

Electrostatic hot spot on DNA-binding domains mediates phosphate desolvation and the pre-organization of specificity determinant side chains

Alpay N. Temiz, Panayiotis V. Benos and Carlos J. Camacho*

Department of Computational Biology, School of Medicine, University of Pittsburgh, 3501 Fifth Avenue, Biomedical Science Tower 3, Pittsburgh, PA 15260, USA

Received September 6, 2009; Revised October 27, 2009; Accepted November 17, 2009

ABSTRACT

A major obstacle towards elucidating the molecular basis of transcriptional regulation is the lack of a detailed understanding of the interplay between non-specific and specific protein–DNA interactions. Based on molecular dynamics simulations of C₂H₂ zinc fingers (ZFs) and engrailed homeodomain transcription factors (TFs), we show that each of the studied DNA-binding domains has a set of highly constrained side chains in preset configurations ready to form hydrogen bonds with the DNA backbone. Interestingly, those domains that bury their recognition helix into the major groove are found to have an electrostatic hot spot for Cl[−] ions located on the same binding cavity as the most buried DNA phosphate. The spot is characterized by three protein hydrogen bond donors, often including two basic side chains. If bound, Cl[−] ions, likely mimicking phosphates, steer side chains that end up forming specific contacts with bases into bound-like conformations. These findings are consistent with a multi-step DNA-binding mechanism in which a pre-organized set of TF side chains assist in the desolvation of phosphates into well defined sites, prompting the re-organization of specificity determining side chains into conformations suitable for the recognition of their cognate sequence.

INTRODUCTION

Understanding the interactions responsible for DNA recognition is critical to reveal the mechanisms of several cellular processes including transcription, replication,

modification and restriction. Although in many cases the target sequence for a given DNA-binding protein is surrounded by a long stretch of non-specific genomic DNA sequences, proteins are able to find their binding sites very efficiently. It has long been accepted that proteins scan binding sites using a mechanism consisting of one-dimensional (sliding) and three-dimensional search (hopping) (1,2).

Initially, specific DNA recognition was thought to involve a limited number of hydrogen bonds between the protein side chains and DNA bases (3). It has now become clear that besides electrostatics, water molecules and solvation effects (4), shape complementarity of protein and DNA, sequence dependent DNA deformability and the physiological environment can also play critical roles in DNA recognition (5–11). The emergence of specific protein–DNA complex structures in recent years has been instrumental in our understanding of how proteins recognize specific DNA sequences (12). Molecular dynamics (MD) simulations of protein–DNA complexes have provided insights on the dynamics of the interactions and the role of water at the complex interface (13–16). However, the molecular basis of the events leading to protein–DNA recognition and binding specificity is not yet fully understood. Part of the problem is that, when DNA is involved, interactions are dominated by charged and polar groups that are highly dependent on the solvent and ionic environment (5,17–24). Positively charged counter ions associate with the negatively charged phosphate groups of nucleic acids, thus maintaining neutrality in solution. Theoretical studies of protein–DNA complexes that concentrate on the effects of counter ions have mainly used two approaches, counter ion condensation (CC) (25,26) and Poisson–Boltzmann (PB) (27) theories. The main difference between these two approaches lies in the description of salt effects around the nucleic acid. CC theory considers two

*To whom correspondence should be addressed. Tel: +1 412 648 3333; Fax: +1 412 648 3163; Email: ccamacho@pitt.edu

distinct layers of counter ion concentration, one uniform layer around the DNA and a distant salt dependent classical ion atmosphere. PB theory, on the other hand, describes the ionic environment as a continuum.

In protein–protein interactions, the role of ions is described by the screening of long-range electrostatics consistent with classical Debye–Huckel theory (28). Indeed, ionic strength has been shown to tune the association rate of some highly optimized receptor–ligand systems by as much as five orders of magnitude (29). Generally, increasing ionic strength results in decreasing binding affinity. Strikingly, Jen–Jacobson and collaborators (20,23) and others (22) have shown that the latter is not necessarily true for protein–DNA interactions, where in several instances it has been shown that *the binding affinity decreases with decreasing ionic strength* (18,19,21,24,30,31) even when the experimental conditions, and individual thermodynamic parameters are different. One example of the role of salt in protein–DNA interactions is catabolite gene activator protein (CAP) binding to the lac promoter region, which also involves DNA bending. For this system, Fried and Stickle showed that for physiologically relevant salt concentrations between 0.05 and 0.2 M, there is a 5-fold *increase* in binding affinity when the salt concentration is increased (22). Similar behavior has been observed in *Escherichia coli* lac repressor–operator complex (18,19), EcoRI (20) and EcoRV (23) in 0–0.1 M range. The aforementioned CC phenomenological picture is actually able to fit the changes in binding affinity as a function of ionic strength, but no molecular/structural mechanism is revealed. Although grouping all of the above studies with one model may lead to oversimplified models, experimental observations still direct us to the question of whether there is a common underlying role played by ions affecting the dynamics and mechanism of specific and non-specific binding at the molecular level.

Motivated by the principle of pre-organization observed in protein–protein interactions (32,33), our goal here is to study its applicability in protein–DNA interactions in the presence of counter ions. Specifically, we start from the observation (33) which predicts that in the absence of their binding partners, side chains that bury a large amount of solvent accessible surface area (SASA) in the acceptor protein sample bound-like conformations for a significant amount of time, in order to facilitate the rapid formation of a bound-like intermediate state as the first step towards the formation of the high affinity complex. The structural properties of the C₂H₂ zinc finger (ZF) (34,35) and engrailed homeodomain (36) transcription factors (TFs) make these proteins ideal to carry out this analysis. Indeed, from a practical point of view, both their size and relatively small backbone structural rearrangement upon binding are crucial for detailed molecular analyses and simulations. Moreover, revealing the structural basis of recognition for these TFs is important in its own right, since they are some of the largest family of nucleic acid binding factors in eukaryotes.

Our hypothesis is that side chains important for recognition (i.e. side chains that contact DNA) have an inherent bound-like dynamics prior to the encounter with DNA.

We show that eight C₂H₂ ZF domains from three TFs and the engrailed homeodomain share this property, though some key solvated side chains are bound like only in the ion-rich environment of DNA. We find that a vast network of side chains that form hydrogen bonds with the DNA backbone is always constrained into bound-like configurations. Based on MD simulations of TFs without DNA, we show that as TFs are exposed to an increasing number of negatively charged Cl[−] counter ions, key side chains rearrange into more bound-like conformations. Analysis of the MD trajectories show that the increasing bound-like ordering of side chains is due to ions that physically interact with the TFs at an electrostatic hot spot located at the same site where negatively charged DNA phosphates bind. In the systems we analyzed, this site is characterized by at least a pair of basic groups on either side of the ion and a third hydrogen bond donor. For ZFs, the latter corresponds to the conserved histidine at helix position +7. These findings suggest a mechanism where TFs in the presence of the DNA backbone will rapidly equilibrate into bound like interactions that lock the phosphates into well-defined hot spots, resulting in a weakly stable non-specific binding complex, while side chains involved in hydrogen bond contacts with DNA bases are also well posed for triggering specific interactions in the presence of their corresponding counterparts. Given the relatively high concentration of counter ions near the DNA, we speculate that Cl[−] might also act as surrogates for the phosphate groups, not only restraining critical side chains in bound-like conformations but also by providing a natural competing substrate for the positive ions that actively neutralize the DNA backbone (17,20,22) prior to binding. Namely, if upon association these counter ions (NaCl) get closer than their solubility threshold of ~6 Å, ions might prefer to move to bulk water, contributing towards the desolvation of the binding interface. In summary, MD simulations provide a more detailed picture of side chains involved in direct physical interactions with DNA, and the role of electrostatics in protein–DNA binding (3,37,38).

METHODS

DNA-binding domains

C₂H₂ zinc fingers and homeodomains are two of the largest families of transcription factors in eukaryotes, both having highly stable folds. More importantly, since their function is mostly to position themselves on the DNA without significantly affecting its structure, it is reasonable to expect that they have a similar binding mechanism. Arguably, enzyme binding might have other requirements.

The C₂H₂ family contains proteins that have two or more ZF domains that work together in a modular fashion to recognize specifically the corresponding DNA target sequences. The classical C₂H₂ ZF domain is composed of a ββ α fold that typically interacts with three to four base-pairs of DNA using key residues in the α -helix (Figure 1). The ZF-fold is held together by a

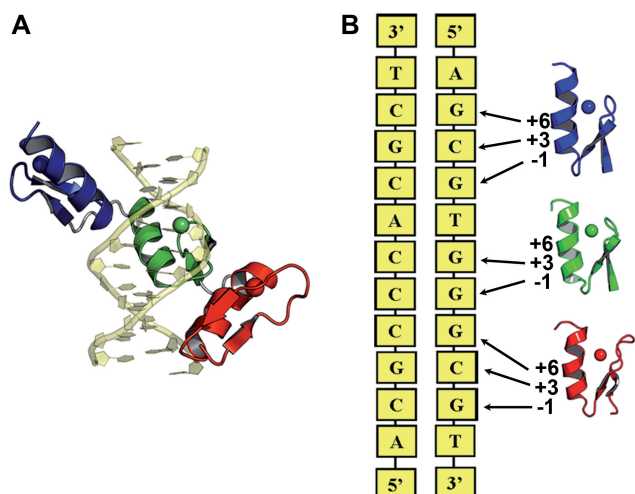


Figure 1. EGR–DNA complex and protein–DNA interactions. (A) Cartoon of EGR bound to DNA [PDB code: 1AAAY (39)]. Fingers I, II and III are color red, green and blue, respectively. (B) Schematic diagram showing key side chain–base interactions between EGR and its target DNA site at positions -1 , $+3$ and $+6$ of binding helix.

tetrahedrally coordinated zinc ion and a small hydrophobic core (39).

Mouse early growth response (EGR) factor. EGR factor (39) (Protein Data Bank (PDB) code 1AAAY) protein has three ZFs (Figure 1A). The α -helix of each finger fits into the major groove of DNA, forming specific contacts with DNA bases. Finger I (FI) binds to the GCG triplet near the 3' end of the primary DNA strand (Figure 1B). FII binds to the TGG triplet in the center and FIII binds to the GCG triplet near the 5' end of the primary DNA strand (Figure 1B). The helical domains of FI and FIII have the same sequence and identical bound structure with DNA. Their critical residues are an Arg preceding the α -helix (Arg $_{-1}$), an Asp on the second position of the α -helix (Asp $_{+2}$), a Glu on the third position (Glu $_{+3}$) and an Arg on the sixth position of the α -helix (Arg $_{+6}$). FII also has an Arg immediately before the helix and an Asp on position 2 of the helix. But it has a His on the third position of the helix (His $_{+3}$) and a Thr on the sixth position of the helix (Thr $_{+6}$). Arg residues in all three fingers form a pair of hydrogen bonds with guanines. EGR also has several contacts with the DNA phosphate backbone (see below). In particular, the first His (His $_{+7}$) coordinating the zinc ion forms a hydrogen bond to a phosphate on the primary strand. A conserved Arg on the second β -strand in each finger also contacts a phosphate on the primary strand.

Xenopus laevis transcription factor IIIA (TFIIIA). TFIIIA has six fingers in the crystal structure (40) (PDB code 1TF6). Fingers I–III wrap around the major groove of DNA like those of EGR. Fingers IV–VI form an open structure. Only FV have contacts with DNA in the major groove. Fingers I, II and V have $+3$ and $+6$ contacts, and the conserved His $_{+7}$ –phosphate contact

similar to those of EGR. In addition to these contacts FIII has an additional $+10$ contact (40).

Human glioblastoma (GLI) protein. GLI protein has five fingers in the crystal structure (41) (PDB code 2GLI). FI does not make any DNA contacts but makes extensive protein–protein contacts with FII (41). FII and FIII have contacts to DNA backbone. FIV and FV make extensive contacts with the DNA. FIV has contacts through $+1$, $+2$, $+3$ and $+6$ positions of the α -helix and FV has contacts through -1 , $+2$, $+3$, $+5$ and $+6$ positions.

Engrailed homeodomain protein. Engrailed homeodomain protein (36) (PDB code: 3HDD) forms a globular fold consisting of an extended N-terminal arm and three α -helices and binds to an optimal TAATTA sequence. The N-terminal arm and the third α -helix make base contacts. The recognition helix binds to the major groove and has base contacts through residues Ile47, Gln50 and Asn51. Arg5 from the N-terminal arm binds to the minor groove. In addition to the base contacts, Thr6, Tyr25, Arg31, Arg53 contact the DNA phosphate backbone.

MD simulations

MD simulations were performed using the MD simulation package GROMACS 3.3.1 (42) on individual fingers of EGR TF Zif268, TFIIIA, GLI and the engrailed homeodomain. FI and FII of GLI were simulated together. In all simulations, based on neutral pH conditions, basic Arg and Lys residues were positively charged, and acidic Asp and Glu residues were negatively charged. His residues coordinating the zinc ion were neutral with the hydrogen atom on the N $^{\delta}$ atom of the His side chain, since the electronegative N $^{\epsilon}$ atom in these histidine residues interact with the zinc ion. Each individual finger or homeodomain was centered in a rhombic dodecahedron box with a 15 Å minimum distance from the protein surface to the box edges. The system was solvated with simple point charge water molecules giving ~ 4600 waters for ZFs and 6100 waters for the engrailed homeodomain. Then, the systems were minimized by using steepest descent method with GROMOS96 (43) force field. Desired numbers of ions ranging from none to 20 were added by replacing water molecules randomly with a minimum distance of 6 Å between the ions and the protein. The temperature was coupled to a bath of 300 K with a weak coupling time constant of 0.1 ps. The pressure was coupled to 1 Bar using Parrinello–Rahman method (44). Initial velocities were generated randomly from a Maxwell distribution at 300 K.

To better deal with charged ions, a twin range cut-off radius of 10 Å was used in the simulations for non-bonded interactions, as opposed to Ewald boundaries that require an overall neutral simulation box. Default GROMOS96 (43) parameters were used for the Zn $^{+2}$, Cl $^{-}$ ions and all other residues. The ion concentration of each system depends on the volume of the simulation box. For example, one Cl $^{-}$ ion corresponds to ~ 11.5 mM in the simulations of EGR FI with a 144.9 nm 3 simulation box.

We note that ionic solvation is hard to treat in most molecular mechanics force fields, as well as PB, where non-linearities have to be taken into account. The above notwithstanding, here we are mostly interested in characterizing the loci of ion sites in the protein surface, and the impact of these ions in side chain dynamics, as opposed to Debye screening or protein–protein interactions that might depend on more subtle details of the ions sampling in the water box.

Since the zinc ion in ZF proteins has a structural role (45), we harmonically constrained the zinc ion and the zinc coordinating residues to keep the tetrahedral coordination using a force constant of $2.4 \text{ kCal/mol/\AA}^2$, as well as N and C atoms of the protein. These constraints are consistent with recent NMR solution structures of various C_2H_2 ZFs which show that the ZF is highly stable (46,47), changing little from the unbound to the bound structure. Supplementary Figure S1 shows a superimposition of fingers from the different ZF proteins including a solution NMR structure. We performed multiple 5 and 9 ns long runs, starting with different counter ion concentrations with a time step of 2 fs using periodic boundary conditions. Coordinates were saved every picosecond. The last 4 or 8 ns of the trajectories are analyzed. Four or more independent runs were performed for each finger at each ion concentration with at least 20 ns of aggregate simulation time. The total simulation time for all fingers was over a microsecond.

Side chain root mean square deviation (RMSD)

The side chain dynamics are analyzed by extracting snapshots from each MD trajectory. The snapshots are overlapped with the bound crystal structure of each domain using the α -helix C^α atoms. The α -carbons of each residue are further translated to coincide with the α -carbon of the side chain in the crystal structure. The RMSDs are calculated with respect to the crystal structure using the side chain heavy atoms starting from C^β atoms. Following the analysis in refs. (33,48), Arg, His, Tyr, Lys and Trp side chains are considered bound-like if the RMSD is under 2 Å. Glu, Gln is considered bound-like if the side chain RMSD is $<1.5 \text{ \AA}$. Asp, Asn, and Leu side chains are considered bound-like with RMSD $<1.25 \text{ \AA}$ and Thr is considered bound-like when RMSD is $<1 \text{ \AA}$. We also cross-checked that side chain RMSD $<2 \text{ \AA}$ correlates with contact distances between H-bonds in both Supplementary Figures S2B and S3B. Supplementary Figure S4 also provides an analysis of top side chain clusters of solvated and buried side chains.

RESULTS

Buried side chains in protein–DNA interactions and their ion dependent dynamics

Residues important for recognition typically bury large amounts of SASA upon binding (33). Table 1 lists amino acids that contact DNA bases, including residues at helix positions -1 , $+3$ and $+6$ in EGR domains (see ‘Methods’ section for description of the

DNA-binding sites on the recognition helices). Also listed is the SASA they bury upon binding, and the percent of time these side chains are in a conformation close to their bound structure in two solvents, one without and one with ions (equivalent to a physiological ion concentration of 150 mM, i.e. 8–20 ions depending on the volume of the simulation box). As expected, the residue that buries the largest amount of SASA in each domain is always contained in this table of specificity determinant contacts. In particular, this data readily identifies Arg $_{+6}$, Arg $_{-1}$ and Arg $_{-1}$ in FI, FII and FIII of EGR, respectively, as the main specificity determinant residues in this complex (in grey), which are relatively free in the unbound state but bury the largest amount of SASA upon binding. MD simulations also show that, *in the absence of DNA*, these side chains *in the presence of Cl^- ions* spend a significant amount of time ($\sim 30\%$ or more in a nanosecond time scale) in configurations similar to the one they acquire in the bound structure with DNA. Table 1 shows that 12 out of the 17 side chains that are more than 80% buried in the complex increase their bound-like behavior under physiological conditions, three side chains remain about the same, and the other two are more than 69% bound-like regardless the number of ions in solution.

Figure 2 shows more in detail the RMSD of the main specificity determinant residue of EGR (49), Arg at position $+6$ in FI, with respect to its bound conformation as a function of time during the last 4 ns MD simulations and increasing number of Cl^- ions in the water box. Figure 2F shows the RMSD of Arg $_{+6}$ with no ions present in the simulation box. In this case, the side chain is found in a bound-like conformation (i.e. $<2 \text{ \AA}$ RMSD from the bound structure) 29% of the time (average of four independent simulations is $23 \pm 6\%$). Addition of counter ions increases this percentage to as much as 79% (Figure 2A). The histograms in the right insets clearly show how the distribution of Arg $_{+6}$ conformations shifts to low RMSDs with respect to the bound structure as a function of Cl^- ion concentration. In addition, we note that the bound-like conformations correlate with the heavy atom charge–charge distances for the protein–DNA contacts (see Supplementary Figures S2B and S3B).

The correlation between counter ions and bound-like behavior was also observed in other specificity determining buried side chains. For instance, in EGR, Arg $_{+6}$ in FIII (Figure 3) and His $_{+3}$ in FII (Table 1). A similar correlation was observed for GLI, TFIIIA and engrailed homeodomain (Table 1), as well as for side chains involved in non-specific binding. In all cases, the percentage of bound-like side chains observed in the simulations appears to saturate beyond a certain number of ions consistent with an ‘effective’ ionic concentration of $\sim 150 \text{ mM}$.

Two relatively buried side chains do not appear to follow the aforementioned trend. Arg62 (Pos. $+6$) in FII of TFIIIA is not as bound-like as one would expect, though the cavity next to it (see Supplementary Figure S2) suggests that it might not be as buried as estimated by NACCESS (50) using a water radius of 1.4 \AA . Moreover, large differences on this side chain between the crystal (40) (PDB:1TF6), a 3.1 \AA resolution and

Table 1. Analysis of buried surface area and bound-like behavior of specificity determinant side chains that contact DNA bases

	Critical residue	Δ SASA ^a (Å ²)	Buried Free ^b (%)	Buried complex ^c (%)	Bound-like (ions) ^d (%)	Bound-like (no ion) ^e (%)	
EGR	FI	R118 (-1)	97	27	75	23 ± 5	27 ± 16
		E121 (+3)	44	64	97	76 ± 1	94 ± 1
		R124 (+6)	129	20	84	63 ± 9	24 ± 6
	FII	R146 (-1)	113	38	94	41 ± 4	41 ± 5
		H149 (+3)	74	42	93	52 ± 7	40 ± 9
		T152 (+6)	41	69	75	88 ± 2	77 ± 3
	FIII	R174 (-1)	134	27	93	69 ± 8	82 ± 6
		E177 (+3)	42	61	93	87 ± 7	64 ± 9
		R180 (+6)	99	13	62	46 ± 14	28 ± 17
TFIIIA	FI	K26 (-1)	85	38	90	63 ± 0.2	55 ± 11
		W28 (+2)	107	25	76	43 ± 14	66 ± 13
	FII	H58 (+2)	62	29	71	61 ± 5	41 ± 4
		H59 (+3)	60	52	93	94 ± 1	82 ± 8
		R62 (+6) ^f	116	35	93	19 ± 0.1	6 ± 1
	FV	L148 (-1)	45	20	52	24 ± 2	0
		R151 (+3)	80	41	81	30 ± 3	0
		R154 (+6)	87	33	76	0	0
	GLI	FII	Y155 (+2)	104	25	83	28 ± 3
D216 (+3)			46	37	82	71 ± 4	64 ± 6
K219 (+6)			59	48	84	93 ± 6	29 ± 3
Homeodomain		R5 ^g	171	0	81	13 ± 3	11.5 ± 1
		I47	59	43	85	68 ± 4	24 ± 1
		Q50	71	41	94	45 ± 6	48 ± 3
		N51	62	22	80	49	32 ± 2

Highlighted rows indicate residues that are mostly free in the unbound state, but are >80% buried in the complex, burying the largest amount of solvent area upon complexation. Also included in the Table are Positions +3 and +6 for EGR.

^aChange in solvent accessible surface areas (SASA) upon complexation. SASA are calculated using the program NACCESS (50).

^bFraction of buried area in the free protein with respect to the tri-peptide Ala-X-Ala.

^cFraction of buried area in the protein-DNA complex with respect to the tri-peptide Ala-X-Ala.

^dFraction of bound-like conformations (RMSD <2 Å) for each side chain in the presence of counter ions at physiological concentrations (150–160 mM corresponding to 8–20 ions in a simulation box). See 'Materials and Methods' section.

^eFraction of bound-like conformations for each side chain in the absence of counter ions (RMSD <2 Å). See 'Materials and methods' section.

^fThe simulation results for the RMSD of R62 using the X-ray crystal is 12 ± 5% with ions and 11 ± 8% with no ions.

^gR5 is not considered as specificity determinant and therefore not highlighted. See text for details.

NMR (51) (PDB:1TF3) structures make the definition of bound-like rather difficult. Based on our own energetic analysis (4), we find that if Arg62 is buried then the NMR conformation will be more stable since Arg62.N^e forms an intra-molecular hydrogen bond with the backbone of His59 (Supplementary Figure S2). On the other hand, if Arg62.N^e is not buried then it can form a protein-solvent hydrogen bond, and both crystal and NMR configurations will have very similar energies. Note that burying a free NH group is energetically unfavorable. MD runs of the NMR structure show that Arg62 is bound-like 20% of the time in the presence of ions, forming the aforementioned backbone bond during this time. Another special side chain is Arg5 in engrailed homeodomain (PDB: 3HDD). This side chain is part of the fully flexible N-terminal domain of the protein. Contrary to a side chain in a structured domain, this side chain can always undergo induced fitting once the main recognition helix docks into the major groove, without interfering with the binding process. Side chains that do not contact DNA and are solvated in the complex do not sample bound-like conformations. Supplementary Figure S4 shows the relation between bound-like RMSD and probabilities of rotamers from

the rotamer library (52) that belong to the same clusters for three side chains of EGR FI. Arg114 precedes the hydrophobic core residue Phe and is facing towards the solvent in the EGR complex (Supplementary Figure S4B). The MD simulations show that this side chain is bound-like only 1% of simulation time. The MD clusters of Arg114 do not correlate with rotamer probabilities. For Arg118 and Arg124, there is also little correlation between MD clusters and rotamer probabilities [see ref. (33) for further evidence of the disagreement between MD clusters and rotamer probabilities].

Buried versus non-buried side chains imply different evolutionary constraints

Rearranging a misfolded side chain at the core of the binding interface is much more difficult than to do it on the periphery. Strikingly, we find that the dynamics of the two side chains that bury the largest amount of SASA in EGR, Arg₊₆ and Arg₋₁ in FI and FIII, respectively, is more than 60% bound-like, while the partially exposed side chains capping the N- and C-terminals (Arg₋₁ and Arg₊₆ in FI and FIII, respectively; see Figure 1) is

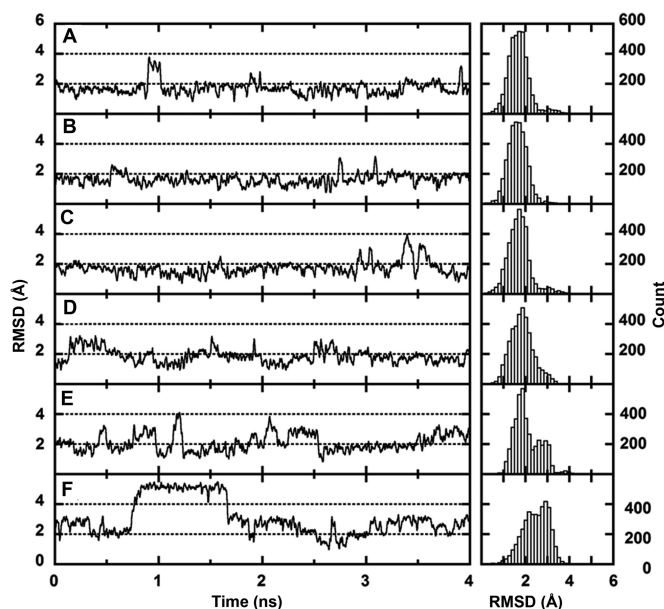


Figure 2. Increasing bound-like behavior of EGR specificity determinant residue Arg₊₆ in FI as a function of ion concentration. Change in RMSD of Arg₊₆ with respect to its bound conformation as a function of time and increasing number of counter ions in the simulation box: (A) 20 ions; (B) 16 ions; (C) 12 ions; (D) 8 ions; (E) 4 ions; (F) No ions. The histograms show the distribution of the RMSDs in each simulation, reflecting the increasing amount of bound-like behavior of Arg₊₆ as a function of ion concentration.

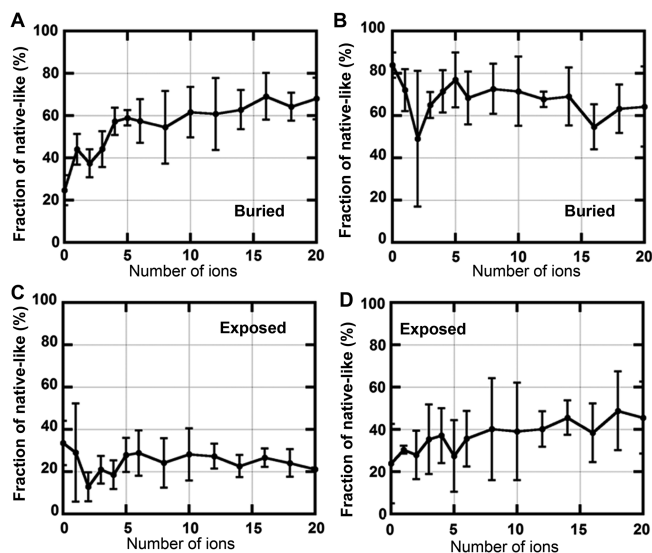


Figure 3. Differential ion dependence of bound-like behavior of key side chains in EGR fingers I and III. The effective ionic strength in millimolar corresponds to the number of ions \times 11.5. The conformations are considered bound-like if the RMSD from the crystal structure conformation is under 2 Å. (A) Fraction of bound-like conformations for buried Arg₊₆ in finger I. (B) Fraction of bound-like conformations for buried Arg₋₁ in finger III. (C) Fraction of bound-like conformations for exposed Arg₊₆ in finger I. (D) Fraction of bound-like conformations for exposed Arg₊₆ in finger III. Error bars are the direct standard deviation from three or more independent four nanosecond MD simulations. Note that in the co-crystal both Arg₋₁₅ and Arg₊₆ have identical structures. However, side chain dynamics is more bound-like if side chain is found buried in the complex relative to if it is found exposed to solvent.

between 20 and 50% bound-like. Hence, despite the fact that Arg₋₁ and Arg₊₆ in FI and FIII domains have an identical crystal structure and helical binding sequence (i.e. RDER/GCG), subtle longer range interactions streaming from differences in the sequences of the β -strands have led to very different dynamics: Buried side chains are strongly native-like, non-buried side chains are less so. These differences in side chain dynamics are, perhaps, a glimpse of the extent that evolution has tuned the complementarity of these interactions.

Ions sit at the loci of the phosphates in the protein–DNA complex

The impact of ions in the dynamics of side chains is no accident, but in fact it correlates with a weak binding site for Cl⁻ ions located at the same position where the negatively charged phosphates bury the largest amount of SASA in the complex structures (see highlighted rows in Table 2). Further supporting this observation is the distribution of the distance between the position of these

Table 2. SASA buried by phosphates and the corresponding residence time of Cl⁻ ions on their binding site during MD

	DNA	Δ SASA ^a (Å ²)	Ion residence ^b (%)
EGR	FI G7	72	70 \pm 12
	FI G8	69	2 \pm 1
	FII G4	60	35 \pm 4
	FII T5	39	4 \pm 2
	FIII Model	–	68 \pm 7
	FIII G2	36	0
TFIIIA	FI G26	53	36 \pm 8
	FI G34	51	4 \pm 1
	FI A27	37	15 \pm 3
	FII ^c T23	67	18 \pm 2
	FII ^c A22	28	0
	FV T8	66	20 \pm 4
	FV C7	49	0
	FV A53	27	2 \pm 1
	FIII G20	54	2 \pm 1
	FIII C40	40	1
FIII A22	13	0	
GLI	FII A65	77	30 \pm 6
	FII G66	50	1
	FIV C59	68	24 \pm 5
	FIV C7	40	4 \pm 1
	FIV C58	40	0
	FIV T8	22	4 \pm 1
	FV G56	55	7 \pm 1
	FV A57	45	1
	FIII C62	47	6 \pm 1
	FIII A64	34	1
FIII T6	20	3 \pm 1	
Homeodomain	G326	53	0
	G327	66	79 \pm 5
	T328	55	0

High residence times are observed on the loci of phosphates that bury the largest amount of SASA (highlighted rows).

^aChange in SASA upon complexation. Surface areas of phosphate groups are calculated in the absence and presence of the protein.

^bThe fraction of ion residence time corresponds to the fraction of the simulation time a counter ion is observed within a 3 Å radius sphere from the position of the phosphorus atom at physiological concentrations (150–160 mM corresponding to 8–20 ions in a simulation box).

^cFII from the solution NMR complex structure (51).

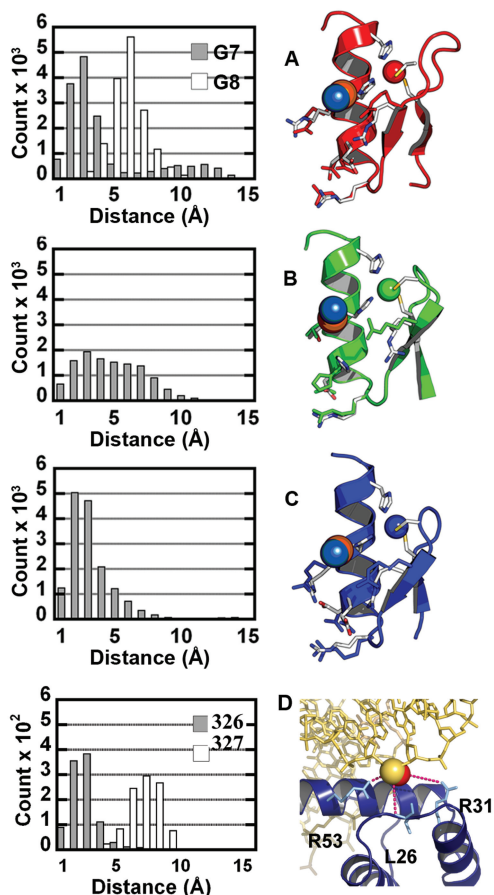


Figure 4. Ions occupy the phosphate binding site in the protein–DNA complex. Distribution of counter ion positions in fingers I–III of EGR and in the engrailed homeodomain relative to the phosphorus atom position in their corresponding co-crystals. The cartoon representations of each domain are on the right: DNA phosphorus atom shown in orange spheres; Cys, His residues and binding motif residues are shown as sticks; CPK colored sticks show the crystal conformation; and, colored sticks show a snapshot of the position of key side chains. A blue sphere shows for each domain the position of the Cl^- in the simulation that had the largest residence time using a 3 Å clustering radius. (A) Finger I (red), for comparison we also show the distribution with respect to a second phosphate binding site; (B) Finger II (green); (C) Finger III (blue); (D) Engrailed homeodomain, phosphorus atom is shown in yellow sphere. DNA is shown as yellow sticks. The two phosphate backbone contacting residues are Leu26, Arg31 and Arg53 are shown in light blue sticks. The contacts are depicted as pink dashed lines. For comparison we also show the distribution with respect to a second phosphate binding site. We find only one ion hot spot per domain, corresponding to the locus of the phosphate that buries the largest amount of SASA upon binding.

phosphates in the crystals and their nearest Cl^- ions during the MD simulations (Figure 4). For comparison, Figure 4A and D also show no significant residence time of Cl^- ions near the loci of the second most buried phosphate in the corresponding crystals.

It is interesting to note that in seven out of a total of eight ZFs the ion binding site corresponds to the pocket supported by the conserved N^δ of His_{+7} , which is also the most highly conserved ZF–DNA backbone contact (35). The sole exception is the pocket of the G26 phosphate of FI of TFIIIA that makes a bond with the OH group of

Tyr24. Note that most ZFs have the highly conserved Phe24 at this position (35). Fully consistent with the aforementioned correlation, our simulations also show that Cl^- ions shift their interaction site to the Tyr24 site.

Overall, MD simulations of 15 different DNA-binding domains (including one homeodomain) detected a robust Cl^- ion interaction site in nine of them, and did not observe one in six. Crystal structures show that all nine domains bury their recognition helix deep into the major groove [as most DNA-binding proteins do (53)], whereas none of the six domains bury their helix as deep (see crystals of FI, FIII and FV of GLI and FIII, FIV and FVI of TFIIIA). These findings suggest that an ‘electrostatic hot spot’ capable of trapping negatively charged ions is important for burying the recognition helices deep into the major groove. On the other hand, domains that do not have this electrostatic hot spot do not trap either Cl^- ions or contact phosphate groups. The percent of simulation time that Cl^- ions spend in the phosphate binding site (i.e. within 3 Å of the phosphorus atom) is 18–79% (Table 2), compared to 0–7% for phosphates that do not bind their helix into the major groove. Ions residence times anywhere on the surface of ZFs that do not contact DNA, i.e. FI of GLI, and FIV and FVI of TFIIIA, were in the range of 0–5% of the full MD simulations. The 3 Å clustering radius around the phosphorus position is quite stringent considering the four oxygens bound to the phosphorus atom. Note that in FIII of EGR the corresponding phosphate was modeled since it is missing from the crystal structure.

For ZFs, the Cl^- ions interact in a pocket formed by the conserved His_{+7} (or Tyr24 in FI of TFIIIA) and two basic groups flanking this residue, one on the α -helix (often Pos. +6) and the other on the second β strand one residue before the conserved Phe in the hydrophobic core, usually an Arg or a Lys (see, e.g. Figure 4). For engrailed homeodomain, the ions also interact with two basic groups, Arg53 and Arg31, as well as Leu26 backbone (Figure 4D). The two basic groups together with a third hydrogen bond donor appear to be a common feature of the electrostatic hot spot. Interestingly, although these residues are involved in forming the electrostatic pockets, not all of them end up forming hydrogen bonds with the phosphates in the complex structure.

The key observation here is that the binding site of the Cl^- ions on the surface of the TF domains corresponds to the same locus where a phosphate group buries the largest amount of SASA upon complexation (see Table 2 and Figure 4). Hence, the ‘re-ordering’ of Arg_{+6} and other side chains into bound-like conformations (in the absence of DNA) reflect in part the phosphate-like electrostatic interactions mimicked by the Cl^- ions (see also Supplementary Figure S3). Since clustering of Cl^- positions on the full surface of the protein domains did not reveal other preferred interacting sites, we conclude that these sites are evolutionary designed to desolvate charged groups. We also checked the sensitivity of the hot spot for acetate (another charge -1 molecule), finding similar propensities for protein association than Cl^- . Three examples of the dynamics of the Cl^- ion

around the phosphate binding site are shown in Supplementary Figures S1–S3.

Non-specific contacts

About two-thirds of protein–DNA contacts are with the DNA backbone. Table 3 lists all backbone contacting side chains, with 30 out of 37 contacting side chains showing a significant amount of bound-like behavior. Interestingly, every single DNA-binding domain shows at least one or more of its side chains forming a hydrogen bond with a DNA backbone phosphate in a highly constrained bound-like configuration. Several side chains are partially buried in the free state, significantly constraining their conformations to bound-like states, e.g. His₊₇ in ZFs. Other side chains show a moderate increase of bound-like conformations in response to Cl[−] ions, e.g. Arg114 in FI of EGR improves its bound-like behavior from 30 to 60%,

and Arg53 in homeodomain goes from 10 to 83% bound-like in the presence of ions. Overall, however, the role of ions for phosphate backbone contacting residues is not as striking as for those side chains forming hydrogen bonds with bases. Finally, some side chains such as Lys29 in FI of TFIIIA and Arg146 in FII of GLI are more than 60% buried prior to binding, requiring only a small rotation to rearrange and make their corresponding Hydrogen bonds.

The Lys residues in the conserved linker sequence (TG-E/Q-KP) in many ZF proteins (35) also contact and stabilize the protein–DNA complex (51,54,55). These hinge regions play a critical role capping the helical domains and become rigid upon DNA binding (55). Simulations of consecutive ZFs, FI–FII in EGR (using the same simulation protocol) show that in the canonical binding structure the linker Lys is 84% of the time in a bound-like conformation (data not shown).

Table 3. Residues contacting the phosphate backbone in EGR, TFIIIA, GLI and engrailed homeodomain

	DNA contact	Residue	Δ SASA ^a (Å ²)	Buried free ^b (%)	Buried complex ^c (%)	Native-like(ions) ^d (%)	Native-like (no ion) ^e (%)		
EGR	F1	G8	R103	40	36	56	52 ± 8	42 ± 8	
		G7	R114	43	35	57	58 ± 9	30 ± 9	
		G7	H125 (+7)	19	84	97	100	100	
	FII	T5	R142	53	23	48	0	0	
		G4	H153 (+7)	24	77	93	100	100	
	FIII	G2	R170	53	32	59	96 ± 2	98 ± 0.5	
		–	H181 (+7)	0	84	84	100	100	
	TFIIIA	FI	G34	R12	65	34	66	0	0
			G34	Y13	44	48	73	100	100
G26			Y24	44	68	92	100	100	
A27			K29 (+3)	54	63	96	0	0	
FII ^f			T23	H63 (+7)	14	23	95	100	100
FIII		A22	T66 (+10)	43	43	86	29 ± 7	27 ± 7	
		A22	T85	88	8	94	46 ± 19	52 ± 5	
		C40	K87 (+1)	36	28	50	0	0	
FV		T8	K144	45	30	58	30 ± 6	0	
		T8	H155 (+7)	17	82	93	100	100	
		C7	V158 (+10)	51	20	65	21 ± 4	0	
GLI		FII	G66	R146	37	66	85	0	0
			G66	K152	56	20	54	0	0
			A65	H160 (+7)	16	89	100	100	100
		FIII	C62	Y181	57	64	96	100	100
	A64		R183	142	12	83	0	0	
	T6		K188 (+5)	29	51	68	97	100	
	FIV	C7	Y200	78	39	83	77 ± 4	80 ± 8	
		T8	R217 (+4)	24	68	80	71 ± 4	88 ± 9	
		C59	H220 (+7)	24	79	95	100	100	
	FV	C58	T224 (+11)	45	37	81	84 ± 0.2	85 ± 9	
		G56	Y242	51	69	98	100	98	
		A57	T243	108	0	96	94 ± 3	92 ± 9	
	Homeodomain	A212	W48	18	85	93	100	100	
		A213	T6	5	47	42	45	30	
		G326	R31	27	70	84	33	52	
G327		K46	31	49	68	59	72		
G327		R53	42	71	91	82	10		
T328		Y25	38	36	57	53	45		
T328		R53	42	71	91	82	10		

^aChange in solvent accessible surface areas (SASA) upon complexation. SASA are calculated using the program NACCESS(50).

^bFraction of buried area in the free protein with respect to the tri-peptide Ala-X-Ala.

^cFraction of buried area in the protein–DNA complex with respect to the tri-peptide Ala-X-Ala.

^dFraction of native-like conformations for each side chain in the presence of counter ions. See ‘Materials and methods’ section.

^eFraction of native-like conformations for each side chain in the absence of counter ions. See ‘Materials and methods’ section.

^fFII from the solution NMR complex structure (51).

For the cases studied here, most of the highly buried residues had a tendency to be bound-like prior to the encounter with DNA, suggesting that they fold in conformations conducive to a smooth binding. The time scales (MD) for this bound-like dynamics are on the order of hundreds of picoseconds to nanoseconds, a time consistent with the lifetime of an encounter complex in protein interactions (33). Other interface residues not directly contacting bases, such as Glu₊₃ in both FI and FIII also form their bound-like hydrogen bond with the backbone of Arg₋₁, prior to the encounter with DNA. Interestingly, in a previous article (4), we have noted that an unconstrained Glu₊₃ side chain will clash with DNA bases of many tri-nucleotides, becoming a serious obstacle for non-specific association.

DISCUSSION

Most side chains that form hydrogen bonds with the DNA phosphate backbone (30 out of 37; Table 3) are between 20 and 100% bound-like, and 21 contacts are between 50 and 100% bound-like regardless of the number of ions present in solution. The latter group of highly constrained side chains encompasses all DNA-binding domains such that in the presence of a stretch of DNA, TFs can quickly form a non-specific complex with the DNA backbone. These complexes are unlikely to require a very precise complementarity since phosphates are relatively easy targets compared to hydrogen bonds to DNA bases. This efficient non-specific binding mechanism is consistent with association rate constants on the order of the diffusion limit, $10^9 \text{ M}^{-1} \text{ s}^{-1}$ for EGR (24).

Side chains that bury the largest amount of SASA in the bound state form bonds to DNA bases. Even before encountering the DNA, these groups are often found in rotamer conformations similar to those acquired in their complex structure. However, several of these side chains are also steered into these 'native-like' configurations by the presence of negatively charged ions in well defined 'electrostatic hot spots'. Interestingly, in a non-specific complex, the DNA backbone phosphates are expected to sit in the same hot spots. Hence, we conclude that already in a non-specific complex, specificity determinant side chains are predisposed to be close to their bound-like rotamer conformations such that if presented with the appropriate partner they will rapidly form a higher affinity complex.

Molecular simulations of three ZF proteins and the engrailed homeodomain show that domains that bind into the DNA major groove also have one electrostatic hot spot for a negatively charged Cl⁻ ion characterized by at least two basic groups in opposite sites of the phosphate and a relatively constrained third hydrogen bond. Crystal structures show that domains missing this substrate do not bury the recognition helix deep into the major groove, and have non-conventional binding modes (see TFIIIA and GLI). Since this site is also the locus of the phosphate that buries the largest amount of SASA upon binding in each domain, we suggest that without these hot spots proteins are not able to overcome the

large penalty entailed by the desolvation of the phosphate from solvent and their own positive counter ions (17,20,22).

We emphasize that a favorable electrostatic site on the DNA-binding proteins is not unexpected. However, the picture emerging from our analysis is that subtle interactions work cooperatively in order to trap DNA backbone phosphates, reflecting evolutionary constraints (see, e.g. Figure 3) that presumably should be taken into consideration in the design of novel TFs with both specific and non-specific binding capabilities.

One might argue that the widespread bound-like behavior of buried residues and other side chains in Tables 1 and 3 is due to the harmonic constraints of the N and C atoms to the bound structure. However, these constraints are consistent with structural evidence showing that the ZF fold does not depend much in sequence [see Supplementary Figure S1 and Figure 1 in ref. (47)] and does not change before and after binding DNA (47). Moreover, some side chains that appear exposed in the co-crystals, mostly at the 3' and 5' end of the complexes, and are not highly buried upon complexation do not show this bound-like behavior (e.g. R118 from FI of EGR and R154 from FV of TFIIIA). Finally, it is important to note that the traditional induced fit theory (56) would suggest that the conformations of most of these side chains should respond to intermolecular interactions. Our findings clearly show that this is not necessarily the case; in fact suitable conformations are imprinted on the TFs fold even in the absence of DNA.

The predisposition of key side chains of TFs for suitable rotamer conformations suggests that functional TFs have evolved structural motifs designed for a fast and efficient formation of a non-specific complex around the DNA backbone. Moreover, the fact that these phosphate backbone contacts are in all fingers is consistent with the experimental observation that two or more ZF domains work together to recognize specific DNA sequences (57,58). Multi-non-specific binding domains may also allow for partial dissociations and rapid reattachments of individual ZF as they diffuse from phosphate-to-phosphate along the DNA. This simple mechanism could reconcile the 'sliding' of TFs along DNA (2,19,59) first suggested by Winter *et al.* (19) by means of non-specific extended desolvated states, while bound-like specificity determinant side chains are ready to stall the 1D diffusion process at their cognate sequence.

Many of the key side chains forming bonds with bases are found buried deep in the binding interface, a natural question to ask is whether the binding process would also benefit from these side chains being bound-like prior to forming the non-specific complex. This could happen if Cl⁻ ions are available to sit in the electrostatic hot spot that phosphates occupy in the crystals, prior to binding. Experiments seem to suggest as much given that the binding affinity can be a factor of four higher in the presence of salt, ~150 mM ionic concentration (22,33). One might argue that any potential benefit for this mechanism will be offset by the extra barrier of desolvating the negative ions upon complexation. However, if the TF with

its negative counter ion and the DNA with its positive (Na^+) ion (17,20,22) get closer and closer, say, within 6 Å. Then, the effective local concentration of Na^+ and Cl^- increases to ~ 6.15 M, i.e. the solubility limit of NaCl. Hence, under these conditions, ions might prefer to leave the interface and move to bulk water, providing an interesting possibility to reconcile the apparent ease that ions are removed from the protein–DNA interface. It is also worth mentioning that an alternative model of protein–DNA association (60) has recently suggested that the role of counter ions in DNA is to bias ‘hopping’ versus ‘sliding’. In as much as removing counter ions from the DNA backbone is a natural barrier for binding, a mostly ‘hopping’ mechanism should still rationalize how these ions are being efficiently removed after each hop. In summary, more experiments are certainly needed to fully resolve these questions, and the origin of the subtle increase in binding affinity as a function of ion concentration.

In conclusion, the dynamics of protein side chains contacting DNA strongly suggest a two-stage mechanism where association is first benefitted by a vast network of side chains that are preset to lock onto the DNA backbone; and, subsequently, the dynamics of ion dependent specificity determinant residues indicates that ion/charge–protein interactions play a role in side chain dynamics conducive to an efficient DNA binding. Of the DNA-binding domains studied here, only those with an electrostatic hot spot correlate with tight binding of the recognition helix. These domains cover the main B-DNA conformational forms BI and BII (61) and account for GG, CG, TG, CT, AA and CC dinucleotide steps (61), implying that our findings are not limited to particular DNA sequences. Collectively, these observations suggest a mechanism by which an electrostatic hot spot mediates the non-specific desolvation of the highly charged/polar DNA backbone, providing a cogent view of how TFs might recognize a small number of DNA sequences with high efficiency and specificity.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Science Foundation (MCB-0444291) and Pittsburgh Supercomputer Center (MCB049958P). Funding for open access charge: National Science Foundation (MCB-0444291).

Conflict of interest statement. None declared.

REFERENCES

- Berg, O.G., Winter, R.B. and von Hippel, P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **20**, 6929–6948.
- Halford, S.E. and Marko, J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.*, **32**, 3040–3052.
- Pabo, C.O. and Sauer, R.T. (1984) Protein-DNA Recognition. *Ann. Rev. Biochem.*, **53**, 293–321.
- Temiz, N.A. and Camacho, C.J. (2009) Experimentally based contact energies decode interactions responsible for protein-DNA affinity and the role of molecular waters at the binding interface. *Nucleic Acids Res.*, **37**, 4076–4088.
- Anderson, C.F. and Record, M.T. Jr (1995) Salt-nucleic acid interactions. *Annu. Rev. Phys. Chem.*, **46**, 657–700.
- Jen-Jacobson, L. (1997) Protein-DNA recognition complexes: conservation of structure and binding energy in the transition state. *Biopolymers*, **44**, 153–180.
- Norberg, J. (2003) Association of protein-DNA recognition complexes: electrostatic and nonelectrostatic effects. *Arch. Biochem. Biophys.*, **410**, 48–68.
- Jayaram, B., McConnell, K., Dixit, S.B., Das, A. and Beveridge, D.L. (2002) Free-energy component analysis of 40 protein-DNA complexes: a consensus view on the thermodynamics of binding at the molecular level. *J. Comput. Chem.*, **23**, 1–14.
- Jayaram, B. and Jain, T. (2004) The role of water in protein-DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 343–361.
- Deremble, C. and Lavery, R. (2005) Macromolecular recognition. *Curr. Opin. Struct. Biol.*, **15**, 171–175.
- Spyrakakis, F., Cozzini, P., Bertoli, C., Marabotti, A., Kellogg, G.E. and Mozzarelli, A. (2007) Energetics of the protein-DNA-water interaction. *BMC Struct. Biol.*, **7**, 4.
- Patikoglou, G. and Burley, S.K. (1997) Eukaryotic transcription factor-DNA complexes. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 289–325.
- Beveridge, D.L., Dixit, S.B., Byun, K.S., Barreiro, G., Thayer, K.M. and Ponomarev, S. (2004) Molecular dynamics of DNA and protein-DNA complexes: progress on sequence effects, conformational stability, axis curvature, and structural bioinformatics. In Stellwagen, N.C. and Mohanty, U. (eds), *ACS Symposium #884: Curvature and Deformation of Nucleic Acids: Recent Advances, New Paradigms*. American Chemical Society, Washington DC, pp. 13–64.
- Cheatham, T.E. III (2004) Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr. Opin. Struct. Biol.*, **14**, 360–367.
- Tsui, V., Radhakrishnan, I., Wright, P.E. and Case, D.A. (2000) NMR and molecular dynamics studies of the hydration of a zinc finger-DNA complex. *J. Mol. Biol.*, **302**, 1101–1117.
- Zakrzewska, K., Lavery, R. and Leszczynski, J. (1999) Modelling Protein-DNA Interactions. In Politzer, P. and Maksic, Z.B. (eds), *Computational Molecular Biology*, Vol. 8. Elsevier, pp. 441–483.
- Record, M.T. Jr, Lohman, M.L. and De, H.P. (1976) Ion effects on ligand-nucleic acid interactions. *J. Mol. Biol.*, **107**, 145–158.
- Winter, R.B. and von Hippel, P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The *Escherichia coli* repressor-operator interaction: equilibrium measurements. *Biochemistry*, **20**, 6948–6960.
- Winter, R.B., Berg, O.G. and von Hippel, P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor-operator interaction: kinetic measurements and conclusions. *Biochemistry*, **20**, 6961–6977.
- Jen-Jacobson, L., Kurpiewski, M., Lesser, D., Grable, J., Boyer, H.W., Rosenberg, J.M. and Greene, P.J. (1983) Coordinate ion pair formation between EcoRI endonuclease and DNA. *J. Biol. Chem.*, **258**, 14638–14646.
- Ebright, R.H., Ebright, Y.W. and Gunasekera, A. (1989) Consensus DNA site for the *Escherichia coli* catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the *E. coli* lac DNA site. *Nucleic Acids Res.*, **17**, 10295–10305.
- Fried, M.G. and Stickle, D.F. (1993) Ion-exchange reactions of proteins during DNA binding. *Eur. J. Biochem.*, **218**, 469–475.
- Engler, L.E., Welch, K.K. and Jen-Jacobson, L. (1997) Specific binding by EcoRV endonuclease to its DNA recognition site GATATC. *J. Mol. Biol.*, **269**, 82–101.

24. Hamilton, T.B., Borel, F. and Romaniuk, P.J. (1998) Comparison of the DNA binding characteristics of the related zinc finger proteins WT1 and EGR1. *Biochemistry*, **37**, 2051–2058.
25. Manning, G.S. (1977) Limiting laws and counterion condensation in polyelectrolyte solutions: IV. The approach to the limit and the extraordinary stability of the charge fraction. *Biophys. Chem.*, **7**, 95–102.
26. Record, M.T. Jr, Anderson, C.F. and Lohman, T.M. (1978) Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Quart. Rev. Biophys.*, **11**, 103–178.
27. Sharp, K.A. and Honig, B. (1995) Salt effects on nucleic-acids. *Curr. Opin. Struct. Biol.*, **5**, 323–328.
28. Debye, P. and Huckel, E. (1923) Zur Theorie der Elektrolyte. *Physik. Z.*, **24**, 185–206.
29. Schreiber, G. and Fersht, A.R. (1996) Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol.*, **3**, 427–431.
30. Romaniuk, P.J. (1990) Characterization of the equilibrium binding of *Xenopus* transcription factor IIIA to the 5 S RNA gene. *J. Biol. Chem.*, **265**, 17593–17600.
31. Moraitis, M.I., Xu, H. and Matthews, K.S. (2001) Ion concentration and temperature dependence of DNA binding: comparison of PurR and LacI repressor proteins. *Biochemistry*, **40**, 8109–8117.
32. Kimura, S.R., Brower, R.C., Vajda, S. and Camacho, C.J. (2001) Dynamical view of the positions of key side chains in protein-protein recognition. *Biophys. J.*, **80**, 635–642.
33. Rajamani, D., Thiel, S., Vajda, S. and Camacho, C.J. (2004) Anchor residues in protein-protein interactions. *Proc. Natl Acad. Sci.*, **101**, 11287–11292.
34. Pabo, C.O., Peisach, E. and Grant, R.A. (2001) Design and selection of novel Cys(2)His(2) zinc finger proteins. *Ann. Rev. Biochem.*, **70**, 313–340.
35. Wolfe, S.A., Nekudova, L. and Pabo, C.O. (2000) DNA recognition by Cys(2)His(2) zinc finger proteins. *Ann. Rev. Biophys. Biomol. Struct.*, **29**, 183–212.
36. Fraenkel, E., Rould, M.A., Chambers, K.A. and Pabo, C.O. (1998) Engrailed homeodomain-DNA complex at 2.2 Å resolution: a detailed view of the interface and comparison with other engrailed structures. *J. Mol. Biol.*, **284**, 351–361.
37. von Hippel, P.H. and Berg, O.G. (1986) On the specificity of DNA-protein interactions. *Proc. Natl Acad. Sci. USA*, **83**, 1608–1612.
38. Kumar, N.V. and Govil, G. (1984) Theoretical studies on protein-nucleic acid interactions. I. Interaction of positively charged amino acids with nucleic acid fragments. *Biopolymers*, **23**, 1979–1993.
39. Elrod-Erickson, M., Rould, M.A., Nekudova, L. and Pabo, C.O. (1996) Zif268 protein-DNA complex refined at 1.6 Å resolution: a model system for understanding zinc finger-DNA interactions. *Structure*, **4**, 1171–1180.
40. Nolte, R.T., Conlin, R.M., Harrison, S.C. and Brown, R.S. (1998) Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc. Natl Acad. Sci. USA*, **95**, 2938–2943.
41. Pavletich, N.P. and Pabo, C.O. (1993) Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science*, **261**, 1701–1707.
42. Van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J.C. (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem.*, **26**, 1701–1718.
43. Scott, W.R.P., Hunenberger, P.H., Tironi, I.G., Mark, A.E., Billeter, S.R., Fennel, J., Torda, A.E., Huber, T., Krüger, P. and van Gunsteren, W.F. (1999) The GROMOS biomolecular simulation program package. *J. Phys. Chem. A*, **103**, 3596–3607.
44. Parrinello, M. and Rahman, A. (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.*, **52**, 7182–7190.
45. Frankel, A.D., Berg, J.M. and Pabo, C.O. (1987) Metal-dependent folding of a single zinc finger from transcription factor IIIA. *Proc. Natl Acad. Sci.*, **84**, 4841–4845.
46. Omichinski, J.G., Clore, G.M., Appella, E., Sakaguchi, K. and Gronenborn, A.M. (1990) High-resolution three-dimensional structure of a single zinc finger from a human enhancer binding protein in solution. *Biochemistry*, **29**, 9324–9334.
47. Lu, D. and Klug, A. (2007) Invariance of the zinc finger module: A comparison of the free structure with those in nucleic-acid complexes. *Prot. Struct. Funct. Bioinform.*, **67**, 508–512.
48. Camacho, C.J. (2005) Modeling side-chains using molecular dynamics improve recognition of binding region in CAPRI targets. *Proteins*, **60**, 245–251.
49. Elrod-Erickson, M. and Pabo, C.O. (1999) Binding studies with mutants of Zif268 - Contribution of individual side chains to binding affinity and specificity in the Zif268 zinc finger-DNA complex. *J. Biol. Chem.*, **274**, 19281–19285.
50. Hubbard, S.J., Campbell, S.F. and Thornton, J.M. (1991) Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J. Mol. Biol.*, **220**, 507–530.
51. Wuttke, D.S., Foster, M.P., Case, D.A., Gottesfeld, J.M. and Wright, P.E. (1997) Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *J. Mol. Biol.*, **273**, 183–206.
52. Dunbrack, R.L. Jr and Cohen, F.E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.*, **6**, 1661–1681.
53. Garvie, C.W. and Wolberger, C. (2001) Recognition of specific DNA sequences. *Mol. Cell*, **8**, 937–946.
54. Choo, Y. and Klug, A. (1993) A role in DNA binding for the linker sequences of the first three zinc fingers of TFIIIA. *Nucleic Acids Res.*, **21**, 3341–3346.
55. Laity, J.H., Dyson, H.J. and Wright, P.E. (2000) DNA-induced [α]-helix capping in conserved linker sequences is a determinant of binding affinity in Cys2-His2 zinc fingers. *J. Mol. Biol.*, **295**, 719–727.
56. Koshland, D.E. (1958) Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl Acad. Sci. USA*, **44**, 98–104.
57. Fairall, L., Harrison, S.D., Travers, A.A. and Rhodes, D. (1992) Sequence-specific DNA binding by a two zinc-finger peptide from the *Drosophila melanogaster* Tramtrack protein. *J. Mol. Biol.*, **226**, 349–366.
58. Fairall, L., Schwabe, J.W., Chapman, L., Finch, J.T. and Rhodes, D. (1993) The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature*, **366**, 483–487.
59. von Hippel, P.H. (2004) Biochemistry. Completing the view of transcriptional regulation. *Science*, **305**, 350–352.
60. Givaty, O. and Levy, Y. (2009) Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.*, **385**, 1087–1097.
61. Svozil, D., Kalina, J., Omelka, M. and Schneider, B. (2008) DNA conformations and their sequence preferences. *Nucleic Acids Res.*, **36**, 3690–3706.