Cell Reports, Volume *42*

Supplemental information

DNA selection by the master transcription factor PU.1

J. Ross Terrell, Samuel J. Taylor, Amelia L. Schneider, Yue Lu, Tyler N. Vernon, Suela Xhani, Ryan H. Gumpper, Ming Luo, W. David Wilson, Ulrich Steidl, and Gregory M.K. Poon

SUPPLEMENT INDEX

SUPPLEMENTAL TABLES

Table S1. Summary of co-crystallographic structures, related to Figure 2

Table S2. Refinement statistics of co-crystallographic models, related to STAR Methods

[**accompanying spreadsheet**]

Table S3. Co-crystallographic search for the DNA-binding site in the *CD11b* promoter, related to Figure 6

Table S1. Summary of co-crystallographic structures, related to Figure 2

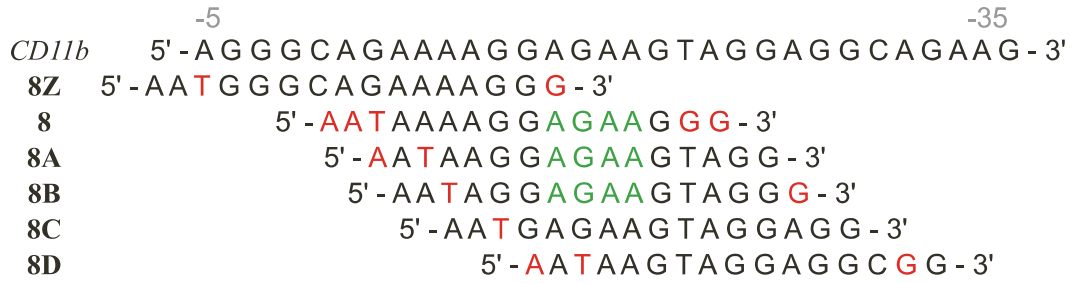
The DNA duplexes used for co-crystallization are based on a cassette as shown for **1H** in the following. The termini of the cassette consist of invariant overhangs (in black) that are critical for co-crystallization. Sequences shown in table S1 represent the 5´-GGAA-3´ strand of the variable portion of shown in blue. The only exceptions are structures **8** and **8B** in which the variable region is shifted; their DNA sequences are given in full.

```
5′ - A A T A A G C G G A A G T G G G - – 3'
3′ - - T A T T C G C C T T C A C C C T – 5'
```

Complex	Protein*	DNA sequence	Comment
1H	WT ΔN165	5′-AGCGGAAGTG-3′	High-affinity (CpG bearing)
1L	WT ΔN165	5′-AGCGGAATGG-3′	Low-affinity (CpG bearing)
1H(BT)	WT ΔN165	5′-AGCGGAAGTG-3′	**1H** co-crystallized in Bis-Tris, pH 5.4
1C	WT ΔN165	5′-ACCGGAAGTG-3′	DNA site is the Class I consensus motif
2H	WT ΔN165	5′-AGAGGAAGTG-3	High-affinity
2L	WT ΔN165	5′-AGAGGAATGG-3′	
3H	WT ΔN165	5′-AAAGGAAGTG-3′	λB motif of the Igλ2-4 enhancer
3L	WT ΔN165	5′-AAAGGAATGG-3′	Low-affinity
4H	Q226E	5′-AGCGGAAGTG-3′	Protein mutation as found in 1PUE
4H2	Q226E	5′-AAAGGAAGTG-3	λB motif
4L	Q226E	5′-AAAGGAATGG-3′	Protein mutation as found in 1PUE
4C	Q226E	5′-AGCGGAAGTG-3′	
5T	WT ΔN165	5′-AATGGAAGTG-3′	
5L	WT ΔN165	5′-AATGGAATGG-3	
5U	WT ΔN165	5′-AAUGGAATGG-3′	
6F	WT ΔN165	5′-AG^{5m}CGGAAGTG-3′ 5′-CACTTCCGCT-3′	Hemi-methylated **1H**, forward strand
6R	WT ΔN165	5′-AGGGAAGTG-3′ 5′-CACTTC^{5m}CGCT-3′	Hemi-methylated **1H** reverse strand
6D	WT ΔN165	5′-AG^{5m}CGGAAGTG-3′ 5′-CACTTC^{5m}CGCT-3′	Dimethylated **1H**
7	WT ΔN165	5′-AGCGGATGTG-3′	5′-GGAT-3′ core with **1H** flanks
8	WT ΔN165	5′-AATAAAAGGAGAAGGG-3′ 5’- TCCCTTCTCCTTTTAT-3'	PU.1 site in CD11B as reported by [S1]
8A	WT ΔN165	5′-AGGAGAAGTA-3′	
8B	WT ΔN165	5′-AATAGGAGAAGTAGGG-3′ 5’-TCCCTACTTCTCCTAT-3'	

* Human numbering of residues. For murine PU.1, +2 *e.g.*, ΔN167, Q228E. mPU.1 is only one-residue (Leu) different from human (His) at position 269, but the C-terminal IDR is not resolved by crystallography.

Table S3. Co-crystallographic search for the DNA-binding site in the *CD11b* promoter, related to Figure 6. Complementary strands of the co-crystallizing sequences contained a 5′ T overhang and terminate at the penultimate position of forward strands. Nucleotides in red deviate from the native sequence and are the result of maintaining the 5′ and 3′ termini of the crystallization cassette. Nucleotides in green represent the apparent core in co-crystallized structures. Sequences **8C**, **8D**, and **8Z** did not yield diffracting crystals under conditions that were amenable to **8**, **8A**, and **8B**. Note the directionality in the presentation of the purine strand (*right to left* towards the TSS).



SUPPLEMENTAL FIGURES

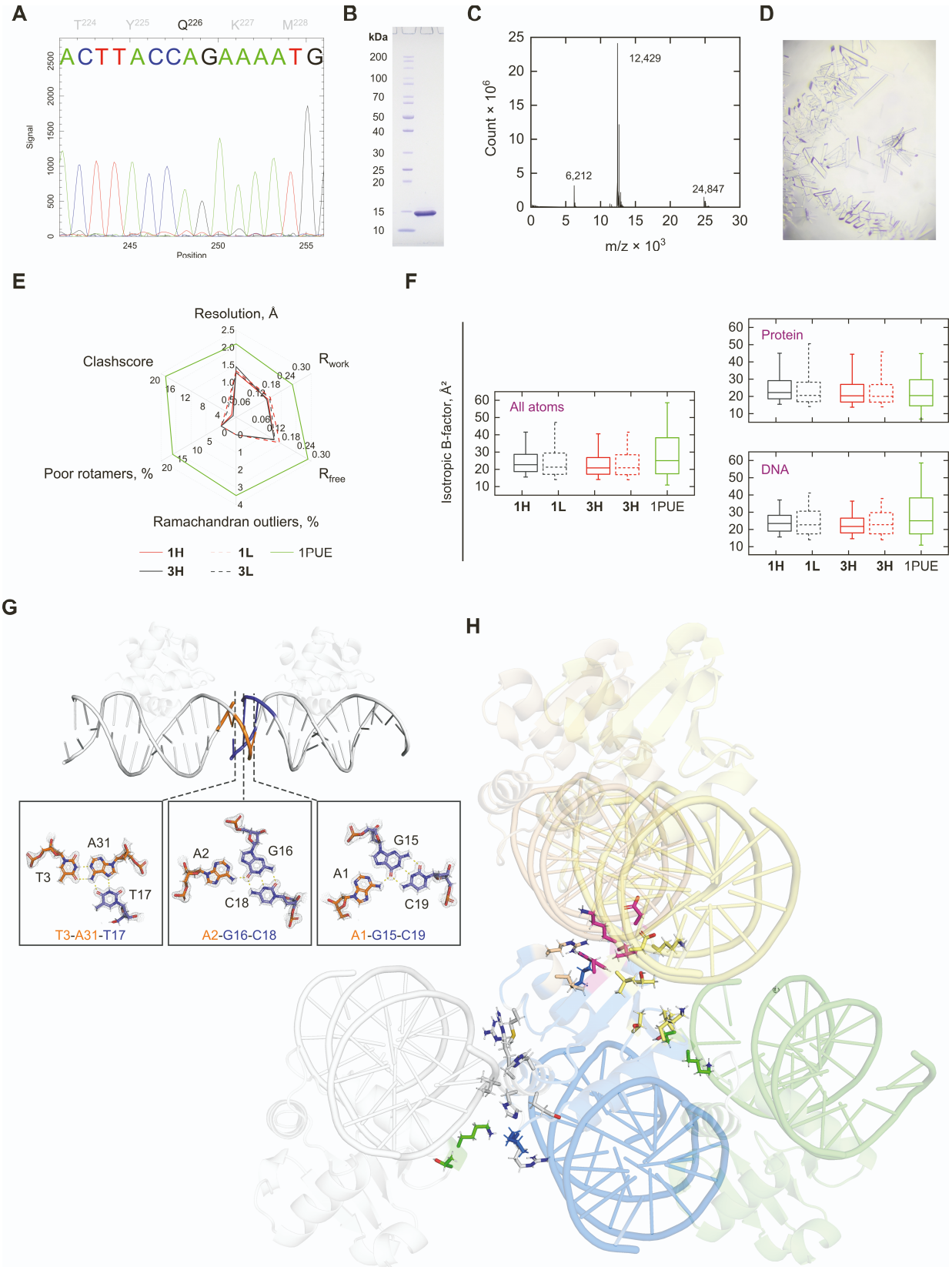

A
T224 Y225 Q226 K227 M228
ACTTACCAGAAAATG
Signal
Position
B
kDa
200
100
70
50
40
30
25
20
15
10
C
Count × 10^6
12,429
6,212
24,847
m/z × 10^3
D
E
Resolution, Å
Clashscore
R_work
Poor rotamers, %
R_free
Ramachandran outliers, %
1H
1L
1PUE
3H
3L
F
Isotropic B-factor, Å²
All atoms
Protein
DNA
1H 1L 3H 3H 1PUE
G
T3
A31
T17
T3-A31-T17
A2
G16
C18
A2-G16-C18
A1
G15
C19
A1-G15-C19
H

Figure S1. Characteristics of the high-resolution PU.1/DNA co-crystals, related to Figure 2 and STAR Methods

A, Sanger sequencing of the DNA construct encoding the hPU.1 ETS domain (ΔN165), focusing on the neighborhood of Q226.

B, Qualification of the recombinant ΔN165 used for co-crystallization with DNA: SDS-PAGE analysis of 100 ng of purified protein. The gel was stained with Coomassie Blue.

C, MALDI-ToF mass spectrometric analysis. The expected mass of ΔN165 was 12,406 Da. In addition to the primary peak, secondary peaks corresponding to a $z = +2$ ion and a gas-phase +1 dimer were observed.

D, Micrograph of co-crystals with high-affinity DNA, **1H**, at 1:1 ratio (400 μM).

E, Quality metrics of selected refined structures as computed by MolProbity. The corresponding values for 1PUE are shown for comparison. See also Table S2.

F, Distributions of atomic B-factors for all atoms or the specified subpopulations. Boxes represent the median ± 25/75 percentile and whiskers represent the 5/95 percentiles. For 1PUE, both complexes in the asymmetric unit are included.

G, Packing contacts in PU.1/DNA co-crystals. Shown is the **1H** complex; the others are homologous. DNA-DNA crystal contacts: Each triplex consisted of a WC pair from one complex interacting in the major groove with a third base from another complex. The distal base triplet exhibited canonical Hoogsteen base pairing of the AT*T type. Mesh represents the 2mFo-DFc density at 1.5 σ. A terminal protruding nucleotide (T32) was displaced and unresolved.

H, Crystal contacts involving protein. Residues form crystal contacts with adjacent asymmetric units, defined by a cutoff of 4.5 Å from a symmetry neighbor. The blue unit is taken as the reference. Residues engaged in crystal contacts are colored according to the contacted neighbor; magenta residues are located within the cutoff for two neighboring complexes. The green residues in the white unit represent the corresponding ones contacting the green unit.

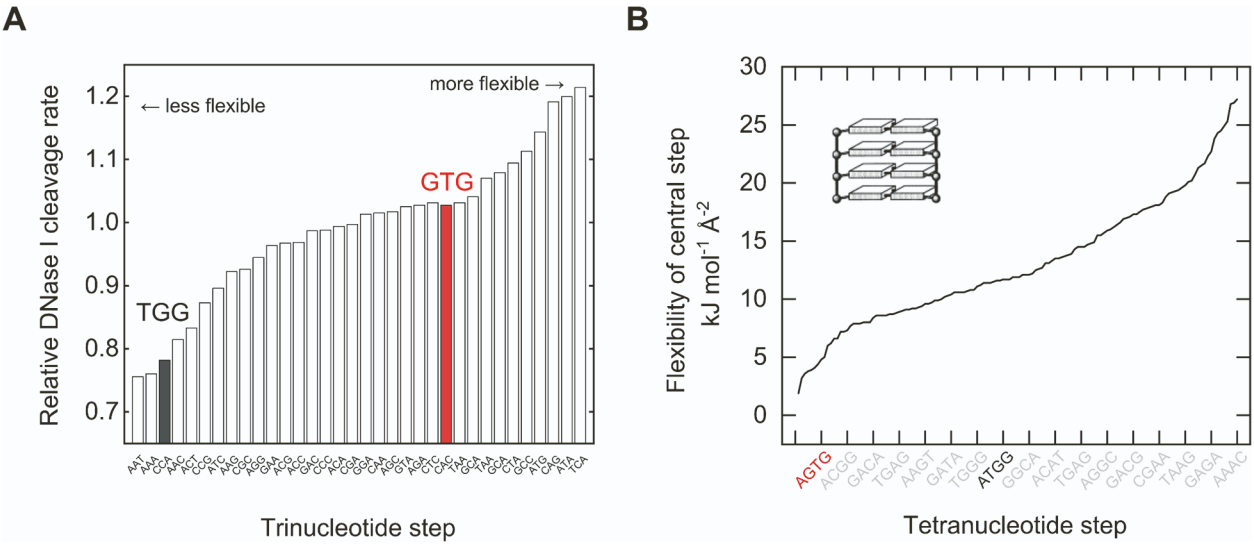


Figure S2. GTG is an intrinsically more flexible base step than TGG: evidence from the literature, related to Figure 2.

A, Experimental DNase I cleavage data as a function of all 34 unique trinucleotide steps, in order from least to most flexible [S2]. DNase I is a minor-groove binding enzyme that prefers and widens flexible and minor grooves to catalyze hydrolysis. The differential reactivity of GTG to TGG to DNase I cleavage is +1.3-fold.

B, Flexibility from molecular mechanics calculations for all 136 unique tetranucleotide steps, defined as the curvature in the potential energy surface with respect to slide of the central base step at the global energy minimum [S3]. Slide is a major translational parameter that is coupled to angular roll variations in protein-induced helical bending [S4]. The differential flexibility of GTG over TGG is -7 kJ mol^{-1} $Å^{-2}$. Every 8^{th} tetranucleotide step is shown along the abscissa.

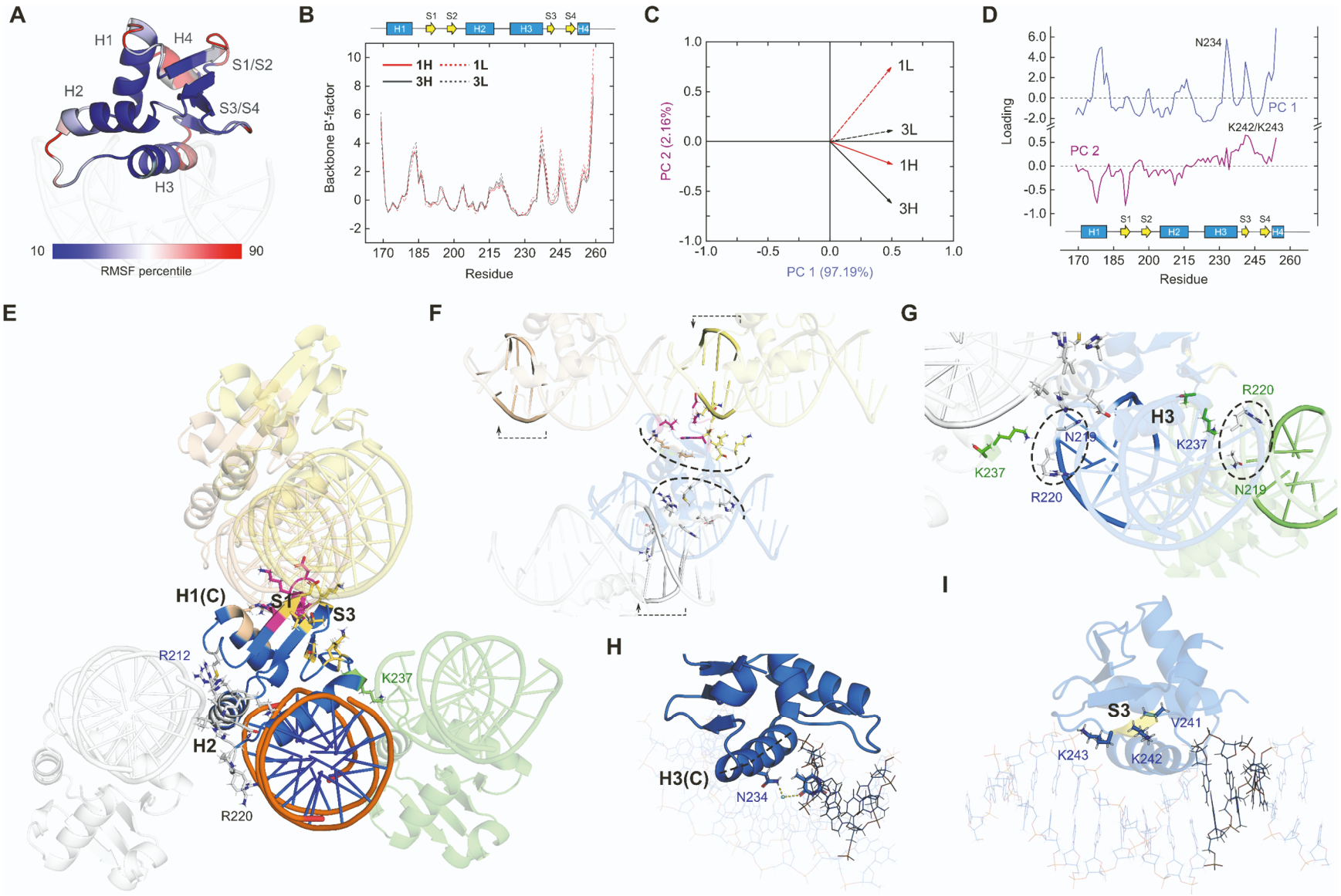

A
H1
H4
S1/S2
H2
S3/S4
H3
10
90
RMSF percentile
B
S1
S2
S3
S4
H1
H2
H3
H4
1H
1L
3H
3L
Backbone B'-factor
Residue
C
1L
3L
1H
3H
PC 2 (2.16%)
PC 1 (97.19%)
D
N234
PC 1
K242/K243
PC 2
Loading
Residue
E
H1(C)
S1
S3
R212
K237
H2
R220
F
G
R220
H3
N219
K237
K237
R220
N219
H
H3(C)
N234
I
S3
V241
K243
K242

Figure S3. Analysis of the backbone B-factors in DNA-bound ΔN165, related to Figure 3.

A, Overlaid (*z*-normalized) B′-factor profiles of the protein backbone in high- and low-affinity DNA complexes.

B, The implied dynamics were reproduced in the **1H** complex by molecular dynamics simulations, colored from the 10^{th} and 90^{th} percentile in backbone RMSF. See *Supplementary Methods* for methodological details.

C, Principal component analysis of B′-factor profiles, excluding disordered residues at the termini. The dominant principal component PC 1 captures the overall trend of the profiles. The next principal component PC 2 qualitatively distinguishes the high- and low-affinity complexes.

D, Per-residue contribution (loading) to the basis profiles corresponding to PC 1 and PC 2.

E, Mapping of residues in **1H** forming crystal contacts within 3.5 Å from adjacent asymmetric units. Residues are colored according to the contacted neighbor, or magenta if within the cutoff for two neighboring complexes. Most of the residues with significant loading in PC 2 (positive or negative) are among these crystal contacts.

F, N-terminally to the recognition helix H3, crystal contacts are clustered near 3′-flanking DNA bases (opaque, with dashed arrows to indicate polarity) in neighboring units (bracketed for clarity).

G, The lone high-loading crystal contact with protein is K237 at the C-terminus of H3. K237 contacts N219 and R220 which are involved in DNA contacts with a neighboring complex. Charge-charge repulsion with R220 results in low sidechain occupancy for K237.

H, N234 near the 3′-flanking DNA (opaque sticks) shows the highest B′ -factor for an internal residue but is not a crystal contact. It interacts only through water, most closely with the final T in the core consensus. At this position, H3 transitions from an α- (*i* to *i*+4) to a π-helix (*i* to *i*+3), bending the helix.

I, C-terminally to H3, S3 (yellow) represents the major high-loading element in PC 2 that does not make crystal contacts. K243 makes ionic interactions with 5′-flanking DNA. Opaque sticks represent 3′-flanking DNA.

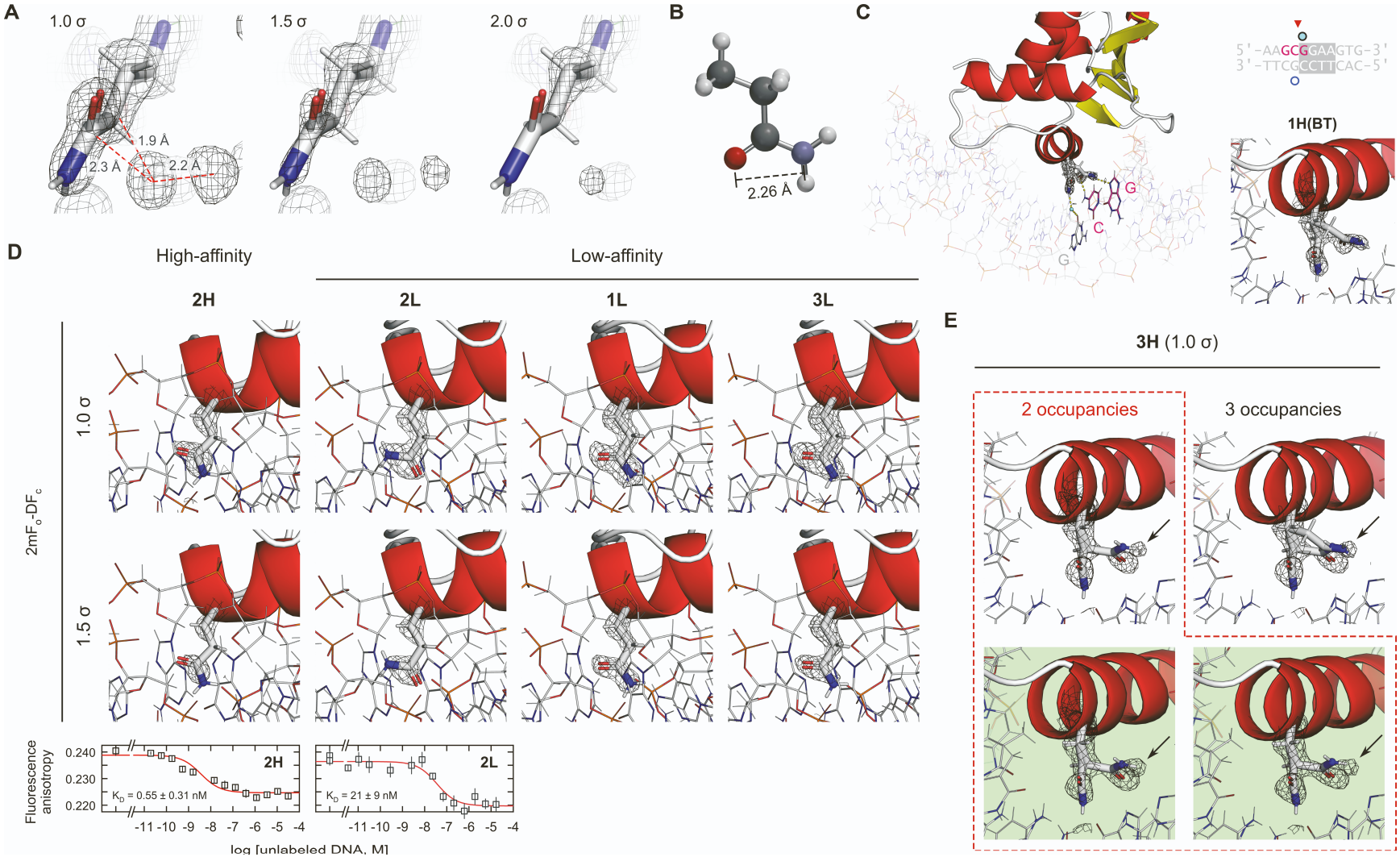

A
1.0 σ
1.5 σ
2.0 σ
1.9 Å
2.3 Å
2.2 Å
B
2.26 Å
C
5'-AAGCGGAAGTG-3'
3'-TTCGCCTTCAC-5'
1H(BT)
G
C
G
D
High-affinity
Low-affinity
2H
2L
1L
3L
1.0 σ
1.5 σ
2mF$_o$-DF$_c$
Fluorescence anisotropy
0.240
0.230
0.220
2H
K$_D$ = 0.55 ± 0.31 nM
2L
K$_D$ = 21 ± 9 nM
-11 -10 -9 -8 -7 -6 -5 -4
log [unlabeled DNA, M]
E
3H (1.0 σ)
2 occupancies
3 occupancies

Figure S4. Heterogeneity in the discrete electron densities around Q226 in high-affinity complexes, related to Figure 3

A, The 2mFo-DFc map around Q226 in the **1H** complex at the indicated σ cutoffs. The marked distances are too short for two oxygen atoms in H-bonded water.

B, The excess densities were consistent with the geometry of the heteroatoms of a glutamine sidechain, as computed by a quantum-mechanical calculation. Shown is the ground-state geometry of propionamide optimized at the MP2/aug-CC-pVTZ level using Gaussian 16 (Gaussian).

C, To formally eliminate the possibility of a non-bonded component such as acetate from the crystallization solution, a **1H** complex was crystallized in Bis-Tris buffer at pH 5.4. In the absence of acetate, Q226 still exhibited discrete densities (1.0 σ shown) in the same manner.

D, Q226 sidechain occupancies and binding affinities in **2H** and **2L**. In the high-affinity complex 2H, the very weak electron density around the Q226 sidechain indicates lower occupancy of the modeled down conformation than other high-affinity complexes **1H** and **3H**. The absence of disconnected densities corresponding to alternate conformations suggest that they are manifold, so only the down conformation was modeled. The same 3′-flanking variation (GTG → TGG) that generates full down Q226 occupancy in **1L** and **3L** appears to have less effect in **2L**, with correspondingly (~5-fold) lower selectivity for **2H**. Points represent mean ± S.D. of triplicate experiments.

E, The excess electron densities around Q226 in **3H** suggests additional up occupancies. With only a single up occupancy, one of the excess densities (arrows) is offset from the amide-N. This density is unlikely the amide-H, a light atom, as evidenced by the absence of density for the corresponding H atom in the down conformation at the same 1.0 σ cutoff. Fitting a second up conformation resolves this inconsistency, but the lack of bridging density for the second up conformation does not improve *R*-values globally (beyond 0.5%). Replicate **3H** crystals (with shaded background) at similar resolution and *R*-values demonstrate the reproducibility of the offset densities.

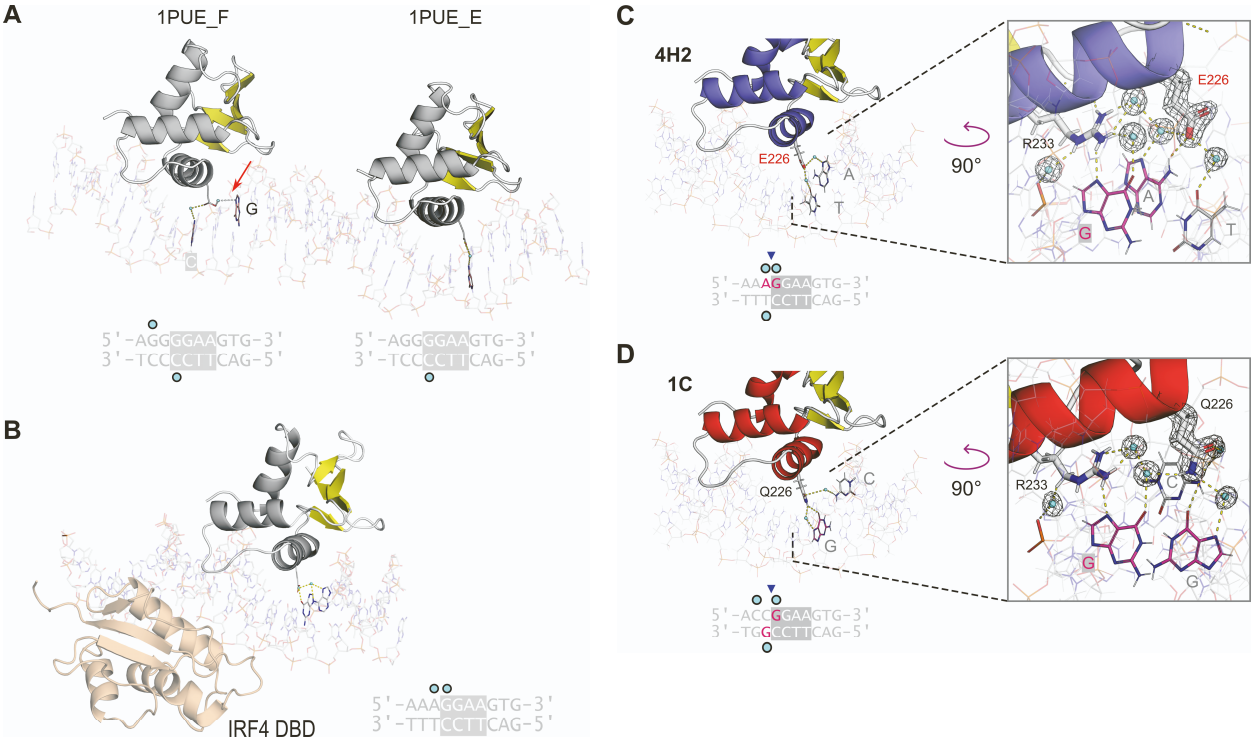


Figure S5. Water acts as H-bond adapters in Q226E and sub-optimal wildtype PU.1/DNA complexes, related to Figure 4

A, Sidechain conformations of E228 in the extant co-crystal structure of murine PU.1, 1PUE. 1PUE contains two complexes per asymmetric unit. E228 (murine numbering) are both in down conformations and make strictly water-mediated contacts as shown. The water bridge in one of the complexes (chain F) to N7 of G6 is marked with an arrow. It is a relatively long H-bond at 3.4 Å (O···N7) distance.

B, Ternary complex with the λB motif and IRF4 (colored in beige). E228, also in a down conformation, makes water-mediated contacts with the 5′-GGAA-3′ strand but fails to contact G6 (underlined). Coordinates courtesy of Drs. C. Escalante and A. Aggarwal [S5].

C, The fully occupied down conformation of E226 in **4H2** cannot H-bond with N7 of A_{-2} (both H-bond acceptors) in the λB motif and misses that contact entirely. The observed contact is with A_{-1}. The E226 sidechain in this complex is farther from R223 than in the other Q226E structures. As a result, the connecting water molecule that links E226 with R223 in the other Q226E structures is connected by a minimum of two water molecules. 2mFo-DFc maps are rendered at a level of 1.0 σ selected sidechains in addition to well-ordered water near that sidechain.

D, In the wildtype complex **1C**, Q226 cannot H-bond with the exocyclic NH_2 of C (both H-bond donors) at the -2 position, and the contact is again complemented by a bridging water. Coupling with R233 is achieved by at least two bridging water molecules as in **4H2**.

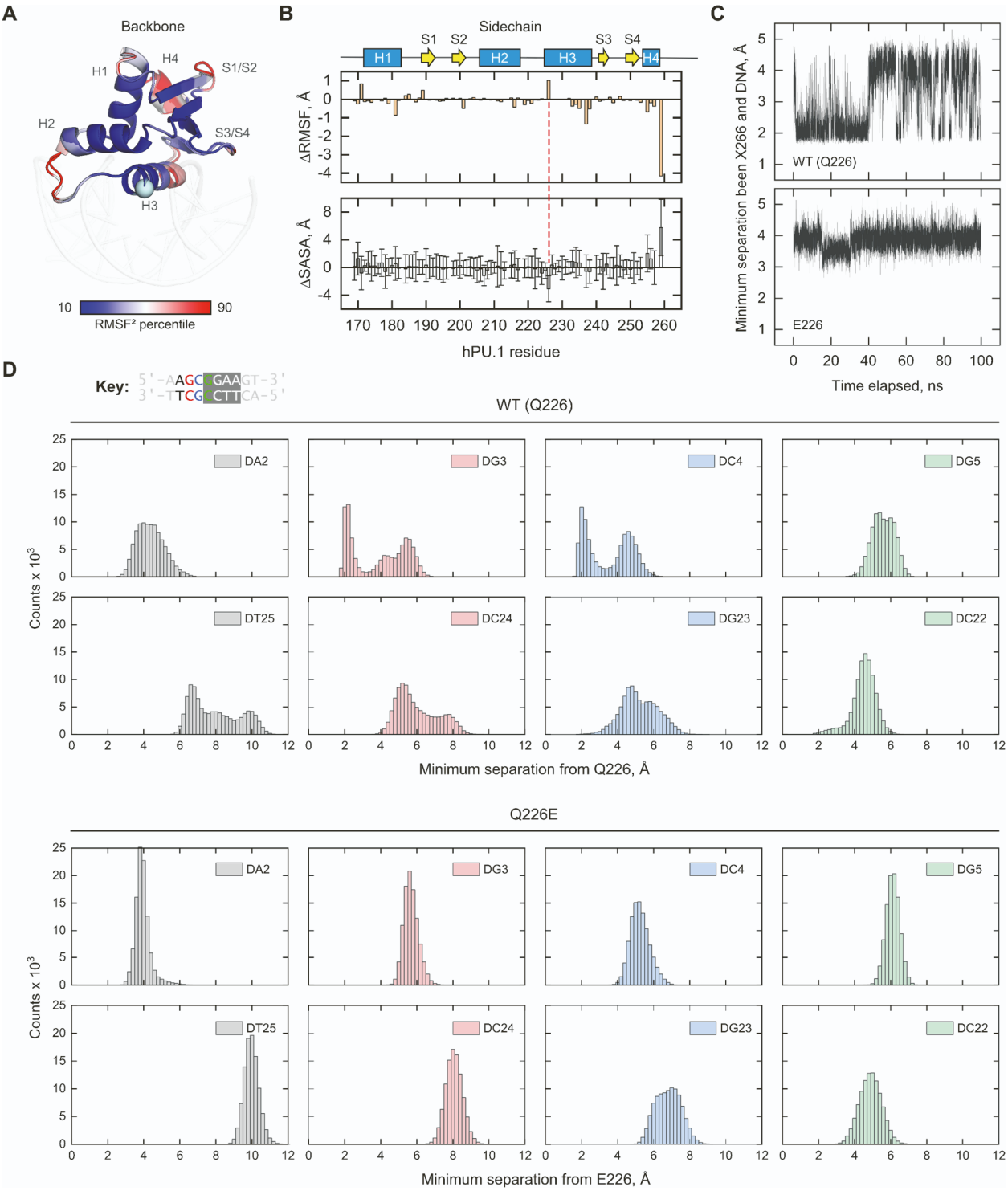


Figure S6. Molecular dynamics of the wildtype and Q226E ΔN165 complexed with high-affinity DNA, related to Figure 4. Backbone and sidechain dynamics were evaluated separately from triplicate equilibrated trajectories.

A, Backbone dynamics in terms of per-residue RMS fluctuation (RMSF) showed negligible relative differences between the two complexes. Position 226 scored at the bottom quartile of backbone RMSF[2] regardless of residue identity.

B, Comparison of sidechain dynamics RMSF (heavy atoms only) and solvent-exposed surface area (SASA), expressed as the difference of wildtype minus Q226E. RMSF values computed from the concatenated trajectories of triplicate simulations. ΔSASA values were computed from time-averages ± SD. The red dashes indicate position 226. Excluding the terminal residues, sidechain RMSF differences were not correlated with altered solvent exposure, with the notable exception at the mutated position 226. The wildtype Gln residue at this position was significantly more excluded from solvent (ΔSASA < 0) while exhibiting increased conformational dynamics.

C, Minimum separation from the residues at position 226 to DNA over an illustrative 100-ns period.

D, Minimum separation distances from the indicated DNA base positions for the trajectory shown in Panel C. The time trajectories clearly showed a significantly broader ensemble of DNA contacts by the wildtype residue relative to E226.

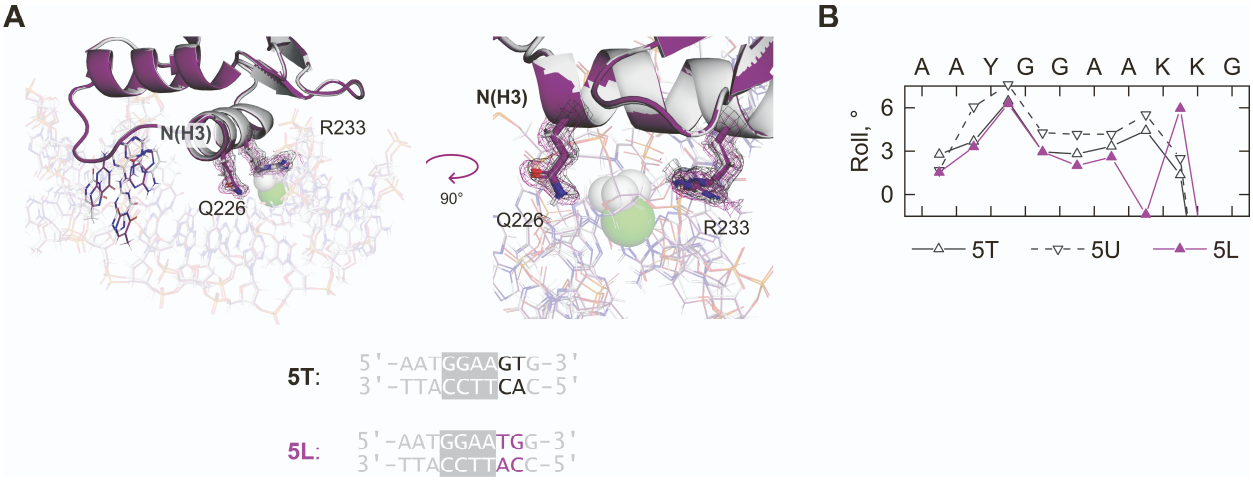


Figure S7. Steric disruption the Q226-R233 couple by T at the -1 position, related to Figure 5.

A, The 5-methyl substituent (spheres) in T at the -1 position sterically displaces the sidechains of Q226 and R233. The two structures are aligned by the protein, ΔN165. Electron density (2mFo-DFc) maps are rendered at 1.0 σ.

B, Sequence variation at the 3′-flanking region as manifest in the roll angles, which perturbs Q226 in structures without T_{-1}, but now exerts no significant effect. The profile for **5U** is included for comparison.

SUPPLEMENTAL REFERENCES

[S1] H.L. Pahl, R.J. Scheibe, D.E. Zhang, H.M. Chen, D.L. Galson, R.A. Maki, D.G. Tenen, The proto-oncogene PU.1 regulates expression of the myeloid-specific CD11b promoter, J Biol Chem, 268 (1993) 5014-5020. 10.1016/S0021-9258(18)53496-7.

[S2] I. Brukner, R. Sanchez, D. Suck, S. Pongor, Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides, EMBO J, 14 (1995) 1812-1818. 10.1002/j.1460-2075.1995.tb07169.x.

[S3] M.J. Packer, M.P. Dauncey, C.A. Hunter, Sequence-dependent DNA structure: tetranucleotide conformational maps, J Mol Biol, 295 (2000) 85-103. 10.1006/jmbi.1999.3237.

[S4] W.K. Olson, A.A. Gorin, X.J. Lu, L.M. Hock, V.B. Zhurkin, DNA sequence-dependent deformability deduced from protein-DNA crystal complexes, Proc Natl Acad Sci U S A, 95 (1998) 11163-11168. 10.1073/pnas.95.19.11163.

[S5] C.R. Escalante, A.L. Brass, J.M. Pongubala, E. Shatova, L. Shen, H. Singh, A.K. Aggarwal, Crystal structure of PU.1/IRF-4/DNA ternary complex, Mol Cell, 10 (2002) 1097-1105. 10.1016/s1097-2765(02)00703-7.