The EMBO Journal, Vol. 1 No. 2, pp.255–261, 1982

The DNA-binding protein of Pf1 filamentous bacteriophage: amino-acid sequence and structure of the gene

Kayo Maeda[1], G. Geoff Kneale[1], Akira Tsugita[1*], Nicholas J. Short[2], R.N. Perham[2], Diana F. Hill[3,4], and George B. Petersen[3,4]

[1]European Molecular Biology Laboratory, Heidelberg, FRG, [2]Department of Biochemistry, University of Cambridge, Cambridge, UK, and [3]MRC Laboratory of Molecular Biology, Hills Road, Cambridge, UK

Communicated by Sir John Kendrew
Received on 12 February 1982

The amino-acid sequence of the single-stranded DNA-binding protein of bacteriophage Pf1 and the nucleotide sequence of the corresponding gene have been determined. The protein has 144 amino acids and a molecular weight of 15 400; the gene consists of 435 nucleotides. The amino-acid sequence was determined by Edman degradation, carboxypeptidase A, B, and P digestion of intact protein and of peptides derived by chymotrypsin, *Staphylococcus aureus* V8 protease, and trypsin digestion. The nucleotide sequence was determined by the dideoxy method after random cloning of fragments of Pf1 DNA into M13. No sequence homology could be established between the amino-acid sequence of the DNA-binding protein of *Pseudomonas aeruginosa*-specific bacteriophage Pf1 and bacteriophage fd of *Escherichia coli*.

Key words: DNA sequence/Pf1 phage/protein sequence/single-stranded DNA-binding protein

Introduction

Filamentous bacteriophage may be subdivided into two classes, exemplified by fd (M13, f1) which infects *Escherichia coli* and Pf1 which infects *Pseudomonas aeruginosa* (Denhardt *et al.*, 1978). The virions of both classes consist of a single-stranded circular DNA molecule with 6500 nucleotides (± 15%, depending on the species) encapsulated by a protein coat, comprised of largely α-helical subunits (Marvin and Wachtel, 1976). On infection of the host bacterium, the viral DNA strand is converted to a double-stranded circular form, which in turn acts as a template for synthesis of progeny viral strands (Ray, 1969). The synthesis of progeny single-stranded DNA depends on the presence of a viral DNA-binding protein (in bacteriophage fd, the product of viral gene 5) which forms a complex with the nascent DNA strand. This nucleoprotein complex is an intermediate between DNA replication and phage assembly, and can be isolated from the cytoplasm of infected cells (Pratt *et al.*, 1974; Kneale and Marvin, 1982). The structure of the Pf1 nucleoprotein complex has been investigated by electron microscopy (Gray *et al.*, 1982) and by X-ray fibre diffraction (Kneale *et al.*, in press). During the assembly of mature virions at the bacterial cell membrane, the DNA-binding protein is displaced from the viral DNA by the coat protein (Marvin and Wachtel, 1976). The DNA-binding protein thus plays a major role in both viral DNA replication and assembly of the virion.

The DNA-binding protein of bacteriophage fd (the gene 5 protein) has a mol. wt. of 9800 and its amino-acid sequence is known (Nakashima *et al.*, 1974). The analogous protein of bacteriophage Pf1 has a mol. wt. of ~16 000 and a very different amino-acid composition from that of the fd protein (Kneale and Marvin, 1982). Both the structure and the salt stability of the respective *in vivo* nucleoprotein complexes also differ substantially (Gray *et al.*, 1982; Kneale and Marvin, 1982). In this paper we present the results of two lines of investigation. The amino-acid sequence of the purified DNA-binding protein of bacteriophage Pf1 and the complete nucleotide sequence of the corresponding gene have been determined.

[4]Permanent address: Department of Biochemistry, University of Otago, Dunedin, New Zealand

*To whom reprint requests should be sent.

Results

Amino-acid composition and C-terminal, N-terminal sequences

The amino-acid composition of the total protein is listed in Table I.

The N-terminal amino-acid sequence was determined by Edman degradation as Met-Asn-Met-Phe-Ala[5]-Thr-Gln-Gly-.

Table I. Amino-acid composition of DNA-binding protein of bacteriophage Pf1

Amino acid	Determined from amino-acid analysis[a]			Determined from sequence
	Hydrolysis time 25 min	Hydrolysis time 50 min	Integer value	
Asp	10.8	10.6	11	4
Asn				7
Thr	11.2	10.8	12	12
Ser	5.4	4.9	6	6
Glu	20.9	20.6	21	8
Gln				13
Pro	12.5	13.3	13	13
Gly	12.2	12.0	12	11
Ala	18.0	18.0	18	18
Val	9.2	10.1	10	11[d]
Met	2.5	2.3	2–3	3
Ile	6.3	6.6	7	7
Leu	7.9	7.9	8	7
Tyr	3.0	3.0	3	3
Phe	5.3	5.1	5	5
His	0.6	0.5	0–1	0
Lys	8.2	8.4	8	8
Arg	6.5	6.8	7	6
Cys[b]	1.3		1–2	1
Trp[c]			1–2	1

[a]The number of alanine residues was set at 18 on the basis of the mol. wt. of DNA-binding protein and the molar concentration of alanine. Values for other amino acids were calculated as molar ratios relative to alanine.
[b]Cysteine was determined after performic acid oxidation of the protein (Hirs, 1956).
[c]Data obtained by spectrophotometry (Goodwin *et al.*, 1946).
[d]Ten valines were obtained from the amino-acid sequence.

© IRL Press Limited, Oxford, England. 0261-4189/82/0102-0255$2.00/0

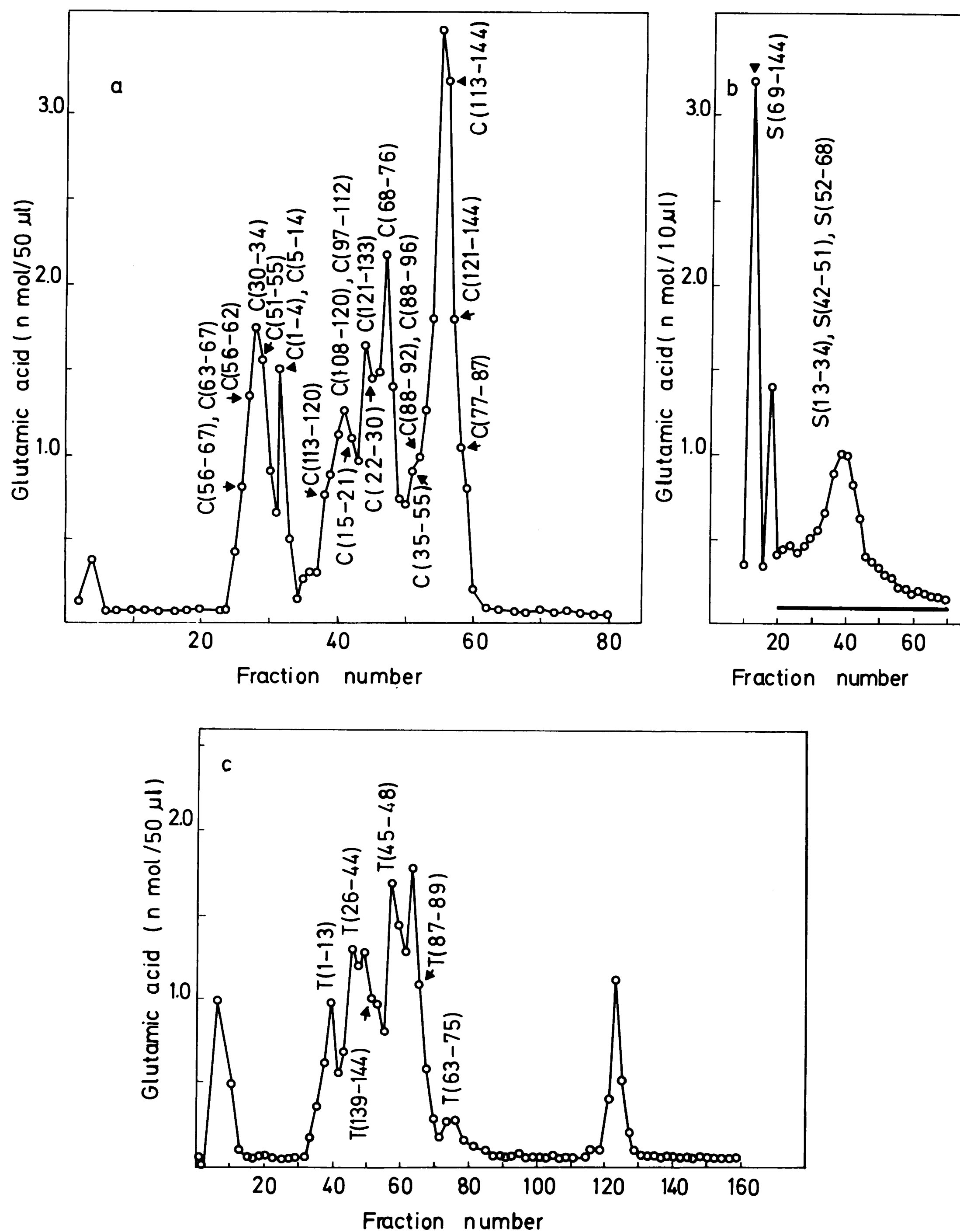

Fig. 1. **a**, Chymotryptic peptides of DNA-binding protein of phage Pf1 on SP-Sephadex C25. The digest of 100 nmol of purified protein was applied to a column (0.6 cm x 15 cm) and eluted with a gradient of pyridine acetate buffers. The gradient was made with 30 ml of each buffer as follows: chamber 1 and 2, 0.1% pyridine acetate (pH 3.8); chamber 3, 0.25% (pH 3.9); chamber 4, 0.75% (pH 4.25); chamber 5, 1.5% (pH 4.47); chamber 6, 2.5% (pH 4.65); chamber 7, 5.0% (pH 4.87); chamber 8, 7.5% (pH 4.96); chamber 9, 20% (pH 5.18). The elution was carried out at a flow rate of 4 ml/h at 37°C and 2 ml fractions were collected. The final elution was carried out with a second gradient of 20% pyridine acetate (pH 5.18) to 30% (pH 6.0) at 50°C. The profile was obtained by acid hydrolysis of 50 µl of eluate from every second fraction followed by amino-acid analysis. **b**, *S. aureus* V8 protease digest of DNA-binding protein of phage Pf1 on Biogel P30. The digest of 120 nmol of purified protein was applied to a 1 cm x 60 cm column and eluted with 70% formic acid at a flow rate of 5 ml/h at room temperature. 1 ml fractions were collected. The profile was obtained by acid hydrolysis of 10µg of eluate from every second fraction. **c**, Tryptic peptides of DNA-binding protein on SP-Sephadex C25. The digest of 120 nmol of purified protein was applied to a column (0.6 cm x 15 cm) and eluted with a gradient of pyridine acetate buffer as described in **a**. Fractions (2 ml) were collected. The profile was obtained by acid hydrolysis of 50 µl of eluate from every second fraction.

The C-terminal amino-acid sequence was -Ala-Lys-Ala, determined by carboxypeptidase A and B digestion.

It is possible to determine the absolute amount of the protein using amino-acid analysis data. According to the sequence, 1 mol protein has 18 mol of alanine residues. The molar concentration of alanine divided by 18 gives the molar concentration of protein. Under our hydrolysis conditions hydrolysis of protein is complete and alanine is not decomposed during hydrolysis (Tsugita and Scheffler, 1982).

Table II. Amino-acid composition of chymotryptic peptides from DNA-binding protein of phage Pf1

	C(1–4)		C(5–14)		C(15–21)		C(22–30)		C(31–34)		C(35–55)		C(51–55)		C(56–62)		C(56–67)		C(63–67)	
Asp	1.0				1.2	1					1.0				1.1		1.7	1	1.1	1
Asn		1										1				1		1		
Thr			0.9	1	2.9	3	2.5	3							1.0	1	0.9	1		
Ser							0.8	1	1.0	1	1.5	2	1.2	1						
Glu			2.0	1			1.0	1	1.2		2.5	3	2.0	2	1.3	1	1.3	1		
Gln				1						1										
Pro											1.9	2			1.0	1	0.9	1		
Gly			2.0	2			1.2	1	0.4		2.7	3	0.4				0.4			
Ala			1.0	1					1.0	1	3.0	3					1.0	1	1.0	1
Val			0.7	1	1.1	1			1.1	1							0.7	1	0.7	1
Met	1.9	2																		
Ile							0.4	1			1.9	3	0.7	1	0.7	1	1.1	2	0.4	1
Leu			0.8	1					0.2						2.0	2	1.6	2		
Tyr					0.5	1	0.7	1			0.2	1	0.7	1						
Phe	1.0	1									0.8	1					0.8	1	1.4	1
His																				
Lys					1.0	1	0.9	1			1.0	1								
Arg											0.8	1								
Cys																				
Trp[a]			1.1	1																

	C(68–76)		C(77–87)		C(88–92)		C(88–96)		C(97–112)		C(108–120)		C(113–120)		C(113–144)		C(121–133)		C(121–144)	
Asp			1.1	1	1.0		1.3		0.9	1					3.2	3			3.1	3
Asn						1		1												
Thr			1.0	1	0.9	1	1.8	2	0.6	1	1.3	1								
Ser			1.0	1	0.4						0.3				1.2	1			1.2	1
Glu	2.3	1	1.0				1.4		1.8		3.7		2.4		8.1		4.7		6.5	
Gln		1		1				1		1		3		2		8		5		6
Pro	1.0	1	1.4	1					2.0	2	3.0	3	2.0	2	5.7	6	2.7	3	4.1	4
Gly	1.8	1			1.3	1	1.3	1	1.4	2	1.3				1.5	1	1.5	1	1.3	1
Ala	0.5		1.0	1					3.0	3	5.0	5	3.0	3	8.0	8	3.5	3	5.0	5
Val	0.9	1	1.3	2			1.1	1	1.8	2	0.4									
Met									0.5	1										
Ile	0.3				0.6	1	0.7	1												
Leu	0.7	1					1.0	1	1.5	2										
Tyr																				
Phe	0.8	1	0.9	1																
His																				
Lys	1.0	1							0.7	1					2.8	3			3.0	3
Arg			2.0	2	0.8	1	1.1	1			1.0	1	1.1	1	1.8	2	1.0	1	0.9	1
Cys	0.6	1																		
Trp[a]																				

Total composition by complete acid hydrolysis (left column) and total compositions from summary of sequence analysis (right column).
[a]Tryptophan content was established by hydrolysis with 3N mercaptoethane sulfonic acid (Penke *et al.*, 1974) using our micro-method (Maeda *et al.*, in preparation).

Amino-acid sequence of the peptides

After digestion by α-chymotrypsin, the peptides were eluted as shown in Figure 1a. The elution pattern shows six peaks. The fractions were rechromatographed using 70% formic acid on Biogel P10 or P4 (200 mesh) columns, depending on the estimated mol. wt. of the peptides. Twenty different peptides were purified.

The protein was digested by *Staphylococcus aureus* V8. SDS-gel electrophoresis showed that the digests consisted of a large peptide (mol. wt. 8000) and several small peptides. Biogel P30 was used for fractionation of the large peptide and the small peptides were pooled (Figure 1b). The small peptides were rechromatographed on SP-Sephadex C25 column (0.6 cm x 15 cm) using conditions similar to those described in the legend of Figure 1a except that each gradient volume was 15 ml. By this procedure four pure peptides were obtained.

The trypsin digest was fractionated as shown in Figure 1c. Further purification of fractions was carried out by Biogel P10 or P6 (200 mesh) columns using 70% formic acid. Tryptic digestion did not yield as many peptides as predicted from the number of lysine and arginine residues in the protein. The amino-acid composition of the purified peptides is listed in Tables II and III and the sequence results are summarized in Figure 2.

Alignment of the peptides

The N-terminal sequence of the protein obtained by Edman degradation was confirmed and extended to the 14th

Table III. Amino-acid composition of peptides obtained from digestion with *Staphylococcus* V8 protease and trypsin

	S(13–34)		S(42–51)		S(52–68)		S(69–144)		T(1–13)		T(26–44)		T(45–48)		T(63–75)		T(87–89)		T(126–135)		T(139–144)	
Asp	1.0	1	1.2		1.7	1	6.4	5	1.2		0.8		0.7		1.3	1			1.4	1	1.5	1
Asn				1		1		1		1				1								
Thr	4.5	6			1.3	1	4.3	4	1.1	1	0.8	1										
Ser	1.6	2			1.4	1	2.3	2	0.8		1.8	2			0.9						0.9	1
Glu	2.0	1	1.4	1	3.0	3	11.7		2.0	1	3.1	2			2.0	1			2.8		0.5	
Gln		1						12		1		1				1				4		
Pro	0.3				0.9	1	9.7	10			1.9	2			0.9	1			1.4	2	0.9	1
Gly	1.4	1	2.8	3	0.5		5.0	5	2.9	2	2.3	2	0.9	1	1.4	1	1.0	1	0.9	1	0.4	
Ala	1.0	1	2.0	2	1.3	1	12.0	12	0.9	1	3.0	3	1.0	1	1.4	1			1.0	1	2.0	2
Val	1.5	2			1.1	1	6.2	6	0.8	1	1.2	1			1.8	2			0.6			
Met							0.9	1	1.5	2												
Ile	0.5	1			1.8	3	1.0	1			1.8	3			0.5	1						
Leu	0.8	1			1.5	2	4.1	4	0.8	1					0.9	1			0.6			
Tyr	1.7	2			0.6	1	0.4				0.7	1										
Phe			0.7	1	0.7	1	2.3	2	0.7	1					0.5	1	1.0	1				
His																						
Lys	1.8	2	1.2	1			5.0	5					1.1	1	0.8	1			0.9	1	1.3	1
Arg			0.9	1			5.1	5			0.8	1					1.5	1				
Cys							1.0	1							0.8	1						
Trp	0.9	1																				

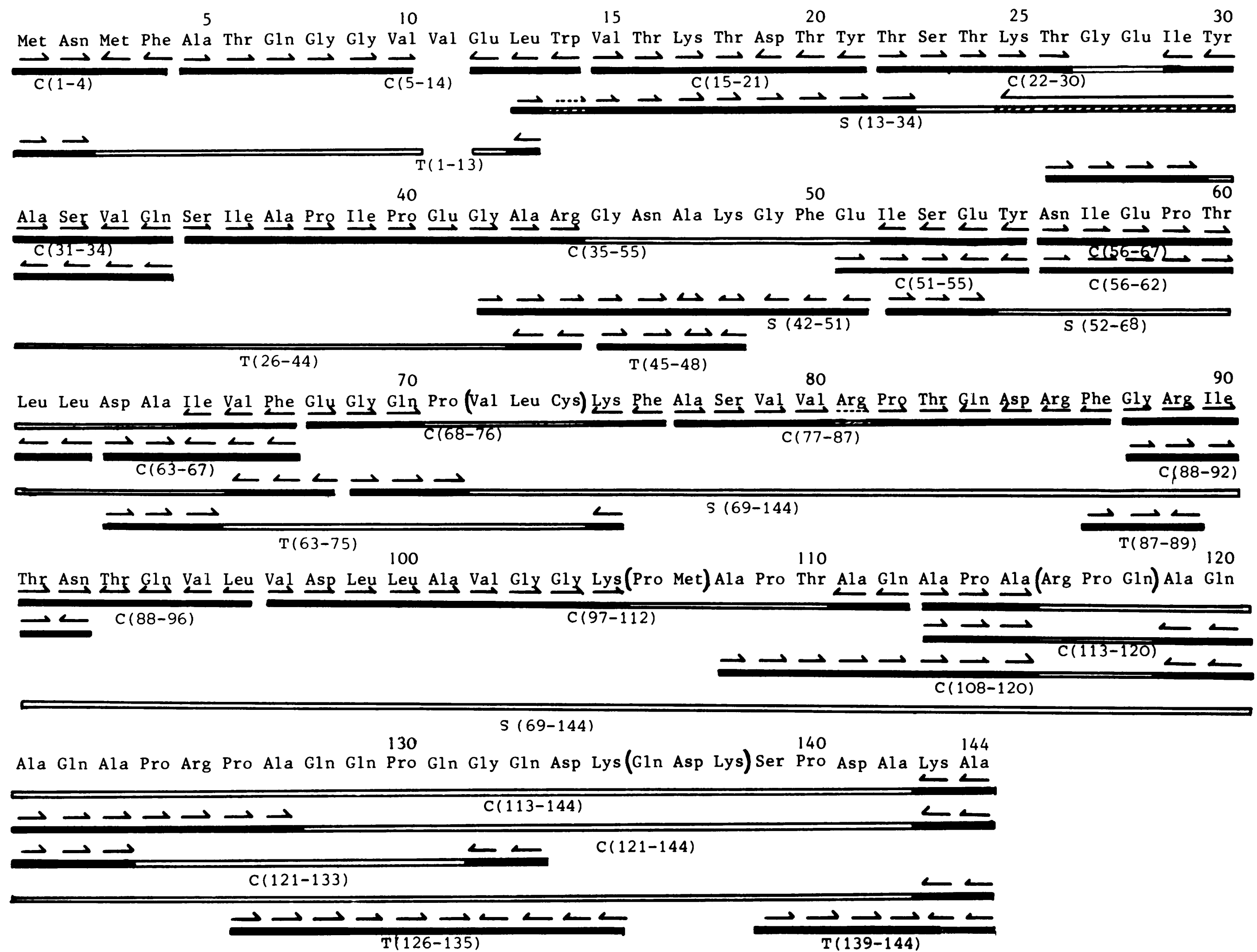

Fig. 2. The partial amino-acid sequence of DNA-binding protein of phage Pf1. ⇀: shows the sequence determined by the Edman degradation and aminopeptidase M digestion. ↼: shows the sequence determined by carboxypeptidase digestion. ( ): this sequence was not determined. The order of amino acids was obtained from the DNA sequence. The filled rectangles represent the region of established sequence; the hatched rectangles represent the regions of suggested sequence; the open rectangles represent the regions of known composition. The abbreviations used for peptides are given in Materials and methods.

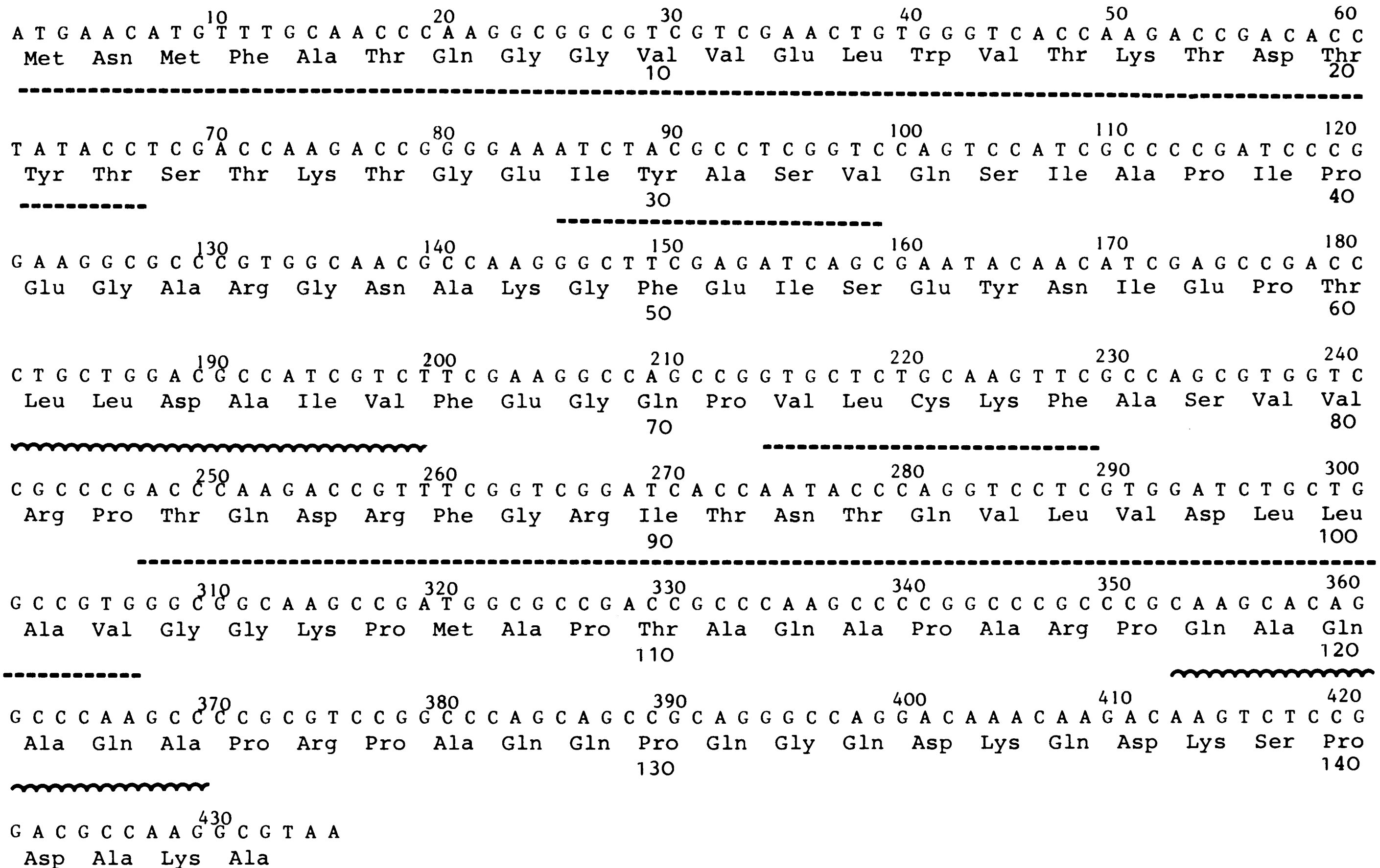

Fig. 3. The nucleotide sequence of the gene and the corresponding amino-acid sequence of the DNA-binding protein of phage Pf1. -----: shows the predicted β-sheet structure. ~~~~ : shows the predicted α-helix structure.

residue by the sequences or compositions of peptides C(1 – 4), C(5 – 14), as well as T(1 – 13). Peptide S(13 – 34) overlaps with C(5 – 14), C(15 – 21), C(22 – 30), and C(31 – 34). Peptide T(26 – 44) overlaps with C(22 – 30), C(31 – 34), C(35 – 55), and S(42 – 51). Peptide S(52 – 68) overlaps with C(35 – 55), C(51 – 55), C(56 – 67), C(56 – 62), C(63 – 67), C(68 – 76), and T(63 – 75). By these overlaps the N-terminal sequence of the protein was determined to Phe^{76} except for the sequence of (^{72}Val,Leu,Cys^{74}).

The tripeptide T(87 – 89) connects C(77 – 87) with C(88 – 96) or C(88 – 92), which provides the sequence between 77 to 96.

Because peptides C(121 – 144), C(113 – 144), and T(139 – 144) have the common C-terminal sequence and share the C-terminal sequence of the protein, those three peptides locate at the C-terminal end of the protein. Peptide C(113 – 144) overlaps with C(108 – 120), which overlaps with C(97 – 112). The sequence from Val^{97} to the C-terminus was therefore established with the exception of the following detailed sequences; (^{106}Pro,Met^{107}), (^{116}Arg,Pro,Gln^{118}) and (^{136}Gln,Asp,Lys^{138}).

Thus, we established the sequence of the three peptide fragments, (N-terminal to 76), (77 to 96), and (97 to C-terminus). The following data further confirm the connection of these three fragments: (1) the N-terminal sequence of peptide S(69 – 144) overlaps with that of C(68 – 76); (2) the total amino-acid composition is in good agreement with the sum of these three fragments (Table I); and (3) the amino-acid composition of peptide S(69 – 144) fits with the amino-acid composition of sequence 68 to C-terminus of the protein.

The amino-acid sequence derived from the DNA sequence data is in complete agreement with the protein sequence established here, with the following discrepancy. From the DNA sequence $Val^{10}Val^{11}$ is predicted, but amino-acid analysis of peptides C(5 – 14), T(1 – 14) indicates that there is only one Val even after extended hydrolysis times. As the Val – Val bond resists acid hydrolysis, the possibility of the second valine cannot be excluded. It should be noted that the $Gln^{34}Ser^{35}$ bond was cleaved by both chymotrypsin and *S. aureus* V8 protease. Because peptide C(31 – 34), which was obtained by chymotrypsin digestion, has Gln as a C-terminus and because of the specificity of chymotrypsin, we would conclude that this Gln was partly deamidated during preparation of the protein or *S. aureus* V8 digestion.

Sequence of the gene coding for the DNA-binding protein

The sequence of the gene coding for the DNA-binding protein has been obtained in the course of the determination of the complete nucleotide sequence of the DNA of bacteriophage Pf1. Full details of these experiments will be described elsewhere in due course.

The consensus nucleotide sequence obtained through the random sequencing approach described in Materials and methods was translated in all phases (Staden, 1977) and the gene coding for the DNA-binding protein identified by scanning the translated sequence for amino-acid sequences corresponding to the N-terminal and C-terminal peptides, C(1 – 4) and T(139 – 144), described in Figure 2. The sequence of 435 nucleotides comprising this gene is given in Figure 3, together with that of the corresponding protein

derived from it. This consensus sequence is derived from the sequences of a total of 16 separate random clones, representing complete sequence data from both strands of the replicative form I (RFI)-DNA.

Discussion

The DNA-binding protein of phage Pf1 binds specifically to single-stranded viral DNA. The complex is seen by electron microscopy to be a long flexible filament similar to the DNA-gene 5 protein complex of fd (Gray *et al.*, 1982). The primary structure of the DNA-binding protein of phage Pf1, determined in parallel and independently from the nucleotide sequence of the gene and the amino-acid sequence of the protein, is shown in Figure 3. By protein sequencing methods, the sequences of the following parts of the molecule were not determined: (72 – 74), (106 – 107), (116 – 118), and (136 – 138). However, the amino-acid compositions of these segments are consistent with the amino-acid sequence deduced from the DNA sequence. The following three points should be noted. (1) By Edman degradation of the protein, we found terminal redundancy namely, the N-terminal mainly starts at methionine but it also occasionally starts from Asn^2. Such a N-terminal redundancy was found in thioredoxin from spinach (Schürmann *et al.*, 1981). (2) The DNA sequence shows that there is a Val – Val bond at positions 10 and 11, but by protein sequencing we could not obtain this dipeptide sequence. As $Val^{79}-Val^{80}$ was sequenced without any difficulty, it appears that the $Val^{10}-Val^{11}$ bond is unusually resistant to acid hydrolysis. (3) The DNA sequence as well as the amino-acid sequence of the chymotryptic peptide C(31 – 34) show that the residue 34 is glutamine. On the other hand, *S. aureus* V8 protease cleaved the peptide bond between 34 and 35 suggesting that the C-terminus is Glu. This cleavage must have occurred at Gln^{34} which had been deamidated during the purification or digestion of the protein. The glutamine of the Gln – Ser bond has been found to be deamidated easily in the case of thioredoxin from *Corynebacterium nephridii* (Meng and Hogenkamp, 1981).

We have predicted the secondary structure of the protein (Figure 3) from its primary sequence by the method of Chou and Fasman (1974). Approximately 40% of the protein is in β-sheet structure, and only 8% exists in a helical conformation. Although there is no apparent sequence homology between the DNA-binding protein of Pf1 and the gene 5 protein of phage fd, the fd protein also has a high content of β-sheet structure (McPherson *et al.*, 1979). Spectroscopic studies on the phage fd gene 5 protein-DNA complex indicate that tyrosine residues are involved in the DNA-protein interaction (Coleman and Oakley, 1980). Fluorescence studies on the Pf1 complex do not provide any evidence on the role of tyrosine in this interaction, but show that the tryptophan is not involved (K.O. Greulich and G.G. Kneale, unpublished data). There are three tyrosine and five phenylalanine residues in the protein. All except Phe^{87} are located in the N-terminal half of the protein. It is known from a study of model peptides that tyrosine stacks preferentially with single-stranded DNA and that this interaction is enhanced by the presence of basic residues nearby (Helene and Maurizot, 1981). From the predicted secondary structure (Figure 3), we suggest that Tyr^{21} may be important for DNA-binding activity; if the area containing Tyr^{21} is in β-sheet structure, two lysines face to the same side of the β-sheet as the tyrosine, and may enhance its intercalation into DNA.

Materials and methods

Isolation of the protein

P. aeruginosa (strain K) was grown to 2 x 10^8 cells/ml and infected at a multiplicity of 10 with bacteriophage Pf1. The cells were harvested after a further 3 h and treated with lysozyme/EDTA, and the nucleoprotein complex purified from cell lysates as described by Kneale and Marvin (1982). Separation of the protein from the DNA was by isopycnic centrifugation in CsCl gradients, followed by dialysis against 10 mM Tris-HCl, pH 7.5 (Kneale and Marvin, 1982). For most experiments, however, the complex was digested with 200 units/ml of micrococcal nuclease (Worthington Biochemicals) in the presence of 5 mM $CaCl_2$ for 1 h at 22°C. This procedure kept the protein more soluble and facilitated its subsequent handling. The presence of oligonucleotides in the complex appears not to affect the amino-acid analysis and the amino-acid composition was the same as that found for the protein alone.

Amino-acid analysis

Purified protein was dialysed against 0.1 M pyridine/collidine acetate buffer, pH 8.2. A sample of the protein (5 μg) was hydrolysed with 300 μl of 2:1 (by volume) mixture of concentrated HCl and trifluoroacetic acid containing 0.005% of phenol at 166°C for 25 min or 50 min in evacuated sealed tubes (Tsugita and Scheffler, 1982). The amino-acid analysis was made with an amino-acid analyser (Durrum, D500, calibrated for 2.2 nmol full scale deflection on the recorder). Peptides (0.1 – 1 nmol) were hydrolysed and analysed under the same conditions.

Chymotryptic digestion and fractionation of the digest

Pf1-DNA binding protein (100 nmol) was digested in 0.1 M pyridine/collidine acetate buffer pH 8.2 with 45 μg of α-chymotrypsin (Worthington Biochemicals) at 37°C for 24 h. The digests were fractionated on a column (0.6 cm x 15 cm) of SP-Sephadex C25. Elution was carried out with a gradient of increasing pH and ionic strength established with a 9-chamber apparatus (Canfield, 1963). The composition of elution buffers and the mode of elution are given in the legend to Figure 1a.

S. aureus V8 protease digestion and fractionation of the digest

The protein (120 nmol) was digested with 20 μg of *S. aureus* V8 protease (Miles Laboratories) in 0.1 M pyridine/collidine acetate buffer pH 8.2 at 37°C for 24 h. The extent of digestion was monitored by SDS-gel electrophoresis, which showed that the digests consisted of a large peptide (mol. wt. 8000) and several small peptides. In order to separate the large peptide from the others, the digests were fractionated on a column (1 cm x 60 cm) of Biogel P30 (Bio-Rad Laboratory) equilibrated with 70% formic acid.

Trypsin digestion and fractionation of the digest

The protein (120 nmol) was digested with 40 μg of L-1-tosylamido-2-phenylethyl chloromethyl ketone (TPCK)-treated trypsin (Worthington Biochemicals) in 0.1 M pyridine/collidine acetate buffer pH 8.2 at 37°C for 24 h. After digestion, fractionation was made on a column (0.6 cm x 15 cm) of SP-Sephadex C25. Elution was carried out as for the fractionation of the α-chymotrypsin digest.

Sequence analysis of peptide

The C-terminal sequence was determined by carboxypeptidase A (treated with diisopropylfluorophosphate, Worthington Biochemicals), B (Sigma), and P (Takara Shuzo) digestion. Peptides or proteins (1 nmol) were dissolved in 0.1 M pyridine/collidine acetate buffer pH 8.2 for carboxypeptidase A and carboxypeptidase B and 0.1 M pyridine formate buffer pH 2.5 for carboxypeptidase P. The reaction was carried out in 100 μl of buffer and 5 μg of carboxypeptidase(s) at 37°C for 2 – 6 h. The digests were dried and analysed directly by the amino-acid analyser.

Edman degradation was carried out on 5 – 10 nmol of peptides or proteins for the N-terminal sequence, according to the procedure of Isobe *et al.* (1978) except that the cyclization step was carried out in 12 N HCl for 5 min at room temperature (Tarr, 1977). The resulting phenylthiohydantoin amino acids were analysed by h.p.l.c. and by t.l.c. on a polyamide sheet (Schürman *et al.*, 1981). The amino-terminal sequences of peptides were also determined by aminopeptidase M (Boehringer, Mannheim) digestion at 37°C in a 0.1 M pyridine/collidine acetate buffer pH 8.2 for 10 min – 6 h.

Nomenclature of peptides

The peptides produced by α-chymotrypsin digestion, *S. aureus* V8 protease digestion, and trypsin digestion are designated "C", "S", and "T", respectively. Numbers after these letters indicate their order in the sequence, starting from the N-terminus of the protein.

Isolation of RFI-DNA for sequence analysis

A 1 l culture of *P. aeruginosa* in tryptone medium (1.6% Bacto-tryptone, 1% yeast extract, 1% NaCl) was grown with vigorous agitation to a cell densi-

ty of 2 x 10^8 cells/ml. Stock bacteriophage Pf1 was added to a multiplicity of infection of 10 and the incubation continued for a further 6 h. Chloramphenicol (40 μg/ml culture) was then added and the cells harvested by centrifugation after a further 30 – 40 min. RFI-DNA was isolated and purified exactly as previously described (Hill and Peterson, 1980), except that the density of the caesium chloride gradient was adjusted to 1.64 g/ml.

Preparation of clones

RFI-DNA was degraded randomly by ultrasonic shearing, the termini of the degraded DNA molecules repaired with phage T4 DNA polymerase and the products separated on the basis of size by electrophoresis on an agarose gel as described by Deininger (to be published). Fragments of DNA of chain length 400 – 600 nucleotides, as judged by comparison with suitable markers, were recovered from the gel by electro-elution, precipitated with ethanol and inserted through blunt-ended ligation with phage T4 ligase into the *Hinc*II site of the phage M13mp7 vector described by Messing *et al.* (1981). The hybrid phage DNA was used to transfect competent cells of *E. coli* JM101 in the presence of isopropyl-thiogalactoside and 5-bromo-4-chloro-indolyl-β-D-galactoside (Messing *et al.*, 1977). These experiments were carried out under Category O conditions. "Clear" plaques were picked with a toothpick and used to infect 1 ml cultures of *E. coli* JM101 and the single-stranded DNA from each clone prepared as described by Sanger *et al.* (1980).

DNA sequencing

The sequences of the random, cloned inserts of RFI-DNA from phage Pf1 were determined by chain extension with DNA polymerase in the presence of dideoxynucleoside triphosphates (Sanger *et al.*, 1977) using a chemically-synthesized heptadecanucleotide (Duckworth *et al.*, 1981) as primer. The sequence data thus obtained were assembled and analysed by computer (Staden, 1977, 1980).

Acknowledgements

D.F.H. and G.B.P. thank Dr. F. Sanger for his hospitality to them while on leave from New Zealand. K.M., G.G.K., and A.T. thank Mr. J.-J. Scheffler for performing the amino-acid analysis, and Dr. D.A. Marvin for his encouragement in this work. D.F.H. was supported by the Medical Research Council of New Zealand. G.B.P. thanks the Royal Society and Nuffield Foundation for a Commonwealth Bursary. N.J.S. thanks the Australian National University for the award of a Postbachelor Travelling Scholarship.

References

Canfield,R.E. (1963) *J. Biol. Chem.*, **238**, 2691-2697.
Chou,P.Y., and Fasman,G.D. (1974) *Biochemistry (Wash.)*, **13**, 222-245.
Coleman,J.E., and Oakley,J.L. (1980) *CRC Crit. Rev. Biochem.*, **7**, 247-289.
Denhardt,D.T., Dressler,D., and Ray,D.S. eds. (1978) *The Single-Stranded DNA Phages*, published by Cold Spring Harbor Laboratory Press, NY.
Duckworth,M.L., Gait,M.J., Goelet,P., Hong,G.F., Singh,M., and Titmas, R.C. (1981) *Nucleic Acids Res.*, **9**, 1691-1706.
Gray,C.W., Kneale,G.G., Leonard,K.R., Siegrist,H., and Marvin,D.A. (1982) *Virology*, **116**, 40-52.
Goodwin,M.T.W., and Morton,R.A. (1946) *Biochem. J.*, **40**, 628-632.
Helene,C., and Maurizot,J.-C. (1981) *CRC Crit. Rev. Biochem.*, **10**, 213-258.
Hill,D.F., and Petersen,G.B. (1980) *J. Virol.*, **34**, 40-50.
Hirs,C.H.W. (1956) *J. Biol. Chem.*, **219**, 611-621.
Isobe,T., Yanagida,M., Boosman,A., and Tsugita,A. (1978) *J. Mol. Biol.*, **125**, 339-356.
Kneale,G.G., Freeman,R., and Marvin,D.A. (1982) *J. Mol. Biol.*, in press.
Kneale,G.G., and Marvin,D.A. (1982) *Virology*, **116**, 53-60.
Marvin,D.A., and Hohn,B. (1969) *Bacteriol. Rev.*, **33**, 172-209.
Marvin,D.A., and Wachtel,E.J. (1976) *Philos. Trans. R. Soc. Lond., B*, **276**, 81-98.
McPherson,A., Jurnak,F., Wang,A., Molineux,I., and Rich,A. (1979) *J. Supramol. Struct.*, **10**, 457-465.
Meng,M., and Hogenkamp,H.P.C. (1981) *J. Biol. Chem.*, **256**, 9174-9182.
Messing,J., Gronnenborn,B., Muller-Hill,B., and Hofschneider,P.H. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 3642-3646.
Messing,J., Crea,R., and Seeburg,P.H. (1981) *Nucleic Acids Res.*, **9**, 309-321.
Nakashima,Y., Dunker,A.K., Marvin,D.A., and Konigsberg,W. (1974) *FEBS Lett.*, **40**, 290-292.
Penke,B., Ferenczi,R., and Kovacs,K. (1974) *Anal. Biochem.*, **60**, 45-50.
Pratt,D., Laws,P., and Griffiths,J. (1974) *J. Mol. Biol.*, **82**, 425-439.
Ray,D.S. (1969) *J. Mol. Biol.*, **43**, 631.
Sanger,F., Nicklen,S., and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H., and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161-178.
Schürmann,P., Maeda,K., and Tsugita,A. (1981) *Eur. J. Biochem.*, **116**, 37-45.
Smilowitz,H., Carson,J., and Robbins,P.W. (1972) *J. Supramol. Struct.*, **1**, 8-18.
Staden,R. (1977) *Nucleic Acids Res.*, **4**, 4037-4051.
Staden,R. (1980) *Nucleic Acids Res.*, **8**, 3673-3694.
Tarr,E.G. (1977) *Methods Enzymol.*, **47E**, 335-357.
Tsugita,A., and Scheffler,J.-J. (1982) *Proc. Jap. Acad. Ser. B*, **58**, 1-4.