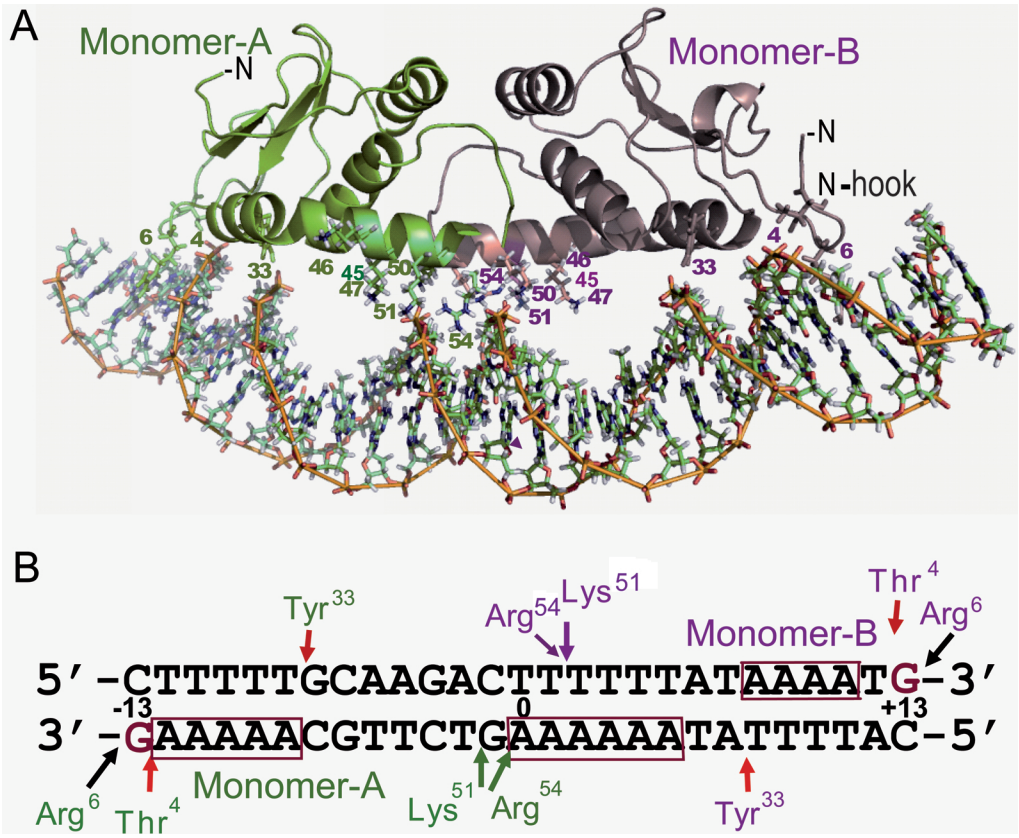
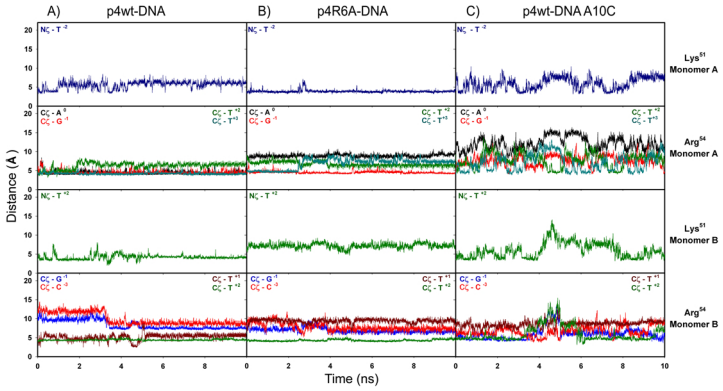
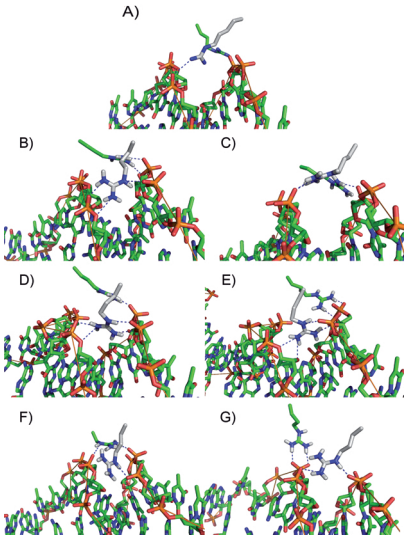
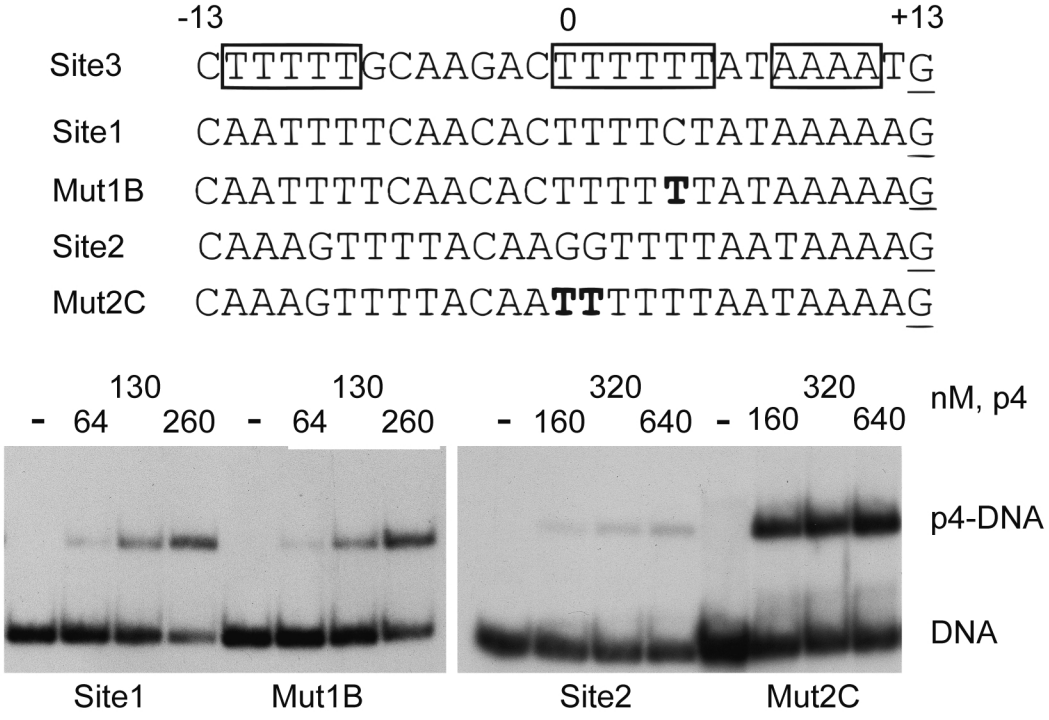
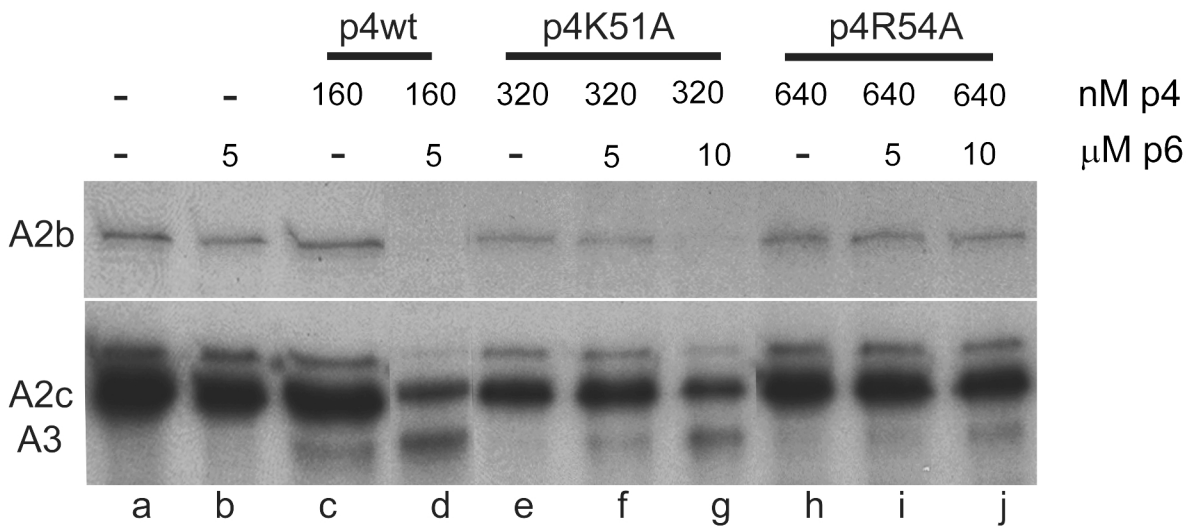
JB Accepts, published online ahead of print on 29 June 2012
J. Bacteriol. doi:10.1128/JB.00677-12

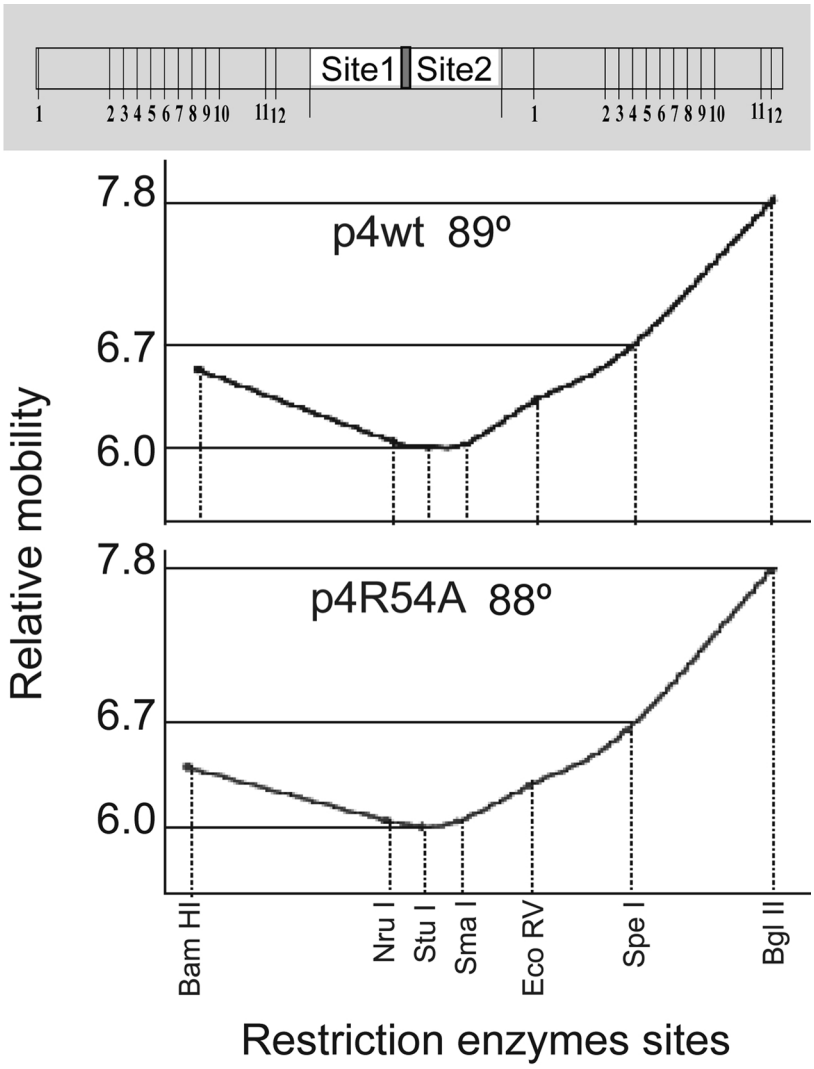


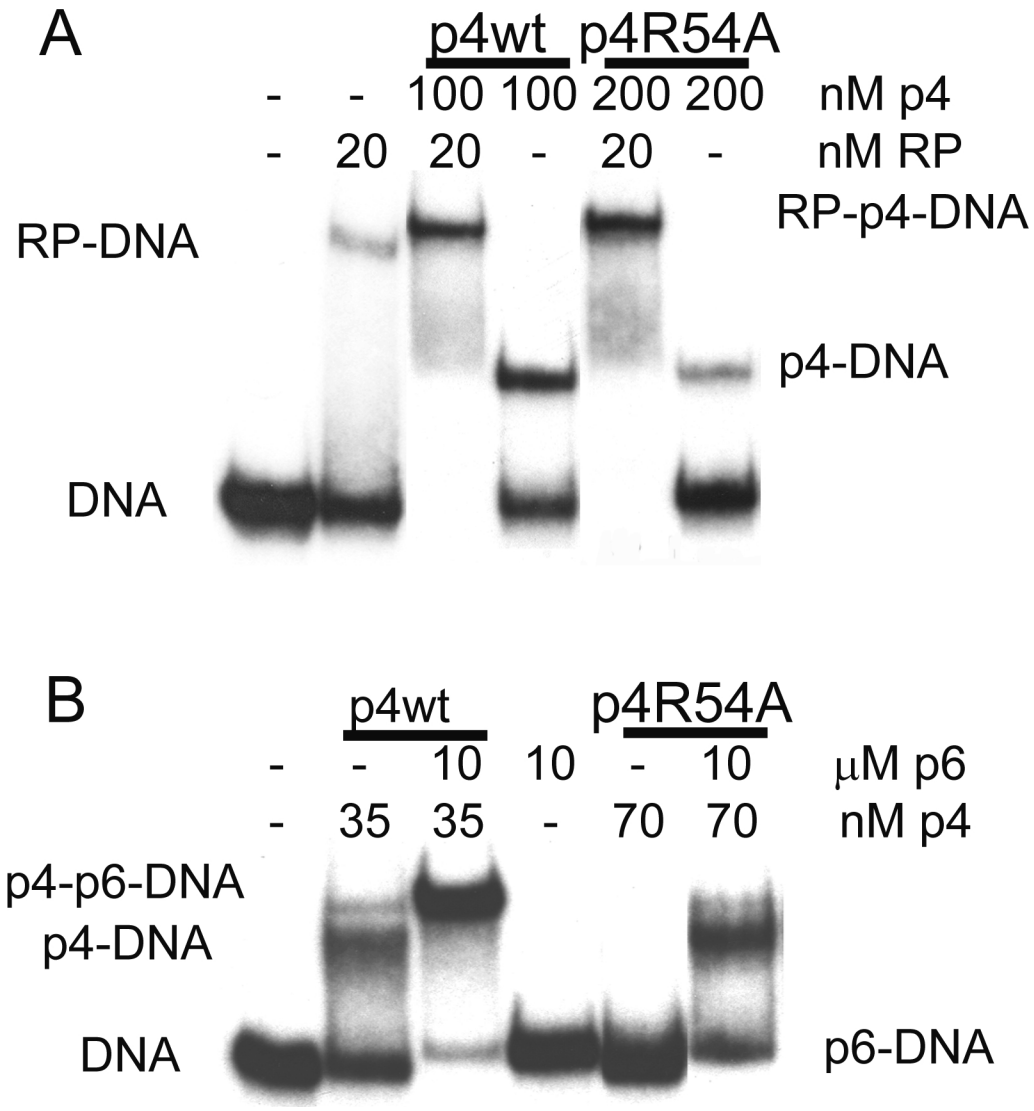


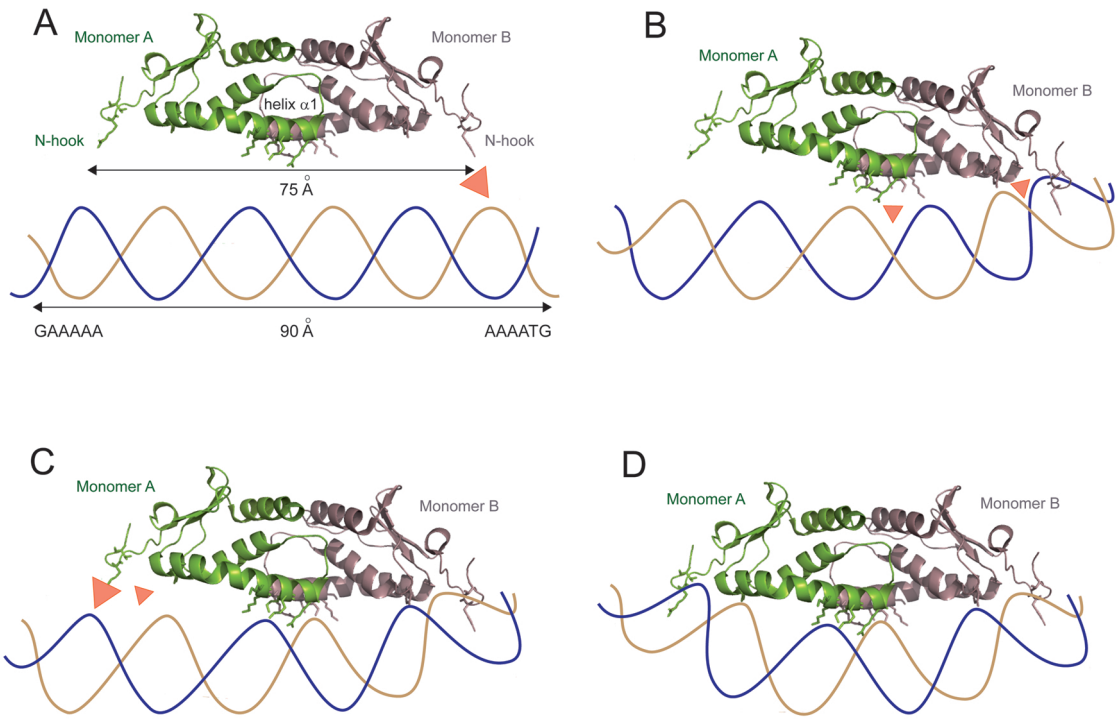












p4dir	5′-GTTTTTTGCGTCGACATCATA-3′
p4rev	5′-CGTTTGTCTGACGTCAATCTCCTTTC-3′.
Primer 1	5′-GATTTCTCTCTGCATCA-3′
Primer 4	5′-CAAAATATCTTCGTGTTCTTC-3′
pBend-1-SalI	5′-CACCCTGGTCGACAATTTTC-3′
pBend-2-SalI	5′-TAAATAGAGAACGTCGACAACAACC-3′
5063	5′-TCTACGTTCTCTATTTAACCACACC-3′
5164	5′-TTGATACATAAGGATTTGTGGGCGTTCTTG-3′
Site3A	5′-CAGATCTCAACTTTTTGCAAGACTTTTTTATAAAATGTTGAGCTAG-3′
Site2a	5′-CAGATCTCAAAGTTTTACAAGGTTTTAATAAAAGTTGAGATCTG-3′
Mut2Ca	5′-CAGATCTCAAAGTTTTACAATTTTTTAATAAAAGTTGAGATCTG-3′
Site1a	5′-CAGATCTCAATTTTCAACACTTTTCTATAAAAAGTTGAGATCTG-3′
Mut1Ba	5′-CAGATCTCAATTTTCAACACTTTTTTATAAAAAGTTGAGATCTG-3′

Functional Specificity of a Protein-DNA Complex Mediated by Two Arginines Bound to the Minor Groove

Jesús Mendieta[1,2], Laura Pérez-Lago[1,3], Margarita Salas[1] and Ana Camacho[1]*

[1]*Instituto de Biología Molecular Eladio Viñuela (CSIC), Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Nicolas Cabrera, 1. Universidad Autónoma de Madrid. 28049 Madrid, Spain.*

[2]*Biomol-Informatics SL, Parque Científico de Madrid, C/ Faraday, 7, Cantoblanco, 28049 Madrid, Spain.*

[3]*Present address: Servicio de Microbiología Clínica y Enfermedades Infecciosas, Hospital General Universitario Gregorio Marañón. 28007 Madrid, Spain.*

Running title: Arginines involved in DNA binding

**Corresponding author*. E-mail address: acamacho@cbm.uam.es

Ana Camacho, Instituto de Biología Molecular Eladio Viñuela (CSIC), Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Nicolas Cabrera, 1. Universidad Autónoma de Madrid. 28049 Madrid, Spain.

Phone: 34-91 196 4463; Fax: 34-91 196 4420

Keywords: transcription-regulator; protein-DNA recognition; A-tract; arginine in minor groove; Molecular Dynamic simulation

Abstract

Bacteriophage Ø29 transcriptional regulator, protein p4, interacts with its DNA target employing two mechanisms: direct readout of the chemical signatures of only one DNA base, and inducing local modification on the topology of short A-tracts (indirect readout). P4 binds as a dimer to targets consisting of imperfect inverted repeats. Here, we use Molecular Dynamic simulation to define interactions of a cluster of 12 positive charged amino acids of p4 with DNA and biochemical assays with modified DNA targets and mutated proteins to quantify the contribution of residues in the nucleoprotein complex. Our results show the implication of Arg54, with base-unspecific interaction in the central A-tract, in p4 binding affinity. Despite being chemically equivalent and in identical protein monomers, the two Arg54 residues differed in their interactions with DNA. We discuss an indirect-readout mechanism for p4-DNA recognition mediated by dissimilar interaction of arginines penetrating the minor groove and the inherent properties of the A-tract. Our finding extent the current understanding of protein-DNA recognition and contributes to the relevance of the sequence-dependent conformational malleability of the DNA, enlightening the role of arginines in binding affinity. Characterization of mutant p4R54A shows that the residue is required for the activity of the protein as transcriptional regulator.

Introduction

Specific interactions between proteins and DNA are fundamental for the regulation of key biological processes such as transcription, replication and recombination. Understanding the mechanisms used by regulatory proteins to discern their target sequence within the DNA genome requires the consideration of the properties of all interactions with their cognate binding sites. Structural and biological studies of contacts between proteins and DNA have led to the conclusion that sequence-specific DNA recognition involves both direct and indirect readout of the target sequences (10, 20, 23). In addition to amino acid-base specific interactions (direct readout), the affinity of a protein for its DNA target involves indirect readout of the target, where sequence-dependent variation leads to recognition of aspects of DNA such as topology of major or minor grooves, local geometry of backbone phosphates, flexibility or malleability, intrinsic curvature or water-mediated hydrogen bonds (18, 42, 54, 57). Indirect readout explains some aspects leading to the stability and affinity of several prokaryotic transcriptional regulators for their target sequences. Among those, the *Escherichia coli* catabolite activator protein (CAP) is selective for a pyrimidine-purine step involved in sequence effects on the energy of kink formation required for bending the DNA around the protein (14, 24, 34). The *trp* repressor recognizes its operator sequence indirectly through its effects on the geometry of the phosphate backbone, which in turn permits the formation of a stable interface by water-mediated contacts (3, 36). Bacteriophage 434 repressor recognizes structural features on the central base pair of its target sequence without amino acid-base pairing (28), while phage P22 c2 repressor induces the B to B´ conformational transition of its target (56).

In the *Bacillus subtilis* phage Ø29 three main promoters, early promoters A2c and A2b and the late promoter A3, are co-ordinately regulated by two viral-encoded proteins: the

transcriptional regulator protein p4 and the nucleoid-associated protein p6 (6, 12). Dimers of p4 bind specifically to four targets with the consensus sequence: 5′-CTTTTT-15 base pairs-AAAATG-3′ (40). In the structure of p4 bound to site 3 (Figure 1A), the N-terminal region of p4, folded into a unique sub-structural motif named N-hook (2), penetrates the DNA major groove where Arg6 establishes hydrogen bonds with guanines at positions + and – 13 (G±13) (Figure 1B). There are additional contacts with phosphates in three consecutive narrow minor grooves containing A-tracks. In the two externally located A-tracts, the N-hook motif residue Thr4 contacts the phosphate of bases at position ±12 while Tyr33, in the N-terminus of helix α1, contacts DNA phosphates at position ±8 across the minor groove. Mutation of Arg6, Thr4 or Tyr33, as well as individual substitutions of the G at positions ±13 or substitutions of the ±10 and 11 A·T base pairs for less deformable base pairs, were deleterious (31). Therefore, p4-specific DNA recognition is driven by both direct readout of G±13 and by indirect readout of the external A-tracts.

In this paper we studied the contribution of a cluster of surface exposed positively-charged amino acids, particularly Lys51 and Arg54 that contact phosphates at the central minor groove, to p4-DNA complex formation and to the function of the protein. To this end, we studied p4-DNA complexes by Molecular Dynamics simulation (MD) and characterized mutant proteins as transcriptional regulators. MD is a suitable tool to explore interactions strengths (17). Previous results on the dynamic behaviour of the p4-DNA complex were in agreement with the biochemical data (31). The results presented suggest that p4 uses the inherent properties of the central A-tract to take advantage of the minor groove environment in order to facilitate p4-DNA interaction. We discuss an additional indirect-readout mechanism in the p4-DNA binding process involving the sequence manipulability of the minor groove associated with the insertion of arginine in the groove. This may be a general

92 mechanism that enables proteins with negligible direct-readout recognition mechanism, such
93 as p4, to use information in the minor grooves to achieve the required DNA-binding by
94 inducing adequate conformations of the sequence.
95

MATERIALS AND METHODS

DNAs and oligodeoxyribonucleotides

The sequences of the synthetic oligonucleotides (Isogene Bioscience BV, The Netherlands) used are listed in Table I. The 368 bp DNA fragment, for the *in vitro* transcription assay, contains promoters A2c, A2b and A3 and was PCR amplified from full-length Ø29 DNA with primers 1 and 4. DNA amplification with primers 5063 and 5164 generate a 102 bp fragment containing promoter A3. Fragments were further purified through NuSieve GTG agarose (FMC) gel electrophoresis. To obtain double-stranded p4 binding sites, two complementary oligodeoxyribonucleotides were used (Table I). One of the oligodeoxyribonucleotides from each pair was 5´-end labelled using [γ-^{32}P] ATP and T4 polynucleotide kinase (44). The labelled strand was purified from unincorporated [γ-^{32}P] ATP through a mini Quick Spin Column (Roche). Complementary oligodeoxyribonucleotides were annealed to yield double-stranded DNA by mixing labelled and unlabelled oligonucleotide in a 1:10 ratio in 80 µl of 25 mM Tris-HCl (pH 7.5), 200 mM NaCl, heated 2 min at 90 ºC, and allowed to cool gradually to 20 ºC.

Cloning of gene 4 and purification of proteins p4 and p6

A 388 bp Ø29 genomic fragment encoding protein p4 was PCR-amplified using primers p4dir and p4rev obtained from Isogen (Table I). The PCR-fragment cleaved with NdeI and EcoRI, was gel purified, ligated to plasmid pT7-7 cut with the same enzymes, and transformed into *E. coli* XL1-Blue MRF´ competent cells. The resulting plasmid pT7-7-p4 expresses protein p4 under the control of the T7 RNA polymerase Ø10 promoter (50). The correct p4 sequence was confirmed by sequencing. Expression of the proteins was carried out in the *E. coli* strain BL21(DE3) pLysS (11, 49). The purification of p4 was done as described (4). The purification of p6 was done as described (29).

Band shift assays

Band shift assays were performed with a fixed amount of DNA and various concentrations of proteins. Binding reactions (20 µl) containing labelled DNA (~100 pM), 25 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 100 mM KCl, 0.5 µg of poly[d(I-C)], 1 µg of bovine serum albumin and the protein(s) were incubated for 15 min at 4 ºC, and analyzed on non-denaturing 6% polyacrylamide gels at 4 ºC. Gels were dried, and the label present in free DNA and protein–DNA complexes, quantified in a Fuji Bass IIIs Imaging Densitometer, gave the amount of DNA-protein complex calculated as a fraction of total DNA.

Circular permutation gel retardation assays

Circular permutation assays were done as described (40). Briefly, p4 sites 1 and 2 were cloned in plasmid pBend2 (21) using oligonucleotides pBend-1-SalI and pBend-2-Sal (Table I). The plasmid was then digested with restriction enzymes to generate DNA fragments having a circular permutated p4 binding sequence. The digested fragments were dephosphorylated with calf intestinal alkaline phosphatase and labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase. The labelled DNA fragments were incubated under the condition described for the band shift assays, with purified p4wt or mutant p4R54A. The p4-DNA complexes were resolved from free DNA onto a non-denaturing 6% polyacrylamide gel and quantified as described above. The p4-induced DNA bend angles and bend centre was calculated by the equation of Thompson and Landy (51).

In vitro transcription assays

Run-off transcription assays (20 µl) contained 25 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 2 mM dithiothreitol, 2 µg of poly[d(I-C)], 100 mM KCl, 10 U of RNasin, 100 µM each GTP,

ATP, CTP and [α-^{32}P]UTP (1 μCi), and ~5 nM DNA template containing promoters A2c, A2b and A3. *B. subtilis* σ^A-RNA polymerase (RNAP), purified as described (47), was added and, after 20 min at 37 ºC, reactions were stopped with SDS (0.15%) and EDTA (2.5 mM). Transcripts were analysed by electrophoresis in 6% denaturing polyacrylamide gels. Quantification of the transcripts was carried out by using a Fuji Bas-IIIs Image analyser.

Molecular Dynamics simulations

MD simulations were performed using the PMEMD module of AMBER11 and the parm 99 parameter set (7, 39)**.** The X-ray structure of p4 (PDB code 2FIO) was used for the MD simulation. The system includes the two p4 monomers (except the residues corresponding to helix α4) and double-stranded DNA with the following sequence: 5'-TAACTTTTTGCAAGACTTTTTTATAAAATGTTGA-3'. Independent simulations were carried out on the wild-type p4-site 3 complex, the mutated protein p4R6A-site 3 complex and the wild-type p4-site 3 mutated A±10C. Counterions and the solvent were added using LEAP module of AMBER. An adequate number of Na^+ ions were added to neutralize the net negative charge of the systems (57 Na^+ ions in the p4wt system and 59 Na^+ ions in the mutated complex). Ions were placed in a shell around the system using a coulomb potential in a grid. Neutralized complexes were then immersed in a truncated octahedron solvent box keeping a distance of 12 Å between the wall of the box and the closest atom of the solute. Initial relaxation of each complex was achieved by performing 10,000 steps of energy minimization using a cut off of 10.0 Å. Subsequently, and to start the MD simulations, the temperature was raised from 0 to 298 ºK in a 200-ps heating phase, and velocities were reassigned at each new temperature according to a Maxwell-Boltzmann distribution. During this period, the positions of the Cα atoms of the solute were restrained with a force constant of 20 kcal $mol^{-1}Å^{-2}$ and the Watson-Crick bonds between all the base pairs of the DNA were

171 constrained. The Cα trace was maintained with a force constant of 10 kcal $mol^{-1}Å^{-2}$ during the
172 equilibration steps to impede a spurious disorganisation of the structure during the heating of
173 the system from 0 ºK to 298 ºK. During the last 100 ps of the equilibration phase of the
174 molecular dynamics the force constant of all the constrains were reduced stepwise until 0
175 except for the distance corresponding to the Watson-Crick hydrogen bonds between the first
176 and the last base pairs of the DNA molecule which was maintained in order to mimic the
177 cooperative stabilizing effect of base pairs present at both DNA ends that are not included in
178 our system. This constraint was maintained during the productive phase of the simulations.
179 The SHAKE algorithm was used throughout to constrain all hydrogen bonds to their
180 equilibrium values so that an integration time step of 2 fs could be employed. The list of non-
181 bonded pairs was updated every 25 steps, and coordinates were saved every 2 ps. Periodic
182 boundary conditions were applied and electrostatic interactions were represented using the
183 smooth particle mesh Ewald method with a grid spacing of about 1 Å. The trajectories length
184 of 10 ns were analysed using the CARNAL module of AMBER11.
185

Results

MD simulation of p4-DNA complexes: Protein contacts and topological modification of the minor groove

The dimerization region of p4 contains a cluster of 12 positively charged amino acids (Figure 1A); of those, Lys51 and Arg54 contact the DNA backbone in the p4-DNA crystal structure and do so asymmetrically in the central minor groove (2) (Figure 1B). Crystal structures identify atomic interactions but they do not always reveal the specificity of those interactions. In order to gain insight into the stabilization of the p4-DNA complex mediated by contacts of the clustered positive amino acids we performed MD simulations of the p4-DNA complex. Simulations were carried out after standard structure equilibrium. Previous MD results (31) showed consistency and stability on the p4 conformation, where fluctuations on its Cα atoms rmsd values did not exceed 1.5Å. Quite the opposite, the rmsd values of the DNA reaching peaks of up to 3.5 Å, reflected dynamic conformations of the DNA that is adapted to fit the protein.

The complex referred to as the wild-type p4-site3 complex (p4wt-DNA) has the atomic coordinates of the 3D structure. After a preliminary analysis of the protein-DNA interactions, the atomic protein-DNA distances were measured along the 10 ns trajectory (Figure 2). From the 12 positively charged residues (Fig 1A) only Lys51 and Arg54 maintain contacts with the DNA along the MD simulation. The distance between the Nε groups of Lys51-monomer-A to the phosphate of T-2 became too large for hydrogen bonding along the 10 ns trajectory, indicating an unstable interaction (Figure 2A). Similarly, the distance between the Nε group of Lys51-monomer-B to phosphate of T+2 was too large (Figure 2A). These results suggest that Lys51-DNA interactions are not critical in the p4-DNA complex. Arg54 maintained stable interactions with several phosphates. Arg54-monomer-A contacted the T+2 phosphate only

during the first 2 ns; however, the residue contacted the T+3 phosphate and phosphates A0 and G-1 located across the minor groove along the entire trajectory (Figure 2A). In contrast, the Arg54-monomer-B maintained a stable interaction only on one DNA strand and mainly with the T+2 phosphate (Figure 2A).

The data also suggested that the guanidinium group of arginine rather than the positive charge in the residue was capable of "bridging" atoms of DNA chains. A closer analysis of the MD simulation showed both Arg54 residues located as in the 3D structure during the first ~2 ns of the trajectory (Figure 3A), superficially positioned over the minor groove and inversely oriented with respect to each other. The position of Arg54-monomer-B was quite stable and could be superimposed fairly well along the trajectory where the NH1 and NH2 groups were contacting alternately the T+2 phosphate (Figure 3B). In contrast, the residue of monomer-A moved occasionally from being superficially positioned toward an asymmetrical location into the groove between residues T+3 and G-1. This conformational change was accompanied by the establishment of hydrogen bonds between Arg54NH2 and the O3´ of A0 located much deeper into the groove. By ns ~4 the arginine returned to its initial position where the NH1 and NH2 groups contacted alternately the phosphate at position -1 (Figure 3C). Hence, in spite of being chemically equivalent and in identical monomers, both arginines have distinguishable DNA interactions.

To gain further insight into the role of arginines 54 in the p4 binding process, we analysed the behaviour of two mutated p4-site 3 complexes with some protein-DNA interactions distorted (31). One complex with the mutated protein p4R6A and wild-type site 3 (p4R6A-DNA; Figure 2B) and the other with wild-type p4 and the A±10C mutated site 3 (p4wt-DNA A10C; Figure 2C). As previously described, the p4 mutant R6A substitution abolishes the amino

236 acid-base interaction while the A±10C substitution disrupts the interaction of p4 at the
237 external A-tracks (31). The MD simulation of the p4R6A-site 3 complex showed a general
238 weakening of the Arg54-monomer-A interactions and a more stable interaction of the Lys51-
239 monomer-A with the T-2 phosphate. The guanidinium group Arg54-monomer-A interacts
240 with the T+3 phosphate in the first 2 ns of the dynamic and with the G-1 phosphate along the
241 entire simulation (Figure 2B). Along the trajectory, the residue moved from being
242 superficially positioned toward an asymmetrical position into the groove, most frequently
243 between G-1 and T+3 (Figure 3, D-E). This movement was accompanied by the establishment
244 of bonds between the NH1 and NH2 groups with the O3´ of A0 and the O3´ of T+2,
245 respectively. Arg54-monomer-B maintained stable interaction with the T+2 phosphate (Figure
246 2B). Arg54-monomer-B localized into the minor groove between T+2/T-2 and T+3/G-1
247 contacting the T+2 phosphate and the O3´ of T+1 (Figure 3, D-E). In the p4R6A-site 3
248 trajectory the linearity of the phosphates backbone presents a local bent, probably due to the
249 insertion of Arginines 54, the width of the groove was irregular (see above).
250
251 In the MD simulation of p4wt-DNA A±10C site 3 complex (Figure 2C), no stable protein-
252 DNA contacts of monomer-A were found, and Lys51-monomer-B did not contact the DNA
253 either. Interactions between Arg54-monomer-B and the phosphates of T+2 and G-1 were
254 stable only for the first ~3 ns. Hence, p4 interactions at the external A-tracks seem to
255 influence central minor groove interactions. However, although Arg54-monomer-A does not
256 maintain stable contacts with phosphates, it may still contribute to some stabilization of the
257 complex since it becomes occasionally inserted straight into the groove between T+2 and T-2
258 and intermittently interacts with the phosphates and with the O3´ of G-1 (Figure 3, F-G). We
259 had previously shown that each of the protein monomers displays different binding entropy
260 due to the asymmetry of the site 3 inverted repeats, with monomer-B having higher binding

entropy (31). Both Arginines 54, although differing in their DNA contacts, seem to be required for p4/DNA interaction.

The minor groove facing p4 narrowed locally from the 11.5 Å of a regular B-DNA width to ~ 9Å, while it widened up on its opposite face (Figure 1A) resulting in an unstable, energetically strained DNA conformation imposed by p4 (31). Measurements of the inter-phosphate distances of the central minor grooves in the MD simulations of wild-type p4-site 3 complex and the p4R6A-site 3 complex are shown in Figure S1. In the course of the p4wt MD, the minor groove was maintained locally narrow between T0 and T+4 (Figure S1) with Arginines 54 superficially located across the minor groove and, when inserted into the groove, their most frequent position was between T+3 and G-1, contacting the O3´of A0 (Figure 3B). It has been shown that enhanced negative potentials of narrowed minor groove attract arginines that could mediate specific protein-DNA interactions (41). Here Arg54 is inserted in the narrowed part of the groove (~9 Å width). The p4R6A-site 3 complex showed manifest local minimum at T0 (~7 Å) with Arg54 most frequently inserted around T+3 and G-1.

Arginines 54 might facilitate the axial displacement of the DNA strands and groove narrowing. DNA molecules show clear non-elastic behaviour when being imposed to forces (15, 22, 25). Hence, the insertion of one Arginine between T+3 and G-1 contacting both DNA strands while the other Arginine inserted at a widened position contacts one DNA strand should modify the topology of the groove. In addition, the fact that Arg54-monomer-A was most frequently inserted into the groove suggests that the residue might play a major role in narrowing the groove.

DNA sequence and p4-DNA complex stability

Arg54 contacts phosphates on the central minor groove that contains an A-tract. This central A-tract, is present from positions 0 to +5 in the higher affinity binding site 3 and is shorter in site 1 (positions 0 to +3) and site 2 (positions +2 to +5) (Figure 4). Protein p4 binds ~2- and ~8-fold less efficiently to site 1 and to site 2, respectively, than to site 3 (40). To determine if the central A-track contributes to p4-DNA complex stability, we tested targets derived from binding sites 1 and 2 with sequences close to those of site 3 on band shift assays (Figure 4). Lengthening of the site 1 A-tract (Mut1B) to position +6 did not affect p4 binding, while substitution of the two C·G base pairs at positions 0 and +1 by A·T pairs on site 2 (Mut2C) improved the affinity of the protein for the target to the level of site 3. Hence, and as would be expected from the MD data, A·T pairs at positions 0 and +1 greatly enhance p4 binding.

The function of p4 residues Lys51 and Arg54

The relevance of Lys51 and Arg54 was studied assaying the effect of mutant proteins at these two residues on the regulation of promoters A2c, A2b and A3 in *in vitro* transcription assays. Ø29 early promoters A2c and A2b are recognized by the *B. subtilis* σ^A-RNA polymerase (RNAP) while late promoter A3 is poorly expressed due to its low consensus at the -35 motif; its activation requires the binding of p4 upstream of promoter A3 (Figure 5, lanes a and c) (30). Cooperative binding of proteins p4 and p6 to the promoters region synchronizes the transition from early to late gene expression by repressing early promoters A2c and A2b and simultaneously activating promoter A3 (Figure 5, lane d) (6). P6 did not affect transcription in the absence of p4 (lane b). Mutant p4K51A was capable of transcription regulation, although less efficiently than wild-type p4 (Figure 5, lanes e to g). On the contrary, mutant p4R54A was unable to repress promoters A2c and A2b and promoter A3 was not fully activated even with 4-fold higher concentration of the protein (lanes h to j), indicating that Arg54 is required for the activity of the protein as a transcriptional regulator.

311

312 The synchronized regulation of promoters A2c, A2b and A3 requires bending of the DNA (40,
313 43) to a hairpin-like conformation (9) as well as functional interactions of p4 with RNA
314 polymerase (35) and p6 (5). We studied these properties on mutant p4R54A. Previous results
315 had shown almost no effect of p4 mutations K51A and R54A when binding to site 3 (2).
316 However, the binding of p4R54A to site 3 reached a value that was 2-fold lower than that of
317 the wild type protein (Figure S2), suggesting that some electrostatic contacts between the
318 protein and the DNA are missing in the mutant complex.

319

320 To investigate whether p4R54A binding induces DNA bending, we employed circular
321 permutation assays (Figure 6). Plasmid pBend2wt containing a 72 bp sequence of Ø29 DNA
322 including p4 sites 1 and 2 (40), was used to determine p4 induced curvature. DNA fragments
323 were produced by restriction enzyme digestions that cleaved twice in the plasmid, resulting in
324 fragments with the sequence of Ø29 located at different positions with respect to the
325 fragments ends. Free DNA migrated with similar mobility, in agreement with the absence of
326 intrinsic curvature as demonstrated previously. In contrast, fragments with a gradient of
327 mobility similar to those of p4wt were generated upon p4R54A binding, with the StuI
328 complex having the slowest mobility followed closely by the NruI- and SmaI-generated
329 fragment complexes, with the BglII fragment-p4 complex having the fastest mobility.
330 Quantification of the induced bending angles indicated values of 89º and 88º for the p4wt-
331 DNA and p4R54A-DNA complexes, respectively. Hence, in both complexes the DNA is bent.

332

333 The ability of the p4R54A mutant to stabilize RNAP at promoter A3 was analysed using a
334 DNA fragment containing promoter A3 and the overlapping promoter A2b (Figure 7A). In the
335 absence of protein p4, RNAP does not bind to A3 with low consensus at the -35 element but

rather to promoter A2b (43) (Figure 7A). Protein p4 interacting with the RNA polymerase α subunit allows closed complex formation at promoter A3 (30). With p4R54A, the p4-DNA complex was reduced with respect to that formed by wild-type p4, but the ternary RNAP-P4-DNA-complex formed was similar to that of p4wt indicating that p4R54A efficiently stimulated binding of the RNA polymerase to promoter A3, despite its failure to produce RNA from promoter A3 (Figure 5).

Cooperativity between p4 and p6 when binding to DNA results in the multimeric nucleoprotein complex, which modifying the topology of the promoters induce the switch from early to late transcription. We assayed p4R54A and p6 cooperativity with the DNA fragment used in the transcription assays, which included promoters A2c, A2b and A3. As shown in Figure 7B, simultaneous addition of p4 and p6 resulted in the formation of a ternary p4-p6-DNA complex characterized by its slower mobility compared with that of the p4-DNA or p6-DNA complexes. Protein p4R54A fails to synergize p6 binding since no stable p4-p6-DNA complex was detected, although p4-DNA binding seems stimulated by the presence of p6. Hence, most probably the hairpin-like DNA bent is impaired and in consequence early promoters A2b, A2c are not properly repressed and late promoter A3 is not fully activated.

354 **DISCUSSION**

355 Protein-DNA complex formation is not a simple encounter between structurally predefined
356 molecules. Several proteins depend on unstructured regions for binding, and DNA can also
357 modify its conformation during the formation of the protein-DNA complex (14, 26, 32, 37, 45,
358 53). Information contained within the DNA is extracted by proteins, which recognize
359 sequences of the DNA by both direct- and indirect-readout mechanisms. The p4-DNA
360 complex presents three patches of amino acids nearby three A-tracts. In the two external
361 patches, Thr4 and Tyr33, contribute to p4 binding specificity through an indirect-readout
362 mechanism that implies remodelling of the external A-tracts (31). A cloud of 12 positive
363 charged amino acids (lysines and arginines) floats over a locally narrow portion of the central
364 minor grove, with Arg54 establishing salt bridges with phosphates. Here, we have shown a
365 role of Arg54 in p4-DNA binding stability, which in turn altered its synergy with protein p6.
366
367 The presence of Arginines in the minor groove is widespread in protein-DNA complexes,
368 isolated as in Brn-5 protein (38) and LEAFY (16), or within a motif with more than one
369 arginine as in phage 434 repressor (46) and Ø29 p4 (this paper). Arginines have been shown
370 to interact at positions of minimal minor-groove width and associated to the presence of short
371 A-tracts (41). Drosophila Hox protein Sex combs reduced (Scr) uses variation in electrostatic
372 potential associated with minor-groove width to distinguish small differences in nucleotide
373 sequence (19).
374
375 MD simulations of p4–DNA complexes permitted us to evaluate the persistence of Arginines
376 54 in the central minor grove, determining the number of hydrogen bonds formed by each
377 residue and simultaneously measuring the groove width. Considering that the 3D structure is
378 the equilibrium state, both Arg54 residues inversely and superficially cross the groove contact

379 phosphates of opposite DNA strands in equilibrium (Figure 3A). Along the p4wt-site 3 MD
380 simulation, the Arg54-monomer-B was touching the phosphates rather than intruding into the
381 groove. The Arg54-monomer-A may confer specificity to the p4-DNA complex by interacting
382 simultaneously with phosphates across the groove and by intruding occasionally into the
383 groove to establish simultaneously interactions with the phosphate on one strand and with the
384 sugar oxygen of the opposite strand. These interactions might contribute to local narrowing of
385 the groove. This picture was evident when the p4-DNA complex was destabilized by
386 disruption of Arg6-G±13 interactions (p4R6A/site 3 complex). In the p4R6A/site 3 complex
387 Arg54-monomer-A lost its superficial interactions with phosphates while the rate of intrusion
388 into the groove greatly increased. Moreover, Arg54-monomer-B is also observed intruding
389 into the groove. In addition, the groove width diminished locally suggesting that Arg54
390 functions narrowing the groove. Variation on the width of the grooves in a controlled manner
391 could lead to local strand deformability. The arginines into the minor groove may use the
392 malleability of the A-track to disrupt the linear helix.
393
394 Aginines participate more frequently than lysines in protein-DNA interactions (33); it has
395 been suggested that this is due to the guanidium group of arginine engaging in more hydrogen
396 bonds than the amino group of lysine (27). Participation of arginines in DNA-histone contacts
397 has been the subject of a number of investigations (8, 13, 52). The arrangement of arginines in
398 the p4-DNA complex differs from that in the complexes of Hox, histones or phage-434
399 repressor. In the nucleosomes, the arginines are positioned asymmetrically frequently bridging
400 O4´ atoms of nucleotides i and i+3 (54). In 434 repressor-DNA complex the arginine is in the
401 centre of the groove with the guanine group bridging the O4´ atoms of nucleotides i and i+2
402 (1). In the p4-DNA complex Arg54 bridged O3´ atoms to phosphates of nucleotides i and i+3.
403

It has been recently shown that bacterial nucleoid-associated protein Fis initially select DNA targets with narrow minor groove separated as dictated by the HTH motifs of the Fis dimer (48). Unlike Fis, p4 target selection is not ruled by curved DNA (31, 40). The stability of the p4-DNA complex is a consequence of p4-induced topological modification of the DNA whereas the primary function of the DNA is its ability to acquire a conformation capable of enhancing interactions with the protein.

The data presented here showed that a short central A-tract and Arg54 are additional requirements for p4 stable binding to DNA and agree with the proposed "zipper model" for p4 DNA recognition and binding (6). The distance from G-13 to G+13 is ~90 Å, while the 75 Å distance from the Arg6 of one of the monomers to the Arg6 of the other monomer is too short to explain simultaneous interaction of both monomers with the inverted repeats of site 3. In our model (Figure 8), one protein monomer interacts first with the higher entropic stability inverted repeat sequence, 5'-AAAATG-3', intruding its N-hook into the major groove and providing Arg6-G+13 specific interaction (monomer-B). Subsequently, the proximal minor groove is narrowed through the interaction of Thr4 and Tyr33 raising the specificity of the nucleoprotein complex and approximating the central A-tract to the Arginines 54. The side chain of Arg54-monomer-A penetrating the groove and contacting simultaneously both DNA strands may induce or stabilize a narrowed groove increasing p4 specificity for its target. This topological modification leads the fit of p4 to one face of DNA while the groove gets wider at its opposite face. Two consecutive minor grooves narrowed at the same face drive the DNA to bend, reducing the target length and, therefore, allowing an adjustment of the target end to end distance to contact the hooks. The bent of the DNA would facilitate the approach of the 5'-AAAAAG-3' inverted repeat to the hook of the second monomer.

429 In conclusion, Ø29 p4 may be considered an additional example of transcriptional regulators
430 which, being a sequence-specific DNA binding protein, does not have relatively simple
431 readout properties.
432

ACKNOWLEDGMENTS

We thank to Professor Lucia B. Rothman-Denes for critical reading of the manuscript and to L. Villar and J.M. Lazaro for proteins purification. This work was supported by the Ministerio de Ciencia e Innovación of Spain [Grants BFU2008-01216/BMC to A.C., BFU2008-00215/BMC to M.S.]; by the Madrid Autonomous Community [Grant S-2009/MAT/1507 to M.S.]; and by an Institutional Grant from the Fundación Ramón Areces to the Centro de Biología Molecular Severo Ochoa.

REFERENCES

1. **Aggarwal AK, Rodgers DW, Drottar M, Ptashne M, Harrison SC**. 1988. Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. Science **242:**899-907.
2. **Badía D, Camacho A, Pérez-Lago L, Escandón C, Salas M, Coll M.** 2006. The structure of the phage Ø29 transcriptional regulator p4-DNA complex reveals a novel binding motif. Molecular Cell **22:**73-81.
3. **Bareket-Samish A, Cohen I, Haran TE.** 1998. Direct versus indirect readout in the interaction of the trp repressor with non-canonical binding sites. J. Mol. Biol. **277:**1071-1080.
4. **Barthelemy I, Lázaro JM, Méndez E, Mellado RP, Salas M.** 1987. Purification in an active form of the Ø29 protein p4 that controls the viral late transcription. Nucleic Acids Res. **15:**7781-7793.
5. **Camacho A, Salas M.** 2001. Mechanism for the switch of Ø29 DNA early to late transcription by regulatory protein p4 and histone-like protein p6. EMBO J. **20:**6060-6070.
6. **Camacho A, Salas M.** 2010. DNA bending and looping in the transcriptional control of bacteriophage Ø29. FEMS Microbiol. Rev. **34:**828-841.
7. **Case DA, Plearman DA, Cadwell JW, Cheatham III T E, Wang J, Ross W, Simmerling CL, Darden TA, Merz KM et al.** 2002. AMBER7, University of California, San Francisco. http://www.amber.ucsf.edu/amber/amber.html).
8. **Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ.** 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. J. Mol. Biol. **319:**1097-10113.

9. **Del Arroyo PG, Vélez M, Piétrement O, Salas M, Carrascosa JL, Camacho A.** 2009. A nucleoprotein-hairpin in transcription regulation. J. Struct. Biol.**168:**444-451.
10. **Drew HR, Travers AA.** 1985. DNA bending and its relation to nucleosome positioning. J. Mol. Biol.**186:**773-790.
11. **Dubendorff JW, Studier FW.** 1991. Creation of a T7 autogene. Cloning and expression of the gene for bacteriophage T7 RNA polymerase under control of its cognate promoter. J. Mol. Biol. **219:**61-68.
12. **Elías-Arnanz M, Salas M.** 1999. Functional interactions between a phage histone-like protein and a transcriptional factor in regulation of Ø29 early-late transcriptional switch. Genes Dev. **3:**2502-2513.
13. **García-Pérez M, Pinto M, Subirana JA**. 2003. Nonsequence-specific arginine interactions in the nucleosome core particle. Biopolymers **69:**432-439.
14. **Gartenberg MR, Crothers DM.** 1988. DNA sequence determinants of CAP-induced bending and protein binding affinity. Nature **333:**824-829.
15. **Gore J, Bryant Z, Nöllmann M, Le MU, Cozzarelli NR, Bustamante C.** 2006. DNA over winds when stretched. Nature **442:**836-839.
16. **Hamès C, Ptchelkine D, Grimm C, Thevenon E, Moyroud E, Gérard F, Martiel JL, Benlloch R, Parcy F, Müller CW.** 2008. Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. EMBO J. **27:**2628-2637.
17. **Hansson T, Oostenbrink C, van Gunsteren W.** 2002. Molecular dynamics simulations. Curr. Opin. Struct. Biol. **12:**190–196.
18. **Hogan ME, Austin RH.** 1987. Importance of DNA stiffness in protein-DNA binding specificity. Nature **329:**263-266.

19. **Joshi R, Passner JM, Rohs R, Jain R, Sosinsky A, Crickmore MA, Jacob V, Aggarwal AK, Honig B, Mann RS.** 2007. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. Cell **131:**530-543.

20. **Kalodimos CG, Biris N, Bonvin AM, Levandoski MM, Guennuegues M, Boelens R, Kaptein R.** 2004. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science **305:**386-389.

21. **Kim J, Zwieb C, Wu C, Adhya S.** 1989. Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. Gene **85:**15-23.

22. **Kosikov KM, Gorin AA, Zhurkin VB, Olson WK.** 1999. DNA stretching and compression: large-scale simulations of double helical structures. J. Mol. Biol. **289:**1301-1326.

23. **Koudelka GB, Mauro SA, Ciubotaru M.** 2006. Indirect readout of DNA sequence by proteins: the roles of DNA sequence-dependent intrinsic and extrinsic forces. Prog. Nucleic. Acid. Re. Mol. Biol. **81:**143-177.

24. **Lindemose S, Nielsen PE, Møllegaard NE.** 2008. Dissecting direct and indirect readout of cAMP receptor protein DNA binding using an inosine and 2,6-diaminopurine in vitro selection system. Nucleic Acids Res. **14:**4797-4807.

25. **Lionnet T, Joubaud S, Lavery R, Bensimon D, Croquette V.** 2006 . Wringing out DNA. Phys. Rev. Lett. **96:**178102.

26. **Love JJ, Li X, Case DA, Giese K, Grosschedl R, Wright PE.** 1995. Structural basis for DNA bending by the architectural transcription factor LEF-1. Nature **376:**791-795.

27. **Luscombe NM, Laskowski RA, Thornton JM.** 2001 Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res. **29:**2860-2874.

28. **Mauro SA, Pawlowski D, Koudelka GB.** 2003. The role of the minor groove substituents in indirect readout of DNA sequence by 434 repressor. J. Biol. Chem. **278:**12955-12960.

29. **Mencía M, Gella P, Camacho A, de Vega M, Salas M.** 2011. Terminal protein-primed amplification of heterologous DNA with a minimal replication system based on phage Ø29. Proc. Natl. Acad. Sci. USA **108:**18655-18660.

30. **Mencía M, Monsalve M, Rojo F, Salas M.** 1996. Transcription activation by phage Ø29 protein p4 is mediated by the interaction with the α subunit of the Bacillus **s**ubtilis RNA polymerase. Proc. Natl. Acad. Sci. USA **93:**6616-6620.

31. **Mendieta J, Pérez-Lago L, Salas M, Camacho A.** 2007. DNA sequence-specific recognition by a transcriptional regulator requires indirect readout of A-tracts. Nucleic Acids Res. **35:**3252-3261.

32. **Mondol T, Batabyal S, Mazumder A, Roy S, Pal SK.** 2012. Recognition of different DNA sequences by a DNA-binding protein alters protein dynamics differentially. FEBS Lett. **586:**258-262.

33. **Nadassy K, Wodak SJ, Janin J.** 1999. Structural features of protein-nucleic acid recognition sites. Biochemistry **38:**1999-2017.

34. **Napoli AA, Lawson CL, Ebright RH, Berman HM.** 2006. Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: recognition of pyrimidine-purine and purine-purine steps. J. Mol. Biol. **357:**173-183.

35. **Nuez B, Rojo F, Salas M.** 1992. Phage Ø29 regulatory protein p4 stabilizes the binding of the RNA polymerase to the late promoter in a process involving direct protein-protein contact. Proc. Natl. Acad. Sci. USA **89:**11401-11405.

36. **Otwinowski Z, Schevitz RW, Zhang RG, Lawson C, Loachimiak A, Marmorstein RQ, Luisi BF, Sigler PB.** 1988. Crystal structure of trp repressor/operator complex at atomic resolution. Nature **335:**321-329.

37. **Parker SC, Tullius TD.** 2011. DNA shape, genetic codes, and evolution. Curr. Opin. Struct. Biol. **21:**342-347.

38. **Pereira JH, Kim SH.** 2009. Structure of human Brn-5 transcription factor in complex with CRH gene promoter. J. Struct. Biol. **167:**159-165.

39. **Pérez A, Lankas F, Luque FJ, Orozco M.** 2008. Towards a molecular dynamics consensus view of B-DNA flexibility. Nucleic Acids Res. **36:**2379-94.

40. **Pérez-Lago L, Salas M, Camacho A.** 2005. A precise DNA bent angle is essential for the function of the phage Ø29 transcriptional regulator. Nucleic Acids Res. **33:**1209-1234.

41. **Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B.** 2009. The role of DNA shape in protein-DNA recognition. Nature **461:**1248-1253.

42. **Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS.** 2010. Origins of Specificity in Protein-DNA Recognition. Annu. Rev. Biochem. **79:**9.1-9.37.

43. **Rojo F, Zaballos A, Salas M.** 1990. Bend induced by the phage Ø29 transcriptional activator in the viral late promoter is required for activation. J. Mol. Biol. **211:**713-725.

44. **Sambrook J, Fritsch EF, Maniatis T.** 1989. Molecular cloning: A Laboratory Manual Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

45. **Schleif R.** 2000. Regulation of the L-arabinose operon of *Escherichia coli*. Trends Genet. **6**:559-565.

46. **Shimon LJ, Harrison SC.** 1993. The phage 434 OR2/R1-69 complex at 2.5 Å resolution. J. Mol. Biol. **232**:826-838.

47. **Sogo JM, Inciarte MR, Corral J, Viñuela E, Salas M.** 1979. RNA polymerase binding sites and transcription map of the DNA of *Bacillus subtilis* phage ϕ29. J Mol Biol **127:** 411-436.

561 48 Stella S, Cascio D, Johnson RC. 2010. The shape of the DNA minor groove direct binding
562 by the DNA-bending protein Fis. Genes & Dev. **24:**814-826.
563 49. **Studier FW, Moffatt BA.** 1986. Use of bacteriophage T7 RNA polymerase to direct
564 selective high-level expression of cloned genes. J. Mol. Biol. **189:**113-130.
565 50. **Tabor S, Richardson CC.** 1992. RNA polymerase/promoter system for controlled
566 exclusive expression of specific genes. Biotechnology **24:**280-284.
567 51. **Thompson JF, Landy A.** 1988. Empirical estimation of the protein-induced DNA bend to
568 lambda site-specific recombination complexes. Nucleic Acids Res. **16:**9687-9705.
569 52. **Tolstorukov MY, Colasanti AV, McCandlish DM, Olson WK, Zhurkin VB.** 2007. A
570 novel roll-and-slide mechanism of DNA folding in chromatin: implications for
571 nucleosome positioning. J. Mol. Biol. **371:**725-738**.**
572 53. **Travers A.** 1998. Transcription: activation by cooperating conformations. Curr. Biol.
573 **8:**R616-618.
574 54. **von Hippel PH.** 1994. Protein-DNA recognition: new perspectives and underlying
575 themes. Science **263:**769-770.
576 55. **Wang D, Ulyanov NB, Zhurkin VB.** 2010. Sequence-dependent Kink-and-Slide
577 deformations of nucleosomal DNA facilitated by histone arginines bound in the minor
578 groove. J. Biomol. Struct. Dyn. **27:**843-859.
579 56. **Watkins D, Srividya M, Gerald B, Williams LD.** 2010. Sequence Recognition of DNA
580 by Protein-Induced Conformational Transitions J. Mol. Biol. **396:**1145-1164.
581 57. **Zhang Y, Xi Z, Hegde RS, Shakked Z, Crothers DM.** 2004. Predicting indirect readout
582 effects in protein-DNA interactions. Proc. Natl. Acad. Sci. USA **101:**8337-8341.
583

584 **FIGURES**

585 **Figure 1.** A) Structure of the p4 dimer in complex with the DNA encompassing the sequence
586 of site 3 used in the MD simulations. Monomers of the p4 dimer are represented in green and
587 purple and each monomer is distinguished as A (green) or B (purple) depending on its position
588 with respect to the sequence of site 3. In sticks and numbered, amino acids involved in DNA
589 recognition and those referred to in this paper. B) Scheme of site 3 showing the protein-DNA
590 interactions from the 3D structure. The centre (position 0) and the extreme locations (positions
591 ±13) of the sequence are marked, guanines referred to in this paper are in red and the three A-
592 tracks are in boxes. Amino acid interactions with the DNA are signalled with arrows; black
593 arrows for the bi-dentated Arg-G interactions; red arrows for Thr and Tyr interactions and
594 green and purple for the interactions studied in this paper of one or the other p4 monomer.
595
596 **Figure 2.** Protein-DNA bond distances along the MD trajectory of the p4/site 3 complex (A),
597 protein p4R6A/site 3 complex (B) and p4wt/site3A10C complex (C). Distances were
598 measured from the Nζ atom of Lys51 or the Cζ atom of Arg54 of both monomers to the
599 phosphorous atom of the phosphate groups corresponding to the nucleotides number shown in
600 the figure.
601
602 **Figure 3.** Rearrangement of Arginines 54 side chains. Snapshots showing the position of
603 Arginines 54 in the MD simulations of the wild-type p4/site 3 complex (A,B,C), the mutated
604 protein p4R6A/site 3 complex (D, E,) and the wild-type p4/A±10C mutated site 3 (F,G).
605 Representative snapshots were selected at 0 ns (A); ~2 ns (B, D, F) and ~4 ns (C, E, G). Arg-
606 monomer-A, grey colour; Arg-monomer-B, green colour. Distances: A) NH1 Arg-monomer-A
607 to G-1 phosphate: 3.16 Å; NH1 and NH2 Arg-monomer-B to T+2 phosphate: 2.56 Å and 2.75
608 Å respectively. B) NH1 Arg-monomer-A to G-1 phosphate: 2.81 Å; NH2 Arg-monomer-A to

609 O3´A0: 3.08 Å; NE Arg-monomer-A to T+3 phosphate: 2.94. NH1 Arg-monomer-B to
610 O3´T+1: 3.07 Å; NH2 Arg-monomer-B to T+2 phosphate: 2.89 Å. C) NH1 and NH2 Arg-
611 monomer- A to G-1 phosphate: 2.37 and 2.78 Å, respectively; NH2 Arg-monomer-B to T+2
612 phosphate: 2.91 Å. D) NH1 Arg-monomer-A to G-1 phosphate: 2.95 Å; NH1 Arg-monomer-
613 A to O3´A0: 3.18 Å; NH2 Arg-monomer-A to -O3´T+2: 2.79 Å; NE Arg-monomer-A to T+3
614 phosphate: 3.0 Å. NH2 Arg-monomer-B to T+2 phosphate: 2.76 Å. E) NH1 Arg-monomer-A
615 to G-1 O3´: 2.91 Å; NH1 Arg-monomer-A to G-1 O4´: 2.96 Å; NH2 Arg-monomer-A to O3´
616 T+2: 2.81 Å; NE Arg-monomer-A to T+3 phosphate: 2.71 Å. NH1 Arg-monomer-B to T+2
617 phosphate: 2.79 Å; NH2 Arg-monomer-B to O3´ T+1: 2.78 Å. In F) the NE group of Arg-
618 monomer-A is at 2.72 Å of T+3 phosphate. The NH2 group of Arg-monomer-B is at 2.75 Å of
619 T+2 phosphate and the NE is at 2.75 Å of T-2 phosphate. In G) the NH2 of Arg-monomer-A
620 is at 2.94 Å of T-2 phosphate and NE is at 2.67 Å of T+2 phosphate. The NH1 and NH2
621 groups of Arg-monomer-B are at 2.97 and 2.89 Å of C-3 phosphate.
622
623 **Figure 4**. Protein p4 binding to modified DNA measured in band-shift assays. A) Sequence of
624 p4 binding sites and its derivative mutants aligned with respect to the central position
625 (position 0), with positions ±13 indicated. External guanines required for target recognition
626 are underlined and the site 3 A-tracts are enclosed in boxes. Sequence modifications are in
627 bold with the corresponding mutant name at the left. B) Acrylamide gels of the protein p4wt
628 binding to each modified sequence with the name of the mutant under the gel. Positions of the
629 p4-DNA complexes and free DNA are denoted at the right side.
630
631 **Figure 5**. Transcription regulation of promoters A2c, A2b and A3 by proteins p4wt, p4K51A
632 and p4R54A. The protein used in each assay is indicated above the autoradiograph.
633 Concentration of the proteins was: p4wt, 160 nM; p4K51A and p4R54A, 320 and 640 nM,

634 respectively; p6, 5 or 10 μM; RNAP, 11 nM. The transcripts corresponding to each promoter
635 are indicated at the left.
636
637 **Figure 6**. DNA bent induced by p4. Data from the band-sift assays of the p4-DNA complexes
638 using circularly permuted DNA fragments. Top lane, diagram of the 72 bp sequence
639 containing p4 binding Sites 1 and 2 cloned on the polylinker of plasmid pBend2 with the
640 restriction enzymes organization around the cloned DNA where the number corresponds to:
641 BamHI (1), SspI (2), NruI (3), StuI (4), SmaI (5), PvuII (6), EcoRV (7), DraI (8), XhoI (9),
642 SpeI (10), NheI (11) and BglII (12). Proteins p4 (160 nM) or p4R54A (160 nM) and purified
643 circularly permuted DNA fragments were incubated for 15 min at 4 ºC and loaded onto 4%
644 acrylamide gels. The relative gel mobilities of the p4-DNA complexes were plotted against
645 the position of restriction sites in the DNA fragment. Bending angles are shown.
646
647 **Figure 7**. Interaction of p4R54A with RNA polymerase and with p6. A) Closed complex
648 formation at promoter A3 in the presence of p4 analysed in band shift assays. Complexes were
649 formed at 4 ºC using the DNA fragment containing promoters A2b and A3, and RNAP and p4
650 wild type or the p4R54A mutant. Nucleoprotein complexes resolved on 4% acrylamide gels
651 are indicated at the side. B) Band-shift assays of the ternary p4-p6-DNA and p4R54A-p6-
652 DNA complexes. Proteins were incubated with a labelled DNA fragment containing p4
653 binding sites 1 to 4. Complexes were resolved on a 4% acrylamide native gel. The position of
654 the complexes are denoted.
655
656 **Figure 8**. Protein p4-DNA binding mechanism: Zipper model. (A) Initial interaction of the
657 hook of monomer-B with the corresponding DNA sequence (5´-AAAATG-3´) is indicated by
658 an orange arrowhead; (B) subsequent interactions between basic residues of both monomers

659 with the central minor groove are indicated as orange arrowheads; (C) the progressive bend of
660 the DNA would approximate the 5'-AAAAAG-3' end to the hook of monomer-A permitting
661 the final interactions between the other monomer hook with the sequence (D).
662
663 **Table I**. Oligonucleotides designed for this study. Pairs of oligonucleotides p4dir and p4rev
664 were designed for cloning and expression of p4. Oligonucleotides 1 and 4 were used to obtain
665 the DNA fragment containing promoters A2c, A2b and A3. Oligonucleotides pBend-1-SalI
666 and pBend-2-SalI were used to clone p4 sites 1 and 2 in plasmid pBend. Oligonucleotides
667 5063 and 5164 were used to obtain the DNA fragment containing promoters A3 and A2b.
668 Other oligonucleotides show the sequences of one of the DNA strands of the p4 binding sites
669 or the mutated binding sites used. Each oligonucleotide was hybridized with its
670 complementary counterpart to generate the double strand sequence used in the binding assays.