DNA phosphate crowding correlates with protein cationic side chain density and helical curvature in protein/DNA crystal structures

Bryce N. Grant, Elizabeth M. Dourlain, Jayme N. Araneda, Madison L. Throneberry and Lori A. McFail-Isom*

Department of Chemistry, University of Central Arkansas, Conway, AR 72035, USA

ABSTRACT

Sequence-specific binding of proteins to their DNA targets involves a complex spectrum of processes that often induce DNA conformational variation in the bound complex. The forces imposed by protein binding that cause the helical deformations are intimately interrelated and difficult to parse or rank in importance. To investigate the role of electrostatics in helical deformation, we quantified the relationship between protein cationic residue density (Cpc) and DNA phosphate crowding (Cpp). The correlation between Cpc and Cpp was then calculated for a subset of 58 high resolution protein–DNA crystal structures. Those structures containing strong Cpc/Cpp correlation (>±0.25) were likely to contain DNA helical curvature. Further, the correlation factor sign predicted the direction of helical curvature with positive (16 structures) and negative (seven structures) correlation containing concave (DNA curved toward protein) and convex (DNA curved away from protein) curvature, respectively. Protein–DNA complexes without significant Cpc/Cpp (36 structures) correlation (-0.25<0<0.25) tended to contain DNA without significant curvature. Interestingly, concave and convex complexes also include more arginine and lysine phosphate contacts, respectively, whereas linear complexes included essentially equivalent numbers of Lys/Arg phosphate contacts. Together, these findings suggest an important role for electrostatic interactions in protein–DNA complexes involving helical curvature.

INTRODUCTION

Gene expression is modulated by intricate cellular processes that often hinge on recognition and binding of upstream DNA sequences by proteins. The binding of proteins to operators controls the expression of genes, regulating processes such as DNA replication, transcription, recombination and repair (1). Far from serving merely rigid and passive roles in these processes, the DNA sequences of operator regions provide docking site that can adapt to form complementary interaction surfaces for protein docking.

Proteins can read DNA sequence by directly probing the floors of the DNA grooves with amino acid side chains [direct readout, (2,3)]. Proteins can also read DNA sequence indirectly by sensing sequence-specific predisposition of DNA to adopt non-canonical conformations [indirect reaout,(4)]. DNA deformations of importance include axial curvature and groove distortion. Attempts to decipher the complex factors contributing to the recognition and binding of proteins to their target DNA have included computational analysis of protein–DNA crystal structure coordinates (5,6). Many proteins use combinations of direct and indirect readout (7,8), whereas some proteins bind almost exclusively by indirect readout (9,10).

Although DNA deformations are readily observed by structural methods such X-ray diffraction and NMR spectroscopy, the forces imparted by a protein on DNA that cause the deformation are subtle and difficult to parse or rank in importance (11–14). DNAs with attached cationic charges or neutralized phosphate backbone show behaviors that suggest the importance of electrostatics in DNA deformation [reviewed in (15)]. Other studies investigating the role of electrostatic interaction in protein–DNA-binding suggest that interactions between the cationic regions of the protein and the DNA phosphate backbone mediate non-specific protein–DNA association (16), although many of these interactions are retained in the sequence-specific bound complex (17). In addition to electrostatic influences, other factors have been shown through crystallographic analysis (18,19), solution (20–22) and computational

*To whom correspondence should be addressed. Tel: +1 501 450 5794; Fax: +1 501 450 3623; Email: lorii@uca.edu

(23–25) studies to influence DNA bending including those imparted by specific DNA sequence combinations that provide curvature and flexibility.

Our findings here suggest important roles for electrostatic interactions between DNA phosphate groups and cationic amino acid side chains in protein–DNA complexes when the primary distortion of DNA is axial curvature. Specifically, we investigated the relationship between cationic protein side chain density (Cpc) around DNA phosphate oxygens (POs), the positioning of those POs with respect to intra- and interstrand neighbors (Cpp) and helical curvature in 58 high resolution protein–DNA crystal structures. We report that structures whose pattern of cationic protein side chain density coincided most consistently with regions of specific PO positioning contained protein–DNA complexes with curved DNA. Further, the direction of correlation between structural regions of cation density and PO position predicted the direction of DNA helical curvature in the complex with positive and negative correlation containing concave and convex curvature, respectively. Protein–DNA complexes without significant correlation contained DNA without curvature.

MATERIALS AND METHODS

Structure selection criteria

The protein/DNA X-ray crystal structures with 2.5 Å resolution or higher were selected for analysis from the Protein Data Bank (26) and screened against selection criteria. Application of these criteria resulted in the final set of 58 protein/DNA structures. The criteria were established to select a final set of structures with sufficient structural detail, cohesiveness and consistency to maximize comparability. Only the highest resolution structure available for each protein was retained and analyzed to prevent over-representation when multiple structures were available. To help ensure structural consistency, only structures including DNA with at least 10 contiguous base pairs and free from significant base modifications, mismatches and unstacking were selected. To minimize the potential influence of local environment on DNA structure, structures including multiple helices in contact with each other within the unit cell, and solvent ions or ligands associated with the grooves or phosphates of the DNA were excluded.

Crowding functions

To investigate the relationship between the positioning of charged protein side chains and negatively charged DNA PO's, the local environment of each PO was quantified. Three crowding functions, Cpp, Cpc and Cpa, were calculated from structure coordinates using Perl scripts. Specifically, for each PO, crowding parameters (Cpx) were calculated by summing the inverse of the distance [Equation (1)] in angstroms between a given PO and each corresponding group, x, within the applied cut off radius [Equation (2)]. Groups were thereby weighted by distance with those further away from the PO contributing less to the total crowding function value than do groups in closer proximity.
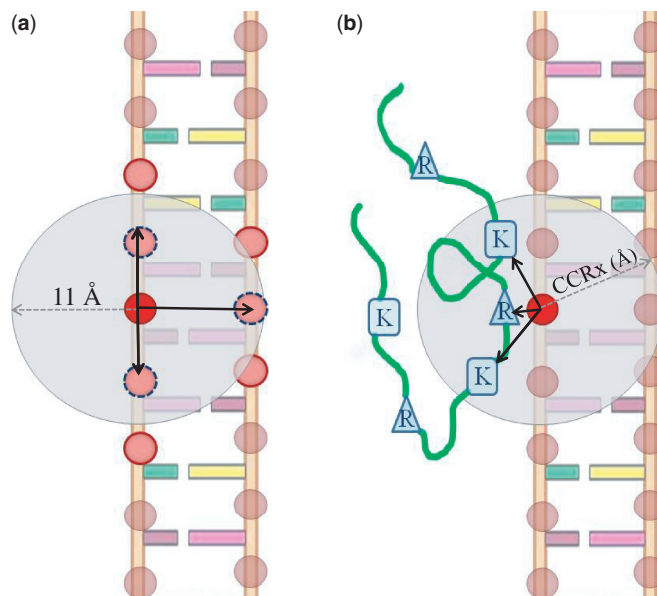

Figure 1. Crowding function (Cpx) schematics. (**a**) The defining components of the Phosphate Crowding function (Cpp). For each PO, the inverse of the distances to all other POs within an 11 Å radius were summed. The 11 Å distance allowed the inclusion of the neighboring intra-strand and inter-strand POs across an average minor groove. (**b**) The defining components of the ion crowding function (Cpx). For each PO, the inverse distances to all charged protein side chains within the selected cut off radius were summed. The cation crowding function (Cpc) included distances between each PO and each Lys and Arg, whereas anion crowding function (Cpa) included distances between each PO and Glu and Asp (Supplementary Figure S2). Values for Cpx were calculated using increasing cut off radii (CCRx) values 1, 3, 5, 7, 9, 11, 13 and 15 Å.

$$d_{px} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \qquad (1)$$

$$Cpx = \sum d_{px}^{-1} \qquad (2)$$

Three crowding parameters, Cpp, Cpc and Cpa, were calculated for each PO. The phosphate crowding function value, Cpp, represents the sum of the inverse distances between a given PO and the other PO's within the cut off radius of 11 Å (Figure 1a). This cut off radius was established based on measurements of the 1.4 Å resolution structure of the B-form DNA dodecamer (PDB 355d) (27) and chosen to include on average the PO of (1) the adjacent intra-strand phosphates and (2) the inter-strand phosphate across the narrowest span of the minor groove. The phosphate crowding function will include additional POs if DNA deformation induces phosphate crowding, placing those POs within the radius.

The ion crowding functions, Cpc and Cpa, represent the sum of the inverse distances between a given PO and the ionic protein side chains within the cut off radius (Figure 1b). Because side chains far from the DNA-binding interface have been shown to affect binding specificity (28), a series of distance cut offs was analyzed for each structure. For the ion crowding functions, analyses were independently conducted for a range of distance cut offs including 1, 3, 5, 7, 9, 11, 13 and 15 Å.

The cationic crowding function, Cpc, includes the distances between each PO and the basic amino acid side chains of arginine and lysine, which are positively charged at physiological pH. For the analysis, the amine nitrogen, NZ, was used for lysine and the central imidizole carbon, CZ, for arginine.

The anionic crowding function, Cpa, represents the sum of the inverse distances between each PO and the acidic amino acid side chains of aspartic acid and glutamic acid containing the negatively charged carboxylate group at physiological pH. Because the negative charge of both carboxylate groups is distributed between the two oxygens by resonance, a point representing the average coordinates for the oxygens is assigned the position of negative charge for aspartate and glutamate side chains.

Data analysis

To investigate the relationship between PO crowding and the corresponding ion density, crowding function values (Cpp, Cpc, Cpa) for each structure were collected, and the correlation coefficients between phosphate crowding (Cpp) and ion density (Cpc or Cpa) were calculated using Equation (3):

$$Correl\ (A,B) = \sum ((a - \overline{a}))(b - \overline{b})/\sqrt{\sum (a - \overline{a})^2 \sum (b - \overline{b})^2} \quad (3)$$

The final truncated Cpx data set for each structure was constructed by removing the Cpx values for the terminal two phosphates of each strand to prevent skewing by end effects. Control data demonstrating Cpx independence from base position/identity are presented in Supplementary Figure S1 and described in Supplementary Methods and Supplementary Results. The truncated Cpx data sets were used for all further analysis contained herein.

Subgroup assignment and characterization

Structures were assigned to one of the three subgroups by Cpp/Cpc correlation value. Two sorting criteria were applied for subgroup assignment (i) absolute maximum and (ii) average correlation over the cut off radius range (5–15 Å). The cut off distances of 1 Å (overall mean corr 0.00) and 3 Å (overall mean corr 0.020) were not included to prevent skewing of the average.

Structures were assigned to various subgroups based on average and maximum correlation with average and absolute maximum correlation coefficient ≥ 0.25 and ≤ -0.25 assigned to the positive and negative subgroups, respectively. The remaining structures with average and maximum correlation falling in the range in between these two extremes (-0.24 to $+0.24$) were assigned to the null subgroup. In 52 of 58 cases, the average and maximum correlation coefficients were consistent for subgroup assignment. Maximum correlation coefficient was used to assign the subgroup for the remaining six structures.

Classification of helical curvature

Our analysis required the classification of helical curvature with respect to qualitative characteristics of the complex, including (i) overall deviation from helical linearity and (ii) the direction of curvature, if present, as toward the protein or away from it. Two methods were used to establish both the presence and direction of DNA helical curvature: visual inspection and calculation using Curves+ (29). As our analysis involved qualitative aspects of curvature, systematic visualization was used for screening and classification. Specifically, in random order without knowledge of previous correlation results, each structure was viewed independently by at least three researchers using either RasMol or Swiss PDB Viewer and classified as containing either curved or linear DNA. If the structure was classified as containing curved DNA, the curvature was designated as described in (15) as either (i) concave in structures containing DNA that curved toward the bound protein or (ii) convex for structures containing DNA that curved away from the protein. Assignment discrepancies were reconciled by consensus among the researchers. The helical curvature of each structure was then calculated using the program Curves+ or curvature value reported in the structure's primary PDB citation.

Counting the number of phosphate interactions with lysine and arginine

Distances between protein arginine CZ or lysine NZ atoms and the phosphorous atom of the phosphate groups of each structure's DNA backbone were calculated. A distance of ≤ 6.0 Å was considered an interaction between the corresponding side chain and the backbone phosphate.

Visualization of crowding function values

Once calculated, the values of Cpc and Cpp were coupled with the corresponding coordinates of each PO in each structure. The POs were then visualized using molecular visualization software and colored with respect to their crowding function value over the cut off radius range. In each case, the PO coloring spectrum ranges from blue, representing the least amount of crowding, to red, representing the most.

RESULTS

Selection criteria produce a 58 X-ray crystallographic structure set for analysis

Application of selection criteria described in 'Materials and Methods' section resulted in 58 crystallographic protein–DNA structures used in subsequent analysis (full list included in Supplementary Table S1).

Cpp/Cpx correlation trends establish three structure subgroups

Patterns of change in correlation coefficient with increasing Cpc cut off radius suggest three structure subgroups. To assess whether phosphate crowding coincided

with proximal cationic side chain density, the correlation coefficient between Cpp and Cpc was calculated for each structure and plotted against the corresponding increasing range of cationic cut off radii. A pattern was detected in the resulting curves, allowing the establishment of three subgroups based on an observed (i) positive, (ii) negative correlation trend and (iii) those with a tendency to maintain a correlation value of zero (null subgroup) as described later in the text and shown schematically in Figure 2a, b and c, respectively. Representative structure data are plotted for each subgroup in Figure 3.

Trends characteristic of each subgroup are revealed in representative and average plots of Cpp/Cpc correlation versus increasing cut off radius. Correlation values for the positive, negative and null subgroups were averaged and plotted against their corresponding cut off radius in Figure 4. The average positive subgroup (16 structures, blue filled circles) coefficient average trend suggests that regions of phosphate crowding have a moderate to strong correlation with regions of cationic side chain density that plateau ∼11 Å at a maximum of 0.45. The negative subgroup (seven structures, filled red triangles) correlation coefficient average suggests that cationic side chains in these complexes tend to surround phosphates that are least crowded displaying an increasingly negative correlation, which levels out ∼13 Å at a maximum of -0.34. The null subgroup (36 structures, filled green squares) coefficient average trend suggests that cationic side chain density is not correlated with phosphate crowding in these complexes.

Control data confirming the absence of correlation relationship ($P = 0.09$) between Cpp/Cpa are presented in Supplementary Figure S2 and described in Supplementary Results.

Correlation subgroups correspond to three distinct categories of DNA helical curvature

To investigate the relationship between the structures comprising the three subgroups, the protein/DNA complexes were visualized. On examination, a subgroup-dependent trend in DNA helical curvature was readily observed. The types of global DNA helical curvature observed in these complexes represent a qualitative difference involving the presence and direction of DNA curvature with respect to the bound protein.

Structures were designated as containing either linear (Figure 5a) or curved DNA. Complexes containing curved DNA were further classified as either (i) concave with DNA curved toward the protein (Figure 5c) or (ii) convex with DNA curved away from the protein (Figure 5e). An example of each type of complex is also given in Figure 5. Structure 1NKP contains DNA without clear qualitative curvature and therefore is designated as linear (Figure 5b). Structure 1B3T contains DNA with concave curvature (Figure 5d), and 2Z3X displays convex curvature (Figure 5f).

The assigned helical curvature classification was confirmed from helical axis-bending values calculated using Curves+ and/or values reported in the structure's
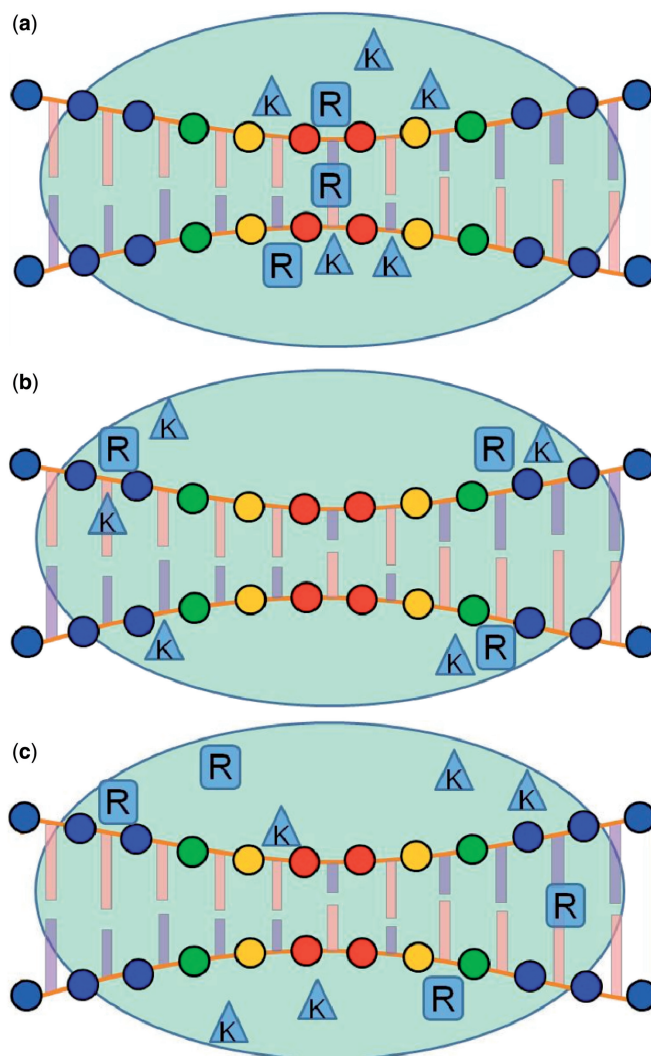

Figure 2. Schematic of types of Cpp/Cpc correlation. For each figure, the POs are represented by circles on the DNA backbone. These PO color reflects the value of Cpp with red representing the most crowded and blue the least. Cationic protein residues are represented by K for Lysine and R for Arginine. (**a**) Positive correlation subgroup. In this subgroup, cationic side chain density (Cpc) is localized around POs with the highest crowding function value (Cpp). (**b**) Negative correlation subgroup. A negative correlation value between Cpp and Cpc means cationic protein side chains are most often clustered around phosphates that are least crowded. (**c**) Null correlation subgroup. Structures displaying a Cpp/Cpc correlation coefficient of ∼0 contain no relationship between phosphate crowding and cationic protein side chain density.

primary citation. The average helical curvature for each bending category was calculated and is given in (Figure 6). The average global helical curvature for complexes classified as concave ($38° \pm 13°$) or convex ($42° \pm 9°$) was statistically greater ($P < 0.001$) than the average helical bend for complexes containing straight DNA ($9° \pm 6°$).

Cpp/Cpc correlation subgroups contain structures with distinct helical curvature classification

Structures with positive correlation between cationic side chain density and phosphate crowding also primarily
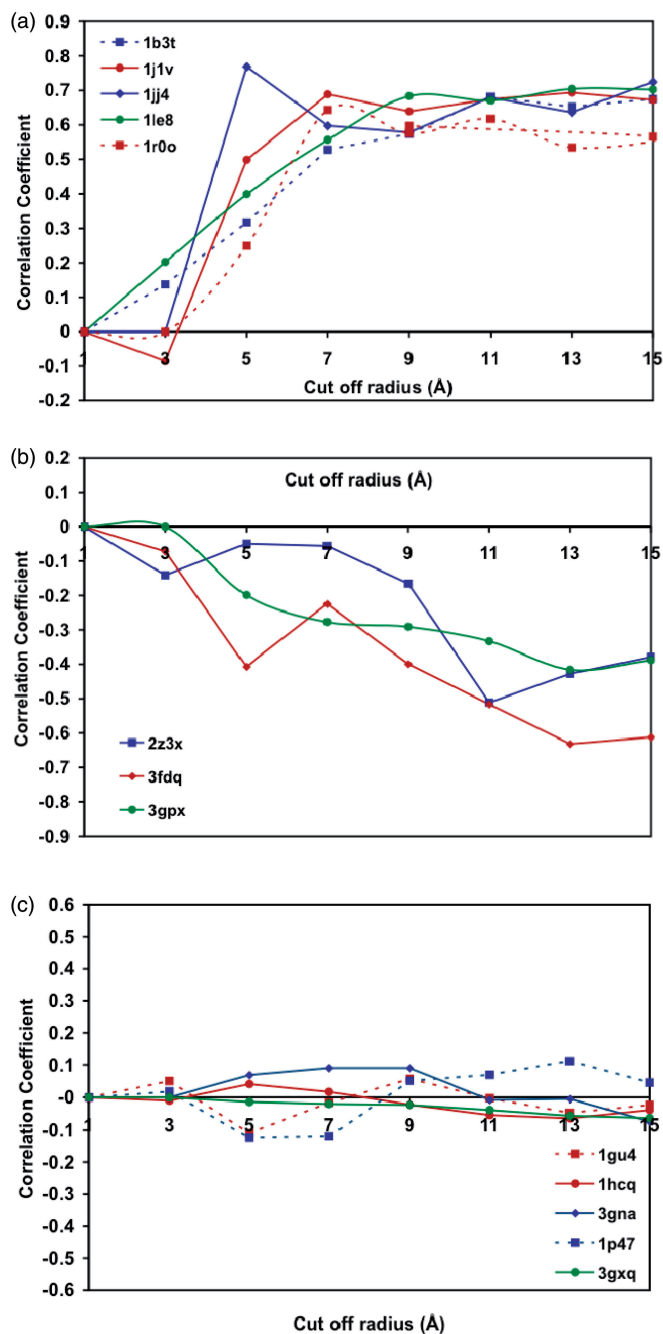
Figure 4. Average correlation coefficient for correlation subgroups. The average Cpc (**a**) or Cpa (**b**) correlation value for all structures classified as positive (circle), null (squares) and negative (triangles) is plotted versus Cpc cut off radius. Error bars reflect the 95% confidence ranges of each average ($P < 0.001$).

Figure 3. Correlation coefficient versus Cpx cut off radius (Å) for structures representative of correlation subgroups. Structures were sorted into correlation subgroups as described in the 'Materials and Methods' section. For the (**a**) positive, (**b**) negative and (**c**) null subgroups, the correlation coefficients for representative structures are plotted versus the Cpx cut off radius (Å).

display concave helical curvature (Figure 7). Specifically, of the 16 structures in the positive correlation subgroup, 14 (88% of subgroup total) also contain concave DNA helical curvature, whereas two (12%) were designated as linear.

Structures with negative correlation between cationic side chain density and phosphate crowding primarily contain DNA with convex helical curvature. Six of the
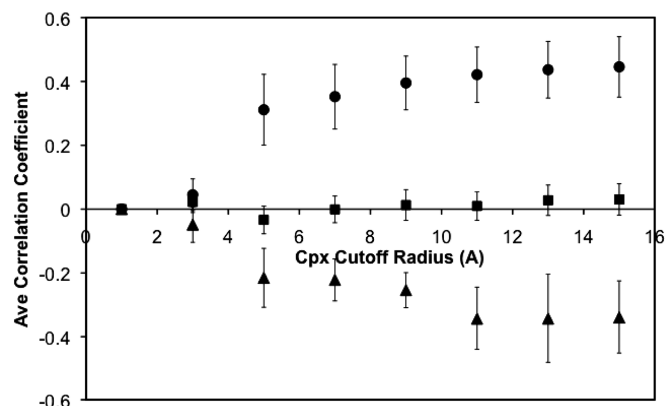
seven (86% of subgroup total) structures in the negative correlation subgroup contained DNA exhibiting negative curvature, whereas one structure (14%) contained positive curvature.

Structures without discernible correlation between cationic side chain density and phosphate crowding tend to contain linear DNA without significant qualitative helical curvature. Of the 35 structures in the null subgroup, 30 of them (86% of subgroup total) contained uncurved DNA. Of the remaining five structures, four (11%) were classified as containing concave and one (3%) as convex curvature.

Crowding function values allow visualization of relationship between Cpp, Cpc and curvature

Visualization of the crowding function values reveal positioning of phosphate crowding, cationic side chain density and helical curvature consistent with subgroup assignment. The corresponding crowding function values were linked to each PO and visualized by spectrum-coding (blue—least, red—most) the crowding function data. One structure from each subgroup is presented in Figures 8, 9 and 10. For each structure, the first frame presents the protein/DNA complex with the protein backbone rendered in green ribbon and the cationic side chains in magenta sticks. The second frame presents the POs spectrum coded by phosphate crowding (Cpp) value. The Cpc data are presented for three cut off radii, each of which is paired with a rendering of the protein/DNA complex with the protein contacts within the specific cut off radius from the DNA shown with cationic side chains in magenta spheres representing the Van der Waals radii of the atoms.. The corresponding Cpp/Cpc correlation coefficient versus Cpc cut off radius graph for each structure is presented in Supplementary Figure S3.

Within the structure with concave curvature, the pattern of phosphate crowding corresponds to the regions of cationic side chain density. The positive correlation subgroup is represented by structure 1le8 (Figure 8). The helix curvature in the protein/DNA complex is concave
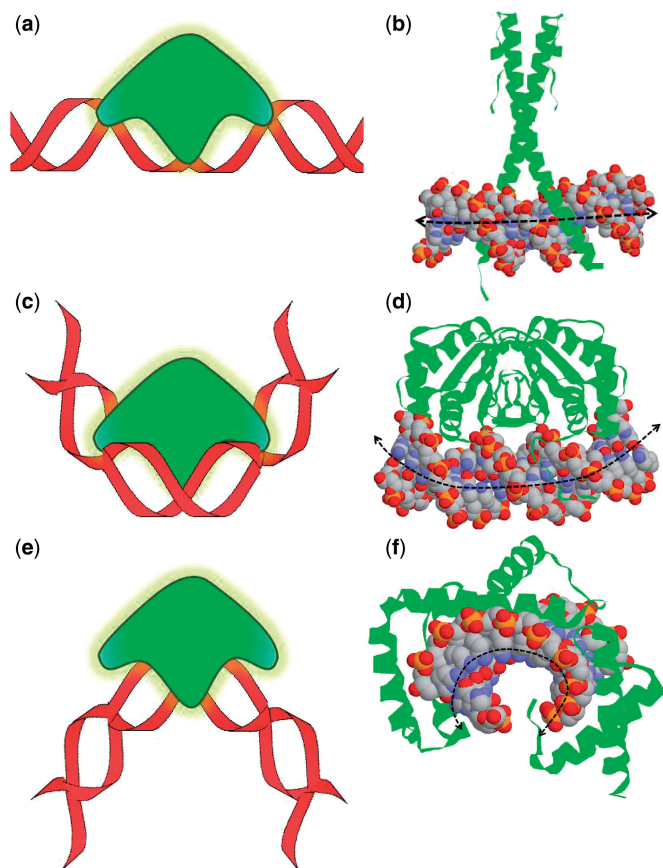
Figure 5. Types of helical curvature in protein–DNA complexes. Structures were sorted by the type of helical curvature present as described in 'Materials and Methods' section. Views of representative structures of each type are given in (**b**), (**d**) and (**f**). In each case, the protein (green ribbon) and the DNA (spacefilled) are shown with an approximation of the helical axis from Curves+ data is shown as a dashed double-sided arrow. (**a**) Linear complexes, including (b) pdbID 2C9I contained DNA without detectable helical curvature. Complexes displaying (**c**) concave and (**e**) convex curvature are represented by (d) pdbID 1B3T and (f) pdbID 2Z3X, respectively.
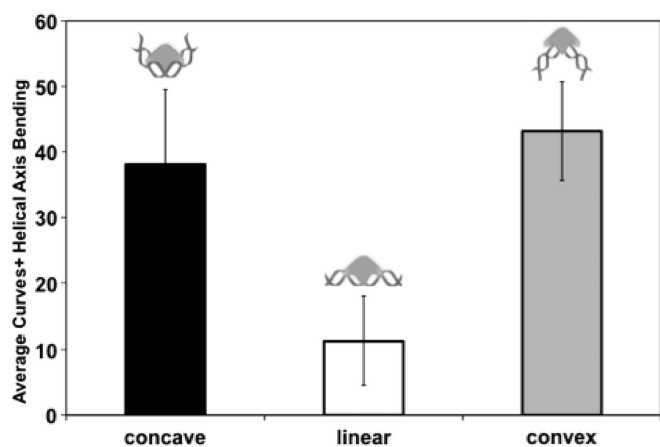

Figure 6. Average calculated global curvature values. Structures classified as containing concave (black bar), linear (white bar) and convex (gray bar) helical curvature. Average of curvature obtained from Curves+ or primary citation ($P < 0.001$). Error bars represent the standard deviation of the averages.
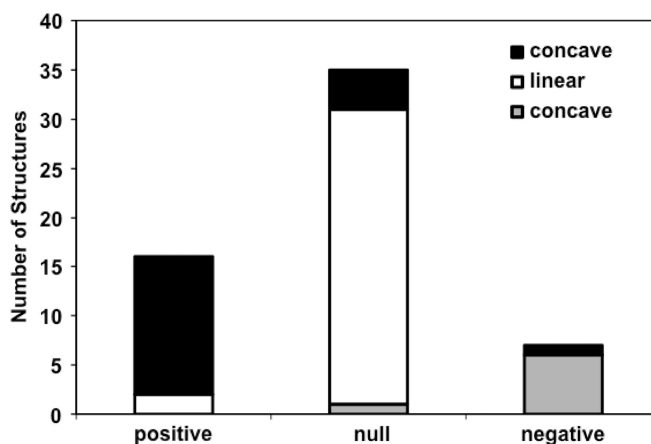

Figure 7. Population of correlation subgroups by helical curvature types. The number of total structures displaying concave (black), linear (white) or convex (gray) within the positive, null and negative correlation subgroups are shown. The positive, null and negative correlation subgroup contains predominantly complexes with concave, linear and convex helical curvature, respectively.
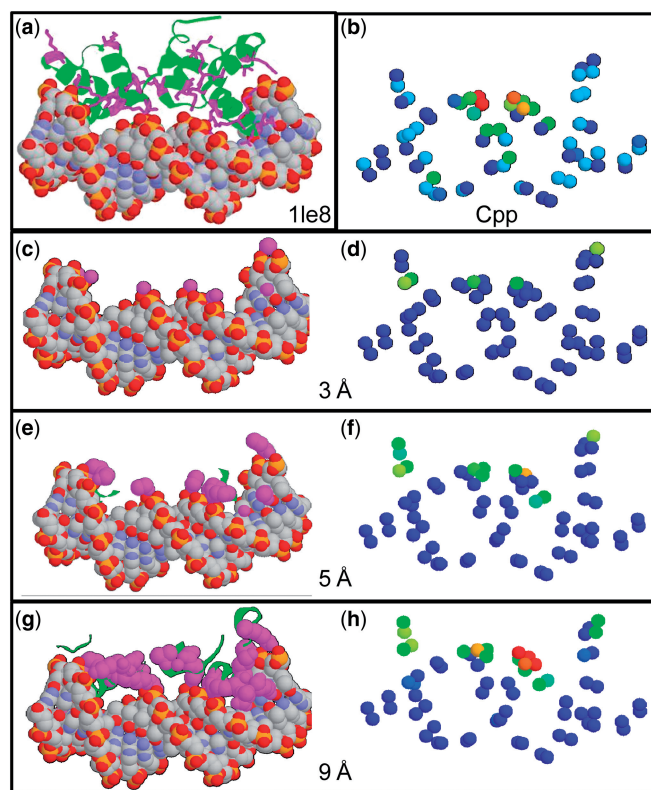

Figure 8. Visualizing the positive Cpp/Cpc correlation present in complexes with concave curvature. For the structure 1le8, (**a**), (**c**), (**e**) and (**g**) display the protein (green ribbon) with cationic residues (magenta stick/spacefilled) for the whole structure (a) or those contributing to Cpc by falling within the specified cut off radius. The relative Cpc distribution for each corresponding cut off radius is presented in (**d**), (**f**) and (**h**). In these figures, the color of the POs (spacefilled) reflect the relative Cpc value with red representing the POs surrounded by the highest cationic side chain density (Cpc) value and blue the lowest. The phosphate crowding function (Cpp) values for the POs are displayed in (**b**) using the same spectrum with high phosphate crowding shown in red and low in blue.
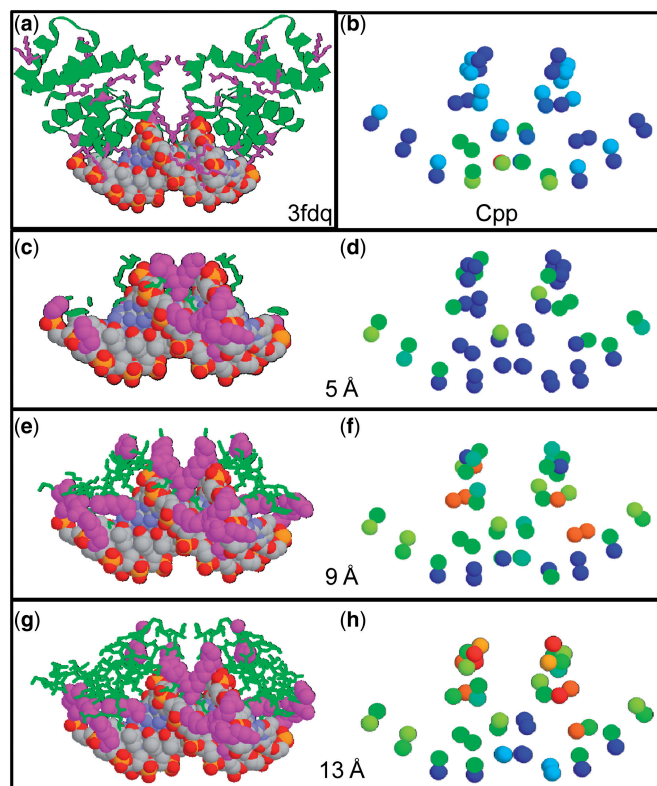
Figure 9. Visualizing the negative Cpp/Cpc correlation present in complexes with convex curvature. For the structure 1fdq, (**a**), (**c**), (**e**) and (**g**) display the protein (green ribbon) with cationic residues (magenta stick/spacefilled) for the whole structure (a) or those contributing to Cpc by falling within the specified cut off radius. The relative Cpc distribution for each corresponding cut off radius is presented in (**d**), (**f**) and (**h**). In these figures, the color of the POs (spacefilled) reflect the relative Cpc value with red representing the POs surrounded by the highest cationic side chain density (Cpc) value and blue the lowest. The phosphate crowding function (Cpp) values for the POs are displayed in (**b**) using the same spectrum with high phosphate crowding shown in red and low in blue.

Figure 10. Visualizing the null Cpp/Cpc correlation present in complexes with linear DNA. For the structure 1hcr, (**a**), (**c**), (**e**) and (**g**) display the protein (green ribbon) with cationic residues (magenta stick/spacefilled) for the whole structure (a) or those contributing to Cpc by falling within the specified cut off radius. The relative Cpc distribution for each corresponding cut off radius is presented in (**d**), (**f**) and (**h**). In these figures, the color of the POs (spacefilled) reflect the relative Cpc value with red representing the POs surrounded by the highest cationic side chain density (Cpc) value and blue the lowest. The phosphate crowding function (Cpp) values for the POs are displayed in (**b**) using the same spectrum with high phosphate crowding shown in red and low in blue.

with the Cpp/Cpc correlation coefficient plateauing ∼0.68 at 9 Å cation cut off radius. When compared with the structure's Cpp values displayed in Figure 8, a consistency between the patterns is evident as early as 5 Å and the pattern continues to clarify to 9 Å.

Within the structure with representing convex curvature, the pattern of phosphate crowding is related inversely to the regions of cationic side chain density. The negative correlation subgroup is represented by structure 3fdq (Figure 9). The helix curvature in the protein/DNA complex is convex with the Cpp/Cpc correlation coefficient reaching its maximum negative value of −0.63 at 13 Å cation cut off radius. When compared with the Cpp values for the structure (Figure 9), the negative correlation pattern is evident as early as 5 Å displaying the most clearly at 13 Å cut off radius.

Within the structure with linear DNA, little correlation between pattern of phosphate crowding and cationic side chain density is present. The null correlation subgroup is represented by structure 1hcr (Figure 10). There is no global helical curvature in the protein/DNA complex, and the Cpp/Cpc correlation coefficient hovers around
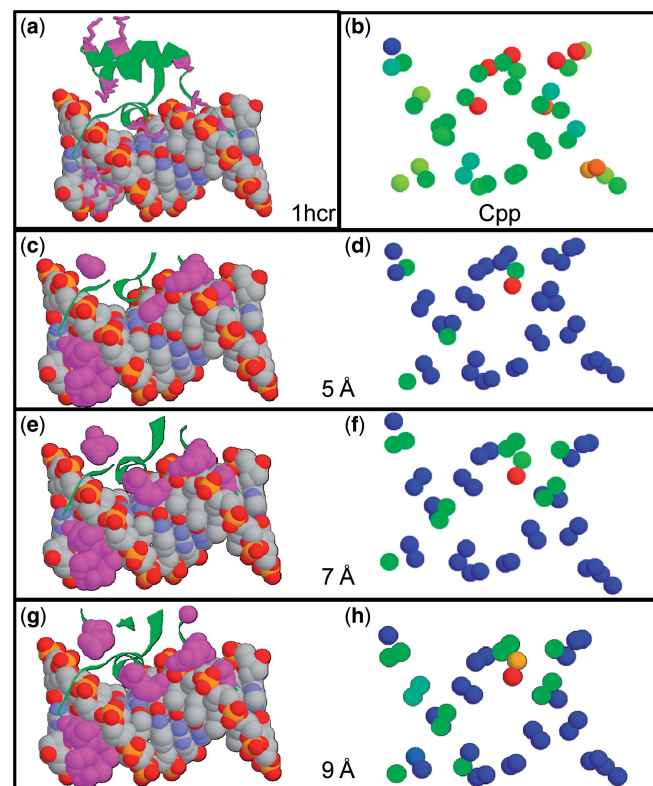
zero over the entire cation cut off radius range. The cationic side chains appear at 5 Å and form the extent of their contacts with the grooves of major and minor grooves of the DNA. Increasing cation cut off radius consistently reveals the extensive contacts between the cationic side chains and the groups presented on the base edges that line the floor of the grooves.

Correlation subgroups differ in predominant interacting basic side chain identity

Interactions between phosphates and basic side chains were compiled for each subgroup. The distances between the basic side chains and the DNA phosphate groups were sorted by correlation subgroup. The average number of contacts was consistent between subgroups with positive containing 10.5, negative 9.7 and null 10.4 contacts per structure on average (Figure 11).

Within positive Cpp/Cpc correlation subgroup, arginine is involved in the closest contacts almost 2:1 over lysine. Of 168 total contacts for this subgroup, 63 (37%) involved lysine, whereas 105 (63%) involved arginine.
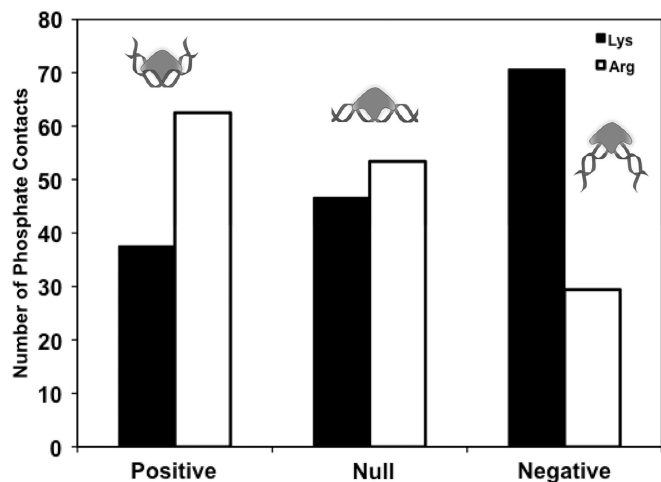
Figure 11. Identity of cationic protein side chain participating in DNA phosphate contacts. The number of contacts between Lysine and Arginine and the DNA phosphates were calculated as described in 'Materials and Methods' section. The percentage of total Lysine (black) and Arginine (white) side chains interacting with phosphate groups are given for concave, linear and convex complexes.

Within the negative Cpp/Cpc correlation subgroup, lysine is involved in the closest contacts almost 3:1 over arginine. Of 68 total contacts for this subgroup, 48 (71%) involved lysine, whereas 20 (29%) involved arginine.

Within the null Cpp/Cpc correlation subgroup, arginine is only slightly more often involved in the closest contacts over lysine. Of 365 total contacts for the null subgroup, 170 (47%) involve lysine, whereas 195 (53%) involve arginine.

DISCUSSION

Our findings suggest an important role for electrostatic interactions between proteins and their target DNA when axial curvature is present. Specifically, our data reveal an intriguing correlation between cationic side chain density, phosphate crowding and DNA curvature. The relationship between regions of phosphate crowding and cationic side chain density reflect general backbone electrostatic interactions and a possibly significant role for cationic side chain/backbone phosphate interaction in stabilizing DNA helical curvature. The location of the cationic regions in the bound protein seem to present a stabilizing scaffold that promotes either (i) the wrapping of the target DNA around the protein in concave complexes by relieving the electrostatic repulsion resulting from the phosphate crowding this conformation induces (15,30–34) or (ii) the dramatic bending of the DNA away from the perching protein by providing the strategically positioned stabilizing lure of favorable backbone contacts resulting from the widened groove that results from this conformation (15).

Structural analysis of our subset of protein/DNA complexes support and extend studies describing the relationship between phosphate neutralization and helical curvature for specific structures (35) and suggest generalization of the relationship to other complexes containing protein–DNA complexes containing helical curvature. These data suggest concave complexes tend to position

cationic side chain density to neutralize the repulsion between negatively charged phosphate groups resulting from crowding, as the helix is curved towards the protein. Structures containing convex helical curvature have higher cationic side chain density around phosphates that are spread relatively further apart as the groove is widened. Therefore, instead of using cation density to relieve electrostatic repulsion, these convex complexes seem to strategically place cationic density to offer attractive electrostatic interaction for phosphates that are able to properly orient themselves accordingly. These potential favorable interactions established by the dramatic DNA distortion in these complexes might serve to promote the conformational transition and to stabilize the resulting complex. Structures containing DNA without significant helical curvature show little correlation between cationic side chain density and phosphate crowding, suggesting a role in these complexes distinct from helical distortion.

Although we are tempted to offer prospective speculation, we would like to clarify that care has been taken not to claim that the analysis herein allows the deciphering of the complex interplay between correlation and causation. Such distinctions will rely on expanded comparisons of high-resolution DNA crystal structures of the same target sequence with protein bound and unbound or by elegant theoretical analyses able to parse multiple factors like those reported in (9,11,24). Although our findings allow clarification and confirmation of the role electrostatic backbone interactions between the DNA and cationic protein side chains, our analysis alone cannot establish whether the DNA in curved complexes is curved before the protein binds or whether the protein induces such curvature on binding. It is unlikely however that the extreme deformation observed in the convex protein–DNA complexes is present in the free DNA.

Consistent with our claim that the regions of cationic density serve a different role in linear complexes, both arginine and lysine are found to interact with the phosphate backbone in almost equal frequencies in these complexes. The difference in cationic side chain identity detected between complexes containing concave and convex curvature is intriguing. We do, however, offer some ideas concerning our observed correlation. Our theory that the positioning of cationic residue density relative to the DNA backbone can stabilize the resulting phosphate orientations in complexes containing either concave or convex helical curvature is enriched by our finding that different cationic residues tend to stabilize these complexes. Concave and convex complexes favor arginine and lysine, respectively. Arginine has been shown to occupy narrow minor grooves regions in protein–DNA complexes (12). The delocalized positive charge offers a less energetically expensive dehydration process on binding (12). The delocalized positive charge distributed among atoms arranged in a planar geometry present a positively charged two-sided paddle-like structure that could offer a diffused screen of neutralizing charge that could relieve the repulsion between approaching POs that groove narrowing requires. In contrast to a diffused screen with greater surface area with arginine, the focused positive charge of lysine residues could lure and

promote distorted convex backbone conformations that would provide stabilizing interactions with the properly positioned POs in complexes whose DNA sequence allows them to comply.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1, Supplementary Figures 1–3, Supplementary Methods, Supplementary Results and Supplementary References [36–91].

REFERENCES

1. von Hippel,P.H. (2007) From "simple" DNA-protein interactions to the macromolecular machines of gene expression. *Annu. Rev. Biophys. Biomol. Struct.*, **36**, 79–105.
2. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B. and Mann,R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.*, **79**, 233–269.
3. Seeman,N.C., Rosenberg,J.M. and Rich,A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA*, **73**, 804–808.
4. Shakked,Z. and Rabinovich,D. (1986) The effect of the base sequence on the fine structure of the DNA double helix. *Prog. Biophys. Mol. Biol.*, **47**, 159–195.
5. Michael Gromiha,M., Siebers,J.G., Selvaraj,S., Kono,H. and Sarai,A. (2004) Intermolecular and intramolecular readout mechanisms in protein-DNA recognition. *J. Mol. Biol.*, **337**, 285–294.
6. Stawiski,E.W., Gregoret,L.M. and Mandel-Gutfreund,Y. (2003) Annotating nucleic acid-binding function based on protein structure. *J. Mol. Biol.*, **326**, 1065–1079.
7. 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.
8. Brown,C., Campos-Leon,K., Strickland,M., Williams,C., Fairweather,V., Brady,R.L., Crump,M.P. and Gaston,K. (2011) Protein flexibility directs DNA recognition by the papillomavirus E2 proteins. *Nucleic Acids Res.*, **39**, 2969–2980.
9. Little,E.J., Babic,A.C. and Horton,N.C. (2008) Early interrogation and recognition of DNA sequence by indirect readout. *Structure*, **16**, 1828–1837.
10. Cherney,L.T., Cherney,M.M., Garen,C.R. and James,M.N. (2009) The structure of the arginine repressor from Mycobacterium tuberculosis bound with its DNA operator and Co-repressor, L-arginine. *J. Mol. Biol.*, **388**, 85–97.
11. Marcovitz,A. and Levy,Y. (2011) Frustration in protein-DNA binding influences conformational switching and target search kinetics. *Proc. Natl Acad. Sci. USA*, **108**, 17957–17962.
12. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
13. Marathe,A., Karandur,D. and Bansal,M. (2009) Small local variations in B-form DNA lead to a large variety of global geometries which can accommodate most DNA-binding protein motifs. *BMC Struct. Biol.*, **9**, 24.
14. 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.
15. Williams,L.D. and Maher,L.J. 3rd (2000) Electrostatic mechanisms of DNA deformation. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 497–521.
16. Pabo,C.O. and Sauer,R.T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.*, **61**, 1053–1095.
17. Luscombe,N.M. and Thornton,J.M. (2002) Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity. *J. Mol. Biol.*, **320**, 991–1009.
18. Olson,W.K., Gorin,A.A., Lu,X.J., Hock,L.M. and Zhurkin,V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl Acad. Sci. USA*, **95**, 11163–11168.
19. Dickerson,R.E. and Chiu,T.K. (1997) Helix bending as a factor in protein/DNA recognition. *Biopolymers*, **44**, 361–403.
20. Marini,J.C., Levene,S.D., Crothers,D.M. and Englund,P.T. (1982) Bent helical structure in kinetoplast DNA. *Proc. Natl Acad. Sci. USA*, **79**, 7664–7668.
21. Hagerman,P.J. (1984) Evidence for the existence of stable curvature of DNA in solution. *Proc. Natl Acad. Sci. USA*, **81**, 4632–4636.
22. Diekmann,S. and Wang,J.C. (1985) On the sequence determinants and flexibility of the kinetoplast DNA fragment with abnormal gel electrophoretic mobilities. *J. Mol. Biol.*, **186**, 1–11.
23. Bertrand,H., Ha-Duong,T., Fermandjian,S. and Hartmann,B. (1998) Flexibility of the B-DNA backbone: effects of local and neighbouring sequences on pyrimidine-purine steps. *Nucleic Acids Res.*, **26**, 1261–1267.
24. Vologodskii,A. (2009) Determining protein-induced DNA bending in force-extension experiments: theoretical analysis. *Biophys. J.*, **96**, 3591–3599.
25. Nair,T.M. (2010) Sequence periodicity in nucleosomal DNA and intrinsic curvature. *BMC Struct. Biol.*, **10(Suppl 1)**, S8.
26. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
27. Shui,X., McFail-Isom,L., Hu,G.G. and Williams,L.D. (1998) The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry*, **37**, 8341–8355.
28. Fuxreiter,M., Simon,I. and Bondos,S. (2011) Dynamic protein-DNA recognition: beyond what can be seen. *Trends Biochem. Sci.*, **36**, 415–423.
29. Blanchet,C., Pasi,M., Zakrzewska,K. and Lavery,R. (2011) CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.*, **39**, W68–73.
30. Mirzabekov,A.D. and Rich,A. (1979) Asymmetric lateral distribution of unshielded phosphate groups in nucleosomal DNA and its role in DNA bending. *Proc. Natl Acad. Sci. USA*, **76**, 1118–1121.
31. Manning,G.S., Ebralidse,K.K., Mirzabekov,A.D. and Rich,A. (1989) An estimate of the extent of folding of nucleosomal DNA by laterally asymmetric neutralization of phosphate groups. *J. Biomol. Struct. Dyn.*, **6**, 877–889.
32. Strauss,J.K. and Maher,L.J. 3rd (1994) DNA bending by asymmetric phosphate neutralization. *Science*, **266**, 1829–1834.
33. Strauss-Soukup,J.K., Rodrigues,P.D. and Maher,L.J. 3rd (1998) Effect of base composition on DNA bending by phosphate neutralization. *Biophys. Chem.*, **72**, 297–306.
34. Tomky,L.A., Strauss-Soukup,J.K. and Maher,L.J. 3rd (1998) Effects of phosphate neutralization on the shape of the AP-1 transcription factor binding site in duplex DNA. *Nucleic Acids Res.*, **26**, 2298–2305.
35. Hardwidge,P.R., Zimmerman,J.M. and Maher,L.J. 3rd (2002) Charge neutralization and DNA bending by the Escherichia coli catabolite activator protein. *Nucleic Acids Res.*, **30**, 1879–1885.