5-4-2020

Characterization of self-regulatory mechanisms and internal dynamics of ETS transcription factor PU.1

Shingo Esaki

CHARACTERIZATION OF SELF-REGULATORY MECHANISMS AND INTERNAL DYNAMICS OF ETS TRANSCRIPTION FACTOR PU.1

by

SHINGO ESAKI

Under the Direction of Gregory M. K. Poon, PhD

ABSTRACT

The ETS family of transcription factors bind to site-specific DNA via DNA-binding domains called the ETS domains. The ETS domains are structurally homologous but divergent in primary sequences. PU.1 is an essential transcription factor and its biological activity is primarily controlled by up- and down-regulation of its expression. Aside from down-regulated expression, only a few inhibitory mechanisms are known for PU.1. The most understood one involves PU.1 forming a heterodimer with other protein partners, such as GATA-1. However, unlike auto-inhibited ETS-family members whose activity is regulated by autoinhibitory elements that reduce the net affinity of binding to specific DNA, PU.1 has no such regulatory mechanism at the protein-DNA level. We report here that PU.1, unlike its auto-inhibited paralog Ets-1, forms a

2:1 complex with site-specific DNA (>10 bp) in a negatively cooperative manner. We also detected potential interface (^{193}DKDK^{196}) of the PU.1 dimer by using heteronuclear single quantum correlation (HSQC) NMR. Self-titration of PU.1 is a negative feedback mechanism at the protein-DNA level. Following these findings, our group found the presence of the IDRs flanking the ETS domain does not change the DNA binding modes of the PU.1 ETS domain, yet the PEST domain modifies DNA recognition by the ETS domain through changing DNA binding affinities. We successfully assigned ~90% or more backbone amide resonances in the ^{1}H-^{15}N HSQC spectra of hPU.1 constructs with and without IDRs, in the absence and presence (1:1 complex) of DNA. Using the fully assigned HSQC spectra, we studied fast (ps to ns) time scale internal dynamics of PU.1 protein. Spin relaxation rates and heteronuclear ^{1}H{^{15}N}-NOE were acquired for the hPU.1 proteins with and without DNA by NMR. We demonstrated that the PEST domain remains disordered but becomes more dynamic upon specific DNA binding. In terms of DNA recognition, the presence of the PEST domain increases the affinity of 1:1 complex of the ETS domain with cognate DNA, without perturbing the structure or changing the fast time scale backbone motions of the ETS domain.

INDEX WORDS: Transcription factors, ETS family, PU.1, Ets-1, autoinhibition, Dimer, Dimeric interface, Protein-protein interactions, Nucleic acids, Site-specific DNA, Nonspecific DNA, Nuclear magnetic resonance, Protein dynamics

CHARACTERIZATION OF SELF-REGULATORY MECHANISMS AND INTERNAL DYNAMICS OF ETS TRANSCRIPTION FACTOR PU.1

by

SHINGO ESAKI

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2020

CHARACTERIZATION OF SELF-REGULATORY MECHANISMS AND INTERNAL DYNAMICS OF ETS TRANSCRIPTION FACTOR PU.1

by

SHINGO ESAKI

Committee Chair: Gregory M. K. Poon

Committee: W. David Wilson

Markus W. Germann

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2020

DEDICATION

I dedicate my dissertation work to my parents and sisters. Without your love, support, and encouragement, I would not have achieved this.

ACKNOWLEDGEMENTS

All the work for these projects was carried out under the supervision of Dr. Gregory M. K. Poon. Through Dr. Poon's guidance and insight, I have gained valuable knowledge and experience. I cannot express my gratitude enough for working and learning in this laboratory.

I thank my committee members Dr. W. David Wilson, Dr. Markus Germann, and (former committer member) Dr. Zhen Huang for the insight and all your help during my studies at Georgia State University.

I would like to thank all my former and present lab members: Suela Xhani, Noa Erlitzki, Dr. Sangchoon Lee, Amanda Albrecht, Kenneth Huang, Hye-Mi Kim, Van Ha, and Dominique Stephens for your assistance and input.

I am also grateful to my collaborators: Dr. Markus Germann and his former graduate student Dr. Marina Evich. I appreciate all the time and effort both of you have devoted to teaching me the knowledge and skills of NMR.

Additionally, I would like to thank my former adviser Dr. Aimin Liu and the lab members. I spent the first three years of my Ph.D. in Liu lab before Dr. Liu's departure from Georgia State University. I thank Dr. Liu for providing me with learning opportunities in various areas of biochemistry. I also gained wonderful memories and precious friendships with lab mates.

And finally, I would like to thank the Molecular Basis of Disease program at Georgia State University for fellowship support for ~4.5 years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

TF, transcription factor; PPI, protein-protein interaction; ITC, isothermal titration calorimetry; EMSA, electrophoretic mobility shift assay; DOSY, diffusion-ordered spectroscopy; HSQC, heteronuclear single quantum coherence; ANS, 8-anilinonaphthalene-1-sulfonate; PDB, Protein Data Bank; IDRs: intrinsically disordered regions

1 INTRODUCTION

1.1 Eukaryotic gene transcription overview

RNA polymerase needs to bind to a promoter sequence, and transcription factors need to bind to enhancer sequences for the initiation of transcription on an opened chromatin template (1). Transcription factors are the proteins necessary for the initiation of transcription but not a part of RNA polymerase. Many roles of transcription factors are known: recognition of (i) *cis*-acting elements of DNA, (ii) other transcription factors, and (iii) RNA polymerase, by forming an initiation complex (1). The mechanism of transcription in eukaryotes is quite different from that in prokaryotes. Prokaryotic transcription occurs on a DNA template, while eukaryotic transcription occurs on a chromatin template (1). Prokaryotic RNA polymerase reads DNA sequences and binds to promoters, but eukaryotic RNA polymerase cannot do the same (1). This is the reason why many eukaryotic transcription factors need to bind to *cis*-acting sites before RNA polymerase binds to DNA. These transcription factors are called basal transcription factors and form a DNA complex, to which RNA polymerase binds to initiate transcription (1).

Only a single RNA polymerase is known for prokaryotes, but three types of RNA polymerase occur in eukaryotes: (i) RNA polymerase I that transcribes 18S/28S rRNA, (ii) RNA polymerase II that transcribes mRNA and a few small RNAs, and (iii) RNA polymerase III that transcribes tRNA, 5S ribosomal RNA, and some small RNAs (1). Basal transcription factors form a complex with DNA at promoters for all three types of RNA polymerase. The structure of the transcription factor/DNA complexes for RNA polymerase I and III is simple, but the one for RNA polymerase II is huge (1). The structure formed by basal transcription factors and RNA polymerase are called basal transcription apparatus (1). The promoters typically lie upstream of

the start point for RNA polymerases I and II, while they are located downstream of the start point for most of RNA polymerase III (1).

RNA polymerases and thousands of proteins, including basal and general transcription factors (TFs), are responsible for gene activation and repression (1). Transcription reactions proceed through three stages: initiation, elongation, and termination. Binding of $TF_{II}D$ to the TATA box or Inr is the first step in initiation (Fig. 1.1). Other TFs subsequently bind to the initiation complex in a defined order. When RNA polymerase II binds to the complex, it initiates transcription. Binding of $TF_{II}E$ and $TF_{II}H$ enables to melt DNA and allow polymerase movement. Initiation is followed by promoter clearance and elongation, which requires phosphorylation of the carboxy-terminal domain (CTD). During termination, synthesized mRNA is released, and RNA polymerase II dissociates from the template DNA.

1.2 Interactions that contribute to DNA recognition by transcription factors

Proteins use similar strategies to recognize nucleic acids (2). The general principles to recognize cognate sites by transcription factors are the same: it is based on the sequence and structure (2). One of the forces that are involved in noncovalent protein-nucleic acid complex formation and contribute most is electrostatic one (2). Nucleic acids are polyanions, and DNA binding domains of transcription factors are typically positively charged due to an abundance of lysine and arginine residues in them (2). Many other weak forces for interactions are also involved in the protein-nucleic acid complex formation, such as hydrophobic and polar ones (direct and water-mediated hydrogen bonds) (3). Therefore, the sum of weak interactions drives complex formation (3).

1.3 helix-turn-helix and winged helix-turn-helix transcription factors

The helix-turn-helix (HTH) is a common motif for DNA recognition in prokaryotes and bacteriophage (2). DNA binding proteins such as bacteriophage λ-repressor and bacterial *Trp* repressor were the first HTH transcription factors (TFs) that were characterized biologically and structurally in prokaryotes and bacteriophage (2). Interestingly, HTH TFs in prokaryotes and bacteriophage are homodimers in general, while eukaryotic HTH TFs are monomeric (2). HTH motif is typically comprised of three core helices that form a right-handed bundle with a tight turn between helix #2 and #3 (so-called H2 and H3) (Fig. 1.2 *A*). A hydrophobic core at the interface of the three helices stabilizes the overall structure as well as serves to present the DNA recognition helix (helix #3; H3). Upon specific DNA binding, the DNA recognition helix is inserted into the DNA major groove, where the H3 sidechains specifically contact both nucleotide bases and sugar-phosphate backbone. Helix #1 (so-called H1) and the turn between H2 and H3 also contact DNA. A variety of orientations of H3 with the DNA major groove are known among HTH TFs, thereby different regions of the DNA recognition helix serve for specific DNA binding (2).

Winged helix-turn-helix (wHTH) motif is a variant of the HTH motif and belongs to Winged helix DNA-binding domain superfamily (EBI entry: IPR036390) (2). The wHTH motif/domain contains one or two wings (W1-2), three α-helices (H1-3), and three β-sheets (S1-3). A typical wHTH domain consists of the three helices bundled from the HTH motif and an additional antiparallel β-sheet located adjacent to the HTH motif and over the DNA minor groove (Fig. 1.2 *B*). This β-sheet makes an additional DNA backbone contact. Many proteins in this superfamily contain a second wing which is comprised of the turn between H2 and H3 or resides by N- or C-terminal extensions to the wHTH domain (2).

The prototypical wHTH motif is seen in HNF-3 protein (2). The structure of 1:1 bound HNF-3γ/DNA complex shows that α-helix #3 (H3) fits well into the major groove of B-form DNA, which mediates sequence-specific DNA contacts, and that two wings (the loop between S1 and S2; the loop after S3 to the C-terminus) mediates DNA backbone contacts to the flanking minor grooves (2).

1.4 ETS Transcription factors

1.4.1 ETS Transcription factors overview

The ETS (E twenty-six) family of transcription factors were originally identified as Ets and Myb genes transduced by E26 virus (4). They are found from sponges to humans (i.e., throughout Metazoa), and 28 human genes are known (5). All family members bind to site-specific DNA via structurally conserved DNA-binding domain called "ETS domain" that consists of ~85 amino acids and exhibits the winged helix-turn-helix (wHTH) motif. ETS TFs bind to specific purine-rich DNA sequences with a consensus sequence (core motif: 5'-GGA(A/T)-3') for ETS proteins (6). And further specificities for each family member are given by the flanking sequences of the 5'-GGA(A/T)-3' core. The ETS domains are structurally similar to each other, but their primary sequences are highly different from one another.

1.4.2 Combinational regulation of ETS Transcription factors

ETS proteins form heterodimers with other transcriptional regulators either through the ETS domain or through regions outside the ETS domain. The dimer formation reinforces site-specific binding to DNA due to so-called combinational regulation (*cf.* Chapter 1.8.1 for details).

The ETS family members SAP-1 and Elk-1 cooperatively bind to the *c-fos* promoter with the MADS-box serum response factor (SRF) (7). Both SAP-1 and Elk-1 have a sequence of 20 amino acids called B-box, which is required for interaction with SRF (7). The cooperativity of

SAP and Elk-1 requires an interaction between the B-box and the DNA binding domain of SRF (7). Direct interactions of the SAP/Elk-1 DNA binding domain with the SRF are also required for DNA recognition by the complex (7). SAP-1 binds to *c-fos* sites efficiently in the absence of SRF, whereas Elk-1 does not. Then, SRF is thought to modify the DNA binding properties of Elf-1. The structure of the ternary SAP-1/SRF/DNA complex was solved (PDB 1HBX; Fig. 1.3 *A*) (8). A direct interaction between the DNA recognition helices of SRF and SAP-1 is visible in this structure. Tyr^{65} that is conserved in almost all the ETS family plays a key role in mediating the extensive interaction (8). SRF reorients the Tyr^{65} residue of SAP-1 for optimal DNA contacts to the GGAA core sequence. Structural comparison of the SAP-1/SRF/DNA complex with site-specific 1:1 Elk-1/DNA complex revealed that the conserved Tyr residue is oriented to prevent interactions with the GGAA consensus sequence. This explains why nascent Elk-1 cannot bind to *c-fos* promoter sites in the absence of SRF (9). Modeling studies of Elk-1/SRF complex with a *c-fos* promoter site further suggested that the conserved Tyr residue in Elk-1 is reoriented to make similar interactions with SAP-1: namely, interactions between Elk-1 and SRF, unless otherwise it is a low-affinity DNA site for Elk-1 (10).

Another example of combinational regulation in the ETS family is the ternary Ets-1/Pax5/DNA complex (PDB 1K78; Fig. 1.3 *B*) (11). The affinity of Ets-1 for the *mb-1* promoter is low in the absence of Pax5, and Pax5 selectively recruits Ets-1 to the promoter. The structure of the Ets-1/Pax5/DNA complex has been compared with that of site-specific 1:1 Ets-1/DNA complex. Then, the side-chain interactions at the Ets-1/Pax5 interface have been found to reorient a conserved Tyr residue (Tyr^{395}) in the DNA recognition helix for optimal DNA contacts. Taken together, a conserved feature of such ternary complexes in the ETS family is the reorientation of residue(s) in the DNA binding domain for optimal DNA contacts.

1.5 Biological roles of ETS Transcription factors

1.5.1 Overview of biological roles of ETS proteins

ETS proteins transcriptionally regulate many viral and cellular genes (12). ETS TFs control gene expression which is important for biological processes such as cellular proliferation and differentiation, cell cycle regulation, cell signaling, hematopoiesis, apoptosis, and metastasis (12). ETS proteins are an important family of transcription factors, thereby aberrant activity of ETS TFs has been found to be associated with a lot of diseases. One of those diseases is cancer that results from the loss of cellular homeostasis, i.e., the balance between cellular proliferation and cell death. And many oncogenes are regulated by ETS target genes (13). Human cancer associated with the activity of ETS TFs includes breast cancer, prostate cancer, leukemia/lymphoma, and Ewing's sarcoma, etc. (13).

1.5.2 Biological roles of PU.1

The transcription factor PU.1 was discovered by Moreau-Gachelin *et al*. in 1988 (14). They reported it is the product of an upregulated gene in murine erythroleukemia, due to proviral integration of the spleen focus forming virus (SFFV) (14). This gene was named *SFFV proviral integration site 1* (*Sfpi1*) in mice and *SPI1* in humans. Klemsz *et al*. isolated the cDNA of the gene coding for an ETS transcription factor and named it PU box binding-1 (PU.1) (9).

PU.1 is an essential transcription factor, and its main biological role is the development of hematopoietic stem cells (HSCs) in the immune system (15). PU.1 is a central transcriptional regulator of HSCs differentiation into lymphocytes and myelocytes, B and T cell development, and HSCs maintenance (16). This function spans from early to late stages of progression in a lineage- and cell type-specific manner; thus, it controls proliferation, terminal differentiation, and maintenance of HSCs (17). Therefore, PU.1 is a key transcriptional regulator within the

hematopoietic system and plays critical roles in both the innate and adaptive immune systems by controlling cell differentiation.

Many target genes of PU.1 have been described since its discovery. Turkistany *et al*. studied the identity of target genes from the published literature (18). Because PU.1's primary role is transcriptional activation, they sorted PU.1 target genes into four criteria. (i) The genes are activated by PU.1 as seen by the changes in mRNA levels in response to PU.1 expression. (ii) The genes have at least one PU.1 binding site that contains the consensus sequence 5'-GGAA-3' (with the exception of 5'-AGAA-3'). (iii) Transient transfection experiments in vitro and/or mutational studies of the predicted binding site(s) can demonstrate that the genes are transcriptionally activated by PU.1. (iv) Site-specific PU.1 binding demonstrated by EMSA and/or ChIP. Thus, they identified 110 PU.1-activated genes based on these criteria. They also found the subcellular location of the 110 target gene products by subsequent bioinformatics analysis (18): 22 in the nucleus; 21 in the cytoplasm; 44 in the plasma membrane; and 23 in extracellular space. Thus, 61% (67 of 110) of the gene products regulated by PU.1 reside in the plasma membrane or are secreted. Therefore, PU.1 plays an important role in regulating cellular communication. Several cytoplasmic proteins regulated by PU.1 also mediate intracellular signaling downstream of plasma membrane proteins such as BTK. Most PU.1-activated nuclear proteins such as GATA-1 and IRF4 are important transcription factors. Thus, PU.1 possibly controls downstream gene networks.

Furthermore, PU.1 activates transcription of genes coding for antibodies, antibody receptors, cytokines such as interleukin 3 (IL-3), cytokine receptors, chemokines, chemokine receptors, and integrins (18).

1.5.3 PU.1 protein and diseases

Deregulation of PU.1 activity has been linked to at least three human diseases: rheumatism, Alzheimer's disease, and hematologic cancers. Thus, PU.1 not only works as an indispensable regulator of normal HSCs but also has pathogenic functions in the hematopoietic immune system. Genome-wide analysis of epigenomic elements by Dozmorov *et al*. provided statistical evidence for PU.1 as a transcriptional regulator of genes associated with rheumatism (19). Gjoneska *et al*. recently reported the upregulation of PU.1 expression in Alzheimer's disease (20).

PU.1 is known as a tumor suppressor in myeloid cells. Inactivating mutations of the *SPI1* gene, which codes for PU.1, have been identified in patients of acute myeloid leukemia (AML) (21). Also, the expression of PU.1 is often suppressed in AML (22). Recent studies have shown that minimal PU.1 expression reduction (35%) is sufficient to induce preleukemic stem cells, which leads to transformation to AML (23). The downregulation of PU.1 expression is also associated with myeloma and classical Hodgkin disease (24,25).

1.6 Structures of PU.1 ETS domain

ETS subfamily belongs to the winged helix-turn-helix (wHTH) superfamily as mentioned above. The ETS domain (DNA binding domain) is the only ordered region for PU.1 protein (Fig. 1.4). Therefore, PU.1 is a Type I transcription factor. Two structures have been determined for the PU.1 ETS domain. One is a co-crystal structure of PU.1 ETS domain and 16-bp high-affinity DNA (PDB: 1PUE) (26). The other is a solution NMR structure of PU.1 ETS domain and all the C-terminal residues, in the absence of DNA (PDB: 5W3G). Interestingly, these structures are almost identical to one another (Fig. 1.5, *A*-*C*). ETS domains in their family are structurally similar to each other, although their primary sequences are mutually far different. For example,

the tertiary structure of the ETS domain of Ets-1 (PDB: 1K79) is superimposable with that of PU.1 (Fig. 1.5 *D*). However, Ets-1 and PU.1 belong to individual family members which are physiologically quite far from each other (27).

As mentioned above, Kodandapani *et al.* solved a co-crystal structure of murine PU.1 ETS domain (171-258 a.a.) and 16-bp DNA (5'-AAAAAGG**GGAA**GTGGG-3') at 2.3 Å resolution (PDB: 1PUE) (26). The PU.1 ETS domain has a globular structure (33 × 34 × 38 $Å^3$) that consists of three α-helices (H1, H2, and H3) and four antiparallel β-sheets (S1, S2, S3, and S4). The structure is typical of a winged helix-turn-helix (HTH) family, consisting of H2-loop-H3 as HTH and another loop between S3 and S4 as a wing. Thus, the ETS domain of PU.1 and other family members have a loop-helix-loop motif.

The DNA-binding site of the ETS domains has four strictly conserved residues: K219, R232, R235, and K245 in murine PU.1. R232 and R235 in H3 directly contact the bases of GGAA (the consensus sequence of the ETS family) in the major groove. These arginine residues also make water-mediated contacts with the bases of the GGAA core. K245 contacts phosphate backbone of the GGAA strand in the minor groove, and K219 contacts phosphate backbone of the other strand. It is notable that R81 and R84 of Fli-1, which correspond to R232 and R235 of PU.1, respectively, do not contact DNA directly. However, intermolecular NOE has been observed between the arginine residues and DNA by NMR (28).

1.7 Regulation/control of PU. 1 activity in the cell

1.7.1 Regulatory mechanisms of PU.1 activity

The biological activity of PU.1, as an essential transcription factor, is primarily controlled by up- and down-regulation of its expression. In addition to downregulation in expression, a few other inhibitory mechanisms for PU.1 are known: formation of a heterodimer with other protein

partners such as GATA-1 (29). Auto-inhibited ETS-family members are regulated by inhibitory helices packing against their DNA-binding domain in the unbound state. However, PU.1 has no such regulatory mechanism.

1.7.2 Roles and functions of autoinhibition for ETS transcription factors

Autoinhibition is a control mechanism of protein activity, whereby inhibitory module or domain of a protein interacts with another part of the protein so that it works for negative regulation (30). Many protein regulation mechanisms are known to proceed through autoinhibition (30). For example, alternative splicing or proteolysis would remove the autoinhibitory module (30). Post-translational modifications (PTM) or protein-protein interactions (PPI) in response to cellular signaling would relieve or reinforce autoinhibition and enable the protein to control downstream events (31).

Autoinhibition has been described as a key regulatory mechanism for ETS transcription factors at the protein/DNA level (30). Most of the 28 paralogs of ETS family in humans, except a few members including PU.1, have been described to possess autoinhibition (31). They typically have a common mechanism in which autoinhibitory elements, typically helices (α- or 3_{10}-helix), adjacent to the ETS domain, make DNA binding unfavorable. Thus, one can determine if an ETS protein is autoinhibited or not by detecting reduced affinity for site-specific DNA of a full-length protein (or a construct harboring both ETS domain and adjacent autoinhibitory elements) by comparison with an isolated ETS domain.

One way to classify the ETS family members is based on the number of autoinhibitory helices on both sides of the ETS domain (31). Ets-1 and Ets-2 have two helices on both sides of the ETS domain (N-terminal HI-1 and HI-2; C-terminal H4 and H5). GABPA and ETV6 have only C-terminal helices (H4 and H5). ELK4 has only one inhibitory helix (H4) at the C-terminal

side of the ETS domain. And several members, including PU.1, FLI1, SPDEF, ELF3, ELF5, ELK1, ELK3, and ELK4, have no autoinhibitory helices adjacent to the ETS domain (31).

Among the ETS family members, Ets-1 has been most characterized about autoinhibition. The autoinhibitory module of Ets-1 consists of the four helices (HI-1, HI-2, H4, and H5) and the interfaces with H1 of the ETS domain (Fig. 1.6, *A* and *B*). Early work on Ets-1 demonstrated that HI-1 unfolds upon binding to a specific DNA. Thus, the autoinhibition is thought to give an energetic penalty to the protein on DNA binding and to reduce net affinity (32). Recent dynamics studies by NMR revealed that both HI-1 and HI-2 unfolds upon binding to both specific and nonspecific DNA (33). Structural data of Ets-1 gives an insight that the inhibitory helices of Ets-1 lie on the distal surface from the DNA binding site. This leads to the understanding that binding to DNA and unfolding of autoinhibitory elements are allosterically coupled (34,35).

In the presence of all the four inhibitory helices, the net affinity of Ets-1 is reduced to a half (36). And the presence of an intrinsically disordered serine-rich region (SRR), which is located at the N-terminus of HI-1 and transiently interacts with both the ETS domain and the inhibitory module, diminishes binding affinity of Ets-1 up to 20-fold (36). Transient interactions with the SRR are enhanced by promoting multisite phosphorylation levels in response to Ca^{2+} signaling. Thus, the autoinhibition of Ets-1 is linked with cellular signaling events mediated by Ca^{2+}-dependent kinases such as CaM kinase II (36). Taken together, Ets-1 autoinhibition is associated with a conformational equilibrium between transcriptionally inactive and active states. Upon specific DNA binding, helices HI-1 and HI-2 unfold, and Ets-1 becomes flexible and in an active state. The inactive state of Ets-1 is favored and therefore stabilized by transient interactions of the SRR with the ETS domain and the autoinhibitory module (the four inhibitory helices). These interactions are dependent on multisite phosphorylation of the SRR.

Various other examples of autoinhibition for ETS proteins have been reported. ETV6 has two C-terminal helices H4 and H5. H5 sterically blocks the DNA-binding interface of its ETS domain, which leads to attenuate DNA binding greatly (37). ETV4 and ETV5 in the PEA3 subfamily have *non-helical* autoinhibitory elements (sequences) on both sides of their ETS domains (38,39). The cooperative binding of USF-1, a binding partner of ETV4, has been reported to enhance the DNA-binding affinity of ETV4 by interacting with the inhibitory elements (38). The autoinhibitory element of ELF3 is an ordered coil sequence, immediately C terminal to S4 of its ETS domain (31).

The autoinhibited ETS proteins mentioned above have autoinhibitory element(s) appended to both sides of or only C-terminal to the ETS domain, whereas distant sequences from the ETS domain autoinhibit two subfamilies (ESE and TCF) of ETS proteins. In the case of ESE subfamily, ELF3 (ESE-1) is autoinhibited by the transactivation domain at the center of the protein, thereby disruption of the transactivation domain enhances specific DNA binding (40). On the other hand, ELF5 (ESE-2) is autoinhibited by the N-terminal sequences of the protein (41). In the case of TCF subfamily, the autoinhibition of ELK1, ELK3, and ELK4 involves interactions between the B-box, the transactivation domain, and the NET inhibitory domain (31,42-44). Intramolecular and intermolecular interactions with the helix-loop-helix Id proteins are also part of the autoinhibition of TCF subfamily (45). Moreover, phosphorylation of the transactivation domain of the TCF subfamily by MAP kinases enhances specific DNA binding (46).

Various autoinhibitory mechanisms have been presented for the ETS TFs. Thus, an interesting question worth discussing is whether common features in the autoinhibition occur in the ETS family. DNA-binding domains for TFs are in general conformationally dynamic, which

is necessary for TFs to search for specific binding sites in a majority of nonspecific DNA, with DNA-scanning quenched upon specific DNA binding (47,48). DNA-binding interfaces in the ETS domains of Ets-1 and ETV6 are also conformationally dynamic as probed by NMR and HDX experiments (33,37), and this feature is likely to be common for ETS domains. Furthermore, such flexibility of the DNA-binding site presumably explains why Ets-1 protein shares the same binding interface for specific and nonspecific DNA binding (33). Collectively, diverse autoinhibitory elements adjacent to the ETS domain presumably reduce these dynamic properties of the ETS family, like the autoinhibitory helices of Ets-1 regulate its DNA-scanning motions.

1.8 Protein-protein interactions (PPIs) of ETS domains

1.8.1 Classification of protein-protein complexes

Classification of protein association into functionally and structurally related classes is important. Both the function and performance of the protein complex need to be considered for classification. Nooren and Thornton defined protein complexes as (i) either obligate or non-obligate and as (ii) either permanent or transient (49). An obligate protein complex is defined as one in which each subunit is too unstable to be found in vivo. A non-obligate protein complex is defined as one in which each subunit is stable and can be found in vivo. The terms permanent and transient classify protein complexes based on their lifetime. A permanent protein complex is defined as very stable associations. A transient protein complex is defined as one in which each subunit associates and dissociates in vivo. Notably, an obligate protein complex is the only permanent one in this definition, while a non-obligate protein complex can be either permanent or transient. In the literature, each of the two terms in each group (group (i) and (ii) above) is not

used to discriminate one from the other (50). They are considered to be interchangeable in protein-protein complexes (51).

While there have been many reports about protein oligomerization, our understanding of its supramolecular assembly and function is still limited compared to our knowledge of tertiary protein structures (52). One way to characterize the relationship between protein quaternary structure and function is to engineer or mutate a protein into a different oligomeric state, as we have studied here. Self-association of proteins is a common feature in the cells, and it has been estimated that 70-80% of cellular proteins are tetramers composed of four monomers associated noncovalently to function as oligomeric proteins (53,54). Various other oligomeric forms of proteins occur from a simple dimer to a complex form composed of many subunits, but the majority of them are either homo-dimers or homo-tetramers (55).

Oligomeric proteins consist of either homo- (identical) or hetero- (non-identical) subunits. Homo-oligomers associate either in an isologous or heterologous manner (52). An isologous association is symmetric and uses the same contact surfaces (residues), while a heterologous association is asymmetric and uses non-identical contact surfaces (52). Hetero-oligomers form only a heterologous assembly by their nature. This classification applies to protein oligomers except for some oligomeric interactions such as domain swapping, which is discussed below. Oligomeric protein structures are also classified into two groups: an obligate or non-obligate interaction (49). Obligate oligomers are usually very stable and exist as oligomers, thus their monomeric components are unstable and not found in vivo. On the other hand, non-obligate oligomers associate and dissociate in vivo. These protein oligomers can be further classified. Dynamic oligomeric orders (equilibrium) and transient interactions are produced by weak interactions and altered by molecular or physiological triggers. Such an equilibrium in the

oligomeric state can be important in protein function. The interfaces of non-obligate oligomers are typically smaller and less hydrophobic than those of obligate oligomers, probably to meet folding and solubility requirements for monomers in non-obligate oligomers (49).

1.8.2 Heterodimers of ETS domains and binding partners

PPIs regulate many eukaryotic transcription factors through their DNA-binding domains or additional subunits to function, typically as non-covalent dimers (56). For ETS transcription factors, dimers can be formed through ETS domains or other domains such as the PNT domain that approximately one-third of ETS family have, for both homodimerization and heterodimerization (31). Heterodimerization enables precise control of tissue-specific transcriptional regulation for the ETS family (57). Thus far, several structures of heterodimers of the ETS domain in (ternary) complex with DNA have been determined: GABPα/β (58), Ets-1/Pax-5 (11), ELK4 (SAP-1)/SRF (59), and PU.1/IRF4 (60).

The heterodimeric structure of the ETS domain of the GABPα subunit in a complex with GABPβ subunit at 2.15 Å resolution (PDB: 1AWC) was reported in 1998 (58). The total buried surface area of the dimerization interface is 1600 $Å^2$, where hydrophobic contacts in the main part and some water-mediated hydrogen bonds are observed (58). The ankyrin repeats of GABPβ insert into a depression of GABPα formed by H1, H4, H5, and the loop between H3 and S3. Even though GABPβ subunit does not have direct DNA contacts, the formation of GABPα/β heterodimer increases DNA-binding affinity compared to monomeric GABPα, presumably due to indirect GABPβ-DNA interaction mediated by hydrogen bonding from Lys^{69} of GABPβ to Gln^{321} of GABPα (58).

Heterodimeric Ets-1/PAX5 complex with the *mb-1* promoter DNA (PDB: 1K78) has only 180 $Å^2$ dimerization interface area because PAX5 binds to its cognate DNA site located on the

opposite side from Ets-1 (11). The DNA has variant 5'-GGAG-3' core. *mb-1* promoter has a low-affinity sequence for Ets-1. Thus, this interaction is critical for binding. It is because the heterodimeric complex forms optimal DNA contacts of the Tyr^{395} side chain in H3 (DNA-recognition helix) of Ets-1 as well as the further formation of van der Waals and salt bridge contacts (11). This is a good example to illustrate that the DNA-recognition helix H3 plays a key role in both protein–DNA and protein-protein interactions.

The interface of the Ets-1/PAX5 complex is similar to that of ELK4 (SAP1)/SRF complex with DNA (PDB: 1K6O (59)). Note that optimal DNA contacts like Ets-1/PAX5 are formed once SRF binds to its cognate DNA and ELK4 protein. In the heterodimeric complex of ELK4/SRF with DNA, further contacts between ELV4 and DNA are formed by reorientation of conserved Tyr and Arg residues of ELK4, after a small hydrophobic pocket in ELK4 H3 accommodates Leu^{155} of SRF (59).

PU.1 forms a heterodimer with the interferon regulatory family transcription factor IRF4 in the presence of immunoglobulin light-chain gene (λB) enhancer DNA (60). Low-affinity DNA binding of IRF4 increases co-operativity 20- to 40-fold, presumably because PU.1 forms a salt bridge with IRF4 and gains binding energy (60). Thus, in the cases of Ets-1, ELK4, and PU.1, dynamic co-operativity upon binding to their heterodimeric binding partners in the presence of DNA are observed either by optimizing binding to low-affinity DNA sequences or by gaining binding energy.

A number of other binding partners of PU.1 have been identified for each domain of PU.1 (61). Proteins such as TFIID, TBP, GATA-1, and GATA-2 interacts with the transactivation domain (N-terminus). PU.1 interacting partner (PIP) and ICSBP are known to interact with the central PEST domain, and phosphorylation of Ser^{148} residue in murine PU.1

plays an essential role in this protein-protein interaction (62). Proteins including c-Jun, c-Myb, GATA-1, GATA-2, and NF-IL6 interact with the C-terminal ETS domain.

In the case of binding through the ETS domain of PU.1, a critical PU.1 coactivator c-Jun interacts with the 'wing' of the winged HTH motif (the S3/S4 region), which is also the binding site for GATA TFs (GATA-1 and GATA-2), and therefore they compete to bind. Through these events, transactivation of both PU.1 and GATA TFs is repressed (63). Interestingly, a structurally homologous ETS family member ERG also binds to the Jun basic domain via its ETS domain, but the 'wing' is not the dimerization interface, which reflects functional diversity of the ETS domain (64). Instead, Arg^{367} and Tyr^{371} in H3 of ERG are critical for the interaction (64). PU.1 also interacts with NF-IL6 via two regions: the 'wing' (the same interface as binding to c-Jun and GATA) and the S2-H2-H3 region (65).

1.8.3 Homodimers of ETS domains in the presence of DNA

In addition to transcriptionally active 1:1 protein/DNA complex, homodimerization of many ETS family members has been reported including Ets-1, Elk-1, ETV1, ETV6, FEV, ERG, and PU.1 (66-70). Here, it is noteworthy that all of these ETS domain homodimers are 2:2 protein/DNA complex except for an example of 2:1 protein/DNA complex of Ets-1 in a non-reducing environment, where two Cys residues from each subunit are likely to form a disulfide bond (71). Among the ETS proteins capable of homo-dimerizing, Ets-1 has been studied most extensively. For example, positively co-operative binding of Ets-1 at 2:2 protein/DNA stoichiometry is observed at repeated (palindromic) specific DNA sites such as stromelysin-1 promoter (72). Such a positively cooperative DNA binding of Ets-1 is known to counteract its autoinhibition (73). Three homodimeric structures of Ets-1 have been determined by crystallization (PDB: 2NNY, 3MFK, and 3RI4) (Fig. 1.7) (72,74,75).

Ets-1 has two dimerization interfaces for its homodimerization. For the structures 2NNY and 3MFK, the binding mode is head-to-head on palindromic specific sites, with HI-2 and the loop between HI-2 and H1 contacting the loop between H2 and H3 reciprocally. For the structures 3MFK and 3RI4, the binding mode is also head-to-head on palindromic specific sites, with HI-1 reciprocally contacting H4, HI-2, and the loop between HI-1 and HI-2. It is notable that 3MFK has two dimerization interfaces.

In the case of Elk-1 (PDB: 1DUX), the homodimer (2:2 protein/DNA complex) has a reciprocal interface involving H1 and the H1/S1 loop of each subunit. The dimerization interface of Elk-1 is similar to that of Ets-1, except that Elk-1 does not have autoinhibitory helices appended to the ETS domain. In the 2:2 complex structure of Elk-1, the DNA-binding site locates on the almost opposite side of the dimerization interface, which is different from Ets-1 that forms head-to-head 2:2 complex. Furthermore, both ETV1 (PDB: 4AVP) and FEV (PDB: 2YPR) structures have a dimerization interface at H1, H4, and the S1/S2 loop, involving an intermolecular disulfide bond, although their orientations and surface positions are different from each other.

1.8.4 Homodimerization of ETS proteins in the absence of DNA

The ETS domain of Elk-1 is known to mediate homodimerization of Elk-1, and the resulting homodimer is given cytoplasmic stability to resist proteasomal degradation as well as localization to the nucleus (68). In the same report, PU.1 in the nucleus forms homodimer(s) in the absence of DNA but is monomeric upon binding to high-affinity DNA (68). This report suggests the biological relevance of PU.1 homodimerization mediated by the ETS domain and also arises a new question of whether PU.1 can dimerize in the presence of DNA.

1.8.5 Potential for dimerization of PU.1 at a single cognate site of DNA

In our previous work, we observed the potential for the ETS domain of PU.1 to dimerize at a single specific site by ITC titration experiments in 2012 (76). When a forward titration was performed by adding PU.1 protein to site-specific DNA, initially stable titration heats (ΔH°) started increasing beyond the 1:1 protein:DNA ratio and returned to baseline after a second equivalence point. Thus, two distinct phases (Fα and Fβ) were observed. Next, when a reverse titration was performed by adding DNA to PU.1 protein, two distinct transitions of ΔH° were observed. Negative ΔΔH° transition (Rα) from the beginning of titration to the first equivalence point (1:1 protein:DNA ratio) and positive ΔΔH° transition (Rβ) from the first to the second equivalence point were observed. It is noteworthy that the magnitudes and signs of the ΔH° are different between the forward and reverse titrations. Collectively, we observed the potential of two distinct DNA binding modes (1:1 and 2:1 protein:DNA ratio) for PU.1. Also, the negative to positive transition in the reverse titration (Rα and Rβ) implies that a 2:1 (PU.1:DNA) complex is formed in a negatively cooperative manner.

1.8.6 Free energy landscape of PU.1 ETS domain in solution

The free energy of the four states of PU.1 (i.e., monomeric and dimeric PU.1 in the absence or presence of site-specific DNA) were further analyzed under standard state conditions, in the same study (76). Among the four states of PU.1, the unbound PU.1 monomer is most unstable, and the 1:1 DNA-bound complex is most stable. Furthermore, taking the free energy between these two monomeric states into account, the DNA-free dimer is less stable than the DNA-bound dimer (i.e., 2:1 PU.1/DNA complex) although the magnitudes of ΔG° between DNA-free and -bound dimer significantly depends on the flanking sequences of the core DNA-binding motif of PU.1 (76). Taken together, the free energy (G°) gradient of the PU.1 ETS

domain is as follows: unbound monomer > PU.1 dimer in the absence of DNA > 2:1 DNA-bound complex > 1:1 DNA-bound complex.

1.9 Intrinsically disordered regions (IDRs) of ETS domains

1.9.1 Intrinsically disordered proteins (IDPs) and regions (IDRs)

A large fraction of any eukaryotic proteomes consists of polypeptides that are unlikely to form well-defined, three-dimensional structures (77). Recent studies support that such protein segments can be functional even in the absence of stable and globular tertiary structures (78-80); these protein segments are referred to as intrinsically disordered regions (IDRs) now. Biased amino acid composition and low sequence complexity are characteristic of IDRs. IDRs cannot form the hydrophobic core necessary for making up fixed tertiary structures due to low proportions of bulky hydrophobic amino acids (81). Proteins that are composed of only disordered sequences and thus have no tertiary structures are called intrinsically disordered proteins (IDPs) (77). However, the majority of eukaryotic proteins consist of both IDRs and structured regions (77).

1.9.2 Emerging roles of IDRs for transcription factors

IDRs are known to be functional since the mid-1990s (82), and the reports on IDRs have changed the classic paradigm of protein structure-function relationship. IDRs are unable to make stable and well-defined tertiary structures. Instead, their disorder is dynamic, and therefore, they can rapidly form a range of conformations (81). Thus, IDRs can display various binding affinity and kinetics due to their dynamic properties (81).

TFs have many advantages with the disorder, which facilitate their function and regulation, such as conformational plasticity and binding promiscuity (81). TFs with disorder can adopt different conformations and interact with multiple binding/interaction partners, which in

turn promotes the assembly of macromolecule complexes (81). This explains why eukaryotic TFs exhibit a significantly higher degree of disorder than prokaryotic ones (83).

1.9.3 IDRs of transcription factors and diseases

IDRs can cause diseases either through dysregulation or aggregation of proteins (84). Diseases associated with TFs that contain IDRs include multiple types of cancer (85-87), neurodegenerative diseases (88-90), cardiovascular diseases (91), and type 2 diabetes (92). One of the ETS family members *FLI1* can be a cause of Ewing's sarcoma by generating an oncogenic fusion protein EWS-FLI1 (86). This is an example of cancer that IDRs drive through dysregulation (chromosomal translocation in this case). Aberrant oligomers of IDRs can assemble into pathological aggregates (amyloids). TFs with IDRs can be dysfunctional as such and cause neurodegenerative diseases such as Alzheimer's disease (93).

1.9.4 PEST sequences

A protein region that destabilizes a protein and its half-life by >100-folds was identified using computational methods in the 1980s (94). This small region is enriched in Pro (P), Glu (E), Ser (S), and Thr (T) and forms a degradation signal, thus it was named the PEST region after the four representative amino acids (95,96). PEST region is hydrophilic and consists of at least 12 amino acids, flanked by positively charged amino acids – Lys, Arg, and His (94). About 10% of proteins have such a sequence, and interestingly, proteins with a shorter life span have higher populations of PEST sequences. Various regulatory proteins such as p53, Jun, Fos, Myc, and protein kinases and phosphatases have PEST sequences (94). Proteins with the PEST sequence get degraded by the proteasome, but the mechanism is still not clear yet (94).

1.9.5 Difficulties in NMR spectral assignment of IDRs

While IDPs are hard to study by crystallography because of their inherent structural flexibility, NMR is a powerful tool to study protein structure, dynamic properties, interactions, and so on. A 2D ^{1}H-^{15}N HSQC spectra can be considered as a protein "fingerprint" since it contains well-dispersed peaks of all backbone amide resonances except Pro residues that do not appear in ^{1}H-^{15}N HSQC spectra. However, residues in IDRs display severe spectral overlap (97). Unfortunately, Pro residues are highly abundant in IDRs, which makes sequential backbone assignment of IDPs even more difficult than we analyze ordered proteins (97).

To overcome such difficulties, measurements at low temperatures and pH are often used for the backbone assignment of IDPs (97). Improved instrumental sensitivities have been achieved for the following reasons (97). (i) Non-uniform sampling technologies allow for high-dimensionality experiments. (ii) Faster acquisition of NMR experiments using cryoprobe allows longitudinal relaxation experiments and direct detection of heteronuclei (^{13}C). The latter has also helped overcome line broadening problems. Signal overlap problems require the future development of NMR techniques (97).

1.9.6 Gaining structures in IDPs

IDPs are sensitive to chemical environments such as pH, temperature, and ligand binding. In a recent study, some IDPs gained more compact structures with higher α-helical content under acidic conditions because electrostatic repulsion of negatively charged residues reduced (98). Thus, IDPs are suggested to be stabilized by favorable electrostatic interactions.

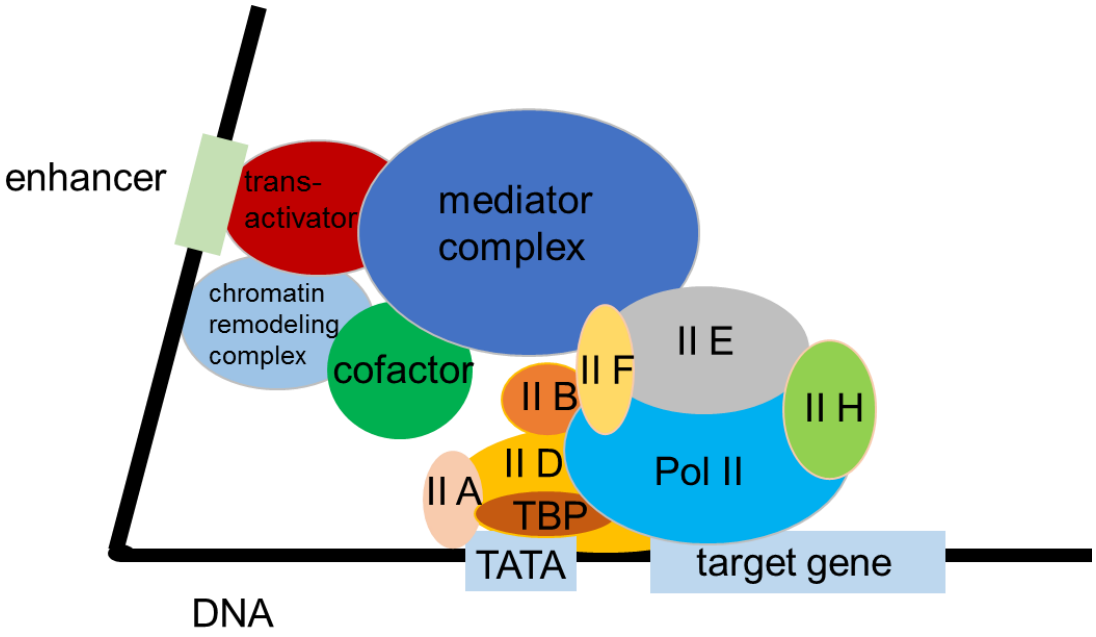


Figure 1.1 Assembly of the preinitiation complex (PIC) in the initiation step of eukaryotic transcription.

In eukaryotes, each gene has its own promoter near and upstream of the gene. RNA polymerase II binds to a promoter sequence of DNA. For promoters that contain TATA box, TBP (TATA box binding protein) binds to the TATA box and initiates transcription complex assembly. Activation signals from mediators and coactivators are sent to transcription activators. The chromatin remodeling complex and coactivators activate chromatin. Subsequently, PIC is assembled by RNA polymerase II, using five general transcription factors (TFIIA, IIB, IIE, IIF, and IIH).

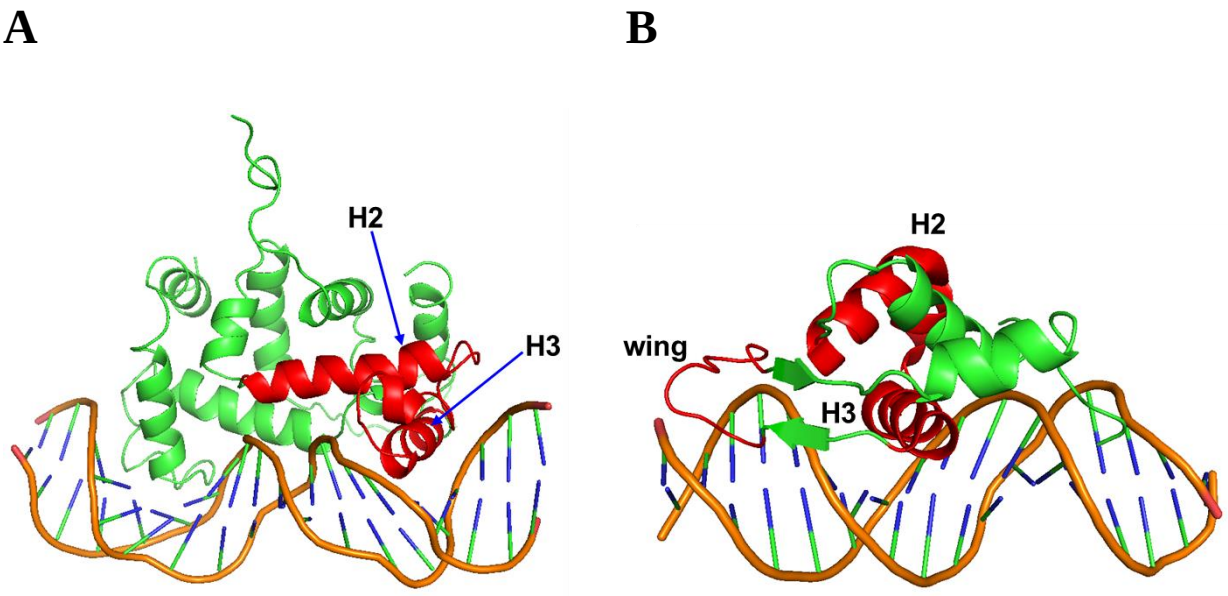
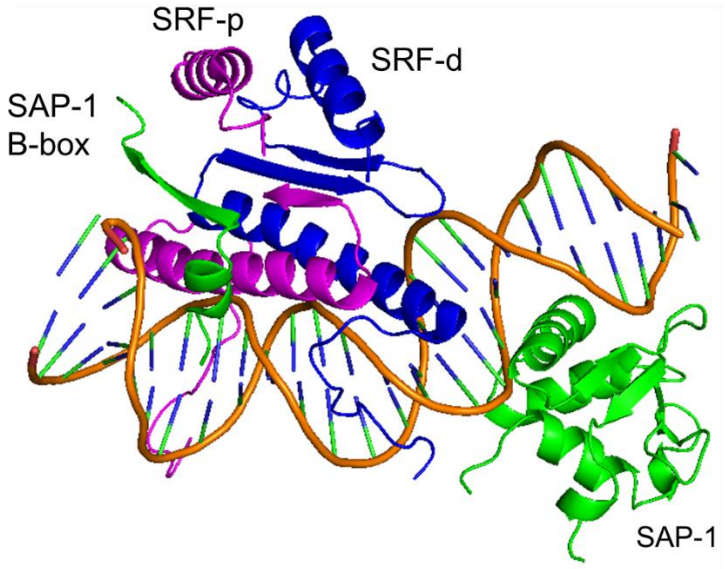


Figure 1.2 Helix-turn-helix and Winged helix-turn-helix transcription factors.

A, a solution NMR structure of the *Trp* repressor and DNA complex (PDB: 1RCS). The protein subunits are shown in green, with the helix-turn-helix (HTH) domain highlighted in red. *B*, a 1:1 complex of HNF-3 (shown in green) and DNA (PDB: 2HDC). The winged-helix-turn-helix (wHTH) domain and the second wing are highlighted in red.

A



B

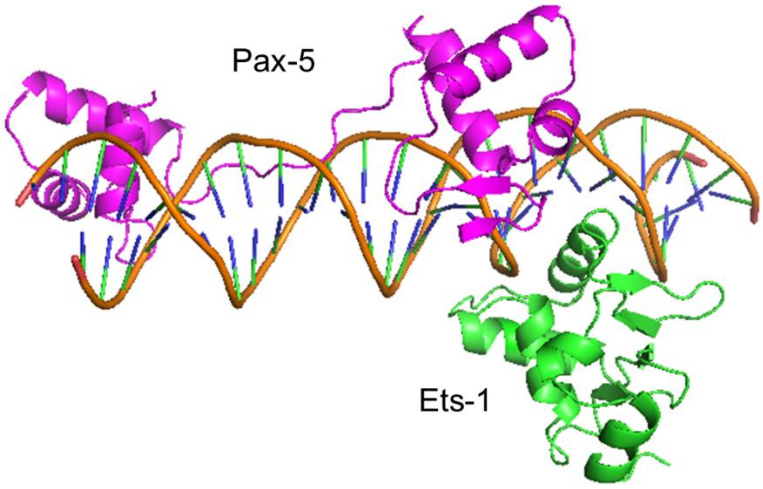


Figure 1.3 Combinational transcriptional regulatory complexes in the ETS family.

A, a crystal structure of ternary SAP-1/SRF/specific DNA complex (PDB: 1HBX). The ETS domain of Sap-1 is shown in green. The MADS-box DNA binding domains of SRF (SRF-d and SRF-p) are in blue and magenta. *B*, a crystal structure of ternary Ets-1/Pax5/specific DNA complex (PDB: 1K78). The ETS domain of Ets-1 is shown in green and Pax-5 is in blue.

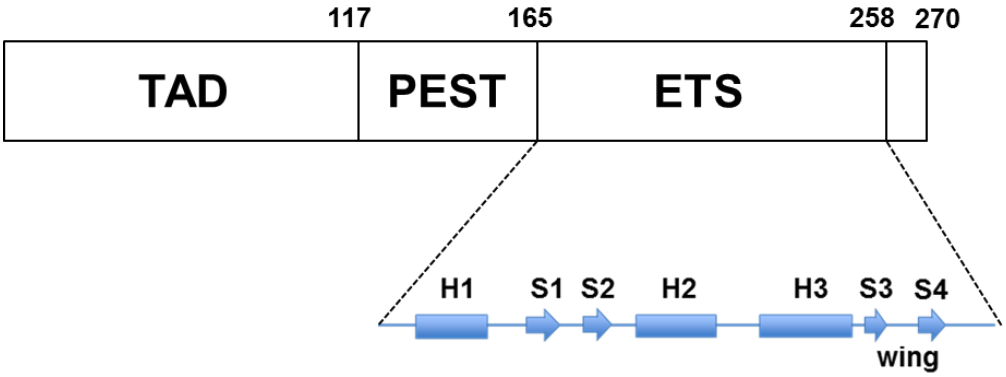


Figure 1.4 Nomenclature and domain organization of hPU.1 protein.

In PU.1 protein, the ETS domain is the only ordered region. hPU.1 protein mainly consists of four parts. N-terminal transactivation domain (1-116), PEST domain (117-164), ETS domain (165-258), and C-terminal disordered residues (259-270). Therefore, the ETS domain is flanked by IDRs.

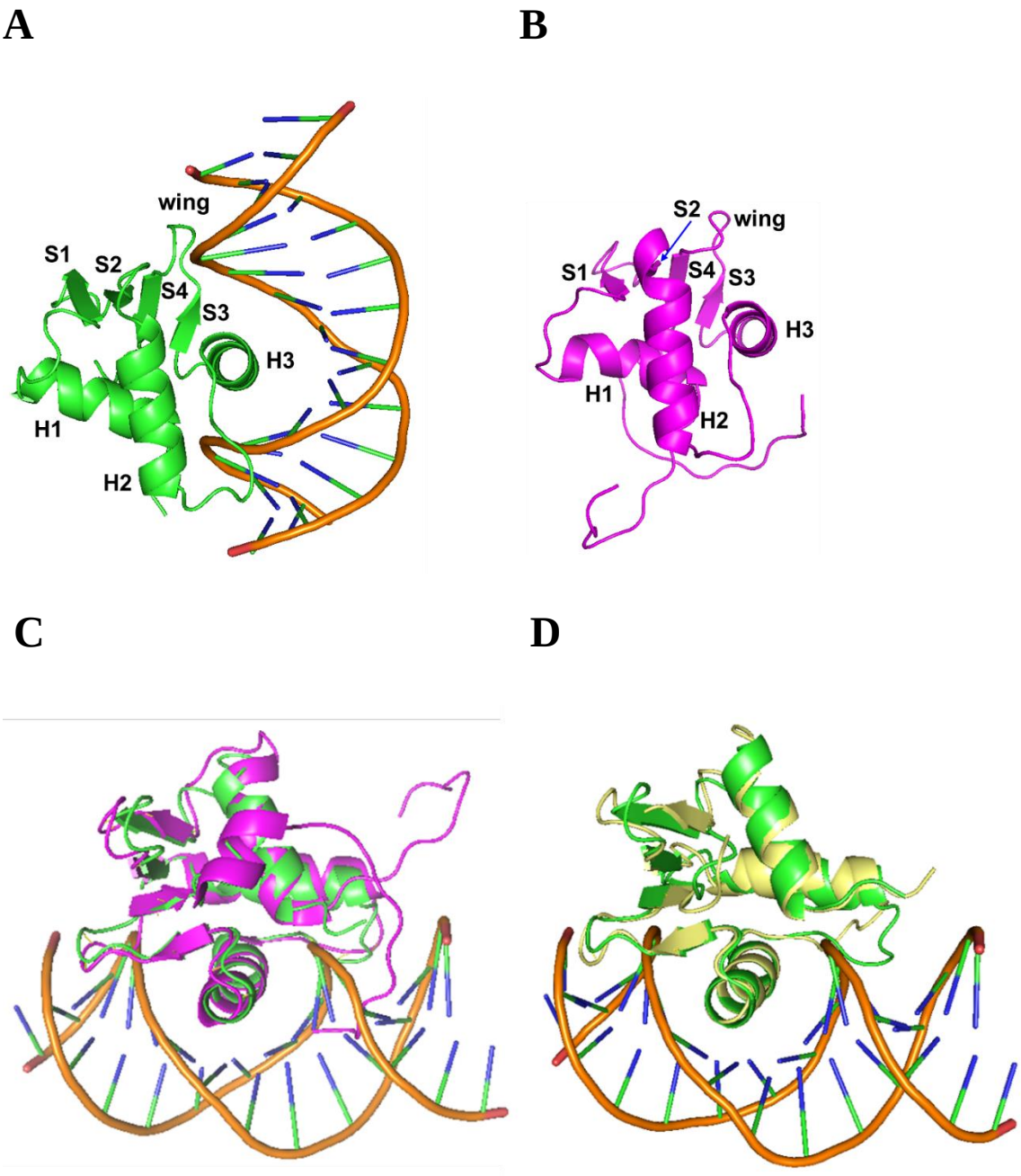
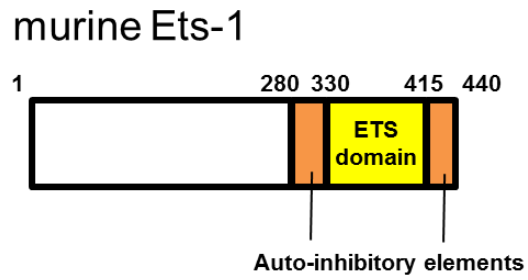


Figure 1.5 PU.1 structures in the presence and absence of DNA, an overlay of PU.1 in the presence and absence of DNA, and an overlay of PU.1 and Ets-1.

A, a crystal structure of PU.1/specific DNA complex (PDB: 1PUE). The ETS domain of PU.1 is shown in green. *B*, a solution NMR structure of unbound PU.1 ETS domain (PDB: 5W3G). The ETS domain of PU.1 including C-terminal IDR is shown in magenta. *C*, an overlay of PU.1 in the presence (green) and absence (magenta) of cognate DNA. *D*, an overlay of PU.1 (green) and Ets-1 (yellow; PDB: 1K79) in the presence of cognate DNA.

A



B

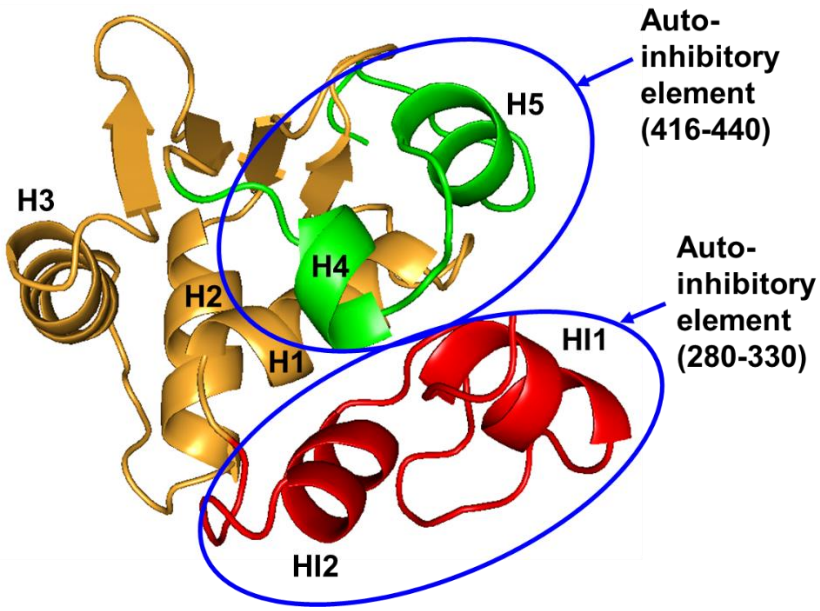


Figure 1.6 Autoinhibitory module of Ets-1.

A, The autoinhibitory elements of Ets-1 flank both termini of the ETS domain. *B*, the auto-inhibitory module of Ets-1 consists of the four helices (HI-1, HI-2, H4, and H5) and the interfaces with H1 of the ETS domain (PDB: 1R36).

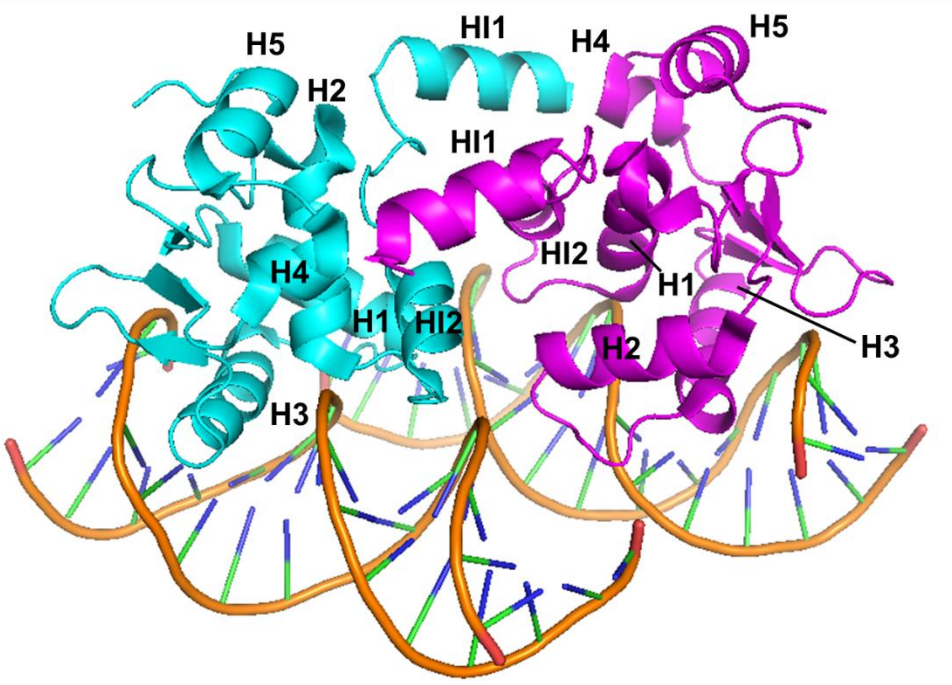


Figure 1.7 2:2 Ets-1/DNA complex.

A crystal structure of $(Ets\text{-}1)_2$/2DNA quaternary complex (PDB: 3RI4), where head-to-head binding between ETS domains on palindromic cognate DNA sites are observed. Two ETS domains of Ets-1 are shown in cyan and magenta. The helices HI-2, H4, and H5 are involved in docking of HI-1 from the other Ets-1 subunit.

2 MULTIPLE DNA-BINDING MODES FOR THE ETS FAMILY TRANSCRIPTION FACTOR PU.1

2.1 Preface

The content in this chapter is based on a peer-reviewed paper: Multiple DNA-binding modes for the ETS family transcription factor PU.1. Esaki S, Evich MG, Erlitzki N, Germann MW, Poon GMK. *Journal of Biological Chemistry.* 2017 Sep 29;292(39):16044-16054 (99). I prepared the protein and DNA samples in this work. I conducted the SNS-dye binding assays and NMR experiments. NMR experiments and analysis were in collaboration with Drs. Markus Germann and Marina Evich. This work was supported by National Science Foundation Grant MCB 15451600 and National Institutes of Health Grant R21 HL129063.

2.2 Abstract

The eponymous DNA-binding domain of ETS (E26 transformation-specific) transcription factors binds a single sequence-specific site as a monomer over a single helical turn. Following our previous observation by titration calorimetry that the ETS member PU.1 dimerizes sequentially at a single sequence-specific DNA-binding site to form a 2:1 complex, we have carried out an extensive spectroscopic and biochemical characterization of site-specific PU.1 ETS complexes. Whereas 10 bp of DNA was sufficient to support PU.1 binding as a monomer, additional flanking bases were required to invoke sequential dimerization of the bound protein. NMR spectroscopy revealed a marked loss of signal intensity in the 2:1 complex, and mutational analysis implicated the distal surface away from the bound DNA as the dimerization interface. Hydroxyl radical DNA footprinting indicated that the site-specifically bound PU.1 dimers occupied an extended DNA interface downstream from the 5'-GGAA-3' core consensus relative to its 1:1 counterpart, thus explaining the apparent site size requirement for sequential

dimerization. The site-specifically bound PU.1 dimer resisted competition from nonspecific DNA and showed affinities similar to other functionally significant PU.1 interactions. As sequential dimerization did not occur with the ETS domain of Ets-1, a close structural homolog of PU.1, 2:1 complex formation may represent an alternative autoinhibitory mechanism in the ETS family at the protein-DNA level

2.3 Introduction

The differentiation of distinct lineages of blood cells from a single progenitor species occurs in a multistep process, termed hematopoiesis, that is intricately controlled at the transcriptional level. The ETS family transcription factor PU.1 ranks among the most essential hematopoietic regulators in ensuring the continued self-renewal of this progenitor, the hematopoietic stem cell (17). PU.1 is also essential for directing correct differentiation of the hematopoietic stem cell to various cell lineages. Crucially, PU.1 governs cell fate specification and functions in a transient, dosage- and cell stage-dependent manner (100). In mature cells, graded PU.1 activity is also required for key cellular processes and the specialization of mature cells into function-specific subtypes. Aberrant PU.1 activity is linked to a spectrum of diseases, including rheumatism (19), hematologic cancers (24,25,101), and Alzheimer's disease (102). Clearly, knowledge of the regulatory mechanisms of PU.1 is essential to an understanding of its role in normal biology and in disease.

Biological modulation of PU.1 activity is generally attributed to up- or down-regulation at the level of expression. With a metabolic half-life of ~50 h, a period that spans the entire cell cycle (103), the cellular persistence of PU.1 means that downregulation of its own expression cannot alone provide a complete description of PU.1 regulation, as additional dampening

mechanisms are required to prevent open-ended escalation of PU.1 activity during its lifetime in the cell.

Outside of down-regulated expression, few inhibitory mechanisms have been described for PU.1. The best understood is the mutual antagonism between PU.1 and the zinc finger transcription factor GATA-1, wherein each protein inhibits DNA binding by the other during myelopoiesis (29). In addition, PU.1 is one of only a few ETS family members that lack so-called autoinhibition, a regulatory mechanism in which helices adjacent to the ETS DNA-binding domain allosterically reduce DNA-binding affinity (Fig. 2.1 *A*) (104). In the case of Ets-1, the archetypal autoinhibited ETS paralog, interactions with partner proteins, such as Pax5 (105) and AML1/RUNX1/CBFα2 (73), relieve autoinhibition and restore high-affinity binding. Thus, in the absence of lineage-specific inhibitory partners such as GATA-1 or some intrinsic regulatory alternative to autoinhibition, PU.1 would be continuously poised in a functionally uncontrolled, transcriptionally permissive state.

In previous work, we have observed in calorimetric titrations the potential for the eponymous DNA-binding domain of PU.1 to dimerize at a single cognate DNA-binding site (76). Whereas dimerization of ETS domains of other ETS family homologs bound to two sites (i.e. 2:2 complexes) has been reported (74,106-109), self-association at a single site is poorly understood. We have carried out an extensive series of spectroscopic and biochemical experiments to directly characterize the variable binding modes of PU.1 as a function of DNA sequence and site size. The results show a site-specific 2:1 complex in exchange between free PU.1 on the one hand and the 1:1 site-specifically bound state on the other, while contacting the DNA over an extended interface beyond the single helical turn observed in the 1:1 co-crystal structure. Sequential dimerization imposes the dual requirements of specific DNA as well as a

site size longer than 10 bp. Nonspecific DNA forms oligomeric but not 1:1 complexes at equilibrium. Sequential dimerization of site-specifically bound PU.1, which sequesters excess circulating PU.1 from accessible DNA target sites, suggests itself as a potential mechanism of negative feedback in the absence of inhibitory binding partners.

2.4 Materials and methods

2.4.1 Proteins

Recombinant constructs representing the ETS domain of murine PU.1 (residues 167–272, designated PU.1ΔN167) and Ets-1 (residues 331–440, designated Ets-1ΔN331) were cloned with a thrombin-cleavable C-terminal His_6 tag as described (110). A similarly tagged construct for autoinhibited Ets-1 (residues 280–440, Ets-1ΔN280) was a gift from Dr. Lawrence P. McIntosh (University of British Columbia). Unlabeled constructs were overexpressed in *Escherichia coli* in LB medium. Uniformly ^{15}N-labeled PU.1ΔN167 was expressed from 5-ml starter cultures in LB broth grown at 37 °C for ~8 h. All of the culture was inoculated into 250 ml of LB broth, grown at 37 °C for ~16 h, and harvested. The cell pellet was resuspended in standard M9 medium containing $^{15}NH_4Cl$, $MgSO_4$, $CaCl_2$, trace metals, minimal essential medium vitamins, and glucose. Protein expression was induced with 0.5 mM isopropyl 1-thio-β-D-galactopyranoside overnight at 25 °C. Both unlabeled and isotopically labeled constructs were purified as described (111). In brief, cleared lysate was first purified by immobilized metal affinity chromatography, cleaved with thrombin, dialyzed against 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.4) containing 0.5 M NaCl, and polished on Sepharose SP (GE Healthcare). Buffers used with Ets-1 constructs, which harbored reduced cysteines, additionally contained 0.5 mM tris(2-carboxyethyl)phosphine-HCl. Protein concentrations were determined by UV absorption at 280 nm using the following extinction coefficients (in M^{-1} cm^{-1}): 22,460 (PU.1ΔN167), 32,430 (Ets-

1ΔN331), and 39,880 (Ets-1ΔN280). The labeling efficiency of ^{15}N-labeledconstructs was >98%, as judged by mass spectrometry (Fig. S1).

2.4.2 Nucleic acids

Synthetic DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA) and annealed to form duplex binding sites harboring the high-affinity 5'-AGC**GGAA**GTG- 3', low-affinity 5'-AAA**GGAA**TGG-3', or nonspecific 5'-AGCGAGAGTG-3' DNA sequence (ETS-specific core consensus in boldface type). Fluorescent DNA probes were constructed by annealing a Cy3-labeled oligonucleotide with excess unlabeled complementary strand as described (112).

2.4.3 Fluorescence polarization titrations

ETS protein binding to fluorescently labeled DNA sites was measured using a Molecular Devices Paradigm plate reader as described (113). In brief, DNA probe (0.5 nM) was incubated to equilibrium with graded concentrations of purified PU.1ΔN167 in a total volume of 30 µl of 10 mM Tris-HCl (pH 7.4) containing 150 mM total Na and 0.1 mg/ml bovine serum albumin. Steady-state fluorescence parallel and perpendicular to the incident polarized light was acquired at 595/35 nm upon excitation at 535/25 nm. Each data point represents the mean ± S.E. of five consecutive measurements as an indication of instrumental noise. Anisotropy data were fitted with a 1:1 or sequential binding model (112) to directly estimate the dissociation constants of the PU.1/DNA 1:1 and 2:1 complexes.

2.4.4 Pulsed field gradient diffusion-ordered NMR (DOSY)

NMR experiments were performed on Bruker Avance 500 and 600 MHz spectrometers, equipped with a 5 mm TBI and QXI probe, respectively (Bruker). Purified PU.1ΔN167 (~250 µM) and DNA (~600 µM high-affinity, low-affinity and non-specific duplexes) were extensively co-

dialyzed against 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.4) containing 150 mM NaCl, lyophilized, and re-dissolved to their previous volumes with 99.996% D_2O. The pH* (meter reading) of the reconstituted samples was 7.6. Protein was titrated with DNA to the indicated ratios in the text. In the case of low-affinity and nonspecific DNA, turbidity was observed at ratios below DNA:protein = 1:3 immediately after titration that resolved overnight at room temperature. A 1D pre-saturation (zgpr) spectrum was measured for each titration prior to diffusion measurements. Diffusion experiment parameters (Δ, δ and gradient strength) were first optimized by running 1D diffusion experiments (stebpgp1sd) at 2 and 95% gradient strengths with 100 ms and 5 ms, Δ and δ diffusion times, respectively for ~10% signal retention. Using these parameters, a pseudo-2D DOSY experiment using stimulated echo with bipolar gradient pulses (stebpgp1s) was acquired with 16k × 20 data points with a spoil gradient of 1.1 ms and 4.0 s relaxation delay from 2 to 95% gradient strength with a linear ramp. Data was processed with Bruker Topspin T1/T2 software using manual peak picking. Care was taken to avoid NMR peaks that potentially overlap with free DNA at 1:1 (protein:DNA) and excess DNA titrations. The intensity I of each picked peak was fitted to the following equation as a function of field gradient strength g:

$$I(g) = I_0 e^{-DQg^2} \tag{E1}$$

where I_0 is the reference (unattenuated) intensity, D the diffusion coefficient, and Q is a constant consisting of fixed parameters specific to the experimental configuration.

2.4.5 2D ^{1}H-^{15}N HSQC NMR

Purified [^{15}N] PU.1ΔN167 (~0.5 mM) was extensively dialyzed together with various duplex DNA constructs (~2 mM) in separate dialysis tubings against 11 mM

NaH_2PO_4/Na_2HPO_4, pH 7.6, 167 mM NaCl, and 0.1% NaN_3 and adjusted to 10% D_2O. DNA was titrated into protein to achieve the desired DNA/protein ratios. ^{1}H-^{15}N correlated measurements were made using a phase-sensitive, double inept transfer with a GARP decoupling sequence, and solvent suppression (hsqcf3gpph19). Spectra were acquired with 1024 × 144 data points and zero-filled to 4096 × 4096.

2.4.6 ANS fluorescence

ANS (ammonium salt, Alfa Aesar) was prepared at 2 mM in 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.4) buffer containing 150 mM NaCl and stored in the dark at 4 °C. Triplicate samples of PU.1 with or without 16-bp high-affinity DNA plus various controls were prepared in the same buffer before the addition of ANS to 200 μM. Final concentrations of PU.1 and DNA were as indicated under "Results." After incubation for 30 min, the fluorescence intensity of each sample was read at 370/530 nm or scanned from 400 to 750 nm with a Paradigm plate reader.

2.4.7 Hydroxyl radical DNA footprinting

A 130-bp DNA fragment harboring a copy of the high-affinity PU.1-binding site 5'-AGC**GGAA**GTG-3' was generated by PCR using two primers, of which the one encoding the 5'-CAC**TTCC**GCT-3' strand had been 5'-end–labeled with [^{32}P]ATP. After purification by agarose gel electrophoresis, the radiolabeled fragment (<1 nM) was incubated to equilibrium with graded concentrations of PU.1ΔN167. Each sample was digested with hydroxyl radical, purified, resolved by denaturing polyacrylamide gel electrophoresis, and digitized by phosphorimagery as described (114). Lane traces were constructed, and bands were indexed using a C ± T chemical sequencing reaction. Peaks were fitted as a superposition of Gaussian distributions, numerically integrated, and normalized to a band outside of the binding site to quantify fractional protection relative to the unbound sample.

2.4.8 Structure-based calculations

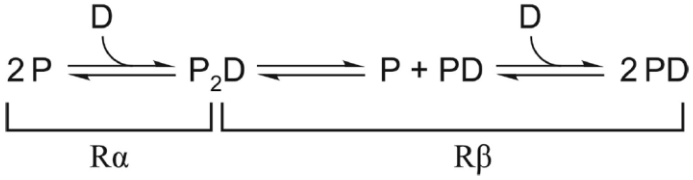
Self-diffusion constants for unbound and 1:1 DNA-bound PU.1ΔN167 were computed using the software HydroPRO (115). DNA-bound and unbound PU.1 structures were templated from the co-crystal structure with DNA (PDB code 1PUE) (26), appended with additional residues present in PU.1ΔN167, and relaxed by all-atom molecular dynamics simulation for 200 ns following our established protocol (113). Computations were carried out using volumetric values for D_2O at 25 °C, namely a density of 1.107 g/ml, viscosity of 1.25 centipoises, and partial specific volume of 0.70 ml/g.

Continuum electrostatics of PU.1 in the co-crystal structure were computed using APBS (116). Calculations were performed for an aqueous solution containing 0.15 M NaCl at 25 °C and rendered on the solvent-accessible surface from -1 to +1 kT/e.

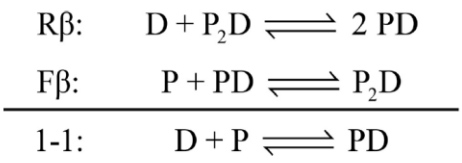
2.4.9 ITC

Purified PU.1 or Ets-1 was dialyzed extensively together with 23-bp DNA harboring the protein's respective optimal target (5'-AGC**GGAA**GTG-3' for PU.1; 5'-GCC**GGAA**GTG-3' for Ets-1) in separate compartments, against 50 mM NaH_2PO_4/Na_2HPO_4, pH 7.4, 150 mM total Na^+, 0.1 mM EDTA, and 1 mM dithiothreitol. Titrations were performed by injecting DNA (initial concentration ~500 μM) into protein (~50 μM) in a Nano ITC instrument (TA Instruments). Data fitting to empirical 1:1 and cooperative models was performed as described (117) only to demonstrate the models qualitatively, not for quantitative estimation of the binding affinities, due to the very strong dissociation constant of the 1:1 complex (10^{-9} M). To compare the calorimetric enthalpies for DNA binding by PU.1 with those for Ets-1, which exhibited strictly 1:1 behavior, the calorimetric enthalpies for PU.1/DNA binding were decomposed as follows to extract the enthalpy changes for the 1:1 complex. Whereas PU.1ΔN167 dimerizes in both unligated and

DNA-bound states, the former occurs at considerably higher concentrations (near 10^{-3} M) (76,118) than those used in the "reverse" DNA-into-protein titrations shown in Fig. 2.1 (*A* and *B*). Under these conditions, the biphasic profile arises from the 1:1 complex being strongly favored and yielding a 2:1 complex only in excess protein at the initial phase (Rα) of the titration as shown in Scheme 1, where P and D represent PU.1ΔN167 and site-specific DNA in their various free and bound states. The two phases Rα and Rβ are marked in the reverse titration shown in Fig. 2.1*B*. Because the transition from the 1:1 to 2:1 PU.1/DNA complex occurs sequentially, both phases are well-defined and extracted according to the technique of "total association at partial saturation" (119).



Scheme 1



Scheme 2

To compare the enthalpy changes meaningfully with the manifestly 1:1 binding for Ets-1 in Fig. 2.1 (*C–F*), the complex heats in the reverse titrations are dissected to account for the thermodynamics of coupled dimerization and dissociation of PU.1 as shown in Scheme 2. The calorimetric enthalpy marked Fβ has been measured previously for PU.1ΔN167 under the same solutions (76). Based on $\Delta H_{R\beta}$ = -44.2 ± 1.4 kJ/mol (cf. Fig. 2.1*B*) and $\Delta H_{F\beta}$ = 17.1 ± 0.7 kJ/mol (76) at 25 °C, the enthalpy change for the formation of the canonical 1:1 complex was -27.1 ± 1.6 kJ/mol. Thus, the enthalpy change for formation of the 1:1 complex from unbound

constituents was larger in magnitude than that for minimal ($\Delta H_{1:1}$ = -12.0 ± 0.4 kJ/mol) and autoinhibited Ets-1 ($\Delta H_{1:1}$ = -8.1 ± 0.4 kJ/mol) at 25 °C.

2.5 Results

A1:1 protein/DNA site stoichiometry is universally observed in co-crystal structures of ETS family transcription factors, including PU.1 (27,30) (Fig. 2.1 *A*). Nevertheless, calorimetric measurements of DNA binding by the PU.1 ETS domain (encoded by the C-terminal fragment, PU.1ΔN167) showed that PU.1 bound site-specific targets with non-1:1 stoichiometry (76). When site-specific DNA was titrated into PU.1ΔN167, the protein bound the DNA in a negatively cooperative manner (Fig. 2.1, *B* and *C*). Dimerization was strictly noncovalent, as PU.1ΔN167 harbored no cysteine residue. To broaden our observations and determine whether 2:1 binding was a class property of ETS domains, we measured high-affinity site-specific binding by the ETS domain of Ets-1 (encoded by the C-terminal fragment Ets-1ΔN331), which contained two free cysteines. Although PU.1ΔN167 and Ets-1ΔN331 represent sequence-divergent ETS members, their backbones are superimposable in their DNA-bound states (120). At comparable concentrations as PU.1ΔN167 (~40 μM) and under reducing but otherwise identical conditions, Ets-1ΔN331 bound site-specific DNA at strictly 1:1 stoichiometry (Fig. 2.1, *D* and *E*). Moreover, the inclusion of flanking elements known to autoinhibit Ets-1 (Ets-1ΔN280) did not affect the binding stoichiometry (Fig. 2.1, *F* and *G*). Thus, dimerization at a single site was not shared by Ets-1 but was particular to PU.1 and possibly other sequence-similar ETS homologs.

Although the ITC titrations could be fitted empirically with model-dependent profiles, the high concentrations (>10^{-5} M) required for the experiments precluded an accurate quantitative determination of binding affinities due to the strong dissociation constant of the 1:1 complex (10^{-}

^{9} M). We therefore titrated a 20-bp Cy3-labeled duplex oligonucleotide harboring the same high-affinity site as used in the ITC experiments and measured binding from changes in fluorescence polarization of the DNA probe (Fig. 2.2 *A*). The binding data yielded a biphasic profile to which a sequential binding model (111) was fitted with dissociation constants of 7.0 ± 1.3 nM and (1.2 ± 0.8) × 10^3 nM, or a (negative) cooperativity parameter of ~170. Constraining the model to 1:1 binding significantly compromised the fit to the data (*green curve* in Fig. 2.2 *A*). To determine whether a single helical turn of contact interface, as observed in the co-crystal PU.1/DNA structure (26), was sufficient to support sequential binding of PU.1, we repeated the titration with a DNA construct in which only the core 10 bp of the cognate site (5'-AGC**GGAA**GTG-3') was duplex. Binding to the 10-bp duplex exhibited monophasic binding that was well-described by a 1:1 model with a ~2-fold reduction in dissociation constant (12 ± 2 nM). To further determine whether the excess binding to the 23-bp construct represented nonspecific interactions, we measured PU.1ΔN167 binding to an isomer of the 20-bp DNA in which the core 5'-GGAA-3' consensus was mutated to 5'- GAGA-3'. In contrast with the specific site, binding to the nonspecific site was >100-fold weaker (2.1 ± 0.2 μM) than either site-specific DNA and yielded a Hill coefficient of well above unity, indicative of concerted binding of two or more equivalents of PU.1ΔN167 (Fig. 2.2 *B*). The titrations therefore showed that *sequential* dimerization of the PU.1 ETS domain, wherein excess protein self-titrated the canonical 1:1 PU.1/DNA complex to form a 2:1 complex, was exclusive to site-specific DNA longer than 10 bp. Although nonspecific DNA bound PU.1 in multiple equivalents, it did not form a 1:1 complex at equilibrium (Fig. 2.2 *B*).

To evaluate site-specific PU.1 dimerization in the presence of excess nonspecific DNA, as would be encountered under genomic conditions, we titrated a 209-bp fragment harboring a

single high-affinity PU.1-binding site with PU.1ΔN167. Resolution of the DNA by native gel electrophoresis (Fig. 2.2 *C*) showed a PU.1-bound band that transitioned to a less mobile band at ~100-fold excess protein with respect to the specific site (1 nM), in agreement with the sequential site-specific dimer observed in the fluorescence polarization titrations. The low mobility and broadness of the dimer peak suggested that this complex was interconverting between free and bound states at rates comparable with electrophoretic separation of the two complexes (121). The dimeric peak occurred in advance of a final nonspecific complex that was not detected in the gel. The latter state was verified with an isomeric nonspecific DNA fragment that failed to yield PU.1ΔN167 at any defined stoichiometry (Fig. 2.2 *D*). Thus, the electrophoretic data showed that the sequential dimerization of PU.1 at a single embedded cognate site effectively resisted competition from excess nonspecific DNA.

2.5.1 Hydrodynamic characterization of PU.1/DNA complexes by NMR spectroscopy

To characterize the solution behavior of the PU.1/DNA complex directly, we interrogated PU.1ΔN167 with site-specific and nonspecific DNA oligonucleotides hydrodynamically by diffusion-ordered NMR spectroscopy (see "Materials and methods"). At 250 μM protein, we measured a self-diffusion coefficient of $(9.1 \pm 0.1) \times 10^{-11}$ m^2/s for unbound PU.1ΔN167 (13.0 kDa) in D_2O at 25 °C. Comparison with a computed value (115) under equivalent conditions for a PU.1ΔN167 monomer derived from the co-crystal structure of the 1:1 PU.1 ETS/DNA complex (see "Materials and methods") found good agreement (8.8×10^{-11} m^2/s). An unbound PU.1 dimer modeled as a pair of rigid spheres would exhibit a diffusion coefficient at 75% of the monomer (122), or ~7×10^{-11} m^2/s. Thus, although unbound PU.1 was known to dimerize at very high concentrations (76,118) in vitro, PU.1ΔN167 was monomeric under the conditions of the DOSY experiments.

We tracked the self-diffusion coefficients of PU.1ΔN167 at graded stoichiometric ratios of DNA (Figs. S2.2 and S2.3; parametric values in Table 1). Titration with 16-bp high- or low-affinity site-specific DNA lowered the apparent diffusion coefficient to a minimum of $(5.9 \pm 0.1) \times 10^{-11}$ m^2/s at, within experimental uncertainty, a DNA/PU.1 ratio of 0.5 (*i.e.* PU.1/DNA 2:1). The subsequent addition of site-specific DNA past this point increased the diffusion coefficient to a stable value of $(7.5 \pm 0.2) \times 10^{-11}$ m^2/s at 1:1 equivalence and beyond (Fig. S2.3 *A*). This biphasic behavior was consistent with the other titration data by ITC, fluorescence polarization, and gel mobility shift. This change in DOSY intensity was not due to simple contributions from added DNA because we had carefully avoided peaks that overlapped with DNA (Fig. S2.2). The measured diffusion coefficient upon reaching molar equivalence also agreed with a computed value (115) of 7.3×10^{-11} m^2/s based on the 1:1 PU.1/DNA co-crystal structure. Finally, the sequential transitions in diffusion coefficients at half and unit molar equivalence were incompatible with a 2:2 complex. Thus, the DOSY titrations indicated that PU.1 formed exclusively a 2:1 complex with site-specific DNA at PU.1/DNA up to 2:1, followed by a 1:1 complex at molar equivalence and above.

In contrast with site-specific DNA, nonspecific binding by PU.1ΔN167 showed qualitatively different behavior (Fig. S2.3 *C*). Specifically, titration of PU.1ΔN167 with 16-bp nonspecific DNA yielded only a single inflection point at a DNA/PU.1 ratio of 0.5 and a stable diffusion coefficient of $(6.6 \pm 0.2) \times 10^{-11}$ m^2/s, between the site-specific 1:1 and 2:1 complexes. Thus, the DOSY titration data pointed to the exclusive formation of a defined dimeric nonspecific 16-bp complex. An alternative scenario in which PU.1ΔN167 formed a mixture of 2:1 and 1:1 complexes was not likely, as the apparent diffusion coefficient would be composition-dependent and change upon continued titration of DNA.

As the fluorescence polarization titration showed that the 10 bp of site-specific DNA bound PU.1ΔN167 in a 1:1 complex but was insufficient to invoke sequential dimerization, we repeated the DOSY titrations with a 10-bp duplexes. In stark contrast with their 16-bp parents, all of the 10-bp complexes regardless of sequence gave indistinguishable hydrodynamic profiles showing single inflections (Fig. S2.3, *D–F*). If the site-specific 10-bp 1:1 complexes maintained the structure observed in the co-crystal structure, their computed (115) diffusion coefficient would be 7.9×10^{-11} m^2/s under the present experimental conditions. Thus, the measured diffusion coefficients of the 10-bp complexes (~6.5×10^{-11} m^2/s), which were significantly lower even than their 16-bp 1:1 counterpart, were unexpected for a 10-bp 1:1 complex. We confirmed the 1:1 stoichiometry of the 10-bp PU.1/DNA complexes by examining the ^1H spectra of the 10-bp high-affinity DNA in the imino region at graded PU.1ΔN167 concentration. Resonances corresponding to free DNA were exhausted by unit molar protein/ DNA ratio (Fig. S2.4). Thus, PU.1 bound 10-bp DNA exclusively as monomers even at excess concentrations, and sequential dimerization of PU.1 was limited to site-specific DNA longer than 10 bp. Moreover, the data implied that the protein underwent significant conformational changes to hydrodynamically larger structures than the same protein bound to 16-bp site-specific DNA.

2.5.2 Structural properties of the site-specific PU.1 ETS dimer

We recorded ^1H-^{15}N HSQC spectra of uniformly ^{15}N-labeled PU.1ΔN167, which showed a structured protein with well-dispersed cross-peaks in the absence of DNA (Fig. 2.3*A*). Upon the addition of 16-bp high-affinity DNA (Fig. 2.3, *B–D*) to a DNA/PU.1 ratio of 1:2, we observed the immediate disappearance of ~80% of the NH resonances and a marked loss of chemical shift dispersion, but no sign of precipitation even after prolonged incubation (~24 h). The addition of a second half-equivalent of DNA to DNA/PU.1 1:1 promptly restored the NH

resonances, with significant chemical shift perturbations compared with the free protein. The further addition of excess DNA produced no further changes to the HSQC spectrum.

When we repeated the HSQC titration using the same high-affinity 10-bp construct as in the DOSY experiments, we observed a progressive disappearance of resonances past the half-equivalence point and no further change past the 1:1 equivalence point (Fig. 2.3, *E–H*). The monotonic transition for the 10-bp DNA tracked the changes in diffusion coefficient (Fig. 2.3, *E–H*), in contrast with the sharply biphasic behavior of the 16-bp site.

To better understand the effect of binding site size on the bound protein's conformation, we probed DNA-bound PU.1ΔN167 with 8-anilino-1-naphthalenesulfonate (ANS), an indicator dye of solvent-exposed hydrophobic moieties. Unbound PU.1 at 50 μM induced strong blue-shifted ANS fluorescence associated with a significant number of untitrated basic residues, which paired with the anionic dye (123), in the absence of DNA (Fig. 2.4 *A*). DNA alone induced a negligible effect on ANS fluorescence. After subtraction of a dye-only control, ANS fluorescence of PU.1 was reduced about 3-fold upon binding a half-equivalent of the 16-bp site-specific DNA (2:1 complex), and another 8-fold at unit equivalence (1:1 complex) (Fig. 2.4 *B*). The higher sensitivity to ANS, together with the NMR DOSY data, suggest that the DNA-bound PU.1 dimer may be less structured than in the 1:1 complex.

2.5.3 Topology of the site-specifically bound PU.1 dimer

The attenuation in NMR signal from the 2:1 complex suggests intermediate exchange between these two states. As a result, although the disappearance of 80% of cross-peaks in the DNA-bound PU.1 dimer (*cf.* Fig. 2.3 *B*) precluded a direct identification of the residues involved in 2:1 complex formation, the remaining resonances still provided valuable clues to the location of the dimerization interface. We overlaid the HSQC spectra for free and bound PU.1 to 16-bp

DNA and identified resonances that showed strong overlap throughout the titration (Fig. 2.5, *A*–*E*). Using a reported ^{1}H-^{15}NHSQCassignment of the PU.1 residues 167–260 (118), which applied well to PU.1ΔN167 (Fig. S2.5), we identified well-resolved, well-overlapped resonances for Arg^{173}, Ala^{210}, Lys^{224}, Gly^{238}, Gly^{241}, Lys^{247}, and Ser^{255}. Resolvable resonances that overlapped only in the free and 1:1-bound states (i.e. no detectable signal in the 2:1 state), including Ser^{184}, Trp^{192}, Trp^{193}, Asp^{197}, and Thr^{200}, mapped to solvent-exposed surfaces away from the DNA (Fig. 2.5 *F*). Thus, the HSQC data implicated the distal surface of PU.1ΔN167 opposite the DNA-binding site as a major part of the dimerization interface of the site-specific 2:1 complex.

To further understand how the distal surface of the PU.1 ETS domain was involved in dimerization, we examined the amino acids that mapped to that surface and noticed a sequence of four alternately charged residues, ^{195}DKDK^{198}, that comprise part of a β-pleated sheet. These residues include (Asp^{197}) or are proximal to residues (Trp^{192}, Trp^{193}, and Thr^{200}) whose resonances became reversibly undetectable in the 2:1 complex (Fig. 2.5, *A* and *D*). The ^{195}DKDK^{198} sequence gave rise to a charge distribution that suggested an electrostatically complementary interface for dimerization (Fig. 2.5 *G*). This hypothesis was further motivated by the low level of sequence conservation in Ets-1 (^{357}TGDG^{360}) and within the ETS family in general (113). We therefore cloned a PU.1ΔN167 mutant harboring ^{195}NINI^{198}, which abrogated the charges but maintained similar side-chain structures and secondary structure propensities (124). In fluorescence polarization and gel mobility shift experiments, the mutant gave titration profiles that showed a single binding mode at up to 10 μM, a concentration at which DNA-bound wild-type PU.1 had undergone two binding transitions (Fig. 2.5 *H*; *cf.* Fig. 2.2). The anisotropy and electrophoretic mobility of the bound mutant corresponded to the 1:1 complex formed by wild-type protein. Thus, the ^{195}NINI^{198} mutant confirmed that the distal surface was involved in

PU.1 dimerization. In addition, the mutant bound DNA ~10-fold more weakly (88 ± 11 nM) than wild-type PU.1ΔN167, suggesting coupling between DNA binding and dimerization of the bound state.

2.5.4 The 2:1 site-specific PU.1/DNA complex occupies an expanded DNA-binding site

To define the contact interface of the 2:1 PU.1/DNA complex, we carried out hydroxyl radical (·OH) footprinting titration of a 130-bp radiolabeled DNA fragment harboring the same high-affinity binding site used in the other experiments. Previous footprinting studies of the 1:1 site-specific PU.1/DNA complex by our group (76,114) and others (125,126) have established that two spaced clusters of minor-groove contacts flanking the 5'-GGAA-3' core consensus generate a highly characteristic ·OH footprint on the 5'-TTCC-3' strand. We therefore used this biochemical signature to track changes in the DNA-binding site at graded concentration of wild-type PU.1ΔN167 (Fig. 2.6 *A*). Upon reaching PU.1ΔN167 concentrations of ~10^{-7} M that saturated 1:1 complex (marked *P1* and *P2* in Fig. 2.6 *B*), additional PU.1ΔN167 gave rise to protected positions near P2 (marked *P**). In total, the footprint of the 2:1 complex spanned ~20 bp of DNA. Quantitation of the protection from ·OH at the protected bases as a function of PU.1 concentration clearly recapitulated the sequential formation of the 1:1 followed by the 2:1 complex observed in the other experiments (Fig. 2.6 *C*). In addition, the hypersensitive positions between P1 and P2, which is also diagnostic of site-specific 1:1 binding, was preserved in the 2:1 complex and tracked the titration profiles produced by fluorescence anisotropy (*cf*. Fig. 2.2 *A*).

Thus, the ·OH footprints showed, at single-nucleotide resolution, that the site-specific PU.1ΔN167 dimer made extended contacts with the DNA minor groove although the dimerization interface was distal from the DNA. The extended footprint exerted by the 2:1

complex also explained the 10-bp specific site's apparent incompatibility with PU.1 dimerization, which required several more flanking bases downstream of the 5'-GGAA-3' core consensus, although its affinity for the 1:1 complex was only modestly compromised relative to longer DNA (*cf.* Fig. 2.2 *A*).

2.6 Discussion

When bound to sufficiently long site-specific DNA, the ETS domain of PU.1 self-associates sequentially to a defined dimer, a behavior that is heretofore unknown for ETS transcription factors (19). The reversibility of dimerization is demonstrated by its independence on directionality. In gel mobility shift, fluorescence anisotropy, or DNA footprinting titrations in which excess protein was titrated into site-specific DNA, the 2:1 complex was produced subsequently to the canonical 1:1 complex. When the titration was reversed (DNA into protein), as was the case in the NMR and ITC studies, limiting concentrations of DNA directly yielded the 2:1 complex. Impressively, the NMR titrations showed that even after prolonged co-incubation, the further addition of site-specific DNA converted the 2:1 complex rapidly and quantitatively into its 1:1 counterpart. Structurally, the identification of the solvent-exposed surface distal from the DNA as the dimerization interface on the one hand, and the expansion of the DNA footprint of the 2:1 complex on the other, suggested an allosteric coupling between the PU.1 dimers and their bound DNA.

Although many DNA-binding domains are known to self-associate when they bind to site-specific DNA (127), this behavior is associated with systems in which the protein protomers bind multiple DNA subsites independently, such as the *Trp* repressor (128), or with positive cooperativity, such as the p53 core domain (129). Ets-1 and several other non-PU.1 ETS members can also bind as homodimers, but only to *two* tandem DNA sites (74,106-109). To our

knowledge, *negatively* cooperative binding to a single DNA site, as the PU.1 ETS domain is able to execute, has not been reported previously.

2.6.1 Flanking sequence length as a specificity determinant of PU.1/DNA binding

We tested a range of DNA lengths to define the site size requirements for PU.1 dimerization in the bound state and to probe the relevance of dimerization in the presence of excess nonspecific DNA. We observed that the binding modes accessible to the ETS domain of PU.1 depended on a threshold length of bases flanking the core 10-bp binding site. Within the range of DNA lengths tested in the various experiments, 16-bp and longer DNA invoked sequential dimerization of bound PU.1. In stark contrast, in the absence of flanking bases, the 10-bp DNA bound PU.1ΔN167 exclusively with 1:1 stoichiometry (Figs. 2.2 *B* and 2.3). Thus, 10 bp of site-specific DNA was insufficient to elicit the full site-specific behavior of the PU.1 ETS domain. Available evidence indicates that flanking sequence identity is not a determinant because we had observed two other site-specific DNA sequences yielding the same ITC profiles for PU.1ΔN167 (76).

The 10-bp complex represented a distinct binding mode as the bound PU.1 monomer was structurally different from its 16-bp counterpart as judged by their HSQC spectra (Fig. 2.3). This observation was unexpected, given the single turn of contacted double helical DNA in the co-crystal structure of the high-affinity PU.1/DNA complex (26) and ETS/DNA structures more generally. Of relevance is the report that DNA with staggered ends was absolutely required for diffraction-quality crystals of the PU.1/DNA complex (130). The overhangs, which paired end-to-end between asymmetric units, would result in essentially continuous DNA in the crystal. Additional interactions with flanking bases that are not part of the core sequence therefore stabilize the bound protein, and without this stabilization, dimerization becomes prohibitive. In

summary, flanking sequence length represents an essential additional determinant to fully specify cognate binding by PU.1 in solution.

2.6.2 Functional relevance of self-titration as a potential negative feedback mechanism for PU.1 transactivation

PU.1 is a highly inducible protein, ranging from <10 to >200 copies of mRNA per cell in murine bone marrow progenitors, depending on the stage of hematopoietic development (131). Under physiologic induction, PU.1 mRNA levels matching and even exceeding that of glyceraldehyde 3-phosphate dehydrogenase, an abundant glycolytic housekeeping enzyme, have been measured in cultured (132) and primary (133) human cells. This inducible expression profile suggests that interactions spanning a large range in affinity are likely to be biologically relevant. For instance, NMR characterization of the functionally essential PU.1/GATA-1 interaction estimated its dissociation constant to be no stronger than 10^{-4} M *in vitro* (134) and did not appear to involve (as judged by chemical shift changes) the dimerization interface of PU.1.

Many ETS family transcription factors, such as Ets-1, ERG, and members of the ETV subfamily, are regulated at the protein/ DNA level by inhibitory helices that pack against their DNA-binding domain in the unbound state (Fig. 2.1 *A*). Perturbing these helices imposes an energetic penalty on DNA-binding that maintains, by default, an autoinhibited state. Binding partners that disrupt the autoinhibitory interactions thus induce a transcriptionally permissive state (30). ETS paralogs, such as PU.1, that lack this mechanism would therefore be locked in a permissive state in the absence of some mechanism for negative regulation. Whereas functionally antagonistic binding partners, such as GATA-1, would serve such an inhibitory role, their expression profiles only partially overlap with that of PU.1 (the common myeloid progenitor in the case of GATA-1, 135). An intrinsic negative feedback mechanism is hitherto

unknown in PU.1. Our observation that PU.1 forms a reversible, negatively cooperative 2:1 complex with site-specific DNA suggests "self-titration" as a potential mechanism of negative feedback: even if the 2:1 complex retains the functional activity of the 1:1 complex, removal of circulating PU.1 alone would attenuate transactivation of target genes. Consistent with this notion, we observed self-titration only with site-specific DNA and not nonspecific DNA. Moreover, we did not observe dimer formation with the structural homolog Ets-1, with or without its autoinhibition helices, when its cysteines were maintained in a reduced state (Fig. 2.1, *D*–*G*). Interestingly, a 2:1 Ets-1/DNA complex was reported under non-reducing conditions (71), reflecting the strong propensity for its two cysteine residues (which are not present in the PU.1 ETS domain) to form nonnative disulfide linkages.

The dissociation constant for binding to oligomeric nonspecific sites (~10^{-6} M) (114), such as that used in our NMR experiments, is only ~10-fold higher than the sequential affinity of the second equivalent of PU.1ΔN167. It might therefore appear that the abundance of nonspecific DNA relative to specific sites would overwhelm self-titration of specific complexes. Our gel mobility data on binding to polymeric DNA (cf. Fig. 2.2 *C*) provide a useful insight into this question. Compared with titration of oligomeric site-specific DNA, formation of the 2:1 complex at an embedded binding site flanked by substantial nonspecific DNA (~100 bp on each side) occurred at ~10-fold lower concentration (~10^{-7} M) and clearly preceded any nonspecific binding. The footprinting data showed the same behavior at a shorter (~130-bp) DNA fragment. This difference reflected the favorable contribution to reaching an embedded site from linear diffusion that was absent for an isolated counterpart. Thus, a complete description of the effect of excess nonspecific flanking DNA (as would be expected under genomic conditions) includes a competitive effect that is more than offset by favorable contributions from linear diffusion.

2.6.3 Nonspecifically bound PU.1 is oligomeric

The 16-bp nonspecific site, involving only the isomeric reversal of two adjacent positions in the core consensus (5'- GGAA-3' to 5'-GAGA-3'), forced the exclusive formation of a dimeric complex. No 1:1 complex was detectable at equilibrium. In the context of self-titration as a potential negative regulatory mechanism, this behavior suggests that the role of site-specific DNA (*i.e.* sequences harboring the core consensus) is not only to provide a much higher-affinity binding site for PU.1 but, perhaps more importantly, to "unlock" the transcriptionally active 1:1 conformation. It may therefore be more appropriate to consider dimeric PU.1, as the default autoinhibited state, which becomes activated, by a coupled dissociation/order transition, upon encountering a specific DNA site at permissively low protein concentrations.

2.7 Conclusion

We report, for the first time, a 2:1 complex formed by PU.1 with a *single* cognate binding site. This complex forms negatively cooperatively with respect to the canonical, transcriptionally active 1:1 complex and resists competition from nonspecific DNA. It is kinetically stable (on the order of many hours) and interconverts efficiently with the 1:1 complex (within minutes) upon the addition of DNA. These biophysical properties of self-titration of PU.1 at site-specific DNA are biologically compatible and, indeed, physiologically appropriate given the significant accumulation of PU.1 under induction ($>10^{-6}$ M), when negative feedback would be most required to dampen its transcriptional response. Self-titration therefore represents a potential buffering mechanism for self-regulation in ETS paralogs, such as PU.1, that lack autoinhibitory elements in their structures.

Table 2.1 Apparent translational self-diffusion coefficients of PU.1/DNA complexes.

Diffusion coefficients ($\times 10^{-11}\ m^2\ s^{-1}$), as plotted in Figure 2.3 in the main text, of PU.1ΔN167 alone and in complex with 16-bp and 10-bp high-affinity (5'-GCAAGC**GGAA**GTGAGC-3'), low-affinity (5'-GCAAAA**GGAA**TGGAGC-3'), and nonspecific DNA (5'-GCAAGCGAGAGTGAGC-3'). The 10-bp DNA sites consist only of the underlined core sequences.

DNA to PU.1 molar ratio	*16-bp high-affinity*	*16-bp low-affinity*	*16-bp nonspecific*	*10-bp high-affinity*	*10-bp low-affinity*	*10-bp nonspecific*
0	9.11±0.17	9.00±0.15	9.40±0.13	9.08±0.17	8.90±0.59	9.29±0.24
1/6	8.15±0.25	7.80±0.25	7.73±0.15			8.26±0.33
1/4				8.30±0.24	8.03±0.45	
1/3	7.27±0.15	6.68±0.11	6.50±0.21			7.68±0.08
1/2	6.21±0.18	5.84±0.16	6.60±0.22	7.11±0.13	7.54±0.32	7.08±0.23
2/3	5.90±0.11	5.92±0.11	6.70±0.22			6.80±0.14
3/4				6.72±0.06	6.69±0.19	
5/6	6.53±0.14	6.56±0.13	6.59±0.14			6.49±0.25
1	7.00±0.14	7.38±0.12	6.70±0.10	6.27±0.10	6.68±0.18	6.88±0.18
1 1/6	7.72±0.27		6.46±0.27			
1 1/4				6.64±0.14	6.43±0.19	7.02±0.10
1 1/3	7.75±0.06		6.51±0.17			
1 1/2	7.53±0.23	7.79±0.08		6.30±0.14	6.62±0.16	7.00±0.18
1 3/4				6.81±0.10		
2	7.56±0.10	7.84±0.15			6.59±0.28	

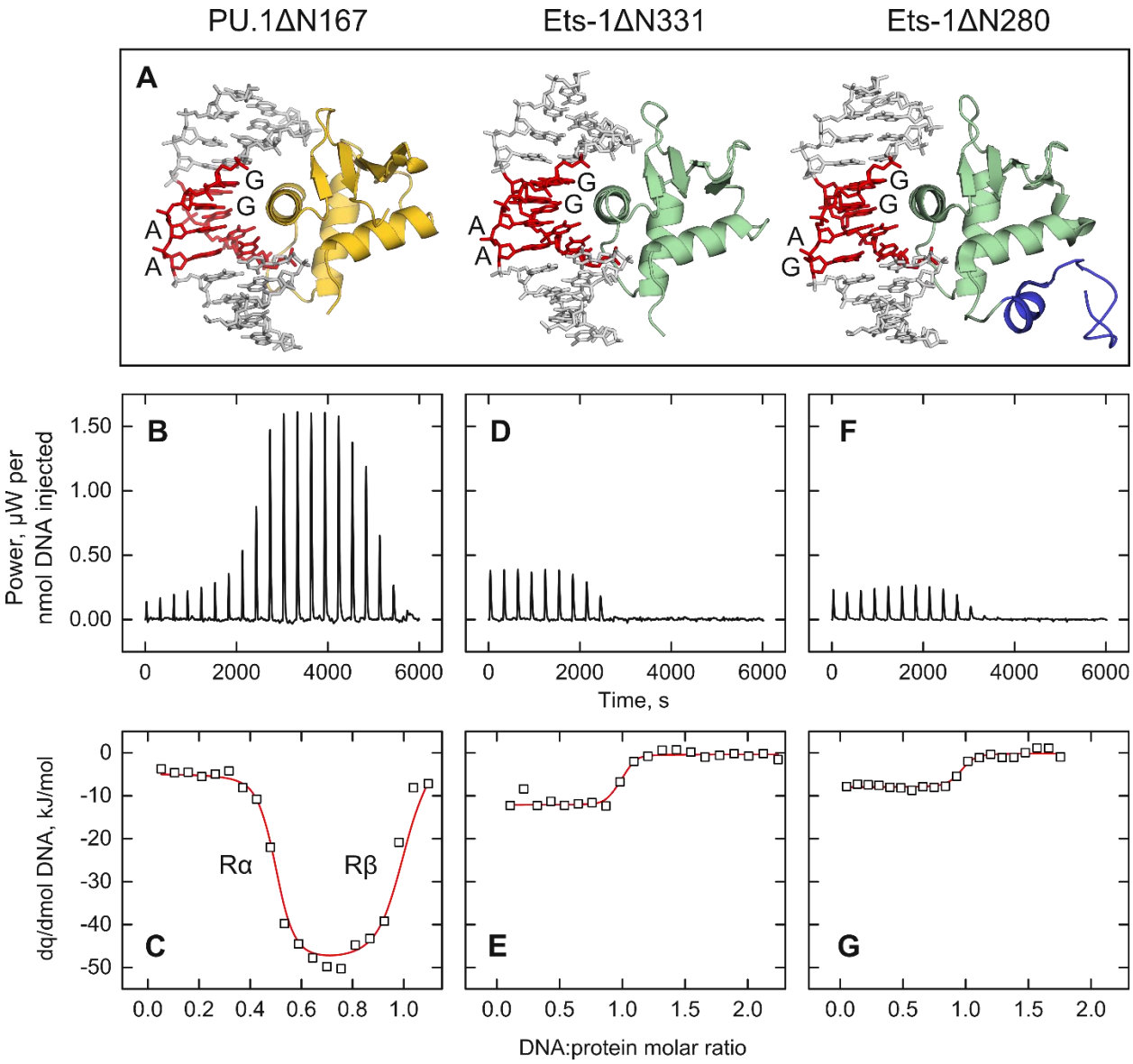


Figure 2.1 Dimerization at a single cognate binding site is intrinsic to the ETS domain of PU.1, but not its structural homolog Ets-1.

A, the co-crystal structures of PU.1 (*gold*; PDB code 1PUE) and Ets-1 (*green*), the latter with (1MDM) or without (1K79) part of its autoinhibitory domain (*blue*). All three structures show the canonical 1:1 binding stoichiometry with oligomeric DNA harboring a core 5'-GGAX-3' consensus (*red*), as labeled. Note that the cognate DNA sequences in the co-crystal structures are not sequence-identical to the experimental sequences in this study. *B*, *D*, and *F*, representative ITC thermograms at 25 °C of DNA-into-protein titrations for the ETS domains of PU.1 (*B*) and Ets-1 (minimal ΔN331 (*D*); autoinhibited ΔN280 (*F*)). The ordinate is baseline-subtracted and normalized to the amount of DNA delivered per injection to aid comparison; exothermic response is upward. *C*, *E*, and *G*, the titration data for PU.1ΔN167 was empirically fitted as a

negatively cooperative interaction. The two phases in the PU.1 titration (marked *Rα* and *Rβ* in *C*) represent the successive formation of a protein/DNA 2:1 complex (protein in excess) followed by the 1:1 complex. For Ets-1ΔN331 (*E*) and Ets- 1ΔN280 (*G*), a 1:1 model was empirically fitted to the data. The stronger and more complex apparent heats associated with the PU.1ΔN167 titrations included the dimerization and binding of PU.1ΔN167 as a 2:1 complex, which dissociates to two 1:1 complexes as DNA reached molar equivalence, in addition to more enthalpically driven 1:1 binding than Ets-1. The details of the thermodynamic deconvolution are provided under “Materials and methods.”

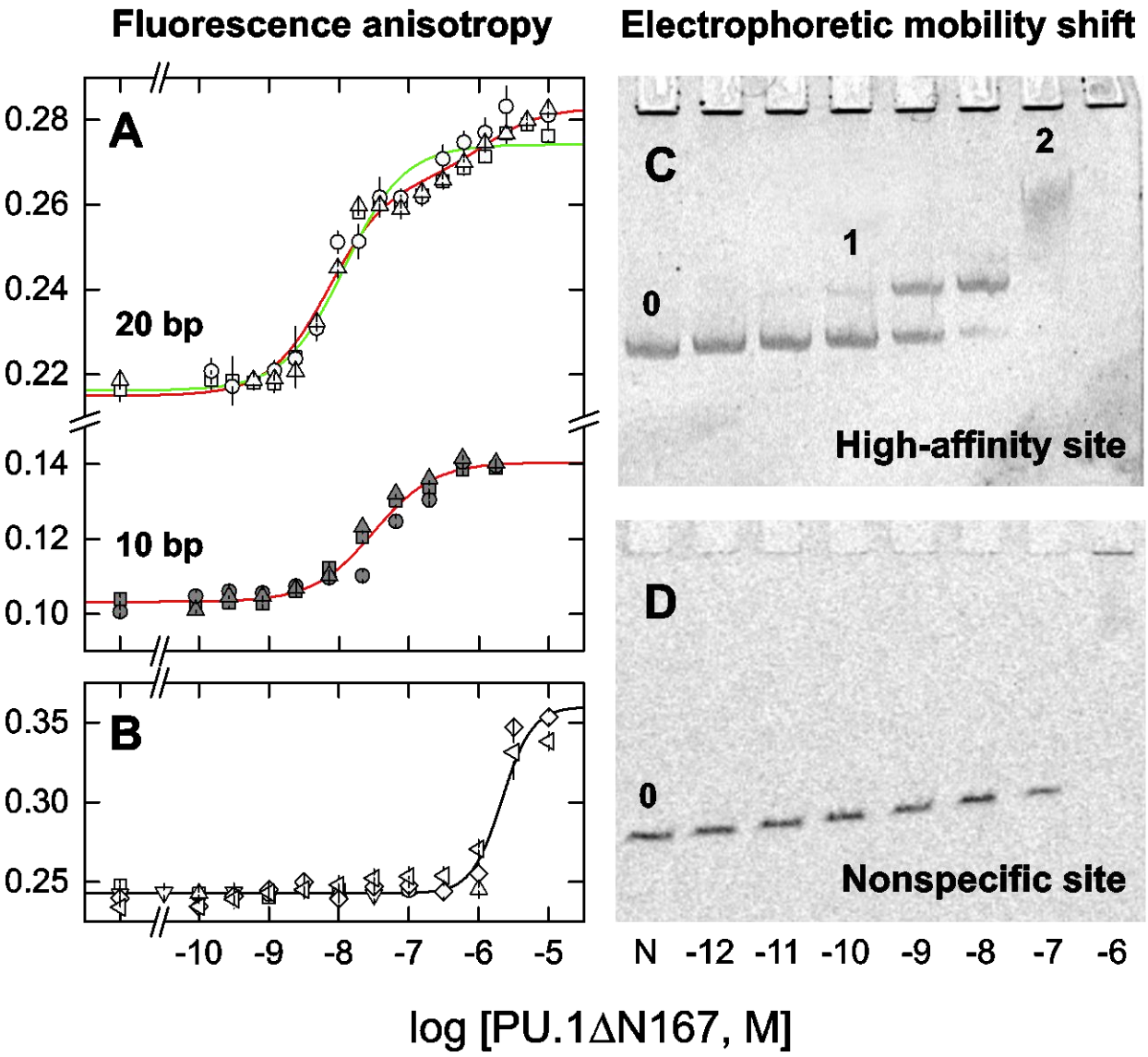


Figure 2.2 Sequence and site size requirements for sequential dimerization of the specific DNA-bound ETS domain of PU.1.

Fluorescence polarization titration of Cy3-labeled 20-bp (open symbols) and 10-bp (gray) DNA probes (0.5 nM) harboring the high-affinity site 5'-AGC**GGAA**GTG-3' (*A*) and its isomeric nonspecific variant 5'-AGCGAGAGTG-3' (*B*) with PU.1ΔN167. *Curves* represent a least-square fit of the data from triplicate experiments to a sequential 2:1 binding model (red) or a constrained to a 1:1 model (green). The latter afforded a poor fit of the data ($p < 1 \times 10^{-4}$, Fisher's *F*-test on sums of squares). The nonspecific data was fitted with the Hill equation (black). *C*, Native electrophoretic mobility shift titration of a 209-bp DNA fragment (1 nM, marked "0") harboring the same high-affinity site with PU.1ΔN167. Following formation of the 1:1 complex (marked "1"), a discrete, low-mobility species was present at 0.1 μM protein (marked "2"). At 1 μM protein

(10^3-fold excess), a nonspecific complex finally forms, which did not enter the gel, as confirmed with a fragment harboring the nonspecific sequence (*D*). The shadows lining the wells in Panel *C* represent an irregular imaging artefact of the stained gel, not protein-DNA complex, as it was observed even in the negative-control lane containing no PU.1 (marked “N”).

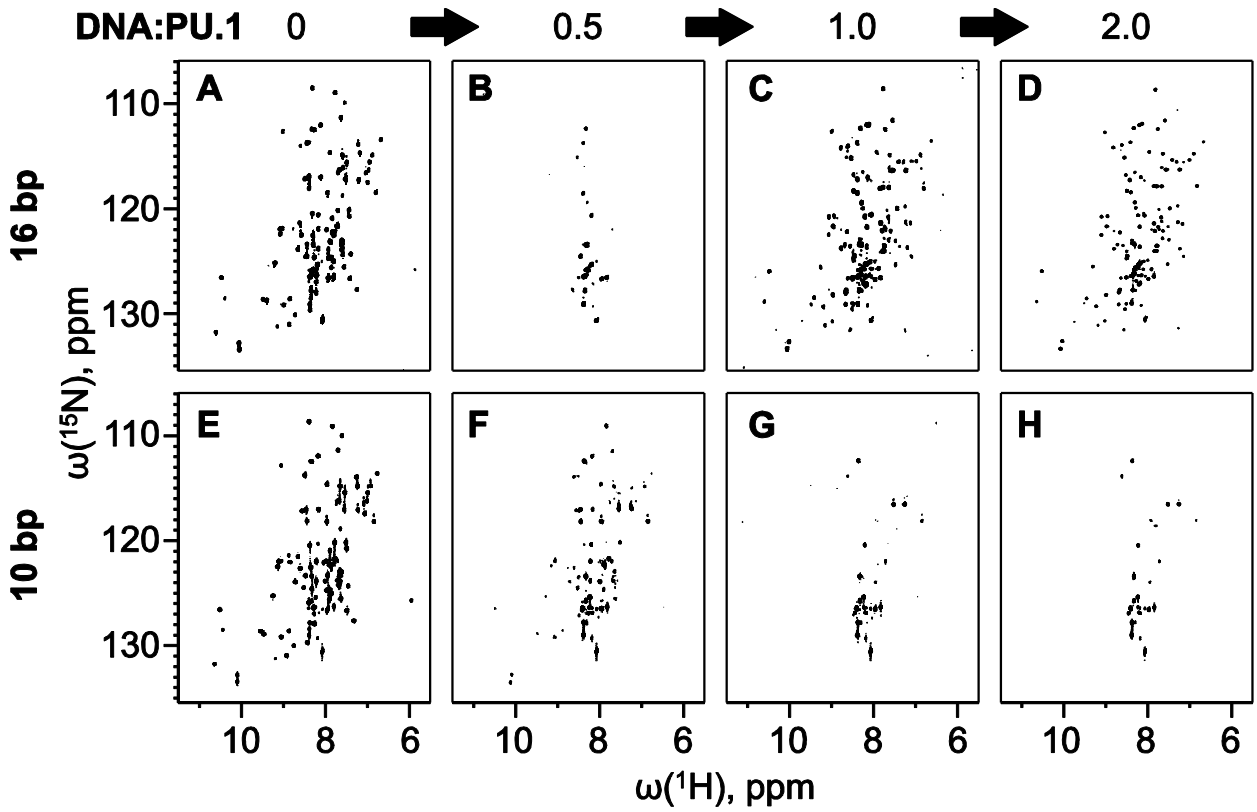


Figure 2.3 ^{1}H-^{15}N HSQC NMR spectroscopy of PU.1/DNA complexes.

Uniformly ^{15}N-labeled PU.1ΔN167 was titrated with a 16- (*A* to *D*) or 10-bp (*E* to *H*) unlabeled high-affinity DNA at the indicated molar ratios. Each series of spectra was acquired with the same sample and intensity adjusted to the same noise level.

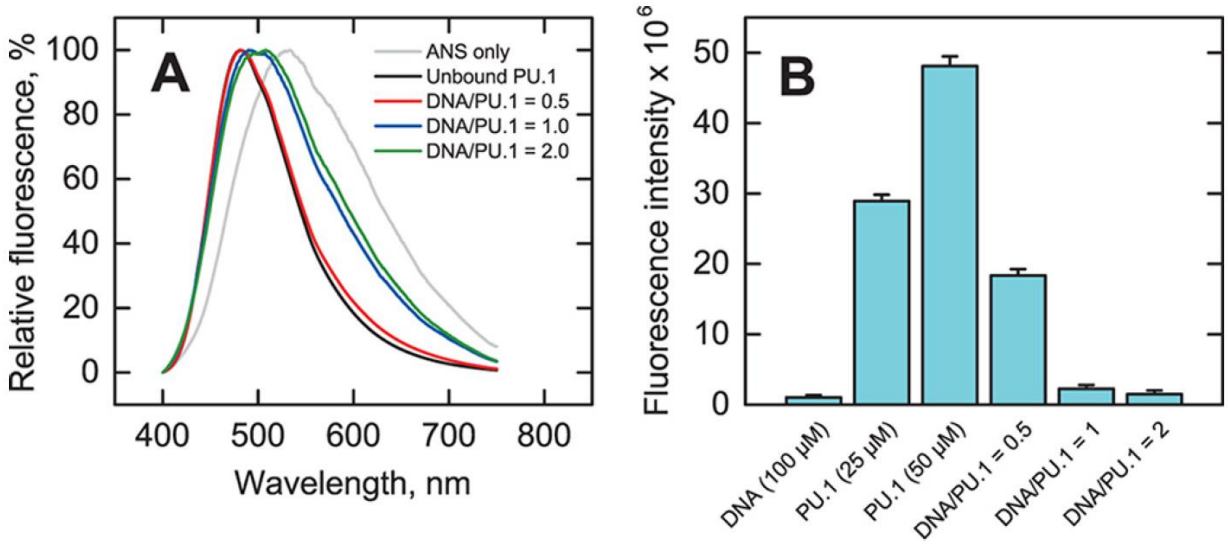


Figure 2.4 Biochemical characterization of PU.1/DNA complex conformation.

A, fluorescence spectra of 50 µM PU.1ΔN167 alone or with 0.5, 1.0, or 2.0 molar eq of 16-bp site-specific DNA after mixing with 200 µM ANS. Samples were excited at 375 nm. *B*, fluorescence intensity at 530 nm of PU.1ΔN167 with or without 16-bp DNA after subtraction of an ANS-only control, shown as average ± S.D. (*error bars*) of triplicate experiments.

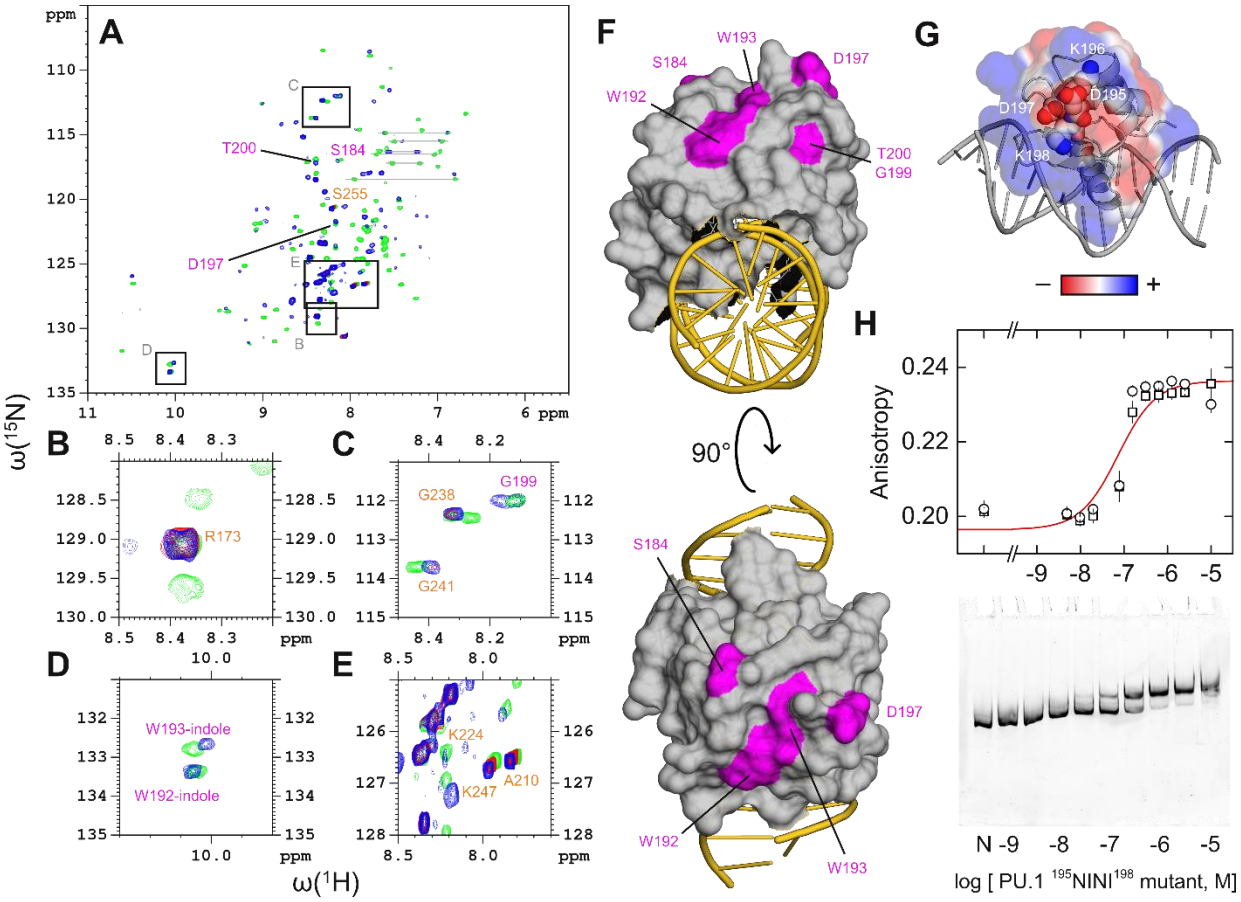


Figure 2.5 Mapping the dimerization interface of the site-specific 2:1 complex.

A, Overlay of ^{1}H-^{15}N HSQC spectra in the absence (*green*) or presence of 16-bp site-specific DNA at 0.5 (*red*) and 1.0 (*blue*) molar ratios. Peaks labeled in orange that showed strong overlap among all three states (*blue*/*red*/*green*) were taken to represent residues not involved in site-specific dimerization. Peaks labeled in *purple* that overlapped *only* in the unbound and 1:1-bound states (*blue*/*green*) were taken to represent residues involved in dimerization. Assigned resonances were as reported for residues 167 to 260 by Jia *et al.* (118). *Boxes* indicate regions that are magnified in Panels *B* to *E*. *F*, Mapping of the (*purple*) residues implicated in PU.1 dimerization to the 1:1 co-crystal structure (PDB: 1PUE). *G*, continuum electrostatic surface potential of PU.1 in the co-crystal structure. The residues ^{195}DKDK^{198} are shown as spheres. *H*, DNA-binding profiles of a ^{195}NINI^{198} mutant of PU.1ΔN167 by fluorescence polarization (20 bp) and gel mobility shift (209 bp) under the same experimental conditions as in Fig. 2.2. Symbols represent replicate experiments; the curve represents a 1:1 fit to the data. *Error bars*, S.E.

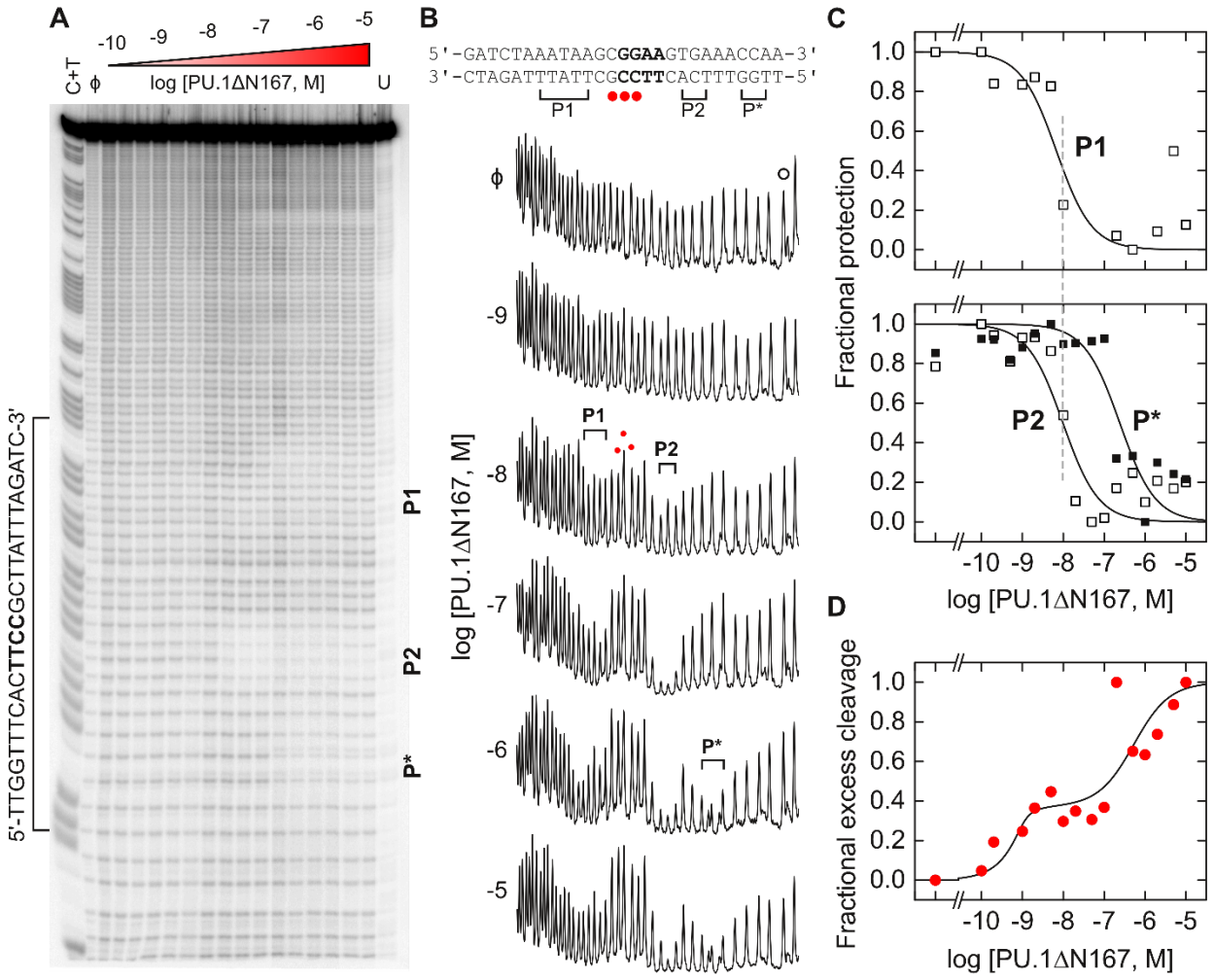


Figure 2.6 Dynamics Expansion of the DNA contact interface in the 2:1 PU.1/DNA complex.

A singly end-radiolabeled DNA fragment was titrated at equilibrium with PU.1ΔN167 and digested with •OH under single-hit conditions. A C+T reaction was included to index the digested DNA following denaturing electrophoresis. *A,* Image of the sequencing gel. N and U denote DNA digested without protein and undigested DNA, respectively. A second footprint was observed at a cryptic binding site (5'-ATG**GGAA**TTC-3') encoded by pUC19 vector further downstream from the cloned high-affinity site. The lower affinity of this site (136) meant that it did not generate the sequential 2:1 complex beyond the 1:1 footprint at the maximum PU.1 concentration used. *B,* Traces of the indicated lanes. Brackets and red dots denote protected and hyper-sensitive positions at the indicated and higher protein concentrations, respectively, relative to a distal control peak marked with a hollow dot (○). *C,* Titration of the summed integrated intensities of the protected bases marked P1 and P2 (□) associated with the 1:1 complex, and P* (■) produced by the 2:1 complex in Panel B, normalized to the control peak intensity and scaled to [0,1]. *Curves* represent empirical fits to the Hill equation. *D,* Titration of the summed

integrated intensities of the hyper-sensitive peaks (●), scaled to [0,1] but normalized to the intensity at the highest PU.1ΔN167 concentration tested. The *curve* represents a fit by a sequential 2:1 binding model.

2.8 Supplemental information

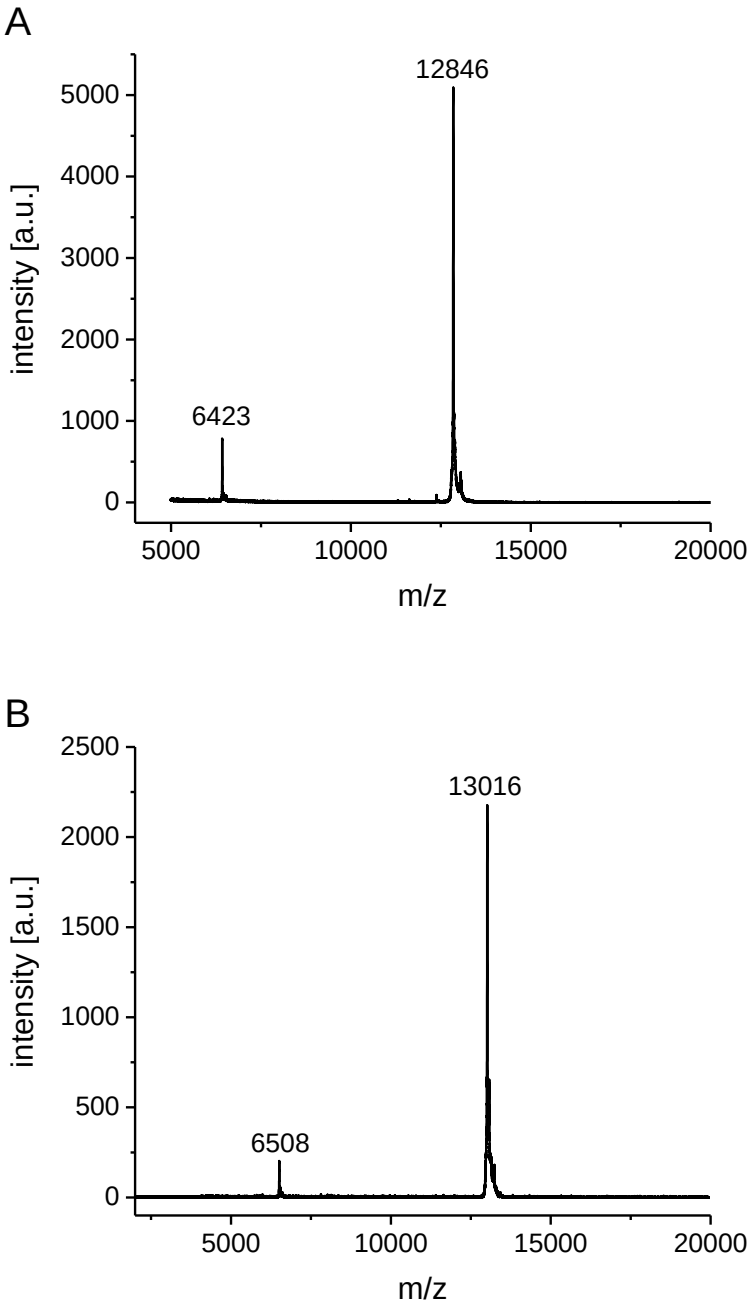


Figure S 2.1 MALDI-ToF spectra of unlabeled and ^{15}N-labeled PU.1ΔN167.

The expected MW for the unlabeled (*A*) and ^{15}N-labeled (*B*) constructs were 12,847 and 13,018 (assuming 99% enrichment), respectively. Both the +1 and +2 ions were detected.

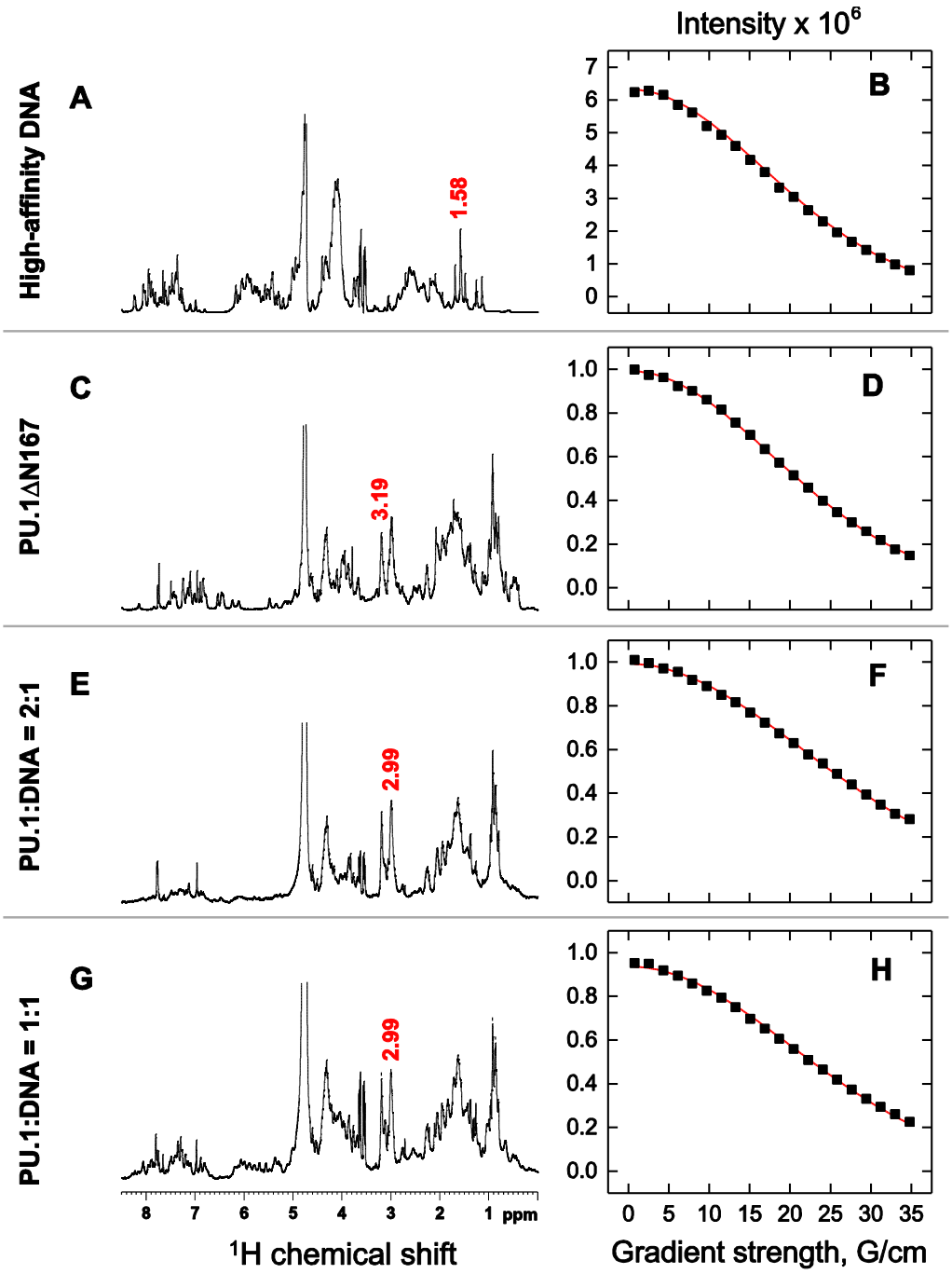


Figure S 2.2 Diffusion ordered NMR (DOSY) spectra of PU.1 ETS domain, target DNA, and their complexes.

Self-diffusion of unlabeled PU.1ΔN167, 16-bp high-affinity DNA, and mixtures of the two at 1:1 and 2:1 molar ratios was determined in solution using pulsed field gradients. Protein concentrations were 250 µM (*C*), 204 µM (*E*), and 173 µM (*G*), and the DNA concentrations were 562 µM (*A*), 102 µM (*E*), and 173 µM (*G*). At each gradient strength, the labeled peaks were individually fitted to Eq. (E1) to estimate the diffusion coefficient and then averaged. Fitted curves

of intensity decay for representative peaks at the indicated chemical shifts are shown in Panels *B*, *D*, *F*, and *H*.

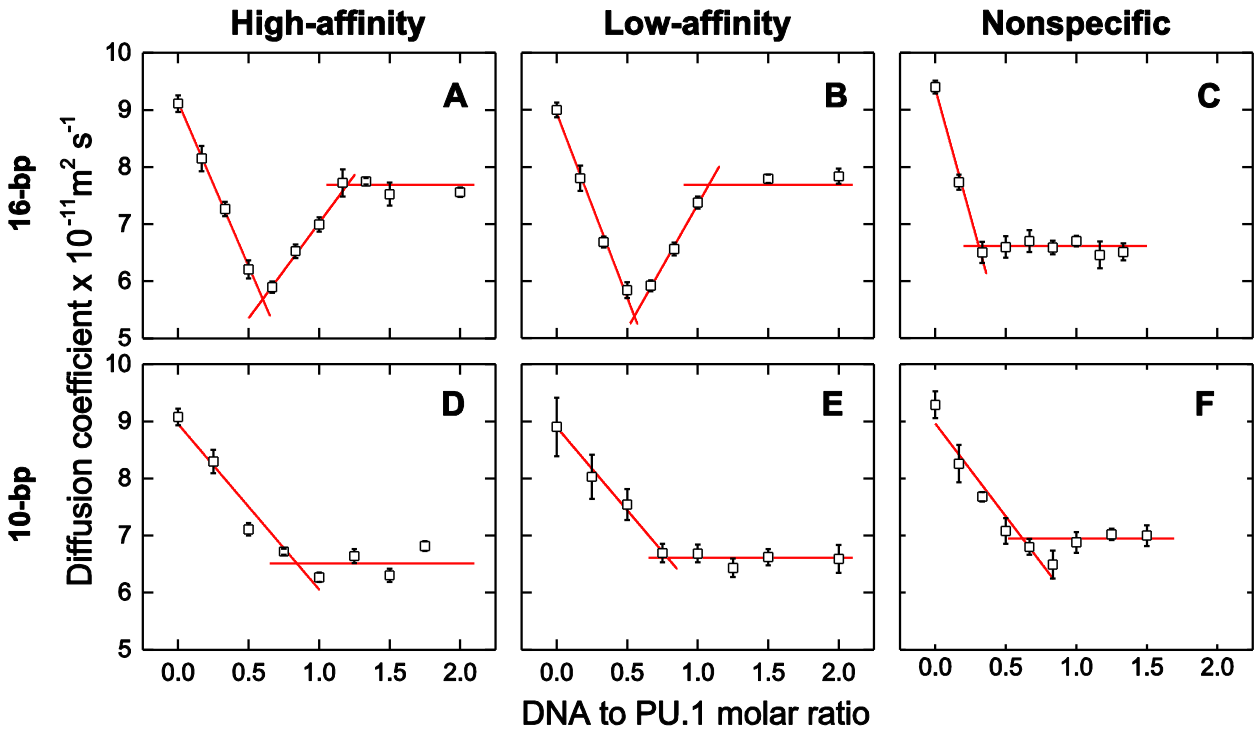


Figure S 2.3 DOSY titrations reveal site requirements for dimerization of DNA-bound PU.1 in solution.

Translational self-diffusion coefficients of PU.1ΔN167 alone and bound to a 16-bp high-affinity (*A*, 5'-GCAAGC**GGAA**GTGAGC-3'), low-affinity (*B*, 5'-GCAAAA**GGAA**TGGAGC-3'), or nonspecific DNA sequence (*C*, 5'-GCAAGCGAGAGTGAGC-3'). Measurements were repeated using 10-bp duplex sites harboring only the underlined sequences under the same solution conditions (*D* to *F*). *Lines* represent linear fits of the data in the indicated ranges. The diffusion coefficients of the 16- and 10-bp DNA alone were (10 ± 1) and (14 ± 1) × 10^{-11} m^2 s^{-1}, respectively. *Error bars*, S.D.

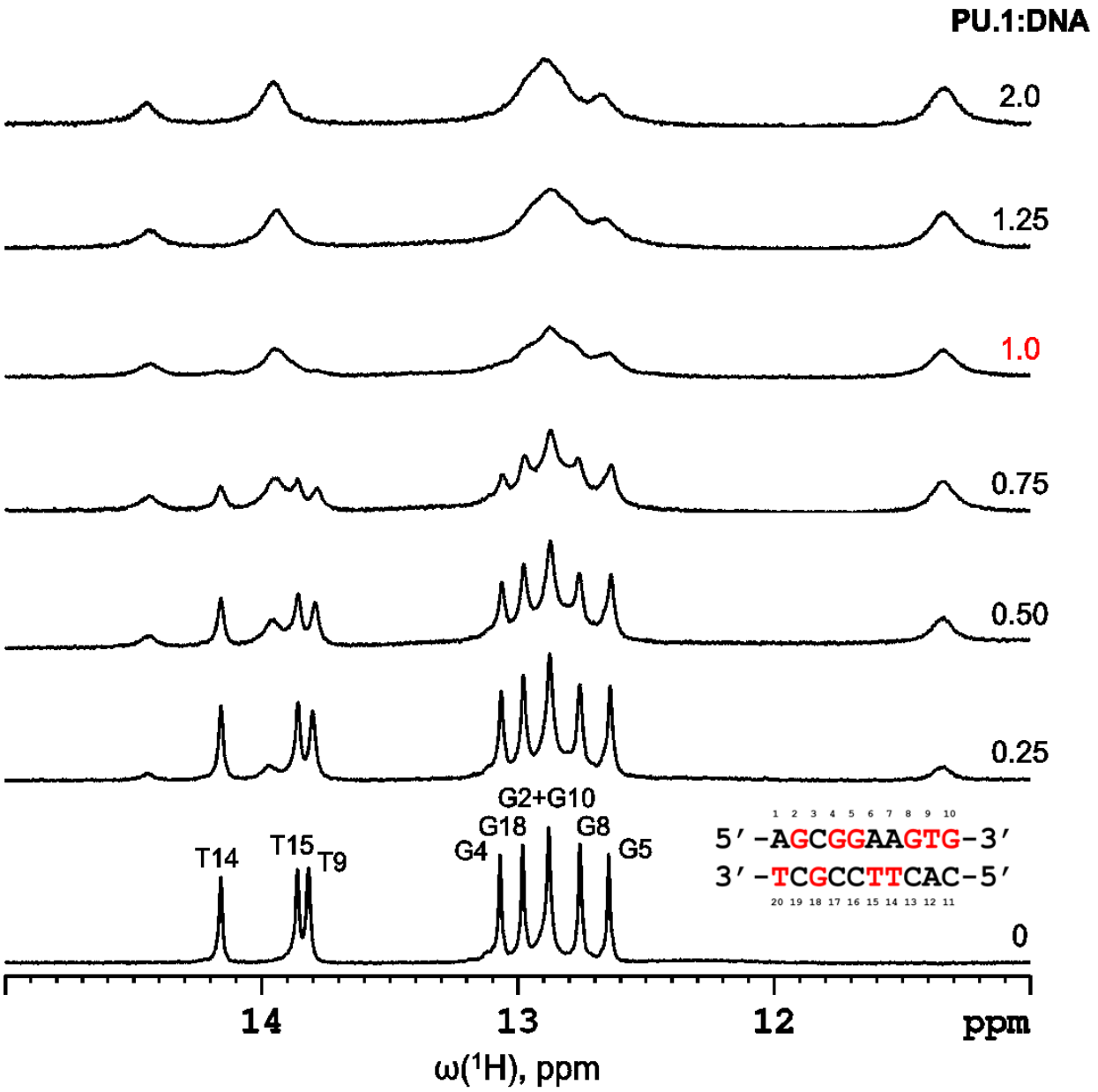


Figure S 2.4 NMR spectroscopic changes to 10-bp site-specific DNA upon titration by the ETS domain of PU.1.

Chemical shift perturbations of imino ^{1}H resonances were monitored upon titration with 250 µM PU.1ΔN167 at the indicated molar ratios at 20°C using 1-1 Jump and Return pulse sequence (600 MHz) (137). Each spectrum was referenced and normalized in intensity to DSS. Since the DOSY titration showed a single transition, resonances from unbound DNA would be expected to be persist up to the stoichiometric ratio of the complex. For the 10-bp high-affinity DNA, the observable imino ^{1}H peaks in the unbound 10-bp high-affinity DNA were fully exhausted by unit molar equivalence.

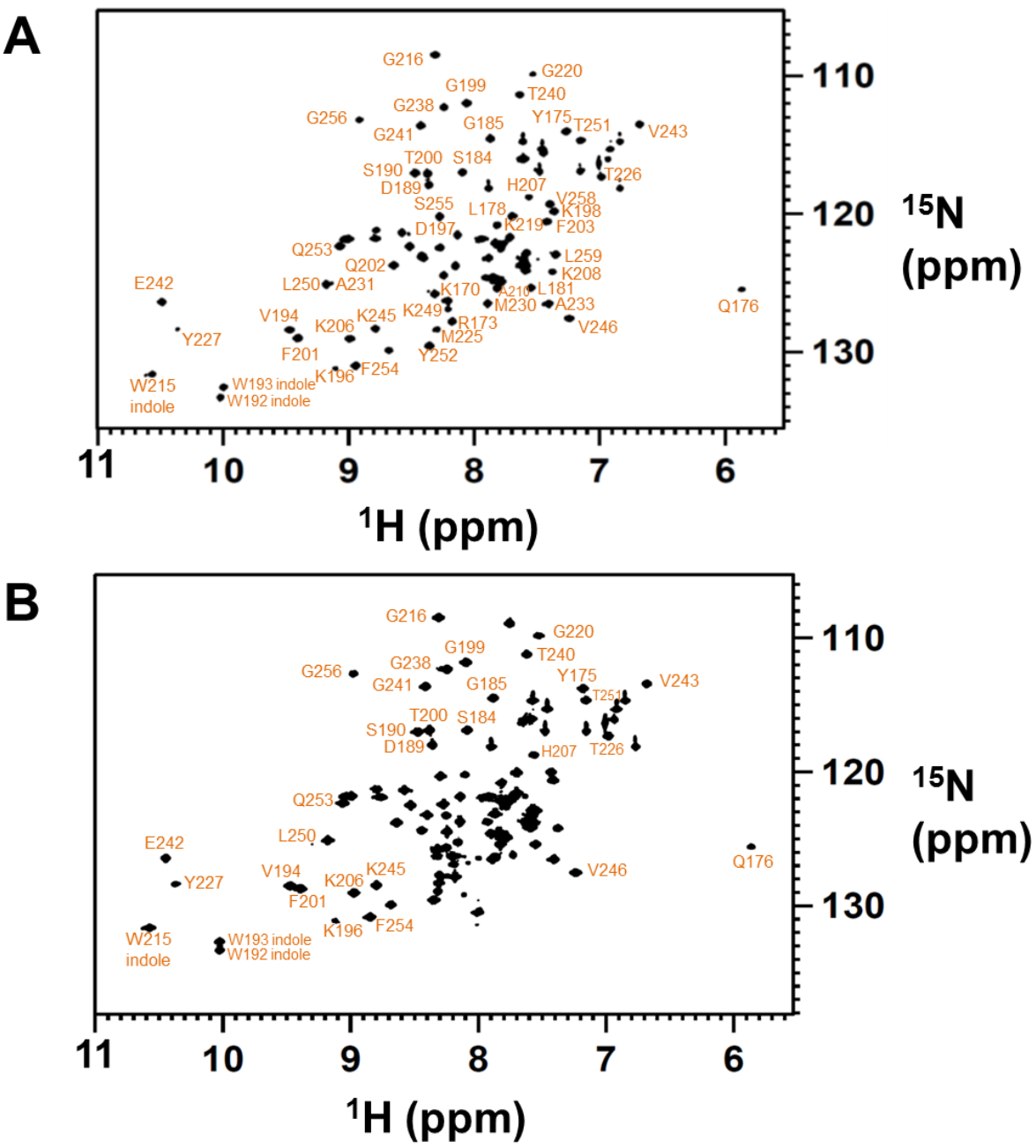


Figure S 2.5 ^{1}H-^{15}N HSQC spectra of unbound murine PU.1(167-260) and PU.1ΔN167.

A truncated PU.1 ETS construct without the final 12 residues in the PU.1ΔN167 used in the experiments described in the main text was cloned and over-expressed as a uniformly ^{15}N-labeled protein in E. coli similarly as PU.1ΔN167, and purified on Sepharose SP. The ^{1}H-^{15}N HSQC of PU.1(167-260) (*A*) closely matched a previously reported spectrum of Jia *et al*. (118) which in turn allowed assignment of many resonances in PU.1ΔN167 (*B*).

3 CHARACTERIZATION OF INTRINSICALLY DISORDERED REGIONS ON INTERNAL DYNAMICS OF THE ETS DOMAIN OF PU.1

3.1 Preface

The objective of the study in this chapter is to characterize the effects of intrinsically disordered regions (IDRs; N-terminal PEST domain and C-terminal 12 residues) flanking the ETS domain on backbone dynamics of PU.1 using NMR. I prepared the protein and DNA samples in this study. I conducted all the NMR experiments with Dr. Markus Germann, except for the hydrogen-deuterium exchange (HDX) experiment. The HDX experiment was in collaboration with Dr. Marina Evich. Our collaborator, Dr. James Aramini at the City University of New York Advanced Science Research Center (CUNY ASRC), assigned the residues of hPU.1 ETS domain (residues 165-258) in the 1:1 complex with 16-bp site-specific DNA. This work was supported by NSF grant MCB 15451600 and NIH grant R21 HL129063.

3.2 Abstract

The presence of the IDRs flanking the ETS domain does not change the DNA binding modes of the PU.1 ETS domain (*cf.* Chapter 2), yet the N-terminal IDR (PEST domain) modifies DNA recognition by the ETS domain through changing DNA binding affinities. We used 3D NMR (HNCO, HN(CA)CO, HNCA, HNCACB, and HN(CO)CACB) to analyze the ^{1}H-^{15}N HSQC spectra of hPU.1 constructs with and without IDRs, and also in the absence and presence (1:1 complex) of DNA. Thus, we successfully assigned ~90% or more backbone amide resonances of hPU.1. Using the fully assigned HSQC spectra, we studied fast (ps to ns) time scale internal dynamics of PU.1 protein. Whole sets of ^{15}N R_1 and R_2 relaxation rates and heteronuclear $^{1}H\{^{15}N\}$-NOE were acquired for all the hPU.1 constructs with and without DNA. We found the PEST domain upon specific DNA binding becomes more dynamic in a disordered

structure. In terms of DNA recognition, the presence of the PEST domain increases the affinity of 1:1 complex of the ETS domain with cognate DNA, without perturbing the structure or changing the fast time scale backbone motions of the ETS domain.

3.3 Introduction

In Chapter 2, we found the ETS domain of PU.1 dimerizes at a single cognate site of DNA in a negatively cooperative manner via the ETS domain, while some ETS family members, including Ets-1, are unable to form a 2:1 complex in physiological conditions (99). This is the first direct demonstration of a 2:1 complex formation with site-specific DNA by an ETS family member to sequester excess protein of its own. Since PU.1 is one of a few members that lack autoinhibition, we proposed that this is a self-regulating and negative feedback mechanism of PU.1 protein instead of autoinhibition. We also detected the dimerization interface of PU.1 in the 2:1 PU.1/DNA complex: the site distal to the DNA-binding interface, including the loop between S1 and S2 where four consecutive charged residues ^{193}DKDK^{196} reside (99).

Following what we discovered in Chapter 2, we planned to investigate the conformations and interactions that the PU.1 dimer displays upon binding with site-specific DNA. However, as we observed in Chapter 2, ~80% of the ^{1}H-^{15}N HSQC resonances disappear in 2:1 PU.1/DNA complex (Fig. 2.3, *B*), which unfortunately prevented us from direct observation of the DNA-bound PU.1 dimer. On the other hand, both unbound PU.1 and 1:1 PU.1/DNA complex (Fig. 2.3, *A* and *C*) show well-dispersed HSQC peaks. Thus, we have decided to study internal dynamics of unbound and 1:1 DNA-bound PU.1 at atomic resolution by NMR, instead of directly examining the 2:1 PU.1/DNA complex.

Moreover, the study in this chapter follows our recent studies of the roles of IDRs on DNA-free PU.1 homodimerization, which was previously observed *in vivo* by Evans et al. (68)

(138). The PU.1 dimers in the absence and presence of DNA antagonize to each other, and IDRs flanking the ETS domain play key roles for this phenomenon (138). We also found that DNA recognition by the ETS domain of PU.1 is modified by the presence of IDR. The N-terminal PEST domain is intrinsically disordered, and this nature is important because it facilitates the 1:1 DNA complex formation. However, the affinity of the 2:1 PU.1/DNA complex is reduced by the disordered PEST (138). The DNA recognition of the ETS domain is not affected by the absence of C-terminal IDR. By contrast, PU.1 is unable to form a DNA-free dimer without C-terminal IDR (138).

Thus, to study how the PEST domain modifies DNA recognition of PU.1, we tested whether or not the PEST domain changes the internal dynamics of the ETS domain using NMR, in this chapter. We, therefore, prepared three constructs of hPU.1 protein that consist of only the ETS domain (sΔN165), ETS domain and C-terminal IDR (ΔN165), and ETS domain and both N- and C-terminal IDRs (ΔN117), respectively (Fig. 3.2 *A*). As a first step, we worked on backbone assignment of PU.1 proteins consisting of the ETS domain and/or flanking IDRs (N-terminal PEST domain and C-terminal 12 residues) in the absence and presence of cognate DNA because these backbone assignments give us opportunities to further investigate the hydrodynamic properties of hPU.1 protein at the residue-by-residue level.

3.4 Materials and methods

3.4.1 Protein and DNA sample preparation

Molecular cloning. DNA fragments of hPU.1ΔN165 (human residues 165–270) and ΔN117 (human residues 117–270) were purchased from Integrated DNA Technologies (Midland, IA), and subcloned directly into the NcoI/HindIII sites of pET28b vector. The hPU.1sΔN165 (human residues 165–258) construct that lacks both N- and C-terminal IDRs

flanking the ETS domain was subcloned directly into the NcoI/NdeI sites of pET15b. All constructs were verified by Macrogen (Rockville, MD) using Sanger sequencing.

Protein expression and purification. Three human PU.1 constructs were used in this study: sΔN165 (ETS domain only), ΔN165 (ETS domain and C-terminal 12 residues), and ΔN117 (N-terminal PEST domain, ETS domain, and C-terminal 12 residues). Uniformly ^{15}N-labeled or $^{15}N/^{13}C$-labeled hPU.1 proteins were overexpressed in BL21(DE3) (for sΔN165 and ΔN165) or BL21(DE3) pLysS (for ΔN117) *E. coli* as previously described (99). In brief, the cell pellet of starter culture in 50 mL LB was harvested, washed, and resuspended in 1 L M9 minimal medium containing $^{15}NH_4Cl$, $MgSO_4$, $CaCl_2$, trace metals, MEM vitamins, and unlabeled or ^{13}C-labeled glucose as required for ^{15}N (or $^{15}N/^{13}C$) labeling. Protein expression was induced with 0.5 mM IPTG for 4 h at 30 °C (for sΔN165 and ΔN165) or 25 °C (for ΔN117). Bacterial pellets were resuspended in 10 mM sodium phosphate (pH 7.4), 500 mM NaCl and 0.1 mM phenylmethanesulfonyl fluoride, and shear-homogenized (Sonic Dismembrator FB-505, Fisher Scientific). The lysate was cleared by centrifugation and directly loaded onto a cation exchange column (HiTrap^{TM} SP HP, GE Healthcare) under the control of a Bio-Rad NGS Quest 10 instrument. After washing out residual impurities, the protein was eluted by a NaCl gradient 0.5-2 M. Purified proteins were extensively dialyzed against 10 mM sodium phosphate buffer (pH 7.0) containing 50 mM (for sΔN165) or 150 mM (for ΔN165 and ΔN117) total [Na^+], 0.5 mM EDTA, and 0.01% NaN_3. Each hPU.1 protein concentration was determined by UV absorption at 280 nm using the extinction coefficients ϵ_{280} = 22,460 M^{-1} cm^{-1} (for sΔN165 and ΔN165) and 23,593 M^{-1} cm^{-1} (for ΔN117).

Nucleic acids. Synthetic DNA oligos were purchased from Integrated DNA Technologies and annealed as described (99). For NMR experiments using PU.1/DNA complex, 16-bp high

affinity DNA (5'-GCAAGC**GGAA**GTGAGC- 3') was co-dialyzed with hPU.1sΔN165 or ΔN165 against 10 mM sodium phosphate buffer (pH 7.0) containing 50 mM (for sΔN165) or 150 mM (for ΔN165 and ΔN117) total [Na^+], 0.5 mM EDTA, and 0.01% NaN_3. The 23-bp high affinity DNA (5'-GCGAATAAGC**GGAA**GTGAAACCG- 3') was co-dialyzed with hPU.1ΔN117 against 10 mM sodium phosphate buffer (pH 7.0) containing 150 mM total [Na^+], 0.5 mM EDTA, and 0.01% NaN_3.

3.4.2 NMR Spectroscopy

Sequential backbone assignment of hPU.1sΔN165, ΔN165, and ΔN117 constructs in the absence of DNA. ^1H-^{15}N correlated 2D HSQC measurements were made as previously described (99). NMR experiments for PU.1 backbone NH groups assignment were done using 976 μM (hPU.1 sΔN165), 682 μM (ΔN165), and 563 μM (ΔN117) protein samples in the absence of DNA, and 600 μM (ΔN165 and ΔN117) in 1:1 complex with cognate DNA. All NMR experiments were performed with a Bruker Avance 600 MHz spectrometer, equipped with a 5 mm QXI probe, at 25 °C (298 K). Signals from backbone ^1H, ^{13}C, and ^{15}N nuclei were assigned from five 3D heteronuclear experiments HNCO, HN(CA)CO, HNCA, HNCACB, and HN(CO)CACB (139). Obtained NMR spectra were processed using a Bruker TopSpin 3.2 or 3.5 pl7, and the data analysis was achieved with NMRFAM-Sparky software (140).

Nuclear spin relaxation measurements. Whole sets of ^{15}N T_1, T_2, and heteronuclear $^1H\{^{15}N\}$-NOE data were acquired using Bruker Avance 600 MHz spectrometer, equipped with a 5 mm HCN triple resonance probe, at 25 °C (298 K) (139). The hPU.1 protein concentrations were 870 μM (unbound sΔN165), 690 μM (sΔN165 in 1:1 complex with DNA), 600 μM (both unbound and 1:1 DNA-bound ΔN165, and 1:1 DNA-bound ΔN117), and 563 μM (unbound ΔN117). A total of 8 data sets were collected to measure T_1 with delay values of: 0.005, 0.05,

0.125, 0.225, 0.350 0.500, 0.750, and 1.000 seconds for all constructs with and without DNA. A total of 7 data sets were collected to measure T_2 with delay values of: 0.01, 0.02, 0.04, 0.07, 0.09, 0.13, and 0.18 seconds (for sΔN165 with and without DNA), 0.01, 0.025, 0.05, 0.08, 0.1, 0.15, and 0.2 seconds (for unbound ΔN165 and ΔN117), 0.005, 0.01, 0.0175, 0.025, 0.03, 0.04, 0.05, 0.06, 0.08, 0.1, 0.15, and 0.2 seconds (for 1:1 DNA-bound ΔN165), or 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.04, 0.06, 0.08, 0.14, and 0.2 seconds (for 1:1 DNA-bound ΔN117). Data were processed with TopSpin 3.2, 3.5 pl7, or 3.6.1 (Bruker) to extract peak intensities and then fit as single exponential decay with Origin 9.1 (OriginLab). Steady-state heteronuclear $^{1}H\{^{15}N\}$-NOE was acquired from the difference between spectra acquired with and without ^{1}H saturation and a total recycle delay of 3s. Heteronuclear $^{1}H\{^{15}N\}$-NOE error was derived by $|NOE|\sqrt{\{(\frac{noise\ Sat}{intensity\ Sat})^2 + (\frac{noise\ Unsat}{intensity\ Unsat})^2\}}$ using backgoround noise level of the spectra.

3.5 Results

3.5.1 The presence of the PEST domain does not change DNA binding modes of PU.1

Our group recently found that the presence of the disordered PEST domain increases the affinity of 1:1 complex of the structured ETS domain with cognate DNA, while it reduces that of 2:1 complex in physiological conditions (138). Based on this finding, we further studied the roles of the N-terminal PEST domain on DNA recognition of the ETS domain using NMR. Translational diffusion constants obtained by DOSY NMR revealed that PU.1 retains its ability to have multiple DNA binding modes with 23-bp site-specific DNA in the presence of the N-terminal intrinsically disordered PEST domain flanking the ETS domain (hPU.1ΔN117 protein) (Fig. 3.1 *A(a)*). The finding is similar to the PU.1 behavior in the absence of the PEST domain with 16-bp specific DNA (i.e., forming both 2:1 and 1:1 PU.1/DNA complex in direct response

to the PU.1/DNA molar ratios) (*cf.* Fig. S 2.3). In contrast, hPU.1ΔN117 titration with 23-bp specific DNA showed a clear difference from that with 16-bp DNA in binding manners (Fig. 3.1 *A(b)*). ΔN117 forms a 2:1 PU.1/DNA complex with 16-bp specific DNA in response to the PU.1/DNA ratio, but 1:1 PU.1/DNA complex formation does not reflect the molar ratio. Interestingly, the binding behavior of hPU.1ΔN117 to 16-bp DNA is also different from what we observed for PU.1 ETS domain (mPU.1ΔN167) to 10-bp specific DNA, as described in Chapter 2 (*cf.* Fig. S 2.3). ΔN117 forms a 2:1 PU.1/DNA complex with 16-bp DNA, but ΔN167 does not with 10-bp DNA. In Chapter 2, we demonstrated that the DNA contact interface in the 2:1 PU.1/DNA complex is longer than that in the 1:1 complex (*cf.* Fig. 2.6). The 16-bp DNA is probably long enough for the PU.1 ETS domain to function fully. Based on the DNA binding manners of ΔN117 to 23- and 16-bp specific DNA, we can suggest that short (namely, 16 bp or less) DNA has a negative impact on DNA binding of PU.1 ETS domain in the presence of the flanking PEST domain.

To monitor the protein "fingerprint" of both the PEST and ETS regions, we tested HSQC titrations in the same manner as described in Chapter 2. We acquired HSQC spectra of hPU.1ΔN117 in a titration with 23-bp specific DNA, at PU.1/DNA molar ratios of 0, 0.5, 1, and 2 (Fig. 3.1 *B*). The ETS crosspeaks that are well-overlapped with our previous HSQC spectrum of mPU.1ΔN167 (*cf.* Fig. 2.3) exhibited almost the same trend at each molar ratio. The HSQC crosspeaks that are not overlapped with the crosspeaks of the ETS residues are most probably the PEST peaks. Such crosspeaks were clustered at ~8.2 ppm on the proton dimension, and these HSQC resonances showed little shifts upon specific DNA binding.

Based on the results of the DOSY NMR described above, we further studied the impact of the DNA length on specific DNA binding of PU.1 ETS domain in the presence of flanking

PEST domain. We overlaid the HSQC spectrum of hPU.1ΔN117 in 1:1 complex with 16-bp specific DNA onto that with 23-bp specific DNA. Most of the crosspeaks overlapped well, but several peaks were found to be shifted. Notably, the peaks of K245-247 showed no overlap between these spectra (Fig. 3.1 *C (a)*). These three consecutive Lys residues are located at the center of the "wing" of PU.1 (Fig. 3.1 *C (b)*). The "wing" binds upstream of the core GGAA sequence of DNA. Thus, if the DNA is so short that the terminal base pairs fray in 1:1 PU.1/DNA complex, these positively charged residues would structurally be perturbed due to charge-charge repulsion with phosphate groups of DNA.

3.5.2 Sequential backbone assignment of three hPU.1 constructs sΔN165, ΔN165, and ΔN117 in the absence and presence of DNA

To assign ^{1}H-^{15}N HSQC resonances of PU.1 backbone amides, we first focused on five specific amino acids Gly, Ala, Ser, Thr, and Pro in PU.1 as these amino acids have characteristic chemical shifts in clear backbone assignment (Fig. 3.2 *A*). Gly residues do not have beta carbons, and Ala residues have exceptionally small beta carbon chemical shifts. Ser/Thr residues have exceptionally large chemical shifts for beta carbons, and Pro residues do not have HSQC peaks of their own backbone amides. We used these residues for the starting points of the assignment. For example, the chemical shift of Ser^{253} beta carbon is much larger than most other beta carbons, and Gly^{254} has no beta carbon. Thus, we were able to find these consecutive residues at the initial stages of the DNA-free hPU.1ΔN165 assignment (Fig. 3.2 *B*). The N-terminal first residue is generally not observed in ^{1}H-^{15}N HSQC spectra. This was the case in our studies.

Then we worked on the backbone assignment of the PU.1 protein containing only the ETS domain (hPU.1sΔN165; residues 165-258) in the absence of DNA. The first three residues at the N-terminus and residues L172, L180, S203, K204, K221, and K222 were unable to be

assigned because the corresponding resonances in the ^{1}H-^{15}N correlation spectra were not observed. The next assignment was done for DNA-free PU.1 protein containing the ETS domain and C-terminal IDR (hPU.1ΔN165; residues 165-270). L172, L180, S203, K204, and K222 were unable to be assigned, in the same manner as the hPU.1sΔN165, as described above. Besides, the assignment of R220 was not achieved. The C-terminal IDR contains two consecutive Pro residues (P268 and P269) that never appear on ^{1}H-^{15}N HSQC spectra due to the absence of ^{1}H attached to its ^{15}N in Pro. Moreover, we were unable to unambiguously assign R265.

Likewise, the assignment was done with unbound hPU.1ΔN117 (residues 117-270) containing N-terminal IDR (PEST domain), the ETS domain, and C-terminal IDR. The same residues as ΔN165 were assigned in the ETS domain and C-terminal IDR. The PEST domain (residues 117-164) is disordered and therefore the ^{1}H-^{15}N HSQC resonances are clustered in the typical amide chemical shift region (~8.2 ppm of ^{1}H). Moreover, this domain contains a lot of Pro residues (P122, 126, 129, 141, 142, 155, 157, and 161) that cannot be assigned. Thus, the sequential assignment of the PEST domain was much more difficult than the other part of the protein. Because N-terminal residues typically exhibit most negative $\{^{1}H\}$-^{15}N NOE (described in detail in 3.5.3 of this chapter), we used the heteronuclear NOE data to distinguish the N-terminal Gly and Ser/Thr residues from the same amino acid residues at other sites. We were eventually able to assign the PEST residues except for L119, Q120, Y121, L124, Q139, D147, E149, in addition to all the ETS residues assigned for ΔN165.

Thus, we successfully assigned 88 of 94 backbone amides of hPU.1sΔN165, 94 of 104 resonances of ΔN165, and 126 of 144 residues of ΔN117 (Fig. 3.2 *A*). Likewise, at least ~90% of HSQC resonances of the same PU.1 proteins were assigned in 1:1 complex with cognate DNA (Fig. 3.2 *B*).

3.5.3 The PEST domain stays disordered but becomes more dynamic upon specific DNA binding

Using the ^{1}H-^{15}N HSQC spectra of PU.1 proteins whose resonances were fully resolved, the well-dispersed crosspeaks of ΔN165 residues were found to be mostly overlapped with ΔN117 residues. Furthermore, all the residues in the PEST region were clustered at ~8.2 ppm on the ^{1}H dimension for both unbound and 1:1-DNA bound ΔN117 (Fig. 3.4 *A* and *B*). This is characteristic of structural disorder of proteins in ^{1}H-^{15}N HSQC spectra. The chemical shifts of these PEST resonances were also similar for both unbound protein and 1:1 complex with specific DNA binding. Thus, ^{1}H-^{15}N HSQC chemical shifts strongly suggest that the PEST domain stays similarly disordered upon specific DNA binding in terms of structural perturbation.

We further performed measurements of ^{15}N relaxation parameters (namely, spin-lattice (R_1) and spin-spin (R_2) relaxation rate and the steady-state heteronuclear $^{1}H\{^{15}N\}$-NOE) for hPU.1ΔN165 and ΔN117 residues both in the absence and presence (at 1:1 molar ratio) of 23-bp specific DNA (Fig. 3.5 *A-L* and Supplemental Table 3.1-3.4). We excluded those which were either overlapping or hardly visible. The NOE values were calculated from the intensity ratios of individual crosspeaks with and without ^{1}H saturation (Fig. 3.5 *F* and *L*). The R_1 and R_2 values of hPU.1 proteins for both unbound and 1:1 DNA-bound forms were obtained by fitting the intensity of each crosspeak with a set of relaxation times (Fig. 3.5 *A,B,D,E,G,H,J,K*).

To evaluate the quality of the spin relaxation data, we obtained the ranges of relative errors. The relative error ranges of R_1, R_2 and NOE are as follows. DNA-free ΔN165: 0.5-25.5%, 1.2-7.9%, and 1.6-49.9%, respectively. DNA-free ΔN117: 1.6-62.2%, 1.0-22.6%, and 2.7-343.0%, respectively. 1:1 DNA-bound ΔN165: 4.3-84.6%, 3.5-24.9%, and 1.3-25.9%, respectively. 1:1 DNA-bound ΔN117: 2.6-101.1%, 2.0-31.2%, and 0.8-597.3%, respectively.

Typical erros in replicated NMR relaxation experiments are 5-10% in the case of Ets-1 from literature (33), where their group conducted multiple NMR relaxation experiments and obtained standard errors. In contrast, our errors (standard deviation) were derived from experiments without replication. Namely, our R_1 and R_2 errors were derived from fitting and our NOE error was from signal intensities relative to background noise level. Thus, the high R_1 or R_2 errors of some PU.1 residues are attributed to both data fitting and significantly weak HSQC signals, and the high NOE errors of some PU.1 residues are due to significantly weak signals in the HSQC specta. We have high errors for some PU.1 residues. Notably, M185 and K245 have high errors because their HSQC signals are weak (Fig. 3.5 M), which resulted in high errors in the spin relaxation measurements.

Large NOE values of the ETS residues and much smaller values of the PEST residues indicate that the ETS domain stays well-ordered, in contrast to the PEST domain which stays disordered upon specific DNA binding (Fig. 3.5 *F* and *L*). Much larger R_1 (Fig. 3.5 *D* and *J*) and much smaller R_2 (Fig. 3.5 *E* and *K*) values of the PEST domain than the corresponding values of the ETS domain also support this observation. Interestingly, $^{1}H\{^{15}N\}$-NOE of unbound PU.1 PEST domain contains both negative and positive values, and transition from highly negative to slightly positive is observed as the residues are closer to the well-ordered ETS domain. On the other hand, $^{1}H\{^{15}N\}$-NOE of the PEST domain of PU.1 in the 1:1 complex is mostly negative, and no such transition of the degree of disorder is observable. Average NOE values of the PEST domain in unbound and 1:1 DNA-bound ΔN117 are -0.09 ± 0.04 and -0.30 ± 0.03, respectively. This suggests that the PEST domain becomes more dynamic upon specific DNA binding. Collectively, the N-terminal disordered region (PEST domain) flanking the DNA-binding domain stays disordered upon 1:1 specific DNA binding but becomes more dynamic.

3.5.4 Effects of the PEST domain on backbone motions of the ETS domain of PU.1 in the absence of DNA and the 1:1 complex with cognate DNA

To examine the effects of the PEST domain on backbone motions of the ETS domain of DNA-free PU.1, we compared heteronuclear $^{1}H\{^{15}N\}$-NOE of ΔN165 and ΔN117. As our group reported recently, ΔN165 is monomeric, and ΔN117 is dimeric at the protein concentrations (~600 μM) used, in the absence of DNA, in physiological conditions (138). We compared the average NOE values of the ETS domain of DNA-free ΔN165 and ΔN117: 0.80 ± 0.02 and 0.74 ± 0.01, respectively. Smaller average NOE of ΔN117 (than that of ΔN165) suggests that the ETS domain of PU.1 is overall more dynamic in the presence of the PEST domain, than in the absence. Furthermore, smaller average R_1 (ΔN117: 1.79 ± 0.07 as compared to ΔN165: 2.10 ± 0.05) and larger average R_2 (ΔN117: 14.59 ± 0.29 as compared to ΔN165: 10.25 ± 0.15) relaxation rates of the ETS domain reflect the molecular mass difference (i.e., between ΔN165 and ΔN117) in addition to the difference in dynamics. We subsequently subtracted heteronuclear NOE values of ΔN117 from ΔN165 in the absence of DNA (Fig. 3.5 *N*). The ETS domain of DNA-free PU.1 dimer (ΔN117) is more dynamic than the monomer (ΔN165) because most residues showed positive ΔNOE values, consistent with average NOE comparison above. We detected the residues with relatively large |ΔNOE| values (> 0.175). The results clearly show that three consecutive residues Asp^{184}, Met^{185}, and Lys^{186}, located on the loop between H1 and S1, show large ΔNOE values, along with some other residues such as Ser^{202} (highlighted in the graph). This is consistent with the exchange broadening observed for the amide of Met^{185} (very weak correlation peak of Met^{185}) in all the ^{1}H-^{15}N HSQC spectra of DNA-free PU.1. These characteristics of Met^{185} residue are consistent with the trends seen in the spin relaxation rates (anomalously large R_2 rate constant) on the picosecond to nanosecond time scale and

conformational exchange dynamics on the millisecond to microsecond time scale (exceptionally large R_{ex} value) for the PU.1 ETS domain from literature (118). Thus, Met^{185} alone or the region around the residue (the loop between H1 and S1) is in conformational equilibria and presumably responsible for PU.1 homodimer formation in the absence of DNA, by contributing to conformational exchange processes between monomer and dimer. On the other hand, the highly negative ΔNOE value of the Lys^{245} residue suggests that the residue is less dynamic in the DNA-free dimer than in the DNA-free monomer. Therefore, the residue is probably important for maintaining the monomeric form of PU.1, which is consistent with our observation that the three consecutive Lys residues (i.e., $Lys^{245-247}$) in the "wing" play an important role in DNA recognition of PU.1, as discussed above (*cf.* in Chapter 3.5.1).

In order to test whether the PEST domain alters internal dynamics of the ETS domain of PU.1 upon specific DNA binding, we compared the average NOE values of the ETS domain of 1:1 DNA-bound ΔN165 and ΔN117: 0.79 ± 0.02 and 0.81 ± 0.03, respectively. Thus, no difference was observed in the average NOE values for ΔN165 and ΔN117, suggesting that the presence of the PEST domain does not change the net dynamics of the ETS domain. Subsequently, we subtracted the heteronuclear NOE values of 1:1 DNA-bound ΔN117 from those of ΔN165 (Fig. 3.5 *O*), in the same manner as the unbound protein, as described above. The ΔNOE values were randomly dispersed in both positive and negative directions. This further suggests that the presence of the PEST domain causes no net change in the internal dynamics of the ETS domain in the 1:1 complex with cognate DNA. Namely, the ETS domain maintains similar levels of fast-time scale internal dynamics in total. This result presents a great contrast to our observation in dimeric (ΔN117) and monomeric (ΔN165) DNA-free PU.1 proteins, as described above. In the case of DNA-free PU.1, the ETS domain of PU.1, as a whole, is more

dynamic in ΔN117 than in ΔN165, primarily because the PEST domain facilitates dynamic homodimer formation of the ETS domain in the absence of DNA.

In summary, the presence of the PEST domain alters the backbone dynamics of the PU.1 ETS domain by facilitating the homodimerization of the ETS domain in the absence of DNA. On the other hand, in the event of site-specific 1:1 DNA-binding, the PEST domain becomes more dynamic but does not change backbone motions of the ETS domain. The well-overlapped ETS domain residues in the ^{15}N-HSQC spectra of ΔN117 and ΔN165 indicate almost no structural perturbations resulting from the presence of the PEST domain. Thus, the presence of the PEST domain increases the affinity of 1:1 complex of the ETS domain with cognate DNA, without perturbing the structure or changing the fast time scale motions of the ETS domain.

3.6 Discussion

3.6.1 Optimal ionic strength and pH for PU.1 backbone assignment

For the backbone assignment of unbound PU.1 ETS domain (hPU.1sΔN165; residues 165-258), we conducted the NMR experiments using a ~0.8 mM protein sample, at 50 mM salt, at pH 7.0, at 25 °C, using a 600 MHz NMR spectrometer. The PU.1 construct was assigned with 50 mM salt because low salt is known to avoid effectively broadening of signals and to make shimming of NMR magnet better. The other assignments (ΔN165 and ΔN117) were done in physiological conditions since no mal-effect by 150 mM salt was found. On the other hand, Jia *et al*. previously assigned backbone amides of PU.1 containing almost the same residues (residues 166-258) as our sΔN165, at 2.5 mM protein sample at 400 mM salt, pH 5.5 and 30 °C, using a 500-MHz NMR spectrometer (139). Except for the N-terminal residue that never appears in ^{1}H-^{15}N HSQC spectra and L174, they successfully assigned 91 of 93 residues. L174 was not detectable in our experiments as well. Furthermore, McIntosh et al. reported an NMR structure

of unbound murine PU.1 ETS domain (PDB: 5W3G), at 150 mM salt, pH 5.5, and 25 °C, using an 850-MHz NMR spectrometer. They successfully assigned all the backbone amides with a 0.3 mM protein sample to solve the structure. Their experimental conditions (salt concentration and temperature) except for pH (5.5) were the same as ours.

We fully assigned the backbone structure of the PU.1 ETS domain with and without flanking IDRs using protein samples at relatively low concentrations (~0.8 mM or lower) in physiological conditions (pH 7.0) and with a 600-MHz NMR spectrometer. However, the experimental conditions of other groups which achieved almost 100% backbone assignments for PU.1 suggest that lowered pH (5.5) effectively reduces exchange-broadening. The charge of the hPU.1 ETS domain (residues 165-270) at pH 7.0 and 5.5 are estimated to be 18.9 and 22.3, respectively (http://protcalc.sourceforge.net/cgi-bin/protcalc). Thus, such a slight increase in the cationic charge density of the protein probably contributes to the local stability of PU.1, and the chemical exchange of PU.1 residues is minimized as a result. High salt generally causes line broadening in NMR, while low salt drives DNA-free homodimerization of the PU.1 ETS domain. Therefore, 150 mM salt probably provides optimal ionic strength for PU.1. Thus, the optimal conditions for NMR experiments of PU.1 backbone assignment we selected are 150 mM salt and pH 5.5.

3.6.2 The presence of the IDR flanking the N-terminus of the ETS domain does not change the DNA-binding interface or dynamics of the ETS domain upon binding specific DNA, for both PU.1 and Ets-1

The ^{1}H-^{15}N HSQC spectra of 1:1 specific DNA-bound hPU.1ΔN117 and ΔN165 superimposed very closely (Fig. 3.4 *B*). This suggests that the PEST domain does not change the structural interactions of PU.1 with site-specific DNA. Our NMR relaxation experiments also

revealed that the presence of the PEST domain does not alter the backbone dynamics of the PU.1 ETS domain on the sub-nanosecond time scale upon 1:1 specific DNA binding. The same trend was reported for the 1:1 complex of Ets-1 and specific DNA by Desjardins et al. (33). The Ets-1ΔN279 protein contains the N-terminal intrinsically disordered region (partially truncated "serine-rich region (SRR)") as well as the ETS domain with the inhibitory module (IM). The ΔN301 protein contains only the ETS domain with the IM. Desjardins et al. demonstrated that the presence of the N-terminal IDR neither perturbs the structure of the ETS domain nor changes dynamics of the Ets-1 upon binding to specific DNA, using ^{1}H-^{15}N HSQC spectra and heteronuclear NOE measurements.

The SRR (residues 244 to 300) of Ets-1 inhibits DNA binding in a phosphorylation-dependent manner, by stabilizing the IM and transiently associating with DNA recognition interface of the ETS domain (74,141-143). The SRR of Ets-1 contains five specific Ser residues (251, 270, 273, 282, and 285) that are the targets of phosphorylation (36). Likewise, the PEST domain of PU.1 contains Ser residues (130, 131, 140, and 146) as phosphorylation targets, whose phosphorylation increases the anionic charge density of the protein and enhances PU.1 binding activity (62,138,144-149). Transition in the degree of disorder, which is the same trend seen in the unbound PU.1 PEST domain as shown above, was reported previously for DNA-free Ets-1 SRR using $^{1}H\{^{15}N\}$-NOE measurements (34,35). Thus, neither of these IDRs (the PEST domain of PU.1 and the SRR of Ets-1) change the structure and dynamics of the ETS domain. Instead, the PEST domain and the SRR build up anionic charges. Consequently, the PEST domain increases transcriptionally active PU.1, and the SRR stabilizes autoinhibition.

The SRR of Ets-1 interacts with the ETS domain via the DNA recognition interface in the absence of DNA, but DNA binding by the ETS domain is favored so that the association of DNA

and SRR is mutually exclusive (150). The chemical shift perturbations observed in the same study revealed that the SRR of Ets-1 also interacts with PU.1 via the same sites (H1 and the wing) that are perturbed by DNA-free dimer formation of PU.1 containing the PEST domain (138,150). Our heteronuclear NOE measurements revealed that the PEST domain is more dynamic in the 1:1 complex with site-specific DNA than in the unbound dimer (*cf.* Fig. 3.5 *F* and *L*). It is presumably because specific DNA binding is favored by the PU.1 ETS domain, and therefore the association of DNA and the PEST domain by the PU.1 ETS domain is mutually exclusive, in the same manner as Ets-1 (the ETS domain and the SRR). Consequently, the "free" PEST domain, which is acidic (pI: 3.5) in contrast to basic ETS domain (pI: 10.5), has charge-charge repulsion with DNA in the 1:1 complex. As a result, the PEST domain becomes more dynamic than that in the DNA-free dimeric form. By contrast, the PEST domain in the DNA-free PU.1 dimer interacts with the ETS domain electrostatically from a distance and therefore stabilizes the homo-dimerization of the ETS domain. Furthermore, such an increase in the affinity of the 1:1 PU.1/DNA complex probably makes the complex more compact, and therefore the rotational correlation time of the complex becomes faster. Thus, faster average R_1 (ΔN117: 1.97 ± 0.10 as compared to ΔN165: 1.53 ± 0.07) and very similar average R_2 (ΔN117: 30.57 ± 1.06 as compared to ΔN165: 29.24 ± 0.65) relaxation rates of the ETS domain of ΔN117 in the 1:1 complex do not directly reflect the molecular weight difference (i.e., between ΔN165 and ΔN117).

3.6.3 PEST domain facilitates 1:1 binding of the ETS domain with specific DNA

Our recent study revealed that the PEST domain drives DNA-free PU.1 dimer formation via the ETS domain (138). Considering the PEST domain is acidic (pI: 3.5) and the ETS domain is highly basic (pI: 10.5), electrostatic interactions of the PEST and ETS domains would greatly

reduce the charge-charge repulsion arising from the homodimer of the ETS domain. Reinforcing negative charges in the PEST domain facilitates the DNA-free dimerization of PU.1 ETS domain, and acidic crowders have the same effect on PU.1 ETS domain in the absence of the PEST domain. Therefore, the anionic charge of the PEST domain probably gives the driving force for DNA-free dimerization of the ETS domain.

In the presence of DNA, the highly cationic ETS domain bound with the PEST domain probably releases it and bind with DNA because phosphate groups of DNA are much more acidic than the PEST domain. Consequently, upon DNA binding of the ETS domain, the released PEST domain becomes more dynamic than in the DNA-free homodimer. This is consistent with our observation by heteronuclear NOE (*cf.* Fig. 3.5 *F* and *L*). Our DNA-binding assays also revealed that the presence of the PEST domain increases affinity in the 1:1 PU.1/DNA complex. This is probably because the "free" PEST domain enhances the ETS:DNA interaction.

3.6.4 Dynamic properties of the ETS domains – PU.1 vs. Ets-1

It is important to understand whether or not autoinhibition changes backbone mobility of the ETS domain in the ETS family. In general, DNA-binding domains of transcription factors are dynamic (47,48). This feature enables DNA-binding domains to search for specific binding sites in a majority of nonspecific DNA until it is quenched upon specific DNA binding (47,48). This feature is probably common for all the ETS domains because DNA-binding interface in the ETS domains of Ets-1 and ETV6 are conformationally dynamic (33,37). Ets-1 shares the same binding interface for specific and nonspecific DNA binding (33), probably due to the flexibility in the DNA-binding interface. However, the flexibility in the ETS domain is likely to compromise due to the autoinhibitory module (IM) adjacent to the ETS domain. Thus, the ETS

domain of PU.1 (non-autoinhibited) is probably more flexible than that of Ets-1 (autoinhibited) in the absence of DNA.

We studied the dynamic properties of PU.1 by amide hydrogen-deuterium exchange (HDX) experiments. An HDX experiment using the PU.1 ETS domain was done in the absence of DNA in physiological conditions. However, almost all the crosspeaks in HSQC disappeared immediately when we prepared the sample in 100% D_2O (only in 15 minutes) (Fig. 3.6). In contrast, successful HDX experiments using unbound Ets-1 have been reported (35,36). Therefore, the unbound PU.1 protein is probably very flexible and not fixed unless it binds to DNA specifically. Thus, the backbone dynamics of PU.1 is quite different from that of Ets-1 in this respect, probably due to the absence of autoinhibition.

3.6.5 Responsible sites of PU.1 ETS domain for PPIs

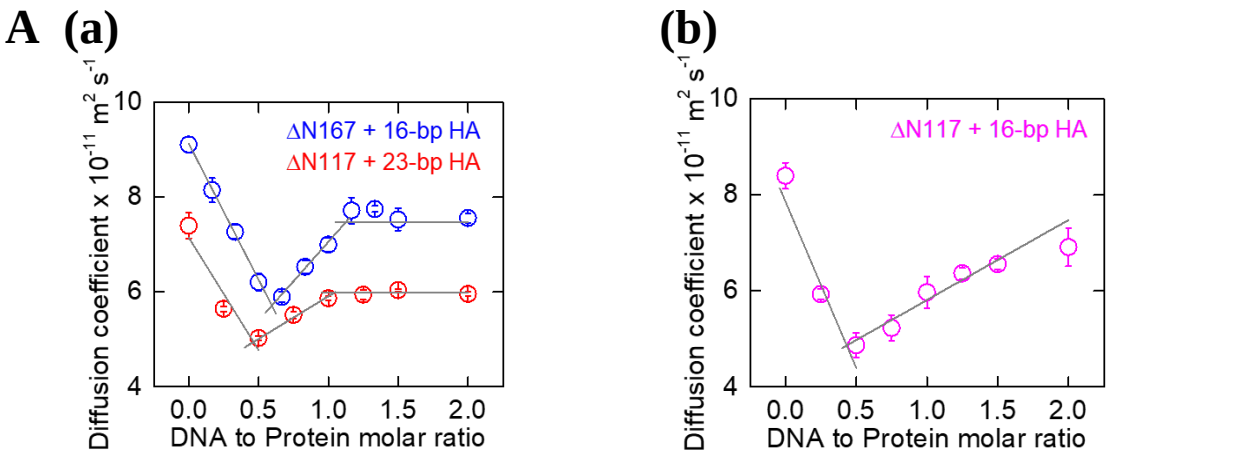
To examine the structural perturbation in the ETS domain upon specific DNA binding, we carried out chemical shift perturbation (CSP) analysis of unbound and 1:1 DNA-bound ETS domain of hPU.1ΔN117. Quite large CSPs were observed for amides in H1, the turn between H2 and H3, and the wing (S3/S4) (Fig. 3.7 *A* and *B*); this is consistent with the trend seen in the CSP analysis of unbound and 1:1 specific DNA-bound Ets-1ΔN301 reported previously (33). Considering the H2 and H3 also exhibit large CSPs for both PU.1 and Ets-1, we suggest that PU.1 and Ets-1 bind specific DNAs via the same interface of the ETS domain.

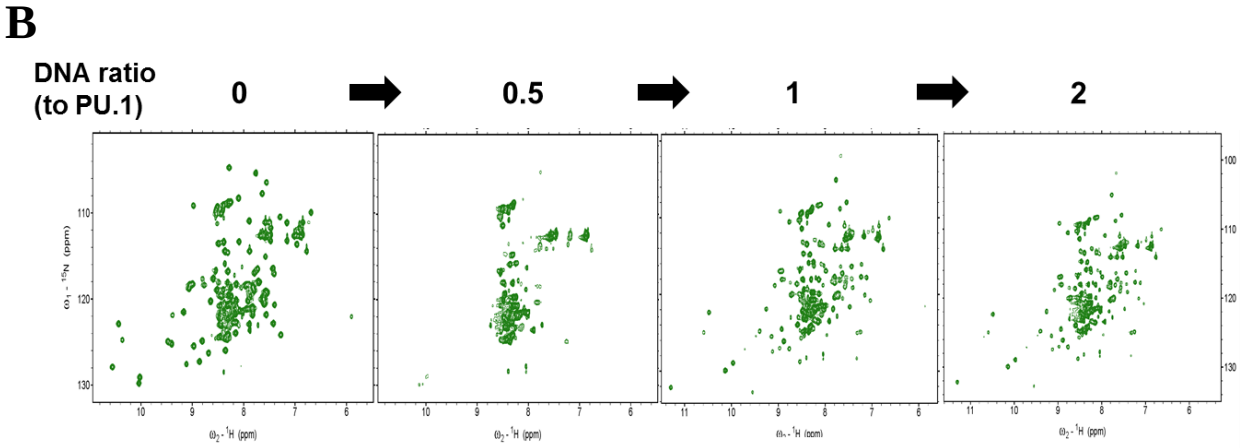
Interestingly enough, H1 and the wing are also the most perturbed sites according to the CSP analyses of DNA-free dimer formation of PU.1 (138) and the SRR moiety of Ets-1 (a peptide) binding to PU.1 (150). Furthermore, Met^{185} alone or the loop between H1 and S1 is presumably responsible for conformational exchange in the absence of DNA as we discussed above. Therefore, H1, the loop between H1 and S1, and the wing are presumably responsible for

ligand binding and protein-protein interactions (PPIs) of PU.1 in the absence of DNA. The four consecutive charged residues (^{193}DKDK^{196}) of hPU.1 on the loop between S1 and S2 distal to the DNA-binding site are at the self-dimerization interface of PU.1 ETS domain in the presence of site-specific DNA as described in Chapter 2. Thus, the site responsible for DNA-free PU.1 homodimerization (H1, the loop between H1 and S1, and the wing) is different from the DNA-bound dimeric interface.

3.7 Conclusion

We successfully assigned at least ~90% of HSQC resonances of hPU.1sΔN165, ΔN165, and ΔN117 proteins both in the absence and presence (1:1 binding) of cognate DNA. This provided us with opportunities to further analyze the PU.1 protein using NMR such as fast (ps to ns) time scale spin relaxation experiments. The protein dynamics study using NMR spectroscopy yielded relaxation and internal dynamics parameters (T_1, T_2, and heteronuclear ^{1}H{^{15}N}c-NOE) for unbound and 1:1 DNA-bound PU.1 proteins. Mainly using the NOE values, we demonstrated that the presence of disordered PEST domain does not change the internal dynamics of the ETS domain upon 1:1 specific DNA binding. The acidic PEST domain has repulsion with DNA in the 1:1 complex and becomes more dynamic than in the DNA-free PU.1, where the PEST domain stabilizes the homo-dimerization and transiently associates with the ETS domain. This study can mark a first step toward the characterization of self-regulatory mechanisms of some ETS family members that lack autoinhibition.


A (a)
(b)
Diffusion coefficient x 10^-11 m^2 s^-1
ΔN167 + 16-bp HA
ΔN117 + 23-bp HA
ΔN117 + 16-bp HA
DNA to Protein molar ratio


B
DNA ratio
(to PU.1)
0
0.5
1
2

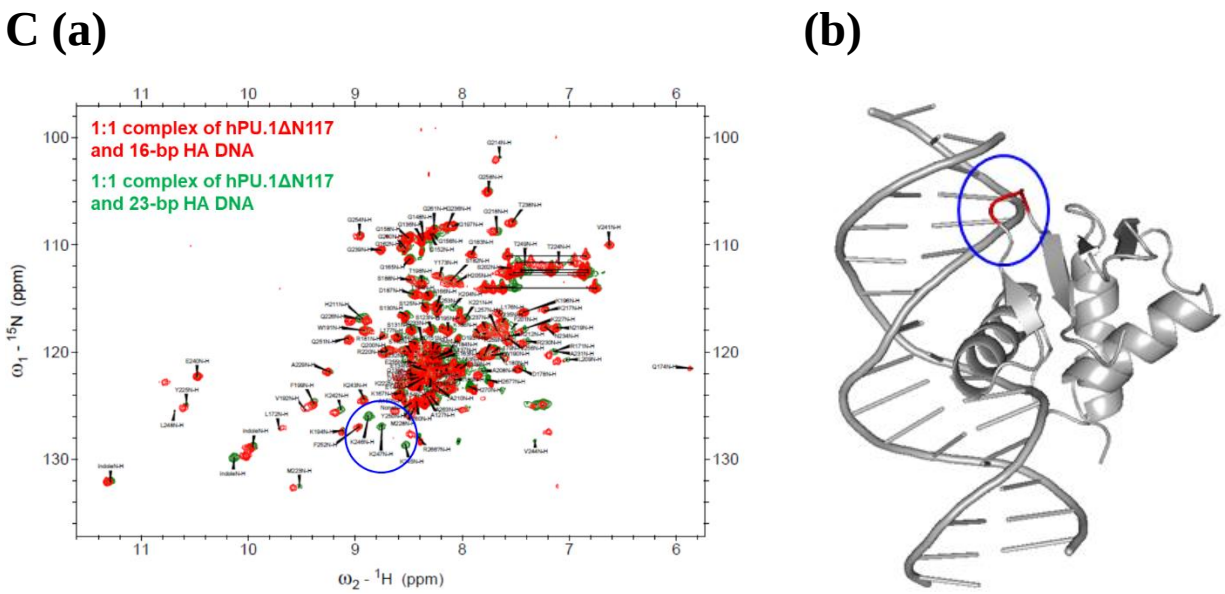
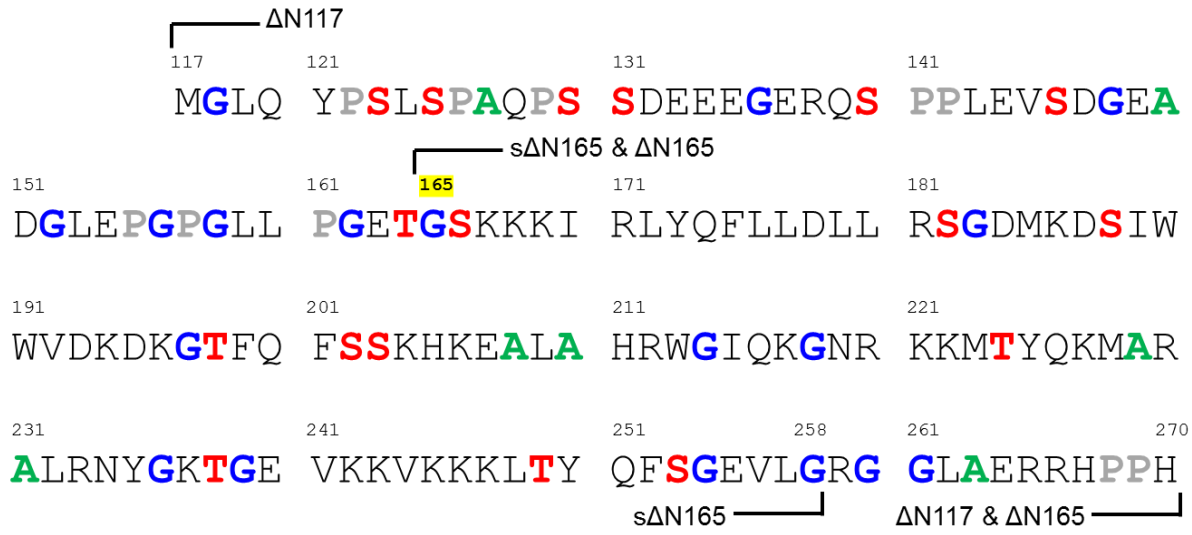

C (a)
(b)
1:1 complex of hPU.1ΔN117
and 16-bp HA DNA
1:1 complex of hPU.1ΔN117
and 23-bp HA DNA
ω1 - 15N (ppm)
ω2 - 1H (ppm)

Figure 3.1 DOSY and ^{1}H-^{15}N HSQC NMR titrations of hPU.1$^{117-270}$ by 23- and 16-bp site-specific DNA.

A, (a) A DOSY NMR titration of hPU.1ΔN117 with 23-bp specific DNA demonstrate two DNA-bound states of PU.1 (dimer at DNA:protein = 0.5 and monomer at 1:1). PU.1 dimer is not a 2:2 DNA complex like its homolog Ets-1 protein as indicated by the absence of a single minimum at DNA:protein = 1:1 when 23-bp DNA was used, as shown in red data points. The DOSY titration data using mPU.1ΔN167 and 16-bp specific DNA, as described in Chapter 2 (*cf.* Fig. 2.3), is shown in blue data points for comparison. (b) As shown in magenta data points, 16-bp DNA was not long enough for hPU.1ΔN117 protein to form the 1:1 and 2:1 PU.1/DNA complexes in direct response to the PU.1/DNA molar ratio. *B*, Uniformly ^{15}N-labeled hPU.1ΔN117 was titrated with a 23-bp unlabeled specific DNA at the indicated molar ratios. *C*, (a) An overlay of the HSQC spectra of 1:1 hPU.1ΔN117 complex with 16-bp HA DNA (red) onto 23-bp HA DNA (green). Three consecutive residues K245-247 shifted a lot between these two states (circled in blue in the HSQC spectrum of the 23-bp DNA). (b) Mapping the K245-247 residues (highlighted in red) on the structure of 1:1 complex (PDB 1PUE). These residues reside at the center of the wing (S3/S4) (circled in blue), and therefore they are highly perturbed by the DNA length.

A


ΔN117
117 121 131 141
MGLQ YPSLSPAQPS SDEEEGERQS PPLEVSDGEA
sΔN165 & ΔN165
151 161 165 171 181
DGLEPGPGLL PGETGSKKKI RLYQFLLDLL RSGDMKDSIW
191 201 211 221
WVDKDKGTFQ FSSKHKEALA HRWGIQKGNR KKMTYQKMAR
231 241 251 258 261 270
ALRNYGKTGE VKKVKKKLTY QFSGEVLGRG GLAERRHPPH
sΔN165
ΔN117 & ΔN165

B

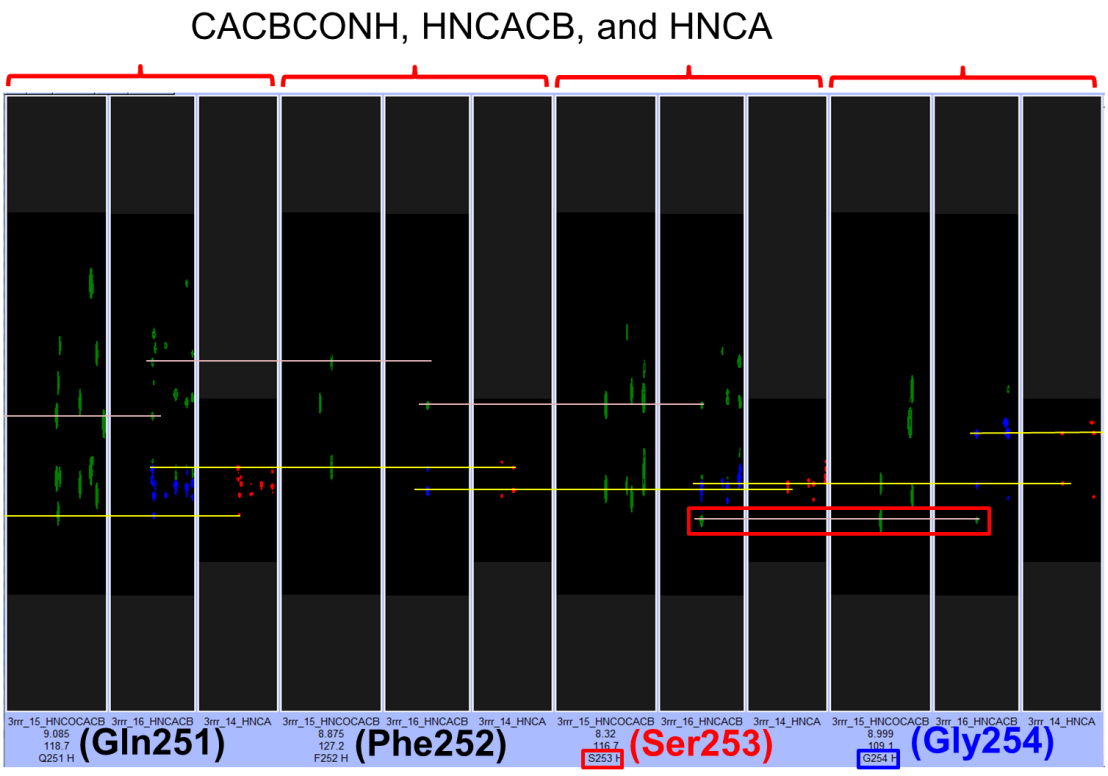
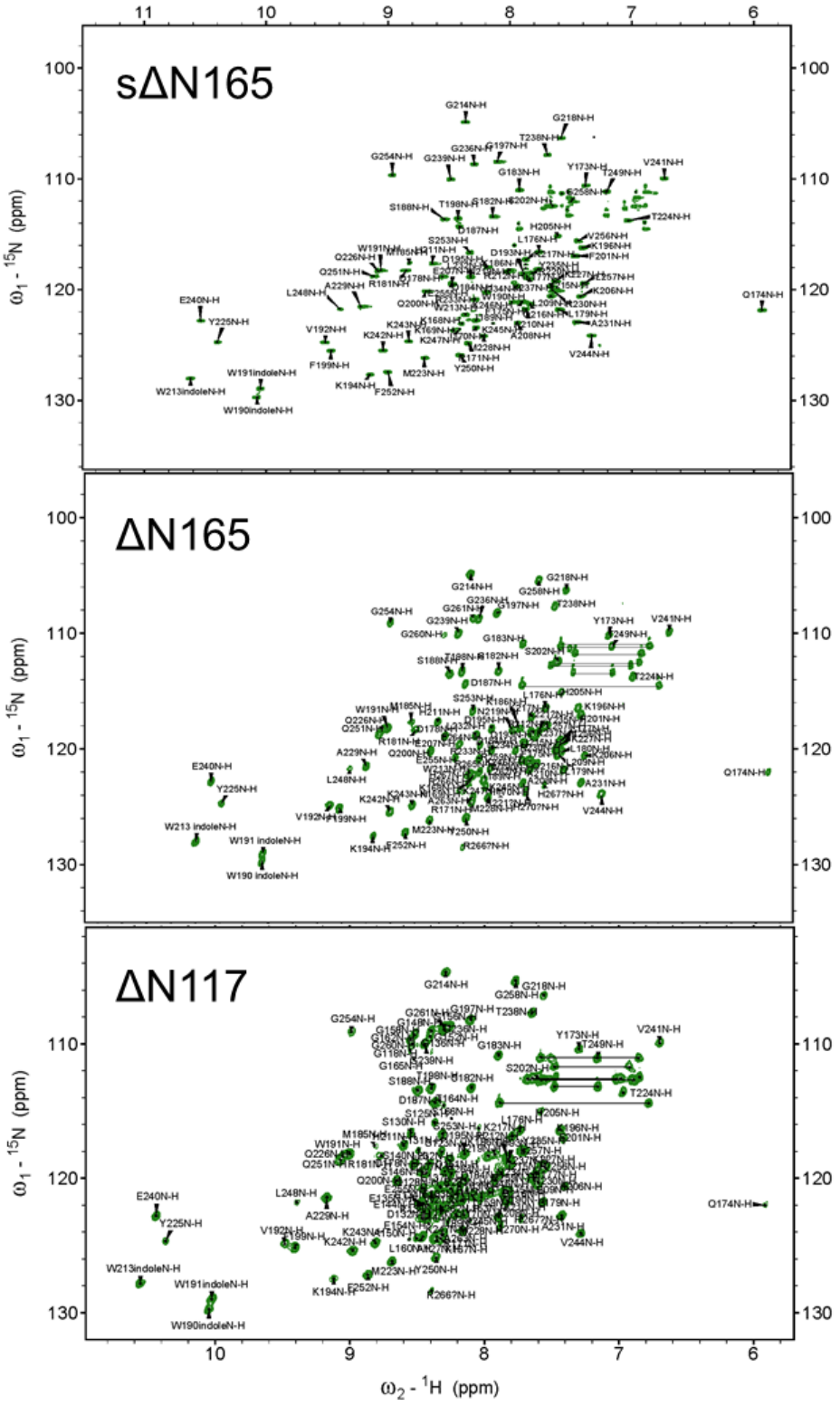

CACBCONH, HNCACB, and HNCA
(Gln251)
(Phe252)
(Ser253)
(Gly254)

Figure 3.2 Primary sequence of hPU.1 and partial strips of 3D NMR spectra used in the sequential backbone assignment.

A, Primary sequence of hPU.1$^{117\text{-}270}$. The N- and C-terminal residues of sPU.1 (165-258 aa), hPU.1ΔN165 (165-270 aa), and hPU.1ΔN117 (117-270 aa) are shown in L-shaped lines. The five specific amino acids that have characteristic carbon chemical shifts (G, A, S, and T) or have no ^{1}H attached to its ^{15}N (P) are shown in color and bold (G: blue; A: green; S and T: red; P: gray). Consecutive residues among these were able to assign directly. Thus, they were used as starting points for sequential assignments. *B*, A representative set of strips of 3D NMR spectra (CACBCONH, HNCACB, and HNCA from left to right) of Q251, F252, S253, and G254 residues of hPU.1ΔN165 in the absence of DNA. Connections of the alpha and beta carbon signals of each residue are shown in yellow lines. The red square indicates the beta carbon signals of S253.

(A) hPU.1 proteins in the absence of DNA



(B) hPU.1 proteins in 1:1 Complex with cognate DNA

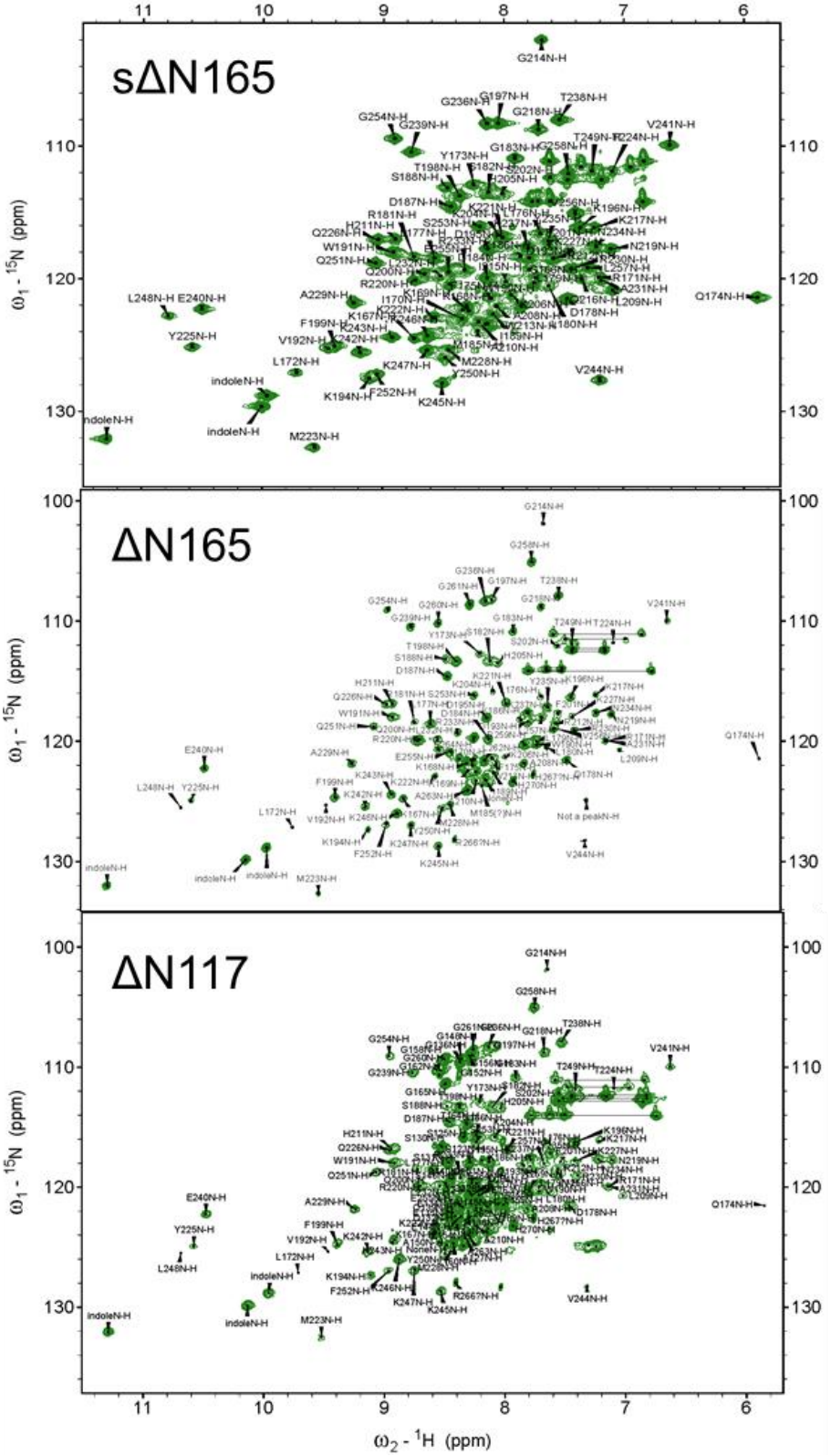
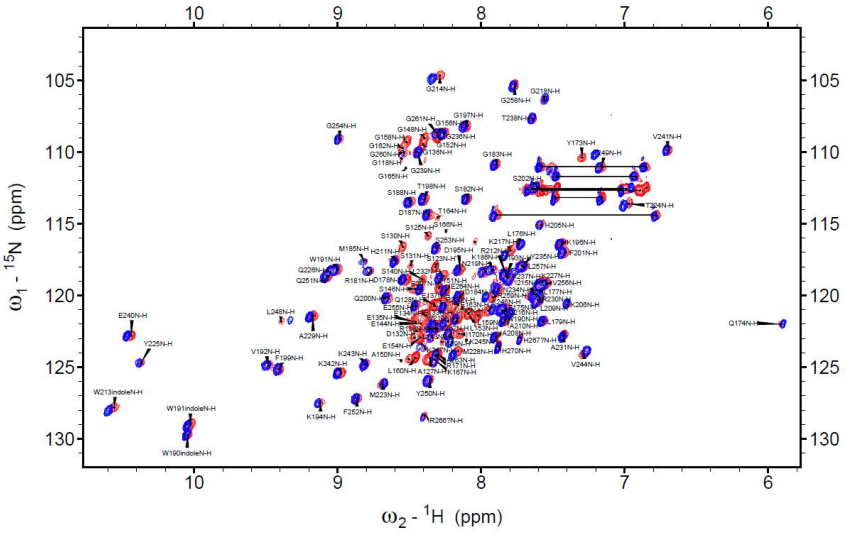


Figure 3.3 ^{1}H-^{15}N HSQC spectra of hPU.1 proteins (sΔN165, ΔN165, and ΔN117) in the absence of DNA and in the 1:1 complex with cognate DNA, with the resonances assigned.

^{1}H-^{15}N HSQC resonances of hPU.1 proteins sΔN165 (top panel), ΔN165 (middle panel), and ΔN117 (bottom panel) in the absence of DNA (*A*), and sΔN165 (top panel), ΔN165 (middle panel), and ΔN117 (bottom panel*)* in 1:1 complex with cognate DNA (*B*) were sequentially assigned using 3D NMR experiments HNCO, HN(CA)CO, HNCA, HNCACB, and HN(CO)CACB. Note that the resonances of hPU.1sΔN165 in the 1:1 complex with 16-bp site-specific DNA were assigned by our collaborator Dr. James Aramini.

A



B

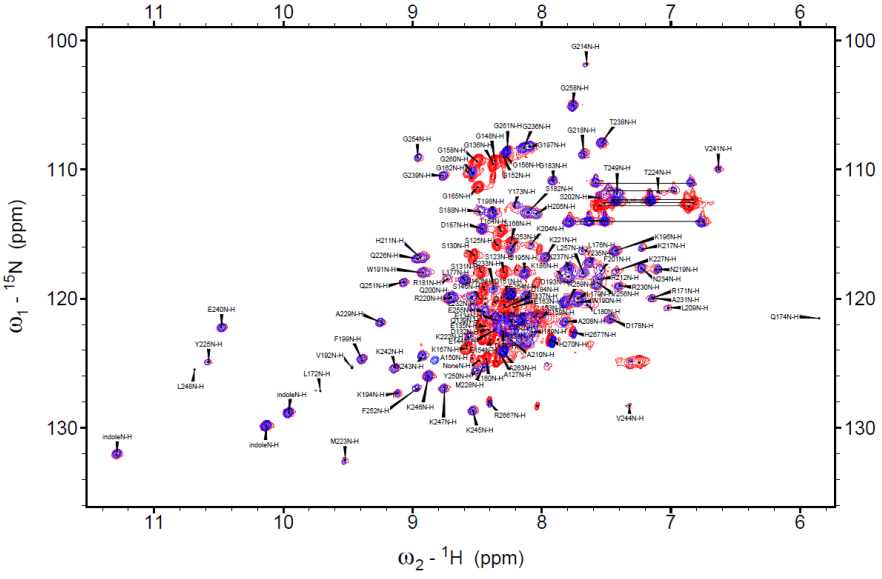
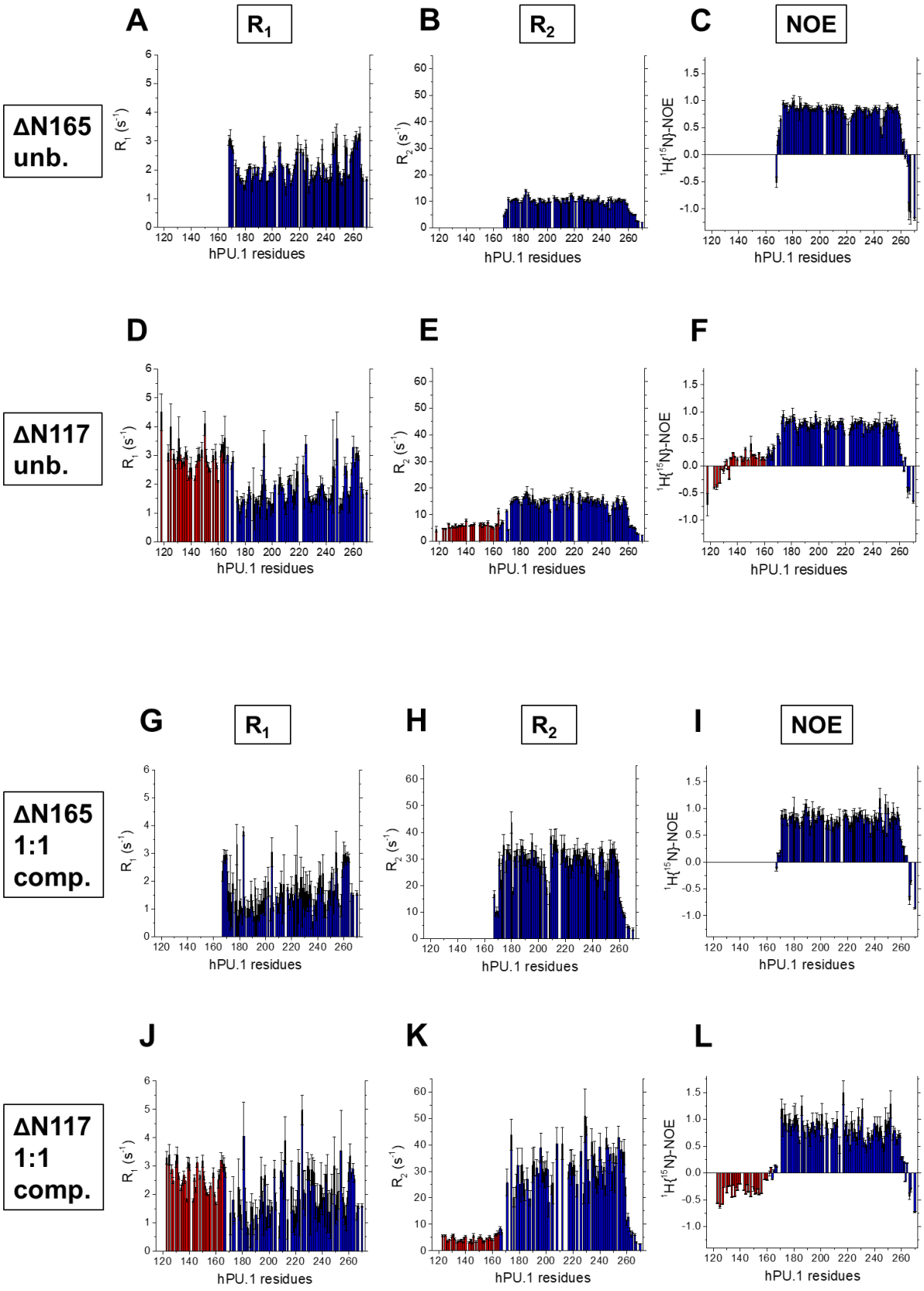
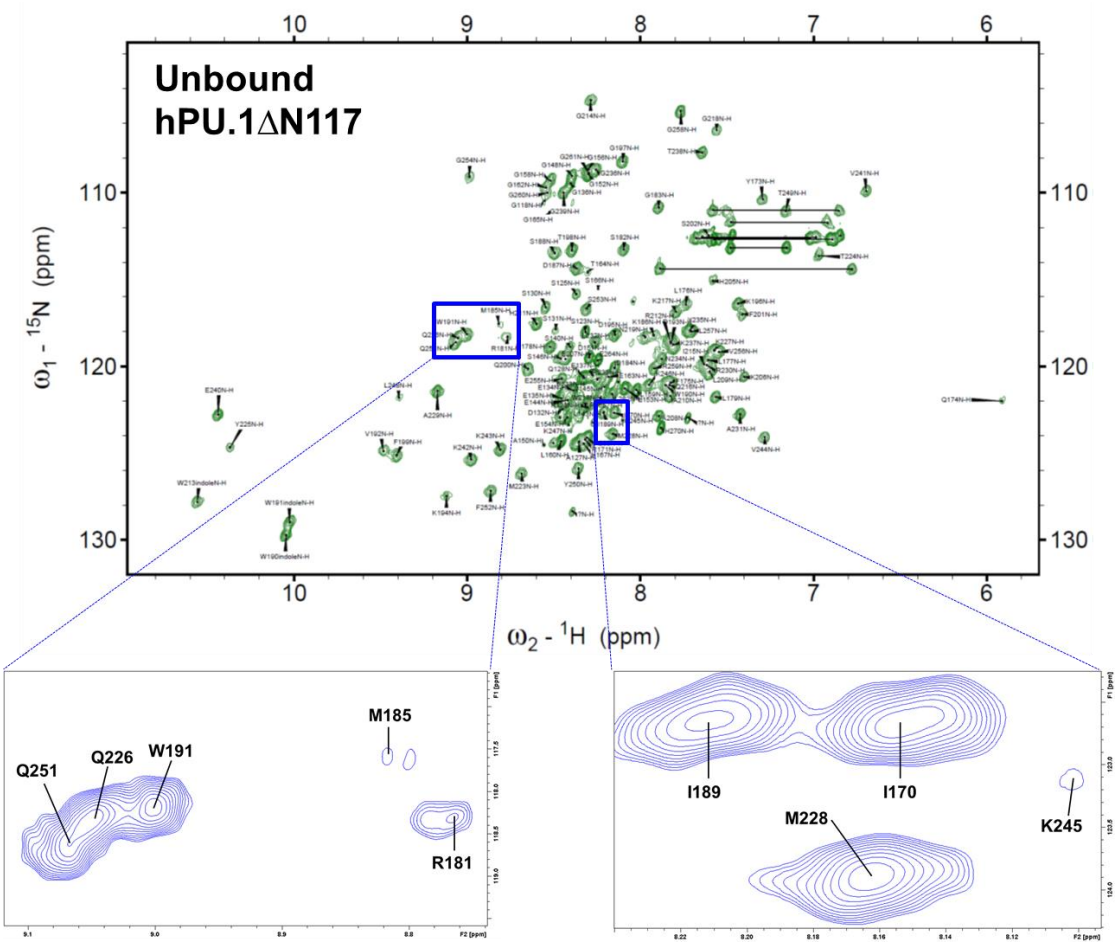


Figure 3.4 The presence of the PEST domain does not perturb the PU.1 ETS domain structurally, and it remains disordered upon DNA binding.

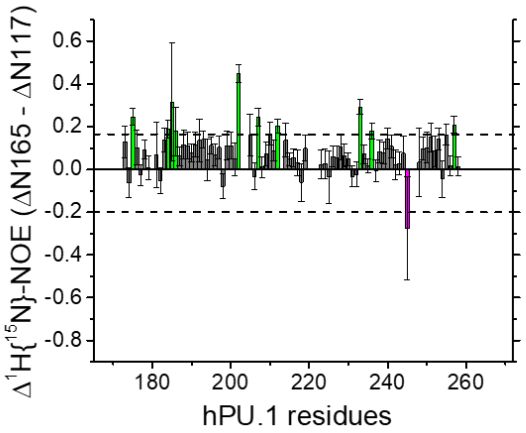
Overlaid HSQC spectra of hPU.1ΔN117 (red) and ΔN165 (blue) in the absence of DNA (*A*) and in the 1:1 complex with cognate DNA (*B*).


A
R1
B
R2
C
NOE
ΔN165 unb.
D
E
F
ΔN117 unb.
G
H
I
ΔN165 1:1 comp.
J
K
L
ΔN117 1:1 comp.
R1 (s-1)
R2 (s-1)
1H{15N}-NOE
hPU.1 residues

M


Unbound
hPU.1ΔN117
ω1 - 15N (ppm)
ω2 - 1H (ppm)
Q251
Q226
W191
M185
R181
I189
I170
M228
K245

N


Δ1Hζ{15Nζ}-NOE (ΔN165 - ΔN117)
hPU.1 residues

O

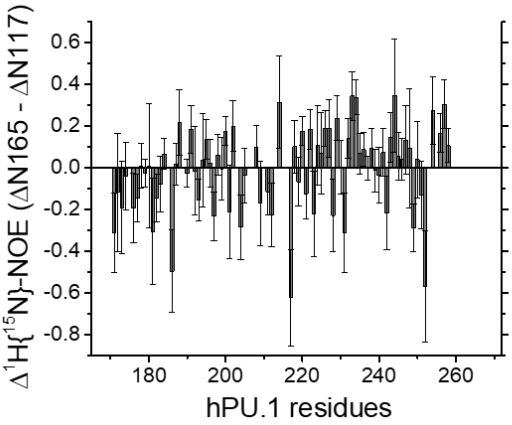

Δ1Hζ{15Nζ}-NOE (ΔN165 - ΔN117)
hPU.1 residues

Figure 3.5 Fast (picosecond to nanosecond) time scale backbone dynamics of unbound and DNA-bound (1:1 complex) PU.1 obtained by ^{15}N spin relaxation measurements.

R_1, R_2, and heteronuclear $^{1}H\{^{15}N\}$-NOE plots of unbound hPU.1ΔN165 (*A-C*) and ΔN117 (*D-F*) and 1:1 complex of ΔN165 (*G-I)* and ΔN117 (*J-L*) with 23-bp specific DNA (5'-GCGAATAAGC**GGAA**GTGAAACCG-3'). The color scheme follows the HSQC in Fig. 3.4. Heteronuclear $^{1}H\{^{15}N\}$-NOE error was derived by $|NOE|\sqrt{\{(\frac{noise\ Sat}{intensity\ Sat})^2 + (\frac{noise\ Unsat}{intensity\ Unsat})^2\}}$ using backgoround noise level of the spectra. Weak HSQC signals of M185 and K245 (*M*) resulted in high errors in the spin relaxation measurements. Subtracted heteronuclear NOE (ΔN165 - ΔN117) of DNA-free PU.1 (*N*) and 1:1 complex of PU.1 with the same DNA (*O*) are also shown.

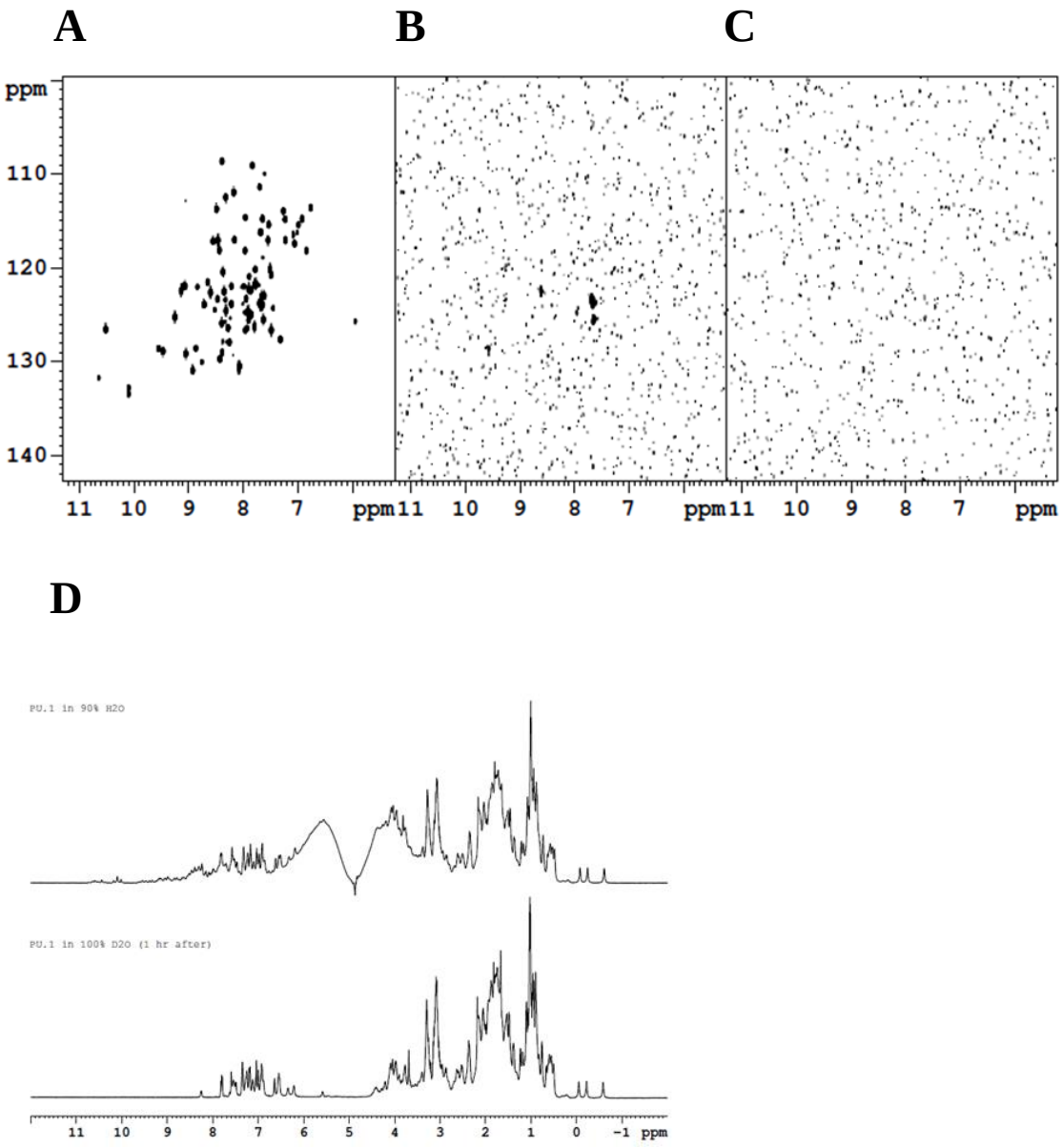
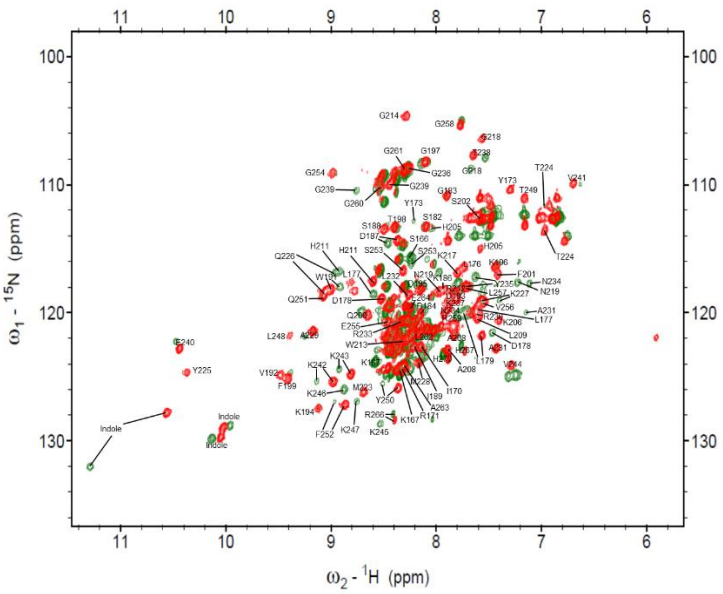


Figure 3.6 Amide hydrogen-deuterium exchange (HDX) of the PU.1 ETS domain.

A ^{1}H-^{15}N HSQC spectrum using mPU.1ΔN167 protein as a reference (*A*). The first ^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N-labeled mPU.1ΔN167 sample in 100% D_2O. Because most of the PU.1 crosspeaks disappeared immediately after bringing up the PU.1 with 100% D_2O and set up an HSQC experiment (only after ~15 min), slow amide-deuterium ^{1}H exchange rates were not able to measure using this sample (*B*). The HSQC spectrum using the same PU.1 sample in 100% D_2O after ~1 h (*C*). 1D ^{1}H NMR of the sample in (*A*) (top) and (*C*) (bottom), respectively. Note that the PU.1 protein in the bottom sample is folded judging from the methyl peaks.

A



B

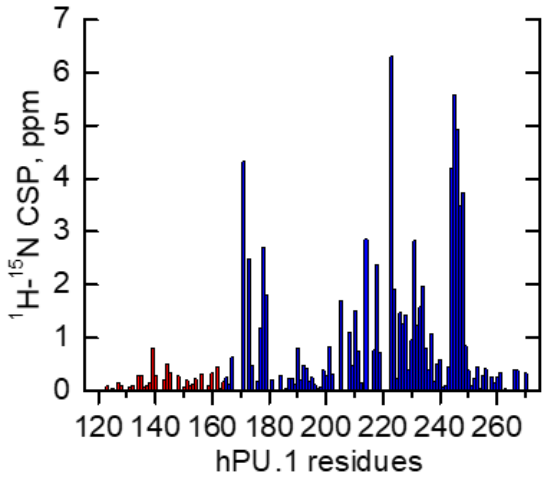


Figure 3.7 ^{1}H-^{15}N Chemical shift perturbations (CSPs) of unbound and DNA-bound (1:1 complex) PU.1 ETS domain.

A, an overlay of 1:1 specific DNA-bound (green) onto unbound (red) hPU.1ΔN117 HSQC spectra. *B*, weighed average of amide (^{15}N and ^{1}H) chemical shift perturbations (CSPs) from unbound and DNA-bound (1:1 complex with 23-bp specific DNA) PU.1 ETS domain (ΔN117), derived by $\Delta\delta = \sqrt{\{\delta^{1}H^{2} + 0.2\ (\delta^{15}N)^{2}\}}$, are plotted as a function of residue number in hPU.1.

Supplemental Table 3.1 Relaxation rates of DNA-free hPU.1ΔN165.

Res.	R_1 (1/s)	R_2 (1/s)	NOE
168	3.04 ± 0.18	4.67 ± 0.37	-0.514 ± 0.099
169	3.09 ± 0.30	5.74 ± 0.33	0.272 ± 0.090
170	2.86 ± 0.15	6.98 ± 0.15	0.465 ± 0.032
171	2.70 ± 0.12	11.03 ± 0.34	0.670 ± 0.046
173	2.24 ± 0.10	9.89 ± 0.26	0.970 ± 0.030
174	1.88 ± 0.09	10.34 ± 0.13	0.892 ± 0.030
175	2.07 ± 0.01	10.63 ± 0.16	0.921 ± 0.022
176	1.70 ± 0.10	10.66 ± 0.39	0.921 ± 0.034
177	1.61 ± 0.10	10.90 ± 0.26	0.802 ± 0.028
178	1.64 ± 0.04	10.50 ± 0.29	0.909 ± 0.020
179	1.47 ± 0.16	9.48 ± 0.16	0.878 ± 0.024
180	1.46 ± 0.15	25.40 ± 1.39	0.970 ± 0.081
181	1.79 ± 0.21	9.88 ± 0.36	1.004 ± 0.092
182	1.95 ± 0.06	10.83 ± 0.47	0.841 ± 0.028
183	2.13 ± 0.07	12.00 ± 0.34	0.870 ± 0.031
184	2.12 ± 0.09	14.22 ± 0.43	0.832 ± 0.025
185	1.65 ± 0.42	20.83 ± 0.48	0.856 ± 0.223
186	1.77 ± 0.09	12.75 ± 0.52	0.959 ± 0.085
187	2.11 ± 0.06	11.27 ± 0.33	0.830 ± 0.023
188	1.83 ± 0.04	9.87 ± 0.30	0.876 ± 0.034
189	1.96 ± 0.13	10.04 ± 0.23	0.917 ± 0.035
190	2.07 ± 0.01	10.63 ± 0.16	0.921 ± 0.022
191	1.63 ± 0.02	9.98 ± 0.24	0.892 ± 0.032
192	1.74 ± 0.13	8.78 ± 0.23	0.861 ± 0.056
193	2.00 ± 0.07	9.22 ± 0.24	0.842 ± 0.032
194	2.98 ± 0.19	10.91 ± 0.29	0.848 ± 0.052
195	2.52 ± 0.28	10.70 ± 0.25	0.839 ± 0.020
196	1.55 ± 0.06	9.73 ± 0.21	0.820 ± 0.028
197	1.61 ± 0.02	10.16 ± 0.61	0.836 ± 0.027
198	1.87 ± 0.06	9.89 ± 0.39	0.875 ± 0.027
199	1.89 ± 0.04	9.10 ± 0.42	0.922 ± 0.037
200	1.83 ± 0.05	9.59 ± 0.28	0.862 ± 0.033
201	1.97 ± 0.10	10.98 ± 0.34	0.877 ± 0.041
202	2.14 ± 0.12	9.79 ± 0.18	0.824 ± 0.032
205	2.71 ± 0.24	11.72 ± 0.47	0.869 ± 0.057
206	2.81 ± 0.14	10.50 ± 0.42	0.781 ± 0.028
207	2.17 ± 0.04	11.10 ± 0.30	0.861 ± 0.024
208	2.10 ± 0.04	10.31 ± 0.26	0.895 ± 0.022
209	1.58 ± 0.09	9.74 ± 0.40	0.820 ± 0.032
210	1.28 ± 0.15	9.10 ± 0.57	0.917 ± 0.025
211	2.15 ± 0.05	10.91 ± 0.30	0.867 ± 0.019

Res.	R_1 (1/s)	R_2 (1/s)	NOE
212	2.03 ± 0.16	10.89 ± 0.27	0.906 ± 0.024
213	1.95 ± 0.07	9.35 ± 0.24	0.786 ± 0.024
214	1.54 ± 0.10	10.91 ± 0.38	0.890 ± 0.037
215	1.71 ± 0.11	11.17 ± 0.35	0.848 ± 0.013
216	2.03 ± 0.14	9.63 ± 0.30	0.888 ± 0.026
217	2.19 ± 0.14	12.45 ± 0.45	0.853 ± 0.026
218	2.63 ± 0.10	12.52 ± 0.60	0.718 ± 0.044
219	2.96 ± 0.24	11.40 ± 0.33	0.713 ± 0.043
221	2.73 ± 0.14	9.90 ± 0.14	0.607 ± 0.053
223	2.61 ± 0.09	11.22 ± 0.47	0.649 ± 0.035
224	1.96 ± 0.03	11.55 ± 0.40	0.734 ± 0.033
225	2.91 ± 0.13	11.23 ± 0.14	0.766 ± 0.056
226	2.40 ± 0.10	12.15 ± 0.27	0.783 ± 0.024
227	1.30 ± 0.12	10.45 ± 0.35	0.880 ± 0.032
228	1.60 ± 0.06	9.84 ± 0.23	0.852 ± 0.025
229	1.98 ± 0.18	10.39 ± 0.23	0.879 ± 0.025
230	1.71 ± 0.11	11.17 ± 0.35	0.848 ± 0.013
231	1.79 ± 0.18	9.59 ± 0.19	0.810 ± 0.020
232	1.82 ± 0.13	10.54 ± 0.27	0.808 ± 0.024
233	1.95 ± 0.13	9.86 ± 0.50	0.860 ± 0.022
234	2.27 ± 0.15	10.39 ± 0.31	0.861 ± 0.021
235	2.09 ± 0.14	10.40 ± 0.35	0.708 ± 0.022
236	1.74 ± 0.06	10.04 ± 0.17	0.831 ± 0.025
237	2.88 ± 0.17	10.95 ± 0.28	0.768 ± 0.039
238	2.18 ± 0.05	11.76 ± 0.45	0.746 ± 0.032
239	1.64 ± 0.06	9.18 ± 0.28	0.803 ± 0.029
240	2.01 ± 0.05	9.55 ± 0.19	0.846 ± 0.023
241	1.64 ± 0.09	10.13 ± 0.28	0.885 ± 0.027
242	1.67 ± 0.06	8.57 ± 0.29	0.838 ± 0.038
243	2.14 ± 0.03	8.30 ± 0.21	0.784 ± 0.027
244	1.76 ± 0.12	9.37 ± 0.23	0.869 ± 0.041
245	2.93 ± 0.39	11.01 ± 0.58	0.517 ± 0.151
246	2.68 ± 0.11	7.01 ± 0.27	0.345 ± 0.029
247	3.10 ± 0.39	11.27 ± 0.57	0.684 ± 0.155
248	3.23 ± 0.37	10.02 ± 0.36	0.707 ± 0.097
249	2.17 ± 0.09	9.50 ± 0.21	0.819 ± 0.032
250	1.89 ± 0.05	10.45 ± 0.44	0.919 ± 0.040
251	1.59 ± 0.09	9.85 ± 0.32	0.882 ± 0.039
252	1.58 ± 0.12	10.61 ± 0.24	0.862 ± 0.036
253	1.89 ± 0.12	11.04 ± 0.23	0.876 ± 0.026
254	3.01 ± 0.20	10.16 ± 0.54	0.813 ± 0.043
255	2.78 ± 0.26	10.30 ± 0.33	0.841 ± 0.025

Res.	R_1 (1/s)	R_2 (1/s)	NOE
256	1.76 ± 0.03	10.51 ± 0.28	0.853 ± 0.027
257	1.78 ± 0.07	10.48 ± 0.24	0.900 ± 0.029
258	2.39 ± 0.12	9.23 ± 0.36	0.788 ± 0.025
259	2.64 ± 0.08	9.02 ± 0.21	0.701 ± 0.033
260	2.78 ± 0.22	6.93 ± 0.28	0.483 ± 0.080
261	2.95 ± 0.20	5.86 ± 0.18	0.237 ± 0.048
262	3.21 ± 0.13	5.42 ± 0.13	0.261 ± 0.025
263	2.86 ± 0.16	4.94 ± 0.17	-0.053 ± 0.026
264	3.13 ± 0.12	4.85 ± 0.13	0.078 ± 0.022
265	3.26 ± 0.23	5.01 ± 0.28	-0.166 ± 0.049
266	2.06 ± 0.13	2.63 ± 0.16	-1.043 ± 0.160
267	1.73 ± 0.12	2.47 ± 0.03	-1.172 ± 0.114
270	1.68 ± 0.07	2.12 ± 0.06	-1.188 ± 0.032

Supplemental Table 3.2 Relaxation rates of DNA-free hPU.1ΔN117.

Res.	R_1 (1/s)	R_2 (1/s)	NOE
118	4.50 ± 0.63	4.49 ± 1.01	-0.717 ± 0.205
123	3.09 ± 0.28	4.82 ± 0.12	-0.410 ± 0.023
125	3.99 ± 0.80	4.69 ± 0.27	-0.396 ± 0.044
127	3.04 ± 0.18	6.57 ± 0.18	-0.318 ± 0.012
128	2.70 ± 0.14	5.03 ± 0.12	-0.067 ± 0.007
130	2.93 ± 0.21	5.65 ± 0.18	-0.110 ± 0.026
131	3.58 ± 0.76	5.14 ± 0.20	0.017 ± 0.059
132	3.03 ± 0.23	6.24 ± 0.40	0.104 ± 0.017
133	2.70 ± 0.14	5.03 ± 0.12	-0.067 ± 0.007
134	2.79 ± 0.10	5.42 ± 0.15	-0.239 ± 0.011
135	2.79 ± 0.15	5.91 ± 0.18	0.146 ± 0.010
136	3.12 ± 0.18	5.45 ± 0.19	0.141 ± 0.015
137	3.01 ± 0.14	6.20 ± 0.21	0.240 ± 0.010
138	2.21 ± 0.07	5.69 ± 0.14	0.168 ± 0.014
139	2.57 ± 0.13	5.71 ± 0.12	
140	2.56 ± 0.37	7.94 ± 0.50	0.129 ± 0.030
143	2.21 ± 0.07	5.69 ± 0.14	0.168 ± 0.014
144	2.69 ± 0.09	6.21 ± 0.15	0.088 ± 0.013
145	3.02 ± 0.21	6.07 ± 0.16	0.144 ± 0.015
146	3.05 ± 0.23	6.58 ± 0.30	0.318 ± 0.026
148	3.19 ± 0.26		0.123 ± 0.020
150	4.11 ± 0.42	6.30 ± 0.55	0.409 ± 0.136
151	3.00 ± 0.22	6.84 ± 0.16	0.194 ± 0.015
152	2.68 ± 0.07	5.34 ± 0.16	0.205 ± 0.009
153	2.58 ± 0.16	6.71 ± 0.37	0.202 ± 0.014
154	2.49 ± 0.06	5.86 ± 0.15	0.144 ± 0.013
156	3.00 ± 0.21	7.26 ± 0.47	0.233 ± 0.012
158	3.13 ± 0.27	5.75 ± 0.08	0.129 ± 0.023
159	2.65 ± 0.10	5.39 ± 0.15	0.163 ± 0.013
160	2.10 ± 0.03	4.87 ± 0.27	0.113 ± 0.013
162	3.16 ± 0.21	6.09 ± 0.54	0.180 ± 0.033
163	3.39 ± 0.29	6.33 ± 0.26	0.332 ± 0.021
164	3.10 ± 0.36	11.37 ± 1.05	0.228 ± 0.060
165	3.60 ± 0.75	5.29 ± 0.67	0.200 ± 0.082
166	0.00 ± 0.00	7.00 ± 0.90	0.247 ± 0.149
167	3.01 ± 0.27	7.16 ± 0.66	0.350 ± 0.030
170	2.65 ± 0.12	11.53 ± 0.47	0.564 ± 0.032
171	2.94 ± 0.19	4.17 ± 0.15	0.451 ± 0.043
173		14.56 ± 0.74	0.841 ± 0.068
174	1.57 ± 0.20	15.14 ± 0.72	0.954 ± 0.064
175	1.30 ± 0.16	15.33 ± 0.95	0.676 ± 0.032

Res.	R_1 (1/s)	R_2 (1/s)	NOE
176	0.86 ± 0.19	15.86 ± 0.63	0.820 ± 0.074
177	1.28 ± 0.29	16.19 ± 0.60	0.829 ± 0.042
178	1.39 ± 0.16	16.20 ± 0.48	0.816 ± 0.041
179	1.48 ± 0.20	14.94 ± 0.68	0.870 ± 0.052
181	1.11 ± 0.26	14.08 ± 0.54	0.937 ± 0.125
182	1.49 ± 0.16	16.29 ± 0.57	0.895 ± 0.054
183	1.92 ± 0.14	16.83 ± 0.82	0.731 ± 0.046
184	1.59 ± 0.12	18.01 ± 0.78	0.644 ± 0.035
185	0.00 ± 0.00	18.48 ± 2.13	0.542 ± 0.163
186	1.65 ± 0.92	15.71 ± 2.03	0.782 ± 0.077
187	0.87 ± 0.45	16.48 ± 0.91	0.770 ± 0.036
188	1.47 ± 0.07	14.41 ± 0.38	0.761 ± 0.057
189	1.25 ± 0.25	17.40 ± 2.22	0.808 ± 0.053
190	1.44 ± 0.14	14.67 ± 0.64	0.845 ± 0.035
191	1.21 ± 0.38	14.96 ± 0.56	0.777 ± 0.049
192	1.78 ± 0.52	12.96 ± 0.37	0.725 ± 0.083
193	1.97 ± 0.50	14.46 ± 0.69	0.703 ± 0.022
194	3.40 ± 0.45	12.72 ± 0.58	0.805 ± 0.083
195	1.98 ± 0.18	15.74 ± 0.18	0.727 ± 0.035
196	1.15 ± 0.38	14.45 ± 0.59	0.752 ± 0.041
197	1.43 ± 0.09	14.65 ± 0.75	0.733 ± 0.046
198	1.38 ± 0.12	15.95 ± 0.39	0.955 ± 0.050
199	1.30 ± 0.40	15.06 ± 0.87	0.811 ± 0.057
200	1.21 ± 0.11	15.14 ± 0.94	0.751 ± 0.054
201	1.35 ± 0.35	15.39 ± 0.66	0.830 ± 0.068
202	1.98 ± 0.12	11.47 ± 0.45	0.376 ± 0.028
205	2.00 ± 0.34	16.28 ± 0.87	0.708 ± 0.077
206	2.25 ± 0.32	15.98 ± 0.64	0.814 ± 0.057
207	1.89 ± 0.13	12.73 ± 1.12	0.617 ± 0.035
208	1.87 ± 0.10	16.38 ± 0.42	0.884 ± 0.043
209	1.41 ± 0.24	15.04 ± 0.36	0.748 ± 0.044
210	1.17 ± 0.22	16.76 ± 0.45	0.750 ± 0.046
211	1.43 ± 0.09	16.92 ± 0.71	0.779 ± 0.043
212	2.13 ± 0.12	14.40 ± 0.54	0.703 ± 0.022
213	2.21 ± 0.07		
214	1.99 ± 0.31	16.33 ± 0.97	0.753 ± 0.071
215	1.49 ± 0.11	17.21 ± 1.00	0.801 ± 0.026
216	1.81 ± 0.67	12.90 ± 0.90	0.834 ± 0.032
217	1.52 ± 0.22	18.03 ± 0.59	0.822 ± 0.056
218	2.21 ± 0.49	17.68 ± 2.39	0.779 ± 0.076
219	2.44 ± 0.50	16.72 ± 1.92	0.615 ± 0.041
223	2.68 ± 0.31	16.37 ± 2.14	0.626 ± 0.057

Res.	R_1 (1/s)	R_2 (1/s)	NOE
224	1.34 ± 0.18	17.97 ± 0.89	0.709 ± 0.061
225	3.39 ± 0.30	12.90 ± 0.95	0.801 ± 0.109
226	2.20 ± 0.09	15.86 ± 0.70	0.723 ± 0.043
227	1.58 ± 0.17	15.23 ± 0.61	0.826 ± 0.042
228	1.47 ± 0.10	15.67 ± 0.70	0.748 ± 0.053
229	1.21 ± 0.18	16.48 ± 1.71	0.814 ± 0.048
230	1.49 ± 0.11	16.92 ± 0.79	0.801 ± 0.026
231	1.41 ± 0.12	14.56 ± 0.28	0.843 ± 0.045
232	1.49 ± 0.14	16.02 ± 0.59	0.831 ± 0.053
233	1.52 ± 0.15	15.16 ± 1.75	0.568 ± 0.025
234	1.83 ± 0.09	14.61 ± 0.53	0.790 ± 0.038
235	1.58 ± 0.31	15.33 ± 0.24	0.692 ± 0.026
236	1.94 ± 0.34	13.72 ± 0.88	0.651 ± 0.028
237	2.02 ± 0.21	15.63 ± 0.38	0.773 ± 0.032
238	1.74 ± 0.22	15.10 ± 0.79	0.664 ± 0.043
239	1.56 ± 0.10	13.52 ± 0.45	0.727 ± 0.043
240	1.53 ± 0.12	15.09 ± 0.41	0.703 ± 0.043
241		16.79 ± 0.72	0.778 ± 0.045
242	1.50 ± 0.38	13.60 ± 0.39	0.817 ± 0.056
243	1.48 ± 0.18	12.57 ± 0.28	0.756 ± 0.045
244	1.67 ± 0.27	14.19 ± 0.76	0.794 ± 0.067
245	2.60 ± 1.62	9.05 ± 1.54	0.792 ± 0.190
246		9.69 ± 0.10	
248	2.53 ± 0.30	12.46 ± 0.73	0.674 ± 0.127
249	3.59 ± 0.92	14.30 ± 0.37	0.722 ± 0.042
250	1.26 ± 0.25	12.26 ± 0.43	0.818 ± 0.060
251	1.31 ± 0.18	15.56 ± 0.89	0.732 ± 0.056
252	1.31 ± 0.18	16.19 ± 0.82	0.777 ± 0.062
253	1.72 ± 0.19	13.80 ± 1.87	0.735 ± 0.044
254	2.63 ± 0.30	12.59 ± 0.60	0.859 ± 0.075
255	2.47 ± 0.25	13.81 ± 0.58	0.680 ± 0.040
256	1.65 ± 0.09	16.11 ± 0.76	0.836 ± 0.040
257	1.31 ± 0.30	15.38 ± 0.28	0.692 ± 0.026
258	1.76 ± 0.15	14.20 ± 0.55	0.775 ± 0.038
259	2.53 ± 0.21	12.36 ± 0.31	0.587 ± 0.033
260	3.29 ± 0.37	8.69 ± 0.73	0.363 ± 0.048
261	2.68 ± 0.07	5.34 ± 0.16	0.205 ± 0.009
262	3.07 ± 0.22	5.95 ± 0.20	0.080 ± 0.014
263	3.09 ± 0.17	5.17 ± 0.20	-0.095 ± 0.011
264	2.99 ± 0.16	5.26 ± 0.12	0.150 ± 0.013
266	2.05 ± 0.16	3.08 ± 0.15	-0.492 ± 0.098
267	1.68 ± 0.14	2.62 ± 0.07	-0.462 ± 0.060
270	1.72 ± 0.06	2.30 ± 0.04	-0.669 ± 0.018

Supplemental Table 3.3 Relaxation rates of 1:1 DNA-bound hPU.1ΔN165.

Res.	R_1 (1/s)	R_2 (1/s)	NOE
167	2.35 ± 0.36	16.70 ± 1.24	-0.136 ± 0.035
168	2.94 ± 0.19	9.17 ± 0.85	0.163 ± 0.008
169	2.87 ± 0.23	10.48 ± 1.38	0.114 ± 0.009
170	2.95 ± 0.16	9.13 ± 0.97	0.195 ± 0.020
171	1.63 ± 0.70	30.05 ± 2.55	0.885 ± 0.071
172	1.25 ± 0.66	22.14 ± 3.54	0.818 ± 0.128
173	1.59 ± 1.35	29.85 ± 1.57	0.894 ± 0.090
174	0.80 ± 0.44	35.22 ± 3.89	0.894 ± 0.079
175	1.92 ± 0.36	29.21 ± 4.22	0.732 ± 0.058
177	1.29 ± 0.42	33.53 ± 1.81	0.859 ± 0.052
178	3.31 ± 0.73	32.39 ± 2.72	0.789 ± 0.060
179	0.74 ± 0.21	31.51 ± 1.91	0.889 ± 0.033
180	1.29 ± 0.74	43.76 ± 4.01	1.023 ± 0.189
181	1.10 ± 0.27	17.84 ± 1.30	0.770 ± 0.138
182	0.73 ± 0.30	30.92 ± 1.80	0.838 ± 0.062
183		29.91 ± 3.01	0.811 ± 0.066
184	1.10 ± 0.35	33.45 ± 2.09	0.680 ± 0.048
186	1.02 ± 0.30	31.17 ± 2.29	0.755 ± 0.072
187	1.01 ± 0.26	34.93 ± 2.27	0.894 ± 0.052
188	0.86 ± 0.13	34.77 ± 2.33	0.977 ± 0.103
189	1.52 ± 0.47	29.90 ± 2.31	1.088 ± 0.078
190	1.08 ± 0.21	31.51 ± 1.91	0.889 ± 0.033
191	0.74 ± 0.17	29.55 ± 2.27	0.948 ± 0.077
192	1.79 ± 0.30	26.59 ± 5.20	0.637 ± 0.137
193	0.76 ± 0.61	30.05 ± 2.03	0.689 ± 0.050
194	1.17 ± 0.56	29.97 ± 3.07	0.983 ± 0.147
195	0.92 ± 0.37	36.29 ± 3.09	0.936 ± 0.062
196	1.16 ± 0.35	30.43 ± 2.90	0.787 ± 0.065
197	1.27 ± 0.43	30.25 ± 2.67	0.780 ± 0.045
198	1.29 ± 0.41	32.76 ± 2.21	0.952 ± 0.059
199	1.39 ± 0.61	28.50 ± 2.05	0.947 ± 0.081
200	1.76 ± 0.40	28.59 ± 1.80	0.819 ± 0.038
201		27.31 ± 1.82	0.880 ± 0.090
202	1.97 ± 0.53	28.98 ± 4.69	0.874 ± 0.121
204	1.03 ± 0.46	29.16 ± 2.37	0.688 ± 0.073
205	3.05 ± 0.52	24.31 ± 2.57	0.787 ± 0.075
206	1.08 ± 0.14	19.17 ± 2.59	0.688 ± 0.088
208	1.46 ± 0.31	17.19 ± 4.28	0.833 ± 0.061
209	1.21 ± 0.36	38.35 ± 2.67	0.655 ± 0.089
210	1.13 ± 0.27	28.58 ± 2.00	0.727 ± 0.067
211	1.56 ± 0.36	35.27 ± 3.41	0.733 ± 0.062

Res.	R_1 (1/s)	R_2 (1/s)	NOE
212	1.89 ± 0.46	36.99 ± 4.08	0.775 ± 0.067
213	1.04 ± 0.56	36.93 ± 4.49	0.685 ± 0.050
214	1.89 ± 1.05	33.81 ± 1.20	0.834 ± 0.155
217	1.77 ± 0.11	33.50 ± 5.49	0.871 ± 0.080
218	1.35 ± 0.20	31.39 ± 2.56	0.900 ± 0.072
219	1.26 ± 0.30	35.78 ± 3.17	0.899 ± 0.064
220	1.76 ± 0.40	28.59 ± 1.80	0.819 ± 0.038
221	1.29 ± 0.23	31.24 ± 2.22	0.687 ± 0.062
222	1.50 ± 0.37	27.25 ± 3.57	0.651 ± 0.079
223	2.24 ± 0.69	30.29 ± 3.33	0.740 ± 0.101
224		25.21 ± 2.51	0.863 ± 0.101
225	0.93 ± 0.45	27.63 ± 2.18	0.888 ± 0.112
226	1.59 ± 0.14	31.42 ± 2.49	0.860 ± 0.052
227	2.13 ± 0.74	33.24 ± 2.24	0.830 ± 0.085
228	1.25 ± 0.89	29.53 ± 2.09	0.822 ± 0.072
229	1.75 ± 0.74	33.68 ± 4.06	0.943 ± 0.063
230	1.30 ± 0.58	31.59 ± 2.22	0.879 ± 0.072
231	1.63 ± 0.70	30.05 ± 2.55	0.885 ± 0.071
232	1.64 ± 0.57	27.81 ± 3.04	0.713 ± 0.063
233	1.21 ± 0.31	30.84 ± 3.20	0.885 ± 0.090
234	1.87 ± 0.50	32.37 ± 1.67	0.785 ± 0.049
235	1.54 ± 0.46	33.21 ± 2.25	0.817 ± 0.056
236	0.54 ± 0.24	33.07 ± 2.69	0.760 ± 0.045
237	1.13 ± 0.21	29.80 ± 1.35	0.502 ± 0.033
238	1.28 ± 0.52	29.65 ± 2.39	0.801 ± 0.056
239	0.95 ± 0.38	25.25 ± 1.76	0.684 ± 0.057
240	1.74 ± 0.20	31.95 ± 1.82	0.889 ± 0.070
241	2.12 ± 0.58	29.03 ± 2.76	0.845 ± 0.065
242	2.21 ± 0.44	24.10 ± 1.98	0.681 ± 0.087
243	1.53 ± 0.25	25.73 ± 1.98	0.827 ± 0.082
244	2.27 ± 0.69	22.04 ± 2.61	1.189 ± 0.183
245	0.96 ± 0.26	31.03 ± 2.89	0.836 ± 0.067
246	1.16 ± 0.48	32.48 ± 1.85	0.601 ± 0.054
247		33.81 ± 2.73	0.675 ± 0.089
248	1.31 ± 0.34	29.68 ± 2.70	1.080 ± 0.183
249	1.67 ± 0.28	16.34 ± 1.48	0.506 ± 0.032
250	1.56 ± 0.43	25.60 ± 3.13	1.006 ± 0.109
251	2.30 ± 0.21	28.14 ± 3.53	0.744 ± 0.069
252	1.63 ± 0.80	25.69 ± 4.02	0.709 ± 0.084
253	1.66 ± 0.21	33.45 ± 3.60	0.849 ± 0.083
254	3.04 ± 0.74	29.33 ± 4.35	1.012 ± 0.096
255	1.91 ± 0.54	33.68 ± 1.92	0.839 ± 0.056

Res.	R_1 (1/s)	R_2 (1/s)	NOE
256	1.22 ± 0.12	33.47 ± 2.85	0.763 ± 0.062
257		30.84 ± 2.78	0.934 ± 0.084
258	1.13 ± 0.39	29.65 ± 2.21	0.826 ± 0.049
259	2.40 ± 0.27	25.95 ± 2.65	0.640 ± 0.028
260	2.96 ± 0.45	14.33 ± 1.04	0.392 ± 0.024
261	2.97 ± 0.31	13.00 ± 0.98	0.293 ± 0.012
262	2.79 ± 0.21	10.62 ± 0.91	0.278 ± 0.008
263	2.88 ± 0.16	9.09 ± 0.83	0.117 ± 0.007
264	2.83 ± 0.15	8.67 ± 0.79	0.121 ± 0.006
266	1.59 ± 0.32	4.75 ± 0.52	-0.717 ± 0.074
267	1.51 ± 0.06	4.21 ± 0.50	-0.377 ± 0.036
270	1.57 ± 0.08	3.53 ± 0.45	-0.868 ± 0.011

Supplemental Table 3.4 Relaxation rates of 1:1 DNA-bound hPU.1ΔN117.

Res.	R_1 (1/s)	R_2 (1/s)	NOE
123	3.30 ± 0.21	5.66 ± 0.39	-0.560 ± 0.010
125	3.40 ± 0.36	5.66 ± 0.39	-0.638 ± 0.015
127	2.89 ± 0.15	3.61 ± 0.35	-0.576 ± 0.006
128	2.48 ± 0.11	4.03 ± 0.27	-0.269 ± 0.002
130	3.19 ± 0.22	5.38 ± 0.30	-0.364 ± 0.007
131	3.38 ± 0.29	5.91 ± 0.36	-0.253 ± 0.010
132	2.65 ± 0.12	4.46 ± 0.43	-0.241 ± 0.003
134	2.23 ± 0.09	3.23 ± 0.38	-0.448 ± 0.005
135	2.40 ± 0.09	3.71 ± 0.36	-0.228 ± 0.004
136	2.68 ± 0.14	3.73 ± 0.37	-0.440 ± 0.004
137	2.48 ± 0.11	4.03 ± 0.27	-0.269 ± 0.002
138	1.78 ± 0.06	3.09 ± 0.25	-0.323 ± 0.004
139	3.13 ± 0.19	4.71 ± 0.44	-0.201 ± 0.007
140	3.08 ± 0.22	5.35 ± 0.30	-0.208 ± 0.008
143	1.78 ± 0.06	3.09 ± 0.25	-0.323 ± 0.004
144	2.59 ± 0.11	4.02 ± 0.31	-0.389 ± 0.006
145	2.16 ± 0.09	2.76 ± 0.32	-0.216 ± 0.005
146	3.13 ± 0.18	5.88 ± 0.72	-0.352 ± 0.007
148	2.68 ± 0.14	3.49 ± 0.27	-0.440 ± 0.004
150	3.17 ± 0.21	4.90 ± 0.31	-0.299 ± 0.042
151	2.66 ± 0.14	5.34 ± 0.58	-0.388 ± 0.006
152	2.30 ± 0.09	4.57 ± 0.29	-0.384 ± 0.003
153	2.07 ± 0.08	3.27 ± 0.27	-0.296 ± 0.005
154	2.01 ± 0.05	3.69 ± 0.29	-0.403 ± 0.006
156	2.30 ± 0.09	4.57 ± 0.29	-0.384 ± 0.003
158	2.77 ± 0.18	5.32 ± 0.41	-0.117 ± 0.006
159	1.99 ± 0.07	4.20 ± 0.34	-0.120 ± 0.006
160	1.70 ± 0.04	4.23 ± 0.34	-0.129 ± 0.006
162	2.53 ± 0.31	5.87 ± 0.40	-0.008 ± 0.045
163	2.70 ± 0.17	6.01 ± 0.39	0.101 ± 0.007
164	3.21 ± 0.26	6.86 ± 0.47	-0.113 ± 0.010
165	3.08 ± 0.27	6.46 ± 0.42	0.047 ± 0.010
166	3.08 ± 0.21	8.54 ± 0.60	0.141 ± 0.010
167	2.77 ± 0.18	6.99 ± 0.97	0.122 ± 0.010
171	1.35 ± 0.78	25.78 ± 5.03	1.197 ± 0.175
172			0.936 ± 0.251
173	1.80 ± 0.83		1.087 ± 0.202
174	1.19 ± 0.14	43.78 ± 5.86	0.935 ± 0.144
176		20.22 ± 3.79	0.925 ± 0.155
177	2.24 ± 1.02		1.004 ± 0.100
178	1.63 ± 0.54	24.47 ± 4.46	0.781 ± 0.088

Res.	R_1 (1/s)	R_2 (1/s)	NOE
179	1.28 ± 0.51	32.24 ± 6.39	0.915 ± 0.057
180			1.015 ± 0.228
181	4.06 ± 1.20	25.21 ± 6.71	1.075 ± 0.216
182	1.41 ± 0.82	32.44 ± 6.31	0.984 ± 0.094
183	0.86 ± 0.87	25.42 ± 4.48	0.891 ± 0.117
184	0.81 ± 0.22	22.89 ± 3.99	0.614 ± 0.057
186	0.93 ± 0.79	27.11 ± 4.12	1.251 ± 0.185
187	1.56 ± 0.46		0.878 ± 0.087
188	1.14 ± 0.56	19.71 ± 2.96	0.761 ± 0.116
190	1.28 ± 0.51	32.24 ± 6.39	0.915 ± 0.057
191	0.98 ± 0.56	33.79 ± 5.29	0.764 ± 0.086
192			0.653 ± 0.163
193	1.48 ± 0.32	31.38 ± 3.07	0.843 ± 0.087
194	2.62 ± 0.70		0.948 ± 0.168
195	1.73 ± 0.40	29.23 ± 4.82	0.801 ± 0.074
196	1.55 ± 0.99	38.90 ± 5.65	0.768 ± 0.095
197	2.67 ± 0.35	30.78 ± 4.80	1.014 ± 0.105
198	0.98 ± 0.23	29.77 ± 4.05	0.892 ± 0.079
199	1.60 ± 0.77	28.49 ± 4.78	0.952 ± 0.127
200	1.31 ± 0.49	31.17 ± 2.69	0.645 ± 0.056
201	1.30 ± 0.47	30.83 ± 6.52	1.094 ± 0.205
202	1.57 ± 0.14	18.35 ± 5.72	0.675 ± 0.021
204	2.90 ± 0.30	24.93 ± 3.53	0.974 ± 0.139
205	1.74 ± 0.30	25.50 ± 7.71	0.825 ± 0.108
208	0.90 ± 0.43	40.53 ± 6.16	0.734 ± 0.084
209	2.84 ± 0.62		0.827 ± 0.182
210	2.61 ± 0.16		0.768 ± 0.150
211	2.53 ± 0.67		0.849 ± 0.095
212	3.90 ± 0.83	40.51 ± 6.02	1.000 ± 0.136
214	1.12 ± 0.69		0.521 ± 0.155
217	2.27 ± 0.92	29.88 ± 2.57	1.494 ± 0.218
218	1.45 ± 0.78	33.38 ± 4.56	0.800 ± 0.102
219	1.08 ± 0.50	31.53 ± 6.68	0.967 ± 0.096
220	1.31 ± 0.49	31.17 ± 2.69	0.645 ± 0.056
221	1.58 ± 0.56	32.20 ± 3.85	0.812 ± 0.106
222	3.12 ± 0.45		0.466 ± 0.053
223	1.81 ± 1.28	25.20 ± 4.93	0.961 ± 0.179
224			0.758 ± 0.162
225	4.98 ± 0.53		0.819 ± 0.160
226	2.16 ± 0.67	37.63 ± 5.54	0.672 ± 0.070
227	2.04 ± 0.80		0.642 ± 0.110
228		21.65 ± 3.95	1.051 ± 0.159

Res.	R_1 (1/s)	R_2 (1/s)	NOE
229	2.99 ± 0.47	51.03 ± 10.17	0.706 ± 0.087
230	2.87 ± 0.38	47.02 ± 3.32	0.877 ± 0.105
231	1.35 ± 0.78	30.56 ± 3.67	1.197 ± 0.175
232	2.86 ± 0.53	26.22 ± 8.18	0.573 ± 0.072
233	2.80 ± 0.50		0.541 ± 0.074
234	2.76 ± 0.44	41.05 ± 5.12	0.448 ± 0.069
235	1.72 ± 0.60	31.87 ± 4.05	0.750 ± 0.081
236	1.55 ± 0.84	36.55 ± 4.89	0.674 ± 0.067
237	1.31 ± 0.27	27.89 ± 4.11	0.504 ± 0.048
238	1.67 ± 0.65	26.74 ± 5.06	0.709 ± 0.076
239	2.31 ± 0.81	27.35 ± 2.49	0.698 ± 0.086
240	1.76 ± 0.46	39.32 ± 5.17	0.927 ± 0.114
241	1.90 ± 0.13	32.00 ± 2.46	0.773 ± 0.092
242	1.78 ± 0.93	26.25 ± 2.78	0.900 ± 0.154
243	2.72 ± 0.30	31.61 ± 1.55	0.681 ± 0.085
244	2.02 ± 1.38	22.85 ± 1.85	0.844 ± 0.199
245	1.39 ± 0.74	42.52 ± 4.09	0.779 ± 0.093
246	1.24 ± 0.80	39.22 ± 0.80	0.561 ± 0.085
247	2.19 ± 0.53	31.74 ± 2.52	0.543 ± 0.139
248		36.57 ± 2.36	0.988 ± 0.220
249	3.01 ± 1.02	26.35 ± 3.62	0.795 ± 0.108
250	1.30 ± 0.33	36.86 ± 3.69	0.967 ± 0.149
251	1.99 ± 0.91	26.89 ± 4.87	0.875 ± 0.145
252	2.10 ± 0.62	35.99 ± 3.06	1.280 ± 0.253
254	3.55 ± 1.43	42.87 ± 4.19	0.738 ± 0.131
255	1.89 ± 0.13		0.795 ± 0.041
256	1.72 ± 0.54	38.22 ± 3.25	0.598 ± 0.070
257	1.26 ± 0.33	33.51 ± 3.81	0.630 ± 0.083
258	1.17 ± 0.46	37.31 ± 4.69	0.720 ± 0.063
259	1.90 ± 0.22	24.19 ± 3.24	0.702 ± 0.040
260	2.78 ± 0.32	10.52 ± 0.95	0.271 ± 0.019
261	3.35 ± 0.43	12.73 ± 1.35	0.293 ± 0.013
262	2.68 ± 0.11	7.88 ± 0.73	0.207 ± 0.008
263	2.91 ± 0.17	5.53 ± 0.56	-0.157 ± 0.005
264	2.55 ± 0.13	7.04 ± 0.42	0.178 ± 0.006
266	1.35 ± 0.26	6.12 ± 0.26	-0.434 ± 0.062
267	1.59 ± 0.08	3.19 ± 0.24	-0.318 ± 0.037
270	1.61 ± 0.06	2.64 ± 0.16	-0.731 ± 0.010

4 CONCLUSIONS AND FUTURE DIRECTIONS

The ETS family of transcription factors has a so-called ETS domain on which site-specific DNA is bound. The ETS domains are structurally homologous but their primary sequences are divergent; for example, two ETS family members PU.1 and Ets-1 share only 30% sequence homology. PU.1 is an essential transcription factor and its main biological role is the development of hematopoietic stem cells (HSCs) in the immune system (15). PU.1 is a central transcriptional regulator of differentiation of HSCs into lymphocytes and myelocytes, B and T cell development, and maintenance of HSCs (16). This function spans from early to late stages of progression in a lineage- and cell type-specific manner; thus, it controls proliferation, terminal differentiation, and maintenance of HSCs (17). Therefore, PU.1 is a key transcriptional regulator within the hematopoietic system and plays critical roles in both the innate and adaptive immune systems by controlling cell differentiation.

The biological activity of PU.1 is primarily controlled by up- and down-regulation of its expression. However, because the metabolic half-life of PU.1 spans the entire cell cycle (~50 h) (103), downregulation of its expression alone is probably not sufficient for PU.1 regulation. Therefore, PU.1 activity during its lifetime in the cell needs to be regulated as well. Only a few inhibitory mechanisms are known for PU.1 other than down-regulated expression. The most understood one involves the nature of PU.1 itself forming a heterodimer with other protein partners, such as GATA-1 (29).

Autoinhibition is a regulatory mechanism of protein activity, whereby inhibitory module or domain of a protein interacts with another part of the protein so that it works for negative regulation (30). Protein expression is known to be regulated via autoinhibition. For example, autoinhibitory modules are removed by alternative splicing or proteolysis. Post-translational

modifications (PTM) or protein-protein interactions (PPI) in response to cellular signaling can relieve or reinforce autoinhibition and enable the protein to control downstream events (31). For ETS transcription factors, autoinhibition has been described as a key regulatory mechanism at the protein/DNA level (30). Autoinhibition in ETS family has a common mechanism in which autoinhibitory elements, typically helices (α- or 3_{10}-helix), adjacent to the ETS domain, make DNA binding unfavorable. Most of the 28 paralogs of ETS family in humans have been found to possess autoinhibition, while PU.1 is one of a few members that are not autoinhibited (31).

In addition to transcriptionally active 1:1 protein/DNA complex, homodimerization of many ETS family members, including Ets-1, Elk-1, ETV1, ETV6, FEV, ERG, and PU.1, has been reported (66-70). Interestingly, all these ETS domain homodimers are 2:2 protein/DNA complex except for a 2:1 protein/DNA complex of Ets-1 in a non-reducing environment, where two Cys residues from each subunit are likely to form a disulfide bond (71). For example, positively co-operative binding of Ets-1 at 2:2 protein/DNA stoichiometry is observed at repeated (palindromic) specific DNA sites such as stromelysin-1 promoter (72). Such a positively cooperative DNA binding of Ets-1 is known to counteract to its autoinhibition (73). Three homodimeric structures of Ets-1 have been determined by crystallization (PDB: 2NNY, 3MFK, and 3RI4) so far (72,74,75). In the case of Elk-1, homodimerization mediated by its ETS domain gives Elk-1 cytoplasmic stability to resist proteasomal degradation as well as localization to the nucleus (68). PU.1 in the nucleus forms DNA-free homodimer(s) but is monomeric upon specific DNA binding (68). Thus, PU.1 homodimerization mediated by the ETS domain is biologically relevant. This also raises a new question of whether PU.1 can dimerize in the presence of DNA.

Before the present study was undertaken, our group had observed the potential for the ETS domain of PU.1 to dimerize at a single specific site by ITC titration experiments (76). We observed two distinct DNA binding modes (1:1 and 2:1 protein:DNA ratio) for PU.1 protein. Also, the negative to positive transition in the reverse titration (namely, adding DNA to PU.1) of ITC implied that a 2:1 (PU.1:DNA) complex is formed in a negatively cooperative manner. The free energy of each of the four states of PU.1 (i.e. monomeric and dimeric PU.1 in the absence or presence of site-specific DNA) under standard state conditions revealed that the free energy (G°) gradient of the PU.1 ETS domain is described as unbound monomer > PU.1 dimer in the absence of DNA > 2:1 DNA-bound complex > 1:1 DNA-bound complex.

In the present study (*cf.* Chapter 2 of this dissertation), we established a DNA-binding model of PU.1 using diffusion-ordered NMR spectroscopy (DOSY). Namely, the ETS domain of PU.1 dimerizes at a *single* cognate site in a negatively cooperative manner, unlike its auto-inhibited family member Ets-1. We also detected a potential interface of DNA-bound PU.1 dimer by using heteronuclear single quantum correlation (HSQC) NMR. We detected four consecutive charged residues (namely, ^{193}DKDK^{196} in hPU.1) on the loop between β-sheets S1 and S2 at the potential interface of DNA-bound dimer of PU.1, by overlaying the HSQC spectra. To assess the effect of electrostatic interactions on PU.1 dimerization in complex with DNA, we generated ^{193}NINI^{196} mutant designed to remove the charges but maintain similar side-chain structures. This mutation abolished site-specific 2:1 PU.1/DNA binding. Furthermore, the DNA contact interface of the 2:1 PU.1/DNA complex made extended contacts with the DNA compared to the 1:1 complex. Taken together, the 2:1 PU.1/DNA complex is presumably a transcriptionally inactive form, and its formation is potentially a self-regulatory mechanism of PU.1 at the protein-DNA level, instead of auto-inhibition for other ETS proteins.

Expression levels of PU.1 in the cells are as high as those of housekeeping genes (133). Moreover, the estimated half-life of PU.1 in the cells is long enough as the lifespan (~50 hours) of the cells (*cf.* Chapter 2). Therefore, we proposed that excess PU.1 proteins are sequestered to form DNA-bound PU.1 dimers, as potentially a negative feedback mechanism for PU.1 activity (*cf.* Chapter 2). Our recent study demonstrated it using mammalian cells (138). We established a reporter gene system in HEK293 cells to measure PU.1 transactivation levels using tandem copies of specific ETS binding site (EBS) spaced by 20 bp. Since PU.1 is not expressed in HEK293 cells, the reporter is not activated unless we induce a plasmid encoding full-length PU.1 that yields EGFP fluorescence. The fluorescence signals of PU.1 transactivation levels showed a bell-shaped response to the dose of the PU.1 plasmid. This suggests that excess PU.1 was used for negative feedback in the cells. Therefore, this study demonstrated that the formation of a 2:1 PU.1/DNA complex using excess PU.1 is a self-regulatory mechanism.

Our recent study also revealed the roles of IDRs on DNA-free PU.1 homodimerization, which was previously observed *in vivo* by Evans et al. (68,138). We found the PU.1 dimers in the absence and presence of DNA antagonize to each other, and the IDRs flanking the ETS domain (i.e., N-terminal PEST domain and C-terminal 12 residues) play key roles for this phenomenon (138). We also found that the presence of IDR modifies DNA recognition by the ETS domain of PU.1. The N-terminal PEST domain, which is intrinsically disordered, increases the affinity of 1:1 DNA complex but reduces that of 2:1 PU.1/DNA complex (138). On the other hand, the absence of C-terminal IDR does not affect DNA recognition of the ETS domain, but PU.1 is unable to form the DNA-free dimer without it (138).

In the present study (*cf.* Chapter 3 of this dissertation), we tracked translational diffusion constants in a titration with 23- and 16-bp site-specific DNA by DOSY NMR in the same way as

described in Chapter 2. The result indicated that the PU.1 ETS domain retains its ability to form a DNA-bound homodimer in a negatively cooperative manner, even in the presence of the N-terminal IDR (the PEST domain) flanking the ETS domain (namely, hPU.1ΔN117) (*cf.* Fig. 3.1 *A*). This result is consistent with our observation for the PU.1 ETS domain in the absence of the PEST domain (namely, mPU.1ΔN167) with 16-bp specific DNA (*cf.* Fig. S 2.3). We also observed 23-bp specific DNA is long enough, but 16-bp is not, for ΔN117 to form the 1:1 and 2:1 complexes in direct response to the PU.1/DNA molar ratios, presumably because the three positively charged residues (K245-247) in the "wing" are structurally perturbed due to charge-charge repulsion with phosphate groups of DNA if the DNA is short (*cf.* Fig. 3.1 *C*).

We also tested whether or not the PEST domain changes the internal dynamics of the ETS domain using NMR to study how the PEST domain modifies DNA recognition of PU.1. We successfully assigned ~90% or more HSQC resonances of hPU.1sΔN165, ΔN165, and ΔN117, both in the absence and presence (i.e., 1:1 binding) of cognate DNA. The backbone assignments provided us with opportunities to perform measurements of ^{15}N relaxation parameters (namely, spin-lattice (R_1) and spin-spin (R_2) relaxation rate and the steady-state heteronuclear ^{1}H{^{15}N}-NOE) for both DNA-free and 1:1 DNA-bound PU.1 proteins. The chemical shifts of the assigned HSQC resonances and the spin relaxation measurements suggest that the PEST domain remains disordered but becomes more dynamic upon 1:1 specific DNA binding of PU.1. The chemical shifts of the assigned HSQC resonances also suggest that the PEST domain does not structurally perturb the ETS domain upon specific DNA binding. Using the ^{15}N relaxation parameters, we observed the presence of the PEST domain does not change the internal dynamics of the ETS domain upon 1:1 specific DNA binding. Taken together, we propose a role of the disordered PEST domain on 1:1 specific DNA binding of the PU.1 ETS domain. In the presence of DNA,

the highly cationic ETS domain bound with the basic PEST domain probably releases it and binds with DNA instead. The released PEST domain upon DNA binding of the ETS domain becomes more dynamic than in the DNA-free form. As a result, the presence of the PEST domain increases affinity in the 1:1 PU.1/DNA complex because the "free" PEST domain presumably enhances the ETS:DNA interaction.

I propose future directions and studies as described below, on the basis of PU.1 studies done in this dissertation. It is reasonable to assume that the interfaces in PPIs have been evolved to optimize their functional requirement (44). In particular, weak and nonspecific interactions at the interfaces have been adjusted properly to survive against selective pressure. In the ETS family, autoinhibition has been lost through evolution: namely, PU.1 is evolutionary the newest and not autoinhibited. Therefore, it would be interesting to study the relationship between quaternary structures (or properties in dimeric interfaces) and the efficiency of self-regulation (i.e., inhibition) through dimerization *in vivo* in the ETS family. If the optimization of the physics of association at the dimeric interface has evolved by selective pressure (namely, in inverse proportion to autoinhibition), self-regulation of the protein through dimerization has probably been achieved most effectively in PU.1 among the ETS family members.

We will be able to do NMR experiments further to characterize the DNA-free PU.1 dimer. In particular, relaxation dispersion NMR to detect motions in the intermediate (i.e., microsecond to millisecond) time scale will be very useful. A conformational exchange rate constant R_{ex} obtained by this experiment explains the chemical exchange processes that contribute to the decay in the transverse magnetization (R_2). Large R_{ex} values generally are derived from line broadening due to chemical (conformational) exchange processes between two states typically in the microsecond to millisecond time scale. Interpretation of the R_{ex} constant is

somewhat complicated because this parameter includes exchange rate, chemical shift difference, and fractional populations at the exchange sites. Nevertheless, large R_{ex} values suggest that dynamic exchange occurs at the corresponding residues in the microsecond to millisecond time scale. As discussed above, we have detected the PU.1 residues that reflect structural perturbation between two states (DNA-free monomer and dimer) using chemical shift perturbations (CSPs) (138). Also, as discussed in Chapter 3, we have detected the PU.1 residues that are dynamic in the fast (i.e., picosecond to nanosecond) time scale, using heteronuclear NOE in the absence of DNA: These PU.1 residues are presumably responsible for DNA-free dimer formation. Therefore, once R_{ex} values are obtained for a DNA-free PU.1 dimer sample, it helps us further characterize the dynamic properties of PU.1 and detect the residues that are responsible for the dimerization.

To further characterize the interface of the DNA-bound dimer of PU.1, we will be able to test ^{193}AAAA^{196} and ^{193}NANA^{196} mutant to study the effects of hydrophobicity in the corresponding region on the PPI in the 2:1 DNA-bound dimer. Even though the ^{193}NINI^{196} mutant was designed to abrogate charges but maintain similar side-chain structures and secondary structure propensities to WT protein, the ^{193}NINI^{196} surface is much more hydrophobic than that of WT since Ile is one of the most hydrophobic amino acids. Therefore, by using the above mutants, we can estimate the contribution from the hydrophobicity of the side chain (or hydrophobic interactions introduced by the Ile side chain) to affect the 2:1 binding. If the ^{193}NANA^{196} mutant does not abolish the 2:1 binding but the ^{193}AAAA^{196} mutant does, then we could estimate the effect of hydrophilicity of the Asp residues in this region for the 2:1 complex formation. We could subsequently test if hydrophilicity is more important than electrostatic interactions for this protein-protein interaction.

If we can detect other component(s) of the dimeric interface in the 2:1 PU.1/DNA complex besides the ^{193}DKDK^{196} site, then such a component would be very useful to characterize the DNA-bound PU.1 dimer. One of the ideas to test this is to generate "^{193}NINI^{196} + R230A/R233A" mutant and to conduct the same DNA binding experiment like the one using the R230A/R233A mutant (138), as described above. Thus, we are probably able to designate the binding order: WT PU.1 always binds to DNA and the mutant binds to the 1:1 PU.1/DNA complex. If the mutant forms a 2:1 complex without forming a 1:1 complex, then it provides us with proof that the presence of the other dimeric interface. Furthermore, capturing the 2:1 complex in an HSQC spectrum would also be very useful to characterize the DNA-bound PU.1 dimer. We will be able to reach this goal by making the 1:1 PU.1/DNA complex first using unlabeled PU.1 and subsequently by adding the equivalent amount of ^{15}N-labeled R230A/R233A mutant. Under the conditions, each monomeric components of the DNA-bound PU.1 dimer are not probably interchangeable because the mutant PU.1 cannot bind DNA. Therefore, the crosspeaks of the 2:1 complex will be visible, and therefore we should be able to see the crosspeaks of the 2:1 complex. If this experiment works, then we could plan to label WT PU.1 by ^{15}N and do the same experiment. Then, we will be able to do spin relaxation measurements. If we successfully obtain R_{ex} rates for PU.1 residues, we can detect the residues responsible for dimerization.

Moreover, we could try to determine the crystal structure of the 2:1 PU.1/DNA complex. We will need to use the PU.1 protein without IDRs (sΔN165) for the study because the presence of intrinsically disordered regions makes protein crystallization difficult and also does not facilitate the DNA-bound dimer formation. Our data show that the two monomers of PU.1 in this complex are interconvertible (*cf.* Fig. 2.3 *B*). Namely, PU.1 in the 2:1 complex is highly

dynamic, and crystallization may be difficult. If the 2:1 complex using WT PU.1 is not crystallized by any means, then the R230A/R233A mutant can be analyzed by NMR. If we mix WT PU.1, R230A/R233A mutant, and specific DNA at 1:1:1 molar ratio, then we will obtain a 2:1 PU.1/DNA complex. It is reasonable to assume that well-dispersed HSQC crosspeaks of PU.1 in the 2:1 complex will be obtained because the interconversion of the two PU.1 monomers is disturbed in this experiment, as discussed above.

REFERENCES

1. Krebs, J. E., Lewin, B., Kilpatrick, S. T., and Goldstein, E. S. (2014) *Lewin's genes XI*, 11th ed., Jones & Bartlett Learning, Burlington, Mass.
2. Rice, P. A., and Correll, C. C. (2008) *Protein-nucleic acid interactions: structural biology*, RSC Pub., Cambridge
3. Kuriyan, J., Konforti, B., and Wemmer, D. (2013) *The molecules of life : physical and chemical principles*, Garland Science, Taylor & Francis Group, New York ; London
4. Watson, D. K., Ascione, R., and Papas, T. S. (1990) Molecular analysis of the ets genes and their products. *Crit Rev Oncog* **1**, 409-436
5. Wang, Z., and Zhang, Q. (2009) Genome-wide identification and evolutionary analysis of the animal specific ETS transcription factor family. *Evol Bioinform Online* **5**, 119-131
6. Wei, G. H., Badis, G., Berger, M. F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A. R., Yan, J., Talukder, S., Turunen, M., Taipale, M., Stunnenberg, H. G., Ukkonen, E., Hughes, T. R., Bulyk, M. L., and Taipale, J. (2010) Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J* **29**, 2147-2160
7. Norman, C., Runswick, M., Pollock, R., and Treisman, R. (1988) Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* **55**, 989-1003
8. Price, M. A., Rogers, A. E., and Treisman, R. (1995) Comparative analysis of the ternary complex factors Elk-1, SAP-1a and SAP-2 (ERP/NET). *EMBO J* **14**, 2589-2601
9. Shore, P., and Sharrocks, A. D. (1994) The transcription factors Elk-1 and serum response factor interact by direct protein-protein contacts mediated by a short region of Elk-1. *Mol Cell Biol* **14**, 3283-3291
10. Park, S., Boder, E. T., and Saven, J. G. (2005) Modulating the DNA affinity of Elk-1 with computationally selected mutations. *J Mol Biol* **348**, 75-83
11. Garvie, C. W., Hagman, J., and Wolberger, C. (2001) Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell* **8**, 1267-1276
12. Sementchenko, V. I., and Watson, D. K. (2000) Ets target genes: past, present and future. *Oncogene* **19**, 6533-6548
13. Seth, A., and Watson, D. K. (2005) ETS transcription factors and their emerging roles in human cancer. *Eur J Cancer* **41**, 2462-2478
14. Moreau-Gachelin, F., Tavitian, A., and Tambourin, P. (1988) Spi-1 is a putative oncogene in virally induced murine erythroleukaemias. *Nature* **331**, 277-280
15. Oikawa, T., Yamada, T., Kihara-Negishi, F., Yamamoto, H., Kondoh, N., Hitomi, Y., and Hashimoto, Y. (1999) The role of Ets family transcription factor PU.1 in hematopoietic cell differentiation, proliferation and apoptosis. *Cell Death Differ* **6**, 599-608
16. Iwasaki, H., Mizuno, S., Arinobu, Y., Ozawa, H., Mori, Y., Shigematsu, H., Takatsu, K., Tenen, D. G., and Akashi, K. (2006) The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages. *Genes Dev* **20**, 3010-3021
17. Iwasaki, H., Somoza, C., Shigematsu, H., Duprez, E. A., Iwasaki-Arai, J., Mizuno, S., Arinobu, Y., Geary, K., Zhang, P., Dayaram, T., Fenyus, M. L., Elf, S., Chan, S., Kastner, P., Huettner, C. S., Murray, R., Tenen, D. G., and Akashi, K. (2005) Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. *Blood* **106**, 1590-1600

18. Turkistany, S. A., and DeKoter, R. P. (2011) The transcription factor PU.1 is a critical regulator of cellular communication in the immune system. *Arch Immunol Ther Exp (Warsz)* **59**, 431-440
19. Dozmorov, M. G., Wren, J. D., and Alarcon-Riquelme, M. E. (2014) Epigenomic elements enriched in the promoters of autoimmunity susceptibility genes. *Epigenetics* **9**, 276-285
20. Gjoneska, E., Pfenning, A. R., Mathys, H., Quon, G., Kundaje, A., Tsai, L. H., and Kellis, M. (2015) Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* **518**, 365-369
21. Bonadies, N., Pabst, T., and Mueller, B. U. (2010) Heterozygous deletion of the PU.1 locus in human AML. *Blood* **115**, 331-334
22. Vangala, R. K., Heiss-Neumann, M. S., Rangatia, J. S., Singh, S. M., Schoch, C., Tenen, D. G., Hiddemann, W., and Behre, G. (2003) The myeloid master regulator transcription factor PU.1 is inactivated by AML1-ETO in t(8;21) myeloid leukemia. *Blood* **101**, 270-277
23. Will, B., Vogler, T. O., Narayanagari, S., Bartholdy, B., Todorova, T. I., da Silva Ferreira, M., Chen, J., Yu, Y., Mayer, J., Barreyro, L., Carvajal, L., Neriah, D. B., Roth, M., van Oers, J., Schaetzlein, S., McMahon, C., Edelmann, W., Verma, A., and Steidl, U. (2015) Minimal PU.1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia. *Nat Med* **21**, 1172-1181
24. Tatetsu, H., Ueno, S., Hata, H., Yamada, Y., Takeya, M., Mitsuya, H., Tenen, D. G., and Okuno, Y. (2007) Down-regulation of PU.1 by methylation of distal regulatory elements and the promoter is required for myeloma cell growth. *Cancer Res* **67**, 5328-5336
25. Jundt, F., Kley, K., Anagnostopoulos, I., Schulze Probsting, K., Greiner, A., Mathas, S., Scheidereit, C., Wirth, T., Stein, H., and Dorken, B. (2002) Loss of PU.1 expression is associated with defective immunoglobulin transcription in Hodgkin and Reed-Sternberg cells of classical Hodgkin disease. *Blood* **99**, 3060-3062
26. Kodandapani, R., Pio, F., Ni, C. Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R. A., and Ely, K. R. (1996) A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. *Nature* **380**, 456-460
27. Poon, G. M. K., and Kim, H. M. (2017) Signatures of DNA target selectivity by ETS transcription factors. *Transcription* **8**, 193-203
28. Liang, H., Mao, X., Olejniczak, E. T., Nettesheim, D. G., Yu, L., Meadows, R. P., Thompson, C. B., and Fesik, S. W. (1994) Solution structure of the ets domain of Fli-1 when bound to DNA. *Nat Struct Biol* **1**, 871-875
29. Nerlov, C., Querfurth, E., Kulessa, H., and Graf, T. (2000) GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription. *Blood* **95**, 2543-2551
30. Hollenhorst, P. C., McIntosh, L. P., and Graves, B. J. (2011) Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu Rev Biochem* **80**, 437-471
31. Hollenhorst, P. C., McIntosh, L. P., and Graves, B. J. (2011) Genomic and biochemical insights into the specificity of ETS transcription factors. *Annual review of biochemistry* **80**, 437-471
32. Petersen, J. M., Skalicky, J. J., Donaldson, L. W., McIntosh, L. P., Alber, T., and Graves, B. J. (1995) Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an alpha helix. *Science* **269**, 1866-1869

33. Desjardins, G., Okon, M., Graves, B. J., and McIntosh, L. P. (2016) Conformational Dynamics and the Binding of Specific and Nonspecific DNA by the Autoinhibited Transcription Factor Ets-1. *Biochemistry* **55**, 4105-4118
34. Garvie, C. W., Pufall, M. A., Graves, B. J., and Wolberger, C. (2002) Structural analysis of the autoinhibition of Ets-1 and its role in protein partnerships. *J Biol Chem* **277**, 45529-45536
35. Lee, G. M., Donaldson, L. W., Pufall, M. A., Kang, H. S., Pot, I., Graves, B. J., and McIntosh, L. P. (2005) The structural and dynamic basis of Ets-1 DNA binding autoinhibition. *J Biol Chem* **280**, 7088-7099
36. Lee, G. M., Pufall, M. A., Meeker, C. A., Kang, H. S., Graves, B. J., and McIntosh, L. P. (2008) The affinity of Ets-1 for DNA is modulated by phosphorylation through transient interactions of an unstructured region. *J Mol Biol* **382**, 1014-1030
37. Green, S. M., Coyne, H. J., 3rd, McIntosh, L. P., and Graves, B. J. (2010) DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J Biol Chem* **285**, 18496-18504
38. Greenall, A., Willingham, N., Cheung, E., Boam, D. S., and Sharrocks, A. D. (2001) DNA binding by the ETS-domain transcription factor PEA3 is regulated by intramolecular and intermolecular protein.protein interactions. *J Biol Chem* **276**, 16207-16215
39. de Launoit, Y., Baert, J. L., Chotteau, A., Monte, D., Defossez, P. A., Coutte, L., Pelczar, H., and Leenders, F. (1997) Structure-function relationships of the PEA3 group of Ets-related transcription factors. *Biochem Mol Med* **61**, 127-135
40. Kopp, J. L., Wilder, P. J., Desler, M., Kinarsky, L., and Rizzino, A. (2007) Different domains of the transcription factor ELF3 are required in a promoter-specific manner and multiple domains control its binding to DNA. *J Biol Chem* **282**, 3027-3041
41. Oettgen, P., Kas, K., Dube, A., Gu, X., Grall, F., Thamrongsak, U., Akbarali, Y., Finger, E., Boltax, J., Endress, G., Munger, K., Kunsch, C., and Libermann, T. A. (1999) Characterization of ESE-2, a novel ESE-1-related Ets transcription factor that is restricted to glandular epithelium and differentiated keratinocytes. *J Biol Chem* **274**, 29439-29452
42. Dalton, S., and Treisman, R. (1992) Characterization of SAP-1, a protein recruited by serum response factor to the c-fos serum response element. *Cell* **68**, 597-612
43. Janknecht, R., and Nordheim, A. (1992) Elk-1 protein domains required for direct and SRF-assisted DNA-binding. *Nucleic Acids Res* **20**, 3317-3324
44. Sharrocks, A. D. (2002) Complexities in ETS-domain transcription factor function and regulation: lessons from the TCF (ternary complex factor) subfamily. The Colworth Medal Lecture. *Biochem Soc Trans* **30**, 1-9
45. Stinson, J., Inoue, T., Yates, P., Clancy, A., Norton, J. D., and Sharrocks, A. D. (2003) Regulation of TCF ETS-domain transcription factors by helix-loop-helix motifs. *Nucleic Acids Res* **31**, 4717-4728
46. Yang, S. H., Shore, P., Willingham, N., Lakey, J. H., and Sharrocks, A. D. (1999) The mechanism of phosphorylation-inducible activation of the ETS-domain transcription factor Elk-1. *EMBO J* **18**, 5666-5674
47. Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, 386-389

48. Kalodimos, C. G., Boelens, R., and Kaptein, R. (2004) Toward an integrated model of protein-DNA recognition as inferred from NMR studies on the Lac repressor system. *Chem Rev* **104**, 3567-3586
49. Nooren, I. M., and Thornton, J. M. (2003) Diversity of protein-protein interactions. *EMBO J* **22**, 3486-3492
50. De, S., Krishnadev, O., Srinivasan, N., and Rekha, N. (2005) Interaction preferences across protein-protein interfaces of obligatory and non-obligatory components are different. *BMC Struct Biol* **5**, 15
51. Ansari, S., and Helms, V. (2005) Statistical analysis of predominantly transient protein-protein interfaces. *Proteins* **61**, 344-355
52. Matthews, J. M. (2012) Protein dimerization and oligomerization in biology. in *Advances in experimental medicine and biology v 747*, Springer Science+Business Media; Landes Bioscience, New York, New York; Austin, Texas
53. Beernink, P. T., and Tolan, D. R. (1996) Disruption of the aldolase A tetramer into catalytically active monomers. *Proc Natl Acad Sci U S A* **93**, 5374-5379
54. Traut, T. W. (1994) Dissociation of enzyme oligomers: a mechanism for allosteric regulation. *Crit Rev Biochem Mol Biol* **29**, 125-163
55. Marianayagam, N. J., Sunde, M., and Matthews, J. M. (2004) The power of two: protein dimerization in biology. *Trends Biochem Sci* **29**, 618-625
56. Latchman, D. S. (2008) *Eukaryotic transcription factors*, 5th ed., Elsevier/Academic Press, Amsterdam; Boston
57. Ravasi, T., Suzuki, H., Cannistraci, C. V., Katayama, S., Bajic, V. B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., Carninci, P., Daub, C. O., Forrest, A. R., Gough, J., Grimmond, S., Han, J. H., Hashimoto, T., Hide, W., Hofmann, O., Kamburov, A., Kaur, M., Kawaji, H., Kubosaki, A., Lassmann, T., van Nimwegen, E., MacPherson, C. R., Ogawa, C., Radovanovic, A., Schwartz, A., Teasdale, R. D., Tegner, J., Lenhard, B., Teichmann, S. A., Arakawa, T., Ninomiya, N., Murakami, K., Tagami, M., Fukuda, S., Imamura, K., Kai, C., Ishihara, R., Kitazume, Y., Kawai, J., Hume, D. A., Ideker, T., and Hayashizaki, Y. (2010) An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140**, 744-752
58. Batchelor, A. H., Piper, D. E., de la Brousse, F. C., McKnight, S. L., and Wolberger, C. (1998) The structure of GABPalpha/beta: an ETS domain- ankyrin repeat heterodimer bound to DNA. *Science* **279**, 1037-1041
59. Mo, Y., Ho, W., Johnston, K., and Marmorstein, R. (2001) Crystal structure of a ternary SAP-1/SRF/c-fos SRE DNA complex. *J Mol Biol* **314**, 495-506
60. Escalante, C. R., Brass, A. L., Pongubala, J. M., Shatova, E., Shen, L., Singh, H., and Aggarwal, A. K. (2002) Crystal structure of PU.1/IRF-4/DNA ternary complex. *Mol Cell* **10**, 1097-1105
61. Gupta, P., Gurudutta, G. U., Saluja, D., and Tripathi, R. P. (2009) PU.1 and partners: regulation of haematopoietic stem cell fate in normal and malignant haematopoiesis. *J Cell Mol Med* **13**, 4349-4363
62. Pongubala, J. M., Van Beveren, C., Nagulapalli, S., Klemsz, M. J., McKercher, S. R., Maki, R. A., and Atchison, M. L. (1993) Effect of PU.1 phosphorylation on interaction with NF-EM5 and transcriptional activation. *Science* **259**, 1622-1625

63. Zhang, P., Behre, G., Pan, J., Iwama, A., Wara-Aswapati, N., Radomska, H. S., Auron, P. E., Tenen, D. G., and Sun, Z. (1999) Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. *Proc Natl Acad Sci U S A* **96**, 8705-8710
64. Verger, A., Buisine, E., Carrere, S., Wintjens, R., Flourens, A., Coll, J., Stehelin, D., and Duterque-Coquillaud, M. (2001) Identification of amino acid residues in the ETS transcription factor Erg that mediate Erg-Jun/Fos-DNA ternary complex formation. *J Biol Chem* **276**, 17181-17189
65. Yang, Z., Wara-Aswapati, N., Chen, C., Tsukada, J., and Auron, P. E. (2000) NF-IL6 (C/EBPbeta) vigorously activates il1b gene expression via a Spi-1 (PU.1) protein-protein tether. *J Biol Chem* **275**, 21272-21277
66. Carrere, S., Verger, A., Flourens, A., Stehelin, D., and Duterque-Coquillaud, M. (1998) Erg proteins, transcription factors of the Ets family, form homo, heterodimers and ternary complexes via two distinct domains. *Oncogene* **16**, 3261-3268
67. Drewett, V., Muller, S., Goodall, J., and Shaw, P. E. (2000) Dimer formation by ternary complex factor ELK-1. *J Biol Chem* **275**, 1757-1762
68. Evans, E. L., Saxton, J., Shelton, S. J., Begitt, A., Holliday, N. D., Hipskind, R. A., and Shaw, P. E. (2011) Dimer formation and conformational flexibility ensure cytoplasmic stability and nuclear accumulation of Elk-1. *Nucleic Acids Res* **39**, 6390-6402
69. Jousset, C., Carron, C., Boureux, A., Quang, C. T., Oury, C., Dusanter-Fourt, I., Charon, M., Levin, J., Bernard, O., and Ghysdael, J. (1997) A domain of TEL conserved in a subset of ETS proteins defines a specific oligomerization interface essential to the mitogenic properties of the TEL-PDGFR beta oncoprotein. *EMBO J* **16**, 69-82
70. Buchwalter, G., Gross, C., and Wasylyk, B. (2004) Ets ternary complex transcription factors. *Gene* **324**, 1-14
71. Samorodnitsky, D., Szyjka, C., and Koudelka, G. B. (2015) A Role for Autoinhibition in Preventing Dimerization of the Transcription Factor ETS1. *J Biol Chem* **290**, 22101-22110
72. Babayeva, N. D., Wilder, P. J., Shiina, M., Mino, K., Desler, M., Ogata, K., Rizzino, A., and Tahirov, T. H. (2010) Structural basis of Ets1 cooperative binding to palindromic sequences on stromelysin-1 promoter DNA. *Cell Cycle* **9**, 3054-3062
73. Goetz, T. L., Gu, T. L., Speck, N. A., and Graves, B. J. (2000) Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. *Mol Cell Biol* **20**, 81-90
74. Lamber, E. P., Vanhille, L., Textor, L. C., Kachalova, G. S., Sieweke, M. H., and Wilmanns, M. (2008) Regulation of the transcription factor Ets-1 by DNA-mediated homo-dimerization. *EMBO J* **27**, 2006-2017
75. Babayeva, N. D., Baranovskaya, O. I., and Tahirov, T. H. (2012) Structural basis of Ets1 cooperative binding to widely separated sites on promoter DNA. *PLoS One* **7**, e33698
76. Poon, G. M. (2012) DNA binding regulates the self-association of the ETS domain of PU.1 in a sequence-dependent manner. *Biochemistry* **51**, 4096-4107
77. Wright, P. E., and Dyson, H. J. (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* **16**, 18-29
78. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197-208
79. Marsh, J. A., Teichmann, S. A., and Forman-Kay, J. D. (2012) Probing the diverse landscape of protein flexibility and binding. *Curr Opin Struct Biol* **22**, 643-650

80. Mittag, T., and Forman-Kay, J. D. (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* **17**, 3-14
81. van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., and Babu, M. M. (2014) Classification of intrinsically disordered regions and proteins. *Chem Rev* **114**, 6589-6631
82. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **293**, 321-331
83. Tsafou, K., Tiwari, P. B., Forman-Kay, J. D., Metallo, S. J., and Toretsky, J. A. (2018) Targeting Intrinsically Disordered Transcription Factors: Changing the Paradigm. *J Mol Biol* **430**, 2321-2341
84. Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* **37**, 215-246
85. Gong, H., Yang, X., Zhao, Y., Petersen, R. B., Liu, X., Liu, Y., and Huang, K. (2015) Amyloidogenicity of p53: a hidden link between protein misfolding and cancer. *Curr Protein Pept Sci* **16**, 135-146
86. Kühne, T. (2011) Ewing Sarcoma Family of Tumors. *Pediatric Oncology*, 161-165
87. Mahner, S., Baasch, C., Schwarz, J., Hein, S., Wolber, L., Janicke, F., and Milde-Langosch, K. (2008) C-Fos expression is a molecular predictor of progression and survival in epithelial ovarian carcinoma. *Br J Cancer* **99**, 1269-1275
88. Panza, F., Seripa, D., Solfrizzi, V., Imbimbo, B. P., Santamato, A., Lozupone, M., Capozzo, R., Prete, C., Pilotto, A., Greco, A., and Logroscino, G. (2016) Tau aggregation inhibitors: the future of Alzheimer's pharmacotherapy? *Expert Opin Pharmacother* **17**, 457-461
89. Takeuchi, T., and Nagai, Y. (2017) Protein Misfolding and Aggregation as a Therapeutic Target for Polyglutamine Diseases. *Brain Sci* **7**
90. Sweeney, P., Park, H., Baumann, M., Dunlop, J., Frydman, J., Kopito, R., McCampbell, A., Leblanc, G., Venkateswaran, A., Nurmi, A., and Hodgson, R. (2017) Protein misfolding in neurodegenerative diseases: implications and strategies. *Transl Neurodegener* **6**, 6
91. Pattison, J. S., and Robbins, J. (2008) Protein misfolding and cardiac disease: establishing cause and effect. *Autophagy* **4**, 821-823
92. Hong, F., Si, C., Gao, P., Cederbaum, A. I., Xiong, H., and Lu, Y. (2016) The role of CYP2A5 in liver injury and fibrosis: chemical-specific difference. *Naunyn Schmiedebergs Arch Pharmacol* **389**, 33-43
93. Citron, B. A., Dennis, J. S., Zeitlin, R. S., and Echeverria, V. (2008) Transcription factor Sp1 dysregulation in Alzheimer's disease. *J Neurosci Res* **86**, 2499-2504
94. Afzal, A., Sarfraz, M., Li, G. L., Ji, S. P., Duan, S. F., Khan, N. H., Wu, D. D., and Ji, X. Y. (2019) Taking a holistic view of PEST-containing nuclear protein (PCNP) in cancer biology. *Cancer Med* **8**, 6335-6343
95. Whiteside, S. T., Ernst, M. K., LeBail, O., Laurent-Winter, C., Rice, N., and Israel, A. (1995) N- and C-terminal sequences control degradation of MAD3/I kappa B alpha in response to inducers of NF-kappa B activity. *Mol Cell Biol* **15**, 5339-5345
96. Rechsteiner, M., and Rogers, S. W. (1996) PEST sequences and regulation by proteolysis. *Trends Biochem Sci* **21**, 267-271

97. Kosol, S., Contreras-Martos, S., Cedeno, C., and Tompa, P. (2013) Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* **18**, 10802-10828
98. Oldfield, C. J., Cheng, Y., Cortese, M. S., Romero, P., Uversky, V. N., and Dunker, A. K. (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* **44**, 12454-12470
99. 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
100. Mak, K. S., Funnell, A. P., Pearson, R. C., and Crossley, M. (2011) PU.1 and Haematopoietic Cell Fate: Dosage Matters. *Int J Cell Biol* **2011**, 808524
101. Rosenbauer, F., Wagner, K., Kutok, J. L., Iwasaki, H., Le Beau, M. M., Okuno, Y., Akashi, K., Fiering, S., and Tenen, D. G. (2004) Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat Genet* **36**, 624-630
102. Gjoneska, E., Pfenning, A. R., Mathys, H., Quon, G., Kundaje, A., Tsai, L.-H., and Kellis, M. (2015) Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* **518**, 365-369
103. Kueh, H. Y., Champhekar, A., Nutt, S. L., Elowitz, M. B., and Rothenberg, E. V. (2013) Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science* **341**, 670-673
104. Pufall, M. A., and Graves, B. J. (2002) Autoinhibitory domains: modular effectors of cellular regulation. *Annu Rev Cell Dev Biol* **18**, 421-462
105. Fitzsimmons, D., Lukin, K., Lutz, R., Garvie, C. W., Wolberger, C., and Hagman, J. (2009) Highly cooperative recruitment of Ets-1 and release of autoinhibition by Pax5. *J Mol Biol* **392**, 452-464
106. Baillat, D., Begue, A., Stehelin, D., and Aumercier, M. (2002) ETS-1 transcription factor binds cooperatively to the palindromic head to head ETS-binding sites of the stromelysin-1 promoter by counteracting autoinhibition. *The Journal of biological chemistry* **277**, 29386-29398
107. Hou, C., and Tsodikov, O. V. (2015) Structural Basis for Dimerization and DNA Binding of Transcription Factor FLI1. *Biochemistry* **54**, 7365-7374
108. Mimeault, M. (2000) Structure-function studies of ETS transcription factors. *Crit Rev Oncog* **11**, 227-253
109. Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K. R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., Palin, K., Vaquerizas, J. M., Vincentelli, R., Luscombe, N. M., Hughes, T. R., Lemaire, P., Ukkonen, E., Kivioja, T., and Taipale, J. (2013) DNA-binding specificities of human transcription factors. *Cell* **152**, 327-339
110. Wang, S., Linde, M. H., Munde, M., Carvalho, V. D., Wilson, W. D., and Poon, G. M. (2014) Mechanistic heterogeneity in site recognition by the structurally homologous DNA-binding domains of the ETS family transcription factors Ets-1 and PU.1. *J Biol Chem* **289**, 21605-21616
111. 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

112. Stephens, D. C., Kim, H. M., Kumar, A., Farahat, A. A., Boykin, D. W., and Poon, G. M. K. (2016) Pharmacologic efficacy of PU.1 inhibition by heterocyclic dications: a mechanistic analysis. *Nucleic acids research* **44**, 4005-4013
113. Xhani, S., Esaki, S., Huang, K., Erlitzki, N., and Poon, G. M. (2017) Distinct Roles for Interfacial Hydration in Site-Specific DNA Recognition by ETS-Family Transcription Factors. *J Phys Chem B* **121**, 2748-2758
114. Poon, G. M. (2012) Sequence discrimination by DNA-binding domain of ETS family transcription factor PU.1 is linked to specific hydration of protein-DNA interface. *The Journal of biological chemistry* **287**, 18297-18307
115. Ortega, A., Amoros, D., and Garcia de la Torre, J. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J* **101**, 892-898
116. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic acids research* **32**, W665-667
117. Poon, G. M. (2010) Explicit formulation of titration models for isothermal titration calorimetry. *Analytical biochemistry* **400**, 229-236
118. Jia, X., Lee, L. K., Light, J., Palmer, A. G., 3rd, and Assa-Munt, N. (1999) Backbone dynamics of a short PU.1 ETS domain. *J Mol Biol* **292**, 1083-1093
119. Bains, G., and Freire, E. (1991) Calorimetric determination of cooperative interactions in high affinity binding processes. *Analytical biochemistry* **192**, 203-206
120. He, G., Tolic, A., Bashkin, J. K., and Poon, G. M. (2015) Heterogeneous dynamics in DNA site discrimination by the structurally homologous DNA-binding domains of ETS-family transcription factors. *Nucleic acids research* **43**, 4322-4331
121. Cann, J. R. (1996) Theory and practice of gel electrophoresis of interacting macromolecules. *Analytical biochemistry* **237**, 1-16
122. Altieri, A. S., Hinton, D. P., and Byrd, R. A. (1995) Association of Biomolecular Systems Via Pulsed-Field Gradient Nmr Self-Diffusion Measurements. *J Am Chem Soc* **117**, 7566-7567
123. Matulis, D., and Lovrien, R. (1998) 1-Anilino-8-naphthalene sulfonate anion-protein binding depends primarily on ion pair formation. *Biophys J* **74**, 422-429
124. Costantini, S., Colonna, G., and Facchiano, A. M. (2006) Amino acid propensities for secondary structures are influenced by the protein structural class. *Biochem Biophys Res Commun* **342**, 441-451
125. Groß, P., Arrowsmith, C. H., and Macgregor, R. B., Jr. (1998) Hydroxyl radical footprinting of DNA complexes of the ets domain of PU.1 and its comparison to the crystal structure. *Biochemistry* **37**, 5129-5135
126. Groß, P., Yee, A. A., Arrowsmith, C. H., and Macgregor, R. B., Jr. (1998) Quantitative hydroxyl radical footprinting reveals cooperative interactions between DNA-binding subdomains of PU.1 and IRF4. *Biochemistry* **37**, 9802-9811
127. Siggers, T., and Gordan, R. (2014) Protein-DNA binding: complexities and multi-protein codes. *Nucleic acids research* **42**, 2099-2111
128. Jin, L., Yang, J., and Carey, J. (1993) Thermodynamics of ligand binding to trp repressor. *Biochemistry* **32**, 7302-7309

129. Rippin, T. M., Freund, S. M., Veprintsev, D. B., and Fersht, A. R. (2002) Recognition of DNA by p53 core domain and location of intermolecular contacts of cooperative binding. *J Mol Biol* **319**, 351-358
130. Pió, F., Ni, C. Z., Mitchell, R. S., Knight, J., McKercher, S., Klemsz, M., Lombardo, A., Maki, R. A., and Ely, K. R. (1995) Co-crystallization of an ETS domain (PU.1) in complex with DNA. Engineering the length of both protein and oligonucleotide. *The Journal of biological chemistry* **270**, 24258-24263
131. Warren, L., Bryder, D., Weissman, I. L., and Quake, S. R. (2006) Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proc Natl Acad Sci U S A* **103**, 17807-17812
132. Zhang, G., Zhou, B., Li, S., Yue, J., Yang, H., Wen, Y., Zhan, S., Wang, W., Liao, M., Zhang, M., Zeng, G., Feng, C. G., Sassetti, C. M., and Chen, X. (2014) Allele-specific induction of IL-1beta expression by C/EBPbeta and PU.1 contributes to increased tuberculosis susceptibility. *PLoS Pathog* **10**, e1004426
133. Imoto, A., Okada, M., Okazaki, T., Kitasato, H., Harigae, H., and Takahashi, S. (2010) Metallothionein-1 isoforms and vimentin are direct PU.1 downstream target genes in leukemia cells. *The Journal of biological chemistry* **285**, 10300-10309
134. Liew, C. W., Rand, K. D., Simpson, R. J. Y., Yung, W. W., Mansfield, R. E., Crossley, M., Proetorius-Ibba, M., Nerlov, C., Poulsen, F. M., and Mackay, J. P. (2006) Molecular Analysis of the Interaction between the Hematopoietic Master Transcription Factors GATA-1 and PU.1. *The Journal of biological chemistry* **281**, 28296-28306
135. Arinobu, Y., Mizuno, S., Chong, Y., Shigematsu, H., Iino, T., Iwasaki, H., Graf, T., Mayfield, R., Chan, S., Kastner, P., and Akashi, K. (2007) Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. *Cell Stem Cell* **1**, 416-427
136. Poon, G. M., and Macgregor, R. B., Jr. (2003) Base coupling in sequence-specific site recognition by the ETS domain of murine PU.1. *J Mol Biol* **328**, 805-819
137. Clore, G., Kimber, BJ, Gronenborn, AM. (1983) The 1-1 Hard Pulse: A Simple and Effective Method of Water Resonance Suppression in FT ^{1}H NMR. *J. Magn. Reson.* **54**, 170-173
138. 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., Germann, M. W., Poon, G. M. K. (2020) Intrinsic disorder controls two functionally distinct dimers of the master transcription factor PU.1. *Sci Adv.* **In press**
139. Xin, J., Larry, K. L., James, L., Arthur, G. P., and Nuria, A.-M. (1999) Backbone dynamics of a short PU.1 ETS domain. *Journal of Molecular Biology* **292**, 1083-1093
140. Lee, W., Tonelli, M., and Markley, J. L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325-1327
141. Pufall, M. A., Lee, G. M., Nelson, M. L., Kang, H. S., Velyvis, A., Kay, L. E., McIntosh, L. P., and Graves, B. J. (2005) Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* **309**, 142-145
142. Cowley, D. O., and Graves, B. J. (2000) Phosphorylation represses Ets-1 DNA binding by reinforcing autoinhibition. *Genes Dev* **14**, 366-376
143. Liu, H., and Grundstrom, T. (2002) Calcium regulation of GM-CSF by calmodulin-dependent kinase II phosphorylation of Ets1. *Mol Biol Cell* **13**, 4497-4507

144. Perkel, J. M., and Atchison, M. L. (1998) A two-step mechanism for recruitment of Pip by PU.1. *J Immunol* **160**, 241-252
145. Lodie, T. A., Savedra, R., Jr., Golenbock, D. T., Van Beveren, C. P., Maki, R. A., and Fenton, M. J. (1997) Stimulation of macrophages by lipopolysaccharide alters the phosphorylation state, conformation, and function of PU.1 via activation of casein kinase II. *J Immunol* **158**, 1848-1856
146. Rieske, P., and Pongubala, J. M. (2001) AKT induces transcriptional activity of PU.1 through phosphorylation-mediated modifications within its transactivation domain. *J Biol Chem* **276**, 8460-8468
147. Azim, A. C., Wang, X., Park, G. Y., Sadikot, R. T., Cao, H., Mathew, B., Atchison, M., van Breemen, R. B., Joo, M., and Christman, J. W. (2007) NF-kappaB-inducing kinase regulates cyclooxygenase 2 gene expression in macrophages by phosphorylation of PU.1. *J Immunol* **179**, 7868-7875
148. Carey, J. O., Posekany, K. J., deVente, J. E., Pettit, G. R., and Ways, D. K. (1996) Phorbol ester-stimulated phosphorylation of PU.1: association with leukemic cell growth inhibition. *Blood* **87**, 4316-4324
149. Hamdorf, M., Berger, A., Schule, S., Reinhardt, J., and Flory, E. (2011) PKCdelta-induced PU.1 phosphorylation promotes hematopoietic stem cell differentiation to dendritic cells. *Stem Cells* **29**, 297-306
150. Perez-Borrajero, C., Lin, C. S., 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
151. Stephens, D. C., and Poon, G. M. (2016) Differential sensitivity to methylated DNA by ETS-family transcription factors is intrinsically encoded in their DNA-binding domains. *Nucleic Acids Res* **44**, 8671-8681
152. Rodier, F., Bahadur, R. P., Chakrabarti, P., and Janin, J. (2005) Hydration of protein-protein interfaces. *Proteins* **60**, 36-45
153. Monecke, P., Borosch, T., Brickmann, J., and Kast, S. M. (2006) Determination of the interfacial water content in protein-protein complexes from free energy simulations. *Biophys J* **90**, 841-850
154. Janin, J. (1999) Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. *Structure* **7**, R277-279
155. England, P., Bregegere, F., and Bedouelle, H. (1997) Energetic and kinetic contributions of contact residues of antibody D1.3 in the interaction with lysozyme. *Biochemistry* **36**, 164-172

APPENDICES

Appendix A: Introduction

PU.1 expression levels in the cells are as high as those of housekeeping genes (133), and the estimated PU.1 half-life in the cells is as long as the lifespan of the cells (~50 hours) (*cf.* Chapter 2). Thus, excess DNA-free PU.1 should be sequestered by forming a presumably inactive 2:1 PU.1/DNA complex, as a negative feedback mechanism for PU.1. Our mammalian cell study recently provided us with proof that excess PU.1 was used for negative feedback in the cells (138). Therefore, our recent study demonstrated that the formation of 2:1 PU.1/DNA complex with excess PU.1 is by a negative feedback mechanism for PU.1.

As we directly demonstrated using the diffusion coefficients of PU.1 by titration with site-specific DNA (> 10-bp), the DNA-bound PU.1 dimer is not a 2:2 complex (Fig. S 2.3). The 2:2 ETS/DNA complex is often seen in the ETS family (66-70). This finding also suggests that the DNA-bound dimer of PU.1 ETS domain is not formed by simple electrostatic interactions between two PU.1 molecules using the charges of ^{193}DKDK^{196} side-chains. If the ^{193}DKDK^{196} sites of two PU.1 molecules interact to form a DNA-bound dimer, a 2:2 complex should eventually be formed because the DNA-binding surface (namely, the surface opposite to the ^{193}DKDK^{196} site) is available for both monomers of PU.1. Thus, the DNA-bound dimer of PU.1 ETS domain is asymmetric (i.e., not a “head-to-head” binding), and the dimeric interface consists of the ^{193}DKDK^{196} site and another site of the ETS domain. A mutant study in our recent report confirmed the asymmetric configuration of the DNA-bound PU.1 dimer (138). We mutated two Arg residues to Ala (R230A/R233A) in the DNA-recognition helix H3 and confirmed this mutant alone does not bind to site-specific DNA. In the presence of a negligible concentration of WT PU.1, this mutant PU.1 only formed a 2:1 PU.1/DNA complex upon

specific DNA binding, at similar binding affinity to WT protein. In this assay, the sub-saturating concentration of WT PU.1 was used to form a 1:1 WT PU.1/DNA complex, where the mutant PU.1 protein was added to form a heterodimeric PU.1 dimer in complex with DNA.

Moreover, nonspecific DNA binding by PU.1 ETS domain is quite different from specific binding (*cf*. supplemental Fig. S2.3 *C*). In a DOSY titration of PU.1 with 16-bp nonspecific DNA, only an inflection point was observed at a DNA/PU.1 ratio of ~0.5, and it had a stable diffusion coefficient at a higher molar ratio. Thus, nonspecific DNA binding of PU.1 yields only a DNA-bound dimer, which is distinct from the DNA-bound dimer with specific DNA, judging from the diffusion coefficients. In the present study, we further characterized the nonspecific DNA binding of PU.1. The results provide us with a comparison of PU.1 with Ets-1.

Appendix B: Materials and methods

Proteins. DNA fragments of mutant PU.1 ETS domain (i.e., mPU.1ΔN167) were obtained by PCR amplification and subcloned directly into the NcoI/HindIII sites of pET28b vector. All the constructs were verified by Sanger sequencing (Macrogen). The protein samples (namely, wildtype hPU.1ΔN165, mutant mPU.1ΔN167, and Ets-1ΔN280) were expressed and purified in the same way as described in Chapters 2 and 3.

Nucleic acids. 16-bp nonspecific DNA (5'-GCAAGCGAGAGTGAGC- 3') was purchased from Integrated DNA Technologies as synthetic DNA oligos and annealed as described in Chapter 2. Fluorescent DNA probes were constructed by annealing a Cy3-labeled oligo with excess unlabeled complementary strand, as described in Chapter 2.

NMR spectroscopy. Uniformly ^{15}N-labeled mPU.1ΔN167 (0.75 mM) or Ets-1ΔN280 (~0.3 mM) was dialyzed with 16-bp nonspecific DNA (~2 mM) in separate dialysis tubings against 22 mM MES, pH 6.5, 55 mM NaCl, 0.5 mM EDTA, 0.02% NaN_3, and 5 mM DTT, and

D_2O was added at a final concentration of 10%. Note that the DNA was titrated into protein to achieve the desired DNA/protein ratios. ^{1}H-^{15}N correlated measurements were made using a phase-sensitive, double inept transfer with a GARP decoupling sequence, and solvent suppression (hsqcf3gpph19). Spectra were acquired with 1024 × 144 data points and zero-filled to 4096 × 4096.

Fluorescence polarization titrations. DNA binding experiments by fluorescence anisotropy measurements of a Cy3-labeled DNA probe were performed as described (111,113,151). Briefly, graded concentrations of WT hPU.1ΔN165 or mutants were incubated to equilibrium with a Cy3-labeled 23-bp DNA duplex oligo harboring the high-affinity PU.1 target site 5'-AGCGGAAGTG-3'. The binding assay samples were made in 10 mM Tris-HCl (pH 7.4) buffer with 0.15 M total [Na^+], and 0.1 mg/mL BSA.

Circular dichroism spectroscopy. Purified PU.1 proteins at graded concentrations were scanned for far-UV (200 to 250 nm) spectra in 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.4), 50 mM total [Na^+] at 25°C using a Jasco J-810 instrument.

Appendix C: Results and discussion

Appendix C.1: Asymmetric configuration of the PU.1 dimer in the presence of DNA

In Chapter 2, we observed ~80% of the HSQC crosspeaks of the DNA-bound PU.1 dimer disappeared (*cf.* Fig. 2.3 *B*). We presumed that it was due to the conformational exchange of PU.1, but not due to disorder. To confirm it, we assigned the ^{1}H-^{15}N resonances of PU.1 ETS domain in complex with nonspecific DNA because most (>90%) of the crosspeaks also disappeared in a titration of mPU.1ΔN167 protein with 16-bp nonspecific DNA at protein:DNA molar ratio of 1:0.5 and more (Appendix Fig. 1 *A*). By overlaying the HSQC spectrum at 1:0.5 molar ratio with that of fully-assigned unbound PU.1 resonances shown in Chapter 3 (*cf.* Fig.

3.3), we were able to assign the remaining resonances. The assigned peaks were only in the N- (before K171) and C-terminal (after G262) loops (Appendix Fig. 1 *B*-*C*). Thus, all the PU.1 ETS domain residues in the structured region disappeared upon nonspecific DNA binding, while the residues in the intrinsically disordered regions did not. This indicates that the HSQC peak disappearance of PU.1 upon DNA binding is due to conformational exchange, but not disorder of the protein. This confirmed our idea that chemical exchange is the reason for the HSQC crosspeak disappearance in the 2:1 complex of PU.1 with site-specific DNA (*cf.* Fig. 2.3 *B*).

To further characterize the DNA-bound dimer that was studied in Chapter2, secondary structures of PU.1 in the DNA-bound dimer were examined using CD spectroscopy. Negligible changes in the secondary structure content were observed in the titration of PU.1 ETS domain with cognate DNA (Appendix Fig. 2). DNA-free (PDB: 5W3G) and 1:1 DNA-bound (PDB: 1PUE) PU.1 structures are similar, as seen by solved structures (*cf.* Fig. 1.5). Thus, the CD data above indicate that the DNA-bound PU.1 dimer consists of a DNA-free and a 1:1 DNA-bound monomer. Namely, the structures of the PU.1 subunits (i.e., PU.1 monomers) in the DNA-bound dimer are probably similar to each of the solved structures (i.e., PDB: 5W3G and 1PUE). Furthermore, each subunit of the DNA-bound PU.1 dimer presumably interconverts with each other. Therefore, this data provided us with proof that the disappeared HSQC crosspeaks of the DNA-bound dimer (*cf.* Fig. 2.3 *B*) are due to conformational exchange between the two states of PU.1.

In sharp contrast, the CD spectra of the DNA-free PU.1 dimer exhibits completely different secondary structures content due to a great contribution by random coils, compared to the DNA-bound dimer (138). The DNA-bound PU.1 dimer is therefore conformationally distinct from the DNA-free dimer. Our group also found that the DNA-bound dimer is asymmetric, as

described above (138). Thus, we tested whether the DNA-free dimer is symmetric or not. We generated a mutant that is suitable to discuss this, namely the ^{193}DKCDK$^{197'}$ mutant. This mutant is obligated to form a dimer of head-to-head ETS domains using the crosslink between the inserted Cys residues in non-reducing conditions. The CD spectrum of the ^{193}DKCDK$^{197'}$ dimer (Appendix Fig. 3 *A*) is utterly different from that of WT PU.1 monomer, whereas it is similar to that of DNA-free PU.1 dimer (WT ΔN165 at ~800 μM at 50 mM salt concentration) which we reported recently (138). Therefore, the DNA-free dimer of PU.1 is suggested to be symmetric, in contrast with the DNA-bound dimer. The ^{193}DKCDK$^{197'}$ PU.1 dimer binds to cognate DNA >100-fold weakly than WT (Appendix Fig. 3 *B*), which further suggests that the DNA-free and -bound PU.1 dimers are conformationally distinct.

Appendix C.2: Electrostatic components responsible for DNA-bound PU.1 dimerization

The ^{193}NINI^{196} (in hPU.1) mutant study indicated that electrostatic interactions via the side-chain charges of ^{193}DKDK^{196} are important for PU.1 to form the 2:1 complex, as described in Chapter 2 (*cf.* Fig. 2.5 *H*). Therefore, we introduced more mutations at this site to further study which electrostatic components are responsible for DNA-bound PU.1 dimer formation. We used three mutants ^{193}AKAK^{196}, ^{193}DADA^{196}, and ^{193}TGDG^{196}. The first two mutants were designed to remove either all the positive or negative charges. The third mutant was designed from the same site (^{357}TGDG^{360}) of Ets-1, which is a close structural homolog of PU.1 but does not form a 2:1 complex with DNA in non-reducing conditions. DNA binding assays of fluorescence anisotropy using these mutants showed none of these mutants abolished the 2:1 DNA binding (Appendix Fig. 4 *A (a)-(d)* and Appendix Table 1). ^{193}DADA^{196} mutation did not change in binding affinity of the 1:1 and 2:1 complex. In sharp contrast, ^{193}AKAK^{196} and ^{193}TGDG^{196}

mutations made the binding affinities of both 1:1 and 2:1 complex lower than that of WT. Namely, for ^{193}AKAK^{196} mutation, K_{D1} was ~3-fold and K_{D2} was ~8-fold lower. For ^{193}TGDG^{196} mutation, K_{D1} was ~4-fold and K_{D2} was ~12-fold lower. Furthermore, ^{193}NINI^{196} mutation (i.e., abrogating all the charges from the ^{193}DKDK^{196} site) abolished the 2:1 binding (*cf.* Fig. 2.5 *H*). Taken together, PU.1 forms the 2:1 complex with a similar dissociation constant to WT even if the cationic charge is completely lost from the ^{193}DKDK^{196} site. In contrast, losing at least one Asp residue from the ^{193}DKDK^{196} site compromises the 2:1 binding. Thus, anionic charge in the ^{193}DKDK^{196} site is important for the 2:1 complex formation. The CD signals of the mutants suggest that ^{193}DADA^{196} and ^{193}AKAK^{196} mutation do not change the structure of PU.1, but ^{193}TGDG^{196} mutation causes a significant change in the structure (Appendix Fig. 4 *B*). Thus, ^{193}TGDG^{196} mutation not only abrogates three of four charged residues but also it changes PU.1 structure significantly. This may be the reason why removing one Asp residue by the ^{193}TGDG^{196} mutation compromises the 2:1 binding more than removing two Asp residues by the ^{193}AKAK^{196} mutation.

As a potential major driving force for the PU.1 self-association other than electrostatic interactions, the hydrophilicity of the ^{193}DKDK^{196} surface may be important for the association of two PU.1 molecules if water-mediated contact drives the association at the interface of PPI. In general, hydration greatly contributes to protein packing and association in general, thus water molecules play important roles in PPIs (152). The reliability of information about solvent molecules depends on the resolution in crystallographic structures. There are on average 1.0 interfacial water molecules per 100 Å^2, discovered from an analysis of a dataset of homo- protein complexes (<2.6 Å resolution) (153). The majority of these water molecules make hydrogen bonds with both partners of the association (154). It has been proposed that PPIs involve similar

levels of water-mediated contacts to direct contacts (154). Some interfacial water molecules are conserved among structurally homologous protein complexes, according to high-resolution crystal structures (152). Some amino acids of one protein at a PPI interface make both direct and water-mediated contacts with the other protein partner. Water-mediated contacts have been demonstrated to contribute to the energy of interaction (155). Thus, interfacial water molecules could facilitate interactions and recognition between protein partners in PPIs.

Appendix C.3: Nonspecific DNA binding of Ets-1 primarily perturbs the autoinhibitory module and H3

In striking contrast to the titration of PU.1 with nonspecific DNA as described above (*cf.* Appendix fig. 1 *A*), the ^{1}H-^{15}N HSQC crosspeaks of Ets-1ΔN280 do not disappear upon nonspecific DNA binding (Appendix Fig. 5 *A(a)*). Using the DNA-free and -bound Ets-1 backbone assignments reported previously (33,36), we were able to assign the HSQC resonances of both unbound Ets-1ΔN280 and its complex with 16-bp nonspecific DNA (at 1:1 Ets-1/DNA molar ratio) (Appendix Fig. 5 *A(b)*). Then, ^{1}H-^{15}N CSPs were calculated from the assigned HSQC resonances (Appendix Fig. 5 *B*). The CSPs as a whole are small, suggesting that nonspecific DNA binding marginally changes the structure of Ets-1, consistent with the literature (33). The residues that have large CSPs were mainly in HI-2, H3, H4, and H5. This indicates that nonspecific DNA binding of Ets-1 primarily perturbs the autoinhibitory module and H3, which is also consistent with the literature (33).

Our study of Ets-1 binding with nonspecific DNA confirmed the trend seen in the Ets-1 complex with nonspecific DNA, reported by Desjardins et al. (33). They previously reported that HI-1 and HI-2 of Ets-1 become predominantly unfolded, yet the protein is still ordered upon DNA binding, regardless of specific or nonspecific binding (33). Their NMR studies

demonstrated that nonspecific DNA binding of Ets-1 also contrasts with specific DNA binding, in that the structural and dynamic changes for Ets-1 are much larger with specific DNA than with nonspecific DNA (33). Their amide chemical shift perturbation mapping showed that dynamic electrostatic interactions drive association of Ets-1 with both specific and nonspecific DNA, through the same canonical interface of the ETS domain, whereas the formation of well-ordered complexes is driven by hydrogen bonding, with specific DNA (33).

For transcription factors in general, DNA-binding domains are dynamic to search for specific binding sites in a majority of nonspecific DNA until the DNA-scanning is quenched upon specific DNA binding (47,48). DNA-binding interface in the ETS domains of Ets-1 and ETV6 has also been demonstrated to be conformationally dynamic (33,37), and this feature is probably common for all the ETS domains. Such flexibility in the DNA-binding interface also explains why Ets-1 protein shares the same binding interface for specific and nonspecific DNA binding (33). However, the autoinhibitory module adjacent to the ETS domain is likely to compromise the flexibility in the ETS domain (or the DNA-binding interface), suggesting that the ETS domain of PU.1 (non-autoinhibited) is more flexible than that of Ets-1 (autoinhibited). Because of the flexibility of the ETS domain of PU.1, HSQC resonances of the ETS domain in complex with nonspecific DNA disappeared, in contrast to nonspecific DNA binding of Ets-1 (*cf*. Both Ets-1 (33) and PU.1 form a dimer with nonspecific DNA.) (*cf*. Fig. S2.3 *C*, Appendix Fig. 1 *A*, and Appendix Fig. 5 *A(b)*).

Appendix D: Conclusion

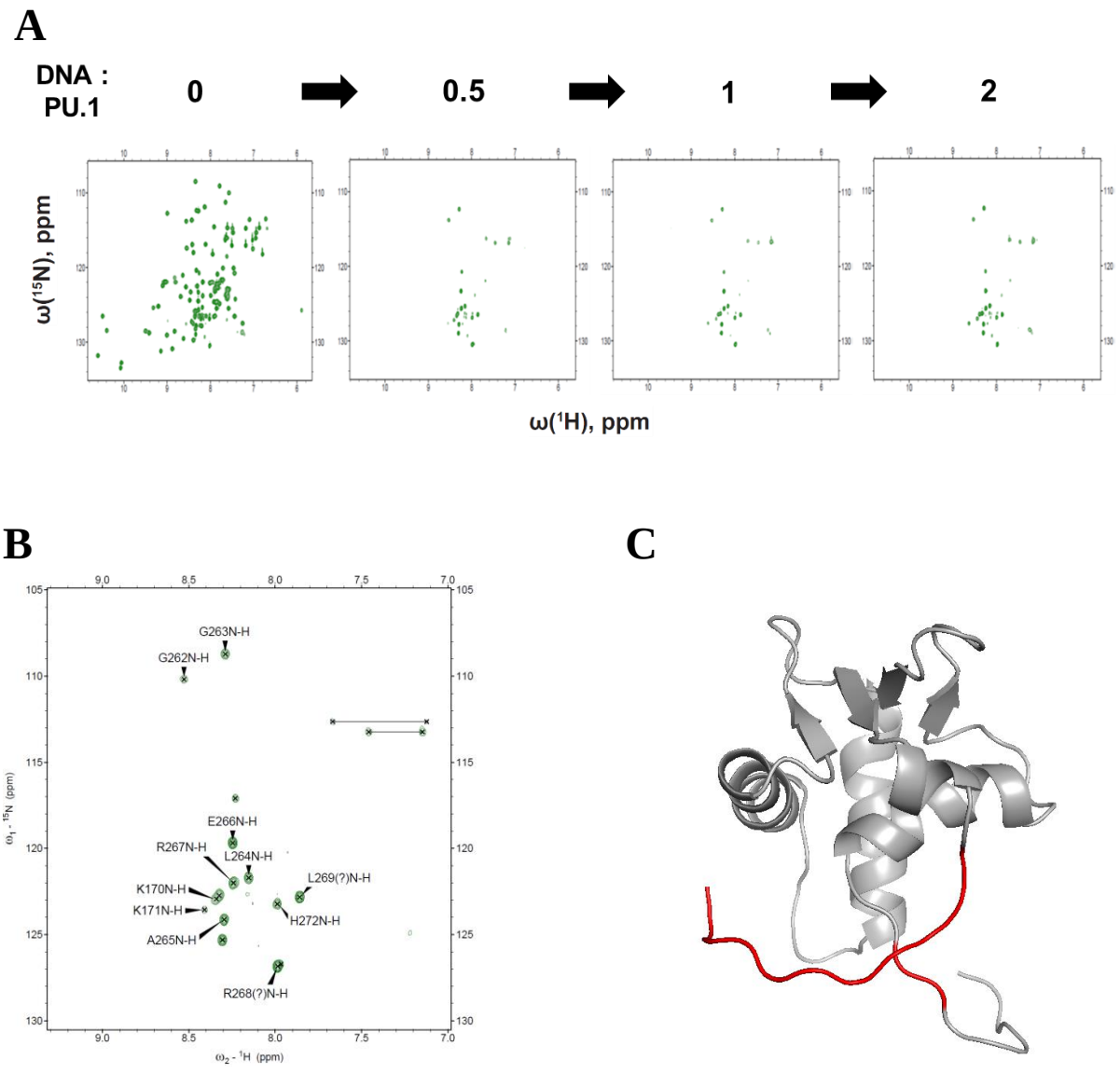
We observed the HSQC resonances of all the PU.1 ETS domain residues in the structured region (i.e., ETS domain) disappeared upon nonspecific DNA binding in a titration of PU.1 with nonspecific DNA, while the residues in the disordered regions did not. Thus, the HSQC peak

disappearance of PU.1 upon DNA binding was confirmed to be due to conformational exchange, but not disorder of the protein. We also studied which electrostatic components at the dimeric interface (^{193}DKDK^{196}) are responsible for DNA-bound PU.1 dimer formation. Anionic charge, but not cationic, in the ^{193}DKDK^{196} site is important for the 2:1 complex formation. The DNA-free PU.1 dimer (symmetric) and DNA-bound one (asymmetric) are conformationally distinct. Furthermore, we observed HSQC resonances of the PU.1 ETS domain in complex with nonspecific DNA disappeared, in contrast to nonspecific DNA binding of Ets-1, because of the flexibility of the ETS domain of PU.1. Even though nonspecific DNA binding of Ets-1 marginally changes the structure of Ets-1, it primarily perturbs the autoinhibitory module and H3, through the same binding interface as for specific DNA binding.

Appendix Table 1 Dissociation constants of WT and mutants of PU.1 ETS domain from fluorescent anisotropy at 150 mM total [Na^+].

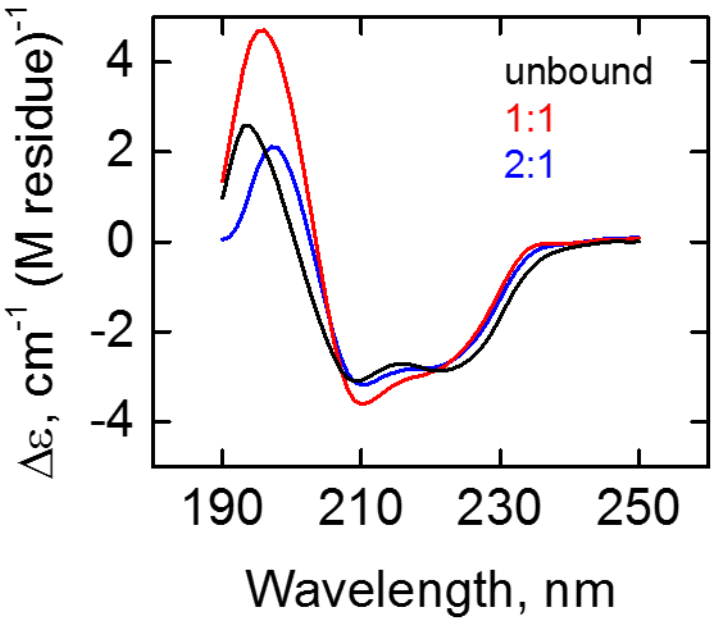
Dissociation constants of WT and mutants of PU.1 ETS domain at 150 mM total [Na^+] were obtained using fluorescent anisotropy. Note that the dissociation constant for the ^{193}DKCDK$^{197'}$ dimer was obtained by Ms. Suela Xhani.

	K_{D1}, M	K_{D2}, M
WT	$(1.1 \pm 0.5) \times 10^{-8}$	$(1.6 \pm 1.0) \times 10^{-6}$
^{193}AKAK^{196} mutant	$(3.2 \pm 1.4) \times 10^{-8}$	$(1.2 \pm 1.3) \times 10^{-5}$
^{193}DADA^{196} mutant	$(7.9 \pm 1.7) \times 10^{-9}$	$(1.1 \pm 0.7) \times 10^{-6}$
^{193}TGDG^{196} mutant	$(4.5 \pm 1.3) \times 10^{-8}$	$(2.0 \pm 2.5) \times 10^{-5}$
^{193}DKCDK$^{197'}$ mutant (dimer)	$(3.1 \pm 0.9) \times 10^{-6}$	



Appendix Figure 1 Assigned HSQC resonances revealed that the ETS domain of PU.1 in complex with nonspecific DNA is exchange-broadened, but not disordered.

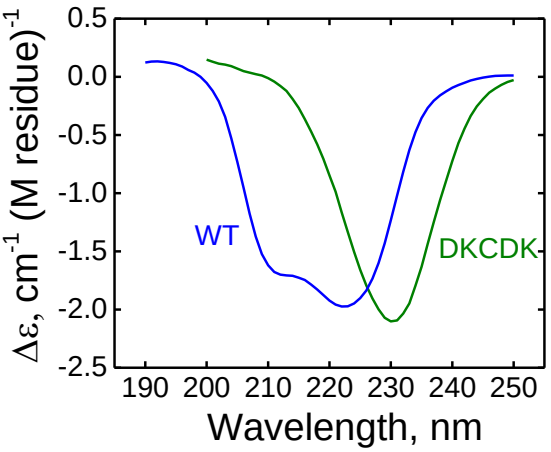
A, ^{1}H-^{15}N HSQC titration of mPU.1ΔN167 with 16-bp nonspecific DNA. *B*, assigned HSQC resonances of PU.1 in complex with 16-bp nonspecific DNA at 1:1 PU.1/DNA molar ratio. *C*, assigned HSQC resonances were mapped on the PU.1 structure (PDB: 5W3G). The assigned PU.1 residues are only in the N- and C-terminal disordered regions.



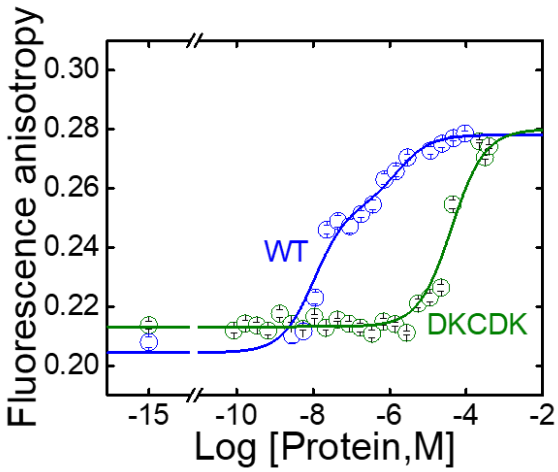
Appendix Figure 2 Far-UV CD spectra of the DNA-bound mPU.1ΔN167 upon subtracting the spectrum of the DNA acquired under identical conditions.

Far-UV CD spectra of the DNA-bound mPU.1ΔN167 in the unbound (black), 2:1 DNA-bound (blue), and 1:1 DNA-bound (red) form, upon subtracting the spectrum of the DNA acquired under identical conditions (75 μM, 0.15 M [Na^+]). The CD-detected structure of PU.1 showed negligible changes upon titration by DNA.

A



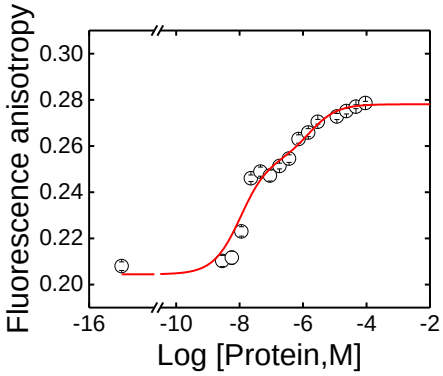
B



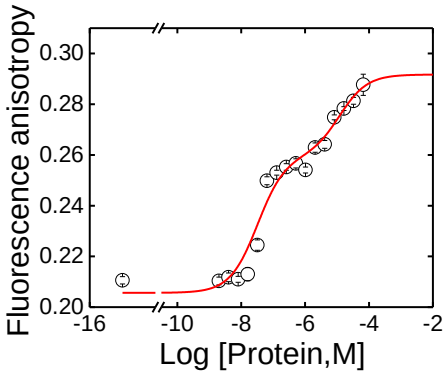
Appendix Figure 3 The ^{193}DKCDK$^{197'}$ dimer of hPU.1ΔN165, mimicking the symmetric DNA-free dimer of PU.1 ETS domain, is conformationally distinct from the asymmetric DNA-bound PU.1 dimer.

CD spectrum of WT ΔN165 (blue) and the ^{193}DKCDK$^{197'}$ dimer (green) (~400 μM proteins at 0.05 M [Na^+]) (A). The ^{193}DKCDK$^{197'}$ dimer is not conformationally similar to DNA-bound PU.1 dimer but resembles the DNA-free dimer (138). Fluorescence anisotropy of cognate DNA binding by WT hPU.1ΔN165 (blue) and the ^{193}DKCDK$^{197'}$ mutant (green) (B). Note that the CD measurements and the DNA binding assay using ^{193}DKCDK$^{197'}$ were conducted by Ms. Suela Xhani.

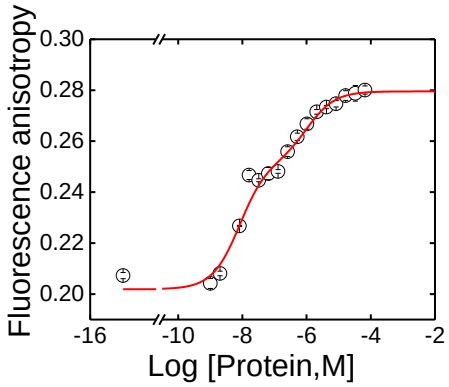
A (a) WT



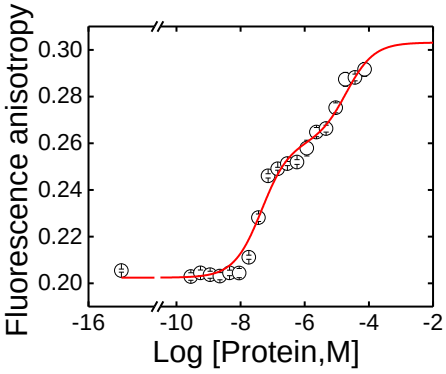
(b) AKAK



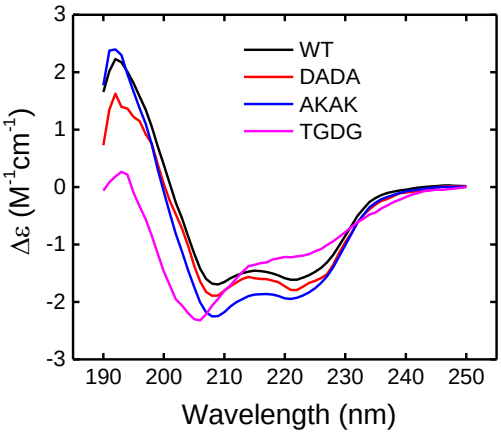
(c) DADA



(d) TGDG

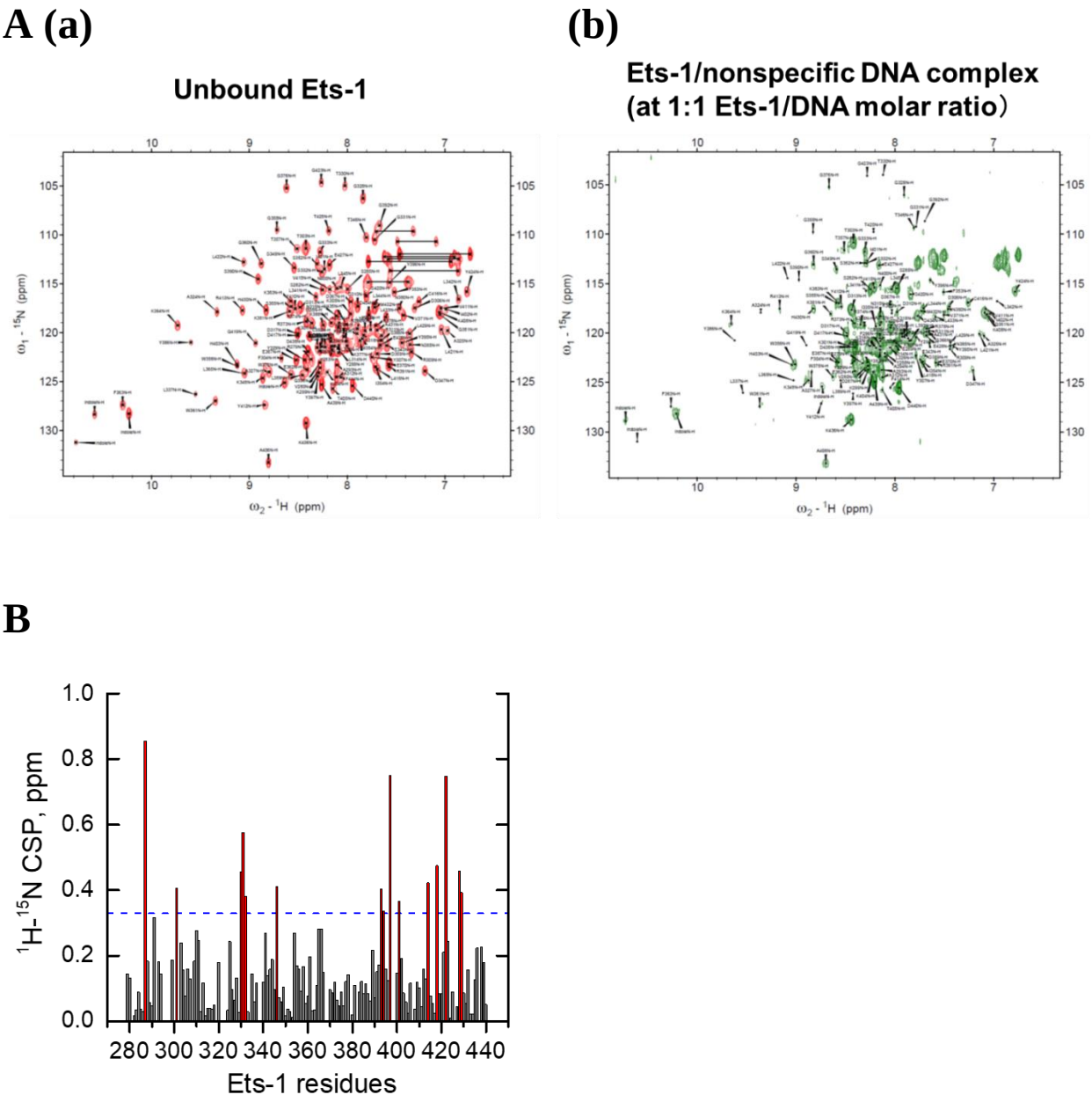


B



Appendix Figure 4 Fluorescence anisotropy of cognate DNA binding by WT hPU.1ΔN165, ^{193}AKAK^{196}, ^{193}DADA^{196}, and ^{193}TGDG^{196} mutants, and CD spectra of these PU.1 proteins.

A, DNA binding assays of fluorescence anisotropy using (a) wild-type hPU.1ΔN165 and three PU.1 ETS domain mutants (b) ^{193}AKAK^{196}, (c) ^{193}DADA^{196}, and (d) ^{193}TGDG^{196} with a 23-bp cognate DNA. The ^{193}AKAK^{196} and ^{193}DADA^{196} mutants were designed to remove either all the positive or negative charges from the wild-type sequence (^{193}DKDK^{196}). The ^{193}TGDG^{196} mutant was designed from the same site (^{357}TGDG^{360}) of Ets-1. *B*, Far-UV CD spectra of DNA-free PU.1 WT (black), ^{193}AKAK^{196} (blue), ^{193}DADA^{196} (red), and ^{193}TGDG^{196} (magenta) at 150 μM and 0.15 M [Na^+]. The CD-detected structures of ^{193}AKAK^{196} and ^{193}DADA^{196} showed negligible changes, but ^{193}TGDG^{196} showed significant changes, in comparison with WT.



Appendix Figure 5 Nonspecific DNA binding of Ets-1 primarily perturbs H3 and the autoinhibitory module.

A, ^{1}H-^{15}N HSQC resonances of (a) unbound and (b) 1:1 nonspecific DNA-bound Ets-1ΔN280. Resonances were assigned based on previous reports on Ets-1 (33,36). B, weighed average of amide (^{15}N and ^{1}H) chemical shift perturbations (CSPs) from unbound and 1:1 nonspecific DNA-bound Ets-1ΔN280, derived by $\Delta\delta = \sqrt{\{\delta^{1}H^{2} + 0.2\ (\delta^{15}N)^{2}\}}$, are plotted as a function of residue number.