ELSEVIER

Biophysical Chemistry 50 (1994) 183–189

Biophysical Chemistry

# DNA structure in the nucleoprotein complex that activates replication of phage Ø29

José M. Hermoso *, Raimundo Freire, Alicia Bravo, Crisanto Gutiérrez, Manuel Serrano [1], Margarita Salas

*Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM), Universidad Autónoma, Canto Blanco, 28049 Madrid, Spain*

(Received 5 January 1994)

**Abstract**

Initiation of phage Ø29 DNA replication is activated by the viral protein p6 which forms a nucleoprotein complex at the replication origins, located at the linear genome ends. The complex consists of a DNA right-handed superhelix wrapped around a multimeric protein core. We have determined the superhelical path of the DNA in the complex, measuring the change in linking number induced by the protein, the surface-related helical repeat and the compaction of the DNA. One superhelical turn has ≈ 63 bp (2.6 p6 dimers). Furthermore, we have determined that the DNA binding domain of protein p6 is located at the N-terminal region, predicted to form an amphipathic α-helix. We have obtained, by site-directed mutagenesis, protein p6 mutants in the polar side of the putative helix in which their DNA binding and replication activation properties were impaired or undetectable, in agreement with *in vivo* results.

*Key words:* Nucleoprotein complex; DNA binding motif; Supercoiled DNA; Initiation of DNA replication; Phage Ø29

## 1. Introduction

*Bacillus subtilis* phage Ø29 has a linear, double-stranded DNA molecule of 19285 bp with a terminal protein (TP) covalently linked to the 5′ ends. Phage Ø29 initiates the replication of its genome by a protein-priming mechanism, in which the viral DNA polymerase interacts with a free molecule of TP forming a heterodimer that recognizes the origins of replication located at both DNA ends. The initiation complex, formed by the covalent linkage of dAMP to the TP, is further elongated by the phage DNA polymerase by a strand-displacement mechanism (for a review, see ref. [1]). In addition to the genes coding for TP (gene 3) and DNA polymerase (gene 2), viral genes 1, 5, 6, and 17 are also involved in Ø29 DNA replication *in vivo* [2,3]. *In vitro* studies have demonstrated that protein p5 is a SSB protein [4,5], and that protein p6 binds to dsDNA forming a multimeric nucleoprotein complex [6],

* Corresponding author. Phone: 1-397 8429, Fax: 1-397 4799.
[1] Present adress: Cold Spring Harbor Laboratory, PO box 100, Cold Spring Harbor, NY 11724, USA.


*SSDI* 0301-4622(94)00019-G

in which the DNA adopts a right-handed toroidal conformation winding around a protein core [7]. The formation of this complex at the origins of replication is required for the *in vitro* activation of the initiation of Ø29 DNA replication [8,9] and may also explain the repression of the Ø29 C2 early promoter, observed both *in vivo* and *in vitro* [10,11]. The *in vitro* formation of multiple, scattered protein p6–DNA complexes through virtually the entire Ø29 genome, led us to propose that protein p6 plays a structural role in the organization of the viral genome into a compact nucleoprotein complex [12].

## 2. Structure of the protein p6–DNA complex

Protein p6 forms dimers in solution as determined by analytical centrifugation [13], and by glutaraldehyde crosslinking [14], and binds cooperatively to dsDNA, but not ssDNA, as determined by a gel retardation assay [6]. The nature of this interaction has been further studied by footprinting experiments. Since protein p6 activates Ø29 DNA replication at the initiation step [15], and the origins of replication are located at the ends of the linear genome, footprinting studies of the protein p6–DNA complex have been mostly carried out with Ø29 DNA terminal fragments. In these fragments, protein p6 produces a specific DNase I digestion pattern, characterized by strong hypersensitive bands regularly spaced all along the fragment ($\approx$ 270 bp) with a periodicity of 24, flanking protected regions [6]. When longer Ø29 DNA terminal fragments ($\approx$ 450–650 bp) were used, it was observed that the complex was extended 250–300 bp from both genome ends [8]. The interaction of protein p6 with the Ø29 replication origins has been further studied by hydroxyl radical footprinting, so the contacts of the protein with the DNA backbone could be determined with a high degree of resolution [16]. Protein p6 protected regions of 3–4 bp regularly spaced with a periodicity of 12, spanning through the whole fragment. The protections in both DNA strands were staggered 4 nucleotides toward the 3′ end, suggesting that the protein binds to the DNA through the minor groove [17]. Both footprinting results imply the formation of a regular, multimeric nucleoprotein complex, since the combined patterns of DNase I digestion and hydroxyl radical protection reveal the repetition of a single motif (protein p6 binding unit). The protein p6 binding unit is formed by a protein dimer bound to a 24 bp DNA segment, in which the centers of the two monomer binding sites are located 12 bp apart. Every monomer binding site is flanked by two DNase I hypersensitive sites at alternating strands so that, in each DNA strand, the hypersensitive sites have a periodicity of 24 [17]. This strongly suggests that protein p6 binds to DNA as a dimer. The strong DNase I hypersensitivities can be explained in terms of strong DNA bends or kinks in which the DNA minor groove is widened at the outside of the curve, favoring the enzyme cleavage [18]. Thus, a protein monomer would contact and bend the DNA every $\approx$ 12 bp, suggesting a model in which the DNA would wrap around a multimeric core of protein p6. In addition, since the helical repeat of the DNA wrapped on the protein surface is 12 (see ref. [19] for topological concepts on surface-wrapped DNA) and therefore larger than the absolute helical repeat of the DNA (10.5), a right-handed superhelix would be generated [20]. This model [17] also accounts for the positive supercoiling restrained by protein p6 on closed circular DNA, observed by gel electrophoresis [6,7].

As a consequence of DNA wrapping, protein p6–DNA complexes show a considerable reduction in length, with respect to naked DNA. By electron microscopy after glutaraldehyde fixation, it has been shown that the compaction of DNA in the complex is 4.2-fold using a spreading technique [7], and 6.5-fold using an adsorption method [12].

As it will be seen below, protein p6 binds with high-affinity to DNA fragments containing tandem repeats of a 24 bp long sequence (protein p6 high-affinity binding unit) present in the Ø29 DNA left terminus. The availability of plasmids containing different number of protein p6 high-affinity binding units, allowed us to calculate the change of linking number induced by a single protein p6 dimer, $\Delta L_k \approx 0.1$ [7].

The values obtained for the protein p6-induced change of linking number and the DNA length-reduction in the nucleoprotein complex, together with that of the surface-related helical repeat, define the DNA superhelix. Applying the expression of the linking number change developed for surface-wrapped DNA molecules [21], we calculated that one superhelical turn has 63 bp (2.6 protein p6 dimers) [7]. By geometrical calculations, considering a DNA compaction of 4.2-fold, the parameters that define the path followed by the DNA in the protein p6 complex were obtained, namely a pitch of 5.1 nm and a diameter of 6.6 nm [7]. Consequently, the DNA appears strongly bent (66° every 12 bp) and underwound (11.5 bp/turn) [7].

Besides protein p6–DNA, the nucleosome is the only nucleoprotein complex in which the ds-DNA path has been determined. One superhelical turn, in the nucleosome, has more bp (81), and is shorter (2.8 nm pitch) and wider (8.6 nm diameter) [22] than in the protein p6 complex. However, the main differences are that it is a multiprotein complex of fixed size (146 bp) in which the DNA is negatively supercoiled.

## 3. Signals in the DNA recognized by protein p6

One of the key questions to study the formation of the nucleoprotein complex was to determine the DNA signals recognized by protein p6. Protein p6 recognition regions were mapped by deletion analysis, in which a set of DNA fragments containing Ø29 terminal sequences of different lengths were tested for complex formation. Though the recognition regions are not unique, the main ones mapped between positions 62 and 125 from the Ø29 DNA right end, and between positions 46 and 68 from the left one [8]. Therefore, the nucleoprotein complex does not assemble from the genome ends, since the nucleation sites are located further inside the genome. In agreement with this, the positioning of protein p6 was not changed when the terminal 10 bp were deleted from a Ø29 DNA terminal fragment [23].

A computer search for nucleotide sequence homology between both Ø29 DNA terminal sequences did not indicate the existence of even a degenerated repetition, suggesting that protein p6 does not recognize directly a specific sequence, but rather a sequence-dependent DNA structural feature. Since protein p6–DNA complex formation involves strong DNA bending, we studied the bendability properties of Ø29 DNA terminal sequences by applying the algorithm developed from nucleosomal DNA [24]. The main protein p6 recognition regions contained sequences predicted to have bendable properties that would favor complex formation [8], that is to say, a marked facility of the DNA to be bent every 12 bp. The idea that protein p6 recognizes bendability was supported by the fact that protein p6 binds with higher affinity to tandem repeats of a bendable 24 bp sequence present in the main recognition region, than to a Ø29 DNA terminal fragment [7]. Furthermore, the location of the protein array in this concatemeric sequence, as deduced from the DNase I footprint pattern, is the same as in the Ø29 genomic fragment, strongly suggesting that this sequence acts as a nucleation site for protein p6–DNA complex formation [7].

## 4. Protein p6 binding to the Ø29 genome

Protein p6 constitutes about 4% of the total mass of *B. subtilis* proteins at late time of phage Ø29 infection. Taking into account that the molecular weight of protein p6, deduced from the nucleotide sequence, is 11873 [25], it was estimated that about $3 \times 10^6$ protein p6 monomers are present per cell [23]. Assuming that the average dimensions of a *B. subtilis* cell are 3.3 μm of length and 1 μm of diameter [26], the intracellular concentration of protein p6 monomers may be as high as 2 mM. A protein p6 monomer binds to 12 bp, thus, there is enough protein to completely cover more than $1.6 \times 10^3$ Ø29 DNA molecules, an amount that would suffice to bind all the intacellular viral DNA. Since the known functions of protein p6 in DNA replication and transcription could be accomplished by its interaction with the terminal 250–300 bp of Ø29 genome, the high intracellular level of protein p6 is intriguing,

and strongly suggests that protein p6 may have further roles. This led us to study the *in vitro* formation of protein p6 complexes with the whole Ø29 genome. Under conditions that greatly favor protein p6–DNA interactions, and at protein p6 saturation, Ø29 DNA molecules fully covered with protein p6 were visualized by electron microscopy after glutaraldehyde fixation. The nucleoprotein complexes appeared as rod-like, rigid and homogeneous structures. To analyze at higher resolution protein p6–Ø29 DNA complexes we used electron microscopy after psoralen crosslinking [27]. Binding of protein p6 to dsDNA prevents extensive psoralen crosslinking, therefore under denaturing conditions, complexes are detected as ssDNA bubbles [7]. Thus, this technique allowed us to analyze the number, size and position of protein p6 complexes along the Ø29 genome. Again, under conditions that greatly favor protein p6–DNA interactions, and at protein p6 saturation, most of the whole Ø29 genome appeared covered by an irregularly spaced array of complexes mostly ≈ 200–300 bp long, although their sizes ranged from ≈ 130 bp up to ≈ 2 Kb [12]. Protein p6 multimeric complexes were also analyzed by micrococcal nuclease digestion, and these results confirmed those obtained by electron microscopy. The size of the protected fragments was heterogeneous, showing a minimal size (≈ 80–90 bp), suggesting that smaller complexes are not stable [12]. The conclusion reached from these studies is that protein p6 has the potential to play a structural role in Ø29 genome organization, holding the DNA in an appropriate conformation, and providing the adequate structural framework for multiple processes, as it happens with other multimeric nucleoprotein complexes [20].

## 5. Biological significance of protein p6 restraining of positive supercoiling

Besides the strongly distorted conformation of the DNA and its compactness, the most striking feature of protein p6–DNA complex is perhaps the handedness (right-handed or positive) of the superhelix, specially since most of supercoiled DNAs, either plectonemic or toroidal, found in nature are negative.

There are, however, some exceptions in extremely thermophilic archaebacteria; for instance, in *Sulfolobus acidocaldarius*, a reverse gyrase that introduces positive supercoiling on DNA has been described [28] and positively supercoiled plasmids have been isolated [29], and in *Methanotermus fervidus*, the histone-related protein HMf forms complexes *in vitro* in which the DNA is constrained in positive toroidal supercoils [30]. In these cases, positive supercoiling has been proposed to give a higher DNA stability against thermal denaturation.

The biological significance of the protein p6-restraining of positive supercoiling could be the discrimination between viral and bacterial DNAs, provided that intracellular protein p6 binds to Ø29 but not to host DNA. The bendable properties characterized as DNA signals recognized by protein p6 do not seem to be exclusive of Ø29 DNA sequences, and protein p6 binds to other DNAs as well, when they are covalently closed circles in the presence of topoisomerase I [6], or nicked circles [7]. DNA topological constraints, therefore, may provide the specificity for protein p6 binding. Protein p6 binding should be strongly impaired in negatively supercoiled DNAs, since the restraining of positive supercoils would increase the negative superhelicity of the rest of the molecule, and this would be an energetically unfavorable process. In support of this hypothesis, protein p6 binds exclusively to high-affinity sites cloned in pUC19 DNA, provided that it is covalently closed, and it binds all along the molecule, even when no high-affinity sites are present, when the plasmid is linearized [7]. It is likely that host DNA has a higher negative superhelicity than the linear Ø29 DNA *in vivo*, thus, protein p6 could bind preferentially to the viral genome.

## 6. DNA binding domain of protein p6

To characterize functional domains in protein p6, deletion mutants were obtained [9,31]. When

five N-terminal amino acids were deleted, the protein showed a reduced DNA binding affinity, and no activity was detected when the deletion was extended to 13 amino acids [31]. This result suggested that the N-terminal region of protein p6 was involved in DNA binding. Secondary structure predictions showed that this region had a strong tendency to form an amphipatic α-helix. Based on the fact that the distance between two adjacent amino acids in an α-helix would allow them to interact with two contiguous phosphate groups in B DNA, a model for the interaction of a α-helix with the DNA minor groove was proposed [32]. It was, therefore, tempting to speculate that protein p6 interacted with the DNA minor groove by hydrogen bonding and/or electrostatic interactions between amino acids of the polar side of the α-helix and the phosphate groups. This hypothesis was tested designing protein p6 mutants in which basic or polar amino acids of the polar side of the predicted α-helix were replaced by alanine residues, to preserve the secondary structure of the protein [33]. In addition, polar residues were substituted by lysines to study the effect of the increase of positive charge. Ten protein p6 mutants were constructed, purified and assayed for DNA binding and activation of Ø29 DNA initiation of replication *in vitro*. The DNA binding activity closely matched the activation of Ø29 DNA initiation of replication in all the cases studied, and they did not increase with the positive charge. Both activities, DNA binding and replication stimulation were impaired to different extent depending on the particular mutation, except in mutant p6N14A that showed wild-type activity. Mutant p6R6A was completely inactive, indicating that arginine at position six plays a crucial role in DNA binding [14].

We have developed an *in vivo* functional assay for protein p6 based on the ability of protein p6-producing *B. subtilis* $su^-$ cells to support growth of a Ø29 *sus* 6 mutant phage. This *trans*-complementation assay has been used to study the effect on *in vivo* viral DNA synthesis of missense mutations introduced into the protein p6 N-terminal region. The two mutant proteins studied, p6K2A and p6R6A, showed a viral DNA synthesis strongly impaired and undetectable, respectively, in full agreement with the *in vitro* results [34].

## 7. Activation by protein p6 of initiation of Ø29 DNA replication

A fundamental step in the initiation of replication is the melting of a sequence precisely located near the replication origins. Formation of multimeric nucleoprotein complexes at replication origins that induce strand separation in AT-rich adjacent regions have been described in a variety of prokaryotic systems [35]. In eukaryotics, the binding of T-antigen to the SV40 origin of replication also leads to the melting of approximately 8 bp, located in a pyrimidine-rich region [36]. The role of protein p6 in the activation of Ø29 DNA initiation of replication could be to facilitate the required DNA unpairing. It is likely that the strong distortion of the DNA structure in the protein p6–DNA complex could favor this process. In fact, as mentioned above, the DNA engaged in the protein p6–DNA complex is underwound, and this could constitute a destabilizing factor decreasing the energy required for strand separation. In agreement with this hypothesis, protein p6 stimulation of the initiation reaction was much greater at 0–10°C, when the strand-separation process is disfavored, than at 30°C [37].

The nucleation site for complex formation that, as mentioned above, is not located at the very genome ends, defines the phase in which the protein moiety is arranged. As deduced from hydroxyl radical footprinting, the contact site of the terminal monomer of protein p6 is centered 11 bp away from each Ø29 DNA end. An additional question is whether complex formation is enough for activation of replication, or a precise positioning of protein p6 with respect to the origins of replication is further required. To answer this question, 4 bp were inserted between the nucleation site and the minimal origin of replication [38] in the right Ø29 DNA terminal fragment. When this fragment was used as template in the initiation reaction, no activation by

protein p6 was observed, while in a 24 bp insertion mutant, therefore with the original phase restored, the initiation of replication was activated to the same extent as that of the wild-type fragment [17]. These results indicate that the activation of initiation, not only requires the formation of the nucleoprotein complex, but a precise positioning of protein p6 with respect to the replication origin, strongly suggesting that the proteins involved in initiation, namely Ø29 DNA polymerase and TP, either directly interact with protein p6, or recognize a precisely located DNA conformational change induced by protein p6.

The mechanism of activation of initiation of Ø29 DNA replication could involve the following steps: (1) protein p6 recognition of a signal in the DNA, from which the complex is assembled with a specific disposition of the protein array, strongly distorting the DNA conformation, (2) stereospecific interaction of TP–DNA polymerase with the protein p6–Ø29 DNA complex, (3) local and transient melting of a DNA sequence, at or near the replication origin. The presence, at a specific site, of a sequence with low helical stability could be an additional element required for activation [20]. In agreement with this, protein p6 did not activate the formation of the initiation complex, using as template a Ø29 DNA right terminal fragment with a deletion of 24 bp (protein p6 binding unit) between the nucleation site and the minimal replication origin, suggesting that this region is also required for activation [23].

## Acknowledgement

This work has been supported by grants 5R01 GM27242-14 from the National Institutes of Health, BIOT-CT 91-0268 from the European Economic Community and PB90/0091 from the Dirección General de Investigación Científica y Técnica. The institutional help of Fundación Ramón Areces is also acknowledged. A. Bravo was holder of a postdoctoral fellowship from Ministerio de Educación y Ciencia and R. Freire of a predoctoral fellowship from Comunidad Autónoma de Madrid.

## References

[1] M. Salas, Ann. Rev. Biochem. 60 (1991) 37.
[2] A. Talavera, M. Salas and E. Viñuela, Eur. J. Biochem. 31 (1972) 367.
[3] J.L. Carrascosa, A. Camacho, F. Moreno, F. Jiménez, R.P. Mellado, E. Viñuela and M. Salas, Eur. J. Biochem. 66 (1976) 229.
[4] G. Martín, J.M. Lázaro, E. Méndez and M. Salas, Nucl. Acids Res. 17 (1989) 3663.
[5] C. Gutiérrez, G. Martín, J.M. Sogo and M. Salas, J. Biol. Chem. 266 (1991) 2104.
[6] I. Prieto, M. Serrano, J.M. Lázaro, M. Salas, and J.M. Hermoso, Proc. Natl. Acad. Sci. USA 85 (1988) 314.
[7] M. Serrano, C. Gutiérrez, M. Salas and J.M. Hermoso, J. Mol. Biol. 230 (1993) 248.
[8] M. Serrano, J. Gutiérrez, I. Prieto, J.M. Hermoso and M. Salas, EMBO J. 8 (1989) 1879.
[9] M.J. Otero and M. Salas, Nucl. Acids Res. 17 (1989) 4567.
[10] H.R. Whiteley, W.D. Ramey, G.B. Spiegelman and R.D. Holder, Virology 155 (1986) 392.
[11] I. Barthelemy, R.P. Mellado and M. Salas, J. Virol. 63 (1989) 460.
[12] C. Gutiérrez, R. Freire, M. Salas and J.M. Hermoso, EMBO J. 13 (1994) 269.
[13] R. Pastrana, J.M. Lázaro, L. Blanco, J.A. García, E. Méndez and M. Salas, Nucl. Acids Res. 13 (1985) 3083.
[14] R. Freire et al., in preparation.
[15] L. Blanco, J. Gutiérrez, J.M. Lázaro, A. Bernad and M. Salas, Nucl. Acids Res. 14 (1986) 4923.
[16] T.D. Tullius and B.A. Dombroski, Proc. Natl. Acad. Sci. USA 83 (1986) 5469.
[17] M. Serrano, M. Salas and J.M. Hermoso, Science 248 (1990) 1012.
[18] D. Suck, A. Lahm and C. Oefner, Nature 332 (1988) 464.
[19] J.H. White, R.M. Gallo and W.R. Bauer, Trends Biochem. Sci. 17 (1992) 7.
[20] M. Serrano, M. Salas and J.M. Hermoso, Trends Biochem. Sci. 18 (1993) 202.
[21] J.H. White, N.R. Cozzarelli and W.R. Bauer, Science 241 (1988) 323.
[22] T.J. Richmond, J.T. Finch, B. Rushton, D. Rhodes and A. Klug, Nature 311 (1984) 532.
[23] M. Serrano, C. Gutiérrez, R. Freire, A. Bravo, M. Salas and J.M. Hermoso, Biochimie, submitted.
[24] S.C. Satchwell, H.R. Drew and A.A. Travers, J. Mol. Biol. 191 (1986) 659.
[25] C.L. Murray and J.C. Rabinowitz, J. Biol. Chem. 257 (1982) 1053.
[26] S.E. Luria, in: The bacteria, Vol. 1, eds. I.C. Gunsalus and R.Y. Stainer, (Academic Press, New York, 1960) p. 9.
[27] J.M. Sogo and F. Thoma, Methods Enzymol. 170 (1989) 142.
[28] A. Kikuchi and K. Asai, Nature 309 (1984) 677.

[29] M. Nadal, G. Mirambeau, P. Forterre, W.D. Reiter and M. Duguet, Nature 321 (1986) 256.
[30] D.R. Musgrave, K.M. Sandman and J.N. Reeve, Proc. Natl. Acad. Sci. USA 88 (1991) 10397.
[31] M.J. Otero, J.M. Lázaro and M. Salas, Gene 95 (1990) 25.
[32] W.G. Turnell, S.C. Satchwell and A.A. Travers, FEBS Letters 232 (1988) 263.
[33] B.C. Cunninham and J.A. Wells, Science 244 (1989) 1081.
[34] A. Bravo, J.M. Hermoso and M. Salas, Nucl. Acids Res., submitted for publication.
[35] D. Bramhill and A. Kornberg, Cell 54 (1988) 915.
[36] J.A. Borowiec and J. Hurwitz, EMBO J. 10 (1988) 3149.
[37] M. Salas, J. Méndez, J.A. Esteban, M. Serrano, C. Gutiérrez, J.M. Hermoso, A. Bravo, M.S. Soengas, J.M. Lázaro, M.A. Blasco, R. Freire, A. Bernad, J.M. Sogo and L. Blanco, in: Viral strategies. Molecular biology and pathogenesis, eds. W. Doerfler and P. Böhm, (VCH, Weinheim, 1993) p. 3.
[38] J. Gutiérrez, C. Garmendia and M. Salas, Nucl. Acids Res. 16 (1988) 5895.