Positively Charged Residues in DNA-Binding Domains of Structural Proteins Follow Sequence-specific Positions of DNA Phosphate Groups

A. G. Cherstvy*

Institut für Festkörperforschung, Theorie-II, Forschungszentrum Jülich, 52425 Jülich, Germany

Received: November 13, 2008; Revised Manuscript Received: January 12, 2009

We study electrostatic charge complementarity along interfaces of DNA−protein complexes. We use the Protein Data Bank atomic coordinates of DNA−protein complexes for some DNA-binding proteins to study the distribution of positively charged protein residues in the close contact with DNA. We show that large structural proteins reveal a peculiar nonuniform distribution of Arg, Lys, and His amino acids in the frame of negatively charged DNA phosphate strands. We study the nucleosome core particles, DNA complexes with prokaryotic DNA-bending histone analogues, but also the basic binding motifs of small DNA-binding proteins. For large DNA−protein complexes, where extensive DNA wrapping around protein cores occurs, we show that positive amino acids on the proteins track sequence-specific positions of individual DNA phosphates. This specificity of electrostatic interactions can contribute to DNA recognition by DNA-binding proteins, which is governed for many DNA−protein complexes primarily by the hydrogen bond formation between protein residues and DNA bases.

I. Introduction

Recognition of DNA by DNA-binding proteins is vital for many biological processes.[1,2] DNA-binding proteins are extremely diverse in their structure, function, and binding motifs to DNA. Several kinds of DNA−protein interactions with different degree of sequence-specificity are often mentioned, with hydrogen bonding, electrostatic, and hydrophobic interactions being most abundant.[3−5] Proteins often bind to a variety of DNA base pair (bp) sequences with very different affinities that are controlled by peculiarities of sequence-specific DNA structure, the set of interactions formed with DNA, many environmental factors, as well as by the presence of helper proteins. All this complicates the construction of a transparent DNA−protein recognition code,[6−9] making it a probabilistic problem,[10] rather than one-to-one correspondence between the sequence of DNA bases and most suited protein residues bound to it.

Many DNA-binding proteins are capable of diffusing along DNA via sliding in steps of ∼100 to 1000 DNA bp before unbinding from the helix.[11] Some proteins find their targets on DNA at amazingly fast rates, up to 100−1000 times faster what is allowed by thermal diffusion in 3D solution, the so-called facilitated protein diffusion.[12−17,55] Some DNA-processing proteins such as RNA polymerase can perform a spiral-like motion along DNA duplex, instead of pure sliding.[18]

Electrostatic interactions are known to contribute substantially to the formation of both nonspecifically and specifically bound DNA−protein complexes. For the lac repressor, for instance,[19] the binding constant decreases dramatically upon addition of simple salt in solution and it depends strongly on the presence of divalent salts in the buffer. Many simple, relatively small (∼30 to 50 Å in diameter) DNA-sliding proteins often form about a dozen of charge−charge contacts with DNA phosphate groups, for example, the lac repressor.[20]

Many DNA-binding proteins are designed by Nature so that they contact negatively charged DNA phosphates mainly by the positively charged residues (Arg and Lys), placing negatively charged amino acids (Asp, Glu) further away from the contact with DNA,[21] see also Figure 1. For many DNA−protein complexes, the DNA binding domains of proteins are the top scores among protein patches with the most positive electrostatic potential. Even net negatively charged proteins often possess positive patches on their surfaces[22,23] that provide positive electrostatic affinity upon DNA binding to this area of the protein. DNA binding sites on protein surfaces often reveal remarkable shape complementary with the surface of DNA duplex and have electrostatic potentials more positive than the average value.[24] These charge−charge interactions are, however, often claimed to be mainly nonspecific with respect to DNA bp sequence.[25] They are thought to provide a close proximity of proteins to DNA, thereby allowing the formation of stronger, sequence-specific DNA−protein interactions, for example, hydrogen bonds with DNA bases.[26]

Recently, we have constructed a mechanism of DNA−protein electrostatic sequence recognition based on complementarity of charge patterns on DNA and DNA-recognition domains of a model protein.[27] The charges on the model protein (all positive) and DNA in this simple model were positioned on quasiperiodic one dimensional (1D) parallel arrays, with random displacements in charge positions. These random variations are intended to mimic the sequence specificity of DNA and protein molecular structure. For DNA, the helical structure is known to be quite corrugated and nonideal depending on DNA bp composition.[28] In the model, the protein charges thus tend to follow the periodic charge positioning of the "unwound spirals" of DNA phosphates. Electrostatic recognition of a particular DNA fragment by the protein charge lattice was introduced via identical charge displacements on this DNA target bp sequence and on the protein. The protein is thus attracted more strongly to this recognizable DNA segment, and a potential well for protein binding is formed in the vicinity of this DNA sequence. The potential well calculated in the model was shown to be deep enough to slow down the protein diffusion along DNA. The times of protein localization in the well were shown to be long

* E-mail: a.cherstvy@googlemail.com.

10.1021/jp810009s CCC: $40.75
Published on Web 03/03/2009

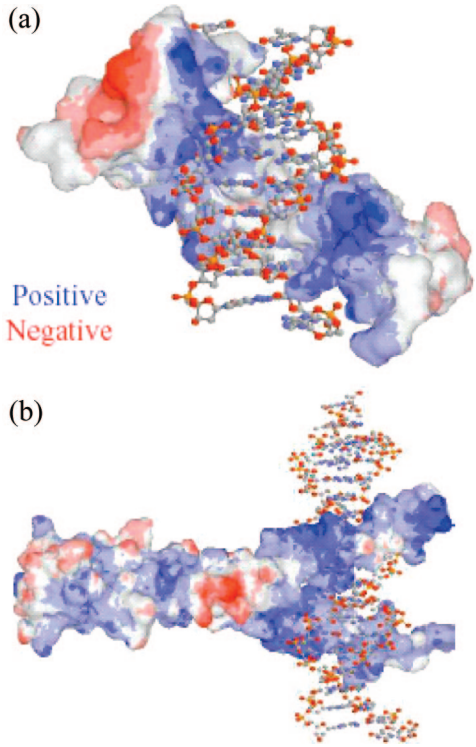

Figure 1. Electrostatic potential distribution on DNA−protein complexes 1aay (a) and 1ysa (b), as visualized by the MDL Chime and Protein Explorer programs.

enough to provoke a reorganization of the protein structure and allowing subsequent formation of stronger bonds with DNA. This model was a first attempt to elucidate the strength of possible bp specificity of DNA−protein electrostatic interactions.

In this paper, we study to what extent the helical symmetry of DNA charges can predispose the pattern of charges existing on some DNA-binding proteins in their complexes with DNA. Below, we test the assumption of our previous theoretical model about quasiperiodic charge patterns on the proteins, studying real DNA-binding domains of DNA-binding proteins from the Protein Data Bank (PDB) data files. We show that not only positively charged protein residues are much more abundant near the protein−DNA interface, but also that the positions of individual DNA phosphates can be tracked by positive protein charges in proximity of DNA. We perform a computational search of PDB data files for many DNA−protein complexes, concentrating on positions of positively charged side chain nitrogens of Arg, Lys, and His residues with respect to closest DNA phosphate groups.

II. Model and Calculations

We use a home-written Mathematica 6 program (available upon request) to operate with PDB files of atomic coordinates of DNA and amino acids in DNA−protein complexes. We extract the positions of side-chain nitrogens of positively charged residues (atoms NZ on Lys, NH1 on Arg, and ND1 on His) and assign a unit positive charge e_0 to each of them.[29] Negative charges are put on oxygens OD1 of Asp and OE1 of Glu. For both positive and negative residues we assign a unit charge rather than the fractional charges ($+0.5e_0$ on NH1 and NH2 of Arg, $-0.5e_0$ on OD1 and OD2 of Asp, $-0.5e_0$ on OE1 and OE2 of Glu). We do so for simplicity and in order to have in the model only a single contact with DNA for every DNA-close charged protein residue. As coordinates of negative DNA groups, we use PDB data for phosphate P atoms. Note that many PDB structures of complexes are obtained from crystals and thus involve DNA/protein deformations due to lattice constrains. Also note that the salt conditions, temperature, and buffers used upon crystallization can be quite different for DNA−protein complexes studied.

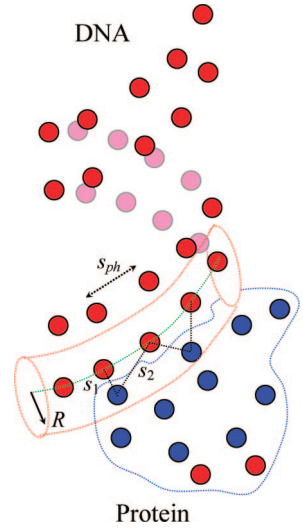

Figure 2. The definition of distances $s_{1,2}$ for protein positive charges (shown in blue) which are closer than some critical distance $r \approx l_B$ to negative charges of DNA phosphates (shown in red).

We separate the set of protein positive charges positioned closer than some critical distance to one of DNA phosphates. This distance is set to be from $r = 7$ to 11 Å, defining an artificial tube of radius $R \approx r$ around each helical DNA strand. The separation of ~7 Å is the Bjerrum length in water, the distance between two unit point charges interacting with energy k_BT. We use this distance as a "measure" of strong versus weak charge−charge DNA−protein contacts. Also, as the height of the DNA minor groove is ~14 Å, the tube radius of 6−7 Å allows us to assign the protein charges unambiguously to a particular closest DNA phosphate strand. Note that for thicker tubes the difference in interaction energy with two neighboring phosphates will decrease and bp specificity of these contacts will thus be smeared out. In the final set of close charges, the repetitive coordinates of protein positive charges that appear for particularly large tube radii are filtered out. We calculate the separations $s_{1,2}$ from every N^+ charge in this set to the two closest phosphates on the same DNA strand and their difference, $s_1 - s_2$, Figure 2, keeping track of the direction of DNA strands.[30] This DNA directionality tracking produces both positive and negative differences of the two separations calculated.

If the closest positive charges on the protein were distributed uniformly in space between neighboring DNA phosphates, then the histogram of $s_1 - s_2$ distributions must have been nearly constant in the vicinity of $s_1 - s_2 = 0$, being symmetric with respect to zero and decaying to zero after some critical value of the difference that is between $s_1 - s_2 \approx (s_{ph}^2 + R^2)^{1/2} - R$ and $s_1 - s_2 \approx s_{ph}$. Here, s_{ph} is the average phosphate−phosphate separation along the DNA helical strand that is ~7 Å for randomly sequenced B-DNA.

On the other hand, if protein positive charges prefer to be localized in the vicinity of closest phosphates, the distribution of $s_1 - s_2$ should reveal a well at $s_1 - s_2 = 0$, some barriers at a finite $s_1 - s_2$, and again decrease to zero at large enough $s_1 - s_2$. This well is an indication that the population of protein charges equally separated from the neighboring phosphates is

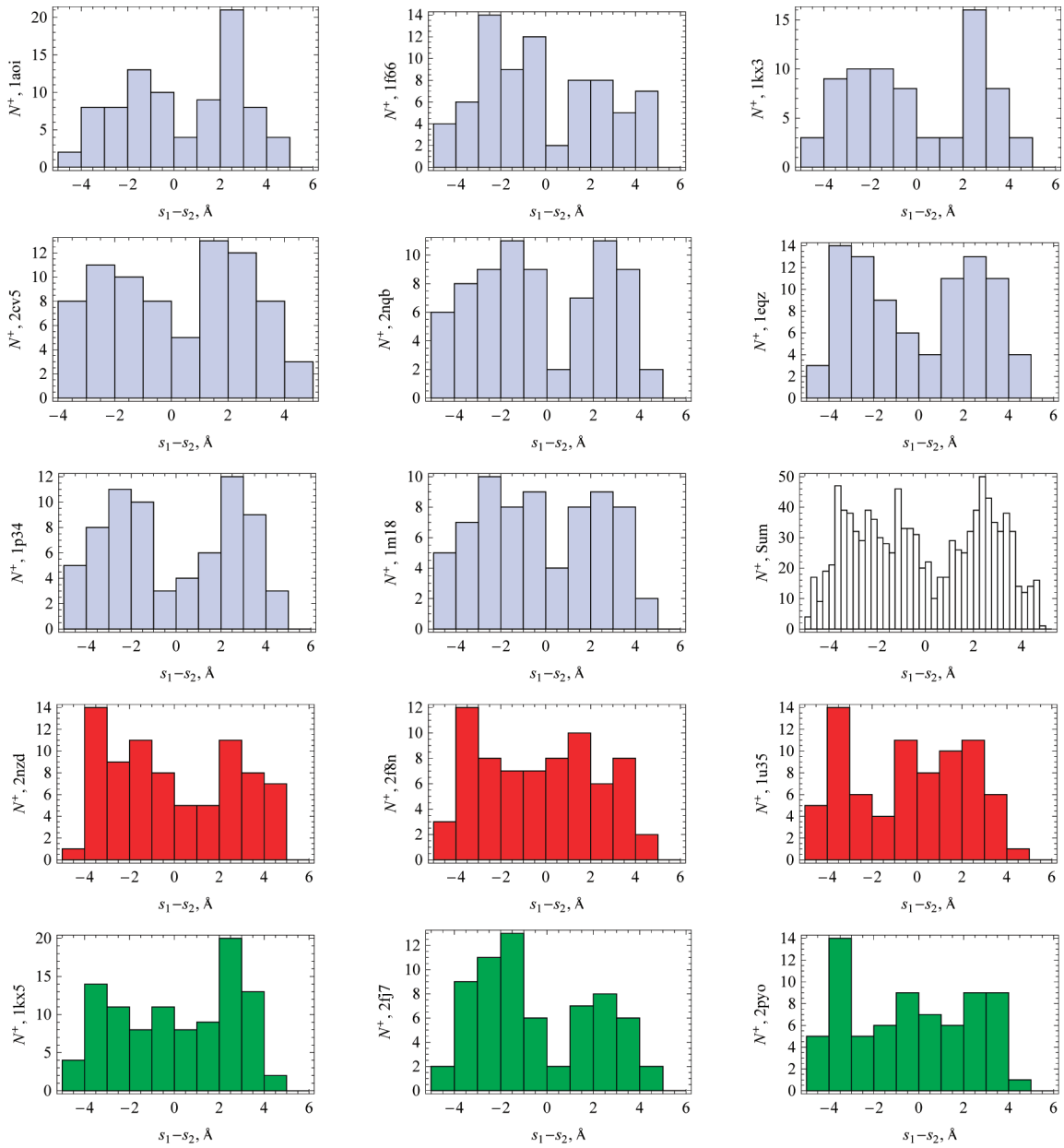

Figure 3. The distribution of histones positive charges on N^+ side chains that are closer than $r = 7$ Å to DNA phosphate groups, plotted as a function of $s_1 - s_2$. For 146 bp nucleosomes, the histograms reveal depleted populations between two neighboring DNA phosphates. We observe that for 1aoi, for NCP with modified histones H2AZ 1f66, and for 1kx3, all from African frog; human NCP 2cv5, fruit fly 2nqb, and chicken 1eqz nucleosomes (all shown in blue). The 145 bp DNA-histone complexes 1nzd, 2f8n, and 1u35 show no pronounced wells in the distributions (in red). Much more uniform $s_1 - s_2$ histograms are also obtained for 147 bp NCPs 1kx5, 2fj7, and 2pyo (in green). The overall distribution of closest positive histone charges for all NCP complexes studied above has a well at $s_1 - s_2 = 0$ (in white).

depleted. Ideally, the proteins should contain large enough number of positive residues close to DNA in order to accumulate a proper statistics on charge−charge contacts, which is not always the case for examples studied below.

III. Results

First, we consider the distribution of positive charges on the histone proteins in the basic unit of DNA compactification in eukaryotic cells, the nucleosome core particle (NCP).[31] DNA in these relatively large protein complexes is wrapped around a sphere-cylinder-like octamer of net positively charged histones in a left-handed superhelical fashion. DNA makes about 1 and 3/4 turns, with the radius of the DNA superhelix of ∼45 Å and separation between DNA turns of ∼28 Å. The coordinates of basic histones are well resolved in crystal structures, while the charges on Lys- and Arg-rich flexible histone tails, protruding from the nucleosome core through tight DNA superhelix to the outside, are typically not included in PDB data files.

DNA-histone complexation occurs partly/mainly because of electrostatic attraction between net positively charged core proteins and the stretch of DNA, with intact NCPs being stable only in a limited range of salinities of solution. DNA−histone recognition and positioning of NCPs on genomic DNAs is a complicated issue that involves sequence-specific DNA bendability and DNA−protein interactions. The later are of primary importance for understanding of DNA unwrapping from NCPs.[32] This unwrapping is likely to be required for DNA transcription in living cells, and its properties are important for mobility and equilibration of NCPs on long DNA in vitro.[33]

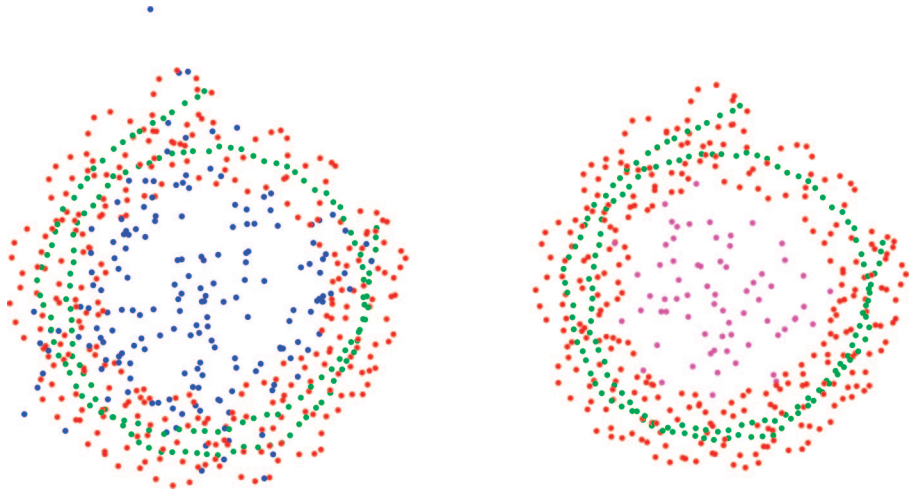

Figure 4. The distribution of positive (blue) and negative (magenta) histone charges for 146 bp NCP 1aoi, as viewed from the top. DNA phosphates are shown in red; the DNA axis defined as the middle between two phosphates on the same DNA base is in green. Positive protein charges are concentrated near the DNA, while the negative residues are located primarily inside the nucleosomal core.

We detect a clear well in distributions of positive charges on histones in NCPs with 146 bp DNA fragments at $s_1 - s_2 = 0$, see Figure 3 plotted for the critical distance of $r = 7$ Å. The height of every bar in histograms in this figure corresponds to the number of protein positive charges with $s_1 - s_2$ in corresponding $s_1 - s_2$ interval. Similar tendencies with the well at $s_1 - s_2 = 0$ and humps at nonzero $s_1 - s_2$ are obtained for larger "tube radii" R corresponding to $r = 11$ Å. If we do not include the charges of His ND1 atoms in calculations, then the histograms change only slightly because for these 146 bp DNA NCPs not typically more than 2−4 His ND1 are closer than 7 Å to DNA phosphates. The number of close positive histone charges is typically ∼75 to 100, whereas there are ∼160 to 220 positive amino acids in total on the histone octamers of NCPs studied.

Humps and wells are much less pronounced or disappear completely for noncanonical NCPs with 145 and 147 bp DNA fragments, red and green histograms, respectively, in Figure 3. The reasons for such differences in shape of distributions for 146 vs 145 and 147 NCPs are to be clarified by detailed future analysis. The majority of well-resolved canonical DNA−histone NCP complexes reveal, however, a clear two-peak distribution of $s_1 - s_2$. We believe that this tendency is of general importance for binding specificity and overall stability of these large structural DNA−protein complexes. The peculiarities in shape of protein charge distribution for the whole family of NCPs, shown in Figure 3 as the histogram in white, can slightly charge after new NCP PDB structures are analyzed and added.

Positive and negative histone charges are not distributed uniformly in the NCP core. Positively charged amino acids are located predominantly in the vicinity of DNA, whereas negative histone charges are concentrated in the core of nucleosomes, Figure 4. There are typically only ∼3 to 6 negative residues per NCP closer than 7 Å from DNA phosphates, out of ∼70 negative charges in total per NCP. For the NCPs considered, the distribution of closest O^- on Asp and Glu residues with respect to the neighboring phosphates is pretty uniform, in contrast to two-peak histograms for N^+ protein charges. The statistics of this O^- distribution is, however, quite poor because negative protein charges are typically farther away from the DNA.

Another class of proteins we consider is the family of DNA-bending proteins of prokaryotes.[34] Similarly to DNA compactification in eukaryotes via the nucleosome formation, these structural proteins bend DNA sharply on small length-scales in extreme U-turn-like fashion. Also, these architectural proteins are likely to use asymmetric neutralization of DNA phosphates that takes place along the protein−DNA contact surface and thus facilitates the electrostatically induced DNA wrapping around the protein core.[35−37]

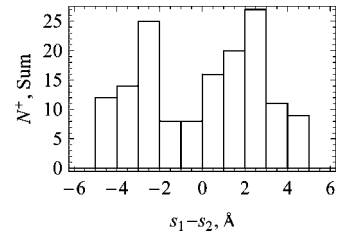

Figure 5. The distribution of DNA-closest positive protein charges extracted from DNA-bending prokaryotic complexes 2np2, 1ihf, 1p51, 1p71, 1p78, 1ouz, 1owf, and 1owg, calculated for $r = 7$ Å.

Complexes of these proteins with DNA have relatively extended contacts, and the crystals are often obtained with long enough DNA fragments. The complexes contain typically 15−25 closest protein charges (at $r = 7$Å), 30−40 protein positive charges in total, and 20−35 bp long fragments of bent DNA. This allows us to perform the analysis of charge−charge separations, although the statistics for every individual complex is poor. We extract $s_1 - s_2$ values for several DNA-bending complexes that possess a high degree of structural and functional homology of protein residues and reveal very similar structures of bent DNA fragments. Similar to NCPs, we observe a two-peak distribution of $s_1 - s_2$ for prokaryotic DNA-bending proteins bound to a stretch of DNA, Figure 5. In some of these complexes, one or two DNA strands are discontinuous. Also, sometimes positively charged N atoms on Lys and Arg are not resolved and are missing in PDB files, and we do not account for such amino acids in calculations.

Next, we consider typical DNA−protein complexes containing some basics motifs of protein binding to the DNA: (a) zinc fingers,[38] (b) leucine zippers,[39] and (c) helix-turn-helix-like complexes.[40] These are relatively small DNA−protein complexes that control and regulate DNA transcription in cells. The α-helices of these proteins in contact with the DNA typically recognize (a) a stretch of DNA major groove via fitting α-helices into it, (b) two locations in the major groove on opposite sides

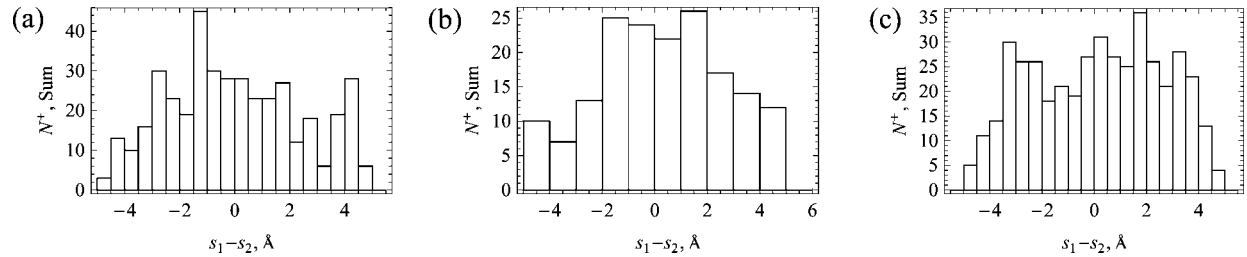

Figure 6. The same as in Figure 5 but for (a) zinc finger containing DNA−protein complexes 1aay, 1a1l, 1p47, 1jk1, 1jk2, 1a1f, 1a1g, 1a1j, 1a1k, 1a1h, 1a1i, 1zaa, 1g2f, 1g2d, 1f2i, 1llm, 1mey, 1ubd, 1tf3, 2jp9, 2gli, 3dfx, calculated with $r = 8$ Å; (b) leucine zipper based DNA−protein complexes 1ysa, 2c9l, 2c9n, 2h7h, 1d66, 1fos, 1gu5, 1hjb, 1jnm, at $r = 7$Å. (c) lac-, lambda-, 434-, cro-, arc-, and repressor-like DNA−protein complexes (1osl, 1l1m, 2bjc, 1cjg; 1lmb, 3bdn, 6cro, 1lli, 1rio; 1par, 1bdt, 1bdv, 2bnz, 2cax; 1au7, 2or1, 1per, 3cro, 1rpe, 2p5l, 1gt0, 1hf0, 1ic8, 1o4x, 2r1j) and CAP proteins (1cgp, 1zrc, 1zrd, 1zre, 1zrf, 1o3q, 1o3r, 1o3s, 1j59, 1run, 2cgp), at $r = 8$ Å.

of DNA, and (c) two successive major grooves on one side of DNA. The DNA recognition mechanism in these complexes is known to be based on formation of hydrogen bonds with DNA bases, hydrogen bonds and electrostatic contacts with the phosphates, and probably van der Waals interactions with exposed DNA bases.[41] The helix-turn-helix motif is characteristic for Cro, lac, and lambda repressors as well as CAP proteins that often share similar structure and amino acid sequences in DNA binding regions. We consider below some representative complexes in each of these structural domain families, without any classification as to the strength and mode of binding (specific vs nonspecific complexes) to a particular DNA fragment.

For zinc finger complexes with DNA analyzed, the distribution $s_1 - s_2$ for positive protein charges does not have a two-peak character, Figure 6a. The complexes analyzed typically have 15−20 close positive protein charges (at $r = 8$ Å), with 25−40 positive charges in total, and 12−20 DNA bp cocrystallized with the protein unit. These short DNA lengths can lead to inaccuracy in the results because for some protein charges the closest phosphates are not included in PDB files. Also, for some complexes not a single DNA−protein unit is given in the PDB crystals, but rather a dimer or trimer of crystal units.

For DNA binding proteins based on the leucine zipper motif, the combined charge distributions also do not show a well at $s_1 - s_2 = 0$ being rather one-hump-like ones, Figure 6b. The complexes analyzed have typically 12−16 close positive protein charges (at $r = 7$Å), with 33−40 positive charges in total, and 12−20 DNA bp-long fragments.

For the helix-turn-helix like complexes, the distribution of N^+ on Lys, Arg, and His does not show a clear well either, Figure 6c. For some repressor−DNA complexes we take only one from several structures available in PDB files, for example, for lac repressor 1l1m and 1osl complexes.

IV. Discussion

We have shown that in some DNA−protein complexes, in particular in the NCPs, the N^+ atoms on the protein Arg and Lys residues prefer to be localized in the vicinity of closest DNA phosphates. As 3D spatial distribution of DNA phosphates is strongly specific to DNA sequence, the positive charges of these proteins track the sequence of the underlying DNA. This indicates sequence specificity of electrostatic binding of these proteins to DNA. For small, simple DNA−protein structures studied, on the contrary, we do not observe this preference, and the distribution of protein positive charges along DNA phosphate strands is pretty uniform, without any localization near the individual phosphates.

Note that Arg and Lys N^+ proximity to DNA phosphates might not necessarily be a prerequisite of sequence-specific electrostatic DNA−protein interactions. It can be a consequence of strongly bp-specific hydrogen bond formation between protein residues and DNA bases.[42] The formation of hydrogen bonds is known to depend much stronger upon the distance and orientation of reacting groups, as compared to charge−charge interactions. The analysis of correlations of hydrogen bonding and van der Waals contacts along the interface of DNA−protein complexes is a natural continuation of this "purely electrostatic" project and will be studied separately.

Indeed, the preferences for hydrogen bond formation between protein residues and DNA bases have been discovered through detailed statistical analysis of the varieties of DNA−protein complexes.[43] In particular, positively charged Arg and Lys were shown to have extremely strong binding preference to the DNA Guanine (interactions of Arg NH1/NH2 and Lys NZ atoms with O6 and N7 atoms of Guanine), Asn and Gln amino acids revealed stronger affinity to the DNA Adenine, whereas negatively charged Glu and Asp tend to bind more frequently to the DNA Cytosine. Also, Arg and Lys are involved more often in water-mediated hydrogen bonds than is anticipated from random DNA−protein docking.

The extension of this study to other classes of DNA−protein complexes as well as a categorization of complex families based on specificity of their electrostatic interactions with DNA are also possible targets for the future research. The criteria might include strongly versus weakly sequence-specific protein binding, nonspecifically versus specifically bound complexes, and strong versus weak salt dependence of the protein binding to DNA. Not only general tendencies in positioning of protein charges with respect to DNA phosphates in DNA−protein complexes can be studied, but also the electrostatic binding affinity and specificity for cognate versus noncognate DNA sequences[44−46] can be analyzed in the future. The results can be compared with binding to ideally helical, "model" DNA fragment. This task, however, involves a complicated issue of the dielectric "constant" in space between DNA and protein,[47,48] the value of which dramatically affects the resulting strength of electrostatic interactions. Also, one should account for the fact that the protein charges "contacting" DNA and positioned on the protein−water boundary will give different contributions to DNA−protein binding strength as compared to charges at the same distance from DNA but inside the protein core.

Tracking of individual DNA phosphates predicted in this paper in particular for NCPs is akin to electrostatic complementarity known to take place upon formation of protein−protein complexes[49,50] and in protein−protein docking.[51] The interface of protein−protein complexes works in many cases as a charge zipper. Namely, positively charged patterns on one protein oppose predominantly negatively charged patches of residues

on the surface of another protein along their contact area. The native protein−protein structures are designed to have one of the best possible complementarity and electrostatic affinity of the components, as compared to the non-native structures.[52]

The idea of electrostatic complementarity also finds its applications in studying electrostatic interactions of DNA duplexes.[53,54] Namely, for two parallel juxtaposed DNA fragments the regions of negative electrostatic potential close to DNA phosphate strands on one molecule can be brought into a perfect match with the regions of positive potential in the grooves of another molecule along the DNA−DNA contact (provided a strong adsorption of counterions into DNA grooves takes place). As a result, attractive DNA−DNA intermolecular electrostatic forces emerge at separations of about 10−20 Å between DNA surfaces, even for randomly sequenced DNAs with a residual net negative charge.

V. Conclusions

We have shown that the distribution of positively charged protein amino acids along the DNA strands in DNA−protein structural complexes is nonuniform. In particular, for histone−DNA complexes in 146 bp NCPs and for prokaryotic histone-like proteins, a pronounced two-peak distribution of Arg and Lys N^+ in the frame of two neighboring DNA phosphates is observed. This indicates that not only the histone positive charges are concentrated in vicinity of DNA, as one could expect from the preference of their attractive electrostatic interactions, but also the positions of individual phosphates, known to be strongly dependent on DNA bp sequence, are tracked by histone positive amino acids in their vicinity. Whether this effect of sequence-dependent electrostatic interactions predicted for NCPs can be observed in experiments and whether it exists for other families of DNA−protein complexes is to be clarified. In contrast, for many simple, small DNA−protein complexes that use basic motifs of binding to DNA, the statistics of distribution of positive charges with respect to neighboring DNA phosphates is rather uniform. This indicates that these complexes statistically have no build-in preference for structuring of positive charges in protein architecture, although particular complexes can reveal a different degree of electrostatic binding specificity to DNA.

Acknowledgment. I acknowledge the support by the Deutsche Forschungsgemeinschaft, DFG Grant CH 707/2-1.

References and Notes

(1) Pabo, C. O.; Nekludova, L. *J. Mol. Biol.* **2000**, *301*, 597.
(2) Paillard, G.; Lavery, R. *Structure* **2004**, *12*, 113.
(3) von Hippel, P. H. *Science* **1994**, *263*, 769.
(4) Honig, B.; Nicholis, A. *Science* **1995**, *268*, 1144.
(5) Branden C.; Tooze J., *Introduction to Protein Structure*; Garland Publishing, Inc., New York and London, 1991.
(6) Lavery, R. *Q. Rev. Biophys.* **2005**, *38*, 339.
(7) Deremble, C.; Lavery, R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 171.
(8) von Hippel, P. H. *Annu. Rev. Biophys. Biomol. Str.* **2007**, *36*, 79, and references cited therein.
(9) Widom, J. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16909.
(10) Benos, P. V. *Bioessays* **2002**, *24*, 466.
(11) Wang, Y. M.; Austin, R. H.; Cox, E. C. *Phys. Rev. Lett.* **2006**, *97*, 048302.
(12) Winter, R. B.; Berg, O. G.; von Hippel, P. H. *Biochemistry* **1981**, *20*, 6961.
(13) Slutsky, M.; Kardar, M.; Mirny, L. A. *Phys. Rev. E* **2004**, *69*, 061903.
(14) Halford, S. I.; Marko, S. I. *Nucleic Acids Res.* **2004**, *32*, 3040.
(15) Hu, T.; Grosberg, A. Y.; Shklovskii, B. I. *Biophys. J.* **2006**, *90*, 2731.
(16) Bonnet, I.; et al. *Nucleic Acids Res.* **2008**, *36*, 4118.
(17) van den Broek, B.; et al. *PNAS* **2008**, *105*, 15738.
(18) Sakata-Sogawa, K.; Shimamoto, N.; et al. *PNAS* **2004**, *101*, 14731.
(19) Record Jr, M. T.; deHaseth, P. L.; Lohman, T. M. *Biochemistry* **1981**, *16*, 4791.
(20) Kalodimos, C. G.; et al. *Science* **2004**, *305*, 386.
(21) Jones, S.; et al. *Nucleic Acids Res.* **2003**, *24*, 7189.
(22) Stawiski, E. W.; et al. *J. Mol. Biol.* **2003**, *326*, 1065.
(23) Fogolari, F.; et al. *J. Mol. Recogn.* **2002**, *15*, 377.
(24) Tsuchiya, Y.; et al. *Proteins* **2004**, *55*, 885, and references cited therein.
(25) von Hippel, P. H.; Berg, O. G. *PNAS* **1986**, *83*, 1608.
(26) Cheng, A. C.; et al. *J. Mol. Biol.* **2003**, *327*, 781.
(27) Cherstvy, A. G.; Kolomeisky, A. A.; Kornyshev, A. A. *J. Phys. Chem. B* **2008**, *112*, 4741.
(28) Olson, W. K.; et al. *PNAS* **1998**, *95*, 11163.
(29) The charge state of an ionizable group on protein side chains depend on many factors. The pK_a value of the group, the pH value, ionic strength, local surrounding on the protein surface (charged, polar, hydrophobic), the group's position with respect to the protein−water interface (local dielectric permittivity), and protein shape in its vicinity, all can affect the final charge state of titratable groups in DNA−protein and protein−protein complexes. We apply a very simplified approach in this study. The NZ atoms of Lys and NH1/NH2 atoms of Arg have pK_a values of about 10.5 and 12.5, respectively, and thus typically carry a unit positive charge in complexes. The pK_a of ND1 of His is about 6.5, i.e., it can typically either be positively charged or neutral. For simplicity, we put a positive charge on it in all DNA−protein complexes studied.
(30) Instead of calculating $s_1 - s_2$, one could in principle perform a local 2D Fourier transform for the closest protein positive charges in the frame of DNA. This involves an axial shift and azimuthal rotation of N^+ coordinates near every DNA bp. If protein charges do track DNA phosphate positions, the Fourier-transformed N^+ positions in DNA frame ($\sim 2\pi$ rotation per $\sim$34 Å along the DNA axis) should reveal a peak in the k-space at $\sim 1/s_{ph} \approx 1/(7$ Å). The DNA in DNA−protein complexes is, however, strongly deformed, being helically nonideal (bent and twisted). The definition of DNA local helical axis, the bp-specificity of DNA twist angles, and DNA irregularities originating from other bp-specific DNA parameters (shift, rise, etc.) make, however, such a Fourier method difficult to implement for DNA−protein complexes. We therefore use a more transparent model for treatment of bp-specificity in charge positioning along DNA−protein interfaces.
(31) Luger, K.; et al. *Nature* **1997**, *389*, 251.
(32) Arcesi, L.; La Penna, G.; Perico, A. *Biopolymers* **2007**, *86*, 127.
(33) Widom, J.; et al. *Nature* **2006**, *442*, 772.
(34) Rice, P. A.; et al. *Cell* **1996**, *87*, 1295.
(35) Mirzabekov, A. D.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 1118.
(36) Matthew, J. B.; Ohlendorf, D. H. *J. Biol. Chem.* **1985**, *260*, 5860.
(37) Cherstvy, A. G.; Winkler, R. G. *J. Phys. Chem. B* **2005**, *109*, 2962.
(38) Elrod-Erickson, M.; et al. *Structure* **1996**, *4*, 1171.
(39) Ellenberger, T. E.; et al. *Cell* **1992**, *71*, 1223.
(40) Jordan, S. R.; Pabo, C. O. *Science* **1988**, *242*, 893.
(41) Garvie, C. W.; Wolberger, C. *Mol. Cell* **2001**, *8*, 937.
(42) Mandel-Gutfreund, Y.; et al. *J. Mol. Biol.* **1995**, *253*, 370.
(43) Luscombe, N. M.; et al. *Nucleic Acids Res.* **2001**, *29*, 2860.
(44) Stormo, G. D.; Fields, D. S. *Trends Biochem. Sci.* **1998**, *23*, 109.
(45) Spolar, R. S.; Th, M.; et al. *Science* **1994**, *263*, 777.
(46) Bakk, A.; Metzler, R. *FEBS Lett.* **2004**, *563*, 66.
(47) Warshel, A.; Russell, T. S. *Q. Rev. Biophys.* **1984**, *17*, 3.
(48) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301.
(49) Kumar, S.; Nussinov, R. *Chem. Bio. Chem.* **2002**, *3*, 604.
(50) Sheinerman, F. B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153.
(51) McCoy, A. J.; et al. *J. Mol. Biol.* **1997**, *268*, 570, and references cited therein.
(52) Norel, R. *Protein Sci.* **2001**, *10*, 2147.
(53) Cherstvy, A. G.; Kornyshev, A. A.; Leikin, S. *J. Phys. Chem. B* **2002**, *106*, 13362.
(54) Cherstvy, A. G.; Kornyshev, A. A. *J. Phys. Chem. B* **2005**, *109*, 13024.
(55) Givaty, O.; Levy, Y. *J. Mol. Biol.* **2009**, *385*, 1087.

JP810009S