsmouth PO1 2DT, U.K.

Received October 6, 1992; Revised Manuscript Received December 21, 1992

ABSTRACT: The structure of the gene 5 protein of filamentous bacteriophage Pf1 and its interaction with viral DNA have been investigated by a series of limited proteolysis experiments. The ability of purified proteolytic fragments of the Pf1 gene 5 protein to bind oligonucleotides and polynucleotides was monitored by gel retardation and fluorescence. The results show the presence of a compact DNA-binding "core" domain consisting of residues 1–112 of the protein, which is protected from proteolysis in the nucleoprotein complex. Digestion of the free gene 5 protein with subtilisin produces a smaller fragment (residues 7–102) which can no longer bind DNA. Although the N-terminal "core" domain shows full DNA binding activity by fluorescence, the gel retardation experiments suggest reduced kinetic stability of this domain in complexes with oligonucleotides, resulting from the removal of residues 113–144 from the C-terminus of the protein. The sequence of the C-terminal 32 amino acid residues is unusual, with a high proportion of alanine, glutamine, and proline residues which may be related to the role of this sequence in stabilizing the complex.

Single-stranded (ss) DNA binding proteins perform a wide variety of functions in the cell, principally in the processes of DNA replication and recombination, but also in controlling the transcription and translation of specific genes (Kneale, 1992). This class of protein is well exemplified by the gene 5 proteins of filamentous bacteriophages Fd (Alberts et al., 1972) and Pf1 (Maeda et al., 1982). The gene 5 protein is one of the major cellular proteins synthesized in phage-infected cells. It forms the principal component of the intracellular nucleoprotein complex in which the gene 5 protein is complexed to the single-stranded circular viral DNA forming a regular helical assembly (Gray et al., 1982; Kneale et al., 1982). The gene 5 protein folds the viral DNA into a form suitable for packaging by the viral coat protein (at the bacterial cell membrane), and the gene 5 protein is recycled in the cell. In addition to its role in viral assembly, the gene 5 protein also functions as a translational repressor (Yen & Webster, 1982), binding specifically to the leader sequence of the gene 2 mRNA (Michel & Zinder, 1989). The combination of both these activities has the net effect of switching off double-stranded (RF) synthesis at a critical stage in the replication cycle and promoting viral strand DNA synthesis.

The gene 5 proteins of Pf1 and Fd share little overall homology, other than in the region of the DNA-binding wing of the protein (residues 15–30 for Pf1) which is homologous to the sequence 12–26 in Fd gene 5 protein (Plyte & Kneale, 1991). The Pf1 gene 5 protein is considerably more resistant to dissociation by salt than the Fd protein (Kneale, 1983), and the helical parameters of the two gene 5 protein–viral DNA complexes differ significantly (Gray et al., 1982). The Pf1 gene 5 protein consists of 144 amino acid residues (M_r = 15 400) and is 57 residues longer than the analogous Fd gene 5 protein (Maeda et al., 1982); however, all the amino acids so far identified as having a functional role in DNA binding are found in the N-terminal half of the Pf1 gene 5 protein sequence (Tsugita & Kneale, 1985; Plyte & Kneale, 1991), suggesting the possibility of a DNA-binding domain similar in size to the Fd gene 5 protein.

Limited proteolysis is a well-established technique for probing the tertiary structure of proteins and for investigation of the protection conferred by DNA binding (Plyte & Kneale, 1993). The existence of domains in DNA binding proteins is widespread, and their role has been established in a number of single-stranded DNA binding proteins. The adenovirus DNA binding protein can be separated into two distinct domains by digestion with chymotrypsin (Tsernoglou et al., 1985), of which the C-terminal domain is active in DNA binding. The T4 gene 32 protein can be cleaved into three functionally distinct domains (Spicer et al., 1979); the core domain is essential for single-stranded DNA binding, and all the amino acids required for binding are found in this domain. However, cooperative association of the gene 32 protein along the ssDNA is not possible without the N-terminal domain. The C-terminal domain of the gene 32 protein is thought to be involved in the interaction with other cellular proteins (Chase & Williams, 1986).

In order to investigate the possibility of structural subdomains in the Pf1 gene 5 protein, we have employed limited proteolysis. Fragments obtained by limited proteolysis of the native protein and its complex with viral DNA were characterized to establish the accessible regions of the protein and to determine the minimum size of the protein fragment required for DNA binding.

EXPERIMENTAL PROCEDURES

Growth, Infection, and Lysis of Bacteria. *Pseudomonas aeruginosa* (strain K) was grown to midlog phase in a 5-L fermentor, infected with Pf1 phage, and grown for a further 4 h. The phage-infected cells were harvested by centrifugation at 4000 rpm 4 °C, for 30 min. The cells were resuspended in 200 mL of 100 mM NaCl and 10 mM Tris-HCl, pH 7.5, and the centrifugation was repeated twice to remove all traces of the growth medium.

Lysozyme (4.4 mg/g of cells) and 100 mM PMSF (4 μL/g of cells) were added to the resuspended cells (4 mL of lysing

† The work has been supported by research grants from SERC (U.K.) and the provision of a Leverhulme Trust Senior Research Fellowship from the Royal Society (to G.G.K.).

* Author for correspondence.

‡ Present address: Ontario Cancer Research Institute, Toronto M4X 1K9, Canada.

0006-2960/93/0432-3623$04.00/0

buffer/g of cells; 100 mM NaCl, 1 mM EDTA, and 10 mM Tris-HCl, pH 7.5) and stirred at room temperature for 5 min. Streptomycin sulfate (0.04 g/g of cells) was then added to precipitate bacterial DNA and the mixture stirred for a further 30 min at room temperature. The cell debris was pelleted by centrifugation at 15 000 rpm, 4 °C, for 30 min. Further streptomycin (0.09 g/g of cells) was added to the supernatant and stirred for 10 min at room temperature. Bacterial DNA was removed by centrifugation at 15 000 rpm, 4 °C for 1 h. The nucleoprotein complex was pelleted by ultracentrifugation at 50 000 rpm, 4 °C for 3 h. Pellets were resuspended overnight in 2 mL of buffer A (10 mM Tris-HCl, pH 7.5, and 1 mM EDTA) at 4 °C. At this stage the resuspended pellets could be used for further purification of the protein or nucleoprotein complex (Kneale, 1983).

Protein Purification. The resuspended nucleoprotein complex was clarified by centrifugation at 15 000 rpm, 4 °C, for 15 min and dissociated by addition of 5 M $MgCl_2$ (to a final concentration of 1.2 M); the ssDNA was removed by ultracentrifugation at 50 000 rpm, 4 °C, for 2 h. The supernatant was desalted down a G-25 (Pharmacia) gel filtration column into 10 mM Tris-HCl, pH 7.5, followed by ammonium sulfate precipitation. The protein was resuspended in 25 mM imidazole, pH 7.6, and dialyzed extensively against this buffer. The dialyzed protein was clarified and applied to a small (10- × 150-mm) chromatofocusing column (packed with PBE 94 resin; Pharmacia) previously equilibrated in 25 mM imidazole, pH 7.6. The protein was eluted with Polybuffer 74 (Pharmacia), pH 4.5 (diluted 1:8), at a flow rate of 40 mL/h. The pure protein was recovered by ammonium sulfate precipitation, resuspended at 4 °C in 10 mM Tris-HCl, pH 7.5, and dialyzed against this buffer. The various fragments produced by limited proteolysis of the Pf1 gene 5 protein (or its complex with DNA) were purified by chromatofocusing using a similar procedure.

Purification of the Pf1 Nucleoprotein Complex. The purification protocol for the Pf1 nucleoprotein complex was based on the procedure of Kneale and Marvin (1982). Resuspended complex (see above) was clarified, loaded onto 5–33% exponential sucrose gradients (1.5 mL of sample loaded per 37-mL gradient), and run in a Beckman SW28 rotor at 27 000 rpm, 4 °C, for 4 h. Pure fractions were identified by SDS–PAGE. The sucrose was removed by extensive dialysis against 10 mM Tris-HCl, pH 7.5, and the nucleoprotein complex pelleted at 50 000 rpm, 4 °C, for 2 h in a Beckman 70.1 Ti rotor. The nucleoprotein complex was resuspended in buffer A and clarified prior to storage at –20 °C.

Limited Protelysis of the Pf1 Gene 5 Protein and Nucleoprotein Complex. The protein (or nucleoprotein complex) was desalted into the appropriate digestion buffer prior to the addition of protease. Digestions with trypsin (TPCK treated; Sigma), α-chymotrypsin (TLCK treated; Sigma), or subtilisin (protease VIII from *Bacillus subtilis*; Sigma) were performed in 10 mM Tris-HCl and 1 mM EDTA, pH 7.4. The papain (from papaya latex; Boehringer Mannheim) was activated by incubation in 10 mM β-mercaptoethanol, 5 mM EDTA, and 10 mM PIPES-HCl, pH 6.8, for 15 min prior to proteolysis. The protein was desalted into 10 mM PIPES-HCl, pH 6.8, prior to digestion with papain. Digestion of the protein (2–3 mg/mL) with the various proteases was performed at 37 °C at a variety of enzyme/substrate ratios. Digestion with trypsin, α-chymotrypsin, and subtilisin was stopped by addition of PMSF to 1 mM; digestion with papain was terminated by the addition of $ZnCl_2$ to 3 mM. Samples were removed at 15-min intervals during digestion for SDS–PAGE analysis on 12.5% vertical slab gels (Plyte & Kneale, 1993). Gels were stained with Coomassie Brilliant Blue (R-250).

Fluorescence Spectroscopy. Fluorescence spectra were recorded on a Perkin Elmer LS5B luminescence spectrophotometer in the ratio mode. The excitation and emission slits were set to 2.5 nm unless stated otherwise. A sample of protein (250 uL) was excited at 295 nm and the emission spectrum recorded between 310 and 410 nm. DNA binding curves were determined by plotting the ratio of the fluorescence intensity at 350 nm to that at 330 nm against the ratio of ssDNA:protein concentration, expressed as nucleotides per protein monomer (Kneale & Wijnaendts, 1985).

Fluorescence polarization titrations were performed by the addition of protein to ssDNA or vice versa. Polarizers were attached to the excitation (10-nm slit width) and emission (20-nm slit width) slits. The sample (250 μL) was excited at 295 nm and the emission recorded at 340 nm. The horizontal and vertical components of the fluorescence emission were measured and fluorescence titrations curves were plotted as described by Carpenter and Kneale (1991).

Gel Retardation Assays. Oligonucleotides were radioactively labeled with ^{32}P and added to a series of tubes containing 0–30 μM protein (9 μL) and incubated at room temperature for 30 min. The nondenaturing gel (4% acrylamide top; 16% bottom) was preelectrophoresed at 23 mA (250–350 V) in 10 mM Tris-HCl, pH 7.5, 10 mM $NaBO_4$, and 1 mM EDTA for 30 min. Glycerol (2 μL of 50%) was added to the samples immediately prior to electrophoresis. The samples were loaded onto the gel and electrophoresed at 23 mA (250–350 V) until the tracking dye reached the interface between the 4% and 16% gels. The gel was then blotted onto Whatman 3MM paper. The dried gel was exposed to preflashed X-ray film (with an intensifying screen) at –20 °C for 2–12 h.

N-Terminal Sequencing. Samples were routinely purified by HPLC in an acetonitrile gradient and lyophilized from acetonitrile prior to sequencing. Automated Edman degradation was performed on an Applied Biosystems 477A pulse liquid amino acid sequencer fitted with a 120A separation system for analysis of the PTH-derivatized amino acids (University of Southampton).

Mass Spectrometry. Mass Spectrometry was carried out on a Finnegan MAT Lazermat mass analyzer at the Ludwig Institute for Cancer Research (London).

C-Terminal Sequencing. The samples were routinely purified by HPLC and lyophilized from acetonitrile. The samples were dissolved in 50 mM trisodium citrate, pH 6.0, to a concentration of ~8 mg/mL. Carboxypeptidase Y ("sequencing grade"—Boehringer Mannheim; 0.7 mg/mL in 50 mM sodium citrate, pH 6.0) was added to an enzyme/substrate ratio of 1:100 (w/w) and incubated at 30 °C. An aliquot (5 μL) was removed immediately (as a control), frozen in liquid nitrogen, and lyophilized overnight. Samples (5 μL) were then removed at 30-min time intervals, frozen, and lyophilized for dabsyl amino acid analysis on HPLC.

Dabsylation and Quantitation of Peptides. Dabsyl chloride (Fluka) was recrystallized from boiling acetone prior to use. The procedure described is for the derivatization of an amino acid standard solution (AA-S-18, 2.5 μmol/mL each amino acid; Sigma). A sample (100 μL) of the amino acid mixture in 0.1 M HCl was dried in a vacuum centrifuge. The sample was then redissolved in 200 μL of dH_2O (ELGA) and evaporated three times to remove all traces of HCl before finally being dissolved in 200 μL of 150 mM $NaHCO_3$, pH 9.0 (adjusted to pH 9.0 with Na_2CO_3). An equal volume of freshly prepared dabsyl chloride (15 mM in acetone: 4.84

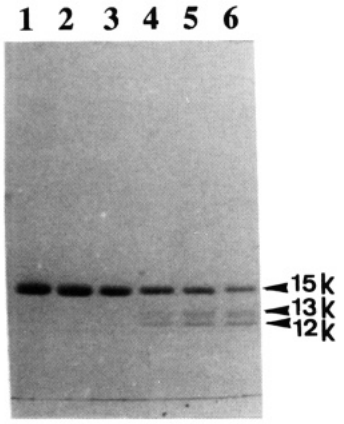

FIGURE 1: Endogenous proteolysis of Pf1 gene 5 protein. Samples of the protein were left at room temperature in 10 mM Tris-HCl buffer (pH 7.5) for 0, 7, 14, 28, 34, and 42 days (tracks 1–6) and analyzed on 12.5% SDS–polyacrylamide gels.

mg/mL) was added and the mixture vortexed vigorously for 1 min. The mixture was incubated at 70 °C in a tightly stoppered tube for 15 min. The dabsylated amino acids were diluted with 40 mM sodium phosphate, pH 6.5/ethanol (1:1 v/v), clarified, and stored at −20 °C ready for amino acid analysis. The dabsylated amino acids were separated by HPLC on an Ultrasphere ODS 5-μm column (Beckman) using a procedure similar to that reported by Vandrell and Aviles (1986).

Electron Microscopy. Pf1 nucleoprotein complex was digested with papain until digestion to the limit peptide was complete (as judged by SDS–polyacylramide gel electrophoresis). The proteolyzed complex was centrifuged at 50 000 rpm in a Beckman 70.1 Ti rotor and resuspended in buffer A. Samples of papain-digested nucleoprotein complex were diluted and dried onto carbon-coated copper grids (400 mesh, 3.05 × 3.05 mm). The samples were stained by incubation in 1% uranyl acetate solution for 2 min. Samples were visualized on a Phillips 300 TEM. Measurements of helical pitch were determined from the negatives on an optical densitometer.

RESULTS

Endogenous Proteolysis. Storage of the Pf1 gene 5 protein over a period of months leads to its degradation into two specific fragments of approximately 12 and 13 kDa (Figure 1). The digestion appears to proceed via the larger fragment (A1) since this eventually disappears at the expense of the smaller fragment, A2. In an attempt to prevent the degradation, PMSF, EDTA, or $ZnCl_2$ were added to the protein and the degradation was monitored over a period of weeks. Only the presence of EDTA resulted in a significant decrease in the rate of degradation (results not shown). The source of degradation is most probably a contaminating protease requiring divalent cations for activity.

The degraded fragments A1 and A2 were purified for further characterization. Separation of the fragments was achieved by chromatofocusing on PBE 94. The two fragments eluted from the column with apparent p*I*'s of 5.8 and 5.5, respectively, and were well resolved from any residual undigested protein. Since it was subsequently established that both A1 and A2 are able to bind DNA in equilibrium binding assays (results not shown), only the smaller fragment A2 was characterized further.

The N-terminal sequence of fragment A2 was determined by Edman degradation and is identical to that of the native protein (Table I). To determine the sequence of the C-terminal end of the fragment A2, carboxypeptidase Y digestion was employed, followed by quantitative determination of the amino acids released from the C-terminus by HPLC separation of the dabsylated amino acids. The major cutting point was identified as alanine 115. The amino acid composition of A2 was subsequently determined, and this was also consistent with the sequence corresponding to residues 1–115 of the intact protein.

Limited Proteolysis of the Pf1 Gene 5 Protein. Attempts were then made to generate fragments of the protein under more controlled conditions. Limited proteolysis of the protein and nucleoprotein complex was performed with papain, trypsin, α-chymotrypsin, and subtilisin to further explore the domain structure of the Pf1 gene 5 protein. Digestion of the free protein with trypsin, α-chymotrypsin, and the papain failed to generate a protease-resistant domain. Digestion of the protein at enzyme/substrate ratios varying from 1:100 to 1:1000 for various times resulted in either complete degradation or no degradation at all; no stable intermediate could be identified by SDS–PAGE. In contrast, limited digestion of the free protein with subtilisin generated a resistant domain with an apparent molecular mass (as judged by gel electrophoresis) in the region of 12 kDa under a wide variety of conditions (Figure 2).

Subsequent purification by chromatofocusing revealed a major peak (fragment S1) with an apparent p*I* of 4.6. Automated Edman degradation of fragment S1 indicated that six amino acids had been cleaved from the N-terminus by subtilisin, since the N-terminal sequence started at Glu 7 of the intact protein (Table I). Carboxypeptidase Y analysis of S1 indicated that the C-terminal sequence terminated close to residue 102. Thus the most likely sequence of S1 corresponds to residues 7–102 of the native protein.

Limited Proteolysis of the Pf1 Nucleoprotein Complex. In order to assess whether specific regions of the Pf1 gene 5 protein were protected further in the complex with viral DNA, a similar series of proteolytic experiments was conducted on the purified nucleoprotein complex. In contrast to the results obtained on the free protein, protease-resistant fragments were produced by all four enzymes upon digestion of the purified Pf1 gene 5 protein–DNA complex (Figure 3). Digestion of the nucleoprotein complex with trypsin produced three peptide fragments, the smallest having a molecular mass of approximately 12 kDa as judged by SDS–PAGE. Papain, subtilisin, and α-chymotrypsin digestion of the complex each produced a band at ca. 12 kDa (i.e., with a similar mobility to the band resulting from subtilisin digestion of the free protein). It is clear that the nucleoprotein complex is very resistant to proteolysis, implying considerable protection of a globular domain in the high molecular weight complex. In one experiment, the nucleoprotein complex was digested with trypsin at an E/S ratio of 1:25 (w/w) for 24 h at 37 °C without further digestion beyond the 12-kDa fragment.

Fragments resulting from papain digestion of the Pf1 nucleoprotein complex were isolated by dissociation of the complex and were purified for further characterization. Purification by chromatofocusing showed a major peak (P1) which eluted with an apparent p*I* of 4.5. The N-terminus of P1 was found to be identical to that of the intact protein (Table I), although there is some evidence of partial degradation of the N-terminal methionine. The amino acids released by carboxypeptidase Y digestion of P1 indicated that the C-terminal residue corresponded to a position in the sequence 112–113 of the native gene 5 protein. However, due to the

Table I: Characterization of Proteolytic Fragments of the Pf1 Gene 5 Protein[a]

fragment	mass (kDa)	N-terminus	C-terminus	location	act.
A1	~13	nd[b]	nd	nd	+
A2	~12	MNMFATQGGVVEL...	...PTAQAPA	1–115	+
S1	~12	QGGVVEL...	...QVLDVLLAV	7–102	–
P1	11.9	MNMFATQGGVVEL...	...PMAPTAQ(A)	1–112	+

[a] N-Terminal analysis was by Edman degradation, C-terminal analysis by carboxypeptidase Y digestion. Approximate molecular masses of the fragments were estimated from their mobility on SDS gels or, in the case of P1, accurately determined by mass spectrometry. The proposed location of these fragments in the intact Pf1 gene 5 protein sequence is shown. DNA binding activity was monitored by fluorescence titrations using poly(dT). [b] Not determined (nd).

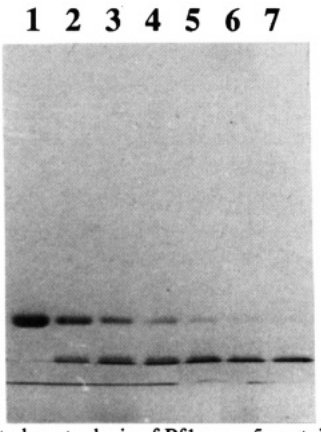

FIGURE 2: Limited proteolysis of Pf1 gene 5 protein with subtilisin. Digestion was performed at an enzyme/substrate ratio of 1:2000 in 10 mM Tris-HCl (pH 8.0) at 37 °C for 0, 15, 30, 45, 60, 75, and 90 min (tracks 1–7) and analyzed on 12.5% SDS–polyacrylamide gels.

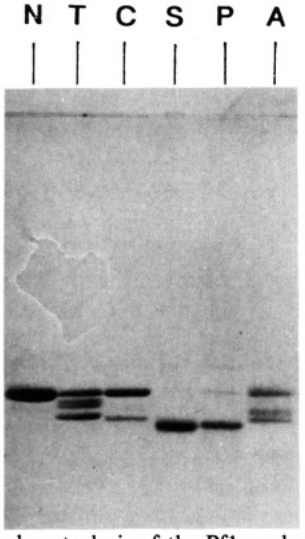

FIGURE 3: Limited proteolysis of the Pf1 nucleoprotein complex. Digestion of the purified Pf1 nucleoprotein complex was performed with the following enzymes (and enzyme:substrate ratios): T, trypsin (1:100); C, α-chrymotrypsin (1:50); S, subtilisin (1:500); P, papain (1:1000). All digestions were performed at 37 °C for 90 min. Tracks N and A show the native protein band and the products of endogenous proteolysis, respectively. Electrophoresis was performed on 12.5% SDS–polyacrylamide gels.

unusual amino acid composition in this region of the sequence, the precise location of the C-terminus could not be identified with certainty. The presence of a methionine residue at position 107 in the gene 5 protein sequence allowed cleavage of P1 with cyanogen bromide, which produced several peptides including one from the C-terminus. Upon automated Edman degradation of this peptide, Glu 112 was the last recorded residue. As a final confirmation of the sequence of P1, mass spectrometry was performed and gave a molecular mass for the fragment of 11 931 Da. Taken together, the results

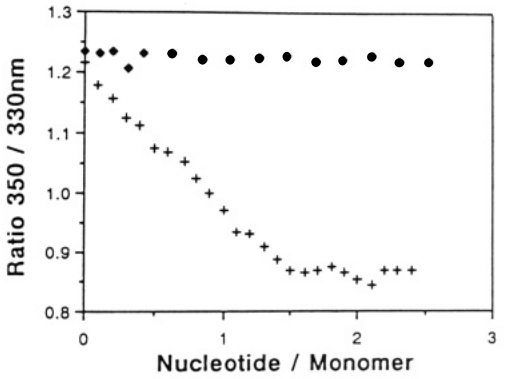

FIGURE 4: Fluorescence spectral shift titrations of fragments S1 and P1 with poly(dT). Protein fluorescence from S1 (♦) and P1 (+) was measured in 0.1 M NaCl and 10 mM Tris-HCl (pH 8.0) at a protein concentration of 20 μM with increasing additions of poly(dT). The ratio of fluorescence at 350 nm to that at 330 nm was recorded using an excitation wavelength of 295 nm.

establish that P1 corresponds to amino acid residues 1–112 (or possibly 2–112) of the Pf1 gene 5 protein sequence.

DNA Binding Studies on the Protease-Resistant Fragments. The purified proteolytic fragments S1 and P1 were characterized for DNA binding activity. Titrations of S1 and P1 with poly(dT) were initially performed by measuring the spectral shift arising from Trp 14 in the gene 5 protein, as described elsewhere for the intact protein (Kneale & Wijnaendts, 1985). The results show that the fluorescence emission ratio ($\alpha = F_{350}/F_{330}$) of S1 remains unchanged upon addition of poly(dT), indicating that this fragment does not bind cooperatively to polynucleotides (Figure 4). In contrast, the binding curve obtained for P1 is essentially identical to that obtained for the native protein. In the reverse titration [protein added to poly(dT)], binding was monitored by the increase in polarization of fluorescence. This technique gives additional information on the mode of binding since, under these conditions, the protein is able to bind DNA at both sites of the dimer (Carpenter & Kneale, 1991). Again fragment P1 binds DNA identically to the native protein with the expected stoichiometry of 4 nucleotides per subunit (Figure 5), whereas fragment S1 shows no evidence of DNA binding. The binding constant for P1 is estimated as 9×10^7 M^{-1}, comparable with that determined for the native protein (Carpenter & Kneale, 1991).

Gel retardation studies were performed on fragments S1 and P1 with a single-stranded oligonucleotide fragment (Figure 6). Binding of P1 to the oligonucleotide is apparent from the shift in the position of the radioactively labeled oligonucleotide. However, binding is clearly weaker than for the intact protein, since only a small proportion of the oligonucleotide is retained as a high molecular weight complex. The majority of the oligonucleotide is seen as a smear, suggesting a less stable complex which dissociates during the electrophoretic separation. Gel retardation experiments with S1, however, gave

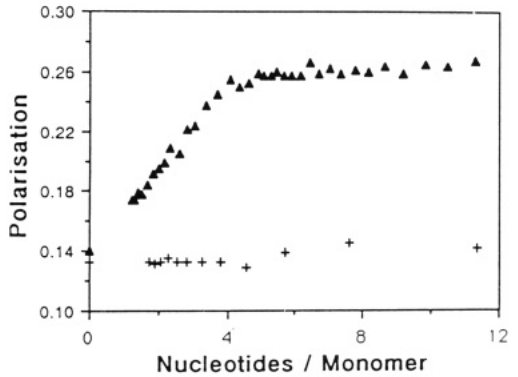

FIGURE 5: Fluorescence polarization titrations of fragments S1 and P1 with poly(dT). The polarization ratio was determined after each addition of S1 (+) or P1 (▲) to a solution of 11.3 μM poly(dT) in 0.1 M NaCl and 10 mM Tris-HCl (pH 8.0). Excitation was at 295 nm, and emission was recorded at 340 nm.

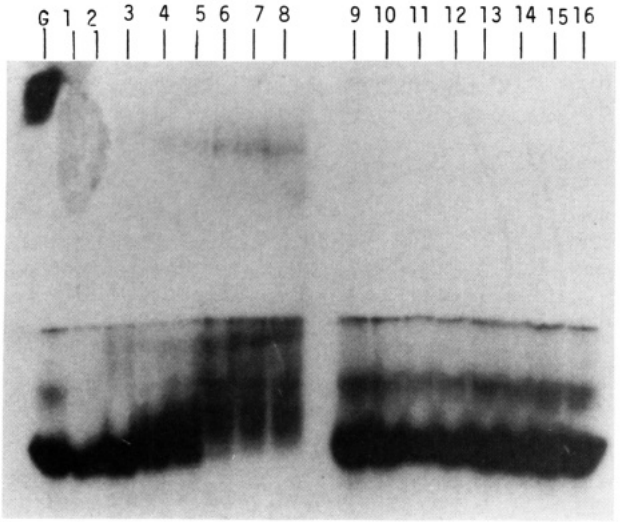

FIGURE 6: Gel retardation analysis of fragments S1 and P1. The single-stranded ^{32}P-labeled oligonucleotide d(TACTTAGCCG-GCACGAGGCG) was subjected to electrophoresis in a nondenaturing polyacrylamide gel (4–16%) after incubation with 0, 1.9, 3.8, 7.7, 11.5, 15.4, 23, and 31 μM protein (P1, lanes 1–8; S1, lanes 9–16). The high molecular weight complex formed with the intact Pf1 gene 5 protein (31 μM) is shown in the lane marked G. The oligonucleotide concentration in each reaction was 1.35 μM.

no indication of complex formation even at the highest protein concentrations used, in agreement with the fluorescence results.

Electron Microscopy of Proteolyzed Nucleoprotein Complex. Electron microscopic analysis of the Pf1 nucleoprotein complex has shown that it is a tightly wound helical complex having a helical pitch of approximately 60 Å (Gray et al., 1982). Electron microscopy was used in order to examine the morphology and gross structural parameters of the nucleoprotein complex after limited proteolysis with papain, under conditions that generate the fragment P1. Figure 7 shows a region of the proteolyzed complex, showing the helical nature of the complex. The dimensions of the proteolyzed complex are comparable with those of the native complex, showing that no gross conformational change has taken place.

DISCUSSION

The Domain Structure of the Pf1 Gene 5 Protein. The results of the current investigation suggest that the N-terminal 112 residues of the Pf1 gene 5 form a tightly folded globular structure that closely associates with other subunits along the DNA lattice. The C-terminal sequence 113–144 is found outside of the compact DNA-binding domain and is accessible

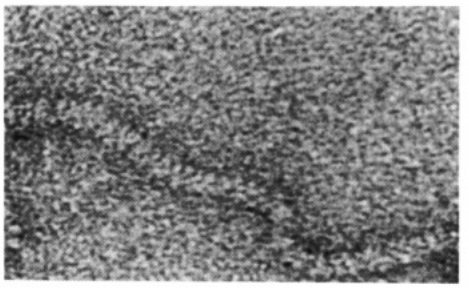

FIGURE 7: Electron microscopy of proteolyzed Pf1 nucleoprotein complex. The complex was digested for 1 h with papain (enzyme/substrate ratio 1:1000) at room temperature and, after termination of the reaction, centrifuged at 50000g for 2.5 h. The pelleted complex was resuspended and negatively stained with uranyl acetate. Degradation of the gene 5 protein to the limit peptide (12 kDa) was confirmed by SDS–PAGE. The scale bar shown on the photograph represents 50 nm.

to a variety of proteases in the protein–DNA complex. There is some evidence that this structure persists in the isolated protein dimers, since limited digestion of the free protein with subtilisin gives rise to a fragment that is partially resistant to further cleavage. In this case, however, the resulting fragment is somewhat smaller, lacking the 6 N-terminal residues, as well as a further 10 residues from the C-terminus.

It is likely that both these regions (residues 1–6 and 103–112) are accessible in isolated protein dimers, but are buried by protein–protein or protein–DNA contacts when forming a helical array in the complex with viral DNA. Moreover, at least one of these regions must be critical for DNA binding, since when both are removed, the interaction with DNA is completely abolished. If direct interactions of amino acid residues in these two sequences are made with DNA, then likely candidates would be aromatic and basic residues such as Phe 4 or Lys 105. Alternatively, these regions of the polypeptide may be required for maintaining the structural integrity of the intact nucleoprotein complex.

The N-terminal core domain (1–112) of the Pf1 gene 5 protein is clearly capable of binding to single-stranded DNA with a high affinity, and this domain is indistinguishable from the intact protein during fluorescence titrations with polynucleotides. Furthermore, electron microscopy reveals that the helical structure of the nucleoprotein complex persists when the C-terminal sequence is removed. The results of gel retardation analysis, however, suggest that the C-terminal sequence of the protein does play a role in DNA binding, offering additional stabilization of the nucleoprotein complex, at least in the complex with oligonucleotides.

The Role of the C-Terminal Domain. The oligonucleotide binding characteristics of P1, as judged by gel retardation analysis, differ from those of the native protein. The oligonucleotide is smeared throughout the lane rather than retained at the top of the gel as a discrete complex. This phenomenon is indicative of the formation of complexes with a relatively short lifetime, in which the equilibrium is disturbed in the time taken to enter the gel (Fried, 1989). The C-terminal sequence (residues 113–144) appears to stabilize the complex. We propose that the removal of this sequence results in an increase in the rate of dissociation of the bound protein from the DNA (k_{off}), probably with a concomitant increase in the rate of association (k_{on}), since the equilibrium binding constant ($K_{eq} = k_{on}/k_{off}$) is not significantly changed.

Removal of the 32-residue C-terminal sequence does not abolish the ability of the Pf1 gene 5 protein to bind

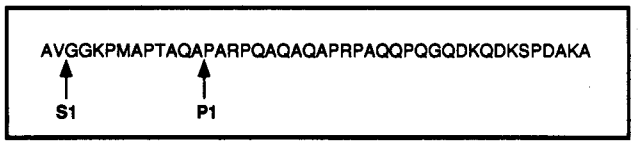

FIGURE 8: C-Terminal sequence of the Pf1 gene 5 protein. Residues 101–144 of the native Pf1 gene 5 protein sequence (Maeda et al., 1982) with the proposed C-terminal cutting points for subtilisin (S1) and papain (P1) indicated.

cooperatively to polynucleotides, since the clustering of proteins along the DNA lattice is the basis of the fluorescence assay (Greulich et al., 1985; Kneale & Wijnaendts, 1985). Similarly, the C-terminal sequence cannot be responsible for maintaining the dimeric structure of the protein, since both single- and two-site binding is observed for the fragment P1, as for the native protein (Carpenter & Kneale, 1991). A possible role for the C-terminal sequence of the Pf1 gene 5 protein is to clamp the DNA, once bound, thus stabilizing the complex. Alternatively, this sequence could be responsible for increased protein–protein interactions between adjacent protein dimers in the nucleoprotein complex. The lack of such a C-terminal tail in the Fd gene 5 protein may be responsible for its decreased stability to dissociation by salt, compared with the Pf1 gene 5 protein (Kneale, 1983).

The sequence of the 32 amino acid residues at the C-terminus is unusual; it lacks any of the larger hydrophobic amino acids and is rich in alanine, proline, and glutamine which together account for 22 of the 32 residues (see Figure 8). The remaining 10 residues include 5 basic and 3 acidic residues. A number of proteins are now known to possess unusual sequences of this type which do not form part of the globular structure. For example the interdomain segments that link the lipoyl domains of the multisubunit enzyme pyruvate dehydrogenase are very rich in alanine and proline, together with a number of polar and charged amino acids; these segments have been shown to possess an unusual degree of conformational flexibility (Radford et al., 1989). Preliminary results of NMR spectroscopy suggest that the C-terminal sequence of the Pf1 gene 5 protein is likewise highly flexible, in agreement with its enhanced susceptibility to proteolysis. Further investigation of the structure and dynamics of the C-terminal region of the Pf1 gene 5 protein will be required to understand its role in stabilizing the nucleoprotein complex formed with viral DNA.

ACKNOWLEDGMENT

We are grateful to Dr. M. Gore (University of Southampton) for performing automated N-terminal amino acid sequencing and to N. Totty (Ludwig Institute) for mass spectroscopic analysis. We thank Dr. A. Thorne for advice and discussion and B. Hartfree for skilled assistance with electron microscopy.

REFERENCES

Alberts, B., Frey, L., & Delius, H. (1972) *J. Mol. Biol. 68*, 139–152.
Carpenter, M. L., & Kneale, G. G. (1991) *J. Mol. Biol. 217*, 681–689.
Chase, J., & Williams, K. (1986) *Annu. Rev. Biochem. 55*, 103–136.
Fried, M. (1989) *Electrophoresis 10*, 366–376.
Gray, C. W., Kneale, G. G., Leonard, K., Siegrist, H., & Marvin, D. A. (1982) *Virology 116*, 40–52.
Greulich, O., Wijnaendts, R., & Kneale, G. G. (1985) *Eur. Biophys. J. 11*, 195–201.
Kneale, G. G. (1983) *Biochim. Biophys. Acta 739*, 216–224.
Kneale, G. G. (1992) *Curr. Opin. Struct. Biol. 2*, 124–130.
Kneale, G. G., & Marvin, D. A. (1982) *Virology 116*, 53–60.
Kneale, G. G., & Marvin, D. A. (1983) *J. Mol. Biol. 171*, 229–232.
Kneale, G. G., & Wijnaendts, R. (1985) *Eur. J. Biochem. 149*, 85–93.
Kneale, G. G., Freeman, R., & Marvin, D. A. (1982) *J. Mol. Biol. 156*, 279–286.
Maeda, K., Kneale, G. G., Tsugita, A., Short, N. J., Perham, R. N., Hill, D. F., & Peterson, G. B. (1982) *EMBO J. 1*, 255–261.
Michel, B., & Zinder, N. D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4937–4941.
Plyte, S. E., & Kneale, G. G. (1991) *Protein Eng. 4*, 553–560.
Plyte, S. E., & Kneale, G. G. (1993) in *Methods in Molecular Biology: Protein–Nucleic Acid Interactions* (Kneale, G. G., Ed.) Humana Press, Totowa, NJ.
Radford, S. E., Laue, E. D., Perham, R. N., Martin, S., & Appella, E. (1989) *J. Biol. Chem. 254*, 767–775.
Spicer, E., Williams, K. & Konigsberg, W. (1979) *J. Biol. Chem. 254*, 6433–6436.
Tsernoglou, D., Tsugita, A., Tucker, A., & van der Vliet, P. (1985) *FEBS Lett. 206*, 119–132.
Tsugita, A., & Kneale, G. G. (1985) *Biochem. J. 228*, 193–199.
Vendrell, J., & Aviles, F. X. (1986) *J. Chromatogr. 358*, 401–413.
Yen, T. S. B., & Webster, R. E. (1982) *Cell 29*, 337–345.