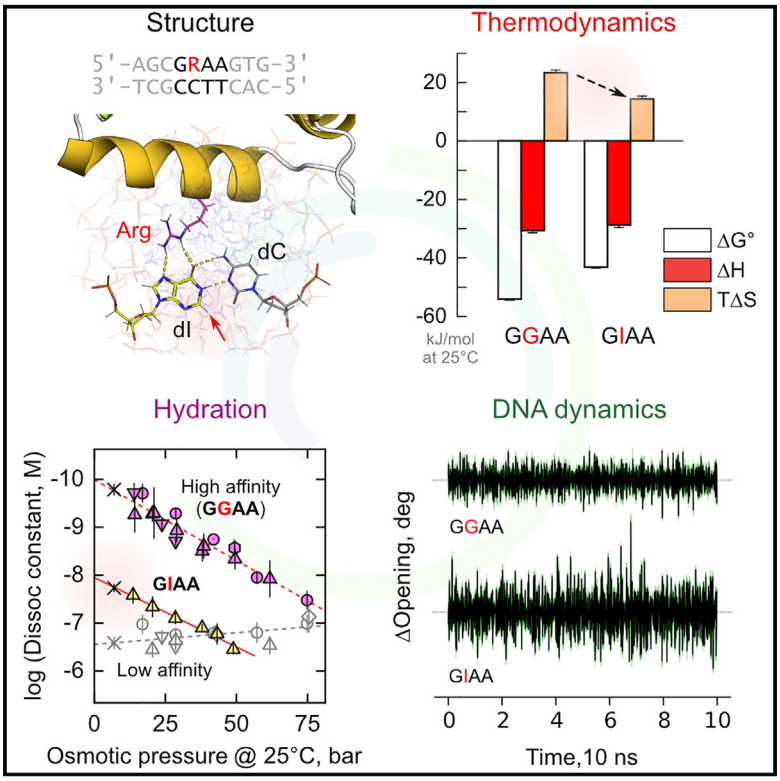
Article

Structure

Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors

Graphical abstract



Highlights

- Inosine substitution in an Arg×GC motif isolates dynamic contributions to binding
- Arg×IC impacts affinity via altered DNA dynamics in the unbound state
- Application to osmotically active PU.1 elucidates the role of hydration in specificity
- Renewed potential of modified bases as targeted structural thermodynamics probes

Authors

Tyler N. Vernon, J. Ross Terrell, Amanda V. Albrecht, Markus W. Germann, W. David Wilson, Gregory M.K. Poon

Correspondence

mwg@gsu.edu (M.W.G.), wdw@gsu.edu (W.D.W.), gpoon@gsu.edu (G.M.K.P.)

In brief

Vernon et al. combined crystallography, solution biophysics, and molecular dynamics (MD) simulations to dissect the structural thermodynamics of the arginine sidechain-guanine contact, a ubiquitous motif among protein/DNA complexes. Analysis of this contact in the PU.1/DNA complex reveals the structural basis of thermodynamic hydration in DNA specificity within the ETS transcription factor family.

Vernon et al., 2024, Structure *32*, 1–14
January 4, 2024
https://doi.org/10.1016/j.str.2023.11.003

CellPress

Please cite this article in press as: Vernon et al., Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors, Structure (2023), https://doi.org/10.1016/j.str.2023.11.003

Structure

CellPress

Article

Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors

Tyler N. Vernon,[1,4] J. Ross Terrell,[1,4] Amanda V. Albrecht,[1] Markus W. Germann,[1,2,*] W. David Wilson,[1,3,*] and Gregory M.K. Poon[1,3,5,*]

[1]Department of Chemistry, Georgia State University, Atlanta, GA 30302, USA
[2]Department of Biology, Georgia State University, Atlanta, GA 30302, USA
[3]Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA 30302, USA
[4]These authors contributed equally
[5]Lead contact
*Correspondence: mwg@gsu.edu (M.W.G.), wdw@gsu.edu (W.D.W.), gpoon@gsu.edu (G.M.K.P.)
https://doi.org/10.1016/j.str.2023.11.003

SUMMARY

Nucleobases such as inosine have been extensively utilized to map direct contacts by proteins in the DNA groove. Their deployment as targeted probes of dynamics and hydration, which are dominant thermodynamic drivers of affinity and specificity, has been limited by a paucity of suitable experimental models. We report a joint crystallographic, thermodynamic, and computational study of the bidentate complex of the arginine side chain with a Watson-Crick guanine (Arg×GC), a highly specific configuration adopted by major transcription factors throughout the eukaryotic branches in the Tree of Life. Using the ETS-family factor PU.1 as a high-resolution structural framework, inosine substitution for guanine resulted in a sharp dissection of conformational dynamics and hydration and elucidated their role in the DNA specificity of PU.1. Our work suggests an under-exploited utility of modified nucleobases in untangling the structural thermodynamics of interactions, such as the Arg×GC motif, where direct and indirect readout are tightly integrated.

INTRODUCTION

It has been recognized for nearly a half-century that the grooves of the double helix present an ordered array of chemical groups whose physicochemical characteristics toward H-bonding and London dispersion interactions encode determinants for sequence-specific binding.[1] Numerous studies have mapped critical contacts in DNA-ligand complexes by adding, removing, or altering atomic groups along the groove floor. The advent of high-throughput experimental techniques has spurred the development of computational frameworks, particularly in machine learning, to infer binding specificities from massive binding datasets.[2–7] In addition to the direct readout of the bases, recent predictive models have incorporated minor groove width and other structural features,[8,9] electrostatic potential,[10] and spatial relationship of protein/DNA contacts[11] into their algorithms. These refinements reflect the recognition that specificity also receives major contributions from other sequence-dependent properties of the double-stranded DNA, such as flexibility and helical geometry, collectively referred to as indirect readout.[12]

Experimental studies of direct readout are facilitated by a palette of non-standard nucleobases, which may be incorporated into DNA enzymatically or synthetically. Since the chemical substituents in nucleobases strongly affect base-pairing and stacking interactions, as well as contacts with ligands, modified nucleobases also serve as useful probes of indirect readout. For example, inosine and diaminopurine (DAP) differ from guanine and adenine only in the absence or addition of an exocyclic 2-NH_2, respectively, in the minor groove. In complexes contacting the DNA major groove, inosine and DAP conserve Watson-Crick (WC) direct readout for guanine and adenine, respectively, while perturbing the intrinsic curvature, groove widths, conformational mechanics, propensity to bend, and hydration properties.[13–15] For minor groove binders, inosine and DAP fulfill the analogous roles for adenine and guanine.[16] In either case, the base substitutions are invoked to rationalize altered binding preferences[17,18] and activities of DNA-modifying enzymes such as restriction endonucleases[19,20] or in DNA repair.[21]

The formal definition of direct versus indirect readout admits a diverse manifestation in nucleoprotein structures, as well as degrees of overlap between the two modes. One type of indirect readout considers whether the DNA is contacted or not by the reading protein. Non-contacted indirect readout is prominent among examples from bacteria and bacteriophages (such as Fis protein,[22] P22 c2 repressor,[23,24] 434 repressor[25]). These proteins form homo-dimeric complexes with dyadic or pseudo-dyadic sequences in which non-contacted central positions are sandwiched between two segments of direct readout.

Please cite this article in press as: Vernon et al., Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors, Structure (2023), https://doi.org/10.1016/j.str.2023.11.003

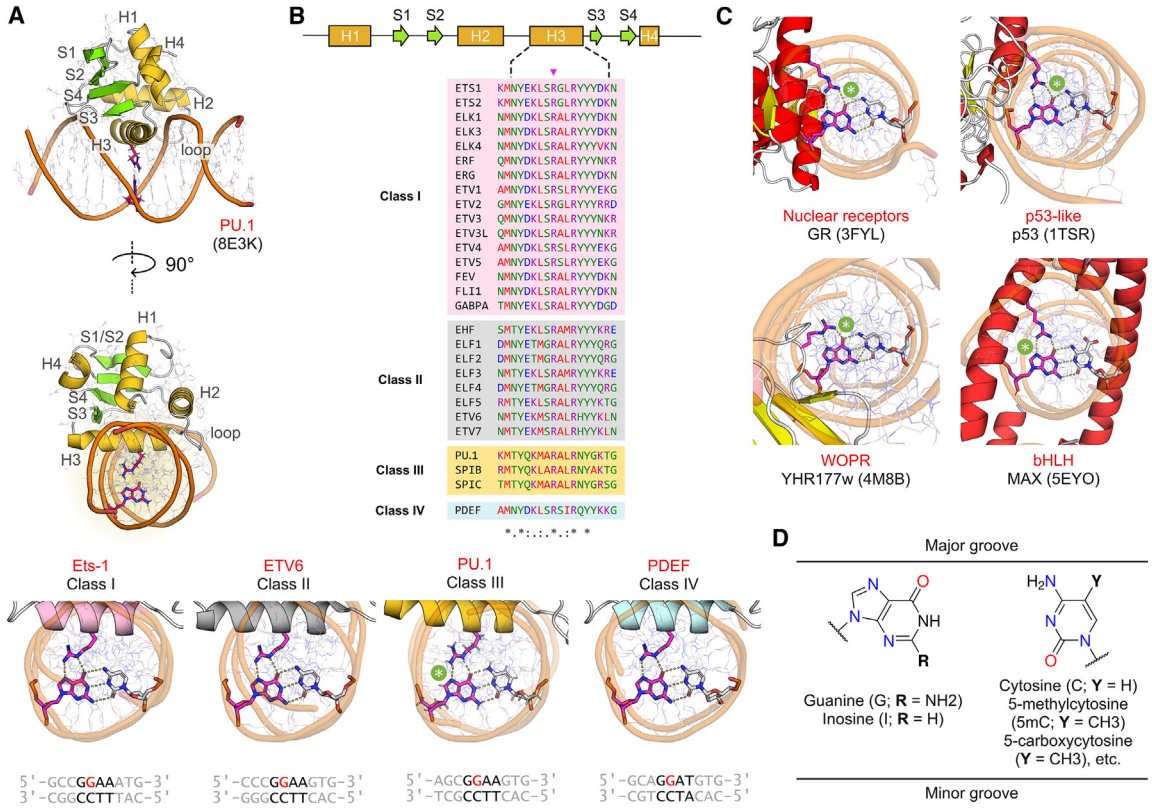


Figure 1. The co-crystal structure of the transcription factor PU.1 presents a broadly representative Arg-guanine interaction motif

(A) The high-affinity co-crystal structure of human PU.1 (PDB: 8E3K).

(B) ETS domains bind DNA motifs harboring a central GGAW core, which formally divide the superfamily into four classes as exemplified by the twenty-eight ETS factors in humans. An absolutely conserved Arg residue (R230 in PU.1, marked with triangle) in the recognition helix H3 of ETS domains mediates an invariant bidentate readout of a core guanine in the DNA major groove. This interaction is labeled with an asterisk.

(C) Conservation of this Arg-guanine interaction motif in selected non-ETS classes of DBDs.

(D) The absence of direct contact on the minor groove aspect permits the chemical and dynamic contributions to DNA readout to be dissected by targeted substitution of the Arg×GC guanine by non-canonical bases.

Among these examples, experimental structures of Fis/DNA complexes harboring inosine and DAP substitutions with matched unmodified controls reveal strong effects of local flexibility and curvature of the non-contacted DNA on complex structure.[26,27] Alternatively, contacted indirect readout may occur through contacts with the compositionally fixed DNA backbone. A well-known example is the trp repressor, which forms a dimeric complex that makes few, if any, significant direct base contacts over the entire operator sequence.[28] Contacted indirect readout may also occur locally, as exhibited for example by the CAP-DNA complex (another homodimer-dyadic DNA structure) in which a 90° kink is mediated by aspartate-phosphate contacts over two consecutive base pairs.[29]

A third, more subtle and not nearly as well-understood mechanism of indirect readout involves the integration of both direct and indirect readout. In general, sequence-specific binding reflects the recognition of intrinsic DNA curvature and induction of protein-preferred geometry in a dynamic environment.[30] Most instances of direct readout are therefore expected to be integrated with local indirect readout by the protein as well. A well-characterized example of integrated readout is the TATA-binding protein (TBP), a minor groove binder. TBP bends DNA by ~90° at either end of a canonical 5′-TATAAAAG-3′ sequence that is contacted by direct readout.[31] Fluorescence energy transfer experiments have shown that the bending angle is DNA sequence-dependent.[32] However, complete substitution of A:T base pairs to I:C to 5′-CICIIIIG-3′ resulted in only a modest change in binding affinity that was attributed to lower hydration in the I:C minor groove.[18]

While TBP poses a striking structure, and is essential in transcription pre-initiation, it does not represent the pervasive form of integrated readout in the nucleoprotein repertoire. The vast majority of transcription factors, by individual count or by class, engage DNA primarily in the major groove.[33] Moreover, there are no pair-matched experimental models in this category (as with Fis for non-contacted indirect readout) for direct structural comparison. To this end, we recently reported[34] the co-crystallization and structural determination of sequence-specific complexes formed by the ETS-family transcription factor PU.1, to resolutions (down to ~1.2 Å) that significantly improve on the reported structures for this family [Figure 1A]. This system is robust to a variety of mutations in the protein as well as the DNA, including modified nucleobases, and crystallizes to an identical crystal form that enables high-resolution comparison. In addition to the favorable crystallographic characteristics, the PU.1/DNA

Please cite this article in press as: Vernon et al., Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors, Structure (2023), https://doi.org/10.1016/j.str.2023.11.003

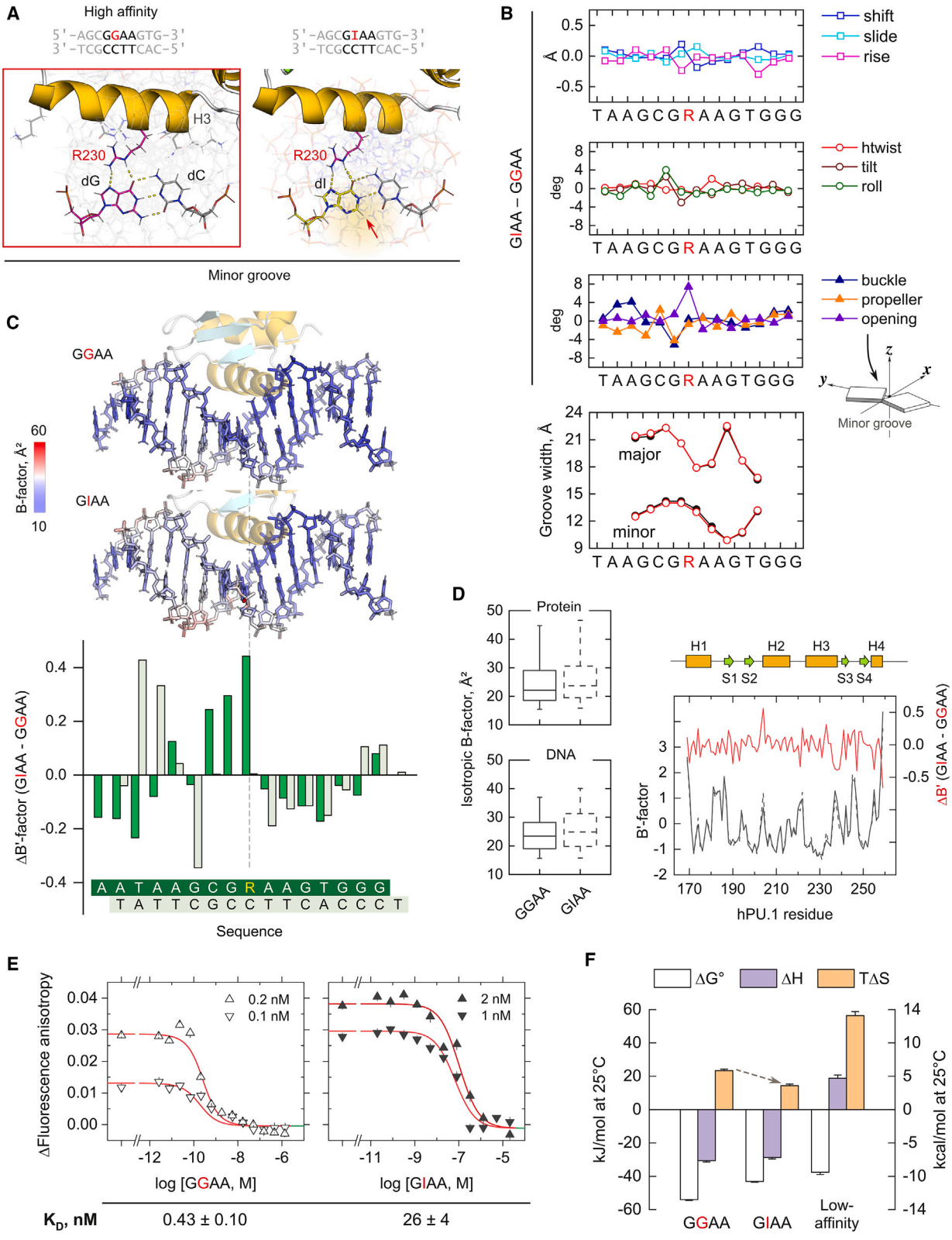


Figure 2. Structural and thermodynamic characterization of a dynamic mutant PU.1/DNA complex

The ETS domain of PU.1 was co-crystallized with a high-affinity sequence or an inosine-substituted variant.

(A) The Arg×GC motif is unaffected by inosine, which base pairs with cytosine in the same orientation as guanine.

(legend continued on next page)

Please cite this article in press as: Vernon et al., Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors, Structure (2023), https://doi.org/10.1016/j.str.2023.11.003

complex provides a high-resolution exemplar of a bidentate H-bond with guanine via a signature Arg residue of the ETS family [Figure 1B]. This conserved motif, which has been proposed as a highly specific interaction for transcription factors,[11,35] is found beyond the ETS family in many other structurally distinct DNA-binding domains (DBDs). The contacting Arg side chain emanates commonly from a groove-inserted α-helix, such as in ETS, nuclear receptors, p53-like domains, and basic helix-loop-helix (bHLH) structures but may also be provided by a loop as in the WOPR domain [Figure 1C].[36] These and other classes of DBDs span all of Eukarya, including plants (e.g., bHLH) and fungi (e.g., WOPR). Moreover, this motif is subject to epigenetic modification at the base-paired cytosine [Figure 1D], such as seen in the MAX/DNA complex (Figure 1C). For these reasons, Arg×GC is a compelling model for high-resolution characterization of DNA readout integration by inosine substitution.

In this article, we report an integrated crystallographic, thermodynamic, and computational study of PU.1/DNA complexes with pair-matched protein mutations and chemically modified DNA, including an inosine substitution for guanine in the core Arg×GC motif. The data reveal at high resolution the significant role of local dynamics in direct readout and strike sharp contrasts with the structural effects of inosine substitution in non-contacted indirect readout, as exhibited by the Fis protein, and integrated readout as exhibited by minor groove-binding TBP. Here, inosine substitution induced a decoupling between the conformational dynamics and hydration properties of PU.1, an osmotically sensitive transcription factor, and illuminated the structural thermodynamics of its target selectivity. The data provide new insight into the integration of direct and indirect readout and suggest an under-exploited utility for modified nucleobases as targeted probes in protein/DNA recognition.

RESULTS

Deoxyinosine substitution of the Arg×GC motif generates a dynamically perturbed ETS/DNA complex

Replacement of the 5′-GGAA-3′ core in an optimal PU.1-binding sequence with 5′-GIAA-3′ led to a highly isomorphous co-crystal (PDB: 8T9U) that diffracted to comparable resolution (1.47 Å) and refinement statistics as the unmodified (1.28 Å; PDB: 8E3K[34]) high-affinity complex [Figure 2A and Table 1]. Crystal contacts, consisting primarily of end-to-end DNA/DNA and secondary protein/DNA contacts, were essentially identical in the two complexes. To detect conformational changes between the two complexes, we evaluate their isomorphous difference (Fo-Fo) map,[37] which combines differences in the observed structural factor amplitudes with a single set of phases derived from the wild-type structure. Difference densities were scattered at low intensity and significant only near the substituted guanine [Figure S1]. Thus, substitution with inosine conserved the global structure and direct readout of the PU.1/DNA complex. Crucially, the potential for crystal contacts to influence DNA conformation[38–40] did not obscure the local conformational perturbations at the Arg×GC motif.

Examining the local helical properties of the PU.1-bound sequences more closely, the dI substitution did not bias base step parameters beyond 0.3 Å for distance-based parameters and 4° for angular parameters [Figure 2B]. Base-pair level parameters exhibited more significant perturbations. An ~8° increase in base opening (widening of a base pair) into the major groove at the dI:dC base pair was accompanied by negative buckle and propeller in the preceding dG:dC pair, as well as additional perturbations in the upstream 5′-flanking positions. Nevertheless, major and minor groove widths were closely preserved (Figure 2B), strongly suggesting that the base-pair level perturbations reflected the protein's preference for the trajectory of the bound DNA. These observations were attested by the atomic B-factors [Figure 2C]. While the absolute B-factors spanned similar ranges between the two complexes, between 10 and 60 Å^2, z-normalized B′-factors revealed a biased redistribution among the same 5′-flanking positions that exhibited pair-level perturbations. For both protein and DNA, absolute B-factors, both high and low, increased to larger values for the protein in the 5′-GIAA-3′ complex due to the difference in resolution between the two structures [Figure 2D]. However, the distribution of protein B-factors was essentially identical in width and without significant difference in per-residue B′-factors. In contrast, B-factors for the DNA were biased toward larger values. The bidentate Arg×IC contact thus conferred increased local dynamics in the 5′ half of the DNA binding site.

To relate these structural observations to binding properties in solution, we measured the affinity of PU.1 binding to DNA harboring the two sequences. Under physiologic ionic (0.15 M Na^+) conditions, the 5′-GIAA-3′ site was bound ~80-fold more weakly than the unmodified sequence [Figure 2E and Table S1]. Thus, a *more* flexible DNA sequence was *less* favorable for complex formation by > 10 kJ/mol at 25°C despite similar helical deformations in the bound structure. To resolve the thermodynamic basis of this difference in affinity, we measured the calorimetric enthalpy changes in forming the two complexes, which were within experimental uncertainty ($\Delta\Delta H$ < 2 kJ/mol at 25°C) [Figures S2 and 2F]. For comparison, binding to a low-affinity site (5′-AAAGGAATGG-3′; K_D = 259 ± 26 μM) is qualitatively differentiated by the sign (positive) of the associated ΔH. The large decrease in free energy of binding for the 5′-GIAA-3′ site was therefore not due to a deficit in favorable interactions, as consistent with the absence of major differences in structure and protein/DNA contacts in the co-crystal structures but rather a *loss* of favorable entropy changes.

(B) Differential base pair and base step parameters and helical groove widths along the protein-bound DNA.
(C) B-factors of the bound DNA mapped to the modeled structures. Per-residue ΔB′-factors for heavy atoms show the dynamics redistribution in each strand.
(D) B-factors of the protein and DNA; box-and-whiskers represent the median ±25/75th and 5/95th percentiles, respectively. Per-residue values of the protein were compared in terms of their z-normalized B′-factors.
(E) Competitive binding assays of unlabeled DNA encoding the two sequences for probe-bound PU.1 (at the indicated concentrations). Error bars and absolute dissociation constants from triplicate experiments were shown.
(F) Thermodynamic binding profiles at 25°C with enthalpy changes determined by ITC. Error bars are fitting errors for ΔH and propagated errors for $\Delta G°$ and $T\Delta S$. The inosine-substituted complex contrasts qualitatively in enthalpy with binding to low-affinity DNA.

Please cite this article in press as: Vernon et al., Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors, Structure (2023), https://doi.org/10.1016/j.str.2023.11.003

Table 1. Refinement statistics of co-crystallographic models

Protein	WT ΔN165	S3 chimera	S3 chimera	S3 chimera
DNA	GIAA	High-affinity	GIAA	Low-affinity
PDB ID	8T9U	8SMH	8SP1	8SMJ
Wavelength	0.99999	1	1	0.9201
Resolution range	33.26–1.47 (1.523–1.47)	33.43–1.37 (1.419–1.37)	33.39–1.62 (1.678–1.62)	24.26–1.39 (1.44–1.39)
Space group	P 1 21 1	P 1 21 1	P 1 21 1	P 1 21 1
Unit cell	43.021 60.622 44.562 90 116.794 90	42.944 60.71 45.066 90 117.279 90	42.824 60.725 44.973 90 117.272 90	42.767 61.193 44.774 90 117.283 90
Total reflections	178445 (16717)	140715 (13452)	86966 (8512)	215184 (14221)
Unique reflections	34426 (3415)	42457 (4177)	25881 (2541)	41218 (2998)
Multiplicity	5.2 (4.9)	3.3 (3.2)	3.4 (3.3)	5.2 (4.7)
Completeness (%)	98.64 (98.13)	98.01 (96.77)	98.86 (97.25)	99.70 (98.00)
Mean I/sigma(I)	14.56 (3.37)	16.94 (2.24)	14.63 (2.92)	12.30 (2.10)
Wilson B-factor	19.61	19.56	20.73	15.42
R-merge	0.0626 (0.3442)	0.0383 (0.4507)	0.04781 (0.4136)	0.0580 (0.5660)
CC1/2	0.996 (0.956)	0.998 (0.892)	0.998 (0.958)	0.998 (0.828)
Reflections used in refinement	34360 (3406)	42457 (4168)	25880 (2515)	41186 (4095)
Reflections used for R-free	2026 (187)	1946 (192)	2018 (190)	1991 (201)
R-work	0.1448 (0.1877)	0.1432 (0.3694)	0.1659 (0.2990)	0.1416 (0.2437)
R-free	0.1810 (0.2437)	0.1772 (0.3862)	0.1993 (0.3358)	0.1540 (0.2434)
CC(work)	0.977 (0.965)	0.974 (0.936)	0.971 (0.903)	0.974 (0.904)
CC(free)	0.958 (0.905)	0.961 (0.861)	0.964 (0.824)	0.971 (0.909)
Number of non-hydrogen atoms	1643	1675	1665	1678
macromolecules	1377	1407	1385	1399
ligands	32	1	32	0
solvent	244	267	258	279
Protein residues	91	92	92	92
RMS(bonds)	0.01	0.008	0.011	0.009
RMS(angles)	1.22	1.08	1.38	1.14
Ramachandran favored (%)	97.75	98.89	98.89	98.89
Ramachandran allowed (%)	2.25	1.11	1.11	1.11
Ramachandran outliers (%)	0	0	0	0
Rotamer outliers (%)	0	0	1.27	1.28
Clashscore	1.58	0.78	0	0.79
Average B-factor	28.03	27.9	25.92	22.14
macromolecules	25.38	25.62	24.48	20.04
Ligands	27.43	45.4	26.29	–
solvent	38.98	39.67	35.46	33.55
Beamline	APS 22-ID	ALS 5.0.2	ALS 5.0.2	NSLS-II AMX 17-ID-1
Detector	Dectris Eiger 16M	Dectris Pilatus 6M	Dectris Pilatus 6M	Dectris Eiger 9M
Oscillation Angle	0.25	0.25	0.25	0.2
Frames Collected	1080	720	720	1350

Inosine substitution raises the local conformational costs of protein binding

In general, the major contributors to entropy changes in binding are configurational entropy and the disposition of solvent molecules and ions.[41] To evaluate these contributions to the *decrease* in favorable entropy associated with binding the inosine-substituted high-affinity site, we first considered the differential changes in conformational degrees of freedom upon forming the two complexes. We performed molecular dynamics (MD) simulations of both DNA sequences alone and in

Please cite this article in press as: Vernon et al., Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors, Structure (2023), https://doi.org/10.1016/j.str.2023.11.003

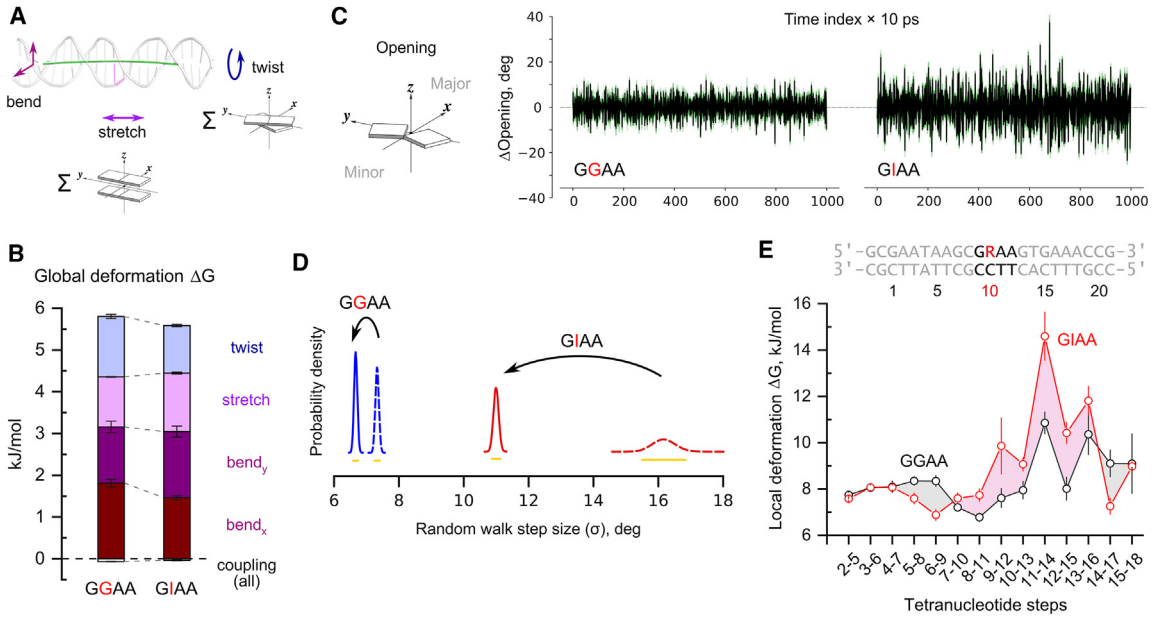


Figure 3. Helical dynamics are differentially quenched in inosine-substituted DNA upon PU.1 binding

The molecular dynamics of free and PU.1-bound DNA ns timescale were simulated in explicit TIP3P water.

(A) Global elastic model of DNA deformation. The four degrees of freedom are bending of a fitted helical axis in generalized *x* and *y* directions, stretching along the contour length of the helix, and twisting.

(B) Change in global deformation free energy upon PU.1 binding. Error bars are computed by block averaging.

(C) Time series analysis of base pair opening at the substituted guanine. The black series show representative 10-ns trajectories of differenced base pair opening in free GGAA and GIAA from the simulations and the green series are every 1,000th random walk in the Markov chain Monte Carlo (MCMC) sampling.

(D) Bayesian estimates of the Gaussian steps in opening for all four states. Dashed and solid distributions represent the free and bound DNA, respectively. The orange bars represent the 95% highest-density intervals.

(E), Local deformation energy computed from the fluctuations over a 4-bp moving window. Error bars are computed by block averaging. The terminal three bp are excluded.

complex with PU.1. Following equilibration at 298 K and 1 bar, the unrestrained dynamics of the free DNA and complexes were rapidly convergent in production, as judged by their root-mean-square (RMS) deviations, by 100 ns and 1.0 μs, respectively [Figure S3A].

In the co-crystal structures, the atomic B-factors (Figure 2D) show that dynamic differences in the two PU.1/DNA complexes occur primarily in the DNA rather than the protein. The simulated dynamics were consistent with this feature as the RMS fluctuations of the DNA-bound proteins showed no systematic differences among the resolved residues in the co-crystal structures [Figure S3B]. We therefore focused on the perturbations in DNA dynamics due to PU.1 binding. We first considered a simple elastic model of global DNA deformation, characterized by bending (in two orthogonal planes), stretching of the contour length (total helical rise), and twisting (total helical twist) [Figure 3A]. Elastic deformation was modeled, using x3_dna software[42] by inversion of a covariance matrix describing the motional fluctuations in the four global degrees of freedom.[43–45] The elastic matrices revealed that both PU.1 binding and inosine substitution primarily perturb the pairwise coupling among the four global degrees of freedom [Figure S4], and mutually offset to negligibly affect the global deformation free energy changes upon binding [Figure 3B]. This result is consistent with the co-crystal structures in that the global (time-averaged) trajectories of the bound DNA, as judged by the groove widths, are highly similar. However, the conservation in groove width also masks significant differences in local helical parameters, thus prompting us to examine local base pair-level dynamics in the DNA.

Structurally, opening is the most sensitive parameter of inosine substitution in a GC base pair (Figure 1B). For this reason, we focused on the opening dynamics at the substituted position of GGAA and GIAA in both free and PU.1-bound states. The opening angles at the IC base pair in free GIAA exhibited significantly greater fluctuations than its GGAA counterpart [Figure 3C]. To model these fluctuations, we treated them as a Gaussian random walk, $dx_t = \sigma dW_t$, where σ is the step size of the standard Gaussian walk i.e., a Wiener process of zero mean and unit variance, providing a direct measure of the fluctuations in *x* as a time-dependent variable. The fitted estimates of σ showed that the fluctuations in opening at the inosine-substituted position for both bound DNA were reduced relative to their free states. The PU.1-bound GIAA site was more dynamic than bound GGAA, in accord with the B-factor analysis, but was eclipsed by the profound *reduction* in σ from the free state for the GIAA site [Figure 3D]. Inosine substitution thus imparted strong local dynamics which were quenched to excess upon protein binding. Fluctuations of a second local parameter, stretch at the same position [Figure S5], were less sensitive to the effects of inosine substitution but showed the same rank order in σ as opening. Thus,

Please cite this article in press as: Vernon et al., Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors, Structure (2023), https://doi.org/10.1016/j.str.2023.11.003

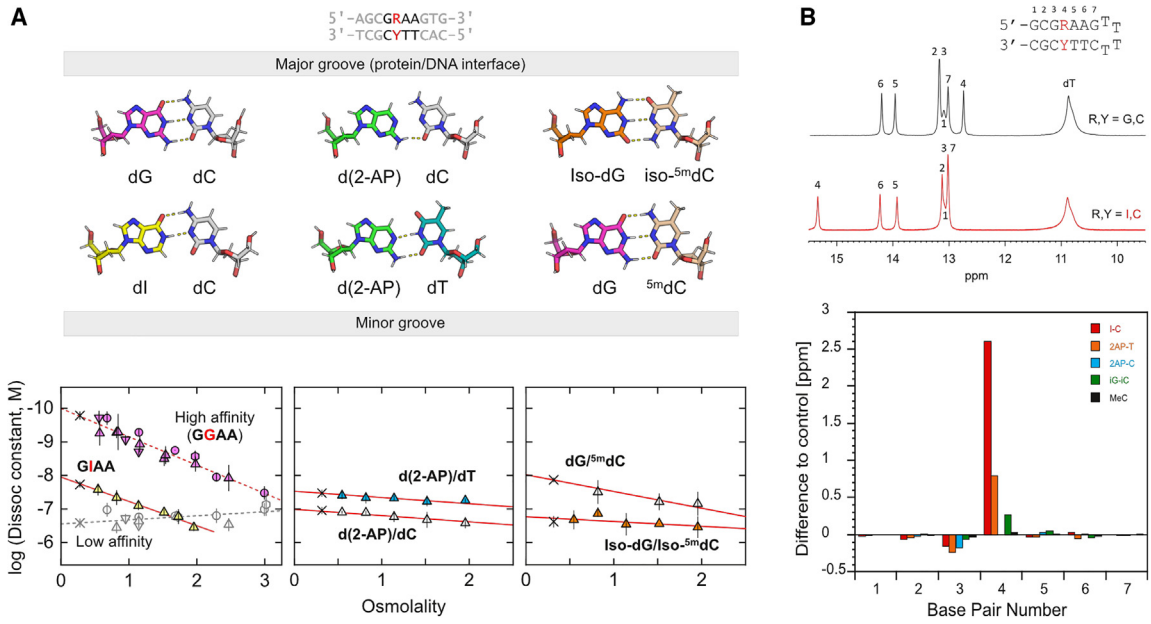


Figure 4. Hydration properties distinguish chemical and dynamic DNA readout by PU.1

(A) Structures and nominal base-pairing of DNA targeting the signature direct contact by ETS proteins are shown. Solution binding affinities of PU.1 ETS domain to various DNA harboring these substitutions in the high-affinity site were determined as a function of osmotic pressure. Data previously reported for high- and low-affinity binding (gray) has established that chemically diverse osmolytes perturb binding in a strictly colligative manner within experimental uncertainty. Symbols represent the average ±S.D. of three or more experiments: ×, no added osmolytes; glycine betaine (up triangle); nicotinamide (down triangle); sucrose (hexagon); triethylene glycol (circle). Given this, betaine was used in the present experiments involving non-canonical nucleobases. Lines represent the linear best fit of the data for each DNA. (B) Differential imino-^{1}H chemical shifts of model DNA hairpins mimicking the core consensus of ETS-binding motifs at 288 K. Spectra for unmodified (control) and inosine-substituted sequences are shown. Spectra for the other substituted DNA are shown in Figure S6.

the *local* deformation energy may be significantly greater for GIAA than for the native counterpart.

By convention, the generalized coordinates for local DNA deformation consist of the six dimer base step parameters: shift, slide, rise, tilt, roll, and twist.[43] Applying the same matrix inversion procedure, we computed (using x3_dna) the local deformation free energy change for these six local degrees of freedom in a four-bp moving window along the internal DNA positions [Figure 3E]. The results show a sharp spike in local deformation energy change for binding to the GIAA site, centered around the substituted core position. The ps-ns timescale dynamics thus supports the notion that inosine substitution represents a significant *negative* contributor to the overall entropy change via an excess quench in local conformational dynamics of the free DNA.

Chemical and dynamic perturbations of DNA readout are distinguished by their hydration effects

In addition to dynamic perturbations, the entropic effects of inosine substitution may also contain hydration contributions. Wild-type PU.1 binding to DNA is uniquely coupled among ETS factors with significant hydration changes in a sequence-specific manner.[46] High-affinity binding is destabilized by a variety of chemically disparate low-MW cosolutes according only to their osmolality, viz-à-viz a colligative property, within experimental uncertainty. In contrast, low-affinity binding is weakly sensitive to osmotic pressure. Compared with high-affinity binding, the osmotic dependence was modestly reduced for the GIAA sequence by about ~25% [Figure 4A]. As the implied reduction in hydration upon binding would predict a more positive entropy change for GIAA, contrary to observation (Figure 2F), we conclude that inosine substitution did not exert its net thermodynamic effects through hydration.

Since inosine substitution fully preserved major groove interactions at the Arg×GC motif, the osmotic stress data suggested that the hydration properties of the complex resided primarily in the protein/DNA interface. To test this idea, we targeted the guanine with other modified nucleobases (Figure 4A). Replacement with 2-aminopurine (2-AP), aimed at disrupting the bidentate H-bond with R230 (Figure 1A), reduced affinity by three orders of magnitude and essentially abolished osmotically sensitive binding, similar to low-affinity sites generated by sequence changes outside the core consensus. To control for the additional destabilization due to wobble pairing of 2-AP with cytosine, we also replaced the complementary base with thymine, which maintains WC geometry with 2-AP.[47] The WC-paired 2-AP showed marginally stronger affinity but did not restore osmotic sensitivity. Calorimetric measurement showed a positive ΔH = +4.5 ± 0.3 kJ/mol at 25°C (Figure S2B), in sharp contrast with GGAA and GIAA binding. Direct readout of the 2-carbonyl by R230 was therefore critical to high-affinity, osmotically sensitive binding.

To evaluate the H-bonding *specificity* of the Arg×GC motif, we replaced guanine with isoguanine (iso-dG), in which the carbonyl and the exocyclic NH_2 are reversed in position. Paired with

Please cite this article in press as: Vernon et al., Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors, Structure (2023), https://doi.org/10.1016/j.str.2023.11.003

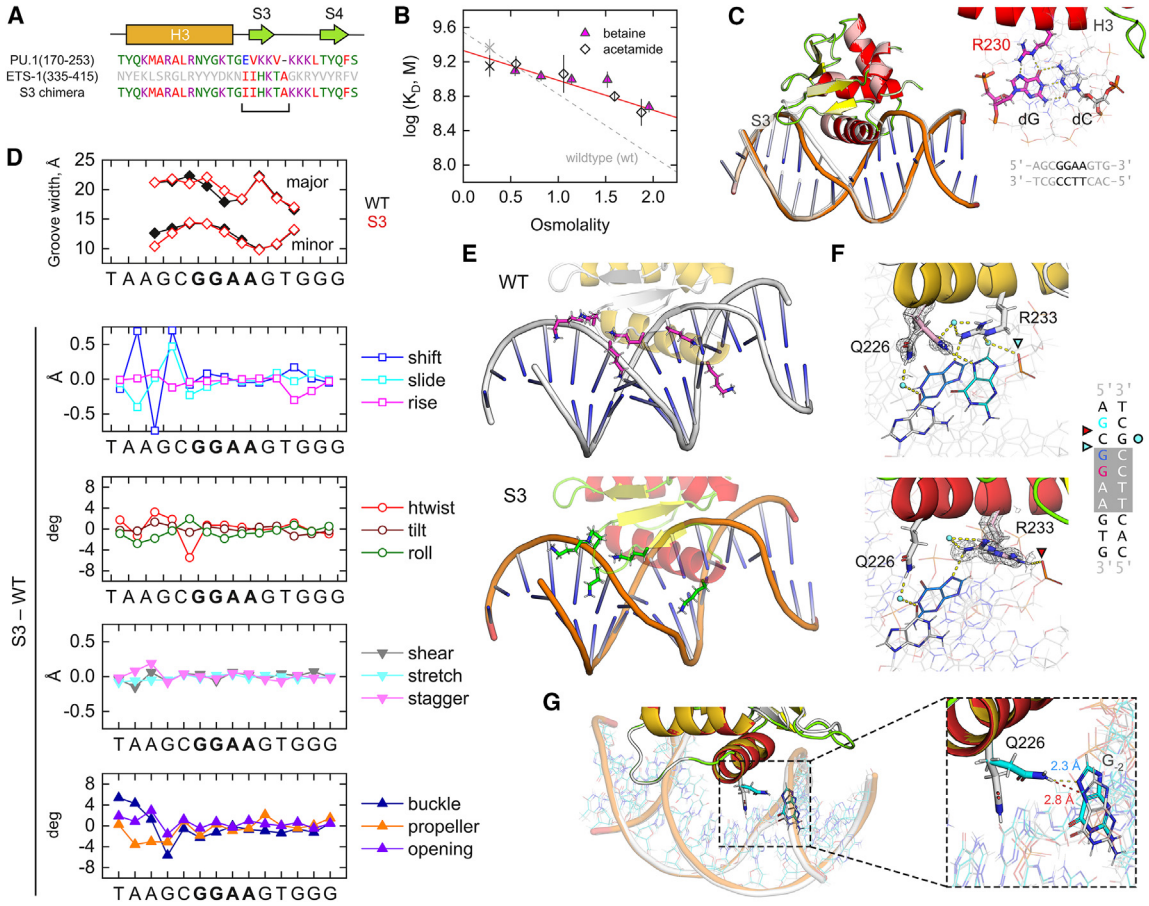


Figure 5. Structural characterization of a hydration mutant of the PU.1/DNA complex

(A) Exchange of the residues encoding S3 in PU.1 and Ets-1, the latter an osmotically insensitive ETS-family member, generates a chimeric construct termed S3 in shorthand.

(B) The reduced osmotic sensitivity in high-affinity binding of the S3 chimera, previously shown with betaine and re-verified here with acetamide, relative to wild-type PU.1. Data points represent the average ±S.D. of three or more experiments.

(C) Co-crystal high-affinity structure of S3 (colored), overlaid with the wild-type complex. The Arg×GC motif is shown in inset. Note the differences in DNA distortion near the S3 element.

(D) Differential groove widths, base pair, and base step parameters along the bound DNA.

(E) Comparison of the electrostatic contacts in the protein-DNA interface.

(F) Multiple occupancies (colored differently by C atoms for visualization) by Q226 and R230 in the wild type and S3 complexes. 2mFo-DFc maps are rendered at a cutoff of 1 σ. The purine at position -2 characteristic of the PU.1 binding motif is colored in cyan.

(G) Loss of H-bonding geometry in the S3 chimera for the canonical high-affinity interactions with the 5′-flanking purine at position -2.

5-methyl-isocytosine (iso-^{5m} dC; the 5-methyl substitution being the commercially available isocytosine), the binding profile was similar to 2-AP. Controlling for the 5-methyl substitution with ^{5m}dC (paired with dG), binding improved to a level similar to GIAA without osmotic stress, but with significant loss of osmotic sensitivity relative to unmodified DNA. Thus, the 2-carbonyl, a strict H-bond acceptor, was specifically essential for direct readout of the Arg×GC motif. To test for secondary effects of the base substitutions on DNA structure, which may affect PU.1 recognition, we examined the NMR chemical shifts of imino protons, which are highly sensitive to their microenvironment, in model DNA hairpins designed to mimic the PU.1-binding sequences [Figure 4B]. Significant chemical shift perturbations relative to the unmodified sequence did not extend beyond one position upstream from the substitution. Substitutions targeted at integrated readout thus retain a high degree of conservation in DNA conformation, whether the primary target is direct or indirect.

Transduction in contacted indirect readout governs hydration properties

The osmotic stress data indicate that mutations targeting direct readout by R230 do not perturb hydration properties of the complex without essentially abolishing high-affinity binding. An alternate route to selective perturbation to hydration was to swap conserved secondary structure elements with ETS homologs that lack osmotic sensitive DNA binding such as Ets-1.[48] The most effective example was a PU.1/Ets-1 chimera, termed S3, in which the eponymous β-strand was replaced in an otherwise wild-type PU.1 structure [Figure 5A]. The S3 chimera bound the

Please cite this article in press as: Vernon et al., Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors, Structure (2023), https://doi.org/10.1016/j.str.2023.11.003

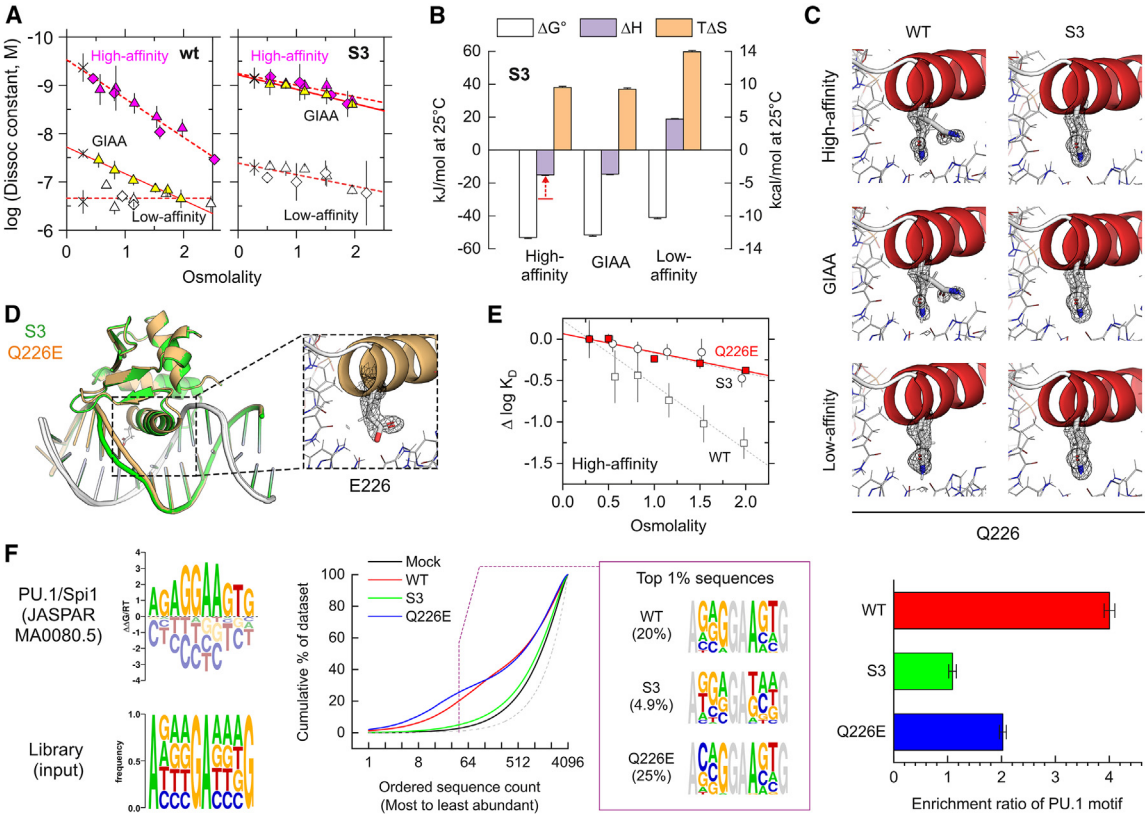


Figure 6. The dynamic basis of osmotically sensitive PU.1/DNA binding and its impact on target specificity

(A) Osmotic sensitivity of DNA binding by the S3 chimera of PU.1. For clarity of presentation, only experiments using betaine (triangles) and acetamide (diamonds) as osmolytes are shown. The data for wild-type PU.1 is included for comparison. Data points represent the average ±S.D. of three or more experiments.

(B) Thermodynamics of S3/DNA binding as measured by ITC. Error bars are fitting errors for ΔH and propagated errors for ΔG° and $T\Delta S$.

(C) Occupancies of the Q226 side chain in wild-type PU.1 and the S3 chimera in co-crystallographic complexes with the indicated DNA. Electron density difference (2mFo-DFc) maps are contoured at 1.0 σ.

(D) Overlay of the Q226E mutant, which conserves the wild-type fold, with the S3 chimera in complex with the same high-affinity DNA. In contrast with wild type, E226 does not adopt multiple occupancies even though there is no structural impediment to do so.

(E), Q226E exhibits the same loss of osmotic sensitive binding as S3 as probed with betaine. Data points represent the average ±S.D. of three or more experiments.

(F), Wild type and mutant PU.1 constructs are screened against a synthetic library with six randomized positions designed around the human PU.1 binding motif. The diversity of the library was 4^6 = 4,096. Biases inherited by the library at synthesis are shown and handled in the analysis. Target selectivity is analyzed in terms of a model-free contraction of sequence space and modeled enrichment in the PU.1-binding motif (duplicate independent samples ±S.D).

optimal PU.1 site with similar affinity (0.70 ± 0.15 nM) as wild type (0.43 ± 0.1 nM) under non-stressed conditions at only one-third of the osmotic dependence [Figure 5B]. As S3 (the structural element) is engaged exclusively in indirect readout of the 5′ flanking positions to the core consensus (Figure 1), structural elucidation of this chimera could provide significant insight into the interplay between conformational dynamics on the one hand, given the inosine-substituted complex, and hydration properties of binding on the other.

We solved the co-crystal structure of S3 in complex with the same high-affinity sequence as wild-type ΔN165 (PDB: 8SMH). S3 co-crystallized in the same space group and diffracted at comparably high resolution (1.37 Å) as wild-type protein, enabling direct comparisons among these models. The bidentate Arg×GC motif was locally preserved, but a superimposition with the wild-type complex revealed that the major groove was locally widened in the S3 complex near the substituted β strand [Figure 5C]. The groove perturbations were accompanied by corresponding changes in local base pair step and base-pair parameters from the wild-type complex, involving primarily angular parameters [Figure 5D]. The S3 element in wild-type ΔN165 contained a pair of consecutive Lys residues at positions 242 and 243 (Figure 5A). The first instance corresponded to a His residue in S3, which was predicted by computational analysis (using PROPKA[49]) of the S3 co-crystal structure to be unprotonated at neutral pH (pKa 6.1). In the wild-type complex, the two Lys residues in the S3 strand and three additional Lys in the downstream loop present a spaced array of ionic contacts with the DNA backbone phosphates along the 5′-flanking positions [Figure 5E]. The disruption of these contacts in S3 results in a local distortion at the missing ionic contact.

Beyond the immediate effects on backbone contacts, the S3 chimera showed transduced perturbations in another region of

Please cite this article in press as: Vernon et al., Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors, Structure (2023), https://doi.org/10.1016/j.str.2023.11.003

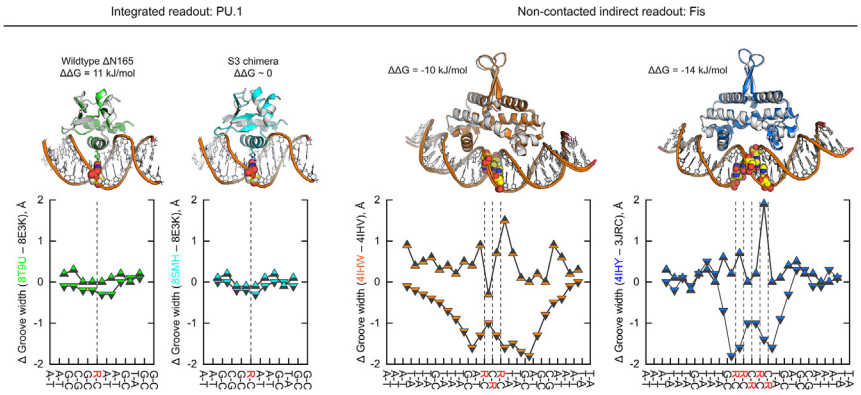


Figure 7. Structural comparison of integrated and non-contacted indirect readout as discerned by inosine-substituted DNA

Integrated readout is represented by the two sets of pair-matched PU.1 complexes (wild type and S3). Two examples of non-contacted indirect readout are afforded by Fis protein. Structural pairs are aligned by the protein backbones. Inosines are rendered in the structures as spheres. Major (up triangles) and minor (down triangles) groove widths (P-P distances) are taken as differences relative to the unmodified sequence (black and white). Thermodynamic data for Fis are taken from Hancock et al.,[26] wherein the two DNA sequences are referred to as F28 (4IHV) and F29 (3JRC).

direct readout by R233, the second signature Arg residues (companion to R230) in ETS domain [Figure 5F]. In the wild-type complex, the side chain of Q226 exhibits excess electron densities that, as we reported recently,[34] correspond to two major occupancies, one of which ("up") directly H-bonds with N7 of a 5′-flanking purine in the 5′-GGAA-3′ strand. This conformation is coupled to R233 via an ordered water molecule, and the coupling between the two residues is characteristic of high-affinity binding by wild-type PU.1. In the S3 complex, Q226 exhibits only a single occupancy in which the Q226-R233 couple is broken. This "down" configuration is also characteristically absent in low-affinity wild-type complexes,[34] even though the S3 complex is as high-affinity in solution as wild type. However, unlike the wild-type low-affinity complex, in which R233 is otherwise well-behaved in a single occupancy, in the S3 complex it exhibits an additional occupancy in contact with a flanking backbone phosphate. Superimposing the S3 and wild-type complexes revealed that the 5′-flanking G_{-2} in the S3 chimera is displaced such that its N7 was no longer within H-bonding distance with Q226 in the up conformation [Figure 5G]. Since the DNA is identical in both structures, we conclude that the up conformation of Q226 is selected by the DNA conformation, not the other way around. The crystallographic evidence thus revealed that the indirect readout by PU.1 at the 5′-flanking positions was directionally transduced to the direct readout interface to regulate chemical recognition of the core consensus.

Dissection of indirect readout reveals the unique role of water in DNA specificity

Since wild-type ΔN165 distinguishes the inosine-substituted optimal site with reduced affinity, we asked whether S3 would exhibit the same response. In stark contrast with wild type, the S3 chimera bound the GIAA sequence identically as the unmodified site within experimental uncertainty across the full range of osmotic pressure [Figure 6A]. Calorimetric measurements in the absence of osmotic stress showed half the magnitude in ΔH for the S3 chimera relative to wild-type ΔN165 [Figure 6B]. As ΔG for both S3 and wild type is similar, S3 binding is more entropically driven, in accord with less hydration water as indicated by osmotic stress. Independently, a similar lack of enthalpic effect from binding GIAA as with wild-type ΔN165 strongly suggests that the perturbations from inosine substitution are similar in both complexes.

The sensitivity of GIAA to hydration mutant S3 from wild-type PU.1 makes it a distinctly useful substrate with which to structurally probe the hydration properties of PU.1/DNA binding. To elucidate the structural basis of this difference, we solved the S3 chimera in complex with GIAA (PDB: 8SP1) and low-affinity DNA (PDB: 8SMJ). All complexes co-crystallized in the identical space group and diffracted to comparably high resolutions. As with wild-type PU.1, inosine substitution did not significantly perturb the conformation of the S3 complex, as shown by a Fo-Fo difference map of the two S3 complexes [Figure S7]. Since the S3/GIAA complex appears to be also dynamically perturbed, and knowing from the wild-type counterpart that the $\Delta\Delta S$ due to dynamics changes was negative, the structural evidence suggests that the negligible $\Delta\Delta S$ for S3 represented the compensation from decreased water uptake.

Comparison of corresponding pairs of DNA complexes (GGAA, GIAA, and low-affinity) bound by wild type and S3 (excluding crystal contacts) pointed to the Q226 side chain as the standout difference for the two complexes [Figure 6C]. Significantly, the up/down occupancies that characterized osmotically sensitive binding by wild-type PU.1 were fully retained in its complex with GIAA. In sharp contrast, all S3 complexes exhibited only the single down formation characteristic of low-osmotic sensitivity binding (including the low-affinity wild-type complex). The consistency with which the behavior of Q226 *in crystallo* associates with sequence-dependent osmotic sensitivity in solution, including inosine-substituted DNA, points to Q226 as the structural nexus of osmotically sensitive DNA recognition by PU.1.

To test this structure-hydration relationship directly, we examined the co-crystal structure of a reported Q226E mutant (PDB: 8EMD).[34] This mutation abolishes H-bonding complementarity with N7 of the 5′-flanking purine at the -2 position. In the Q226E structure, E226 adopts a single down conformation, identically as the low-affinity wild-type complex and the S3 complexes [Figure 6D, *c.f.*, Figure 5G]. In solution, Q226E exhibited indistinguishable osmotic dependence in high-affinity DNA binding as the S3 chimera [Figure 6E]. The structural and thermodynamic data together thus suggest that the hydration properties of PU.1 are coupled to contacted indirect readout in the 5′-flanking positions and are integrated by the Q226 side chain.

The Q226E mutation is known to alter the sequence specificity of PU.1 in genomic occupancy by favoring cognate variants harboring a cytosine in the −2 position, which are targeted by

Please cite this article in press as: Vernon et al., Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors, Structure (2023), https://doi.org/10.1016/j.str.2023.11.003

distantly related ETS factors.[34,50] As the Q226E mutant also shared the low osmotic sensitivity of S3 (Figure 6D) with high-affinity DNA, in contrast with wild type, these strongly suggested that the hydration properties of wild-type PU.1 were thermodynamically responsible for its DNA specificity. To address this question, we screened an oligonucleotide library against wild-type ΔN165, the S3 chimera, and Q226E point mutant under identical non-osmotically stressed solution conditions [Figure 6F]. We designed the library around the most up-to-date binding motif for human PU.1 (curated in the JASPAR database, profile MA0080.5). Based on an expected dataset of $>3 \times 10^5$ reads, we randomized six key positions in the motif to ensure robust sequence read depths.[5] Bound fractions of the DNA library at sub-saturating protein concentrations were resolved by high-throughput sequencing.

As an explicit model-free measure of specificity, we enumerated the distribution of unique sequences bound by each protein. Binding by wild-type ΔN165 resulted in a marked consolidation of sequences relative to the S3 chimera: the top 1% of unique sequences accounted for 20% of the wild-type dataset and were highly enriched in the PU.1-binding motif. In contrast, the top-1% sequences comprised only 5% of the S3-bound ensemble. Remarkably, the Q226E mutant showed even stronger consolidation (25%) of the top-1% sequences than wild type, but instead enriching sequences bearing cytosine in the -2 position. For a modeled approach, we used a standard algorithm (SEA from the MEME Suite) to determine the enrichment of the PU.1-binding motif among the sequence ensembles (Figure 6F). The PU.1 motif was over-represented by 4-fold in the wild-type-bound ensemble, while the S3 chimera was essentially unenriched in PU.1-binding sequences. Enrichment by the Q226E mutant was reduced by half relative to wild type.

In summary, the S3 chimera is substantively devoid of the selectivity of wild-type PU.1, consistent with the former's indifference to inosine substitution (Figure 6A). In contrast, Q226E reshaped the sequence landscape for PU.1 toward one more typical of distant ETS relatives. As both mutants exhibited the same impaired osmotic properties, the library screen data suggested that the hydration properties of PU.1 served to enforce a specification for G at the -2 position, rather than as a general specificity determinant. The side chain of R233, which interacts with Q226 in the wild-type complexes, remains in canonical conformation in the Q226E mutant but is dislocated in the S3 chimera (Figure 5F). The coupling between R233 and Q226 therefore appears to represent the structural mediator of PU.1 target specificity.

DISCUSSION

The affinity and specificity of protein-DNA recognition are thermodynamically specified in terms of equilibrium constants or free energy changes. Despite this standard, the interpretation is frequently framed in terms of contacts in the nucleoprotein complex which makes up only a subset of the thermodynamic system. The emphasis on well-ordered contacts neglects or at least makes implicit the significant contributions from nearby solvent and other solutes that complete the thermodynamic system,[41] as well as the properties of the protein and DNA in their unbound states.

For the bidentate Arg×GC motif, which occurs widely in transcription factor/DNA complexes (Figure 1C), the pleotropic effects of inosine substitution as modeled by PU.1 would remain cryptic if considered solely in terms of contacts in the complexes. For wild-type PU.1, the micro-heterogeneities in helical structure or even dynamics of the complex alone were insufficient to explain the unfavorable $\Delta\Delta S$ in binding the inosine-substituted site. The resultant impact on affinity, ~80-fold weaker relative to wild type, is functionally significant as recently shown for the CSF1R receptor promoter (a native PU.1 target).[34] The thermodynamics of binding the inosine-substituted site could only be rationalized in light of the substitution's effect on the unbound state. For the S3 chimera, the bound structure becomes informative only when considered in conjunction with the change in thermodynamic hydration (as mediated by the Q226 side chain) upon binding. The biological relevance of these thermodynamics is significant: the S3 chimera, which is indifferent to inosine-substituted DNA and whose hydration properties recapitulate the point mutant Q226E, is a poorly sequence-specific protein. The Q226E mutant exhibits off-target occupancy of genomic sites targeted by other ETS proteins and is a recurrent lesion in Waldenström macroglobulinemia, an incurable lymphoma.[50]

The role of thermodynamic water in DNA recognition

The co-crystallographic structures of the wild-type high-affinity PU.1/DNA complex and the hydration mutant S3, which were refined to comparable resolution, present similar numbers of ordered water molecules in the protein/DNA interface. The divergent osmotic dependencies in binding the same DNA site for the two proteins therefore suggest significant differences in thermodynamic hydration, which is not crystallographically resolved. A solution NMR ensemble of unbound ΔN165 [51] shows no significant difference in the folded elements compared with the DNA-bound protein. However, the ensemble exhibits a broad range of conformations adopted by residues flanking the ETS domain, which are also not resolved in the co-crystallographic structures. It is unknown the extent to which the conformational ensemble of the flanking disordered regions is modified in the bound state. Evidence that this may be the case is derived from the effects of these disordered residues on the homodimerization of PU.1 in both bound and unbound states.[52] Differences in their intramolecular interactions with the protein, or possibly the bound DNA in the complex, may therefore be associated with corresponding changes in thermodynamic hydration, as well as the net positive entropy change in DNA binding.

Comparison with non-contacted indirect readout

While inosine (and DAP) substitution for guanine and adenine have been used in many protein/DNA studies, there are few high-resolution models of pair-matched unmodified and substituted complexes for explicit comparisons. Beyond the structures presented in this work, the best examples are afforded by the dimeric Fis protein, which exhibits strictly alternating segments of direct and indirect readout along the binding site.[22] Inosine or DAP substitution in the central segment of the Fis binding site, which is not contacted by protein, markedly perturbs binding affinities as well as the conformation of the Fis-bound DNA.[26] Using the groove widths as representative metrics, inosine-substituted Fis complexes deviate from their unmodified counterparts by as much as 2 Å, an order of magnitude over PU.1 [Figure 7]. These deviations, particularly the minor

groove compressions at the substituted positions, reflect differential curvature in the non-contacted region. The strong structural effects observable in these complexes sharply contrast with PU.1, where direct and indirect readout is tightly integrated. As PU.1 demonstrates, integrated readout is structurally dominated by the conformational dictates of the protein, and the underlying thermodynamic transactions are not disclosed by the complex alone. In addition to the ps-ns helical dynamics sampled by the MD simulations here, inosine substitution for guanine is known to promote conformational dynamics in free DNA across timescales in terms of global helical melting [53], local base pair opening[54] and transient Hoogsteen base pairs.[55]

Conclusion

Direct and indirect readout of DNA denotes formal divisions among overlapping possibilities in structure, dynamics, and thermodynamics (particularly hydration). As inosine substitution of the bidentate Arg×GC motif demonstrates, direct and indirect readout may be profoundly integrated to a degree that is not dissectable from structural characterization of the complexed state alone. Although direct readout devoid of an indirect component may occur locally,[17] a general analysis of direct readout should be framed as integrated readout. For PU.1, direct interrogation of integrated readout has provided the means of dissecting the dynamics and hydration components of DNA recognition, the latter of which mediates the thermodynamic control of specificity beyond other ETS-family relatives.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Protein expression and purification
  - X-Ray crystallography
  - NMR spectroscopy
  - Fluorescence polarization titrations
  - Isothermal titration calorimetry
  - DNA library screening
  - Molecular dynamics simulations
  - Time series analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.str.2023.11.003.

ACKNOWLEDGMENTS

We thank the beamline staff at the Advanced Light Source at Lawrence Berkeley National Laboratory (Berkeley, CA), SER-CAT at the Advanced Photon Source (Lemont, IL), and NSLS-II at Brookhaven National Laboratory (Upton, NY) for their extensive support during the X-ray data collections. G.M.K.P. is grateful to Drs. Michael E. Harris (Northwestern University) and Michael Van Dyke (Kennesaw State University) for helpful suggestions on DNA library screening. This investigation was supported by NIH grants GM111749 (to W.D.W.), GM137160 and HL155178 (to G.M.K.P.) and NSF grant MCB 2028902 to G.M.K.P and M.W.G. This work is dedicated to the memory of Gary Hutton.

AUTHOR CONTRIBUTIONS

Conceptualization, G.M.K.P.; Methodology, M.W.G., W.D.W., and G.M.K.P. Investigation, T.N.V., J.R.T., A.V.A., M.W.G., and G.M.K.P.; Formal data analysis, T.N.V., J.R.T., A.V.A., M.W.G., and G.M.K.P.; Writing and editing, T.N.V., J.R.T., M.W.G., W.D.W., and G.M.K.P.; Data visualization, T.N.V., J.R.T., M.W.G., and G.M.K.P.; Supervision, M.W.G., W.D.W., and G.M.K.P.; Funding acquisition, M.W.G., W.D.W., and G.M.K.P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: August 8, 2023
Revised: October 12, 2023
Accepted: November 7, 2023
Published: December 1, 2023

REFERENCES

1. Seeman, N.C., Rosenberg, J.M., and Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl. Acad. Sci. USA *73*, 804–808.
2. Alipanahi, B., Delong, A., Weirauch, M.T., and Frey, B.J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat. Biotechnol. *33*, 831–838.
3. Stormo, G.D. (2013). Modeling the specificity of protein-DNA interactions. Quant. Biol. *1*, 115–130.
4. Weirauch, M.T., Cote, A., Norel, R., Annala, M., Zhao, Y., Riley, T.R., Saez-Rodriguez, J., Cokelaer, T., Vedenko, A., Talukder, S., et al. (2013). Evaluation of methods for modeling transcription factor sequence specificity. Nat. Biotechnol. *31*, 126–134.
5. Le, D.D., Shimko, T.C., Aditham, A.K., Keys, A.M., Longwell, S.A., Orenstein, Y., and Fordyce, P.M. (2018). Comprehensive, high-resolution binding energy landscapes reveal context dependencies of transcription factor binding. Proc. Natl. Acad. Sci. USA *115*, E3702–E3711.
6. Roven, C., and Bussemaker, H.J. (2003). REDUCE: An online tool for inferring cis-regulatory elements and transcriptional module activities from microarray data. Nucleic Acids Res. *31*, 3487–3490.
7. Rastogi, C., Rube, H.T., Kribelbauer, J.F., Crocker, J., Loker, R.E., Martini, G.D., Laptenko, O., Freed-Pastor, W.A., Prives, C., Stern, D.L., et al. (2018). Accurate and sensitive quantification of protein-DNA binding affinity. Proc. Natl. Acad. Sci. USA *115*, E3692–E3701.
8. Zhou, T., Yang, L., Lu, Y., Dror, I., Dantas Machado, A.C., Ghane, T., Di Felice, R., and Rohs, R. (2013). DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. Nucleic Acids Res. *41*, W56–W62.
9. Mathelier, A., Xin, B., Chiu, T.P., Yang, L., Rohs, R., and Wasserman, W.W. (2016). DNA Shape Features Improve Transcription Factor Binding Site Predictions In Vivo. Cell Syst. *3*, 278–286.e4.
10. Chiu, T.P., Rao, S., Mann, R.S., Honig, B., and Rohs, R. (2017). Genome-wide prediction of minor-groove electrostatic potential enables

biophysical modeling of protein-DNA binding. Nucleic Acids Res. *45*, 12565–12576.

11. Chiu, T.P., Rao, S., and Rohs, R. (2023). Physicochemical models of protein-DNA binding with standard and modified base pairs. Proc. Natl. Acad. Sci. USA *120*. e2205796120.
12. Koudelka, G.B., Mauro, S.A., and Ciubotaru, M. (2006). Indirect readout of DNA sequence by proteins: the roles of DNA sequence-dependent intrinsic and extrinsic forces. Prog. Nucleic Acid Res. Mol. Biol. *81*, 143–177.
13. Bailly, C., and Waring, M.J. (1998). The use of diaminopurine to investigate structural properties of nucleic acids and molecular recognition between ligands and DNA. Nucleic Acids Res. *26*, 4309–4314.
14. Bailly, C., Payet, D., Travers, A.A., and Waring, M.J. (1996). PCR-based development of DNA substrates containing modified bases: an efficient system for investigating the role of the exocyclic groups in chemical and structural recognition by minor groove binding drugs and proteins. Proc. Natl. Acad. Sci. USA *93*, 13623–13628.
15. Cristofalo, M., Kovari, D., Corti, R., Salerno, D., Cassina, V., Dunlap, D., and Mantegazza, F. (2019). Nanomechanics of Diaminopurine-Substituted DNA. Biophys. J. *116*, 760–771.
16. Starr, D.B., and Hawley, D.K. (1991). TFIID binds in the minor groove of the TATA box. Cell *67*, 1231–1240.
17. Lindemose, S., Nielsen, P.E., and Møllegaard, N.E. (2008). Dissecting direct and indirect readout of cAMP receptor protein DNA binding using an inosine and 2,6-diaminopurine in vitro selection system. Nucleic Acids Res. *36*, 4797–4807.
18. Khrapunov, S., and Brenowitz, M. (2004). Comparison of the effect of water release on the interaction of the Saccharomyces cerevisiae TATA binding protein (TBP) with "TATA Box" sequences composed of adenosine or inosine. Biophys. J. *86*, 371–383.
19. Brennan, C.A., Van Cleve, M.D., and Gumport, R.I. (1986). The effects of base analogue substitutions on the cleavage by the EcoRI restriction endonuclease of octadeoxyribonucleotides containing modified EcoRI recognition sequences. J. Biol. Chem. *261*, 7270–7278.
20. McLaughlin, L.W., Benseler, F., Graeser, E., Piel, N., and Scholtissek, S. (1987). Effects of functional group changes in the EcoRI recognition site on the cleavage reaction catalyzed by the endonuclease. Biochemistry *26*, 7238–7245.
21. Sibghat, U., Gallinari, P., Xu, Y.Z., Goodman, M.F., Bloom, L.B., Jiricny, J., and Day, R.S., 3rd (1996). Base analog and neighboring base effects on substrate specificity of recombinant human G:T mismatch-specific thymine DNA-glycosylase. Biochemistry *35*, 12926–12932.
22. Stella, S., Cascio, D., and Johnson, R.C. (2010). The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. Genes Dev. *24*, 814–826.
23. Watkins, D., Hsiao, C., Woods, K.K., Koudelka, G.B., and Williams, L.D. (2008). P22 c2 repressor-operator complex: mechanisms of direct and indirect readout. Biochemistry *47*, 2325–2338.
24. Watkins, D., Mohan, S., Koudelka, G.B., and Williams, L.D. (2010). Sequence recognition of DNA by protein-induced conformational transitions. J. Mol. Biol. *396*, 1145–1164.
25. Mauro, S.A., Pawlowski, D., and Koudelka, G.B. (2003). The role of the minor groove substituents in indirect readout of DNA sequence by 434 repressor. J. Biol. Chem. *278*, 12955–12960.
26. Hancock, S.P., Ghane, T., Cascio, D., Rohs, R., Di Felice, R., and Johnson, R.C. (2013). Control of DNA minor groove width and Fis protein binding by the purine 2-amino group. Nucleic Acids Res. *41*, 6750–6760.
27. Hancock, S.P., Stella, S., Cascio, D., and Johnson, R.C. (2016). DNA Sequence Determinants Controlling Affinity, Stability and Shape of DNA Complexes Bound by the Nucleoid Protein Fis. PLoS One *11*, e0150189.
28. 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.
29. Schultz, S.C., Shields, G.C., and Steitz, T.A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science *253*, 1001–1007.
30. Chen, C., and Pettitt, B.M. (2016). DNA Shape versus Sequence Variations in the Protein Binding Process. Biophys. J. *110*, 534–544.
31. Kim, J.L., and Burley, S.K. (1994). 1.9 A resolution refined structure of TBP recognizing the minor groove of TATAAAAG. Nat. Struct. Biol. *1*, 638–653.
32. Wu, J., Parkhurst, K.M., Powell, R.M., Brenowitz, M., and Parkhurst, L.J. (2001). DNA bends in TATA-binding protein-TATA complexes in solution are DNA sequence-dependent. J. Biol. Chem. *276*, 14614–14622.
33. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. *10*, 252–263.
34. Terrell, J.R., Taylor, S.J., Schneider, A.L., Lu, Y., Vernon, T.N., Xhani, S., Gumpper, R.H., Luo, M., Wilson, W.D., Steidl, U., and Poon, G.M.K. (2023). DNA selection by the master transcription factor PU.1. Cell Rep. *42*, 112671.
35. Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (2001). Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res. *29*, 2860–2874.
36. Lohse, M.B., Rosenberg, O.S., Cox, J.S., Stroud, R.M., Finer-Moore, J.S., and Johnson, A.D. (2014). Structure of a new DNA-binding domain which regulates pathogenesis in a wide variety of fungi. Proc. Natl. Acad. Sci. USA *111*, 10404–10410.
37. Rould, M.A., and Carter, C.W. (2003). Isomorphous Difference Methods. In Methods in Enzymology (Academic Press), pp. 145–163.
38. Jain, S., and Sundaralingam, M. (1989). Effect of crystal packing environment on conformation of the DNA duplex. Molecular structure of the A-DNA octamer d(G-T-G-T-A-C-A-C) in two crystal forms. J. Biol. Chem. *264*, 12780–12784.
39. Heinemann, U., and Alings, C. (1991). The conformation of a B-DNA decamer is mainly determined by its sequence and not by crystal environment. EMBO J. *10*, 35–43.
40. Shakked, Z. (1991). The influence of the environment on DNA structures determined by X-ray crystallography. Curr Opin Struc Biol *1*, 446–451.
41. Spolar, R.S., and Record, M.T., Jr. (1994). Coupling of local folding to site-specific binding of proteins to DNA. Science *263*, 777–784.
42. Kumar, R., and Grubmüller, H. (2015). do_x3dna: a tool to analyze structural fluctuations of dsDNA or dsRNA from molecular dynamics simulations. Bioinformatics *31*, 2583–2585.
43. Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M., and Zhurkin, V.B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc. Natl. Acad. Sci. USA *95*, 11163–11168.
44. Lankas, F., Sponer, J., Langowski, J., and Cheatham, T.E., 3rd (2003). DNA basepair step deformability inferred from molecular dynamics simulations. Biophys. J. *85*, 2872–2883.
45. Landau, L.D., and Lifshitz, E.M. (1980). Fluctuations. In Statistical Physics, L.D. Landau and E.M. Lifshitz, eds. (Butterworth-Heinemann)), pp. 333–400.
46. Poon, G.M.K. (2012). Sequence discrimination by DNA-binding domain of ETS family transcription factor PU.1 is linked to specific hydration of protein-DNA interface. J. Biol. Chem. *287*, 18297–18307.
47. Law, S.M., Eritja, R., Goodman, M.F., and Breslauer, K.J. (1996). Spectroscopic and calorimetric characterizations of DNA duplexes containing 2-aminopurine. Biochemistry *35*, 12329–12337.
48. Albrecht, A.V., Kim, H.M., and Poon, G.M.K. (2018). Mapping interfacial hydration in ETS-family transcription factor complexes with DNA: a chimeric approach. Nucleic Acids Res. *46*, 10577–10588.
49. Olsson, M.H.M., Søndergaard, C.R., Rostkowski, M., and Jensen, J.H. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. J. Chem. Theory Comput. *7*, 525–537.

50. Roos-Weil, D., Decaudin, C., Armand, M., Della-Valle, V., Diop, M.K., Ghamlouch, H., Ropars, V., Hérate, C., Lara, D., Durot, E., et al. (2019). A Recurrent Activating Missense Mutation in Waldenstrom Macroglobulinemia Affects the DNA Binding of the ETS Transcription Factor SPI1 and Enhances Proliferation. Cancer Discov. *9*, 796–811.

51. Perez-Borrajero, C., Lin, C.S.H., Okon, M., Scheu, K., Graves, B.J., Murphy, M.E.P., and McIntosh, L.P. (2019). The Biophysical Basis for Phosphorylation-Enhanced DNA-Binding Autoinhibition of the ETS1 Transcription Factor. J. Mol. Biol. *431*, 593–614.

52. Xhani, S., Lee, S., Kim, H.M., Wang, S., Esaki, S., Ha, V.L.T., Khanezarrin, M., Fernandez, G.L., Albrecht, A.V., Aramini, J.M., et al. (2020). Intrinsic disorder controls two functionally distinct dimers of the master transcription factor PU.1. Sci. Adv. *6*, eaay3178.

53. Watkins, N.E., Jr., and SantaLucia, J., Jr. (2005). Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. Nucleic Acids Research. *33*, 6258–6267.

54. Ferris, Z.E., Li, Q., and Germann, M.W. (2019). Substituting Inosine for Guanosine in DNA: Structural and Dynamic Consequences. Nat. Prod. Commun. *14*. 1934578X1985003.

55. Nikolova, E.N., Stull, F., and Al-Hashimi, H.M. (2014). Guanine to inosine substitution leads to large increases in the population of a transient G.C Hoogsteen base pair. Biochemistry *53*, 7145–7147.

56. Lu, X.J., and Olson, W.K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res. *31*, 5108–5121.

57. Barthels, F., Schirmeister, T., and Kersten, C. (2021). BANDeltaIT: B'-Factor Analysis for Drug Design and Structural Biology. Mol. Inform. *40*, e2000144.

58. Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One *11*, e0163962.

59. Salvatier, J., Wiecki, T.V., and Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. PeerJ Comput. Sci. *2*, e55.

60. Stephens, D.C., Kim, H.M., Kumar, A., Farahat, A.A., Boykin, D.W., and Poon, G.M. (2016). Pharmacologic efficacy of PU.1 inhibition by heterocyclic dications: a mechanistic analysis. Nucleic Acids Res. *44*, 4005–4013.

61. Wells, J.W. (1992). Analysis and interpretation of binding at equilibrium. In Receptor-Ligand Interactions: a Practical Approach, E.C. Hulme, ed. (IRL Press at Oxford University Press), pp. 289–395.

62. Esaki, S., Evich, M.G., Erlitzki, N., Germann, M.W., and Poon, G.M.K. (2017). Multiple DNA-binding modes for the ETS family transcription factor PU.1. J. Biol. Chem. *292*, 16044–16054.

63. Poon, G.M.K. (2012). DNA binding regulates the self-association of the ETS domain of PU.1 in a sequence-dependent manner. Biochemistry *51*, 4096–4107.

64. Wang, S., Poon, G.M.K., and Wilson, W.D. (2015). Quantitative Investigation of Protein–Nucleic Acid Interactions by Biosensor Surface Plasmon Resonance. In DNA-Protein Interactions, B.P. Leblanc and S. Rodrigue, eds. (Springer New York), pp. 313–332.

65. Poon, G.M.K. (2010). Explicit formulation of titration models for isothermal titration calorimetry. Anal. Biochem. *400*, 229–236.

66. Ivani, I., Dans, P.D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A., et al. (2016). Parmbsc1: a refined force field for DNA simulations. Nat. Methods *13*, 55–58.

Please cite this article in press as: Vernon et al., Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors, Structure (2023), https://doi.org/10.1016/j.str.2023.11.003

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E. coli BL21(DE3)pLyS	ThermoFisher	C602003
Chemicals, peptides, and recombinant proteins		
hPU.1 ETS domain, residue 165 to 270 (ΔN165)	Addgene	199796
ΔN165 S3	This manuscript	N/A
ΔN165 Q226E	This manuscript	N/A
Deposited data		
WT ΔN165/GIAA	This manuscript	PDB: 8T9U
S3 Chimera/High-Affinity	This manuscript	PDB: 8SMH
S3 Chimera/GIAA	This manuscript	PDB: 8SP1
S3 Chimera/Low-Affinity	This manuscript	PDB: 8SMJ
Oligonucleotides		
Sequences are shown in Table S2	Integrated DNA Technologies (IDT)	N/A
Software and algorithms		
CCP4	CCP4	https://www.ccp4.ac.uk/
Phenix	Phenix Online	https://phenix-online.org
Coot	University of Cambridge	https://bernhardcl.github.io/coot/wincoot-download.html
AutoProc	Global Phasing	https://www.globalphasing.com/autoproc/
PyMOL	Schrödinger	https://pymol.org
3DNA	Lu et al.[56]	http://web.x3dna.org/
BANΔIT	Barthels et al.[57]	https://bandit.uni-mainz.de/
OriginPro	OriginLab	https://www.originlab.com/
GROMACS	GROMACS	https://www.gromacs.org/index.html
SeqKit	Shen et al.[58]	https://bioinf.shenwei.me/seqkit
MEME Suite	University of Washington	https://meme-suite.org/meme/tools/sea
REDUCE	Roven et al.[6]	http://reducesuite.bussemakerlab.org/
PyMC	Salvatier et al.[59]	https://www.pymc.io/welcome.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Gregory Poon (gpoon@gsu.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Co-crystallographic PU.1/DNA structures and electron densities have been deposited at wwPDB and are publicly available as of the date of publication. Accession numbers (8T9U, 8SMH, 8SP1, and 8SMJ) are listed in Table 1.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Please cite this article in press as: Vernon et al., Dissection of integrated readout reveals the structural thermodynamics of DNA selection by transcription factors, Structure (2023), https://doi.org/10.1016/j.str.2023.11.003

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Chemically competent BL21(DE3) pLysS *E. coli* was purchased from ThermoFisher and used directly without further authentication. Cells were cultured in LB media at 37°C.

METHOD DETAILS

Protein expression and purification

Recombinant ETS domains were expressed in BL21(DE3) pLysS *E. coli* as previously described.[48] Cleared lysates were processed by ion exchange or immobilized metal affinity chromatography and then polished on a HiLoad 16/600 Superdex 75 (Cytiva) in 10 mM HEPES, pH 7.4, containing 0.15 M NaCl. Protein concentrations were determined by UV absorption at 280 nm. Each construct was verified by MALDI-ToF(+) analysis.

X-Ray crystallography

Co-crystallization of PU.1/DNA complexes was performed as described.[34] In brief, equimolar mixtures of purified protein and duplex DNA were crystallized over three to five days by vapor diffusion at 293 K in a hanging drop consisting of a 1:1 mixture of complex with mother liquor containing 100 mM sodium acetate, pH 4.6, and 2% PEG 3350. X-ray diffraction datasets were collected at SER-CAT at the Advanced Photon Source, Chicago, IL, the Advanced Light Source at Lawrence Berkeley National Laboratory, Berkeley, CA, and the National Synchrotron Light Source II at Brookhaven National Laboratory, Upton, NY.

The diffraction data were processed using the XDS package and scaled using Aimless in the CCP4 package or autoPROC. Molecular replacement was performed using the wildtype PU.1 complex (8E3K) as the search model in the PHASER-MR module of PHENIX. Structural refinement was carried out in Phenix.refine. Isomorphous difference (Fo-Fo) maps were generated with the eponymous tool in PHENIX using the structure factors of the complexes being compared and the wildtype complex for phase calculation. DNA helical parameters were computed using x3DNA.[42] Protein B-factors were normalized using BANΔIT.[57] The PU.1/DNA co-crystal structures have been deposited in the RCSB Protein DataBank with the PDB codes: 8T9U, 8SMH, 8SP1, and 8SMJ. For review purposes, coordinates in mmCIF format and accompanying electron density maps (MTZ) and validation reports following wwPDB data deposition are attached as supplements.

NMR spectroscopy

Hairpin-forming oligodeoxynucleotides were dissolved in 0.5 M NaCl to dissociate ionic contaminants and purified on a 5 × 5 mL HiTrap Desalting column (Cytiva). The desalted DNA was lyophilized and dissolved in 20 mM NaH_2PO_4/Na_2HPO_4 containing 50 mM NaCl and 0.5 mM EDTA. D_2O was added to 10% and the pH adjusted to 6.40. NMR experiments were performed on a Bruker Avance I 500 spectrometer equipped with a TBI $\{^1H(^{13}C, X)\}$ *z* gradient probe. For monitoring imino proton resonances a 1-1 jump and return sequence was used to record spectra from 288 K to 308 K. Phase sensitive 1-1 jump and return NOESYs were collected at 288 K with 2048 × 800 data points in the two dimensions and 72 scans per t_1 increment using a 150 ms mixing time and a 1.0 s relaxation delay. 2-D spectra were strip transformed and processed using a 4 K × 2 K matrix. Both dimensions were apodized with shifted $\sin(\pi/2)$ bell functions. Proton chemical shifts were referenced to internal 2,2-dimethyl-2-silapentane-5-sulfonate (DSS).

Fluorescence polarization titrations

Equilibrium protein/DNA titrations were performed as previously described.[52] A Cy3-labeled DNA probe (0.25 nM) and sub-saturating concentrations of protein (10^{-9} to 10^{-8} M) were incubated to equilibrium with unlabeled 23-bp DNA duplexes as described in the text. The binding mixtures contained 10 mM Tris-HCl, pH 7.4, 0.15 M NaCl, 0.1 mg/mL bovine serum albumin, and osmolytes as needed to achieve the desired osmotic pressure. Solution osmolality was determined by vapor pressure measurements on a Wescor VAPRO 5600 instrument. Steady-state anisotropies were measured at 595 nm in a Molecular Dynamics Paradigm micro-plate reader with 530 nm excitation. Operationally, sensitivity was optimized as follows: 1) long integration time (10 s for each polarization direction); 2) wide excitation and emission filter bandwidth; 3) use of a very bright fluorophore (Cy5) that is stable against photobleaching; 4) use of low-volume 384-well black microplates (Corning) to maximize emission at the photomultiplier. Fluorescence anisotropies were computed as mean ± S.D. of three or more experiments.

Model-dependent analysis by non-linear least square methods were performed as previously described[60,61] and summarized as follows. The observed anisotropy r_{obs} was expressed as the fractional bound DNA probe (F_b), scaled by the anisotropy of the probe-bound r_1 and unbound states r_0 as follows:

$$r_{obs} = F_b(r_1 - r_0) + r_0 = F_b \cdot \Delta r + r_0 \quad \text{(Equation 1)}$$

where F_b was fitted to a mutually exclusive model in which PU.1 (P) binds either the probe (D*) or unlabeled DNA competitor (D), but not both. The equilibrium dissociation constants describing the labeled and unlabeled PU.1/DNA complexes are:

$$K_{PD} = \frac{[P][D*]}{[PD*]}$$
$$K_{PD*} = \frac{[P][D]}{[PD]} \quad \text{(Equation 2)}$$

With the independent variable taken as the total titrant concentration, the binding polynomial is a cubic in [PD*][61]:

$$0 = \varphi_0 + \varphi_1[PD*] + \varphi_2[PD*]^2 + \varphi_3[PD*]^3$$

$$\begin{cases} \varphi_0 = -K_{PD*}[D]_t^{\,2}[P]_t \\ \varphi_1 = K_{PD}K_{PD*}[D]_t + K_{PD}[D*]_t[D]_t + K_{PD*}[D]_t^{\,2} + 2K_{PD*}[D]_t[P]_t - K_{PD}[D]_t[P]_t \\ \varphi_2 = -K_{PD}K_{PD*} + K_{PD}^{\,2} - K_{PD}[D*]_t - 2K_{PD*}[D]_t + K_{PD}[D]_t - K_{PD*}[P]_t + K_{PD}[P]_t \\ \varphi_3 = K_{PD*} - K_{PD} \end{cases} \quad \text{(Equation 3)}$$

where the subscript "t" represents the total concentration of the referred species. Root finding for $[PD*] = F_b[D*]_t$ was executed numerically from Equation 3, rather than analytically via the cubic formula, to avoid failure due to loss of significance. To constrain the fit and minimize correlation of the two dissociation constants, K_{PD*} and r_1 were fixed using values independently determined from a direct titration of the probe alone.

Isothermal titration calorimetry

DNA and purified ETS domains were co-dialyzed in 10 mM NaH_2PO_4/Na_2HPO_4 at a pH of 7.4 with 0.15 M NaCl. Dialysate was used in all dilution and washing procedures. DNA was loaded into the cell (975 μL) at 10 μM and protein at 200 μM in the syringe. DNA (or buffer) was titrated with thirty 5.0-μL increments of protein at 25°C with constant stirring (Figure S2A). After baseline subtraction, heat peaks were integrated and fitted as detailed[52,62–65] to determine the enthalpy change for the 1:1 complex (Figure S2B).

DNA library screening

A 121-nt hairpin-forming deoxyoligonucleotide encoding randomized bases at selected positions was used to generate the complete hairpin by isothermal elongation with Taq polymerase at 72°C for 30 min. The product was purified from an agarose gel without chemical processing through a centrifugal device (Ultrafree-DA, Millipore). The purified DNA was titrated with the target PU.1 construct in 10 mM TrisHCl containing 0.15 M NaCl and 1% w/v Ficoll 400, incubated to equilibrium at 25°C, and loaded directly onto a 16% polyacrylamide gel running at 20 V/cm. The gel was stained with GelRed (Biotum) and bound bands corresponding to the partial saturation of the 1:1 complex at the lowest protein concentration were excised. The DNA was eluted by the soak-and-crush method, ethanol-precipitated with 1 μg glycogen (Thermo), and directly redissolved in a reaction mixture for EcoRV (Thermo) to cleave the hairpin at a unique restriction site upstream of the PU.1-binding site. The restriction reaction was then diluted 10-fold into a PCR reaction mixture containing primers that encode adapter sequences for Illumina sequencing. The single 155-bp target was barcoded and sequenced by Azenta Life Sciences (Chelmsford, MA). A typical dataset, consisting of $>3 \times 10^5$ sequences with quality scores over Q20, was analyzed using SeqKit,[58] REDUCE,[6] and SEA (MEME Suite 5.53, University of Washington).

Molecular dynamics simulations

Explicit-solvent simulations were performed with the Amber14SB/parmbsc1 forcefields[66] in the GROMACS 2022.x environment. The co-crystal structure (8E3K) was used as initial coordinates of the wildtype PU.1/DNA complex. Each system was set up in dodecahedral boxes at least 1.0 nm wider than the longest dimension of the solute, solvated with TIP3P water, and neutralized with Na^+ and Cl^- to 0.15 M. Electrostatic interactions were handled by particle-mesh Ewald summation with a 1 nm distance cutoff. All simulations were carried out at 298 K (modified Berendsen thermostat) and 1 bar (Parrinello-Rahman ensemble). A timestep of 2 fs was used and H-bonds were constrained using LINCS. After the structures were energy-minimized by steepest descent, the *NVT* ensemble was equilibrated at 298 K for 1 ns to thermalize the system, followed by another 1 ns of equilibration of the *NPT* ensemble at 1 bar and 298 K. The final *NPT* ensemble was simulated without restraints for up to 2.0 μs, recording coordinates every 1 ps. Convergence of the trajectories were checked by RMSD from the energy-minimized structures, after corrections for periodic boundary effects. Triplicate production runs were carried out using different random seeds in the velocity distribution.

Time series analysis

For RMS fluctuation calculations, concatenated trajectories from the replicas were used. DNA helical fluctuations were analyzed using 3dna[56] and do_x3dna.[42] Model-dependent analysis was performed by Bayesian inference. Trajectories were differenced as $dx_t = x_{i+1} - x_i$ (where the value x at time t is indexed in terms of the recorded frame i) and modeled as univariate stochastic processes as described in the text. Parameter estimation was performed by Markov chain Monte Carlo (MCMC) simulations using PyMC[59] to sample the input and generate posterior estimates of model parameters from defined prior distributions. MCMC simulations typically

involved four chains and 10^5 moves each following 10^3 tuning steps per chain which were discarded as burn-in. Convergence of the simulations were confirmed by inspection of the sampling chains and formally in terms of the Gelman-Rubin metric. Parametric estimates are presented as means of the posterior distribution with the 95% highest posterior density (HPD, or two-tailed credible interval).

QUANTIFICATION AND STATISTICAL ANALYSIS

OriginPro software (OriginLab) was used for statistical analysis. Specific tests, sample sizes, and significance levels are specified in the figure legends and Results.