The EMBO Journal vol.13 no.18 pp.4353–4360, 1994

A new protein domain for binding to DNA through the minor groove

Raimundo Freire, Margarita Salas[1] and José M.Hermoso

Centro de Biología Molecular 'Severo Ochoa' (CSIC-UAM), Universidad Autónoma, Cantoblanco, 28049 Madrid, Spain

[1]Corresponding author

Communicated by M.Salas

Protein p6 of the *Bacillus subtilis* phage φ29 binds with low sequence specificity to DNA through the minor groove, forming a multimeric nucleoprotein complex that activates the initiation of φ29 DNA replication. Deletion analysis suggested that the N-terminal part of protein p6, predicted to form an amphipathic α-helix, is involved in DNA binding. We have constructed site-directed mutants at the polar side of the putative α-helix. DNA binding and activation of initiation of φ29 DNA replication were impaired in most of the mutant proteins obtained. A 19 amino acid peptide comprising the N-terminus of protein p6 interacted with a DNA fragment containing high-affinity signals for protein p6 binding with ~50-fold higher affinity than the peptide corresponding to an inactive mutant. Both wild-type peptide and protein p6 recognized the same sequences in this DNA fragment. This result, together with distamycin competition experiments, suggested that the wild-type peptide also binds to DNA through the minor groove. In addition, CD spectra of the wild-type peptide showed an increase in the α-helical content when bound to DNA. All these results indicate that an α-helical structure located in the N-terminal region of protein p6 is involved in DNA binding through the minor groove.

Key words: α-helix/DNA minor groove binding/DNA replication/low sequence specificity protein/phage φ29

Introduction

Structural studies of DNA binding proteins have allowed the characterization of many protein motifs involved in DNA recognition. Most of them recognize DNA-specific sequences through the major groove mainly by hydrogen bonding with bases and electrostatic interactions with phosphate groups. On the other hand, some DNA binding motifs recognize a DNA structure rather than a specific sequence (reviewed in Churchill and Travers, 1991); a common property of these motifs is their interaction with the DNA through the minor groove. The amount of information available for DNA binding motifs in sequence-specific DNA binding proteins (reviewed in Pabo and Sauer, 1992) is much smaller for proteins of low sequence specificity, despite the fact that some structural motifs have been described. Thus, an exposed loop has been demonstrated by X-ray diffraction of DNase I complexed with DNA (Suck *et al.*, 1988), an antiparallel β-sheet has been proposed to interact with DNA for HU (White *et al.*, 1989), and a β-turn has been suggested for histones H1 and H2B (Suzuki, 1989), HMG-1 (Reeves and Nissen, 1990) and DAT1 protein (Reardon *et al.*, 1993). The TATA binding protein also interacts with the DNA minor groove, although recognizing a specific sequence; the structure of the protein−DNA complex has revealed that the binding domain consists of β-sheets and β-loop protrusions (J.L.Kim *et al.*, 1993; Y.Kim *et al.*, 1993).

Protein p6 from *Bacillus subtilis* phage φ29 is required *in vivo* for viral DNA replication (Carrascosa *et al.*, 1976; Mellado *et al.*, 1980), and activates *in vitro* the initiation of φ29 DNA replication (Blanco *et al.*, 1986), as well as the transition to elongation (Blanco *et al.*, 1988). Phage φ29 replicates its linear, double-stranded (ds) DNA genome by a protein-priming mechanism in which the initiation complex, formed by the covalent linkage of the 5′ terminal nucleotide (dAMP) to the terminal protein (TP), catalyzed by the phage DNA polymerase is further elongated by the same DNA polymerase by a strand-displacement mechanism (reviewed in Salas, 1991). Protein p6 activates the initiation of φ29 DNA replication by forming a multimeric nucleoprotein complex at the replication origins (Prieto *et al.*, 1988; Serrano *et al.*, 1989) located at the genome ends. The path of the DNA in the complex has been characterized as a right-handed DNA superhelix wrapped around a protein p6 multimeric core, and the parameters that define the superhelix have been determined (Serrano *et al.*, 1993). Protein p6 has 103 amino acids (Murray and Rabinowitz, 1982), forms dimers in solution (Pastrana *et al.*, 1985) and binds to DNA through the minor groove (Serrano *et al.*, 1990). It is one of the most abundant proteins in φ29-infected *B.subtilis* cells, with ~3×10^6 copies per cell in the late stages of infection (Serrano *et al.*, 1994), and it has been shown to interact *in vitro* with the whole viral genome suggesting that it may have a more general role in genome packaging and organization (Gutiérrez *et al.*, 1994). The fact that the protein p6−DNA complex can be formed on DNA regions of >1 kb (Gutiérrez *et al.*, 1994) indicates that protein p6 must be able to interact with a large variety of sequences, although it recognizes preferentially those showing a tendency to bend with a defined periodicity (Serrano *et al.*, 1993). Protein p6 thus corresponds to the class of low sequence specificity ds DNA binding proteins that, like proteins such as HU and the HMG-1 family, can act as architectural elements in various DNA transactions (Grosschedl *et al.*, 1994).

Previous studies with protein p6 N-terminal deletion mutants suggested that the N-terminus of protein p6 is involved in DNA binding (Otero *et al.*, 1990). Since this region has been shown to have a strong tendency to form

© Oxford University Press

an amphipathic α-helix, the positive face of the helix was an attractive candidate for binding to the phosphate backbone of the DNA. In this paper we have tested this hypothesis by two experimental approaches: (i) using protein p6 mutants directed to the N-terminus, and (ii) with synthetic peptides corresponding to this region. The results obtained allow us to propose that the protein p6 N-terminal region adopts an α-helical conformation that is involved in DNA interaction, and constitutes a novel protein motif for minor groove DNA binding.

Results

Site-directed mutagenesis at the protein p6 N-terminal region

Protein p6 mutants in which five and 13 N-terminal amino acids were deleted showed impaired and undetectable DNA binding activities, respectively (Otero *et al.*, 1990), suggesting that this region was involved in DNA binding. The N-terminus of protein p6 is rich in basic amino acids, and secondary structure predictions (Chou and Fasman, 1978; Garnier *et al.*, 1978) showed a high tendency to form an α-helix. This putative α-helix is predicted to have a strong amphipathic character with a hydrophobic moment of 0.48 for the first 18 residues (Saier and McCaldon, 1988). To investigate further the involvement of this region in DNA binding, we designed site-directed mutants in which positively charged or polar residues of the hydrophilic side of the predicted helix were substituted by either alanine, to preserve protein secondary structure (Cunningham and Wells, 1989), or lysine, to increase the positive charge. Figure 1 shows the 19 N-terminal amino acid sequence arranged as an α-helix, as well as the amino acid substitutions introduced. An M13-derivative phage containing wild-type gene 6 was constructed and site-directed mutagenesis was performed. The mutated genes were cloned into an expression vector under the control of the bacteriophage λ P_L promoter. Protein p6 mutants were overproduced and purified up to at least 90% homogeneity. The standard procedure used for wild-type protein p6 purification (Pastrana *et al.*, 1985) had to be modified since most of the mutant proteins failed to bind to a phosphocellulose column. This different behavior was in agreement with amino acid changes affecting the DNA binding domain.

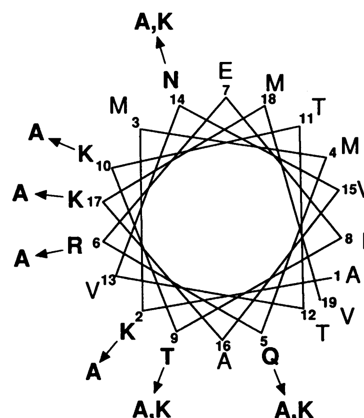

Fig. 1. Protein p6 N-terminus. α-Helical projection of the N-terminal 19 amino acids. The projection down the helix axis displays the location of the side-chain α-carbons along the peptide backbone. The amino acid substitutions performed are indicated by arrows.

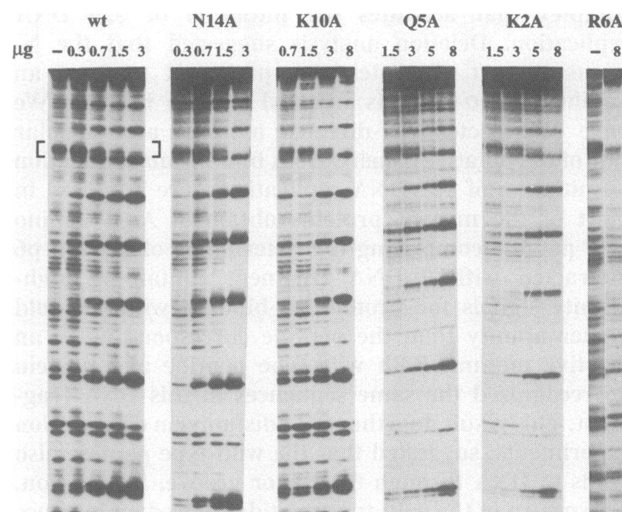

Fig. 2. DNase I footprinting of wild-type and mutant p6 proteins. DNase I footprints of wild-type and the indicated protein p6 mutants were performed with the 273 bp long ϕ29 DNA *Hin*dIII right terminal fragment. The protein amount used in each case is indicated. Square brackets show the region scanned to quantify complex formation. This region was chosen by the preferential DNase I cleavage in the absence of protein and the good protection obtained by wild-type protein p6. The same region was scanned in all the mutant proteins.

DNA binding affinity of protein p6 mutants

DNA gel retardation studies provided the first indication that some of these mutant proteins had altered DNA binding affinities. Thus, a higher amount of protein p6K2A mutant was required to obtain the same DNA retardation as that with the wild-type protein, and no retardation was achieved with the p6R6A mutant protein (results not shown). The formation of the multimeric protein p6–DNA complex gives rise, by DNase I footprinting, to a characteristic pattern of hypersensitive bands regularly spaced every 24 bp flanked by protected regions (Prieto *et al.*, 1988). Hence, DNase I footprinting was the method of choice for DNA binding studies, since it not only allows the estimation of binding affinities but also the assessment of the structure of the nucleoprotein complex formed. Figure 2 shows the results obtained with representative mutants. The complexes formed by all the mutant proteins, except p6R6A, gave the same footprint pattern as the wild-type protein; therefore the mutations did not affect the structure of the complexes. However, the protein concentration required to form the complex was different for each mutant. The p6N14A mutant protein showed the same affinity for DNA as the wild-type protein. The p6K10A, p6Q5A and p6K2A mutant proteins showed a decreasing capacity for complex formation, deduced from the higher amount of protein needed to yield the same digestion pattern. On the other hand, p6R6A mutant protein failed to bind to DNA at a protein concentration at least 10-fold higher than the wild-type protein. The DNA binding affinity was estimated by densitometry of the protected region indicated by square brackets in Figure 2, using the undigested fragment as an internal standard. The values obtained were plotted against protein concentration and

Table I. Summary of DNA binding affinities and initiation of ϕ29 DNA replication activities of protein p6 mutants

Protein p6	DNA binding[a]	Initiation activity protein dose for half-maximal stimulation[b]	Maximal stimulation[c]
wt	1.0	1.0	100
N14A	1.1	0.9	98
N14K	1.6	1.9	72
T9K	3.2	2.1	105
T9A	3.5	4.8	107
K10A	4.2	5.0	39
Q5A	4.5	3.6	98
K2A	5.2	8.2	79
Q5K	6.6	4.7	76
K17A	7.4	13.8	98
R6A	>14	>32	<4

[a]Protein amount necessary for 50% complex formation normalized to that of the wild-type protein (0.85 μg).
[b]Protein amount necessary for half-maximal stimulation in the initiation reaction normalized to wild-type protein (0.25 μg).
[c]Maximal stimulation value obtained in the initiation reaction, relative to that of the wild-type protein.

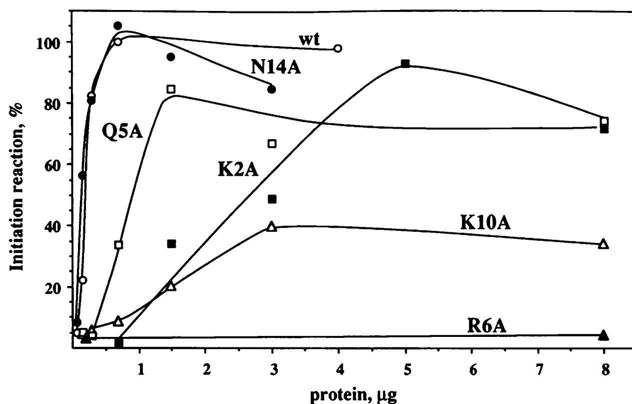

Fig. 3. Activation of initiation of ϕ29 DNA replication by protein p6. The wild-type and indicated protein p6 mutants were assayed, at different protein concentrations, for activation of ϕ29 DNA initiation of replication, shown as the percentage of the maximal stimulation obtained with the wild-type protein p6.

the DNA binding affinity was calculated as the protein concentration required to obtain 50% of full protection. Table I shows the mean values from two different experiments with all the mutant proteins, normalized for those obtained for the wild-type protein. The effective binding constant of wild-type protein p6 to ϕ29 DNA has been estimated to be ~10^5 M^{-1} (Gutiérrez *et al.*, 1994).

Activity of protein p6 mutants in the initiation of replication

The functionality of the protein p6 mutants was also tested by their ability to stimulate the initiation of ϕ29 DNA replication in an *in vitro* assay with purified proteins (Pastrana *et al.*, 1985). Figure 3 shows the results obtained with the mutant proteins shown above (see Figure 2). While p6N14A mutant protein exhibited essentially the same activity as the wild-type protein, the rest of the mutant proteins showed to different extents a decrease in the activation of the initiation reaction. As expected from the DNA binding results, p6R6A mutant protein did not

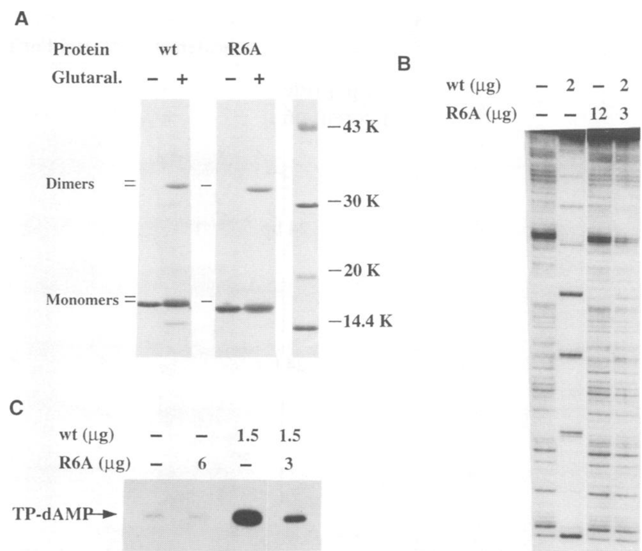

Fig. 4. (**A**) Dimer formation by p6 wild-type and p6R6A mutant proteins. Proteins, 3 μg each, were crosslinked with glutaraldehyde and run in SDS–PAGE along with untreated samples and molecular weight markers. (**B**) Interference assay of p6 wild-type and p6R6A mutant proteins. The indicated amounts of p6 wild-type and p6R6A mutant proteins were used, either separately or together, in DNase I footprinting with the ϕ29 DNA *Hin*dIII right terminal fragment. (**C**) The same as in (B), except that they were assayed *in vitro* for initiation of ϕ29 DNA replication. The mobility of the initiation complex (TP–dAMP) is shown by an arrow.

show any activation even at a concentration 10-fold higher than that required for the wild-type protein to reach maximal activation. The other mutant proteins, with the exception of p6K10A, reached a maximal activation of at least 70% of that of the wild-type protein, although at a higher protein concentration. Table I summarizes the behavior of all the mutant proteins, and shows the maximal stimulation achieved and the protein concentrations needed to obtain half-maximal stimulation. The mean values from at least three different experiments were normalized for those obtained for the wild-type protein. Mutant proteins were also tested in their capacity to stimulate the elongation of the initiation complex (TP–dAMP) (Blanco *et al.*, 1988) in a truncated elongation assay (Blasco *et al.*, 1993), where they showed the same behavior as in the initiation assay (results not shown).

Dimer formation of protein p6 mutants

It has been reported that protein p6 forms dimers in solution (Pastrana *et al.*, 1985), and most probably binds to DNA as a dimer (Serrano *et al.*, 1990). The impaired DNA binding of the protein p6 mutants described above could be explained if mutations were affecting either the general folding of the protein or the dimerization domain, rather than the DNA binding domain itself. To rule out the possibility that dimerization was affected in the mutant proteins, we set up an assay in which protein p6 monomers were crosslinked with glutaraldehyde, and protein p6 dimers were detected by SDS–PAGE. Figure 4A shows that the p6R6A mutant protein forms dimers to the same extent as the wild-type protein. A similar behavior was observed for all the other mutant proteins (not shown). These results indicate that the mutations do not affect the dimerization domain and suggest that there are no major

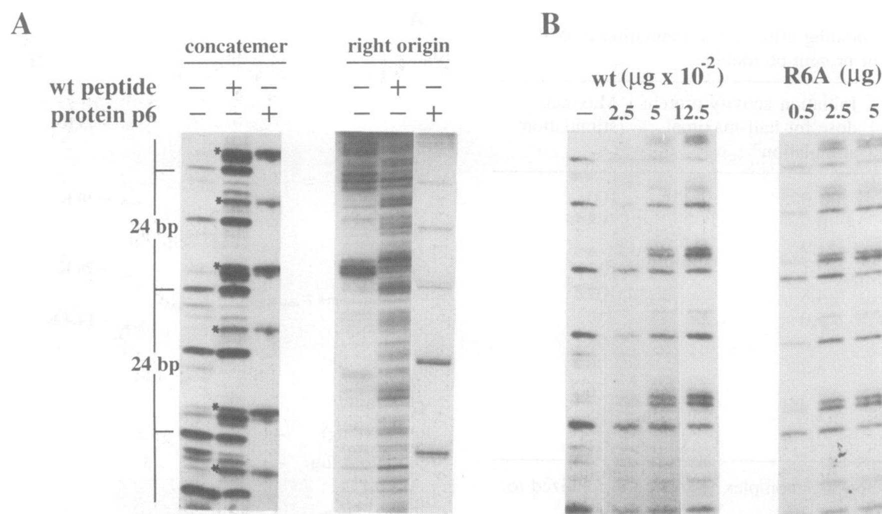

Fig. 5. DNase I footprinting of protein p6 N-terminal peptides. (**A**) DNase I footprint of protein p6 and wild-type peptide with a DNA fragment (concatemer) containing 10 tandem repeats of a 24 bp sequence with high affinity for protein p6 binding (Serrano *et al.*, 1993) or the ϕ29 *Hin*dIII right terminal fragment (right origin). The amounts used with the concatemer and the right origin were for the peptide, 1 and 0.5 μg, and for protein p6, 6 and 2 μg, respectively. (**B**) DNase I footprint of the indicated amounts of either wild-type or R6A peptides with the concatemer fragment.

alterations in the general folding of the protein. Furthermore, crosslinking studies suggested the formation of mixed dimers of wild-type and R6A mutant p6 proteins (not shown). Figure 4B shows DNase I footprints of wild-type together with p6R6A mutant proteins, demonstrating that p6R6A strongly interfered with wild-type protein p6 in DNA binding. This can be explained by the formation of inactive heterodimers, in agreement with the indication that the R6A mutation does not affect dimerization. Figure 4C shows that the same effect is observed in the *in vitro* assay for initiation of ϕ29 DNA replication: the p6R6A mutant protein, that did not activate the formation of the initiation complex (TP−dAMP; see also Figure 3), strongly impaired the activation of TP−dAMP formation by the wild-type protein p6 when they were added together.

DNA binding of N-terminal peptides

The DNA binding capacity of the N-terminal region of protein p6 was investigated further using synthetic oligopeptides of 19 amino acids with the N-terminal sequence of the wild-type protein p6 (wild-type peptide) and that of the p6R6A mutant protein (R6A peptide). Figure 5A shows the DNase I footprint of protein p6 and the wild-type peptide with a DNA fragment containing a protein p6 high-affinity binding sequence, 24 bp long, repeated in tandem (Serrano *et al.*, 1993), hereafter the concatemer fragment. The hypersensitive sites induced by protein p6 were also induced by the peptide (indicated by asterisks), suggesting that the peptide recognizes the same sites in the DNA fragment; however, the formation of the protein p6−DNA complex gives rise to protections to DNase I digestion not observed with the wild-type peptide, indicating that the peptide cannot form such a multimeric complex. The periodicity of the digestion pattern obtained with the wild-type peptide reflects, therefore, the sequence repetition in the DNA fragment. Figure 5A also shows the DNase I footprint of the wild-type peptide, and of protein p6 as control, to the ϕ29 DNA right terminal fragment. The wild-type peptide changed the footprint pattern, mainly inducing hypersensitivities and occasionally a few protections (see also Figure 6). In contrast to the pattern observed with protein p6, the hypersensitivities were not periodic, further indicating that the wild-type peptide does not form a multimeric complex.

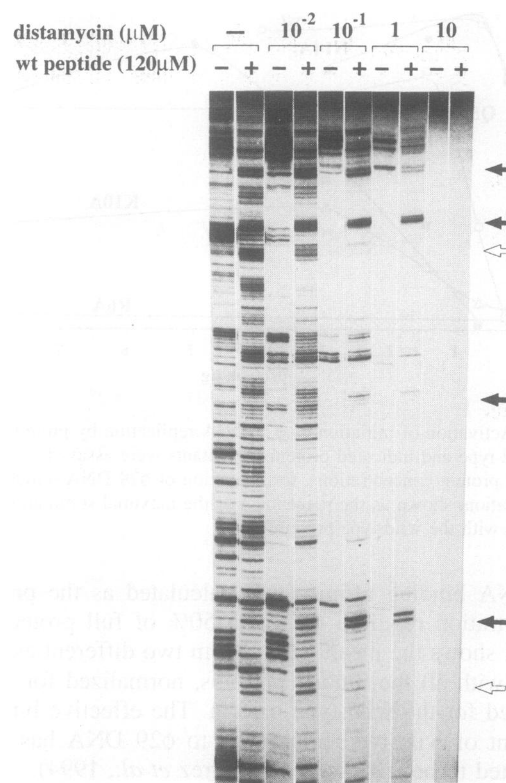

Fig. 6. DNase I footprinting of wild-type peptide and distamycin. DNase I footprinting was performed with the Ø29 DNA *Hin*dIII right terminal fragment, 5 μg (120 μM) of wild-type peptide and the indicated concentrations of distamycin. Filled and empty arrows indicate hypersensitivities induced by the wild-type peptide that remained up to 1 and 10^{-2} μM distamycin, respectively.

Figure 5B shows the binding to the concatemer fragment of different concentrations of wild-type and R6A peptides.

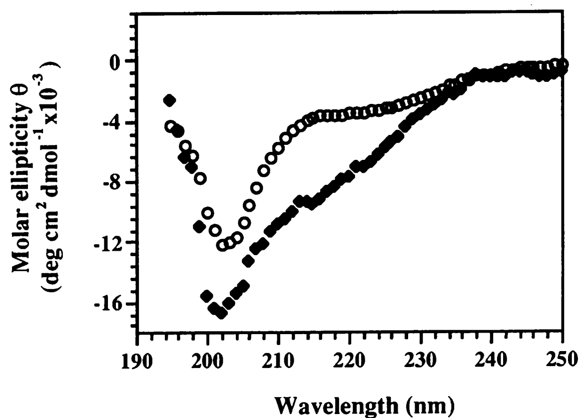

Fig. 7. CD spectra of the wild-type peptide in the absence and presence of DNA. Molar ellipticity (θ) is expressed in °cm^{2}dmol^{-1} at different wavelengths. Spectrum of the wild-type peptide alone (○), or in the presence of DNA after subtraction of the spectrum of the DNA alone (◆).

Although the same DNase I footprint pattern was observed in both cases, the R6A peptide required a concentration ~50-fold higher than that for the wild-type peptide. This result is in agreement with that shown in Figure 2, in which the p6R6A mutant protein was not able to bind DNA with a protein concentration 10-fold higher than that required for the wild-type protein. Furthermore, the R6A peptide bound the ϕ29 DNA right terminal fragment with a 100-fold lower affinity than the wild-type peptide (result not shown).

All these results are in agreement with those obtained with the site-directed protein mutants, supporting the implication of the N-terminal region, and in particular the arginine residue at position 6, in DNA binding.

Competition of wild-type peptide and distamycin for DNA binding

Protein p6 binds to DNA through the minor groove (Serrano *et al.*, 1990). Although the result shown in Figure 5A suggested that the peptide interacted with the concatemer fragment in a way similar to protein p6, to assess whether the wild-type peptide binds to DNA also through the minor groove we challenged the peptide with distamycin, a minor groove DNA binding drug (Coll *et al.*, 1987). Figure 6 shows the DNase I footprint of distamycin to the ϕ29 DNA right terminal fragment, in the absence or presence of the wild-type peptide. Distamycin largely protected the fragment from DNase I digestion, and the protection was total at a concentration of 10 μM. Figure 6 also shows, as a control, the DNase I footprint of the wild-type peptide alone (see also Figure 5A), where induced hypersensitive sites can be clearly observed. However, when the peptide and distamycin were added together, the peptide-induced hypersensitivities gradually decreased as distamycin concentration increased, probably reflecting differences of the peptide affinity for different DNA sites. This result is consistent with a displacement of the DNA-bound peptide by distamycin. As a control, the same experiments were performed with the major groove-binder methyl green (Zimmer and Wahnert, 1986), and no effect on peptide binding was obtained with drug concentrations up to 1 μM (results not shown), despite the fact that distamycin and methyl green have similar DNA binding affinities (Kumar and Muniyappa, 1992). Altogether, these results suggest that the peptide binds to the DNA also through the minor groove.

CD studies of DNA-induced conformational changes of the peptide

Once it was determined that the N-terminal peptide of protein p6 was involved in DNA binding, it was important to study the secondary structure of the peptide in solution and the effect of DNA binding. This was approached by CD, as has been performed for other cases (Zlotnick and Brenner, 1988; Anthony-Cahill *et al.*, 1992; Fisher *et al.*, 1993). The spectrum of the DNA at wavelengths >240 nm was essentially not affected by the presence of the peptide (results not shown), indicating that the structure of the DNA remained unchanged. Therefore, structural changes in the peptide due to the presence of DNA can be directly analyzed by subtracting the spectrum of the DNA alone from that of the DNA together with the peptide. Figure 7 shows the CD spectrum of the peptide alone, in which the ~200 nm minimal peak reflects a mostly random conformation, and the shoulder at ~220 nm suggests a small amount of α-helix. Essentially the same spectrum was obtained with the R6A peptide (not shown); thus, this mutation did not seem to alter the conformation of the wild-type peptide. Figure 7 also shows the CD spectrum of the peptide in the presence of DNA from which the spectrum of the DNA alone was subtracted. It can be observed that the ellipticity at ~220 nm was shifted with respect to that of the peptide alone, indicating that the peptide undergoes a transition to an α-helical structure upon DNA binding.

Discussion

While many protein motifs have been described for sequence-specific recognition through the DNA major groove (Pabo and Sauer, 1992), the available information for non-sequence-specific DNA recognition through the minor groove is scant. Phage ϕ29 protein p6 binds to DNA through the minor groove (Serrano *et al.*, 1990) and it does not recognize a specific nucleotide sequence (Serrano *et al.*, 1989). However, it shows a preferential binding to sequences with defined bendable properties that favor the formation of the multimeric complexes with DNA (Serrano *et al.*, 1993). Therefore, it was interesting to study the structural basis of the binding of protein p6 to DNA.

Protein p6 deletion analysis (Otero *et al.*, 1990) suggested that the N-terminal region was involved in DNA binding. Moreover, secondary structure predictions indicated a strong tendency of this region to adopt an amphipathic α-helical conformation. In this work we have studied, by site-directed mutagenesis, the involvement in DNA binding of amino acid residues of the polar side of the putative α-helix, with the aim of characterizing a protein motif for DNA binding. The DNA binding in most of the mutants studied was impaired to different extents. However, only the p6R6A mutant protein was completely inactive. Therefore, arginine at position 6 seems to establish critical contacts with DNA. The rest of the residues

are not so critical, and the asparagine at position 14 does not seem to be relevant. Mutants in which a positive charge was introduced did not, in general, lead to a higher binding affinity; in fact, a lysine residue at positions 5 or 14 showed lower activities than alanine, though at position 9 the opposite effect was observed. As expected, the DNA affinity was more dependent on the position of the residue changed than on the increase of the positive charge. The possibility that these mutations affected the general folding of the protein or the dimerization domain was ruled out since they showed a similar ability as the wild-type protein to form dimers.

In agreement with previous results (Serrano *et al.*, 1989; Otero *et al.*, 1990), a good correlation between DNA binding affinity and activation of ϕ29 DNA initiation of replication was observed. When assayed for initiation of ϕ29 DNA replication, all the mutant proteins required higher amounts than the wild-type protein to reach maximal stimulation, with the exceptions of p6N14A mutant protein, which behaved as the wild-type, and p6R6A mutant protein, which was inactive. The maximal stimulation values obtained with the mutant proteins were at least 70% of that reached by wild-type protein p6. However, the p6K10A mutant protein leveled off at 40%, despite the fact that it exhibited the same DNase I footprint pattern as the wild-type protein and therefore the structure of the nucleoprotein complex was the same. This result suggests that this residue, besides its role in DNA binding, could have an additional role in the initiation of DNA replication. Activation of ϕ29 DNA replication requires the specific positioning of protein p6 with respect to the ϕ29 replication origin (Serrano *et al.*, 1989), suggesting that the proteins involved in initiation (DNA polymerase and/or TP) could interact directly with protein p6. Though interaction of protein p6 with initiation proteins has not been demonstrated directly, it is conceivable that this interaction could be affected in the p6K10A mutant protein.

An *in vivo* functional assay for protein p6 has been developed, in which protein p6-producing *B.subtilis* cells support DNA replication of a ϕ29 *sus* 6 mutant phage. This assay has been used to analyze the effect of K2A and R6A mutations in ϕ29 DNA replication *in vivo*. In cells producing p6K2A mutant protein, the rate of ϕ29 DNA synthesis was reduced markedly with respect to those producing wild-type protein. The R6A mutation was much more detrimental, and no significant viral DNA synthesis was detected (Bravo *et al.*, 1994). These results are in agreement with those obtained *in vitro*, supporting the conclusion that lysine at position 2 is involved in DNA binding and arginine at position 6 is essential for that activity.

The study of the involvement of the N-terminal region of protein p6 in DNA binding was also approached using a synthetic peptide comprising the 19 N-terminal amino acids of protein p6 and studying its ability to interact with a DNA fragment containing high-affinity binding sites for protein p6 (Serrano *et al.*, 1993). The peptide was able not only to bind to the fragment, but also to recognize the same signals as protein p6. The hypersensitivities induced by the peptide reflect local distortions in the DNA, making it more accessible to DNase I, i.e. widening the DNA minor groove. The same effect has been reported for DNA minor groove binding drugs like distamycin or actinomycin (Fox and Waring, 1984). The competition observed for DNA binding between the peptide and distamycin, together with the fact that the peptide binds to high-affinity sequences in a way similar to that of protein p6, suggest that the peptide binds to the DNA also through the minor groove. The importance of the arginine residue at position 6 in DNA binding is reinforced by the fact that the peptide containing the mutation R6A bound to DNA with 50- to 100-fold less affinity than the wild-type peptide. The secondary structure of the peptide, as shown by CD studies, was mostly random conformation. However, the spectra of the peptide in the presence of DNA showed a decrease of ellipticity at ~220 nm, suggesting an increase in α-helical content upon DNA binding. It is possible that this region is more structured in the protein than in the peptide, and it may have an α-helical conformation stabilized by either DNA interactions, as in RecA (Kumar *et al.*, 1993), or other regions of the protein. It is interesting to note that an extensive computer search has revealed that nearly 10% of all viral proteins contain potential N-terminal α-helices with strong amphipathic character, 55% of which interact functionally with nucleic acids (Saier and McCaldon, 1988).

The DNA binding motifs that have been proposed previously for proteins that bind through the minor groove with low sequence specificity differ from the one proposed here. An exposed loop has been described for DNase I (Suck *et al.*, 1988), the only case in which structural data of the protein–DNA interaction are available. Other motifs have been suggested, such as an antiparallel β-sheet in *E.coli* histone-like protein HU (White *et al.*, 1989) and a β-turn structure (Suzuki *et al.*, 1993) in the SPKK motif found in the N-terminal tails of histones H1 and H2B (Suzuki, 1989), in HMG-1 proteins (Reeves and Nissen, 1990) and in yeast DAT1 protein (Reardon *et al.*, 1993).

In a few instances, α-helical structures have also been suggested. Protamines, with a high content of basic residues, have been proposed to interact with the phosphate groups through both the major and minor grooves (Warrant and Kim, 1978). In RecA, CD studies have shown an increase in α-helical content upon DNA interaction (Kumar *et al.*, 1993), suggesting that the DNA binding domain, probably located in two disordered regions shown in the 3-D structure (Story *et al.*, 1992), has a potential α-helical structure. In the case of the HMG domain of SRY, the partial intercalation of a non-polar side chain in the minor groove led to the prediction of an insertion of an α-helix into a widened minor groove (King and Weiss, 1993). The results presented in this paper, in contrast, indicate a different type of interaction for protein p6. The amino acids shown to be involved in DNA binding are basic or polar, and whereas protein p6 dimer formation increases with the ionic strength (R.Freire, unpublished results), DNA binding decreases (Prieto *et al.*, 1988). Therefore we conclude that protein p6 interaction with DNA is mainly electrostatic. In addition, the proposed α-helix must be positioned outside of the minor groove, since the dimensions of the groove would not allow the accommodation of an α-helix (Kim, 1983), especially in this case in which the DNA is wrapped around a protein core, so protein p6 faces a narrowed minor groove (Serrano *et al.*, 1990).

For a decapeptide motif found in histones H1 and H4,

it has been proposed that on theoretical grounds an α-helix could interact with both DNA strands bordering a relatively narrow minor groove. The amino acids, mostly basics at positions 3, 4, 7 and 8 of the helix, can be arranged in a spatial orientation that is precisely complementary to that of the phosphate oxygen atoms of DNA (Turnell *et al.*, 1988). The amino acid residues of protein p6 found to be involved in DNA binding (Glu5, Arg6, Thr9 and Lys10) can be arranged in the same manner. Thus, the N-terminal region of protein p6 could interact with the DNA backbone in a similar way, constituting the first experimental evidence of the proposed model (Turnell *et al.*, 1988). Two additional amino acids flanking the central part of the motif, Lys2 and Lys17, are also involved in this interaction. This would increase in three turns the length of the helix. The additional contacts could explain the strong distortion imposed by protein p6 in the DNA path (Serrano *et al.*, 1993) and could be related to the DNase I hypersensitivities induced by the protein p6 N-terminal peptide.

In conclusion, we propose that the N-terminal region of protein p6 constitutes a new protein domain for DNA binding in which the polar side of an amphipathic α-helix interacts with the DNA backbone through a narrowed minor groove.

Materials and methods

Chemicals, oligopeptides, oligonucleotides, nucleotides and enzymes

Glutaraldehyde was obtained from Serva. Synthetic oligonucleotides were obtained from Isogen. The peptides NH_2-AKMMQREITKTTVNV-AKMV-amide (wild-type peptide) and NH_2-AKMMQAEITKTTVNVA-KMV-amide (R6A peptide) were obtained from Multiple Peptide System and their purity was >90% as determined by HPLC. The site-directed mutagenesis kit and [α-^{32}P]dATP (400 and 3000 Ci/mmol) were purchased from Amersham International plc. DNase I was obtained from Worthington; restriction endonucleases, Klenow fragment of DNA polymerase I and T4 DNA ligase were acquired from New England Biolabs.

Site-directed mutagenesis and expression of protein p6 mutants

The DNA fragment containing the ϕ29 gene 6 was obtained from plasmid pRP8 (Pastrana *et al.*, 1985) by digestion with *Acc*I and *Eco*RI. The cohesive ends of the gene 6-containing fragment were filled in with the Klenow fragment of DNA polymerase I and then cloned into the *Sma*I site of the M13mp19 replicative form. This construction, named mp6wt, was used as template for site-directed mutagenesis, carried out essentially as described (Sayers *et al.*, 1988). The *Eco*RI–*Hin*dIII fragment of mp6wt derivatives containing the mutant genes was isolated and cloned into the expression vector pPLc28 (Remaut *et al.*, 1981), under the control of the P_L promoter of phage λ. All the mutant genes were entirely sequenced (Sanger *et al.*, 1977) to confirm that no other mutation was present. Overproduction of the different mutant proteins was carried out essentially as described in *E.coli* NFI cells containing the thermosensitive *cI* 857 repressor (Pastrana *et al.*, 1985).

Purification of protein p6 mutants

E.coli cells overproducing protein p6 mutants were disrupted by grinding with alumina, and suspended in buffer A (50 mM Tris–HCl, pH 7.5, 5% glycerol) containing 1 M NaCl; alumina and cell debris were removed by centrifugation. Polyethyleneimine (PEI) was added up to 0.25% to the supernatant, after adjusting absorbance at 260 nm to 120 U/ml, and centrifuged for 10 min at 10 000 *g* to remove DNA. The supernatant was made to 0.35 M NaCl with buffer A, and the protein p6 mutants were recovered in the pellet after centrifugation as above. To eliminate PEI, the pellet was resuspended in 0.4 M NaCl, and the protein was precipitated with ammonium sulfate to 70% saturation. The pellet was dissolved in buffer A and applied to a heparin–agarose column. Protein p6 mutants were eluted with buffer A containing 0.20 or 0.25 M NaCl, depending on the mutant. The samples were further subjected to chromatography through phenyl-Sepharose columns equilibrated in buffer A containing 0.25 M NaCl; protein p6 mutants were collected in the flow-through. After this step most of the mutant proteins were at least 90% homogeneous. In some cases, a 15–30% glycerol gradient was performed to reach that purity. Protein concentration was determined by the Lowry method (Lowry *et al.*, 1951).

Initiation of ϕ29 DNA replication assay

The reaction mixture contained, in 25 μl, 50 mM Tris–HCl, pH 7.5, 10 mM $MgCl_2$, 20 mM ammonium sulfate, 0.1 mg/ml bovine serum albumin, 1 mM dithiothreitol, 10% glycerol, 0.25 μM [α-^{32}P]dATP (400 Ci/mmol), 0.5 μg ϕ29 DNA–TP complex, 20 ng purified TP, 20 ng purified ϕ29 DNA polymerase and the indicated amounts of either wild-type or mutant p6 proteins. After incubation for 5 min at 15°C, the reaction was stopped by adding up to 10 mM EDTA and 0.1% SDS, and the samples were filtered through Sephadex G-50 spin columns in the presence of 0.1% SDS. The initiation complex formed was analyzed by SDS–PAGE as described (Peñalva and Salas, 1982). Quantitation was performed by densitometric scans of the exposed film in a Molecular Dynamics 300A densitometer.

DNase I footprinting

DNase I footprinting was carried out essentially as described (Galas and Schmitz, 1978). The ^{32}P-labeled *Hin*dIII ϕ29 DNA right terminal fragment (2 ng) was incubated in the presence of the indicated amounts of either wild-type or mutant p6 proteins in a final volume of 25 μl of the initiation reaction mixture. In the interference experiment the buffer contained 36 instead of 20 mM ammonium sulfate. For peptide or distamycin footprinting the DNA fragment was incubated, with the indicated amounts of peptide and/or distamycin, in a 20 μl buffer containing 50 mM Tris–HCl, pH 7.5 and 10 mM $MgCl_2$. When the DNA fragment containing direct repeats of the protein p6 high-affinity binding unit (Serrano *et al.*, 1993) was used, 2 ng of the fragment were incubated with 6 μg of wild-type protein p6 or 1 μg of wild-type peptide in 20 μl of a solution containing 50 mM Tris–HCl, pH 7.5, and 10 mM $MgCl_2$. To improve binding, the incubation and DNase I digestion were carried out at 15°C in all cases, except in the distamycin experiment which was performed at 0°C.

Glutaraldehyde crosslinking of protein p6

Glutaraldehyde crosslinking was carried out essentially as described (Balzer *et al.*, 1992). Either wild-type or mutant p6 proteins, 3 μg each, were incubated at room temperature in 40 μl of triethanolamine, pH 8.0, 50 mM NaCl and 500 μM glutaraldehyde. After 1 h incubation, the reaction was stopped by adding Tris–HCl, pH 7.5 up to 150 mM. Samples were analyzed by SDS gel electrophoresis in a 10–20% polyacrylamide gradient.

Circular dichroism

CD spectra were recorded using a Jasco-600 dichrograph. Samples were prepared at room temperature in 10 mM sodium phosphate, pH 7.5, containing 50 mM NaCl. ϕ29 DNA and wild-type peptide were used at concentrations of 50 μg/ml. When both peptide and DNA were incubated together, samples were centrifuged at 5000 *g* for 10 min to remove insoluble material. The DNA and peptide concentrations were calculated, before CD analysis, by absorbance at 260 nm and by amino acid analysis, respectively. Spectra were recorded from 310 to 200 nm using a 1 cm path length cuvette. Molar ellipticity was calculated per amino acid residue assuming a mean M_r of 114.6.

Acknowledgements

We are very grateful to M.Serrano for his contribution in the early steps of this work and for helpful suggestions. We are also indebted to J.M.Lázaro for his advice in protein purification and to J.Gavilanes, A.Martínez del Pozo and V.M.González for their help in CD studies. We also thank A.Bravo for discussions and C.Gutiérrez for critical reading of the manuscript. This work has been supported by grants 5R01 GM27242-15 from the National Institutes of Health, BIOT-CT91-0268 and CHRX-CT92-0010 from the European Economic Community and PB90/0091 from the Dirección General de Investigación Científica Técnica. The institutional help of Fundación Ramón Areces is also acknowledged. R.Freire was a recipient of a predoctoral fellowship from Comunidad Autónoma de Madrid.

References

Anthony-Cahill,S.J., Benfield,P.A., Fairman,R., Wassweman,Z.R., Brenner,S.L., Stafford,W.F.,III, Altenbach,C., Hubbell,W.L. and Degrado,W.F. (1992) *Science*, **255**, 979–983.
Balzer,D., Ziegelin,G., Pansegrau,W., Kruft,V. and Lanka,E. (1992) *Nucleic Acids Res.*, **20**, 1851–1858.
Blanco,L., Gutiérrez,J., Lázaro,J.M., Bernard,A. and Salas,M. (1986) *Nucleic Acids Res.*, **14**, 4923–4937.
Blanco,L., Bernad,A. and Salas,M. (1988) *J. Virol.*, **62**, 4167–4172.
Blasco,M.A., Lázaro,J.M., Blanco,L. and Salas,M. (1993) *J. Biol. Chem.*, **268**, 16763–16770.
Bravo,A., Hermoso,J.M. and Salas,M. (1994) *Mol. Gen. Genet.*, in press.
Carrascosa,J.L., Camacho,A., Moreno,F., Jiménez,F., Mellado,R.P., Viñuela,E. and Salas,M. (1976) *Eur. J. Biochem.*, **66**, 229–241.
Chou,P.Y. and Fasman,G.D. (1978) *Adv. Enzymol.*, **47**, 45–148.
Churchill,M.A.E. and Travers,A.A. (1991) *Trends Biochem. Sci.*, **16**, 92–97.
Coll,M., Frederick,C., Wang,A.H. and Rich,A. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 8385–8389.
Cunningham,B.C. and Wells,J.A. (1989) *Science*, **244**, 1081–1085.
Fisher,D.E., Parent,L.A. and Sharp,P.A. (1993) *Cell*, **72**, 467–476.
Fox,K.R. and Waring,M.J. (1984) *Nucleic Acids Res.*, **12**, 9271–9284.
Galas,D.J. and Schmitz,A. (1978) *Nucleic Acids Res.*, **9**, 3157–3170.
Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97–120.
Grosschedl,R., Giese,K. and Pagel,J. (1994) *Trends Genet.*, **10**, 94–100.
Gutiérrez,C., Freire,R., Salas,M. and Hermoso,J.M. (1994) *EMBO J.*, **13**, 269–276.
Kim,J.L., Nikolov,D.B. and Burley,S.K. (1993) *Nature*, **365**, 520–527.
Kim,S.H. (1983) In Mizobuchi,K., Watanabe,I. and Watson,J.D. (eds), *Nucleic Acid Research: Future Development.* Academic Press, New York, pp. 165–177.
Kim,Y., Geiger,J.H., Hahn,S. and Sigler,P.B. (1993) *Nature*, **365**, 512–520.
King,C.Y. and Weiss, M.A. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 11990–11994.
Kumar,K.A. and Muniyappa,K. (1992) *J. Biol. Chem.*, **267**, 24824–24832.
Kumar,K.A., Mahalakshmi,S. and Muniyappa,K. (1993) *J. Biol. Chem.*, **268**, 26162–26170.
Lowry,O.H., Rosenbrough,N.J., Farr,A.L. and Randall,R.J. (1951) *J. Biol. Chem.*, **193**, 265–275.
Mellado,R.P., Peñalva,M.A., Inciarte,M.R. and Salas,M. (1980) *Virology*, **104**, 84–96.
Murray,C.L. and Rabinowitz,J.C. (1982) *J. Biol. Chem.*, **257**, 1053–1062.
Otero,M.J., Lázaro,J.M. and Salas,M. (1990) *Gene*, **95**, 25–30.
Pabo,C.O. and Sauer,R.T. (1992) *Annu. Rev. Biochem.*, **61**, 1053–1095.
Pastrana,R., Lázaro,J.M., Blanco,L., García,J.A., Méndez,E. and Salas,M. (1985) *Nucleic Acids Res.*, **13**, 3083–3100.
Peñalva,M.A. and Salas,M. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 5522–5526.
Prieto,I., Serrano,M., Lázaro,J.M., Salas,M. and Hermoso,J.M. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 314–318.
Reardon,B.J., Winters,R.S., Gordon,D. and Winter,E. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 11327–11331.
Reeves,R. and Nissen,M.S. (1990) *J. Biol. Chem.*, **265**, 8573–8582.
Remaut,E., Stanssens,P. and Fiers,W. (1981) *Gene*, **15**, 81–93.
Saier,M.H. and McCaldon,P. (1988) *J. Bacteriol.*, **170**, 2296–2300.
Salas,M. (1991) *Annu. Rev. Biochem.*, **60**, 39–71.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
Sayers,J.R., Schmidt,W. and Eckstein,F. (1988) *Nucleic Acids Res.*, **16**, 791–802.
Serrano,M., Gutiérrez,J., Prieto,I., Hermoso,J.M. and Salas,M. (1989) *EMBO J.*, **8**, 1879–1885.
Serrano,M., Salas,M. and Hermoso,J.M. (1990) *Science*, **248**, 1012–1016.
Serrano,M., Gutiérrez,C., Salas,M. and Hermoso,J.M. (1993) *J. Mol. Biol.*, **230**, 248–259.
Serrano,M., Gutiérrez,C., Freire,R., Bravo,A., Salas,M. and Hermoso,J.M. (1994) *Biochimie*, in press.
Story,R.M., Weber,I.T. and Steiz,T.A. (1992) *Nature*, **335**, 318–325.
Suck,D., Lahm,A. and Oefner,C. (1988) *Nature*, **332**, 464–468.
Suzuki,M. (1989) *EMBO J.*, **8**, 797–804.
Suzuki,M., Gerstein,M. and Johnson,T. (1993) *Protein Engng*, **6**, 565–574.
Turnell,W.G., Satchwell,S.C. and Travers,A.A. (1988) *FEBS Lett.*, **232**, 263–268.
Warrant,R.W. and Kim,S.H. (1978) *Nature*, **271**, 130–135.
White,S.W., Appelt,K., Wilson,K.S. and Tanaka,I. (1989) *Proteins*, **5**, 281–288.
Zimmer,C. and Wahnert,L. (1986) *Prog. Biophys. Mol. Biol.*, **47**, 31–112.
Zlotnick,A. and Brenner,S.L. (1988) *J. Mol. Biol.*, **209**, 447–457.

Received on May 11, 1994; revised on June 22, 1994