

# Anomalous DNA binding by E2 regulatory protein driven by spacer sequence TATA

Zhiqun Xi<sup>1</sup>, Yongli Zhang<sup>2</sup>, Rashmi S. Hegde<sup>3</sup>, Zippora Shakked<sup>4</sup> and Donald M. Crothers<sup>1,2,\*</sup>

<sup>1</sup>Department of Chemistry, <sup>2</sup>Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, <sup>3</sup>Division of Developmental Biology, Children's Hospital Research Foundation, University of Cincinnati School of Medicine, 3333 Burnet Avenue, Cincinnati, OH 45229, USA and <sup>4</sup>Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

Received November 20, 2009; Revised February 2, 2010; Accepted February 8, 2010

## ABSTRACT

We have investigated the anomalously weak binding of human papillomavirus (HPV) regulatory protein E2 to a DNA target containing the spacer sequence TATA. Experiments in magnesium ( $Mg^{2+}$ ) and calcium ( $Ca^{2+}$ ) ion buffers revealed a marked reduction in cutting by DNase I at the CpG sequence in the protein-binding site 3' to the TATA spacer sequence. Studies of the cation dependence of DNA-E2 affinities showed that upon E2 binding the TATA sequence releases approximately twice as many  $Mg^{2+}$  ions as the average of the other spacer sequences. Binding experiments for TATA spacer relative to ATAT showed that in potassium ion ( $K^+$ ) the E2 affinity of the two sequences is nearly equal, but the relative dissociation constant ( $K_d$ ) for TATA increases in the order  $K^+ < Na^+ < Ca^{2+} < Mg^{2+}$ . Except for  $Mg^{2+}$ ,  $K_d$  for TATA relative to ATAT is independent of ion concentration, whereas for  $Mg^{2+}$  the affinity for TATA drops sharply as ion concentration increases. Thus, ions of increasing positive charge density increasingly distort the E2 binding site, weakening the affinity for protein. In the case of  $Mg^{2+}$ , additional ions are bound to TATA that require displacement for protein binding. We suggest that the TATA sequence may bias the DNA structure towards a conformation that binds the protein relatively weakly.

## INTRODUCTION

Recognition of DNA by proteins can be divided into direct and indirect readout effects. The former occurs when the nucleic acid bases interact directly with

functional groups on the protein, including hydrogen bonding to purine N3 and N7 atoms, the exocyclic amino groups and carbonyl oxygen, and hydrophobic bonding with the thymine methyl group. Indirect readout falls into the general categories of protein-DNA recognition mediated by DNA structure and DNA deformability which are sequence dependent. This includes the interaction of a protein functional group with a nucleic acid functionality, such as the sugar-phosphate backbone, whose nature is independent of base sequence, but whose precise geometric position and hence interaction energy still depend on sequence as proposed for the first time on the basis of the sequence-specific trp repressor/operator complex (1). A more recent and striking example is provided by *Escherichia coli* integration host factor or IHF protein, which can distinguish differences in twist at individual dinucleotide steps by interaction with the corresponding phosphates (2). There are many other protein-DNA complexes (3) that exemplify indirect readout resulting from intrinsic sequence-dependent variations in DNA structure (4).

DNA distortion forms the basis for a second kind of indirect readout. Since many proteins deform DNA upon binding (5), it is to be expected that protein-DNA association might be facilitated by enhanced DNA deformability, as well as by a match between intrinsic DNA shape in solution and the conformation in the complex. The effect of DNA distortion can be separated from the influence of direct readout interactions when proteins interact with DNA at two or more sites, leaving a stretch of DNA between the contact points that is distorted but not touched by the protein. Two examples are the phage 434 repressor (6) and the human papilloma virus (HPV) regulatory protein E2 (7–9). In both cases, four non-contacted base pairs separate the protein contact sites.

In order to characterize the effects of DNA intrinsic properties on protein-complex stability, Zhang *et al.* (10)

\*To whom correspondence should be addressed. Tel: +1-203-4325204; Fax: +1-203-4326144; Email: Donald.Crothers@yale.edu  
Present address:

Yongli Zhang, Department of Cell Biology, Yale University, New Haven, CT 06520, USA.

used the cyclization kinetic method to measure variations in curvature and flexibility of 4 bp spacer sequences placed between the E2 protein-contact sites. A statistical mechanical theory was then used to calculate the relative ability of the spacers to begin and end in the proper configuration for binding. In spite of the variation of E2-binding affinities by three orders of magnitude, the curvature and flexibility of the unbound DNA sequences could be used to predict the relative binding constants within a factor of 3 for 15 of 16 sequences. The sole exception was for the spacer sequence TATA, whose measured affinity was 44-fold lower than predicted on the basis of the measured properties of the unbound TATA sequence.

The experiments we report here were designed to clarify this discrepancy. Use of DNase I as an enzymatic probe revealed a sequence-dependent structural anomaly of DNA in the protein-recognition region adjacent to the TATA spacer sequence, which was not found with control spacers. This anomaly appeared in the presence of either magnesium ( $Mg^{2+}$ ) or calcium ( $Ca^{2+}$ ) ions. We also investigated the cation dependence of DNA–E2-binding affinities. The presence of the TATA spacer sequence results in release of approximately twice as many  $Mg^{2+}$  ions as the average of the other spacer sequences upon E2 binding. On the other hand,  $Ca^{2+}$  release is smaller and close to the average value. At high (>10 mM) divalent cation concentrations, E2 binding in the presence of  $Mg^{2+}$  is substantially weaker than in the presence of  $Ca^{2+}$ . In the presence of potassium ion ( $K^+$ ) and absence of divalent ions, relative E2 binding to the TATA spacer is about half as strong as to the ATAT spacer, in line with expectation based on the properties of the two DNAs (10). Relative electrophoretic mobility measurements on sequence ladders revealed that  $Mg^{2+}$  increased the curvature of the TATA spacer construct more than for other spacer sequences.

## MATERIALS AND METHODS

### Construction of DNA for DNase I digestion

Oligonucleotides were synthesized by the Keck Foundation (Yale University). For the DNase I digestion experiments, three 68-bp duplex constructs were prepared. Each construct top-strand sequence was 5'-GCAGATATCGATCGCATCACGTTGTAGCCTAGCTTGCA**accgnnnccggtTGCGACTTGCGCTCTAGC**-3', with the E2 site in bold lower case, and the 4bp nnnn spacer sequence is ATAT, TATA or ACGT. 100 pmol of top strand was labelled at the 5'-end by T4-DNA polynucleotide kinase and [ $\gamma$ - $^{33}P$ ]ATP. To obtain double-stranded DNA, the 5'-labelled top strand was annealed with the unlabelled complementary strand (68-mer) at a molar ratio of 1:2. The hybridized labelled strand was purified by electrophoresis on a nondenaturing 10% polyacrylamide gel, the desired DNA band was excised, submerged into tris–ethylenediaminetetraacetic acid (TE) buffer, and incubated at 4°C overnight. DNA was concentrated by ethanol precipitation.

### DNase I digestion and densitometry

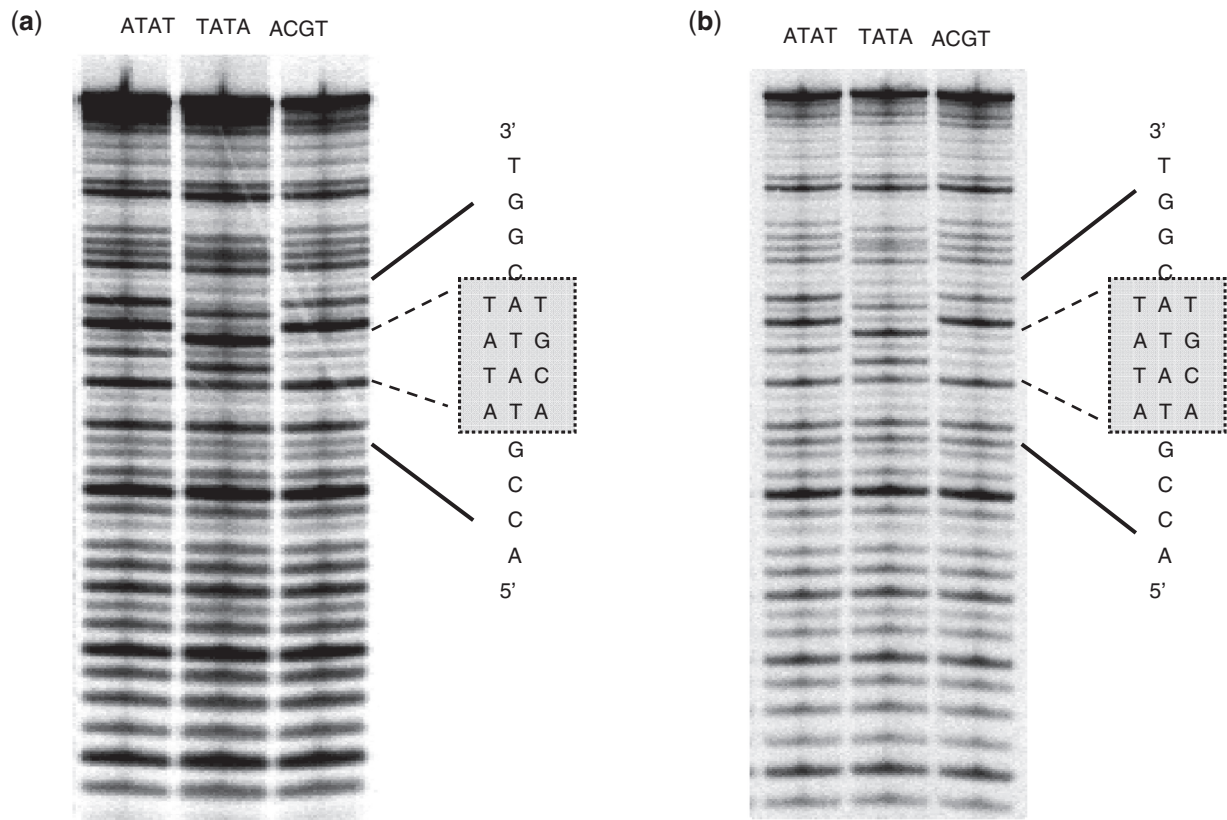
Digestion was carried out in 50  $\mu$ l buffer containing 50 mM Tris–hydrochloric acid (HCl) pH 7.6, 10 mM magnesium chloride ( $MgCl_2$ ) or 10 mM calcium chloride ( $CaCl_2$ ), 10 mM dithiothreitol (DTT), 1 mM ATP and 25  $\mu$ g/ml bovine serum albumin (BSA). This buffer condition is similar to that used in E2 protein-binding buffer; 1 U DNase I (Roche) was added. The reaction was incubated at room temperature for 1 min, and filtered by micropure-EZ (Millipore) to remove DNase I. DNA was precipitated with ethanol and loaded onto a 15% denaturing polyacrylamide gel. Autoradiograms of the gel were scanned on a FLA 5100 (Fuji Film) scanning densitometer.

### DNA gel-migration anomaly assay

Five oligonucleotides (top strands) of 10 nucleotides, i.e. tcggtcgata (for the ATAT-containing sequence), acggtcgat (TATA), tcggtcgaat (AATT) and acggtcgta (TTAA), tcggtcgacg (ACGT) were chemically synthesized, desalted, labelled at 5'-end with  $^{33}P$  using polynucleotide kinase and purified. They were then hybridized with their respective complementary strands to form 8 bp-duplex DNA with two-nucleotide cohesive ends. Then the DNA monomers were ligated in the presence of 30 U/ $\mu$ l T4 DNA ligase (New England Biolabs, MA) at room temperature for 3.5 h. Finally, the ligated products were run on 10% nondenaturing polyacrylamide gel at Tris–borate buffer in the presence or absence of 10 mM  $MgCl_2$ . A reference DNA ladder with no appreciable DNA curvature was similarly made and electrophoresed along with the above ladders of E2-binding sequences. Positions of DNA bands were quantified with phosphorimaging.

### Gel-mobility shift assay

The E2 protein encoded by the HPV type 16 genome (HPV-16) was purified as previously described (11) and used throughout this work. The E2 affinity was measured in the same buffer as previously used (10), i.e. 50 mM Tris–HCl, pH 7.6/10 mM  $MgCl_2$ /10 mM DTT/1 mM ATP/25 mg/ml BSA/0.05% Nonidet P-40/10% glycerol. The E2-binding site was incorporated into the 28-bp duplex region of a DNA hairpin, whose complete nucleotide sequence is GCTTGCA**accgnnnccggtTGCGACTTGCCCCCAAGTCGCAaccgnnnccggtTGCAAGC**, with the E2 site in bold lower case. Chemically synthesized oligos were desalted, 5'-end labelled with [ $\gamma$ - $^{33}P$ ]ATP, purified and scintillation counted. DNA corresponding to 2000 CPM, with final concentration estimated to be 2–10 pM, was mixed with a series of fresh E2 dilutions and incubated at 21°C in a waterbath for 5 h. The final E2 concentration spans from 0.01 to 320 nM in a total 40  $\mu$ l mixture. A sample of 10  $\mu$ l was removed, mixed with around 1.5  $\mu$ l loading dye, and immediately loaded onto a 10% gel. DNA bands were quantified with phosphorimaging. The measured fractions ( $\eta$ ) of E2-bound DNA at different E2 concentrations ([E2]) are fitted with the formula  $\eta = [E2]/(K_d + [E2])$ , yielding the dissociation constant  $K_d$ .



**Figure 1.** DNase I footprinting of the 5'-end labelled complementary strand containing different E2-binding spacer (ATAT, TATA and ACGT) in presence of (a)  $Mg^{2+}$  and (b)  $Ca^{2+}$ . Spacer regions are indicated in boxed area.

## RESULTS AND DISCUSSION

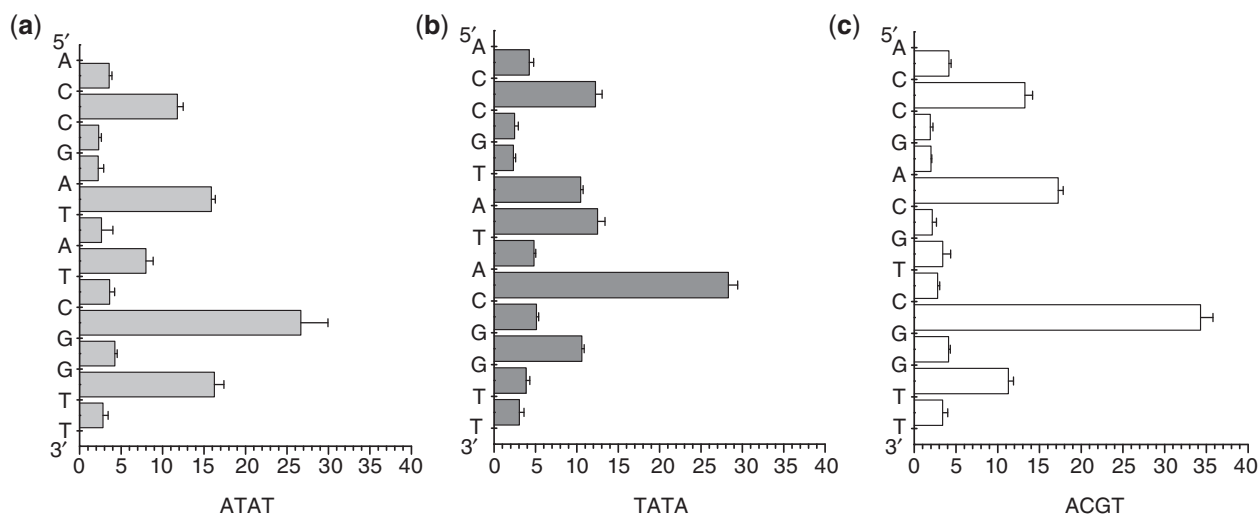
### DNase I digestion of DNA

Bovine pancreatic deoxyribonuclease I (DNase I) attacks the DNA phosphate backbone from the minor groove, with a cleavage rate that decreases where the minor groove is narrow (12,13). It has been widely used to probe detailed sequence-dependent structural variations (14,15). Three 68-mer double-stranded DNAs containing the E2-binding site with different spacer sequences, ATAT, TATA and ACGT, were used in the present study. The 5'- $^{33}P$ -end labelled DNA was digested by DNase I and the hydrolysis products were analysed by electrophoresis on sequencing gels.

An autoradiograph of the patterns of digestion of the different E2 target sequences in the presence of  $Mg^{2+}$  is shown in Figure 1a. It can be seen that DNase I digestion produces essentially the same gel patterns except in regions containing the different binding sites. The intensities of the cleavage products in  $Mg^{2+}$ , measured by scanning densitometry in the binding-site domains, are shown in Figure 2. All three spacer variants display very similar cutting patterns in the 5'-ACCG sequence, which lies 5' to the spacer sequence. As expected, the spacer sequences show distinctive patterns, with relatively strong cleavage at ApT steps in the ATAT and TATA spacers.

In contrast, cutting at the CpG dinucleotide is highly variable. As shown in Figure 2, weak cutting is observed in the 5'-half of the binding site, where the sequences are (C)CpG(A), (C)CpG(T) and (C)CpG(A) in the ATAT, TATA and ACGT spacers, respectively. On the other hand, cutting is strong at CpG in the complementary sequence in the 3'-portion of the binding site for the ATAT and ACGT spacers, for which both sequences are (T)CpG(G). The outlier is the TATA spacer, with sequence (A)CpG(G), for which cutting is weak. A similar pattern of DNase I cutting in the central 8 bp of the ATAT and TATA spacer sequences was observed by Fox (16).

Thus DNase I cleavage at the phosphate group in the CpG 3' binding-site sequence is strongly affected by the neighbouring spacer, reflecting a local alteration of DNA backbone geometry transmitted from spacer nucleotides 5' to the CpG dinucleotide. Reduction in the cleavage rate at CpG presumably reflects a narrowing of the minor groove in the binding site. It is reasonable to infer that this structural effect perturbs the geometry of the major groove, to which an  $\alpha$ -helix from the E2 protein is bound in the complex (17), thereby reducing the binding affinity. Since the binding site is 2-fold symmetric, the CCGT sequence on the other strand should be similarly affected. On the other hand, DNase I cleavage does not reveal any significant spacer-dependent structural



**Figure 2.** Ratio of intensity and total intensities (%) of DNase I cleavage measured by scanning densitometry in the regions of different protein-binding sites (a) ATAT, (b) TATA and (c) ACGT in the presence of 10 mM  $Mg^{2+}$ .

difference in the ACCG strand in the 5'-half of the binding site. It is notable that the C in the CpG step 3' to the spacer is variable in the E2-binding sites on the papilloma virus genomes (18).

The weakly cut CpG in the binding site of the TATA spacer molecule is embedded in an ACG trinucleotide, whereas strong cutting is displayed at the TCG trinucleotide in the other two spacer sequences. A similar effect of weak cutting is also displayed by the ACG trinucleotide of the central ACGT spacer. It is unlikely, however, that the structural effect can be explained entirely on the basis of the ACG trinucleotide, since the spacer sequence TTAA, which contributes this trinucleotide to the adjacent CpG, showed binding affinity within a factor 2 of the value expected on the basis of its curvature and mechanical properties (10).

The effect of  $Ca^{2+}$  on the patterns of cutting by DNase I was also investigated. Figures 1b and 3 reveal clear similarities in cutting pattern to that observed in the presence of  $Mg^{2+}$ , including reduced cutting at CpG adjacent to the TATA spacer.

### Sequence-dependent $Mg^{2+}$ -induced curvature

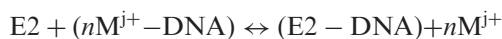
Considerable interest has focused on sequence-specific and ion-dependent divalent cation association with DNA molecules (19–22), which is thought to affect DNA bending and flexibility. Therefore, the interplay between DNA global structure and mechanical properties, cation binding and E2 protein association is intriguing, but has not been specifically investigated in solution.

We compared apparent DNA bending in the absence and presence of  $Mg^{2+}$  utilizing DNA gel migration-anomaly assay for a series of DNA ladders constituted by the E2-binding sites ligated approximately in phase with their helical repeats, with results shown in Figure 4. The presence of  $Mg^{2+}$  retards DNA migration, reflecting reduced net charge. This effect can be removed by normalizing the mobilities to a standard sequence, as

shown in Figure 4. The extent of the relative retardation varies amongst DNA sequences, mainly depending on apparent curvature; TATA exhibits the largest shift when  $Mg^{2+}$  is added, indicating the largest static and/or dynamical structural sensitivity to  $Mg^{2+}$ , whereas ACGT shows the smallest. However, the induced apparent curvature is small when compared with the known (weak) DNA-bending element AATT (Figure 4) as also shown by the crystal structure of the dodecameric E2 DNA target, ACCGAATTCGGT (23).

### Sequence-dependent ion release upon E2 binding

In seeking to understand the weaker binding of E2 to DNA containing the TATA spacer in  $Mg^{2+}$  than predicted on the basis of the curvature and flexibility of the spacer DNA (10), we have investigated the thermodynamic linkage between binding of E2 and a series of ions ( $K^+$ ,  $Na^+$ ,  $Ca^{2+}$  and  $Mg^{2+}$ ) by measuring protein-binding affinity as a function of ion concentration. In its simplest form, in which only cation release is considered, the stoichiometric equation

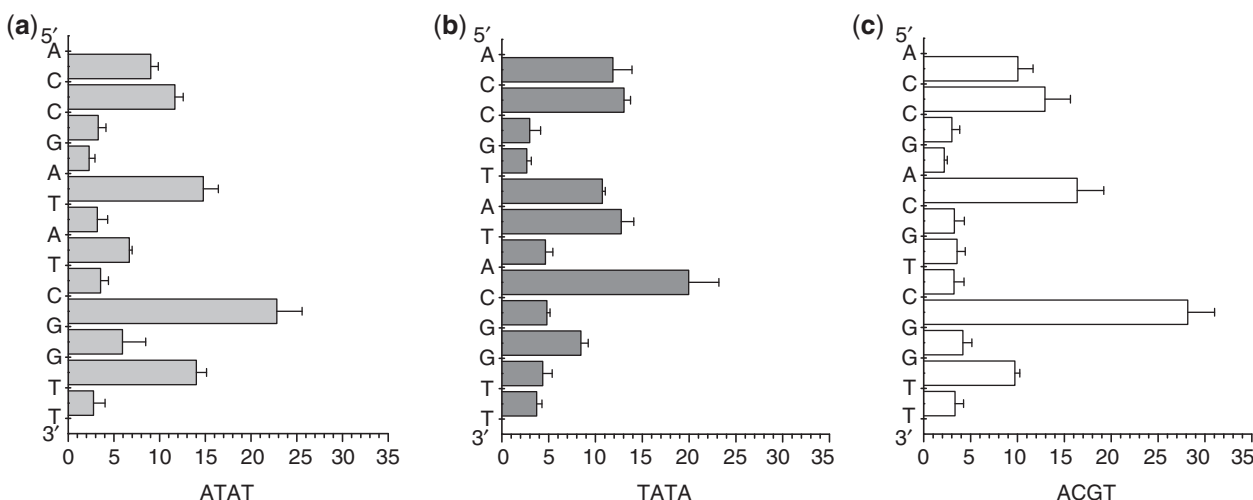


describing the release of  $n$  cations  $M$  of charge  $j^+$  leads to the theoretical dependence of dissociation constant  $K_d$  on ion concentration  $[M]$  according to

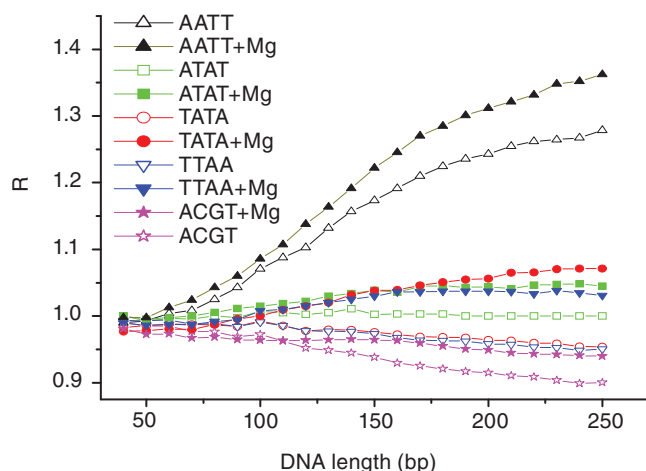
$$\frac{\partial \log K_d}{\partial \log [M]} = n$$

It should be noted that the released ions upon protein binding may come not only from DNA, but also from protein; comparative values from one spacer sequence to another are more significant than absolute numbers.

Figure 5 shows the primary data, and Table 1 summarizes the slopes of the lines in Figure 5, giving the apparent number of ions released. The TATA spacer is clearly exceptional in that the maximal number of  $Mg^{2+}$  is displaced, by a margin that well exceeds experimental



**Figure 3.** Ratio of intensity and total intensities (%) of DNase I cleavage measured by scanning densitometry in the regions of different protein-binding sites (a) ATAT, (b) TATA and (c) ACGT in the presence of 10 mM  $\text{Ca}^{2+}$ .



**Figure 4.** Magnesium enhances DNA curvature. Comparisons of DNA migration anomaly in the absence and presence of magnesium. The gel mobility retardation ( $R$ ) is expressed as ratio of the migration distance of a control DNA ladder to that of indicated DNA molecule. The control DNA sequence used probably has larger curvature than those of several sequences of interest such as ACGT, leading to their mobility  $R < 1$ .

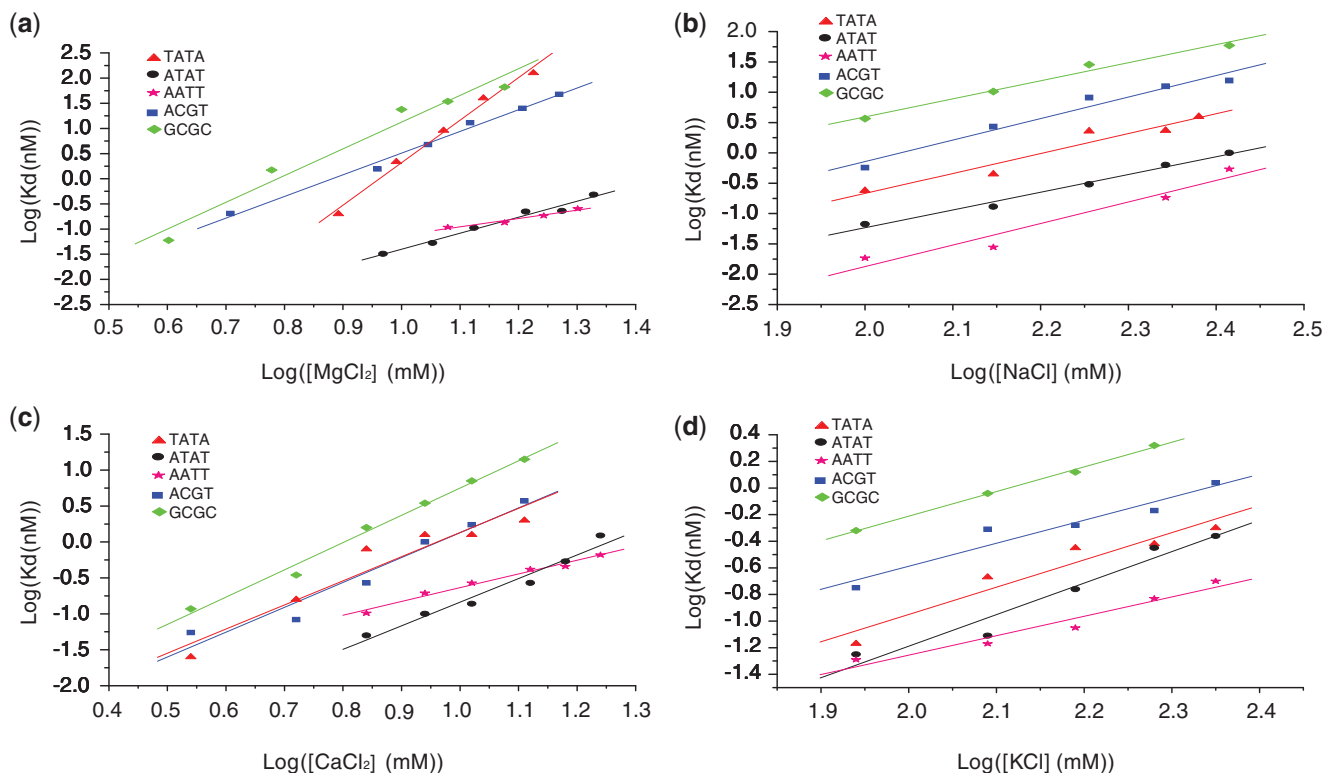
errors. By contrast, in the presence of  $\text{Ca}^{2+}$ , there is no clear difference in  $n$  between TATA and ATAT spacers. Except for TATA, GC-rich spacer sequences tend to release more divalent cations than AT-rich sequences, consistent with the general observation of Chiu and Dickerson (22) that  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  show a preference for binding GC-rich sequence elements compared to AT. Chiu and Dickerson also found that  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  have different modes of binding to localized sequence elements in identical crystal packing environments, which provides a plausible basis for the differential release of  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  from identical spacer sequences.

Divalent cations can better hydrate well-ordered water molecules than can monovalent cations, leading to their sequence-specific hydrogen bonding with DNA bases and

phosphates (20–22), in addition to their direct electrostatic interactions. To test this valence-specific effect, divalent cations in the binding buffer were replaced with monovalent cations ( $\text{Na}^+$  or  $\text{K}^+$ ) and their  $n$  values were determined. As shown in Table 1, ion release varied only slightly over the sequences tested, implying that the possible sequence-dependent variation of the electrical potential of the DNA surface does not result in variations of monovalent cation binding.

Tables 2 and 3, which report the ratio of  $K_d$  values for the closely related TATA and ATAT spacer sequences for the ion series studied, provide further insight into the underlying ion-dependent phenomena. It is notable that the  $K_d$  ratio increases strongly in the order of increasing positive charge density on the ions, in the order  $\text{K}^+$ ,  $\text{Na}^+$ ,  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ . In  $\text{K}^+$ , the relative affinity is within experimental error of the value ( $\sim 2$ ) predicted on the basis of the relative curvature and flexibility values (measured in the presence of  $\text{Mg}^{2+}$ ) (10). For the ions  $\text{K}^+$ ,  $\text{Na}^+$  and  $\text{Ca}^{2+}$ , there is no significant systematic dependence of relative binding constant on ion concentration. We conclude that  $\text{K}^+$ ,  $\text{Na}^+$  and  $\text{Ca}^{2+}$  interfere increasingly with binding to TATA over ATAT, but that this effect is not due to an increase in direct competition for binding sites in one sequence over the other. On the other hand, in  $\text{Mg}^{2+}$  the  $K_d$  ratio increases  $\sim 30$ -fold for a  $\sim 2$ -fold increase in ion concentration, implying that 5–6 more  $\text{Mg}^{2+}$  ions compete with E2 for binding to the TATA sequence compared to ATAT.

Thus, there appear to be two components to the phenomenon at hand. First, ions of increasing positive charge density increasingly distort DNA structure in a way that weakens protein affinity, but protein binding either (or both) does not require dissociation of the  $\text{K}^+$ ,  $\text{Na}^+$  and  $\text{Ca}^{2+}$  ions that induce this effect, or results in equal displacement of these ions from the two sequences. The former could be the case, for example, if the ions interact with the minor groove, while the protein occupies the major groove. Second, in  $\text{Mg}^{2+}$  solution, 5–6 additional



**Figure 5.** Variation of binding constant with (a)  $Mg^{2+}$ , (b)  $Na^+$ , (c)  $Ca^{2+}$  or (d)  $K^+$  concentration. The number of cation release can be calculated by  $n = \partial \log K_d / \partial \log [\text{cation}]$ , or slope of the fitted straight line as shown, with results exhibited in Table 1.

**Table 1.** Number of  $Mg^{2+}$ ,  $Ca^{2+}$ ,  $Na^+$  or  $K^+$  released upon DNA-E2 binding

Sequence	$Mg^{2+}$ released	$Ca^{2+}$ released	$Na^+$ released	$K^+$ released
ACGT	4.3 (0.2)	3.5 (0.4)	3.5 (0.4)	1.7 (0.3)
ATAT	3.2 (0.2)	3.3 (0.3)	3.0 (0.2)	2.4 (0.3)
TATA	8.4 (0.5)	3.4 (0.5)	3.2 (0.5)	2.1 (0.2)
AATT	1.7 (0.3)	1.9 (0.1)	3.5 (0.4)	1.5 (0.2)
GCGC	5.3 (0.6)	5.2 (0.3)	3.0 (0.3)	1.9 (0.1)

**Table 2.** Ratio of dissociation constants for TATA and ATAT spacer constructs

Ion	Concentration range (mM)	$K_d(\text{TATA})/K_d(\text{ATAT})$
$K^+$	100–260	$1.6 \pm 0.7$
$Na^+$	100–260	$4.1 \pm 1.4$
$Ca^{2+}$	8–15	$9.2 \pm 0.3$

ions bind to TATA compared to ATAT in a way that directly competes with binding. This could result, for example, from additional ion binding to the major groove at positions contacted by the protein.

In structural terms, the anomaly shown by the TATA spacer may be related to previous observations that this particular tetranucleotide can adopt an A-DNA-like

**Table 3.** Ratio of TATA and ATAT dissociation constants in  $Mg^{2+}$

$Mg^{2+}$ concentration (mM)	$K_d(\text{TATA})/K_d(\text{ATAT})$
8.0	15
10.0	52
12.0	143
15.0	476

structure characterized by a wide and shallow minor groove either alone, or when bound to the TATA-box-binding protein (24). The structure of the TATA-containing octamer (GGTATACC), in its complex with DNase I, also exhibited A-like features characterized by a wide and shallow minor groove (12) whereas other sequences bound to the enzyme exhibited the B-type conformation (25). Such conformational variations may be related to the different cutting pattern of the TATA spacer in comparison to that of the ATAT and ACGT spacer sequences. The ApC step in the DNase I complex of GGTATACC is the (putative) cutting site (12), analogous to the strong cutting site ApC in the TATA spacer construct. The observation that this site was not cleaved in the complex of the crystal structure could be related to the excess of EDTA and other ingredients used for crystallization (12). We note that the very strong cutting sites in the three constructs seem always to be flanked by quite weak sites, which raises the possibility that there may be local competition effects for binding/cutting. In this

interpretation, the weak cutting at CpG would be a consequence of competition from the strong cutting at the adjacent ApC, supporting the view that the phenomenon results primarily from the shift to A-like geometry induced by the TATA sequence.

The transition from the TATA element structure with A-type characteristics, wide and shallow minor groove, to the one required to form the complex with the E2 protein, namely, a narrow and deep minor groove caused by DNA bending towards the protein, could account for the relatively lower binding affinity of this sequence. The presence of cations may further distort the DNA helix, thus explaining the observed increase in the  $K_d$  and the large release of  $Mg^{2+}$  ions upon E2 binding in comparison to the other sequences. Whereas it is well known that  $Mg^{2+}$  ions are crucial for RNA folding, their roles in mediating DNA structural variation are generally neglected. Our work corroborated results from previous studies and suggested active roles of  $Mg^{2+}$  ions in sequence-specific DNA structures and protein–DNA interactions (26,27).

## FUNDING

National Institutes of Health (GM 21966 to D.M.C.). Funding for open access charge: National Institutes of Health (GM 21966).

*Conflict of interest statement.* None declared.

## REFERENCES

- Otwinowski,Z., Schevitz,R.W., Zhang,R.G., Lawson,C.L., Joachimiak,A., Marmorstein,R.Q., Luisi,B.F. and Sigler,P.B. (1988) Crystal structure of trp repressor/operator complex at atomic resolution. *Nature*, **335**, 321–329.
- Lynch,T.W., Read,E.K., Mattis,A.N., Gardner,J.F. and Rice,P.A. (2003) Integration host factor: putting a twist on protein–DNA recognition. *J. Mol. Biol.*, **330**, 493–502.
- Berman,H.M., Olson,W.K., Beveridge,D.L., Westbrook,J., Gelbin,A., Demeny,T., Hsieh,S.H., Srinivasan,A.R. and Schneider,B. (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.*, **63**, 751–759.
- Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein–DNA recognition. *Nature*, **461**, 1248–1253.
- Dickerson,R.E. (1998) DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.*, **26**, 1906–1926.
- Koudelka,G.B., Harrison,S.C. and Ptashne,M. (1987) Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and Cro. *Nature*, **326**, 886–888.
- Kim,S.S., Tam,J.K., Wang,A.F. and Hegde,R.S. (2000) The structural basis of DNA target discrimination by papillomavirus E2 proteins. *J. Biol. Chem.*, **275**, 31245–31254.
- Sanders,C.M. and Maitland,N.J. (1994) Kinetic and equilibrium binding studies of the human papillomavirus type-16 transcription regulatory protein E2 interacting with core enhancer elements. *Nucleic Acids Res.*, **22**, 4890–4897.
- de Prat-Gay,G., Gaston,K. and Cicero,D.O. (2008) The papillomavirus E2 DNA binding domain. *Front. Biosci.*, **13**, 6006–6021.
- Zhang,Y., Xi,Z., Hegde,R.S., Shakked,Z. and Crothers,D.M. (2004) Predicting indirect readout effects in protein–DNA interactions. *Proc. Natl Acad. Sci. USA*, **101**, 8337–8341.
- Hegde,R.S. and Androphy,E.J. (1998) Crystal structure of the E2 DNA-binding domain from human papillomavirus type 16: implications for its DNA binding-site selection mechanism. *J. Mol. Biol.*, **284**, 1479–1489.
- Weston,S.A., Lahm,A. and Suck,D. (1992) X-ray structure of the DNase I-d(GGTATACC)<sub>2</sub> complex at 2.3 Å resolution. *J. Mol. Biol.*, **226**, 1237–1256.
- Hogan,M.E., Roberson,M.W. and Austin,R.H. (1989) DNA flexibility variation may dominate DNase I cleavage. *Proc. Natl Acad. Sci. USA*, **86**, 9273–9277.
- Suck,D. and Oefner,C. (1986) Structure of DNase I at 2.0 Å resolution suggests a mechanism for binding to and cutting DNA. *Nature*, **321**, 620–625.
- Drew,H.R. and Travers,A.A. (1985) DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.*, **186**, 773–790.
- Fox,K.R. (1992) Probing the conformations of eight cloned DNA dodecamers; CGCGAATTCGCG, CGCGTTAACGCG, CGCGTATACGCG, CGCGATATCGCG, CGCAAATTTGCG, CGCTTAAAGCG, CGCGGATCCGCG and CGCGGTACCGCG. *Nucleic Acids Res.*, **20**, 6487–6493.
- Hegde,R.S., Grossman,S.R., Laimins,L.A. and Sigler,P.B. (1992) Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature*, **359**, 505–512.
- Hegde,R.S. (2002) The papillomavirus E2 proteins: structure, function, and biology. *Annu. Rev. Biophys. Biomol. Struct.*, **31**, 343–360.
- Hud,N.V. and Polak,M. (2001) DNA-cation interactions: The major and minor grooves are flexible ionophores. *Curr. Opin. Struct. Biol.*, **11**, 293–301.
- Subirana,J.A. and Soler-Lopez,M. (2003) Cations as hydrogen bond donors: a view of electrostatic interactions in DNA. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 27–45.
- Hud,N.V. and Plavec,J. (2003) A unified model for the origin of DNA sequence-directed curvature. *Biopolymers*, **69**, 144–158.
- Chiu,T.K. and Dickerson,R.E. (2000) 1 Å crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium. *J. Mol. Biol.*, **301**, 915–945.
- Hizver,J., Rozenberg,H., Frolow,F., Rabinovich,D. and Shakked,Z. (2001) DNA bending by an adenine–thymine tract and its role in gene regulation. *Proc. Natl Acad. Sci. USA*, **98**, 8490–8495.
- Guzikevich-Guerstein,G. and Shakked,Z. (1996) A novel form of the DNA double helix imposed on the TATA-box by the TATA-binding protein. *Nat. Struct. Biol.*, **3**, 32–37.
- Lahm,A. and Suck,D. (1991) DNase I-induced DNA conformation. 2 Å structure of a DNase I-octamer complex. *J. Mol. Biol.*, **222**, 645–667.
- McFail-Isom,L., Sines,C.C. and Williams,L.D. (1999) DNA structure: cations in charge? *Curr. Opin. Struct. Biol.*, **9**, 298–304.
- McConnell,K.J. and Beveridge,D.L. (2000) DNA structure: what's in charge? *J. Mol. Biol.*, **304**, 803–820.