Sequence Discrimination by DNA-binding Domain of ETS Family Transcription Factor PU.1 Is Linked to Specific Hydration of Protein-DNA Interface

Received for publication, January 12, 2012, and in revised form, March 23, 2012 Published, JBC Papers in Press, April 2, 2012, DOI 10.1074/jbc.M112.342345

Gregory M. K. Poon

From the Department of Pharmaceutical Sciences, Washington State University, Pullman, Washington 99164-6534

**Background:** PU.1 recognizes a large number of binding sites with differing affinities.

**Results:** The role of hydration in sequence-specific binding by the ETS domain of PU.1 was determined.

**Conclusion:** Sequence discrimination is linked to the uptake of specifically bound waters at the protein-DNA interface.

**Significance:** The sequence specificity of PU.1 ETS is directly related to the transactivational activity and frequency of in vivo binding sites for the full-length protein.

PU.1 is an essential transcription factor in normal hematopoietic lineage development. It recognizes a large number of promoter sites differing only in bases flanking a core consensus of 5’-GGAA-3’. DNA binding is mediated by its ETS domain, whose sequence selectivity directly corresponds to the transactivational activity and frequency of binding sites for full-length PU.1 in vivo. To better understand the basis of sequence discrimination, we characterized its binding properties to a high affinity and low affinity site. Despite sharing a homologous structural framework as confirmed by DNase I and hydroxyl radical footprinting, the two complexes exhibit striking heterogeneity in terms of hydration properties. High affinity binding is destabilized by osmotic stress, whereas low affinity binding is insensitive. Dimethyl sulfate footprinting showed that the major groove at the core consensus is protected in the high affinity complex but accessible in the low affinity one. Finally, destabilization of low affinity binding by salt is in quantitative agreement with the number of phosphate contacts but is substantially attenuated in high affinity binding. These observations support a mechanism of sequence discrimination wherein specifically bound water molecules couple flanking backbone contacts with base-specific interactions in a sequestered cavity at the core consensus. The implications of this model with respect to other ETS paralogs are discussed.

PU.1 (or Spi-1) is a lineage-restricted member of the ETS family of transcription factors that shares a structurally conserved DNA-binding domain. It binds a large number of cognate sequences (mostly of myeloid and lymphoid origin) harboring a central 5’-GGAA-3’ consensus (1, 2). These sites exhibit significant variation in the flanking bases and range over 400-fold in affinity (3). Although the combinatorial nature of transcriptional control is well established, the intrinsic selectivity of the ETS domain’s intrinsic selectivity per se is also biologically relevant. The latter point is demonstrated by its exact relationship with transactivational activity by full-length PU.1 (4) as well as the close correspondence between high affinity sequences identified in vitro and their prevalence among ETS binding sites in vivo (1, 2). Given current interest in the transcriptional specificity of ETS proteins at the genomic level (5–7), knowledge of the physical determinants for the intrinsic selectivity of ETS proteins is essential to understanding higher order protein-protein and protein-DNA interactions found in the transcription regulatory complex.

In the co-crystal structure of the ETS domain of PU.1 bound to a high affinity binding site (8), the recognition helix makes base-specific contacts (several of which are water-mediated) in the major groove of the 5’-GGAA-3’ core consensus. In the adjacent minor grooves, the protein makes contacts exclusively with DNA phosphate and deoxyribose atoms. Although this organization is consistent with the existence of flanking sequence variants, it does not explain their non-degenerate occurrences that represent a ubiquitous feature of all ETS domains (1). The complexity of flanking sequence discrimination is demonstrated by two complementary lines of experiments. In screening experiments, the occurrence of certain bases at a flanking position is strongly associated with a particular base neighbor (2). In direct measurements of sequence-specific binding, each flanking base position contributes non-additively to the overall affinity (3) as well as the underlying thermodynamics (9). To account for these observations, an “indirect readout” mechanism of DNA structure as encoded by the flanking sequences was proposed (9). However, the physical nature of this “code” remains unresolved. A major missing piece to this puzzle is the lack of biophysical data on low affinity flanking sequence variants. Characterization of these variants is essential, even if they are not highly represented in in vivo binding sites, because this knowledge is ultimately required to complete our understanding of the intrinsic preference of ETS protein for one binding site over another. Unfortunately, comparative studies of core consensus variants 5’-GGAN-3’ (10–13) do not directly answer this question because all four core bases receive base-specific inputs.

---

* This investigation was supported by the College of Pharmacy, Washington State University.

** This article contains supplemental Table S1 and Figs. S1–S5.

1 To whom correspondence may be addressed: P.O. Box 646534, Pullman, WA 99164-6534. Tel.: 509-335-8341; Fax: 509-335-5902; E-mail: gpoon@wsu.edu.
Specific Hydration in PU.1 ETS Binding

To address the physical contribution of the flanking sequences to sequence discrimination by the PU.1 ETS domain, we investigated the structure and binding properties of two PU.1 ETS-bound sites differing strictly in the flanking sequences. Osmotic stress experiments revealed the unusual finding that high affinity specific binding is associated with an apparent net uptake of water molecules. In contrast, low affinity specific binding is manifestly hydration-neutral. Dimethyl sulfate footprinting detected a putative solute-excluding cavity at the major groove of the core consensus in the high affinity complex but not in the low affinity one. DNase I and hydroxyl radical footprinting identified no other major conformational differences along the protein-DNA interface. The two sequence-specific complexes also differ significantly in terms of destabilization by NaCl, which was interpreted in terms of attenuation of phosphate neutralization by specifically bound waters. We propose a model of sequence-specific discrimination wherein flanking backbone contacts are coupled to base-specific interactions between the recognition helix and the core consensus via a network of specifically bound water molecules.

EXPERIMENTAL PROCEDURES

Cloning and Preparation of DNA Fragments—A ~130-bp fragment harboring a single 23-bp ETS binding site (Table 1) was cloned into the EcoRI/HindIII sites of pUC19. For non-radioactive experiments, DNA fragments (206 bp) were amplified by PCR using standard M13 forward and reverse primers. For footprinting experiments, internal primers were used to generate shorter, singly 5′-radiolabeled fragments by PCR. All amplicons were purified from an agarose gel.

Protein Preparation—A synthetic DNA fragment encoding residues 165–272 of murine PU.1 was cloned between the NcoI and BgIII sites of pQE-60 (Qiagen). This construct contains a C-terminal His6 tag separated by a thrombin cleavage site. Protein was overexpressed in Escherichia coli M15[pREP4] and purified by standard affinity chromatography. The His6 tag was cleaved by incubation with 10 units/ml thrombin overnight at room temperature and removed on a Superdex 30/200 column. Detagged, purified PU.1 ETS domain was dialyzed into binding buffer (described below) and quantified by UV spectrometry (ε280 = 22,460 M⁻¹ cm⁻¹).

Buffers and Reaction Conditions—The binding buffer in all experiments was 10 mM sodium phosphate, pH 7.4, prepared by diluting stocks of Na₂HPO₄ and NaH₂PO₄. NaCl was added to achieve the required [Na⁺], taking into account contributions from the buffer salts. Minor salt-induced changes in pH were not expected to influence DNA binding, which is pH-independent between pH 6.7 and 9.0 (14). Samples were extensively incubated under the stated conditions to achieve equilibrium; high viscosity samples were routinely equilibrated for up to 24 h.

Filter Binding Experiments—Experimental procedures were as described (3, 9) with the following modifications. Double-stranded radiolabeled oligonucleotides were prepared by labeling one strand (1 nmol) with [γ-³²P]ATP, quantified by UV spectrometry, and annealed with the complementary strand at a 1.2-fold excess. The specific activity (SA) of the labeled DNA was calculated by scintillation counting of an aliquot (of known total concentration) on the day of the experiment. Observed binding to the filters as a function of total DNA oligonucleotide concentration [D] was fitted to the following equation,

\[ B_{obs} = \epsilon (B_0 + C([D]_l - B_0)) \]  

(Eq. 1)

where \( B_0 \) is the concentration of PU.1 ETS-bound radiolabeled DNA, and \( \epsilon \) is the filter efficiency. Background DNA binding to filters was measured in ETS-free samples and estimated simultaneously with ETS-containing samples to define the linear coefficient \( C. \) \( B_0 \) is computed numerically as the real, physical root of the quadratic equation, \( \varphi_2 B_0^2 + \varphi_1 B_0 + \varphi_0 = 0 \), where \( \varphi_2 = -K_D [D]_l, \varphi_1 = S_A K_D [D]_l ([D]_l + p_S), \) and \( \varphi_0 = -S_A^2 K_D^2 p_S. \) SA is expressed as a fraction of [D]ₗ. When the concentration of PU.1 ETS \( p_S \) and SA enter as constants, no significant correlation was observed among the remaining parameters.

DNA Footprinting—ETS-DNA binding was performed in 50 μl of binding buffer containing ~1 nm singly labeled DNA fragment, PU.1 ETS domain, acetylated BSA (100 ng/μl), and poly(dAT)poly(dAT) (10⁻³ M bp). In some experiments requiring higher sequestration of nonspecific binding, purified salmon testes DNA (0.75 mM bp) was used instead of the synthetic copolymer. At equilibrium, enzymatic or chemical footprinting was performed under conditions that introduced <30% total cleavage in order to maximize the fraction of singly cleaved DNA. Selected chemical sequencing reactions were also performed for base identification. DNase I footprinting was performed following standard procedures. Hydroxyl radical footprinting was performed as described by Tullius and Dombroski (15). For methylation protection footprinting, dimethyl sulfate (DMS)² was diluted 1:5 in water just prior to use. One μl was thoroughly mixed into the ETS-DNA mixture. After 20 s, the reaction was quenched by ethanol precipitation. The recovered DNA was dissolved in 100 μl of T0.1E, phenol-extracted, and washed with diethyl ether to remove protein. Piperidine was added to 10%, heated at 90 °C for 5 min, and reprecipitated with ethanol. Purified DNA was separated by standard denaturing electrophoresis.

Osmotic Stress Experiments—The effect of osmolytes on the binding of PU.1 ETS to various binding sites was measured as a competition for 10 nM PU.1 ETS by 1 nM high affinity DNA fragment and graded concentrations of various DNA oligonucleotides. Acetylated BSA was also present at 100 ng/μl in a 20-μl volume. Osmolytes were dissolved or diluted in binding buffer; samples containing triethylene glycol were adjusted for total [Na⁺]. Stoichiometric molality was converted to osmolality by interpolation of reported values. Data for triethylene glycol, glycerol, and glycerol were taken from Refs. 8, 16, and 24, respectively. At equilibrium, samples were loaded onto a running 6% polyacrylamide gel. After electrophoresis, the gel was stained with SYBR Gold (Invitrogen). No significant changes in quantum yield due to protein binding were observed

² The abbreviations used are: DMS, dimethyl sulfate; NS, nonspecific; Osm, osmolality; osm, osmolar.
TABLE 1
Sequences and affinities of the PU.1 ETS binding sites studied

| Site | [-] flanking segment | Core | [+] flanking segment | $K_D$, M | $\Delta G^{\circ}c$, kJ mol$^{-1}$ |
|------|----------------------|------|----------------------|---------|-------------------------------|
| [-]GC | AGC | GGAA | GTG | (7.3 ± 3.1) x 10$^{-10}$ |
|       | TCG | CCTT | CAC | |
| [+T]G | GCGATA | AAA | TGG | (1.1 ± 0.2) x 10$^{-7}$ | +12.5 ± 1.2 |
|       | CGCTTAT | TTT | ACC | |
| NS   | AGC | GAGA | GTG | (1.5 ± 0.1) x 10$^{-6}$ | +18.9 ± 1.2 |
|       | TCG | CTCT | CAC | |

$^a$ The notations [-] and [+T] refer to positions 5' and 3' from the core in the GGAA strand. The letters refer to base substitutions relative to the AB motif of the Ig2-4 enhancer: 5'AAACGGAAAGTC-3'.

$^b$ Filter binding data as fitted to Equation 1.

$^c$ Relative to the [-]GC site.

reported for the SYBR series of dyes (18). Fractional bound fragment ($f_b$) as a function of total DNA oligonucleotide concentration ([D$_1$]) was fitted to the positive, real root of the following cubic equation,

$$
\varphi_1 f_b^3 + \varphi_2 f_b^2 + \varphi_3 f_b + \varphi_4 = 0
$$

(Eq. 2)

where $\varphi_1 = (K_1 - K_3) [D_1]$, $\varphi_2 = K_2 (K_2 - [D_1] + [D_2] + p_1) - K_1 (K_2 + 2[D_2] + p_1)$, $\varphi_3 = K_2 ([D_1] - p_1) + K_1 [D_2] + 2p_1$, and $\varphi_4 = -K_1 p_1$. This model yields absolute equilibrium dissociation constants for the DNA fragment ($K_i$) and oligonucleotide ($K_j$) from a competition experiment. When total concentrations of DNA fragment ([D$_2$] = 1 nm) and PU.1 ETS domain ($p_1 = 10$ nm) enter as constants, no significant correlation was observed among the fitted values of the dissociation constants.

Computations—All parameters ± S.E. were estimated by nonlinear regression. Electroreoretic data were digitized using a Storm 860 instrument (GE Healthcare). For phosphorylating data, lane traces were fitted to a superposition of Lorentzian functions without base line (19). Stained gels (excitation/emission = 450/520 nm) were fitted with Gaussian peaks on a heuristic base line to account for background staining. Structural models were rendered using PyMOL; water-accessible cavities were identified using Caver (20).

RESULTS

We have carried out a detailed comparison of the structural and binding properties of a high affinity, low affinity, and non-specific PU.1 ETS-DNA complex (Table 1). They were selected from studies in which the affinities of specific binding sites have been directly determined (3, 9, 14). These sequences are variations of the AB motif of the Ig2-4 enhancer, a native high affinity PU.1 binding site (21). Based on the PU.1 ETS-DNA co-crystal structure (8), the altered flanking positions contact the ETS domain via backbone phosphate or deoxyribose atoms only. The nonspecific sequence is derived from [-]GC, in which the 5’-GGAA-3’ consensus has been changed to 5’-GAGA-3’. The affinities of these sequences for PU.1 ETS were determined by filter binding titrations at 150 mm Na$^+$ and pH 7.4 (supplemental Fig. S1). The equilibrium dissociation constants of [-]GC and [+T]G yield a difference of 160-fold or 13 kJ/mol in free energy (Table 1), in agreement with reported values (3). The dissociation constant of the nonspecific (NS) sequence, which has not been characterized, is 1.5 ± 0.1 μM, about 10-fold higher than [+T]G. Relative to the parent [-]GC site, this mutation of the core consensus results in a loss of over 2,000-fold in affinity, equivalent to 19 kJ/mol in binding free energy.

The similarity in affinity between the [+T]G and NS sites calls into question the extent to which low affinity “sequence-specific” binding is differentiated from “nonspecific” binding. Such differences should be manifest in the structure of the protein-DNA interface. DNase I and hydroxyl radical footprinting was therefore performed to characterize the contact interface for the three PU.1 ETS-DNA complexes. Both the [-]GC and [+T]G complexes yielded footprints characteristic of sequence-specific ETS-DNA complexes (Fig. 1 and supplemental Fig. S2). Specifically, the TTCC strand shows a strongly hypersensitive band at N$_1$ (5’-TpCpCpCpN$_1$ pN$_2$-3’) just 3’ to the core consensus. This hypersensitivity is a universal hallmark of sequence-specific ETS-DNA complexes (22–26). It is attributed to a widening of the minor groove by the recognition helix (h3) at the major groove of the core consensus. When probed by hydroxyl radicals, the footprints for both [-]GC and [+T]G sequences are highly similar (supplemental Fig. S2, C-H) and correspond to reported examples of sequence-specific PU.1 ETS-DNA complex (22, 27). Hydroxyl radicals attack deoxyribose hydrogens in the minor groove (28). Accordingly, positions flanking the core consensus are protected, whereas those along the core consensus itself are fully accessible. The nonspecific complex, in contrast, produced a diffuse footprint without DNase I hypersensitivity or protection from hydroxyl radicals at the corresponding flanking positions. We conclude, therefore, that the [+T]G sequence is a low affinity but sequence-specific binding site.

High Affinity PU.1 ETS Binding Is Associated with Net Water Uptake—The presence of water clusters along the interface of the PU.1 ETS-DNA co-crystal structure suggests that structural waters play a dominant role in high affinity sequence-specific binding. To test this possibility, we used the osmotic stress technique (18) to uncover differences in water uptake or
release upon complex formation. Chemically inert osmolytes modulate the behavior of proteins and nucleic acids through differences in their preferential interactions with water and other solutes in the vicinity of the biopolymers versus bulk solvent (reviewed in Refs. 29 and 30). The effect of osmotic stress on complex stability was quantitatively measured in terms of the equilibrium constant. We set up competitive titrations of \([-\text{GC}]\) fragment (206 bp, 10^{-8} M duplexes) and PU.1 ETS (10^{-8} M) by oligonucleotides (23 bp) encoding the \([-\text{GC}], [+\text{TG}], \text{or NS}\) sequence. The data were fitted to a model (Equation 2) formulated to directly yield absolute dissociation constants (not transfer constant) for each sequence encoded by the oligonucleotide.

As shown in Fig. 2, competitive binding to the high affinity \([-\text{GC}]\) oligonucleotide was destabilized in the presence of osmolyte. In contrast, competitive binding to the low affinity but sequence-specific \([+\text{TG}]\) oligonucleotide appeared to be indifferent. The latter observation was not an artifact arising from native electrophoresis of a low affinity complex because this technique was able to resolve the nonspecific complex. In the absence of osmolyte, the fitted dissociation constants agree with those measured under the same solution conditions by filter binding (Table 1). Inclusion of osmolyte (triethylene glycol) in the gel matrix (31) made no discernable difference in the data.

In total, we measured the dependence of the binding constant \(K_B (= 1/K_a \text{ from Equation 2})\) on the osmolality of four osmolytes: triethylene glycol, glycine betaine, nicotinamide, and sucrose. Because osmolality (Osm) corresponds to osmotic pressure \(\pi\) and water activity \(a_w\) via Osm = \(\pi/RT = -55.5\ln a_w\);

\[
\frac{\partial \ln K_B}{\partial \text{Osm}} = \frac{\partial \ln K_B}{\partial \text{Osm}} = -55.5 \frac{\partial \ln a_w}{55.5} \quad \text{(Eq. 3)}
\]

\(\Delta \Gamma_{PW}\) represents the change in preferential hydration of the PU.1 ETS-DNA complex relative to the unbound species (29). Negative values of \(\Delta \Gamma_{PW}\) arise from stabilization of the complex by osmolytes and correspond to a net release of water molecules upon binding. The effects of the four osmolytes on complex stability for \([-\text{GC}],[+\text{TG}],\text{and NS}\) sequences are shown in Fig. 3; estimates of the slopes and \(\Delta \Gamma_{PW}\) are given in Table 2. The sequence-dependent responses to osmolytes as shown in Fig. 2 are reproduced for all four osmolytes. The destabilization of the high affinity \([-\text{GC}]\) site (positive \(\Delta \Gamma_{PW}\)) indicates a net uptake of water upon binding to PU.1 ETS. Curvature is observed with increasing osmolality, but \(\Delta \Gamma_{PW}\) remained linear up to at least 2.5 osm. In contrast, the low affinity, sequence-specific site \([+\text{TG}]\) is essentially insensitive to all osmolytes up to 6 osm, suggesting little, if any, net change in hydration between the bound and unbound states. This apparent coupling of water uptake and high affinity binding suggest that water molecules play a specific, stabilizing role in the sequence-specific complex. In contrast, the nonspecific complex is associated with net water release (negative \(\Delta \Gamma_{PW}\)). It is also more strongly stabilized by nicotinamide than the other osmolytes, which behave similarly to each other with respect to the specific complexes. Clearly, differences in hydration are strongly related to sequence specificity and affinity for PU.1 ETS binding.

Preferential hydration and preferential exclusion of osmolyte in the vicinity of biopolymers are coupled properties.
Effect of osmotic stress on the binding properties of PU.1 ETS

| Oligonucleotide | \( K_d(0) \) | \( \Delta \ln K_d/\Delta \text{Osm} \) | \( \Delta \Gamma_{PW}^b \) |
|-----------------|--------------|-------------------------------|-------------------------|
| \([-]\)GC        | \( 2.4 \pm 0.6 \times 10^{-10} \) | \( -1.8 \pm 0.2 \) | \( 98 \pm 10 \) |
| \([+]\)TG       | \( 2.1 \pm 0.5 \times 10^{-7} \) | \( 0.023 \pm 0.072 \) | \( -0 \) |
| NS              | \( 9.9 \pm 1.3 \times 10^{-6} \) | \( 0.53 \pm 0.05^e \) | \( -30 \pm 3^d \) |

* Equilibrium dissociation constant extrapolated to 0 osm.
* Negative values denote a net release of water.
* Linear fit of the data up to 2.5 osm. For reference, the limiting slope extrapolated to 0 osm for the parabola used to fit the entire range is \(-1.9 \pm 0.3\).
* Excludes data for nicotinamide.
* Nicotinamide data only.

Osmotic stress perturbs PU.1 ETS binding to DNA in a sequence-dependent manner. Binding of PU.1 ETS to a DNA fragment harboring a high affinity specific binding site (\([-]\)GC) was competed with graded concentrations of DNA oligonucleotides encoding the \([-]\)GC site, a low affinity but specific site (\([+]\)TG), or a nonspecific site (NS). ETS-bound and unbound DNA fragments were separated at equilibrium by electrophoretic mobility shift. Shown are best fits of the titration data to Equation 2 were averaged from replicate experiments (up to three for each condition) and converted to \( \ln K_d \pm \text{S.E. (error bars)} \). The lines represent error-weighted fits of the aggregate data for each binding site to a line or parabola (\([-]\)GC, dashed line). For \([-]\)GC, there is a significant increase in the weighted sum of squares if the full range of data is fitted to a line (F test, \( p = 0.00268 \)) but not for the subset up to 2.5 osm (\( p = 0.71 \)). The dependence on nicotinamide for the NS binding site was fitted separately from data for the other osmolytes. The fitted estimates of the slopes are given in Table 2.

Specific Hydration in PU.1 ETS Binding

sequently, the magnitude of \( \Delta \Gamma_{PW}^b \) is expected to differ when the same macromolecular system is probed with osmolytes differing in physicochemical properties and therefore their extent of preferential exclusion. The simplest explanation for the apparent lack of dependence on osmolyte identity is that \( \Delta \Gamma_{PW}^b \) is dominated by the movement of water into or out of a cavity that excludes all of the osmolytes used (29). To check that the osmolytes are indeed acting “osmotically,” the integrity of the protein-DNA interface was confirmed by DNase I footprinting at 1.5 osm (supplemental Fig. S3).

Differential Accessibility of PU.1 ETS/DNA Interface to Osmolytes—Based on our interpretation (which we elaborate at length under “Discussion”) of the osmotic stress data, the low affinity \([+]\)TG complex, which is insensitive to osmolyte, should lack the asymmetrically accessible (to water but not osmolytes) compartment found with the \([-]\)GC site. The co-crystal structure suggests that the most likely candidate for such a cavity is the major groove of the core consensus, where a number of crystal waters are observed. Therefore, we probed the protein-DNA interface by titration footprinting using DMS. DMS methylates the N7-positions of purines, which are accessible only from the major groove. It is also comparable in size with the osmolytes used in the osmotic stress experiments and is therefore well suited to probe the interface on the GGAA strand.

The DMS footprints of the GGAA strand for the \([-]\)GC and \([+]\)TG complexes (Fig. 4) show clearly that PU.1 ETS strongly protects the guanines (\( G_{-1} \) and \( G_0 \)) in the core consensus of \([-]\)GC. In contrast, the corresponding guanines in
Specific Hydration in PU.1 ETS Binding

the low affinity [+]+TG sequence are accessible to DMS. The lack of protection from DMS for the [+]+TG site is not due to its low affinity for PU.1 ETS because DNase I (Fig. 1) and hydroxyl radicals (supplemental Fig. S2) quantitatively score the fractional binding under the identical conditions. For both sites, the major groove at the flanking contacts is solvent-exposed and therefore unprotected from DMS methylation.
Specific Hydration in PU.1 ETS Binding

Salt Dependence of PU.1 ETS-DNA Binding—The quantitative dependence of the binding constant on salt (monovalent cation M⁺) activity is classically interpreted in terms of the polyelectrolyte theory (32) to give the number of phosphate contacts, Z.

\[
\frac{\partial \ln K_B}{\partial \ln [M^+]\psi} = -Z \psi 
\]  
(Eq. 4)

\[\Psi = 0.88\] for double-stranded B-DNA. The salt dependence of several high affinity PU.1 ETS-DNA complexes (\[K_B < 10^{-7} M\]) has been determined previously (9). In all cases, Z is significantly smaller (\(\approx 4\)) than the number of seven phosphate contacts observed in the co-crystal structure. We measured the dependence of PU.1 ETS binding to the [+T]G and NS site on NaCl (Fig. 5). In contrast with high affinity specific sequences, binding to the [+T]G site is significantly more sensitive to salt. Application of Equation 4 yields 9 ± 1 phosphate contacts for [+T]G, which are in close agreement with the co-crystal structure. Similar to other specific sites, linearity in \(\ln K_B\) breaks down below 0.1 M NaCl (9), consistent with instability of the native protein structure at very low salt (33). For the NS site, 5 ± 1 appears to be involved in PU.1 ETS binding, and \(\ln K_B/\ln [M^+]\) linearity below 0.1 M NaCl is maintained.

DISCUSSION

Sequence-specific PU.1 ETS-DNA complexes exhibit structural and thermodynamic heterogeneity that implicate the formation of specifically coordinated water molecules along the protein-DNA interface with high affinity binding. The application of osmotic stress destabilizes the high affinity complex with the [-]GC site but exerts little effect on the low affinity complex with [+T]G. The major groove of the core consensus excludes dimethyl sulfate in the [-]GC complex but remains accessible in the low affinity [+T]G complex. Finally, high affinity complexes are significantly less sensitive to salt than expected for the number of known phosphate contacts in the co-crystal structure, whereas good agreement was found for the low affinity site [+T]G. The ETS domain of PU.1, as with ETS domains in general, exhibits a dispersion of affinities toward a large number of binding sites differing only in the bases flanking the core consensus. If flanking sequence variants share a common structural framework, the present data argue for a physical linkage of flanking and core contacts by specifically bound water molecules.

Protein-Solvent Interactions as Window into PU.1 ETS-DNA Binding—Protein-DNA binding is typically accompanied by dehydration (i.e. water release) of the contact interface. Because macromolecular surfaces in general preferentially exclude solutes (or, equivalently, are preferentially hydrated) (34–36), osmolytes drive complex formation to minimize these solute-exposed surfaces. Moreover, high affinity interactions that appose highly complementary surfaces are expected to displace more water than low affinity and nonspecific complexes (18). Our observation that PU.1 ETS binding to the high affinity [-]GC site is destabilized by osmotic stress is therefore highly unusual. Perturbations due to direct binding of the osmolytes are improbable based on the highly similar effects of all four osmolytes, which vary markedly in physicochemical properties (supplemental Table S1). If the observed reduction in affinity were due to osmolyte binding, the binding constants and capacities would have to be implausibly similar for all four osmolytes. Electrostatic perturbations are also unlikely because these osmolytes differ qualitatively in their effects on the solution dielectric. For both the [-]GC and [+T]G, all four solutes behave identically as a function of osmolality (Fig. 3).

In the paradigm of preferential hydration, the apparent uptake of water by [-]GC implies an increase of solute-exposed surface area upon protein-DNA binding. However, any major folding coupled to DNA binding is incompatible with spectroscopic data indicating that the unbound ETS domain is substantially folded as observed in the co-crystal structure (14, 33). Calculations based on a rigid body association give a loss of >2,400 Å² of solvent-accessible surface area (14), a relatively small but still significant amount of burial.

To reconcile these observations, we propose that sequence-specific binding by PU.1 ETS is dependent on specific hydration. Specifically, optimal interactions within the contact interface require the formation of a network of water-mediated contacts in a cavity that sterically excludes other solutes. The resultant osmotic pressure penalty would account for the observed linear dependence on osmolality up to about 2.5 osm and the lack of dependence of this effect on osmolyte identity. At higher osmolality, osmotic pressures become sufficient to strip these specifically bound water molecules, leading to the observed curvature in Fig. 4. The putative identification of this cavity by DMS protection in the [-]GC complex, but not the low affinity [+T]G one, is consistent with this explanation. An analysis of the co-crystal structure (8) also reveals such a solute-excluding cavity at the core consensus (Fig. 6A). An osmolyte-accessible interface of the [+T]G complex is manifest as inen-
Specific Hydration in PU.1 ETS Binding

sitivity to osmolytes. In the model described in the following, the loss of osmotic responsiveness reflects a suboptimal core interface at the [+]TG site that fails to build specific hydration networks.

If the $\Delta \ln K_{\text{m}} / \partial \text{Osm}$ slope is interpreted strictly in terms of osmotic stress (Equation 3), it translates to a net uptake of $96 \pm 11$ water molecules for the high affinity [\(+]\)GC site relative to the low affinity [\(-\)]TG site (Table 2). This quantity corresponds to associated water molecules that are detectable thermodynamically upon binding. If at least some of the sequestered water molecules participate in ordered interactions within the binding interface, this slope would include high energy contributions (37) in addition to the osmotic work of filling the cavity. Eleven such water molecules are observed at the core consensus in the co-crystal structure (8). Given a nominal molecular volume of $30 \, \text{Å}^3$ for water (assuming normal density), these waters are just accommodated within the volume of the solute-excluding cavity (337 $\, \text{Å}^3$) computed in Fig. 6A.

Model for Site Selection by PU.1 ETS Domain—All of the present data can be accounted for by a model for site selection in which core interactions are coupled to flanking contacts by bridging water molecules. Although the co-crystal structure does not permit an unambiguous assignment of hydrogen atoms, a network of hydrogen bonds can be drawn from backbone to nucleobase contacts via specifically bound water molecules (Fig. 6B). In the [\(-\)] segment, a water molecule that mediates a phosphate oxygen contact from Tyr\(^{252}\) is also coordinated by Arg\(^{235}\), one of the key arginines involved in base-specific contacts in the core consensus. This train of water-mediated contacts extends through Glu\(^{228}\) and Thr\(^{226}\) in the core consensus to Asn\(^{221}\), Lys\(^{223}\), and Lys\(^{229}\) in the [\(+\)] flanking segment.

In the context of specific hydration, indirect readout of DNA conformation (2) relates sequence-dependent effects on helical twist, roll, and rise in the flanking segments to the energetics of establishing ordered water networks. Considering the tight packing within the core cavity (Fig. 6A), such networks must be highly constrained for a given flanking sequence variant. The required cooperativity is consistent with destabilization of the [\(-\)]GC complex by osmolytes being stronger than expected for a simple osmotic penalty alone. It also explains the strong enthalpy-entropy compensation observed among specific PU.1 ETS-DNA complexes (9). Because flanking sequence variants do exhibit a reasonable dispersion of affinities (2, 3), alternative formations are probably tolerated at the expense of binding free energy. The present data demonstrate that the low affinity site [\(+\)]TG, which appears to be water-neutral, does adopt the conformation characteristic of sequence-specific complexes. Assuming that [\(-\)]GC and [\(+\)]TG are the most “hydrated” and “unhydrated” binding sites, respectively, the effective energetic contribution of the structural waters would be reflected by their
affinities, or 13 kJ/mol of free energy under physiological conditions.

In addition to their role as structural waters, we propose that specific hydration plays a related role in controlling DNA curvature. Hypersensitivity to DNase I, a minor groove probe, is an invariant feature of sequence-specific ETS-DNA complexes (22–26). Because minor groove width is defined by residues $i$ and $i + 2$ in the complementary strand (38), the $[-]$ flanking bases in the TTCC strand ($5'$-GCT-3' in $[-]$GC, $5'$-TTT-3' in $[+]$TG) are directly involved in defining the minor groove width of the core consensus. Because effects on groove width are in turn coupled to curvature of the helical axis (8° in the co-crystal structure), the flanking contacts represent key positioning elements for optimizing contact by the recognition helix at the core consensus.

The physical driving force for curvature is provided by the asymmetric neutralization of backbone flanking phosphates (three in the $[-]$ segment, four in the $[+]$ segment of the co-crystal structure). Complete neutralization of these phosphates by chemical derivatization to methylphosphonates induces a 28° bend (39). We believe that water molecules involved in backbone phosphates in the flanking segments exaggerate the bend at the core consensus and enhances its accessibility to (larger) osmolytes. Because the core cavity in a high affinity complex is tightly packed (Fig. 6A), a larger, more distorted cavity would probably result in a suboptimal core interface that debilitates the formation of specific hydration networks bridging the recognition helix and core bases.

Experimentally, our salt dependence data are consistent with this proposal. All measured high affinity PU.1 ETS binding sites depend on $[Na^+]$ to an extent that substantially underestimates the number of phosphate contacts if they are assumed to be fully neutralized (9). The present data on the low affinity site $[+]$TG shows a pronounced dependence on $[Na^+]$, which closely tracks the number of phosphate contacts. Moreover, the polyelectrolyte theory predicts a purely entropic effect of counterion release upon DNA binding (40). An interactive effect with specific hydration would explain thermodynamic observations (9) that salt exerts an enthalpic effect on binding as well.

Based on the foregoing, our model predicts that a specific site of intermediate affinity for the PU.1 ETS domain should exhibit corresponding responses to osmolytes, DMS protection, and salt. To test our model, we used the $\lambda B$ site (AAAAGAAATG) whose affinity for PU.1 ETS is intermediate of the $[-]$GC and $[+]$TG sites (3). Application of osmotic stress destabilized binding with a slope of $-0.90 \pm 0.01$ (Fig. 7), about half the value for the $[-]$GC site (cf. Table 2). PU.1 ETS binding protects the core consensus from DMS attack but less than observed for the $[-]$GC site. Finally, the salt dependence of the $\lambda B$ site is $Z = 4.2 \pm 0.9$ (9), between that for $[-]$GC and $[+]$TG (Table 3). Thus, the linkage connecting specific hydration, formation of an osmolyte-excluding cavity, and electrostatics in sequence-specific PU.1-ETS interactions our model is supported by data from three different specific sites at least in rank order.

**Divergence of Sequence Discrimination of ETS Domains**—ETS domains share relatively weak sequence homology but strong structural conservation and similar target site preference (41). PU.1 is the most sequence-divergent member of the ETS family, a feature that may be correlated with its extensively hydrated protein-DNA interface. "Canonical" ETS-DNA complexes such as Ets-1 (11, 12) and GABPα (42) feature a higher density of direct protein-to-nucleobase contacts than PU.1. Specific hydration might therefore evolve as compensatory

![Figure 7: Testing the linkage between specific hydration and high affinity PU.1 ETS-DNA interactions.](image)

**TABLE 3** Salt dependence of DNA binding affinities of PU.1 ETS

| Site     | $S_{K_{obs}}$ (mM) | $Z$   |
|----------|--------------------|-------|
| $[-]$GC  | $-2.4 \pm 0.2$     | 2.7   |
| $[+]$TG  | $-7.5 \pm 1.1$     | 8.5   |
| NS       | $-4.6 \pm 0.5$     | 5.2   |

*Excludes data acquired below 0.1 mM Na*.
Specific Hydration in PU.1 ETS Binding

mechanism to replace some of the “missing” anhydrous contacts. The involvement of specifically bound water at the backbone contacts by PU.1 is also consistent with the substantially smaller DNA curvature (8°) compared with Ets-1 (27°) and GABPα (18°) despite sharing identical phosphate contacts (25). Finally, high affinity binding for the minimal ETS domain of Ets-1 exhibits dissociation constants at 10−12 M under physiological conditions (43). This 102-fold difference in affinity (cf. Table 1) reflects, in part, the intrinsic free energy cost of building structural waters in high affinity PU.1 complexes.

Nature of Nonspecific PU.1 ETS Complex—Nonspecific binding by the PU.1 ETS domain is clearly distinguished from sequence-specific binding by DNase I and hydroxyl radical footprinting. It is also associated with a net release of water molecules, in contrast with specific binding, and in a manner that depends on the osmolyte’s identity. Therefore, hydration is a major distinguishing feature in sequence discrimination by the PU.1 ETS domain. Our data suggest a nonspecific complex that is structurally defined but very different from that of sequence-specific ETS-DNA complexes.

Acknowledgment—We thank Crystal Van Dyken for technical assistance.

REFERENCES

1. Wei, G. H., Badis, G., Berger, M. F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A. R., Yan, J., Talukder, S., Turunen, M., Taipale, M., Stunnenberg, H. G., Ukkonen, E., Hughes, T. R., Bulyk, M. L., and Taipale, J. (2010) Genome-wide analysis of ETS family DNA binding in vitro and in vivo. EMBO J. 29, 2147–2160
2. Szymczyna, B. R., and Arrowsmith, C. H. (2000) DNA binding specificity studies of four ETS proteins support an indirect read-out mechanism of protein-DNA recognition. J. Biol. Chem. 275, 28363–28370
3. Poon, G. M., and Macgregor, R. B., Jr. (2003) Base coupling in sequence-specific ETS-DNA complexes. J. Biol. Chem. 278, 5469–5473
4. Poon, G. M., and Macgregor, R. B., Jr. (2004) A thermodynamic basis of DNA sequence selectivity by the ETS domain of murine PU.1. J. Mol. Biol. 338, 805–819
5. Li, S. L., Schlegel, W., Valente, A. J., and Clark, R. A. (1999) Critical flanking sequences of PU.1 binding sites in myeloid-specific promoters. J. Biol. Chem. 274, 32453–32460
6. Hollenhorst, P. C., Chandler, K. J., Poulsen, R. L., Johnson, W. E., Speck, N. A., and Graves, B. J. (2009) PLoS Genet. 5, e1000778
7. Hollenhorst, P. C., Jones, D. A., and Graves, B. J. (2004) Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. Nucleic Acids Res. 32, 5693–5702
8. Hagen, J. M. (2007) Genetic analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. Genes Dev. 21, 1882–1894
9. Kodandapani, R., Pio, F., Ni, C. Z., Piccialli, G., Klemets, M., McRae, K. R., and Ely, K. R. (1996) A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. Nature 380, 456–460
10. Poon, G. M., and Macgregor, R. B., Jr. (2004) A thermodynamic basis of DNA sequence selectivity by the ETS domain of murine PU.1. J. Mol. Biol. 335, 113–127
11. Pió, F., Assa-Munt, N., Yguerabide, J., and Maki, R. A. (1999) Protein Sci. 8, 2098–2109
12. Garvie, C. W., Hageman, J., and Wolberger, C. (2001) Structural studies of Ets-1/Pax5 complex formation on DNA. Mol. Cell 8, 1267–1276
13. Reddy, S. Y., Obika, S., and Brucie, T. C. (2003) Conformations and dynamics of Ets-1 ETS domain-DNA complexes. Proc. Natl. Acad. Sci. U.S.A. 100, 15475–15480
14. Obika, S., Reddy, S. Y., and Brucie, T. C. (2003) Sequence-specific DNA binding of Ets-1 transcription factor. Molecular dynamics study on the Ets domain–DNA complexes. J. Mol. Biol. 331, 345–359
15. Poon, G. M., Gross, P., and Macgregor, R. B., Jr. (2002) The sequence-specific association of the ETS domain of murine PU1 with DNA exhibits unusual energetics. Biochemistry 41, 2361–2371
16. Tullius, T. D., and Dombroski, B. A. (1986) Hydroxyl radical “footprinting.” High-resolution information about DNA-protein contacts and application to a repres sor and Cro protein. Proc. Natl. Acad. Sci. U.S.A. 83, 5469–5473
17. Scatchard, G., Hamer, W. J., and Wood, S. E. (1938) Isotonic solutions. I. The chemical potential of water in aqueous solutions of sodium chloride, potassium chloride, sulfuric acid, sucrose, urea, and glycerol at 25 °C. J. Am. Chem. Soc. 60, 3061–3070
18. Sidorova, N. Y., and Rau, D. C. (1996) Differences in water release for the binding of EcoRI to specific and nonspecific DNA sequences. Proc. Natl. Acad. Sci. U.S.A. 93, 12272–12277
19. Shadle, S. E., Allen, D. F., Guo, H., Pogozelski, W. K., Bashkin, J. S., and Tullius, T. D. (1997) Quantitative analysis of electrophoresis data. Novel curve fitting methodology and its application to the determination of a protein-DNA binding constant. Nucleic Acids Res. 25, 850–860
20. Petrek, M., Otyepka, M., Banas, P., Kosinová, P., Koca, J., and Damborsky, J. (2006) CAVER. A new tool to explore routes from protein clefts, pockets, and cavities. BMC Bioinformatics 7, 316
21. Eisenbeis, C. F., Singh, H., and Storb, U. (1993) PU.1 is a component of a multicomponent protein which binds an essential site in the murine immunoglobulin α 2–4 enhancer. Mol. Cell Biol. 13, 6462–6461
22. Gross, P., Arrowsmith, C. H., and Macgregor, R. B., Jr. (1998) Hydroxyl radical footprinting of DNA complexes of the ets domain of PU.1 and its comparison to the crystal structure. Biochemistry 37, 5129–5135
23. Rao, E., Dang, W., Tian, G., and Sen, R. (1997) A three-protein-DNA complex on a B cell-specific domain of the immunoglobulin μ heavy chain gene enhancer. J. Biol. Chem. 272, 6722–6732
24. Graves, B. J., Gillespie, M. E., and McIntosh, L. P. (1996) DNA binding by the ETS domain. Nature 384, 322
25. Gunther, C. V., and Graves, B. J. (1994) Identification of ETS domain proteins in murine T lymphocytes that interact with the Moloney murine leukemia virus enhancer. Mol. Cell Biol. 14, 7569–7580
26. Nye, J. A., Petersen, J. M., Gunther, C. V., Jonsen, M. D., and Graves, B. J. (1992) Interaction of murine ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. Genes Dev. 6, 975–990
27. Gross, P., Yee, A. A., Arrowsmith, C. H., and Macgregor, R. B., Jr. (1998) Quantitative hydroxyl radical footprinting reveals cooperative interactions between DNA-binding subdomains of PU.1 and IRF4. Biochemistry 37, 9802–9811
28. Balasubramanian, B., Pogozelski, W. K., and Tullius, T. D. (1998) DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone. Proc. Natl. Acad. Sci. U.S.A. 95, 9738–9743
29. Timasheff, S. N. (2002) Protein-scaffold preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. Proc. Natl. Acad. Sci. U.S.A. 99, 9721–9726
30. Harries, D., and Röös, J. (2006) A practical guide on how osmolytes modulate macromolecular properties. Methods Cell Biol. 84, 679–735
31. Sidorova, N. Y., Hung, S., and Rau, D. C. (2010) Stabilizing labile DNA-protein complexes in polyacrylamide gels. Electrophoresis 31, 648–653
32. 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. Q. Rev. Biophys. 11, 103–178
33. Jia, X., Lee, L. K., Light, J., Palmer, A. G., 3rd, and Assa-Munt, N. (1999) Backbone dynamics of a short PU.1 ETS domain. J. Mol. Biol. 292, 1083–1093
34. Arakawa, T., and Timasheff, S. N. (1982) Stabilization of protein structure by sugars. Biochemistry 21, 6536–6544
35. Arakawa, T., and Timasheff, S. N. (1983) Preferential interactions of proteins with solvent components in aqueous amino acid solutions. Arch.
36. Courtenay, E. S., Capp, M. W., Anderson, C. F., and Record, M. T. (2000) Vapor pressure osmometry studies of osmolyte-protein interactions. Implications for the action of osmoprotectants in vivo and for the interpretation of “osmotic stress” experiments in vitro. Biochemistry 39, 4455–4471

37. Holdgate, G. A., Tunnicliffe, A., Ward, W. H., Weston, S. A., Rosenbrock, G., Barth, P. T., Taylor, I. W., Paupit, R. A., and Timms, D. (1997) The entropic penalty of ordered water accounts for weaker binding of the antibiotic novobiocin to a resistant mutant of DNA gyrase. A thermodynamic and crystallographic study. Biochemistry 36, 9663–9673

38. Burkhoff, A. M., and Tullius, T. D. (1987) The unusual conformation adopted by the adenine tracts in kinetoplast DNA. Cell 48, 935–943

39. Strauss-Soukup, J. K., and Maher, L. J., 3rd. (1997) Role of asymmetric phosphate neutralization in DNA bending by PU.1. J. Biol. Chem. 272, 31570–31575

40. Record, M. T., Jr., deHaseth, P. L., and Lohman, T. M. (1977) Interpretation of monovalent and divalent cation effects on the lac repressor-operator interaction. Biochemistry 16, 4791–4796

41. Pio, F., Kodandapani, R., Ni, C. Z., Shepard, W., Klemsz, M., McKercher, S. R., Maki, R. A., and Ely, K. R. (1996) New insights on DNA recognition by Ets proteins from the crystal structure of the PU.1 ETS domain-DNA complex. J. Biol. Chem. 271, 23329–23337

42. Batchelor, A. H., Piper, D. E., de la Brousse, F. C., McKnight, S. L., and Wolberger, C. (1998) The structure of GABPα/β. An ETS domain-ankyrin repeat heterodimer bound to DNA. Science 279, 1037–1041

43. Jonsen, M. D., Petersen, J. M., Xu, Q. P., and Graves, B. J. (1996) Characterization of the cooperative function of inhibitory sequences in Ets-1. Mol. Cell Biol. 16, 2065–2073