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

Jesús Mendieta,^{a,b} Laura Pérez-Lago,^{a*} Margarita Salas,^a and Ana Camacho^a

Instituto de Biología Molecular Eladio Viñuela (CSIC), Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Universidad Autónoma de Madrid, Madrid, Spain,^a and Biomol-Informatics SL, Parque Científico de Madrid, Cantoblanco, Madrid, Spain^b

A bacteriophage ϕ 29 transcriptional regulator, protein p4, interacts with its DNA target by employing two mechanisms: by direct readout of the chemical signatures of only one DNA base and by 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 used molecular dynamic simulation to define interactions of a cluster of 12 positively 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 non-base-specific 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 findings extend the current understanding of protein-DNA recognition and contribute to the relevance of the sequence-dependent conformational malleability of the DNA, shedding light on the role of arginines in binding affinity. Characterization of mutant p4R54A shows that the residue is required for the activity of the protein as a transcriptional regulator.

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 stability and the 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 ϕ 29, three main promoters, the early promoters A2c and A2b and the late promoter A3, are coordinately regulated by two virus-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'-CTTTT-15 bp-AAAATG-3' (40). In the

structure of p4 bound to site 3 (Fig. 1A), the N-terminal region of p4, folded into a unique substructural motif named N-hook (2), penetrates the DNA major groove, where Arg6 establishes hydrogen bonds with guanines at positions +13 and -13 ($G \pm 13$) (Fig. 1B). There are additional contacts with phosphates in three consecutive narrow minor grooves containing A tracts. In the two externally located A tracts, the N-hook motif residue Thr4 contacts the phosphate of bases at positions ± 12 , while Tyr33, in the N terminus of helix $\alpha 1$, contacts DNA phosphates at position ± 8 across the minor groove. Mutation of Arg6, Thr4, or Tyr33, as well as individual replacements of G at positions ± 13 or replacements of the ± 10 and 11 A·T base pairs with less-deformable base pairs, were deleterious (31). Therefore, p4-specific DNA recognition is driven both by direct readout of $G \pm 13$ and by indirect readout of the external A tracts.

In this work, we studied the contribution of a cluster of surface-exposed positively charged amino acids, particularly Lys51 and Arg54, which 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 interaction strengths (17). Previous results on the dynamic behavior of the

Received 23 April 2012 Accepted 25 June 2012

Published ahead of print 29 June 2012

Address correspondence to Ana Camacho, acamacho@cbm.uam.es.

* Present address: Laura Pérez-Lago, Servicio de Microbiología Clínica y Enfermedades Infecciosas, Hospital General Universitario Gregorio Marañón, Madrid, Spain.

Supplemental material for this article may be found at <http://jb.asm.org/>.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00677-12

FIG 1 (A) Structure of the p4 dimer in complex with the DNA encompassing the sequence of site 3 used in the MD simulations. Monomers of the p4 dimer are represented in green and purple, and each monomer is distinguished as A (green) or B (purple) depending on its position with respect to the sequence of site 3. In stick form and numbered are amino acids involved in DNA recognition and those referred to in this paper. (B) Scheme of site 3, showing the protein-DNA interactions from the 3D structure. The center (position 0) and the extreme locations (positions ± 13) of the sequence are marked, guanines referred to in this paper are in red, and the three A tracts are in boxes. Amino acid interactions with the DNA are signaled with arrows: black arrows for the bidentated Arg-G interactions, red arrows for Thr and Tyr interactions, and green and purple arrows for the interactions studied in this paper of one p4 monomer or the other.

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

MATERIALS AND METHODS

DNAs and oligodeoxyribonucleotides. The sequences of the synthetic oligonucleotides (Isogene Bioscience BV, The Netherlands) used are listed in Table 1. The 368-bp DNA fragment for the *in vitro* transcription assay contains the 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 generated 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 1). One of the oligodeoxyribonucleotides from each pair was 5' end labeled using [γ - 32 P]ATP and T4 polynucleotide kinase (44). The labeled 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 labeled and unlabeled oligonucleotide in a 1:10 ratio in 80 μ l of 25 mM Tris-HCl (pH 7.5), 200 mM NaCl, heated 2 min at 90°C, and allowed to cool gradually to 20°C.

Cloning of gene 4 and purification of proteins p4 and p6. A 388-bp Ø29 genomic fragment encoding the protein p4 was PCR amplified using the primers p4dir and p4rev, obtained from Isogen (Table 1). The PCR fragment, cleaved with NdeI and EcoRI, was gel purified, ligated to the

TABLE 1 Oligonucleotides designed for this study^a

Name	Sequence
p4dir	5'-GTTTTTTCGCTCGACATCATA-3'
p4rev	5'-CGTTTGTCTGACGTCATCTCCTTTC-3'
Primer 1	5'-GATTTCCTCTGTCATCA-3'
Primer 4	5'-CAAAATATCTTCGTGTTCTTC-3'
pBend-1-Sall	5'-CACCTGGTCGACAAATTTTC-3'
pBend-2-Sall	5'-TAAATAGGAACGTCGACAAACACC-3'
5063	5'-TCTACGTTCTCTATTTAACCACACC-3'
5164	5'-TTGATACATAAGGATTGTGGGCGTCTTG-3'
Site3A	5'-CAGATCTCAACTTTTTCGAAGACTTTTATAAAATGTGAGCTAG-3'
Site2a	5'-CAGATCTCAAAGTTTACAAGTTTATAAAAGTTGAGATCTG-3'
Mut2Ca	5'-CAGATCTCAAAGTTTACAATTTTATAAAAGTTGAGATCTG-3'
Site1a	5'-CAGATCTCAATTTTCAACACTTTTCTATAAAAGTTGAGATCTG-3'
Mut1Ba	5'-CAGATCTCAATTTTCAACACTTTTATAAAAGTTGAGATCTG-3'

^a Pairs of oligonucleotides p4dir and p4rev were designed for cloning and expression of p4. Oligonucleotides 1 and 4 were used to obtain the DNA fragment containing promoters A2c, A2b, and A3. Oligonucleotides pBend-1-Sall and pBend-2-Sall were used to clone p4 sites 1 and 2 in plasmid pBend. Oligonucleotides 5063 and 5164 were used to obtain the DNA fragment containing promoters A3 and A2b. Other oligonucleotides show the sequence of one of the DNA strands of the p4 binding sites or the mutated binding sites used. Each oligonucleotide was hybridized with its complementary counterpart to generate the double-strand sequence used in the binding assays.

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 the protein p4 under the control of the T7 RNA polymerase Φ 10 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 previously (4). The purification of p6 was done as described previously (29).

Band shift assays. Band shift assays were performed with a fixed amount of DNA and various concentrations of proteins. Binding reaction mixtures (20 μ l) containing labeled DNA ($\sim$ 100 pM), 25 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 100 mM KCl, 0.5 μ g of poly[d(I-C)], 1 μ g of bovine serum albumin, and the protein(s) were incubated for 15 min at 4°C and analyzed on nondenaturing 6% polyacrylamide gels at 4°C. Gels were dried, and the label present in free DNA and protein-DNA complexes, quantified in a Fuji Bass IIIa 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 previously (40). Briefly, p4 sites 1 and 2 were cloned in plasmid pBend2 (21) using the oligonucleotides pBend-1-SalI and pBend-2-SalI (Table 1). The plasmid was then digested with restriction enzymes to generate DNA fragments having a circular permuted p4 binding sequence. The digested fragments were dephosphorylated with calf intestinal alkaline phosphatase and labeled with [γ -³²P]ATP and T4 polynucleotide kinase. The labeled DNA fragments were incubated under the condition described for the band shift assays, with purified wild-type p4 (p4wt) or mutant p4R54A. The p4-DNA complexes were resolved from free DNA onto a nondenaturing 6% polyacrylamide gel and quantified as described above. The p4-induced DNA bend angles and bend center were calculated by the equation of Thompson and Landy (51).

In vitro transcription assays. Runoff transcription assays (20 μ l) contained 25 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 2 mM dithiothreitol, 2 μ g of poly[d(I-C)], 100 mM KCl, 10 U of RNasin, 100 μ M (each) GTP, ATP, CTP, and [α -³²P]UTP (1 μ Ci), and $\sim$ 5 nM DNA template containing the promoters A2c, A2b and A3. *B. subtilis* σ^A -RNA polymerase (RNAP), purified as described previously (47), was added, and after 20 min at 37°C, reactions were stopped with SDS (0.15%) and EDTA (2.5 mM). Transcripts were analyzed by electrophoresis in 6% denaturing polyacrylamide gels. Quantification of the transcripts was carried out by using a Fuji Bas-IIIa image analyzer.

Molecular dynamics simulations. MD simulations were performed using the PMEMD module of the AMBER11 software program 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 α 4) and double-stranded DNA with the following sequence: 5'-TAACCTTTTGAAGAC TTTTATATAAAATGTTGA-3'. Independent simulations were carried out with the wild-type p4-site 3 complex, the mutated protein p4R6A-site 3 complex, and the wild-type p4-site 3 mutated A $\pm$ 10C. Counter ions and the solvent were added using the 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 cutoff 10.0 Å. Subsequently, and to start the MD simulations, the temperature was raised from 0 to 298 K 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 C- α 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. The C- α trace was maintained with a force constant of 10 kcal mol⁻¹ Å⁻² during the equilibration steps to impede a spurious disorganization of the structure during the heating of the system

from 0 K to 298 K. During the last 100 ps of the equilibration phase of the molecular dynamics, the force constant of all the constraints was reduced stepwise to 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 constrain all hydrogen bonds to their equilibrium values so that an integration time step of 2 fs could be employed. The list of nonbonded 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 about 1 Å. Trajectory lengths of 10 ns were analyzed using the CARNAL module of AMBER11.

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 (Fig. 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) (Fig. 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 of the p4 conformation, where fluctuations on its C- α atom root mean square deviation (RMSD) values did not exceed 1.5 Å. Quite the opposite: the RMSD values of the DNA reaching peaks of up to 3.5 Å reflected dynamic conformations of the DNA, adapted to fit the protein.

The complex referred to as the wild-type p4-site 3 complex (p4wt-DNA) has the atomic coordinates of the three-dimensional (3D) structure. After a preliminary analysis of the protein-DNA interactions, the atomic protein-DNA distances were measured along the 10-ns trajectory (Fig. 2). From the 12 positively charged residues (Fig. 1A), only Lys51 and Arg54 maintained contacts with the DNA along the MD simulation. The distance between the Ne 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 (Fig. 2A). Similarly, the distance between the Ne group of Lys51-monomer B to the phosphate of T+2 was too large (Fig. 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 (Fig. 2A). In contrast, Arg54-monomer B maintained a stable interaction only on one DNA strand and mainly with the T+2 phosphate (Fig. 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 $\sim$ 2 ns of the trajectory (Fig. 3A), superficially positioned over the minor groove and inversely oriented with respect to each other. The position of Arg54-monomer B was quite

FIG 2 Protein-DNA bond distances along the MD trajectory of the p4/site 3 complex (A), protein p4R6A/site 3 complex (B), or p4wt/site 3 A10C complex (C). Distances were measured from the N_{ζ} atom of Lys51 or the C_{ζ} atom of Arg54 of both monomers to the phosphorous atom of the phosphate groups corresponding to the nucleotide numbers shown.

stable and could be superimposed fairly well along the trajectory where the NH_1 and NH_2 groups were alternately contacting the T+2 phosphate (Fig. 3B). In contrast, the residue of monomer A moved occasionally from being superficially positioned toward an asymmetrical location in the groove between residues T+3 and G-1. This conformational change was accompanied by the establishment of hydrogen bonds between Arg54 NH_2 and O-3' of A0, located much deeper in the groove. By nanosecond ~4, the arginine returned to its initial position, where the NH_1 and NH_2 groups alternately contacted the phosphate at position -1 (Fig. 3C). Hence, in spite of being chemically equivalent and in identical monomers, the two arginines have distinguishable DNA interactions.

To gain further insight into the role of arginines 54 in the p4 binding process, we analyzed the behavior 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; Fig. 2B) and the other with wild-type p4 and the A±10C mutated site 3 (p4wt-DNA A10C; Fig. 2C). As previously described, the p4 mutant R6A substitution abolishes the amino acid-base interaction, while the A±10C substitution disrupts the interaction of p4 at the external A tracts (31). The MD simulation of the p4R6A-site 3 complex showed a general weakening of the Arg54-monomer A interactions and a more stable interaction of Lys51-monomer A with the T-2 phosphate. The guanidinium group Arg54-monomer A interacts with the T+3 phosphate in the first 2 ns of the dynamic and with the G-1 phosphate along the entire simulation (Fig. 2B). Along the trajectory, the residue moved from being superficially positioned toward an asymmetrical position in the groove, most frequently between G-1 and T+3

(Fig. 3D and E). This movement was accompanied by the establishment of bonds between the NH_1 and NH_2 groups with the O-3' of A0 and the O-3' of T+2, respectively. Arg54-monomer B maintained stable interaction with the T+2 phosphate (Fig. 2B). Arg54-monomer B localized in the minor groove between T+2/T-2 and T+3/G-1, contacting the T+2 phosphate and the O-3' of T+1 (Fig. 3D and E). In the p4R6A-site 3 trajectory, the linearity of the phosphate backbone presents a local bend, probably due to the insertion of arginines 54; the width of the groove was irregular (see above).

In the MD simulation of the p4wt-DNA A±10C site 3 complex (Fig. 2C), no stable protein-DNA contacts of monomer A were found, and Lys51-monomer B did not contact the DNA either. Interactions between Arg54-monomer B and the phosphates of T+2 and G-1 were stable only for the first ~3 ns. Hence, p4 interactions at the external A tracts seem to influence central minor groove interactions. However, although Arg54-monomer A does not maintain stable contacts with phosphates, it may still contribute to some stabilization of the complex since it becomes occasionally inserted straight into the groove between T+2 and T-2 and intermittently interacts with the phosphates and with O-3' of G-1 (Fig. 3F, and G). We had previously shown that each of the protein monomers displays different binding entropy 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 ~9 Å, while it widened on its opposite face (Fig. 1A), resulting in an unstable, energetically strained

FIG 3 Rearrangement of arginine 54 side chains. Snapshots showing the position of arginines 54 in the MD simulations of the wild-type p4/site 3 complex (A, B, and C), the mutated protein p4R6A/site 3 complex (D and E), or the wild-type p4/A $\pm$ 10C mutated site 3 (F and G). Representative snapshots were selected at 0 ns (A), $\sim$ 2 ns (B, D, and F), and $\sim$ 4 ns (C, E, and G). Arg-monomer A, gray color; Arg-monomer B, green color. Distances are as follows: for panel A, NH₁ Arg-monomer A to G-1 phosphate, 3.16 Å; NH₁ and NH₂ Arg-monomer B to T+2 phosphate, 2.56 Å and 2.75 Å, respectively; for panel B, NH₁ Arg-monomer A to G-1 phosphate, 2.81 Å; NH₂ Arg-monomer A to O-3'A0, 3.08 Å; NE Arg-monomer A to T+3 phosphate, 2.94 Å; NH₁ Arg-monomer B to O-3'T+1, 3.07 Å; NH₂ Arg-monomer B to T+2 phosphate, 2.89 Å; for panel C, NH₁ and NH₂ Arg-monomer A to G-1 phosphate, 2.37 Å and 2.78 Å, respectively; NH₂ Arg-monomer B to T+2 phosphate, 2.91 Å; for panel D, NH₁ Arg-monomer A to G-1 phosphate, 2.95 Å; NH₁ Arg-monomer A to O-3'A0, 3.18 Å; NH₂ Arg-monomer A to O-3'T+2, 2.79 Å; NE Arg-monomer A to T+3 phosphate, 3.0 Å; NH₂ Arg-monomer B to T+2 phosphate, 2.76 Å; for panel E, NH₁ Arg-monomer A to G-1 O-3', 2.91 Å; NH₁ Arg-monomer A to G-1 O-4', 2.96 Å; NH₂ Arg-monomer A to O-3'T+2, 2.81 Å; NE Arg-monomer A to T+3 phosphate, 2.71 Å; NH₁ Arg-monomer B to T+2 phosphate, 2.79 Å; NH₂ Arg-monomer B to O-3'T+1, 2.78 Å; in panel F, the NE group of Arg-monomer A is at 2.72 Å of T+3 phosphate, the NH₂ group of Arg-monomer B is at 2.75 Å of T+2 phosphate, and the NE is at 2.75 Å of T-2 phosphate; In panel G, the NH₂ of Arg-monomer A is at 2.94 Å of T-2 phosphate, and NE is at 2.67 Å of T+2 phosphate. The NH₁ and NH₂ groups of Arg-monomer B are at 2.97 and 2.89 Å of C-3 phosphate.

DNA conformation imposed by p4 (31). Measurements of the interphosphate distances of the central minor grooves in the MD simulations of the wild-type p4-site 3 complex and the p4R6A-site 3 complex are shown in Fig. S1 in the supplemental material. In the course of the p4wt MD, the minor groove was maintained as locally narrow between T0 and T+4 (see Fig. S1), with arginines 54 superficially located across the minor groove, and when inserted into the groove, their most frequent position was between

FIG 4 Protein p4 binding to modified DNA, measured in band shift assays. (Top) Sequences of p4 binding sites and its derivative mutants, aligned with respect to the central position (position 0), with positions $\pm$ 13 indicated. External guanines required for target recognition are underlined, and the site 3 A tracts are enclosed in boxes. Sequence modifications are in bold, with the corresponding mutant name at the left. (Bottom) Acrylamide gels of the protein p4wt binding to each modified sequence, with the name of the mutant under the gel. Positions of the p4-DNA complexes and free DNA are denoted at right.

T+3 and G-1, contacting O-3' of A0 (Fig. 3B). It has been shown that enhanced negative potentials of a narrowed minor groove attract arginines that could mediate specific protein-DNA interactions (41). Here Arg54 is inserted in the narrowed part of the groove ($\sim$ 9-Å width). The p4R6A-site 3 complex showed a manifest local minimum at T0 ($\sim$ 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 behavior when external forces are imposed (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 higher-affinity binding site 3 and is shorter in site 1 (positions 0 to +3) and site 2 (positions +2 to +5) (Fig. 4). Protein p4 binds $\sim$ 2- and $\sim$ 8-fold less efficiently to site 1 and to site 2, respectively, than to site 3 (40). To determine if the central A tract 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 (Fig. 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. The Ø29 early pro-

FIG 5 Transcription regulation of promoters A2c, A2b, and A3 by the proteins p4wt, p4K51A, and p4R54A. The protein used in each assay is indicated above the autoradiograph. The concentrations of the proteins were as follows: p4wt, 160 nM; p4K51A and p4R54A, 320 and 640 nM, respectively; p6, 5 or 10 μ M; RNAP, 11 nM. The transcripts corresponding to each promoter are indicated at left.

motors A2c and A2b are recognized by the *B. subtilis* σ^A -RNA polymerase (RNAP), while the 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 (Fig. 5, lanes a and c) (30). Cooperative binding of proteins p4 and p6 to the promoter region synchronizes the transition from early to late gene expression by repressing the early promoters A2c and A2b and simultaneously activating promoter A3 (Fig. 5, lane d) (6). p6 did not affect transcription in the absence of p4 (lane b). The mutant p4K51A was capable of transcription regulation, although less efficiently than wild-type p4 (Fig. 5, lanes e to g). On the contrary, the 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.

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

To investigate whether p4R54A binding induces DNA bending, we employed circular permutation assays (Fig. 6). Plasmid pBend2wt, containing a 72-bp sequence of ϕ 29 DNA, including p4 sites 1 and 2 (40), was used to determine p4-induced curvature. DNA fragments were produced by restriction enzyme digestions that cleaved twice in the plasmid, resulting in fragments with the sequence of ϕ 29 located at different positions with respect to the fragment ends. Free DNA migrated with similar mobility, in agreement with the absence of intrinsic curvature, as demonstrated previously. In contrast, fragments with a gradient of mobility similar to those of p4wt were generated upon p4R54A binding, with the StuI complex having the slowest mobility, followed closely by the NruI- and SmaI-generated fragment complexes, with the BglII fragment-p4 complex having the fastest mobility. Quantification of the induced bending angles indicated values of

FIG 6 DNA bend induced by p4. Data from the band shift assays of the p4-DNA complexes using circularly permuted DNA fragments. Top lane, diagram of the 72-bp sequence containing p4 binding sites 1 and 2 cloned on the polylinker of plasmid pBend2, with the restriction enzyme organization around the cloned DNA, where the numbers correspond to BamHI (1), SspI (2), NruI (3), StuI (4), SmaI (5), PvuII (6), EcoRV (7), DraI (8), XhoI (9), SpeI (10), NheI (11), and BglII (12). Protein p4 (160 nM) or p4R54A (160 nM) and purified circularly permuted DNA fragments were incubated for 15 min at 4°C and loaded on 4% acrylamide gels. The relative gel mobilities of the p4-DNA complexes were plotted against the positions of restriction sites in the DNA fragment. Bending angles are shown.

89° and 88° for the p4wt-DNA and p4R54A-DNA complexes, respectively. Hence, in both complexes the DNA is bent.

The ability of the p4R54A mutant to stabilize RNAP at promoter A3 was analyzed using a DNA fragment containing promoter A3 and the overlapping promoter A2b (Fig. 7A). In the absence of protein p4, RNAP does not bind to A3, with a low consensus at the -35 element, but rather binds to promoter A2b (43) (Fig. 7A). The 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 (Fig. 5).

Cooperativity between p4 and p6 when binding to DNA results in the multimeric nucleoprotein complex, which, modifying the topology of the promoters, induces 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 Fig. 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 complex. The protein p4R54A fails to synergize p6 binding, since no stable p4-p6-DNA

FIG 7 Interaction of p4R54A with RNA polymerase and with p6. (A) Closed complex formation at promoter A3 in the presence of p4, analyzed in band shift assays. Complexes were formed at 4°C using the DNA fragment containing the promoters A2b and A3, along with RNAP and p4 wild type or the p4R54A mutant. Nucleoprotein complexes resolved on 4% acrylamide gels are indicated at the side. (B) Band shift assays of the ternary p4-p6-DNA and p4R54A-p6-DNA complexes. Proteins were incubated with a labeled DNA fragment containing p4 binding sites 1 to 4. Complexes were resolved on a 4% acrylamide native gel. The positions of the complexes are denoted.

complex was detected, although p4-DNA binding seems to be stimulated by the presence of p6. Hence, most probably the hairpin-like DNA bend is impaired and in consequence the early promoters A2b and A2c are not properly repressed and the late promoter A3 is not fully activated.

DISCUSSION

Protein-DNA complex formation is not a simple encounter between structurally predefined molecules. Several proteins depend on unstructured regions for binding, and DNA can also modify its conformation during the formation of the protein-DNA complex (14, 26, 32, 37, 45, 53). Information contained within the DNA is extracted by proteins, which recognize sequences of the DNA by both direct- and indirect-readout mechanisms. The p4-DNA complex presents three patches of amino acids near three A tracts. In the two external patches, Thr4 and Tyr33 contribute to p4 binding specificity through an indirect-readout mechanism that implies remodelling of the external A tracts (31). A cloud of 12 positively charged amino acids (lysines and arginines) floats over a locally narrow portion of the central minor groove, with Arg54 establishing salt bridges with phosphates. Here, we have shown a role of Arg54 in p4-DNA binding stability, which in turn altered its synergy with protein p6.

The presence of arginines in the minor groove is widespread in protein-DNA complexes, isolated as in the Brn-5 protein (38) and LEAFY (16) or within a motif with more than one arginine as in the phage 434 repressor (46) and Ø29 p4 (this article). Arginines have been shown to interact at positions of minimal minor-groove width and are associated with the presence of short A tracts (41). The *Drosophila* Hox protein Sex combs reduced (Scr) uses variation in electrostatic potential associated with minor-groove width to distinguish small differences in nucleotide sequence (19).

MD simulations of p4-DNA complexes permitted us to evaluate the persistence of arginines 54 in the central minor groove, determining the number of hydrogen bonds formed by each residue and simultaneously measuring the groove width. Considering that the 3D structure is the equilibrium state, both Arg54 residues inversely and superficially cross the groove and contact phosphates of opposite DNA strands in equilibrium (Fig. 3A). Along the p4wt-site 3 MD simulation, Arg54-monomer B was touching the phosphates rather than intruding into the groove. Arg54-monomer A may confer specificity to the p4-DNA complex by interacting simultaneously with phosphates across the groove and by intruding occasionally into the groove to simultaneously establish interactions with the phosphate on one strand and with the sugar oxygen of the opposite strand. These interactions might contribute to a local narrowing of the groove. This picture was evident when the p4-DNA complex was destabilized by disruption of Arg6-G±13 interactions (p4R6A/site 3 complex). In the p4R6A/site 3 complex, Arg54-monomer A lost its superficial interactions with phosphates, while the rate of intrusion into the groove greatly increased. Moreover, Arg54-monomer B is also observed intruding into the groove. In addition, the groove width diminished locally, suggesting that Arg54 functions in narrowing the groove. Variation in the width of the grooves in a controlled manner could lead to local strand deformability. The arginines in the minor groove may use the malleability of the A tract to disrupt the linear helix.

Arginines participate more frequently than lysines in protein-DNA interactions (33); it has been suggested that this is due to the guanidium group of arginine engaging in more hydrogen bonds than the amino group of lysine (27). Participation of arginines in DNA-histone contacts has been the subject of a number of investigations (8, 13, 52). The arrangement of arginines in the p4-DNA complex differs from that in the complexes of Hox, histones, or phage 434 repressor. In the nucleosomes, the arginines are positioned asymmetrically, frequently bridging O-4' atoms of nucleotides *i* and *i*+3 (54). In the 434 repressor-DNA complex, the arginine is in the center of the groove, with the guanidine group bridging the O-4' atoms of nucleotides *i* and *i*+2 (1). In the p4-DNA complex, Arg54 bridged O-3' atoms to phosphates of nucleotides *i* and *i*+3.

It has recently been shown that the bacterial nucleoid-associated protein Fis initially selects DNA targets with a narrow minor groove, separated as dictated by the helix-turn-helix (HTH) motifs of the Fis dimer (48). Unlike the case with 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 stable binding of p4 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 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 (Fig. 8), one protein monomer interacts first with the higher-entropic-stability inverted repeat sequence, 5'-AAATG-3', intruding its N hook into the major groove and providing Arg6-G+13-specific interaction (monomer B). Subse-

FIG 8 Protein p4-DNA binding mechanism: zipper model. (A) Initial interaction of the hook of monomer B with the corresponding DNA sequence (5'-AAAATG-3') is indicated by an orange arrowhead. (B) Subsequent interactions between basic residues of both monomers with the central minor groove are indicated as orange arrowheads. (C and D) The progressive bend of the DNA would approximate the 5'-AAAAAG-3' end to the hook of monomer A (C), permitting the final interactions between the other monomer hook with the sequence (D).

quently, 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 simultaneously contacting 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 bend of the DNA would facilitate the approach of the 5'-AAAAAG-3' inverted repeat to the hook of the second monomer.

In conclusion, ϕ 29 p4 may be considered an additional example of a transcriptional regulator which, being a sequence-specific DNA binding protein, does not have relatively simple readout properties.

ACKNOWLEDGMENTS

We thank Lucia B. Rothman-Denes for critical reading of the manuscript and L. Villar and J. M. Lázaro for protein purification.

This work was supported by the Ministerio de Ciencia e Innovación of Spain (grants BFU2008-01216/BMC to A.C. and 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

- Aggarwal AK, Rodgers DW, Drott 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.
- Badía D, et al. 2006. The structure of the phage ϕ 29 transcriptional regulator p4-DNA complex reveals a novel binding motif. *Mol. Cell* 22:73–81.
- 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.
- Barthelemy I, Lázaro JM, Méndez E, Mellado RP, Salas M. 1987. Purification in an active form of the ϕ 29 protein p4 that controls the viral late transcription. *Nucleic Acids Res.* 15:7781–7793.
- Camacho A, Salas M. 2001. Mechanism for the switch of ϕ 29 DNA early to late transcription by regulatory protein p4 and histone-like protein p6. *EMBO J.* 20:6060–6070.
- Camacho A, Salas M. 2010. DNA bending and looping in the transcriptional control of bacteriophage ϕ 29. *FEMS Microbiol. Rev.* 34:828–841.
- Case DA, et al. 2012. AMBER12, University of California, San Francisco. <http://www.ambermd.org/>.
- 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.* 319:1097–1113.
- del Arroyo PG, et al. 2009. A nucleoprotein-hairpin in transcription regulation. *J. Struct. Biol.* 168:444–451.
- Drew HR, Travers AA. 1985. DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* 186:773–790.
- 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.
- Elías-Arnanz M, Salas M. 1999. Functional interactions between a phage histone-like protein and a transcriptional factor in regulation of ϕ 29 early-late transcriptional switch. *Genes Dev.* 13:2502–2513.
- García-Pérez M, Pinto M, Subirana JA. 2003. Nonsequence-specific arginine interactions in the nucleosome core particle. *Biopolymers* 69:432–439.
- Gartenberg MR, Crothers DM. 1988. DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nature* 333:824–829.
- Gore J, et al. 2006. DNA over winds when stretched. *Nature* 442:836–839.
- Hamès C, et al. 2008. Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *EMBO J.* 27:2628–2637.
- Hansson T, Oostenbrink C, van Gunsteren W. 2002. Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* 12:190–196.
- Hogan ME, Austin RH. 1987. Importance of DNA stiffness in protein-DNA binding specificity. *Nature* 329:263–266.
- Joshi R, et al. 2007. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* 131:530–543.

20. Kalodimos CG, et al. 2004. Structure and flexibility adaptation in non-specific 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 Res. 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.* 36: 4797–4807.
25. Lionnet T, Joubaud S, Lavery R, Bensimon D, Croquette V. 2006. Wringing out DNA. *Phys. Rev. Lett.* 96:178102. doi:10.1103/PhysRevLett.96.178102.
26. Love JJ, et al. 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. Menciá 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 ϕ 29. *Proc. Natl. Acad. Sci. U. S. A.* 108: 18655–18660.
30. Menciá M, Monsalve M, Rojo F, Salas M. 1996. Transcription activation by phage ϕ 29 protein p4 is mediated by the interaction with the α subunit of the *Bacillus subtilis* RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* 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 ϕ 29 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. U. S. A.* 89:11401–11405.
36. Otwinowski Z, et al. 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: 2379–2394.
40. Pérez-Lago L, Salas M, Camacho A. 2005. A precise DNA bend angle is essential for the function of the phage ϕ 29 transcriptional regulator. *Nucleic Acids Res.* 33:126–134.
41. Rohs R, et al. 2009. The role of DNA shape in protein-DNA recognition. *Nature* 461:1248–1253.
42. Rohs R, et al. 2010. Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* 79:233–269.
43. Rojo F, Zaballos A, Salas M. 1990. Bend induced by the phage ϕ 29 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, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
45. Schleif R. 2000. Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet.* 16:559–565.
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 ϕ 29. *J. Mol. Biol.* 127:411–436.
48. Stella S, Cascio D, Johnson RC. 2010. The shape of the DNA minor groove direct binding by the DNA-bending protein Fis. *Genes Dev.* 24: 814–826.
49. Studier FW, Moffatt BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189: 113–130.
50. Tabor S, Richardson CC. 1992. RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Biotechnology* 24:280–284.
51. Thompson JF, Landy A. 1988. Empirical estimation of the protein-induced DNA bend to lambda site-specific recombination complexes. *Nucleic Acids Res.* 16:9687–9705.
52. Tolstorukov MY, Colasanti AV, McCandlish DM, Olson WK, Zhurkin VB. 2007. A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. *J. Mol. Biol.* 371:725–738.
53. Travers A. 1998. Transcription: activation by cooperating conformations. *Curr. Biol.* 8:R616–618.
54. von Hippel PH. 1994. Protein-DNA recognition: new perspectives and underlying themes. *Science* 263:769–770.
55. Wang D, Ulyanov NB, Zhurkin VB. 2010. Sequence-dependent Kink-and-Slide deformations of nucleosomal DNA facilitated by histone arginines bound in the minor groove. *J. Biomol. Struct. Dyn.* 27:843–859.
56. Watkins D, Srividya M, Gerald B, Williams LD. 2010. Sequence recognition of DNA by protein-induced conformational transitions. *J. Mol. Biol.* 396:1145–1164.
57. Zhang Y, Xi Z, Hegde RS, Shakked Z, Crothers DM. 2004. Predicting indirect readout effects in protein-DNA interactions. *Proc. Natl. Acad. Sci. U. S. A.* 101:8337–8341.