DNA sequence-specific recognition by a transcriptional regulator requires indirect readout of A-tracts

Jesús Mendieta, Laura Pérez-Lago, Margarita Salas and Ana Camacho*

Instituto de Biología Molecular 'Eladio Viñuela' (CSIC), Centro de Biología Molecular 'Severo Ochoa' (CSIC-UAM), Universidad Autónoma, Canto Blanco, 28049 Madrid, Spain

ABSTRACT

The bacteriophage Ø29 transcriptional regulator p4 binds to promoters of different intrinsic activities. The p4–DNA complex contains two identical protomers that make similar interactions with the target sequence 5′-AACTTTTT-15 bp-AAAATGTT-3′. To define how the various elements in the target sequence contribute to p4's affinity, we studied p4 binding to a series of mutated binding sites. The binding specificity depends critically on base pairs of the target sequence through both direct as well as indirect readout. There is only one specific contact between a base and an amino acid residue; other contacts take place with the phosphate backbone. Alteration of direct amino acid–base contacts, or mutation of non-contacted A.T base pairs at A-tracts abolished binding. We generated three 5 ns molecular dynamics (MD) simulations to investigate the basis for the p4–DNA complex specificity. Recognition is controlled by the protein and depends on DNA dynamic properties. MD results on protein–DNA contacts and the divergence of p4 affinity to modified binding sites reveal an inherent asymmetry, which is required for p4-specific binding and may be crucial for transcription regulation.

INTRODUCTION

Phage Ø29 protein p4, in synergy with viral protein p6, effects the transcriptional switch that divides bacteriophage Ø29 infection into early and late phases (1,2). Protein p4 binds to two regions of DNA, each containing two binding sites in tandem, and each binding site (named sites 1–4) consists of imperfect inverted repeats. The consensus binding sequence is: 5′-AACTTTTT-15 bp-AAAATGTT-3′ (see Figure 1; 3–6). Site 3 is the highest affinity one, followed by sites 1 and 2. Site 4, with

the most imperfect inverted repeat sequence, is the lowest affinity-binding site.

Protein p4 crystallizes as a dimer, and each monomer has an α/β fold and a novel N-terminal β-turn substructure, the N-hook, for DNA interaction (7). In the p4–DNA co-crystal, the DNA presents a B conformation with locally narrowed and widened minor grooves. Each p4 monomer hook intrudes into the DNA major groove where Gln5 and Arg6 establish hydrogen bonds with the DNA bases $T \pm 15$ and $G \pm 13$, respectively (Figure 1). In addition, 14 contacts between the protein dimer and the DNA phosphates are observed. Substitutions of Gln5 and Arg6 with Ala provided evidence that the Arg6-G ± 13 interaction is required for p4-binding site recognition through a direct readout mechanism. Since B-DNA sequences have dissimilar distribution of hydrogen-bond donor and acceptor groups on their bases, but similar sugar–phosphate backbones, p4–phosphate backbone contacts seemed insufficient, *a priori*, to explain p4's affinity and sequence specificity. Therefore, recognition of other aspect of the sequence structure (indirect readout) could also account for p4 sequence specificity. It is known that indirect readout of DNA sequences plays a role in determining the stability and/or specificity of many protein–DNA complexes, but precise knowledge of how the sequence modulates DNA structure is limited.

Although DNA–protein crystal structures identify atomic interactions, they do not always reveal the specificity of those interactions, or the induced conformational changes in the protein and DNA. Here, we studied the structural stability of the protein and DNA in the p4–DNA complex, and the relative roles played by direct and indirect readout in sequence-specific recognition by protein p4. Molecular dynamics (MD) simulations (8) represent a highly suitable tool to explore the dynamic behaviour of p4 and its binding site, and the mechanism of these interactions. Therefore, we analysed p4–DNA interactions by MD simulations, and by

*To whom correspondence should be addressed. Tel: 34-91 497 8435; Fax: 34-91 497 8490; Email: acamacho@cbm.uam.es
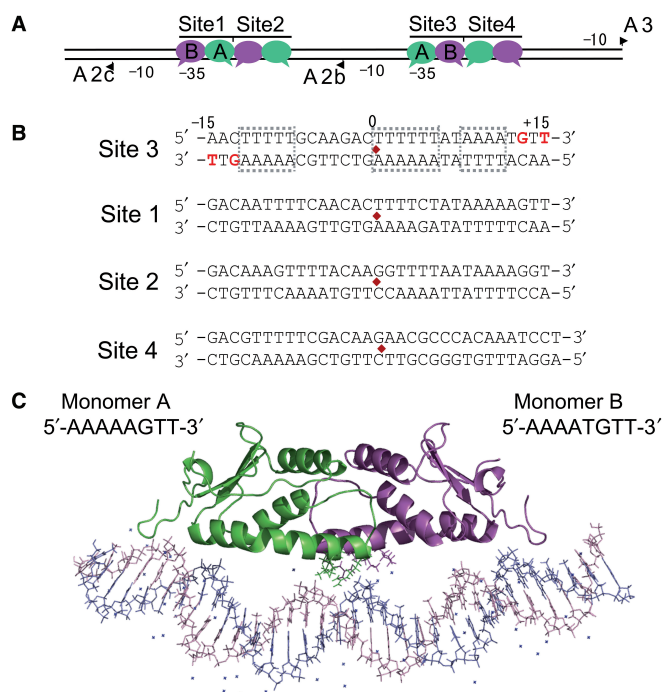
Figure 1. (**A**) Location of the four p4-binding sites with respect to the early (A2b and A2c), and late (A3) bacteriophage Ø29 promoters with the −10 and −35 elements indicated. Protein p4 monomers are drawn as elliptical objects with their N-hook DNA-recognition motifs. Monomers of the p4 dimer are represented in purple and green and each monomer is distinguished as A or B (see text). (**B**) Sequence of the p4-binding sites 1–4 aligned and with the central position of each site indicated by a red diamond. Base pairs corresponding to positions 0 and ±15 are indicated. The guanines and thymines referred to in this article are in red. The three site 3 sequence A-tracts are enclosed in discontinuous lines boxes. (**C**) Structure of the protein p4 in complex with the 34-bp site 3 DNA sequence used in the MD simulations.

focussing on the interaction of p4 to sequence-modified binding sites.

MATERIALS AND METHODS

Protein and DNAs

Bacteriophage Ø29 p4 protein was expressed and purified as described (9). The oligodeoxyribonucleotides (Isogen) used are shown in Figure S1. To obtain each 60-bp double-stranded DNA, two complementary oligodeoxyribonucleotides were used. One of the oligodeoxyribonucleotides from each pair was 5′-end labelled using [γ-^{32}P] ATP and T4 polynucleotide kinase (10). 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 oligonucleotides in a 1:10 ratio in 80 μl of 25 mM Tris-HCl (pH 7.5), 200 mM NaCl, heated for 2 min at 90°C, and allowed to cool gradually (14–24 h) to 20°C.

Band-shift assays

Band-shift assays were performed with a fixed amount of DNA and increasing concentrations of p4 for each of the

DNA substitutions. From the data obtained, experiments were carried out using those p4 concentrations that give rise to a linear response with each DNA. Binding reactions (20 μl) contained labelled DNA (5 fmol), 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 protein p4 at the concentrations indicated in each figure. Incubation was for 15 min at 4°C, and the reaction mixture was loaded onto a non-denaturing 6% polyacrylamide gel after addition of 4% (v/v) glycerol. Electrophoresis was performed at 4°C at 20 mA/gel. Gels were dried, and the label present on free DNA and p4–DNA complexes, quantified in a GS-710 Imaging Densitometer, gave the total amount of DNA; the amount of DNA complexed with the protein was calculated as a fraction of total DNA.

MD simulations

MD simulations were performed using the PMEMD module of AMBER8 and the parm 99 parameter set (11). The X-ray structure (PDB code 2FIO) was used for the MD simulation. The system includes the two p4 monomers (except residues 114–124 corresponding to the flexible helix α4) and the DNA molecule with the following sequence: 5′-TAACTTTTTGCAAGACTTTT TTATAAAATGTTGA-3′. Independent simulations were carried out on the p4–DNA complex (Native complex) and the unbound form of the DNA derived from the structure of the complex but devoid of the protein (free DNA). A third simulation was carried out on the p4–DNA complex where the Arg6 on each monomer of the protein was substituted to Ala (R6A complex). An adequate number of Na$^+$ ions were added to neutralize the net negative charge of the systems (57 Na$^+$ ions in the native system, 65 Na$^+$ ions in the free DNA system and 59 Na$^+$ ions in the R6A system). The counterions were placed in a shell around the system using a Coulombic potential in a grid. The 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. The counterions and the solvent were added using LEAP module of AMBER. 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 298K in a 200-ps heating phase, and velocities were reassigned at each new temperature according to the 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 constrained. This constraint with a force constant of 10 kcal mol^{-1} Å^{-2} was maintained during the equilibration steps to impede a spurious disorganization of the structure during the heating of the system from 0 to 298K. During the last 100 ps of the equilibration phase of the MD, the force constant was reduced stepwise until 0 except for the distance corresponding to the Watson–Crick hydrogen bonds between the first and the last base pairs of the DNA

molecule which was maintained in order to mimic the cooperative stabilizing effect of base pairs present at both DNA ends that are not included in our system. This constraint was maintained during the productive phase of the simulations. The SHAKE algorithm was used throughout to constraint all hydrogen bonds to their equilibrium values so that an integration time step of 2 fs could be employed. The list of non-bonded pairs was updated every 25 steps, and coordinates were saved every 2 ps. Periodic boundary conditions were applied and electrostatic interactions were represented using the smooth particle mesh Ewald method with a grid spacing of ~1 Å. The trajectory length was 5 ns for all the complexes. Analysis of the trajectories was performed using the CARNAL module of AMBER 8.

RESULTS

Affinity of p4 for synthetic binding sites with base substitutions at the A-tracts

The fact that only two bases, G13 and T15, on the target site are bonded by protein p4 suggests that sequence-specific recognition by the protein may involve factors other than direct amino acid side chain–base interactions. The protein contacts backbone phosphates neighbouring three A-tracts (Figure 1). A-tracts have special properties, including bifurcated hydrogen bonds, high propeller twist and buckle of base pairs, suitable for protein recognition (12–15).
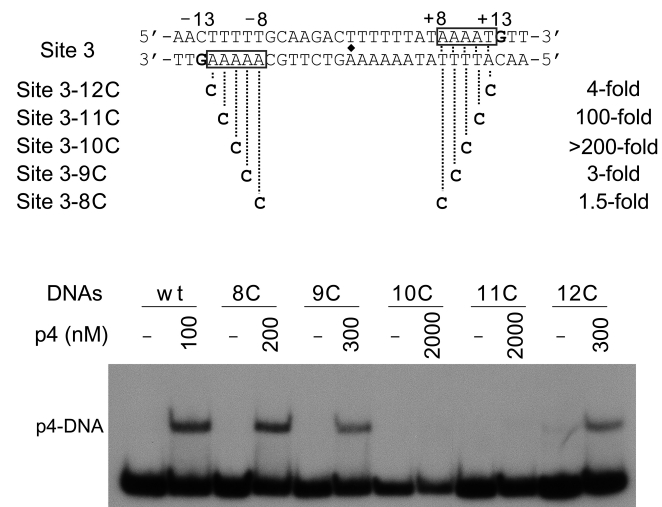

Figure 2. Band-shift assays of p4 and DNA sequences with modified base pairs in the external A-tracts. The site 3 sequence with the two external A-tracts in boxes and base pair at positions ±8 and ±13 is shown. Modified sites 3 assayed are summarized below with the values corresponding to the decrease of relative affinities indicated. Note that a C·G base pair substituted each A·T base pair on both external A-tracts except at position +12 where C·G is substituting a T·A base pair. The p4–DNA complex was formed, and resolved from free DNA through a polyacrylamide gel. The ratio between bound and total DNA was used to calculate the relative affinity of p4 for each modified site 3. The concentration of p4 (in nM dimer) used is indicated on top of the gel.

To elucidate the molecular basis of p4 selectivity for its cognate sequence, we studied the binding affinity of p4 to site 3 sequences containing base modifications at the A-tracts (Figures 2–4). The ability of p4 to bind to modified site 3 bearing C·G base pairs substituting each A·T base pair on the external A-tracts adjacent to G ± 13 was analysed by band-shift assays (Figure 2). Base-pair substitutions at positions ±8 did not affect p4 binding, while the amount of complex formed was reduced by 3- or 4-fold when the base pair was modified at positions ±9 and ±12, respectively. Interestingly, p4 binding was drastically impaired when C·G substituted the A·T base
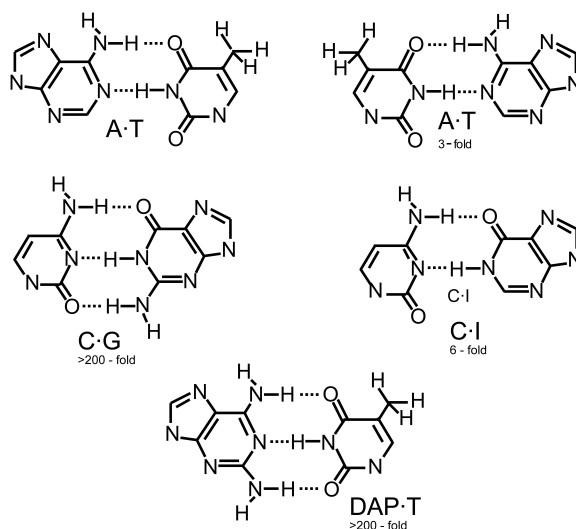

Figure 3. Effect of base substitutions at site 3 position ±10 by nucleotide analogues on the p4 affinity for site 3. The substitution at positions ±10 was made symmetrically, such that the modified base appears at identical positions on both strands. The decrease in the relative affinity of p4 for each modified site 3 is indicated.
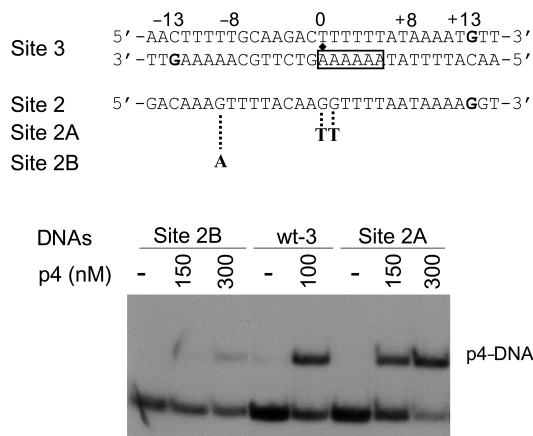

Figure 4. Band-shift assays of p4 with DNA sequences with modified base pairs in the A-tracts of site 2. The site 3 sequence with the central A-tract boxed and the central position denoted by a diamond is shown with the base pairs at positions ±13, ±8 and 0 indicated. The site 2 sequence and the modifications of the sequence are shown below. Other conditions as in Figure 2.

pairs at the centre of the A-tracts (positions ±10 or ±11); a faint band of p4–DNA complex was produced with the DNA modified at positions 11 and no p4–DNA complex was detected when the modification was at position 10. Thus, the affinity decreased 100-fold for substitutions at positions ±11, and >200-fold when the substitution resided at position ± 10. Therefore, the A·T ± 10 and the A·T ±11 base pairs are critical for p4 binding. Since there are no bases or phosphates contacted by the protein at these positions, the effect should be due to specific sequence recognition by indirect readout.

The failure of p4 binding to site 3 bearing C·G substitutions at positions ±10 and ±11 could be explained if, when present, the guanine amino group exerts a detrimental effect on p4 binding through the formation of a third hydrogen bond with the pairing cytosine, which renders the base pair less deformable than the A·T base pair. Alternatively, the 2-amino group of guanine in the minor groove may sterically or electrostatically interfere with p4–DNA interaction. If the presence of the amino group in the minor groove and the increase in the number of hydrogen bonds between the bases interfered with p4 binding, the affinity of p4 for DNA would be drastically decreased upon substitution of adenine with the base analogue diaminopurine (DAP). DAP substitution at positions ±10 reduced p4 binding >200-fold (Figure 3). Conversely, removal of the extra hydrogen bond and the amino group using C·I or T·A base pairs at positions ±10 greatly reduced the deleterious effect (Figure 3). These latter substitutions decreased the affinity of p4 for the DNA by only 3-fold for T·A or 6-fold for C·I. The C·I base pair mimics the A·T base pair in the minor groove, while being identical to the C·G base pair in the major groove (16,17).

Protein p4 has ∼8-fold lower relative affinity for site 2 than for site 3 (6). There are two main differences at the A-tract level: site 2 has only three adenines in one of the external repeats, and its central A-tract is shorter with respect to the homologous site 3 A-tract (Figure 4). To analyse further the role of the external A-tracts, and to elucidate the relevance of the central A-tract, we studied the binding affinity of p4 to the site 2 sequence modified at its A-tracts. The relative affinity of p4 for site 2 did not substantially increase by substituting the G·C base pair by A·T at position −9 (Figure 4, Site 2B), while enlarging the central A-tract by substitutions at positions 0 and +1 (Figure 4, Site 2A) yielded an amount of p4–DNA complex similar to that formed with the site 3 sequence. Therefore, and in agreement with the data from modified site 3, an A·T base pair is not critical at position −9. However, when present at positions 0 and +1, it enhances p4 binding.

MD simulations of p4–DNA complexes

In order to gain insight into the solution conformation of the DNA and the induced conformational changes in the protein and DNA upon complex formation, we explored the p4–DNA complex by MD simulations. Simulations were carried out after standard structural equilibration. The three structures studied are referred to as the native complex (Figure 1C), the p4R6A complex where the

protein has been modified by changing Arg6 into Ala, and the free DNA dynamics of the site 3 sequence with the conformational modification of the p4–DNA crystal structure but devoid of p4.

Stability of the native complex. Cα atoms root-mean-square-deviation (RMSD) values versus simulation time of DNA and p4 in the complexes were tested. Figure 5 provides a quantitative view of these parameters. The most remarkable feature in the p4–DNA (native) trajectory along the simulation is the consistency of the protein conformation with the X-ray structure and its stability over the 5 ns simulation. Indeed, the structural fluctuations did not exceed 1.5 Å. The RMSD values of the DNA stabilized after 1.5 ns to reach peaks of up to 3.5 Å. These RMSD variations reflect dynamic conformations of the DNA and, since neither is followed by alteration of the protein, the DNA must adapt to fit the protein.

The influence of p4 on the conformational adaptation of the DNA in the complex can be studied by following the time course of the MD simulation of free DNA starting from the structure obtained from crystallography, and allowing the DNA to relax to the solution-free form. Conformational changes from DNA bound to p4 to the unbound form are represented in the free DNA simulation of Figure 5, where initial values lower than 2 Å soon reached values up to 8 Å. These results suggested that the protein imposed the DNA conformation in the crystal, if so, an intermediate situation should be found in the p4R6A–DNA complex, where the Arg6–G13 interactions
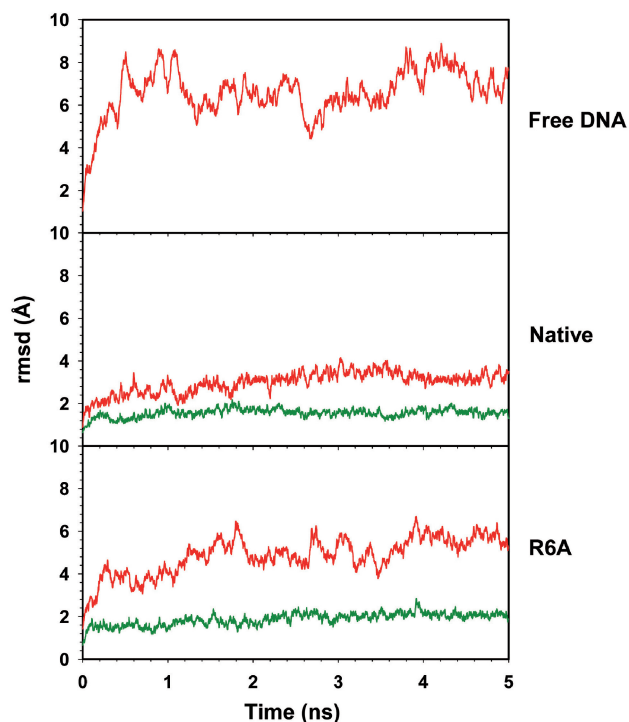

Figure 5. Root-mean-square-deviation (RMSD) in angstroms of MD snapshots of the protein Cα trace (green lines), and the DNA phosphates (red lines) as a function of the simulation time. The different panels are, from top to bottom: free DNA, native protein p4-complex and p4R6A complex.

of both p4 monomers are abolished. The trajectory of the protein in the p4R6A–DNA complex reflects, as was the case of the native complex, the stability of the protein with RMSD values lower than 1.5 Å (Figure 5). However, the RMSD for the DNA reached values higher than those of the native complex (up to 6 Å), as would be expected if some of the protein-dependent constraints were released. Therefore, we focussed our analysis on the obvious changes of the DNA conformation and the consequences of the protein-imposed structural restriction to the DNA.

Minor groove width and curvature. At the three areas of amino acid–phosphate contacts, the minor groove facing p4 narrowed from the 11.5 Å of a regular B-DNA width, while it widened up to 15 Å on the opposite face of the minor groove (Figure 1C). These effects agree with the final structure of the DNA curved towards the protein. To verify the protein-induced structural features (minor groove narrowing and DNA bending) imposed on the DNA, and to determine how much is intrinsic to the sequence-dependent structure of the free DNA, we analysed the degree of DNA curvature and minor groove width variation along each MD simulation (Figure 6). The curvature of the DNA was calculated using the program *Curves* (18). During the MD simulation of the free DNA, the bent structure evolved rapidly into a low curvature (with a mean value of $28.1 \pm 8.1°$). In contrast, the p4-bound form of the DNA (native), with an initial bend angle of 54.2°, reached a curvature mean value of $68 \pm 7.5°$ (Figure S2). Since the free DNA relaxes to a significantly different structural form, the DNA conformation in the complex is an unstable, energetically strained form, and not a stable substrate, that is only stabilized after interaction with p4.

Minor groove parameters were determined measuring the distances between phosphates of T10 and T14, T9 and G13, and T8 and A12 or T12, which correspond to marked width minima. In the native complex, the minor groove width was maintained at 8–9 Å along the 5 ns trajectory at both external A-tracts (for simplicity, we assign monomer A to the sequence 5′-AAAAAGTT-3′ and monomer B to the sequence 5′-AAAATGTT-3′, see Figure 1), the narrowest point being between T9 and G13 (Figure 6, native green lines). In contrast, the free DNA trajectory displayed values consistent with the minor groove width of B-DNA, ~12 Å for the phosphates on the sequence of monomer A and between 10 and 12 Å for the phosphates on the sequence of monomer B (Figure 6, free DNA). Hence, p4 generated local narrowing of the minor groove and this induced conformation may differ upon disruption of the Arg6–G13 interactions. The measurements in the p4R6A trajectory showed variation in minor groove widths depending on the distances and on the sequence measured. On the sequence of monomer B (5′-AAAATGTT-3′), the distances between phosphates 8 and 12 (yellow line) or 9 and 13 (green line) were maintained at 10 Å along the trajectory, while the distance between phosphates 10 and 14 (red line) reached peaks up to 13 Å, which may correlate with the peaks on the DNA RMSD of Figure 5. In contrast, phosphate distances measured on the sequence of monomer A (5′-AAAAAGTT-3′) were uneven. Minor groove width between phosphates 9 and 13 (green line) was maintained at ~9 Å, between phosphates 8 and 12 (yellow line) the width varied between 10 and 12 Å, and the distance between phosphates 10 and 14 (red line) reached up to 16 Å, which followed its DNA RMSD profile (Figure 5). Therefore, Arg6 bonds are required to maintain the DNA conformation, and the data additionally suggest that the asymmetry of the DNA inverted repeats influences
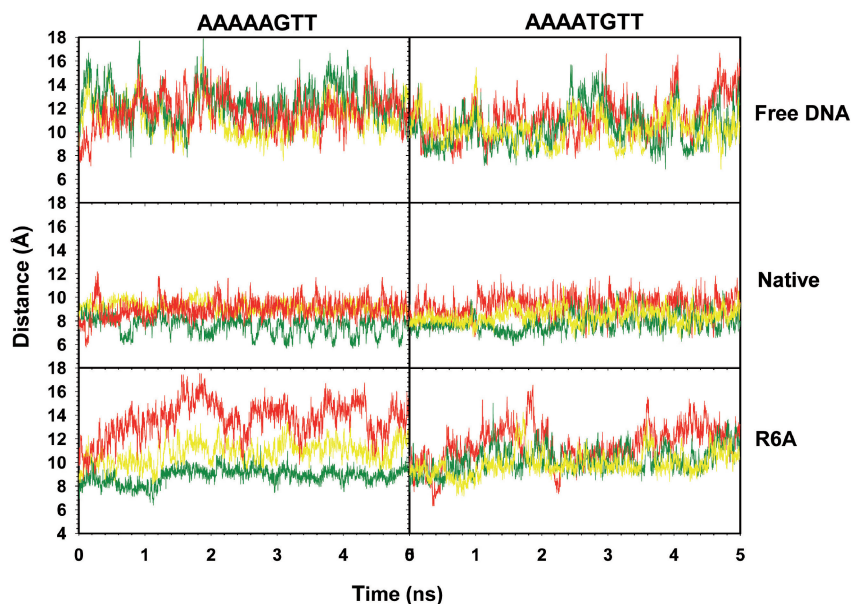

Figure 6. Minor groove width measurements as a function of time. From top to bottom, dynamics of free DNA, of native complex and of p4R6A complex. Left graphs represent measurements between the phosphates of the sequence 5′-AAAAAGTT-3′. Right graphs represent measurements on the sequence 5′-AAAATGTT-3′ (Figure 1). Red lines represent the distances between the T14 and T10 phosphates, green lines show the distances between the G13 and T9 phosphates and yellow lines are the distances between the phosphates of A or T at position 12 and T8.

differentially its interactions with the 2-fold symmetric protein.

Protein–DNA interactions. Since the conformational variation of the DNA is not followed by alterations of the protein, the interactions with the protein must influence
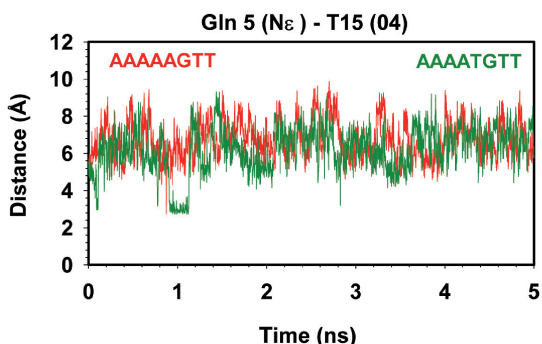

Figure 7. Dynamic profiles of the inter-atomic distance between atom Nε of Gln5 and the atom O4 of T15 as a function of simulation time (nanoseconds). Red line corresponds to the interaction of protein monomer A, green line represents the contact of protein monomer B. In the crystal, the bond distance in monomers A and B are 3.32 and 3.04 Å, respectively.

the changes in DNA. We addressed this question by measuring the distance of protein and DNA atoms involved in interactions in the protein–DNA X-ray structure along the MD trajectories. Each p4 monomer established two types of protein–DNA contacts: amino acid–base interactions (Gln5–T15 and Arg6–G13), and amino acid–backbone interactions (Thr4 and His10 with the G13 phosphate, Tyr33 with the T8 phosphate, Lys36 of monomer A with the G7 phosphate, and Lys36 of monomer B with the A7 phosphate). Mutation of Gln5 to Ala had no effect on, or even improved, p4 binding to DNA (7). To analyse further the Gln5–T15 interaction, we measured the distance between the atoms involved in the MD trajectory of the native complex (Figure 7). From the beginning of the trajectory, and in most of the snapshots, the distances from the Nε group of Gln5 to the O4 of T15 on both monomers were longer than 6 Å, too large for hydrogen bonding, suggesting that this interaction is not necessary to stabilize the p4–DNA complex. In agreement with this result, we found that changing the G·C base pair at position −15 of binding site 1 to A·T (see Figure 1) did not favour p4 binding (results not shown).

In contrast, the amino acid–base interaction Arg6–G13 and the interaction of Thr4 with the phosphate at position ±13 were maintained at ∼3 Å in both monomers along the native complex trajectory (Figure 8A).
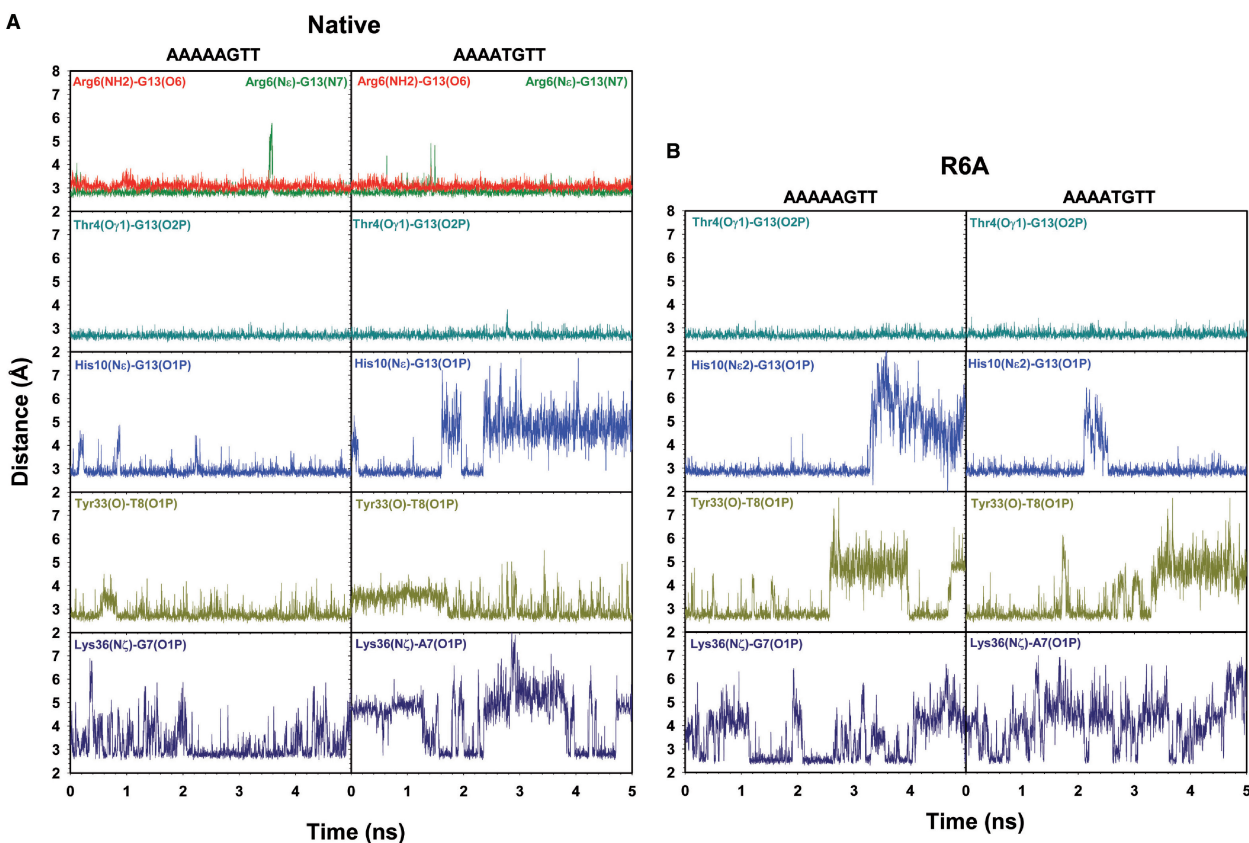

Figure 8. Protein–DNA bond distances along the MD trajectory of the native and p4R6A complexes. (**A**) Dynamics of the p4–DNA native complex. (**B**) Dynamics of the p4R6A–DNA complex. Left, distance measurements between residues of p4 monomer A and the sequence 5′-AAAAAGTT-3′; right, distance measurements between residues of monomer B and the sequence 5′-AAAATGTT-3′. Bond distances measured were: from the NH_2 and Nε atoms of Arg6 to atoms O6 and N7 of ±13G (red and green lines, respectively); from atom Oγ1 of Thr4 to O2P of ±13G (marine); from atom Nε of His10 to O1P of ±13G (blue); from O of Tyr33 to O1P of ±8T (light green); from Nζ of Lys36 to O1P of −7G or +7A (dark blue).

The hydrogen bonds connecting Tyr33 with phosphate 8 were present in ~50% of the trajectory of both monomers. Furthermore, the overall average interactions between the DNA with His10 and Lys36 appeared in a lower percentage of the snapshots of the trajectory of monomer B compared to those of monomer A, where the interaction with Lys36 fluctuates considerably (Figure 8A). Hydrogen bonding between His10 and O1P at positions ±13 and Lys36 with O1P at positions ±7 occurred in a high percentage of the trajectory for the monomer A, but were only transiently present on monomer B. Inspection of the p4R6A complex (Figure 8B), where the Arg6:G13 contacts were abolished, showed stable and persistent hydrogen bonding between Thr4 and the O2P at positions ±13 while the remaining protein–DNA interactions reached values over 3 Å in a high percentage of the snapshots in both monomers after the stabilization time. The distances measured with O2P indicate that there is no switching of partners (not shown). Since a complete protein–DNA dissociation process lies beyond our time-scale, this process would not be detected; the described features display a possible beginning step of the dissociation process. From the results obtained, we conclude that Thr4 is an important stabilizing element of the p4–DNA complex and, while monomer A–DNA contacts remain more constant, contacts on monomer B present fluctuations.

Functional asymmetry on protein binding

To study the functional difference of the sequences contacted by each protein monomer, we analysed p4 binding to site 3 with symmetric inverted repeats. Figure 9 shows that protein p4 has higher affinity (4-fold) for site 3 bearing the sequence 5′-AAAATGTT-3′ (B) than that for
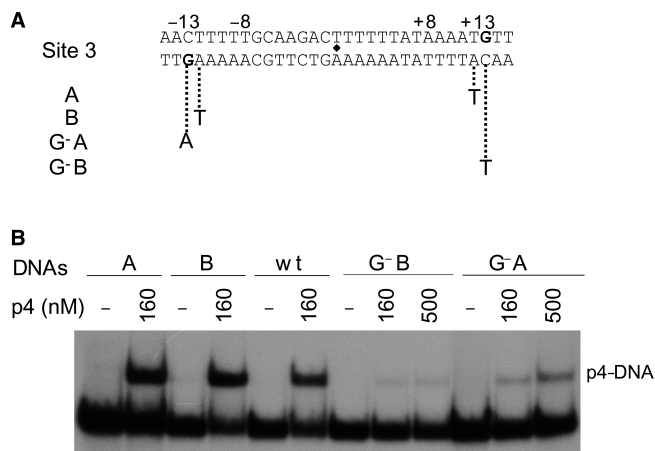

Figure 9. Specificity of p4 for site 3 modified at the inverted repeat sequences. Band-shift experiment of p4 with site 3 sequences with symmetry on the inverted repeats of the external A-tracts (**A** and **B**), or site 3 devoid of one or the other guanine at positions 13 of the inverted repeats (G⁻A and G⁻B). Mutant A has the sequence 5′-AAAAAGTT-3′ at both inverted repeats, while mutant B has the sequence 5′-AAAATGTT-3′ at its inverted repeats. Mutant G⁻A has modified the guanine at position −13 by adenine and in mutant G⁻B the guanine modified by adenine was at position +13. The concentration of p4 (in nM) used is indicated on top of the gel.

site 3 with the sequence 5′-AACTTTTT-3′ (A) in both external A-tracts. To corroborate this data, we substituted individually each guanine at positions ±13 by adenines. Even if p4-binding affinity was clearly diminished when either of these guanines was removed we found that the amount of complex formed with a site 3 devoid of guanine on the monomer A sequence (position −13; G⁻A) was reduced 10-fold. In contrast, there was a greater than 30-fold decrease when the guanine was mutated on monomer B sequence (position +13; G⁻B). In agreement with this finding, single substitution of DAP at position +10 diminished >10-fold the relative affinity of the protein for DNA, while binding was diminished 5-fold when DAP was present only at position −10 (not shown). These data and the MD simulations trajectories support the conclusion that each protein monomer displays different binding entropies due to the slight asymmetry of the inverted repeats on site 3.

DISCUSSION

Protein p4 is a transcriptional regulator that binds to four target sites with different relative affinities (Figure 1). In this article, we studied the principles determining p4 binding specificity. A structural study of the p4–DNA complex by X-ray crystallography showed two direct base–amino acid interactions; Arg6 and Gln5 of each protein monomer contact G13 and T15, respectively. However (i) alanine substitution of each amino acid showed that only the Arg6–G13 interaction is required for p4 binding; (ii) substitution of either guanine at position 13 significantly reduced the relative affinity of p4 for DNA, and simultaneous mutation of both guanines abolished p4 binding (Figures 9 and S4); (iii) MD simulation on the p4–DNA complex showed stable contacts along the trajectory between the Arg6 of both monomers and the guanines at position ±13, while the distances from the Nε of Gln5 of either monomer to the O4 of T15 are excessively large for hydrogen bonding in >90% of the trajectory. It is conceivable that the interaction of Gln5 with T15 in the X-ray structure resulted from the bend of the DNA toward the protein. Hence, specific sequence recognition by direct readout relies exclusively on Arg6–G13 interactions. Taking into account that two guanines on complementary DNA strands separated by 25 bp is a frequent event along the Ø29 genome while p4 binds specifically only to the four target sequences depicted in Figure 1, other characteristics of the p4-binding site sequence should contribute, in addition, to the p4-binding specificity.

Based on the X-ray structure, the p4–DNA complex with the sequence of site 3 includes three patches of amino acids and three A-tracts. Those A-tracts are present in the other p4-binding sites, as well as in the binding sites of the p4 homologous protein of Nf, a phage closely related to Ø29 (19). Residues Thr4 and His10 contact the Ø29 site 3 DNA backbone precisely at one border of the two external A-tracts, with Tyr33 and Lys36 contacting the opposite border of the A-tracts. Substitution of A·T with C·G at each base pair of those A-tracts influences or abolishes the

affinity of p4 for DNA. Our findings show that placing an amino group on the minor groove of the base pair located at the centre of the A-track (positions ±10) abolishes p4–DNA complex formation. The A·T base pair exhibits a higher intra-base pair propeller deformation than C·G or DAP·T base pairs, a consequence of two Watson–Crick hydrogen bonds between base pairs rather than three. DNA bearing the amino group in the minor groove is under-wound with respect to sites lacking this group, which may directly impact the DNA helical twist and twisting flexibility by mechanical occlusion. Either way, the main negative effect results in a less deformable base pair and, therefore, an increase in the energy required to distort the DNA. Furthermore, A·T or T·A base pairs at position 10 allow the DNA to be more easily bent in the direction of the minor groove that faces the protein at this position. Base substitutions at the A-tracts assert both sequence recognition by indirect readout and p4 over-winding of the minor groove at the centre of the A-tracks.

Examination of the unbound and bound DNA conformations by MD provides clues as to the dynamics of the system. The absolute value of the RMSD for DNA in the p4–DNA complex, lower than 2 Å, can be considered small taking into account the large size of the simulated system. However, the RMSD values for DNA have a higher mean value (up to 8 Å) in the free DNA MD, demonstrating considerable DNA immobilization mediated by p4 binding. The RMSD values of the protein with respect to the starting value have a constant mean value of ~1.5 Å along the trajectory, suggesting high stability of the protein moiety in the complex. On the other hand, comparison of the free and DNA-bound crystal structures of p4 does not reveal significant structural changes upon p4 binding with a C$^\alpha$ RMSD of 0.665 Å (7). The DNA in the p4–DNA crystal structure is curved towards the protein due to local compression of the minor grooves at the areas where the protein contacts the backbone phosphates. In the MD simulation, the stable form of free DNA released from p4 does not show overall bending or a stable local narrow minor groove. In contrast, the DNA bound to p4 displays a stably narrowed minor groove between T10 and T14, T9 and G13, and T8 and A12 or T12 phosphates. Since in the absence of p4 the free DNA relaxes to a significantly different conformation, the structure of the DNA in the complex is not a stable or meta-stable substrate but a consequence of the induced conformational modification impressed by p4, which in turn does not reveal significant structural changes upon DNA binding. In fact, p4 binding does not require intrinsically bent DNA (6). The MD data, the disadvantage of sequences containing disrupted A-tracts (this article) for p4 binding to DNA, and the drastic reduction on binding affinity of p4 Thr4Ala and Tyr33Ala mutants (7) indicate that the sequence-dependent characteristic of A-tracts provides an indirect readout by affecting the optimal complementarity both for amino acid–base hydrogen bonding and for precisely positioned interactions between amino acid and DNA phosphates. Hence, the stability of p4–DNA complex is a delicate balance of direct and indirect readout.

The collective results of this study indicate that p4–DNA-binding stability is a consequence of p4-induced conformational modification of the DNA from the canonical B form through an indirect readout mechanism whereas the primary function of the DNA is its ability to acquire a conformation capable of enhancing positive interactions with its cognate protein.

Indirect readout is less well characterized than direct readout. The affinity of protein for its DNA target by indirect readout relies on the fact that B-DNA exhibits a high degree of sequence-dependent structural variation (20–25) which includes recognition of aspects of DNA structure such as intrinsic curvature, topology of major and minor grooves, local geometry of backbone phosphates and flexibility or deformability. These mechanisms have been used to explain some aspects of the affinity of other prokaryotic transcriptional regulators for its target sequences; CAP seems to discriminate between a consensus pyrimidine–purine steps involving sequence effects on the energetic of primary-kink formation (26,27), while bacteriophage 343 repressor recognizes structural features on the central base pair of its target sequence (28,29), and water-mediated contacts are known as an important recognition tool in the trp-repressor operator system (30,31).

The complexity of p4 interaction with its target sequence is compounded by the fact that, despite the 2-fold symmetry of the protein dimer, the protein uses pseudo-inverted repeats to interact with DNA, one monomer (monomer A) interacting with the sequence 5′-AAAAAGTT-3′ and the other (monomer B) with the sequence 5′-AAAATGTT-3′. Moreover, p4 is capable of recognizing other sequences with asymmetric inverted repeats (Figure 1). We demonstrated that each inverted repeat provides different contributions to p4-binding affinity. Additionally, the p4–DNA MD simulation indicates that the hydrogen bonds of Tyr33, His10 and Lys36 with DNA are significantly more variable, both in residence time as well as bonding distance on monomer B than on monomer A, suggesting that interactions on monomer B present higher entropic stability. Since the pyrimidine–purine T/G step is more susceptible to deformation than the A/G step because it has a smaller amount of base overlap, the T/G bases of monomer B may permit a better orientation of the G + 13 for its interaction with Arg6 (Figure S3).

Taking into account that DNA deformation is an important component of the driving force for p4–DNA association, the data presented here provide insights into how the role of DNA sequence may influence a directional binding for p4. We consider that: (1) asymmetry is functionally required for p4–DNA interaction; (2) the MD simulations suggest a net order of p4 binding to DNA sites with minor groove narrowing and curving of the helical axis; (3) the distance from G − 13 to G + 13 is of ~90 Å, while the distance from the Arg6 of one of the monomers to Arg6 of the other monomer of 75 Å is too short for simultaneous interaction of both p4 monomers at the inverted repeats. Therefore, we propose a zipper-binding model where one of the p4 monomers interacts first with the higher entropic stability sequence,

5′-AAAATGTT-3′ followed by local minor groove narrowing. This change in DNA conformation will allow interactions between basic residues of both monomers with the central A-tract, and the progressive bend of the DNA would permit the 5′-AAAAAGTT-3′ inverted repeat to interact with the hook of the second p4 monomer.

Phage Ø29 early promoters A2c and A2b and late promoter A3 are coordinately regulated by a multimeric complex of viral proteins p4 and p6, which elicits the switch from early to late transcription repressing promoters A2c and A2b, and simultaneously activating promoter A3. In the multimeric complex, p4 dimers occupy binding sites 1 and 3, and p6 binds the sequence from sites 1 to 3 synergizing p4 binding (2). Since protein p6 polymerizes from the A-track 5′-AAAAAGTT-3′ of site 1 to the A-track 5′-AAAAAGTT-3′ of site 3, a second functional implication for the site asymmetry could explain p6-mediated stabilization of p4 binding by anchoring the p4 monomer to the lower affinity A-track. This stabilization is critical for the regulation of the promoters and so required for the transition between early to late transition during Ø29 development.

SUPPLEMENTARY DATA

Supplementary data are available at NAR Online.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Elías-Arnanz,M. and 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.*, **13**, 2502–2513.
2. Camacho,A. and 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.
3. Barthelemy,I. and Salas,M. (1989) Characterization of a new prokaryotic transcriptional activator and its DNA recognition site. *J. Mol. Biol.*, **208**, 225–232.
4. Nuez,B., Rojo,F. and Salas,M. (1994) Requirement for an A-tract structure at the binding site of phage Ø29 transcriptional activator. *J. Mol. Biol.*, **237**, 175–181.
5. Camacho,A. and Salas,M. (2004) Molecular interplay between RNA polymerase and two transcriptional regulators in promoter switch. *J. Mol. Biol.*, **336**, 357–368.
6. Pérez-Lago,L., Salas,M. and Camacho,A. (2005) A precise DNA bend angle is essential for the function of the phage Ø29 transcriptional regulator. *Nucleic Acids Res.*, **33**, 126–134.
7. Badia,D., Camacho,A., Pérez-Lago,L., Escandón,C., Salas,M. and Coll,M. (2006) The structure of phage Ø29 transcription regulator p4-DNA complex reveals an N-hook motif for DNA. *Mol. Cell*, **22**, 73–81.
8. Hansson,T., Oostenbrink,C. and van Gunsteren,W. (2002) Molecular dynamics simulations. *Curr. Opin. Struct. Biol.*, **12**, 190–196.
9. Barthelemy,I., Lázaro,J.M., Méndez,E., Mellado,R.P. and Salas,M. (1989) Purification in an active form of the Ø29 protein p4 that controls the viral late transcription. *Nucl. Acids Res.*, **15**, 7781–7793.
10. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
11. Case,D.A., Plearman,D.A., Cadwell,J.W., Cheatham,T.E. III, Wang,J., Ross,W., Simmerling,C.L., Darden,T.A., Merz,K.M. *et al.* (2002) AMBER7, University of California, San Francisco. http://www.amber.ucsf.edu/amber/amber.html.
12. Coll,M., Frederick,C.A., Wang,A.H. and Rich,A. (1987) A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl. Acad. Sci. USA*, **84**, 8385–8389.
13. Nelson,H.C., Finch,J.T., Luisi,B.F. and Klug,A. (1987) The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature*, **330**, 221–226.
14. Crothers,D.M. (1998) DNA curvature and deformation in protein-DNA complexes: a step in the right direction. *Proc. Natl. Acad. Sci. USA*, **95**, 15163–15165.
15. Goodsell,D.S. and Dickerson,R.E. (1994) Bending and curvature calculations in B-DNA. *Nucleic Acids Res.*, **22**, 5497–5503.
16. Saenger,W. (1983) *Principles of Nucleic Acids. Advanced Texts in Chemistry*. Springer, New York.
17. Xuan,J.C. and Weber,I.T. (1992) Crystal structure of a B-DNA dodecamer containing inosine, d(CGCIAATTCGCG), at 2.4 Å resolution and its comparison with other B-DNA dodecamers. *Nucleic Acids Res.*, **20**, 5457–5464.
18. Lavery,R. and Sklenar,H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.*, **6**, 655–667.
19. Pérez-Lago,L., Salas,M. and Camacho,A. (2005) Homologies and divergences in the transcription regulatory system of two related *Bacillus subtilis* phages. *J. Bacteriol.*, **187**, 6403–6409.
20. Hogan,M.E. and Austin,R.H. (1987) Importance of DNA stiffness in protein-DNA binding specificity. *Nature*, **329**, 263–266.
21. von Hippel,P.H. (1994) Protein-DNA recognition: new perspectives and underlying themes. *Science*, **263**, 769–770.
22. Harrington,R.E and Winicov,I. (1994) New concepts in protein-DNA recognition: sequence-directed DNA bending and flexibility. *Prog. Nucleic Acid Res. Mol. Biol.*, **47**, 195–270.
23. Dickerson,R.E. and Chiu,T.K. (1997) Helix bending as a factor in protein/DNA recognition. *Biopolymers*, **44**, 361–403.
24. Kalodimos,C.G., Biris,N., Bonvin,A.M., Levandoski,M.M., Guennuegues,M., Boelens,R. and Kaptein,R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science*, **305**, 386–389.
25. Gromiha,M.M., Siebers,J.G., Selvaraj,S., Kono,H. and Sarai,A. (2005) Role of inter and intramolecular interactions in protein-DNA recognition. *Gene*, **364**, 108–113.
26. Gartenberg,M.R. and Crothers,D.M. (1988) DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nature*, **333**, 824–829.
27. Napoli,A.A., Lawson,C.L., Ebright,R.H. and Berman,H.M. (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.
28. Koudelka,G.B. and Carlson,P. (1992) DNA twisting and the effects of non-contacted bases on affinity of 434 operator for 434 repressor. *Nature*, **355**, 89–91.

29. Mauro,S.A., Pawlowski,D. and Koudelka,G.B. (2003) The role of the minor groove substituents in indirect readout of DNA sequence by 434 repressor. *J. Biol. Chem.*, **278**, 12955–12960.

30. Otwinowski,Z., Schevitz,R.W., Zhang,R.G., Lawson,C.L., Joachimiak,A., Marmorstein,R.Q., Luisi,B.F. and Sigler,P.B. (1988) Crystal structure of trp repressor/operator complex at atomic resolution. *Nature*, **335**, 321–329.

31. Bareket-Samish,A., Cohen,I. and Haran,T.E. (1998) Direct versus indirect readout in the interaction of the trp repressor with non-canonical binding sites. *J. Mol. Biol.*, **277**, 1071–1080.