12-2020

What Controls the Controller: Structure and Function Characterizations of Transcription Factor PU.1 Uncover Its Regulatory Mechanism

Suela Xhani Georgia State University

WHAT CONTROLS THE CONTROLLER: STRUCTURE AND FUNCTION CHARACTERIZATIONS OF TRANSCRIPTION FACTOR PU.1 UNCOVER ITS REGULATORY MECHANISM

by

SUELA XHANI

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

ABSTRACT

The ETS family transcription factor PU.1/Spi-1 is a master regulator of the self-renewal of hematopoietic stem cells and their differentiation along both major lymphoid and myeloid branches. PU.1 activity is determined in a dosage-dependent manner as a function of both its expression and real-time regulation at the DNA level. While control of PU.1 expression is well established, the molecular mechanisms of its real-time regulation remain elusive. Our work is focused on discovering a complete regulatory mechanism that governs the molecular interactions of PU.1. Structurally, PU.1 exhibits a classic transcription factor architecture in which intrinsically

disordered regions (IDR), consisting of 66% of its primary structure, are tethered to a well-structured DNA binding domain. The transcriptionally active form of PU.1 is a monomer that binds target DNA sites as a 1:1 complex. Our investigations show that IDRs of PU.1 reciprocally control two separate inactive dimeric forms, with and without DNA. At high concentrations, PU.1 forms a non-canonical 2:1 complex at a single DNA specific site. In the absence of DNA, PU.1 also forms a dimer, but it is incompatible with DNA binding. The DNA-free PU.1 dimer is further promoted by phosphomimetic mutants of IDR residues that are phosphorylated in B-lymphocytic activation. These results lead us to postulate a model of real-time PU.1 regulation, unknown in the ETS family, where independent dimeric forms antagonize each other to control the dosage of active PU.1 monomer at its target DNA sites. To demonstrate the biological relevance of our model, cellular assays probing PU.1-specific reporters and native target genes show that PU.1 transactivation exhibits a distinct dose response consistent with negative feedback. In summary, we have established the first model for the general real-time regulation of PU.1 at the DNA/protein level, without the need for recruiting specific binding partners. These novel interactions present potential therapeutic targets for correcting de-regulated PU.1 dosage in hematologic disorders, including leukemia, lymphoma, and myeloma.

INDEX WORDS: Transcription factors, Intrinsically disordered regions, Dimer, ETS family, Regulation, Negative feedback, PU.1, Hydration, Selectivity

WHAT CONTROLS THE CONTROLLER: STRUCTURE AND FUNCTION CHARACTERIZATIONS OF TRANSCRIPTION FACTOR PU.1 UNCOVER ITS REGULATORY MECHANISM

by

SUELA XHANI

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

WHAT CONTROLS THE CONTROLLER: STRUCTURE AND FUNCTION CHARACTERIZATIONS OF TRANSCRIPTION FACTOR PU.1 UNCOVER ITS REGULATORY MECHANISM

by

SUELA XHANI

Committee Chair: Gregory M. K Poon

Committee: Binghe Wang

W. David Wilson

Markus W. Germann

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2020

DEDICATION

This work is dedicated to my parents Arben Xhani and Miranda Xhani.

ACKNOWLEDGEMENTS

I want to acknowledge my advisor, Dr. Gregory Poon, for his guidance and support throughout the years. I appreciate him taking me under his wing and personally training me in numerous laboratory techniques ranging from purification of proteins to differential scanning calorimetry. Additionally, I would like to thank him for his daily ongoing moral and academic support. When I first joined the lab, I was terrified of public speaking and presenting. His constant encouragement has pushed me to put myself outside of my comfort zone and expand my scientific knowledge by attending various conferences. His belief in my abilities to overview many projects has given me confidence and a platform to lead. This work would not be possible without his help.

Next, I would acknowledge my committee members Dr. David Wilson, Dr. Markus Germann, and Dr. Binghe Wang. I would like to thank Dr. Wilson for always believing in me and provide me with the confidence to succeed in pursuing a Ph.D. I will never forget your great advice and guidance during our joint weekly meetings. Next, I would like to thank Dr. Germann for his continuous support and NMR teaching as a student during his lectures and as a graduate student. On numerous occasions, Dr. Germann has listened to my frustrations and patiently has pointed out the humor in situations. I would also like to thank Dr. Wang for graciously accepting to be part of my committee and providing thoughtful feedback on my research. Dr. Wang's input has helped me look at my research from a different lens that has improved my presentation delivery.

Additionally, I would like to extend my acknowledgments to our collaborators Dr. James Aramini at CUNY and Dr. Siming Wang at GSU. Dr. Aramini has patiently answered numerous questions, as well as running NMR experiments on our behalf. I want to thank Dr. Wang for always running our MS samples and taking her time to teach me how to analyze the ESI results.

I would like to thank the current and past members of the Poon lab: Kenneth Huang, Amanda Albrecht Reed, HyeMi Kim, Dr. Sangchoon Lee, Van Ha, Shingo Esaki, Mahtrab Khanezarrin, and Giselle Fernandez. A special thank you goes to Amanda and Ken for always being my companions throughout this journey. I will forever cherish our inspiring scientific conversations and the non-scientific ones discussed over lunch and coffee breaks.

I want to thank my parents Arben Xhani and Miranda Xhani, for being the world's best parents. Your love, compassion, and support both morally and financially have eased my journey. These past five years have been tough being away from you, and every time I thought grad school was hard, I remember the day you accompanied me to the airport when I left for Atlanta. I promised myself that day that I will make you proud and work hard to accomplish my goals. You have sacrificed immensely for a better opportunity for life, and I will forever be grateful. You have lead by example, to be honest, fair, and self-sufficient, and I wouldn't be standing where I am without your support. I love you both and will always strive to make you proud.

Additionally, I have to thank my sister Xhensila for always being a phone call away every time I felt stuck or just bored and wanted to chat. Thank you for being by my side and inspiring me to be less scared of the unknown and embrace new opportunities and adventure. Also, I like to thank my brother Dorian for his encouraging words throughout these years.

Furthermore, I would like to thank my husband, Ryan, for his constant love, support, and understanding, even in trying times. You have been my rock every time I cried over a failed experiment, and your words and support have inspired me to keep trying until I succeeded. It's the simple memories that I will fondly cherish during the graduate school years, such as our numerous nights discussing scientific ideas over a glass of wine and the morning drives in heavy Atlanta traffic (as annoying as those were). Your cheer, positivity, and confidence in me have given me

the strength to believe in myself and shaped me into the person I am today. I wouldn't choose anyone else to go through this graduate school journey, and I cannot wait for our next adventure together!!

Finally, I would like to thank the funding sources of the Molecular Basis of Diseases and the Harry P. Hopkins, Jr. Scholarship in Physical Chemistry and the Alfons L. Baumstark Fellowship for their financial support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Transcription Factors (TF)

Pre-initiation complex (PIC)

RNA polymerase II (RNA pol II)

DNA binding domain (DBD)

Transactivation domain (TAD)

Proline, Glutamic acid, Serine, Threonine (PEST)

Tryptophan repressor protein (TR)

E-twenty-six-specific (ETS)

Pointed domain (PNT)

PU-box binding-1 (PU.1)

Intrinsically disordered regions (IDR)

TATA-binding protein (TBP)

Sterile alpha motif (SAM)

Zinc finger (ZNF)

Homeodomains (HD)

Helix-turn-helix (HTH)

Helix-loop-helix (HLH)

Basic protein –leucine-zipper (bZIP)

Winged helix-turn-helix (wHTH)

Phorbol 12-myristate 13-acetate (PMA)

Protein Data Bank (PDB)

Dithiothreitol (DTT)

Chemical shift perturbations (CSPs)

Bovine serum albumin (BSA)

Differential scanning calorimetry (DSC)

Heat capacity change (ΔCp)

Solvent accessible surface area (SASA)

Interferon regulatory factor, (IRF4)

Liquid-liquid phase separation (LLPS)

1 INTRODUCTION

1.1 Eukaryotic gene transcription

The blueprint of life is the DNA genetic code which contains the instructions for a cell to sustain itself. Eukaryotic gene transcription is much more complicated than prokaryotic gene transcription as the first one is carried out in the nucleus of the cell and translated into the cytoplasm. This spatial separation allows eukaryotic gene expression to be regulated differently, contributing to the vast eukaryotic forms and function (1). An additional level of complexity of eukaryotic gene expression comes from the different promoter elements such as the TATA box, CAAT box, initiation elements, and the downstream promoter elements (2), which can be recognized other proteins than RNA polymerase II (1). Eukaryotes have three types of RNA polymerases I, II, and III, while the prokaryotic expression only has one (3).

1.2 The central dogma of molecular biology

The central dogma of molecular biology shown in [Figure 1.1] demonstrates the whole process from gene sequence to protein product. The concept was introduced by Francis Crick in 1957 and restated in 1970 in a Nature review (4).

There are three significant steps involved in molecular biology's central dogma: DNA replication, transcription, and translation. DNA replication occurs when the genetic material of the mother cell is passed down to its daughter cells. This replication process is highly regulated as it is one of the most crucial steps of the survival of the cell, and (5) is controlled by a complex regulatory machine involving numerous enzymes and proteins called the replisome (6). The process requires a helicase to unwind the double-stranded DNA and a single-stranded protein (SSP) that will inhibit it from reassembling. RNA primase will then create an RNA primer-template for each strand, giving DNA polymerase III the green light to read the existing 3’ to 5’

template and add new nucleotides from the 5' to 3'of the new daughter strand. DNA polymerase I will then remove the RNA primers leaving DNA ligase to join the two fragments by phosphodiester bonds, creating an exact copy of the chromosomal DNA (5).

Transcription is making a messenger RNA (mRNA) that encodes the synthesis of the desired protein from DNA in the nucleus of the cell. Transcription in eukaryotic systems has three main stages: initiation, elongation, and termination. The initiation step begins with the assembly of the pre-initiation complex (PIC), which is composed of several general transcription factors (GTFs) and mRNA polymerase II (Pol II) [Figure 1.2] (7). The second step of transcription is elongation which is a repetitive bond formation at each base pair in the DNA-track system (8). The last step is termination, where the RNA polymerase II will fall off, causing a complete dissociation of the mRNA produced from the DNA-RNA hybrid (9). Although all three transcription stages are crucial for regulation, initiation is directly affected by proteins (called transcription factors) that bind to a specific DNA binding site and modulate RNA polymerase II activity (10).

Eukaryotic transcription initiation includes many vital components such as gene-specific transcription factors (activators/repressors), co-factors such as histone modifiers, chromatin remodelers, general transcription factors, and the assembly of PIC (11) [Figure 1.2]. Transcription initiation starts when a gene-specific transcription factor (TF) binds to an enhancer region of DNA, allowing for a TBP (TATA-binding protein) binding to the TATA region (promoter region) and induces a ~90^{o} bend on the DNA (12,13). Next, general transcription factors TFIIA and TFIIB are recruited following by Pol II and TFII, with the last one stabilizing DNA along with the branched polymerase II (14). Finally, TFIIH and TFIIE are recruited to complete the eukaryotic PIC (15). The PIC assembly and the TF at the enhancer region are then stabilized in place by a mediator

complex through protein-protein interactions, and (16) once everything is in place, Pol II unwinds four DNA turns with help from TFIIH and is ready to start transcription (17).

The last step of translation follows the conversion of mRNA into a protein by the ribosome in the cell cytoplasm. Just like transcription, translation has three stages: initiation, elongation, and termination. The first step begins when the ribosome assembles around mRNA at the AUG codon (start codon) containing a Methionine-transfer RNA (met-tRNA) and reads mRNA from the 5' to 3' (18). The elongation step then takes over, which decodes codons and forms peptide bonds. The whole elongation process has four sub-steps to sequentially add amino acids to the nearest peptide's carboxy-terminal end and, therefore, complete the formation of a polypeptide (19). The termination step is signaled by the presence of a STOP codon such as UGA, resulting from the release of the polypeptide chain (20). The newly formed polypeptide chain undergoes post-translational modification to become a fully functional protein that will go on to carry out its function.

1.3 Eukaryotic transcription factors

1.3.1 Classification of transcription factors

Roughly 10% of the eukaryotic genome encodes for the expression of transcription factors (21-23). Historically, the term transcription factor (TF) was given to any protein that contributed to transcription or gene expression. As more data has emerged, the term can be more specific to only proteins capable of 1) binding sequence-specific DNA and 2) regulating transcription (24,25). In humans, roughly 300 TFs belong to the necessary transcriptional machinery, while approximately 3000 TFs are sequence-specific DNA binding proteins (25). Some specific binding TFs are found virtually in all organisms and cells because they are responsible for transcribing "housekeeping" genes (11). In contrast, others serve as an activator or repressor of particular genes

(26). In some cases, the same TF acts as both activator and repressor, such as ETS family transcription factor ETV6 (27). TFs regulate genes by binding cis-acting enhancers (a segment of DNA ~100 bps) and recruit co-activators/repressors and, in some cases, RNA polymerase II to target genes (28). In embryonic stem cells, gene expression is surprisingly controlled by a small number of transcription factors referred to as master transcription factors (28) or pioneering factors due to their ability to overcome chromatin restrictions (29). PU.1 is considered a master transcription factor due to its involvement in the regulation of both lymphoid and myeloid progenitor cells in mice (30,31), B cells (32), as well as human monocytes and macrophages (33).

Transcriptional misregulation is responsible for many diseases and cancers (25), often due to mutations in the DNA binding domain (DBD) of the transcription factors such as HOXD13, which leads to limb malformation. Other times these diseases manifest with a strong immune dependent component with B cell and T cell-specific genetic signals (34). The link between transcription regulation and human diseases is clear. Therefore, it is important to stress the efforts of understanding transcription factors real-time regulatory mechanisms.

1.3.2 Structure of transcription factors

The primary feature of transcription factors is their DNA binding domain (DBD), which recognizes specific DNA sequences within their enhancer region of the genes they control. The interaction of the DBD and the DNA occurs by the sidechains of the amino acids of the DBD interacting with the nucleotide bases located in the double-helical DNA structure. Although minor groove binders are reported, the interactions between the DBD and nucleotide bases allow transcription factors to target DNA's major groove (35). The interactions between the DBD and DNA is held in place by hydrogen bonding. Van der Waals contacts determine the transcription factor's specificity and affinity to its DNA target site (36-39). Ionic interactions also play a role in

the stability of the positively charged residues of the DBD with the DNA phosphate backbone. Additionally, the topology of the DNA (its curvature) influences the affinity a transcription factor has to the target site (40), with some TF inducing DNA bending upon binding (41).

Transcription factors can be divided into families according to their DBD structural characteristics [Figure 1.3]. These families include the zinc-finger (ZNF) motif (it is the most common DBD motif of TFs with ~50% of known TFs (25)), the Homeodomains (HDs), Helix-turn-helix (HTH), Helix-loop-helix (HLH), Basic protein-leucine-zipper (bZIP) and the winged helix-turn-helix (wHTH). [Figure 1.3] shows an example of a transcription factor of each of the most common motifs.

The second essential domain is the transactivation domain (TAD), which serves as a contact for transcription co-activators (42). It is suggested the TAD facilitates interactions of the recruitment of the transcription machinery during initiation. Furthermore, it has been observed that TADs tend to be intrinsically disordered with a high number of acidic and hydrophobic residues, giving it the name of "acidic blobs and negative noodles" (43). An additional optional domain is a signal-sensing domain such as the PEST domain of ETS family transcription factor PU.1, which serves as a prosomal recruiting signal for metabolic turnover (44). The PEST domain stands for the abundance of the proline (P), glutamic acid (E), serine (S), and threonine (T), and similarly to the TAD domain, it is highly acidic and intrinsically disordered. Eukaryotic transcription factors are composed of a higher percentage of intrinsically disordered regions than prokaryotic transcription factors (45). As a further breakdown, there are broadly two types of eukaryotic transcription factors [Figure 1.4]. Type I only has the DBD as a well-folded domain tethered by intrinsically disordered regions on both N-terminal and C-terminal. In contrast, Type II has other well-structured domains serving many functions in addition to the DBD, which are connected by

intrinsically disordered regions (45). This leads one to argue that eukaryotic transcription factors have evolved to incorporate intrinsically disordered regions in their function. Therefore, it is important to study these regions to understand their contribution to gene regulation fully.

1.3.3 Intrinsically disordered properties of transcription factor

A genome-wide analysis study showed that ~25% of eukaryotic proteins have intrinsically disordered regions with more than 50 residues (46). For a while, prior biological significance was only attributed to proteins with a well-defined tertiary structure (47). However, it was discovered that intrinsically disordered proteins (IDP), even though lacking well-folded domains, have several motifs that interact with RNA, DNA, or other proteins (48). The amount of IDRs these proteins possess in larger than 200-300 residues (~80% of the protein) and are determined by the presence of specific amino acids (P, E, S, Q, and K) (49). There are different time scale dynamics that proteins undertake depending on the motion involved, such as partial folding (slower timescale) or vibrational and loop motions (faster time scale) (50). One dynamic process that required changes in torsion angles of the backbone polypeptide is responsible for conformational changes (51). Only two rotations are allowed for a polypeptide backbone the N-Cα bond named ϕ and Cα-C bond called ψ (52). These torsional dynamics are fundamental for any protein dynamics and can be mapped by Ramachandran ϕ-ψ conformational space. Typical α-helix and β-sheet secondary structure have angles that cluster around ~60 ψ/-60ϕ and ~ 120 ψ/-120 ϕ, respectively [Figure 1.7]. IDPs or proteins with IDRs dynamics arise from the Ramachandran dihedral angle fluctuation due to their random-coil similarities and a lack of well-defined secondary structure. IDPs have been shown to have high biological relevance as they are found in transcription factors, ribosomal proteins, and ribonucleoproteins complexes (47) and are associated with different functions and diseases (47,53,54). A challenging aspect of characterizing IDPs or proteins with IDRs are the

same features that make them interesting in the first place: lack of structure. The techniques to analyze these type of proteins are limited to a handful of examples including circular dichroism (CD) and NMR (47).

CD spectroscopy is a technique that characterizes the protein by the secondary structure (alpha helix, beta-sheets, and random coil) by monitoring in the far UV region (170-250nm) (55). CD provides structural data of IDPs and IDRs. Alpha helical proteins have a distinct double negative trough at 208nm and 222 nm and a peak at 190 nm. Beta sheets proteins also have distinct trough patterns at 217 nm and a peak at 198 nm (47). While IDPs or proteins rich in IDRs have a strong negative signal at 200 nm, very similar to the random coil, supporting the idea that IDPs and IDRs lack a well-folded secondary structure (56-58).

NMR spectroscopy is another technique that provides excellent insight into the nature of long IDRs (59). Typically, a triple-resonance (^{13}C, 15N, and ^{1}H nuclei) experiment is used to find the neighboring residues' sequential connection. For globular proteins, these residues are dispersed throughout due to their chemical environment. While, IDRs tend to cluster around 8.2 ± 0.2 ppm in ^{15}N-^{1}H spectra, creating a spectral overlap and, therefore, difficult to distinguish residue signals (60). This feature is very typical of IDRs and IDPs, reinforcing the idea that they are dynamic. ^{1}H-^{15}N Heteronuclear nuclear Overhauser effects (NOEs) is a method often used to identify the dynamics of IDRs residues as they have low signal or entirely negative signals. In contrast, residues in the structured domain have high positive values. NOE values come from the ratio of the saturated spectrum's peak intensities to the unsaturated spectrum (61).

1.3.4 Regulation of transcription factors

Gene regulation is crucial in living organisms, as the misregulation of the initiation stage of transcription often leads to a broad range of diseases (62). Gene expression is a complicated

process involving the organization of numerous multi-level events of regulation. The primary regulators of gene expression are transcription factors, which beg the question: what regulates these regulators upon binding to their specific DNA sequence? The multi-level complexity of understanding the interactions between TFs and DNA arises from the involvement of other cofactors, cooperative binding of other transcription factors such as interferon regulatory factor (IRF4) with PU.1 (63), as well as epigenetic modifications like DNA methylation, post-translation modification, chromatin remodeling and histone modification (64). However, TFs can also utilize its structure to regulate the DNA binding affinity by auto-inhibition. ETS-1 is an example of a TF that undergoes a structural transition of unfolding helix HI-1 upon binding DNA (65). Additional regulation arises from the surrounding environmental stress response that diverges the attention from proliferation to stress protection such as DNA damage, stressed repressed genes, defense against reactive oxygen species, and osmotic stress (66). Other times TFs will form a homo-dimeric complex to regulate their DNA binding activity. Tryptophan repressor protein (TR) is an excellent example of a TF that utilizes dimerization and the environmental response to repress genes when tryptophan concentration is high. TR does so by using hydration as part of its mechanism where interfacial waters are fundamental for its thermodynamic stability and the active complexation of TR dimer with its specific target DNA (67).

Nature has evolved eukaryotic transcription factors to have a diverse sense of real-time regulatory mechanisms, as described above. However, considering that most of these transcription factors are enriched within intrinsically disordered regions, it is interesting to explore their role in self-regulation at the DNA level. Due to their disordered nature, these domains are often left unstudied, leaving the full characterization only to the well-folded DBD. This presents an incomplete representation of gene regulation and a gap in basic biophysical knowledge of each

domain of the transcription factor's role. We have selected transcription factor PU.1, a member of the ETS family, as an ideal model to study how hydration (response to osmotic pressure) and intrinsically disordered regions modulate the real-time regulation at the DNA/protein level.

1.3.5 Hydration of proteins

Water is the most abundant molecule on Earth, covering ~70% of the surface. Water has shaped the way life on Earth has evolved as hydration interactions highly influence protein molecules. Protein hydration, its dynamics, and packing have affected the way proteins can fold and bind (68); therefore, characterizing the hydration patterns of proteins will determine the energetics, stability, and recognition.

It has been observed that ordered waters are present in crystallographic structures (69) and that they are involved in many regulatory processes such as DNA recognition (70), substrate specificity (71), and catalysis (72). From the first solved structure of a protein bound to DNA, it was apparent what elements play a role in the interaction: electrostatic interactions, ion release, conformation rearrangement, hydrogen bonding, hydrophobic effects, direct contacts between protein and DNA, and water-mediated contacts (73-75). The role of water in both free and protein-bound DNA has exhaustively been studied by both experimental and computational techniques as different mechanisms have been proposed based on interaction and recognition (76). Water is abundantly present and affects the specificity of the protein/DNA complex (77). At the local interface of the protein/DNA waters can act as an intermediate bridge between protein/DNA or interact dynamically. The contribution of hydration is due to the difference between the bulk water medium and the waters in the local interface of the protein/DNA complex (78). The restriction endonuclease EcoRI is an excellent example of how critical hydration contribution is in DNA specific binding as the number of waters present determines its ability to bind high affinity or

nonspecific DNA (79). The hydration patterns of protein/DNA ensembles are critical in forming the complex and the ability to bind specific high affinity DNA.

Advances in NMR and X-ray crystallography techniques have allowed for solving numerous protein/DNA structures. Many crystallographic complexes have revealed that the minor groove of DNA is the most hydrated (80), and simulation studies have supported these crystallographic/NMR observations (81). However, a current restriction with crystallography and NMR is flexibility and size, respectively (82). Flexible proteins are hard to crystalize, and therefore identifying the proteins' hydration patterns in those specific areas remains a mystery. The obstacle with NMR and larger proteins (>30 KDa) has to do with increases in spectral complexity and sensitivity loss (83). Other techniques such as hydrodynamics (ex: analytical ultra-centrifugation AUC, densimetry) and solution scattering (ex: small-angle X-ray scattering, (SAXS)) are often used when crystallography and NMR are limited. These techniques have their advantages as they can be used at very dilute solutions. There is no size limit; however they are low resolution, and the interpretation of the data depends on the parameters involved (82).

Volumetric measurement techniques such as densimetry and hydrostatic pressure are often used to provide insight into proteins' hydration. Interactions of charged residues of protein with surrounding water molecules can change thermodynamics properties of bulk water and therefore distinguish waters important for protein hydration (68).

1.3.5.1 Volumetric analysis to measure hydration in protein/DNA binding

The hydration patterns of a protein/DNA complex can be interpreted thermodynamically. Micro-calorimetry gives accurate heat changes of macromolecules in an aqueous solution, reflecting hydration at a specific condition (84). Thermodynamics profiles can be also be acquired by placing the protein or complex (protein/DNA) under non-ideality conditions such as osmotic

pressure and exclusion volumes (85). Pressure can be applied to bio-molecular systems [Equation 1] (protein-protein, protein-DNA) to study the complex conformational and supramolecular structure without altering the bond angles and bond length.

$$P + D \underset{K_D}{\rightleftharpoons} PD \tag{1}$$

where P stands for protein and D for DNA. The experiments' pressure is typically ranging from 1 bar to 10 kbar (86). The effects of pressure on chemical equilibrium can be described by [Equation 2]

$$\left(\frac{\partial \ln K}{\partial P}\right)_T = -\frac{\Delta V}{RT} \tag{2}$$

Where K is the pressure-dependent- equilibrium constant, and ΔV is the associated reaction volume. The volume change is measured by the partial molar volume of the products subtracting the reactants (86,87). [Equation 3]

$$\Delta V = V_{DNA:\mathrm{Protein}} - V_{DNA} - V_{\mathrm{Protein}} \tag{3}$$

Densimetry is a technique used to calculate the partial molar volume V^o [Equation 4]

$$V^o = \frac{M}{\rho_o} - \frac{\rho - \rho_o}{\rho_o C} \tag{4}$$

M is the solute's molecular weight, and ρ_o is the density of the buffer, ρ is the density measured, and C is the concentration of the molar solute. The change in volume includes the volume of complex formation as well as the hydration volume, and it can be expressed in three additive parts [Equation 5]

$$\Delta V = \Delta V_M + \Delta V_T + \Delta V_I \tag{5}$$

Where V_M is the intrinsic volume of the solute and ΔV_M is the change is solvent excluded volume. ΔV_T is the change in thermal volume (configurational dynamics). Finally, ΔV_I is

interaction volume, which changes hydration in binding (Δn_h). For the hydration of a protein binding to DNA this is the most critical parameter as it gives us insight into how many water molecules are released: where the denominator (V_h-V_o) are taken to be -1.8 cm^3/mol [Equation 6]

$$\Delta n_h = \frac{\Delta V_I}{V_h^o - V_o^o} \tag{6}$$

Densimetry is a useful technique in determining the protein's density changes as it undergoes different interactions (self-association or protein-DNA). However, densimetry is also limited to high concentrations as sensitive as other techniques such as osmotic pressure.

Hydrostatic pressure is another technique used to study proteins under stressful conditions posed by pressure. Hydrostatic pressure is a significant biological feature as it essential for many extremophile organisms (88) that populate the ocean floor where the pressure is 1,100 atm (89,90). Some bacterial species have also been found deep in the Earth's crest, where the temperature is 102^{o}C, and the pressure is 3,000 atm (91). These discoveries of organisms living in such a harsh environment have led many to question how proteins, DNA, and RNA can alter their physical chemistry to hydrostatic pressure (90). Hydrostatic pressure perturbs the volume of the system without changing the internal energy (92). Hydrostatic pressure experiments are carried out by applying excessive pressure to the system of about 10 kbar (9.87 atm) as the goal often is to promote dissociation and disturbance of globular structure (92). Increasing hydrostatic pressure disturbs the equilibrium between native, N, and unfolded, U, states of the protein [Equation 7].

$$N \rightleftarrows U \tag{7}$$

For a two-state equilibrium process, the pressure alteration is defined by the volume change upon unfolding (90) [Equation 8].

$$\left(\frac{\partial \ln K}{\partial P}\right)_T = \Delta V \tag{8}$$

One particular limitation of hydrostatic pressure is that you require prior knowledge of a structure for better accuracy. Additionally, the instrumentation available to carry the experiments is costly, and accurate measurements require a constant sample density throughout.

1.3.5.2 Osmotic pressure on protein/DNA binding

Osmotic pressure (93) experiments can be used to determine the number of water molecules released upon binding as well as determine specific and nonspecific DNA binding (94,95). For example, osmotic pressure studies of hydration changes have been useful in studying binding substrate to hexokinase (96). As the substrate binds to hexokinase, its estimated dependence affinity on osmotic pressure is visible from the osmotic work done to remove the waters from the osmolyte-excluded cavity. Therefore, the volume of waters removed is calculated by osmotic pressure and osmotic work (97). The osmolyte-excluded cavity refers to the volume of the osmolytes that are excluded (97). Osmolytes are osmotic agents, such as urea, which denatures proteins, crowders that exclude the volume of proteins or complexes (ex. Polyethylene glycol PEG), and non-ionic co-solutes (ex: betaine) (98). Depending on the experiments, different osmolytes for different outcomes are required.

Our focus is to determine ETS family protein hydration patterns, as is demonstrated in chapter 2 between PU.1 and ETS-1. A widely neutral compatible non-ionic co-solute betaine (99) was chosen to apply osmotic pressure on the protein/DNA complex. The data was analyzed by interpreting the effects of betaine on the equilibrium constant under increasing osmotic pressure (100). Assuming complete exclusion in binding from osmolyte (101) and the binding constant ($-\log K_D$) as the equilibrium constant, we can use the equation in terms of osmolality where 55.5 is molal of pure water while ΔVw is the hydration change between free and bound forms. Positive

change in hydration is interpreted as water uptake and negative values of change in hydration as net release [Equation 9].

$$-\frac{\partial \log K_D}{\partial Osm} = -\frac{\Delta V_w}{55.5 \ln 10} \quad (9)$$

1.3.6 It takes two to tango: dimerization in transcription factors

Dimerization is the process in which two proteins interact with one another. This collaboration between two proteins could be heterotopic (proteins deriving from two different gene families) or homotopic (proteins interacting with members of the same family) (102). Dimerization is used as a regulatory mechanism depending on the partner choice and cellular system such as in the cell cycle, immunity, development, metabolism, homeostasis, and programmed cell death (102). For this reason, there is a high interest in understanding how these dimeric properties evolved to develop therapeutic targets (103). Functional implications of dimerization allow recognition of different DNA elements while the monomer-to-dimer transition can act as an additional regulatory mechanism for transcription factors. Dimerization is also linked to protein concentration, synergy, and allostery (104,105), as well as non-linear behavior such as oscillations (106). A monomeric transcription factor could have different binding partners that interact with one another in different ways (concentration, phosphorylation, or affinity). This promiscuous behavior allows for extra layers of different gene regulations (102). Dimerization could occur in solution as the two monomers associate and then bind to DNA, also referred to as dimer pathway, or they bind to DNA as a monomer and then dimerize (monomer pathway) (107).

1.4 ETS family of transcription factors

The first *ets* oncogene was discovered in 1983 in avian erythroblast virus, E26, which also gave it its name E-twenty-six-specific sequence (108). The ETS family is present throughout the

whole animal kingdom with 28 ETS genes in humans classified into 12 subgroups based upon the DNA-binding domain, also called the ETS domain (109) [Figure 1.4]. The ETS domain is composed of ~85 residues that form a winged helix-turn-helix (wHTH) motif with three α-helices and four β-sheets that recognize a core consensus GGAA/T [Figure 1.5]. Helix 3 is the recognition helix as it makes contact with the major groove of DNA where the consensus GGAA/T is located. Two arginines in Helix 3 make contact with two guanines in the ETS binding site. The binding specificity of each member of the ETS family is determined by the flanking sequence of the DNA as the core consensus is similar for all of them (110) as well as interaction with other nuclear factors. (111).

Another conserved domain present in 11 out of 28 members is the pointed (PNT) domain. The PNT domain has a well-structured sterile alpha motif (SAM) and is shown to be involved in homo/hetero-dimerization as well as protein-protein interactions (111).

1.4.1 The biological role of ETS transcription factors and diseases

To date, there have been 700 ETS target genes identified (111). For a long time, it was believed that the ETS members were either activators or repressors, while now we have a clearer understanding that some of them can function as both depending on the type of promoter, making a very diverse set of transcription factors. So far, 11 members are embryonic lethal, with the majority demonstrating their role in hematopoiesis or combination with other lineage defects (111). Hematopoietic regulators include PU.1, which is a key regulator for the differentiation of both myeloid and lymphoid lineages (112), ETS-1(113) and Elf-4 (114), which are principal regulators of T cells production, and Fli-1 (115) along with Etv-6 (116) and Etv-2 (117), which display functions in angiogenesis. The ETS members involved in lethal non-hematopoietic defects include Etv1 and Fev for defective neurogenesis (111), and Etv-5 for male fertility (118).

Deregulated *ets* gene expression is linked to numerous diseases in humans, including leukemia, thyroid, liver, prostate, colon, and breast carcinomas (119). The ETS transcription factors have been found to regulate the expression of genes responsible for cancer cells that invade the nearby tissues or move into the bloodstream to invade tissues all over the body called malignant and metastatic processes (111). Examples of these processes include cell proliferation such as the case for cdks and cyclins, invasion (ex: urokinase-type plasminogen activator; uPA), extravasation (Integrins), motility (hepatocyte growth factor, HGF) as well as the maintenance of new blood vessels that deliver oxygen to the tissues called angiogenesis (120). Correlation analysis further demonstrates a connection between *ets* genes and the ETS transcription factors with the progression of prostate cancer (121). SPDEF an ETS member has been shown by functional data to negatively regulate uPA (122), and both have been found to be inversely correlated in colon cancer (123).

1.4.2 Regulation of ETS transcription factors

ETS protein regulation depends on the complexes they form with different transcription factors as it strengthens their activity interactions and specificity to the binding of DNA (111). Many cofactors of ETS members have their DNA binding sites adjacent to the ETS DNA binding site (124,125). Some of these cofactors include AP1 (*fos/Jun*) IRF4/8, PAX5, GATA1, to name a few (111). ETS members Etv-1, Etv-4, Etv-5, and ERG have been shown to interact with AP1 transcription factors to activate MAPK in prostate cells in the absence of MAPK activators (126). PU.1 interacts with IRF4 and IRF8 through the Ets/interferon consensus elements (EICE) to regulate early B cell development (127). ETS-1, in collaboration with PAX5 allows for binding to a non-consensus of ETS binding domain in early B-cell specific promoter (128). Another well-

formed partnership is Fli1 and GATA-1, as they both work together to activate multiple megakaryocytic genes such as *gpIX, gpVI,* and *gpIb* (129).

In addition to partner proteins, ETS factors are regulated by phosphorylation, acetylation, ubiquitination, and sumoylation (111). Perhaps the most studied regulatory mechanism of ETS proteins is phosphorylation, as it is crucial for protein-protein interaction, DNA binding, and transcriptional activation. Many members (ETV-1, ETV-4, GABPA, ETS-1, ETS-2, SPIB, ELK1, ELK-3, and ELK-4) can be phosphorylated by MAPKs to activate transcription (130). At the same time, the ETV6 DNA binding affinity is weekend by MAPK phosphorylation as it is negatively regulated (111). PU.1 utilizes phosphorylation of its S148 in the PEST region for protein-protein interaction with IRF4 (131) as well as S41 in its TAD region is induced by AKT signal in a phosphoinositide 3-kinase-dependent manner (132). Acetylation is another common form of regulatory mechanism for ETS factors. For example, ETS-1 is acetylated as a response from the TGFβ signal dissociating it from the CBP/p300 complex (133). Oppositely Etv1 has an increase in DNA binding affinity when it is acetylated, activating its transcriptional genes. (134) Additionally, ETS TFs get regulated by ubiquitination and proteasome degradation as well as sumoylation as the latter shows a decrease in transcriptional activity for ETS-1, Etv-4, and Etv-5. (111).

1.4.3 Dimerization of ETS proteins

Dimerization in the ETS family occurs differently for specific members. 11 members possess a PNT domain, with ERG, SPDEF, and ELF subfamilies showing no reported activity. Etv6 and its Drosophila homolog Yan dimerize through its PNT domain, which is a well-structured domain. (109) The PNT domain allows Etv6 to self-associate (homodimer) (135) utilizing its helical homo-polymers in a head-to-tail orientation (136-138). Homotypic interactions allow

ETV6 to increase DNA binding affinity. Additionally, Etv6 forms heterodimeric complexes through the PNT of other proteins such as NAD-dependent malic enzyme (*Mae*) to facilitate the phosphorylation of Etv6 during MAP kinase Rolled (a Drosophila homolog of human MAPK3) (109). These interactions (either homotopic or heterotopic) through the PNT domain are in response to different signals.

Another way the ETS members dimerize is through their ETS domain. ETS-1 is a good example of this as its DNA binding domain (ETS domain) has additional helices in the N-terminus that act as an auto-inhibitory region. These helices (HI-1 and HI-2) will unfold as ETS-1 binds cooperatively as a homotopic dimer to DNA in a palindromic head to head fashion offsetting auto-inhibition (139). Experiments with a short form of ETS-1 lacking the auto-inhibitory regions proved that the protein does not form a homotopic dimer meaning that the N-terminal elements are crucial for the cooperativity mechanism of ETS-1 (140).

Another family member that utilizes the ETS domain to dimerize is GABPA. It has two partners GABPB-1 and GABPB-2, to form a hetero-tetramer bound to DNA (141). The tetrameric interface forms a GABPA-$(GABPB)_2$- GABPA. However, the structure of this trimeric complex, as well as knowledge of its ability to distinguish specific sites in promoters, have not been determined yet.

The dimerization of the ETS members has been proven to regulate DNA binding as well as phosphorylation mechanisms. However, thus far, the only members characterized have contained additional well-structured regions, either adjacent to the ETS domain (ex: HI-1/HI-2 of ETS-1) or the whole PNT domain. It is in our interest to study members that only have the ETS domain structured and the rest of the protein intrinsically disordered. Will the intrinsically

disordered regions of these proteins have the same effect as the structured regions in mediating dimerization and therefore regulating gene expression?

1.5 Transcription factor PU.1

ETS family member PU.1 was first discovered in 1988 (142). It encodes for the *spi1* gene in humans and the *spfi1* gene in mice. (131). Its name derives from PU-box binding-1 as the protein binds to a purine-rich region "PU-box" of nucleotides GGAA (143). PU.1 has 270 residues in humans [Figure 1.6]. The ETS domain is made of ~85 residues near the C-terminus, while adjacent to the ETS domain on the N-terminus is the PEST domain that is composed of 48 residues. This is followed by a Q-rich region of about 35 residues and the transactivation region (TAD) of 80 residues. PU.1 only has its ETS domain as a well-folded domain, while the rest of the protein is disordered and tethered around the DNA binding domain.

1.5.1 Biological role PU.1 transcription factor and diseases

A knockout technique of the *spfi1* gene has been used to study the function of PU.1. Many knockout alleles of the *spfi1* gene have been generated and analyzed throughout the years (144-149), demonstrating that PU.1 is critical to generate B cells, macrophages, dendritic cells, and neutrophils. PU.1 is partially critical in generating T cells and erythroid cells and not important at all for the production of megakaryocytes. In addition to the knockout experiments, "knock-in" experiments have also been performed in the *spfi1* locus in mice to analyze the expression of PU.1. (112,150-153). These studies showed that PU.1 expressed the highest in the myeloid linage in a tissue-specific manner. PU.1 expression is controlled in a dosage depended manner increasing during granulocytes-macrophages differentiation and decreasing during megakaryocytes and erythrocytes. In summary, PU.1 plays a critical role in regulating the differentiation of hematopoietic lineages (lymphoid and myeloid), and it's highly expressed in immune systems

cells, particularly for the production of B cells (154). Disruption of PU.1 regulation leads to many diseases such as lymphomas, leukemia, myeloma, Alzheimer's disease, and tissue fibrosis (155), making PU.1 an attractive therapeutic target. Thus far, the Poon lab, in collaboration with other labs, has worked at pharmacological mediations in targeting PU.1 with therapeutic compounds for fibrotic diseases and acute myeloid leukemia (156-158).

1.5.2 Hydration in PU.1

Emerging evidence also links the ETS family to the hydration theory with the heterogeneity of DNA targets (159). Many ETS members DBD have been solved by crystallography bound to their respective high-affinity DNA and deposited in the protein data bank (PDB). Interestingly when observing these co-crystals, we can see that some proteins such as PU.1 have a high abundance of interfacial water-mediated contacts in comparison to other co-crystals such as ETS-1 (143,160). The difference between PU.1 and ETS-1 continues as the two proteins are the most phylogenetically distant class of ETS proteins (109). Considering that ETS-1 is a more ancient protein (161) and it has fewer water contacts between its DBD and DNA while PU.1 has a large amount of water-mediated contacts suggests that hydration could be an evolutionary trait of the ETS family. The water-mediated contacts by PU.1 are replaced by direct contacts by ETS-1. This fact raises the question of whether PU.1 has a regulatory mechanism that involves hydration.

The Poon lab, in collaboration with the Wilson lab, has worked on this same question for many years to establish how hydration plays a role in PU.1 regulation. Experiments in which both PU.1 and ETS-1 bound to DNA (high affinity and low affinity) undergo osmotic stress by various physiologically compatible osmolytes have been performed (162) as well as swapping structural elements of one protein to another (78) to assess their reaction to osmotic stress. The data has

revealed that PU.1 is indeed osmotically sensitive when bound to high-affinity DNA while ETS-1 is not.

This brings us to a new question regarding PU.1 hydration mechanism: what residues are responsible for the osmotic sensitivity? In chapter 2 we address this same question as both PU.1 and ETS-1 (along with the whole family) have a handful of residues that have the same orientation when bound to DNA. We wanted to mutate a highly conserved residue in the family to see the effects of hydration in both paralogs.

1.5.3 Dimerization of PU.1

PU.1 structure bound to DNA (1:1) has already been solved (143) and it is presumed to be the active state of PU.1. However, previous work by the Poon lab has shown that PU.1 can dimerize at a single cognate DNA binding site (2:1) (163,164). This property of PU.1 dimer on one single cognate DNA is interesting as no other ETS member has demonstrated it thus far. Biophysical data indicates that PU.1 forms a sequential dimer as a potential negative feedback mechanism in the absence of inhibitory cofactors such as GATA-1 (164). This information gives an insight into how PU.1 self-regulates its activity and a new way of ETS proteins controlling their genes. In Chapter 3 we demonstrate how this self-sufficient mechanism regulates PU.1.

1.5.4 Using X-Ray Crystallography for protein structure

Over 125 years ago Wilhelm Roentgen discovered X-rays and almost two decades later Max von Laue showed x-ray diffraction patterns of crystals (165). Soon after William Lawrence Bragg derived the equation that describes the principle of image formation by X-Ray diffraction (166). William Lawrence Bragg and his father William Henry Bragg won a Nobel Prize in Physics in 1915 *"For their services in the analysis of crystal structure by means of X-rays."*

$$n\lambda = 2d \sin \theta \qquad (10)$$

[Figure 1.8A] shows the diagram of Equation 10 where two waves in phase with each other are separated by a distance (d) between the atomic planes. Crystallography is a technique that is highly favorable in getting a molecular structure from a crystal that diffracts when exposed to an X-ray beam. The diffracting pattern is used as a map to determine the symmetry and the size of the protein. The intensity of the spots diffracted help construct an electron density map (heavy atoms have more density) to build a molecular structure of the protein from amino acids. Finally, the structure is refined to adopt the most favorable conformation (167).

X-ray crystallography has developed into one of the most powerful structure determination tools for protein, DNA, viruses, and organic/inorganic material. This year alone (2020), there are more than 7000 x-ray structures deposited in the Protein Data Bank: 5x more than electron microscopy (EM) and 29x more than NMR.

For the purpose of this thesis I am going to focus on protein: DNA complexes. The first step of getting crystals is the purification and screening techniques of the right condition. In [Figure 1.8B] the Art Robbins Gryphon LCP crystallization system uses the Index and JCSG screening kits from Hampton Research to screen potential conditions for crystallization of PU.1/DNA complex. We set out to crystalize PU.1 and DNA in a 1:1 complex at a higher resolution then previously resolved for in depth data analysis. I will cover the findings and significance of the co-crystal structure in Chapter 4 for more detail. After finding the correct conditions we set to optimize them for better results utilizing the hanging-drop method (168) [Figure 1.8C-D]. Days after the hanging-drops were set, crystals started to form [Figure 1.8E]. An interesting fact should be mentioned at this point that there is no correlation between the appearance of the crystal under the microscope with its capability to diffract (168).

To acquire the data the crystal gets bombarded by X-rays from synchrotron radiation which can vary in wavelengths from 0.5- 3.0 Å, depending on the size of the crystal (168). The diffraction is recorded by a detector and then is subjected to data analysis. From the diffraction pattern one can determine the unit cell dimensions which is the smallest repeating unit. The measurements of the unit cell are calculated from the spacing of the spots from the diffraction pattern and given in lengths (a, b and c) and angles (α, β and γ) (167). The data then gets scaled according to amplitude (to relate the intensities of each data point) (167) and then phased by isomorphous replacement (169) or the molecular replacement method (170). For the PU.1:DNA complex, X-ray Detector Software (XDS) is used for amplitude scaling and phasing was done by molecular replacement in Python-based Hierarchical ENvironment for Integrated Xtallography (Phenix). The amplitude and phasing data is then used to reconstruct an electron-density map which is limited by its resolution. Resolution is dependent upon the distance between the atomic planes [e.i Figure 1.8A], the crystal packing, the brightness of the x-ray radiation source, and the detectors sensitivity (171).

It is important to remember that crystal structures are snapshots of a refined average conformation frozen in time (165). In biological solutions these proteins and/or protein: DNA complexes are more dynamic and could occupy more distinct conformations. Here, is where crystallography fails to determine the dynamic nature of macromolecules. Along that line intrinsically disordered regions (IDRs) of proteins do not show up on the density map due to their intrinsic flexibility. In these cases, NMR techniques are used to determine the dynamic solution structure of proteins and/or molecular dynamics (MD) simulations to capture conformational changes (165,172).

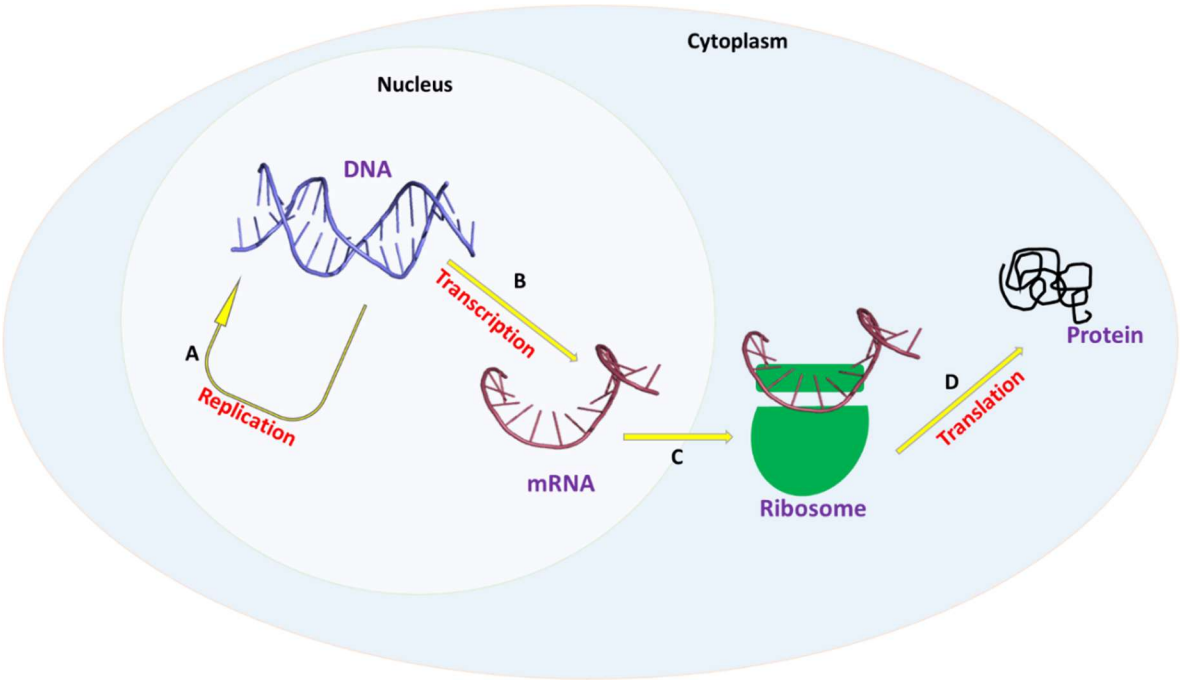


Figure 1-1 The central dogma of molecular biology

A The replication step is when a cell is involved in making a copy of its genomic information before it divides. B, the transcription step is in which the information of the gene required will be transcribed into portable messenger RNA (mRNA) by RNA polymerase II. C, the portable mRNA leaves the nucleus for the cytoplasm with the destination to the ribosome. D, the ribosome reads the mRNA and translates the information into making proteins. These proteins will undergo post-translation modification before carrying their function into the body

.

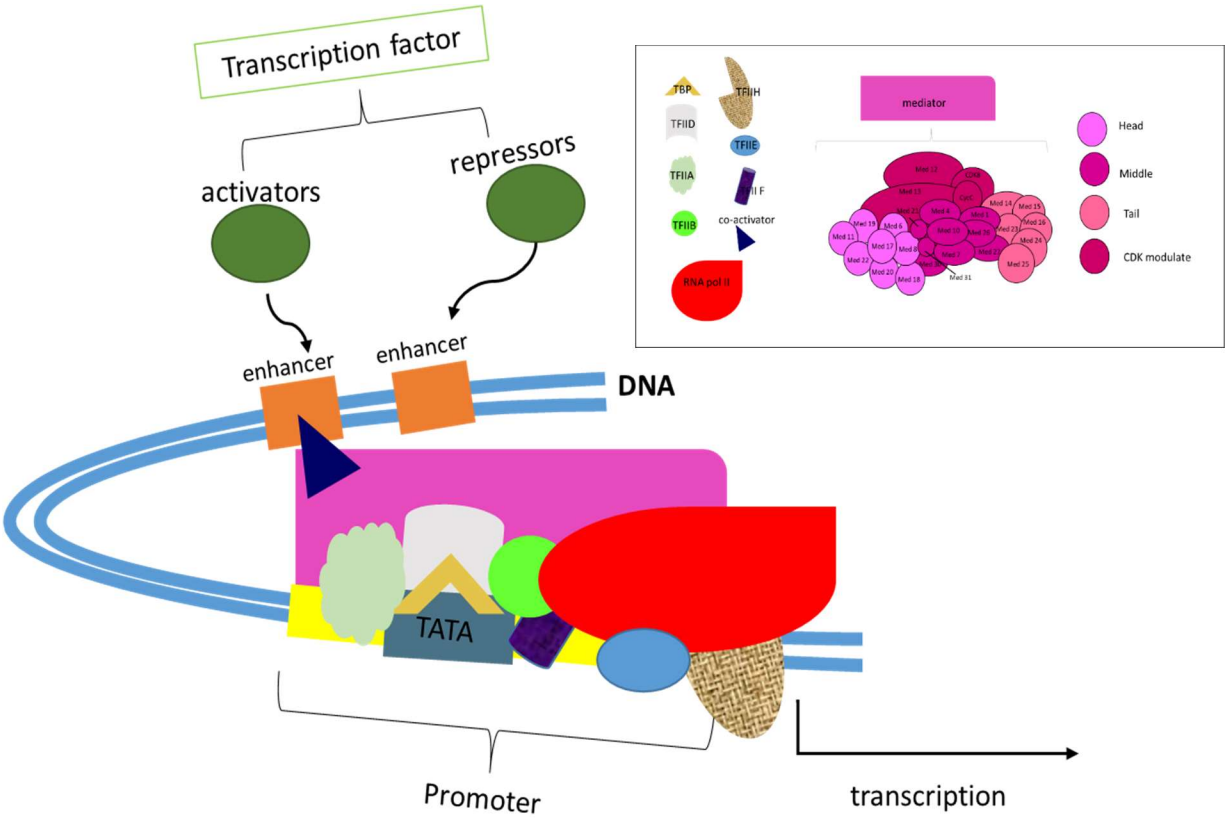


Figure 1-2 Initiation of transcription

The first step in transcription is called initiation, it involves a variety of transcription elements. Once the signal is given and a transcription factor (either activator or repressor) binds to its target DNA binding region called the enhancer region (orange) will then set in motion the initiation cascaded of events. The TATA-binding protein (TBP in mustard yellow) will bind to the TATA region (dark blue) that is located in the promoter (~100 bp in bright yellow) and will induce a 90° bend in the DNA. This then recruits the general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (color breakdown insert in the figure) which will bind to specific promoter sequences to recruit RNA polymerase II (red) to initiate transcription. A mediator (purple) is then added to maintain the pre-initiation complex (General TF and RNA pol II) to transmit signals between the transcription factors and the polymerase (173) This figure was made by inspiration from(174) (175)

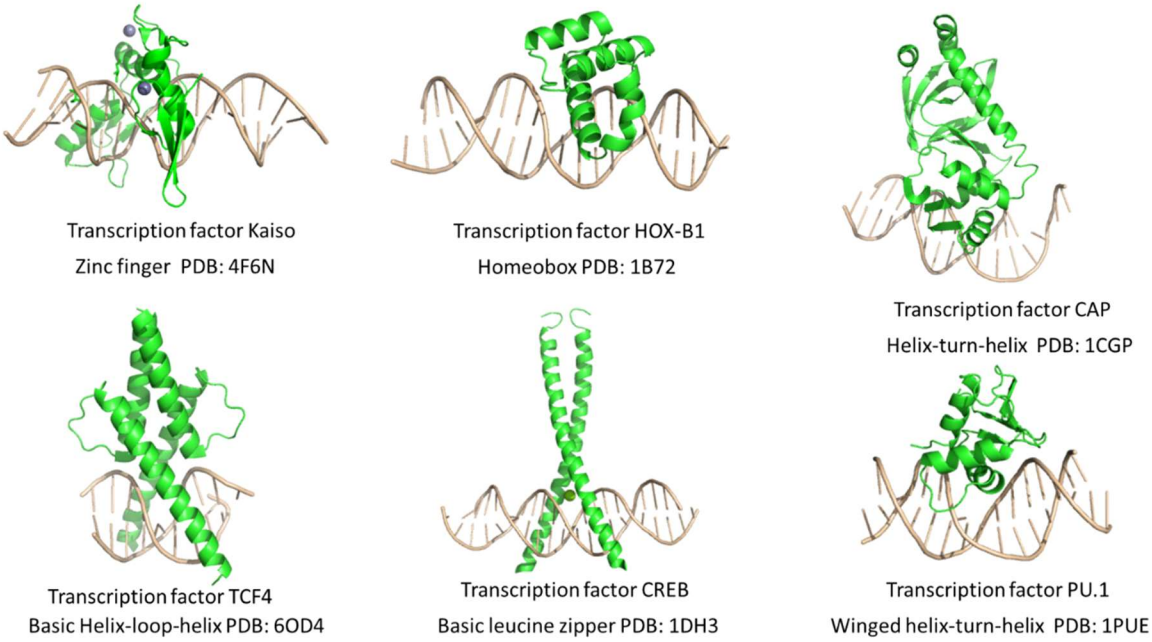


Figure 1-3 The most common transcription factor DNA binding motifs

Representation of the different motifs is in (green) bound to their respective DNA (light brown). The Zinc-finger representative is transcription factor Kaiso which is stabilized by a Zn^{+2} ion (blue) that gives the family its name. The second transcription factor is HOX-B1 that belongs to the Homeobox family. The whole domain is made of the three-helices bundle, two are parallel to one another and the third recognizes DNA. The next representative is transcription factor CAP with a helix-turn-helix motif. This motif involves the helix that recognizes DNA and a turn that connects it to another helix. The following representative is transcription factor TCF4 belonging to the basic helix-loop-helix motif. This motif is identified by two helices mediating DNA recognition and connected by a loop. Next is the transcription factor CREB with the basic leucine zipper motif. The interaction of the two helices is held in place by leucine residues that form hydrophobic stable interactions when bound to DNA. Lastly, there is the transcription factor PU.1 representing the winged helix-turn-helix motif. The characteristic of this motif is the insertion of one helix into the major groove and it is connected by a turn with another helix and connected on the other side by a beta-sheet creating a floppy "wing". All the structures were taken from the PDB (176) and the figure was created using Pymol (The PyMOL Molecular Graphics System, Open Source, Schrodinger, LLC

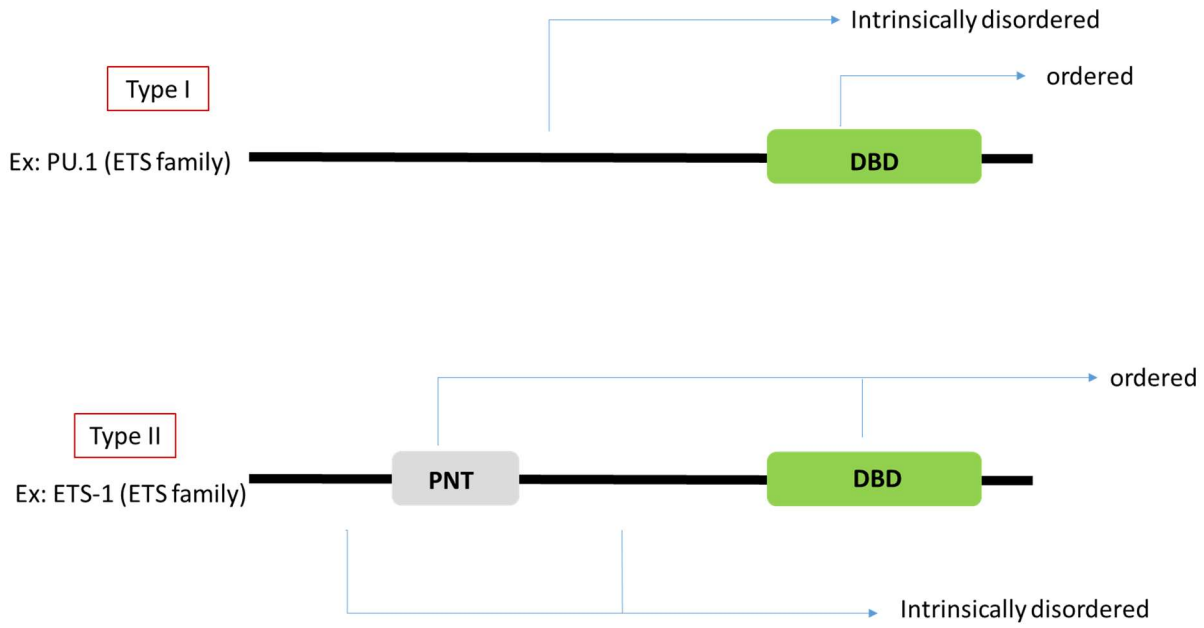


Figure 1-4 Transcription factors have an abundance of intrinsically disordered regions
A schematic representation of a narrowing down the different types of transcription factors. Type I incorporates the transcription factors that only have their DBD well-structured and the rest of the protein is disordered both on the N-terminus and the C-terminus of the DBD. The example shown is of PU.1 a transcription factor belonging to the ETS family. Type II incorporates all the transcription factors that have another or many other domains that are well structured in addition to their DBD. The example given is of ETS-1 of the ETS family indicating that proteins of the same family can have different architecture. Figure inspired by (45)

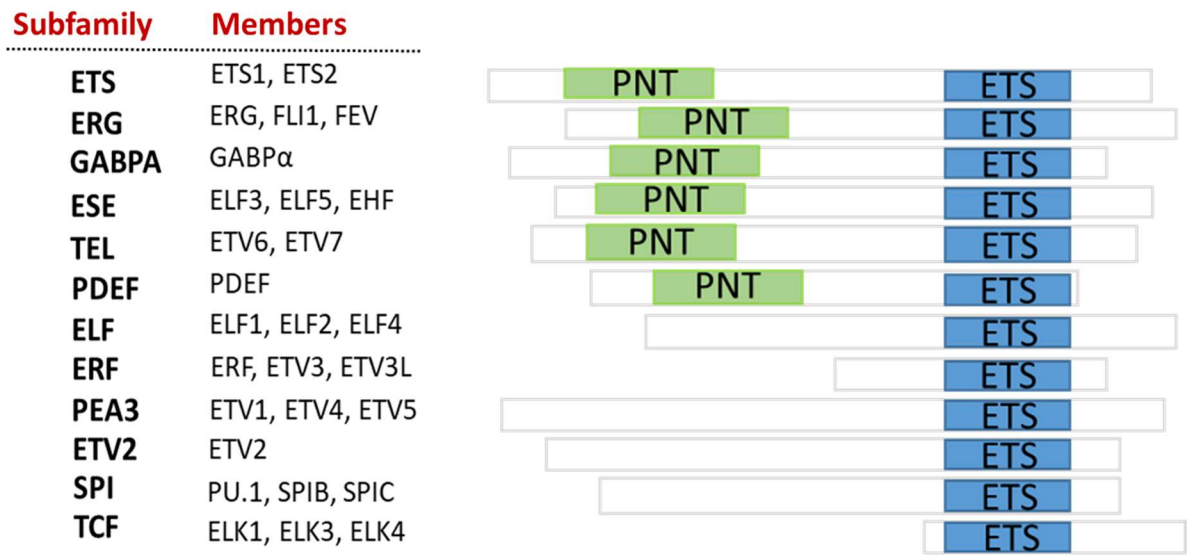


Figure 1-5 The ETS family of transcription factors

The representation of the ETS family demonstrated by the 12 subfamilies and the 28 members. The schematic representation of the length and the domain position of the ETS in blue and the PNT in green have been inspired by (109)

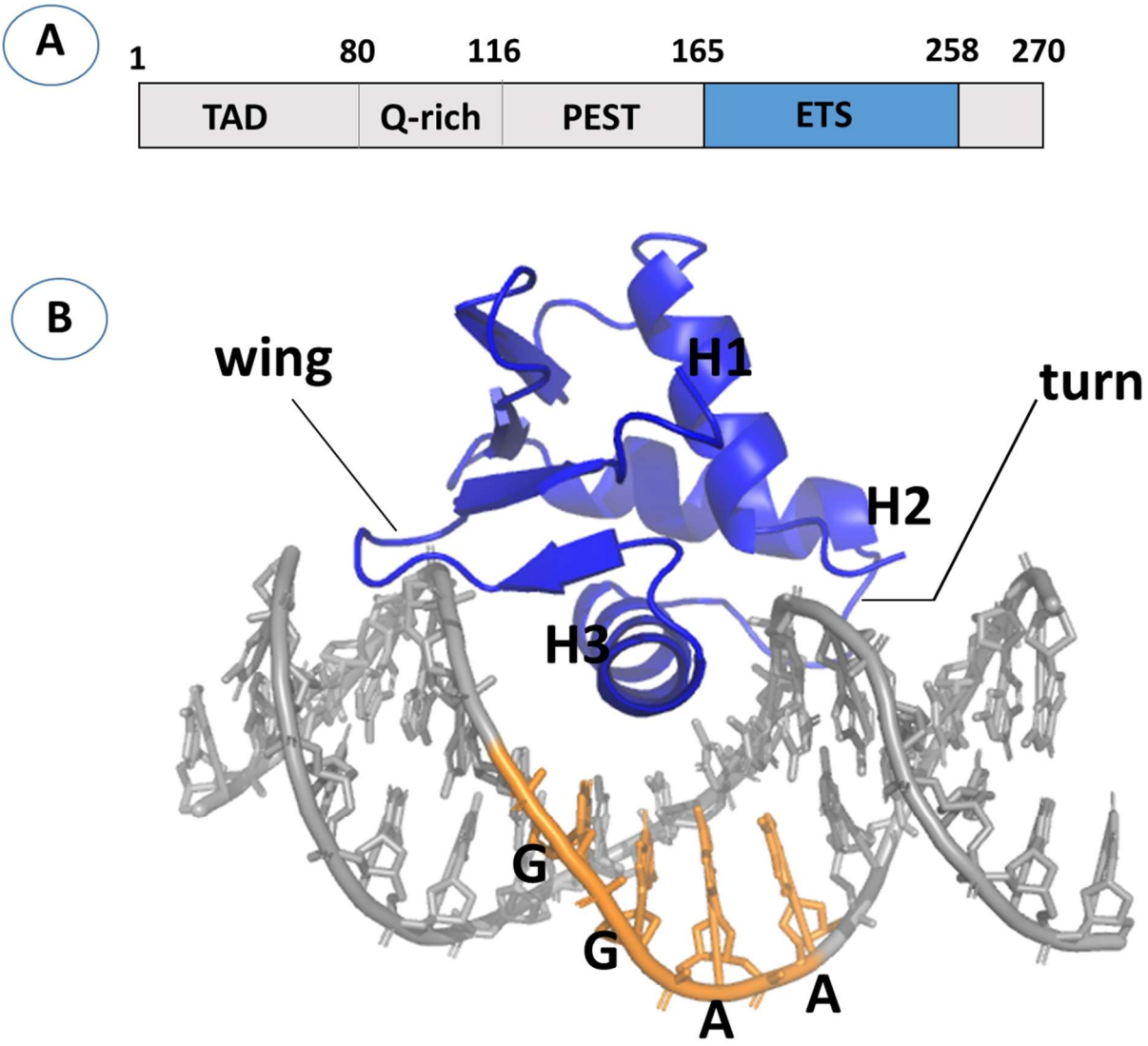


Figure 1-6 A closer look at transcription factor PU.1

A is the detailed schematic of PU.1 representing the different regions of the protein. In blue is the ETS domain and in gray are the disordered regions divided by the specific names of the domains, TAD, Q-rich, and PEST domain (155). B is the zoomed-in PDB of PU.1 (1PUE) (143) that has a wHTH motif as the Helix 3 is inserted into the major groove making contact with the consensus GGAA nucleotides (orange)

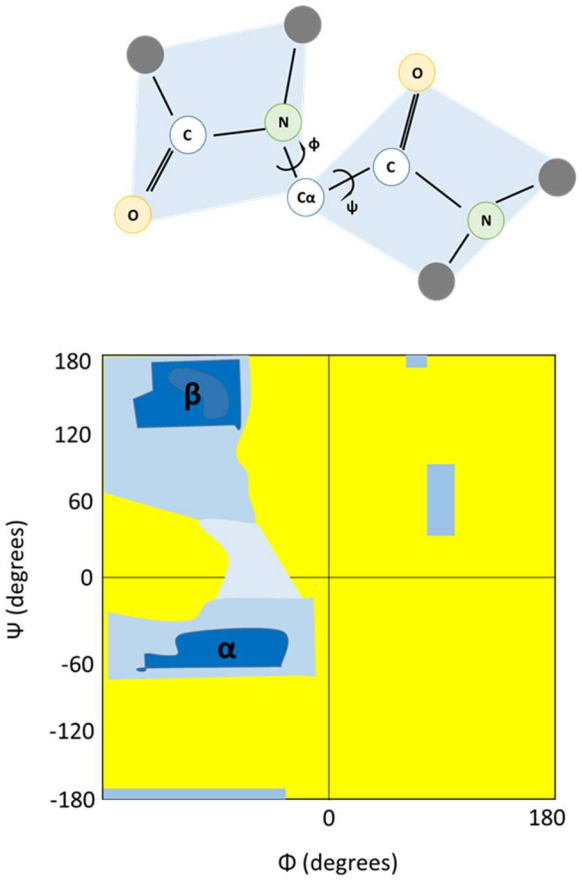


Figure 1-7 Protein dihedral angles

The peptide backbone can rotate by the N-Cα bond called ϕ and by the Cα-C bond called ψ. The bond angles of the proteins can be plotted in Ramachandran plot deriving its name from the scientist that discovered it, G. N Ramachandran, in 1963. Torsion angles are among the most important feature indicating protein folding. The torsion angles determine the conformation of the residues (in blue) where the typical secondary structure of alpha-helix and beta-sheet clusters of the protein would be. The lighter blue regions indicate residues that are allowed to come closer together and yet they don't have a defined secondary structure (random coil). Many conformations however are not possible due to steric hindrance (yellow). The figure was inspired by (177)

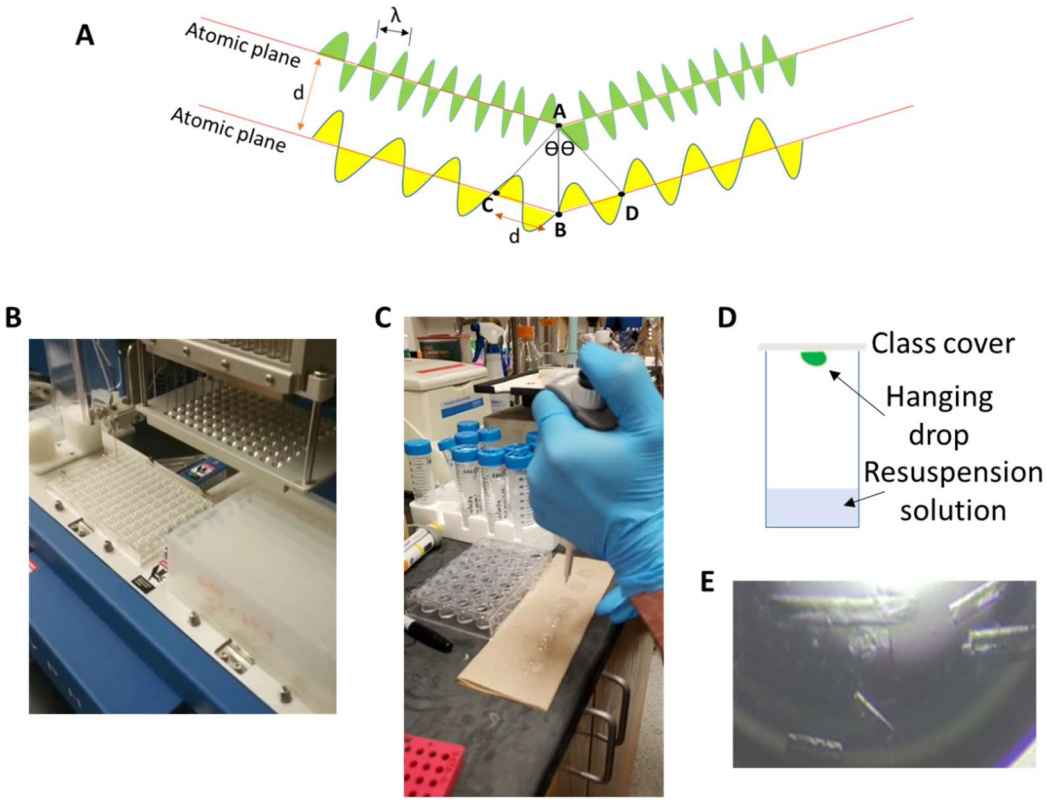


Figure 1-8 PU.1/DNA complex crystals settings

A: The diagram of two waves in phase with a distance (d) between the atomic plates. The angle Θ is the reflected angle and the path between points C, B and D be a whole number (n= 2 above) of wavelengths (λ). The distance from C-B is the same as B-D therefore we get Braggs equation nλ= 2dsin Θ. The figure is inspired from Encyclopedia Britannica. B. The robot screening conditions for PU.1: DNA complex. C. Ross Terrell (collaborator from Ming Luo's lab) setting hanging drops of PU.1: DNA complex. D. Cartoon of vapor diffusion hanging-drop mode. E. Crystals of PU.1: DNA complex

2 DISTINCT ROLES FOR THE INTERFACIAL HYDRATION IN SITE-SPECIFIC DNA RECOGNITION BY ETS-FAMILY

 [The Journal of Physical Chemistry B 121, issue 13, pages 2748-2758] 15 March, 2017 DOI: 10.1021/acs.jpcb.7b00325

2.1 Abstract

The ETS family of transcription factors is a functionally heterogeneous group of gene regulators that share a structurally conserved, eponymous DNA-binding domain. Unlike other ETS homologues, such as Ets-1, DNA recognition by PU.1 is highly sensitive to its osmotic environment due to excess interfacial hydration in the complex. To investigate interfacial hydration in the two homologues, we mutated a conserved tyrosine residue, which is exclusively engaged in coordinating a well-defined water contact between the protein and DNA among ETS proteins, to phenylalanine. The loss of this water-mediated contact blunted the osmotic sensitivity of PU.1/DNA binding, but did not alter binding under normal-osmotic conditions, suggesting that PU.1 has evolved to maximize osmotic sensitivity. The homologous mutation in Ets-1, which was minimally sensitive to osmotic stress due to a sparsely hydrated interface, reduced DNA-binding affinity at normal osmolality but the complex became stabilized by osmotic stress. Molecular dynamics simulations of wild type and mutant PU.1 and Ets-1 in their free and DNA-bound states, which recapitulated experimental features of the proteins, showed that abrogation of this tyrosine-mediated water contact perturbed the Ets-1/DNA complex not through disruption of interfacial hydration, but by inhibiting local dynamics induced specifically in the bound state. Thus, a configurationally identical water-mediated contact plays mechanistically distinct roles in mediating DNA recognition by structurally homologous ETS transcription factors.

2.2 Introduction

The ETS family of transcription factors binds site-specific DNA via eponymous, structurally conserved DNA-binding domains that share low sequence homology. To date, a plethora of co-crystals of site-specific binary ETS/DNA complexes, as well as ternary structures in combination with other protein binding partners, show a highly conserved binding mode in which a recognition helix is inserted into the major groove of target DNA harboring the core sequence 5′-GGAA/T-3′, with additional interactions along the DNA backbone at flanking minor groove positions. Despite the apparent homogeneity at the macromolecular level, heterogeneous levels and patterns of hydration pervade the protein/DNA interface. For instance, in the cocrystal of the PU.1 ETS domain with site-specific DNA, the contact interface is densely hydrated with several interfacial residues engaging in exclusively water-mediated contacts with the target DNA (143). In contrast, the cocrystal structure for Ets-1 shows a sparsely hydrated interface where most of the corresponding residues make direct contact with the DNA (160). The differences in crystallographic hydration between the two ETS domains, whose backbone trajectories are superimposable in the DNA-bound complex, have been reproduced in solution studies that perturbed ETS/DNA binding by osmotic pressure (162,178). These hydration differences are profoundly correlated to their binding kinetics, conformational dynamics, and site discrimination (179). In turn, target discrimination by ETS homologues, all of which share overlapping DNA sequence preferences (110), continues to be a major area of interest in understanding how biological specificity is achieved in vivo (180,181). Since the topology of the ETS/DNA contact interface is highly conserved among ETS homologues, differences in conformational dynamics are presumably important in determining the hydration of ETS/DNA complexes. While such differences have been observed globally in ensemble experiments by dynamic light scattering,

structurally resolved data do not yet exist. To address the correlation between interfacial hydration and conformational dynamics of ETS/ DNA complexes, we targeted conserved ETS residues involved only in water-mediated contacts and not any other direct interactions with either itself or the DNA. In PU.1, the side chain −OH of Tyr252 coordinates a bridging water with a universally conserved arginine (Arg235) and a backbone phosphate of the target DNA. This is a universally conserved configuration of interfacial hydration in all ETS homologues, including Ets-1, that harbor this equivalent Tyr. We mutated this Tyr in both PU.1 and Ets-1 and compared their DNA site recognition under osmotic pressure as well as in silico by molecular dynamics simulations. The two orthogonal approaches provide a deeper insight into the mechanism of DNA recognition with respect to the absence of extensive interfacial hydration of Ets-1 (and sequence-similar homologues such as Ets-2, Fli-1, and GABPα) versus PU.1 (and its phylogenetic relatives such as ETV6).

2.3 Results

Despite a well-defined structural homology in terms of backbone trajectories, and overlapping DNA sequence tolerance, the DNA-binding domains of ETS transcription factors are divergent in amino acid sequence. Of the limited set of residues that are strictly or highly conserved, most are engaged in direct protein-to-DNA contacts, and alanine scans typically abolish DNA binding. The role of other conserved residues, such as an interfacial tyrosine near the C-terminus in the primary sequence of ETS domains [Figure 2.1A], has been less clear. This tyrosine is found in 26 of the 28 known murine (and human) ETS homologues, and all co-crystals of these homologues show a water-mediated contact coordinating the side chain −OH of the tyrosine, a universally conserved arginine in the recognition helix, and the backbone phosphate of the base upstream from the 5′-GGAA-3′ core consensus. Moreover, this well-defined water-mediated

"motif" is conserved whether the interface is otherwise extensively hydrated (such as in PU.1) (143) or not (such as in Ets-1) (160) [Figures 2.1B−G].

The strong conservation of this exclusively water-coordinated tyrosine among ETS homologues suggests that it plays an important role in mediating ETS/DNA binding. In the cocrystal structure of PU.1, this tyrosine anchors one end of an extensive interfacial water network, and its mutation to phenylalanine (loss of the side chain −OH) represents an attractive approach to probing this water network. We therefore mutated this Tyr to Phe in the ETS domain of PU.1 (PU.1ΔN167) and Ets-1 (Ets-1ΔN331) [Figure 2.2A], and compared their recognition of optimal site-specific DNA under normal-osmotic and osmotically stressed solution conditions relative to their wild type counterparts

In previous studies,(162,178) we have extensively characterized the perturbative effects of compatible co-solutes on DNA binding by the ETS domains of both PU.1 and Ets-1. Using a palette of co-solutes spanning a broad range in physical chemistry, such as betaine, disaccharides (sucrose, maltose), and glycols (triethylene glycol), we observed that their effect on site-specific binding affinity correlated only with their osmolality over an extended range of co-solute concentrations. The key conclusion from this colligative behavior is that the co-solutes did not exhibit co-solute specific preferential interactions with the macromolecules and that their perturbations on binding could be interpreted in terms of hydration changes upon binding..(93,100,182,183) For our studies here we used betaine, a physiologically compatible osmolyte, out of convenience given its relatively neutral effect on bulk viscosity and dielectric constant.

We carried out competitive fluorescence polarization titrations of ETS-bound, TYE 563-labeled DNA probe with an unlabeled competitor (harboring the optimal sequence for each

respective protein) with or without graded concentrations of betaine [Figure 2.2B]. The data was analyzed with a mechanistic model (184) to estimate the absolute binding constant (not IC50's) for the unlabeled complex of interest. The in-solution affinities for wild type PU.1 and Ets-1 show quantitative agreement with previous values obtained by electrophoretic mobility shift (162,178) and surface plasmon resonance (185,186).

2.3.1 Mutation of a Conserved Interfacial Tyrosine Inducing Markedly Different Effects on DNA Recognition by the Structurally Homologous ETS Domains of PU.1 and Ets-1

Measurement of the apparent dissociation constants for the wild type and mutant PU.1ΔN167 and Ets-1ΔN331 showed that both mutants retained high-affinity binding to their wild types' optimal DNA targets ($K_D < 10^{-8}$ M) under normal-osmotic conditions (150 mM Na^+, 0.29 osmolal). However, osmotic stress to 2.5 osmolal showed distinct mutational effects on the affinity and osmotic sensitivity of DNA recognition by the ETS homologues.

In the case of PU.1, both wild type and mutant protein were negligibly different in their affinities for high-affinity site-specific DNA, but the mutant bound its target more strongly than wild type with increasing osmotic pressure. Over the measured range of osmotic pressure, the osmotic dependence (slope) of the binding affinity was ~25% lower in magnitude for the PU.1ΔN167Y252F mutant than wild type PU.1ΔN167 [Figure 2.2C], consistent with a deficit of interfacial hydration in the mutant/DNA complex relative to wild type. Quantitatively, the effect was more pronounced than if the nine water-mediated contacts seen in the cocrystal structure contributed equally to the binding free energy. In stark contrast to PU.1, the ETS domain of Ets-1 exhibited altogether different behavior when probed by osmotic pressure. At normal-osmotic pressure, binding of Ets-1ΔN331Y412F to high-affinity site-specific DNA was ~60-fold weaker

relative to wild type, although still >10-fold stronger than nonspecific binding (187). As observed previously, (162) unlike PU.1, the site-specific complex of wild type Ets-1ΔN331 was slightly stabilized by osmotic pressure. In comparison, the osmotic dependence of the affinity (slope) for Ets-1ΔN331Y412F was ∼35% more positive than wild type, inferring relative dehydration of the mutant Ets-1/DNA complex relative to wild type.

As binding affinity reflects the free energy change of the components from the unbound to bound state, differential perturbations of the mutations on the unbound proteins may also contribute to heterogeneity in DNA binding. Therefore, we probed wild type and mutant PU.1ΔN167 and Ets-1ΔN331 in the absence of DNA by circular dichroism spectroscopy to determine the secondary structure contents of each construct at 25 °C under normal-osmotic conditions [Figure 2.3A] and measured their stability by thermal melting [Figure 2.3B]. Decomposition of their CD spectra showed an apparent loss of 15 to 20% β-sheet in both mutants relative to wild type [Table 1], which was the secondary structure at the mutated positions. Despite this difference, wild type and mutant PU.1ΔN167 unfolded reversibly with experimentally indistinguishable melting temperatures and enthalpy changes. In contrast, Ets1ΔN331Y412F unfolded at a higher melting temperature than wild type by ∼8 °C. Although thermal denaturation was not reversible for either Ets-1 construct, as previously noted (188), the alignment of the signal in the cooling run with the post transition baseline of the heating run for both Ets-1 constructs indicated that irreversible change occurred after denaturation. Since Ets-1 harbored free cysteine (but not PU.1), we confirmed by size-exclusion chromatography that the mutations did not cause protein oligomerization under the solution conditions of the CD experiments [Figure 2.3C]. All wild type and mutant constructs eluted quantitatively as a single species according to their nominal molecular weight (∼13 kDa, at a higher elution volume than myoglobin at 17 kDa). To further

probe the structural differences between wild type and mutant Ets-1ΔN331, we acquired and compared their $^{1}H-^{15}N$ HSQC NMR spectra [Figure 2.3D], which showed complete or partial overlap of >90% of resolved resonances. Using the solution NMR structure of Ets-1ΔN301 (188), about half of the peaks showing the largest shifts (partial or complete) are ±20 residues from the mutated position (Y412). Thus, wild type and mutant Ets-1ΔN331 showed substantially similar folding but exhibited some local and distal conformational changes in the unbound state.

In summary, one-atom mutation of a highly conserved interfacial Tyr to Phe produced distinct effects on the DNA binding and conformational properties of the ETS domains of PU.1 and Ets-1. This heterogeneity was unexpected given the strong structural homology of the two ETS domains in general (109), an identical configuration in coordinating a water-mediated contact in their cocrystal structures [cf. Figures 2.1F and 2.1G], and the similar solvent-accessible topology of the mutated Tyr.

2.3.2 Molecular Dynamics Simulations of Wild type and Mutant ETS Domains of PU.1 and Ets-1 Recapitulated Experimental Data.

While the free (188) and DNA-bound (160,189) structures of the Ets-1 ETS domain are known, only the DNA bound PU.1 structure1 currently exists. To understand the experimental differences among - and mutant PU.1 and Ets-1 in greater detail, we carried out all-atom molecular dynamic simulations of the four constructs in their free and site-specifically DNA-bound states. Unbound Ets-1 was taken from a crystallographic structure (PDB: 1GVJ) and unbound PU.1 from the cocrystal structure. The protein/DNA complexes were constructed by modifying the cocrystal structures to match the experimental sequences. All crystallographic water molecules were retained and treated as TIP3P water. Following equilibration in explicit solvent containing 0.15 M NaCl, and confirming that the crystal waters at the interface in the complexes had not exchanged

with bulk solvent, each system was simulated to convergence as indicated by RMSD from the initial coordinates [Figure 2.4, parts A and B, for unbound proteins, 100 ns]. Enumeration of H-bonds with solvent showed identical hydration between wild type and mutant unbound PU.1 (within 1%) in agreement with their similar melting temperatures and unfolding enthalpies [Figure 2.4C]. However, wild type Ets-1 showed >4% (average) more solvent H-bonds than the mutant, indicating less solvent exposure in the mutant [Figure 2.4, parts C and D] and consistent with its relative thermal stability over wild type Ets-1 [cf. Figure 2.3B]. In addition, the shallower slope of the mean square displacement with respect to time (which is proportional to the self-diffusion constant) for wild type Ets-1 (0.2×10^{-9} m^2 /s) relative to wild type PU.1 (0.6×10^{-9} m^2 /s) [Figure 2.4, parts E and F] recapitulated published dynamic light scattering measurements showing a 14% larger hydrodynamic radius for Ets-1ΔN331 over PU.1ΔN167 (179).

In the case of the DNA-bound complexes (200 ns simulations), both wild type and mutant complexes formed by PU.1 with its 14-bp DNA remained well-behaved through the simulation. However, the wild type complexes formed by Ets-1 with a 14-bp site-specific target exhibited significant fraying at the DNA ends and interactions between the protein and the frayed ends. To account for this, both wild type and mutant Ets1 complexes were re-simulated with longer DNA duplexes matching the full 23-bp experimental construct. Bases proximal to the protein remained duplex throughout the simulations with the longer DNA. Although the wild type and mutant complexes for both ETS homologues did not differ significantly in terms of RMSD from the initial coordinates [Figure 2.5A], there was a substantial bias in dynamic fluctuations in favor of wild type over mutant Ets-1 complexes. Comparison of the per-residue RMS fluctuations (RMSF; standard deviations from average positions) of the DNA-bound states indicated significant local dynamics in wild type Ets-1 over the Y412F mutant [Figure 2.5B]. The average excess RMSF

(wild type over mutant) was +0.025 nm per residue for Ets-1 and lower than +0.01 nm for PU.1. To evaluate the validity of the simulated dynamics, we examined the available solution NMR structures of free (188) and DNA bound Ets-1 (189). The experimental structures show excess dynamics in DNA-bound Ets-1 that are localized in elements contacting the minor groove DNA backbone flanking the 5′- GGAA-3′ core consensus, as well as two loops distal from DNA binding interface [Figure 2.5C].[1] Although the agreement was not quantitative, the simulations correctly reproduced the presence and positions of the induced experimental dynamics in DNA-bound Ets-1. These dynamic hotspots were discontinuous in primary sequence, clustering primarily in the N-terminal half and distally from C-terminal position of Tyr^{412}, indicating that the dynamics were transmitted through the protein's tertiary structure and probably through the DNA "substrate." In summary, the simulations' agreement with the experimental data for wild type PU.1 and Ets-1 reported here and by others provided evidence that the simulations semi quantitatively described the solution behavior of the proteins in their free and bound states.

2.3.3 Hydration Dynamics in the Wild type and Mutant Phe-to-Tyr PU.1 and Ets-1 Complexes.

To probe interfacial hydration in the simulated structures, H-bonds formed by solvent to protein and nucleotide bases participating in water mediated contacts in the contact interface were enumerated for each system based on combined distance and angle criteria from triplicate independent production runs [Figure 2.6]. The hydration of the tripartite water-mediated contact in the crystal structures was evaluated using the functional groups involved: the side chain −OH of Tyr^{252}/Tyr^{412} (PU.1/Ets-1), the guanidinium N of Arg^{235}/Arg^{394} (PU.1/Ets-1), and the backbone phosphate at the −1 position (relative to the 5′- GGAA-3′ core consensus). To assess interfacial

[1] See also Addendum at the end of this Chapter, Section 2.7.

hydration more generally, the number of solvent H-bond contacts made by guanidinium N of Arg^{235}/Arg^{394} and the major groove-facing heteroatoms of the nucleobases in the 5′-GGAA-3′ core consensus (i.e., O^6 and N^7 for G; N^7 and N^6 for A; O^4 in T, and N^4 in C) were also calculated. While these groups represent only a subset of all the interfacial residues, they are fully conserved between PU.1 and Ets-1, providing a chemically fair and equivalent basis for comparison between the two systems. Starting from the interfacial hydration observed in the PU.1/ DNA cocrystal structure, the number of solvent H-bond contacts made by the wild type complex remained steady throughout the simulation. In particular, the 4-OH group of Tyr^{252} was associated with a time-averaged value of ∼0.9 solvent H-bond, consistent with its role in coordinating the tripartite water-mediated contact with Arg^{235} and a DNA backbone phosphate as observed crystallographically. Without the tyrosyl 4-OH group, hydration of Arg^{235}, the phosphate at the −1 position, and the 5′-GGAA-3′ interface all became partially depleted in the mutant complex. Nonetheless, the deficit of solvent H-bonds for interfacial residues in the mutant PU.1 complex did not significantly exceed the loss (∼1 contact) already accounted for by Arg^{235}, which was present in both sets of residues. Depletion of interfacial hydration in the PU.1/ DNA complex due to the Y252F mutation therefore appeared to be limited primarily to the loss of water associated with the tyrosyl 4-OH group. In the case of wild type Ets-1, the time-averaged hydration at Tyr412 was significantly lower (<0.3 H-bonding contact) relative to PU.1 despite an identical configuration of water-coordinating residues, suggesting that the tripartite water-mediated contact was less persistent in Ets-1 compared to PU.1. In agreement with the cocrystal structures, hydration at the 5′-GGAA-3′ interface in the wild type Ets-1 complex was also lower, at down to half the level observed in the wild type PU.1 complex [cf. Figure 2.1, parts D and E]. However, abolishing the tyrosyl 4-OH contact in the Y412F mutant did not significantly perturb the hydration near the residue or in the ETS-1/DNA interface,

suggesting that, like the mutant PU.1/DNA complex, any dehydration was also limited to the mutated residue.

2.3.4 Structural Interpretation of Interfacial Phe-to-Tyr Mutations on Affinity and Osmotic Sensitivity of DNA Binding.

Osmotic stress experiments showed that loss of a conserved, interfacial H-bond donor in PU.1 and Ets-1 generated altogether different responses to osmotic pressure in their DNA complexes. In PU.1, whose wild type high-affinity DNA complex was profoundly destabilized by osmotic stress, the Y252F mutant was less sensitive. In Ets-1, the wild type complex was essentially insensitive to osmotic stress, but the Y412F mutant became more sensitive in the direction of increased water release. However, the experiments also showed that DNA recognition by PU.1 under norm-osmotic conditions were not affected by the partial loss of interfacial hydration in the mutant, whereas affinity for Ets-1 was reduced. The simulations provided a structural correlation in terms of induced localized dynamics in the wild type Ets-1/DNA complex, which agreed semi quantitatively with experimental NMR data of free and DNA-bound Ets-1 (cf. Figure 2.5C). If induced dynamics in the bound state are specific to Ets-1, as consistent with thermodynamic observations that site-specific Ets-1/DNA binding is more entropically driven than for PU.1 (162), the dampened fluctuations exhibited by the Y412F mutant would reduce the favorable entropic contribution to its binding affinity. Excess dynamics in the DNA-bound state would also contribute to the skewing of osmotic sensitivity toward water release for the Y412F mutant, because hydration of dynamic elements with increased solvent exposure in the wild type complex would be absent in the mutant complex. Thus, the analogous Tyr-to-Phe mutation in PU.1 and Ets-1 reveals different selection pressures at work (osmotically sensitivity for PU.1 vs

dynamics induction for Ets-1) in the divergent evolution of ETS homologues, while maintaining a highly conserved fold characteristic of this family of transcription factors.

2.4 Discussion

Mutation of a single conserved tyrosine that comprises an exclusively water-mediated interfacial contact with site-specific DNA, to phenylalanine, causes markedly heterogeneous effects on the affinities of the ETS domains of PU.1 and Ets-1. In addition to dissimilar effects on DNA recognition under norm-osmotic conditions, osmotic stress reveals disparate disposition of hydration water between the two structural homologues. In this study, molecular dynamics simulations provided a high-resolution interpretation of the experimental data. To this end, the constructs and conditions of the simulations were configured to match experimental (normal-osmotic) conditions as closely as possible using extant structural data as templates. The extensive agreement of the simulations with experimental data reported here and by others from CD spectroscopy [cf. Figure 2.3], dynamic light scattering, (179) titration calorimetry, (162), and solution NMR [cf. Figure 2.5]) provided confidence that the force field (CHARMM36) and the results represent an accurate semi quantitative description of ETS/DNA interactions

2.4.1 PU.1 Tuned to Optimize Sensitivity to the Osmotic Environment in DNA Recognition.

One of the most intriguing features of the PU.1ΔN167^{Y252F} mutant is its recovery from partial loss of interfacial hydration to bind site-specific DNA equally as tightly as wild type PU.1 under norm-osmotic conditions. This unexpected observation suggests that alternative, less hydrated binding modes are accessible to PU.1. However, to what extent are interfacial hydration configurations adaptive? The very few ETS homologues that harbor a Phe at the corresponding Y252 position in PU.1 offer an insight. In the cocrystal structure of ETV6 (190), which binds target

DNA sites equally as well as any other ETS domain (10^{-10} to 10^{-9} M under norm-osmotic conditions), Phe^{395} is the wild type residue at the corresponding Tyr^{252} and Tyr^{412} position in PU.1 and Ets-1 [cf. Figure 2.1A]. Similar to PU.1, the ETV6/ DNA interface is well hydrated, albeit in a different configuration by one fewer than the hydration network in the wild type PU.1/DNA interface. It is also significant that ETV6 is the closest phylogenetic relative of PU.1 (but less distal from Ets-1) (109), suggesting that biological mechanisms may exist that discriminate specific levels of interfacial hydration in cognate ETS/DNA complexes.

Functionally, PU.1ΔN167^{Y252F} binds as well as wild type to site-specific DNA in the absence of osmotic stress. On the basis of affinity alone, it would seem that the wild type protein is at best functionally neutral with respect to the Y252F mutant under norm-osmotic conditions. What may be a selective advantage of wild type PU.1? The evolutionary maintenance of the wild type as a more osmotically sensitive phenotype, in light of the hydration properties of ETV6, suggests that responsiveness to osmotic pressure is a functional advantage for PU.1. Biologically, PU.1 is a lineage-restricted transcription factor and critical regulator of hematopoiesis, the multistep process by which stem and progenitor cells in bone marrow self-renew and differentiate into terminal lineages of blood cells (145,150,191-193). As physiological osmotic stress has been established in lymphoid tissues in vivo (194) due to the very high metabolic rate within rapidly dividing cells, (195) an emerging hypothesis is that strong osmotic responsiveness by PU.1 may be a favorable regulatory phenotype. We have previously discovered through a bioinformatics analysis that PU.1 targets are over-represented among known osmotically sensitive genes (162). The present mutagenesis data indicating that DNA binding by PU.1 is biophysically optimized for osmotic sensitivity lends further support to this hypothesis.

2.4.2 Role of Interfacial Hydration in DNA Recognition by ETS-1.

In stark contrast with PU.1, the loss of the analogous Tyr^{412} water-mediated contact in Ets-1 reduces its DNA binding affinity (though >10-fold above nonspecific levels) (187) under norm-osmotic conditions. Thermal denaturation of wild type and mutant proteins in the absence of DNA showed that the Tyr-to-Phe mutation was more perturbative to the conformational stability of unbound Ets-1 than PU.1 [cf. Figure 2.3B]. Molecular dynamics simulations suggested that the Y412F mutation in Ets-1 dampened dynamic fluctuations in the free protein, as judged by differences in their relative hydration and diffusive properties [cf. Figure 2.4, parts D and F], and attenuated induced fluctuations in the DNA-bound state that appeared to be crucial for Ets-1 binding. Thus, the interfacial tyrosine in Ets-1 (Y412), even though it shows the same hydration pattern as PU.1 (Y252), is not directly linked to hydration, but rather part of the structural dynamics that are specific to Ets-1. This interpretation is consistent with experimental dynamics from solution NMR of free and unbound Ets-1, titration calorimetry (Ets-1/DNA binding being more entropically driven than PU.1) (162), and the skewing of osmotic sensitivity toward water release by the Y412F mutant (abrogation of hydration of dynamic elements in wild type Ets-1). Strong dynamics in the N-terminal helices have also been reported in a previous molecular dynamics study of DNA-bound wild type Ets-1 (196), but that study did not include a paired simulation of the unbound protein or any mutant for comparison. Taken together, the structural observation that Ets-1/DNA interface is sparsely hydrated and maintained primarily by direct contacts between protein and DNA (160,162) should now be understood mainly in terms of induced dynamics in the DNA-bound state, rather than the lack of an adhesive role for interfacial water-mediated contacts as with PU.1 and its proximal phylogenetic ETS relatives.

2.5 Conclusion

By way of mutating an interfacial water-coordinating tyrosine that is highly conserved among known members of the ETS family, we showed that interfacial hydration could be modulated in a protein/DNA complex. Presumably, additional strategic point mutations could achieve further fine-grained tuning of hydration in a strongly osmotically sensitive system such as PU.1. However, interfacial hydration plays distinct roles in this family, as PU.1ΔN167^{Y252F} binds its target DNA site just as well as wild type PU.1 under norm-osmotic conditions but affinity for Ets-1 ΔN331^{Y412F} is reduced. This disparity arises from DNA-bound Ets-1's low dependence on interfacial hydration in the first place and the coupling of a specific water-mediated contact (via Tyr^{412}) to induced dynamics in the DNA-bound state. The present observations add to a growing line of evidence attesting that ETS homologues, despite their structurally superimposable DNA-bound structures, recognize target DNA sites by distinct mechanisms.

2.6 Materials and methods

2.6.1 Nucleic acids

Synthetic DNA oligos were purchased from Integrated DNA Technologies (Coralville, IA) and annealed to form duplex binding sites. DNA sites were the optimal binding sequences for PU.1 (5′-AGCGGAAGTG-3′) (197) and Ets-1 (5′GCCGGAAGTG-3′, termed SC1); (198) the ETS-specific core consensus is in bold. Fluorescent DNA probes were constructed by annealing a TYE 563-labeled oligo with excess unlabeled complementary strand as described. (184) Concentrations of unmodified oligos were determined spectrophotometrically using nearest-neighbor extinction coefficients. (199)

2.6.2 Molecular cloning

The cloning of the ETS domains of murine PU.1 (residues 167 to 272, designated PU.1ΔN167) and Ets-1 (residues 331 to 440, designated Ets-1ΔN331) have been described previously. (162,178) The point mutants Y252F for PU.1 and Y421F for Ets-1 were constructed by standard PCR mutagenesis. Briefly, for each construct separate PCR reactions were performed using mutagenic primers to produce two DNA fragments with overlapping sequences that harbored the intended Tyr-to-Phe mutation. A second PCR amplification of the first-round products with external primers generated the final construct that was cloned into pET28b vector and propagated in DH5α Escherichia coli cells under kanamycin selection (50 μg/mL). Clones were verified by Sanger sequencing and transformed into BL21*(DE3) E. coli for expression

2.6.3 Protein Expression and Purification

Wild type or mutant ETS constructs were overexpressed in Escherichia coli as fusions with a thrombin-cleavable C-terminal 6xHis tag and purified as previously described. (162) In brief, cleared lysate from sonicated cell pellets were first purified on Co-NTA 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 (for PU.1) or 0.15 M (for Ets-1) NaCl, and polished on Sepharose SP (GE). Buffers used with Ets-1 constructs, which harbored reduced cysteine, additionally contained 0.5 mM TCEP. Purified constructs were homogeneous as judged by Coomassie-stained SDS-PAGE. Protein concentrations were determined by UV absorption at 280 nm using the following extinction coefficients (in M^{-1} cm^{-1}): 22460 (wild type PU.1ΔN167), 20970 (PU.1ΔN167Y252F), 32430 (wild type Ets-1ΔN331), and 30940 (Ets-1ΔN331Y412F).

2.6.4 Fluorescence Polarization Titrations

ETS protein binding to fluorescently labeled DNA sites was measured in a solution using a Molecular Devices Paradigm plate reader. TYE-labeled DNA probe (0.5 nM) was incubated to equilibrium with purified wild type (PU.1ΔN167 or Ets-1ΔN331) or mutant protein (PU.1ΔN167Y252F or Ets-1ΔN331Y412F) and graded concentrations of unlabeled high affinity site in 30 μL of total volume. The solution was 10 mM TrisHCl (pH 7.4) containing 150 mM total Na+, 5 mM DTT, 0.1 mg/mL acetylated bovine serum albumin (Promega), and betaine as indicated. Solution osmolality was measured using a freezing point depression osmometer (Osmomat 3000, GonoTec) calibrated with commercial NaCl standards. Immediately before fluorescence measurement, samples were transferred to black 384-well plates (Corning) and excited at 535/25 nm. Steady-state fluorescence parallel and perpendicular to the incident polarized light were acquired at 595/35 nm. Dark counts of a buffer-only control were subtracted from each emission count before conversion to the anisotropy using a grating factor that was independently determined with sulforhodamine B ($\langle r \rangle$ = 0.025 at 25 °C). Anisotropy data from multiple experiments were fitted simultaneously with a mechanistic binding model to directly estimate the dissociation constants ± SE of each protein complex with unmodified DNA. (184)

2.6.5 Circular Dichroism Spectroscopy

Purified proteins were extensively dialyzed against 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.4) containing 150 mM NaCl before scanning from 300 to 190 nm at 10 μM in a 1 mm-path length quartz cuvette at 25 °C using a Jasco J-810 instrument. Per-residue normalized spectra were decomposed using BeStSel, (200) without scaling, to estimate their secondary structure contents. Melting experiments were carried out with 25 μM of each construct, detected at 222 nm while

heated at 45 °C/h to 70 °C, followed by cooling at the same rate. Melting and refolding data were simultaneously fitted to a uni-molecular two-state transition with independent linear baselines.

2.6.6 Size Exclusion Chromatography

Purified proteins (3 nmol) were injected and eluted at 1.3 mL/min with 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.4) containing 150 mM NaCl and 0.5 mM TCEP in a Superdex 75 10/300 GL column (GE) under the control of a Bio-Rad NGC instrument. The column was calibrated with bovine serum albumin (66 kDa), carbonic anhydrase (30 kDa), and myoglobin (17 kDa). Eluate was detected by UV absorption at 280 and 409 nm, the latter specific for the Soret band of myoglobin. The void volume of the column was 8 mL.

2.6.7 2D $^{1}H–^{15}N$ HSQC NMR

Uniformly ^{15}N-labeled wild type Ets-1ΔN331 and Ets-1ΔN331Y412F were expressed in M9 minimal media containing $^{15}NH_4Cl$ and purified identically as unlabeled proteins. Samples (~0.3 mM) were extensively dialyzed against 11 mM NaH_2PO_4/Na_2HPO_4, pH 7.4, 167 mM NaCl, and 0.1% NaN_3 and adjusted to 10% D_2O. $^{1}H–^{15}N$ correlated measurements were made using a phase-sensitive, double inept transfer with a garp decoupling sequence and solvent suppression (hsqcf3gpph19). Spectra were acquired with 1k × 144 data points and zero-filled to 4k × 4k

2.6.8 Molecular Dynamics Simulations

Free and site-specifically bound wild type and mutant ETS domains of PU.1 and Ets-1 were simulated using GROMACS 2016.1 with the CHARMM36 force field. Starting coordinates for the wild type ETS domains of PU.1 and Ets-1 were taken from available (co)crystal structures and trimmed to the same length after alignment of their protein sequences. The DNA was mutated to the experimental sequences without altering any existing backbone or sugar coordinate using 3DNA. (201) The target Tyr residues were mutated to Phe using CHARMM-GUI. (202) All crystal

waters were retained as TIP3P waters. Each system was placed in a dodecahedral box of TIP3P water with a minimum of 1.0 nm from its periodic boundary and neutralized with 0.15 M NaCl. The systems were then minimized by steepest descent. The energy-minimized systems were further relaxed, in order, as NVT and NPT ensembles for 100 ps at 2 fs time steps at 298 K with restraints on the macromolecules. Afterward, restraints were removed from the equilibrated systems and run for 100 ns for unbound proteins and 200 ns for DNA-bound complexes at 2 fs time steps at 298 K. Bond constraints were introduced using the LINCS algorithm at order 4. Coordinates, velocities, and energies were saved at 10 ps intervals. Trajectories were then imaged and analyzed using GROMACS tools. In cases where the lengths of the bound DNA were different, RMS measurements were limited to corresponding protein in the structures. Hydrogen-bond enumerations were carried out using geometric cut-offs of 0.35 nm for distance and 150° for angle (between donor-H and donor−acceptor heavy atoms).

2.7 **Addendum in dissertation**

As detailed in the legend to Figure 2.5C, a per-residue B-factor was computed as a measure of the RMSF in the simulation and experimental models (1R36 and 2STT), via the relationship: $B = \frac{8\pi^2}{3}(\mathrm{RMSF})^2$. Strictly speaking, NMR ensembles represent a set of structures that satisfy the experimental distance and angular constraints imposed by NOE and other NMR observables. Our interpretation of locally loose constraints in terms of dynamics, rather than an artefact of the NMR data, is based on the quality analyses of 1R36 and 2STT curated by the RCSB. The neighborhoods of loose constraints exhibit levels of clashes and Ramachandran/sidechain outliers consistent with the overall (high) quality of the models. The close spatial correspondence with the *unrestrained* simulations as shown in Figure 2.5C lends further credence to our interpretation.

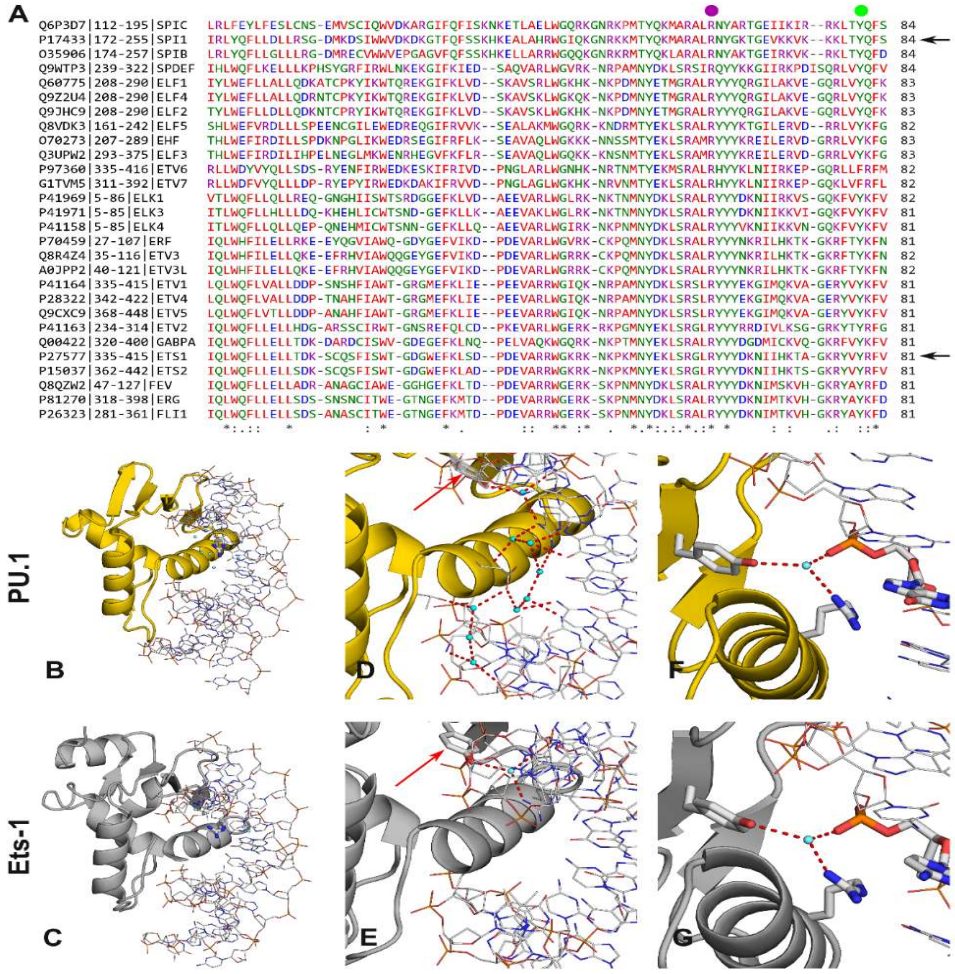


Figure 2-1 Highly conserved interfacial tyrosine coordinating a tripartite water-mediated ETS/DNA contact.

(A) Sequence alignment of the 28 murine members of the ETS transcription factor family, in which an interfacial Tyr (green ●) is conserved in 26 homologs including PU.1 (Spi1; Tyr252) and Ets-1 (Tyr412). Residues are colored by amino acid type. Symbols denote identity (∗), strong (:), or partial (·) conservation. (B−G) Differential interfacial hydration in the cocrystal structures of PU.1 (PDB 1PUE) and Ets-1 (1K79). (B and C) Hydration water within 5 Å of a conserved interfacial Arg (Arg235 in PU.1, Arg394 in Ets-1; (purple ●) in panel A) is shown as cyan spheres. (D and E) Water-mediated contacts (within 3.4 Å) in the protein/DNA interface of PU.1 and Ets-1, the latter contacting DNA primarily via direct contacts. Tyr252 and Tyr412 are marked by red arrows. (F and G) Expanded view of Tyr252/Tyr412 coordinating a highly conserved water-mediated contact with Arg235/Arg394 and a specific DNA backbone phosphate

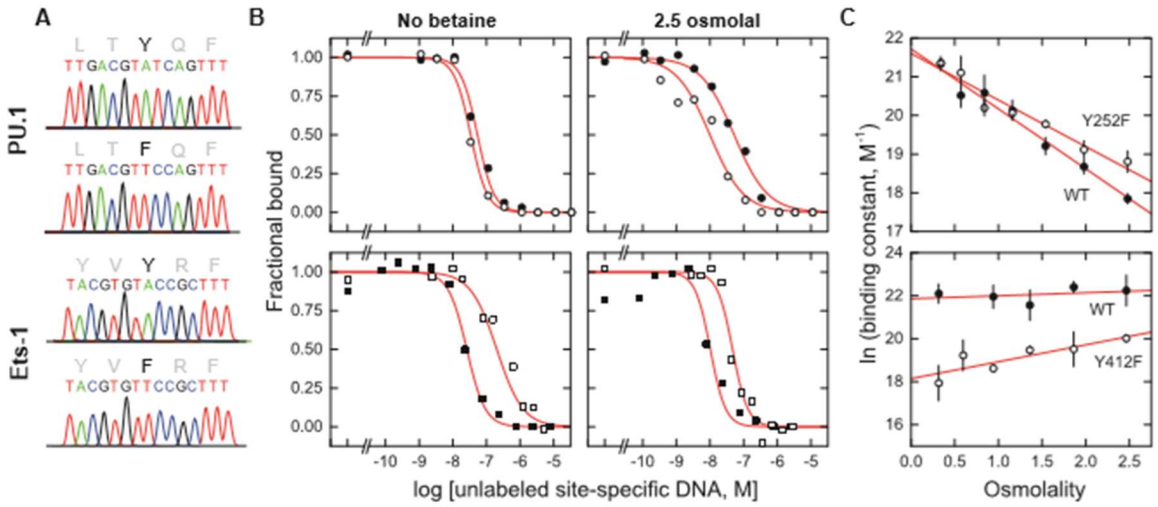


Figure 2-2 Mutation of an interfacial tyrosine to phenylalanine perturbing DNA-binding affinity and sensitivity to osmotic pressure by the ETS domains of PU.1 and Ets-1
(A) Electropherograms from Sanger sequencing of recombinant ETS constructs harboring the one-atom-different Tyr-to-Phe mutations in PU.1ΔN167 and Ets-1ΔN311. (B) Representative fluorescence polarization titrations of ETS-bound, TYE-labeled DNA with unlabeled, optimal site-specific DNA under normal-osmotic (0.29 osmolal) and osmotically stressed (2.5 osmolal) conditions. DNA probe was present at 0.5 nM and protein between 10 to 100 nM, depending on construct and conditions, to establish pre and post-titration baselines. Wild type and mutant protein binding are denoted by solid and open symbols, respectively. Curves represent fits of the data to a 1:1 competitive model. (C) Osmotic dependence of the binding constant shows sharply different responses by the two ETS homologs and their mutants to osmotic pressure. Lines represent weighted linear fits for each data set.

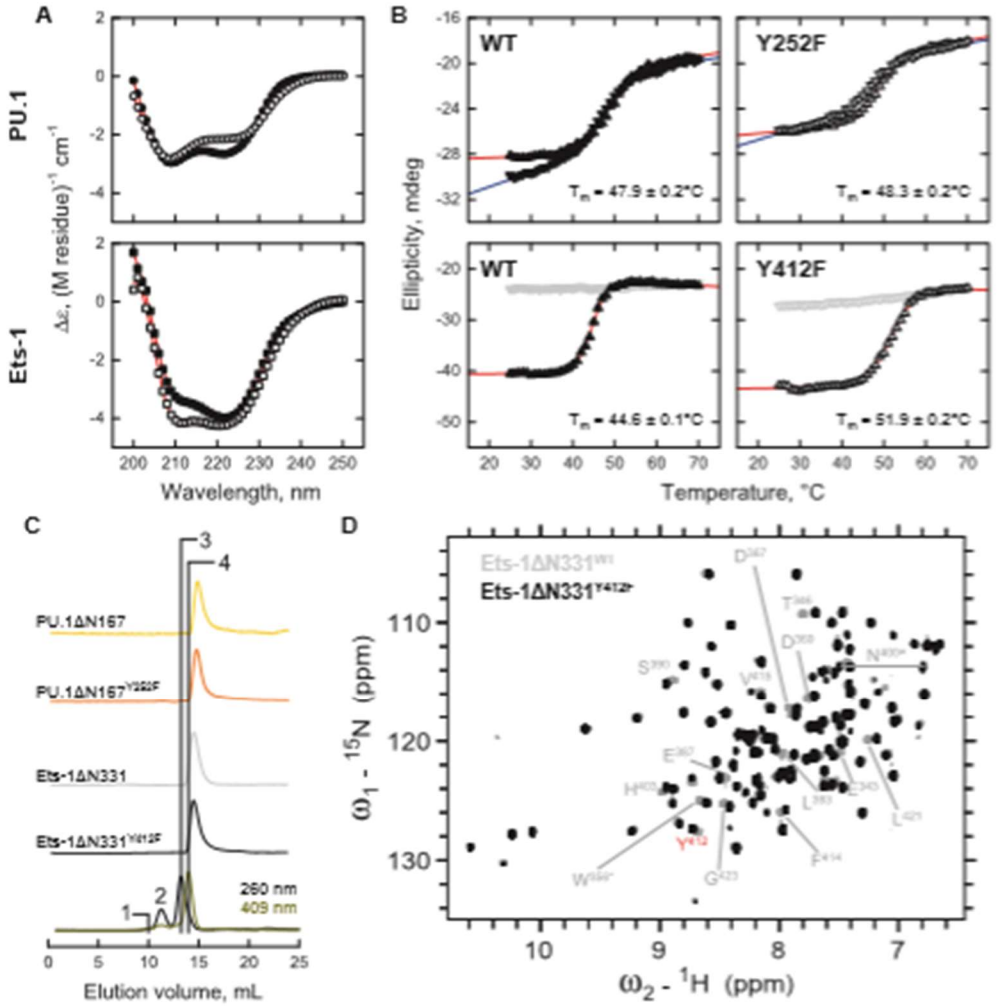


Figure 2-3 Analogous Tyr-to-Phe mutations which are differentially perturbative to the unbound PU.1 and Ets-1 ETS domains

(A) Far-UV circular dichroism scans of wild type (solid symbols) and mutant (open) PU.1ΔN167 and Ets-1ΔN331 at 25 °C, fitted with BeStSel (curves) to estimate secondary structure content. Parametric values are given in Table 1. (B) Thermal melting (up triangles) and cooling curves (down triangles) of wild type and mutant PU.1ΔN167 and Ets-1ΔN331, monitored at 222 nm. Neither ETS-1 construct unfolded reversibly (gray down triangles). Heating and, if reversible, cooling data was globally fitted with a uni-molecular two-state model (curves). Values are given in Table 1. (C) Size-exclusion chromatography of purified wild type and mutant PU.1ΔN167 and Ets-1ΔN331 on a Superdex 75 column, which was calibrated with bovine serum albumin (peaks 1 and 2; a 132 kDa dimer and 66 kDa monomer), carbonic anhydrase (peak 3, 30 kDa), and myoglobin (peak 4, 17 kDa, detected via its Soret band at 409 nm). (D) Overlaid solution HSQC spectra of wild type (gray) and Ets-1ΔN331Y412F (black). Where possible, chemical shift assignments of wild type resonances that exhibited the largest chemical shift changes (partial or complete lack of overlap) were taken from the solution NMR structure (PDB: 1R36). (188) Identity in the case of partially overlapped mutant residues was not assumed. Asterisks denote side-chain resonances.

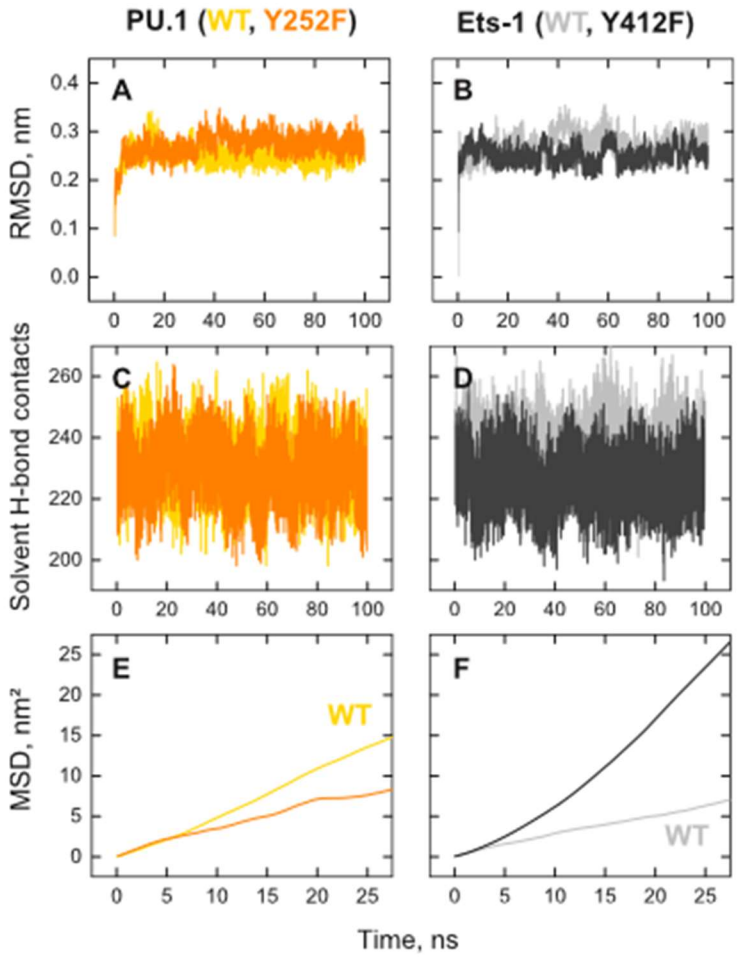


Figure 2-4 Molecular dynamics simulations recapitulate experimental properties of the unbound ETS domains of PU.1 and Ets-1

Following all-atom simulations for 100 ns at 298 K, the trajectories and structures of the ETS domains of PU.1 and Ets-1 were analyzed in terms of RMS deviation from initial coordinates (RMSD, parts A and B), hydrogen bonds with solvent (parts C and D), and mean square displacement (MSD, parts E and F). The RMSD data showed rapid convergence in the dynamics of both sets of proteins. Enumeration of H-bond contacts with solvent showed a 4% bias in favor of wild type Ets-1, consistent with its thermal instability relative to the Y412F mutant. Finally, the lower MSD of wild type Ets-1 relative to PU.1 reflects dynamic light scattering measurements showing a 14% larger hydrodynamic radius for ETS-1. The MSD trajectories were divergent out to 100 ns.

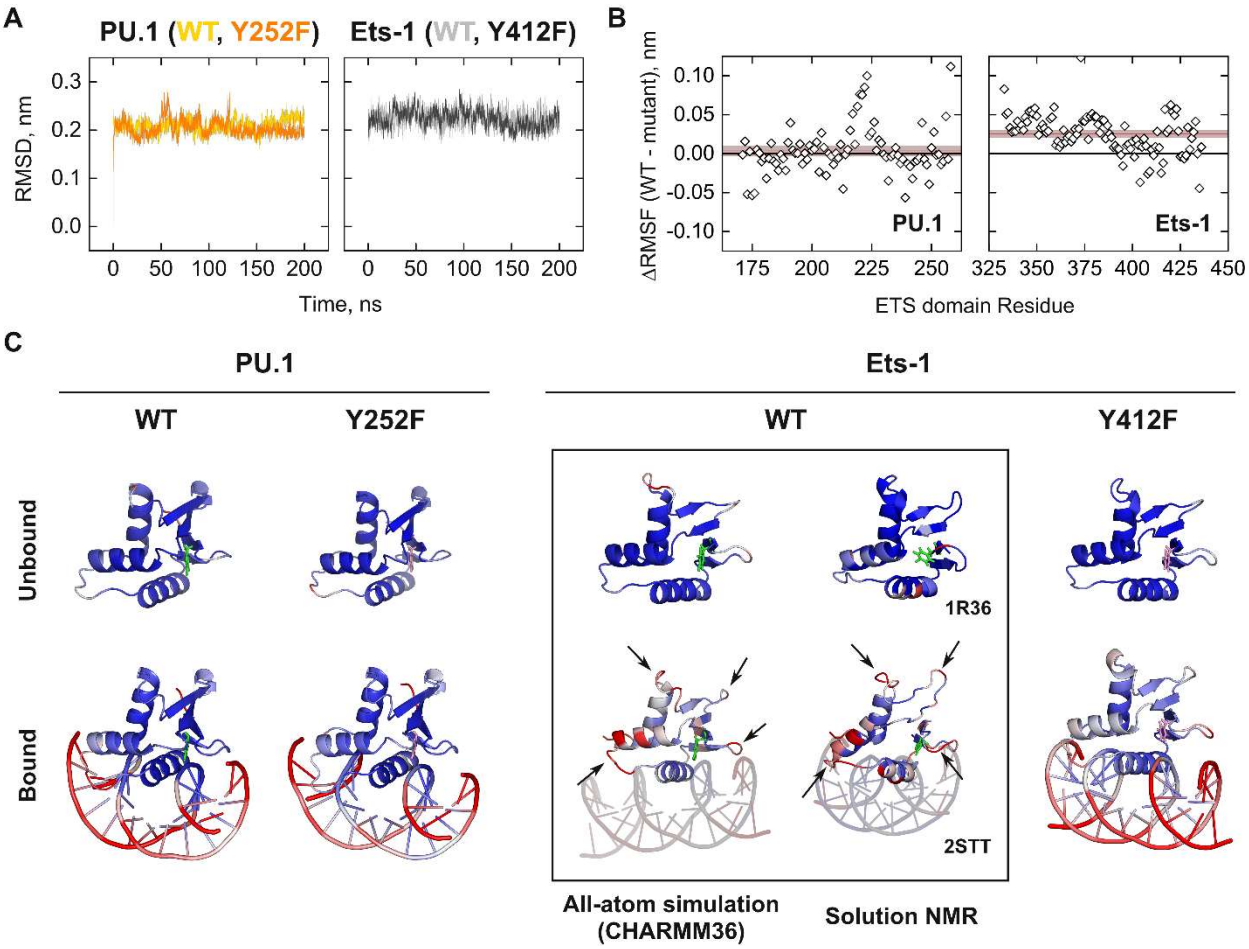


Figure 2-5 Molecular dynamics simulations recapitulating induced experimental dynamics in the wild type Ets-1 complex and revealing a basis for destabilization by Y412F mutation
(A) All-atom RMS deviation from initial coordinates for wild type and mutant PU.1 and Ets-1 in their DNA-bound states. (B) Per-residue RMS fluctuations, computed from the final 50 ns of the simulations, revealed significant dynamic bias in wild type Ets-1 (positive values indicate excess dynamics in wild type), but not PU.1. The red line represents the average RMSF ± SE. Dynamics of the bound DNA, not shown, exhibited identical trends. (C) Comparison with experimental NMR structures shows that the simulations reproduced the localized dynamics in the wild type Ets-1/DNA complex. Both experimental structures were averaged from 25 conformers in the PDB ensemble; the simulated averaged structure consisted of 5000 models. Fluctuations are colored by residue based on a "B-factor" (proportional to RMSF) on a scale from 0 to 100 in a blue-white-red continuum, computed from the fluctuations from average positions. This scaling is intended only to facilitate a comparison of the isotropic simulated and experimental fluctuations on a common scale. Corresponding sites of induced dynamics between the simulated and experimental wild type Ets-1/DNA complexes are marked by arrows (the DNA is rendered semitransparent for clarity). The secondary structure assignments of the simulated structures are inherited from their crystallographic templates and differ at places from the solution NMR structure. Tyr412 (wild type) and Phe412 (mutant) are shown in green and pink, respectively.

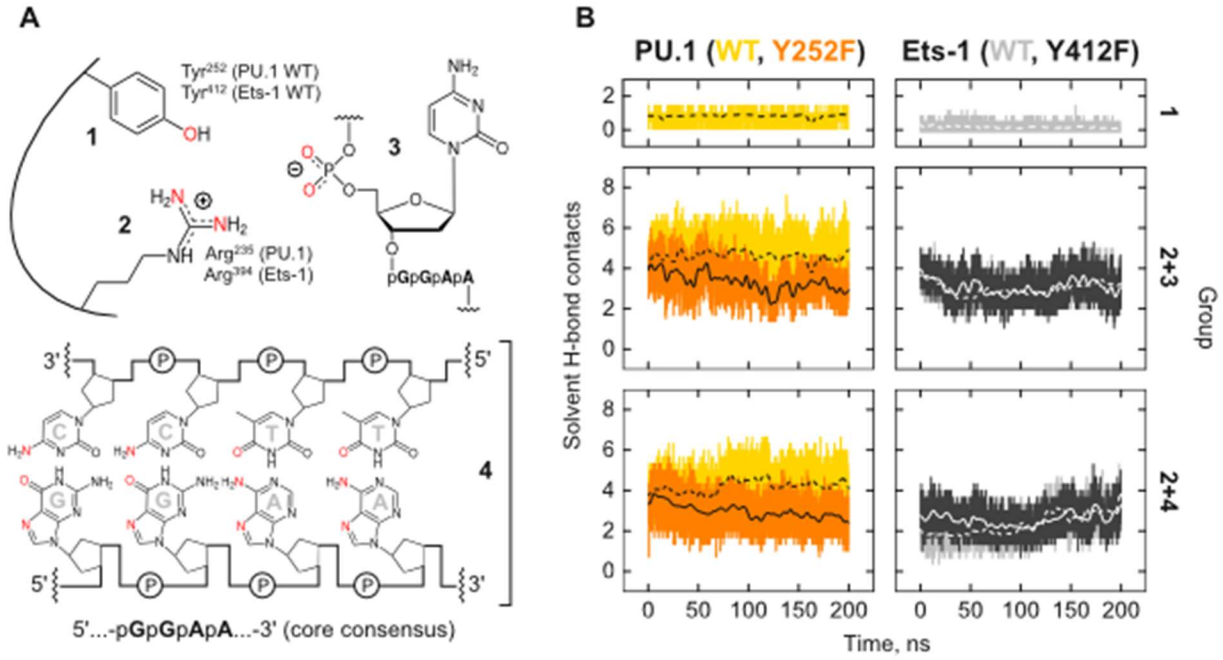


Figure 2-6 Molecular dynamics simulations of DNA-bound PU.1 and Ets-1 ETS domain, which provide a window into the interfacial hydration of wild type and mutant PU.1 and Ets-1
(A) Groups in the ETS/DNA interface are used to probe interfacial hydration. Heteroatoms used as probes for solvent H-bonds are highlighted in red. Note that the H-bonding atoms in Group 4 (representing the 5′-GGAA-3′ core consensus) specifically face the major groove toward the protein; backbone sugar and phosphates are schematized. (B) Solvent H-bond contacts were made by the indicated groups given as the average of three independent simulations. The curves (dashed and solid for wild type and mutant, respectively) represent FFT-smoothed averages based on a 2.5 ns window.

Table 1 Conformational Properties of Free Wild type and Mutant ETS Domains of PU.1 and Ets-1[a]

	Secondary structure content, %				Conformational stability	
	α helix	β sheet	Turns	Other	T_m, °C	$\Delta H(T_m) \times 10^5$, kJ/mol
PU.1ΔN167^{WT}	22.9	31.3	13.9	31.9	47.9 ± 0.2	26.6 ± 1.5
PU.1ΔN167^{Y252F}	22.4	26.2	15.0	36.4	48.2 ± 0.2	29.2 ± 1.6
Ets-1ΔN331^{WT}	25.0	36.5	9.3	29.2	44.6 ± 0.1[b]	44.2 ± 1.4[b]
Ets-1ΔN331^{Y412F}	24.2	32.0	14.2	29.6	51.9 ± 0.2[b]	30.1 ± 1.4[b]

[a]Secondary structure content for each construct was estimated by decomposition of their CD spectra using the BeStSel algorithm (200). Conformational stability was measured by CD-detected thermal denaturation and renaturation and fitted globally for both heating and cooling (where possible) by a uni-molecular two-state model. [b] Determined from heating curve only.

3 INTRINSIC DISORDER CONTROLS TWO FUNCTIONALLY DISTINCT DIMERS OF THE MASTER TRANSCRIPTION FACTOR PU.1

 [Science Advances Volume 6, Number 8] 21 Feb 2020 DOI: 10.1126/sciadv.aay3178

3.1 Abstract

Transcription factors comprise a major reservoir of conformational disorder in the eukaryotic proteome. The hematopoietic master regulator PU.1 presents a well-defined model of the most common configuration of intrinsically disordered regions (IDRs) in transcription factors. We report that the structured DNA binding domain (DBD) of PU.1 regulates gene expression via antagonistic dimeric states that are reciprocally controlled by cognate DNA on the one hand and by its proximal anionic IDR on the other. The two conformers are mediated by distinct regions of the DBD without structured contributions from the tethered IDRs. Unlike DNA-bound complexes, the unbound dimer is markedly destabilized. Dimerization without DNA is promoted by progressive phosphomimetic substitutions of IDR residues that are phosphorylated in immune activation and stimulated by anionic crowding agents. These results suggest a previously unidentified, nonstructural role for charged IDRs in conformational control by mitigating electrostatic penalties that would mask the interactions of highly cationic DBDs.

3.2 Introduction

Eukaryotic transcription factors are highly enriched in intrinsically disordered regions (IDR), which are sequences that do not adopt a stably structured conformation but are nevertheless essential for activity. Compared with only ~5% in prokaryotes and archaea, more than 80% of eukaryotic transcription factors have extended IDRs (203). In the unicellular baker's yeast (Saccharomyces), transcription factors comprise the most prodigious functional category of

disorder encoding proteins (204). In multicellular organisms, ~50% of all residues in eukaryotic factors from model animals (humans, Drosophila) and plants (Arabidopsis) map to disordered regions (45). Clearly, IDRs constitute a major component in eukaryotic gene regulation, and it is therefore important to define their contributions to the molecular properties of transcriptional factors.

While IDRs are generally diverse in sequence, charge characteristics confer specific properties to transcription factor IDRs. For example, positively charged tails mediate diffusion along DNA (205) and ubiquitination by E3 ligases of several transcription factors, notably p53 (206). More common, however, are negatively charged ("acidic") IDRs such as transactivation domains, which recruit basal factors such as TFIIB and TATA binding protein to the promoter (207,208), and signaling moieties such as PEST domains that are rich in Glu and Asp residues (209,210). While IDRs exhibit sequence-dependent conformational preferences on their own, these preferences are also modified by folded domains to which they are tethered (211). In transcription factors, IDRs are highly enriched around DNA binding domains (DBDs) (212), which display electrostatically biased surfaces to their surroundings. Their DNA contact surfaces are typically rich in positively charged residues while exposing neutral or even negatively charged residues elsewhere. Because DBDs alone represent an incomplete context in functional regulation, our aim is to elaborate the mechanism by which charged, particularly acidic IDRs regulate the recognition of tethered DBDs with each other as well as with target DNA.

As a model system for understanding the impact of intrinsically disordered tethers in transcription factors, the ETS family protein PU.1 exemplifies the most common (known as type I) configuration (45), in which its eponymous DBD of ~90 residues comprises the only well-folded structure. The remaining ~170 and 12 residues that flank N- and C-terminally, respectively, are

intrinsically disordered sequences. The extended N-terminal IDR consists of an acidic transactivation domain (human residues 1 to 80), a Q-rich domain (residues 81 to 116), and a highly negatively charged PEST domain (residues 117 to 165), all of which are characteristic disordered regions in eukaryotic factors (210). This domain architecture, conserved among PU.1 orthologs, commends PU.1 as an ideal model from which more complex transcription factor architectures may be approached.

In addition to the canonical attributes representative of eukaryotic transcription factors, PU.1 is also specifically required for life. During hematopoiesis, all circulating blood cells are ultimately derived from a small population of self-renewing stem cells. PU.1 is a master regulator that is required for the renewal of the hematopoietic stem cells (213) and, in collaboration with other factors, directs their differentiation to every major myeloid and lymphoid lineage. Aberrant PU.1 activity is associated with lymphomas (214), myeloma (215), leukemia (216), and Alzheimer's disease (217). Most recently, PU.1 was also identified as the key trigger for tissue fibrosis (156). Genetic and pharmacologic interventions targeted at PU.1 have established its therapeutic potential in acute myeloid leukemia (157,158) and fibrotic diseases (156). Mechanisms that govern the molecular interactions of PU.1 are therefore relevant to developmental genetics and multiple therapeutic areas including hematology/oncology, immunology, neurology, and rheumatology.

Despite its biological significance, detailed knowledge of the molecular properties of PU.1 has been limited to its structured ETS domain. PU.1 therefore exemplifies the "incomplete context problem" in structural biology, which we have now tackled by addressing the role of the N- and C-terminal IDRs in the behavior of the ETS domain. The data reveal that these tethered IDRs critically control the propensities of the ETS domain to form discrete dimers with and without

cognate DNA. These dimeric states, which are conformationally distinct, establish a novel regulatory mechanism that enables negative feedback in PU.1 transactivation. In addition to implications on PU.1 autoregulation in vivo, these results address a general class of problems in which negatively charged IDRs, which are abundant in transcription factors as transactivation and other functional domains, exert direct functional control at the protein/ DNA level.

3.3 Results

The DNA binding (ETS) domain of PU.1 represents its only structured domain, whose 1:1 complex with cognate DNA [Fig. 3.1A] is structurally conserved in this family of transcription factors. However, in contrast with other ETS members, the ETS domain of PU.1 (ΔN165) forms 2:1 complexes at single DNA cognate sites in biophysical assays (163,164). Measurements of self-diffusion by protein-observed diffusion ordered spectroscopy (DOSY) nuclear magnetic resonance (NMR) showed that inclusion of the N-terminal PEST domain (ΔN117) maintained the DNA binding modes accessible to the ETS domain [Fig. 3.1B]. Specifically, DOSY titrations of both ΔN117 and ΔN165 with DNA oligomers harboring a single cognate binding site showed two distinct bound states, with the minima in diffusion coefficient occurring sharply at a DNA: protein ratio of 0.5, corresponding to a 2:1 complex. The single minima at DNA: protein = 0.5 excluded the formal possibility of a 2:2 complex or nonspecific binding. If PU.1 were binding DNA nonspecifically beyond the 1:1 complex at equilibrium (i.e., in an un-saturable manner proportional to the concentration of free protein), then the minima in diffusion coefficient would occur at the lowest DNA: protein ratios where protein would be at greatest excess relative to DNA. Independently, protein-into-DNA titrations showed that ΔN117 enhanced the affinity of the 1:1 complex (K_{D1}) by more than twofold and reduced the affinity of the 2:1 complex (K_{D2}) relative to ΔN165 by about fourfold [Fig. 3.1B]. Taking the ratio K_{D2}/K_{D1} as an index of cooperativity in

DNA binding, 2:1 complex formation by ΔN117 was therefore more negatively cooperative than ΔN165. The PEST domain, therefore, preserved the intrinsic binding modes of the ETS domain of PU.1, namely, a 1:1 and 2:1 complex with cognate DNA, while modulating their affinities in solution.

3.3.1 PU.1 is self-regulated by negative feedback

In tissues that natively express PU.1, such as macrophages, PU.1 activity is highly inducible (218). The 1:1 complex formed by ETS domains represents the established trans-regulatory complex for ETS transcription factors. Little is understood about the functional nature of the 2:1 complex, for which no ETS analog is known, although its negative cooperative relationship with the 1:1 complex suggests an inactive species [Fig. 3.1C]. To solve this puzzle, we measured PU.1 transactivation in cells using an enhanced green fluorescent protein (EGFP) reporter gene under the control of various synthetic enhancer elements consisting only of tandem copies of the λB motif [Fig. 3.1D], a PU.1-specific ETS binding site (EBS) derived from the lymphoid Igλ2-4 enhancer (GenBank X54550). Each consecutive site was spaced by 20 base pairs (bp), or two helical turns, such that bound proteins were arrayed on the same helical face to facilitate the recruitment of the transcriptional machinery. In addition, as the 2:1 complex was known to require an extended site size relative to the monomer (164), presenting the bound protein along one helical face would amplify site-site interactions and DNA perturbations, thus rendering most manifest the functional effects of the 2:1 complex.

When transiently transfected into PU.1-negative human embryonic kidney (HEK) 293 cells, the reporters were negligibly activated by endogenous transcription factors, including other ETS family proteins [Fig. 3.1E]. Co-transfection of an expression plasmid encoding full-length PU.1, which was independently tracked by a co-translating infrared fluorescent protein (iRFP)

marker, yielded EGFP fluorescence in a dose-dependent manner. We established a dosing range for the PU.1 expression plasmid that gave a linear variation in PU.1 abundance in HEK293 cells within the physiologically inducible range found in PU.1-expressing myeloid cells [Fig. 3.2]. In this configuration, PU.1-dependent transactivation was quantified as the fraction of iRFP-positive cells that were also EGFP positive [Fig. 3.1E]. The functional outcome of an inactive, negatively cooperative 2:1 complex would be a bell-shaped reporter dose-response as the enhancer, which varied in density and spacing of EBS (i.e., cis regulatory syntax), became saturated with nonproductively (2:1) bound PU.1. In the alternative, the reporter signal would dose-dependently settle to a saturable level, depending on the level at which the 2:1 complex retained activity relative to the 1:1 complex. The synthetic λB reporters were therefore well suited to interrogate cellular PU.1 activity, free from the requirement or interference from other promoter-specific cofactors, at the protein/DNA level.

At equivalent PU.1 doses, all enhancer configurations showed graded reporter expression in step with the density of EBS at each enhancer [Fig. 3.1F]. This was consistent with an expected multivalent effect with respect to PU.1 binding sites. However, EGFP expression increased monotonically only with enhancers harboring tandem 3× and 5× EBS. Upon peaking at intermediate PU.1 doses, the 1× and 2× enhancers were repressed by further increases in PU.1. To determine whether the reversal in transactivation involved PU.1 interaction at the enhancer, we mutated the even-numbered sites in the 5×EBS reporter to generate a 3×EBS variant in which the cognate sites doubled in spacing [Fig. 3.1D]. The resultant 3×-alt-EBS reporter exhibited lower transactivation than the more densely spaced 3×EBS, and its reporter signal also no longer increased monotonically [Fig. 3.1F]. The spacing effect, therefore, demonstrated that the functional reversal could not be due solely to PU.1 interaction away from the DNA, which would

be inert to syntax changes at the DNA. The observation of bell-shaped dose response for the 1× and 2× enhancers, but not the 3× or 5×EBS enhancers, suggested additive perturbations of the local DNA structure, which were amplified by the helical spacing of the sites. This interpretation was supported by previous DNA footprinting of the PU.1 ETS domain, which showed strong differences between the singly and doubly PU.1-bound DNA (164). Alternatively, binding at the higher-density sites might exhaust a required co-repressing factor for the 2:1 complex. However, this possibility was discounted by the different dose responses exhibited by the 3×EBS and 3×-alt-EBS, which had the same site density, and the occurrence in a cell line (HEK293) that does not natively use PU.1 in gene regulation. Because net transactivation activity was reduced under conditions corresponding to population of the 2:1 complex, the evidence suggested that the 2:1 complex lost activity relative to the 1:1 complex. Thus, manipulation of enhancer syntax (density and spacing) demonstrated negative feedback in PU.1 transactivation in a manner consistent with self-titration of the transcriptionally active 1:1 complex by an inactive dimer bound to DNA

To extend our functional results to a more physiologic context, we evaluated the impact of graded PU.1 inhibition on three PU.1 target genes in THP-1 cells, a widely used human monocyte/macrophage model. Cells were treated with a PU.1 inhibitor [Fig. 3.3], as a function of dose or incubation period, before stimulation with phorbol 12-myristate 13-acetate (PMA) to mimic PU.1 induction during myeloid differentiation. As PU.1 targets, we examined the *pu.1 (Spi-1)* gene itself, which is auto-regulated (219); *csf1ra*, a PU.1 target that encodes the α subunit of the colony-stimulating factor receptor; and *e2f1*, which is negatively regulated by PU.1(220). We first tested the effect of dose-dependent inhibition of PU.1 for a fixed period of 2 hours on the transcription of these genes by reverse transcription polymerase chain reaction (RT-PCR) (table S2). Expression of *pu.1* and *csf1ra*, both positively regulated PU.1 targets, was increased by lower

doses of inhibitor before marked reduction to ~50% at higher doses, yielding bell-shaped profiles [Fig. 3.1G]. In the case of negatively regulated *e2f1*, expression was further inhibited across the dosage range of inhibitor tested. Trans-regulation of all three genes upon dose-dependent inhibition of PU.1 was consistent with an increase in PU.1 activity associated with the relief of negative feedback.

To assess the impact of PU.1 inhibition temporally, we tested a fixed dose of inhibitor (20 μM) over time, up to 16 hours before PMA induction. While PU.1 expression gave a bell-shaped dose response at 2 hours of inhibitor exposure, continued exposure at an intermediate (de-repressing) dose became strictly inhibitory [Fig. 3.1G]. In contrast, de-repression in csf1ra expression continued for 8 hours. Expression of the negative-regulated *e2f1* gene, which was dose-dependently reduced at 2 hours of PU.1 inhibition, began to increase by 8 hours of inhibitor exposure. These results thus demonstrated a dynamic nature to the negative feedback that corresponded to the specific effect of PU.1 on the target gene (peaks in activated genes or troughs in repressed genes). The opposing behavior of *csf1ra* and *e2f1* expression, in accordance to their opposite dependence on PU.1, supported the physiologic relevance of PU.1 negative feedback. Last, the latency exhibited by the two target genes relative to the auto-regulated *pu.1* gene suggested a combined effect between changes in PU.1 availability at the expression level and competition for binding at the DNA level.

In summary, the expression profiles of *pu.1*, *csf1ra,* and *e2f1* showed that graded PU.1 inhibition led to non-monotonic changes in trans-regulatory activity in a manner consistent with depression of negative feedback. Together with the dependence of the synthetic λB reporter on PU.1 dose and enhancer syntax (site density and spacing), the data support the biophysically

observed 2:1 complex as a functionally relevant species in the cell and motivate specific interest in how the ETS domain dimerizes in its native structural context

3.3.2 The PU.1 PEST domain is an IDR that modulates the stability of the 2:1 DNA complex

Comparison of DNA binding by ΔN117 and ΔN165 shows that the N-terminally tethered PEST domain enhanced the affinity of the 1:1 complex but reduced the affinity of the 2:1 complex [Fig. 3.1B]. To better understand the influence of the PEST domain on DNA recognition by PU.1, we first established whether the PEST domain was disordered in the cognate complex by comparing the ^{1}H-^{15}N heteronuclear single quantum coherence spectroscopy (HSQC) fingerprint region of DNA-bound ΔN165 and ΔN117 [Fig. 3.4A]. As with ΔN165 (164), the unbound and 1:1 complex gave well-dispersed spectra, while >80% of the cross peaks for 2:1-bound ΔN117 were broadened out [Fig. 3.5]. The similar behavior by the two constructs indicated that broadening was not due to the larger size of the 2:1 complex, in which case broadening would be exacerbated for ΔN117. In 1:1-bound ΔN165, whose resonances were well resolved, 88 of the 95 assigned residues overlapped with ΔN117, with all PEST residues clustered around 8.2 ± 0.2 parts per million (ppm) on the 1 H dimension, a chemical shift characteristic of disordered structures. Because this region also represented the residues that were detected in HSQC of the 2:1 complex [Fig. 3.5], the evidence suggested similar changes in the chemical environment for the structured ETS domain (represented by the dispersed resonances in intermediate exchange) between free and DNA-bound states of ΔN117 and ΔN165. Thus, the local structure of the ETS domain was not altered upon DNA binding by the flanking residues, and the PEST domain behaved as a disordered tether in the ETS/DNA complex.

Ligand binding to DNA is generally sensitive to electrostatic interactions. To better understand the impact of the disordered PEST domain on 2:1 complex formation, we probed the electrostatic contribution to site-specific binding by ΔN165 and ΔN117 [Fig. 3.4B]. Reducing Na^+ concentration from 0.15 M (as shown in Fig. 3.1B) to 0.10 M did not affect 2:1 binding by ΔN117. However, the biphasic binding indicative of strongly negatively cooperative formation of the 2:1 complex for ΔN165 was abolished as the anisotropy values showed. A further reduction to 0.05 M salt resulted in monophasic transitions to the 2:1 complex by both constructs. Notably, binding weakened with decreasing Na^+ concentration and therefore could not reflect simple electrostatic effects on DNA binding. These observations indicated that additional unbound species must regulate DNA recognition by PU.1 and that these species were salt sensitive and controlled by the disordered PEST domain

3.3.3 The disordered PEST promotes PU.1 homo-dimerization

The isolated ETS domain, ΔN165, forms a feeble dimer without DNA, as judged by heteronuclear NMR (221) as well as static and dynamic light scattering (163). To determine the role of the disordered PEST domain in PU.1 dimerization without DNA, we examined several hydrodynamic parameters, which are highly sensitive to self-association, of ΔN117 as a function of concentration [Fig. 3.6A]. DOSY NMR spectroscopy revealed a marked concentration dependence for the apparent diffusion coefficient. The profile was described by a two-state monomer-dimer equilibrium (detailed in Materials and Methods) with a dissociation constant below 10 μM (table S1). To assess concentrations below 50 μM, which was limiting for NMR, we performed intrinsic Trp fluorescence anisotropy measurements, which is sensitive to rotational diffusion. ΔN117 exhibited a substantial change in steady-state anisotropy that was also described by a two-state dimer with a dissociation constant at below 10 μM. In contrast, ΔN165 showed no

change. The localization of the three Trp residues in the structured ETS domain of both constructs represented further evidence that the concentration-dependent changes in ΔN117 involved the ETS domain. Last, high-precision densimetry showed a concentration-dependent transition by ΔN117 that was again described by two-state dimerization. (Because density varies directly with concentration, density-detected transitions sit on sloped baselines as opposed to the flat baselines in spectrometric titrations.) Relative to the diffusion probes, the densimetric titration gave a higher dissociation constant, 35 ± 15 μM. As a control, ΔN165 gave a concentration-independent partial specific volume (from the slope, see Materials and Methods) of 0.77 ± 0.01 ml/g, a value characteristic of structured globular proteins. Multiple orthogonal probes therefore described a reversible ΔN117 dimer that was considerably more avid than ΔN165.

We pause to note that concentration dependence of the equilibrium constant (and melting temperatures) rules out monomolecular interactions, such as conformational changes without association. Local conformational changes can and do produce changes in diffusion and volumetric parameters, but this behavior without an intermolecular component cannot depend on total concentration at thermodynamic equilibrium. Artefacts such as aggregation during the experiments are unlikely based on the linear post-transition baselines for all three probes (DOSY NMR, fluorescence anisotropy, and density). Independent evaluation of purified ΔN117 by SDS–polyacrylamide gel electrophoresis (PAGE) and mass spectrometry (MS) [Fig. 3.7] also confirmed the absence of detectable contamination and aggregation. At a deeper level of analysis, the two-state self-association model, given by Eq. 7 in Materials and Methods that fitted the titration data in [Fig. 3.6A], is a nth-order polynomial, where n is the stoichiometry of the oligomer. The value of n (= 2 for dimer), which is fixed in the fitting, imposes a severe constraint on the shape of the titration to which the model may adequately fit. As detailed elsewhere (222),

oligomers $n \geq 3$ invariably show sigmoidal (S-shaped) transitions. Only a two-state dimer exhibits non-sigmoidal profiles on linear concentration scales, precisely as constructed in [Fig. 3.6A] and observed in the data. On this basis, the biophysical evidence is unambiguous in showing homo-dimerization of PU.1 without DNA, and the range of dissociation constants yielded by the different probes reflected the distinct molecular properties they sampled.

To further strengthen this evidence, we resolved ΔN117 by electrospray ionization (ESI)–MS up to a concentration of 840 μM. Using an established maximum entropy procedure (223), peaks corresponding to both monomeric and dimeric PU.1 were observed in de-convoluted zero-charge mass spectra [Fig. 3.6B]. The integrated intensities of the two species were quantitative, but they did not correspond to solution conditions in the other experiments. This was due to the technique's requirement for a volatile buffer (NH_4HCO_3), species-dependent ionization efficiency, and the potential for ionization-induced dissociation of the complex. Notwithstanding, the ratio of dimer-to-monomer intensities varied in favor of the dimeric species with increasing total protein concentration [Fig. 3.6B, bottom]. The concentration dependence excluded the possibility that either species could represent a static contaminant but rather corresponded to a ΔN117 monomer and dimer at dynamic equilibrium.

To gain insight into the conformational structure of the free PU.1 dimer, we interrogated ΔN165 and ΔN117 by circular dichroism (CD) and NMR spectroscopy. At an identically low concentration (25 μM), a net contribution of coil content due to the PEST domain was apparent [Fig. 3.6C]. With increasing concentration, ΔN165 showed a spectral shift but without an endpoint at 300 μM. In contrast, the corresponding spectra for ΔN117 (weighted by contributions from the disordered PEST domain) underwent a non-sigmoidal transition that, unlike ΔN165, was substantially completed at 300 μM [Fig. 3.6D]. Model fitting of the far–ultraviolet (UV) CD

spectra, which are sensitive to secondary structure content, to a two-state dimer yielded a dissociation constant of K2 = 46 ± 19 μM. As an analysis of full CD spectra by singular value decomposition rendered more structural information than the other titration probes in Fig. 3A, we will use the CD-fitted K2 for comparison with other PU.1 constructs and solution conditions. To probe the local structure of the PU.1 dimer, we compared the ^{1}H-^{15}N HSQC fingerprint of 400 μM ΔN117 and ΔN165, concentrations at which the preceding experiments showed that ΔN117 was predominantly dimeric, while ΔN165 remained monomeric [Fig. 3.6E; compare to Fig. 3.6A]. Dispersed cross-peaks for the two constructs mostly overlapped within experimental uncertainty (inset). PEST domain residues were clustered at 8.2 ± 0.2 ppm. ^{1}H-^{15}N-NOE (nuclear Overhauser effect) measurements confirmed that the ETS residues in ΔN117 remained well ordered throughout, similarly as ΔN165, while PEST residues exhibited much lower values as a group [Fig. 3.6E, inset]. Thus, the ΔN117 dimer was a fuzzy complex in which the PEST domain did not deviate from a tethered IDR to the structured ETS domain.

3.3.4 Dimeric forms of PU.1 with and without DNA are nonequivalent

The 2:1 complex formed by PU.1 at a single cognate site suggested that the PU.1 dimer was asymmetric, as a symmetric dimer that exposes the DNA contact surfaces would logically yield a 2:2 complex. However, this stoichiometry was excluded by the DOSY titration data, which showed two inflections with the least diffusive species at a DNA: protein ratio of 1:2, corresponding to the 2:1 complex [Fig. 3.1B]. Unbound PU.1 also formed a homodimer, which could logically arise only if the complex was symmetric. Experimentally, a symmetric dimer was strongly inferred by a single set of ^{1}H-^{15}N signals for unbound ΔN117 at high concentrations [Fig. 3.6E]. Moreover, the CD-detected structure of PU.1 showed negligible changes upon titration

by DNA [Fig. 3.8A], in contrast with the self-titration in the absence of DNA [Fig. 3.6D]. These clues suggested that DNA-bound and free PU.1 dimerized into distinct conformers.

To test these notions, we constructed a constitutive ETS dimer via insertion of a single Cys residue into ΔN165, which did not harbor this amino acid, between residues 194 and 195 [Fig. 3.8B]. We targeted this position given its turn conformation in the known structures of the ETS monomer [Protein Data Bank (PDB): 5W3G] and the 1:1 complex (1PUE), and its reported involvement in 2:1 complex formation by heteronuclear NMR (164). Purification of this mutant, termed DKCDK, by ion exchange chromatography under non-reducing conditions eluted monomer and its cysteine-linked dimer at >1 M NaCl [Fig. 3.8C]. Fractions containing predominantly monomer or dimer were separately dialyzed into a buffer containing 0.15 M NaCl with or without 5 mM dithiothreitol (DTT), respectively. In the absence of DNA, the far-CD spectrum of the DKCDK monomer (maintained with 5 mM DTT) overlapped closely with the spectrum of ΔN165 [Fig. 3.8D] and formed the 1:1 complex with cognate DNA similarly as wild-type ΔN165, indicating that the Cys insertion was non-perturbative in the DKCDK monomer [Fig. 3.8E]. In stark contrast, the cysteine-linked DKCDK dimer exhibited a CD spectrum that was altogether unlike PU.1 at equivalent molar concentrations (400 μM). It bore some similarity to a spectrum for ΔN165 at the highest concentration available (800 μM), which contained a greater fraction of dimeric PU.1 (dashed spectrum in Fig. 3.8D). However, the dimeric DKCDK spectrum was further redshifted by ~7 nm and ~15% more intense. Moreover, the DKCDK dimer bound cognate DNA >100-fold more poorly than wild-type ΔN165 [Fig. 3.8E]. Thus, the DKCDK mutant showed that a symmetric configuration was severely perturbed in conformation without DNA and unlike DNA-bound wild-type ΔN165 [compare to Fig. 3.8A]. Together with a deficiency

in DNA binding, the DKCDK mutant demonstrated that the wild-type DNA-bound dimer was not a symmetric species in contrast with the unbound PU.1 dimer.

To assess the feasibility of an alternative, asymmetric configuration in forming the 2:1 complex, which would involve the DNA contact surface, we then examined an R230A/R233A mutant in the DNA-recognition helix H3 of PU.1 [Fig. 3.8B]. The double R→A mutant retained an indistinguishable CD spectrum as wild-type ΔN165 [Fig. 3.8F]. At a sub-saturating concentration of wild-type ΔN165, the addition of the mutant at a concentration that showed no DNA binding on its own nevertheless produced strong DNA loading [Fig. 3.8G]. Such a result would most simply arise if the R→A mutant associated with the wild-type 1:1 complex to drive the 2:1 hetero-complex. The data thus pointed to an asymmetric PU.1 dimer in the 2:1 complex, in which the secondary structure content of PU.1 did not change significantly. Both features contrast sharply with the symmetric conformation required by the DNA-free dimer.

A synthesis of the evidence leads us to propose a model for PU.1 dimerization in the presence and absence of DNA [Fig. 3.8H]. In terms of affinity, the 1:1 active complex is strongly favored (>102 -fold) over either the 2:1 complex or the unbound dimer. Excess PU.1 drives one or the other dimeric state depending on the presence of DNA. The key cornerstone of this model is the nonequivalence of the two dimeric states. Specifically, the incompatibility of the free dimer with DNA binding means that a preexisting dimer cannot serve as an intermediate for the 2:1 complex. Thermodynamic insulation of the two dimeric species leads to a mutually antagonistic relationship, in which the formation of one species is favored at the expense of the other. ΔN117 illustrates this antagonism, as relative to ΔN165, the N-terminal PEST domain promotes dimerization without DNA and reduces the affinity of 2:1 complex formation [Fig. 3.1B]. Together

with enhancing the apparent affinity for 1:1 binding, the result is a widened concentration window for the 1:1 complex for ΔN117.

3.3.5 The electrostatic basis of PU.1 dimerization

The ETS domain as embodied by ΔN165 is highly enriched in Lys and Arg residues, with an isoelectric point (pI) of 10.5. Dimerization should, therefore, be highly sensitive to salt concentration. Contrary to the expectation that the dimer would be stabilized at high salt, which would screen electrostatic repulsion, the opposite was observed. CD-detected self-titration of ΔN165 at 50 mM Na^+ showed a nearly complete two-state transition [Fig. 3.9A] but not at 150 mM Na^+ [compare to Fig. 3.6A]. The low-salt spectra, extended in wavelength to 190 nm and protein concentration to 800 μM because of the reduced Cl^- level, showed the same transition characteristics as acquired at 150 mM Na^+ [Fig. 3.10], indicating that the same transition was inspected at both salt concentrations. Although the transition at 50 mM Na^+ corresponded to a dissociation constant of ~200 μM, it was still fivefold higher than that for ΔN117 in 150 mM Na^+ [Fig. 3.9B]. The data, therefore, reaffirmed the stimulatory role of the disordered PEST domain in dimerization of the ETS domain while revealing an electrostatic basis in the unbound PU.1 dimer.

The sensitivity of the ETS dimer to salt allowed us to access the local structural changes in the DNA-free ETS dimer by NMR spectroscopy. ^{1}H-^{15}N HSQC spectra of ΔN165 with 0.5 to 0.025 M NaCl [Fig. 3.9C] revealed a panel of residues with significant chemical shift perturbations (CSPs). Taking the spectrum acquired in 0.5 M NaCl as the reference for monomeric PU.1, the CSPs exhibited a well-ordered salt dependence [Fig. 3.9C, inset]. The salt-induced CSPs were plotted as a function of residues [Fig. 3.9D], and a cutoff of 0.05 was applied to identify the residues most affected by electrostatic interactions. These perturbed residues were spatially diffuse, as a formal mapping to the unbound PU.1 structure demonstrated [Fig. 3.9E] and did not

overlap with the known residues involved in 2:1 complex formation (164) We also examined the transverse spin relaxation (T_2) properties of the methyl proton peaks in the ^{1}H spectra as a global representation of the tumbling of ΔN165 at different NaCl concentrations [Fig. 3.11]. The effective T_2* relaxation values for the three characteristic methyl ^{1}H peaks at 0.025 M NaCl were up to ~25% lower than at 0.5 M NaCl and well beyond experimental error. This result indicated that the salt-induced CSPs reflected the formation of a slower tumbling dimer.

To correlate the NMR data with the CD-detected changes, we used the heteronuclear chemical shifts to infer secondary structure via the chemical shift index (CSI) (224). The CSI results corroborated the CD-detected loss of α-helical and gain in β/coil content and furthermore localized these changes to helix 1 (H1) and the loop between β sheet 3 (S3) and β sheet 4 (S4) (Fig. 5F). Local H1 unwinding accounted for the CSPs observed near the N terminus of ΔN165, including the particularly strong CSP at Y173, while the loop between S3 and S4 gained β-sheet structure.

3.3.6 Dimeric PU.1 is conformationally destabilized relative to the constituent monomer

The N-terminal IDR promotes a structurally perturbative PU.1 dimer in the absence of DNA. To reveal the underlying conformational thermodynamics of the PU.1 dimer, we performed thermal melting experiments over a range of protein concentrations, using the near-UV CD spectrum from 250 to 300 nm as a probe. The thermal transition was analyzed from a singular value decomposition of the full spectra at each concentration and fitted to a two-state model. The apparent melting temperature (Tm) dropped with increasing concentration in step with the propensity for dimer formation [Fig. 3.9G]. ΔN117 suffered a larger drop than ΔN165 over a ~10-fold increase in concentration. A reduction in salt concentration, which drove dimerization,

similarly caused a larger drop in Tm for ΔN165 (0.15 versus 0.05 M Na^+; [Fig. 3.9G], dashed line). The presentation of [Fig. 3.9G] as Tm^{-1} versus the logarithm of concentration implies that steeper slopes correspond to lower enthalpies (heats) of dissociation/unfolding, which relate to the quality of conformational interactions. To rigorously define the conformational thermodynamics of the PU.1 dimer, we performed differential scanning calorimetry (DSC) experiments on ΔN165 under conditions (salt and protein concentrations) where the quantitatively major population was either monomer or dimer [Fig. 3.9H, all values on a per-mole monomer basis]. The thermograms showed a much greater calorimetric molar enthalpy (area under the curve) for the monomer (300 μM at 0.15 M Na^+) than dimer (500 μM at 0.05 M Na^+). In addition to enthalpy, DSC yields heat capacity changes (ΔCp, difference in the pre- and post-transition baselines) that inform on changes in water-accessible surface area. The ΔN165 monomer exhibited a ΔCp of 3.1 ± 0.3 kJ/(mol K), in good agreement with the structure-based value of 3.3 kJ/(mol K) from the NMR structure of the PU.1 monomer (225). In contrast, the ΔN165 dimer exhibited a significantly reduced ΔCp of 0.97 ± 0.27 kJ/ (mol K). Assuming identical thermally unfolded states, the differences in heat capacity changes indicated that ΔN165 was less well folded than the monomer.

To probe the effect of the N-terminal IDR on the conformational stability of the PU.1 ETS domain, we performed chemical denaturation experiments with urea, which could be reported by intrinsic tryptophan fluorescence at much lower protein concentrations than DSC. At a strictly monomeric concentration (1 μM) at 0.15 M Na^+, ΔN117 was only slightly more sensitive to urea, as judged by the urea concentration at 50% unfolding, than ΔN165 [Fig. 3.9I]. At 100 μM concentration, at which ΔN117 is mostly dimeric but ΔN165 remains monomeric, ΔN117 became significantly more sensitive to urea, suggesting highly perturbative interactions between the PEST and ETS domains. A conformationally perturbed ΔN117 dimer was also implied by its volumetric

properties. The post-transition density slope in [Fig. 3.6A] yields a partial specific volume of ΔN117 of 0.85 ± 0.01 ml/g, which is atypically high for structured globular proteins and suggests altered molecular packing and hydration properties. ΔN165 was more stable at 1 μM at 0.05 M Na^+ than at 0.15 M Na^+, an observation consistent with the ~2°C higher apparent Tm for 10 μM ΔN165 over the same Na^+ concentrations [Fig. 3.9G]. The ΔN165 monomer and dimer were therefore opposite in their conformational stabilities with respect to salt, underlining the structure perturbation by salt or the anionic PEST domain.

In summary, spectroscopic and calorimetric measurements showed that the PU.1 ETS dimer was destabilized with respect to unfolding relative to its monomeric constituents. Structural considerations aside, conformational destabilization contributes to the DNA binding deficiency of the apo ETS dimer. A destabilized dimeric state implied favorable concentration-dependent interactions within the unfolded ensemble over the folded state. The ability of the anionic PEST domain to promote formation of the unbound dimer in ΔN117 therefore further suggests a basis in mitigating the electrostatic repulsion among the cationic ETS domains.

3.3.7 The C-terminal IDR is required for PU.1 dimerization without DNA

In addition to the N-terminal IDR, the structured ETS domain of PU.1 is also tethered at the C terminus to a shorter, 12-residue disordered segment (residues 259 to 270), as apparent in the unbound PU.1 monomer structure [Fig. 3.12A] (225). Far-UV CD spectra at 0.15 M Na^+ showed that hPU.1(117-258) and hPU.1(165-258), termed sΔN117 and sΔN165, respectively [Fig. 3.12B], lacked the secondary structure changes characteristic of ΔN117 and Δ165 across comparable concentrations [Fig. 3.12C; compare to Fig. 3.6D]. sΔN117 was also much less sensitive to urea over the same protein concentration range as ΔN117, and sΔN165 showed no change relative to ΔN165. In contrast to their dimeric deficiency without DNA, sΔN117 and

sΔN165 were both intact with respect to dimerization with cognate DNA [Fig. 3.12D]. While sΔN117 formed 1:1 and 2:1 DNA complexes with the same affinities as ΔN117 in 0.15 M NaCl within experimental error, sΔN165 was a significantly poorer DNA binder than ΔN165 [Table 3]. In particular, 2:1 complex formation was less negatively cooperative for sΔN165, with the concentration window (K_{D2}/K_{D1}) for the 1:1 complex only ~65% that for ΔN165 [Fig. 3.12D]. Last, unlike ΔN165 at 0.05 M Na^+, sΔN165 showed a negligible propensity to dimerize and exhibited biphasic binding with cognate DNA [Fig. 3.12E; compare to Fig. 3.4B]. The divergent impact of removing the C-terminal IDR on dimerization with and without DNA stood in clear agreement with our concept of nonequivalent dimeric states for PU.1 and the structural distinctiveness of the two states.

3.3.8 Phosphomimetic substitutions of the N-terminal IDR reinforce the dimeric propensity of the DNA-free PU.1 dimer

Characteristic of many IDRs flanking DBDs, the N-terminal PEST domain in PU.1 is highly enriched in Glu and Asp residues (pI 3.5), in sharp contrast with the positively charged DBD (pI 10.5) to which it is tethered. The foregoing structural and thermodynamic evidence strongly suggests that the acidic IDR interacts with the ETS domain and shifts it toward dimer formation. Functional studies have established a panel of Ser residues in the PEST domain, including residues 130, 131, 140, and 146 (human numbering), which are multiply phosphorylated in cells (226,227). Phosphoserines at these positions would enhance the anionic charge density by a substantial amount from −11 (3 Asp + 8 Glu) to −17 (~−1.5 per phosphoserine). Because these residues are disordered, we made phosphomimetic substitutions of these residues to Asp, generating a di-substituted (140 and 146, termed D2ΔN117) and tetra-substituted mutant (termed D4N117), to probe their general charge-dependent effects [Fig. 3.13A]. Far-UV CD spectra

showed that the phosphomimetic substitutions progressively drove the affinity of the DNA-free dimer, and the resultant dimers appeared to harbor greater random coil content than their wild-type counterpart [Fig. 3.13B; compare to Fig. 3.6D]. In DNA binding experiments, the di-substituted mutant D2ΔN117 behaved approximately as wild-type ΔN117, while the affinity of the 2:1 complex for the tetra-substituted mutant D4ΔN117 was ~15-fold lower than that for wild-type ΔN117 [Table 3]. Stimulation of the unbound dimer was therefore associated with a marked reduction in the affinity of the 2:1 complex by D4ΔN117 [Fig. 3.13C]. As a result, the selective effect on the 2:1 complex in D4ΔN117 resulted in greater negative cooperativity (i.e., increasing K_{D2}/K_{D1}) in the dimerization of DNA-bound PU.1. In turn, the concentration window for the 1:1 complex widened more than fourfold for D4ΔN117 relative to wild-type ΔN117.

The reinforcing effects of multiple phosphomimetic substitutions in the disordered PEST domain strongly suggest that it influences the behavior of the ordered ETS domain via a generally electrostatic, non-structurally specific mechanism. To further establish this notion, specifically the absence of dependence on structurally specific interactions, we tested the effect of crowding concentrations (in the range of 102 g/liter) of ovalbumin or bovine serum albumin (BSA) on DNA binding by ΔN165 [Fig. 3.13D]. These two anionic proteins share pIs (pI = 5.2 and 4.7 for albumin and BSA, respectively) that are close to the PEST domain (pI = 3.5) but present well-formed globular structures. If PEST/ETS interactions involved structurally specific interactions between the two domains, the anionic crowders should differ significantly from the PEST domain in their effects on DNA recognition by the ETS domain. DNA binding in the presence of up to 20% (w/v) ovalbumin showed little effect on the 1:1 complex [Fig. 3.13D, inset] while progressively decreasing the affinity of 2:1 binding. This behavior mirrored closely the phosphomimetic mutants, and similarly, the more pronounced biphasic appearance in the presence of ovalbumin

was a result of the increased negative cooperativity and widening concentration window for the 1:1 complex. With BSA, an even more anionic crowder, the effect was correspondingly more pronounced. A concentration of 5% suppressed 2:1 binding at 10^{-5} M, an almost 105-fold molar excess of ΔN165 over DNA. Only the 1:1 complex was formed (inset). In contrast, the neutral crowder PEG 8K preserved biphasic DNA binding [Fig. 3.14A], showing that the effects of BSA and ovalbumin were not due to volume exclusion from crowding alone and highlighting the importance of charge. To test our model's prediction that BSA would, therefore, promote the PU.1 dimer, we evaluated ΔN165 labeled with 5-fluoroTrp by ^{19}F NMR in the presence of BSA. The three tryptophan residues in ΔN165 underwent distinct CSPs with 5% BSA under conditions that gave monomers in dilute solution [Fig. 3.14B]. These changes reflected conformational perturbations consistent with unbound dimer formation. Thus, phosphomimetic substitutions and acidic crowding supported non-microstructural electrostatic field interactions on the ETS domain as the basis of the PEST-stimulated dimerization in the absence of DNA.

3.4 Discussion

PU.1 is a markedly inducible transcription factor during hematopoiesis and immune stimulation (218). Open-source repositories such as the Human Protein Atlas show that the expression of PU.1 varies among a panel of resting cell lines by ~25-fold. Independently, single-cell cytometry shows that the abundance of PU.1 transcript in unstimulated murine bone marrow cells ranges from less to 5% to ~50% that of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (228), a housekeeping glycolytic enzyme that is present at ~70 μM in the cell (229). Induction by ligands such as retinoic acid (230) or bacterial endotoxins (231) increases PU.1 expression another 10-fold or more. Depending on the combination of cell line, physiology, and the presence of stimulatory ligands, cellular PU.1 abundance varies by a multiplier comparable to

the ratio of the two dissociation constants K_{D2}/K_{D1} (102 - to 103 -fold) of the 1:1 and 2:1 PU.1/DNA complexes.

In patients and animal models, PU.1 dosage is well established as critical to hematopoietic physiology and dysfunction in vivo (216,232). Dosage effects have been extensively defined in terms of expression, but relatively little is understood about direct dosage effects on transactivation at the protein/DNA level. In this study, manipulation of enhancer syntax in HEK293 cells, which do not use PU.1, demonstrated negative feedback in ectopic PU.1 trans-regulation, independent of modifying interactions with tissue-specific coactivators. The recapitulation of negative feedback, manifested by dose-dependent de-repression of endogenous PU.1 in myeloid THP-1 cells, strongly supports functional relevance in native PU.1-dependent gene regulation. Characterization of the attributable species, a 2:1 DNA complex, revealed two nonequivalent dimeric states that are reciprocally controlled by DNA and the IDRs tethered to the structured DBD. Under physiologic salt conditions, structural alterations that bias unbound PU.1 toward dimerization (e.g., full phosphomimetic substitutions of the N-terminal IDR) oppose dimerization of DNA-bound PU.1. Conversely, alterations that abrogate PU.1 dimerization (e.g., truncation of the C-terminal IDR) promote the formation of the 2:1 DNA complex. Only at low salt conditions (e.g., 50 mM Na^+) are the DNA-free and DNA-bound dimers both favored, and the 1:1 complex is not populated. The tethered IDRs do not appear to become part of the structured dimer but determine preference between the two dimeric states such that the DNA-free dimer remains essentially cryptic without both terminal IDRs.

If the 2:1 complex represents the structural basis of negative feedback, what functional role does the unbound PU.1 dimer play? The thermodynamic relationships among the various states accessible to PU.1 [Fig. 3.8H], critically the nonequivalent free and DNA-bound dimer,

suggest a novel "push-pull" mechanism of PU.1 autoregulation by distinct pools of dimeric protein. By antagonizing the 2:1 complex, we postulate that the unbound dimer suppresses negative feedback and dynamically increases the circulating dose of transcriptionally active PU.1. This model affords for the first time a unifying basis for the PU.1-activating effects of PEST phosphorylation by casein kinase II and protein kinase C (227,233,234), as well as the PU.1-inactivating effects of phosphorylation inhibition by oncogenic transcription factors (235), by biasing PU.1 conformations toward or away from the unbound dimer.

Earlier in vitro studies reporting on dimers of the ETS domain (163,164,221), including our own, did not appreciate their functional significance. The solution NMR structure of the unbound monomer (5W3G) reflects the incomplete context afforded by the ETS domain alone (i.e., ΔN165) at physiologic ionic strength (0.15 M Na^+/K^+). The dissociation constant for the PU.1 dimer in dilute solution (10^{-5} M) should not be misinterpreted as denoting a physiologically irrelevant interaction. The complex formed by PU.1 and its partner GATA-1 is functionally critical in cell lineage specification during myeloid differentiation in vivo (236), but its equilibrium dissociation constant in dilute solution was 10^{-4} M as determined by NMR spectroscopy (237). This and other examples show how volume exclusion and other crowding effects favor interactions in vivo relative to dilute solution. Facilitated diffusion along genomic DNA is also expected to promote cognate occupancy beyond the affinity for oligomeric targets from free solution.

As the tethered IDRs remain disordered in fuzzy PU.1 dimers with and without DNA, their formation is therefore unrelated to paradigms such as induced fit, conformational selection, or fly-casting mechanisms that involve order-disorder transitions by the IDRs (238,239). Instead, the charged intrinsic disorder in PU.1 is involved in through space electrostatic interactions. In the absence of the N-terminal IDR, dimerization of the highly cationic ETS domain (ΔN165) is

favored by low salt. Conformational destabilization of the resultant dimer under these conditions suggests an electrostatic penalty arising from charge-charge repulsion of the DBDs. The stimulatory effect of the negatively charged IDR on the DNA-free dimer, therefore, arises from attenuation of this repulsive penalty of association. A non-structurally specific basis is borne out by the similarly favorable effects of reinforcing negative charges via phosphomimetic substitutions in ΔN117 as anionic crowders on ΔN165. As phosphoserines confer higher charge density (~−1.5) than carboxylates, phosphorylation is expected to exert an even greater effect than the phosphomimetic substitutions. Notably, despite its nominal designation as a proteasome-recruiting signal, the PEST domain does not target PU.1 for metabolic turnover, but it is associated with a local role in dimerization and protein-protein partnerships such as with the lymphoid-specific factor IRF4 (240).

The properties of the flanking IDRs on DNA binding as revealed in this study highlight the divergent roles played by intrinsic disorder within the ETS transcription factor family, which is united by eponymous, their structurally homologous DBDs. Many ETS members are controlled by auto-inhibition, a mechanism that specifically involves short flanking helices in the unbound state that unfolds and disrupts DNA binding allosterically or at the DNA contact interface (241). High-affinity binding to native promoters requires coactivators or homo-dimerization at tandem sites to displace auto-inhibitory helices, forming with positive cooperativity 2:2 complexes (242). The regulatory strategy is activation through recruitment by other coactivators. In Ets-1, the paradigm auto-inhibited ETS member, disordered elements in the serine-rich region (SRR) domain upstream of the auto-inhibitory helices further modulate the regulatory potency of the auto-inhibitory helices. Progressive phosphorylation of the SRR domain reinforces auto-inhibition (240). PU.1, and likely its proximal ETS relatives, upends this paradigm. Lacking auto-inhibitory

domains, high-affinity DNA binding is the default behavior. The disordered elements flanking its structured ETS domain regulate DNA binding by modifying negative feedback. PU.1 is the recruiter in protein-protein partnerships such as IRF4 (63). Phosphorylation of the disordered PEST domain promotes the persistence of the active 1:1 complex and has been established as broadly stimulatory (226,227,233,235). These contrasting features help frame in molecular and functional terms the evolutionary divergence in the ETS family, one of the most ancient families of transcription regulators in metazoan evolution.

As a final remark, both dimeric forms of PU.1 represent highly novel structures. Asymmetric DNA-bound dimers are known in the case of the zinc finger protein HAP1 (243). Zinc fingers are obligate dimers with a 2:2 subunit-to-DNA sub-site configuration. Asymmetry in the HAP1 dimer is directed by the polarity of the DNA sub-sites bound to the subunits. The asymmetric 2:1 complex with PU.1 involves only a single DNA site without significant change in conformation. Functional deficiency of the 2:1 complex as evidenced by the cellular experiments in Fig. 1, therefore, suggests perturbation of DNA structure relative to the singly bound state or denial of specific surfaces of the 1:1 complex to form the transcriptional machinery. In contrast with the localized surface implicated in the 2:1 complex, NMR evidence shows that the residues involved are diffusely distributed with many buried in the PU.1 monomer, leading to a conformationally destabilized dimer. The CSPs observed in the DNA-free dimer, namely, H1 and the wing (S3/S4), were also recently observed for the interaction of PU.1 with a disordered peptide from the SRR domain of Ets-1 (225). These regions may, therefore, represent interaction hotspots for protein/protein partnerships for PU.1 in the absence of DNA. Beyond the minimal ETS domain, the short C-terminal IDR acts in concert with the PEST domain to reinforce the dimerizing properties of the ETS domain. Structurally, this suggests that the two IDRs likely interact

physically, either antagonistically in the monomeric state or cooperatively in the dimeric state. In the cytoplasm, the ETS domain mediates nuclear import of PU.1 (244), so dimerization may also help regulate subcellular trafficking. Further studies to solve the structures of dimeric PU.1 and map their distributions in subcellular compartments will define their dynamics in vivo and contributions to target gene expression.

3.5 Methods and materials

3.5.1 Molecular cloning

DNA encoding fragments of human PU.1 encompassing the ETS domain with and without various segments of its N- and C-terminal IDRs were synthesized by Integrated DNA Technologies (IDT) (Midland, IA) and sub-cloned into the NcoI/HindIII sites of pET28b (Novagen). For truncated constructs harboring the PEST domain (i.e., ΔN117, sΔN117, D2ΔN117, and D4ΔN117), Cys118 was mutated to Ser to facilitate purification and biophysical experiments. Full-length PU.1 used in cell-based experiments was fully wild type. Various PU.1- sensitive enhancer sequences as described in the text were also purchased from IDT and inserted between the AgeI/BglII sites of pD2EGFP (Clontech, CA). All constructs were verified by Sanger sequencing.

3.5.2 Cell culture

THP-1 and HEK293 cells were purchased from the American Type Culture Collection and were routinely cultured in RPMI 1640 and Dulbecco's modified Eagle's medium, respectively, supplemented with 10% heat-inactivated fetal bovine serum. Where indicated, cells were induced with a single dose of PMA at 16 nM for 72 hours (final dimethyl sulfoxide concentration: 0.1%, v/v). All cell lines were maintained at 37°C under 5% CO_2.

3.5.3 Cellular reporter assays

Cellular PU.1 transactivation was measured using a PU.1-dependent EGFP reporter construct under the control of a minimal enhancer harboring only cognate binding sites for PU.1. In PU.1-negative HEK293 cells, the reporter was transactivated in the presence of an expression plasmid encoding wild-type full-length PU.1 and a co-translating iRFP marker (157). Cells (7 × 104) were seeded in 24- well plates and co-transfected with a cocktail consisting of the EGFP reporter plasmid (250 ng) and up to 25 ng of expression plasmids for full-length PU.1, using jetPRIME reagent (Polyplus, Illkirch, France) according to the manufacturer's instructions. The total amount of plasmid was made up to 500 ng with empty pcDNA3.1(+) vector. Twenty-four hours after transfection, cells were trypsinized and analyzed by flow cytometry using an FCS Fortessa instrument (BD Biosciences). Live cells were gated for iRFP and EGFP fluorescence using reporter and full-length PU.1 only controls, respectively, in FlowJo (BD Biosciences) before computing the total fluorescence of the dually fluorescent population.

3.5.4 RT-PCR experiments

Following extraction of total RNA using a spin column kit (Omega) and RT (Thermo Fisher Scientific), RT-PCR reactions were performed on a QuantStudio 3 instrument (Applied Biosystems) with SYBR Green PCR Master Mix (Thermo Fisher Scientific). Expression levels of genes were normalized to *gapdh*. The primer sequences used for *pu.1, csf1ra, e2f1*, and *gapdh* are given in table 2.

3.5.5 Protein expression and purification

Heterologous overexpression in BL21(DE3)pLysS Escherichia coli was performed as previously described (164). In brief, expression cultures in LB or M9 media (the latter containing $^{15}NH_4HCl$ or U-$^{13}C_6$-glucose as required) were induced at an optical density (OD_{600}) of 0.6 with

0.5 mM isopropyl β-d-1-thiogalactopyranoside for 4 hours at 25°C. Uniformly ^{15}N- and $^{15}N/^{13}C$-labeled constructs were expressed in appropriate M9-based media. Harvested cells were lysed in 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.4) containing 0.5 M NaCl by sonication. After centrifugation, cleared lysate was loaded directly onto a HiTrap Sepharose SP column (GE) in 10 mM NaH_2PO_4/ Na_2HPO_4 (pH 7.4) containing 0.5 M NaCl. After extensive washing in this buffer, the protein was eluted in a gradient at ~1 M NaCl in phosphate buffer. Purified protein was dialyzed extensively into various buffers, as described in the text, and diluted as needed with dialysate. Protein concentrations were determined by UV absorption at 280 nm.

3.5.6 DNA binding experiments

DNA binding by protein was measured by steady-state fluorescence polarization of a Cy3-labeled DNA probe encoding the optimal PU.1 binding sequence 5'-AGCGGAAGTG-3'. In brief, 0.5 nM of DNA probe was titrated with protein in a 10 mM TrisHCl buffer (pH 7.4) containing 0.1% w/v BSA and NaCl at concentrations as stated in the text. Steady-state anisotropies $\langle r \rangle$ were measured at 595 nm in 384-well black plates (Corning) in a Molecular Dynamics Paradigm plate reader with 530 nm excitation. The signal represented the fractional bound DNA probe (F_b), scaled by the limiting anisotropies of the *i*-th bound $\langle r_i \rangle$ and unbound states $\langle r_0 \rangle$ as follows:

$$\langle r \rangle = F_b \left(\sum_{i=1}^{n} \langle r_i \rangle - \langle r_0 \rangle \right) + \langle r_0 \rangle = F_b \sum_{i=1}^{n} \Delta r_i + \langle r_0 \rangle \quad (1)$$

where F_b as a function of total protein concentration was fitted to various models as follows. In all cases, the independent variable was the total titrant concentration as taken.

For DNA binding, the two stepwise dissociation constants describing the formation of the 1:1 and 2:1 PU.1/DNA complexes are:

$$\begin{aligned} K_{\mathrm{D1}} &= \frac{[\mathrm{P}][\mathrm{D}]}{[\mathrm{PD}]} \\ K_{\mathrm{D2}} &= \frac{[\mathrm{PD}][\mathrm{P}]}{[\mathrm{P_2D}]} = \omega K_{\mathrm{D1}} \end{aligned} \tag{2}$$

where P and D denote PU.1 and DNA. In this analysis, the binding affinities were not further constrained by interactions of the unbound states. The ratio of $K_{\mathrm{D2}} / K_{\mathrm{D1}} = \omega$ defines the nature of the cooperativity of the 2:1 complex in the paradigm of McGhee and von Hippel. Values of ω below, equal to, or above unity denote the positively cooperative, non-cooperative, and negatively cooperative formation of the 2:1 complex concerning the 1:1 complex, respectively.

In direct titrations of the DNA probe by PU.1, the observed anisotropy change represented the summed contributions of the two complexes as expressed by Eq (1). The most efficient approach is to determine binding in terms of the unbound protein, P. The solution, which is cubic in [P], is:

$$\begin{aligned} 0 &= \varphi_0 + \varphi_1[\mathrm{P}] + \varphi_2[\mathrm{P}]^2 + \varphi_3[\mathrm{P}]^3 \\ &\begin{cases} \varphi_0 = K_{\mathrm{D1}}K_{\mathrm{D2}}[\mathrm{P}]_{\mathrm{t}} \\ \varphi_1 = K_{\mathrm{D1}}K_{\mathrm{D2}} - K_{\mathrm{D2}}[\mathrm{P}]_{\mathrm{t}} + K_{\mathrm{D2}}[\mathrm{D}]_{\mathrm{t}} \\ \varphi_2 = K_{\mathrm{D2}} - [\mathrm{P}]_{\mathrm{t}} + 2[\mathrm{D}]_{\mathrm{t}} \\ \varphi_3 = 1 \end{cases} \end{aligned} \tag{3}$$

where the subscript "t" represents the total concentration of the referred species. [P] was solved numerically from Eq. (3), rather than analytically via the cubic formula, to avoid failure due to loss of significance. With [P] in hand, [D], [PD], and [P_2D] were computed from Eq. (2) and the corresponding equations of state. In the limit of no formation of the 2:1 complex (*i.e.,* $K_{D2} \rightarrow \infty$), Eq. (3) simplifies to a quadratic, corresponding to the formation of only the 1:1 complex:

$$0 = K_{D1}[P]_t + \left(K_{D1} - [P]_t + [D]_t\right)[P] + [D]^2 \quad (4)$$

3.5.7 Competition ESI-MS

Analyses were performed on a Waters Q-TOF (quadrupole orthogonal acceleration–time-of-flight) micro mass spectrometer equipped with an ESI source in positive ion mode. Samples were dialyzed extensively against 0.01 M NH_4HCO_3 (pH 8) and introduced into the ion source by direct infusion at a flow rate of 5 µl/min. The instrument operation parameters were optimized as follows: capillary voltage of 2800 V, sample cone voltage of 25 V, extraction cone voltage of 2.0 V, de-solvation temperature of 90°C, source temperature of 120°C, and collision energy of 3.0 V. Nitrogen was used as nebulizing and drying gas on a pressure of 50 and 600 psi, respectively. MassLynx 4.1 software was used for data acquisition and deconvolution. A multiply charged spectra were acquired through a full scan analysis at mass range from 300 to 3000 Da and then de-convoluted by a maximum entropy procedure (223) to the zero-charge spectra presented. Samples were diluted with dialysate to different concentrations for acquisition and data processing under the same condition.

3.5.8 CD spectroscopy

Spectra were acquired in 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.4) plus NaCl as a function of concentration or temperature as indicated in the text in a Jasco J-810 instrument. Thermal denaturation experiments were performed at 45°C/hour with a response time of 32 s. The path length for near-UV scans was typically 1 or 0.1 mm for far-UV scans. Spectral analysis following blank subtraction and normalization concerning path length and concentration was performed by singular value decomposition as follows. For each experiment, a matrix A with column vectors represented by CD intensities at each protein concentration was factorized into the standard decomposition

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^{\mathbf{T}} \tag{5}$$

Where the left-singular unitary matrix **U** contained the orthonormal basis spectra $\mathbf{u}_i$ scaled by the singular values σ_i from the diagonal matrix **Σ**. The row vectors in the right-singular unitary matrix $\mathbf{V}^{\mathbf{T}}$ gave the concentration- or temperature-dependent contribution of each basis spectrum to the observed data and is termed as *transition vectors* $\mathbf{v}^{\mathbf{T}}$ in the text. For ease and clarity of presentation, the scaling due to Σ is captured into the transition vector i.e., $\mathbf{a} = \mathbf{u}(\sigma\mathbf{v}^{\mathbf{T}})$ (matrix multiplication is associative) which does not affect the fitted parameters. The transition vectors were fitted to titration models describing a two-state transition with dissociation constant *K* as follows

$$X = F_{1n}\left(X_n - X_1\right) + X_1 = F_{1n}\Delta X + X_1 \tag{5}$$

where $X = \sigma\mathbf{v}^{\mathbf{T}}$ and the subscripts "1" and "n" refer to monomer and oligomer (n = 2 for dimerization), respectively. F_{1n} is given by

$$K_n = np_t^{n-1}\frac{(1-F_{1n})^n}{F_{1n}} \tag{6}$$

where F_{1n} is the fractional two-state 1-to-n oligomer at equilibrium, p_t is total protein concentration. As detailed elsewhere (222), a fundamental feature of Eq. (6) is that dimerization is uniquely non-sigmoidal on linear scales, which is diagnostic for two-state dimers. Any higher-order oligomers processes are invariably sigmoidal on linear scales.

3.5.9 Nuclear magnetic spectroscopy

NMR experiments were conducted at 25°C using Bruker BioSpin 500, 600, or 800 MHz spectrometers. For DOSY experiments, unlabeled protein and DNA were co-dialyzed in separate compartments against the required buffer, lyophilized, and reconstituted to 250 μM in 100% D_2O before data acquisition at 500 MHz with a 5-mm TBI probe. For 2D/3D experiments, uniformly labeled ΔN165 and ΔN117 (± unlabeled DNA) were dialyzed against the required buffer at 11/10 excess concentration and adjusted 10% D_2O at 400 to 700 μM protein. The dependence of the DOSY-derived self-diffusion coefficients on total protein concentration were fitted using Eqn. (5) with X corresponding to the diffusion coefficients of the oligomer and monomer.

^{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 1k × 144 data points and zero-filled to 4k × 4k. Steady-state heteronuclear $^{1}H^{15}N$-NOE was acquired at 600 MHz from the difference between spectra acquired with and without ^{1}H saturation

and a total recycle delay of 3 s. The data was processed with TopSpin 3.2 to extract peak intensities and fitted as single exponential decays.

Spectra were assigned with purified $^{13}C/^{15}N$-labeled constructs in a standard suite of 3D experiments: HNCA, HNCACB, HN(CO)CACB, HNCO and HN(CA)CO at 800 MHz using a 5-mm TCI cryoprobe for bound protein to DNA and at 600 MHz using a 5-mm QXI probe for unbound protein. Spectra were processed using NMNRFx software, referenced to DSS, and peak picked/analyzed with NMRFAM-Sparky. Automated Assignments were made using the NMRFAM Pine server, and verified manually.

3.5.10 Fluorescence-detected self-association and protein denaturation

The intrinsic fluorescence from three tryptophan residues in the PU.1 ETS domain was excited at 280 nm and detected at 340 nm with a slit for excitation of 15 nm and 20 nm for emission. Intensity data recorded in the vertical and horizontal polarizer positions were corrected for the grating factor and by blank subtraction. Concentration-dependent anisotropies were fitted to Eqn. (5) and (6), where $X = \langle r \rangle$. For denaturation studies, PU.1 at 100 μM and 1 μM in 10 mM TrisHCl buffer (pH 7.4) containing either 0.15 M or 0.05 M NaCl was titrated with urea. Blank-subtracted intensity data was directly fitted with the linear extrapolation method.

3.5.11 Differential scanning calorimetry

Protein samples were exhaustively dialyzed in against 10 mM NaH_2PO4/Na_2HPO_4, pH 7.4, and 0.05 M NaCl over 48 h with at least three buffer changes. The final dialysate was reserved and used to rinse and fill the reference cell as well as a diluent for the samples. Thermal scans were carried out at 45°C/h from 10°C to 80°C using a Microcal VP-DSC instrument (Malvern). All scans were carried out only when the baseline was reproducibly flat. Thermograms were fitted to two-state transition models. Nonpolar and polar solvent accessible surface area (SASA) for

monomeric ΔN165 was estimated from the solution NMR structure 5W3G (225) based on a 1.4-Å probe. SASA for the unfolded state ensemble was provided by the ProtSA algorithm. The change in SASA in Å was converted to heat capacity change in kJ mol^{-1} K^{-1} using coefficients as follows:

$$\Delta C_p^0 = (0.32 \pm 0.04)\,\Delta A_{\text{nonpolar}} - (0.14 \pm 0.04)\,\Delta A_{\text{polar}} \quad (7)$$

3.5.12 High-precision densimetry

Solution densities ρ were measured in 10 mM TrisHCl, pH 7.4 at 25°C, containing 150 mM NaCl using an Anton Paar Model DMA-5000 vibrating tube densimeter with a precision of 1.5×10^{-6} g/mL. The partial molar volume of the solute $V°$ is determined from the following relationship

$$\rho = \rho_0 + (M - V°\rho_0)c \quad (8)$$

where ρ_0 is the density of the buffer, c is the molar solute concentration, and M is the molecular weight of the solute. For a two-state dimeric species, the observed density is analyzed as follows:

$$\rho_{\text{obs}} = F_{1n}\rho_2 + (1 - F_{1n})\rho_1 \quad (9)$$

Where F_{1n} is as defined by Eq (6), with $n = 2$. Since the observed density varies with the concentration of any species ρ_1 and ρ_2 is each treated as linear functions as described by Eq. (8)

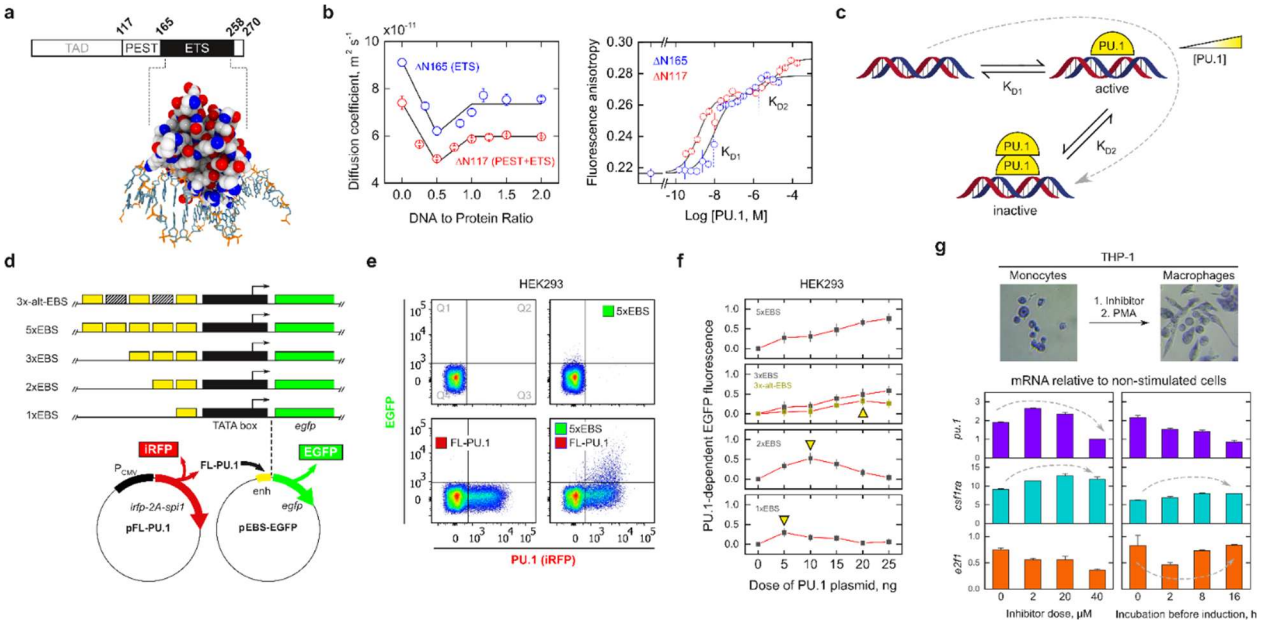


Figure 3-1 PU.1 transactivation is regulated by negative

(A) The ETS domain (ΔN165) is the only structured and minimal DNA binding unit (PDB: 1PUE). (B) Left: DOSY NMR titrations of ΔN117 and ΔN165 with cognate DNA yielding equivalence points at DNA: protein=0.5 and 1.0, corresponding to 2:1 and 1:1 complexes, respectively. The absence of single global minima at DNA: protein = 1:1 formally excludes the possibility of a 2:2 complex. Right: Fluorescence anisotropy titrations with labeled cognate DNA. Both ΔN117 and ΔN165 form a 2:1 complex at a single DNA site with different negative cooperativity as defined by the ratio of the two sequential dissociation constants K_{D1} and K_{D2} (dashed lines; see Materials and Methods). Parametric values are given in table S1. (C) Scheme of negative feedback in PU.1 trans-regulation. A mechanistic link between dimerization and negative feedback predicts a reduction in PU.1 activity under conditions permissive of an inactive 2:1 complex. (D) Synthetic PU.1-dependent EGFP reporters. A minimal TATA box was driven by enhancers composed only of tandem EBS (yellow blocks) spaced 20 bp apart. Hatched blocks represent mutated sites. (E) Representative flow cytometric data of untreated HEK293 cells and upon transfection with a constant dose of the 5×EBS reporter and/or up to 25 ng of an expression plasmid encoding full-length PU.1 (see Materials and Methods). Quadrant Q2 contained the EGFP-positive cells to be counted out of all PU.1-expressing cells (Q2 +Q3). (F) EGFP fluorescence in Q2 taken over the summed fluorescence in Q2 +Q3 at 24 hours after co-transfection of the EGFP reporter plasmid and the indicated dose of PU.1 expression plasmid. Each data point represents the means ± SE of triplicate or more samples. (G) RT-PCR measurements of *pu.1, csf1ra, and e2f1* mRNA abundance (relative to *gapdh*) in THP-1 cells induced with PMA following exposure to either doses of a PU.1 inhibitor for 2 hours (left) or a fixed dose of 20 μM inhibitor for various periods (right). Cells were visualized at ×40 magnification after Giemsa staining.

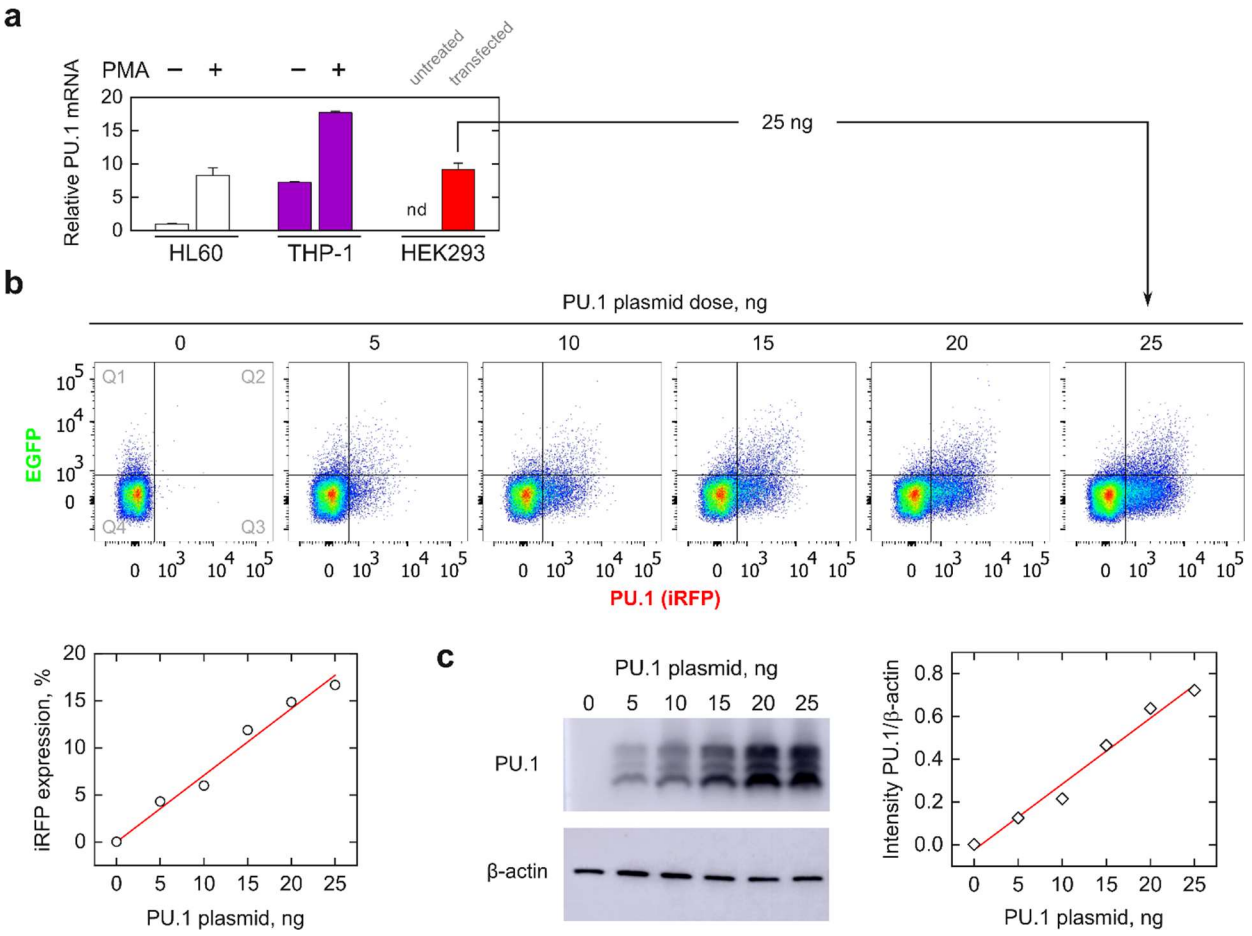


Figure 3-2 Calibration of transgenic PU.1 dosage

a, to determine the dosage of PU.1 transgene that would yield physiologically relevant levels of PU.1, transcripts of the *pu.1* gene were quantified in two model human myeloid cell lines, under normal proliferative conditions as well as under induction by 16 nM of phorbol 12-myristate (PA). Consistent witMh the strong inducibility of PU.1 in primary tissue, PMA induced PU.1 mRNA transcripts in both cell lines. All samples were presented as means ± SE of at least 3 biological replicates relative to *gapdh*. The data establish that a maximum dose of 25 ng (per well of a 24-well plate) of the PU.1 expression plasmid in HEK293 (see Materials and Methods) would yield an expression level comparable to induced HL-60 and non-induced THP-1, but well below levels in that cell line under PMA induction. b, we then determined whether 25 ng of the PU.1-expression plasmid would yield a linear dosing range for the λB reporters following transient transfection in HEK293 cells. PU.1 expression was quantified by its co-translating PU.1 iRFP marker by flow cytometry. Cells in quadrant Q2 and Q3 represented PU.1-expressing cells. The sum of their counts was plotted against the plasmid dosage. The line represents a linear fit to the data. c, to cross-validate the flow cytometric approach, PU.1 abundance was also measured by immunoblotting transfected HEK293 lysates at 10 µg per lane. PU.1 was probed with a rabbit polyclonal antibody (Cell Signaling, #2266) followed by an HRP-coupled mouse anti-rabbit secondary antibody (Santa Cruz). β-actin (loading control) was probed with a mouse monoclonal antibody (Santa Cruz) with an HRPcoupled rabbit anti-mouse secondary antibody (Santa Cruz). Blots were visualized using an Amersham Imager 600 and quantified using ImageJ software.

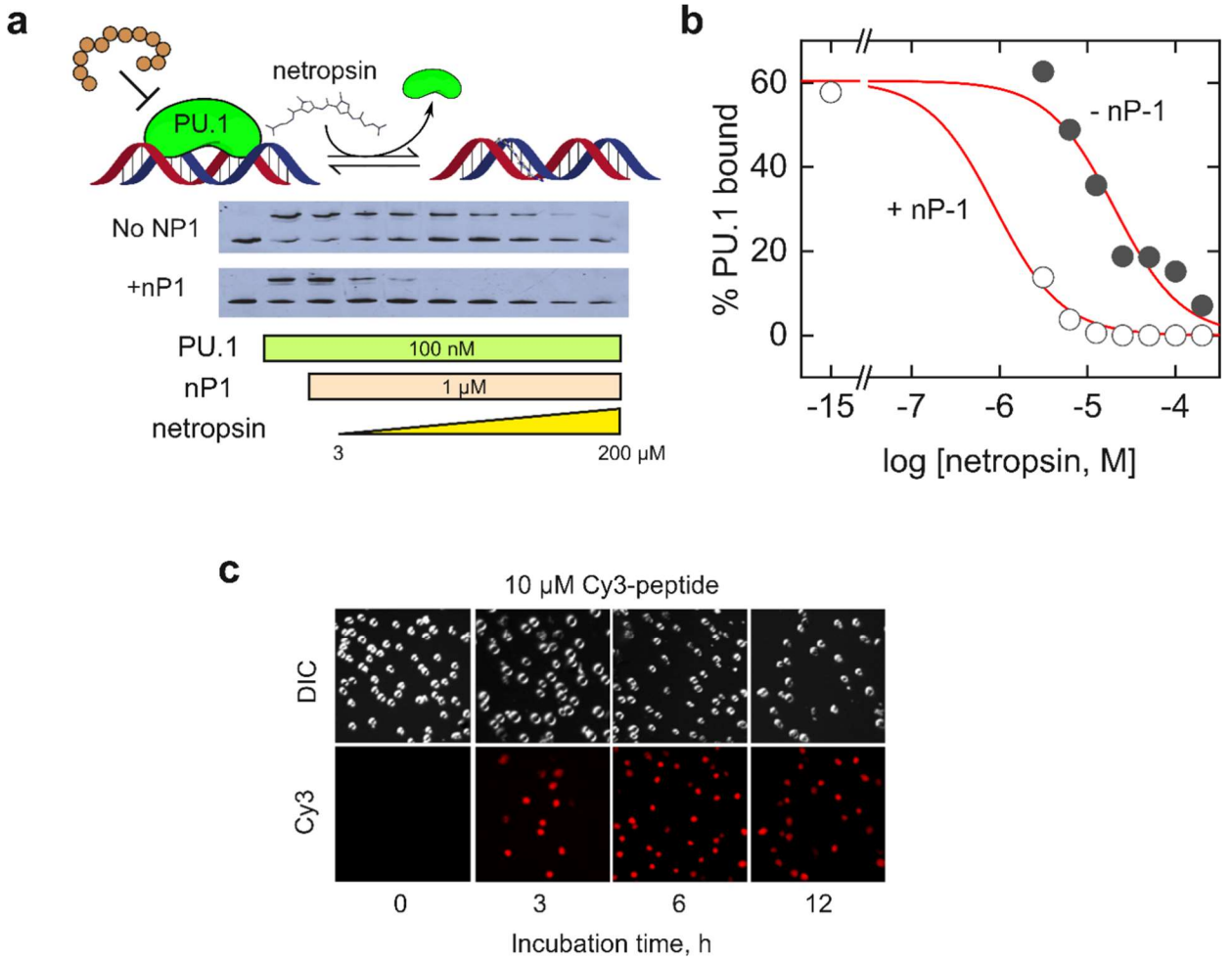


Figure 3-3 Characterization of a peptide-based PU.1 inhibitor

The synthetic dodecameric peptide nP-1 (Ac-RDYHPRDHTATW-NH2) inhibited cognate DNA binding by PU.1 in a competitive gel-mobility shift assay using netropsin. Each netropsin-containing lane represented a 2-fold increase in concentration. In the negative control without nP-1, PU.1 was displaced by netropsin binding to the DNA minor groove adjacent to the 5'-GGAA-3' consensus. b, Quantification of the mobility shift data shows that nP-1 at 1 μM was highly synergistic with netropsin, reducing the apparent IC_{50} by >20 fold. c, Uptake of Cy3-labeled peptide by THP-1 cells at 10 μM, imaged at 63× magnification.

Table 2 Primers in RT-PCR experiments

Gene	⇆	Sequence	Amplicon, bp	T_m, °C
pu.1	F	5′-CTC CAG TAC CCA TCC CTG TC-3′	158	64
	R	5′-CGG ATC TTC TTC TTG CTG CC-3′		
csf1ra	F	5′-CAG AGC CTG CTG ACT GTT GA-3′	263	60
	R	5′-TTG CCC TCA TAG CTC TCG AT-3′		
e2f1	F	5′-AAG GGA AGG AGT CTG TGT GG-3′	214	64
	R	5′-CGA AAG TGC AGT TAG AGC CC-3′		
gapdh	F	5′-CGG AGT CAA CGG ATT TGG TCG-3′	225	60
	R	5′-TCT CGC TCC TGG AAG ATG GTG AT-3′		

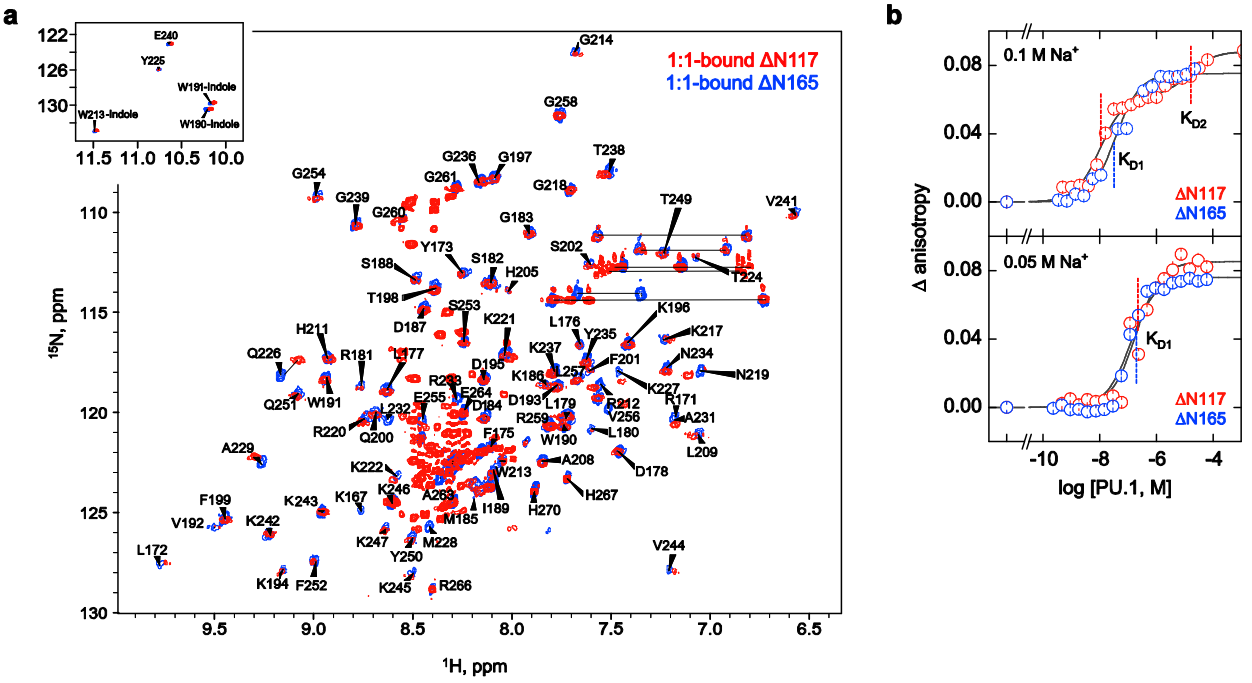


Figure 3-4 The intrinsically disordered PEST domain modifies DNA recognition by PU.1 (A) ^{1}H-^{15}N HSQC of ΔN117 and ΔN165 in the 1:1 complex with cognate DNA. The assignment of the ΔN165 spectrum was 90% complete. (B) DNA binding by ΔN165 and ΔN117 in 0.1 M and 0.05 M NaCl, showing the impact of the PEST domain on the cooperativity of 2:1 complex formation. Parametric values of the equilibrium dissociation constants are given in table 3.8.

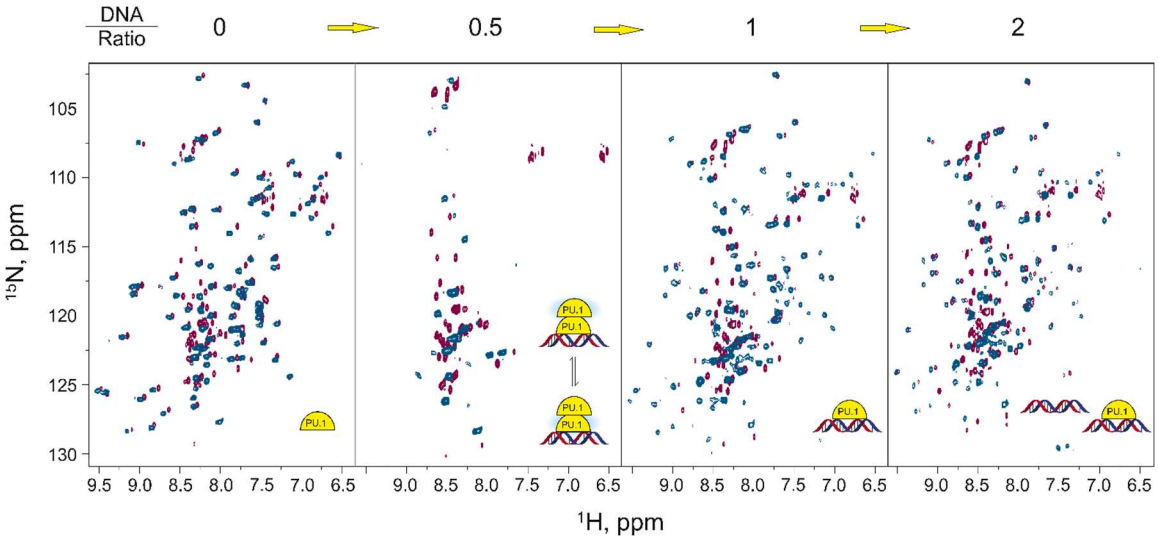


Figure 3-5 ^{1}H-^{15}N HSQC-detected titration of ΔN117 and ΔN165 by cognate DNA
ΔN117 is in red, ΔN165 in blue. Conditions were essentially the same as the DOSY titration in Figure 1a (0.15 M NaCl), except uniformly ^{15}N-labeled protein was used at 400 μM.

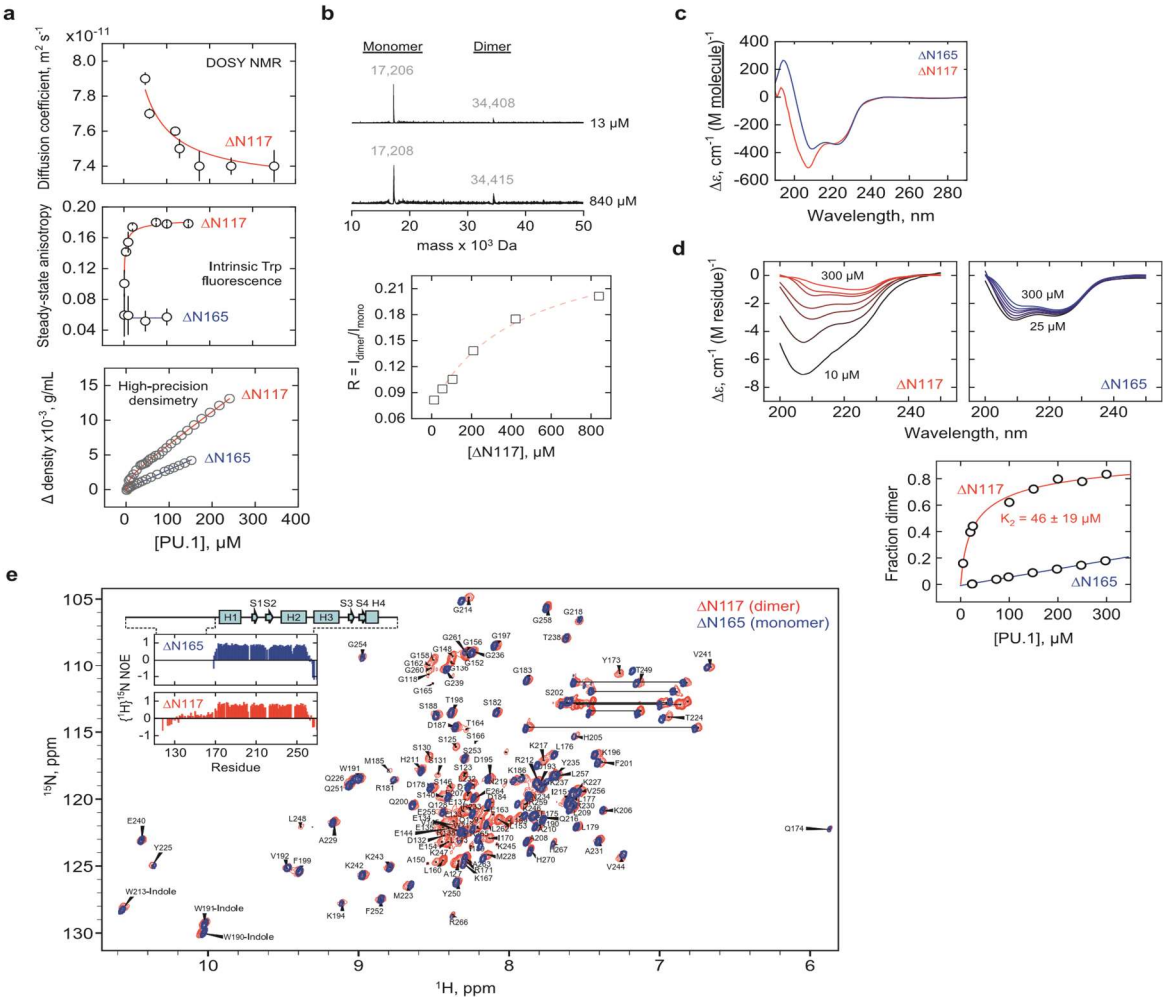


Figure 3-6 The disordered PEST domain in PU.1 drives dimerization in the absence of DNA

a, Concentration-dependent changes in hydrodynamic and volumetric properties by DOSY NMR, intrinsic Trp fluorescence anisotropy, and high-precision densimetry of ΔN117 in 0.15 M Na^+ at 25°C. Red *curves* represent fits of the data to a two-state monomer-dimer transition. b, Representative zero-charge ESI mass spectra of ΔN117 at 13 and 840 μM total concentration, normalized to the height of the monomer (17 kDa) peak. The ratios of the integrated dimer-to monomer intensities (MW showed) were French-curved to guide the eye. c, Far-UV CD spectra of ΔN117 and ΔN165 at 25 μM, plotted on a per-molecule basis to highlight the contribution of the N-terminal residues. d, Concentration-dependent, per-residue spectra of ΔN117 and ΔN165 (left). Dimerization as revealed by the singular value decomposition of the ΔN117 spectra and fitted to a two-state transition. e, ^1H-^{15}N HSQC of 400 μM ΔN117 and ΔN165 at 0.15 NaCl. Under these conditions, ΔN117 was predominantly dimeric and ΔN165 was monomeric. The assignments shown are for ΔN117. *Inset*, ^1H-^{15}N-NOE for ΔN117 and ΔN165.

DNA binding data were fitted to titration models as detailed in *Materials and Methods*. The choice of fitting with one or two dissociation constants was made based on the magnitude of anisotropy change and statistically by the Fisher's *F*-test on sums of squares at $p = 0.05$. Estimates in μM are given as means ± SE of three or more replicate experiments. n.d., not detectable.

Table 3 DNA-binding and self-association equilibrium constants of PU.1 constructs

K_{D1} and K_{D2}, μM	$[Na^+]$, M		
	0.15	0.10	0.05
ΔN117	0.0038 ± 0.0019 7.1 ± 3.3	0.010 ± 0.001 15 ± 5	0.19 ± 0.05
ΔN165	0.0091 ± 0.0015 1.9 ± 0.9	0.036 ± 0.003	0.13 ± 0.02
sΔN117	0.0051 ± 0.0007 6.0 ± 1.3		0.0074 ± 0.0022 3.5 ± 1.8
sΔN165	0.11 ± 0.03 15 ± 9		0.13 ± 0.02 32 ± 24
D_2ΔN117	0.0056 ± 0.0008 6.7 ± 3.2		
D_4ΔN117	0.013 ± 0.004 108 ± 39		
DKCDK monomer	0.013 ± 0.03 2.5 ± 0.4		
DKCDK dimer	3.1 ± 0.9 n.d.		
R232A/R235A	>0.5		

Dissociation constants of 1:1 and 2:1 complexes in the presence of co-solutes

K_{D1} and K_{D2}, M	[Co-solute], % w/v		
	15%	20%	
Ovalbumin	$(3.8 \pm 0.7) \times 10^{-9}$ $(9.5 \pm 1.1) \times 10^{-6}$	$(6.3 \pm 0.6) \times 10^{-9}$ $(3.2 \pm 0.5) \times 10^{-5}$	
	5%	10%	
BSA	$(7.9 \pm 1.5) \times 10^{-11}$ $>10^{-4}$	$(3.0 \pm 1.9) \times 10^{-11}$ $>10^{-4}$	
	10%		10%
PEG 8K	$(1.6 \pm 0.7) \times 10^{-8}$ $(2.1 \pm 1.5) \times 10^{-6}$	Ethylene glycol	$(7.0 \pm 1.6) \times 10^{-9}$ $(6.2 \pm 1.6) \times 10^{-6}$

[NaCl] = 0.15 M

Two-state dissociation constants of ΔN117 and ΔN165 constructs by CD spectroscopy

(Affinities at 0.15 M NaCl unless otherwise indicated)

Construct	K_2, μM
ΔN117 (wildtype)	46 ± 20
sΔN117	≫300
D_2ΔN117	25 ± 3
D_4ΔN117	13 ± 1
ΔN165 (wildtype)	≫800 202 ± 72 (0.05 M NaCl)
sΔN165	n.d.

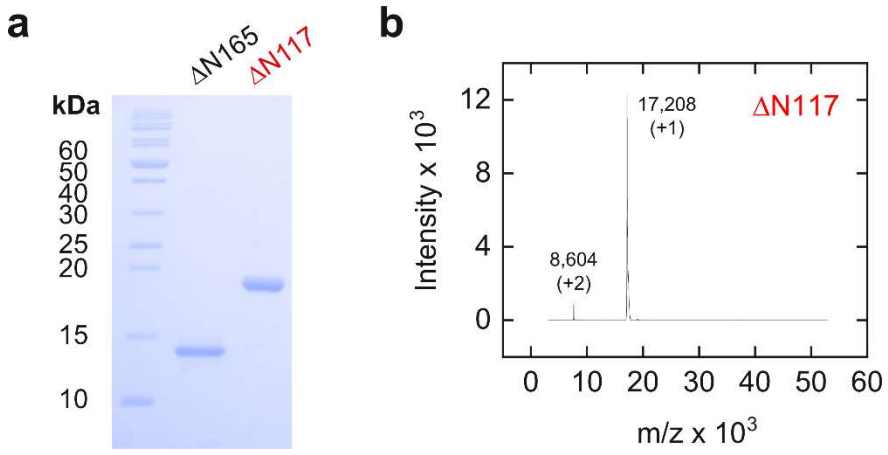


Figure 3-7 Purity of recombinant PU.1 constructs
a, one µg of each purified PU.1 construct was resolved in a 15% polyacrylamide gel (29:1) under standard denaturing (0.1% SDS) and reducing conditions (0.5 mM TCEP). b, MALDI-ToF mass spectrum of ΔN117.

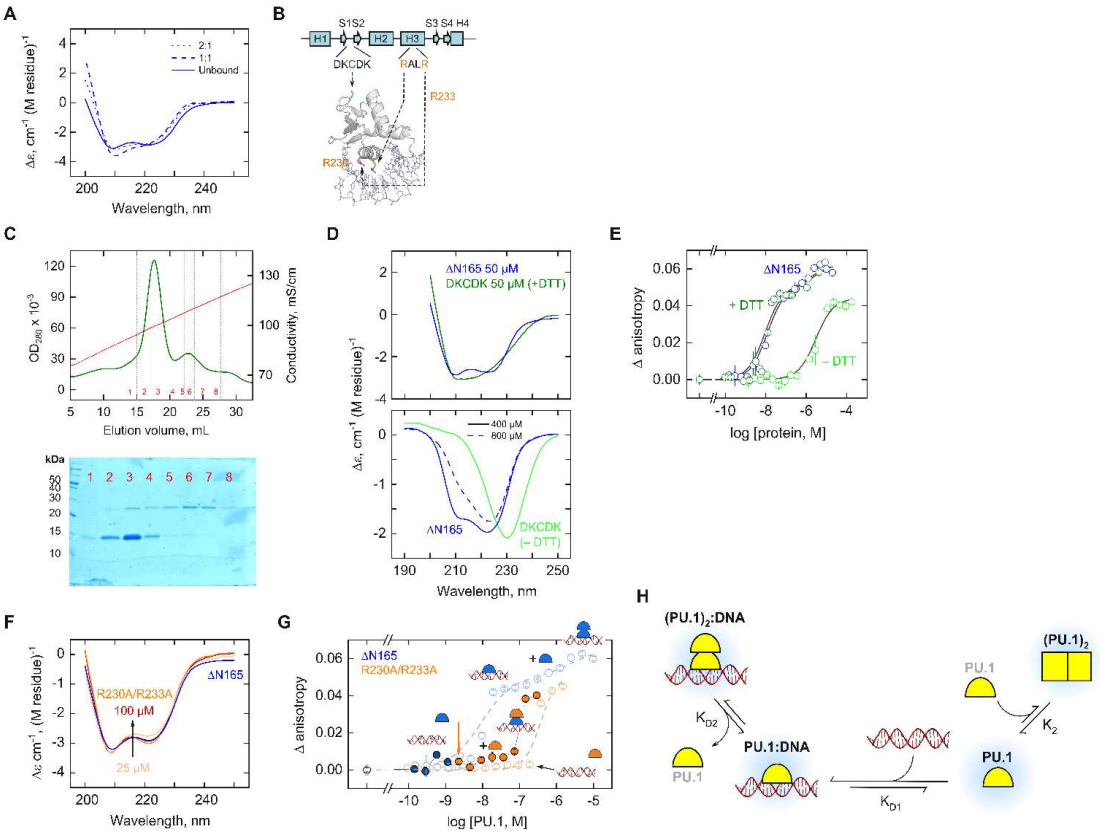


Figure 3-8 Mutations demonstrate non-equivalent PU.1 dimers with and without of DNA
a, Far-UV CD spectra of the DNA-bound ΔN165 upon subtracting the spectrum of the cognate DNA acquired under identical conditions (75 μM, 0.15 M Na^+). b, Residues involved in the DKCDK and the binding-deficient mutant (R230A/R233A). The structure is homology-modeled against the co-crystal 1PUE. c, Purification of the DKCDK mutant by ion-exchange chromatography under non-reducing conditions. The lysate was loaded at 0.5 M NaCl and extensively washed before elution over a linear gradient to 2 M NaCl. SDS-PAGE of purified fractions are shown. Fractions containing primarily monomer (e.g., 1 and 2) or dimers (e.g., 5 onwards) were concentrated and dialyzed separately into a buffer containing 0.15 M NaCl with or without 5 mM DTT, respectively. d, CD spectra of the DKCDK monomer (*top*) and dimer (*bottom*) under various conditions with wild type ΔN165 as reference. The spectrum for the DKCDK monomer was less well resolved due to the presence of DTT which contributed to the total absorption of the sample at 50 μM protein. See the text for details. e, Fluorescence anisotropy measurements of cognate DNA binding by monomeric and dimeric DKCDK with wild type ΔN165 as reference. f, CD spectrum of 25 to 100 μM of the R230A/R233A mutant, with ΔN165 at 25 μM as reference. g, DNA loading by the R230A/R233A mutant in the presence of wild type ΔN165 (solid symbols). Concentrations of the mutant and wild-type protein that individually failed to bind DNA collaborated to bind DNA as a hetero-complex. h, Proposed model for the formation of two non-equivalent PU.1 dimer: an asymmetric one in the 2:1 DNA complex and a symmetric one without DNA.

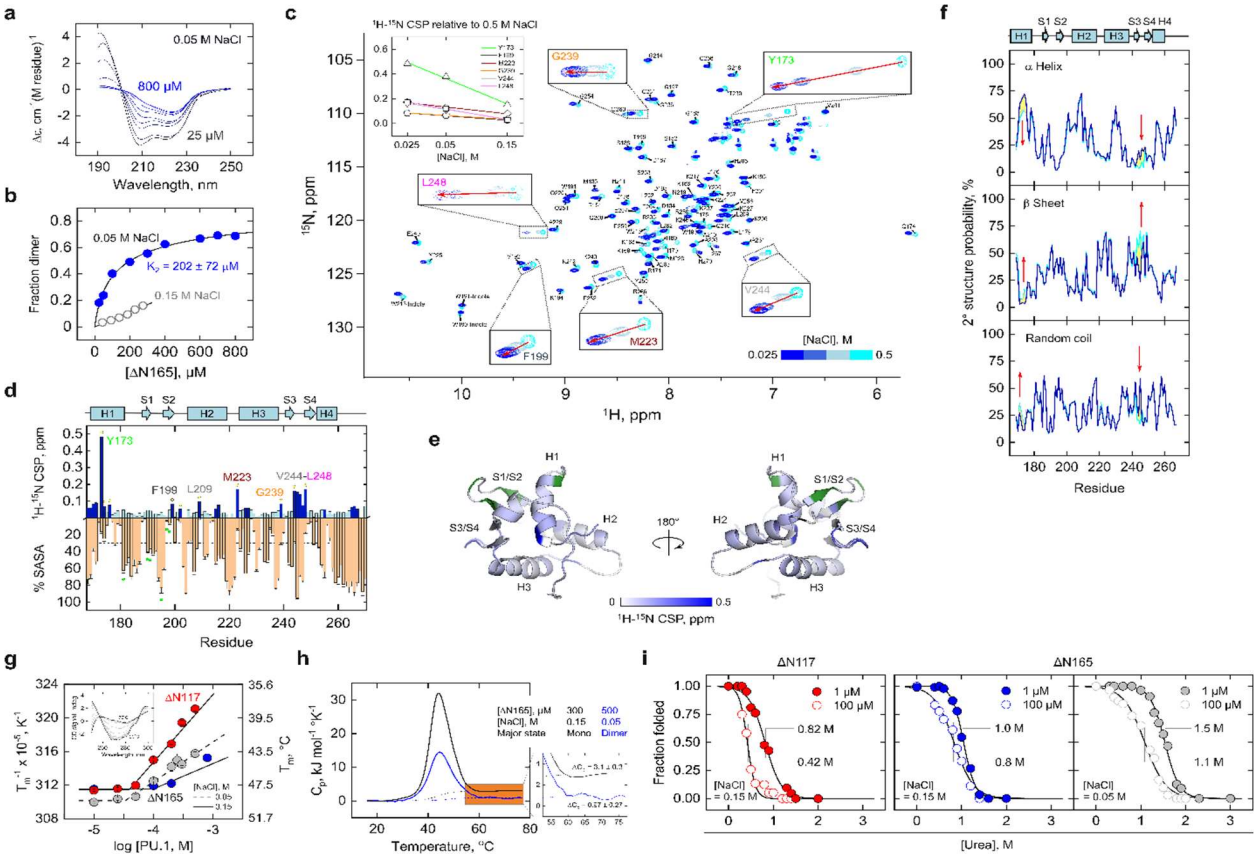


Figure 3-9 Dimerization of the DNA-binding domain of PU.1 is electrostatically mediated and conformationally destabilizing

a, CD-detected titration of ΔN167 at 50 mM NaCl from 25 to 800 µM. b, Analysis of the titration by singular value decomposition yields a two-state transition with a dissociation constant K_2 of 202 ± 72 µM. c, ^{1}H-^{15}N HSQC as a function of salt from 25 to 500 mM NaCl. Residues with the strongest ^{1}H-^{15}N chemical shift perturbations (CSPs; Y173, M223, G239, V244, and L248) are boxed. *Inset*, salt dependence of the CSPs of these residues. d, Summary of the residue CSPs with the average %SASA from the unbound NMR monomer, 5W3G. Residues above a 0.5 ppm cutoff are colored in dark blue and the subset of internal residues (<35% SASA, based on the termini) are marked with yellow circles. Residues implicated in the *DNA-bound* dimer are marked with green circles. e, Mapping of the high-CSP residues to 5W3G. Green residues mark known residues involved in the DNA-bound dimer (*20*). f, Chemical shift-derived secondary structure prediction via the chemical shift index using ^{1}H and ^{15}N signals. The color scheme follows the HSQC in *Panel C.* Regions with significant changes in secondary structure are marked by arrows. g, Near-UV CD-detected thermal melting of ΔN117 and ΔN165. Two salt concentrations were evaluated for ΔN165 (blue and grey). *Inset*, Representative near-UV CD spectra. h, DSC thermograms (solid) for ΔN165 under conditions (salt and concentrations) in which the protein was primarily monomeric or dimeric. The ΔC_p values are given in kJ (mol monomer)$^{-1}$ K^{-1}. Dotted *curves* represent the two-state transition for a monomer (black) and dimer (blue). i, Trp fluorescence-detected denaturation by urea of ΔN117 and ΔN165 at two monomer concentrations. *Curves* represent fit to the linear extrapolation model for a two-state dimer. The marked concentrations represent urea concentration at 50% unfolding.

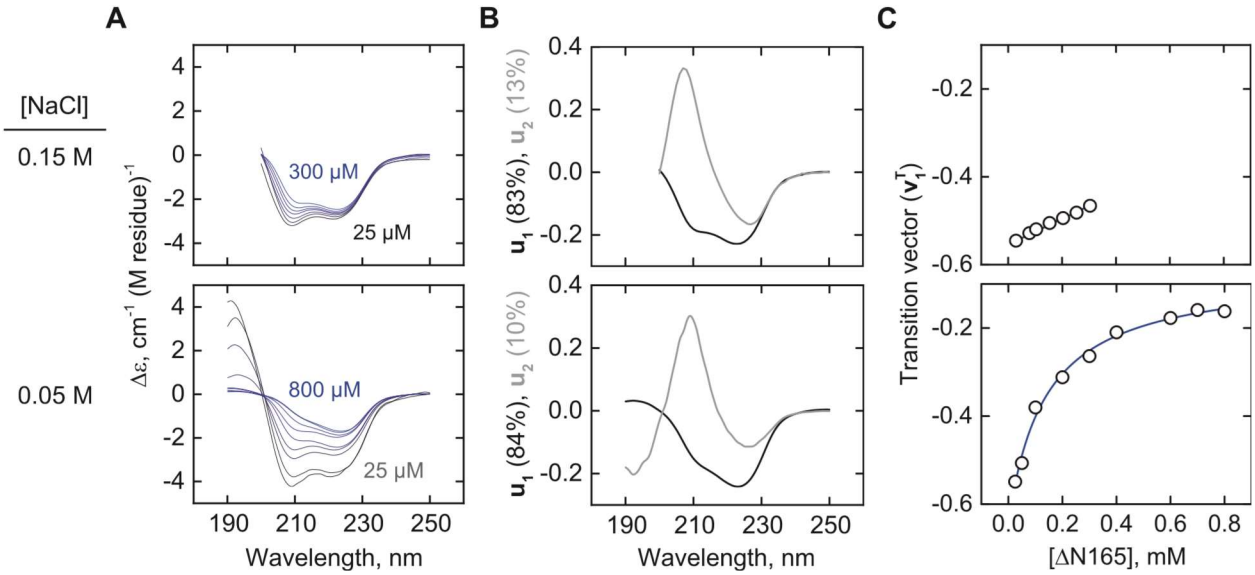
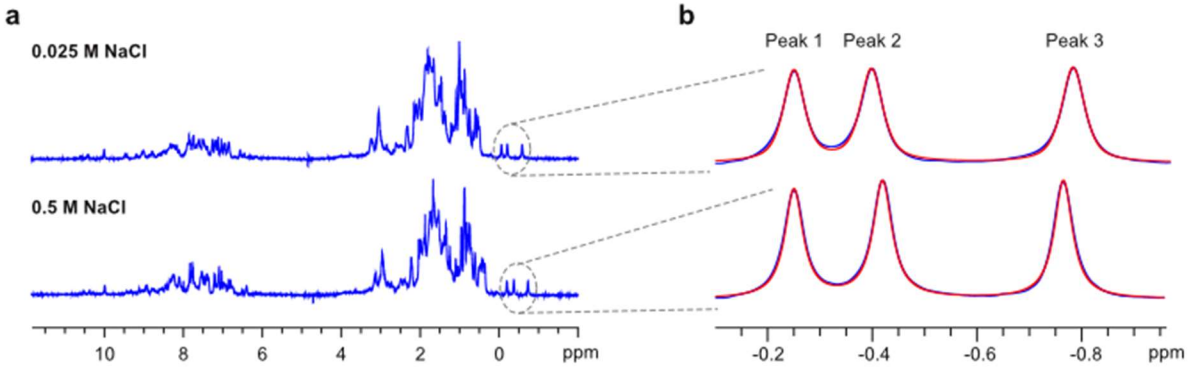


Figure 3-10 Spectral analysis of far-UV CD of ΔN165 in 0.15 and 0.05 M NaCl
a, Blank subtracted, concentration-corrected spectra. Absorption by Cl^- limits the useable ranges of wavelengths and protein concentration in 0.15 M NaCl to those shown. b, following singular value decomposition (see *Materials and Methods* for details), the two most dominant basis vectors in U. c, Progress curves of the concentration dependence as represented by the transition vectors V^T. The 0.05 M transition was fitted to a two-state dimer. The dissociation constant is 202 ± 72 µM.



c

	0.025 M NaCl		0.5 M NaCl	
	$\Delta\nu_{½}$, Hz	T^*_2, s	$\Delta\nu_{½}$, Hz	T^*_2, s
Peak 1	24.2 ± 0.2	0.0132 ± 0.0001	21.9 ± 0.1	0.0145 ± 0.0001
Peak 2	25.3 ± 0.2	0.0125 ± 0.0001	22.9 ± 0.1	0.0139 ± 0.0001
Peak 3	27.6 ± 0.2	0.0115 ± 0.0001	21.9 ± 0.1	0.0145 ± 0.0001

Figure 3-11 Salt-dependent line broadening of methyl protons in ΔN165

a, Methyl protons located at negative chemical shifts of ^{1}H spectra, acquired at 500 MHz, of identical concentrations (350 μM) of ΔN165 in the presence of 0.025 and 0.5 M NaCl, respectively. b, the linewidths, converted to Hz = s^{-1}, were fitted to a sum of Lorentzian peaks (red) from which the full widths at half maximum ($\Delta\nu_{½}$) were estimated by nonlinear regression. c, the effective T_2^* for each peak was then computed by:

$$T_2^* = \frac{1}{\pi \Delta \nu_{½}}$$

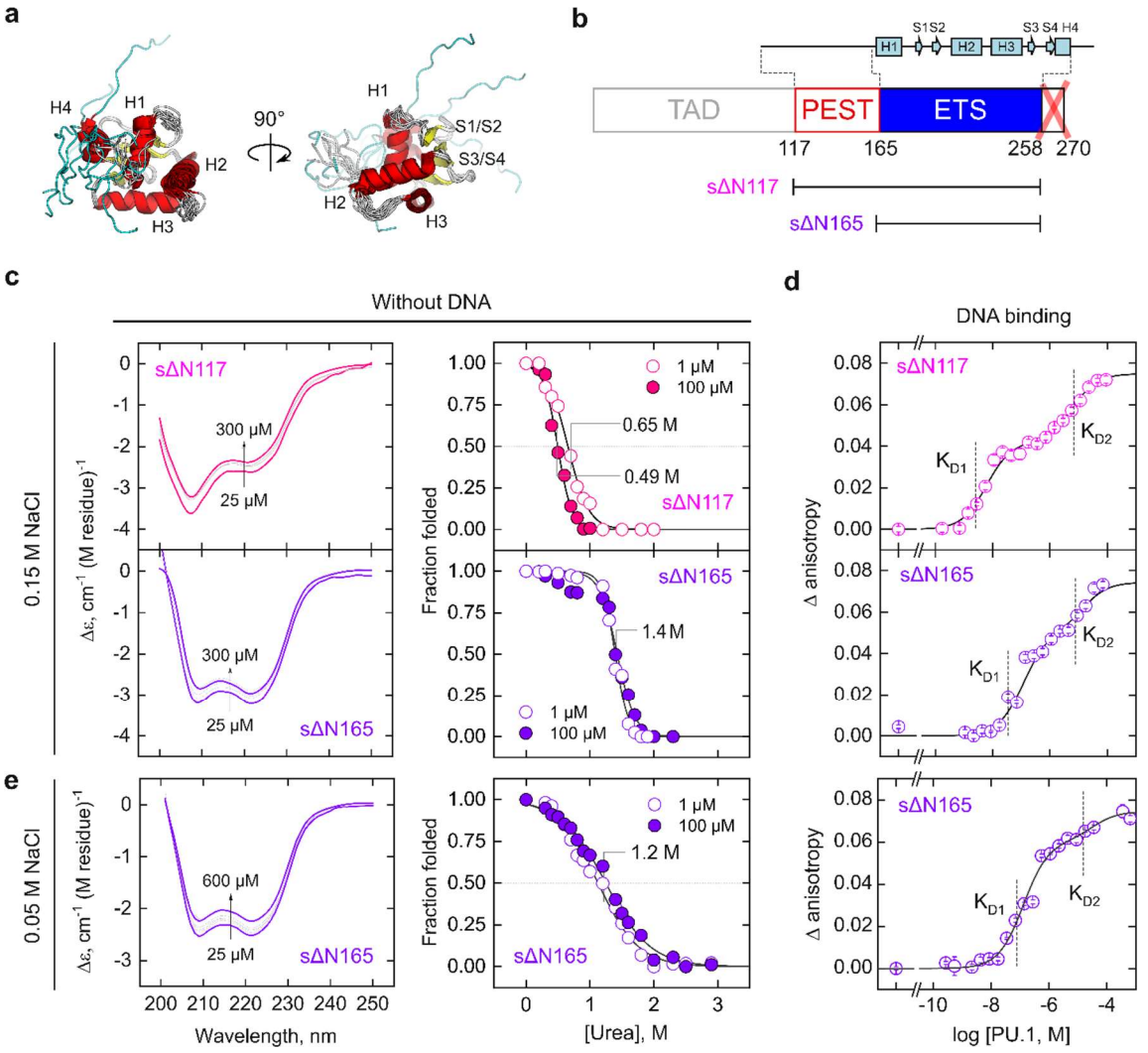


Figure 3-12 The short C-terminal IDR is required for DNA-free PU.1 dimerization
a, Superposition of the 10 models in the solution NMR structure of the PU.1 monomer, 5W3G. The C-terminal IDR is colored cyan. b, Schematic of PU.1 constructs, sΔN117, and sΔN165, without the C-terminal 12-residue IDR. c, Far-UV CD spectra (left) and urea denaturation curves (right) of sΔN117 and sΔN165 at 0.15 M Na^+. The molarity stated represents urea concentrations at 50% unfolding. d, Cognate DNA binding by sΔN117 and sΔN165 at 0.15 M Na^+. e, Corresponding data for sΔN165 at 0.05 M Na^+. In contrast with ΔN165 [*c.f.* Figures 3.11a and b], sΔN165 shows a negligible propensity for dimerization without DNA but exhibits biphasic DNA binding.

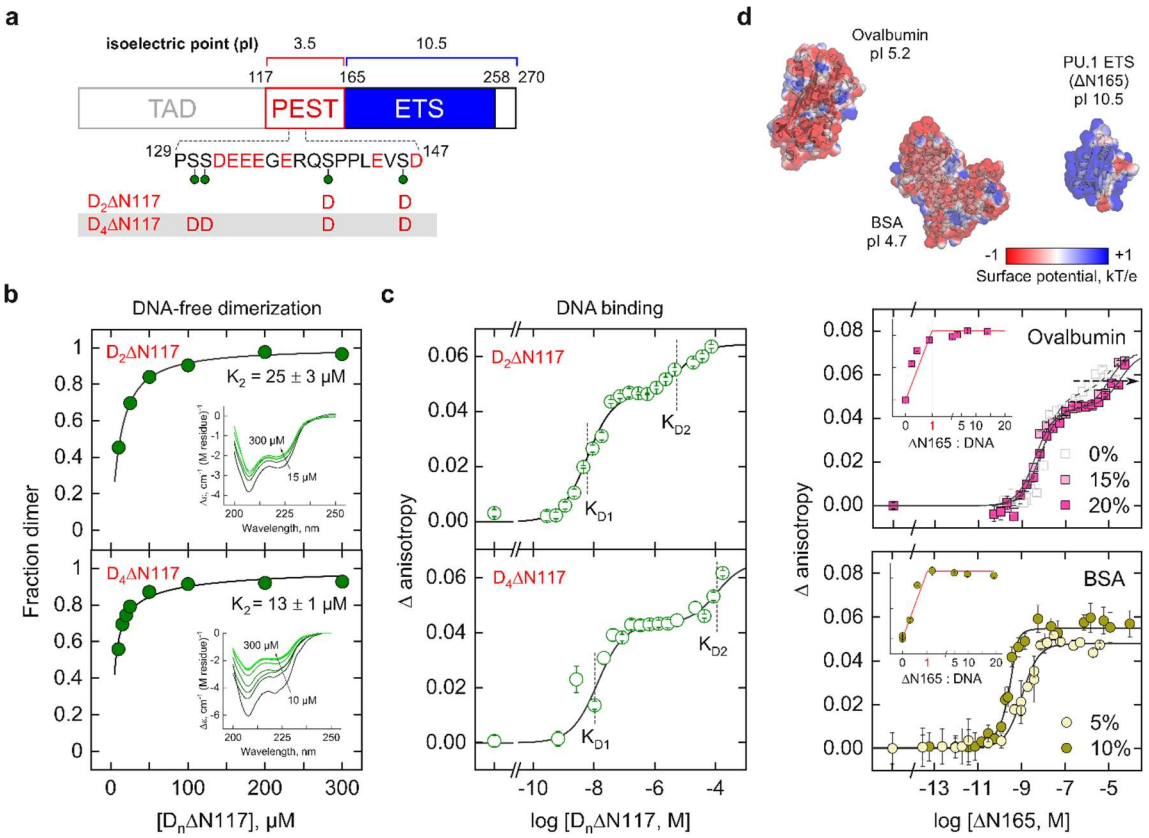


Figure 3-13 Phosphomimetic substitutions at the PEST domain and charged crowding demonstrate a general electrostatic basis of IDR/ETS interactions

a, Schematic of the phosphorylated Ser residues in human PU.1, marked by green pins. The Ser → Asp substituted positions in $D_2\Delta N117$ and $D_4\Delta N117$ are shown. b, $D_2\Delta N117$ and $D_4\Delta N117$ exhibit enhanced dimeric propensities without DNA relative to ΔN117 [*c.f.* Figure 3.7d]. c, $D_2\Delta N117$, and $D_4\Delta N117$ are progressively impaired in 2:1 DNA complex formation. d, the anionic crowders ovalbumin, and BSA modulate DNA binding by ΔN117 in a similar way as the phosphomimetic substitutions on ΔN117, to an extent that correlates with their sizes and low isoelectric points. *Inset*, stoichiometric determination using 1 μM DNA (first binding transition in the case of ovalbumin). The spacing of the ordinates is identical to the main plots. The surface potentials of the structures were computed by APBS at 0.15 NaCl.

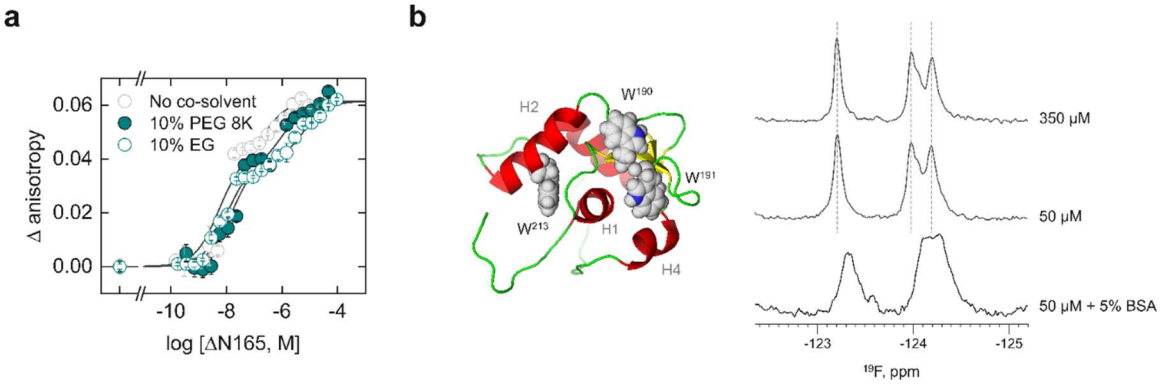


Figure 3-14 Effect of macromolecular crowding on dimerization of ΔN165 in the unbound and DNA-bound state

a, Charge-neutral crowding with PEG 8K on cognate DNA binding by ΔN165. Representative titrations in the presence of 10% w/v ethylene glycol (EG) showed destabilization consistent with osmotic stress as previously reported (Poon, J Biol Chem 287, 18297-18307, 2012). The presence of the same mass concentration of polymeric PEG 8K was similar. In both cases and stark contrast with the acidic crowder BSA, the 2:1 complex was preserved. b, ^{19}F NMR reveals conformational perturbations in ΔN165 consistent with unbound dimer formation. The three Trp residues (190, 191, and 213) in ΔN165 were isotopically labeled with 5-fluoroindole in minimal M9 media. Chemical shifts were acquired at 0.15 M Na^+ in D_2Oreconstituted buffer and referenced against the published value for 5-fluoroindole in 90% D_2O (Sarker et al., Biochemistry 55, 3048-3059, 2016). No concentration-dependent chemical shift perturbation (CSP) was detectable at 350 and 50 μM at ΔN165, in contrast with the presence of 5% w/v BSA. Line broadening in the BSA-containing sample was expected due to increased viscosity.

4 DYNAMIC MECHANISMS OF DNA TARGET SELECTIVITY BY THE ETS-FAMILY TRANSCRIPTION FACTOR PU.1

This work is still in the writing phase prior for submission

4.1 Introduction

Transcription factors represent a significant functional component of the eukaryotic proteome. In virtually all human tissue types, transcription factors alone comprise ~6% of expressed genes (25) As DNA-binding proteins, DNA sequence selectivity, summarized bioinformatically as "motifs," represents core attributes in our knowledge base of transcription factors (34). These factor-binding motifs, for which comprehensive collections and catalogs have been curated, are indispensable for the functional assignment of non-protein-coding sequences. At the same time, motif matching with enriched sequences in ChIP-seq data (245) has established the importance of intrinsic target specificity in transcription factor function *in vivo (246).* The structural origins of the DNA sequence preferences of transcription factors (39) as primarily embodied by their DNA-binding domains (DBDs, are therefore fundamental to our understanding of gene regulation.

A long-standing challenge in transcription factor/DNA recognition is posed by the selectivity mechanisms of proteins harboring DBDs that are highly structurally homologous. The ETS family of transcription factors, of which 28 members are expressed in humans and are found in every major tissue, has been an important model for this problem (110,247,248) ETS proteins share a structurally conserved DNA binding domain, also referred to as the ETS domain of ~85 residues (109). The ETS domain is a winged helix-turn-helix motif (wHTH) where the recognition helix (H3) inserts into the DNA major groove harboring the core consensus 5'-GGA(A/T)-3' and makes additional backbone-mediated contacts with sequences flanking both ends of the consensus.

The flanking sequence variation of ETS-binding motifs represents the primary basis of classification of ETS family members into four classes (110). DNA preferences among ETS domains correlate with their amino acid sequence alignment, which is strongly conserved near the consensus-contacting recognition helix H3 [Figure 4.1], but significantly divergent elsewhere.

To date, the mechanisms of flanking sequence discrimination by ETS proteins remain enigmatic. The markedly segmented pattern of ETS/DNA contacts, where flanking bases are bound via backbone phosphate or sugar contacts, have suggested an indirect readout mechanism in which the sequence-dependent propensity of DNA conformation (249). While this interpretation dovetails with the currently favored emphasis in DNA shape (250) structural definition of this propensity has remained unclear. Moreover, several co-crystallographic (4MHG, of ETV6) and solution NMR structures (2STW, of Ets-1) of ETS members show evidence of contact with flanking nucleobases. These examples of direct readout at flanking positions involve residues near the transition of H3 and a preceding extended loop (the turn-helix in wHTH). In NMR structures, the sidechains of these residues are notably dynamic, exhibiting various conformations that alternately contact a 5'-flanking nucleobase or avoid DNA altogether. These features, together with the absence of a viable indirect readout mechanism, prompted us to ask whether dynamic direct readout might serve as the mechanism of flanking sequence selection. To this end, an attractive model is PU.1, an ETS-family relative (Class III) that exhibits preferences for 5'-flanking sequences that differ distinct from the vast majority of its ETS relatives (Classes I and II).

DNA-bound PU.1 was the first ETS domain whose structure (1PUE) was solved by crystallography (143). Unfortunately, the recombinant crystallized domain carried a Gln→Glu point mutation (murine residue 228) in the same neighborhood (loop-H3) where direct readout of 5'-flanking regions was implicated. In a recent solution NMR structures of the unbound state

(5W3G), which harbored the native Gln, this residue exhibited strong sidechain conformational dynamics (225). To resolve the role of this Gln in flanking sequence selection, we therefore re-examined the DNA-bound complex crystallographically. We co-crystallized the sequence-correct ETS domain of human PU.1 (which is identical to the murine ortholog) with the highest-affinity DNA target (197), with a view of maximizing the potential for capturing direct readout. This new co-crystallographic structure of PU.1/DNA (PDB code 7K1T), solved to a resolution of 1.3 Å, unveiled how PU.1 recognized a highly preferred flanking sequence by direct readout via its native loop-H3 Gln residue (human position 226).

4.2 Results

4.2.1 Comparison of high-affinity PU.1/DNA co-crystal structures

The new bound structure of PU.1 (7K1T) at 1.34 Å [Figure 4.2A] [Table 4] resolution presents a thorough profile on the binding properties of PU.1, hydration patterns of the complex, and putative placement of hydrogens during crystallographic refinement [Figure 4.3]. The 7K1T complex can be easily compared to its predecessor complex 1PUE solved in 1996 (143) [Figure 4.2B] and to some extent to the unbound NMR structure (PDB: 5W3G) solved in 2017 [Figure 4.2C]. The high resolution structure 7K1T is composed of PU.1 (human residues 169-259) and DNA 16 bp (5’ AATAAGCGGAAGTGGG, 3’ TATTCGCCTTCACCCT), 1PUE represents PU.1 (human residues 169-256) and 16 bp DNA (5’AAAAAGGGGAAGTGGG, 3’ TTTTCCCCTTCACCCT) and 5W3G represents unbound PU.1 (human residues 165-270).

The two crystallographic complexes differ in many aspects: 7K1T is a monomer in the asymmetric unit, with a known biological DNA flanking sequence of 5’AGC and a longer sequence of protein by three residues at the C-terminus that initiate a formation of Helix 4. In comparison 1PUE has two monomeric complexes in the asymmetric unit, a DNA flanking

sequence optimized for crystallography, and the formation of helix 4 at the c-terminus is not present in the structure. Interestingly the crystal packing of 7K1T [Figure 4.4] and 1PUE differs with 7K1T crystallizing in P 1 21 1 space group (density of unit cell being 3.84g/cm^3) and 1PUE crystallizing in C 1 2 1 space group (density of unit cell being 1.54 g/cm^3) partially giving rise to the difference in resolution.

The differences in DNA sequence and crystal packing between 7K1T and 1PUE drive slight changes in the B-factor distribution on the DNA [Figure 4.2D]. The DNA for 1PUE gives a sharp contrast in the distribution of B-factors (red: high B-factors; blue: low B-factors) where the high B-factors occur in the flanking bases and the low B-factors in the nucleotides adjacent or complementary to the core consensus 5'-GGAA-3'. By comparison, 7K1T does not exhibit a sharp transition, instead a gradual change throughout the DNA backbone is observed. Comparing the individual base pairs of 1PUE and 7K1T [Figure 4.5] revealed the base pairs of the core consensus sequence overlap closely, while the flanking nucleotides are not superimposable. A closer look at the crystal packing interface reveals that the DNA in 7K1T form three non-Watson and Crick (WC) base pairs [Figure 4.6]. Along with the differences in sequence, as well as the formation of the non-WC base pairs, this evidence serves as an explanation for the minor differences in DNA B-factors as well as structure.

To investigate whether the changes in crystal packing and DNA sequence impacted the protein structure of 7K1T, we aligned 7K1T and 1PUE and found the RMSD to be 0.260 Å. The different crystal packing environment along with different DNA sequences, indicates that the DNA binding domain of PU1 is structurally agnostic to both the DNA flanking sequences and protein environment. This would suggest that small changes in side chains would be mechanistically responsible for selectivity and changes in flanking sequences.

Seeing as the difference in sequence and crystal packing did not give rise to large changes in the structure of the DNA or protein, we decided to create a Ramachandran plot to compare the two DNA-bound structures (7K1T and 1PUE) with the unbound NMR structure [Figure 4.2E]. Comparing the three structures it was noticed that most of the residues overlapped, indicating a stable structural core indifferent to experimental conditions, giving further evidence that small conformational changes of the amino acid side chains are responsible for DNA binding and selectivity. Most of the unmatched residues are all located in loops: D184 and M185 are present in the loop between α helix 1 (H1) and β-sheet 1, G218 and R220 are present in the loop between α helix 2 (H2) and α helix 3 (H3). The remaining residues (G254, E255, V256, G258 and R259) are located at the C-terminus of the protein in α-helix 4 (H4). A mapping of the residues [inset Figure 4.2E] visually indicated that the clustering displayed certain patterns. Furthermore, comparing the ψ and ϕ angles of 7K1T and 1PUE [Figure 4.2F] all residues with the exception to those previously mentioned (D184, M185, G254, E255, V256) fall on the correlation line indicating that the backbone dihedrals angles of both structures are similar. Seeing how at the macroscopic level, the structural similarity in the backbone dihedrals, not only lends credence to what can visually be observed, but also provides concrete proof the DNA selectivity is born from the orientation of the amino acid side chains.

In order to investigate the differences in the sidechains, we examined the sidechain dihedral angles between 7K1T and 1PUE [Figure 4.2G], and plotted the difference of the χ angles (7K1T-1PUE). As a representative, we have shown the residues that are different in at least three out of four χ angles [Figure 4.2H]. Residue K196 is the only residue that is different in both proteins with a difference in χ1-4 angles. Residue K222 and R212 are different in three χ angles. Furthermore, it was noticed during this analysis that R212 and Q226 holds two distinct conformations in 7K1T,

giving evidence that these residues may be important for driving the selectivity of DNA binding. With the overall structure remaining unchanged across multiple experimental conditions, the differences in the amino acid conformations must play a role in DNA recognition and selectivity.

4.2.2 Persistent waters aid DNA selectivity

The biggest difference between the two crystallographic complexes is the mutation of residue Q226 to E226 in 1PUE while there are no such mutations in 7K1T. Interestingly, this Q226 in 7K1T has two separate conformations (up 42% and down 58%). When aligned with 1PUE and the average structure of the 5W3G ensemble, we noticed that the "down" position of Q226 in 7K1T aligns with E226 in 1PUE and Q226 of 5W3G. [Figure 4.7]. However, considering that 5W3G is a NMR solution structure it was of interest to investigate the individual conformation that Q226 occupies in solution [Figure 4.8]. The two conformations of Q226 in 7K1T occupy the conformations shown in the NMR solution structure. Furthermore, in the crystallographic "up" position of 7K1T, Q226 hydrogen bonds to DG6 and DG8 on the 5' portion of the DNA. This secondary position, could act as a potential scanning mechanism to recognize sequences upstream of the core consensus sequence. Additionally, residue Q226 is only conserved in Class III ETS proteins, with E/D being prevalent in the other classes suggesting that it is important in selectively facilitating DNA binding in a sequence specific manner.

The high resolution of 7K1T allows us to place more crystallographic waters and is a good opportunity to identify water bridging contacts of PU.1 to DNA [Figure 4.9A]. The crystallographic waters also allow us to compare with 1PUE to identify structurally conserved waters that are important for DNA binding. Utilizing the clustering algorithm DBSCAN, we compared waters with different cutoffs from 1.0 Å to 0.3 Å [Figure 4.9B]. As predicted, as the cutoff for clusters decreased, the number of clusters decreased from 15 to 3 clusters. The lowest

cutoff allowed us to easily identify waters that were positioned in the same spot of the crystal structure, while the cutoff of 0.7 Å was chosen so as to start identifying water networks. Interestingly, at the strictest cutoff of 0.3 Å, one of the three water clusters interacts with Q226 [Figure 4.9C] in its "up" position when it binds to DNA. This indicates that the position of Q226 that binds to DNA and the water mediated contact is highly important to the stability of the protein for DNA binding. This water also interacts with R233 a universally conserved residue [Figure 4.9D] that is known to bind to the second guanosine of the consensus GGAA (7K1T and 1PUE). This structural evidence suggests that water is important for maintaining Q226 and R233 interaction while they respectively bind to DNA.

4.3 Discussion

PU.1 recognizes DNA through its recognition helix 3 (H3) where R230 and R233 directly interact with the cognate DNA sequence GGAA. PU.1 has a preference of DNA in the 5' flanking sequence (197) for high affinity binding. Its inclination to bind purines 5'AXC/A (where X is G or A) has been demonstrated by DNA binding experiments (197). However, identifying a specific residue and binding mechanism has proven to be elusive. We set to determine the selectivity of PU.1 by co-crystalizing the ETS domain (human residues 169-259) and high affinity 16 bp DNA 5'AGC. The crystal structure PDB code 7K1T demonstrated that the residue Q226 occupies two distinct conformations to which one of them makes direct contact with DNA at the Guanine base (DG6). While the 42% occupancy of Q226 interacts with DNA the other 58% occupancy interacts with a dynamic water in the downward position. Q226 is far from the crystal contact which means that the dynamics of the conformation observed are not a crystal packing artifact. This was also demonstrated by the comparison of the NMR structure 5W3G where the motion of Q226 in solution shifts between the two crystallographic conformations.

The high resolution of the new structure 7K1T gives insight into the hydration patterns missed previously by 1PUE. Waters in PU.1 serve as water-bridges between protein and DNA (251) as part of the high affinity DNA binding. It is important to distinguish dynamic waters vs persistent waters in crystallographic experiments as the later one is crucial for DNA binding. Comparing the two structures (7K1T and 1PUE) we determined at 0.3 Å cutoff the persistent waters that serve as links between residues and base pairs. A persistent water was found to bridge Q226 and R233 together while they individually contact DNA. This connection leads us to propose a selectivity mechanism by Q226. On the downward position Q226 scans DNA through dynamic waters and once it recognizes the purines in the 5' flanking sequence it allows for R233 to bind to the core consensus first guanine GGAA. The water network in the DNA major groove between protein and DNA creates stability once PU.1 recognizes its DNA target. These persistent waters are also present in 1PUE however, there is no connection between the water, point mutation E226, and R233. PU.1 and its subfamily members of Class III are the only ones with a glutamine residue in that position while the rest of the family (90% of family) has a glutamic acid or aspartate. This distinction makes it easier to understand why PU.1 prefers a dynamic glutamine to be able to scan DNA for a purine in the 5'flanking sequence while interacting with water that bridges a universally conserved residue such as R233.

4.4 Methods and materials

4.4.1 Nucleic acids

Deoxynucleotides for co-crystallization were prepared by solid state synthesis and purified by reverse-phase HPLC by Integrated DNA Technologies (Coralville, IA). Lyophilized DNA was re-dissolved and annealed in Buffer H150 (10 mM HEPES and 0.15 M NaCl adjusted to pH 7.4 with NaOH). The molecular cloning of the wild type ETS domain of human PU.1 (residues 165 to

270, termed ΔN165) has been previously described (155). Point mutants of ΔN165 were generated following a PCR-based strategy and cloned into pET28b or pCDF-1b vectors (Novagen) as previously described (78). Constructs were verified by Sanger sequencing.

4.4.2 Protein purification

Heterologous over-expression in BL21(DE3)pLysS *Escherichia coli* and purification was performed as described previously (155). In brief, cultures in LB medium were induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside at an OD_{600} of 0.6 for 4 hours at 25°C. Harvested cells were re-suspended in Buffer P (10 mM NaH_2PO_4/Na_2HPO_4, pH 7.4, and 0.5 M NaCl) containing 1 mM PMSF and lysed by sonication. The lysate was cleared by centrifugation and loaded onto Sepharose SP (GE HiTrap) equilibrated with Buffer P. After washing, the protein was eluted along a linear NaCl gradient in Buffer P under the control of a Bio-Rad NGC instrument. Following characterization by SDS-PAGE and mass spectrometry, the purified protein was exhaustively dialyzed against Buffer H150. PU.1 concentration was determined by UV absorption at 280 nm based on an extinction coefficient of 22,460 M^{-1} cm^{-1}.

4.4.3 Crystallization

The PU.1ΔN165: DNA complex was formed by gentle equimolar mixing of PU1 and duplex DNA at 400 μM each. Crystals were grown by vapor diffusion at 298 K in a hanging drop comprised of a 1:1 mixture of protein: DNA complex with mother liquor containing 100 mM sodium acetate and 2.5% PEG 3350. Prior to freezing, 2 μL of cryoprotectant solution containing 100 mM sodium acetate, 2.5% PEG 3350, and 20% glycerol was laid on top of the hanging drop and the well closed for 1 hour of incubation (4 μL total volume, 10% glycerol concentration). After 1 hour, crystals were transferred to the above 20% glycerol solution prior to freezing.

4.4.4 Data collection

X-ray diffraction data sets were obtained using a Pilatus3 6M 25 Hz detector at the Berkeley Center for Structural Biology (BCSB) Advanced Light Source on the 5.0.1 beamline at a fixed wavelength of 0.977408 Å with an oscillation angle of 0.25°.

4.4.5 Data processing

The diffraction data was processed using the XDS package (252) and was scaled using Aimless in the CCP4 package (253). Molecular replacement was then performed using 1PUE (downloaded from the RCSB protein databank) as the search coordinates in the PHENIX suite (254) with maximum-likelihood procedures in PHASER. Twenty-five rounds of refinement was then carried out using phenix refine (254) and model building was done with Coot (255). A summary of crystallographic data is shown in Table 4. Structures of atomic coordinates were rendered using PyMOL (v2.4.0, Schrodinger).

4.4.6 Data analysis

4.4.6.1 Analysis of water clusters

The analysis of the water clusters was done using PDBs (1PUE and 7K1T). The protein chain of 1PUE was aligned to the single chain of 7K1T using the align function in PyMol v2.4.0 with careful consideration of the waters associated to each chain. Separate PDBs were made for each of the aligned chains which included all of the waters. The following analysis was done with code written in Python v. 3.7 using the libraries Biopandas and SciKitLearn (256,257). Biopandas allows for the conversion of a pdb file into an easily workable pandas dataframe, while SciKitLearn was used for DBSCAN the clustering algorithm. In order to identify water clusters only between different pdb files the DBSCAN algorithm was employed with different eps (distance cutoff) and

minimum number of samples. All of the code for the analysis can be found in the repository www.github.com/rgumpper/water_analysis

4.4.6.2 Analysis of Dihedral angles

The dihedral angles were calculated using the R programming language (v4.0.0) and the bio3d library (258). The torsion.pdb function was used to calculate the dihedral angles and were output into an excel file for further plotting and analysis.

4.4.6.3 Average NMR structure

The average structure of 5W3G was done by calculating the average position of each atom across each frame of the deposited pdb file. Biopython (259) was used to aid in reading the structure and calculating average positions. The code for this calculation can be found in the repository www.github.com/rgumpper/average_pdb

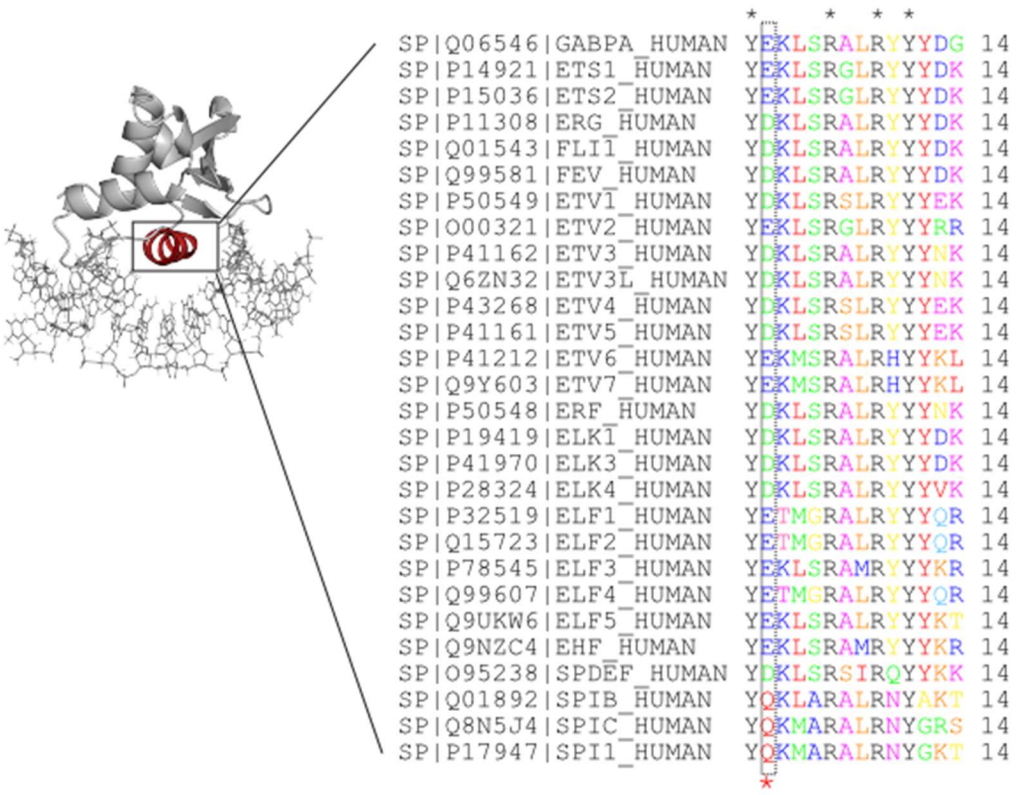


Figure 4-1 Recognition helix of ETS family alignment

All the ETS members are aligned for comparison of the recognition helix (H3). There are only 4 residues universally conserved (assigned black stars, *) throughout the helix. The residue Q226 (human numbers for PU.1) only present in Class III is shown with a red star, *

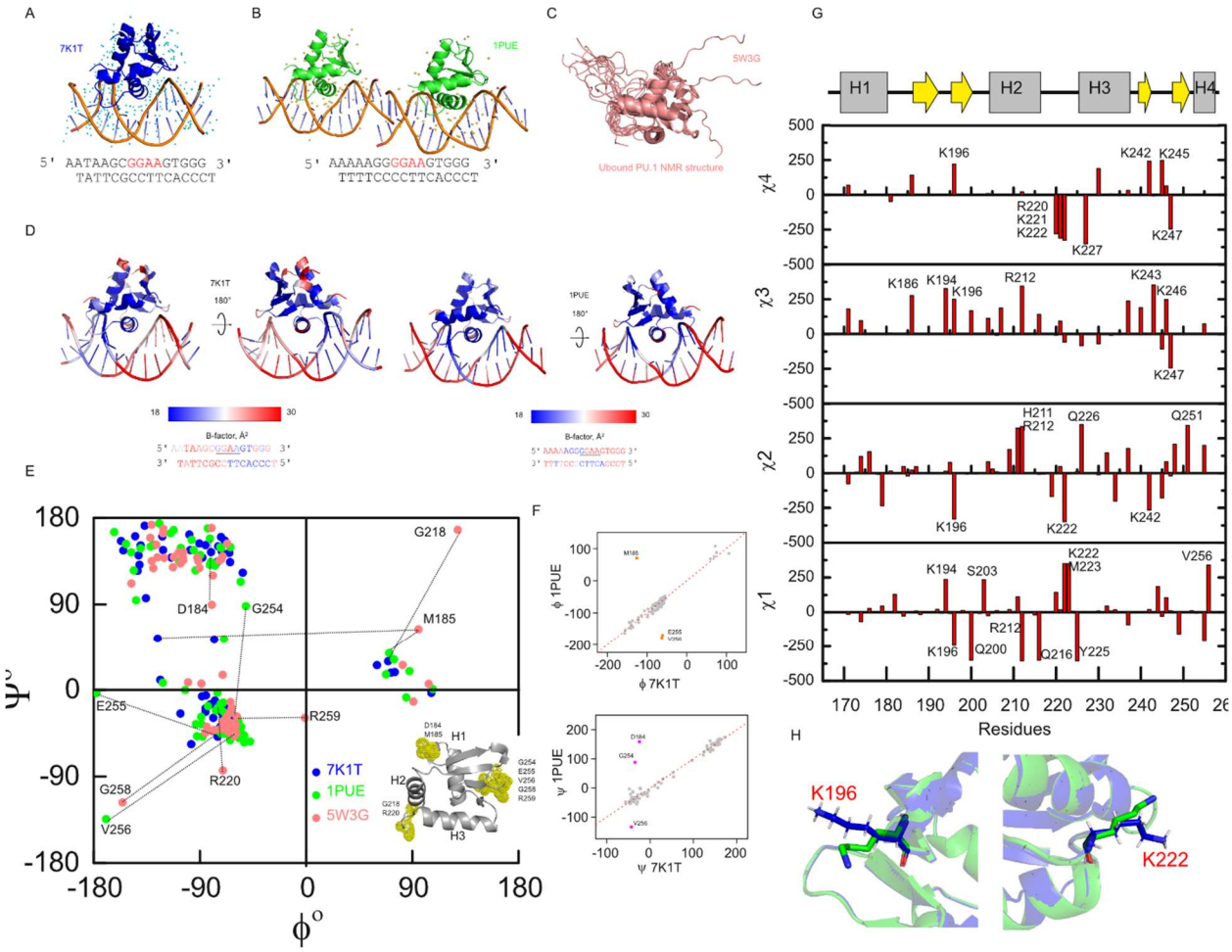


Figure 4-2 PU.1 crystal structure comparison

A, The new PU.1 crystal structure (PDB code 7K1T). B, previously solves structure of PU.1, (PDB code 1PUE) which has two monomeric asymmetric units. C, the solution NMR structure 5W3G. D, B-factor putty representation of 7K17 and 1PUE

E, The Ramachandran plot of backbone dihedrals of 7K1T, 1PUE and 5W3G. F, Dihedral angles ψ (7K1T vs. 1PUE) and ϕ (7K1T vs 1PUE) G, sidechain dihedral angles of all the residues 7K1T -1PUE plotted H, Residues K196, K222 as representative residues with the most χ angle differences between 7K1T and 1PUE

Table 4 Crystallographic table

	7K1T
Wavelength (Å)	0.977408
Resolution range (Å)	39.7 - 1.34 (1.388 - 1.34)
Space group	P 1 21 1
Unit cell (Å)	42.914 60.587 44.484
Unit cell (°)	90 116.818 90
Total reflections	85650 (5321)
Unique reflections	44340 (3591)
Multiplicity	1.9 (1.5)
Completeness (%)	96.97 (78.81)
Mean I/sigma(I)	37.50 (2.66)
Wilson B-factor (Å^2)	17.38
R-merge	0.009791 (0.3066)
R-meas	0.01385 (0.4336)
R-pim	0.009791 (0.3066)
CC1/2	1 (0.852)
CC*	1 (0.959)
Reflections used in refinement	44282 (3589)
Reflections used for R-free	1950 (166)
R-work	0.1298 (0.2195)
R-free	0.1607 (0.2573)
CC(work)	0.980 (0.942)
CC(free)	0.972 (0.906)
Number of non-hydrogen atoms	1688
• **macromolecules**	1399
• **ligands**	1
• **solvent**	288
Protein residues	91
RMS(bonds) (Å)	0.009
RMS(angles) (°)	1.13
Ramachandran favored (%)	97.75
Ramachandran allowed (%)	2.25
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashscore	3.53
Average B-factor (Å^2)	26.07
• **macromolecules**	22.9
• **ligands**	63.62
• **solvent**	41.33

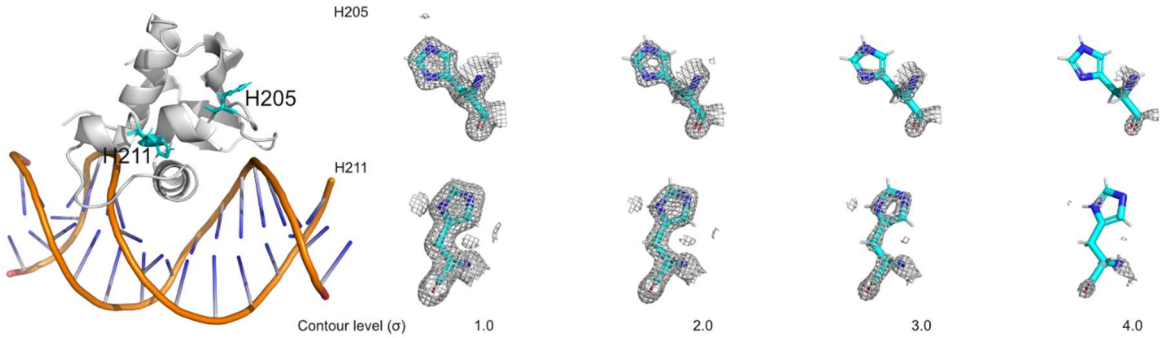


Figure 4-3 Refined hydrogens in Histidine

There are two Histidine residues in PU.1 ETS domain; H205 and H211. Due to the high resolution of the crystal structure 7K1T we added hydrogens in the refinement. By increasing the countering levels, we can see that both H205 and H211 are deprotonated.

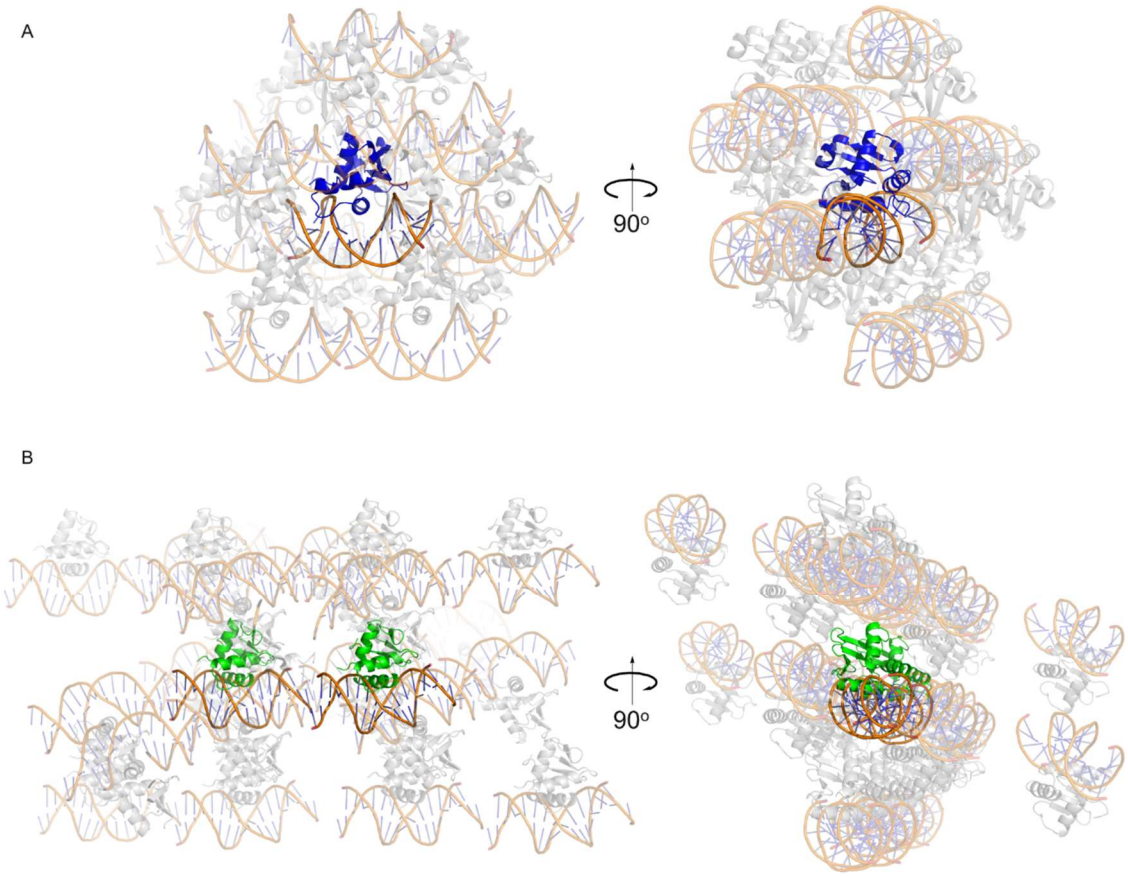


Figure 4-4 Crystal packing of 7K1T and 1PUE

A. Crystal packing of 7K1T shown with the single symmetric unit in color (blue) and the packing in transparent gray. The figure is rotated 90 degrees on the y-axis for a better side visual packing. B, previously solves structure of PU.1, (PDB code 1PUE) which has two monomeric asymmetric units. The figure is rotation of 90 degrees for a side view. The two asymmetric units are shown in green and the packing in transparent gray.

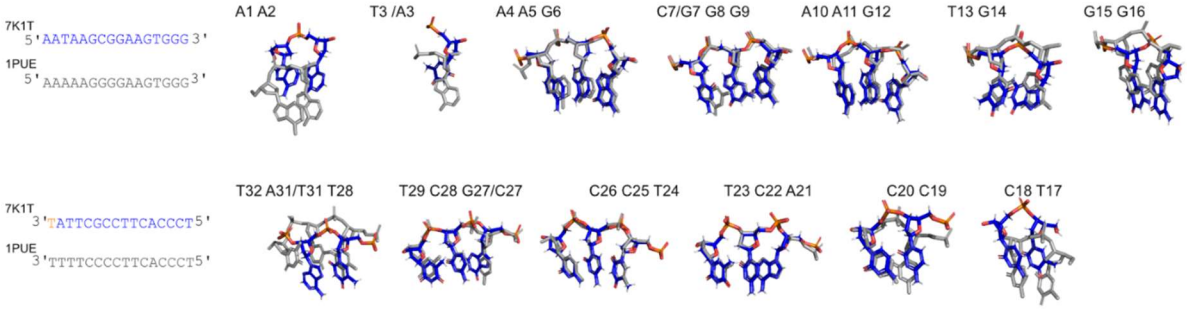


Figure 4-5 DNA comparison of 7K1T vs 1PUE

The backbone and bases of the two DNA (7K1T and 1PUE) compared indicate that the core consensus of GGAA overlap well with one another while the flanking region does not.

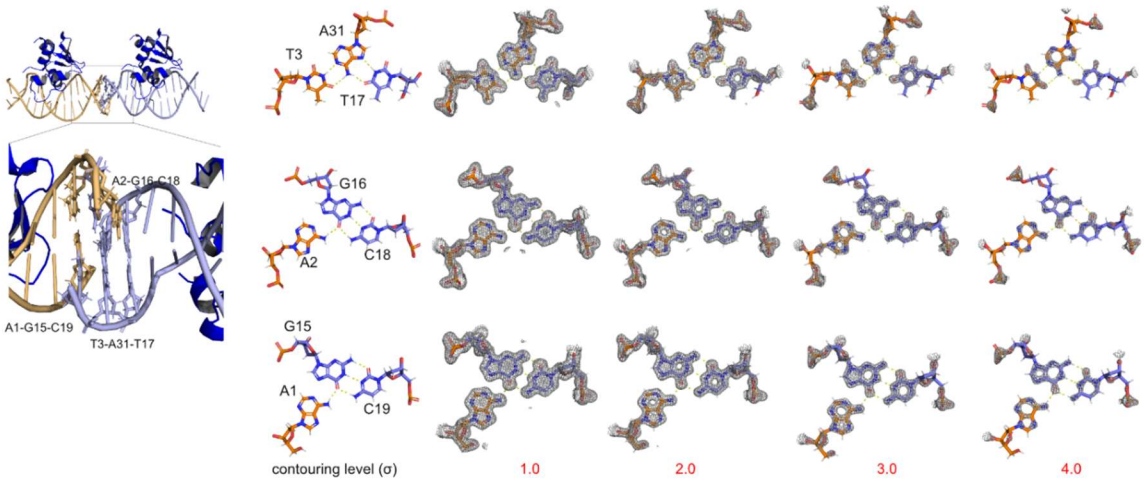


Figure 4-6 PU.1 7K1T shows structural differences due to the non-Watson and Crick base pairs: Two adjacent crystallographic subunits of 7K1T with DNA colored light blue and light orange. The 2Fo-Fc map is shown with increasing contouring level (σ) indicates that a charge cannot be unambiguously assigned.

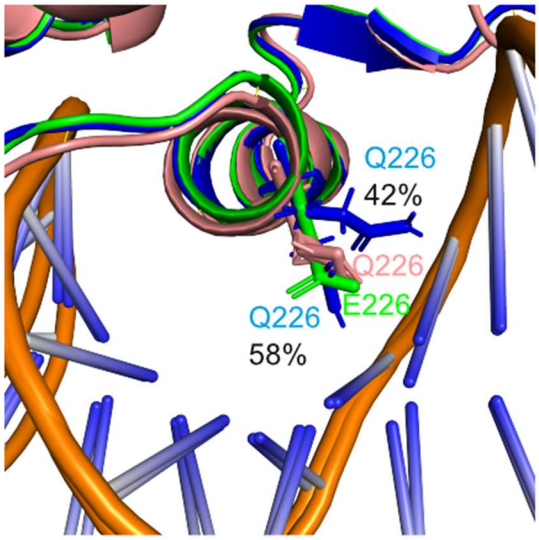


Figure 4-7 Q226 has two distinct conformations

All three; 5W3G in salmon color, 1PUE in green and 7K1T in blue were superimposed for comparison of residue Q226. In both WT structures 5W3G and 7K1T there is a Q226 while in the mutation of 1PUE there is an E226 7K1T shows two conformations of Q226, one at 42% in which the residue interacts with DNA at position DG6 and the other at 58% where the residue does not interact with DNA.

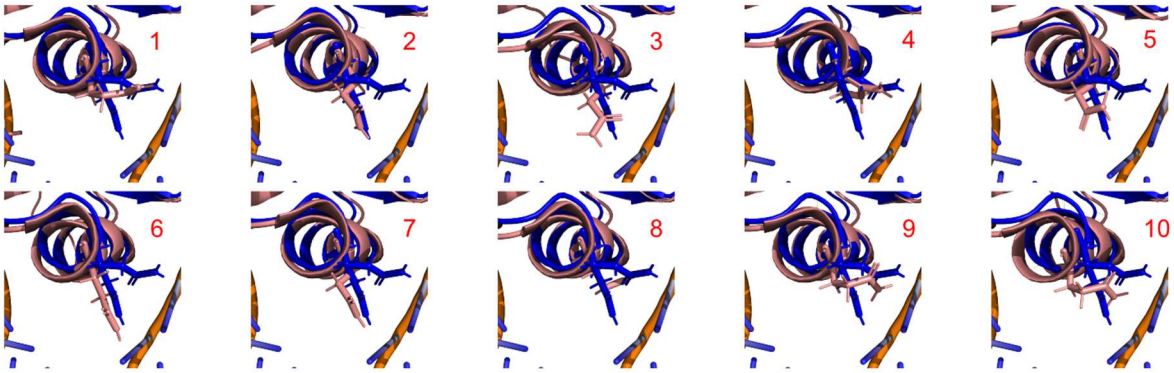


Figure 4-8 NMR ensemble aligned to 7K1T

NMR solution 5W3G has 10 ensembles (shown in salmon color) which are overlapped with 7K1T to monitor only Q226 that ranges between the two boundaries of the crystal structure.

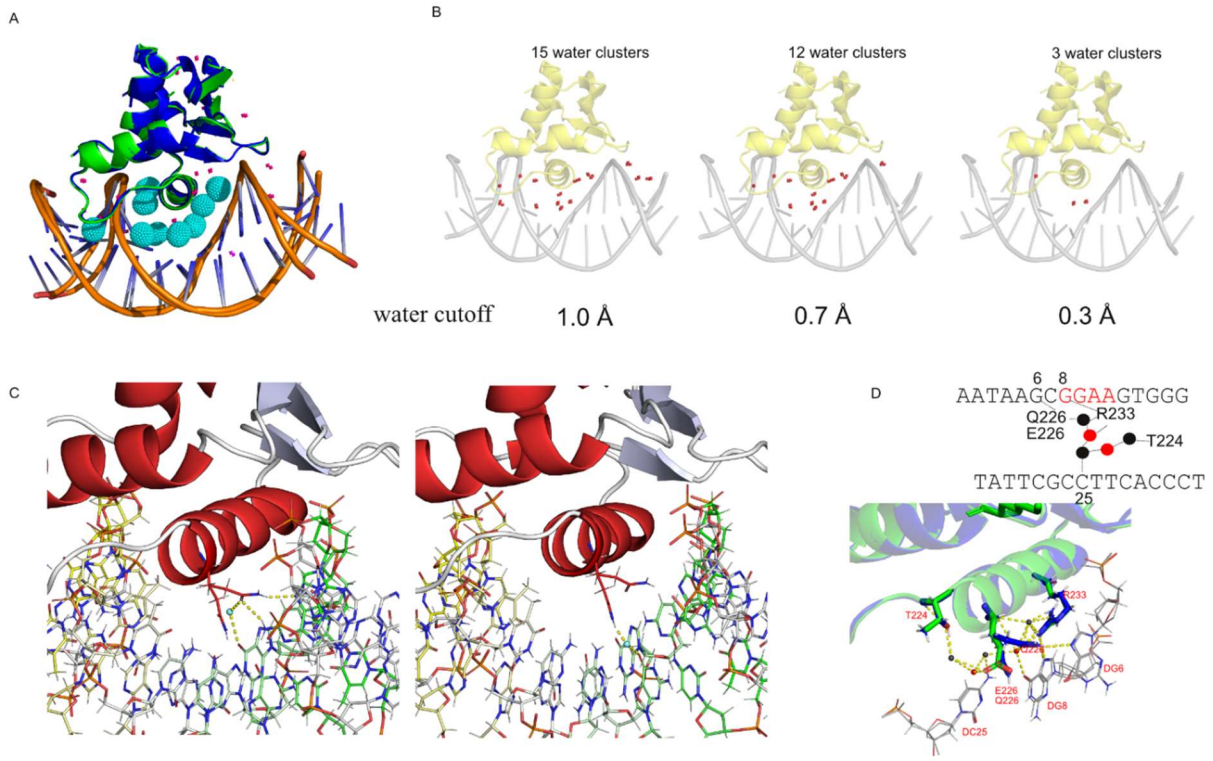


Figure 4-9 Water mediated contacts stabilize the 7K1T complex

A, All the waters bridging protein and DNA shown cyan balls, appear in the major groove around the H3. B, Comparing the waters between 7K1T and 1PUE at different cutoff (from 1.0 Å to 0.3 Å) monitoring the persistent water cluster. C, the interaction of Q226 in the upward position contacting DNA and a water that is persistent at 0.3 Å. In the downward position Q226 does not interact with DNA but a dynamic water. D, a schematic of water network and PU.1 interaction with DNA, through Q226 sharing a water bridge with universally conserved R233.

5 SUMMARY

PU.1 is a vital transcription factor as it is responsible for the self-renewal of hematopoietic stem cells as well as the differentiation of both branches of lymphoid and myeloid pathways (213). Homozygous PU.1 is lethal to the fetus (260), while heterozygous PU.1 patients will survive with underlying conditions. Experiments in which the gene of PU.1 was "knocked-out" in mice showed an early onset of acute myeloid leukemia (AML) (216) a likely reflection on human AML. The de-regulated activity of PU.1 is also been shown in a handful of diseases such as lymphoma, myeloma, Alzheimer's disease to name a few (156-158,214-217). Therefore, the activity of PU.1 is extremely important for normal hematopoietic cells to sustain life. The activity of a transcription factor can be determined by two important factors: the expression of that factor and the real-time regulation upon binding to its target DNA. The expression of PU.1 has been exhaustively studied as I have demonstrated from the introduction of this dissertation as well as in the introduction of both chapters 2 and 3. However, the mechanisms for the real-time regulation of PU.1 has been lacking in the literature. We set out to determine what the regulatory mechanism of PU.1 is at its DNA/protein level and its molecular interactions.

In chapter 2, we focused on the hydration of PU.1 in comparison to paralog ETS-1, both proteins belonging to the same ETS family. In numerous papers comparing ETS family members, PU.1 and ETS-1 have always shared the spotlight as the two have the sharpest contrast. Their DBD is the most phylogenetically distant in the family (159) but while their tertiary structure looks identical when bound to DNA (179). As an experimental model, both PU.1 and ETS-1 have a highly conserved tyrosine (Y) in their DNA binding affinity motif, Y252, and Y412 respectively, which make a water-mediated contact with the phosphate backbone of DNA.

Therefore, disturbing that contact by mutating the tyrosine into phenylalanine (Y->F) and observing the effect of both proteins under osmotic stress will provide insight into how these mutants behave in osmotic stress. For PU.1 at normal osmotic pressure the mutation, the Y252F showed no difference in DNA binding between the wild-type and mutant, indicating that other forms of PU.1 may exist. While the ETS-1 Y412F mutation perturbs binding at normal osmotic pressure. Under high osmotic pressure, the DNA binding of the mutant is weaker than the wild type, indicating that PU.1Y252F is an osmotic sensor. In contrast, ETS-1 under high osmotic pressure is sparsely hydrated, and binding is maintained primarily by direct contacts, so the Y412F mutant induces dynamics to the protein at a distal site away from the DNA binding site.

The mutation of the conserved Tyrosine residues gave us an insight into the hydration patterns of both ETS proteins with lessons to remember; the ETS homologs, despite their similar DNA-bound structures, recognize target DNA sites by distinct mechanisms.

In chapter 3 we tackled the mammoth hypothesis that PU.1 forms a dimeric complex at a single specific DNA binding site in a sequential manner (163,164) as part of its regulatory mechanism in real-time. Like the majority of transcription factors, PU.1 possesses a single well-structured domain, the DBD (45). The rest of the protein is intrinsically disordered as it tethers the ETS domain. On the N-terminus of the ETS domain, there is the PEST region which is highly acidic while the C-terminus there are 12 residues that do not seem to undergo any conformational structure. The experiments carried out were various utilizing synthetic enhancer elements consisting only of tandem copies of the EBS, cell lines that natively express PU.1, and a diverse array of biophysical techniques to demonstrate that PU.1 forms an inactive complex when bound as a dimer to DNA. This inactive 2:1 complex is apparent through negative feedback and the active 1:1 complex is an intermediate that cannot be bypassed. Furthermore, we found that PU.1

also can form an unbound dimer that is electrostatically mediated. The unbound dimer is symmetrical and does not involve the PEST region as it remains disordered in the dimeric complex. Additionally, the unbound dimer is thermodynamically less stable than the counterpart monomer indicating that it serves as a reservoir form of PU.1 ready to dissociate when needed.

A balancing regulatory mechanism surfaces between the inactive 2:1 complex, acting as negative feedback, and the inactive unbound dimer as a reservoir of protein when concentrations are high. These methods of regulation are all mediated by the intrinsically disordered regions of PU.1. Environmental factors favor one side of the balance (DNA) or the other (unbound) such as salt, concentration, and crowding agents. When truncating the C-terminal 12 residues we see that the balance shifts towards the DNA favored dimer while when phosphomimetic mutations of select PEST residues favor the apo dimeric complex. Furthermore, we demonstrated that these two dimeric complexes (with or without DNA) are not structurally the same as the unbound is symmetrical and the bound asymmetrical. Interestingly enough the protein will not undergo a transition of unbound dimeric PU.1 straight to dimeric complex with DNA creating a typical thermodynamic closed system. Instead, PU.1 undergoes an unfolded state ensemble explained by the principle of microscopic reversibility as it clarifies why the dimer is destabilized compared to the unfolded concerning the monomer. This also explains how the unfolding of the ETS domain justifies the unfolding ΔN117 dimer by mitigating the electrostatic repulsion.

This insight into the real-time regulatory mechanism of PU.1 gives rise to many questions regarding Type I (45) proteins and how PU.1 may exemplify a general class of electrostatic interactions by tethered IDR/DBD a mechanism. These IDRs are a novel drug target as they shift the equilibrium between active and inactive complexes. Finally, on a much bigger scale, we can

apply our current knowledge of PU.1 dimer formation to target gene therapy in heterozygous patients removing the faulty allele from the protein pool.

5.1 Where to next?

As often happens in life the moment a question is answered, another one immediately arises. We proposed a model for self-regulation of PU.1 by its intrinsically disordered regions and clarified some key residues crucial for osmotic sensitivity. The next big task is to identify the residues in both the PEST region and ETS domain that interact with one another to form the dimeric complexes. The PEST region is highly acidic and the ETS domain is basic with the highest cluster of basic residues in helix 3 (H3), which makes direct contact with the major groove of DNA. We hypothesize that the PEST residues interact with H3 in the unbound form, protecting H3 from binding to any other DNA in the nucleus. This hypothesis could be proven with a swap of charges: making the PEST region basic and the H3 acidic to see the effect of both these properties by DNA binding experiments.

With our proposed model for PU.1 regulation and dimeric complex formation along with the new structure of PU.1 solved in chapter 4, 7K1T. We could utilize docking to analyze the interface of PU.1 bound dimer. 7K1T (deleting the DNA) would dock into the 7K1T complex using a docking tool such as RosettaCommons (Version 3.12) to yield the protein-protein interface. Additional non-computational experiments could be a cysteine mapping of the ETS domain. Strategic cysteine (similar to the DKCDK) placement would give insight into the region of the protein that is involved in the 2:1 complex. DNA binding experiments would demonstrate the importance of those regions as it will have a poor binding affinity sue to its symmetry. While the cysteine in the areas that PU.1 forms an unbound dimer (symmetric dimer) will have yield a unique CD spectra similar to the once we observed in Chapter 3.

Recently it was proposed that transcriptional control may be driven by liquid-liquid phase separation (261) Liquid-liquid phase separation (LLP) is a process when a homogenous mixture separates into two liquids apart (262) a concentrated macromolecule phase (coacervation) (263) and a dilute macromolecule phase. LLP have been associated with various biological roles such as elastin whose polymeric assembly depends on phase separation (264) RNA metabolism (265) and chromatin rearrangement (266). These LLPs formations are deemed to occur as weak interactions mediated by the protein's intrinsically disordered regions (IDRs) (267). Eukaryotic transcription factors, as described in the introduction section, are abundant in IDRs (45,268) with an abundance of both positive charges (clustered in the DNA binding domain) and negative charges dispersed throughout the IDRs (269,270). The architecture of these TF allows them to be highly dynamic adapting various conformations, form fuzzy complexes (155,271) and facilitate phase separation of condensates (272-275) Due to their weak interactions associations/dissociations serves as a regulatory mechanism in gene expression including chromatin structure organization (276).

In Chapter 3 we demonstrated that transcription factor PU.1 forms a weak dimeric interaction in the absence of DNA serving as a reservoir of active PU.1 (155). Additionally, PU.1 has pioneering properties as it engages with closed chromatin and consequently recruiting the transcription machinery for intuition (29,32,33). Considering the data and the essential role of PU.1 in biology it is an easy assumption to think it undergoes LLPS as a mechanism for self-preservation. Experiments to prove such a transition occurs would have to be carried out with both constructs ΔN117 and ΔN165 as well as a well-known protein already demonstrating LLPS such as lysozyme at high concentration. Lysozyme and ΔN165 would serve as negative controls as they don't not possess IDRs present in ΔN117 which facilitate LLPS formation. The experiments would

require a high concentration of protein to see the droplet formation both with and without DNA. Experiments to scan optimal conditions (buffer and protein) would give clear transitional state of LLPS formation. From Chapter 3 we also know that the ETS domain alone can form dimers under low salt conditions, which could be explored further by trying to form LLPS droplets at different salt ranges. The experiments would be pretty straight forward as the droplets are big enough to see in a bright-field microscopy.

REFERENCES

1. Berg, J. M., Tymoczko, J. L, Stryer, L. (2002) Eukaryotic Transcription and Translation Are Separated in Space and Time. in *Biochemistry. 5th edition*. pp
2. Gromak, N., West, S., and Proudfoot, N. J. (2006) Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol Cell Biol* **26**, 3986-3996
3. Weis, L., and Reinberg, D. (1992) Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *FASEB J* **6**, 3300-3309
4. Crick, F. (1970) Central dogma of molecular biology. *Nature* **227**, 561-563
5. Yao, N. Y., and O'Donnell, M. (2010) SnapShot: The replisome. *Cell* **141**, 1088, 1088 e1081
6. Benkovic, S. J., Valentine, A. M., and Salinas, F. (2001) Replisome-mediated DNA replication. *Annu Rev Biochem* **70**, 181-208
7. Venters, B. J., and Pugh, B. F. (2009) How eukaryotic genes are transcribed. *Crit Rev Biochem Mol Biol* **44**, 117-141
8. Imashimizu, M., Shimamoto, N., Oshima, T., and Kashlev, M. (2014) Transcription elongation. Heterogeneous tracking of RNA polymerase and its biological implications. *Transcription* **5**, e28285
9. Komissarova, N., Becker, J., Solter, S., Kireeva, M., and Kashlev, M. (2002) Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Mol Cell* **10**, 1151-1162
10. Cooper, G. M. (2000) Regulation of Transcription in Eukaryotes. *A Molecular Approach* **2nd edition**
11. Brivanlou, A. H., and Darnell, J. E., Jr. (2002) Signal transduction and the control of gene expression. *Science* **295**, 813-818
12. Kim, J. L., Nikolov, D. B., and Burley, S. K. (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520-527
13. Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-520
14. He, Y., Fang, J., Taatjes, D. J., and Nogales, E. (2013) Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481-486
15. Han, Y., and He, Y. (2016) Eukaryotic transcription initiation machinery visualized at molecular level. *Transcription* **7**, 203-208
16. Allen, B. L., and Taatjes, D. J. (2015) The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* **16**, 155-166
17. Krishnamurthy, S., and Hampsey, M. (2009) Eukaryotic transcription initiation. *Curr Biol* **19**, R153-156
18. Cooper, G. M. (2000) Translation of mRNA. *The Cell: A Molecular Approach* **2nd edition**
19. Hernandez, G., Proud, C. G., Preiss, T., and Parsyan, A. (2012) On the Diversification of the Translation Apparatus across Eukaryotes. *Comp Funct Genomics* **2012**, 256848
20. Kapp, L. D., and Lorsch, J. R. (2004) The molecular mechanics of eukaryotic translation. *Annu Rev Biochem* **73**, 657-704
21. Levine, M., and Tjian, R. (2003) Transcription regulation and animal diversity. *Nature* **424**, 147-151

22. Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., Hannett, N. M., Tagne, J. B., Reynolds, D. B., Yoo, J., Jennings, E. G., Zeitlinger, J., Pokholok, D. K., Kellis, M., Rolfe, P. A., Takusagawa, K. T., Lander, E. S., Gifford, D. K., Fraenkel, E., and Young, R. A. (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99-104
23. Reece-Hoyes, J. S., Deplancke, B., Shingles, J., Grove, C. A., Hope, I. A., and Walhout, A. J. (2005) A compendium of Caenorhabditis elegans regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biol* **6**, R110
24. Fulton, D. L., Sundararajan, S., Badis, G., Hughes, T. R., Wasserman, W. W., Roach, J. C., and Sladek, R. (2009) TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol* **10**, R29
25. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., and Luscombe, N. M. (2009) A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**, 252-263
26. Latchman, D. S. (1993) Transcription factors: an overview. *Int J Exp Pathol* **74**, 417-422
27. Lopez, R. G., Carron, C., Oury, C., Gardellin, P., Bernard, O., and Ghysdael, J. (1999) TEL is a sequence-specific transcriptional repressor. *J Biol Chem* **274**, 30132-30138
28. Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., and Young, R. A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319
29. Pham, T. H., Minderjahn, J., Schmidl, C., Hoffmeister, H., Schmidhofer, S., Chen, W., Langst, G., Benner, C., and Rehli, M. (2013) Mechanisms of in vivo binding site selection of the hematopoietic master transcription factor PU.1. *Nucleic Acids Res* **41**, 6391-6402
30. Wilson, N. K., Foster, S. D., Wang, X., Knezevic, K., Schutte, J., Kaimakis, P., Chilarska, P. M., Kinston, S., Ouwehand, W. H., Dzierzak, E., Pimanda, J. E., de Bruijn, M. F., and Gottgens, B. (2010) Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* **7**, 532-544
31. Zhang, J. A., Mortazavi, A., Williams, B. A., Wold, B. J., and Rothenberg, E. V. (2012) Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* **149**, 467-482
32. Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C. L., Ragoussis, J., and Natoli, G. (2010) Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* **32**, 317-328
33. Pham, T. H., Benner, C., Lichtinger, M., Schwarzfischer, L., Hu, Y., Andreesen, R., Chen, W., and Rehli, M. (2012) Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states. *Blood* **119**, e161-171
34. Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., and Weirauch, M. T. (2018) The Human Transcription Factors. *Cell* **172**, 650-665
35. Manassero, N. G., Viola, I. L., Welchen, E., and Gonzalez, D. H. (2013) TCP transcription factors: architectures of plant form. *Biomol Concepts* **4**, 111-127

36. Shimoni, L., and Glusker, J. P. (1995) Hydrogen bonding motifs of protein side chains: descriptions of binding of arginine and amide groups. *Protein Sci* **4**, 65-74
37. Suzuki, M., Brenner, S. E., Gerstein, M., and Yagi, N. (1995) DNA recognition code of transcription factors. *Protein Eng* **8**, 319-328
38. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* **29**, 2860-2874
39. Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., and Mann, R. S. (2010) Origins of specificity in protein-DNA recognition. *Annu Rev Biochem* **79**, 233-269
40. Pan, Y., Tsai, C. J., Ma, B., and Nussinov, R. (2010) Mechanisms of transcription factor selectivity. *Trends Genet* **26**, 75-83
41. van der Vliet, P. C., and Verrijzer, C. P. (1993) Bending of DNA by transcription factors. *Bioessays* **15**, 25-32
42. Warnmark, A., Treuter, E., Wright, A. P., and Gustafsson, J. A. (2003) Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation. *Mol Endocrinol* **17**, 1901-1909
43. Ravarani, C. N., Erkina, T. Y., De Baets, G., Dudman, D. C., Erkine, A. M., and Babu, M. M. (2018) High-throughput discovery of functional disordered regions: investigation of transactivation domains. *Mol Syst Biol* **14**, e8190
44. Reverte, C. G., Ahearn, M. D., and Hake, L. E. (2001) CPEB degradation during Xenopus oocyte maturation requires a PEST domain and the 26S proteasome. *Dev Biol* **231**, 447-458
45. Minezaki, Y., Homma, K., Kinjo, A. R., and Nishikawa, K. (2006) Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J Mol Biol* **359**, 1137-1149
46. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C., and Brown, C. J. (2000) Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* **11**, 161-171
47. Na, J. H., Lee, W. K., and Yu, Y. G. (2018) How Do We Study the Dynamic Structure of Unstructured Proteins: A Case Study on Nopp140 as an Example of a Large, Intrinsically Disordered Protein. *Int J Mol Sci* **19**
48. Dosztanyi, Z., Meszaros, B., and Simon, I. (2010) Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. *Brief Bioinform* **11**, 225-243
49. Oldfield, C. J., and Dunker, A. K. (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **83**, 553-584
50. Das, D., and Mukhopadhyay, S. (2018) Studying backbone torsional dynamics of intrinsically disordered proteins using fluorescence depolarization kinetics. *J Biosci* **43**, 455-462
51. Ramachandran, G. N., and Sasisekharan, V. (1968) Conformation of polypeptides and proteins. *Adv Protein Chem* **23**, 283-438
52. Carugo, O., and Djinovic-Carugo, K. (2013) Half a century of Ramachandran plots. *Acta Crystallogr D Biol Crystallogr* **69**, 1333-1341
53. Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., and Lansbury, P. T., Jr. (1996) NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* **35**, 13709-13715

54. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197-208
55. Johnson, W. C., Jr. (1990) Protein secondary structure and circular dichroism: a practical guide. *Proteins* **7**, 205-214
56. Arai, M., Sugase, K., Dyson, H. J., and Wright, P. E. (2015) Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc Natl Acad Sci U S A* **112**, 9614-9619
57. Maestro, B., Galan, B., Alfonso, C., Rivas, G., Prieto, M. A., and Sanz, J. M. (2013) A new family of intrinsically disordered proteins: structural characterization of the major phasin PhaF from Pseudomonas putida KT2440. *PLoS One* **8**, e56904
58. Hamdi, K., Salladini, E., O'Brien, D. P., Brier, S., Chenal, A., Yacoubi, I., and Longhi, S. (2017) Structural disorder and induced folding within two cereal, ABA stress and ripening (ASR) proteins. *Sci Rep* **7**, 15544
59. Mittag, T., and Forman-Kay, J. D. (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* **17**, 3-14
60. Konrat, R. (2014) NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J Magn Reson* **241**, 74-85
61. Metcalfe, E. E., Zamoon, J., Thomas, D. D., and Veglia, G. (2004) (1)H/(15)N heteronuclear NMR spectroscopy shows four dynamic domains for phospholamban reconstituted in dodecylphosphocholine micelles. *Biophys J* **87**, 1205-1214
62. Lee, T. I., and Young, R. A. (2013) Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237-1251
63. Perkel, J. M., and Atchison, M. L. (1998) A two-step mechanism for recruitment of Pip by PU.1. *J Immunol* **160**, 241-252
64. Gibney, E. R., and Nolan, C. M. (2010) Epigenetics and gene expression. *Heredity (Edinb)* **105**, 4-13
65. Cowley, D. O., and Graves, B. J. (2000) Phosphorylation represses Ets-1 DNA binding by reinforcing autoinhibition. *Genes Dev* **14**, 366-376
66. Pereira Silva, L., Alves de Castro, P., Dos Reis, T. F., Paziani, M. H., Von Zeska Kress, M. R., Riano-Pachon, D. M., Hagiwara, D., Ries, L. N., Brown, N. A., and Goldman, G. H. (2017) Genome-wide transcriptome analysis of Aspergillus fumigatus exposed to osmotic stress reveals regulators of osmotic and cell wall stresses that are SakA(HOG1) and MpkC dependent. *Cell Microbiol* **19**
67. Brown, M. P., Grillo, A. O., Boyer, M., and Royer, C. A. (1999) Probing the role of water in the tryptophan repressor-operator complex. *Protein Sci* **8**, 1276-1285
68. Chalikian, T. V. (2003) Volumetric properties of proteins. *Annu Rev Biophys Biomol Struct* **32**, 207-235
69. Royer, W. E., Jr., Pardanani, A., Gibson, Q. H., Peterson, E. S., and Friedman, J. M. (1996) Ordered water molecules as key allosteric mediators in a cooperative dimeric hemoglobin. *Proc Natl Acad Sci U S A* **93**, 14526-14531
70. Joachimiak, A., Haran, T. E., and Sigler, P. B. (1994) Mutagenesis supports water mediated recognition in the trp repressor-operator system. *EMBO J* **13**, 367-372
71. Quiocho, F. A., Wilson, D. K., and Vyas, N. K. (1989) Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature* **340**, 404-407

72. Singer, P. T., Smalas, A., Carty, R. P., Mangel, W. F., and Sweet, R. M. (1993) The hydrolytic water molecule in trypsin, revealed by time-resolved Laue crystallography. *Science* **259**, 669-673
73. Jordan, S. R., and Pabo, C. O. (1988) Structure of the lambda complex at 2.5 A resolution: details of the repressor-operator interactions. *Science* **242**, 893-899
74. Brennan, R. G., Roderick, S. L., Takeda, Y., and Matthews, B. W. (1990) Protein-DNA conformational changes in the crystal structure of a lambda Cro-operator complex. *Proc Natl Acad Sci U S A* **87**, 8165-8169
75. Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991) Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science* **253**, 1001-1007
76. Spyrakis, F., Cozzini, P., Bertoli, C., Marabotti, A., Kellogg, G. E., and Mozzarelli, A. (2007) Energetics of the protein-DNA-water interaction. *BMC Struct Biol* **7**, 4
77. Janin, J. (1999) Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. *Structure* **7**, R277-279
78. Albrecht, A. V., Kim, H. M., and Poon, G. M. K. (2018) Mapping interfacial hydration in ETS-family transcription factor complexes with DNA: a chimeric approach. *Nucleic Acids Res* **46**, 10577-10588
79. Sidorova, N. Y., and Rau, D. C. (2004) Differences between EcoRI nonspecific and "star" sequence complexes revealed by osmotic stress. *Biophys J* **87**, 2564-2576
80. Williams, H. E., and Searle, M. S. (1999) Structure, dynamics and hydration of the nogalamycin-d(ATGCAT)2Complex determined by NMR and molecular dynamics simulations in solution. *J Mol Biol* **290**, 699-716
81. Jayaram, B., and Jain, T. (2004) The role of water in protein-DNA recognition. *Annu Rev Biophys Biomol Struct* **33**, 343-361
82. Harding, S. E. (2001) The hydration problem in solution biophysics: an introduction. *Biophys Chem* **93**, 87-91
83. Frueh, D. P., Goodrich, A. C., Mishra, S. H., and Nichols, S. R. (2013) NMR methods for structural studies of large monomeric and multimeric proteins. *Curr Opin Struct Biol* **23**, 734-739
84. Cooper, A., Johnson, C. M., Lakey, J. H., and Nollmann, M. (2001) Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophys Chem* **93**, 215-230
85. Winzor, D. J., Carrington, L. E., and Harding, S. E. (2001) Analysis of thermodynamic non-ideality in terms of protein solvation. *Biophys Chem* **93**, 231-240
86. Winter, R. (2019) Interrogating the Structural Dynamics and Energetics of Biomolecular Systems with Pressure Modulation. *Annu Rev Biophys* **48**, 441-463
87. Akasaka, K. (2006) Probing conformational fluctuation of proteins by pressure perturbation. *Chem Rev* **106**, 1814-1835
88. Horikoshi, K. (1998) Barophiles: deep-sea microorganisms adapted to an extreme environment. *Curr Opin Microbiol* **1**, 291-295
89. Yayanos, A. A. (1995) Microbiology to 10,500 meters in the deep sea. *Annu Rev Microbiol* **49**, 777-805
90. Chen, C. R., and Makhatadze, G. I. (2017) Molecular determinant of the effects of hydrostatic pressure on protein folding stability. *Nat Commun* **8**, 14561

91. Mason, O. U., Nakagawa, T., Rosner, M., Van Nostrand, J. D., Zhou, J., Maruyama, A., Fisk, M. R., and Giovannoni, S. J. (2010) First investigation of the microbiology of the deepest layer of ocean crust. *PLoS One* **5**, e15399
92. Robinson, C. R., and Sligar, S. G. (1995) Hydrostatic and osmotic pressure as tools to study macromolecular recognition. *Methods Enzymol* **259**, 395-427
93. Parsegian, V. A., Rand, R. P., and Rau, D. C. (2000) Osmotic stress, crowding, preferential hydration, and binding: A comparison of perspectives. *Proc Natl Acad Sci U S A* **97**, 3987-3992
94. Garner, M. M., and Rau, D. C. (1995) Water release associated with specific binding of gal repressor. *EMBO J* **14**, 1257-1263
95. Sidorova, N. Y., and Rau, D. C. (2001) Linkage of EcoRI dissociation from its specific DNA recognition site to water activity, salt concentration, and pH: separating their roles in specific and non-specific binding. *J Mol Biol* **310**, 801-816
96. Rand, R. P., Fuller, N. L., Butko, P., Francis, G., and Nicholls, P. (1993) Measured change in protein solvation with substrate binding and turnover. *Biochemistry* **32**, 5925-5929
97. Highsmith, S., Duignan, K., Cooke, R., and Cohen, J. (1996) Osmotic pressure probe of actin-myosin hydration changes during ATP hydrolysis. *Biophys J* **70**, 2830-2837
98. Auton, M., Rosgen, J., Sinev, M., Holthauzen, L. M., and Bolen, D. W. (2011) Osmolyte effects on protein stability and solubility: a balancing act between backbone and side-chains. *Biophys Chem* **159**, 90-99
99. Capp, M. W., Pegram, L. M., Saecker, R. M., Kratz, M., Riccardi, D., Wendorff, T., Cannon, J. G., and Record, M. T., Jr. (2009) Interactions of the osmolyte glycine betaine with molecular surfaces in water: thermodynamics, structural interpretation, and prediction of m-values. *Biochemistry* **48**, 10372-10379
100. Parsegian, V. A., Rand, R. P., and Rau, D. C. (1995) Macromolecules and water: probing with osmotic stress. *Methods Enzymol* **259**, 43-94
101. Erlitzki, N., Huang, K., Xhani, S., Farahat, A. A., Kumar, A., Boykin, D. W., and Poon, G. M. K. (2017) Investigation of the electrostatic and hydration properties of DNA minor groove-binding by a heterocyclic diamidine by osmotic pressure. *Biophys Chem* **231**, 95-104
102. Amoutzias, G. D., Robertson, D. L., Van de Peer, Y., and Oliver, S. G. (2008) Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem Sci* **33**, 220-229
103. Darnell, J. E., Jr. (2002) Transcription factors as targets for cancer therapy. *Nat Rev Cancer* **2**, 740-749
104. Desvergne, B., Michalik, L., and Wahli, W. (2006) Transcriptional regulation of metabolism. *Physiol Rev* **86**, 465-514
105. Germain, P., Chambon, P., Eichele, G., Evans, R. M., Lazar, M. A., Leid, M., De Lera, A. R., Lotan, R., Mangelsdorf, D. J., and Gronemeyer, H. (2006) International Union of Pharmacology. LXIII. Retinoid X receptors. *Pharmacol Rev* **58**, 760-772
106. Smolen, P., Baxter, D. A., and Byrne, J. H. (2000) Mathematical modeling of gene networks. *Neuron* **26**, 567-580
107. Kohler, J. J., and Schepartz, A. (2001) Kinetic studies of Fos.Jun.DNA complex formation: DNA binding prior to dimerization. *Biochemistry* **40**, 130-142

108. Blair, D. G., and Athanasiou, M. (2000) Ets and retroviruses - transduction and activation of members of the Ets oncogene family in viral oncogenesis. *Oncogene* **19**, 6472-6481
109. 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
110. 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
111. Findlay, V. J., LaRue, A. C., Turner, D. P., Watson, P. M., and Watson, D. K. (2013) Understanding the role of ETS-mediated gene regulation in complex biological processes. *Adv Cancer Res* **119**, 1-61
112. Colucci, F., Samson, S. I., DeKoter, R. P., Lantz, O., Singh, H., and Di Santo, J. P. (2001) Differential requirement for the transcription factor PU.1 in the generation of natural killer cells versus B and T cells. *Blood* **97**, 2625-2632
113. Mouly, E., Chemin, K., Nguyen, H. V., Chopin, M., Mesnard, L., Leite-de-Moraes, M., Burlen-defranoux, O., Bandeira, A., and Bories, J. C. (2010) The Ets-1 transcription factor controls the development and function of natural regulatory T cells. *J Exp Med* **207**, 2113-2125
114. Lacorazza, H. D., Miyazaki, Y., Di Cristofano, A., Deblasio, A., Hedvat, C., Zhang, J., Cordon-Cardo, C., Mao, S., Pandolfi, P. P., and Nimer, S. D. (2002) The ETS protein MEF plays a critical role in perforin gene expression and the development of natural killer and NK-T cells. *Immunity* **17**, 437-449
115. Spyropoulos, D. D., Pharr, P. N., Lavenburg, K. R., Jackers, P., Papas, T. S., Ogawa, M., and Watson, D. K. (2000) Hemorrhage, impaired hematopoiesis, and lethality in mouse embryos carrying a targeted disruption of the Fli1 transcription factor. *Mol Cell Biol* **20**, 5643-5652
116. Wang, L. C., Swat, W., Fujiwara, Y., Davidson, L., Visvader, J., Kuo, F., Alt, F. W., Gilliland, D. G., Golub, T. R., and Orkin, S. H. (1998) The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev* **12**, 2392-2402
117. Ferdous, A., Caprioli, A., Iacovino, M., Martin, C. M., Morris, J., Richardson, J. A., Latif, S., Hammer, R. E., Harvey, R. P., Olson, E. N., Kyba, M., and Garry, D. J. (2009) Nkx2-5 transactivates the Ets-related protein 71 gene and specifies an endothelial/endocardial fate in the developing embryo. *Proc Natl Acad Sci U S A* **106**, 814-819
118. Chen, C., Ouyang, W., Grigura, V., Zhou, Q., Carnes, K., Lim, H., Zhao, G. Q., Arber, S., Kurpios, N., Murphy, T. L., Cheng, A. M., Hassell, J. A., Chandrashekar, V., Hofmann, M. C., Hess, R. A., and Murphy, K. M. (2005) ERM is required for transcriptional control of the spermatogonial stem cell niche. *Nature* **436**, 1030-1034
119. Seth, A., and Watson, D. K. (2005) ETS transcription factors and their emerging roles in human cancer. *Eur J Cancer* **41**, 2462-2478
120. Sementchenko, V. I., and Watson, D. K. (2000) Ets target genes: past, present and future. *Oncogene* **19**, 6533-6548
121. Tomlins, S. A., Laxman, B., Dhanasekaran, S. M., Helgeson, B. E., Cao, X., Morris, D. S., Menon, A., Jing, X., Cao, Q., Han, B., Yu, J., Wang, L., Montie, J. E., Rubin, M. A., Pienta, K. J., Roulston, D., Shah, R. B., Varambally, S., Mehra, R., and Chinnaiyan, A.

M. (2007) Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* **448**, 595-599
122. Findlay, V. J., Turner, D. P., Yordy, J. S., McCarragher, B., Shriver, M. R., Szalai, G., Watson, P. M., Larue, A. C., Moussa, O., and Watson, D. K. (2011) Prostate-Derived ETS Factor Regulates Epithelial-to-Mesenchymal Transition through Both SLUG-Dependent and Independent Mechanisms. *Genes Cancer* **2**, 120-129
123. Moussa, O., Turner, D. P., Feldman, R. J., Sementchenko, V. I., McCarragher, B. D., Desouki, M. M., Fraig, M., and Watson, D. K. (2009) PDEF is a negative regulator of colon cancer cell growth and migration. *J Cell Biochem* **108**, 1389-1398
124. Li, R., Pei, H., Watson, D. K., and Papas, T. S. (2000) EAP1/Daxx interacts with ETS1 and represses transcriptional activation of ETS1 target genes. *Oncogene* **19**, 745-753
125. Verger, A., and Duterque-Coquillaud, M. (2002) When Ets transcription factors meet their partners. *Bioessays* **24**, 362-370
126. Hollenhorst, P. C., Ferris, M. W., Hull, M. A., Chae, H., Kim, S., and Graves, B. J. (2011) Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes Dev* **25**, 2147-2157
127. Pang, S. H., Minnich, M., Gangatirkar, P., Zheng, Z., Ebert, A., Song, G., Dickins, R. A., Corcoran, L. M., Mullighan, C. G., Busslinger, M., Huntington, N. D., Nutt, S. L., and Carotta, S. (2016) PU.1 cooperates with IRF4 and IRF8 to suppress pre-B-cell leukemia. *Leukemia* **30**, 1375-1387
128. Fitzsimmons, D., Lutz, R., Wheat, W., Chamberlin, H. M., and Hagman, J. (2001) Highly conserved amino acids in Pax and Ets proteins are required for DNA binding and ternary complex assembly. *Nucleic Acids Res* **29**, 4154-4165
129. Szalai, G., LaRue, A. C., and Watson, D. K. (2006) Molecular mechanisms of megakaryopoiesis. *Cell Mol Life Sci* **63**, 2460-2476
130. Charlot, C., Dubois-Pot, H., Serchov, T., Tourrette, Y., and Wasylyk, B. (2010) A review of post-translational modifications and subcellular localization of Ets transcription factors: possible connection with cancer and involvement in the hypoxic response. *Methods Mol Biol* **647**, 3-30
131. Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C., and Maki, R. A. (1990) The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. *Cell* **61**, 113-124
132. 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
133. Czuwara-Ladykowska, J., Sementchenko, V. I., Watson, D. K., and Trojanowska, M. (2002) Ets1 is an effector of the transforming growth factor beta (TGF-beta) signaling pathway and an antagonist of the profibrotic effects of TGF-beta. *J Biol Chem* **277**, 20399-20408
134. Goel, A., and Janknecht, R. (2003) Acetylation-mediated transcriptional activation of the ETS protein ER81 by p300, P/CAF, and HER2/Neu. *Mol Cell Biol* **23**, 6243-6254
135. Zhang, J., Graham, T. G., Vivekanand, P., Cote, L., Cetera, M., and Rebay, I. (2010) Sterile alpha motif domain-mediated self-association plays an essential role in modulating the activity of the Drosophila ETS family transcriptional repressor Yan. *Mol Cell Biol* **30**, 1158-1170

136. Kim, C. A., Phillips, M. L., Kim, W., Gingery, M., Tran, H. H., Robinson, M. A., Faham, S., and Bowie, J. U. (2001) Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *EMBO J* **20**, 4173-4182
137. Tran, H. H., Kim, C. A., Faham, S., Siddall, M. C., and Bowie, J. U. (2002) Native interface of the SAM domain polymer of TEL. *BMC Struct Biol* **2**, 5
138. Qiao, F., Harada, B., Song, H., Whitelegge, J., Courey, A. J., and Bowie, J. U. (2006) Mae inhibits Pointed-P2 transcriptional activity by blocking its MAPK docking site. *EMBO J* **25**, 70-79
139. 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. *J Biol Chem* **277**, 29386-29398
140. Leprivier, G., Baillat, D., Begue, A., Hartmann, B., and Aumercier, M. (2009) Ets-1 p51 and p42 isoforms differentially modulate Stromelysin-1 promoter according to induced DNA bend orientation. *Nucleic Acids Res* **37**, 4341-4352
141. 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
142. Moreau-Gachelin, F., Tavitian, A., and Tambourin, P. (1988) Spi-1 is a putative oncogene in virally induced murine erythroleukaemias. *Nature* **331**, 277-280
143. 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
144. Back, J., Dierich, A., Bronn, C., Kastner, P., and Chan, S. (2004) PU.1 determines the self-renewal capacity of erythroid progenitor cells. *Blood* **103**, 3615-3623
145. Scott, E. W., Simon, M. C., Anastasi, J., and Singh, H. (1994) Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* **265**, 1573-1577
146. Carotta, S., Dakic, A., D'Amico, A., Pang, S. H., Greig, K. T., Nutt, S. L., and Wu, L. (2010) The transcription factor PU.1 controls dendritic cell development and Flt3 cytokine receptor expression in a dose-dependent manner. *Immunity* **32**, 628-641
147. 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
148. McKercher, S. R., Torbett, B. E., Anderson, K. L., Henkel, G. W., Vestal, D. J., Baribault, H., Klemsz, M., Feeney, A. J., Wu, G. E., Paige, C. J., and Maki, R. A. (1996) Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. *EMBO J* **15**, 5647-5658
149. Dakic, A., Metcalf, D., Di Rago, L., Mifsud, S., Wu, L., and Nutt, S. L. (2005) PU.1 regulates the commitment of adult hematopoietic progenitors and restricts granulopoiesis. *J Exp Med* **201**, 1487-1502
150. Nutt, S. L., Metcalf, D., D'Amico, A., Polli, M., and Wu, L. (2005) Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors. *J Exp Med* **201**, 221-231
151. Back, J., Allman, D., Chan, S., and Kastner, P. (2005) Visualizing PU.1 activity during hematopoiesis. *Exp Hematol* **33**, 395-402

152. Nelsen, B., Tian, G., Erman, B., Gregoire, J., Maki, R., Graves, B., and Sen, R. (1993) Regulation of lymphoid-specific immunoglobulin mu heavy chain gene enhancer by ETS-domain proteins. *Science* **261**, 82-86
153. Hromas, R., Orazi, A., Neiman, R. S., Maki, R., Van Beveran, C., Moore, J., and Klemsz, M. (1993) Hematopoietic lineage- and stage-restricted expression of the ETS oncogene family member PU.1. *Blood* **82**, 2998-3004
154. 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
155. 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., and Poon, G. M. K. (2020) Intrinsic disorder controls two functionally distinct dimers of the master transcription factor PU.1. *Sci Adv* **6**, eaay3178
156. Wohlfahrt, T., Rauber, S., Uebe, S., Luber, M., Soare, A., Ekici, A., Weber, S., Matei, A. E., Chen, C. W., Maier, C., Karouzakis, E., Kiener, H. P., Pachera, E., Dees, C., Beyer, C., Daniel, C., Gelse, K., Kremer, A. E., Naschberger, E., Sturzl, M., Butter, F., Sticherling, M., Finotto, S., Kreuter, A., Kaplan, M. H., Jungel, A., Gay, S., Nutt, S. L., Boykin, D. W., Poon, G. M. K., Distler, O., Schett, G., Distler, J. H. W., and Ramming, A. (2019) PU.1 controls fibroblast polarization and tissue fibrosis. *Nature* **566**, 344-349
157. Antony-Debre, I., Paul, A., Leite, J., Mitchell, K., Kim, H. M., Carvajal, L. A., Todorova, T. I., Huang, K., Kumar, A., Farahat, A. A., Bartholdy, B., Narayanagari, S. R., Chen, J., Ambesi-Impiombato, A., Ferrando, A. A., Mantzaris, I., Gavathiotis, E., Verma, A., Will, B., Boykin, D. W., Wilson, W. D., Poon, G. M., and Steidl, U. (2017) Pharmacological inhibition of the transcription factor PU.1 in leukemia. *J Clin Invest* **127**, 4297-4313
158. Durual, S., Rideau, A., Ruault-Jungblut, S., Cossali, D., Beris, P., Piguet, V., and Matthes, T. (2007) Lentiviral PU.1 overexpression restores differentiation in myeloid leukemic blasts. *Leukemia* **21**, 1050-1059
159. Poon, G. M. K., and Kim, H. M. (2017) Signatures of DNA target selectivity by ETS transcription factors. *Transcription* **8**, 193-203
160. Garvie, C. W., Hagman, J., and Wolberger, C. (2001) Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell* **8**, 1267-1276
161. Laudet, V., Niel, C., Duterque-Coquillaud, M., Leprince, D., and Stehelin, D. (1993) Evolution of the ets gene family. *Biochem Biophys Res Commun* **190**, 8-14
162. 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
163. 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
164. 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
165. Shi, Y. (2014) A glimpse of structural biology through X-ray crystallography. *Cell* **159**, 995-1014
166. Bragg, W. L. (1913) The diffraction of short electromagnetic waves by a crystal. *Cambridge: University Press* **17**, 43-57

167. Smyth, M. S., and Martin, J. H. (2000) x ray crystallography. *Mol Pathol* **53**, 8-14
168. Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2013) Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J* **280**, 5705-5736
169. Ke, H. (1997) Overview of isomorphous replacement phasing. *Methods in Enzymology* **276**, 448-461
170. Rossmann, M. G. (1990) The molecular replacement method. *Acta Crystallogr A* **46 (Pt 2)**, 73-82
171. DiMaio, F., Kondrashov, D. A., Bitto, E., Soni, A., Bingman, C. A., Phillips, G. N., Jr., and Shavlik, J. W. (2007) Creating protein models from electron-density maps using particle-filtering methods. *Bioinformatics* **23**, 2851-2858
172. Ota, M., Koike, R., Amemiya, T., Tenno, T., Romero, P. R., Hiroaki, H., Dunker, A. K., and Fukuchi, S. (2013) An assignment of intrinsically disordered regions of proteins based on NMR structures. *J Struct Biol* **181**, 29-36
173. Poss, Z. C., Ebmeier, C. C., and Taatjes, D. J. (2013) The Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol* **48**, 575-608
174. Berk, A. J. (2012) Yin and yang of mediator function revealed by human mutants. *Proc Natl Acad Sci U S A* **109**, 19519-19520
175. Toth-Petroczy, A., Oldfield, C. J., Simon, I., Takagi, Y., Dunker, A. K., Uversky, V. N., and Fuxreiter, M. (2008) Malleable machines in transcription regulation: the mediator complex. *PLoS Comput Biol* **4**, e1000243
176. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242
177. N, N., C, G. P., Chakraborty, C., V, K., D, T. K., V, B., R, S., Lu, A., Ge, Z., and Zhu, H. (2016) Mechanism of artemisinin resistance for malaria PfATP6 L263 mutations and discovering potential antimalarials: An integrated computational approach. *Sci Rep* **6**, 30106
178. 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. *J Biol Chem* **287**, 18297-18307
179. 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 Res* **43**, 4322-4331
180. Hollenhorst, P. C., Chandler, K. J., Poulsen, R. L., Johnson, W. E., Speck, N. A., and Graves, B. J. (2009) DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genet* **5**, e1000778
181. Hollenhorst, P. C., Shah, A. A., Hopkins, C., and Graves, B. J. (2007) Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. *Genes Dev* **21**, 1882-1894
182. Timasheff, S. N. (1998) In disperse solution, "osmotic stress" is a restricted case of preferential interactions. *Proc Natl Acad Sci U S A* **95**, 7363-7367
183. Timasheff, S. N. (2002) Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. *Proc Natl Acad Sci U S A* **99**, 9721-9726

184. 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
185. Munde, M., Poon, G. M., and Wilson, W. D. (2013) Probing the electrostatics and pharmacological modulation of sequence-specific binding by the DNA-binding domain of the ETS family transcription factor PU.1: a binding affinity and kinetics investigation. *J Mol Biol* **425**, 1655-1669
186. Myszka, D. G., Jonsen, M. D., and Graves, B. J. (1998) Equilibrium analysis of high affinity interactions using BIACORE. *Anal Biochem* **265**, 326-330
187. 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
188. 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
189. Werner, M. H., Clore, G. M., Fisher, C. L., Fisher, R. J., Trinh, L., Shiloach, J., and Gronenborn, A. M. (1997) Correction of the NMR structure of the ETS1/DNA complex. *J Biomol NMR* **10**, 317-328
190. De, S., Chan, A. C., Coyne, H. J., 3rd, Bhachech, N., Hermsdorf, U., Okon, M., Murphy, M. E., Graves, B. J., and McIntosh, L. P. (2014) Steric mechanism of auto-inhibitory regulation of specific and non-specific DNA binding by the ETS transcriptional repressor ETV6. *J Mol Biol* **426**, 1390-1406
191. Dahl, R., and Simon, M. C. (2003) The importance of PU.1 concentration in hematopoietic lineage commitment and maturation. *Blood Cells Mol Dis* **31**, 229-233
192. DeKoter, R. P., and Singh, H. (2000) Regulation of B lymphocyte and macrophage development by graded expression of PU.1. *Science* **288**, 1439-1441
193. 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
194. Go, W. Y., Liu, X., Roti, M. A., Liu, F., and Ho, S. N. (2004) NFAT5/TonEBP mutant mice define osmotic stress as a critical feature of the lymphoid microenvironment. *Proc Natl Acad Sci U S A* **101**, 10673-10678
195. Ho, S. N. (2006) Intracellular water homeostasis and the mammalian cellular osmotic stress response. *J Cell Physiol* **206**, 9-15
196. Reddy, S. Y., Obika, S., and Bruice, T. C. (2003) Conformations and dynamics of Ets-1 ETS domain-DNA complexes. *Proc Natl Acad Sci U S A* **100**, 15475-15480
197. 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
198. Nye, J. A., Petersen, J. M., Gunther, C. V., Jonsen, M. D., and Graves, B. J. (1992) Interaction of murine ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. *Genes Dev* **6**, 975-990
199. Tataurov, A. V., You, Y., and Owczarzy, R. (2008) Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids. *Biophys Chem* **133**, 66-70
200. Micsonai, A., Wien, F., Kernya, L., Lee, Y. H., Goto, Y., Refregiers, M., and Kardos, J. (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc Natl Acad Sci U S A* **112**, E3095-3103

201. Lu, X. J., and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* **31**, 5108-5121
202. Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859-1865
203. Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., and Dunker, A. K. (2006) Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873-6888
204. Singh, G. P., and Dash, D. (2007) Intrinsic disorder in yeast transcriptional regulatory network. *Proteins* **68**, 602-605
205. Vuzman, D., Azia, A., and Levy, Y. (2010) Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *J Mol Biol* **396**, 674-684
206. Lohrum, M. A., Woods, D. B., Ludwig, R. L., Balint, E., and Vousden, K. H. (2001) C-terminal ubiquitination of p53 contributes to nuclear export. *Mol Cell Biol* **21**, 8521-8532
207. Roberts, S. G., Ha, I., Maldonado, E., Reinberg, D., and Green, M. R. (1993) Interaction between an acidic activator and transcription factor TFIIB is required for transcriptional activation. *Nature* **363**, 741-744
208. Friedman, J. S., Khanna, H., Swain, P. K., Denicola, R., Cheng, H., Mitton, K. P., Weber, C. H., Hicks, D., and Swaroop, A. (2004) The minimal transactivation domain of the basic motif-leucine zipper transcription factor NRL interacts with TATA-binding protein. *J Biol Chem* **279**, 47233-47241
209. Rogers, S., Wells, R., and Rechsteiner, M. (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **234**, 364-368
210. Chevaillier, P. (1993) Pest sequences in nuclear proteins. *Int J Biochem* **25**, 479-482
211. Mittal, A., Holehouse, A. S., Cohan, M. C., and Pappu, R. V. (2018) Sequence-to-Conformation Relationships of Disordered Regions Tethered to Folded Domains of Proteins. *J Mol Biol* **430**, 2403-2421
212. Guo, X., Bulyk, M. L., and Hartemink, A. J. (2012) Intrinsic disorder within and flanking the DNA-binding domains of human transcription factors. *Pac Symp Biocomput*, 104-115
213. Staber, P. B., Zhang, P., Ye, M., Welner, R. S., Nombela-Arrieta, C., Bach, C., Kerenyi, M., Bartholdy, B. A., Zhang, H., Alberich-Jorda, M., Lee, S., Yang, H., Ng, F., Zhang, J., Leddin, M., Silberstein, L. E., Hoefler, G., Orkin, S. H., Gottgens, B., Rosenbauer, F., Huang, G., and Tenen, D. G. (2013) Sustained PU.1 levels balance cell-cycle regulators to prevent exhaustion of adult hematopoietic stem cells. *Mol Cell* **49**, 934-946
214. Yuki, H., Ueno, S., Tatetsu, H., Niiro, H., Iino, T., Endo, S., Kawano, Y., Komohara, Y., Takeya, M., Hata, H., Okada, S., Watanabe, T., Akashi, K., Mitsuya, H., and Okuno, Y. (2013) PU.1 is a potent tumor suppressor in classical Hodgkin lymphoma cells. *Blood* **121**, 962-970
215. 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
216. 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
217. 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

218. 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
219. Chen, H., Ray-Gallet, D., Zhang, P., Hetherington, C. J., Gonzalez, D. A., Zhang, D. E., Moreau-Gachelin, F., and Tenen, D. G. (1995) PU.1 (Spi-1) autoregulates its expression in myeloid cells. *Oncogene* **11**, 1549-1560
220. Solomon, L. A., Podder, S., He, J., Jackson-Chornenki, N. L., Gibson, K., Ziliotto, R. G., Rhee, J., and DeKoter, R. P. (2017) Coordination of Myeloid Differentiation with Reduced Cell Cycle Progression by PU.1 Induction of MicroRNAs Targeting Cell Cycle Regulators and Lipid Anabolism. *Mol Cell Biol* **37**
221. 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
222. Poon, G. M. (2007) Enhancement of oligomeric stability by covalent linkage and its application to the human p53tet domain: thermodynamics and biological implications. *Biochem Soc Trans* **35**, 1574-1578
223. Ferrige, A. G., Seddon, M. J, Green, B. N, Jarvis, S. A, Skilling, J, Staunto, J. (1992) Disentangling electrospray spectra with maximum-entropy. *Rapid Commun. Mass Spectrom* **6**, 707–711
224. Berjanskii, M. V., and Wishart, D. S. (2008) Application of the random coil index to studying protein flexibility. *J Biomol NMR* **40**, 31-48
225. 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
226. Huang, W., Horvath, E., and Eklund, E. A. (2007) PU.1, interferon regulatory factor (IRF) 2, and the interferon consensus sequence-binding protein (ICSBP/IRF8) cooperate to activate NF1 transcription in differentiating myeloid cells. *J Biol Chem* **282**, 6629-6643
227. 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
228. 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
229. Furfine, C. S., and Velick, S. F. (1965) The Acyl-Enzyme Intermediate and the Kinetic Mechanism of the Glyceraldehyde 3-Phosphate Dehydrogenase Reaction. *J Biol Chem* **240**, 844-855
230. Mueller, B. U., Pabst, T., Fos, J., Petkovic, V., Fey, M. F., Asou, N., Buergi, U., and Tenen, D. G. (2006) ATRA resolves the differentiation block in t(15;17) acute myeloid leukemia by restoring PU.1 expression. *Blood* **107**, 3330-3338
231. Pedchenko, T. V., Park, G. Y., Joo, M., Blackwell, T. S., and Christman, J. W. (2005) Inducible binding of PU.1 and interacting proteins to the Toll-like receptor 4 promoter during endotoxemia. *Am J Physiol Lung Cell Mol Physiol* **289**, L429-437
232. 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
233. 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
234. 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
235. Seshire, A., Rossiger, T., Frech, M., Beez, S., Hagemeyer, H., and Puccetti, E. (2012) Direct interaction of PU.1 with oncogenic transcription factors reduces its serine phosphorylation and promoter binding. *Leukemia* **26**, 1338-1347
236. 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
237. Liew, C. W., Rand, K. D., Simpson, R. J., 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. *J Biol Chem* **281**, 28296-28306
238. Fuxreiter, M. (2018) Fuzziness in Protein Interactions-A Historical Perspective. *J Mol Biol* **430**, 2278-2287
239. Shoemaker, B. A., Portman, J. J., and Wolynes, P. G. (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A* **97**, 8868-8873
240. 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
241. Currie, S. L., Lau, D. K. W., Doane, J. J., Whitby, F. G., Okon, M., McIntosh, L. P., and Graves, B. J. (2017) Structured and disordered regions cooperatively mediate DNA-binding autoinhibition of ETS factors ETV1, ETV4 and ETV5. *Nucleic Acids Res* **45**, 2223-2241
242. 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
243. King, D. A., Zhang, L., Guarente, L., and Marmorstein, R. (1999) Structure of a HAP1-DNA complex reveals dramatically asymmetric DNA binding by a homodimeric protein. *Nat Struct Biol* **6**, 64-71
244. Kwok, J. C., Perdomo, J., and Chong, B. H. (2007) Identification of a monopartite sequence in PU.1 essential for nuclear import, DNA-binding and transcription of myeloid-specific genes. *J Cell Biochem* **101**, 1456-1474
245. Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-1502
246. 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
247. Graves, B. J., and Petersen, J. M. (1998) Specificity within the ets family of transcription factors. *Adv Cancer Res* **75**, 1-55
248. Hollenhorst, P. C., Jones, D. A., and Graves, B. J. (2004) Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. *Nucleic Acids Res* **32**, 5693-5702
249. Szymczyna, B. R., Arrowsmith, C. H. (2000) DNA binding specificity studies of four ETS proteins support an indirect read-out mechanism of protein-DNA recognition. *The Journal of biological chemistry* **275**, 28363-28370
250. Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., and Honig, B. (2009) The role of DNA shape in protein-DNA recognition. *Nature* **461**, 1248-1253
251. 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
252. Kabsch, W. (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132
253. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242
254. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221
255. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501
256. S, R. (2017) Biopandas: Working with molecular structures in pandas dataframes. *The Journal of Open Source Software* **2**
257. Pedregosa. F VG, G. A., Michel. V, Thirion. B, Grisel. O, Blondel. M, Prettenhofer. P, Weiss. R, Dubourg. V, Vanderplas. J, Passos. A, Cournapeau. D, Brucher. M, Perrot. M, Duchesnay. E. (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825-2830
258. Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A., and Caves, L. S. (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695-2696
259. Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M. J. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423
260. DeKoter, R. P., Kamath, M. B., and Houston, I. B. (2007) Analysis of concentration-dependent functions of PU.1 in hematopoiesis using mouse models. *Blood Cells Mol Dis* **39**, 316-320

261. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., and Sharp, P. A. (2017) A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13-23
262. Li, Q., Peng, X., Li, Y., Tang, W., Zhu, J., Huang, J., Qi, Y., and Zhang, Z. (2020) LLPSDB: a database of proteins undergoing liquid-liquid phase separation in vitro. *Nucleic Acids Res* **48**, D320-D327
263. BUNGENBERG DE JONG, H. G., KRUYT, H. R. (1929) Coacewation (Partial miscibility in colloid systems). *Proc. Acad. Sci. Amst.* **43**, 849-856
264. Muiznieks, L. D., Sharpe, S., Pomes, R., and Keeley, F. W. (2018) Role of Liquid-Liquid Phase Separation in Assembly of Elastin and Other Extracellular Matrix Proteins. *J Mol Biol* **430**, 4741-4753
265. Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A. S., Yu, T., Marie-Nelly, H., McSwiggen, D. T., Kokic, G., Dailey, G. M., Cramer, P., Darzacq, X., and Zweckstetter, M. (2018) RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol* **25**, 833-840
266. Tatavosian, R., Kent, S., Brown, K., Yao, T., Duc, H. N., Huynh, T. N., Zhen, C. Y., Ma, B., Wang, H., and Ren, X. (2019) Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. *J Biol Chem* **294**, 1451-1463
267. Zhang, G., Wang, Z., Du, Z., and Zhang, H. (2018) mTOR Regulates Phase Separation of PGL Granules to Modulate Their Autophagic Degradation. *Cell* **174**, 1492-1506 e1422
268. Peng, Z., Yan, J., Fan, X., Mizianty, M. J., Xue, B., Wang, K., Hu, G., Uversky, V. N., and Kurgan, L. (2015) Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci* **72**, 137-151
269. Campen, A., Williams, R. M., Brown, C. J., Meng, J., Uversky, V. N., and Dunker, A. K. (2008) TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept Lett* **15**, 956-963
270. Williams, R. M., Obradovi, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., Takayama, S., Brown, C. J., and Dunker, A. K. (2001) The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac Symp Biocomput*, 89-100
271. Tompa, P., and Fuxreiter, M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* **33**, 2-8
272. Banani, S. F., Lee, H. O., Hyman, A. A., and Rosen, M. K. (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298
273. Alberti, S. (2017) The wisdom of crowds: regulating cell function through condensed states of living matter. *J Cell Sci* **130**, 2789-2796
274. Hyman, A. A., Weber, C. A., and Julicher, F. (2014) Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol* **30**, 39-58
275. Shin, Y., and Brangwynne, C. P. (2017) Liquid phase condensation in cell physiology and disease. *Science* **357**
276. Gibson, B. A., Doolittle, L. K., Schneider, M. W. G., Jensen, L. E., Gamarra, N., Henry, L., Gerlich, D. W., Redding, S., and Rosen, M. K. (2019) Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **179**, 470-484 e421