Understanding ETS transcription factors:

from ordered domains to disordered sequences

by

Desmond Ka Wing Lau

B.Sc. (Honours), Simon Fraser University, 2005

M.Sc. Simon Fraser University, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2017

Abstract

ETS (E26 transformation specific) transcription factors play critical roles in regulating cellular growth, development and differentiation. They share a conserved ETS domain that interacts with specific DNA sequences and a subset of ETS proteins also contain a PNT domain responsible for protein partnerships. My research initially focused on the PNT domain of *Drosophila* Pointed-P2, with the goal of understanding the impact of phosphorylation on the activation of gene expression. Using a battery of NMR spectroscopic approaches, I demonstrated that the Pointed-P2 PNT domain contains a dynamic N-terminal helix H0 appended to a core conserved five-helix bundle. This helix must be displaced to allow docking of the PNT domain with the ERK2 MAP kinase Rolled, which in turn phosphorylates three N-terminal phosphoacceptor sites.

The second part of my thesis focuses on three members of the ETS family called the ETV1/4/5 sub-group: ETV1 (Er81), ETV5 (Erm), and ETV4 (PEA3 (polyoma enhancer activator 3)). Using an extensive set of ETV4 deletion fragments, the DNA binding autoinhibitory sequences at both N- and C-terminal to the ETS domain were identified. Through detailed NMR spectroscopic studies, I confirmed that the inhibitory sequences are predominantly disordered and transiently interact with a coarsely defined surface on the ETS domain. This surface overlaps the DNA-recognition interface, thus indicating a steric mechanism of autoinhibition. Overall, my studies help define the molecular mechanisms underlying ETV1/4/5 factors autoinhibition, and may inspire new anti-cancer strategies.

Finally, I also investigated the stability and dynamics of several uninhibited ETS domains, including ETV4, PU.1, Ets1, and ETV6. Using NMR spectroscopy, I determined the structure of the PU.1 ETS domain and identified an appended, C-terminal helix. Similar to Ets1 and ETV6, the DNA-recognition helix H3 of ETV4 and PU.1 are dynamic as evidenced by amide hydrogen exchange (HX). I also utilized molecular dynamics simulations to map the motions of the four ETS domains and identified several critical pathways that may impact their stabilities and possibly, the DNA-binding abilities. Overall,

the data presented in my thesis will provide further understanding of the structure and regulation of the ETS transcription factors.

Lay summary

Proteins such as the ETS transcription factors are biological "lego-blocks" that make up the "molecular machines" used by our cells to read the genes encoded within our DNA. By determining the three-dimensional shapes (structures) of these blocks, we can understand how they work correctly in normal cells and explain why cancers result when mutations occur. The goal of my research was to investigate how the shapes and motions of several ETS factors affect their functions. Using a technique called nuclear magnetic resonance (NMR) spectroscopy as a "molecular microscope", I characterized how a chemical modification (phosphorylation) influences the shape of the ETS factor Pnt-P2 and thereby defines how it interacts with other transcription regulators. I also investigated how DNA-reading by three closely related members of the ETS family, called the ETV1/4/5 factors, are auto-inhibited when one part of each protein changes the shape of its own DNA-reading surface.

Preface

Chapter 2 of this thesis is based on a slightly reformatted version of the paper: **Desmond K.W. Lau**, Mark Okon, Lawrence P. McIntosh (2012) *"The PNT domain from Drosophila Pointed-P2 contains a dynamic N-terminal helix preceded by a disordered phosphoacceptor sequence."* **Protein Science**. 21:1716-1725. I carried out all experiments with Mark Okon's assistance to record NMR data. Dr. McIntosh and I wrote and edited the paper.

Chapter 3 is based on an extensively reformatted version of the paper: Simon L. Currie*, **Desmond K.W. Lau***, Jedediah J. Doane, Frank G. Whitby, Mark Okon, Lawrence P. McIntosh, Barbara J. Graves (2017) *"Structured and disordered regions cooperatively mediate DNA-binding autoinhibition of ETS factors ETV1, ETV4 and ETV5."* **Nucleic Acids Research**. 45:2223-2241 (* co-first authors). All experiments regarding EMSA DNA binding assays, limited proteolysis and *in vitro* acetylation were carried out by Simon Currie in the laboratory of Dr. Barbara Graves at the University of Utah. X-ray crystallography were also performed by Simon Currie with Dr. Whitby's assistance. I performed all the NMR experiments presented in chapter 3, as well as the design and cloning of the *trans* peptide system. Mark Okon provided assistance to record NMR data. Dr. Currie, Dr. Graves, Dr. McIntosh and I wrote and edited the paper.

I performed all of the NMR experiments and thermal stability assays in chapter 4, with the exception that in spectral assignments and dynamics studies of PU.1 were done with the help of a summer student, Arion Lochner. MD simulations were performed by a colleague, Florian Heinkel, in a collaboration with Dr. Joerg Gsponer at UBC. A manuscript summarizing this research is under preparation.

In collaboration with Dr. Michael Cox, Dr. Paul Rennie and Dr. Artem Cherkasov at the BC Prostate Center, we identified small molecules that inhibit ERG transcription. My contribution was to characterize the interaction between the small molecules and the ERG

ETS domain using NMR spectroscopy and EMSA. Miriam S. Butler*, Mani Roshan-Moniri*, Michael Hsing*, **Desmond K.W. Lau***, Ari Kim, Paul Yen, Marta Mroczek, Mannan Nouri, Scott Lien, Peter Axerio-Cilies, Kush Dalal, Clement Yau, Fariba Ghaidi, Yubin Guo, Takeshi Yamazaki, Sam Lawn, Martin E. Gleave, Cheryl Y. Gregory-Evans, Lawrence P. McIntosh, Michael E. Cox, Paul S. Rennie, and Artem Cherkasov. (2017) *"Discovery and characterization of small molecules targeting the DNA-binding ETS domain of ERG in prostate cancer."* **Oncotarget**. Advance publication 17124. (* co-first authors)

Table of contents

List of tables

List of figures

Glossary

ATP	adenosine 5'-triphosphate
BP	base pair
CamKII	calmodulin kinase II
CBP	CREB binding protein
CD	circular dichroism
CID	C-terminal inhibitory domain
D_2O	deuterium oxide
Da	Dalton
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
ESI-MS	electrospray ionization mass spectrometry
ETS	E26 transforming specific
ERK2	mitogen activated protein kinase 2
Gd(DTPA-BMA)	gadolinium(III) 5,8-bis(carboxylatomethyl)-2-[2-(methylamino)-2-oxoethyl]-10- oxo-2,5,8,11-tetraazadodecane-1-carboxylate hydrate (trade name: gadodiamide or omniscan)
GST	glutathione S-transferase
HAT	histone acetyl transferase
HDAC	histone deacetylase
HDX	protium deuterium exchange
HSQC	heteronuclear single quantum correlation
HX	hydrogen exchange
IDPR	intrinsically disordered protein region
IM	inhibitory module
IPTG	isopropyl-βD-thiogalactopyranoside
K_D	equilibrium dissociation constant
MALDI-TOF	matrix-assisted laser desorption/ionization-time of flight

MAPK	mitogen activated protein kinase
MD	molecular dynamics
MES	2-(N-morpholino)ethanesulfonic acid
MESNA	2-mercaptoethane sulfonate Na (Na being the symbol for sodium)
MICS	motif identification from chemical shift
MTSL	S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate
NID	N-terminal inhibitory domain
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOESY	nuclear Overhauser enhancement spectroscopy
PCR	polymerase chain reaction
PDB	protein database (http://www.rcsb.org/pdb/)
PNT domain	pointed domain
PTM	post-translational modification
PRE	paramagnetic relaxation experiments
S^2	squared order parameter
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SRR	serine-rich region
SSP	secondary structure propensity
SUMO	small ubiquitin related modifier
TCEP	tris(2-carboxyethyl)phosphine
TF	transcription factor
TOCSY	total correlation spectroscopy
TROSY	transverse relaxation-optimized spectroscopy
wHTH	winged helix-turn-helix

Acknowledgements

The reason that I can finish the work presented in this thesis is because I was grateful to have so many people provided numerous guidance to help me develop and improve as a scientist over years in Lawrence's laboratory. First and foremost, I would like to thank my senior supervisor Lawrence McIntosh for first hiring me as a technician and later having me as a student. Under his guidance, I learn how to think critically and to expand my scientific knowledge especially in the field of structural biology. I would like to thank Eric, Genevieve, Jerome, Patrick, and Shaheen for sharing science and other aspects of life when I first joined the lab. I am grateful to Hanso for teaching and showing me around the lab and be a part of the exciting PNT domain project. I wish to thank the current and past members of the lab (Cecilia, Jacob, Florian, Chloe, Karl, Stacy, Ben, Miriam, Soumya, Laura, Adrienne, and Helen) for being there for coffee no matter how my science turned out. Without Mark, collecting NMR experiments would not been as smooth. The exciting results of ETV1/4/5 projects would not be possible without the collaboration with Simon Currie in Barbara Graves's lab. I would also like to thank Mani, Miriam B, and Michael for bringing me to the BC Prostate Center to learn about *in vivo* assays.

For my family, I owe tremendously to my wife Karina. She is always there to listen and to encourage me to continue no matter how my day ended. I wish to thank my two daughters, Kassidy and Brienne, for giving me a big smile and hug when I come home and to remind me that how awesome I am, at least to them. Lastly I would like to thank my parents and my in-laws for providing care to my daughters and still support me when I go through my ups and downs in life. Thank you everyone for your support through this long journey, I wouldn't make this far without any of you.

Chapter 1.   Introduction

1.1. Eukaryotic gene expression

Maintaining cell growth, differentiation and homeostasis, including responding to stimuli, requires a complex and delicately balanced gene expression program. Many regulatory elements act to tightly control the expression of any given gene, and to prevent unnecessary gene products that could subsequently lead to a broad range of diseases processes, including cancer. In general there are two main categories of regulatory elements. Cis-regulatory elements refer to the promoter or enhancer regions that are encoded in the DNA sequences of the genome and dictate which genes are to be transcribed. Trans-regulatory elements include transcription factors or co-regulators that occupy specific DNA sequences and activate/repress target genes. Overall, it is vital to manage the interplay between these regulatory elements, as well as the recruitment and activation of the RNA polymerase complex, in order to maintain normal cellular activity.

1.1.1. Transcription regulation

In eukaryotes, genes are transcribed into mRNA by RNA polymerase II (Djebali et al. 2012). The regulation of this polymerase is a highly complex process, involving a myriad of potential protein-protein and protein-DNA interactions. Central to transcription is the binding of sequence-specific transcription factors to cis-regulatory elements. These include proximal promoters that are usually located ~ 70 to 200 bp upstream to the core promoter (~ 30 bp before the transcription start site), as well as distal enhancer elements, that can be a few kb upstream or downstream from the 5' of the start site of transcription. In addition to DNA binding, sequence-specific transcription factors also interact with co-activator (e.g. p300, Mediator) and co-repressor complexes (e.g., Sin3), including those involved in chromatin modeling (Juven-Gershon & Kadonaga 2010; Malik & Roeder 2010; Sikorski & Buratowski 2009; Taatjes 2010). They also serve as scaffolds or bridges to recruit other regulatory proteins. Ultimately, this directs the general transcription factors,

such as TATA-binding protein (TBP) to bind the TATA-box of the core promoter regions. The binding of TBP in turn recruits other general transcription factors, including TFIIA, TFIIB, and RNA polymerase II. Subsequently, TFIIE and TFIIH join RNA polymerase II and the Mediator complex to form the Pre-Initiation Complex (PIC) and initiate transcription.
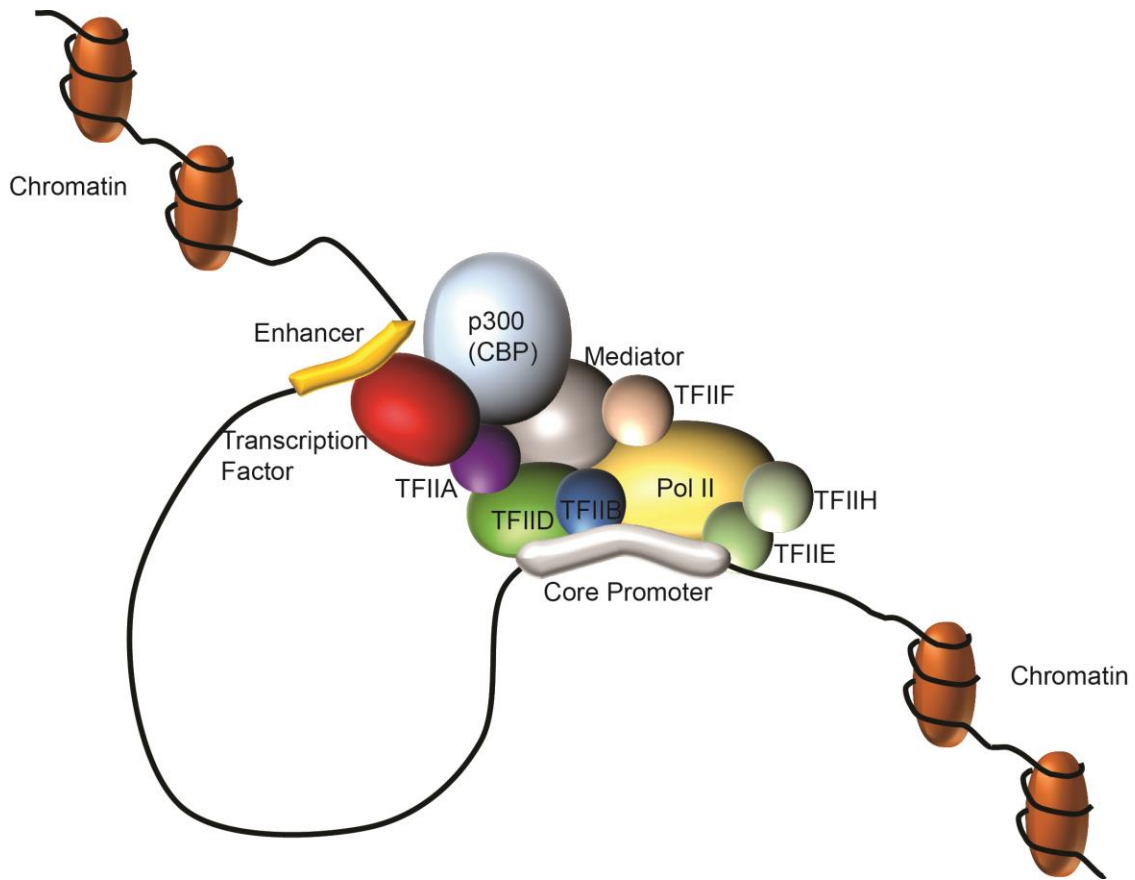
Figure 1-1 Overview of eukaryotic transcription.

A cartoon representation illustrates protein complex formation during the activation of the eukaryotic transcription. Specific regulatory proteins (transcription factors) bind to the upstream enhancer region. Transcriptional co-regulator (CBP) and mediators are then recruited to serve as scaffold, and provide acetyltransferase activity (HAT) for histone modification (acetylation). Collectively, this recruits RNA polymerase II along with other general transcription factors to form the Pre-initiation Complex (PIC). A key step of PIC formation is initiated by the binding of the TBP subunit of TFIID to the TATA box element in the core promoter region.

1.2. Sequence-specific transcription factors

Sequence-specific transcription factors are proteins involved in transcribing DNA into mRNA by aiding the recruitment RNA polymerase II and other co-factors. The ability of these transcription factors to bind specific cis-regulatory sites on DNA and thereby control the expression of given genes is crucial to cellular differentiation and development (Levine & Tjian 2003; Pan et al. 2010). These transcription factors possess at least one DNA-binding domain, as well as additional functional modules including possible protein- and ligand-interaction domains, and trans-activation or repression domains. They are also regulated at numerous levels, including their own expression and nuclear localization, as well as ligand binding and a wide range of post-translational modifications. As such, sequence-specific transcription factors detect and integrate information from multiple signal transduction pathways. Several transcription factors often cluster within promoter and enhancer regions and recruit other co-activators and chromatin-remodeling proteins to form a higher-order complex named the "enhanceosome" (Panne et al. 2007). The combinatorial assembly of various transcription factors within enhanceosomes helps drive the specificity of gene expression in a spatial and temporal manner.

In contrast to the relatively few general transcription factors that are common to expression of all genes, there are a remarkably high number (>2000) of specific transcription factors (Brivanlou & Darnell 2002). The latter can be classified into different families based on conserved DNA-binding motifs. Well-characterized DNA-binding motifs include, among others, leucine zippers (Figure 1-2A), basic helix-loop-helix domains (Figure 1-2B), zinc finger domains, winged helix-turn-helix domains (Figure 1-2C), and helix-turn-helix motifs (Figure 1-2D). In general, DNA-binding motifs make contacts with DNA via electrostatic, hydrophobic and hydrogen bonding interactions, as well as through water-mediated contacts. Specific DNA sequences are recognized through hydrogen bonds and van der Waals interactions between amino acid side chains and the nucleobases ("direct readout") or via interactions that are sensitive to the sequence-dependent shape of the phosphodiester backbone ("indirect readout") (Rohs et al. 2010). For example, transcription factor c-FOS belongs to the basic leucine zipper family and it

interacts with DNA by inserting "basic" arginine-rich helices into the major groove of the DNA, allowing specific hydrogen bonds to be formed between the guanidinium sidechains and the nucleobases (Figure 1-2A) (Pogenberg et al. 2014). Similarly, the transcription factor MAX contains a basic helix-loop-helix motif whereby positively charged residues make specific contacts along the major groove of the DNA. The members belonging to either of these families can form homotypic or heterotypic dimers through leucine zipper or helix-loop-helix motifs, immediately adjacent to the DNA-binding helices (Figure 1-2B).
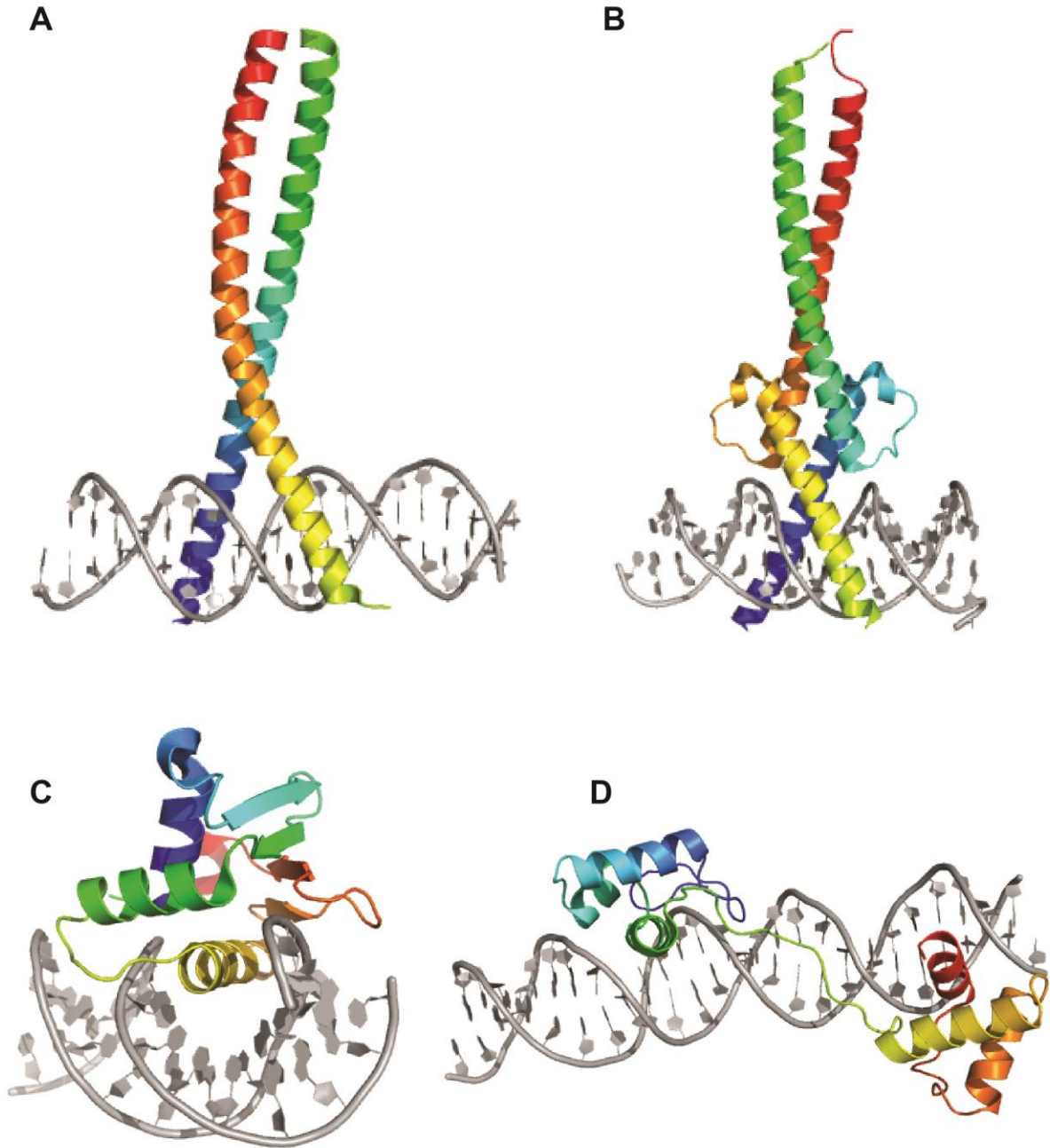
Figure 1-2 Examples of various DNA-binding motifs.

(A) Leucine zipper from Fos/Jun complex (PDB: 1fos). (B) Helix-loop-helix from Myc/Max complex (PDB: 1nkp). (C) Winged helix-turn-helix from ETV6 (PDB: 4mhg). (D) Helix-turn-helix from Pax5/Ets1 complex, with Ets1 removed for simplicity (PDB: 1mdm).

1.3. ETS transcription factor family

ETS (E26 transformation specific) transcription factors play critical roles in regulating cellular growth, development and differentiation (Hollenhorst, McIntosh, et al. 2011). The founding member of ETS transcription factor family, Ets1, was identified more than three decades ago within the open reading frame of an avian oncogenic virus responsible for erythroblastosis in chickens (Nunn et al. 1983). There are 28 identified human ETS paralogs, exhibiting both common and diverse properties (Figure 1-3). All ETS factors are classified by the presence of a conserved ETS domain responsible for interacting with DNA. The ETS family is divided further into sub-groups based on sequence conservation and the presence of additional functional domains. For example, several members of the ETS family contain a PNT domain, which is an ETS-specific variant of the widespread SAM domains that mediate diverse protein-protein and protein-RNA interactions (Figure 1-3). Other regions of the ETS factors appear to be intrinsically disordered, including those functioning as transactivation domains (Augustijn et al. 2002; Macauley et al. 2006; Lens et al. 2010).

ETS factors can be divided into sub-families based on their phylogenic relationships (Hollenhorst, McIntosh, et al. 2011) (Figure 1-3). Not surprisingly, members within a sub-family of ETS transcription factors have higher levels of sequence conservation of their ETS and PNT domains than members from different sub-families (Laudet et al. 1999). ETS sub-family members also often exhibit redundant functions. For example, in hematopoietic stem cells, the related Erg and Fli1 bind to a largely redundant set of targets (Wilson et al. 2010). Genome-wide analyses of ETS factors occupancy also reveal redundant occupancy of Ets1 and other ETS factors at promoters of "housekeeping" genes in T cells, likely because of common sequence preferences for a $^{5'}GGA(A/)T^{3'}$ consensus motif (Hollenhorst et al. 2009). Additional studies have shown that Ets1, GABPA, Elf1, and Spi1 have overlapping occupancies in different cell lines and species, with striking enrichment of the rather short sequence $^{5'}CCGGAAGT^{3'}$ (Hollenhorst et al. 2007; Boros et al. 2009). These ETS binding motifs are most frequently found ~ 20 to 40 bp upstream of the transcription start site of housekeeping genes that are genes

7

expressed in all cell types and function in normal cell growth and homeostasis. However, ETS factors also have specific targets. This specificity conundrum can be explained partly by the distinct biochemical properties of additional domains or sequences beyond the ETS domain, combined with additional factors, including post-translational modifications, protein partnerships and even the patterns of their tissue-specific expression and cellular localization.

Post-translational modifications are central to regulating gene expression. They influence the temporal and spatial activation of transcription factors in response to a molecular effector (e.g., a hormone) or other type of signals (e.g., stress). Many of the components involved in transcription initiation can be post-translationally modified. For example, acetylation of specific lysines provide binding motifs for proteins with bromodomains (Sanchez & Zhou 2009), whereas methylation recruits proteins with chomodomains. These recruited protein complexes often remodel the nucleosome and lead to gene activation or repression (Tajul-Arifin et al. 2003). Importantly, phosphorylation is one of the most common ways to regulate the activity of an ETS transcription factor. The ETS protein ELK-1 plays a critical role in chromatin remodeling and gene activation that is dependent upon its phosphorylation by MAPK (Li et al. 2003). Also, phosphorylation of ELK-1 at the transactivation domain results in enhanced DNA binding (Yang et al. 1999). Among other phosphorylation pathways, phosphorylation by Ras/MAPK kinase or CaM kinase II are two key regulators for the activation of ETS proteins, including Ets1, Ets2, ELK-1/3/4, GABPA, SPIB, and ETV1/4/5 (Hollenhorst, Ferris, et al. 2011). Ets1 and Ets2 factors can activate gene transcription by occupying tandem Ets-AP1 binding sites upon phosphorylation (Yang et al. 1996). The orthologs of human Ets1/Ets2 and ETV6 in *Drosophila*, named Pointed-P2 and Yan, respectively, are also linked to the "son-of-sevenless" Ras/MAPK signaling cascade that dictates eye development.

Aberrant activities of ETS transcription factors often result in altered gene expression leading to disease development, including oncogenesis. Therefore, the activation/repression of ETS transcription factors is tightly regulated. Chromosomal translocations that place the *Erg* or *ETV1/4/5* genes under control of the androgen-

responsive, prostate-specific *TMPRSS2* promoter result in their overexpression, which in turn drives early prostate neoplastic development (Huang & Waknitz 2009). Similarly, chromosomal translocations that fuse fragments of the genes encoding a functional PNT domain of ETV6 with a receptor tyrosine kinase domain, such as that of NTRK3, produces the chimeric oncoproteins that drive leukemogenesis (De Braekeleer et al. 2012). Since the RAS/MAPK signaling pathway is often activated in cancer, it suggests a link to the oncogenic nature of some of the ETS factors (Hollenhorst, Ferris, et al. 2011). Given the importance and frequent involvement in cancer development, one goal of my thesis research is to elucidate the molecular mechanisms underlying control of ETS factors by post-translational modifications, DNA-binding autoinhibition, and protein dynamics.
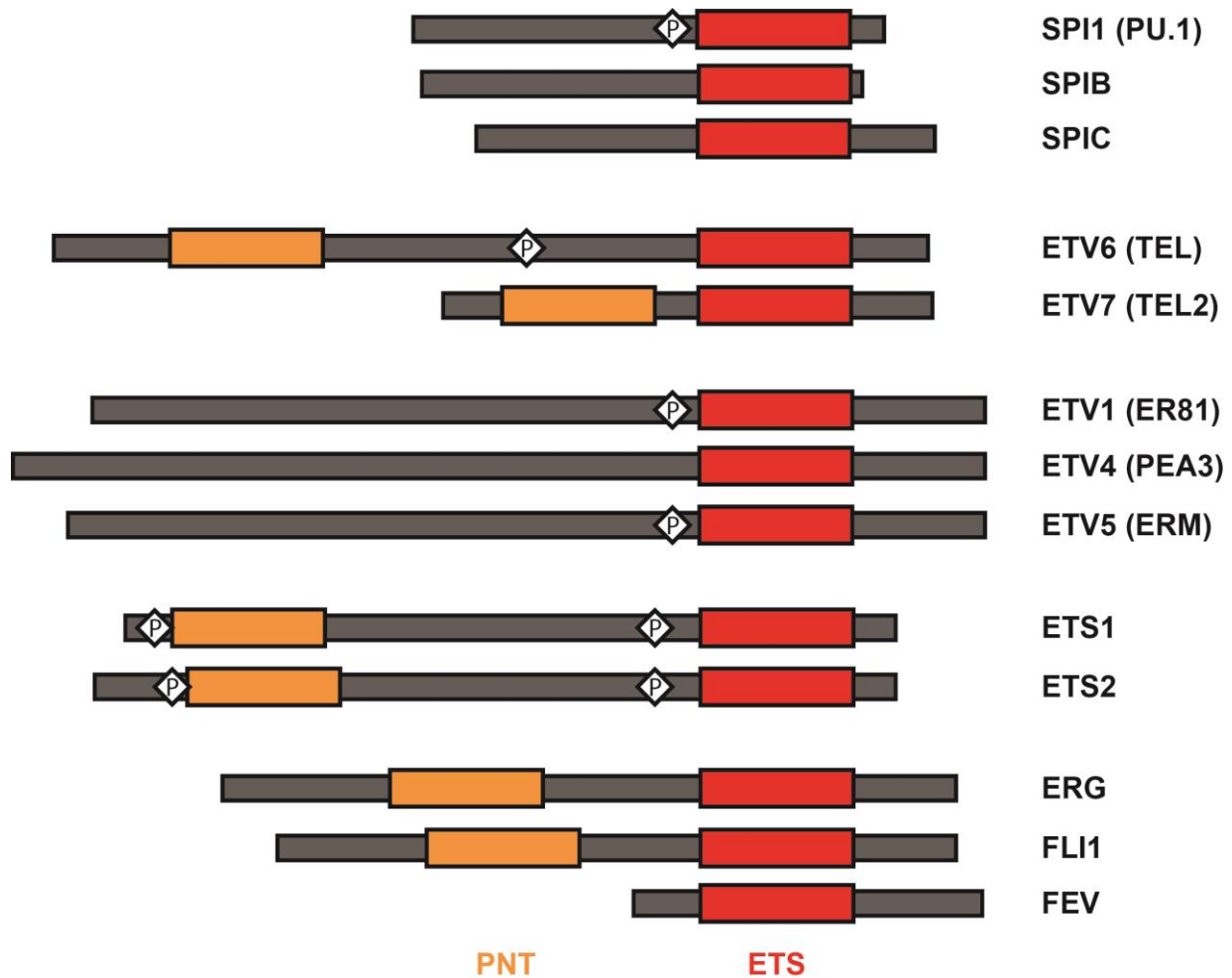
Figure 1-3 ETS transcription factor family.

This figure highlights some of the ETS proteins discussed in this thesis. They are grouped according to their phylogenetic sub-families, as listed in (Hollenhorst, McIntosh, et al. 2011). These sub-families are SPI (SPI1, SPIB, SPIC), TEL (ETV6, ETV7), PEA3 (ETV1, ETV4, ETV5). ETS (Ets1, Ets2), and ERG (ERG, FLI1, FEV). The boxes indicate the DNA-binding ETS domains (red) and PNT domains (orange). The diamond P indicates known phosphorylation sites discussed in the text.

1.3.1. ETS domain

The ETS factors have a modular architecture including a highly conserved ETS domain composed of ~ 85 amino acids. This "winged helix-turn-helix" domain folds as three α-helices on a four-stranded, antiparallel β-sheet scaffold (Figure 1-2C, Figure 1-4A). A less conserved fourth helix, H4, is also present on most, if not all, ETS domains. Helices H2 and H3 form the helix-turn-helix motif. ETS domains bind DNA over a region spanning 12 to 15 bp, but display sequence preference for only ~ 9 bp with a central, invariant $^{5'}$GGA(A/T)$^{3'}$ core. Specific contacts (direct readout) can form between side chains located in helix H3 and the nucleobases in the major groove of DNA. Crystal structures of ETS domain in complex with DNA reveal that the most important of these direct interactions are between two invariant arginines in the recognition helix H3 and the two guanines of the $^{5'}$GGA(A/T)$^{3'}$ core. Bases flanking the consensus core provide a small degree of specificity, despite the lack of direct contacts with the ETS domain. This indirect readout utilizes the "wing" between β-strand S3 and S4, and loop between α-helices H2 and H3 to recognize sequence-dependent positioning of the phosphodiester backbone. Despite having conserved ETS domains and conserved DNA target sites, detailed investigations show that ETS factors can interact with DNA via different mechanisms. For example, Ets1 binds DNA with a "dry" mechanism, whereas PU.1 interacts with DNA through hydration with specific water molecules that couple to backbone contacts (Wang et al. 2014). Also, as discussed below, additional helices appended to the ETS domains of various sub-families provide routes for regulation and transcriptional specificity.

1.3.2. PNT domain

Approximately one-third of all ETS factors also contain a PNT domain. The PNT domains are a subset of the widespread SAM domains, which were initially identified by Ponting on the basis of the conservation of ~ 70 amino acids domain in 14 eukaryotic proteins (Ponting 1995). The predicted helical structure of this domain and its presence in yeast proteins that are essential for sexual differentiation gave rise to the name sterile alpha motif (SAM). The PNT domains share a common core architecture of four to six α-helices,

11

yet exhibit different association states and function in a wide variety of protein-protein and protein-RNA interactions (Qiao & Bowie 2005). Indeed, PNT domains are usually found in the context of larger multidomain proteins that may be present in all cellular compartments (Qiao & Bowie 2005). One source of regulation of the ETS factors arises from helices appended on their PNT domains. Yan, ETV6, ERG, Elf3 and Fli1 contain only the minimal helical bundle, GABPA and SPDEF have one additional N-terminal helix (Mackereth et al. 2004) and those of Ets1 and Ets2 have two (Figure 1-4B) (Nelson et al. 2010). This variation generates different surface features (Meruelo & Bowie 2009). This gives rise to the self-association of the PNT domains of ETV6 and Yan (Kim et al. 2001; Qiao et al. 2004) while other PNT domains are monomeric in isolation. In addition, post-translational modifications of PNT domains enable regulation. For example, Ras/MAPK signaling kinase phosphorylates Ser127 of Yan, a residue located C-terminal to the PNT domain to abrogate Yan's transcriptional repressor activity and to facilitate Crm-1adapted translocation out of the nucleus (Tootle et al. 2003; Rebay & Rubin 1995).
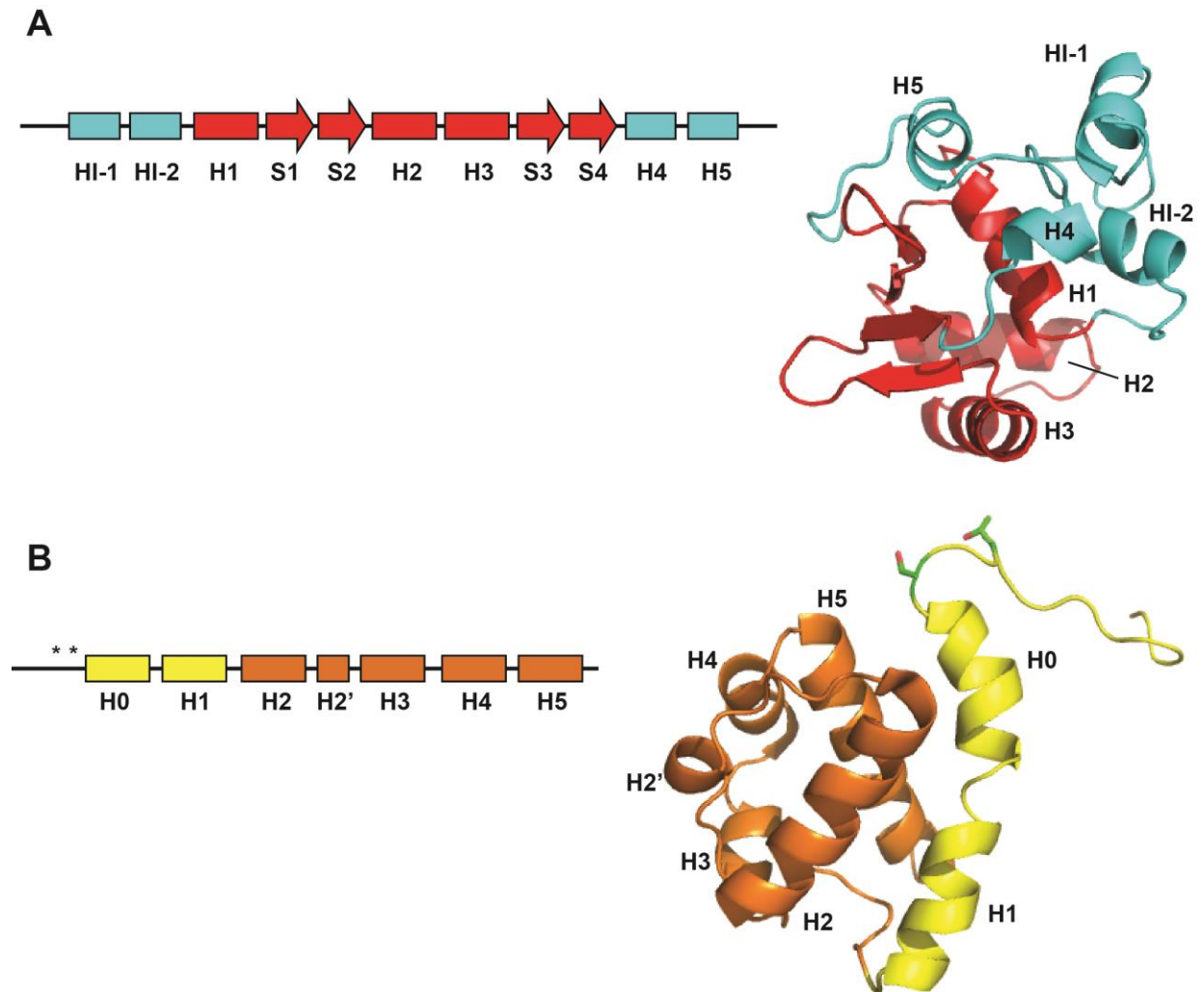
Figure 1-4 Structural overview of Ets1.

(A) The conserved Ets1 ETS domain (H1 – S4; red) and the DNA-binding inhibitory module (HI-1/HI-2/H4/H5; cyan). Helix H3 is inserted into the major groove of the DNA (not shown) to make specific contacts between two invariant arginines on H3 and the bases of the specific DNA. The marginally stable HI-1 and HI-2 are located distal to the DNA-binding interface and are unfolded upon DNA binding. (B) The PNT domain of Ets1 (orange) with its appended dynamic regulatory elements (H0/H1; yellow). The two asterisks correspond to the phosphoacceptors (Thr38 and Ser41) and are shown as stick format (carbon, green; oxygen, red) on the cartoon. The dynamic H0 is displaced upon phosphorylation of Thr38 and Ser41 to allow high affinity binding to its substrate such as Taz1.

1.4. Common regulatory mechanisms of ETS transcription factors – autoinhibition and phosphorylation-dependent interactions

Autoinhibition refers to a regulatory mechanism for repressing protein function via intramolecular (monomeric systems) or intracomplex (oligomeric systems) interactions (Pufall & Graves 2002). In general, one region of a protein or protein complex interacts with another to negatively regulate its activity. This can be achieved sterically or allosterically. Autoinhibition has several advantages associated with the "on-site" presence of a regulatory module that is at a constant high local concentration relative to separated inhibitory and functional proteins. Autoinhibition can also be regulated through a variety of mechanisms, including post-translational modifications, ligand binding, and proteolysis. As summarized below, the DNA binding activities of several ETS factors has been reported to be autoinhibited via several distinct mechanisms. This also provides a route to ETS factor specificity.

1.4.1. ETV6 (or TEL)

ETV6 is a member of the ETS transcription factor family with a conserved DNA-binding ETS domain, as well as a self-associating PNT domain (Figure 1-3). ETV6 was initially discovered as TEL (Translocation Ets Leukemia) in which a translocation that fused *ETV6* with *PDGFRB* gene was found in a patient with acute myeloid leukemia (AML) (Bohlander 2005). It was later renamed to ETV6 in an effort to systematize ETS factor nomenclature; this also avoids confusion with the abbreviation for telomere. ETV6 is unique as it was identified as a transcriptional repressor (Hiebert et al. 1996; Chakrabarti & Nucifora 1999). The transcriptional repression activity is mediated through the PNT domain and the central region of ETV6 via distinct mechanisms. The central region recruits corepressors including SMRT and mSin3A and repression can be relieved by inhibiting histone deacetylases (HDAC). In contrast the PNT domain likely polymerizes to create extended complexes that are postulated to wrap around DNA and cause transcriptional repression (Bohlander 2005). Indeed, the distinct DNA-binding properties of ETV6 arise from the fact that the PNT domain can form a very stable head-to-tail polymer (Kim et al. 2001; Qiao &

Bowie 2005), thereby facilitating cooperative DNA binding with multiple ETS binding sites (Figure 1-5) (Green et al. 2010). The ability of ETV6 to polymerize via the PNT domain thus provides a distinct route to counteract autoinhibition.

The DNA-binding autoinhibition of ETV6 occurs via a well characterized steric mechanism, in which an appended C-terminal inhibitory domain (helix H5) directly blocks the DNA-binding interface of the adjacent ETS domain (Figure 1-6A) (Green et al. 2010; Coyne et al. 2012). Detailed investigations revealed that helix H5 has low helical propensity and is only marginally stable. As expected, helix H5 unfolds upon binding to both specific and non-specific DNA. Conversely, mutations engineered to stabilize helix H5 re-enforce autoinhibition, and a disulfide bond that "staples" H5 to the ETS domain severely impairs DNA binding (De et al. 2014).

The mechanisms by which autoinhibition contribute to the regulation of ETV6 activity remain to be established. At the simplest level, autoinhibition could act to slow the kinetics of DNA association (which requires helix H5 unfolding) while ensuring that bound ETV6 has a long residency time (slow dissociation) at target DNA sites. This would also work well with the polymerization of native ETV6 via the head-to-tail association of its PNT domain. That is, polymerization could compensate for the low affinity caused by autoinhibition to favor cooperative binding to tandem, rather than isolated, target DNA sites. At a more complex level, a linker inhibitory domain (LID), located N-terminal to the ETS domain has been proposed to de-repressing autoinhibition (Green et al. 2010). Furthermore, the LID also contains phosphoacceptor sites and thus may enable regulation via undefined kinase signaling pathways. However, the biological validity and molecular mechanisms underlying these potential phenomena have not been established.
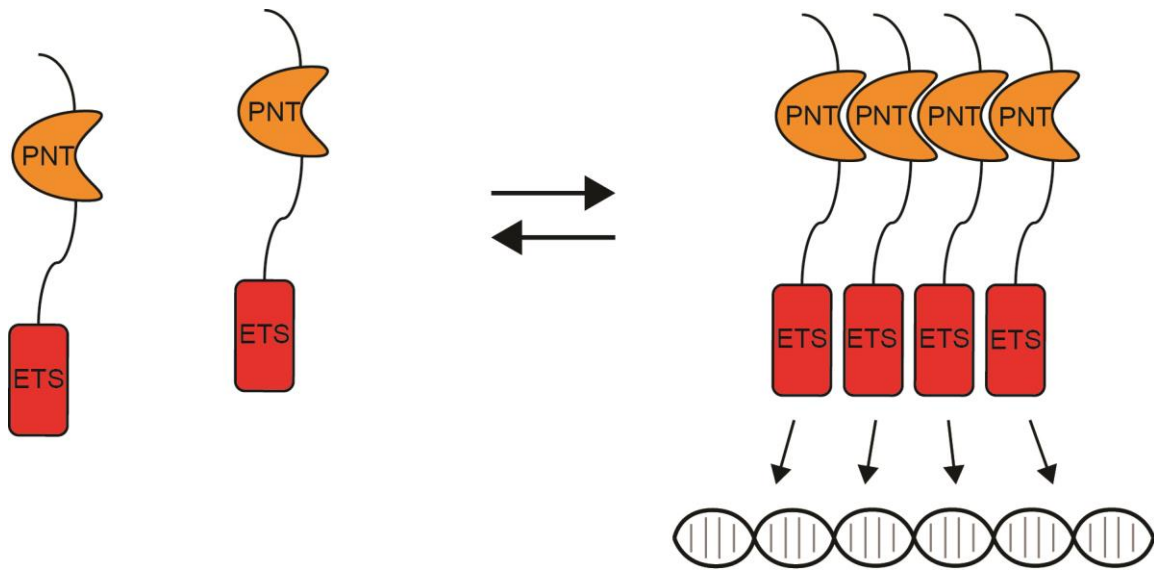
Figure 1-5 ETV6 can form polymers in a head-to-tail fashion.

This is a modified version of the original figure from (Green et al. 2010). Head-to-tail polymerization of the PNT domain enables cooperative binding of the ETS domain to tandem DNA sites.
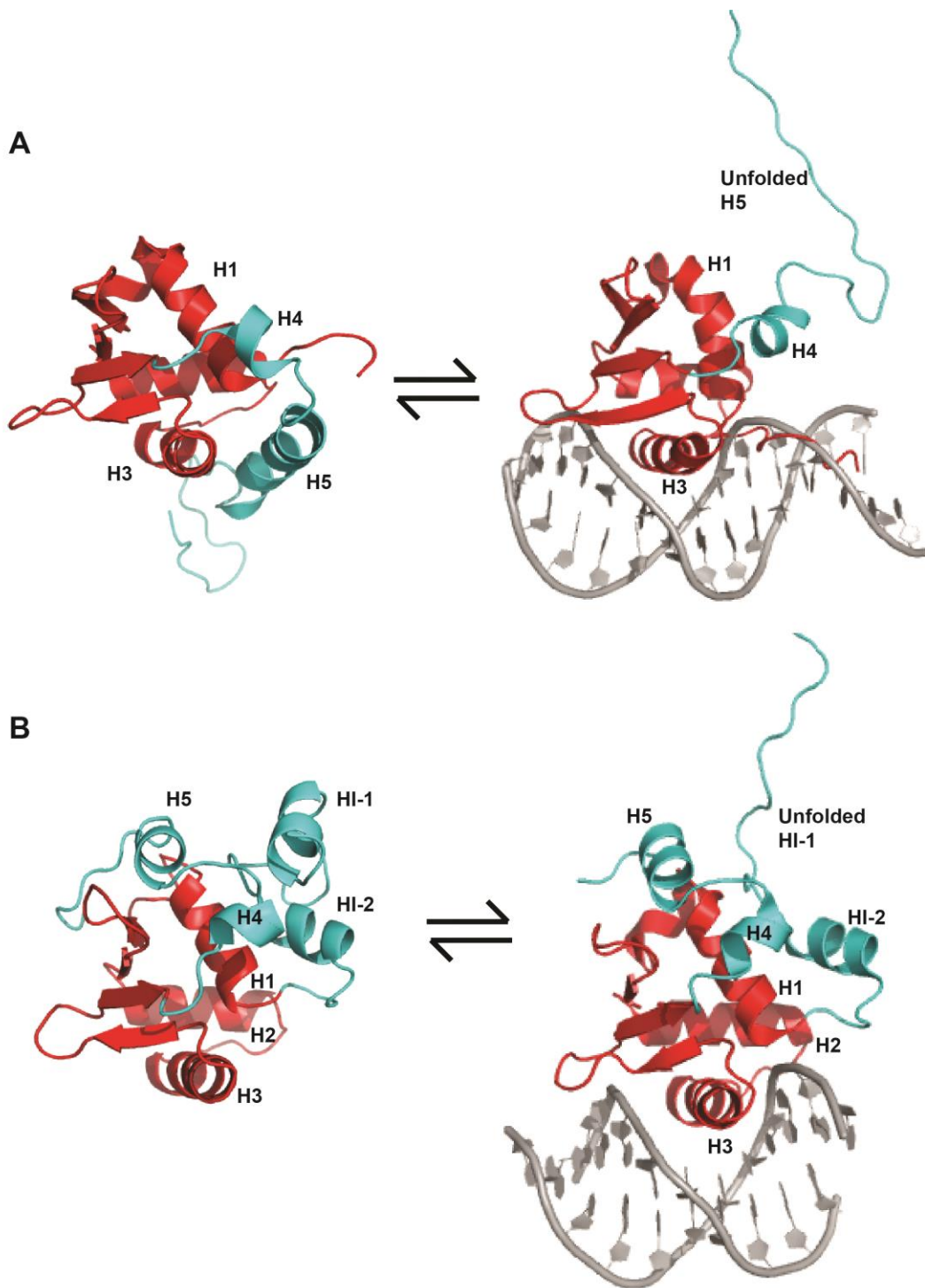
Figure 1-6 Autoinhibition mechanism of ETV6 and Ets1.

(A) ETV6 has an α helix H5 near the DNA-binding helix H3 that sterically interfere with DNA binding. (B) Inhibitory module HI-1 and HI-2 of Ets1 are distal from the DNA-binding site. In both cases, the inhibitory module unfolds upon binding with DNA.

1.4.2. Ets1

Ets1, as the founding member of the family, has been extensively characterized. Ets1 is closely related to v-Ets, originally identified as one of the transforming components of the E26 avian leukemia retrovirus. Cellular Ets1 plays roles in cell proliferation, differentiation, lymphoid cell development, transformation, invasiveness, angiogenesis and apoptosis (R. Li et al. 2000). Ets1 protein is expressed at high levels in the lung, spleen, and thymus during embryonic and post-natal development (Hollenhorst et al. 2007). Ets1 is a 50 kDa protein that contains 440 residues and is highly conserved (>80%) between different species. Ets1 is composed of three modular domains: the N-terminal protein interaction PNT domain, the central transactivation domain (TAD), and the C-terminal DNA-binding ETS domain. Structures of the PNT domain and ETS domain of Ets1 have been solved (Nelson et al. 2010; Lee et al. 2005; Shiina et al. 2014; Garvie et al. 2002). Along with detailed biochemical and biological studies, these structures provide important mechanistic insights into regulation of Ets1 via a complex interplay of post-translational modifications and protein-protein interactions.

Extensive work has been done to understand the function of Ets1 PNT domain. Importantly, disordered regions immediately preceding the PNT domains of closely related mammalian Ets1/Ets2 and *Drosophila melanogaster* Pointed-P2 contain phosphoacceptor sites for the orthologous Ras-activated MAP kinase ERK2 and Rolled, respectively (Nelson et al. 2010; Klämbt 1993; Wasylyk et al. 1997). Briefly, phosphorylation of Thr38 and Ser41 displaces the marginally stable helix H0 from the core helical bundle (H2-H5) (Figure 1-7). This exposes a negatively charged surface that contributes to the electrostatically driven interaction of the PNT domain with the positively charged TAZ1 domain of the general transcriptional co-activator acetyltransferase CBP (Nelson et al. 2010; Foulds et al. 2004; Yang et al. 1998). Intriguingly, the PNT domain serves as both a docking site for ERK2 to enable phosphorylation of these residues, as well as in interface for binding the TAZ1 domain. The flexibility of PNT domain is therefore speculated to be crucial in regulating its activity and specificity. One goal of my thesis

research was to explore the Ras signaling mechanism of the *Drosophila* homolog, Pointed-P2.
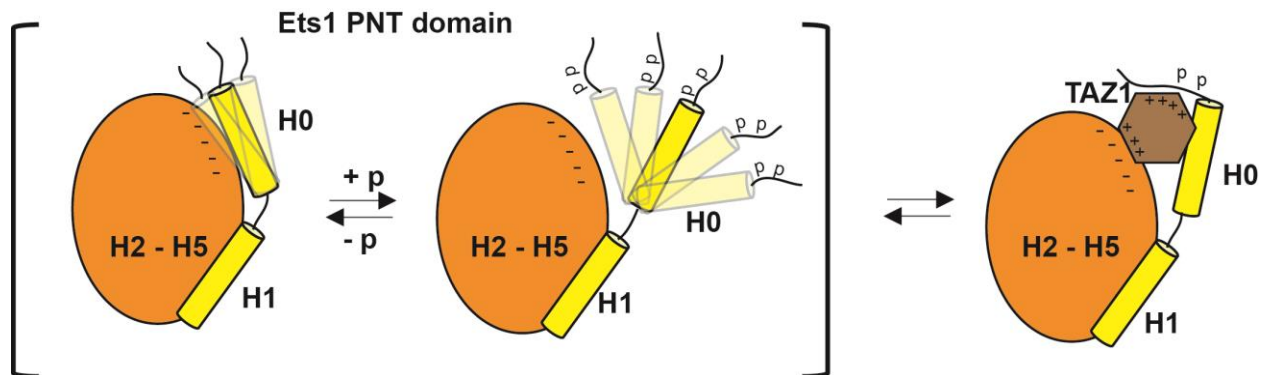
Figure 1-7 The phospho-switch model for the interaction of Ets1 PNT domain and TAZ1 domain of CBP.

The PNT domain exists in a conformational equilibrium with the dynamic helix H0 packs against the PNT domain (closed state) and away from the PNT domain (open state). Phosphorylation shifts the equilibrium towards the open state, which is favored for TAZ1 binding via complementary electrostatic interactions.

The ETS domain of Ets1 is autoinhibited for DNA binding by an appended helical bundle (Petersen et al. 1995; Lee et al. 2005). This bundle is formed by two N-terminal (HI-1 and HI-2) and two C-terminal (H4 and H5) helices interfaced with helix H1 of the ETS domain. Upon specific and non-specific DNA binding, the marginally stable helices HI-1 and HI-2 unfold (Figure 1-6B). In contrast to the case of ETV6, these helices are distal to the DNA-binding interface, indicating that autoinhibition follows an allosteric mechanism. However, the inhibitory helices only attenuate DNA binding by ~ 2-fold, and the full ~ 20-fold effect exhibited by wild-type Ets1 requires the presence of an adjacent intrinsically disordered serine rich region (SRR). The SRR transiently interacts with the inhibitory module, increasing its stability and dampening motions in the ETS domain. This shifts Ets1 from a more flexible state that is active for DNA binding to a more rigid inactive state (Pufall et al. 2005). In addition to this allosteric effect, the SRR also interacts directly with the DNA-

19

binding interface of Ets1 to sterically reinforce autoinhibition (Pufall et al. 2005). The transient interactions of the SRR with the ETS domain and inhibitory module are progressively strengthened upon multi-site phosphorylation of the SRR by calcium/calmodulin-dependent protein kinase II. Furthermore, protein partnerships, such as with Runx1, or even with Ets1 itself, relieves autoinhibition. Collectively this provides a fascinating route for integrating cellular signaling to enable the graded (or "rheostatic") control of Ets1 at the level of DNA binding.

Beyond the structured PNT and ETS domains, Ets1 appears to be intrinsically disordered. Within these disordered regions lie several sites of post-translational modifications. For example, sumoylation of Lys15 near the PNT domain or Lys227 in the TAD represses the transcriptional activity of Ets1, likely by recruitment of histone deacetylases (HDAC) or death-associated protein (DAXX) (Hahn et al. 1997; Macauley et al. 2006; M. Li et al. 2000; Ji et al. 2007).

1.4.3. ETV1/4/5

ETV1/4/5, also known as the PEA3 group, is composed of three ETS transcription factors: ER81 (ETV1), PEA3 (ETV4), and ERM (ETV5). The founding member of this family, PEA3 (Polyomavirus Enhancer Activator 3), was first identified as an adenovirus E1A enhancer-binding protein to regulate viral DNA replication (Xin et al. 1992). Endogenous PEA3 protein in human HeLa cells occupies the promoter sites upstream of genes involves in cell growth, migration, and differentiation (Gutman & Wasylyk 1991). All three PEA3 group members contain a DNA-binding ETS domain that is conserved and highly identical (~ 95%) to each other, as well as two transactivation domains (TAD) near the N-terminus and C-terminus of the protein (de Launoit et al. 1997). As with all ETS factors, ETV1/4/5 bind to the ETS consensus motif encompassing the core $^{5'}$GGAA/T$^{3'}$. The sequence conservation of the ETV1/4/5 proteins is consistent with their similar functions as transcriptional activators. Their biological roles include promoting muscle cell differentiation (Taylor et al. 1997) and the development of sensory neurons (Lin et al. 1998).

These ETV1/4/5 factors are of great medical interest given their frequent involvement in prostate cancer and diseases. Chromosomal rearrangements that place the *ETV1/4/5* genes with the androgen responsive promoter *TMPRSS2* result in overexpression of full-length or truncated versions of the *ETV1/4/5* genes in patients diagnosed with prostate cancer (Hollenhorst, Ferris, et al. 2011). The overexpression of *ETV1* and *ETV5* subsequently increases the invasiveness of prostate cell lines (Cai et al. 2007). In addition, >40% of melanomas and most gastrointestinal stromal tumors express high levels of *ETV1* (Chi et al. 2010; Jané-Valbuena et al. 2010).

The ETV1/4/5 factors are also autoinhibited for DNA binding. Early studies indicated that this autoinhibition was mediated by sequences flanking their ETS domains (Bojović & Hassell 2001; Greenall et al. 2001). In contrast to the cases of Ets1 and ETV6, proline scanning mutagenesis hinted that the inhibitory sequences did not adopt helical structures. As presented in Chapter 3 of my thesis, we now know that DNA-binding by ETV1/4/5 results from the cooperative interactions of intrinsically disordered sequences N-terminal to the ETS domain and an appended helix C-terminal to the ETS domain. This divergent mechanism for autoinhibition may contribute to the specificity of the ETV1/4/5 factors and provide distinct routes for their regulation. If so, then small molecules that reinforce autoinhibition could become lead compounds for the development of specific drugs against PEA3-related cancers.

ETV1/4/5 factors are also regulated through protein partnerships. For example, the interaction of ETV4 with USF-1 helps control the expression of the *bax* gene that is involved in apoptosis (Firlej et al. 2005). The interaction is thought to relieve DNA-binding autoinhibition of ETV4 in a similar mechanism to the Ets1-Runx1 complex, whereby the two transcription factors disrupt each other's inhibitory modules upon cooperatively binding a composite promoter site (Greenall et al. 2001). As a second example, ETV5 associates with the basal transcription complex proteins $TAF_{II}60$, TBP and $TAF_{II}40$ for transcription initiation (Defossez et al. 1997). Interestingly, ETV5 has a dual role as a transcriptional repressor by interacting with the androgen receptor (Schneikert et al. 1996)

and as a transcriptional activator when associating with the AP1 complex protein c-Jun (Nakae et al. 1995). In addition, ETV1 and ETV4 are able to interact with the p300 transcriptional coactivator histone acetyltransferase domain. This coactivator is essential in chromatin remodeling (Goel & Janknecht 2003; Liu et al. 2004).

A second route for modulating the transcriptional activity of ETV1/4/5 is via post-translational modifications. The most common modification found in the ETV1/4/5 proteins is phosphorylation. ETV1/4/5 can be phosphorylated by two independent signaling cascades: growth factor dependent ERK1/2 (extracellular signal related kinase) and stress factor dependent PKA (protein kinase A) (O'Hagan et al. 1996). In particular, ETV1 and ETV5 are phosphorylated at Ser334 and Ser367, respectively, by PKA to increase their transcriptional capacities (Baert et al. 2002; Bosc et al. 2001). In the case of ETV5, Ser367 is located at the N-terminus of the ETS domain and phosphorylation significantly reduces its DNA-binding affinity, possibly due to the addition of a negatively charged phosphate to compete with the DNA (Cooper et al. 2015). Notably, ETV4 lacks the conserved serine residue, suggesting that PKA would not inhibit its DNA-binding. However, transcriptional activation of zebrafish ETV4 is potentiated by the PKA pathway, indicating that phosphorylation by PKA not only affects DNA binding, but also provide an alternative mechanism for transcriptional activation (Brown et al. 1998). Also, ETV1/4/5 are targets of the MAPK pathway including Ras, Raf-1, MEK, ERK1 and ERK2. These kinases specifically phosphorylate threonine and serine residues on ETV1/4/5 that are situated outside of the ETS domain, and they generally increase the transactivation capacity of the protein (Figure 1-3) (de Launoit et al. 2006).

Another type of post-translational modification to regulate the transcriptional activity of ETV1/4/5 is acetylation of specific lysine residues. For instance, Lys33 and Lys116 of ETV1 are acetylated by transcriptional coactivator p300. This subsequently enhances the ability of ETV1 to transactivate gene expression. (Goel & Janknecht 2003). As will be shown in Chapter 3, acetylation of Lys226 and Lys260 also counteracts autoinhibition to increase the affinity of ETV4 for DNA. Since ETV5, but not ETV4 possesses a lysine residue homologous to Lys116 in ETV1, DNA-binding of ETV5 is likely to increase upon

acetylation by p300 (Guo et al. 2011). Lysine residues of ETV1 and ETV5 are also post-translationally modified by conjugating with the ubiquitin-like protein SUMO (Gocke et al. 2005; Degerny et al. 2005). SUMO modification of Lys89, Lys263, Lys293, and Lys350 of ETV5 causes inhibition of transcription without affecting the subcellular localization, stability, or DNA-binding capacity of the protein (Degerny et al. 2005). Furthermore, ETV4 and ETV5 are degraded via the 26S ubiquitin-proteasome pathway, potentially downregulating these transcription factors (Baert et al. 2007; Takahashi et al. 2005).

1.5. Intrinsically disordered proteins

With the rapid advances of computational and DNA sequencing technologies, genomic sequences and predicted full-length protein sequences are readily available. However, a large proportion of these sequences lack any useful function annotation (Raes et al. 2007). Available annotations are largely based on the detection of sequence homology with well-characterized protein segments. For the most part, this information is limited to proteins that have been structurally characterized by X-ray crystallography or NMR spectroscopy, and thus corresponds to domains with well-defined folded three-dimensional structure.

Over the past decade, many reports have shown that proteins often execute their functions through regions that lack any defined three-dimensional conformation under physiological conditions (reviewed in (Lee et al. 2014)). Protein structures can be described in a continuum ranging from fully structured to completely disordered, with intermediate states that could include compact molten globules containing extensive secondary structure, and unfolded states with transiently populated local elements of secondary structure (Figure 1-8A). Proteins are often composed of combinations of structured and intrinsically disordered regions (IDRs). However, some lack any structured domains, and are thus referred to as intrinsically disordered proteins (IDPs). IDRs or IDPs are most commonly found in eukaryotes. For example, ~ 40% of human protein-coding genes contain disordered segments of > 30 amino acids in length (Oates et al. 2013).

IDRs and IDPs can often be identified based on their low complexity sequences. These sequences tend to be biased towards charged and polar amino acids and generally lack bulky hydrophobic groups (Uversky et al. 2000; Romero et al. 2001; Lise & Jones 2005; Weathers et al. 2004). Such biased amino acid compositions weaken the hydrophobic forces that normally drive the folding of polypeptides into compact tertiary structures. In general, IDRs can be described as an ensemble of rapidly interconverting conformations, but depending on the distribution of charged and polar amino acids, IDRs can exist as disordered globule or swollen coils (Mao et al. 2010; Müller-Späth et al. 2010). IDRs are highly dynamic and often adopt induced structures upon interacting with another protein (Wright & Dyson 2009). Indeed, coupled folding and binding is one of the important functions of an IDR. For example, the phosphorylated kinase-inducible domain (pKID) of the transcription factor cyclic-AMP-response-element-binding protein (CREB) is unstructured in its apo form, but it rapidly folds into an α-helix upon complexation with the KID-binding (KIX) domain of CREB-binding protein (CBP) (Radhakrishnan et al. 1997).

1.5.1. Intrinsically disordered protein function

The importance of intrinsically disordered proteins often lies with the regulation of biological processes. Their functions include regulation of transcription and translation, cellular signal transduction, regulation of self-assembly of large multiprotein complexes and many more such processes (Dunker et al. 2002). IDRs also frequently contain post-translational modification sites. A single protein may consist of several disordered regions that exhibit different functions. In the simplest case, an IDR can function as a flexible linker to allow movement of flanking structured domains, forming so-called "beads-on-a-string". For example, the linker regions of CBP bridge several functionally important domains (TAZ1/2, KIX, Bromo, and HAT), and provide flexibility for the assembly of the transcriptional machinery. Detailed investigation of the linkers reveals that they are often conserved between species in terms of overall amino acid composition, but not in exact sequence and length (Dyson & Wright 2005). However, within linker regions, one often finds short conserved patches of amino acids, rich in charged, hydrophobic or proline residues. Such sequences often function as linear interaction motifs, recognizable by a

plethora of structured docking domains. For example, the nuclear-receptor interaction domain located at the N-terminus of CBP contains the disordered KHKXLXXLL motif that mediates binding to the nuclear receptor (Heery et al. 1997).

Post-translational modifications of IDRs are particularly relevant in transcriptional regulation and signaling. As described above, the IDR N-terminal to helix H0 of the Ets1 PNT domain serves as a docking site for the kinase ERK2, and subsequent phosphorylation of two conserved serines leads to increased affinity for the TAZ1 domain of CBP (Nelson et al. 2010). As a second example, the linker between the KIX domain and Bromo domain of CBP is rich in charged residues and contains two SUMO-modification sites that are required for transcriptional repression (Girdwood et al. 2003). The flexibility of IDRs allows for easy access and recognition of the post-translational modifications by "writer", "eraser" and "reader" proteins. The interaction between IDRs and it partner is often transient and weak, but specific. This is because upon binding, the flexible and disordered region loses conformational freedom; the associated entropic cost weakens the overall free energy of binding (Kriwacki et al. 1996; Wright & Dyson 1999; Dyson & Wright 2005).

1.5.2. Disordered proteins and fuzzy complexes

Intrinsically disordered sequences often adopt ordered conformations upon binding to macromolecular partners, including proteins and nucleic acids. Emphasis in the literature has been placed upon defining whether this occurs through a conformational selection process, whereby the IDR transiently adopts a binding-competent conformation in the absence of its partner, or through an induced fit mechanism, whereby folding occurs after binding (Wright & Dyson 2009). However, these are two extreme views of a more likely continuum of possible binding/folding pathways.

Remarkably, some IDRs appear to retain a significant amount of disorder even in a bound state. Disorder in the bound state is referred to as fuzziness and fuzzy complex conformations cover a structural continuum ranging from static to dynamic, and from fully

to partially disordered (Tompa & Fuxreiter 2008). Some examples of fuzzy complexes are illustrated in Figure 1-8B-E. It has been proposed that the bound, yet disordered regions, can impact the interaction affinity and specificity of the complex and tune interactions of folded regions with proteins or DNA (Fuxreiter et al. 2011).

A

Disorder → Order

extended | transient secondary structure | compact globule | molten globule | disordered loop | folded protein
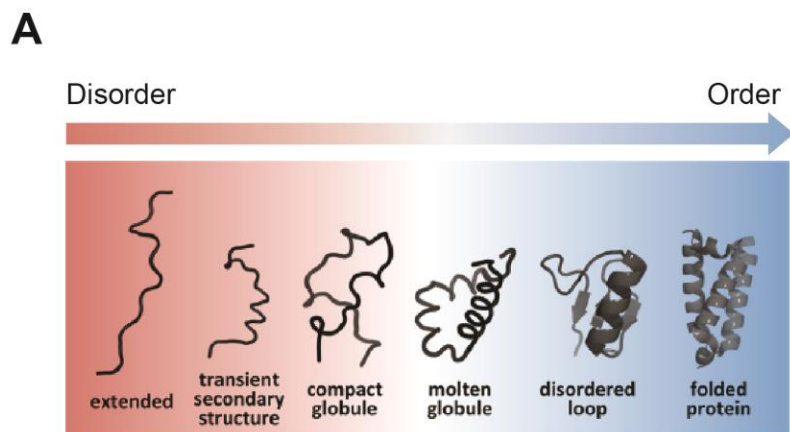
B
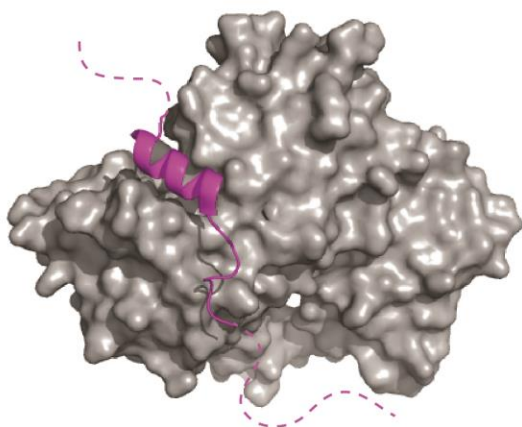
C
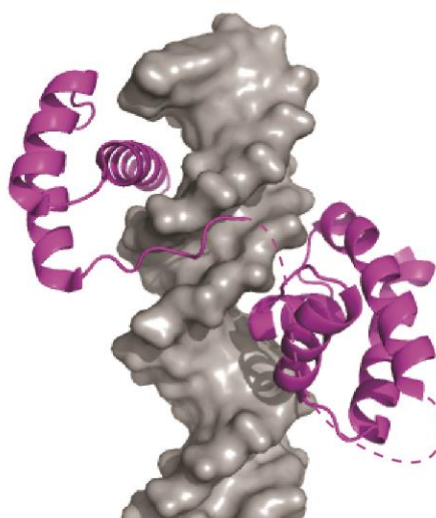
D
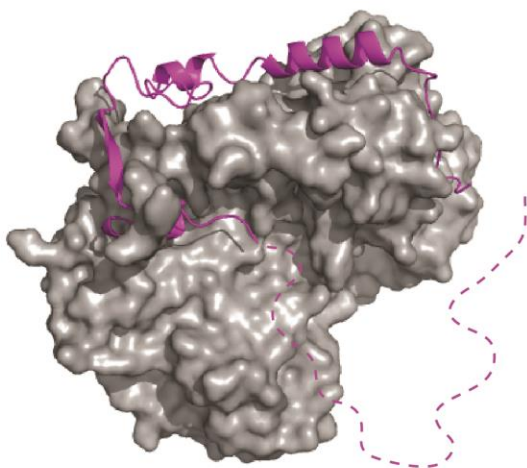
E
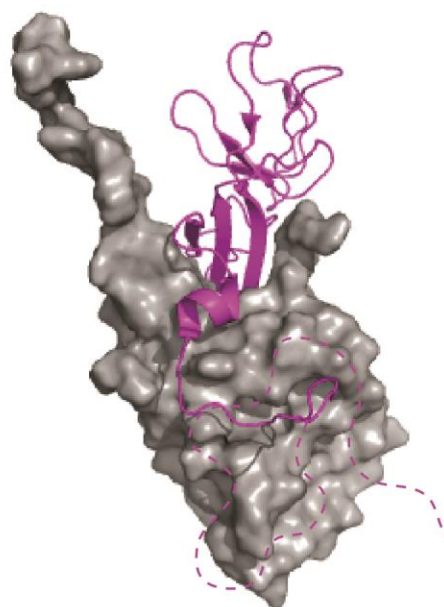
Figure 1-8 Schematic representation of the continuum of protein structures with examples of intrinsically disordered regions forming interactions.

This is a modified version of the original figure from (Lee et al. 2014)**.** (A) Color gradient represents a continuum of conformational states ranging from highly dynamics, extended conformations (red) to highly order, fully folded states (blue). (B-E), examples of different types of fuzzy complexes. (B) The WH2 domain (gray) of ciboulot interacts with actin via an 18-residue segment (magenta) (PDB: 3U9Z) with flanking regions remaining dynamically disordered (dashed lines). (C) The Oct-1 transcription factor (magenta) has a bipartite DNA recognition motif. The two globular binding domains are connected by a 23-residue long disordered linker (PDB: 1HF0), DNA (gray). (D) A cell-cycle kinase inhibitor (magenta) binds to the cyclin-Cdk2 complex (gray) (PDB: 1JSU). The kinase binding site is flanked by a long disordered linker. (E) UmuD2 is a dimer (PDB: 1I4V) but retains a significant amount of random coil character.

1.6. Protein dynamics

Biochemical thinking is often dominated by the simplistic view that the functions of proteins can be understood based upon their static three-dimensional structures. However, it is also well recognized that proteins are dynamic and often change their conformations in order to carry out functions including among many others, ligand binding, enzymatic catalysis, and allosteric regulation. Indeed, protein structures are stabilized by large networks of weak interactions, and thus they are easily rearranged by thermal motions at biologically relevant temperatures (Tavernelli et al. 2003). This provides proteins with an intrinsic dynamic character that is better represented by an ensemble of inter-converting conformers, rather than one single mean structure. Since Emil Fisher's "lock and key" model of 1894 (Fischer 1894), the structure-function paradigm is constantly evolving as we better understand how to measure and describe the time-dependence of protein conformations.

Although well-folded proteins have stable low free-energy three-dimensional structures, these molecules are actually very dynamic and can adopt different conformations to effectively perform their functions. The flexibility or motions of a protein can be thought of as the sampling of different conformations away from the lowest energy state. The amplitudes and timescales of these conformational transitions span from high frequency bond vibrations to slow global unfolding. Flexible proteins have lower energy barriers between different conformational states, whereas rigid proteins have higher barriers.

The structures of proteins fluctuate over a wide range of timescales (femtoseconds to hours) and with various amplitudes (pm to potentially μm), and these fluctuations often play important roles in their biological function. Conformational flexibility can be as subtle as a single bond rotation or vibration to re-adjust the side chain or backbone, or as dramatic as folding/unfolding of the protein. Enzymes must have a well-folded and ordered structure to allow the stabilization of transition states by their active sites, yet dynamic enough to permit alternative conformations required for subsequent catalytic

steps (Campbell et al. 2016). Slower and global motions in the millisecond or beyond modulate allostery and conformational transitions (Figure 1-9). The slowest motions on the timescale are characteristic of protein folding and complex formation. It is important to note that pictures of conformational ensembles do not completely represent protein dynamics because they usually provide information about the amplitude differences, but not rates of inter-conversion, between conformers.

1.6.1. Experimental and computational approaches to investigate protein dynamics

There are numerous experimental techniques that allow detailed investigations of protein conformational dynamics. X-ray crystallography is the most established and accurate method of determining the three-dimension structure of a protein. Although often giving the false impression that protein structures are static and rigid, dynamics can often be inferred from "snapshots" of the protein with differing conformations due to crystallization conditions, post-translational modifications, ligand binding, and so-forth. The flexibility of a protein is also reflected in the crystallographic Debye-Waller temperature factors or B-factors. Loosely speaking, B-factors represent the deviations of atomic electron densities around their equilibrium positions and thus are dependent upon protein dynamics. However, B-factors also reflect static disorder within the crystal lattice, as well as errors in model building. Crystallography is also restricted to proteins that adopt defined conformations, and thus has limited use for the study of IDRs and IDPs. Nevertheless, X-ray crystallography provides the "gold standards" for understanding the protein structure-function relationship.

Nuclear magnetic resonance (NMR) spectroscopy provides a complementary view of proteins and other biomolecules. In general, NMR-derived structures are restricted to small systems (< 25 kDa or so) and are usually less precise and accurate than those determined by X-ray crystallography. This is because NMR-derived structures rely heavily on dihedral angle and interproton distance restraints obtained from a variety of sources, including chemical shifts, scalar couplings, and the nuclear Overhauser effect (NOE), that

are often challenging to record and interpret. For example, relaxation properties of large proteins causes severe line boardening yielding to poor signal to noise. Peaks that overlap each other often happens and thus reduces the quality of the assignment. Importantly, protein NMR studies lead to conformational ensembles, rather than single structures, that are similarly consistent with all experimental restraints. Strictly speaking, structural variation within a conformational ensemble simply reflects the amount and quality of experimental data, and the computational protocols, used to determine that ensemble. However, dynamic processes impact these structural restraints and thus contribute to conformational deviations observed within these ensembles.

Beyond providing a route for determining protein structures, NMR spectroscopy is uniquely suited to studying dynamical processes because it extracts site-specific information for a large variety of motions that span many timescales (Figure 1-9). NMR experiments that determine amide spin-lattice (T_1), spin-spin (T_2) relaxation rate constants, as well as $^1H/^{15}N$ heteronuclear NOE values, are routinely used to analyze the fast dynamics (ps to ns) and global tumbling correlation times of proteins. These relaxation parameters can also be used in model-free analyses to express the amplitudes of motions as squared order parameters, S^2 (Lipari & Szabo 1982). S^2 values decrease from 1 to 0 as the fast timescale mobility of an amide ^{15}N-$^1H^N$ pair increases. Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments provide a tool to investigate slower motions (µs to ms) that normally involve larger conformational transitions with transient intermediates (Anthis & Clore 2015; Kleckner & Foster 2011). These experiments, which detect the additional line broadening of NMR signals due to conformational exchange, can be analyzed to extract the exchange rate constants and populations of interconverting conformers. For characterizing slower motions (ms to hour) that often represent local and global folding/unfolding of the protein, amide hydrogen exchange is a well-suited tool. The labile amide protons exchange with water protons (or deuterons) over time and this process is limited by their hydrogen bonding and solvent accessibility. Rate constants for $^1H/^1H$ or $^1H/^2H$ exchange can be determined using CLEANEX magnetization transfer or simple time-course ^{15}N-HSQC measurements, respectively. These rate constants can be compared to those predicted for a random coil

protein with the same sequence (corrected for pH, temperature, and isotope effects) and expressed as protection factors. These factors are often interpreted as the inverse of the equilibrium constants for conformational fluctuations that allow the detected amide hydrogen exchange.

Molecular dynamics simulations provide a theoretical link between structure and dynamics by enabling the exploration of the conformational energy landscape accessible to protein molecules (Figure 1-9) (Karplus & Kuriyan 2005). Due to optimized force fields and vastly improved software protocols for conformational sampling and hardware for computations, simulations in explicit solvents can model the amplitudes and timescales of protein motions in a realistic way, taking into account the fundamental role of water-mediated interactions. Indeed, it is now routine to run simulations of relative large proteins with the duration approaching the microsecond timescale.
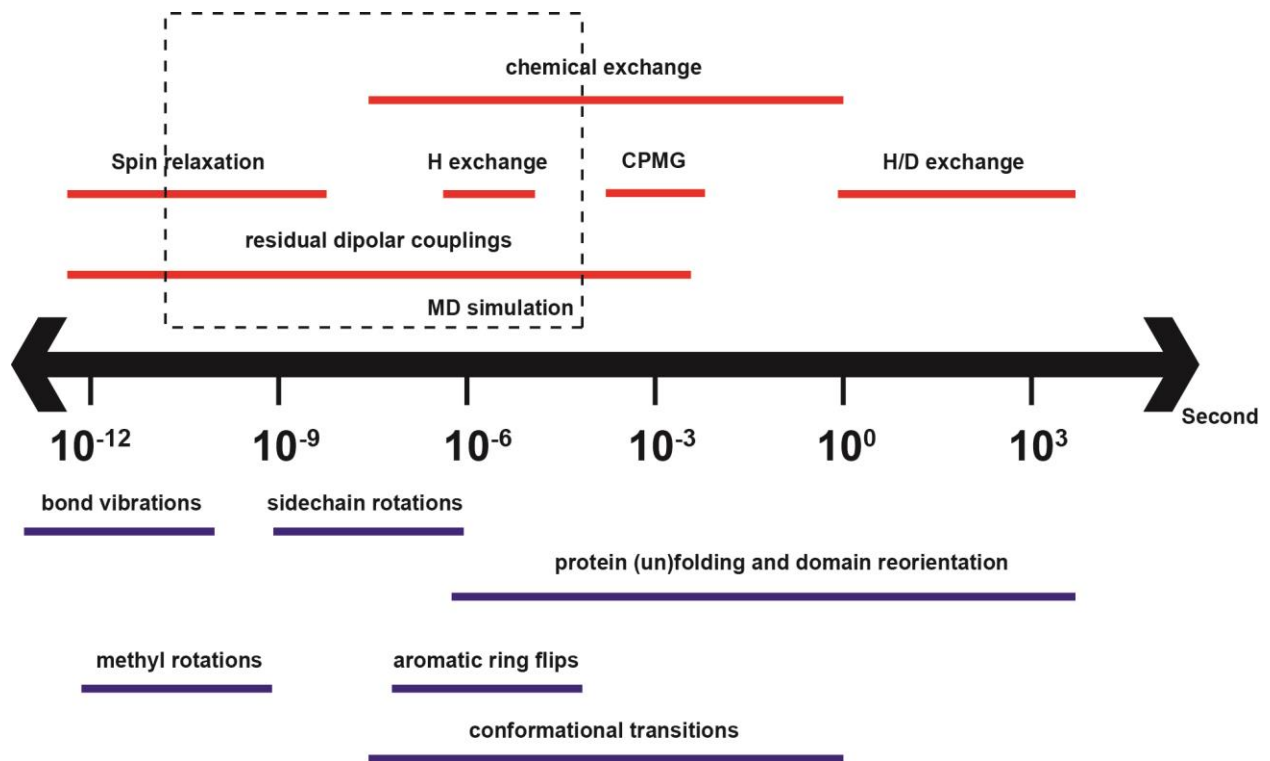
Figure 1-9 Timescales for protein dynamics and methods for their detection.

Proteins exhibits conformational dynamics (blue lines) ranging from atomic vibrational motions on the picosecond timescale (leftmost) to exchanging conformational substrate of rotameric side chains, to loop motions, to protein (un)folding at millisecond or even longer timescale (rightmost). NMR experimental techniques to probe structure and dynamics are listed on red lines. MD simulation highlighted in dashed box can aid in interpreting a collection of NMR data from pm-ms timescale (van den Bedem & Fraser 2015).

1.7. Goals and thesis overview

The ETS transcription factor family has 28 human paralogs exhibiting both common and diverse properties. However, these proteins share both conserved DNA-binding domains and consensus DNA sequences. This raises a conundrum for understanding the biological specificity of ETS transcription factors.

Phospho-dependent regulation of the PNT domain

Part of the answer lies within the PNT domain that is found within a subset of ETS factors. For example, Ras/MAPK signaling cascade can activate Ets1 that leads to phosphorylation-enhanced binding of the acetyltransferase co-activator CBP (Foulds et al. 2004). This is achieved through displacing the marginally stable helix H0 of Ets1 PNT domain away from its core bundle and shifting the equilibrium from a closed to an open state upon phosphorylation of the phosphoacceptors Thr38/Ser41 (Figure 1-7).

The first goal of my thesis research was to test if this phosphor-switch mechanism is shared by Pointed-P2, the *Drosophila* ortholog of human Ets1 and Ets2. That is, what are the similarities and differences between the PNT domain and the phosphoacceptor regions of these two proteins? As described in chapter 2, I used NMR spectroscopy and other biochemical methods to monitor the changes in structure and dynamics of Pointed-P2 PNT domain upon its phosphorylation. Based on main chain chemical shifts, the Pointed-P2 PNT domain also contains additional helices H0/H1 appended to a core SAM-like helical bundle. The dynamic properties of the PNT domain were investigated using a series of NMR relaxation and amide HX experiments. Similar to Ets1, I found that the appended helices H0/H1 of the Pointed-P2 PNT domain are dynamic and only marginally stable. In addition, the conserved phosphoacceptor Thr151 that lies within a MAP kinase consensus sequence (Pro-x-Ser/Thr-Pro), I identified two additional phosphorylation sites on Pointed-P2 (Thr145 and Thr 154) using complementary NMR spectroscopic methods. These sites are located in the disordered region immediately preceding helix H0 and their phosphorylation only had minor effects on the secondary structure and dynamics of the

PNT domain. NMR-monitored titrations also revealed that the phosphoacceptors and helix H0, as well as the region of the core helical bundle identified previously by mutational analyses as a kinase docking site, are selectively perturbed upon binding of the ERK2 MAP kinase Rolled by Pointed-P2. Based on a homology model derived from the ETS1 PNT domain, helix H0 is predicted to partially occlude the docking interface. Therefore, this dynamic helix must be displaced to allow both docking of the kinase, as well as binding of Mae, a *Drosophila* protein that negatively regulates Pointed-P2 by competing with the kinase for its docking site. Finally, I examined whether phosphorylation of Pointed-P2 enhances its binding to TAZ1 domain of Neijre, the *Drosophila* ortholog of CBP. Unfortunately, soluble versions of Neijre could not be obtained for these studies. Regardless, this work confirmed that the PNT domains and phosphoacceptor sites of Ets1 and Pointed-P2 share many common structural, dynamic, and functional features.

Autoinhibition of the ETS domain

A second route to impart specificity and regulate the transcriptional activity of ETS factors lies with autoinhibition of the ETS domain. As demonstrated by Ets1 and ETV6, intrinsically disordered sequences as well as helices, appended on their ETS domains autoinhibit DNA-binding via steric and allosteric mechanisms (Figure 1-6). The inhibitory helices of Ets1 and ETV6 are only marginally stable and are poised to unfold upon DNA binding. Furthermore, phosphorylation-enhanced autoinhibition results from a fuzzy complex between an intrinsically disordered SRR sequence and the Ets1 ETS domain.

In chapter 3 of my thesis, I expanded our understanding of the autoinhibition repertoire of ETS transcription factors by investigating the DNA-binding properties of ETV1/4/5 sub-family, with a particular focus on ETV4. In 2001, Greenall and colleagues reported that DNA-binding of ETV4 is autoinhibited by sequences flanking their ETS domain, and coarsely defined that the autoinhibitory sequences are unlikely to form α-helices or β-strands by proline-scanning mutagenesis. Therefore, we hypothesize that ETV1/4/5 utilize a distinct mechanism of autoinhibition through structural dynamics and fuzzy interaction between the ETS domain and its flanking sequences. In collaboration with Dr.

Graves at the University of Utah, we used various biochemical methods along with X-ray crystallography and NMR spectroscopy to define and characterize autoinhibited fragments of ETV1/4/5. Together, we have found that the inhibitory domains of ETV4 reside both N- and C-terminal of the ETS domain and cooperate to inhibit DNA binding. The C-terminal inhibitory domain (CID) is a dynamic α-helix that packs against the ETS domain and perturbs the relative positioning of the DNA recognition helix H3. The N-terminal inhibitory domain (NID) is an intrinsically disordered region that transiently interacts with the ETS domain and the CID. In addition to sterically blocking DNA-binding, the NID may also reinforce the inhibitory effects of the CID. Our preliminary results also indicate that lysine acetylation of the NID increases the transcriptional activity of ETV4 by counteracting autoinhibition.

As discussed throughout this introductory chapter, the dynamic properties of ETS transcription factors play key roles in regulating their biological activities. The inhibitory modules of Ets1, ETV6 and ERG dampen motions of their ETS domain and stabilize the ETS domain detectable by amide HX and ^{15}N relaxation dispersion (Pufall et al. 2005; Lee et al. 2008; Coyne et al. 2012; Regan et al. 2013). As a result, an overarching hypothesis of my research is that ETS factors exist in conformational equilibria between flexible active and rigid inactive states. Accordingly, in chapter 4 of my thesis, I investigated the dynamic properties of the ETS domain from several ETS factors to understand their common and distinct features. The ETS domains of Ets1, ETV6, ERG, ETV4 and PU.1 were studied because they displayed a wide range of autoinhibitory mechanisms. More importantly, despite sharing a common fold, these ETS domains exhibit a ~ 100-fold difference in affinity for binding a common consensus DNA motif with a $^{5'}$GGAA$^{3'}$ core (Currie et al. 2017). Therefore, I hypothesize that there are key structural or dynamic differences that influence their DNA binding properties. Circular dichroism spectroscopy, MD simulations, and a combination of NMR methods were thus used to probe these differences. As a pre-requisite step, I determined the structure of the isolated PU.1 ETS domain via NMR methods and identified an appended dynamic helix H4. Although the functional role of this helix remains to be determined, it appears to be a general feature of most, if not all, ETS domains. Amide HX measurements revealed that

the DNA-recognition helix H3 of each of these ETS domains is also dynamic. The conservation of this dynamic feature suggests that conformational flexibility is needed for processes such as "scanning non-specific" genomic DNA and opting high affinity complexes with specific target DNA sites. MD simulations also were employed to obtain a unified description of the motions of these ETS domains. It led to an observation that the dynamics of the "turn" between helix H2 and H3, as a part of the DNA-binding interface, is dependent on the length, the glycine/proline content, and electrostatic characteristics. In addition, several critical pathways that relay "information" between the DNA recognition helix H3 to other parts of the ETS domain were identified. These may provide avenues of allosteric control for ETS domains. For example, this may help to explain the unfolding of the inhibitory helix HI-1 of Ets1 as an allosteric response to DNA binding by the distal recognition helix H3.

In summary, the studies presented in my thesis help expand our knowledge of the regulation of ETS transcription factors. This is achieved by detailed comparison of the PNT domain between Pointed-P2 and Ets1 upon phosphorylation. Building upon the theme of ETS DNA-binding autoinhibition, I demonstrated that ETV1/4/5 are also autoinhibited by a N-terminal intrinsically disordered sequence and a flexible C-terminal helix. These studies highlight the important relationships between post-translational modifications, dynamics, and function of the ETS factors.

Chapter 2. Identification of phosphoacceptor sites on *Drosophila* Pointed-P2 PNT domain

Chapter 2 is a modified version of the article: Lau, D.K.W., Okon, M., and McIntosh, L.P. (2012) *"The PNT domain from Drosophila Pointed-P2 contains a dynamic helix preceded by a disordered phosphoacceptor sequence"* **Protein Sci.** 21: 1716-1725. The modifications include reformatting and the incorporation of supplementary material into the main text. I performed all experimental work, with assistance from Mark Okon for NMR data collection. Lawrence McIntosh and I contributed to data analysis and manuscript preparation.

Overview

Pointed-P2, the *Drosophila* ortholog of human ETS1 and ETS2, is a transcription factor involved in Ras/MAP kinase regulated gene expression. In addition to a DNA-binding ETS domain, Pointed-P2 contains a PNT (or SAM) domain that serves as a docking module to enhance phosphorylation of an adjacent phosphoacceptor threonine by the ERK2 MAP kinase Rolled. Using NMR chemical shift, relaxation, and amide hydrogen exchange measurements, we demonstrate that the Pointed-P2 PNT domain contains a dynamic N-terminal helix H0 appended to a core conserved five-helix bundle diagnostic of the SAM domain fold. Neither the structure nor dynamics of the PNT domain are perturbed significantly upon *in vitro* ERK2 phosphorylation of three threonine residues in a disordered sequence immediately preceding this domain. These data thus confirm that the *Drosophila* Pointed-P2 PNT domain and phosphoacceptors are highly similar to those of the well characterized human ETS1 transcription factor. NMR-monitored titrations also revealed that the phosphoacceptors and helix H0, as well as a region of the core helical bundle identified previously by mutational analyses as a kinase docking site, are selectively perturbed upon ERK2 binding by Pointed-P2. Based on a homology model derived from the ETS1 PNT domain, helix H0 is predicted to partially occlude the docking

interface. Therefore, this dynamic helix must be displaced to allow both docking of the kinase, as well as binding of Mae, a *Drosophila* protein that negatively regulates Pointed-P2 by competing with the kinase for its docking site.

2.1.   Introduction

Gene expression and signal transduction require tightly controlled macromolecular interactions and post-translational modifications involving both modular domains and intrinsically disordered regions of regulatory proteins. This paradigm is well exemplified by the ETS transcription factors. In addition to a conserved DNA-binding ETS domain, ~1/3 of all ETS factors also contain a PNT domain (Hollenhorst, McIntosh, et al. 2011), which is an ETS-specific member of the widespread family of SAM domains. Although sharing a common core architecture of four α-helices and a fifth small α - or 3_{10}-helix, SAM domains exhibit remarkably diverse association states and function in a wide variety of protein-protein and protein-RNA interactions (Qiao & Bowie 2005). Defining the molecular basis for this diversity remains an important challenge. In the case of the ETS factors, one source of specificity is provided by additional helices appended to the core SAM domain. The PNT domains of Yan, ETV6 (or Tel), ERG, ELF3, and FLI1 contain only the minimal helical bundle, whereas those of GABPα and SPDEF have one additional N-terminal helix (Mackereth et al. 2004) and those of ETS1 and ETS2 have two (Nelson et al. 2010). Furthermore, due to differing surface features (Meruelo & Bowie 2009), the PNT domains of Yan and ETV6 self-associate (Kim et al. 2001; Qiao et al. 2004), whereas the remainder are monomeric in isolation. Also, in disordered regions immediately preceding the PNT domains of closely related mammalian ETS1/ETS2 and *Drosophila melanogastar* Pointed-P2 are phosphoacceptors for the orthologous Ras-activated MAP kinases ERK2 and Rolled, respectively (Klämbt 1993; Nelson et al. 2010; Wasylyk et al. 1997).

The role of the ETS1 PNT domain in Ras/MAP kinase signaling has been investigated extensively using combination of cell-based and biophysical measurements. The

monomeric PNT domain is a docking module for ERK2, enhancing the efficiency of phosphorylating Thr38 and Ser41 in a flexible region preceding this structured domain (Rainey et al. 2005; Seidel & Graves 2002). The docking interfaces on the PNT domain and the kinase have been mapped coarsely through mutagenesis, NMR spectroscopic, chemical footprinting, and competition studies (Abramczyk et al. 2007; Callaway et al. 2010; Callaway et al. 2006; Piserchio et al. 2011; Seidel & Graves 2002). In order to both accommodate the proposed docking mechanism and position the phosphoacceptors in the catalytic site of the kinase, a significant conformational change in the ETS1 PNT domain, such as the unfolding of the appended N-terminal helix H0 (residues Lys42-Thr52), appears to be required (S. Lee et al. 2011). Indeed, NMR relaxation and amide hydrogen exchange (HX) measurement have shown this helix is only marginally stable and structurally flexible (Nelson et al. 2010). Furthermore, upon phosphorylation, this helix H0 remains folded, but adopts a broad distribution of conformations displaced from the core helical bundle (H2-H5). This in turn contributes to enhanced electrostatically-driven interactions with the TAZ1 domain of the general transcriptional co-activator acetyltransferase CBP, and ultimately, increased expression of Ras-responsive ETS1 target genes (Foulds et al. 2004; Nelson et al. 2010).

Pointed-P2, the *Drosophila melanogaster* ortholog of ETS1, plays a similar role in the well characterized EGF and Sevenless receptor tyrosine kinase/Ras-mediated signal transduction pathway to control fly eye development (Brunner et al. 1994; O'Neill et al. 1994; Tootle & Rebay 2005; Vivekanand & Rebay 2006). Monomeric Pointed-P2 is a transcriptional activator and antagonist to Yan, an ETS family transcriptional repressor that polymerizes via head-to-tail self-association of its PNT/SAM domain (Qiao et al. 2004). Both control the expression of a common set of target genes, including a crucial regulator called Mae (Vivekanand et al. 2004). Upon stimulation of the receptor tyrosine kinase signaling cascade, the MAP kinase Rolled is activated and enters the nucleus to phosphorylate Yan, thereby leading to its CRM1-mediated export and subsequent cytoplasmic degradation (Song et al. 2005; Tootle et al. 2003). This is facilitated by the SAM domain of Mae (Baker et al. 2001), which acts as a tight-binding heterotypic partner of the Yan PNT/SAM domain to favor its depolymerization and to expose a critical

phosphoacceptor site (Qiao et al. 2004). In parallel, Rolled phosphorylation of Pointed-P2 leads to transcriptional activation of the genes previously repressed by Yan. The PNT domain of Pointed-P2 is a docking module for the kinase to enhance phosphorylation of the adjacent phosphoacceptor Thr151 (Qiao et al. 2006). However, as part of a negative feedback mechanism, Mae also heterodimerizes with the PNT domain of Pointed-P2 to block kinase docking and hence down regulating phosphorylation-enhanced gene expression (Qiao et al. 2006).

In this chapter, we have used NMR spectroscopy to investigate further the similarities between the PNT domains and phosphoacceptor regions of ETS1 and Pointed-P2 (Figure 2-1A). Based on mainchain chemical shifts, the PNT domain of Pointed-P2 also contains additional helices H0/H1 appended to the core SAM helical bundle. NMR relaxation and amide HX studies confirm that these appended helices are dynamic and only marginally stable. Furthermore, ERK2 phosphorylates *in vitro* a Pointed-P2 fragment at three acceptor sites (Thr145, Thr151, and Thr154) in the unstructured region N-terminal to the PNT domain, and these post-translational modifications have only modest effects on the secondary structure and dynamics of the PNT domain. NMR-monitored titrations of Pointed-P2 with ERK2 reveal that both the phosphoacceptor region and the PNT domain interact with the kinase. Based on these similarities, it is likely that the ETS1 and Pointed-P2 PNT domains also share similar mechanisms of MAP kinase docking and recruitment of transcriptional co-activators.

2.2.   Results

2.2.1. Pointed-P2 PNT domain contains a dynamic helix H0

The well-dispersed ^{15}N-HSQC spectrum of PntP2$^{142-252}$ confirms that the monomeric PNT domain-containing fragment of Pointed-P2 adopts an independently folded structure (Figure 2-1B). With the exception of Glu142, Val143, Phe166 and the N-terminal Gly-Ser-His-Met tag, almost complete ^{1}H^{N}, ^{15}N, ^{13}C$^{\alpha}$, ^{13}C$^{\beta}$, and ^{13}C^{o} resonance assignments were

obtained for this construct. These chemical shifts were used to identify the secondary structural elements of PntP2$^{142-252}$ with the MICS (Motif Identification by Chemical Shift) algorithm (Shen & Bax 2012). Based on this analysis, PntP2$^{142-252}$ contains 6 helical regions that closely match those identified in Ets1$^{29-138}$ (Figure 2-3A,B), as well as Ets2$^{69-172}$ (not shown (Nelson et al. 2010)). This strongly suggests that the PNT domains of Pointed-P2, ETS1, and ETS2 also share a common tertiary structure of a SAM core (helices H2-H5) with appended N-terminal helices H0/H1. In the NMR-derived structural ensemble of Ets1$^{29-138}$, these two appended helices are essentially continuous, yet bend at Phe53 to allow extended packing against the core helices H2 and H5. The corresponding residue in PntP2$^{142-252}$ is Phe166.
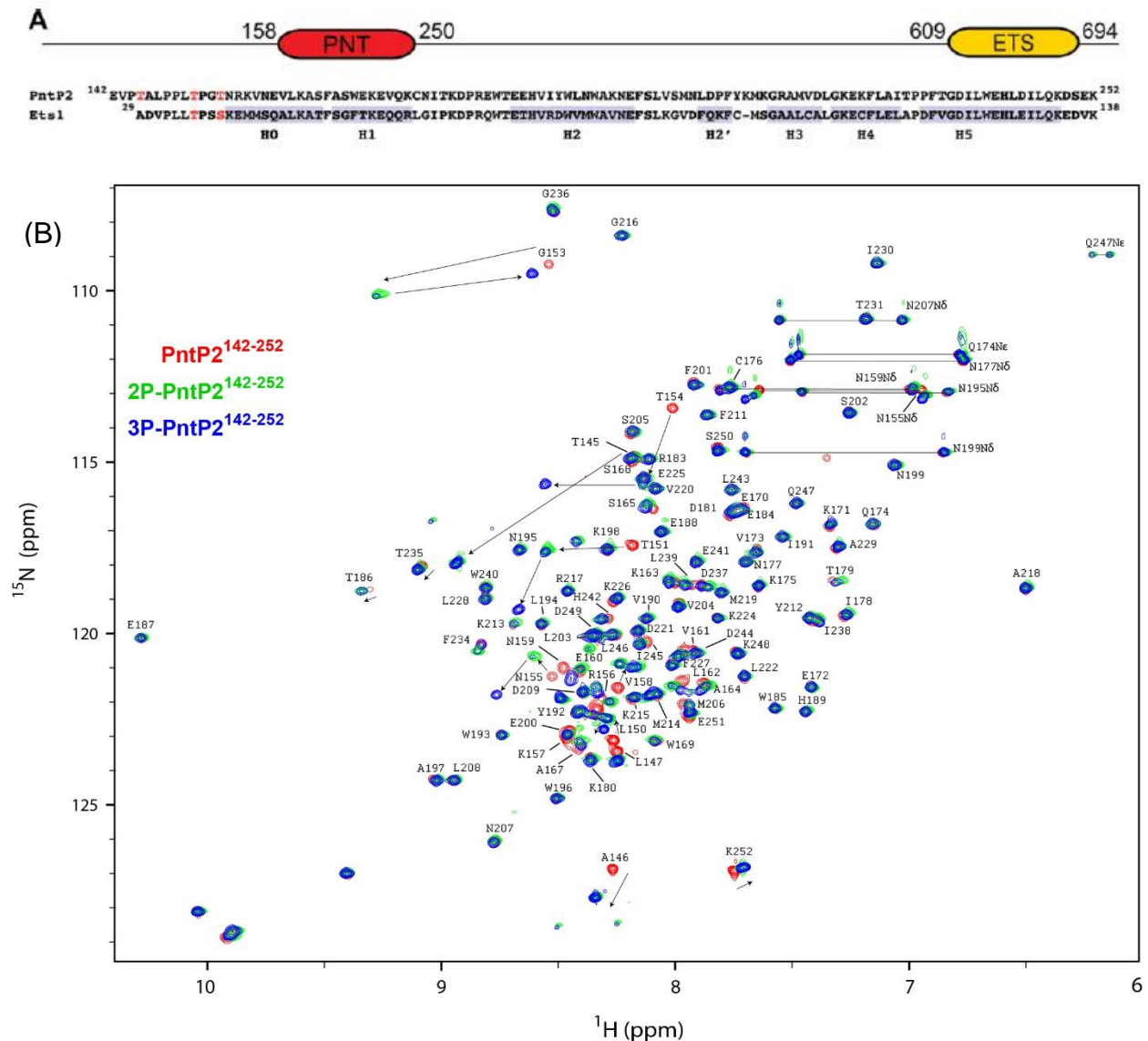
Figure 2-1 NMR spectroscopic characterization of PntP2$^{142–252}$

(A) Cartoon of the 718-residue Pointed-P2 transcription factor showing the structured PNT domain and DNA-binding ETS domain, along with aligned sequences of PntP2$^{142–252}$ and Ets1$^{29–138}$. The identified ERK2 phosphoacceptors are in red, and the observed consensus α-helices (or 3$_{10}$-helix for H2') in the NMR-derived structural ensembles of Ets1$^{29–138}$ (2jv3.pdb) and 2P-Ets1$^{29–138}$ (2kmd.pdb) are shaded in light blue. (B) Superimposed ^{15}N-HSQC spectra of PntP2$^{142–252}$ (red), 2P-PntP2$^{142–252}$ (green), and 3P-PntP2$^{142–252}$ (blue). The arrows indicate the progressive spectral perturbations upon increased phosphorylation.
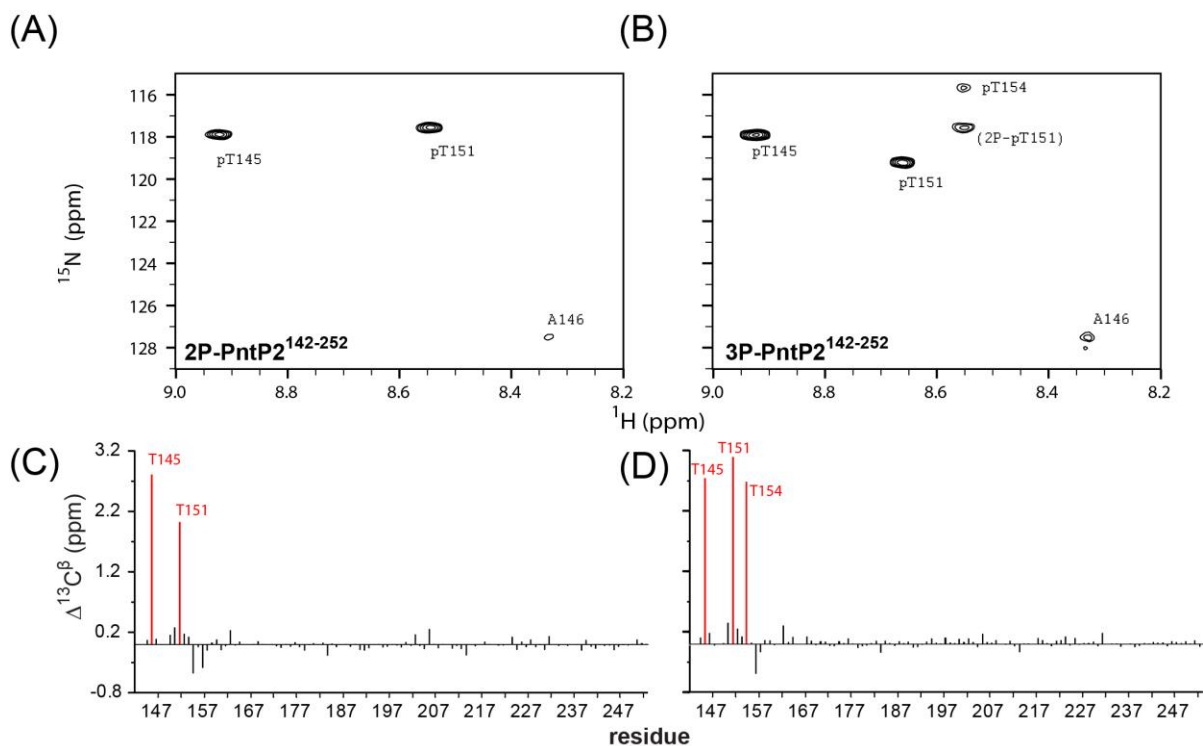
Figure 2-2 Direct detection of phosphothreonines of PntP2$^{142–252}$ using NMR spectroscopy

Amide signals from phosphothreonines are observed selectively in the ^{1}H–^{15}N faces of ^{31}P-edited HNCA spectra of (A) 2P-PntP2$^{142–252}$ and (B) 3P-PntP2$^{142–252}$. Although the latter experiment also detects the residue immediately following a phosphothreonine (McIntosh et al. 2009), only Ala146 is observed as Pro152 lacks an amide ^{1}H^N and Asn155 has a relatively weak ^{15}N-HSQC signal. Owing to incomplete separation by anion exchange chromatography, resolved signals from contaminating 2P-PntP2$^{142–252}$ were observed in the spectra of 3P-PntP2$^{142–252}$ shown in (Figure 2-1B and Figure 2-2B). The phosphothreonines in (C) 2P-PntP2$^{142–252}$ and (D) 3P-PntP2$^{142–252}$ were also identified based on a large diagnostic downfield ^{13}C^β shift (Bienkiewicz & Lumb 1999) relative to the corresponding unmodified amino acid in PntP2$^{142–252}$.

45
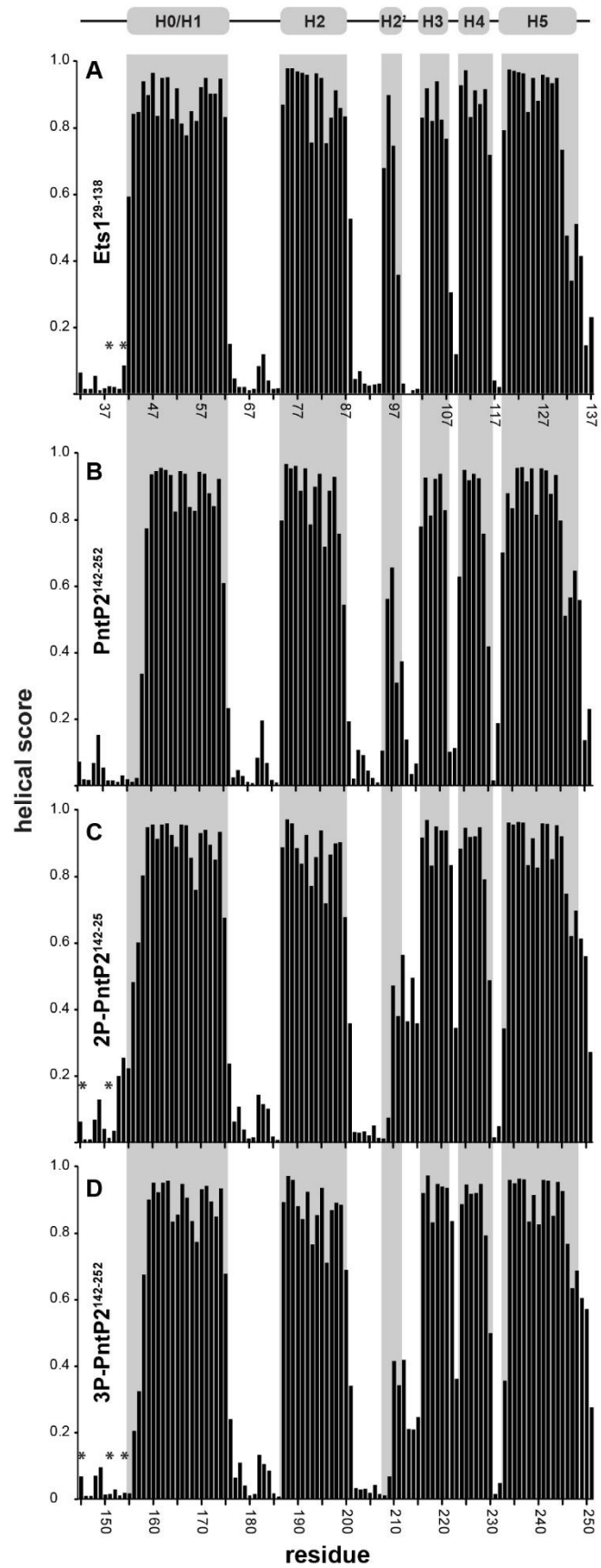
Figure 2-3 The PNT domains of PntP2 and Ets1 share similar helical secondary structures.

Shown are the predicted helical scores for (A) 2P-Ets1$^{29–138}$ (data from (Nelson et al. 2010)), (B) PntP2$^{142–252}$, (C) 2P-PntP2$^{142–252}$, and (D) 3P-PntP2$^{142–252}$, based on an analysis of $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and/or $^{13}C^{'}$ chemical shifts using the program MICS (Shen & Bax 2012). Consideration of phosphorylation-dependent chemical shift changes (Bienkiewicz & Lumb 1999) does not significantly alter the scores for the phosphoacceptor serines/threonines (asterisks). The top cartoon and the gray rectangles indicate the observed consensus α-helices (or 3$_{10}$-helix for H2') in the NMR-derived structural ensembles of Ets1$^{29–138}$ (2JV3.pdb) and 2P-Ets1$^{29–138}$ (2KMD.pdb) (Nelson et al. 2010).

The dynamic properties of PntP2$^{142-252}$ were examined using ^{15}N relaxation and amide HX measurements. As shown in Figure 2-4A, the heteronuclear ^{15}N-NOE values of residues in helix H0 progressively decrease toward its N-terminus. A model-free analysis of the ^{15}N T_1, T_2, and NOE values of PntP2$^{142-252}$ shows a similar trend with lower S^2 order parameters for amides near the start of this helix, as well as at the C-terminus of helix H5 (Figure 2-5). These data are indicative of enhanced backbone mobility on a ns–ps timescale relative to the well-ordered SAM domain core. However, it is difficult to estimate the nature of the conformations sampled by these fast motions from ^{15}N relaxation measurements alone (Kleckner & Foster 2011). More importantly, rapid amide HX was detected for residues throughout helices H0/H1 (Figure 2-6A). The measured HX rate constants for these residues are comparable to those predicted for an unstructured polypeptide under similar conditions (Bai et al. 1993). The only other regions of the PntP2$^{142-252}$ showing such behavior are the disordered termini and exposed loops, as the remaining amides in the structured PNT domain exchange too slowly to be detected by the ΔEX magnetization transfer approach. This clearly demonstrates that helices H0/H1 are only marginally stable and must undergo substantial local unfolding to allow facile exchange with water. Very similar ^{15}N relaxation and HX behavior was observed for helix H0 in Ets1$^{29-138}$, indicating that the PNT domains of these two ETS family members also exhibit common dynamic properties (Nelson et al. 2010).
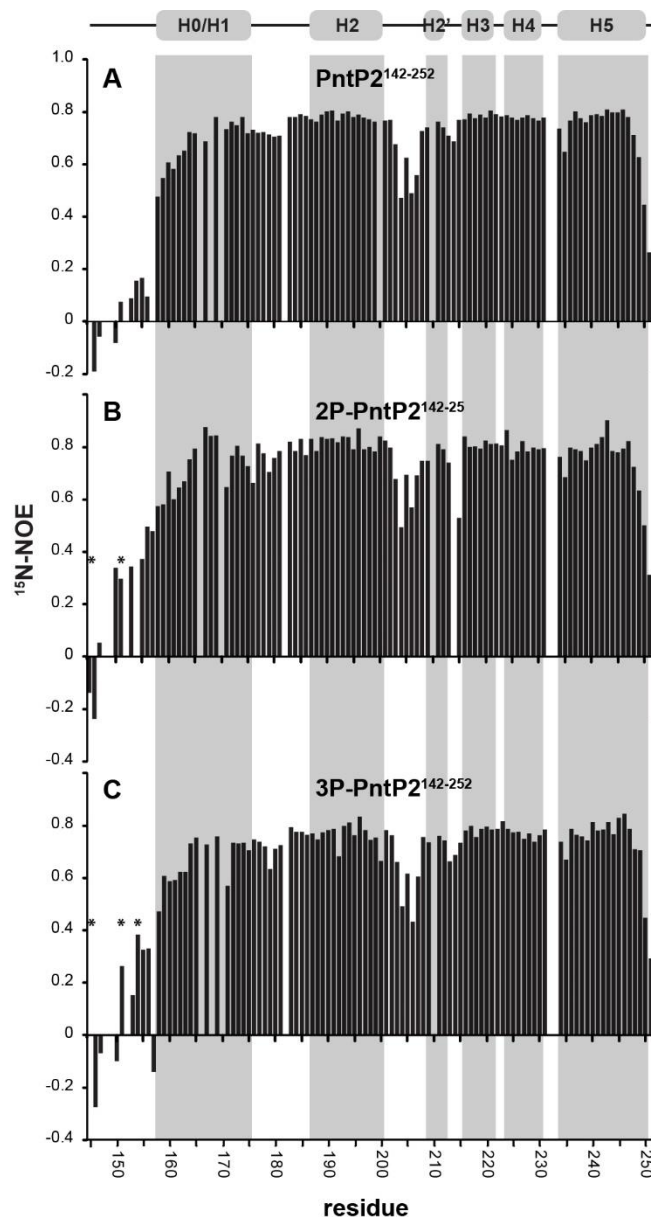
Figure 2-4 PntP2$^{142-252}$ phosphoacceptors and helix H0 are flexible.

Steady-state heteronuclear ^{15}N-NOE values for (A) PntP2$^{142-252}$, (B) 2P-PntP2$^{142-252}$, and (C) 3P-PntP2$^{142-252}$. Well-ordered residues have NOE values of ~ 0.8, whereas decreasing values indicate increasing mobility of the ^{15}N–^{1}H^{N} bond vector on the ns–ps timescale (Farrow et al. 1994). The top cartoon and the gray rectangles indicate the helices, based on the consensus MICS scores, for the three PntP2$^{142-252}$ species [H0/H1, 158–175; H2, 187–200; H2' , 209–212; H3, 216–221; H4, 224–230; H5, 234–250; Figure 2-3(B–D)], and the phosphothreonines are identified with asterisks. Missing data correspond to prolines or residues with overlapping signals. The error is ~5%.
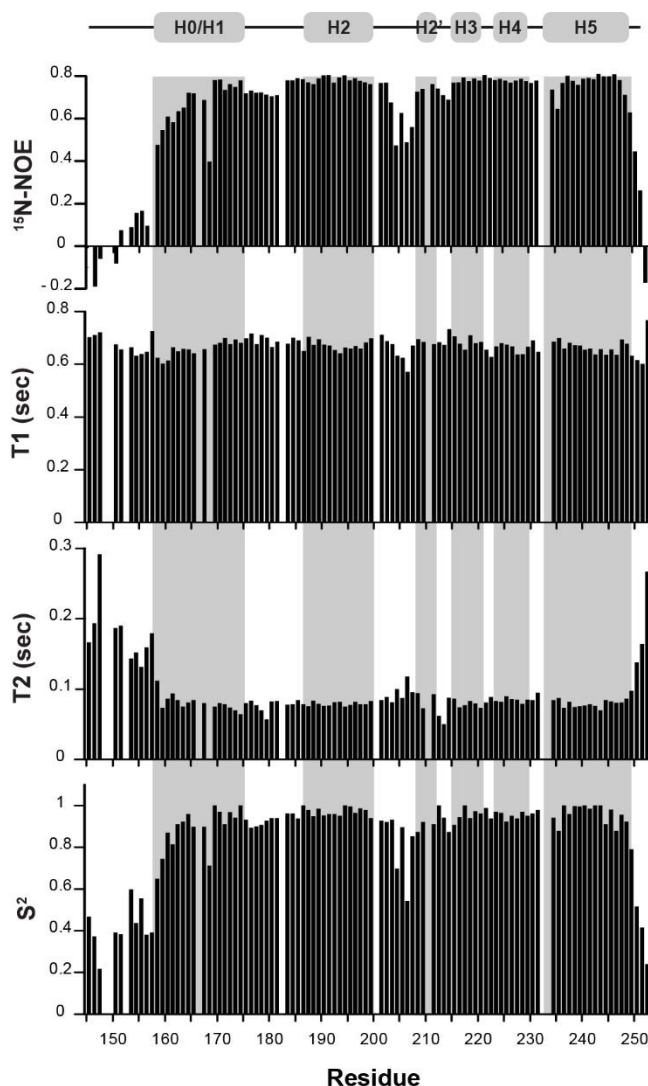
Figure 2-5 Amide ^{15}N T$_1$, T$_2$, and steady-state heteronuclear ^{15}N-NOE values for PntP2$^{142\text{-}252}$, recorded at 25 °C with a 600 MHz NMR spectrometer.

Model-free analysis with TENSOR2 yielded an isotropic correlation time for global tumbling of 8.7 ± 0.1 ns and the illustrated anisotropic generalized S^2 order parameters. The latter were fit with the homology model of PntP2$^{142\text{-}252}$ generated using Ets1$^{29\text{-}138}$ (2JV3.pdb) as a template. However, isotropic order parameters are similar (not shown). The cartoon and grey rectangles indicate the helices based on the consensus MICS scores for the three PntP2$^{142\text{-}252}$ species (Figure 2-4 B-D), and missing data correspond to prolines or residues with overlapping signals. Note that a reduced ^{15}N-NOE value and increased T$_2$ lifetime reflects increased mobility of the amide ^{15}N- ^{1}H^N bond vector on the ns-ps timescale, and hence a lower order parameter.

50
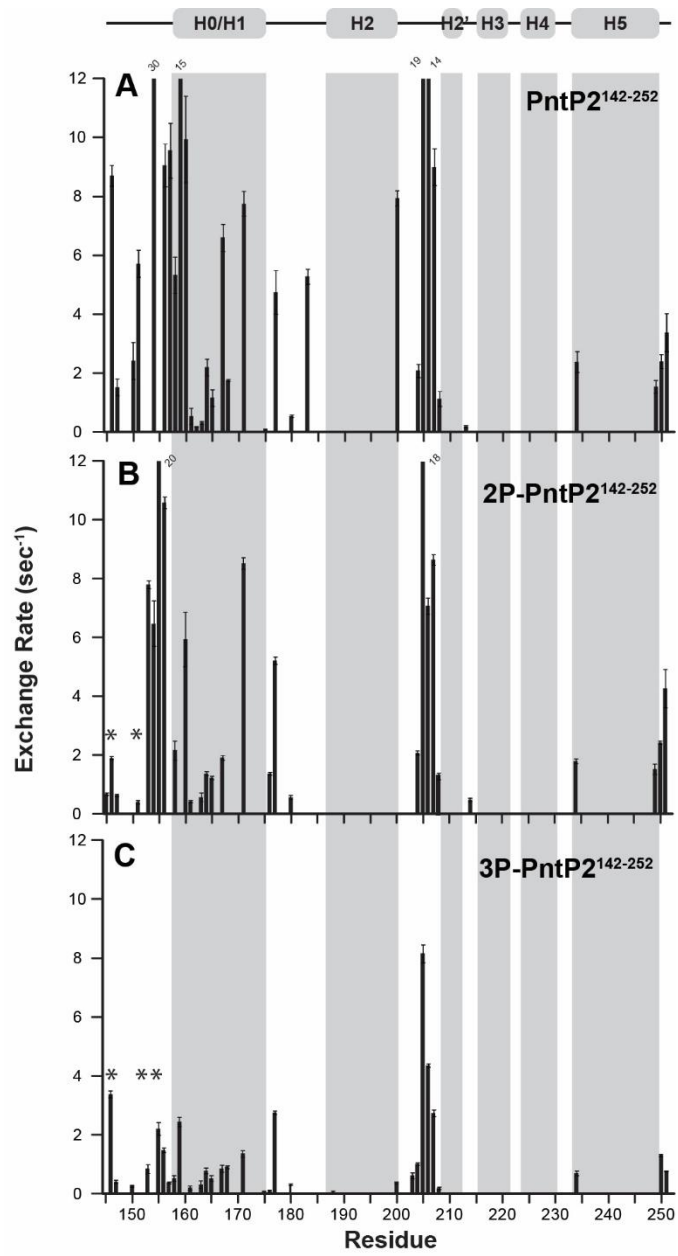
Figure 2-6 Helices H0/H1 are marginally stable with little protection from HX.

Amide HX rate constants were determined from CLEANEX experiments recorded at 25°C for (A) PntP2$^{142-252}$ (pH 6.7 and pH 7.5), (B) 2P-PntP2$^{142-252}$ (pH 6.7), and (C) 3P-PntP2$^{142-252}$ (pH 7.1 and 7.5), and normalized to pH 6.7, assuming a first-order dependence on [OH^{-}]. For better comparison, the bars for several data points were truncated and the rate constants indicated by the numbers. Phosphothreonines are identified with asterisks. Missing data points correspond to prolines, amides with overlapping ^{15}N-HSQC signals, and amides with HX rates too slow to be measured by the CLEANEX approach under the sample pH conditions examined (i.e., $k_{ex} < 0.5$ s^{-1}). Most amides fall in the latter category owing to their presence in stable, hydrogen-bonded structural elements of the protein. However, in the case of 3P-PntP2$^{142-252}$, the three phosphothreonines likely exchange rapidly, but were not included in (C) owing to ambiguous spectral assignments at elevated sample pH values.

2.2.2. Pointed-P2 PNT domain phosphoacceptors are disordered

The effect of phosphorylation on the properties of Pointed-P2 was examined using active ERK2 to phosphorylate PntP2$^{142-252}$ *in vitro*. Treatment with the kinase yielded products with two or three phosphorylated residues. The sites of modification were identified unambiguously via two NMR spectroscopic methods. The first method exploits a weak 2-bond ^{31}P-^{13}C$^{\alpha}$ scalar coupling to selectively detect the amide ^{1}H^{N}-^{15}N signals from a phosphorylated serine/threonine (i) and its (i+1) neighbor via a ^{31}P-edited HNCA spectrum (Figure 2-2B,C) (McIntosh et al. 2009). The second relies on the observation that the ^{13}C$^{\alpha}$ signal of a random coil serine/threonine shifts downfield by ~ 2 - 4 ppm upon phosphorylation (Figure 2-2D,E) (Bienkiewicz & Lumb 1999). Based on these complementary approaches, 2P-PntP2$^{142-252}$ is clearly phosphorylated at both Thr145 and Thr151, whereas 3P-PntP2$^{142-252}$ contains an additional modification at Thr154. Of these phosphoacceptors, only Thr151 is within a MAP kinase consensus sequence (Pro-x-Ser/Thr-Pro) (Gonzalez et al. 1991; Songyang et al. 1996).

The phosphoacceptor threonines of PntP2$^{142-252}$ are within the disordered region N-terminal to the helical PNT domain. Similar to Ets1$^{29-138}$ (Nelson et al. 2010), the conformational flexibility of these residues is evident from both random coil chemical shifts (Figure 2-3B), low heteronuclear ^{15}N-NOE values (Figure 2-4B) and rapid HX (Figure 2-6A). Furthermore, phosphorylation does not induce any predominant secondary structure for this region of the Pointed-P2 fragment (Figure 2-3C,D) and only slightly dampens fast ns-ps timescale motions of pThr151 and pThr154 detectable via ^{15}N-NOE measurements (Figure 2-4B,C).

Phosphorylation of PntP2$^{142-252}$ has also no pronounced effect on the structure or dynamics of the PNT domain. A comparison of ^{15}N-HSQC spectra reveals that amide chemical shift perturbations owing to phosphorylation are localized to residues near the phosphoacceptors (Figure 2-1, Figure 2-2, Figure 2-7). Thus, the structure of the PNT domain is not altered upon modification of the threonines. A small increase in the MICS-helical scores of residues 155-157 is noted in 2P-PntP2$^{142-252}$, presumably due to

phosphorylation of Thr151, yet these values decrease in 3P-PntP2$^{142-252}$ with the subsequent modification of Thr154 (Figure 2-3). The increase of helical score of 2P-PntP2$^{142-252}$ is likely due to the stabilization of helix H0 by the negatively charged phosphate (helix capping), while 3P-PntP2$^{142-252}$ is less stable could be because of charge repulsion.

Within experimental error, the heteronuclear ^{15}N-NOE values of the unmodified and modified forms of PntP2$^{142–252}$ are similar, indicating that phosphorylation does not dampen any fast timescale motions of amides in helix H0 (Figure 2-4). Importantly, CLEANEX measurements also show that several residues in helix H0/H1 still undergo rapid HX in 2P-PntP2$^{142–252}$ and 3P-PntP2$^{142–252}$, albeit at a slightly reduced rate relative to the unmodified protein (Figure 2-6). The average ~ 2.5-fold reduction in HX rate constants for corresponding residues in helix H0/H1 of 3P-PntP2$^{142–252}$ relative to the unmodified protein suggests that, in particular, pThr154 might marginally stabilize these helices by acting as an N-terminal cap (Andrew et al. 2002). However, chemical shift analyses by the MICS algorithm do not detect such a predominant role for either pThr151 or pThr154 (Figure 2-3). Furthermore, the modest changes (particularly when viewed on a free energy scale) could be owing to electrostatic or inductive effects of a phosphothreonine on the intrinsic exchange rates of its neighboring residues (Bai et al. 1993), or simply to subtle differences in experimental conditions as other amides in loop regions and at the C-terminus of the 3P-PntP2$^{142–252}$ also showed a comparable reduction in measured HX rate constants. Regardless, these experiments indicate that the structure and dynamics of PntP2$^{142–252}$ are perturbed minimally by phosphorylation.
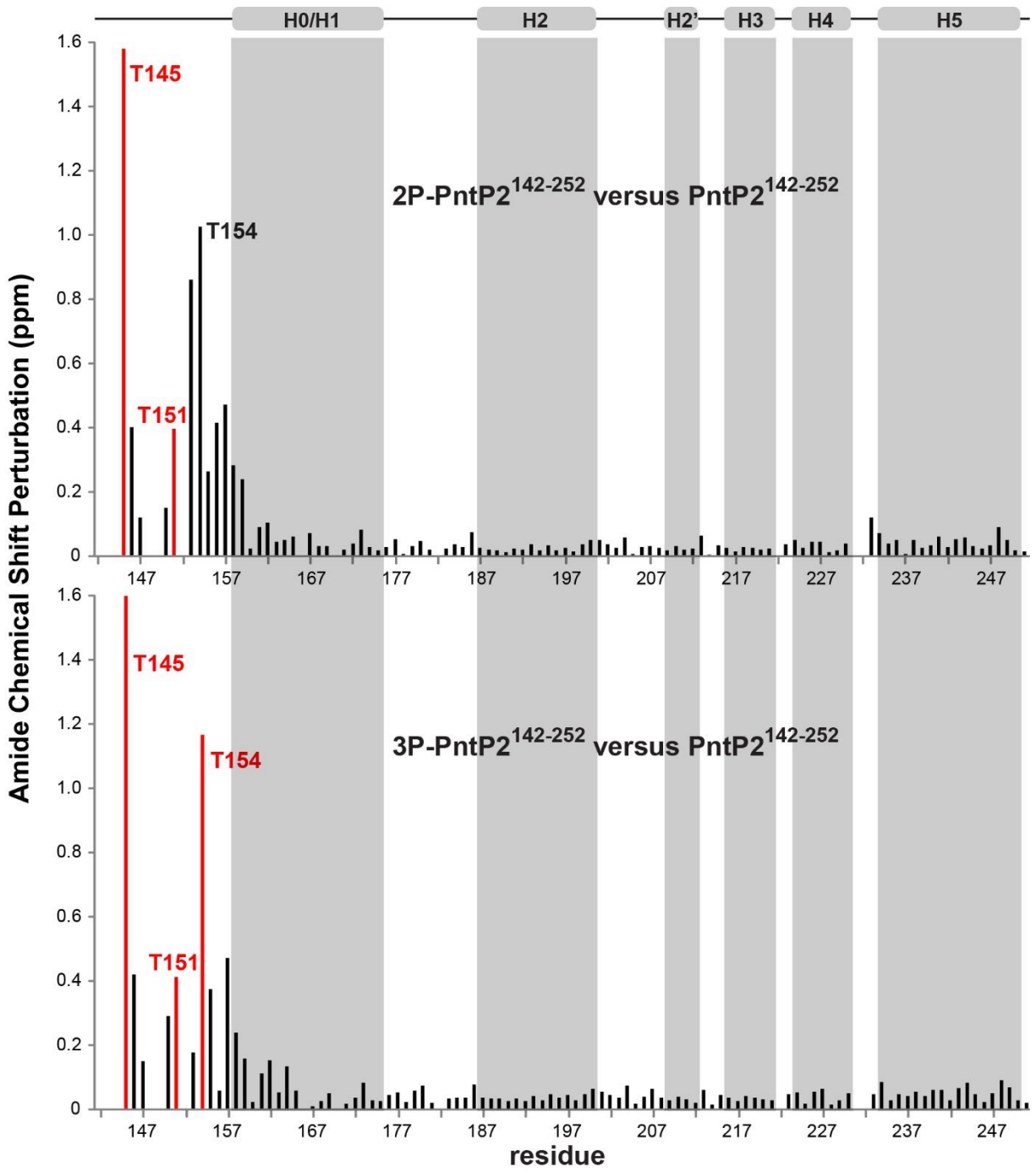
Figure 2-7 Amide chemical shift perturbations upon phosphorylation are localized to residues near the phosphoacceptor threonines.

Shown are values for (top) 2P-PntP2$^{142\text{-}252}$ and (bottom) 3P-PntP2$^{142\text{-}252}$ versus unmodified PntP2$^{142\text{-}252}$. The cartoon and grey rectangles indicate the helices based on the consensus MICS scores for the three PntP2$^{142\text{-}252}$ species (Figure 2-3 B,C,D), and the phosphoacceptors are highlighted in red. Missing data correspond to prolines or residues with overlapping signals. The amide chemical shift perturbations were calculated as $\{(\Delta\delta(_{1HN}))^2 + (\Delta\delta(_{15N})/5)^2\}^{1/2}$. Parenthetically, these data and the spectra of Figure 2-1 B and Figure 2-2 show that great caution must be exercised when using ^{15}N-HSQC spectra alone to identify sites of phosphorylation. For example, the perturbations of Thr154 presumably due to phosphorylation of Thr151 in 2P-PntP2$^{142\text{-}252}$ are comparable to those due to its own modification in 3P-PntP2$^{142\text{-}252}$. Similarly, Thr151 is also perturbed upon phosphorylation of Thr154. These effects might result in part from electrostatic interactions between the phosphate moieties. Also, Gly153 experiences compensating shift perturbations due to phosphorylation of Thr151 followed by Thr154 such that its signal almost overlaps in the ^{15}N-HSQC spectra of PntP2$^{142\text{-}252}$ and 3P-PntP2$^{142\text{-}252}$.

2.2.3. MAP kinase docking by Pointed-P2

The interaction of PntP2$^{142-252}$ with ERK2 was also examined using ^{15}N-HSQC-monitored titrations. Upon addition of the unlabeled kinase, an overall decrease in the ^{1}H^{N}-^{15}N signal intensities of amides throughout PntP2$^{142-252}$ was observed (Figure 2-8A,B). This is attributed to faster ^{1}H^{N} and ^{15}N transverse relaxation due to formation of a high molecular mass complex with the kinase. Based on kinetic studies and equilibrium binding measurements, the K_m or K_D values ERK2 kinases and PNT domains are in the 10-100 μM range (Qiao et al. 2006; Rainey et al. 2005; Seidel & Graves 2002), and thus partial saturation of PntP2$^{142-252}$ is expected under these experimental conditions. More importantly, residues in the phosphoacceptor region, as well as helices H0 and H5, showed substantially greater intensity perturbations. This suggests that these residues undergo pronounced exchange broadening due either to direct contacts with the kinase, as would be expected for the phosphoacceptors, or to indirect conformational perturbations. Indeed, when mapped on a homology model of PntP2$^{142-252}$, the residues in helices H0 and H5 are both in close proximity and adjacent to sidechains in the PNT domain shown previously by mutagenesis to be involved in docking interactions with the ERK2 Rolled (Figure 2-8C) (Qiao et al. 2006). Similar results have also been observed for the interaction of the ETS1 PNT domain with ERK2 (Seidel & Graves 2002).

57
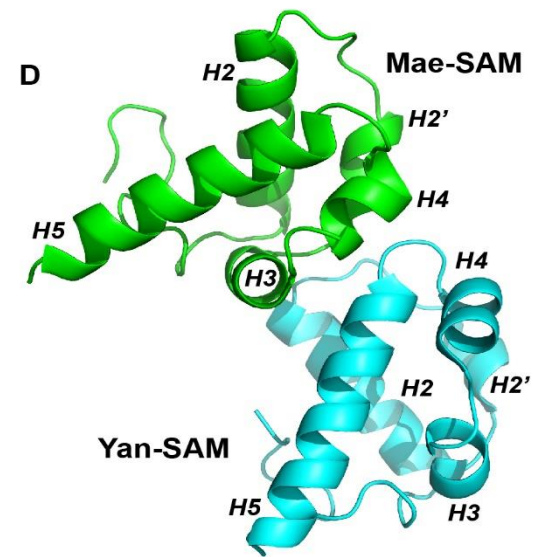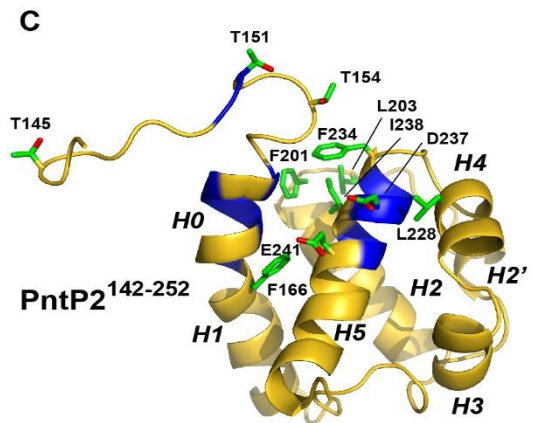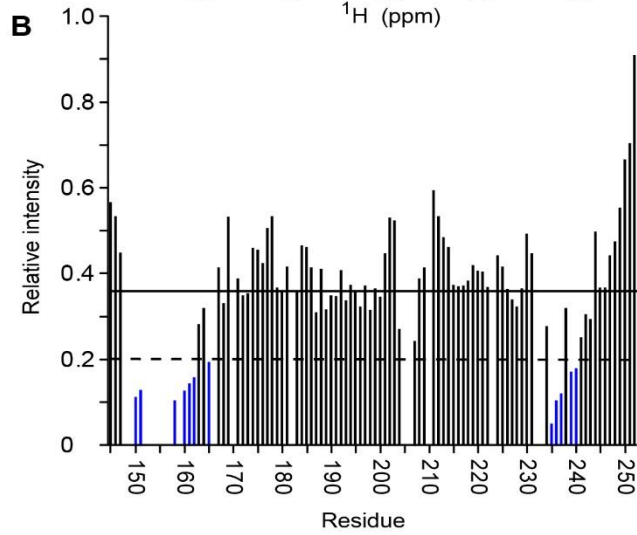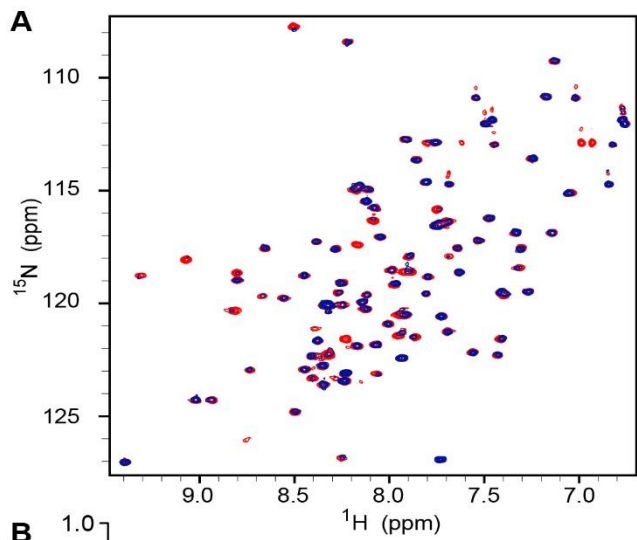
Figure 2-8 ERK2 interacts with the PNT domain and phosphoacceptor region of PntP2$^{142–252}$.

(A) Superimposed ^{15}N-HSQC spectra of PntP2$^{142–252}$ at pH 7.5 in the absence (red) and presence (blue) of a 0.25 molar equivalent of active ERK2. (B) Addition of EKR2 leads to a significant reduction in the relative signal intensities of specific amides in the phosphoacceptor region, as well as in helices H0 and H5 of PntP2$^{142–252}$ (blue histogram bars). There is also an overall reduction in ^{15}N-HSQC signal intensities to an average value of ~0.4 (horizontal solid line) attributed to sample dilution and increased global relaxation rates owing to the formation of a high-molecular-mass complex. Missing data correspond to prolines, amides with overlapping signals, or residues not observed owing to rapid HX at the elevated sample pH of 7.5 that was required to prevent ERK2 aggregation. (C) The homology model of PntP2$^{142–252}$ generated with SwissModel37 using Ets1$^{29–138}$ (2JV3.pdb) as a template. Amides showing the largest change in signal intensity upon ERK2 binding (below the horizontal dashed line in (B)) are highlighted in blue. Also shown in stick format (carbon, green; oxygen, red) are the side chains of residues identified by mutagenesis to be important for kinase docking (Qiao et al. 2006), as well as the three phosphoacceptors (Thr144, Thr151, and Thr154) and Phe166 at the H0/H1 bend. (D) X-ray crystallographic structure of the heterodimeric complex formed by the SAM/PNT domains of Mae (green) and Yan (cyan) (1SV0.pdb) (Qiao et al. 2004). Binding of Mae to the corresponding region of PntP2$^{142–252}$ to prevent Rolled ERK2 docking would require displacement of the dynamic helix H0. The structural figures were rendered with PyMol (Delano & Bromberg 2004).

2.3.   Discussion

Using NMR spectroscopy, we have examined the structural and dynamic properties of a fragment of Pointed-P2 encompassing its phosphoacceptor region and adjacent PNT domain. Based on mainchain chemical shifts, ^{15}N-NOE relaxation, and rapid amide HX measurements, the PNT domain of Pointed-P2 closely resembles that of ETS1 with a dynamic helix H0/H1 appended to a SAM domain core (helices H2-H5). Phosphorylation of three threonines in the disordered N-terminal region of PntP2$^{142\text{-}252}$ does not significantly perturb the structure of its PNT domain and only slightly increases the protection of residues in helix H0/H1 against HX. Very similar subtle effects were observed for Ets1$^{29\text{-}138}$ when phosphorylated at Thr38 and Ser41 (Nelson et al. 2010). However, as evidenced by changes in residual dipolar couplings and interproton NOE interactions, the dynamic helix H0 of 2P-Ets1$^{29\text{-}138}$ adopts a broad distribution of conformations that are more displaced from the PNT/SAM core than in the structural ensemble of unmodified Ets1$^{29\text{-}138}$ (Nelson et al. 2010). This increased displacement is likely owing to electrostatic repulsion between pThr38/pSer41 and several negatively-charged residues in helices H2 and H5. Given the sequence similarity of the two ETS family members, we speculate that such a conformational shift may also occur when Pointed-P2 is phosphorylated. Testing this hypothesis will, of course, require more detailed tertiary structural analyses of PntP2$^{142–252}$ in its unmodified and modified forms.

In addition to confirming that Thr151 is phosphorylated by ERK2, we also identified Thr145 and Thr154 as previously unrecognized phosphoacceptors adjacent to the PNT domain of Pointed-P2. Mutational studies have demonstrated that phosphorylation of Thr151 is critically required for the *in vivo* function of Pointed-P2 (Brunner et al. 1994; O'Neill et al. 1994; Tootle & Rebay 2005), whereas the biological roles, if any, of these additional non-consensus sites have not been examined. It is certainly possible that these modifications are an artifact resulting from using a large scale *in vitro* kinase system to produce milligram quantities of modified PntP2$^{142\text{-}252}$ for NMR spectroscopic studies, and/or due to differences between mammalian ERK2 and *Drosophila* Rolled. However,

Ets1[29-138] is also phosphorylated *in vitro* by ERK2 at the corresponding Thr38 and Ser41 (Fig. 1F), and *in vivo* tests have confirmed that both these residues contribute to Ras-enhanced transactivation by Ets1 (Foulds et al. 2004; Nelson et al. 2010). Thus, the potential roles of Thr145 and Thr154 in the control of gene expression by Pointed-P2 remain to be evaluated.

The results of this study have several implications for understanding the role of Pointed-P2 in *Drosophila* signal transduction. Based on a mutational analysis, the Bowie group (Qiao et al. 2006) identified several residues centered around Phe234 at the start of helix H5 that contribute to the docking of Pointed-P2 and the ERK2 Rolled (Figure 2-8C). Consistent with this analysis, amides within helix H5 showed pronounced spectral perturbations upon titration of PntP2[142-252] with ERK2. Furthermore, by comparison with the X-ray crystallographic structure of the Mae-SAM/Yan-SAM heterodimer (Figure 2-8 D) (Qiao et al. 2004), these residues are also within the expected interface for Mae, thus leading to the proposal that Mae attenuates the activity of Pointed-P2 by sterically blocking kinase docking (Qiao et al. 2006). A homology model of PntP2[142-252], generated from the NMR spectroscopically-derived structure of Ets1[29-138], predicts that helix H0 and the adjacent phosphoacceptors would partially occlude this interface (Figure 2-8C). If so, then helix H0 must be displaced to allow the binding of either ERK2 Rolled or Mae. This could lead to the ^{15}N-HSQC signal losses for residues in helix H0 observed when ERK2 was added to PntP2[142-252]. The conformational flexibility and marginal stability of helix H0, detected by NMR relaxation and HX studies, would allow such displacement to readily occur. A similar proposal has been made for the interaction of Ets1 and ERK2, and leads to the hypothesis that helix H0 must unfold to allow both the interaction of the PNT domain with two distinct docking sites on the kinase and binding of the phosphoacceptors within its active site (S. Lee et al. 2011; Nelson et al. 2010). Unfortunately, we found the isolated Mae PNT domain to be very insoluble *in vitro* and thus were unable to examine its predicted effect on PntP2[142-252] using NMR spectroscopy.

We speculate that the conformational flexibility and marginal stability of helix H0, detected by NMR relaxation and HX studies, might facilitate the phosphorylation of Pointed-P2 in

two related ways. First, when folded, helix H0 positions the phosphoacceptors near the docking surface of the PNT domain, perhaps enhancing the initial association with the MAP kinase. Subsequently, the facile unfolding of this helix could then allow simultaneous interactions of the PNT domain with a docking site on the kinase and the phosphoacceptors with its active site to enable proximity-enhanced catalysis (Rainey et al. 2005). A similar proposal has been made for the interaction of Ets1 and ERK2 (Nelson et al. 2010; S. Lee et al. 2011).

It is well established that phosphorylation of Pointed-P2 at Thr151 is necessary for the activation of its target genes (Brunner et al. 1994; O'Neill et al. 1994; Tootle & Rebay 2005; Vivekanand & Rebay 2006). However, the molecular mechanisms underlying this process have not been defined. By analogy to Ets1 (Foulds et al. 2004; Nelson et al. 2010), it is plausible that phosphorylation of Pointed-P2 leads to enhanced recruitment of Nejire, the *Drosophila* ortholog of the mammalian co-activator CBP. Although studies have shown that Nejire functions during successive stages of *Drosophila* eye development (Kumar et al. 2004), such a direct interaction with Pointed-P2 has not been reported. In an effort to test this hypothesis, we expressed the predicted TAZ1 domain of Nejire using a range of methods established for the TAZ1 domain of CBP and p300 (De Guzman et al. 2005). Unfortunately, we were unsuccessful in obtaining a soluble, folded protein fragment as required for NMR-monitored binding studies with PntP2[142-252]. Therefore, future investigations will be required to uncover the link between Pointed-P2 phosphorylation and transcriptional activation. Our demonstration that the PNT domains and adjacent phosphoacceptors of ETS1 and Pointed-P2 share very similar structure and dynamics guide this research.

2.4. Materials and methods

2.4.1. Protein expression

The gene encoding residues 142-252 of *D. melanoganstar* Pointed-P2 (Genbank NM_079737.2) was cloned by PCR methods into the pET28a vector for expression in *Escherichia coli* BL21 (λDE3) cells as a His$_6$-tagged construct. The single cysteine (Cys250) was mutated to serine to avoid potential oxidation. Following established protocols (Nelson et al. 2010), samples of ^{15}N/^{13}C-labeled protein were produced in M9 minimal media, containing 1 g/L ^{15}NH$_4$Cl and 3 g/L ^{13}C$_6$-glucose, with 1 mM IPTG induction overnight at 30 °C. Harvested cells were resuspended and homogenized in lysis buffer (20 mM sodium phosphate, 500 mM NaCl, 5 mM imidazole, 2 mM DTT, pH 7.4) and the expressed protein isolated by Ni^{+2}-affinity chromatography, followed by thrombin digestion to remove the His$_6$-tag. After further purification using S75 size exclusion chromatography, the protein was dialyzed against NMR sample buffer (20 mM MOPS, 10 mM NaCl, 2 mM DTT, pH 6.7) and concentrated by ultrafiltration. The resulting construct contains a non-native N-terminal Gly-Ser-His-Met from the cleavage site and is denoted as PntP2$^{142-252}$.

2.4.2. *In vitro* phosphorylation

PntP2$^{142-252}$ was phosphorylated by overnight incubation at 30 °C in a 40:1 molar ratio with ERK2 kinase (125 mM Tris, 5 mM DTT, 50 mM MgCl$_2$, 100 mM NaCl, pH 7.5), as described previously for Ets1$^{29-138}$ (Nelson et al. 2010). The active rat ERK2 was prepared from *E. coli* BL21 (λDE3) grown in TB media with 0.8% glycerol, 0.1 mg/ml carbenicillin, 0.08% glucose and 0.04 mM IPTG for induction, according to published methods (Seidel & Graves 2002; Foulds et al. 2004). The resulting 2P- and 3P-PntP2$^{142-252}$ were separated by anion exchange chromatography on a Mono Q column (20 mM Tris pH 7.5, 10% glycerol, 2 mM DTT, gradient 0-1 M NaCl), confirmed by MALDI-ToF mass spectrometry, and dialyzed into NMR sample buffer.

2.4.3. NMR spectroscopy

Spectra were obtained for proteins in NMR sample buffer with 10% D_2O at 25 °C using 600 MHz Varian Inova or Bruker Avance III spectrometers. Data were processed with NMRpipe (Delaglio et al. 1995) and analyzed using Sparky (Lee et al. 2015). Main chain resonance assignments were obtained using standard ^{15}N-HSQC, HNCO, HN(CA)CO, CBCA(CO)NH, HNCACB, and C(CO)TOCSY-NH experiments recorded with $^{1}H/^{13}C/^{15}N$ cryogenic probes (Sattler et al. 1999). The ^{1}H-^{15}N spectrum of a ^{31}P-edited HNCA experiment was recorded using a room temperature $^{1}H/^{13}C/^{15}N/^{31}P$ QXI probe. Steady state $^{1}H/^{15}N$ heteronuclear NOE relaxation (Farrow et al. 1994) and CLEANEX amide HX (Hwang et al. 1998) measurements were recorded and analyzed as described previously (Nelson et al. 2010).

Chapter 3.   Structured and disordered regions cooperatively mediate DNA-binding autoinhibition of ETS factors ETV1, ETV4, and ETV5

Chapter 3 is a close collaborative work with Dr. Simon Currie and Dr. Barbara Graves at the University of Utah. It is a modified version of the article: Currie, S.L.*, Lau, D.K.W.*, Doane, J.J., Whitby, F.G., Okon, M., McIntosh, L.P., Graves, B.J. (2017) *"Structured and disordered regions cooperatively mediate DNA-binding autoinhibition of ETS factors ETV1, ETV4, and ETV5"* **Nucleic Acid Res**. 45:2223-2241 (* co-first authors). The modifications include reformatting, the incorporation of supplementary material into the main text, and the addition of ancillary data not published in the original manuscript. Briefly, the DNA-binding affinity measurements and X-ray crystallography studies were done by Simon Currie at the University of Utah. I performed all NMR spectroscopic studies at the University of British Columbia. Simon Currie, Desmond Lau, Lawrence McIntosh and Barbara Graves all contributed to data analysis and manuscript preparation.

Overview

Autoinhibition enables spatial and temporal regulation of cellular processes by coupling protein activity to surrounding conditions, often via protein partnerships or signaling pathways. We report the molecular basis of DNA-binding autoinhibition of ETS transcription factors ETV1, ETV4 and ETV5, which are often overexpressed in prostate cancer. Inhibitory elements that cooperate to repress DNA binding were identified in regions N- and C-terminal of the ETS domain. Crystal structures of these three factors revealed an α-helix in the C-terminal inhibitory domain that packs against the ETS domain and perturbs the conformation of its DNA-recognition helix. NMR spectroscopy demonstrated that the N-terminal inhibitory domain is intrinsically disordered, yet utilizes transient intramolecular interactions with the DNA-recognition helix of the ETS domain to

mediate autoinhibition. Acetylation of selected lysines within the N-terminal inhibitory domain activates DNA binding. This investigation revealed a distinctive mechanism for DNA-binding autoinhibition in the ETV1/4/5 subfamily involving a network of intramolecular interactions not present in other ETS factors. These distinguishing inhibitory elements provide a platform through which cellular triggers, such as protein-protein interactions or post-translational modifications, may specifically regulate the function of these oncogenic proteins.

3.1.   Introduction

Autoinhibition occurs in diverse proteins and allows for spatiotemporal modulation of biological processes in response to various inputs such as signaling pathways and macromolecular interactions (Pufall & Graves 2002). This self-dampening behavior can influence the equilibria between the active and inactive states of proteins by serving as the integration point for post-translational modifications and protein partnerships. Partaking in alternative intramolecular and intermolecular interactions is often the key attribute for an autoinhibitory element (Kim et al. 2000; Morreale et al. 2000). Both structured regions with dynamic character and intrinsically disordered regions (IDRs) can be effective inhibitory elements (Trudeau et al. 2013). Conformational flexibility, and even disorder, allows for distinct and adaptable recognition of intramolecular interfaces, as well as surfaces on diverse interacting proteins (Wright & Dyson 2015).

The *ETS* gene family, which encodes 28 human transcription factors, has provided a model system to develop an understanding of autoinhibition of sequence-specific DNA binding (Hollenhorst, McIntosh, et al. 2011). The conserved ETS domain is autoinhibited in several family members, yet by different mechanisms. For example, a serine-rich IDR allosterically inhibits DNA binding of ETS1 through transient phosphorylation-enhanced interactions with the structured ETS domain and flanking N- and C-terminal inhibitory α-helices (Lee et al. 2005; Pufall et al. 2005). In contrast, a single flanking C-terminal α-

helix sterically inhibits DNA binding of ETV6 (Green et al. 2010; Coyne et al. 2012; De et al. 2014). In the biological context, autoinhibition of a particular ETS factor provides distinct routes to specific regulation, such as via post-translational modifications (Pufall et al. 2005; Lee et al. 2008) and protein-protein interactions (Greenall et al. 2001; Shrivastava et al. 2014; Shiina et al. 2014; Garvie et al. 2002). Collectively, this has led to the hypothesis that divergent modes of autoinhibition involving regions flanking the ETS domain help enable specific gene regulation by individual ETS factors (Hollenhorst, McIntosh, et al. 2011).

The involvement of the *ETS* genes of the ERG and ETV1/4/5 (also known as PEA3) subfamilies in prostate cancer motivated our goal to expand the mechanistic understanding of autoinhibition to these ETS factors. Chromosomal rearrangements involving ERG and ETV1/4/5 subfamilies are observed in the majority of prostate cancer tumors (Tomlins et al. 2005; Tomlins et al. 2006; Helgeson et al. 2008). There is aberrant expression of these full-length, or nearly full-length, ETS proteins upon rearrangement with a prostate-specific or a constitutively expressed promoter (Tomlins et al. 2007). In addition, ETV1 and ETV4 mediate PI3-kinase and Ras signaling pathways, resulting in aggressive and metastatic disease phenotypes (Aytes et al. 2013; Baena et al. 2013). Previous studies suggested that the ETV1/4/5 subfamily also displays autoinhibition of DNA binding (Greenall et al. 2001; Laget et al. 1996; Bojović & Hassell 2001); however, detailed characterization, including structural mapping of inhibitory elements, and mechanistic insights are lacking. We propose that a full understanding of the autoinhibition of ERG and ETV1/4/5 and its regulation by cellular processes will enable insights into the roles of these factors in prostate cancer progression and provide windows of opportunity for targeted therapeutic interventions.

In this study we describe the molecular basis of DNA-binding autoinhibition in the ETV1/4/5 subfamily of ETS factors. Using ETV4 as a model for this subfamily, we found that inhibitory domains reside both N- and C-terminal of the ETS domain and cooperate to inhibit DNA binding. Crystal structures identified the C-terminal inhibitory domain (CID) as an α-helix that packs against the ETS domain and perturbs the relative positioning of

its DNA-recognition helix. NMR spectroscopy demonstrated that the N-terminal inhibitory domain (NID) is an IDR that transiently interacts with the ETS domain and the CID. Lysine acetylation of the NID relieves autoinhibition, likely through disruption of these intramolecular interactions. Mutational analyses revealed specific intramolecular linkages among the regulatory elements. From these findings we propose a model for autoinhibition in the ETV1/4/5 subfamily in which structured and disordered regions regulate the DNA-recognition helix.

3.2. Results

3.2.1. Identification of the inhibitory sequences boundary

In collaboration with Simon Curie and Dr. Barbara Graves at the University of Utah, we initially sought to determine the magnitude of autoinhibition in the ETV1/4/5 subfamilies of ETS factors. Towards this aim, they measured the DNA binding affinities (equilibrium dissociation constant, K_D) of the full-length proteins, different truncation fragments and nearly-minimal DNA-binding domains (DBD) for ERG, FLI1, ETV1, ETV4, and ETV5 (Figure 3-1A and Table 3-1). Overall, robust autoinhibition was observed in ETV4 with the full-length proteins displaying ~10-fold weaker binding than their minimal DBDs (Figure 3-1). These levels of autoinhibition are comparable to those previously reported for ETS1 (Lee et al. 2005) and ETV6 (Green et al. 2010). In contrast, ERG and its subfamily member FLI1 displayed modest 2- to 3-fold autoinhibition, as also previously reported (Regan et al. 2013). Interestingly, the K_D values cluster in a pattern that reflects their subfamily phylogenetic classifications (Figure 3-1E) (Hollenhorst, Ferris, et al. 2011). The ~ 100-fold range of K_D values for ETS DBDs suggests that there are key differences that influence DNA binding despite the high overall sequence conservation in the ETS domain of these factors. Additionally, the known inhibitory elements from ETS1 and ETV6 are not conserved in the ETV1/4/5 or ERG subfamilies, and the poor sequence conservation outside of the ETS domain among these factors indicates that the mechanism of

autoinhibition is likely different for the ETV1/4/5 and ERG subfamilies (Figure 3-2). Based on the larger magnitude of autoinhibition observed with ETV1, ETV4, and ETV5, as compared to ERG and FLI1, we focused on the ETV1/4/5 subfamily for mechanistic studies.
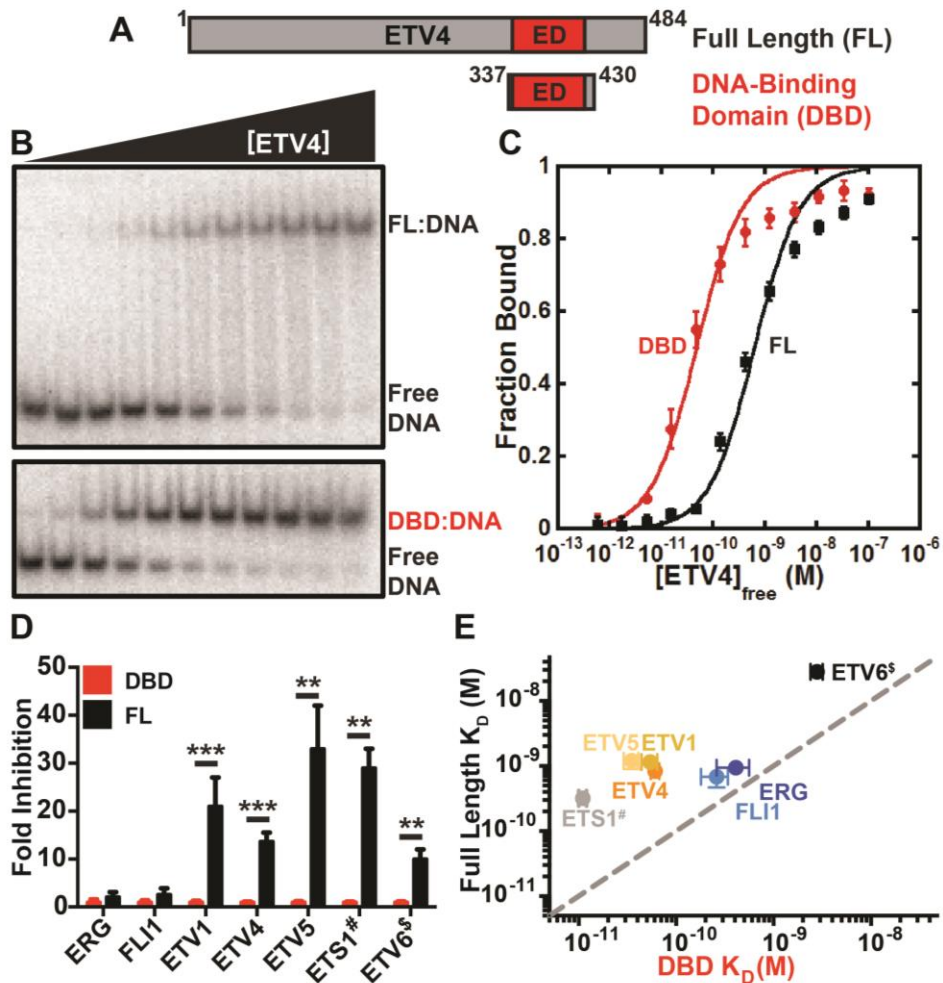
Figure 3-1 ETV1/4/5 are autoinhibited.

Autoinhibition in the ERG and ETV1/4/5 subfamilies. (A) Schematic of full-length protein, FL, and nearly minimal DNA-binding domain, DBD, for ETV4. Based on the sequences of all ETS factors, the conserved ETS domain, ED, is noted in red. (B) Representative examples of EMSA gels for ETV4 FL or DBD with a double-stranded DNA duplex containing a core ETS binding site. (C) Binding isotherms for ETV4 FL and DBD. Data points and error bars refer to the mean and the standard error of the mean with 4 replicates for each protein. See methods for details. (D) Fold inhibition of ERG, FLI1, ETV1, ETV4, and ETV5, calculated as K_D (FL or DBD) / K_D (DBD). ETS1 (Pufall et al. 2005)[#] and ETV6 (Green et al. 2010)[$] data are included for comparison. Mean and standard error of the mean from at least three replicates are plotted; "**" and "***" indicate $p < 0.01$ and $p < 0.001$, respectively. See Table 3-1 for K_D values and numbers of replicates. (E) K_D values of FL versus DBD for each of the ETS factors tested. The dashed diagonal line represents no autoinhibition [i.e., K_D (FL) = K_D (DBD)].

Table 3-1 Equilibrium dissociation constants (K_D) and fold-inhibition values for ETS factors

ETS Factor	Fragment	K_D (x10^{-11} M)[a]	Fold-inhibition[a,b]	p[c]	n
ERG	DBD$^{307-400}$	40 ± 10	1.0 ± 0.5	-	3
	FL$^{1-479}$	94 ± 9	2.3 ± 0.9	0.05	3
FLI1	DBD$^{277-370}$	26 ± 8	1.0 ± 0.4	-	7
	FL$^{1-452}$	70 ± 20	3 ± 1	0.1	3
ETV1	DBD$^{332-425}$	5.4 ± 1.0	1.0 ± 0.3		6
	FL$^{1-479}$	110 ± 20	21 ± 6	0.0006	10
ETV4	DBD$^{337-430}$	6.1 ± 0.6	1.0 ± 0.1	-	25
	FL$^{1-484}$	83 ± 8	14 ± 2	3 x 10^{-7}	35
ETV5	DBD$^{364-457}$	3.6 ± 0.4	1.0 ± 0.2	-	4
	FL$^{1-510}$	140 ± 30	39 ± 9	0.003	8
ETS1[d]	DBD	1.1 ± 0.1	1.0 ± 0.1	-	3
	FL	32 ± 4	29 ± 4	0.002	3
ETV6[e]	DBD	280 ± 40	1.0 ± 0.2	-	4
	FL	2,800 ± 400	10 ± 2	0.004	4

[a] Mean and standard error of the mean are given for K_D and fold-inhibition values.
[b] The DBD is set as uninhibited and used as a reference for calculating fold inhibition as K_D (FL or DBD) / K_D (DBD).
[c] The p-values were calculated using a two-tailed heteroscedastic t-test and compare the DBD and FL fragments for each ETS factor.
[d] Data included for comparison from reference (Pufall & Graves 2002).
[e] Data included for comparison from reference (Kim et al. 2000).

```
ETV6    ----------------------------------------------------------------  ----------------------------------------------------------------
ETV7    ----------------------------------------------------------------  ----------------------------------------------------------------
ETV4    MERRMKAGYLDQQVPYTFSSKSPGNGSLREALIGPLGKLMDPGSLPPLDSEDLFQDLSHFQETWLAEAQVPDSDEQ--FVPDFHS-----------ENLAFH-------SPT-----TRI
ETV1    ------MDGFYDQQVPYMVTNSQRGRNCNEKPTNVRKRKFINRD--LAHDSEELFQDLSQLQEAWLAEAQVPDNDEQ--FVPDYQA-----------ESLAFH-----GLP-----LKI
ETV5    -----MDGFYDQQVPFMVPGKSRSEECRGRPVIDRKRKFLDTD--LAHDSEELFQDLSQLQEAWLAEAQVPD-DEQ--FVPDFQS-----------DNLVLH------APP----PTKI
ERG     ---------MIQTVPDP------------------------AAHIKEALSVVSEDQSLFECAYGT-PHLAKTEMTASSSSDYGQTSKMSPRVPQQD---W-----L--SQPPARVTIKM
FLI1    ---------------M-------------------------DGTIKEALSVVSDDQSLFDSAYGAAAHLPKADMTASGSPDYGQPHKINPLPPQQE---W-----I--NQPVR-V--NV
ETS1    -----------------------------------------------MKAAVDLKPTLTI---------------IKTEKVDL
ETS2    -------------------------------------------MNDFGIK----NMDQVAPVANSYRGTLKRQPAFDTFDGSLFAVFPSLNEEQTLQEVPTGL

ETV6    ---------MSETPAQCSIKQERISY--TPPESPVPSYA-SSTPLHVPVPRALRMEEDSIRLPAHLRLQPIYWSRDDVAQWLKWAENEFSLRPIDSNT----F-EMNGK----ALL--LL
ETV7    ------------------MQEGELAISPISPVAAMPPLG-THVQARCEAQINLLGEGGICKLPGRLRIQPALWSREDVLHWLRWAEQEYSLPCTAEHG----F-EMNGR----ALC--IL
ETV4    KK--EPQSPRTDPALSCSRKPPLPYH---HGEQCLYSS-AYD----PPRQIAIK------------SPAPGALGQSPLQPFP-----R-------------------------------A
ETV1    KK--EPHSPCSEISSACSQEQPFKFS---YGEKCLYNVSAYD----QKPQVGMR------------PSNPPTPSSTPVSPL------HHASPNSTHTP-------------------KP
ETV5    KR--ELHSPSSEL-SSCSHEQALGAN---YGEKCLYNYCAYD-----RKPPSGFK------------PLTPPTTPLSPTHQNPLFPPPQATLPTSGHAPAAGPVQGVGPAPAPHSLPEPGP
ERG     ECNPSQVNGSRNSPDECSVAKGGKMV--GSPDTVGMNYGSYMEEKHM-PPPNMTTNERRVI----VPADPTLWSTDHVRQWLEWAVKEYGLPDVNILL----FQNIDGK----ELC--KM
FLI1    KREYDHMNGSRESPVDCSVSKCSKLV--GGGESNPMNYNSYMDEKNGPPPPNMTTNERRVI----VPADPTLWTQEHVRQWLEWAIKEYSLMEIDTSF----FQNMDGK----ELC--KM
ETS1    ELFPSPDMECADVPL---LTPSSKE---------------MMSQALKATFSGFTKEQQRLG----IPKDPRQWTETHVRDWVMWAVNEFSLKGVDFQK----F-CMNGA----ALC--AL
ETS2    DSI-SHDSANCELPL---LTPCSKA---------------VMSQALKATFSGFKKEQRRLG----IPKNPWLWSEQQVCQWLLWATNEFSLVNVNLQR----F-GMNGQ----MLC--NL

ETV6    TKEDFRYR-SPHSGDVLYELLQHILKQRKPRILFSPF------FHPGNSIHTQPEVILHCNHEEDNCVQRTPRPSVDNVHHNPPTIELLHRSRSPITTNHRPSPDPEQRPLRSPL---DN
ETV7    TKDDFRHR-APSSGDVLYELLQYIKTQRRA-LVCGPF------FGGIFRLKTPTQ-------------HSPVPPE----------EVT-------------GP--S------QM---DT
ETV4    EQRNFLRSSGTS-----------------------QP------HPGHG--YLGEHSSVFQQPLDICHSFTSQGGGREPLPAP--YQHQL--SEPCP---------PYPQQSFKQEYHDP
ETV1    D-RAFPAH-------------------------------LPPSQSIPDSS--YPMDH-RFRRQLSEPCNSFPPLPTMPREGRPM--YQRQM--SEPNI---------PFPPQGFKQEYHDP
ETV5    QQQTFAVPRPPH-------------------------QPLQMPKMMPENQ--YPSEQ-RFCRQLSEPCHPFPPQPGVPGDNRPS--YHRQM--SEPIV------PAAPPPPQGFKQEYHDP
ERG     TKDDFQRLTPSYNADILLSHLHYLRETPLPHLTSDD----VD----------------KALQNSPRLMHARNTGGAAFIFPN---TSVY------------PEATQR-I---------
FLI1    NKEDFLRATTLYNTEVLLSHLSYLRESSLLAYNTTS-----------------------------------H------------TDQSSR-L---------
ETS1    GKDCFLELAPDFVGDILWEHLEILQKEDVKPY----QVNGVNPAYPESR--YTSDYFISYGIEHAQCVP-PSEFSEPSFITES--YQTLH---------PISSEELLSLKYE-ND-
ETS2    GKERFLELAPDFVGDILWEHLEQMIKENQEKTEDQYEENSHLTSVP--H--WINSNTLGFGTEQAPYGMQTQNYPKGGLL------DSMC------------PASTPSVLSSEQEFQM-

ETV6    MIRRLSPAERAQGPRPHQENNHQESYPLSVSPMENNHCPASSESHP--------------K------PSS-PRQESTRVIQLMPSPIMHPLILNPRHSVDFKQSRLSEDGLHR----EGK
ETV7    RRG-----HLLQPPDPGLTSN---------------------------  ------------------------------------FGHLDDPGLARWTPGKEE
ETV4    LYEQAGQPAVDQGGVN---GHRYPGAGVVIKQEQTDF-AYDSDVTGCASMYL-----------------------------------HTEGFSGPSPGD-G
ETV1    VYEHNTMV----GSAA---SQSFP-PPLMIKQEPRDF-AYDSEVPSCHSIYM-----------------------------------RQEGFLAHPS---R
ETV5    LYEHGVPG---MPGPP---AHGFQ-SPMGIKQEPRDY-CVDSEVPNCQSSYM-----------------------------------RGGYFS---S---S
ERG     ------------------------------------TTRPD------L------------------PYEPPRR--SAWTG-HGHPTPQSKA----------AQPS----------
FLI1    ------------------------------------SVKED------P------------------SYDSVRR--GAWGN-NMNSGLNKSP---------PLGG----------
ETS1    -YPSV----------------IL-RDPLQTDTLQNDYFAIKQEVVTPDNMCMGRTSRGKL-GGQDSFESIESYDSCDRLTQSWSS-QS--SFNSLQ---------RVPSYDSFDSED-Y
ETS2    -FPKS----------------R-------LSSVSVTYCSVSQDFPGS-NLNLLTNNSGTPKDHDSPENGADSFESSDSLLQSWNS-QS--SLLDVQ---------RVPSFESFEDD--C

ETV6    PINLSHREDL-AYMNHIMVSV----------SPPEEHAMPIGRIADCRLLWDYVYQLLSDSRYENFIRWEDKESKIFRIVDPNGLARLWGNHKNRTNMTYEKMSRALRHYYKLNIIRKEP
ETV7    SLNLCHCAEL-GCRTQG-VCS----------FPAMPQAPIDGRIADCRLLWDYVYQLLLDTRYEPYIKWEDKDAKIFRVVDPNGLARLWGNHKNRVNMTYEKMSRALRHYYKLNIIKKEP
ETV4    AMGYGYEKPLRPFPDDVCVVPEKFEGDIKQEGVGAFREGPPYQRRGALQLWQFLVALLDDPTNAHFIAWTGRG-MEFKLIEPEEVARLWGIQKNRPAMNYDKLSRSLRYYYEKGIMQKVA
ETV1    TEGCMFEKGPRQFYDDTCVVPEKFDGDIKQE-PGMYREGPTYQRRGSLQLWQFLVALLDDPSNSHFIAWTGRG-MEFKLIEPEEVARRWGIQKNRPAMNYDKLSRSLRYYYEKGIMQKVA
ETV5    HEGFSYEKDPRLYFDDTCVVPERLEGKVKQE-PTMYREGPPYQRRGSLQLWQFLVTLLDDPANAHFIAWTGRG-MEFKLIEPEEVARRWGIQKNRPAMNYDKLSRSLRYYYEKGIMQKVA
ERG     PSTVPKT-------EDQRPQLDPY----QILGPTSSR--LANPGSGQIQLWQFLLELLSDSSNSSCITWEGTN-GEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVH
FLI1    AQTISKN-------TEQRPQPDPY----QILGPTSSR--LANPGSGQIQLWQFLLELLSDSANASCITWEGTN-GEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVH
ETS1    PAALPNHKPKGTFKDYVRDRAD-LNK-DKPVIPAAAL--AGYTGSGPIQLWQFLLELLTDKSCQSFISWTGDG-WEFKLSDPDEVARRWGKRKNKPKMNYEKLSRGLRYYYDKNIIHKTA
ETS2    SQSLCLNKPTMSFKDYIQERSDPVEQ-GKPVIPAAVL--AGFTGSGPIQLWQFLLELLSDKSCQSFISWTGDG-WEFKLADPDEVARRWGKRKNKPKMNYEKLSRGLRYYYDKNIIHKTS

ETV6    GQRLLFRFMKTPDEIMSGRT-DRLEHLESQELDEQI------------YQEDEC-----------------------------------
ETV7    GQKLLFRFLKTPGKMVQDKH-SHLEPLESQEQDRIE------------FKDKRPEISP-----------------------------------
ETV4    GERYVYKFVCEPEALFSLAFPDNQRPALKAEFDRPV----------SEEDTVPLS---HLDESPAYLPELAG-----------PAQPFGPKGGYSY
ETV1    GERYVYKFVCDPEALFSMAFPDNQRPLLKTDMERHI----------NEEDTVPLS---HFDESMAYMPEGGC-------------CNPHPYNEGYVY
ETV5    GERYVYKFVCDPDALFSMAFPDNQRPFLKAESECHL----------SEEDTLPLT---HFEDSPAYLLDMDR-----------CSSLPYAEGFAY
ERG     GKRYAYKFDFHGIAQALQPHPPESSL-YKYPSDLPYMGSYHAHPQKMNFVAPHPPALPVTSSSFFAAPNPYWNSPTGGIYPNTRL---PTSHMPSHLGTYY
FLI1    GKRYAYKFDFHGIAQALQPHPTESSM-YKYPSDISYMPSYHAHQQKVNFVPPHPSSMPVTSSSFFGAASQYWTSPTGGIYPNPNVPRHPNTHVPSHLGSYY
ETS1    GKRYVYRFVCDLQ--SLLGYTPE---------ELHAMLDV---------KPDADE-----------------------------------
ETS2    GKRYVYRFVCDLQ--NLLGFTPE---------ELHAILGV---------QPDTED-----------------------------------
```

Figure 3-2 Sequence alignment of ETS factors tested for autoinhibition.

The full-length sequences for ETS factors tested for autoinhibition were aligned using Clustal Omega (Sievers et al. 2011). Sequences for ETV7 and ETS2 were included as these factors belong to the same subfamilies as ETV6 and ETS1, respectively. The ETS domain (ETV4[339-420]) is shaded red and flanking α-helices and known inhibitory regions are shaded cyan and labeled per previous nomenclature (Pufall et al. 2005; Coyne et al. 2012; Regan et al. 2013). Residues discussed in this study are indicated by an arrow; ETV4 Y220, F225, K226, Y229, L233, Y234, W344, Y401, Y403, I407, F414, A426, and L430. These factors are highly conserved within the ETS domain, and are highly divergent outside of the ETS domain. Additionally, known inhibitory regions from ETV6 (H4 and H5) and ETS1 (SRR, HI-1, HI-2, H4, and H5) are not conserved in the ETV1/4/5 or ERG subfamilies.

We chose ETV4 as a model factor to further investigate autoinhibition in the ETV1/4/5 subfamily. Initially, our collaborators in Utah used partial proteolysis to aid the design of truncation boundaries for mapping inhibitory elements (Figure 3-3A-C). They found that the predominant trypsin-resistant fragment, spanning amino acids 165-484, retained levels of autoinhibition comparable to full-length ETV4 (Figure 3-4A and Table 3-2). Subsequent deletion studies revealed that amino acid residues both N- and C-terminal of the ETS domain inhibit DNA binding independently, but also act cooperatively to yield higher than additive levels of inhibition (Figure 3-4A). Hereafter, these regions will be denoted as the NID (N-terminal inhibitory domain) (ETV4$^{165-336}$) and the CID (C-terminal inhibitory domain) (ETV$^{427-436}$), whereas the nearly-minimal DBD will be denoted as an uninhibited species. We hypothesized that the ETV1/4/5 NID and CID function through direct interactions with the ETS domain and/or with each other to cooperatively inhibit DNA binding (Figure 3-4C).
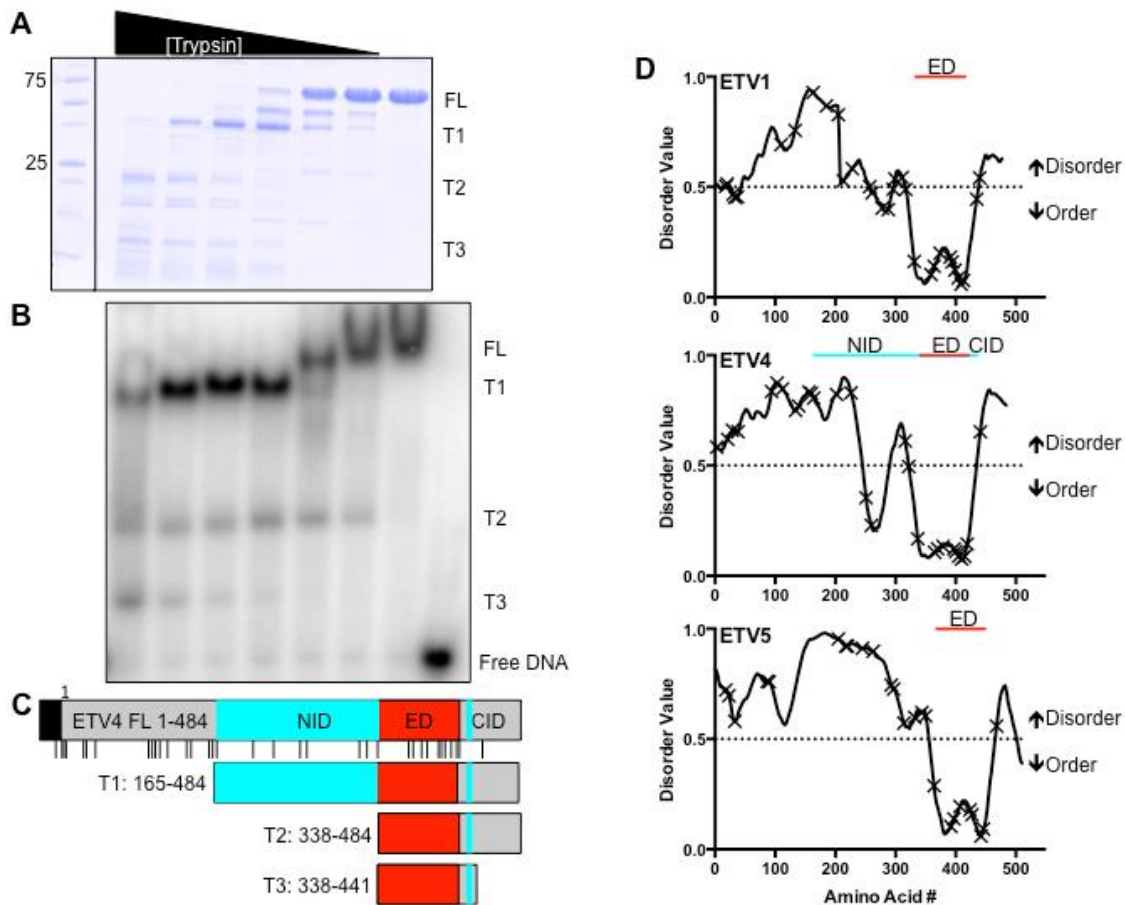
Figure 3-3 ETV4$^{165-484}$ is a trypsin-resistant fragment.

(A) SDS-PAGE gel of partial trypsin proteolysis of ETV4. The leftmost lane contains protein molecular weight standards, and next seven lanes show products from two minute digestion with 450, 150, 45, 15, 4.5, 1.5, and 0 ng of trypsin. A representative example of three independent experiments is displayed. (B) Electrophoretic mobility shift assay with tryptic fragments from (A). The far-right lane is a DNA-only control. (C) Schematic of ETV4 full length (FL) and tryptic fragments retaining the ETS domain as identified by electrospray ionization mass spectrometry (ESI-MS). The predominant DNA-binding tryptic fragments are named T1, T2, and T3. The black bar refers to an N-terminal His$_6$-tag encoded by the pET28 vector and the vertical lines mark potential trypsin digestion sites as predicted by ExPASY Peptide Cutter (Gasteiger et al. 2005). The ETS domain (ED) is noted in red, and N-terminal inhibitory domain (NID) and C-terminal inhibitory domain (CID), as identified for ETV4 (Figure 3-4), are noted in cyan. (D) Predicted disorder values are plotted over the full length of ETV1 (top), ETV4 (middle), and ETV5 (bottom). These values, calculated using Predictor of Naturally Disordered Regions (PONDR) VL3 (Radivojac et al. 2003), range from 0 (likely ordered) to 1 (likely disordered). Potential trypsin digestion sites are denoted by "X". Red lines refer to residues that span the ETS domain (ED), cyan lines in ETV4 refer to the NID and CID as identified for ETV4 (Figure 3-4).
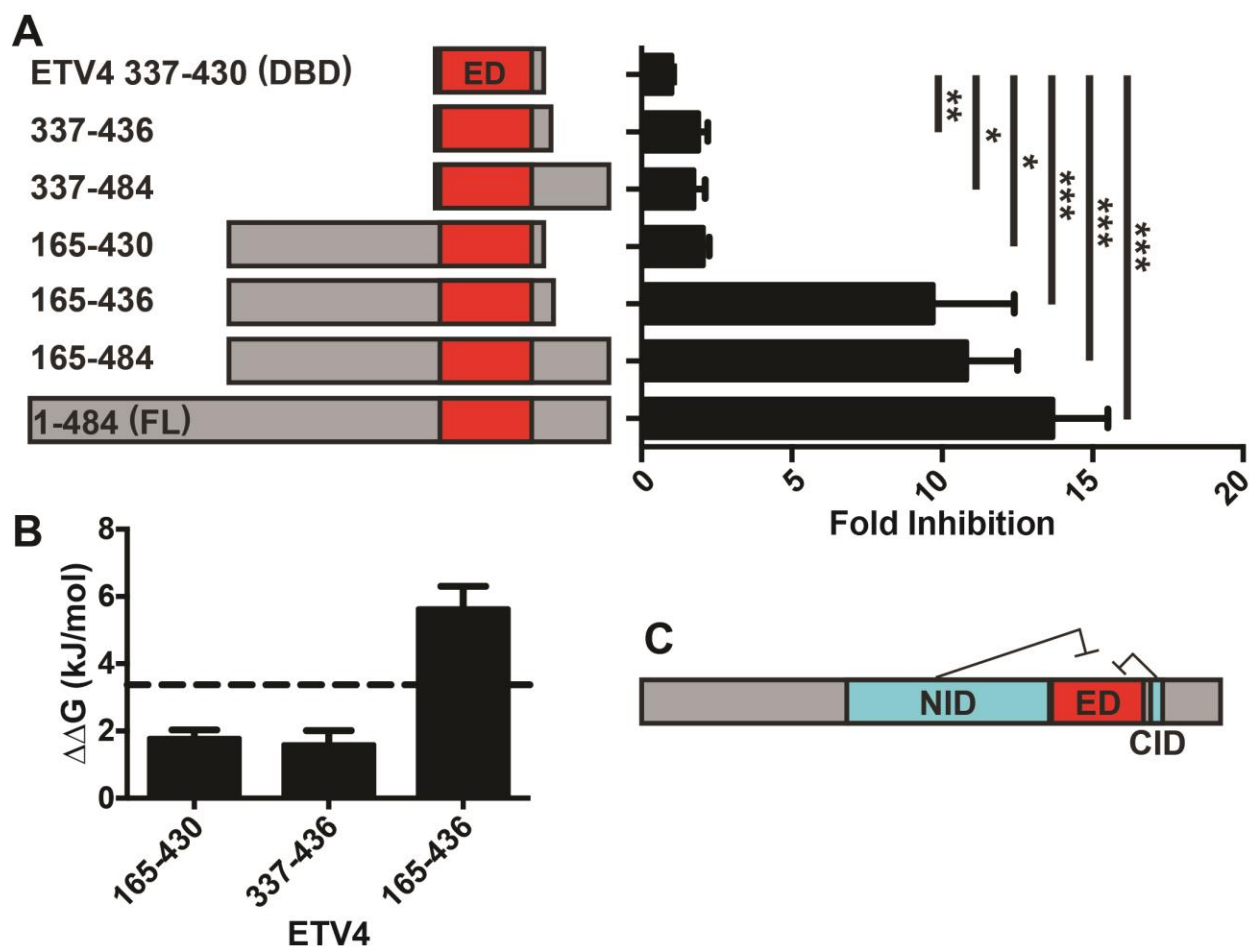
Figure 3-4 NID and CID cooperate to inhibit ETV4 DNA binding: Mapping autoinhibition through deletion analyses.

(A) Fold inhibition of the ETV4 fragments with mean and standard error of the mean displayed. Fold inhibition calculated as K_D (fragment) / K_D (337-430). "*", "**", and "***" indicate $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. See Table 3-2 for K_D values and numbers of replicates. (B) $\Delta\Delta G$ = RT ln [K_D ETV4 inhibited fragment / K_D ETV4$^{337-430}$] measured for fragments containing the NID (165-430), the CID (337-436), or both (165-436). The dotted line indicates the sum of the $\Delta\Delta G$ values for 165-430 and 337-436. (C) Schematic of ETV4 autoinhibition depicting cooperative inhibitory contributions from both the NID and CID, cyan. The ETS domain (ED) is noted in red.

Table 3-2 Equilibrium dissociation constants, K_D, and fold-inhibition values for ETV4 fragments

ETV4 Fragment	K_D (x10^{-11} M)[a]	Fold-inhibition[a,b]	p[c]	n
337-430 (DBD)	6.1 ± 0.6	1.0 ± 0.1	-	25
337-436	12 ± 2	2.0 ± 0.4	0.009	23
337-484	11 ± 2	1.8 ± 0.4	0.04	4
165-430	12 ± 1	2.1 ± 0.2	0.03	3
165-436	60 ± 10	10 ± 3	3 x 10^{-7}	11
165-484 (T1)	66 ± 9	11 ± 2	3 x 10^{-7}	18
1-484 (FL)	83 ± 8	14 ± 2	4 x 10^{-7}	35

[a] Mean and standard error of the mean are given for K_D and fold-inhibition values.
[b] ETV4$^{337-430}$ (DBD) , the uninhibited fragment, was used as a reference for calculating fold inhibition as K_D (fragment or full length) / K_D (ETV4$^{337-430}$).
[c] The p-values were calculated with ETV4$^{337-430}$ as the reference.

3.2.2. CID interactions perturb the DNA-recognition helix H3 to mediate autoinhibition

To elucidate the mechanism of autoinhibition by the CID, our collaborators used X-ray crystallography to determine the structures of the partially inhibited fragments of ETV1$^{332-435}$ and ETV4$^{337-441}$. These proteins contain both the ETS domain and the CID, as mapped in ETV4. Their structures were very similar with a root mean square deviation (RMSD) of 0.16 Å for backbone alignment of their ETS domains (Figure 3-5A,B and Figure 3-6A). The CID includes an α-helix, termed H4, which packs on one face of the ETS domain. In ETV4, Ala426 and Leu430 in H4 lie in a hydrophobic groove along the ETS domain in proximity to the conserved residues Trp344 from H1, Ile407 from the loop between H3 and β-strand S3, and Phe420 from S4 (Figure 3-5C). Homologous residues had similar interactions in ETV1. Replacing Leu430 with an alanine resulted in a reduction in autoinhibition (activation in DNA binding), whereas mutation to methionine, the

homologous amino acid in ETV1 and ETV5, does not affect DNA binding (Figure 3-5D). These structural and functional data demonstrated that the CID inhibits DNA binding through intramolecular contacts between H4 and the ETS domain, mediated in part by a leucine or methionine in this helix.
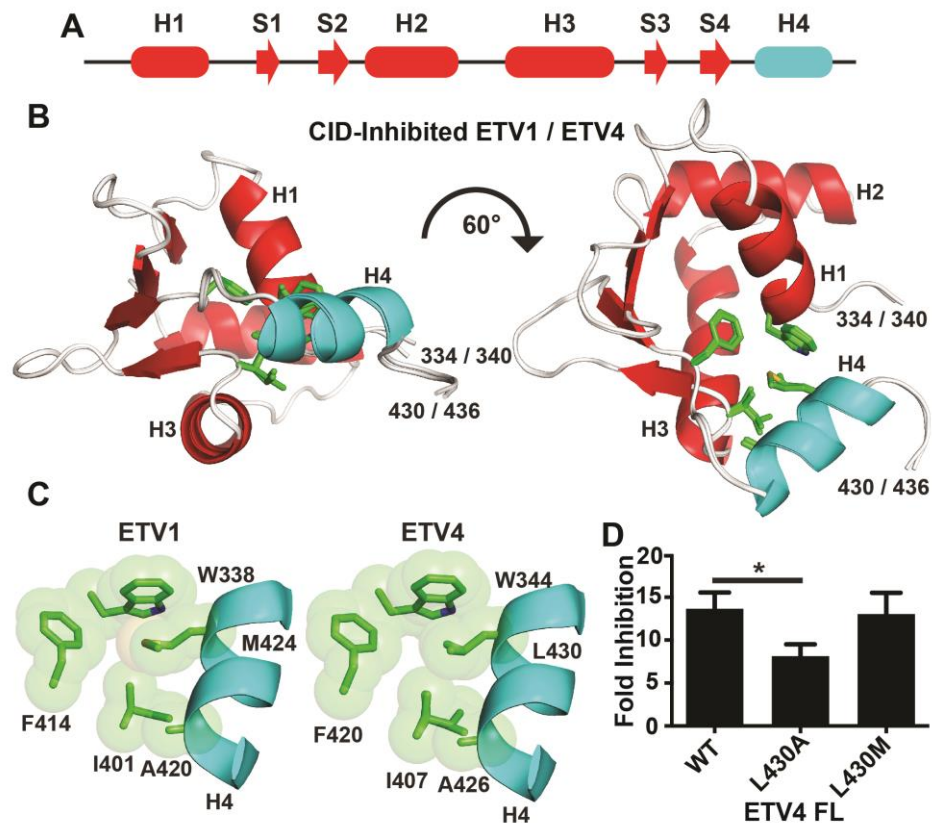
Figure 3-5 The CID inhibits DNA binding through hydrophobic contacts between α-helix H4 and the ETS domain.

(A) Schematic of ETS domain, H1-H3 and S1-S4, and α-helix H4 of ETV1, ETV4, and ETV5. ETS domain, red; inhibitory elements, cyan; α-helices, cylinders; β-strands, arrows. (B) Cartoon representations of the closely aligned structures for the ETS domain and CID of ETV1$^{332-435}$ and ETV4$^{337-441}$. Displayed in stick format are Ala426 and Leu430 from α-helix H4 in ETV4, and the analogous Ala420 and Met424 from ETV1, as well as the conserved amino acids in the ETS domain that form a hydrophobic cluster. Numbering for homologous amino acids and endpoints denoted as ETV1/ETV4. (C) Portions of the ETV1, left, and ETV4, right, structures, in van der Waals sphere format to show hydrophobic interactions between the ETS domain and H4. There is clear evidence for two conformations of the δ^1-methyl of Ile407 in the electron density map of ETV4. (D) Fold inhibition of ETV4 FL in its wild-type form, WT (n=35), or with point mutations L430A (n=11) or L430M (n=3). "*" Indicates $p < 0.05$.
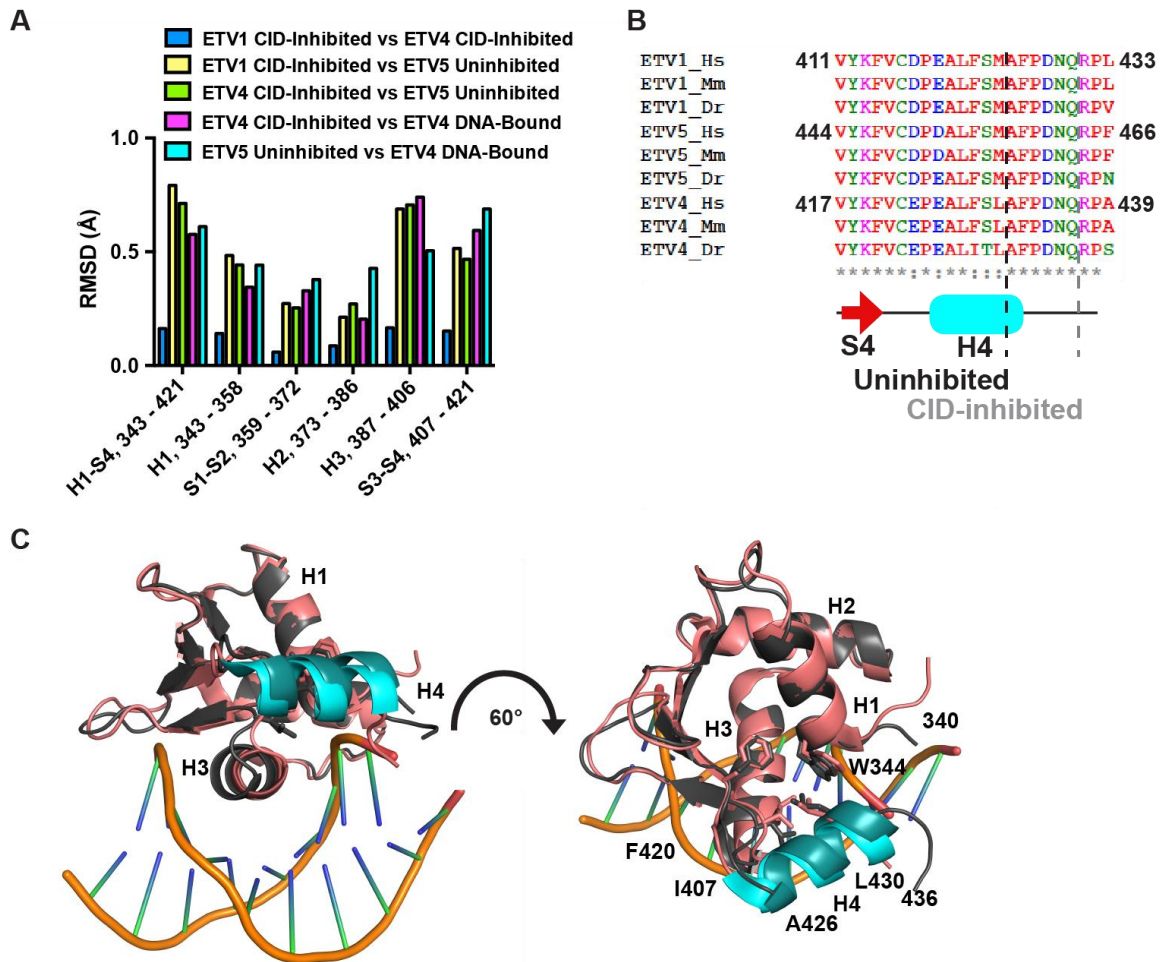
Figure 3-6 Structural comparison of CID-inhibited ETV1 and ETV4 with uninhibited ETV5 and DNA-bound ETV4.

(A) Root mean square deviations (RMSDs) were calculated for backbone atoms to compare the crystal structures of uninhibited ETV5[364-457], with CID-inhibited ETV1[331-435], and ETV4[337-441], and DNA-bound ETV4[337-441] (4UUV.pdb) (Cooper et al. 2015). Secondary structural elements are defined as in Figure 3-5 and the numbering on the x-axis refers to ETV4. For subsections of the entire structure (e.g., H1, 343-358), the different structures were realigned based on that particular subsection and RMSD values correspond to backbone atoms within that subsection. The CID-inhibited ETV1 and ETV4 structures are very similar and have low RMSD values. The ETS domain overall (H1-S4), as well as most subsections (H1, S1-S2, H2, and S3-S4), have similar RMSD values for the remaining comparisons. In contrast, the RMSD value for H3 is lower for the uninhibited ETV5 versus DNA-bound ETV4 comparison than for the CID-inhibited ETV1/ETV4 versus uninhibited ETV5 or the CID-inhibited ETV4 versus DNA-bound ETV4 comparisons. This indicates that H3 is more similar in the uninhibited and DNA-bound states than in the CID-inhibited state. (B) Sequence alignment of ETV1/4/5 helix H4 from *H. sapiens* (Hs), *M. musculus* (Mm), *and D. rerio* (Dr) colored according to Clustal Omega (Sievers et al. 2011). The red arrow and cyan cylinder indicate β-strand S4 of the ETS domain and α-helix H4, respectively. The vertical dashed black and gray lines identify truncation

endpoints that cause activation or retain CID inhibition, respectively. (C) CID-inhibited ETV4 in its free (this study) and DNA-bound forms (4UUV.pdb) (Cooper et al. 2015) were aligned based on the entire protein sequence. ETS domain and inhibitory residues are colored gray and dark teal, respectively, for the free ETV4 and pink and cyan, respectively, for the DNA-bound ETV4. Selected side chains are displayed in stick format as in Figure 3-5. Comparison with the free form demonstrates that there are subtle shifts of backbone atoms in the C-terminus of α-helix H3, as well as H4.

Table 3-3 X-ray crystallography data collection and refinement statistics

	ETV1 (332-435)	ETV4 (337-441)	ETV5 (364-457)
Data Collection			
Processing software	HKL2000	HKL2000	HKL2000
Beamline	SSRL 7-1	SSRL 7-1	SSRL 7-1
Wavelength	1.0000	1.0000	1.1271
Detector type	Q315 CCD	Q315 CCD	Q315 CCD
Collection date	2/7/13	2/7/13	1/12/13
Space group	$P3_121$	$P3_121$	$C222_1$
Unit cell	(50.2, 50.2, 69.3)	(50.9, 50.9, 68.6)	(57.5, 65.7, 53.0)
Resolution (Å)	55.00 - 1.40	45.00 - 1.10	30.00 - 1.80
Resolution (Å)	1.45 - 1.40	1.13 - 1.10	1.86 - 1.80
# Reflections measured	705,596	1,577,832	50,220
# Unique reflections	20,493	42,215	9,566
Redundancy	34.4	37.4	5.2
Completeness (%)	100.0 (100.0)[a]	100.0 (100.0)[a]	99.2 (97.3)[a]
$<I/\sigma I>$	16 (1.9)[a]	5 (0.9)[a]	9 (1.0)[a]
Mosaicity (°)	0.4	0.2	1.3
R(pim)	0.018 (0.243)[a]	0.020 (0.676)[a]	0.039 (0.363)[a]
Refinement			
Refinement software	PHENIX.REFINE	PHENIX.REFINE	PHENIX.REFINE
Resolution (Å)	30.0 - 1.40	45.00 - 1.10	30.0 - 1.80
Resolution (Å)	1.47 - 1.40	1.13 - 1.10	2.05 - 1.80
# Reflections used for refinement	20,457	42,112	8163
# Reflections in R_{free} set	967	1,988	410
R_{cryst}	0.157 (0.217)[a]	0.181 (0.361)[a]	0.186 (0.247)[a]
R_{free}	0.178 (0.237)[a]	0.201 (0.388)[a]	0.234 (0.285)[a]
RMSD: bonds (Å) / angles (°)	0.006 / 1.175	0.005 / 1.047	.008 / 1.456
$<B>$ (Å^2): All protein atoms / # atoms	16.1 / 890	16.5 / 1013	29.7 / 851
$<B>$ (Å^2): water molecules / # water	32.8 / 114	28.9 / 125	37.1 / 81
Ramachandran favored (%)	87.5	91.8	87.7
Ramachandran additionally allowed (%)	12.5	8.2	12.3
Protein Data Bank ID	5ILS	5ILU	5ILV

[a]Values in parentheses are for highest-resolution shell.
One crystal was used to measure the data for each structure.
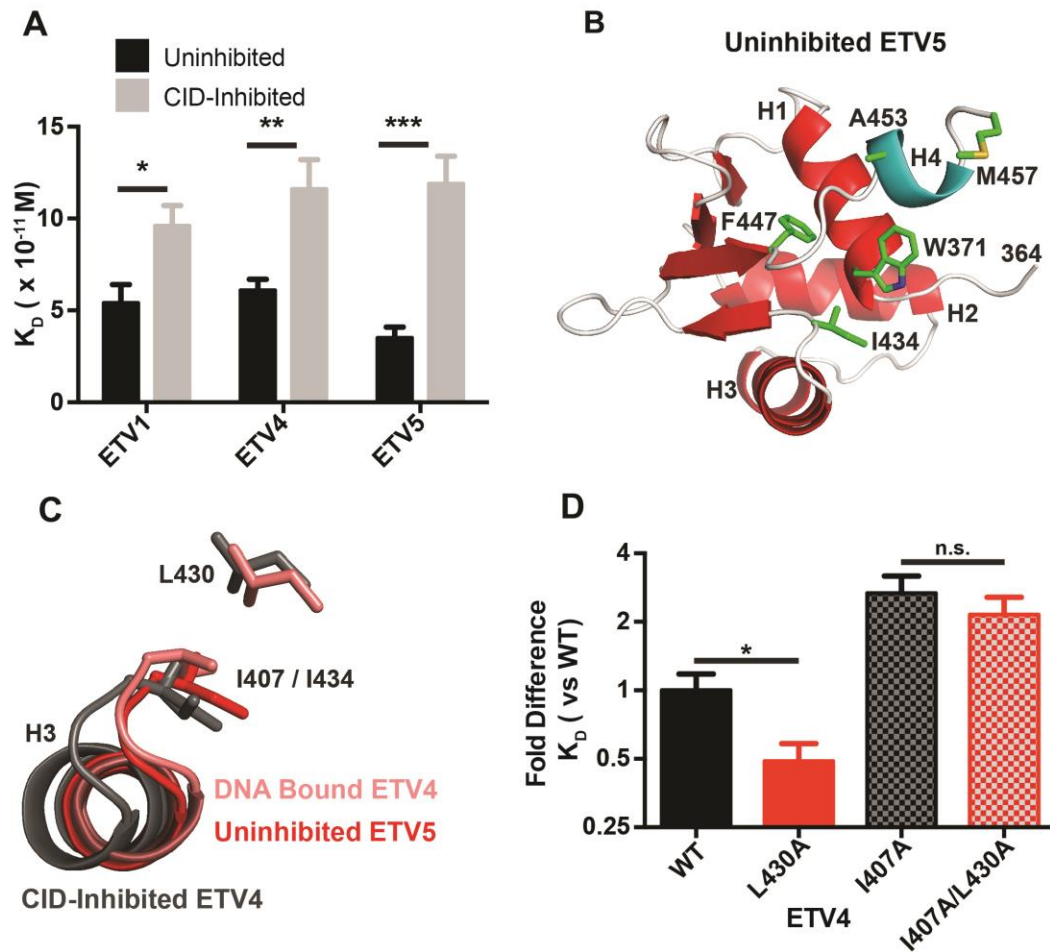
Figure 3-7 Interactions between the CID and the ETS domain affect DNA-recognition helix H3 positioning.

(A) Equilibrium dissociation constant, K_D, values for uninhibited ETV1[332-425]; n=6, ETV4[337-430]; n=25 and ETV5[364-457]; n=4 versus CID-inhibited ETV1[332-430]; n=7, ETV4[337-436]; n=23, and ETV5[364-463]; n=7. "*", "**", and "***" indicate $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. (B) Crystal structure of uninhibited ETV5[364-457], showing the truncated H4 and the same selected sidechains as in Figure 3-5. (C) H3 positioning from CID-inhibited ETV4 (gray), uninhibited ETV5 (red), and ETV4 bound to DNA (pink, 4UUV.pdb) (Cooper et al. 2015). Structures were aligned to DNA-bound ETV4 across the entire protein sequence (Figure 3-6C). Met457 of ETV5, the homologous residue to Leu430 in ETV4, is not in frame due to the repositioning of H4 in the uninhibited ETV5 crystal structure. (D) Comparison of K_D values for ETV4 FL in its wild-type form, WT (n=35), or with point mutations L430A (n=11), I407A (n=4), or both I407A and L430A (n=4). "*" Indicates $p < 0.05$ and "n.s." indicates $p > 0.05$. Fold difference for K_D values are relative to WT ETV4 FL.

81

Based on the crystal structures of CID-inhibited ETV1 and ETV4, we noted that the uninhibited, minimal DBD fragments used for demonstrating autoinhibition in ETV1, ETV4, and ETV5 were predicted to have a shorter or possibly unfolded helix H4 (Figure 3-6B). As with ETV4, loss of these homologous residues in ETV1 and ETV5 also activated DNA binding (Figure 3-7A). Therefore, an intact and full-length H4 is a necessary and conserved feature of the CID.

To understand the structural nature of the residues mapped to H4 within the context of uninhibited ETV1, ETV4, and ETV5, we attempted to crystalize these fragments with success for ETV5$^{364-457}$ (Figure 3-7B). Despite the deletion of amino acids mapped to the intact H4, the α-helix is retained, albeit truncated. However, the shorter H4 is rotated ~ 60° away from the ETS domain relative to the position of the full-length H4 in ETV1 and ETV4. This alternate position is accommodated in the crystal by intermolecular contacts between the truncated H4 and the ETS domain of a neighboring molecule (Figure 3-8). With H4 in this alternate position, Met457 is unable to form the intramolecular inhibitory contacts with the ETS domain observed for the homologous Met424 and Leu430 in the CID-inhibited structures of ETV1 and ETV4, respectively, potentially explaining the loss of autoinhibition of this fragment (compare Figure 3-5B and Figure 3-7B). In conclusion, the relief of autoinhibition by the partial truncation of H4 and by disruption of an intramolecular contact between H4 and the ETS domain demonstrated the role of H4 in autoinhibition. In addition, while the folding of truncated helix H4 and its alternate position is potentially a consequence of crystallization, we propose that this repositioning reflects an intrinsic mobility of the CID. This idea is supported by NMR spectroscopy studies, presented below.

To further our structural studies of ETV1/4/5, we compared our crystal structures of the uninhibited ETV5 with a truncated H4 to that of the CID-inhibited ETV1 and ETV4 with a full-length H4. In comparison to the highly similar CID-inhibited ETV1 and ETV4 structures (backbone RMSD of 0.16 Å), the ETS domain from uninhibited ETV5 was distinct with RMSD values of 0.79 Å and 0.72 Å when aligned to ETV1 and ETV4, respectively (Figure 3-6A). Closer examination of subsections of the ETS domain revealed that the differences

between uninhibited and CID-inhibited structures were most pronounced around the DNA-recognition helix H3, as well as β-strands S3 and S4. Visually, the backbone of the C-terminal half of the DNA-recognition helix H3 is shifted about 2 Å between the inhibited and uninhibited structures, relative to the rest of the ETS domain (Figure 3-7C). Further comparison with the previously reported structure of ETV4 in complex with DNA (Cooper et al. 2015) demonstrated that in the DNA-bound form, H3 of ETV4 is also shifted to a similar position as observed for uninhibited ETV5 (Figure 3-7C and Figure 3-6C). We speculate that in the ETV1/4/5 subfamily, the active state of a DNA-bound ETS domain requires this shift of H3 and, thus, matches the conformation of uninhibited ETV5 determined by X-ray crystallography.

Having observed the activation of the ETV4 mutant L430A (Figure 3-5D) and the variable positioning of the DNA-recognition helix H3 in our crystal structures (Figure 3-7C), we hypothesized that helix H4 inhibits DNA binding by modulating H3 through an interaction between Leu430 in H4 and Ile407 in the H3-S3 loop. We tested this postulate by mutating Ile407 and Leu430 to alanine, separately and in combination. The ETV4 mutant I407A had a reduction in DNA-binding affinity compared to the wild-type protein indicating that Ile407 contributes to DNA binding. Importantly, the I407A mutation also abrogated the activating nature of L430A in the double mutant I407A/L430A (Figure 3-7D). We conclude that H3 and the CID are coupled through the Ile407-Leu430 interaction and propose that CID-mediated autoinhibition functions by shifting a conformational equilibrium of H3 towards a state that is less competent for DNA binding.
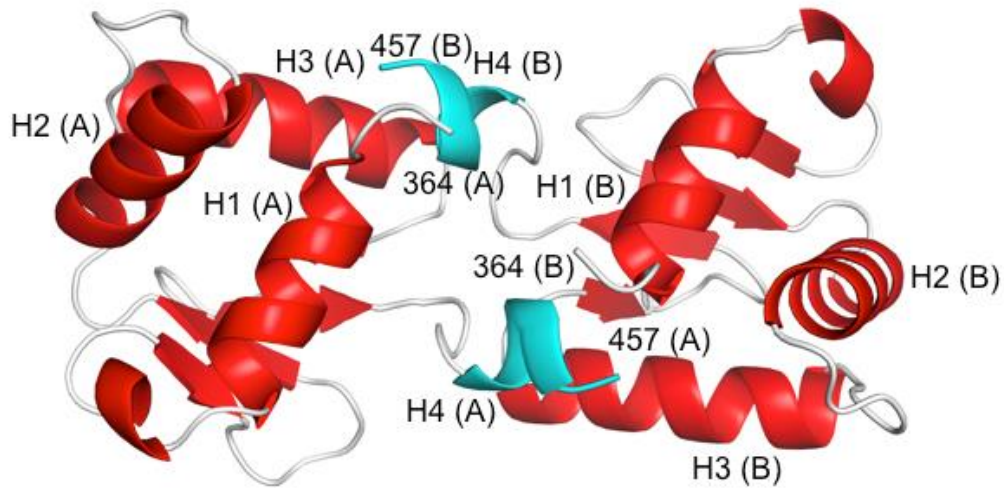
Figure 3-8 Crystal packing of uninhibited ETV5[364-457] influences the positioning of the truncated helix H4.

The labels (A) and (B) distinguish the two molecules of uninhibited ETV5. The contacts between (A) and (B) may affect the position of truncated α-helix H4 (cyan) as compared to the position in solution or in the intact H4 in inhibited ETV5.

3.2.3. Dynamic features of CID autoinhibition mechanism

To further investigate the CID mechanism of autoinhibition, we utilized NMR spectroscopy to compare uninhibited and CID-inhibited species (Figure 3-9A). Although differing slightly at their N- and C-terminal boundaries, the two ETV4 fragments displayed the same affinities for DNA as the corresponding species discussed above (Figure 3-10A). Based on mainchain chemical shifts, residues from the truncated H4 in the uninhibited ETV4[328-430] and the full-length H4 in the CID-inhibited ETV4[313-446] both adopted folded α-helical conformations under solution conditions (Figure 3-10B). However, relative to ETV4[313-446], the C-terminal residues in the shorter H4 of ETV4[328-430] exhibited reduced chemical shift-derived helical propensities and increased mobility as detected by amide ^{15}N relaxation measurements (Figure 3-10B and Figure 3-12). Nevertheless, these NMR spectroscopic studies indicate that the truncated α-helix H4 observed by X-ray crystallography is not an artifact of the crystallization process (Figure 3-7B).

A comparison of the $^{1}H^{N}$ and ^{15}N chemical shifts of the uninhibited and CID-inhibited species demonstrated that amides near the C-terminal end of H1 and H3, and throughout H4 were most affected by the presence of the full-length versus truncated H4 (Figure 3-9B). The amino acids in H3 that showed chemical shift perturbations match closely to those undergoing the backbone realignment observed in the comparison of the crystal structures of CID-inhibited ETV1 and ETV4 versus uninhibited ETV5. Thus, the interaction between H4 and the ETS domain, as well as the H4-dependent perturbations of H3, observed in the crystal structures are also retained in solution.

Additional NMR-monitored amide HX experiments were used to probe the dynamics of CID-inhibited ETV4 (Figure 3-11). Residues within H1, H2 and the β-sheet displayed relatively large protection factors (>10^{4}), indicating that they form the stable core of the ETS domain. In contrast, residues preceding the ETS domain and in loop regions had lower protection factors, as expected based on their solvent exposure and lack of any persistent hydrogen-bonded secondary structure. Most interestingly, many residues in the DNA-recognition α-helix H3 and in the inhibitory CID α-helix H4, displayed

intermediate protection factors (100-10,000) indicative of conformational dynamics to sample partially unfolded states detectable by HX. Similar behavior is observed with the DNA-recognition and inhibitory helices of ETS1 (Pufall et al. 2005; Lee et al. 2008) and ETV6 (Coyne et al. 2012; De et al. 2014). These NMR experiments extend our hypothesis that the CID autoinhibitory mechanism involves a conformational equilibrium involving interactions between helices H3 and H4, by demonstrating the dynamic nature of the helices.
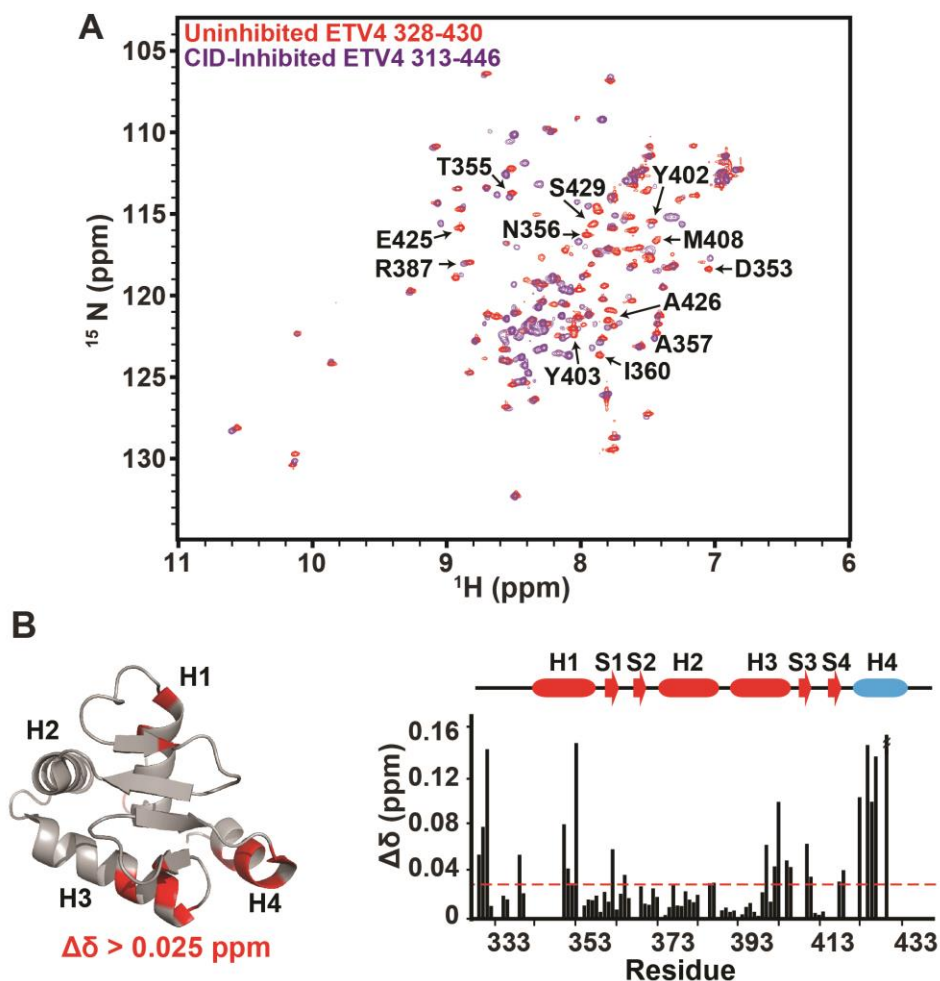
Figure 3-9 The CID perturbs the dynamic DNA-recognition helix H3.

(A) Overlaid ^{15}N-HSQC spectra of uninhibited ETV4$^{328-430}$; red, and CID-inhibited ETV4$^{313-446}$; purple. Selected assignments are indicated. Despite minor differences in the boundaries of these constructs, they bind to DNA with similar affinities as the previously described uninhibited and CID-inhibited fragments (Figure 3-10A). (B) Amide chemical shift perturbations ($\Delta\delta = [(\Delta\delta_H)^2 + (0.2\Delta\delta_N)^2]^{1/2}$) for corresponding residues in the spectra of (A) are plotted as a histogram and mapped onto the crystal structure of ETV4. Perturbed residues with $\Delta\delta > 0.025$ ppm (dashed line), are highlighted in red on the structure.
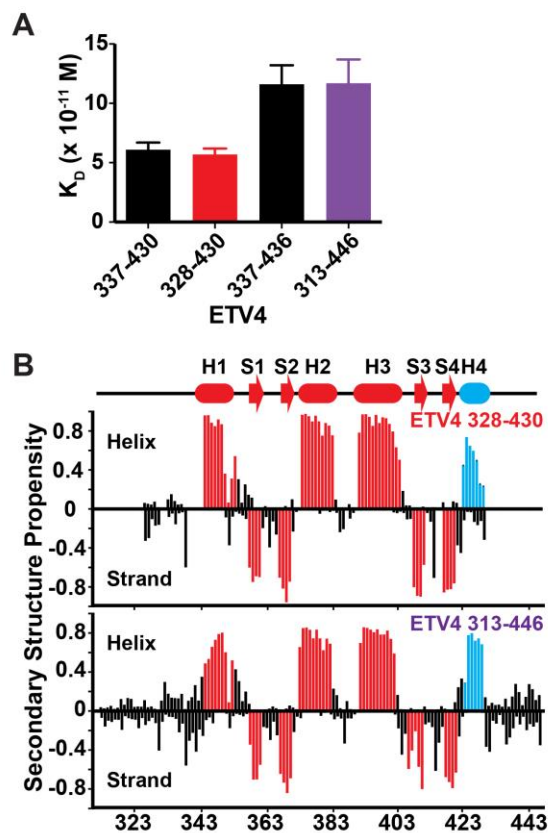
Figure 3-10 ETV4 fragments used for NMR spectroscopic studies have the same affinities for DNA and secondary structures as similar sized fragments used for X-ray crystallography.

(A) K_D values for the uninhibited (328-430, n=4, red) and inhibited (313-446, n=4, purple) ETV4 fragments used for NMR spectroscopy compared to those used for X-ray crystallography, black [n=25 and n=23 for ETV4[337-430] and ETV4[337-436], respectively]. (B) Secondary structure propensities for the two NMR-characterized ETV4 fragments calculated from their $^1H^N$, ^{15}N, $^{13}C^\alpha$, $^{13}C^\beta$, ^{13}CO chemical shifts using the algorithm MICS (Shen & Bax 2012). Helix, strand (shown as negative values), and coil (not shown) propensities sum to 1. Colored histogram bars identify amides in helices or strands of the ETS domain, red, and CID, cyan, as observed in the X-ray crystal structure of inhibited ETV4[337-441] (Figure 3-5B). Although truncated, residues corresponding to H4 still adopt a folded α-helical conformation when ETV4[328-430] is in the solution conditions used for NMR spectroscopic studies. However, the chemical shift-derived helical propensities of residues towards the C-terminus of the truncated H4 are reduced relative to the full helix in ETV4[337-446]. Amide ^{15}N relaxation measurements (Figure 3-12) also indicate that the C-terminal residues of ETV4[328-430] are more mobile than those in the N-terminal portion of the truncated H4.
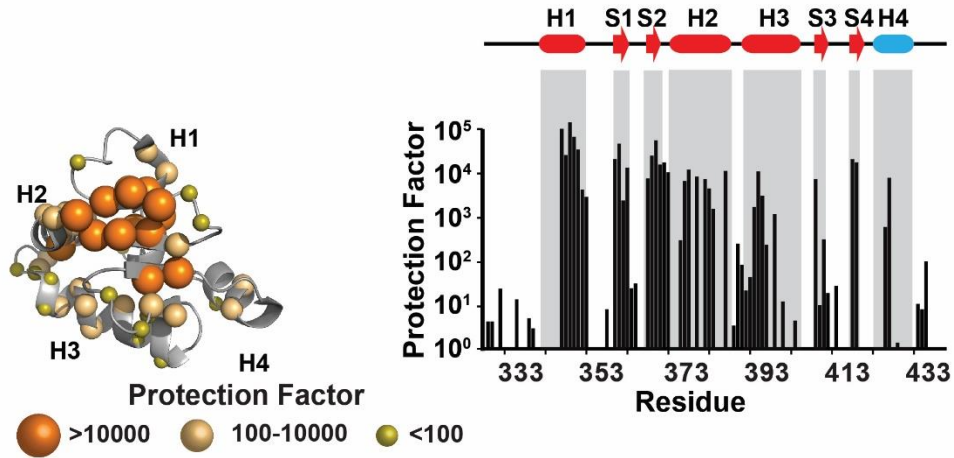
Figure 3-11 The DNA-binding helix H3 and CID are dynamics indicated by moderate amide HX protection factors.

Amide HX protection factors of ETV4[328-436], are plotted as a histogram and mapped onto the crystal structure of ETV4 using spheres with the indicated size/color scale. Although ETV4[328-436] is a combined name to denote data merged from HX studies of uninhibited ETV4[328-430] and CID-inhibited ETV4[337-436], the additional six C-terminal residues, which form the full CID, did not measurably change the protection factors of amides within the ETS domain. Missing values correspond to prolines, residues with unassigned or overlapped NMR signals, and residues exchanging too slowly to be measured with CLEANEX-PM (sec timescale), yet too fast to measure via ^{1}H/^{2}H exchange (> hours).
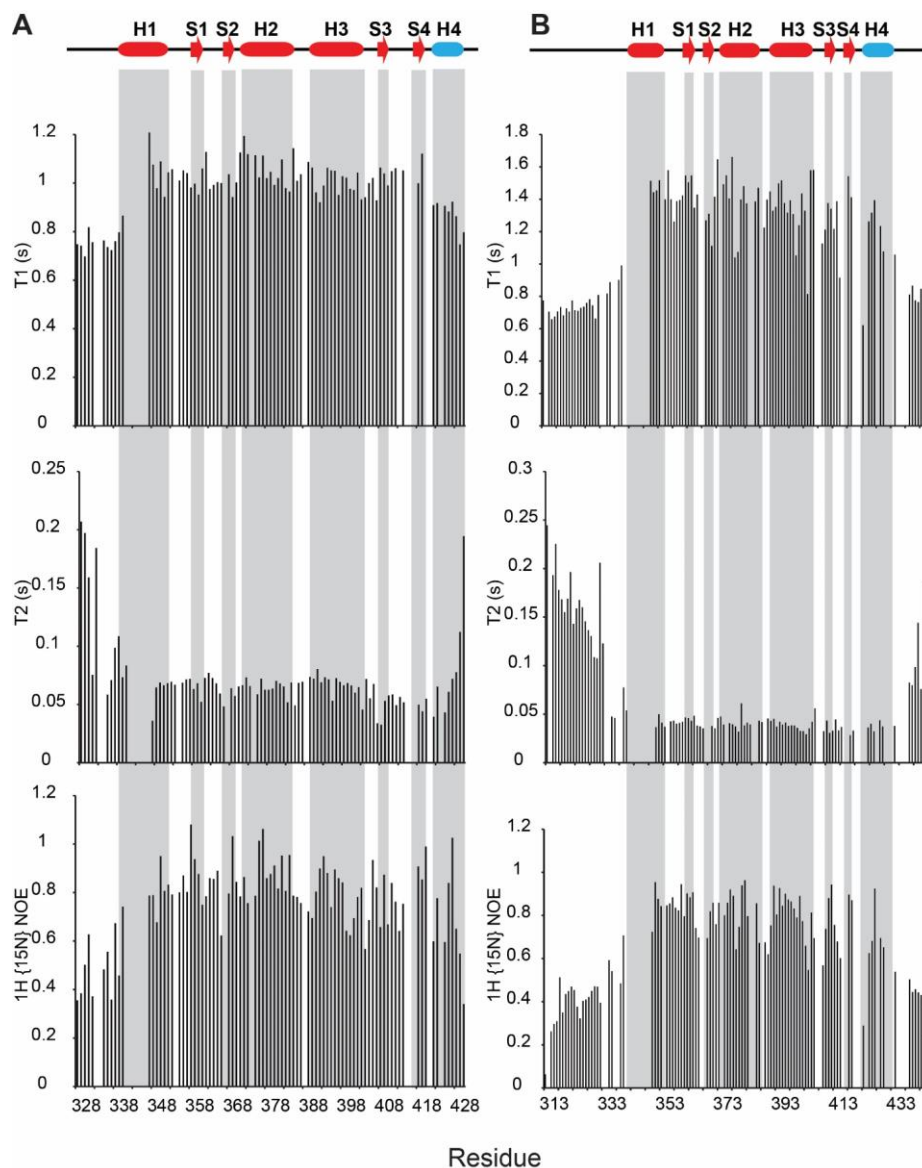
Figure 3-12 ¹⁵N amide relaxation indicates that the truncated CID is flexible.

Shown are amide relaxation data for (A) uninhibited ETV4$^{328-430}$ and (B) CID-inhibited ETV4$^{313-446}$. Top panel; T_1 relaxation, middle panel; T_2 relaxation, bottom panel; heteronuclear $^1H\{^{15}N\}$ NOE recorded with an 850 MHz NMR spectrometer. In contrast to the relatively uniform relaxation results for ordered core ETS domain, elevated T_2 lifetimes and reduced NOE values indicate that amides within loops and at the termini are conformationally mobile on the sub-ns timescale. The truncated CID in uninhibited ETV4$^{328-430}$ also shows increased mobility relative to the full CID in inhibited ETV4$^{313-446}$. The longer average T_1 and shorter average T_2 values of core residues ETV4$^{313-446}$ versus ETV4$^{328-430}$ indicates slower tumbling of the larger protein fragment. Fitting of these data with the Model-Free formalism yielded global rotational diffusion (tumbling) correlation times of 11 ns for ETV4$^{328-430}$ (12.2 kDa) and 17 ns for ETV4$^{313-446}$ (15.7 kDa).

3.2.4. Inhibitory properties of the NID map to intrinsically disordered sequences

As the next step towards a mechanistic understanding of the ETV4 autoinhibition, we investigated the NID using biophysical approaches. We initially compared the ^{15}N HSQC of several truncation fragments of the ETV4 encompassing different lengths of the NID, as well as the sequences following the CID. The latter were included to ensure full autoinhibition. The minimal ETS domain fragment (ETV4$^{328-430}$) gave a dispersed ^{15}N-HSQC spectrum, indicative of a well-folded protein (Figure 3-13). Longer fragments (ETV4$^{313-446}$, ETV4$^{295-484}$, and ETV4$^{266-484}$) produced reasonable NMR spectra, yet tended to precipitate heavily over time. More importantly, the ^{15}N-HSQC spectra of ETV4$^{295-484}$ and ETV4$^{266-484}$ contained many overlapping peaks with limited ^{1}H^{N} dispersion. These data hinted strongly that the NID is intrinsically disordered and predominantly adopts random coil conformations. This result was not unexpected, given the predicted disordered nature of the NID (Figure 3-3D) and the fact that efforts to crystallize fragments of ETV4 encompassing the NID proved unsuccessful.

To circumvent the challenges in interpreting the complex NMR spectra of large inhibited ETV4 fragments, an alternative strategy was pursued. First, the isolated NID (ETV4$^{165-336}$) was expressed and characterized. The ^{15}N-HSQC spectrum of this polypeptide displayed limited ^{1}H^{N} chemical shift dispersion, yet was amenable to assignment via standard ^{1}H/^{13}C/^{15}N correlation experiments (Figure 3-14A, left). An analysis of its assigned mainchain ^{1}H, ^{13}C, and ^{15}N chemical shifts confirmed that the isolated NID predominantly samples random coil conformations and thus is indeed an IDR (Figure 3-15). Circular dichroism spectroscopy added additional evidence for the overall disordered character of this species (data not shown).
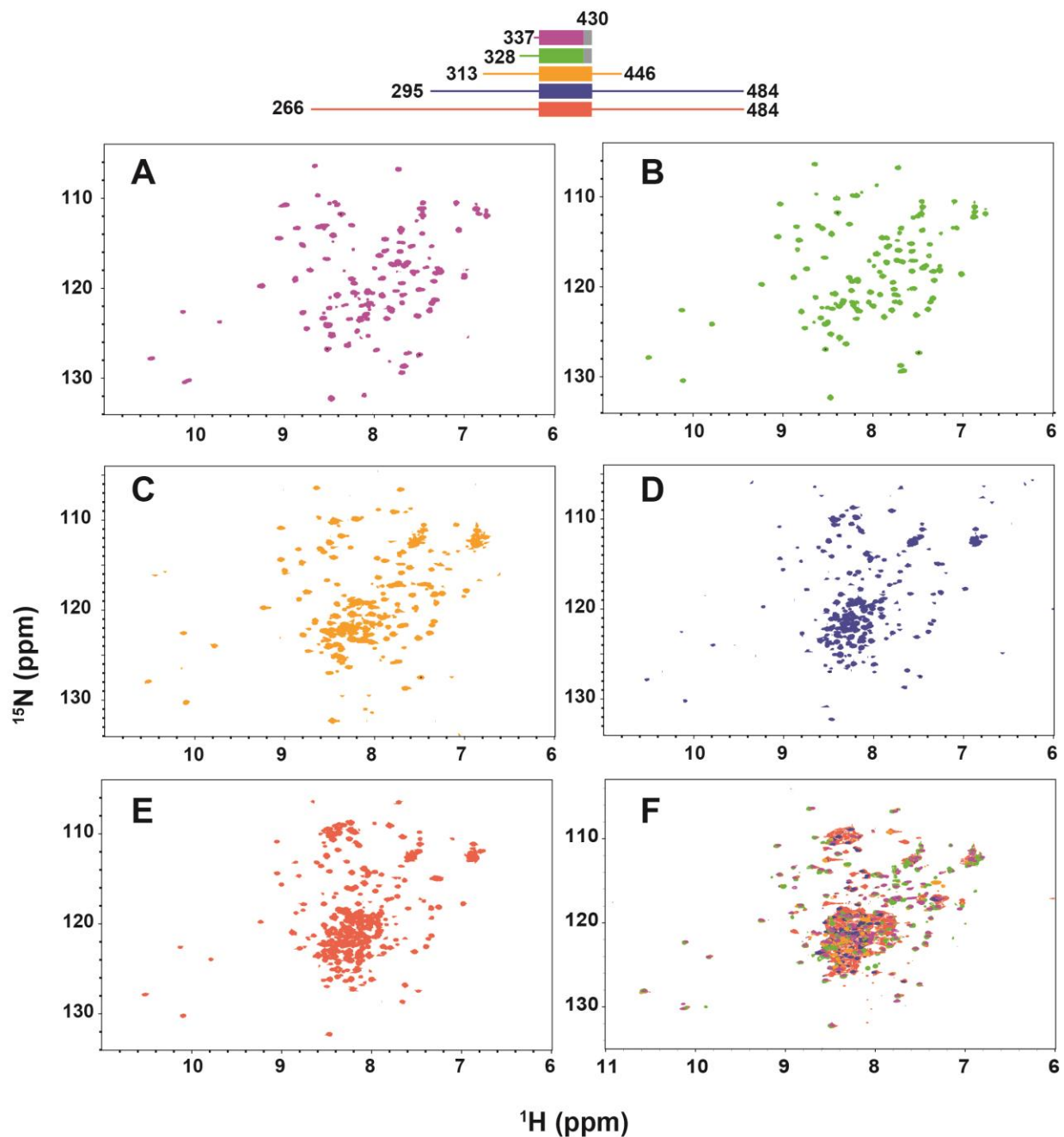
Figure 3-13 NMR spectroscopic characterization of ETV4 deletion fragments.

Shown are the separate (A-E) and superimposed (F) ^{15}N-HSQC spectra of a series of ETV4 deletions fragments. The progressive addition of residues flanking the ETS domain and CID results in an increased number of amide signals with random coil ^{1}H^{N} chemicals shifts. This indicates that these residues are predominantly disordered.
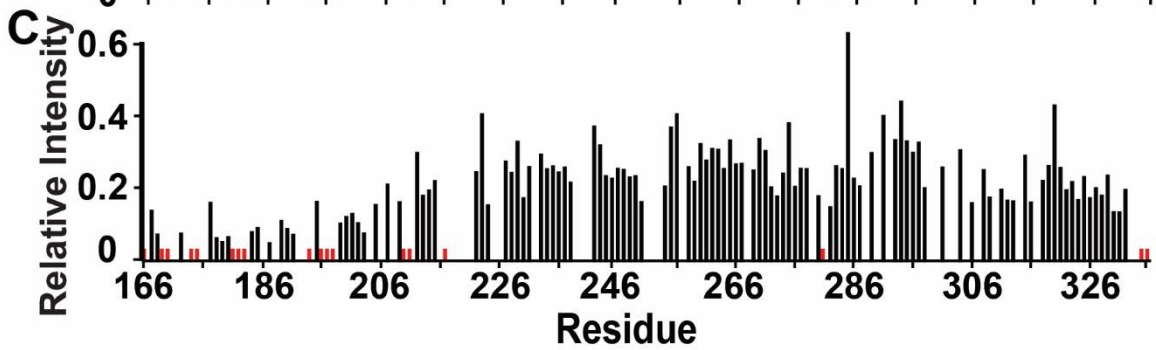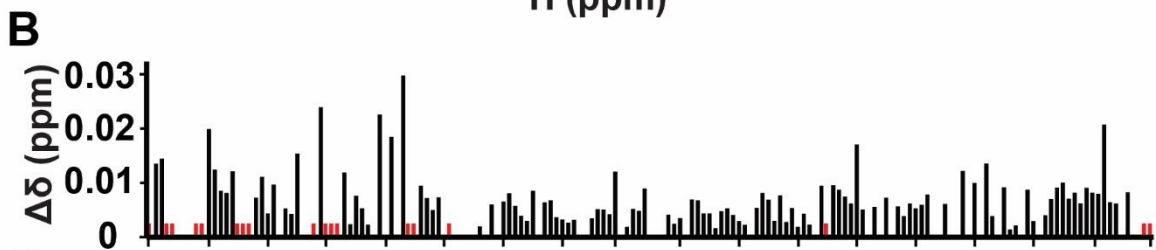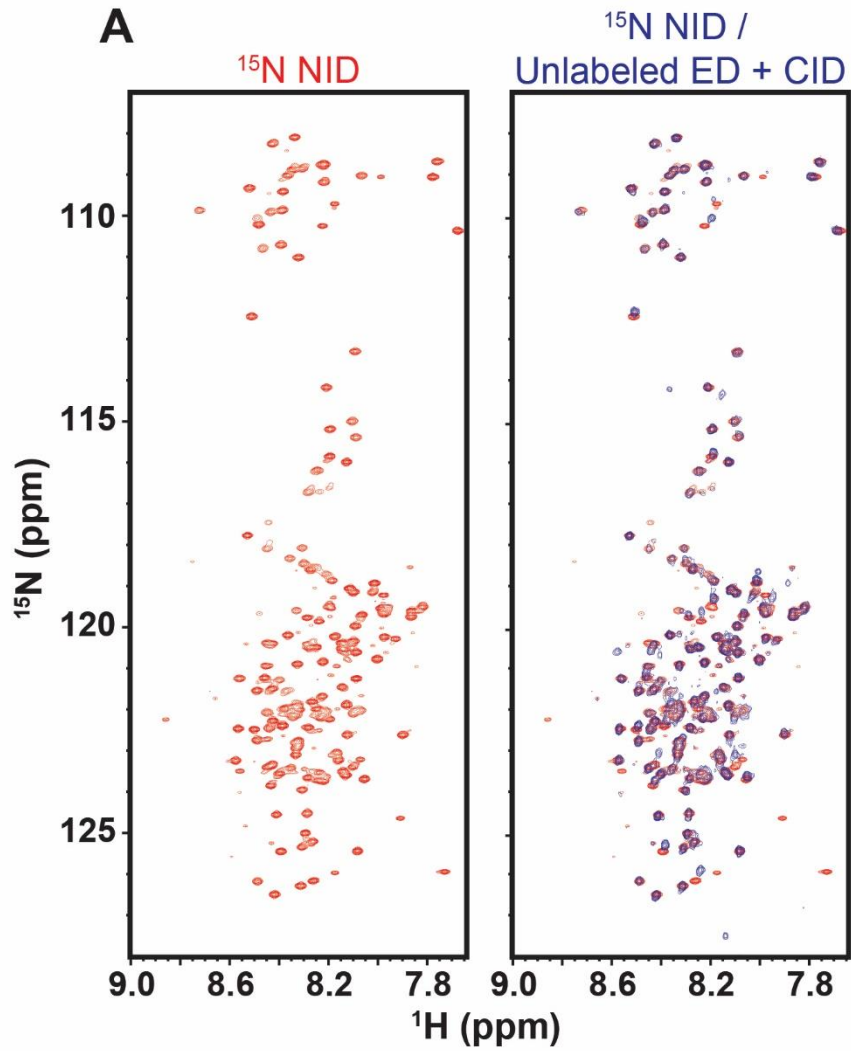
Figure 3-14 The NID is intrinsically disordered whether in isolation or linked "in cis" to the ETS domain and CID.

(A) Both panels show the ^{15}N-HSQC spectrum of the isolated NID (ETV4$^{165-336}$) in red. The right panel also shows the overlapped spectrum of the ^{15}N-labled NID intein-ligated to the unlabeled ETS domain and CID (337-436) in blue. Consistent with its limited ^{1}H^N dispersion, analysis of the assigned main chain chemical shifts (^{1}H^N, ^{15}N, ^{13}C$^\alpha$, ^{13}C$^\beta$, ^{13}CO) of the isolated NID with the algorithms CSI 2.0 (Hafsa & Wishart 2014) and δ2D (Camilloni et al. 2012) confirms that it is an IDR (Figure 3-15). Comparisons of ^{15}N-HSQC amide chemical shift perturbations ($\Delta\delta = [(\Delta\delta_H)^2 + (0.2\Delta\delta_N)^2]^{1/2}$), (B), and relative peak intensities, (C), for the NID alone versus attached to the ETS domain and CID. The simple peak intensity ratios were not normalized for differences in sample concentration and spectral acquisition times. Amides broadly localized near the N-terminus of the NID showed small chemical shift and relative intensity perturbations due to the ETS domain and CID "in cis". The ^{15}N-HSQC spectrum of the intein-ligated species was assigned by comparison with that of the isolated NID, and red bars indicate amide signals that could not be confidently identified in both proteins. Missing histogram bars correspond to prolines and unassigned amides.

A

```
166  SGTSQPHPGHGYLGEHSSVFQQPLDICHSFTSQGGGREPLPAPYQHQLSE  215
166  CCCCCCCCCCCCCCCCCCCCCCCCCCCHHHCCCCCCCCCCCCCCCCCCCCC  215

216  PCPPYPQQSFKQEYHDPLYEQAGQPAVDQGGVNGHRYPGAGVVIKQEQTD  265
216  CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC  265

266  FAYDSDVTGCASMYLHTEGFSGPSPGDGAMGYGYEKPLRPFPDDVCVVPE  315
266  CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC  315

316  KFEGDIKQEGVGAFREGPPYQ  336
316  CCCCCCCCCCCCCCCCCCCCC  336
```

B
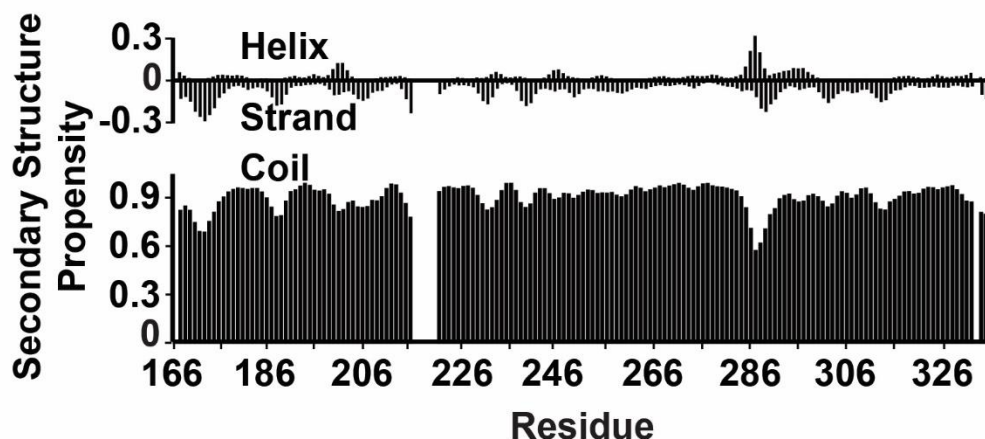
Figure 3-15 The secondary structure propensity revealed that the NID is intrinsically disordered.

(A) Secondary structure propensities calculated from the assigned mainchain chemical shifts ($^1H^N$, ^{15}N, $^{13}C^\alpha$, $^{13}C^\beta$, ^{13}CO) of the isolated NID (ETV4[165-336]) with CSI 2.0 (C, coil; H, helix) (Hafsa & Wishart 2014). (B) Normalized secondary structure propensities for α-helical (top, positive values), β-strand (top, negative values) and random coil or polyproline-II conformations (bottom), calculated from main chain chemical shifts using δ2D (Camilloni et al. 2012). Although differing in scoring criteria and output format, both algorithms reveal that the NID is predominantly disordered.

Many IDRs, while disordered in isolation, take on a more structured character in the presence of a binding partner through a coupled "folding and binding" mechanism (Dyson & Wright 2002). Therefore, as the second step of our analysis strategy, we asked whether the NID is still disordered in the presence of the ETS domain and CID. To address this, we used intein and sortase ligation technologies to covalently link the ^{15}N-labeled NID (ETV4$^{165-336}$) to an unlabeled ETS domain and CID of ETV4$^{337-436}$. We confirmed that this ligated fragment displayed comparable autoinhibition to the native protein fragment (Figure 3-16). The NID spectrum retained limited ^{1}H^{N} chemical shift dispersion, indicating the lack of any detectable structure induced upon covalent-linkage to the ETS domain and CID (Figure 3-14A). Although small changes in the chemical shifts or relative intensities of the ^{1}H^{N}-^{15}N signals from amides localized near the N-terminus of the NID were observed (Figure 3-14B,C), no obvious segment of residues interacted with the ETS domain and/or CID of ETV4 with sufficient affinity to adopt a persistent conformation detectable by NMR spectroscopy. Thus, even when the NID is attached "in cis" to the ETS domain and CID, it remains predominantly disordered.

In parallel, we interrogated which regions of the NID are important for inhibition. A truncation series indicated that the progressive inclusion of residues N-terminal to the ETS domain provided progressively greater autoinhibition of DNA binding (Figure 3-17). The region spanning residues 203-287 contributed the largest effect of autoinhibition, but other regions of the NID also contributed towards the overall inhibitory effect. The lack of a clear boundary for the inhibitory residues is consistent with the intrinsic disorder of the NID and the absence of any identifiable cluster of residues that interact strongly with the ETS domain and CID.
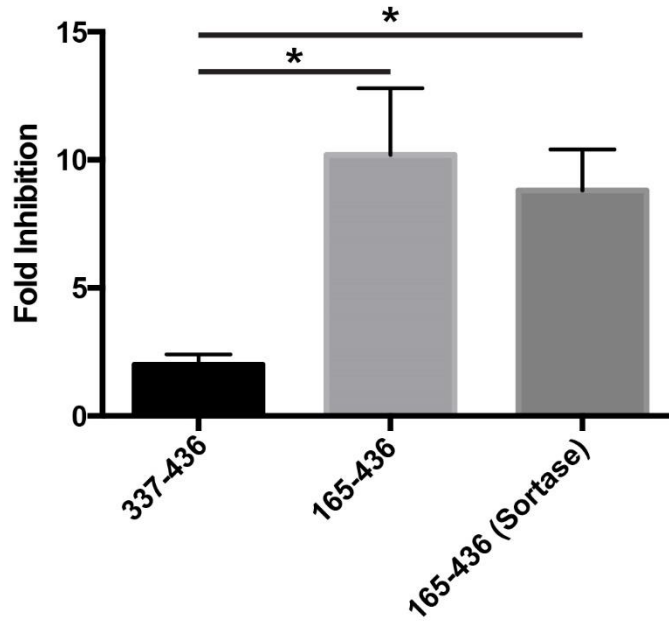
Figure 3-16 Sortase-linked ETV4$^{165\text{-}436}$ retains autoinhibition.

Fold inhibition for ETV4$^{337\text{-}436}$; ED + CID versus (ETV4$^{165\text{-}436}$; NID + ED + CID) expressed as a single protein or generated by sortase ligation of independently expressed (165-336) and (337-436) fragments. "*" indicates $p < 0.05$. These data indicate that the process of sortase ligation does not disrupt the autoinhibition of ETV4$^{165\text{-}436}$.
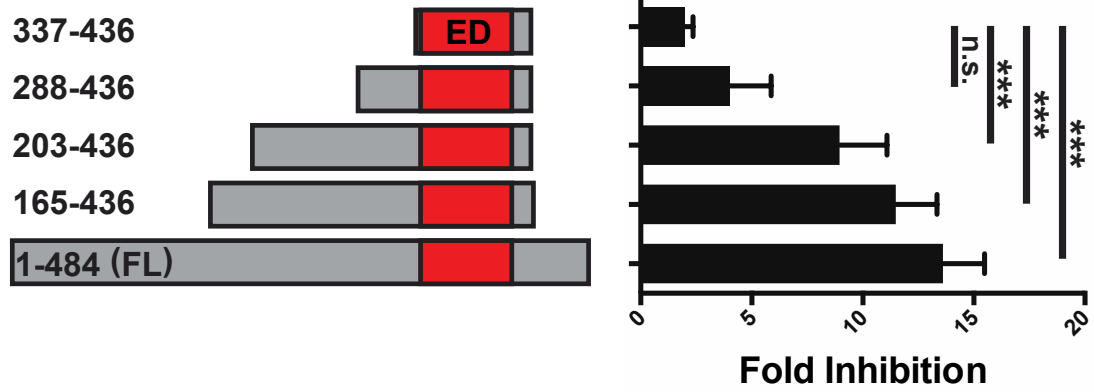
Figure 3-17 Multiple regions within the NID contribute to the autoinhibition of ETV4.

Fold-inhibition values for ETV4 fragments with various N-terminal truncations of the NID. Fold inhibition was calculated by comparing proteins to uninhibited ETV4$^{337-430}$, as in Figure 3-1A and Figure 3-4A. Bars and error bars refer to the mean and the standard error of the mean for the following number of replicates: ETV4$^{337-436}$, n = 23; ETV4$^{288-436}$, 8; ETV4$^{203-436}$, 8; ETV4$^{165-436}$, 11; ETV4$^{1-484}$, 35. "***" indicates $p < 0.001$ and n.s. indicates $p > 0.05$.

3.2.5. Intramolecular interactions of NID with the ETS domain and CID

To define the possible interactions of the NID with the ETS domain and CID, we initially prepared a set of polypeptides corresponding to NID deletion fragments (ETV4[312-336], ETV4[295-336], and ETV[165-336]). These unlabeled species were then used for ^{15}N-HSQC-monitored titrations with a sample of ^{15}N-labeled ETS domain (ETV4[337-436]). However, none caused any meaningful spectral perturbations (not shown). Thus, "in trans", the NID fragments do not detectably interact with the ETV4 ETS domain.

However, within the context of native ETV4, the NID and ETS domain are covalently linked as a continuous polypeptide chain. Therefore we used intein technology to ligate unlabeled NID (165-336) to a ^{15}N-labeled fragment containing the ETS domain and CID (337-436). This served to reconstruct the fully inhibited species while retaining the simplified spectrum assigned previously for isolated ETV4[337-436] (Figure 3-18A). In addition to the expected changes at the N-terminus of H1, the ligated NID weakly perturbed the ^{1}H^{N}-^{15}N signals of amides in H2, the C-terminal region of H3 and the surface-exposed face of the CID (Figure 3-18B,C). These data suggest that the NID may inhibit DNA binding by transiently interacting with the DNA-recognition helix H3 and/or by interacting with and reinforcing the inhibitory position of the CID.

The largest chemical shift perturbations, besides those near the N-terminus ligation site, were observed for Tyr401 and Tyr403 on the DNA-recognition helix H3. Therefore, we tested whether these tyrosine residues are functionally important for autoinhibition by serving as an interaction site within the ETS domain for the NID. Tyr401 and Tyr403 were mutated to asparagine and glycine, respectively, in the CID-inhibited fragment of ETV4[337-436]. These alternative residues were chosen due to their presence in homologous positions in other ETS factors, suggesting that there would be less chance of structural perturbations. However, due to their position at the DNA interface, mutation of both residues substantially impaired DNA binding (Figure 3-19A). Importantly, full-length ETV4 and the CID-inhibited fragment of ETV4 with Y401N and Y403G substitutions had identical affinity for DNA, indicating a loss of NID-mediated inhibition (Figure 3-18D and

Figure 3-19A). These data indicate that Tyr401 and Tyr403 are required for NID-mediated inhibition and may serve as part of a transiently occupied interface. Due to the role of these two tyrosines in direct contact to DNA, this NID intramolecular interaction may be acting sterically by masking the DNA interface.

Another potential NID-interaction site was mapped to the surface of the CID. However, mutation of several glutamate or phenylalanine residues along the surface-exposed face of the CID (E423K, E425K or F428A/F432A) did not affect NID-mediated inhibition (Figure 3-19B). Therefore, two possibilities exist to explain the NMR spectroscopically detected perturbation of the CID by the NID. The potential NID-CID interface may be sufficiently broad such that it is resilient to individual mutations, or the CID and the NID indirectly interact via the direct interactions between the NID and H3, and between H3 and the CID.
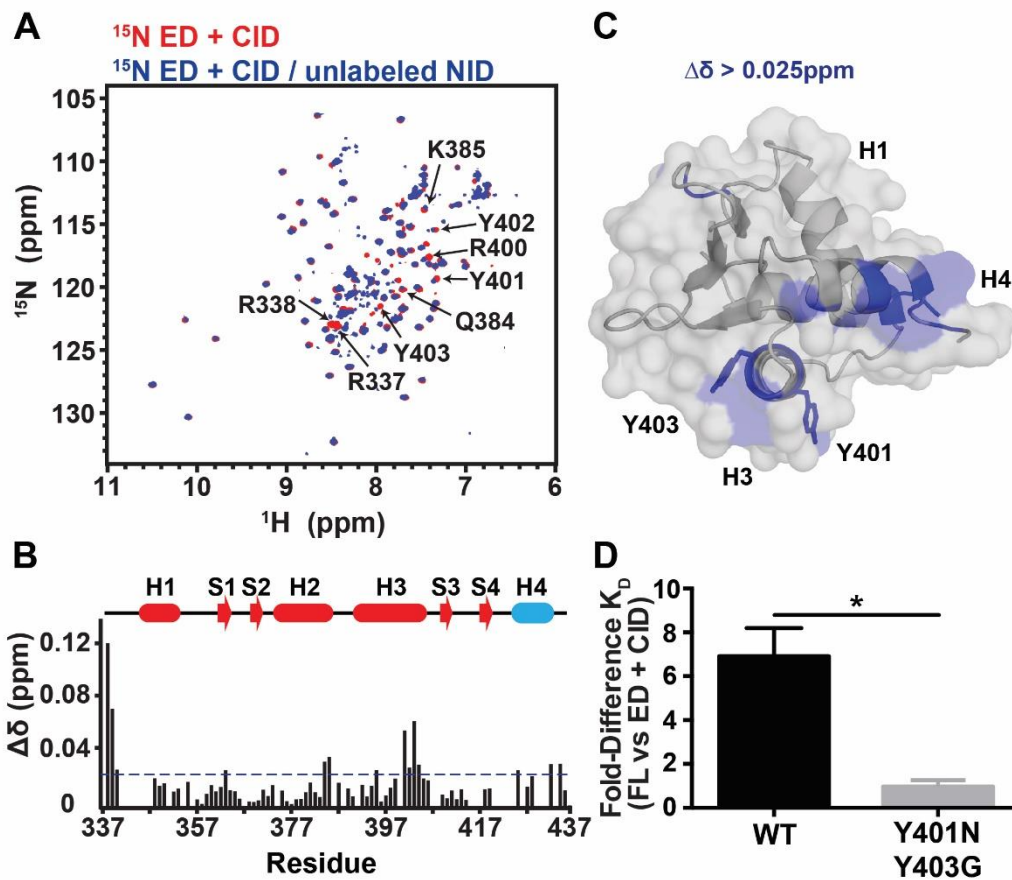
Figure 3-18 The NID interacts with the CID and the DNA-recognition helix H3

(A) Overlaid ^{15}N-HSQC spectra of ^{15}N-labeled ETV4 ETS domain and CID alone (337-436, red), and with the unlabeled NID (165-336, blue) joined via intein ligation. Selected peaks are labeled. (B, C) The amide chemical shift perturbations, ($\Delta\delta= [(\Delta\delta_H)^2 + (0.2 \Delta\delta_N)^2]^{\frac{1}{2}}$), resulting from the ligated NID are displayed in histogram format and mapped onto the structure of ETV4 ETS domain and CID (blue, $\Delta\delta > 0.025$ ppm; grey, $\Delta\delta < 0.025$ ppm, prolines, or residues with unassigned NMR signals). (D) Fold difference of K_D values between full length (FL: residues 1-484) and CID-inhibited (ED + CID: 337-436) ETV4 for wild type proteins or proteins with both Y401N and Y403G mutations. Bars and error bars represent the mean and the standard error of the mean for the following number of replicates: FL Y401N/Y403G, 6; ED + CID Y401N/Y403G, 5. "*" indicates $p < 0.05$.
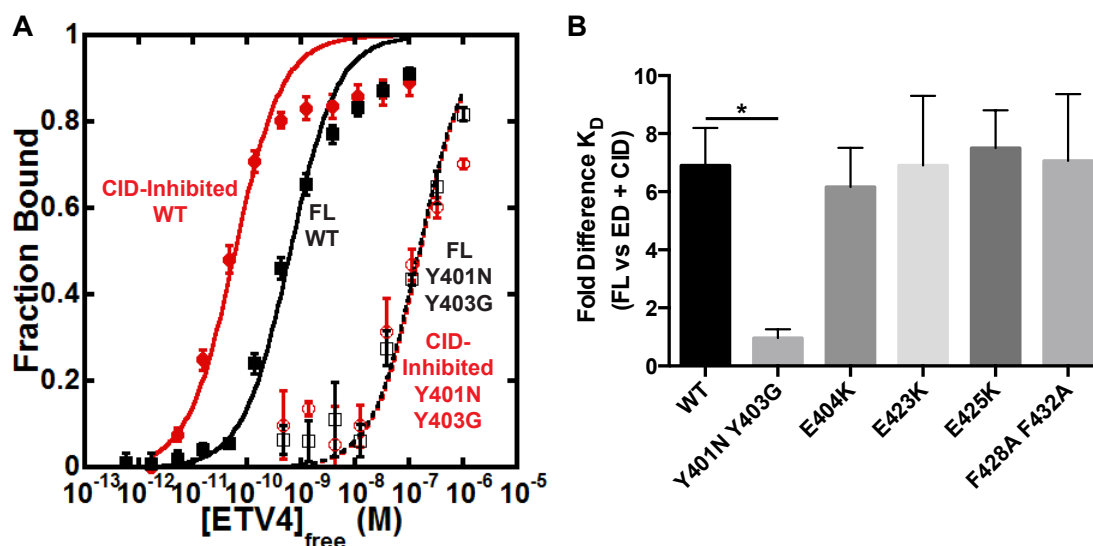
Figure 3-19 Tyr401 and Tyr403 in H3 are required for NID-mediated inhibition.

(A) Binding isotherms of CID-Inhibited ETV4 (red: 337-436) and full-length ETV4 (black: 1-484), solid lines and data points, and the same fragments with Y401N and Y403 point mutations, dashed lines and open data points. Data points and error bars represent the mean and the standard error of the mean for the following number of replicates: CID-inhibited, 7; FL, 14; CID-inhibited Y401N/Y403G, 3; FL Y401N/Y403G, 4. Mutating Tyr401 and Tyr403 weakens the affinity for DNA (compare CID-inhibited and CID-inhibited Y401N/Y403G). Importantly, these mutations also relieve autoinhibition from the NID (compare CID-inhibited Y401N/Y403G and FL Y401N/Y403G). (B) Fold difference of K_D values comparing full-length ETV4 (FL) and CID-inhibited ETV4 (337-436, ED + CID) for wild type (WT) proteins and with the indicated point mutants. Data points and error bars represent the mean and the standard error of the mean for the following number of replicates: WT, 23; Y401N Y403G, 5; E404K, 4; E423K, 4; E425K, 4; F428A F432A, 6. "*" indicates $p < 0.05$. These data demonstrate that Tyr401 and Tyr403 are required for NID-mediated autoinhibition and suggest that these residues are critical interaction sites for the NID. Although not altering autoinhibition, the E404K mutation strengthens DNA binding equally for both FL and CID-inhibited ETV4, whereas the E423K, E425K, and F428A/F432A substitutions have no effect on DNA binding (data not shown).

3.2.6. Probing transient NID interactions using paramagnetic relaxation enhancement experiments

The interactions between the NID and the ETS domain/CID led to rather modest NMR spectral perturbations and thus a rather coarse mapping of interfacial residues. Accordingly, I utilized paramagnetic relaxation enhancement (PRE) approaches as a potentially more sensitive method to detect the weak association of these ETV4 domains (Battiste & Wagner 2000). Initially, a MTSL nitroxide spin label was covalently linked to the single cysteine (Cys422) in ETV4[313-446], a partially inhibited species containing the CID and a truncated NID. Due to the presence of the spin label, the ^{15}N-HSQC signals of ETS domain and CID amides that are spatially near Cys422 (< 20 Å or I_{para}/I_{dia} ratio < 0.4) showed lower intensities relative to a control spectrum recorded after reduction of the paramagnetic nitroxide to a diamagnetic hydroxylamine and/or cleavage of the disulfide linkage joining the MTSL to the cysteine sidechain (Figure 3-20A). These PRE effects are consistent with the crystal structure of the CID-inhibited ETV4 ETS domain. Most importantly, the NID amides were not markedly perturbed, indicating that NID residues are not persistently localized near Cys422.

In a complementary set of experiments, amino acid 422 was mutated from cysteine to serine, and a lone cysteine was introduced at position 312 (M312C/C422S) in the truncated NID. As shown in Figure 3-20B, the presence of the spin label in the NID enhanced relaxation of ETS domain/CID amides within a broadly localized surface spanning the strand regions, as well as the end of helix H3 and most of helix H4. This surface generally matched that mapped in Figure 3-18 via amide ^{1}H^N-^{15}N chemical shift perturbations resulting from ligation of the full length NID on an ETS domain and the CID. It is also noteworthy that the spin-label did not completely eliminate the signals from amides within this surface, as would be expected if this portion of the NID containing Cys312 bound with a long lifetime to a well-defined position on the ETS domain. Collectively, these PRE experiments support our conclusion that the NID transiently interacts with a broad surface of the DNA-binding interface and thereby leads to a steric mechanism of autoinhibition (Figure 3-20B).
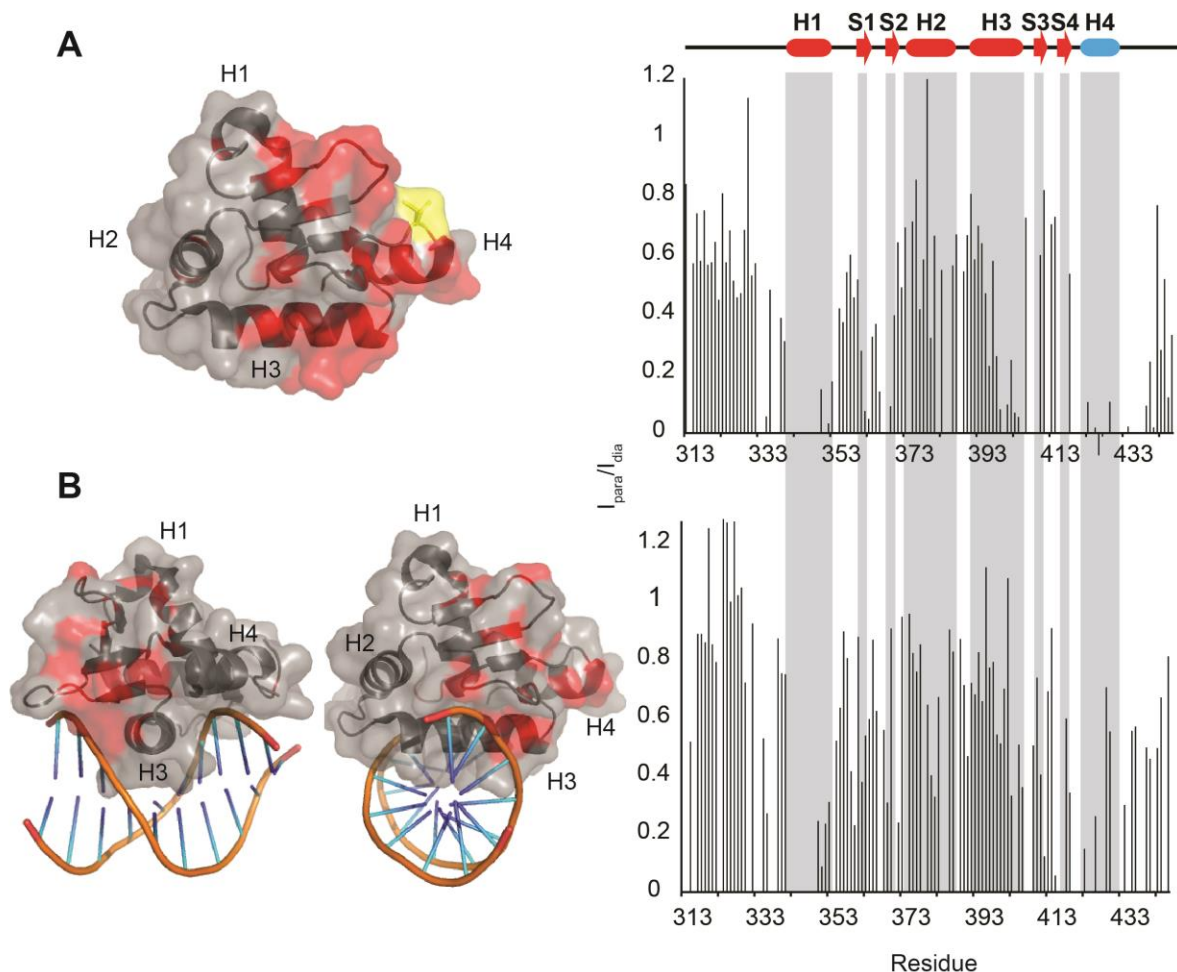
Figure 3-20 Paramagnetic relaxation enhancements (PRE) helps define the intramolecular interaction of partially autoinhibited of ETV4[313-446].

In (A), a paramagnetic spin label is covalently attached to Cys422 (yellow) on ETV4[313-446]. The relative intensities of amide ^{15}N-HSQC signals before (I_{para}) and after reduction of the nitroxide (I_{dia}) are plotted as in histogram format. Residues with (I_{para}/I_{dia}) values < 0.5 are mapped in red onto the crystal structure of ETV4. Amides within the ETS domain and the CID that are proximal to Cys422 show reduced intensities due to enhanced paramagnetic relaxation. However, NID residues are not perturbed relative to the protein average, and thus do not localize near Cys422. In (B), the spin label is covalently attached to the mutated Cys312 (in a M312C/C422S mutant). Residues showing (I_{para}/I_{dia}) values < 0.5 are also mapped in red on the crystal structure of ETV4. These data indicate that Cys312 in the truncated NID at least transiently localizes to regions of the ETV4 ETS domain including helix H3. A cartoon representation of DNA is include i to highlight the potential steric clash between the NID and the DNA-binding interface. Blank histogram values correspond to prolines and amides with unassigned or overlapping residues.

104

3.2.7. Potential self-association of ETV4

A different PRE approach was used to characterize the interaction between the inhibitory sequences and the ETS domain. In this method, a highly inert water-soluble paramagnetic compound, Gd(DTPA-BMA) (trademarked as Omniscan), was added to samples of either the uninhibited ETV4$^{328-430}$ or the partially inhibited ETV4$^{313-446}$. The paramagnetic compound will enhance the relaxation of amides close to or exposed on a protein's surface (Johansson et al. 2015; Pintacuda & Otting 2002). We thus compared the ^{15}N-HSQC spectra of the two ETV4 fragments in the absence and presence of Gd(DTPA-BMA). Unexpectedly, both the uninhibited ETV4$^{328-430}$ (Figure 3-21A) and partially inhibited ETV4$^{313-446}$ (Figure 3-21B) were found to have a surface spanning helix H1 and helix H2 that was well protected from the soluble PRE compound. This is not due to the presence of the partial NID. One possible explanation is that this protected surface reflects a dimerization interface for the ETV4 species.

Although not rigorously investigated, size exclusion chromatography indicated that various ETV4 constructs are dimeric (or oligomeric) under some experimental conditions. Also, these proteins tended to aggregate, particularly when highly concentrated in low ionic conditions. Furthermore, the correlation times for global tumbling extracted from ^{15}N relaxation measurements (Figure 3-12) were 12 ns for ETV4$^{328-430}$ (12.2 kDa) and 17 ns for ETV4$^{313-446}$ (15.7 kDa). The latter value, in particular, is indicative of a dimeric species.

Potential dimerization of ETV4$^{313-446}$ was also investigated by microscale thermophoresis. Based on a dilution series of the fluorescently tagged protein, a K_d value of > 10 μM was estimated for self-association (not shown). This value indicates that the very dilute samples of the ETV4 constructs used for EMSA DNA-binding studies were most certainly monomeric, and thus autoinhibition is not an artifact of self-association. Also, amide ^{15}N relaxation studies on the sortase ligated ETV4$^{165-436}$ yielded a global tumbling time of 7 ns, consistent with a monomer (not shown). However, this fully inhibited protein was studied at a concentration of 30 μM, versus more typical concentrations of 150 - 300 μM used for most NMR experiments. In conclusion, ETV4 has a propensity to self-associate,

possibly via the surface spanning helix H1 and helix H2 that was identified through solvent PRE measurements with Gd(DTPA-BMA). However, such potential dimerization does not lead to autoinhibition and does not complicate the key conclusions drawn from NMR spectroscopic and X-ray crystallographic studies of ETV4.
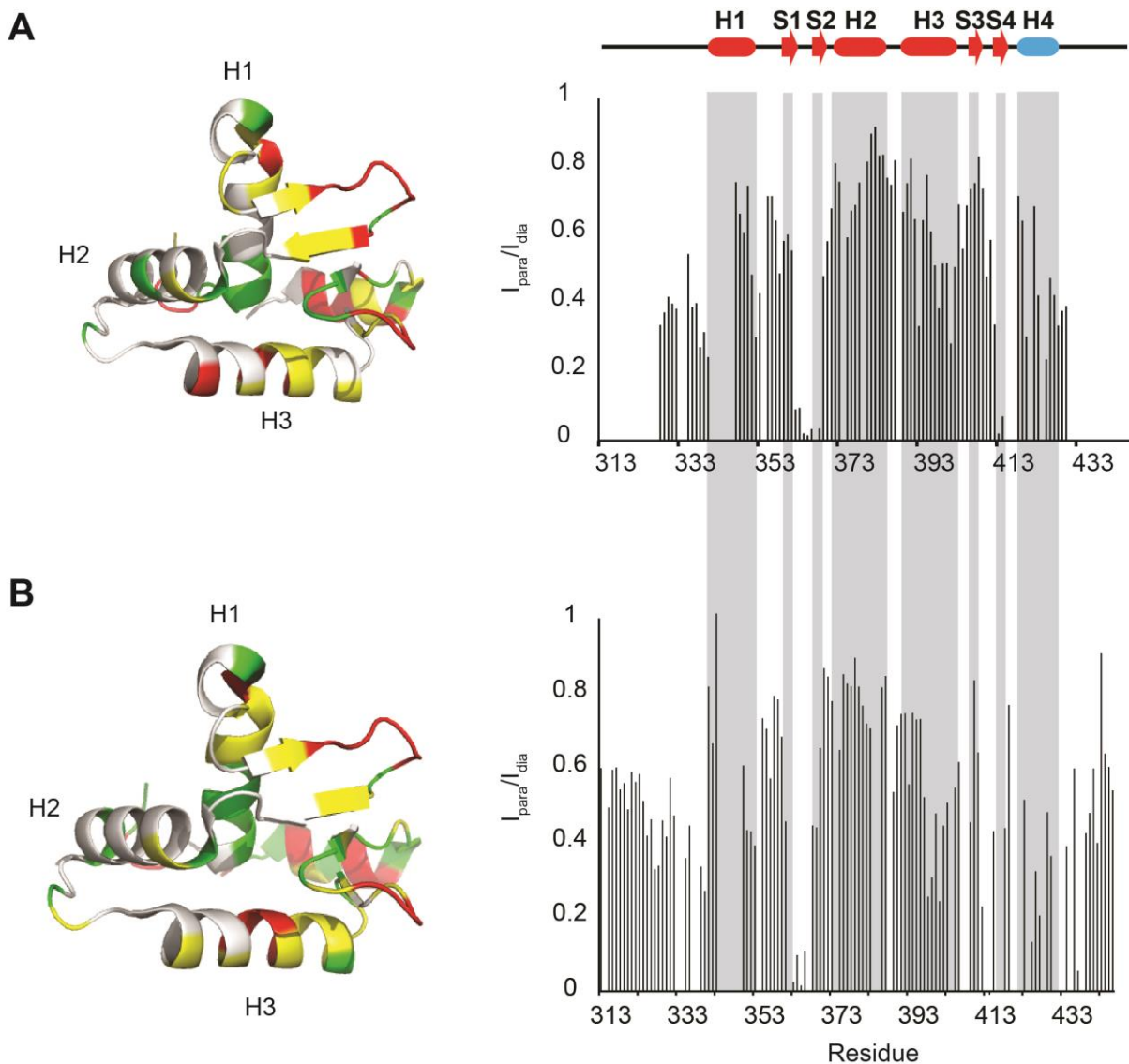
Figure 3-21 Probing the accessible surface of ETV4 via solvent PRE measurements.

The inert, water-soluble paramagnetic compound Gd(DTPA-BMA) was added in a 10:1 molar ratio to sample of ETV4[328-430] and ETV4[313-446]. The relative amide ^{15}N-HSQC peak intensities in the presence (I_{para}) versus absence (I_{dia}) of Gd(DTPA-BMA) are plotted in histogram format, and mapped onto the structure of the ETV4 ETS domain and CID. (white = protected ($I_{para}/I_{dia} > 0.6$), yellow = small protection (I_{para}/I_{dia}, 0.4-0.6), red = not protected ($I_{para}/I_{dia} < 0.4$), green = no assignment).

3.2.8. Acetylation of the NID counteracts DNA-binding autoinhibition

Widespread acetylation of lysine residues activates the DNA binding of ETV4, and two known sites of acetylation, Lys226 and Lys260, reside within the NID (Goel & Janknecht 2003; Guo et al. 2011; Kim et al. 2014). Therefore, we tested whether acetylation of these residues is sufficient for activating ETV4 DNA binding. Acetylation of either Lys226 or Lys260, independently, resulted in a decrease of DNA binding autoinhibition by 2.8- or 1.6-fold, respectively (Figure 3-22). First, we hypothesized that the positive charge of these lysine residues may be important for inhibition. However, mutation of Lys226 and Lys260 to glutamate or glutamine failed to recapitulate the activating nature by acetylation of these residues (Figure 3-23). Next we tested whether hydrophobic forces provided intramolecular interactions between the NID and the ETS domain and CID, such that the added bulk of acetyllysine might disrupt such interactions formed by nearby aromatic residues in the NID. Conserved aromatic residues that are located proximally to Lys226 in the NID were substituted with alanines, singly and in combination. However, this mutagenesis did not activate DNA binding (Figure 3-23). Therefore, while the exact nature of the inhibiting residues within the NID remains unclear, several lines of evidence support the occurrence of intramolecular interactions that might include a steric mechanism of DNA-binding autoinhibition.
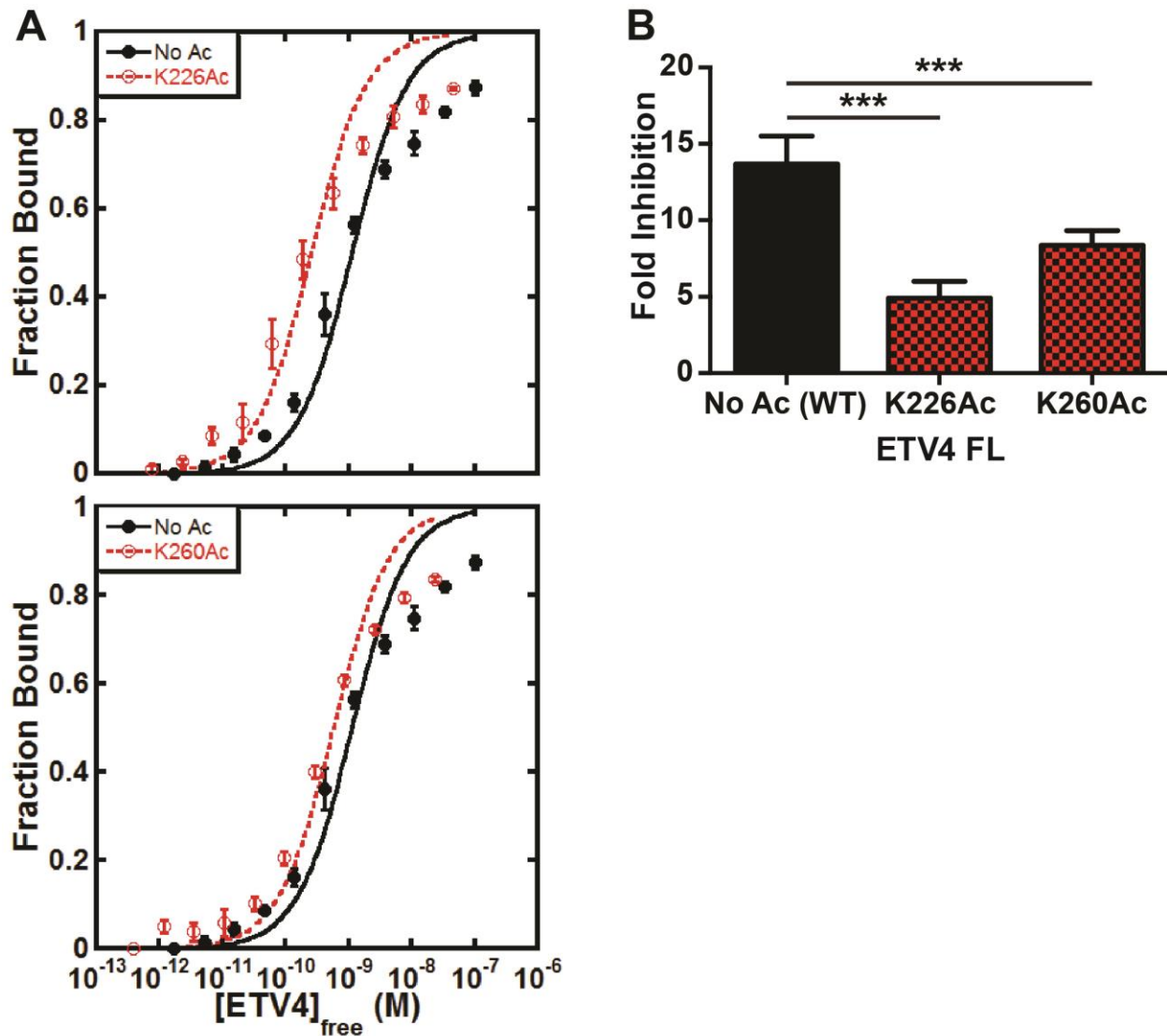
Figure 3-22 Acetylation of Lys226 or Lys260 relieves NID-dependent autoinhibition.

(A) Binding isotherms for full length ETV4 (FL) in its unacetylated form (No Ac; black), and acetylated at Lys226 (top, red) or Lys260 (bottom, red). Data points and error bars correspond to the mean and standard error of the mean from four replicates. (B) Quantification of fold inhibition relative to uninhibited ETV4[337-430] as in Figure 3-1D and Figure 3-4A. The DNA binding of ETV4 Lys226Ac (K_D, $30 \pm 6 \times 10^{-11}$ M) and ETV4 Lys260Ac (K_D, $51 \pm 3 \times 10^{-11}$ M) was inhibited 5 ± 1-fold and 8 ± 1-fold, respectively, whereas, unmodified ETV4 was inhibited 14 ± 2 fold. "***" Indicates $p < 0.001$.
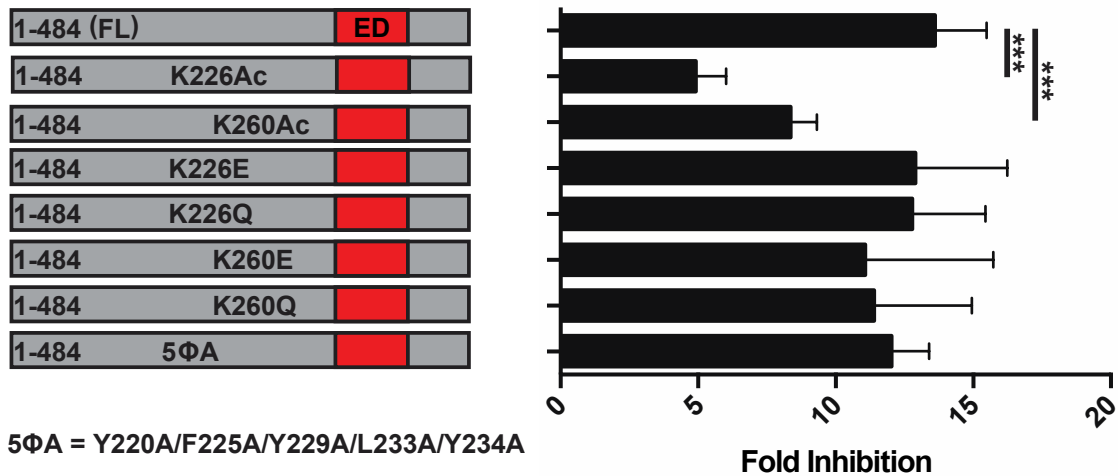
Figure 3-23 Acetylation at Lys226 and Lys260 activates the DNA binding of ETV4.

Fold-inhibition values for ETV4 with the indicated acetylated lysine residues and point mutations. Fold inhibition was calculated by comparing proteins to uninhibited ETV4[337-430], as in Figure 3-1A and Figure 3-4D. Bars and error bars refer to the mean and the standard error of the mean for the following number of replicates: 1-484, 35; 1-484 K226Ac, 4; 1-484 K260Ac, 4; K226E, 6; K226Q, 9; K260E, 3; K260Q, 7; Y220A F225A Y229A L233A Y234A, 3. "***" indicates p < 0.001.

3.2.9. Testing ETV4 autoinhibition *in vivo*

In a preliminary set of experiments to probe the *in vivo* roles of autoinhibition, I tested the transcriptional activity of ETV4 in the absence and presence of the inhibitory sequences. I first confirmed the over-expression of the ETV4 protein in the prostate cancer cell line PC3 (data not shown). Along with increasing amount of NID peptide (ETV4$^{165\text{-}336}$), these cells were transiently transfected with an Endoglin E3 promoter-derived ETS-responsive firefly luciferase reporter (pETS-luc) construct containing 3 conserved ETS recognition motifs. In a dose dependent manner, the pET-luc reporter activity dropped by up to ~ 20% due to presence of the exogenous NID peptide (Figure 3-24A). In a complementary approach, I cloned the genes encoding the NID (ETV4$^{165\text{-}336}$) and ETV4$^{437\text{-}484}$ into pcDNA3.1 vector for endogenous over-expression. Plasmids expressing the NID, or one of the controls, ETV4$^{437\text{-}484}$ (the non-inhibitory C-terminal segment of ETV4) or siAR (small interfering RNA targeting androgen receptor), were transfected into PC3 cells along with the pET-luc reporter. Endogenous over-expression of the proteins was confirmed using a Western blot (data not shown). Consistent with the isolated peptide transfections, the presence of the NID reduced ETV4 transcriptional activity (~ 30%), whereas ΔN437 and siAR had no effect (Figure 3-24B). Together, these very initial studies hint that the *in vivo* activity of wild-type ETV4 might be attenuated via an "in trans" interaction with the isolated NID.
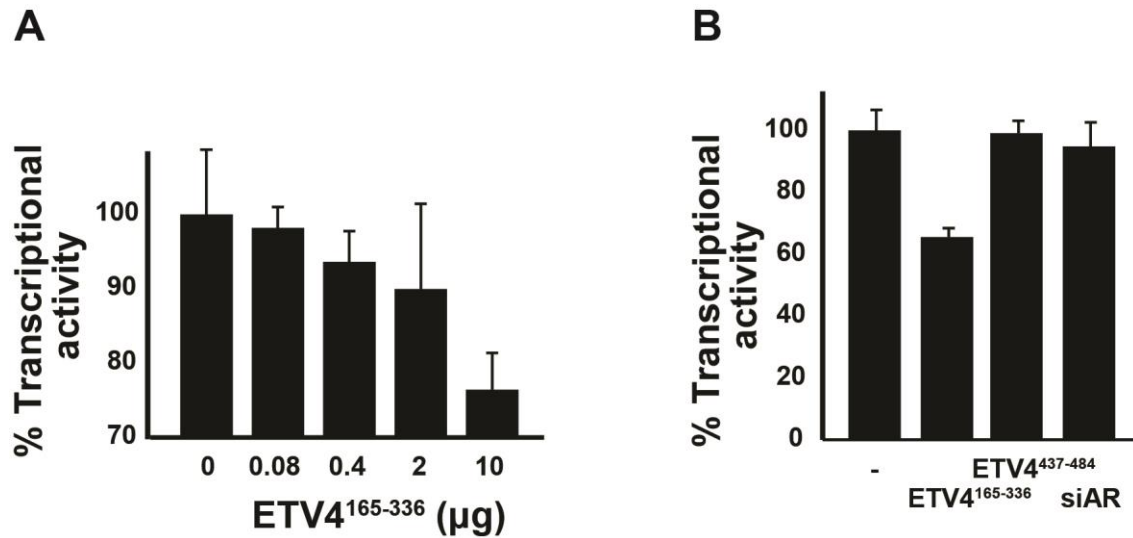
Figure 3-24 NID reduced ETV4 transcriptional activity *in vivo*.

The activity of endogenously overexpressed ETV4 in the prostate cancer cell line PC3 was assayed using an ETS-responsive firefly luciferase reporter. (A) Intracellular delivery of ETV4$^{165-336}$ using the Pro-Ject cationic lipid mixture. (B) Endogenous over-expression of transfected plasmids expressing the NID (ETV4$^{165-336}$), or the controls ETV4$^{437-484}$ and siAR. Error bars correspond to the mean and standard error of the mean from six replicates.

3.3. Discussion

3.3.1. Mechanistic model of autoinhibition

Here, we demonstrated that members of the ETV1/4/5 subfamily of ETS factors have regions N- and C-terminal of their ETS domains that act together to impinge upon helix H3 and inhibit DNA binding (Figure 3-26A). We propose that the CID allosterically shifts the conformational equilibrium of the DNA-recognition helix H3 towards a state less competent for binding. In contrast, the NID works, at least in part, in a steric manner to occlude the DNA binding interface, thus also requiring a conformational change for DNA binding.

The CID functions by influencing the position of helix H3, as supported by structural and mutational analysis. Amide HX experiments revealed that the CID helix H4 and the DNA-recognition helix H3 are both dynamic. These helices also exist in distinct conformations in crystallographic structures. Importantly, the uninhibited and the DNA-bound conformations of H3 match one another, but are distinct from the CID-inhibited conformation (Cooper et al. 2015). We propose that the dynamic nature of helices H3 and H4, detected by HX measurements, reflect the sampling of these multiple conformations. Direct interactions between Ile407-Leu430 couple these two helices allowing the CID to "push" H3 towards a state with lower affinity for DNA binding (Figure 3-26A).
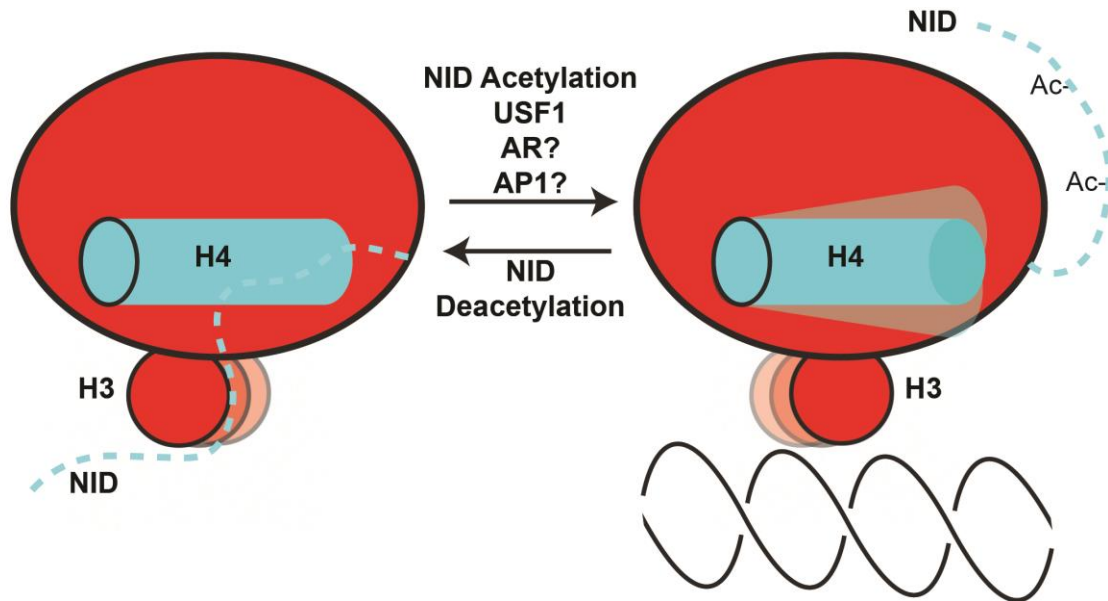
Figure 3-25 ETV1/4/5 subfamily factors are in equilibrium between forms that are more or less competent for binding to DNA.

The N-terminal inhibitory domain (NID) and C-terminal inhibitory domain (H4) inhibit the ETS domain (red oval) of ETV1/4/5 subfamily factors. H4 "pushes" the DNA-recognition helix (H3) towards a position that is less competent for DNA binding. The NID makes direct contact with the DNA-recognition helix to sterically inhibit DNA binding, and may also reinforce the inhibitory position of H4. Acetylation of lysine residues in the NID partially relieves autoinhibition,likely by disrupting NID interactions with the ETS domain. USF1 relieves ETV4 autoinhibition (Greenall et al. 2001), and we speculate that interactions with other factors, such as DNA binding factors AR or AP1, may also regulate ETV4 DNA-binding activity.

114
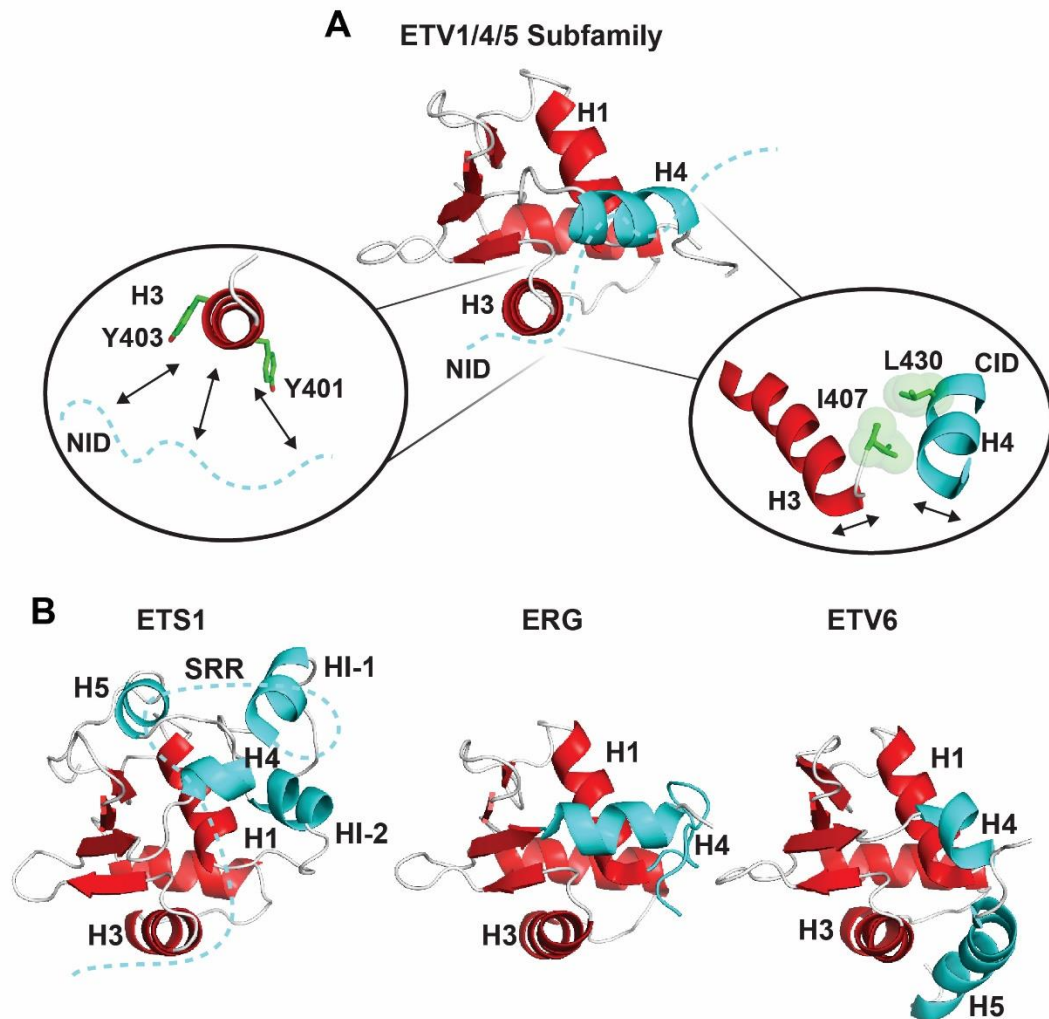
Figure 3-26 Autoinhibition in ETS family of transcription factors (ETS domain, red; inhibitory elements, cyan).

(A) Structural and molecular elements of ETV1/4/5 subfamily autoinhibition. The CID is an α-helix, H4, that interacts with the ETS domain primarily through Leu430 (ETV4 numbering). In particular, Leu430 interacts with Ile407 to influence the positioning of H3 (right inset). The NID, cyan dotted line, is intrinsically disordered and interacts via multiple regions with H3 of the ETS domain, as well as the CID. Tyr401 and Tyr403 of H3 are required for NID-mediated autoinhibition (left inset). (B) Examples of structurally characterized autoinhibited ETS factors: ETV1/4/5 subfamily (this study), ETS1 (Lee et al. 2005; Pufall et al. 2005; Lee et al. 2008; Garvie et al. 2002), ERG (Regan et al. 2013), and ETV6 (Green et al. 2010; De et al. 2014; Coyne et al. 2012). Dashed cyan lines refer to the disordered NID of ETV1/4/5 or the SRR of ETS1.

The NID is predominantly intrinsically disordered and inhibits the ETS domain through interactions with helix H3, and possibly with the CID. These interactions are weak and transient as evidenced by the small NMR spectral perturbations accompanying ligation of the NID, the lack of detectable structural changes in the NID due to presence of the ETS domain or CID, and the contribution of multiple regions of the NID to autoinhibition. However, multiple weak (or "fuzzy") interactions (Fuxreiter 2012) may lead to the overall inhibitory effect of the NID on the ETS domain and/or CID of ETV4. Tyr401 and Tyr403 in H3 are critical for NID-mediated autoinhibition and these residues directly contact DNA base pairs, suggesting that the NID sterically occludes part of the DNA binding interface. The NID also influences the CID, either through direct interaction that utilizes a broad interface or indirectly through the composite of the NID-H3 and H3-CID interactions. The tyrosine residues from H3 are not conserved in all ETS domains, and the sequence and positioning of CID helix H4 is unique to the ETV1/4/5 subfamily; therefore, the putative NID-interaction interface is specific to the DNA-binding domain of this subfamily (Figure 3-26A). Based on these collective findings, we propose that ETV1, ETV4, and ETV5 are in a dynamic equilibrium between conformations with different competencies for binding to DNA and that the NID and CID shift the equilibrium towards the less competent state (Figure 3-25).

The binding affinities of ETV1/4/5 fragments with either amino or carboxyl truncations suggest that the NID and CID work cooperatively, rather than additively, to inhibit DNA binding. Broadly speaking, this cooperation is supported by our structural and mutational data as the NID and the CID both impact the same location of the ETS domain, including the C-terminal portion of the DNA-recognition helix H3. In contrast to CID inhibition, there is insufficient understanding of the NID to ascertain the full basis of its inhibitory effects. Nevertheless, we speculate that, in addition to a simple steric mechanism, as evidenced by direct perturbations of the DNA binding interface, the NID also reinforces the CID-driven conformation of helix H3. This is consistent with a cooperative mechanism of autoinhibition and the NID-induced perturbations of residues in both the ETS domain and the CID, detected by NMR spectroscopy.

3.3.2. Autoinhibition in ETS family of transcription factors

The characterization of autoinhibition in the ETV1/4/5 subfamily adds to the diversity of molecular mechanisms utilized in inhibiting DNA binding by ETS factors (Figure 3-26B). The DNA binding of ETS1 is allosterically inhibited by an α-helical module that flanks the ETS domain and by an IDR termed the serine-rich region (SRR) (Lee et al. 2005; Pufall et al. 2005). Partial unfolding of the inhibitory module is linked to DNA binding (Petersen et al. 1995; Desjardins et al. 2016). In contrast, a single C-terminal α-helix sterically inhibits the DNA-binding interface of ETV6, and unfolds to allow DNA binding (Green et al. 2010; Coyne et al. 2012; De et al. 2014). An α-helix and an IDR are reported to allosterically inhibit the DNA binding of ERG (Regan et al. 2013). Although these inhibitory elements show no sequence similarity and are structurally distinct between ETS factors (Figure 3-2), in all of these cases helix H4 interacts with a conserved hydrophobic surface on the ETS domain (Figure 3-27). As this interaction is important for coupling the CID to the DNA-recognition helix H3 in ETV4, this conserved hydrophobic contact may reflect a common mechanism of inhibition among ETS factors. Beyond this conserved contact the diversity of appended helices likely facilitates distinct intramolecular interactions with inhibitory IDRs and intermolecular interactions with unique protein partners.

Although the inhibitory domains of ETV1/4/5 are distinct from the previously characterized examples of ETS1, ETV6, and ERG, the cooperation of inhibitory elements is most reminiscent of ETS1 autoinhibition. Four α-helices flanking the ETS domain of ETS1 provide a small 2-fold level of inhibition (Jonsen et al. 1996), and this autoinhibition is reinforced to ~ 20-fold by the SRR (Pufall et al. 2005; Lee et al. 2008). As is the case for the proposed interaction between the NID and the ETS domain/CID of ETV4, the dynamic SRR also interacts transiently with both the flanking inhibitory α-helices of ETS1 and its ETS domain. Furthermore, tyrosine and phenylalanine residues, amino acids that are usually depleted within IDRs (Williams et al. 2001), are present in the SRR of ETS1 (Desjardins et al. 2014) and in the NID of ETV1, ETV4, and ETV5. However, in ETS1, these aromatic residues reside in a repeating Ser-(Tyr/Phe)-Asp pattern, and these repeats are key to the transient interactions that respond to signaling-induced

117

phosphorylation and mediate inhibition. Such a repeat unit is not observed in the NID of ETV1/4/5, and mutation of aromatic residues in the NID of ETV4 did not influence autoinhibition. Therefore, although inhibitory IDRs are present in ETS1 and ETV1/4/5, these IDRs appear to modulate DNA binding via different intramolecular interactions with their corresponding ETS domains.

Divergent biological pathways and protein partnerships regulate the inhibitory elements of the ETS factors. Serine phosphorylation of ETS1, which targets the Ser-(Tyr/Phe)-Asp repeat, enhances the DNA-binding autoinhibition (Desjardins et al. 2014; Pufall et al. 2005). In contrast, serine phosphorylation of ETV1 does not impact autoinhibition (Wu & Janknecht 2002). Conversely, the relief of ETV4 autoinhibition by lysine acetylation, reported here, has not been observed for ETS1, ETV6, or ERG. Similarly, disparate protein partnerships regulate the DNA-binding autoinhibition of ETS1 and ETV1/4/5. For example, RUNX1 (Goetz et al. 2000; Shrivastava et al. 2014) and PAX5 (Garvie et al. 2002) counter ETS1 autoinhibition, and USF-1 relieves ETV4 autoinhibition (Greenall et al. 2001). The RUNX1-ETS1 partnership results in ETS1-specific regulation of ETS1-RUNX composite sites in T-cells (Hollenhorst et al. 2007), suggesting that the specific regulation of DNA-binding autoinhibition for an individual ETS factor can form the basis for that factor's unique biological function.

With a mechanistic foundation now in place for ETV1/4/5 autoinhibition, potential regulatory routes may be discovered, thus, providing insight into how these factors function in prostate cancer. We identified one possible route of ETV4 DNA binding activation through acetylation of lysines in the NID. Interestingly, the expression of p300, one of the acetyltransferases that modifies ETV1/4/5 factors (Goel & Janknecht 2003; Guo et al. 2011), correlates with prostate cancer progression and high levels of p300 are prognostic of biochemical recurrence in prostate cancer patients (Debes et al. 2003; Isharwal et al. 2008). Additionally, protein-protein partnerships may regulate the DNA binding of ETV1/4/5 factors. Investigating the effect on ETV1/4/5 DNA-binding autoinhibition by other transcription factors that bind to proximal genomic sites and act in the prostate will be of particular interest. Besides USF-1 (Greenall et al. 2001), candidates

include the AP1 factors (Hollenhorst, Ferris, et al. 2011) and the androgen receptor (Baena et al. 2013; Massie et al. 2007; Chen et al. 2013). Finally, inhibitory IDRs, and the signaling pathways that regulate them, are a potential therapeutic target as they differ between ETS factors and could provide factor-specific interventions. Despite the difficulty in rationally inhibiting IDRs (Y. Zhang et al. 2015), recent successes suggest that IDRs are a tractable small-molecule target (Hammoudeh et al. 2009; Krishnan et al. 2014; Pop et al. 2014; Z. Zhang et al. 2015).
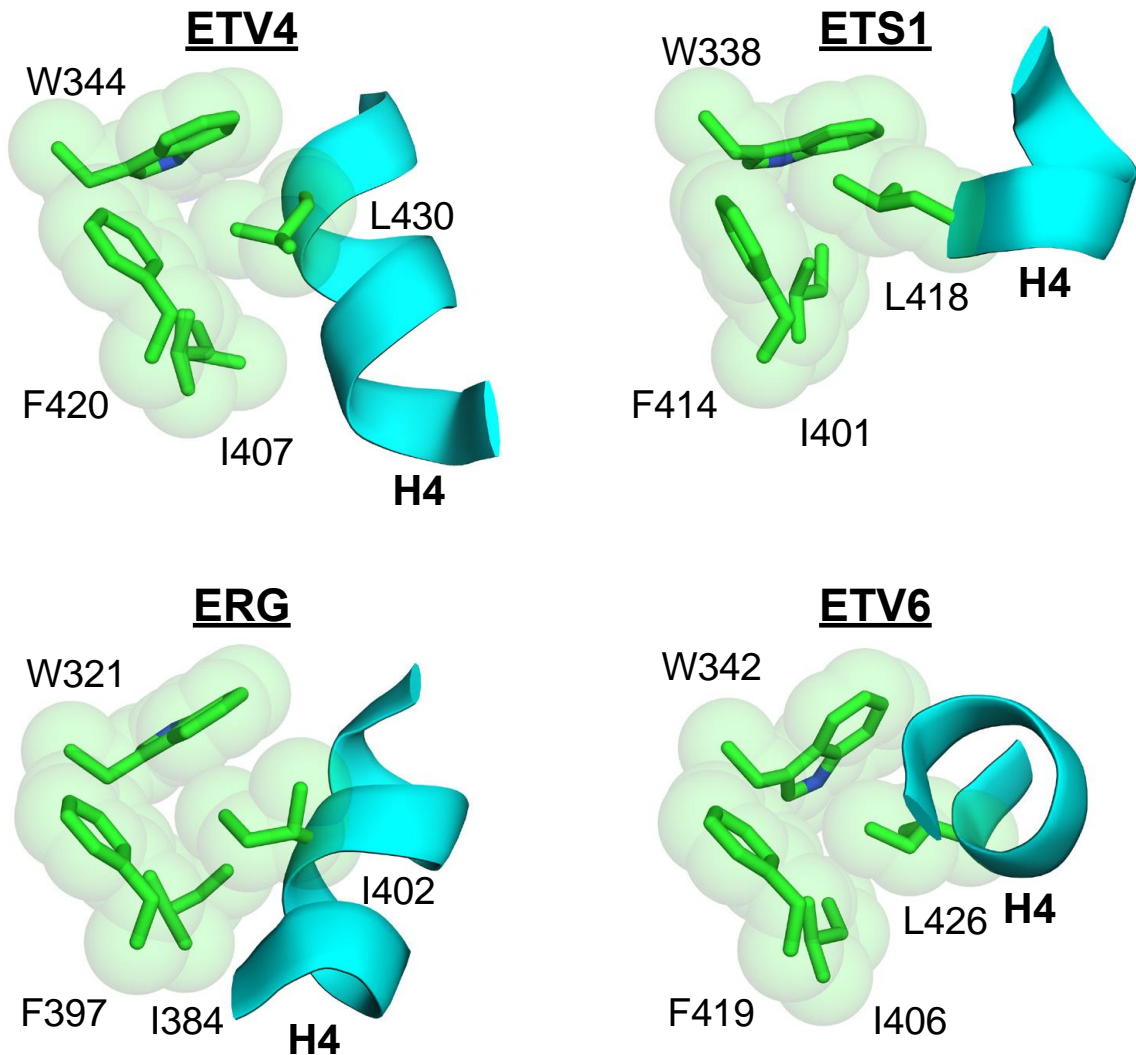
Figure 3-27 H4 is distinct in different ETS factors, but makes similar hydrophobic contacts with the ETS domain.

ETV4 residues Trp344, Ile407, and Phe420 from the ETS domain and Leu430 from α-helix H4 are shown in van der Waals sphere format to illustrate the hydrophobic contacts between H4 and the ETS domain. ETS1, ERG, and ETV6 are formatted in the same way and shown at the same angle. In the ETV1/4/5 subfamily of factors we propose that the Ile407-Leu430 interaction inhibits DNA-binding by "pushing" the DNA-recognition helix H3 into a conformation that is less competent for binding to DNA (Figure 3-7). The distinct, and subfamily-specific, versions of helix H4 in other ETS factors make similar hydrophobic contacts with this conserved surface on the ETS domain. Therefore, H3-H4 coupling may be a conserved mechanism of autoinhibition in ETS factors.

3.3.3. Autoinhibition as a route to transcription factor specificity

Many transcription factors are encoded by gene families and share a conserved DNA-binding domain (Vaquerizas et al. 2009). Disparate roles in development and disease indicate that individual transcription factors carry out specific functions and are not completely redundant with all other factors from the same family (Hollenhorst, McIntosh, et al. 2011; Rezsohazy et al. 2015). ETS factors, as a prototype for investigating this conundrum, have provided insight into how such specificity could have evolved. Previous work had established that ETS factors have distinct inhibitory domains that are specific to an individual factor, or subfamily of factors (Lee et al. 2005; Pufall et al. 2005; Green et al. 2010; Coyne et al. 2012; Regan et al. 2013). Cellular triggers can specifically regulate an individual ETS factor by integrating different signaling pathways or protein partnerships into these distinctive inhibitory features (Garvie et al. 2002; Shrivastava et al. 2014; Shiina et al. 2014; Goetz et al. 2000; Hollenhorst et al. 2007). We have extended this knowledge by describing an additional inhibitory module in the ETV1/4/5 subfamily with a distinct mechanism of inhibition and mode of cellular regulation. The divergent inhibitory domains of the four subclasses of ETS factors studied to date contact, in part, conserved regions on the ETS DNA-binding domain. Thus, subfamily-specific α-helices that flank the ETS domain serve as "adapters" that generate unique intra- and intermolecular interaction surfaces. Subfamily-specific IDRs interact with these surfaces to inhibit DNA binding via steric and/or allosteric mechanisms. These unstructured regions also provide diverse sites of post-translational modification that can inhibit or activate a factor in response to cellular regulation. Thus, the modest sequence variability among related DNA-binding domains could have been leveraged during evolution to enhance biological specificity by the coordinated divergence of structured inhibitory regions that flank the DNA-binding domain and more distal inhibitory IDRs.

3.4. Materials and methods

3.4.1. Expression plasmids

Human *ETV1*, *ETV4*, *ETV5*, *ERG*, and *FLI1* cDNAs corresponding to full-length or truncated proteins were cloned into the bacterial expression vector pET28 (Novagen) using standard sequence and ligation independent cloning strategies (Li & Elledge 2012). Point mutations were introduced into the ETV4 plasmid using the QuikChange site-directed mutagenesis protocol (Stratagene). For acetylation studies, codons en- coding Lys226 or Lys260 in the full-length ETV4 gene were mutated to an amber codon (UAG), and the natural amber stop codon was mutated to an opal codon (UGA). Mutated ETV4 cDNA was then cloned from the pET28 plasmid into a pCDF plasmid (kind gift from Dr. Jason Chin) for expression (Neumann et al. 2008).

3.4.2. Expression and purification of proteins

All proteins were produced in *Escherichia coli* (λDE3) cells. Uninhibited ETS factor DNA-binding domains and the ETV1/4/5 fragments not containing the NID were efficiently expressed into the soluble fraction. Cultures of 1 L Luria broth (LB) were grown at 37 °C to OD_{600} ~ 0.7 – 0.9, induced with 1 mM isopropyl-β -D-thiogalactopyranoside (IPTG), and grown at 30 °C for ~ 3 hr. To produce isotopically enriched proteins, expression was carried out using M9 minimal media supplemented with 3 g/L ($^{13}C_6$, 99%)-D-glucose and/or 1 g/L (^{15}N, 99%)-NH_4Cl.

Harvested cells were resuspended in 25 mM Tris pH 7.9, 1 M NaCl, 5 mM imidazole, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 2 mM 2-mercaptoethanol (BME), and 1 mM phenylmethanesulfonylfluoride (PMSF). Cells were lysed by sonication and centrifuged at 125,000 x g for at least 30 min at 4 °C. After centrifugation, the soluble supernatants were loaded onto a Ni^{2+} affinity column (GE Biosciences) and eluted over a 5 – 500 mM imidazole gradient. Fractions containing purified protein were pooled, combined with ~ 1

U thrombin / mg of purified protein, and dialyzed overnight at 4 °C into 25 mM Tris pH 7.9, 10% glycerol (v:v), 1 mM EDTA, 50 mM KCl, and 1 mM dithiothreitol (DTT). After centrifugation at 125,000 x g and 4 °C, the soluble fraction was loaded onto a SP-sepharose cation exchange column (GE Biosciences) and eluted over a 50 – 1000 mM KCl gradient. Fractions containing the ETS proteins were loaded onto a Superdex 75 gel filtration column (GE Biosciences) in 25 mM Tris pH 7.9, 10% glycerol (v:v), 1 mM EDTA, 300 mM KCl and 1 mM DTT. Eluted fractions were analyzed by SDS-PAGE. The final, purified proteins were then concentrated on a 10-kDa molecular weight cut-off (MWCO) Centricon device, snap-frozen with liquid nitrogen, and stored at -80 °C in single-use aliquots for subsequent EMSA studies.

Full-length ETS factors and ETV4 truncations containing the NID generally expressed more efficiently in the insoluble fraction using an autoinduction protocol (Studier 2005). Briefly, bacteria in 250 mL of autoinduction media were grown in 4 L flasks at 37 °C to an $OD_{600} \sim 0.6 - 1$. The temperature was then reduced to 30 °C and cultures were grown for another $\sim 12 - 24$ hr. Final OD_{600} values were typically $\sim 6 - 12$, indicating robust autoinduction. Harvested cells were resuspended as described above, sonicated and centrifuged at 31,000 x g for 15 min at 4 °C. The soluble fraction was discarded and this procedure was repeated with the pellet / insoluble fraction twice more to rinse the inclusion bodies. The final insoluble fraction was resuspended with 25 mM Tris pH 7.9, 1 M NaCl, 0.1 mM EDTA, 5 mM imidazole, 2 mM BME, 1 mM PMSF, and 6 M urea. After sonication and incubation for ~ 1 hr at 4 °C, the sample was centrifuged at 125,000 x g for at least 30 min at 4 °C. The soluble fraction was loaded onto a Ni^{2+} affinity column (GE Biosciences) and refolded by immediately switching to a buffer with the same components as above, except lacking urea. After elution with 5 – 500 mM imidazole, the remaining purification steps using ion-exchange and size-exclusion chromatography were performed as described above. However, a Q-sepharose anion-exchange column was used instead of a SP-sepharose cation-exchange column due to differing isoelectric points of the desired proteins.

Acetyllysine was incorporated into defined locations in the amino acid sequence and acetylated full-length ETV4 proteins were expressed according to a published protocol (Neumann et al. 2008). Briefly, expression was induced with IPTG, as de-scribed above, but in the presence of 10 mM acetyllysine, 20 mM nicotinamide, and a plasmid expressing an am- ber tRNA that has been mutated to recognize acetyllysine. Acetylated proteins were purified as outlined above for unacetylated full-length ETV4.

ETV4 proteins prepared for NMR spectroscopy were purified using protocols slightly different from above. Harvested cells were resuspended in 50 mM sodium phosphate, 500 mM NaCl, 10 mM imidazole, 6 M guanidinium HCl, pH 7.4 and lysed by at least one round of freeze/thaw, followed by passage 5 times through an EmulsiFlex-C5 homogenizer at 10 kPa, and finally, 15 min of sonication. The cell lysate was spun down by centrifuging at 25,000 x g for 1 hr at 4 °C. The supernatant containing ETV4 was then loaded onto Ni^{2+} affinity column (GE Biosciences), washed with 30 mM imidazole and eluted with 1000 mM imidazole and 6 M guanidinium HCl. Eluted fractions containing the desired protein were dialyzed against 3 L of refolding buffer (50 mM sodium phosphate, 1 M NaCl, 2 mM DTT and 1 mM EDTA, pH 7.5) at 4 °C overnight. The His_6-tag of the refolded proteins was cleaved by adding 1 U of thrombin/mg or TEV protease at a TEV/protein ratio of 1/200 (w/w). The mixture was loaded onto another Ni^{2+} affinity column, and the flow-through containing the tag-free ETV4 fragment was concentrated using a 3 kDa MWCO Centricon device to 2 mL. Size exclusion chromatography with Superdex 75 was used for a last purification step. Eluted fractions were assessed using SDS-PAGE and those containing the purified protein were pooled and dialyzed against NMR sample buffer (20 mM sodium phosphate, 200 – 1000 mM NaCl, 2 mM DTT, 0.1 mM EDTA, pH 6.5).

Protein concentrations were determined by measuring the absorbance at 280 nm using predicted ε_{280} values or at 595 nm after mixing 20 µL of protein with 1 mL of Bio-Rad Protein Assay Dye Reagent (diluted 1:5 in deionized water) and comparing to a bovine

serum albumin standard curve. Molecular weights for each ETS protein were predicted using the Peptide Property Calculator (Northwestern University) or using the ExPASY web server (Gasteiger et al. 2005).

3.4.3. Expressed protein ligation and purification

The DNA encoding ETV4 ETS domain and CID (337-436) was sub-cloned into bacterial expression vector pEM5B (kind gift from Dr. Pierre Barraud, Université Paris Descartes) between *Xho*I and *Bam*HI restriction sites. This enabled the addition of the required cysteine and TEV cleavage site (ENLYFQC) preceding the ETS domain, as described for the segmental labeling and expressed protein ligation protocol (Barraud & Allain 2013). The protein construct was expressed in LB media (unlabeled) or M9 media (^{15}N-labeled), purified under denaturing conditions, and refolded as described above. The protein was concentrated to 0.3 mM and stored in the inactive reaction buffer (50 mM HEPES, 200 mM NaCl, 0.1 mM TCEP, pH 7).

The DNA encoding ETV4 NID (165-336) was sub-cloned into pEM9B (kind gift from Dr. Pierre Barraud) between *Nde*I and *Sap*I restriction sites. The pEM9B expression vector also encodes a C-terminal *Mxe* GyrA intein. Nine additional amino acids (GGGHM preceding and GSSC following the NID) were introduced as a result of cloning and to enable protein ligation. The protein construct was expressed in LB media (unlabeled) or M9 media (^{15}N-labeled), cells were harvested and resuspended in native buffer (50 mM sodium phosphate, 500 mM NaCl, 10 mM imidazole, pH 7.4), and lysed by cell homogenization and sonication, as described above. The supernatant containing the desired protein was purified first by loading onto the Ni^{2+} affinity column, washed by 30 mM imidazole and eluted with 1000 mM imidazole. The protein was concentrated to 0.5 mM and stored in the inactive reaction buffer (50 mM HEPES, 200 mM NaCl, 0.1 mM TCEP (tris(2-carboxyethyl)phosphine) , pH 7).

Purified protein samples containing ^{15}N-labeled ETV4 ETS domain and CID and unlabeled ETV4 NID were mixed in a 1:2 molar ratio. The reaction was activated by adding 100 mM 2-mercaptoethanesulfonate (MESNA) and TEV protease at a TEV/protein ratio of 1/200 (w/w). The reaction mixture was incubated at 16 °C for 5 days. Time points were collected and analyzed on SDS-PAGE to monitor the ligation efficiency. TEV protease-cleaved products and intein self-cleaved products were purified on a chitin column equilibrated with 50 mM HEPES, 200 mM NaCl, pH 7. The flow-through of the chitin column containing the ligated product was purified on either ion-exchange chromatography (Mono Q) equilibrated with 50 mM HEPES pH 7 and eluted with 0 – 1000 mM NaCl gradient, and/or size exclusion chromatography (Superdex 75) equilibrated with NMR sample buffer (20 mM sodium phosphate, 200 mM NaCl, 2 mM DTT, 0.1 mM EDTA, pH 6.5). Fractions containing the final product were verified by SDS-PAGE and MALDI-ToF mass spectrometry on a Voyager-DE STR (Applied Biosystems) with a sinapinic acid matrix. The final product was dialyzed against NMR buffer. For the ligation reaction using ^{15}N-labeled ETV4 NID and unlabeled ETV4 ETS domain and CID, equal molar ratio were mixed (100 μM) to minimize aggregation due to highly concentrated ETV4$^{337-436}$. The reaction was initiated and the final product was purified and confirmed, as described above.

3.4.4. Segmental isotope labeling using sortase A

The DNA encoding the Sortase A peptidase was a kind gift from Dr. Michael Sattler (Institute of Structural Biology, Helmholtz Zentrum München). A mutation of Gly to Ala immediately after the TEV cleavage site was made to optimize ligation efficiency. The plasmids encoding Sortase A and ETV4$^{337-436}$ were transformed into *E. coli* (λDE3) cells for protein expression at 37 °C with 1 mM IPTG induction. Sortase A was purified by Ni^{+2} affinity and size exclusion chromatography. ^{15}N-labelled ETV4 (337-436) with an N-terminal glycine was prepared as described above, dialyzed into 50 mM Tris, 150 mM NaCl, pH 8.0, concentrated to 0.1 mM and stored at -80 °C. ETV4$^{165-336}$ was sub-cloned into pET28a between *NdeI* and *XhoI* restriction sites, with modification in the C-terminal to include the Sortase recognition sequence LPQTG plus a C-terminal His$_6$-tag. ETV4^{165-}

126

[336] was expressed and purified using the same protocol as for ETV4[337-436], except without thrombin digestion. Both Sortase A and ETV4[165-336] were dialyzed into 50 mM Tris, 150 mM NaCl, pH 8.0 and concentrated to 0.5 mM for storage at -80 °C. The ligation reaction was carried out as previously described (Freiburger et al. 2015). Briefly, the ^{15}N-labelled ETV4 fragment, the unlabelled ETV4 fragment, and Sortase A were combined in a 2:6:1 molar ratio. The mixture was centrifuged at 2000 rcf in a 3 kDa MWCO Centricon device with reaction buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 10 mM $CaCl_2$) at 20 °C for 4 hours. The ligated product was purified by passing twice through a Ni^{+2} affinity column with thrombin cleavage after the first purification step, as described above. The final Sortase ligated sample of unlabelled ETV4[165-336] linked to ^{15}N-labeled ETV4[337-436] was verified by MALDI-ToF mass spectrometry and SDS-PAGE.

3.4.5. Electrophoretic mobility shift assays (EMSA)

DNA-binding assays of ETS factors utilized a duplexed 27-bp oligonucleotide with a consensus ETS binding site: 5'-TCGACGGCCAAGCC**GGAA**GTGAGTGCC-3' (arbitrarily assigned as "top" strand) and 5'-TCGAGGCACTCACTTCCGGCTTGGCCG-3' ("bottom" strand). Boldface **GGAA** indicates the core ETS binding site motif. Each of these oligonucleotides, at 2 µM as measured by absorbance at 260 nM on a NanoDrop 1000 (Thermo Scientific), were labeled with [γ-^{32}P] ATP using T4 polynucleotide kinase at 37 °C for ~ 30 – 60 min. After purification over a Bio-Spin 6 chromatography column (Bio-Rad), the oligonucleotides were incubated at 100 °C for ~ 5 min, and then cooled to room temperature over 1 – 2 hr. The DNA for EMSAs was diluted to 1×10^{-12} M and held constant, whereas protein concentrations ranged ~ 6 orders of magnitude. For full binding isotherms, the exact concentration range was chosen according to the K_D of particular protein fragments. Protein concentrations were determined after thawing each aliquot of protein, using the Protein Assay Dye Reagent. Equivalent starting amounts (0.2 µg) of each protein utilized on a given day were run on an SDS-PAGE gel to confirm their relative concentrations. The binding reactions were incubated for 45 min at room temperature in a buffer containing 25 mM Tris pH 7.9, 0.1 mM EDTA, 60 mM KCl, 6 mM $MgCl_2$, 200 µg/mL BSA, 10 mM DTT, 2.5 ng/µL poly(dIdC), and 10% (v:v) glycerol, and then resolved

on an 8% (w:v) native polyacrylamide gel at room temperature. The ^{32}P-labeled DNA was quantified on dried gels by phosphorimaging on a Typhoon Trio Variable Mode Imager (Amersham Biosciences). Equilibrium dissociation constants (K_D) were determined by nonlinear least squares fitting of the total protein concentration $[P]_t$ at each titration point versus the fraction of DNA bound ($[PD]/[D]_t$) to the equation $[PD]/[D]_t = 1/[1 + K_D/[P]_t]$ using Kaleidagraph (v. 3.51; Synergy Software). Due to the low concentration of total DNA, $[D]_t$, in all reactions, the total protein concentration is a valid approximation of the free, unbound protein concentration. Reported K_D values represent the mean of at least three independent experiments and the standard error of the mean. Two-tailed, heteroscedastic t-tests were used to compare K_D values of different proteins.

Although there were deviations between the EMSA titration data and the non-linear regression at higher protein concentrations, this had essentially no effect on calculated K_D values. For comparison we fit all data using a one-site specific binding model with a Hill coefficient (Prism, version 6), and treating the maximum fraction of DNA bound as a variable. Use of this alternative approach, which eliminated the deviation between the data and the curve fit at higher protein concentrations, led to K_D values that were uniformly stronger by about 1.3 – 1.7-fold depending on the protein tested, with Hill coefficients ranging from 0.9 – 1.1. Importantly, all fold-inhibition values changed less than 1.5-fold, and thus all conclusions made from EMSA studies were supported by either curve fitting approach.

To test protein activity, binding reactions with known concentrations of radiolabeled duplex DNA (as described above) were titrated against a fixed concentration of protein corresponding to ~ 50-fold greater than the K_D value of each individual protein, as previously described (Jonsen et al. 1996). All proteins analyzed demonstrated high activity levels (> 95 %), indicating that the autoinhibition of DNA binding observed for the larger protein species is not due to a large fraction of the protein being an inactive binding species.

3.4.6. Partial proteolysis

For tryptic digestion studies, 20 µl of full length (FL) ETV4 at 20 µM was incubated with 1.5–450 ng of trypsin (Sigma) in a buffer containing 25 mM Tris pH 7.9, 10 mM $CaCl_2$, and 1 mM DTT. After 2 min of incubation, the reaction was quenched with 1 % (v:v) acetic acid (final volume). The resulting samples were analyzed by SDS-PAGE and ESI- MS (total mixture analyzed), and used for EMSA studies.

3.4.7. Crystallization and structure determination

Purifed proteins were dialyzed overnight in 10 mM Tris pH 7.9 and 50 mM NaCl, and then concentrated to 5 mg/ml. Crystals were grown by vapor diffusion in sitting drops of 2:1 protein:reservoir (v:v). CID-inhibited ETV1[332–435] was crystallized against a reservoir of 30% (w:v) PEG 5000 monomethyl ether, 0.1 M MES sodium salt and 0.2 M am- monium sulfate at pH 6.5 and 20°C. CID-inhibited ETV4[337–441] was crystallized against a reservoir of 1 M diammonium phosphate and 0.1 M sodium acetate at pH 4.5 and 20°C. Uninhibited ETV5[364–457] was crystallized against a reservoir of 0.2 M diammonium phosphate and 20% PEG 3350 at pH 5.0 and 4°C.

Crystals were immersed briefly in mother liquor containing 20% glycerol, and then cryo-cooled by plunging into liquid nitrogen. Diffraction data were collected on a Q315 CCD using Stanford Synchrotron Radiation Lightsource (SSRL) beamline 7-1 with X-rays at 1.0000 Å (ETV1 and ETV4) or 1.1271 Å (ETV5). The resulting data were integrated and scaled using HKL2000 (Otwinowski & Minor 1997). Phases were determined by molecular replacement with Phaser-MR (McCoy et al. 2007) using the ETS domain of ETS1 (1MD0.pdb) as a search model. Models were built with COOT (Emsley et al. 2010) and refined with PHENIX (Adams et al. 2010). PyMOL (Schrödinger, LLC) was used to render molecular structure figures.

Model geometries were analyzed by MolProbity (Chen et al. 2010) within PHENIX. For ETV1 (1.4 Å resolution data), 87.5% of residues have favorable backbone dihedrals and 12.5% fall into allowed regions. Residues 332–333 and 435 were not visible in the electron density. For ETV4 (1.1 Å), 91.8% of residues have favorable backbone dihedrals and 8.2% of residues fall into allowed regions. Residues 337–339 and 337–441 were not visible in the electron density. For ETV5 (1.8 Å), 87.7% of residues have favorable backbone dihedrals and 12.3% of residues fall into allowed regions. Residues 364–365 were not visible in the electron density. X-ray crystallography data collection and refinement statistics are provided in Table 3-3. Structural coordinates have been deposited in the RCSB Protein Data Bank under ID codes 5ILS (ETV1), 5ILU (ETV4) and 5ILV (ETV5).

3.4.8. Circular dichroism spectroscopy

Aliquots of frozen ETV4$^{165-336}$ (NID), expressed and purified as described above, were thawed, dialyzed overnight into 20 mM sodium phosphate, 50 mM NaCl, pH 7.9, and diluted to 25 µM concentration. CD spectra were recorded at 4 °C through the wavelength range of 190-260 nm with a 1 nm wavelength step. A baseline reference, consisting of buffer only, was subtracted from the CD spectra. Three scans were collected in series and averaged after visually verifying their consistency. Data were converted to molar ellipticity as described (Greenfield 2006).

3.4.9. NMR spectroscopy

NMR data were recorded at 25 °C on cryoprobe-equipped 500, 600, and 850 MHz Bruker Avance III spectrometers. Proteins were in NMR sample buffer (plus 10% lock D_2O) with 1 M NaCl for spectral assignments and with 200 mM NaCl for all other experiments. The elevated ionic strength reduced slow aggregation over long-term measurements. Data were processed and analyzed using NMRpipe (Delaglio et al. 1995) and Sparky (Lee et al. 2015). Signals from mainchain and sidechain ^1H, ^{13}C, and ^{15}N nuclei were assigned by standard multi-dimensional heteronuclear correlation experiments, including ^{15}N-

HSQC, HNCO, HN(CA)CO, CBCA(CO)NH, and HNCACB (Sattler et al. 1999). Amide $^1H/^2H$ hydrogen exchange (HX), after transfer into ~ 99% D_2O NMR sample buffer via a Sephadex G25 spin column, and CLEANEX-PM $^1H/^1H$ HX measurements were recorded using 850 MHz NMR spectrometer and analyzed as described previously (Coyne et al. 2012; Hwang et al. 1998). The two approaches detect slow (minutes-days) and fast (seconds) timescale exchange, respectively. Initially, HX measurements were carried out at pH 6.5 and 25 °C with uninhibited ETV4 (328-430). However, a more complete set of data were obtained for CID-inhibited ETV4 (337-436) via $^1H/^2H$ exchange measurements at pH 5.7 and 20 °C and CLEANEX-PM $^1H/^1H$ exchange at pH 7.5 and 25 °C. Protection factors were calculated as a ratio of the predicted HX rate constant for each amide in an unstructured polypeptide with the sequence of ETV4 versus the corresponding measured HX rate constant. The predicted values, corrected for pH, temperature and isotope effects, were obtained with the program Sphere (http://landing.foxchase.org/research/labs/roder/sphere/) (Bai et al. 1993). The protection factors for the two proteins, studied under several conditions, were merged and reported using the combined name ETV4[328-436] (Figure 3-11). For amides with HX quantitated under more than one condition, the highest protection factor is shown.

3.4.10. Paramagnetic relaxation enhancement

ETV4[313-446] contains a native single cysteine at position 422 which was used to covalently link the nitroxide spin label MTSL (S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methylmethanesulfonothioate) in one of the experiments. QuickChange mutagenesis protocol (Stratagene) was used to generate the mutant (M312C/C422S) for labeling the MTSL in the truncated NID. The proteins were expressed and purified as described above. The proteins were extensively dialyze into NMR buffer without DTT to remove any trace of reducing agent. Ten molar excess of MTSL was added to the protein and incubated overnight at room temperature to incorporate the spin label onto the cysteine. The modification was verified by MALDI-ToF mass spectrometry. The reaction mixtures were buffer exchanged to remove unreacted MTSL using an Amicon ultrafiltration device and concentrated to ~ 50 µM. ^{15}N HSQC were recorded on the spin-labeled ETV4 as the

paramagnetic state and the samples were subsequently reduced by adding 10 mM DTT for 24 hr. Another ^{15}N HSQC was recorded as the diamagnetic state. Ideally, a full T2 relaxation set should be collected for all PRE samples in their paramagnetic and diamagnetic states to report the changes in R2 values. However, the low concentration of the samples prevented such experiments and instead the reported PRE values are the amide peak intensity ratios (I_{para}/I_{dia}) in the two states. PRE intensity of the oxidized and reduced states were fitted with Sparky in order to obtain the intensity ratio.

3.4.11. Microscale thermophoresis (MST)

MST experiments were performed on a NanoTemper Monolith NT.115 instrument with blue/red channels. ETV4$^{313-446}$ was labeled using the Monolith NT Protein Labeling kit RED-MALEIMIDE (NanoTemper Technologies GmbH) according to the supplied protocol. Samples were prepared in NMR buffer (50 mM Na$_2$HPO$_4$, 200 mM NaCl, 2 mM DTT, 0.1 mM EDTA, pH 6.5), loaded into premium coated capillaries and measurements were performed at 20 % LED and 40% MST power. Laser off/on times were 5 and 30 s, respectively. The fluorescently labelled ETV4$^{313-446}$ were used at concentrations of 100 μM and was serially diluted to ~ 12 nM. The signals were fitted to the following formula to calculate the K$_D$ for the dissociation of ETV4$^{313-446}$.

$$f(c) = unbound + \frac{(bound-unbound)}{2(FluoConc+c+KD)-\sqrt{((FluoConc+c+KD)^2-4(FluoConc*c))}}$$

3.4.12. Cell culture and dual reporter luciferase assay

PC-3 cells were tested previously at the Vancouver Prostate Centre for ETV4 overexpression and were used in our assays. PC-3 cells were maintained in RPMI 1640 medium (Life Technologies) supplemented with 5 % (v/v) fetal bovine serum (FBS). Cells were grown in a humidified, 5 % CO$_2$ incubator at 37 °C.

Dual reporter luciferase assay was performed using the ETV4-overexpressing cell line PC-3. Three thousand cells in 150 μL per well of a 96 well plate were seeded and after a

24 hr incubation were transfected with 50 ng of an Endoglin E3 promoter-derived ETS-responsive Firefly luciferase reporter containing –507/–280 of the (E3) promoter inserted into luciferase reporter vector (Signosis) and 5 ng of the Renilla luciferase reporter (pRL-tk, Promega) using TransIT 20/20 transfection reagent (Mirus, USA). After 16 hr incubation, cells were treated with peptides using the Pro-Ject Protein Transfection Reagent (Thermo Scientific) for a further 24 hr. In this case, the peptides were first expressed in *E. Coli* and purified as mentioned above and were stored in 20 mM Na_2HPO_4, 200 mM NaCl, 2 mM DTT, 0.1 mM EDTA, pH 6.5. Luciferase and Renilla activity were measured using a TECAN M200Pro plate reader. Data were normalized first to Renilla and then to the protein transfection reagent control. The luciferase assays were repeated with different amount of (0.08 to 10 µg) peptides. A different attempt to look for the effect of the inhibitory peptides was to transiently transfect plasmids encoding the inhibitory peptides into PC-3 cells and allow for endogenous expression. The NID (ETV4[165-337]) and ∆N437 (ETV4[437-484]) were cloned into pCDNA3.1 with and without N-terminal HIS tags. One hundred nanograms of the plasmids encoding the inhibitory peptides were co-transfected with a firefly luciferase reporter as mentioned above and incubated for 24 hr for expression. The expression of the peptides was confirmed with anti-HIS antibody and western blot and luciferase and renilla activity was measured with cells transfected with non-tagged inhibitory peptides.

Chapter 4.   ETS domain dynamics

Overview

Protein dynamics is inherently related to function. A functional protein exhibits a wide range of motions including bond vibrations, sidechain rotations, conformational changes, and even complete (un)folding of the protein. ETS transcription factors regulate gene expression through their ETS domains, which recognize specific DNA sequences. ETS domains are dynamic and their motions are dampened in the presence of their autoinhibitory modules. Here, I report the common and distinct features of the motions of ETS domains ETV4, PU.1, Ets1, and ETV6, which are evolutionary distinct and have distinct mechanisms of DNA-binding autoinhibition. Thermal denaturation studies revealed that PU.1 and ETV4 had a mid-point unfolding temperature of 48 °C whereas Ets1 and ETV6 were more stable. Using NMR spectroscopy, I determined the structure of PU.1 (an ETS factor that is not autoinhibited) and identified an appended dynamic helix H4. Hydrogen exchange experiments also revealed that the DNA-recognition helices of all these factors are less protected compared to their core ETS domains. MD simulation and ^{15}N relaxation data demonstrated that the "turn" and "wing" at the DNA-binding interfaces are dynamic and this is likely due to the sizes, glycine/proline content, and charge states of the loops. Dynamical network analysis revealed the β-strands may serve as a central hub that relays "information" through the ETS domain. The motions of the ETS domains may thus provide the necessary contacts to distinguish their DNA-binding affinity and specificity.

4.1.   Introduction

Protein dynamics are central to ETS domain DNA-binding autoinhibition. Based on detailed studies of several ETS factors by our group and others, a general model has arisen in which autoinhibition involves modulating a conformational equilibrium between a more flexible state that is active for DNA binding, and a more rigid inactive state

(Hollenhorst, McIntosh, et al. 2011). Flexibility may contribute at several levels spanning the rapid searching of non-specific DNA sequences to the adoption of high affinity complexes with cognate DNA sites.

Ets1 represents the best characterized example linking autoinhibition and dynamics. As discussed throughout this thesis, Ets1 autoinhibition results from a helical inhibitory module, as well as an adjacent intrinsically disordered serine rich region (SRR), appended onto its core ETS domain (Figure 3-26) (Lee et al. 2005; Pufall et al. 2005; Lee et al. 2008). Through detailed hydrogen exchange (HX) studies by NMR spectroscopy, it was revealed that the inhibitory helices HI1 and HI2 are only marginally stable and thus poised to unfold (Lee et al. 2005). Upon binding specific and non-specific DNA, the inhibitory helices unfold, suggesting that the energetic penalty of this conformational change is linked to autoinhibition (Desjardins et al. 2016). However, this conformational change is only accountable for the ~ 2-fold difference in affinity between Ets1 and the minimal ETS domain. Full autoinhibition (~ 20-fold) is recapitulated only with the presence of SRR (Pufall et al. 2005). The SRR is phosphorylated by CaM kinase II in response to Ca^{+2} signaling, and increasing levels of multi-site phosphorylation progressively increase autoinhibition to ~ 500-fold. NMR analyses demonstrated that the SRR is intrinsically disordered and phosphorylation increases its transient interactions with the inhibitory helices and ETS domain (Lee et al. 2008; Desjardins et al. 2014). Importantly, the DNA recognition helix H3 of the ETS domain also has reduced HX protection indicative of local flexibility, and partakes in a network of ms-μs motions linked to the inhibitory helices. These motions are dampened by the SRR. This has led to a model of Ets1 autoinhibition in which multi-site phosphorylation acts as a "dimmer switch" to regulate transcription at the level of DNA binding (Pufall et al. 2005).

Dampened motions of the ETS domain due to the presence of inhibitory elements are also seen with ERG and ETV6. In the case of ERG, unstructured sequences appended N-terminal to the ETS domain (NID) and a C-terminal inhibitory helix H4 together yield a modest 2- to 5-fold DNA-binding autoinhibition. The presence of these inhibitory elements suppress ms-μs motions of the ETS domain, as determined by NMR relaxation dispersion

measurements (Regan et al. 2013). In the case of ETV6, two inhibitory helices are appended C-terminal to the ETS domain. The inhibitory helix H5 sterically blocks the DNA-recognition interface of the ETS domain (Coyne et al. 2012). NMR relaxation and HX measurements also revealed that helix H5 is only marginally stable and poised to unfold (Coyne et al. 2012; De et al. 2016). Furthermore, similar to Ets1, the DNA-recognition helix H3 of ETV6 has limited HX protection, and exhibits ms-µs timescale conformational fluctuations detectable by NMR relaxation dispersion experiments. The inhibitory helices dampen these motions and increase HX protection (Coyne et al. 2012). This also supports a model of ETV6 autoinhibition involving a conformational equilibrium between a flexible active and a rigid inactive state.

My goals in this final section of my thesis are to investigate the dynamic properties of the ETS domains from several ETS factors in order to understand their common and distinct features. These include Ets1, ETV6, ETV4, PU.1, and ERG, which exhibit a range of autoinhibitory mechanisms. The ETS domains of Ets1 and ETV6 have been extensively characterized by the McIntosh and Graves groups, and ETV4 was discussed in detail in the previous chapter. For further comparison, I also focused on PU.1, a divergent ETS factor that does not have any known autoinhibitory properties. These five ETS domains encompass three out of the four sub-families of ETS factors defined by their specificities for variant DNA sequences (Wei et al. 2010). Also, as described in chapter 3, they exhibit an ~ 100-fold range in K_D values for binding a common consensus DNA with a $^{5'}GGAA^{3'}$ core. This suggests that, despite the overall sequence conservation of their ETS domains, there are key structural or dynamic differences that influence their DNA binding properties. Insights from these studies may thus help explain the specific functions exhibited by these ETS factors in a cellular context.

```
        HI-1                    HI-2              H1              S1
PU.1^167-272   ----------DGEADGLEPGPGLLPGE------TGSKKKIRLYQFLLDLLRSGDMKDSIWWVDKD
ETV6^335-426   YMN-HI-----------MVSVSPPEEHAMPIGRIADCRLLWDYVYQLLSDSRYENFIRWEDKE
ETV4^337-436   PFPDDVCVVPEKFEGDIKQEGVGAFR--EGPPYQRRGALQLWQFLVALLDDPTNAHFIAWTGRG
Ets1^331-440   TFKDYVRDRAD----LNKDKPVIPAA--ALAGYTGSGPIQLWQFLLELLTDKSCQSFISWTGDG
ERG^307-407    ----TEDQRPQ----LDPYQILGPTS--SRLANPGSGQIQLWQFLLELLSDSSNSSCITWEGTN
                                                    *:::: **  .    *  *.

        S2        H2              H3          S3      S4   H4
PU.1^167-272   KGTFQFSSKHKEALAHRWGIQKGNRKKMTYQKMARALRNYGKTGEVKKVK-KKLTYQFSGEVLG
ETV6^335-426   SKIFRI--VDPNGLARLWGNHK-NRTNMTYEKMSRALRHYYKLNIIRKEPGQRLLFRFMKTPDE
ETV4^337-436   -MEFKL--IEPEEVARLWGIQK-NRPAMNYDKLSRSLRYYYEKGIMQKVAGERYVYKFVCEPEA
Ets1^331-440   -WEFKL--SDPDEVARRWGKRK-NKPKMNYEKLSRGLRYYYDKNIIHKTAGKRYVYRFVCDLQS
ERG^307-407    -GEFKM--TDPDEVARRWGERK-SKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFHGIA
                *::   .  : :*: ** :* .:  *.*:*::*.** *   :  *   ::  ::*

        H5
PU.1^167-272   RGGLAERRHPPH------------------------------------------------
ETV6^335-426   IM---SGRTDRLEH-----------L-------ESQELDEQIYQEDEC----------
ETV4^337-436   LF---SLAFPDNQR--------PAL--------KAEFDRPVSEEDTVPLSHLDESP
Ets1^331-440   -----LLGYTPE--------ELHAMLDV----------KPDADE------------
ERG^307-407    QA---LQPHPPESSLYKYPSDLPYMGSYHAHPQKMNFVAPHPPA--LPVTSSS---
```

Figure 4-1 Sequence alignment of ETS domains characterized in this chapter.

Shown are the sequences of the ETS domain containing constructs of the ETS factors characterized in this chapter (aligned and color coded by Clustal Omega (Sievers et al. 2011)). Shaded in light orange are the boundaries of each protein characterized in this chapter. Non-native residues from the expression vector are also present at the N-termini as follows (PU.1: HIHM, ETV4/ETV6/ERG: GSHM). Secondary structural elements determined experimentally by X-ray crystallography and/or NMR spectroscopy are highlighted (α-helices, red shade; β-strands, blue shade)

4.2.   Results

4.2.1. Thermal stability parameters of ETS domains

Initially, I measured the thermal stabilities of the minimal uninhibited (or nearly uninhibited) ETS domain-containing fragments of ETV4[337-436], ERG[307-407], Ets1[301-440], ETV6[335-426] and PU.1[167-272] using circular dichroism (CD) spectroscopy. To facilitate comparison, the exact sequences and secondary structures of these uninhibited or weakly inhibited species are shown in Figure 4-1.

As expected, each CD spectrum at 20 $^\circ$C showed broad and negative signal spanning from 208 to 222 nm, indicative of a folded protein with α-helical secondary structure Figure 4-2A). Accordingly, to monitor their unfolding transitions, CD signals at 222 nm were recorded as a function of temperature from 15 °C to 95 °C (Figure 4-2B). Although recorded under moderately different conditions of ionic strength and pH (see Methods), the ETV6 ETS domain was found to be most stable with a midpoint unfolding temperature T_m of 66 °C, whereas those of ETV4 and PU.1 were the least stable with T_m values of 48 °C (Figure 4-2C and Table 4-1). However, in contrast to the other four proteins, which showed relatively sharp cooperative unfolding transitions, ETV6[335-426] exhibited a very broad denaturation curve and hence low ΔH and ΔS values. The origin of this unusual, and reproducible, behavior is unclear. Although such broad transitions are often associated with molten globule-like behaviour, ETV6[335-426] has a very well-defined structure at 25 $^\circ$C (Coyne et al. 2012; De et al. 2014). It should also be noted that none of the proteins refolded upon cooling and thus the fit parameters in Table 4-1 do not reflect reversible conformational changes.
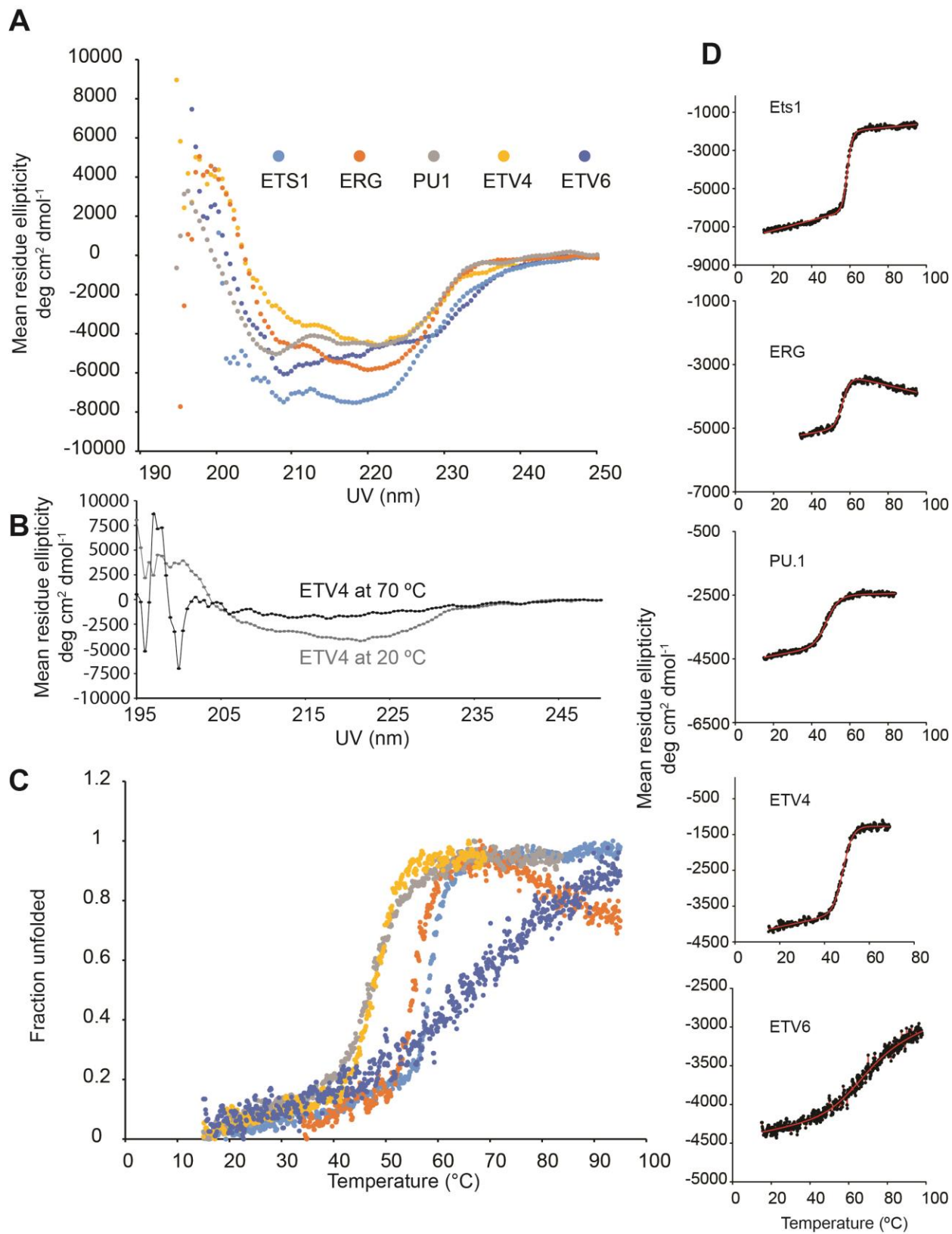
Figure 4-2 Circular dichroism spectra and thermal denaturation curves of five ETS domains.

(A) The CD spectra of the ETS domains were measured at 20 °C from 190 nm to 260 nm and converted to mean residue ellipticity for comparison. The differences in the spectra may reflect structural differences, such as the presence of appended inhibitory sequences, as well as errors in protein concentration determination and base line correction. (B) The CD spectra of ETV4 at 20 °C and 70 °C indicative of the signal loss upon thermal denaturation. (C) Superimposed thermal denaturation curves of the ETS domains, monitored at 222 nm, indicate different T_m values and the unusual unfolding transition of ETV6. For comparison, the curves were scaled to fraction unfolded. Fraction unfolded = $S - S_{min}$ / $S_{max} - S_{min}$ where S is the CD signal. (D) Individual thermal denaturation curves of mean residue ellipticity versus temperature fit to a two-state transition model (red line). See Table 4-1 for fit values.

Table 4-1 Thermodynamics parameters for ETS domain unfolding[a]

	Ets1	ETV4	ERG	ETV6	PU.1
T_m (°C)	59 ± 1	48 ± 1	57 ± 1	66 ± 2	48 ± 1
ΔH (kcal mol^{-1})	165 ± 2	93 ± 2	108 ± 2	24 ± 2	65 ± 1
ΔS (kcal K^{-1} mol^{-1})[b]	0.50 ± 0.05	0.29 ± 0.05	0.33 ± 0.05	0.07 ± 0.05	0.19 ± 0.02

[a] The constructs were Ets1[301-440] (pH 6.5), ETV4[337-436] (pH 6.5), ERG[307-407] (pH 6.5), ETV6[335-426] (pH 6.5) and PU.1[167-272] (pH 5.5). See Methods for exact buffer conditions.

[b] From the curve fitting, $\Delta S = \Delta H / T_m$ and both values correspond to the mid-point unfolding T_m temperature for the given protein.

4.2.2. Structural characterization of the PU.1 ETS domain

To expand the scope of my comparative studies of ETS domain dynamics, I used NMR spectroscopy to characterize the ETS factor PU.1. This transcription factor belongs to the SPI sub-family of ETS proteins, which have a preference for AT-rich sequences flanking the core ETS consensus motif (Munde et al. 2014). Importantly, PU.1 is evolutionarily divergent from other ETS family members and has not been reported to exhibit DNA-binding autoinhibition (Hollenhorst, McIntosh, et al. 2011). As a result, PU.1 provides a valuable reference for comparisons with previously characterized autoinhibited ETS family members.

The well-dispersed ^{15}N-HSQC spectrum of PU.1$^{167-272}$ confirmed that the protein is folded and stable in solution (Figure 4-3A). With the aid of a summer student, signals from the main-chain and side-chain ^{1}H, ^{13}C, ^{15}N nuclei of PU.1$^{167-272}$ were assigned via standard heteronuclear correlation NMR methods. Based on an analysis of its ^{1}H^{N}, ^{15}N, ^{13}C$^{\alpha}$, ^{13}C$^{\beta}$, and ^{13}C$^{'}$ chemical shifts with the MICS algorithm, PU.1$^{167-272}$ clearly had the secondary structure of three α-helices and four β-strands common to all ETS domains. In addition, a short C-terminal α-helix H4 is formed by residues 256-260 (Figure 4-3B). Due to the use of C-terminally truncated proteins, this helix was not present in the previously determined X-ray structure of PU.1$^{171-259}$ in complex with DNA (Kodandapani et al. 1996) or detected through NMR spectroscopic studies of unbound PU.1$^{169-260}$ (Jia et al. 1999). However, a report by Escalante *et al* in 2002 noted the presence of helix H4 in the unreleased X-ray crystallographic structure of PU.1$^{172-262}$ in complex with DNA and a partner transcription factor, IRF-4 (Escalante, Shen, et al. 2002; Escalante, Brass, et al. 2002). Residues corresponding to helix H4 have not been associated with any functional role of PU.1, and are perhaps best viewed as a structural feature of ETS domains in general. Indeed, it appears that most, if not all ETS factors contain a helix H4 of variable length and orientation appended to their ETS domains (Figure 4-1).

The structural ensemble of PU.1$^{167-272}$ was determined with CYANA 3.0 (Güntert 2004) and Ponderosa (W. Lee et al. 2011) using nuclear Overhauser enhancement (NOE)-

derived distance and chemical-shift derived dihedral angle restraints (Table 4-2). As expected, residues 174-254 displayed the conserved architecture of an ETS domain comprised of three helices and a four stranded antiparallel β-sheet scaffold (Figure 4-4; H1: 174-184; S1: 191-195; S2: 199-203; H2: 205-219; H3: 227-240; S3: 243-245; S4: 251-254). A helical turn is also present between helix H1 and strand S1 and between strand S1 and S2. The core ETS domain structure superimposed well upon the X-ray structure of DNA-bound PU.1[171-259] (PDB 1PUE). Consistent with the MICS analysis of main chain chemical shifts, residues 256-260 formed the newly identified helix H4. This short amphipathic helix is composed of Val258 and Leu259 facing the core of the ETS domain and Glu257 exposed to the solvent. As expected from their chemical-shifts derived RCI-S^2 (random coil index squared order parameter) values (Figure 4-3), the remaining N-terminal (167-173) and C-terminal (261-272) regions of PU.1[167-272] are conformationally disordered.
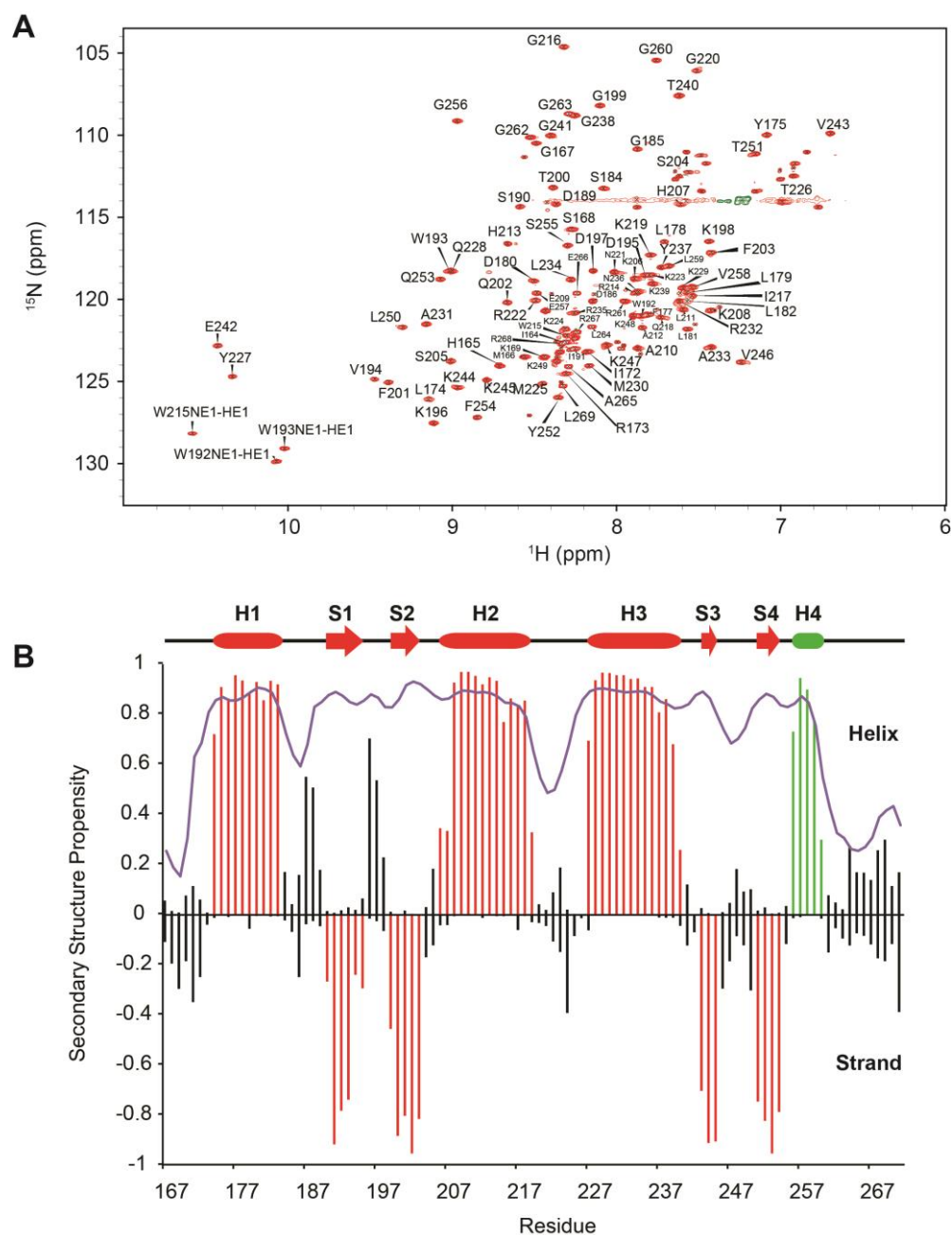
Figure 4-3 Backbone amide assignment of PU.1[167-272].

A) Assigned ^{15}N-HSQC spectrum of PU.1[167-272]. (B) Secondary structure propensities calculated from main chain chemical shifts using the program MICS (Shen & Bax 2012). Values above 0 indicate α-helical propensity and those below 0 indicate β-strand propensity. In addition to the characteristic helices and strands described previously for the PU.1 ETS domain (red), a short C-terminal helix H4 was also identified (green). Also shown are the RCI-S^2 values (purple line; decreasing value indicates increasing flexibility).
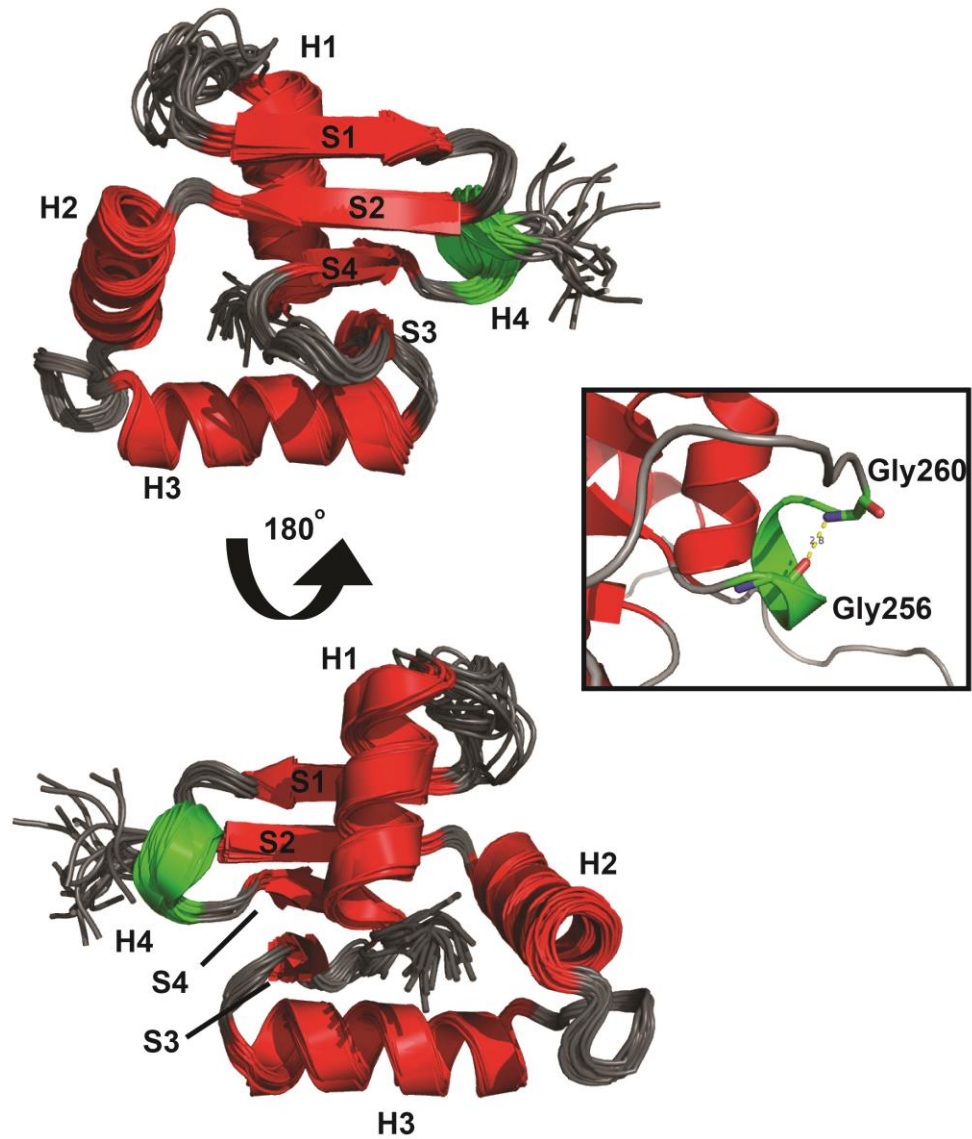
Figure 4-4 Structural ensembles of PU.1[167-272].

Cartoon representations of the refined 20-member structural ensemble of PU.1[167-272]. Disordered N- and C-terminal residues are not shown for clarity. Core α-helices and β-strands are in red, and the newly identified α-helix H4 is in green. Grey represents loop regions. Box highlights the hydrogen bond in helix H4 identified in all 20 structures between the carboxyl of Gly256 and the amide of Gly260.

Table 4-2 NMR refinement statistics for PU.1[167-272] structural ensemble

	PU.1[167-272]
NMR distance and dihedral restraints	
Distance restraints	
Total NOE	984
Intra-residue	300
Inter-residue	684
Sequential ($|i - j| = 1$)	539
Medium-range ($|i - j| \leq 4$)	156
Long-range ($|i - j| \geq 5$)	289
Dihedral angle restraints	
Φ, Ψ	77,77
Structure statistics	
Violations (mean ± SD)	
Distance restraints (Å)	0.048 ± 0.001
Dihedral angle restraints (°)	0.212 ± 0.055
Max. dihedral angle violation (°)	1.56 ± 0.38
Max. distance restraint violation (Å)	0.14 ± 0.05
Residues located within the generously allowed regions of the Ramachandran plot (%)	94.9
Average pairwise rmsd (Å)[a]	
All heavy atoms	1.14 ± 0.22
Backbone only	0.66 ± 0.16

[a] alignment between residue 174-261

4.2.3. Fast timescale dynamics of the ETV4 and PU.1 ETS domains

Having determined the structure of the PU.1 ETS domain, I next used ^{15}N relaxation measurements to characterize the fast timescale dynamics of PU.1$^{167-272}$ and uninhibited ETV4$^{328-430}$. These studies complement previously reported relaxation measurements of Ets1$^{301-440}$, ETV6$^{335-426}$ and ERG$^{307-407}$ (Lee et al. 2005; Coyne et al. 2012; Regan et al. 2013).

The amide ^{15}N T$_1$ and T$_2$ lifetimes and steady-state heteronuclear NOE values of PU.1$^{167-272}$ and ETV4$^{328-430}$ are shown in the histograms of Figure 4-5. The latter was presented previously in Figure 3-12 and is included for comparison. The heteronuclear NOE in particular, and to a lesser extent, the T$_2$ lifetimes are very sensitive indicators of amide mobility on a sub-ns timescale, with decreasing and increasing values, respectively, indicating increasing flexibility. The NOE values and T$_2$ lifetimes of the amides at the ends of proteins were lower and higher, respectively, than those in the core ETS domains, indicating that their terminal residues are flexible. For both ETS domains, the NOE values were also reduced for residues in the "turn" between H2 and H3, and the "wing" between S3 and S4, indicating partial flexibility on this fast timescale (Figure 4-5). This is also seen with unbound Ets1 (Desjardins et al. 2016) and ETV6 (Coyne et al. 2012), whereas in the presence of specific DNA, the "turn" and "wing" become more ordered due to contacts with the phosphodiester backbone (Desjardins et al. 2016). In contrast, amides within the α-helices and β-strands of the two proteins had relatively uniform relaxation behaviors, indicating that their ETS domains have well ordered secondary structural elements. In the case of PU.1$^{167-272}$, helix H4 is also ordered on the sub-ns timescale. In the case of ETV4$^{328-430}$, helix H4 is actually truncated relative to a full length, inhibitory helical CID, and does show some evidence of mobility by ^{15}N relaxation (Figure 3-12). Indeed, as described in Chapter 3, truncated and full length helix H4 adopt different conformations in X-ray crystallographic structures of ETV4 fragments (Chapter 3).

The amide relaxation data for the two proteins were fit to the model-free model with Tensor2 (Dosset et al. 2000) to obtain a generalized order parameter S^2 for each residue

Figure 4-6AB). This parameter describes the mobility of the N-H bond and decreases as spatial restrictions decrease (Lee et al. 2005). Overall, secondary structure elements showed uniformly high S^2 values, whereas terminal and loop regions were more mobile. One exception is the truncated helix H4 of ETV4, which exhibited decreasing S^2 values towards its C-terminal end. In contrast, helix H4 of PU.1 is well ordered. It is also noteworthy that the S^2 and RCI-S^2 values obtained for PU.1[167-272] from ^{15}N relaxation (Figure 4-6A) and main chain chemical shifts (Figure 4-3B), respectively, agree well.

This model-free analysis also yielded correlation times for the isotopic global tumbling of PU.1[167-272] (8.1 ± 0.1 ns) and ETV4[328-430] (12.2 ± 0.1 ns). These values indicate that PU.1[167-272] is predominantly monomeric in solution, whereas ETV4[328-430] appears to be self-associated to dimeric or oligomeric forms under these experimental conditions. As discussed in chapter 3, this is consistent with the observation that ETV4[328-430] showed a propensity to aggregate at the relatively high protein concentrations used for NMR analysis. Such aggregation was reduced using higher ionic strength buffer conditions.
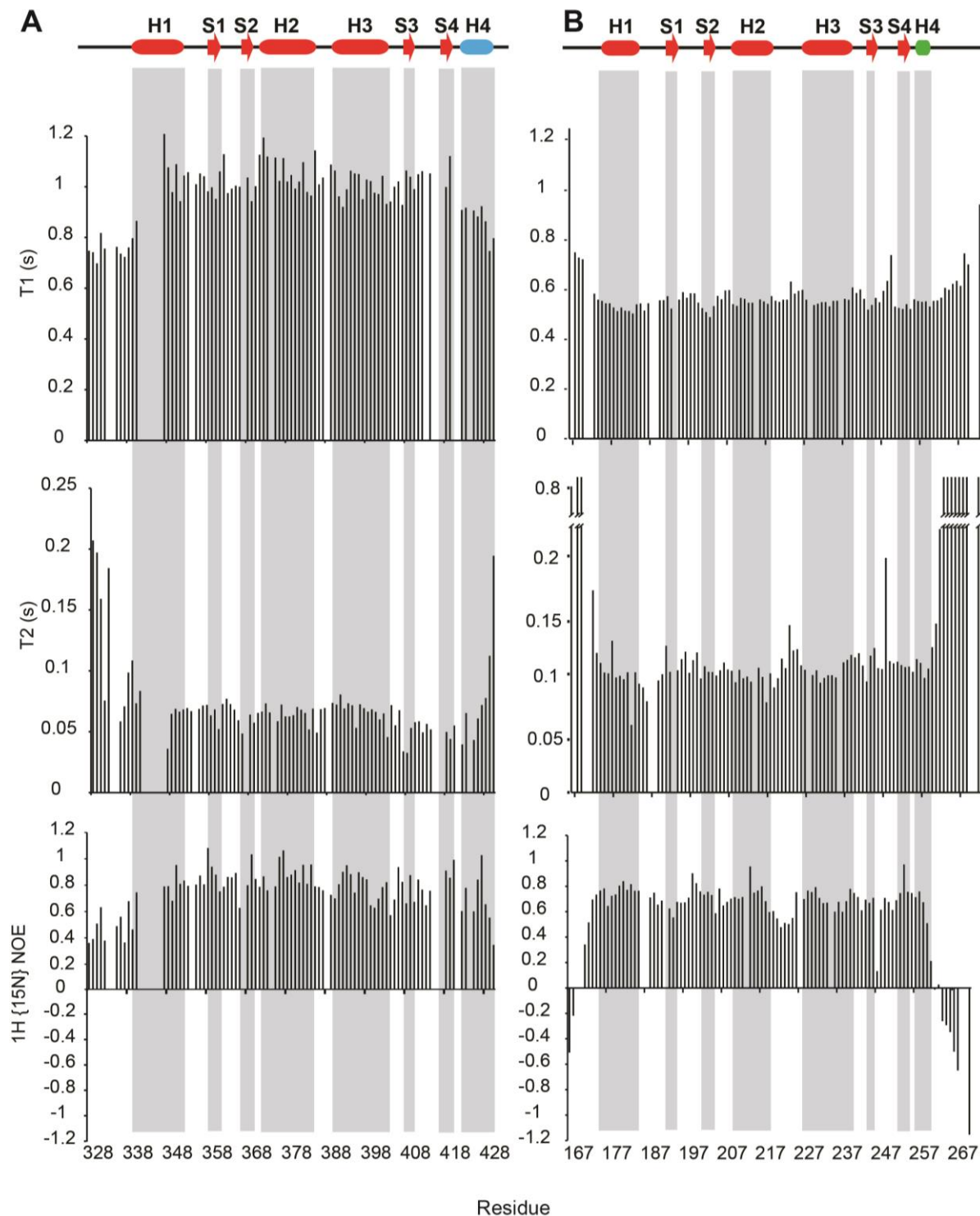
Figure 4-5 Fast timescale dynamics of the ETS domain of ETV4$^{328-430}$ and PU.1$^{167-272}$.

Amide ^{15}N T$_1$ (top) and T$_2$ (middle) lifetimes and steady-state heteronuclear NOE values of (A) ETV4$^{328-430}$ with a partially truncated H4 and (B) PU.1$^{167-272}$. The standard deviations of the fit exponential decays are approximately 5%.
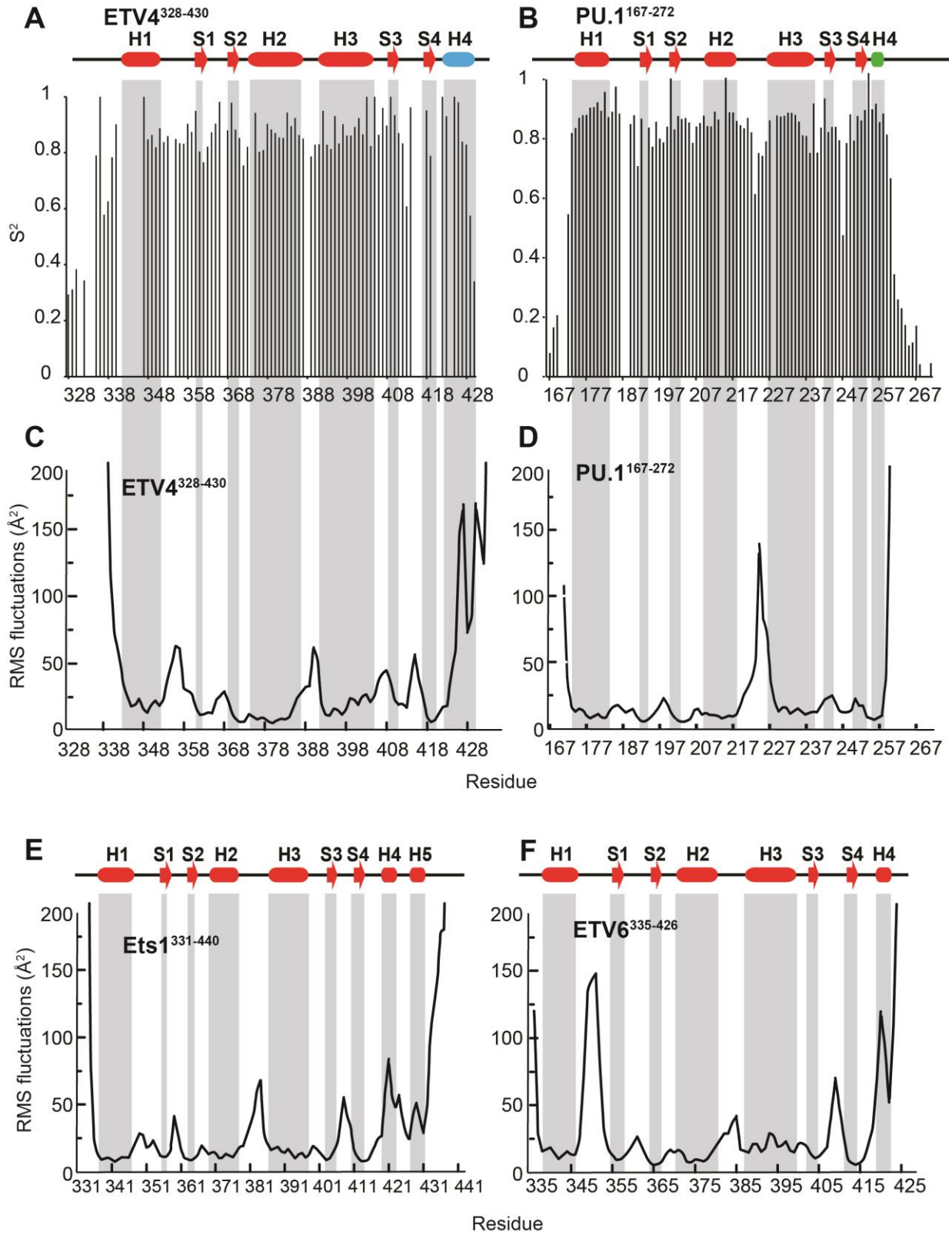
Figure 4-6 Experimental order parameter (S^2) are in good agreement with MD simulations revealing flexibility of the ETS domain.

Model-free order parameter (S^2) of (A) ETV4[328-430] and (B) PU.1[167-272] calculated from amide ^{15}N heteronuclear NOE, T_1 and T_2 data. Decreasing S^2 values indicate increasing mobility on the sub-ns timescale. Also shown are mean squared fluctuations calculated from MD simulations. (C) ETV4[337-436] displays high fluctuation around helix H4 and (D) PU.1[167-272] has significant fluctuations in the loop between helix H2 and H3 that correspond to the lower S^2. Mean squared fluctuations of (E) Ets1[331-440] and (F) ETV6[335-426] have similar patterns with increased flexibility in loop regions. However, the magnitudes of these mean squared values differ between the four ETS proteins.

4.2.4. Probing ETS domain stability and dynamics with amide HX

To further characterize the dynamic properties of the ETS domains, we investigated ETV4 and PU.1 using amide HX experiments. Rapid amide ^{1}H/^{1}H HX was detected by the CLEANEX magnetization transfer approach (Hwang et al. 1998). This approach requires that exchange occurs on the seconds timescale, and thus a range of sample pH values (5.7 to 8.5) was used to characterize amides with protection factors (PFs) spanning from ~ 1 to 1000. For more protected amides, ^{1}H/^{2}H exchange was measured using short (min-hr) ^{15}N-HSQC spectra recorded over several days. Combining data from the two approaches enabled me to determine the PFs for most amides in the ETS domain of ETV4 and PU.1 (Figure 4-7). A PF is the ratio k_{pred}/k_{ex}, where k_{ex} is the measured HX rate constant for a given amide and k_{pred} is the predicated rate constant for a corresponding random-coil polymer under the same conditions of pH, temperature, and solvent (Coyne et al. 2012). Note that merging data from several experiments requires that exchange occurs in the commonly observed pH-dependent EX2 regime and that the stability and dynamics of the protein do not change significantly over the pH range examined. Although not rigorously demonstrated, k_{ex} values were approximately first-order in hydroxide concentration (pH), as expected for the EX2 regime. In this regime, protein conformational fluctuations occur faster than exchange, and thus a PF is the inverse of an equilibrium constant between a "closed" state, where HX cannot occur due to factors such as hydrogen bonding, and an "open" exchangeable state (Li & Woodward 1999). PFs can provide a residue-specific measure of the free energy landscape allowing exchange [$\Delta G°_{HX} = RT\ln(PF)$] and thus insights into the local and global stability of the protein. A positive change in free energy indicates the unfolding of the protein.

HX measurements revealed very similar behaviors for ETV4$^{328-436}$ and PU.1$^{167-272}$ (Figure 4-7). In both proteins, the stable core of the ETS domain is formed by residues in helix H1 and strands S1 and S2. These secondary structural elements have the highest protection factors (PF ~ 6.4×10^4 for ETV4$^{328-436}$ and ~3.5×10^4 for PU.1$^{167-272}$) and likely exchange via a global unfolding pathway (Woodward & Li 1998). The corresponding $\Delta G°_{HX}$ ~ 6.5 kcal/mol and ~ 6.2 kcal/mol for ETV4$^{328-436}$ and PU.1$^{167-272}$, respectively,

provide the lower limits on the unfolding free energy of the proteins under "native" conditions. These results are consistent with previous studies of Ets1[301-440] and ETV6[335-426], which also demonstrated that helix H1 and strands S1 and S2 form the stable core their ETS domains (Figure 4-7). The PFs and $\Delta G°_{HX}$ values for all four ETS factors are summarized in Table 4-3. Due to incomplete amide HX data collection of ERG, the PF and thus the $\Delta G°_{HX}$ were not determined.

A comparison of the HX profiles of the four ETS factors yields several additional insights into the local dynamics of these proteins (Figure 4-7). As expected, residues at the termini of the fragments and in exposed loop regions showed little protection. This is consistent with several complementary measures of dynamics, including chemical shifts, ^{15}N relaxation, and high rmsd values in NMR-derived structural ensembles. The appended N-terminal and C-terminal helices, such as H4 showed intermediate PFs, indicative of reduced local stability relative to the ETS domain core. Perhaps most interestingly, the DNA recognition helix H3 of all four ETS factors had PF values ~ 10^3. Thus, although well defined in structural models determined by X-ray crystallography and NMR spectroscopy, these helices undergo local conformational transitions detectable by HX. This appears to be a conserved feature of the ETS domains, and as discussed below, likely reflects a plasticity needed for DNA binding.
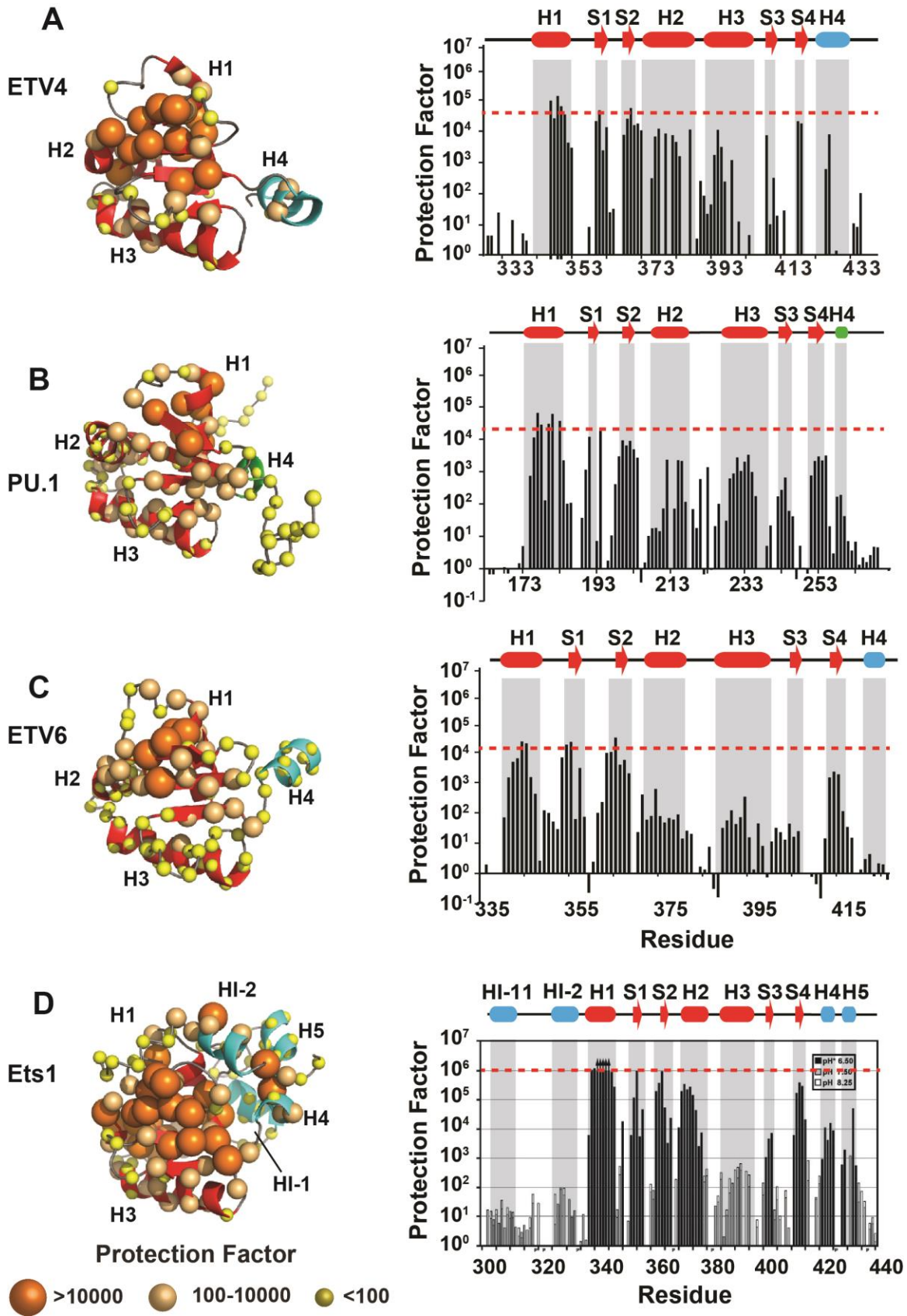
Figure 4-7 Protection factors reveal the stable core of the ETS domain and the dynamic helices H3 and H4.

Amide HX protection factors of (A) ETV4[328-436], (B) PU.1[167-272] (C) ETV6[335-426] and (D) Ets1[301-440] are plotted as histograms and mapped onto the crystal structures of the proteins using spheres with the indicated size/color scale. Dashed red lines represent the average max PFs used to calculate $\Delta G°_{HX}$ listed in Table 4-3. ETV4 protection factors are reproduced from Figure 3-11 in Chapter 3. ETV6 PFs are reproduced from (Coyne et al. 2012) and Ets1 are reproduced from (Lee et al. 2005). The arrowheads in the latter plot indicate lower limits. Missing values correspond to prolines, residues with unassigned or overlapped NMR signals, and residues exchanging too slowly to be measured with CLEANEX-PM (sec timescale), yet too fast to measure via $^{1}H/^{2}H$ exchange (> hours).

Table 4-3 Protection factors and $\Delta G°_{HX}$ values for the ETS factors

	Ets1[301-440]	ETV4[328-436]	ERG	ETV6[335-426]	PU.1[167-272]
T_m (°C) [a]	59 ± 1	48 ± 1	57 ± 1	66 ± 2	48 ± 1
PF_{max} [b]	>1.0 x 10^6	6.4 x 10^4	NA	2.2 x 10^4	3.5 x 10^4
$\Delta G°_{HX}$ (kcal mol^{-1}) [c]	>8	6.5	NA	5.9	6.2

[a] T_M values for thermal denaturation from Table 4-1
[b] For each protein, the 10 highest PFs were averaged to obtain PF_{max} ± standard deviation.
[c] $\Delta G°_{HX} = RT\ln(PF_{max})$ at 25 °C.

4.2.5. MD simulations of ETS domains also reveal backbone dynamics

NMR relaxation and HX experiments demonstrated that ETS domains possess an intrinsic ability to undergo conformational transitions. Importantly, the DNA-recognition helix H3 and the inhibitory helices are dynamic (Coyne et al. 2012; Lee et al. 2005). To integrate these data into a unified description of the structural/functional properties of ETS domains, we turned to molecular dynamics (MD) simulations. These simulations were carried out to ~ 900 ns for the uninhibited ETS domains of Ets1$^{331-440}$, ETV6$^{335-426}$, ETV4$^{337-434}$ and PU.1$^{171-258}$. These fragments were chosen because their structure is readily available and had the closest boundaries to our studied fragments. After an initial equilibration period, the conformations sampled over the course of the simulations remained stable and did not differ substantially from the energy-minimized structures (average rmsd ~1.5 Å of N, Cα, CO) (Figure 4-8).

The MD trajectories were converted to residue profiles of the backbone positional fluctuations. These fluctuations are shown as plots in Figure 4-6C-F and mapped on the structures of the ETS domains in Figure 4-9. Over the ~ 900 ns sampled, the secondary structural elements of all four proteins underwent minimal conformational changes, indicating that they form a stable core of the ETS domains. In contrast, residues at the termini and loop regions of the ETS factors showed a range of fluctuations in the MD simulations. This is particularly noticeable for those in the "turn" between helix H2 and H3 and the "wing" between strands S3 and S4 (Figure 4-6C-F). A qualitative comparison of the mean-squared fluctuation plots indicates that the motions of the "turn" decrease in the order PU.1 >> ETV4 ~ Ets1 > ETV6, whereas those of the "wing" decrease as ETV6 > Ets1 ~ ETV4 >> PU.1. The distinct behavior of PU.1 might result from the fact that it has an extra glycine and three positively charged residues in the "turn" (GNRKKM), while also lacking an otherwise conserved glycine in the "wing" (Figure 4-1). In contrast, the other ETS factors have fewer charged residues, and/or a proline, in their "turns", plus an additional glycine in their "wings" (Figure 4-1). Although PU.1 has more charged residues in the "wing" that has less motion, this suggest that the extra glycine contributed to the fluctuations seen in MD simulation. This is consistent with the fact that glycine allows a

large degree of dihedral angles that permit different conformations. According to crystallographic studies, the conformations of the "turn" and the "wing" change upon major groove binding by the DNA-recognition helix H3 in order to provide flanking contacts to the DNA phosphodiester backbone (Garvie et al. 2002; Garvie et al. 2001). Thus, as will be discussed below, the differing preferences of these ETS factors for flanking nucleotides outside of the core $^{5'\text{-}}$GGAA$^{-3'}$ (Wei et al. 2010) may reflect the conformational flexibility of their "turn" and "wing" loops.

The loop between helix H1 and strand S1 of ETV6 also fluctuated significantly. This is seen to a lesser extent with ETV4, but not PU.1 nor Ets1. The enhanced fluctuations in this region of ETV6 may corresponded to the greater length of the loop (Figure 4-1). However, the H1-S1 loop has not been implicated in any functional role.

Overall, the results of the MD simulations qualitatively parallels trends in S^2 values from ^{15}N relaxation and RCI-S^2 values from chemical shift analyses (Figure 4-6A,B). Thus MD calculations appear to recapitulate the fast timescale dynamics of ETS domains observed under real experimental conditions. However, it is noteworthy that the DNA-recognition helix H3 does not show enhanced mobility relative to the other core helices or strands. This differs from the results of HX measurements and most certainly reflects limited conformational sampling over the relatively short time scale of the MD simulations.
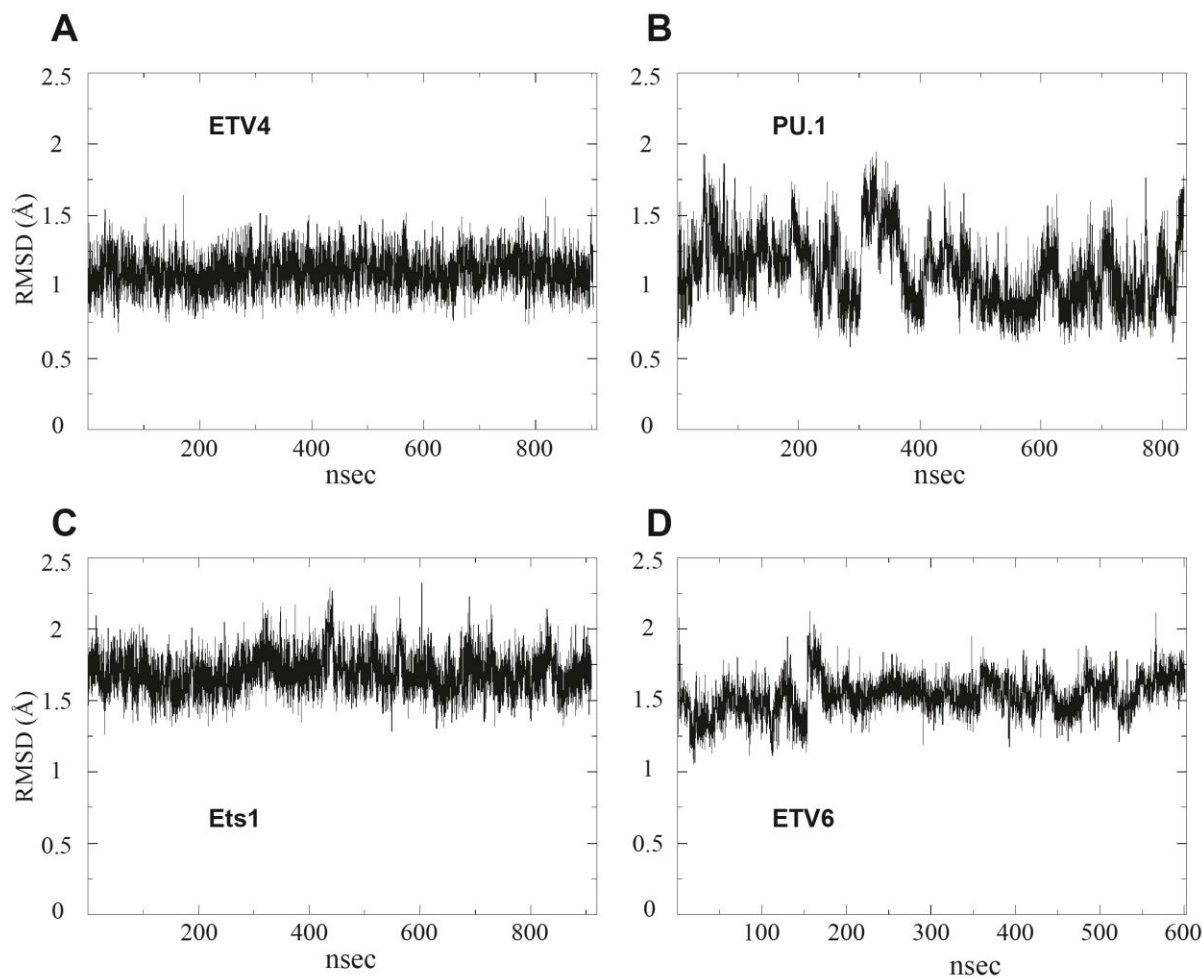
Figure 4-8 Time profile RMSD fluctuations for the ETS domains.

The amount of RMSD changes of backbone atoms (N, Cα, CO) during MD simulations. Each protein was initially equilibrated (see methods) and their backbone atoms fluctuations remain stable before the full MD production calculations.
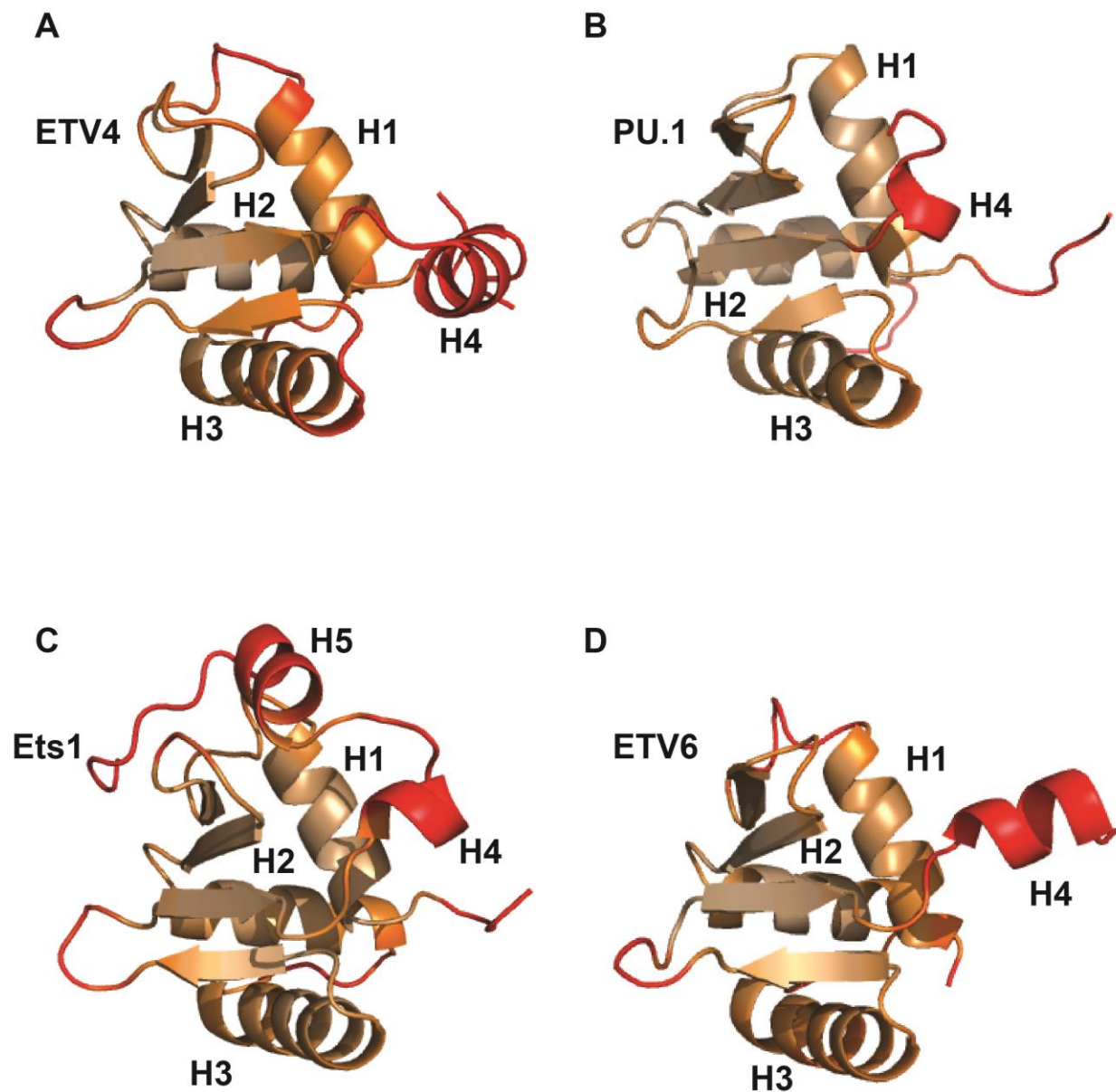
Figure 4-9 RMS fluctuations mapped onto the ETS domains.

The RMS fluctuation plotted in Figure 4-6 are mapped onto the structures of (A) ETV4[337-434], (B) PU.1[167-262], (C) Ets1[331-440], and (D) ETV6[335-426]. Light gray-orange indicates RMS fluctuation values < 30 Å^2; orange, RMS 30-50 Å^2: and red, RMS >50 Å^2.

4.2.6. MD simulations indicate that motions within the ETS domain are coupled

We used cross-correlation analysis (Sethi et al. 2009) to understand how motions throughout the ETS domain might be coupled to the DNA-recognition helix H3. Such correlated dynamics might provide routes for autoinhibition or other forms of allosteric regulation. Coupled motions in the ETS domains were identified by normalizing the cross-correlation matrixes of atomic fluctuations over the lengths of the simulations. A strong positive correlation indicates that the two atoms move in the same direction on the same axis, while anti-correlation means the two atoms move in the opposite direction on the same axis. Two atoms moving in other directions on different axes are considered not correlated. As expected, for each ETS domain, there was strong local correlation within each secondary structural element. This is seen along the diagonals of cross-correlation maps for α-helices and as anti-diagonals for anti-parallel β-strands. Besides local correlations, the more global motions of Ets1 were the least coupled throughout the protein whereas ETV4, PU.1 and ETV6 had various (anti)correlation regions (Figure 4-10). However, the effects appear modest and differ with each protein. With the focus on the DNA-recognition helix H3 (bracketed with lines), the MD-calculated motions are mostly anti-correlated (blue color) with other parts of the ETS domain. For example, motion of helix H3 of PU.1 is moderately coupled to the strand S3 and S4 region (Figure 4-10B).

Although correlation analysis provides insight into possible allosteric effects, the communication pathways between different parts of the ETS domain cannot be elucidated by solely using these methods. Therefore we sought to identify the pathways and the residues critical for communication by using a dynamic network analysis (Bui & Gsponer 2014). Dynamic network analysis specifies a node or community to represent residues that move together. Each node is connected by an edge if they are in contact during a majority of the simulation. In the dynamic network, the edges are weighted by the correlation values from the simulations so that the width of the edges increases as the

161

correlation (or energy of interaction) between the nodes increases (Sethi et al. 2009). The ETS domains from the four ETS factors are split into 5 communities with overall similar community structures as shown in Figure 4-11 and Figure 4-12. In general, each α-helix is represented with a community. The β-strands are grouped into one community for ETV4 and ETV6, and into two closely connected communities for Ets1 and PU.1. As illustrated in Figure 4-11 and Figure 4-12, the β-sheet acts as a central hub connecting all other communities. Thus, the start of strand S3 and the end of strand S4 form one central critical pathway for relaying dynamical information to the DNA-recognition helix H3.

In the case of ETV4, Phe420 located near the end of strand S4 is identified as a critical residue that relays information from helix H4 to the DNA-recognition helix H3 (Figure 4-11A). And as previously demonstrated (chapter 3), helix H4 of ETV4 regulates its DNA-binding ability by shifting the conformational positioning of the DNA-recognition helix H3. It is likely that altering Phe420 or the stability of the β-sheet would change the dynamics of the ETS domain and possibly the communication between the DNA-binding helix H3 and other parts of the ETS domain. This idea is further supported by the "in silico" mutation of the conserved Phe420 to Ala420 using Rosetta Design (Lyskov et al. 2013). The single mutation of F420A is predicted to alter the free energies of the ETS domain near the N-terminal of helix H1, C-terminal of helix H3 and helix H4 (Figure 4-13).
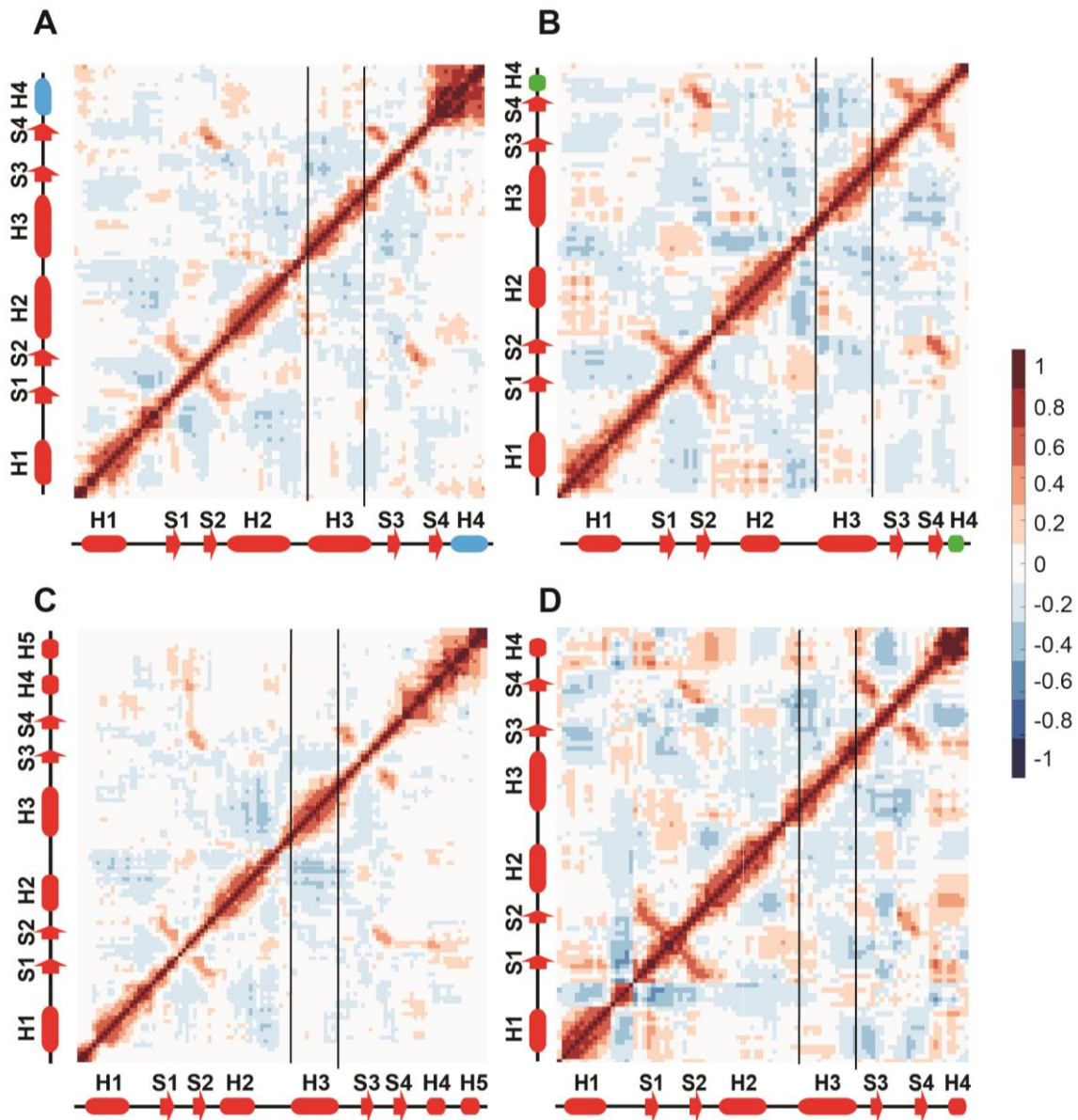
Figure 4-10 Cross-correlation map of ETS domains highlight the correlated/anti-correlated motions in MD simulations.

(A) ETV4$^{337-434}$. (B) PU.1$^{167-272}$. (C) Ets1$^{331-440}$. (D). ETV6$^{335-426}$. Motions that are (anti)correlated are indicated with red (blue). The vertical black lines delineate the DNA recognition helix H3.
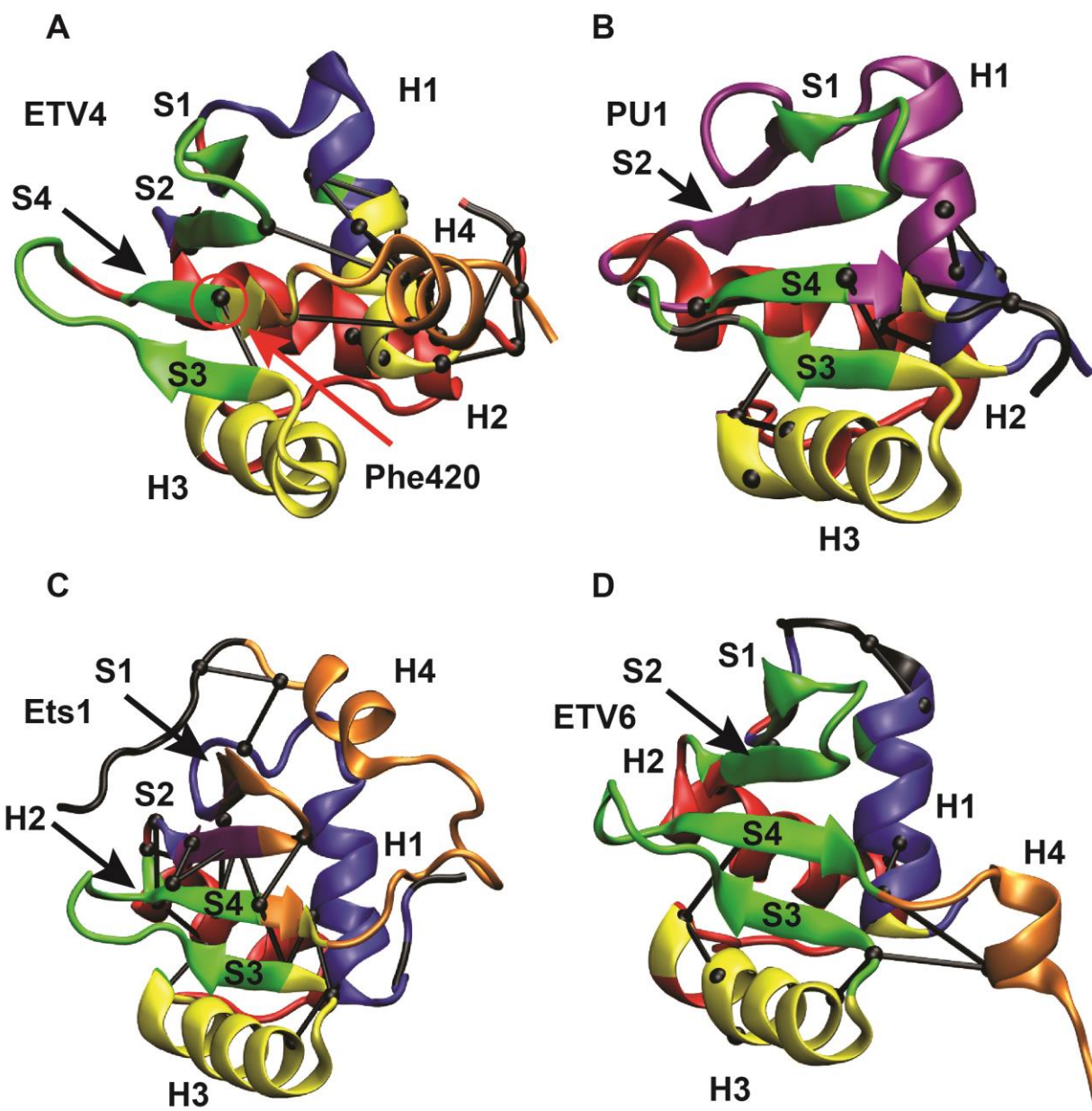
Figure 4-11 Dynamical network analysis highlight ETS domain in dynamic communities and critical pathway.

(A) ETV4$^{337-434}$. (B) PU.1$^{167-272}$. (C) Ets1$^{331-440}$. (D). ETV6$^{335-426}$. Each color represent one community of residues that moved together in MD simulations (see also Figure 4-12). Residues that are part of small communities (3 or less residues) are colored in black. Critical paths with the highest "betweenness" pathway connecting two communities are colored as black nodes and edges. Betweenness is defined as the number of shortest paths that cross a given edge.
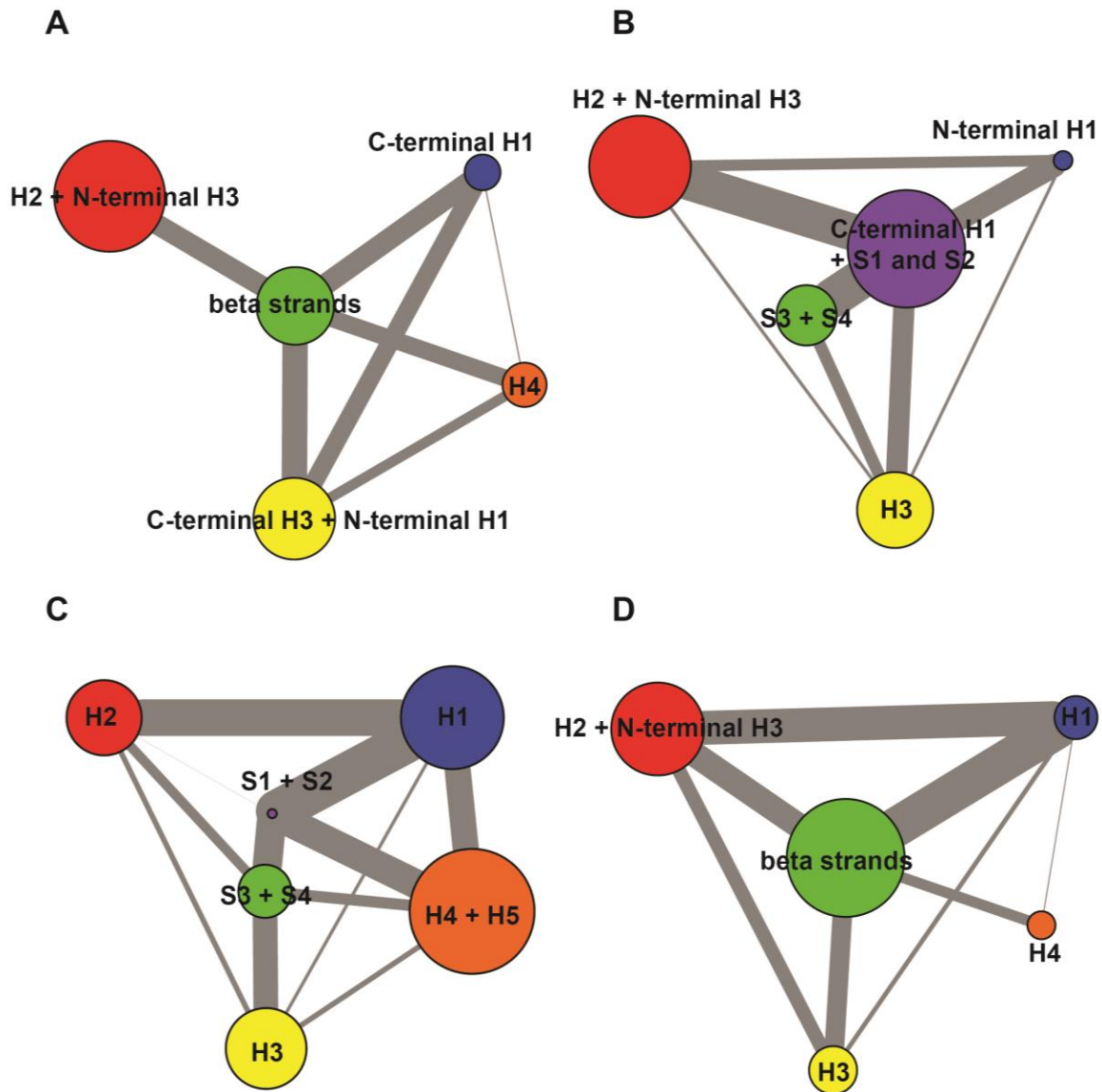
Figure 4-12 Dynamic network analysis showing communities and communication pathways.

(A) ETV4$^{337-434}$. (B) PU.1$^{167-272}$. (C) Ets1$^{331-440}$. (D). ETV6$^{335-426}$. Simple graphical representation of community and connectedness. The size of nodes represents the size of a community. The width of edges represents "betweenness" on the critical path connecting the two communities. The same color coding was used for Figure 4-11 and Figure 4-12.
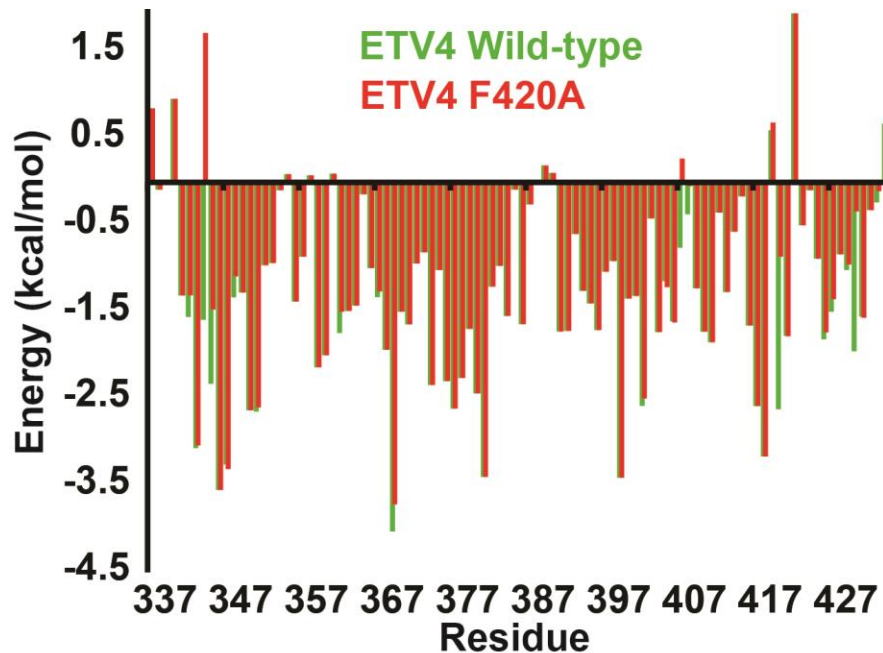
Figure 4-13 Rosetta design predicted F420A to destabilize the ETS domain.

Mutation from phenylalanine to alanine at position 420 of ETV4 (red) reduces the hydrophobic packing (see also Figure 3-27). The energies at position 420 and residues nearby in space (~ 340 and ~ 427) are reduced compared to wild-type (green) as calculated by Rosetta Design (Liu & Kuhlman 2006) (red is less negative than green). This suggests that the critical residue F420 identified by dynamical network analysis plays an important role in stabilizing the protein.

4.3.   Discussion

The structural and functional roles of the ETS domain have been extensively characterized (Hollenhorst, McIntosh, et al. 2011). The ETS domain is a DNA-binding module with a winged helix-turn-helix motif that recognizes specific DNA sequences. Although the core ETS domain is highly conserved, flanking sequences are more divergent, with conservation limited to subfamily members. In particular, helices appended on the ETS domain contribute to functions including DNA-binding autoinhibition. In this study, I chose to study fragments from 4 different ETS factors corresponding to their minimal or near minimal ETS domains. These fragments exhibited no or very modest DNA-binding autoinhibition.

The protein structure-function paradigm is constantly evolving as the critical roles of conformational dynamics are better elucidated. In this chapter, I have focused on the relationship between ETS domain structure/dynamics and the potential implications for DNA-binding. I initially measured the thermal stabilities of the uninhibited ETS domains of ETV4, PU.1, Ets1, ETV6 and ERG. Combining the results with HX experiments, the least stable ETS domain of ETV4 and PU.1 have T_m values of 48 °C and global unfolding free energies ($\Delta G°$) of ~ 6.5 kcal/mol and ~5 kcal/mol, respectively. I have determined the tertiary structure of PU.1 in solution and identified a short α-helix H4 appended to the conserved ETS domain. Through molecular dynamic simulation, we found that the C-terminal helices of ETV4, PU.1, Ets1 and ETV6 ETS domain are flexible and that motions of the DNA-recognition helix H3 are relayed through the β-strands. MD simulations also revealed that the motions "turn" between helix H2 and H3, as part of the DNA-binding interface, differs between ETS domains. This is also seen with NMR relaxation experiments and may be correlated with the size of the turns, their glycine/proline content and their electrostatic features (Figure 4-14). Lastly, several critical residues including Phe420, are hypothesized to stabilize the ETS domain of ETV4 and potentially impact DNA-binding autoinhibition by altering the dynamics.
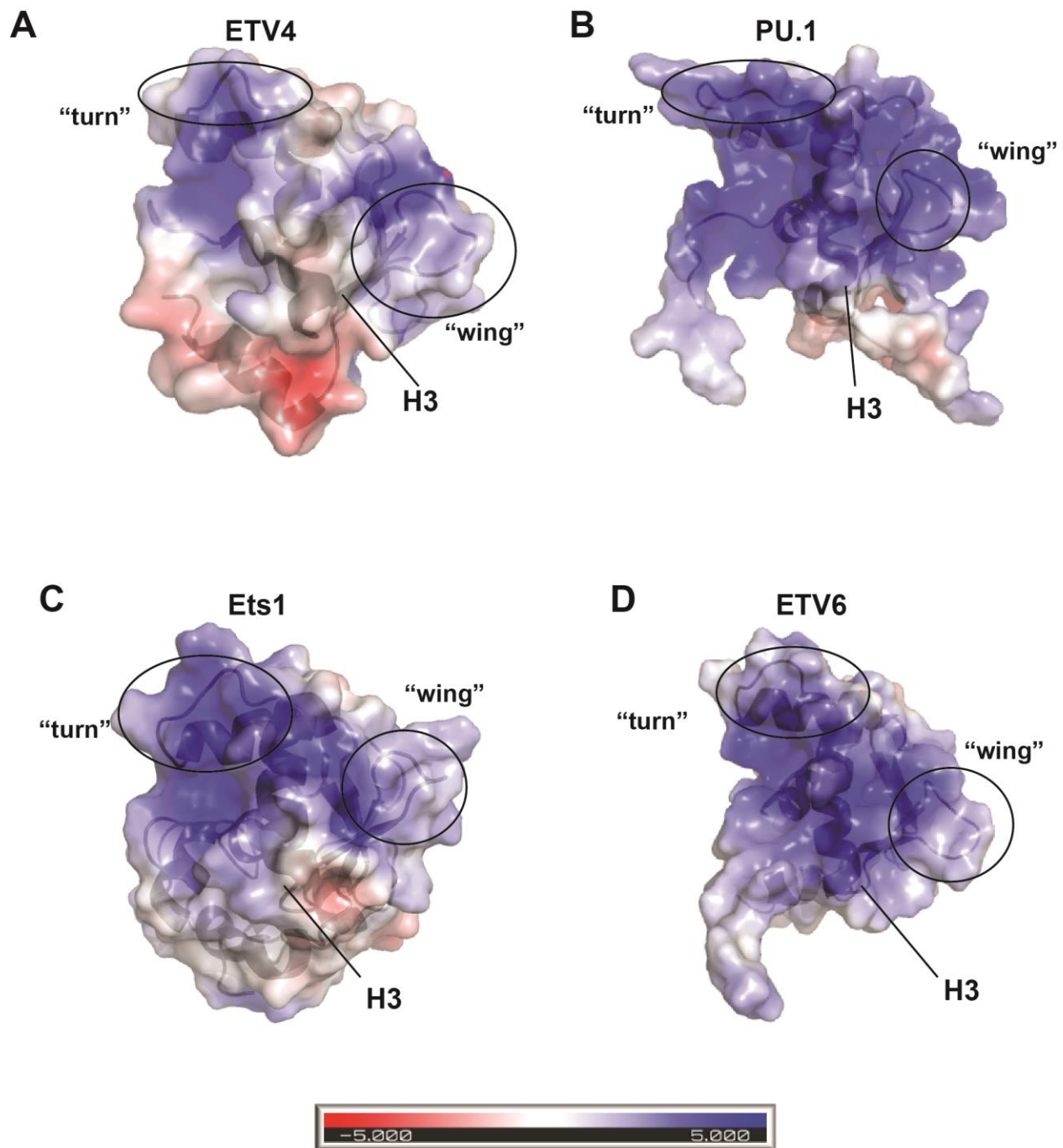
Figure 4-14 Electrostatic map of the ETS domains.

(A) ETV4$^{337-434}$. (B) PU.1$^{167-272}$. (C) Ets1$^{331-440}$. (D) ETV6$^{335-426}$. With the DNA-binding helix H3 facing front, it shows that the DNA-binding site is mostly positively charged (blue). Circled regions are the "turn" between helix H2 and H3 and the "wing" between strand S3 and S4.

4.3.1. Structure of the PU.1 ETS domains

Using NMR spectroscopic methods, I determined the solution structure of PU.1 and identified an α-helix H4 appended C-terminal to the ETS domain. This helix H4 spans from residues Gly256 to Gly260 and was not reported in the X-ray structure of PU.1 bound to DNA (PDB: 1PUE) (Kodandapani et al. 1996) likely because their PU.1 fragment ended at residue 259. However, helix H4 of PU.1 was reported in complex with transcription factor IRF4 and DNA (Escalante, Shen, et al. 2002; Escalante, Brass, et al. 2002). This helix H4 also has limited protection factors, indicative of frequent local unfolding. However, PU.1 is not known to be autoinhibited and a fourth helix or helical turn is seen with most ETS domain (Hollenhorst, McIntosh, et al. 2011), so it is best viewed as part of the complete ETS domain fold.

4.3.2. Relative stabilities of the ETS domains

Despite their conserved sequences and structures, I observed a substantial ~ 20 °C difference in the T_m values (~ 20 °C range) and $\Delta G°_{HX}$ values (~ 2 kcal mol^{-1}) between the four ETS factors. The trends in the mid-point unfolding temperatures obtained using circular dichroism spectroscopy to monitor global unfolding correlated with the trends in the free energies calculated from HX protection factors. In general, a lower T_m value corresponds to a lower $\Delta G°_{HX}$ detected by amide HX (Figure 4-15). One notable exception is seen with ETV6, with a rather high T_m value. However, the thermal denaturation data of the ETS domains were evaluated assuming a two-state transition in which the proteins are assumed to exist only in their native or denatured states. The unusually broad denaturation curve observed by circular dichroism spectroscopy for ETV6 suggests that the folding/unfolding transition of this protein is not as cooperative as the other ETS domains (Malhotra & Udgaonkar 2016). The reasons for this difference is difficult to define as all of the ETS factors adopted well-defined structures under non-denaturing conditions.
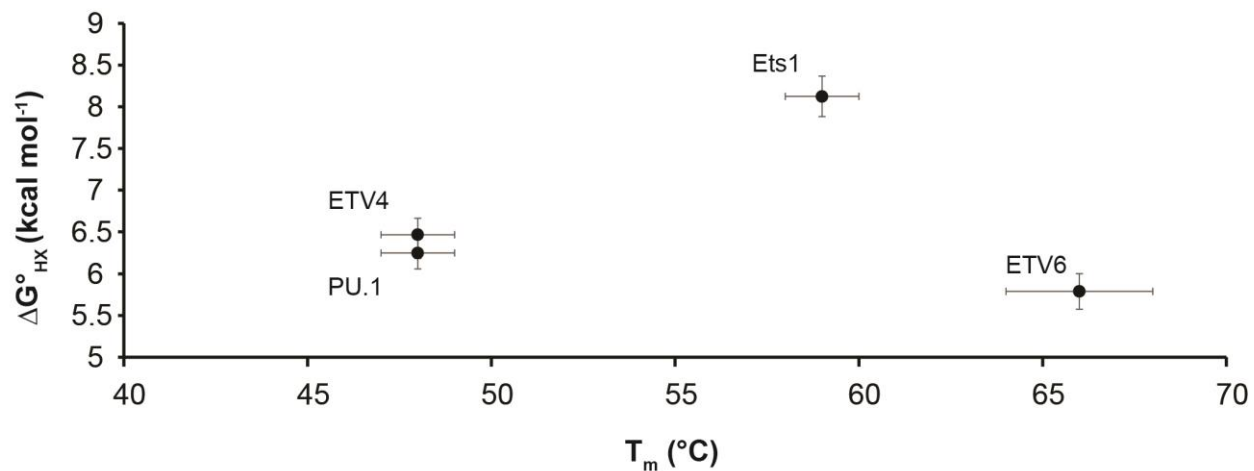
Figure 4-15 The midpoint unfolding temperature vs ΔG°HX of unfolding for the four ETS factors.

Data from Table 4-3. The T_m errors are from the curve fitting to a two-state transition model. A 5% error is assumed for ΔG°HX.

Understanding the structure, stability and dynamics of a protein can provide fundamental knowledge to aid future research to manipulate its function. For example, protein engineering with the use of directed evolution identified stable mutants of apoptotic protein IFI16 with enhanced protein function to destabilize double-stranded DNA (Lau et al. 2016). Indeed, mutations introduced to stabilize the C-terminal inhibitory helix H5 of ETV6 led to reinforced DNA-binding autoinhibition (De et al. 2016). This suggest that autoinhibition of Ets1 and ETV4 might also be modulated through changes, including mutations and posttranslational modifications.

4.3.3. ETS dynamics and DNA binding

A thorough understanding of a protein's function requires the investigation of its dynamics; that is, its time-dependent conformational changes (Henzler-Wildman & Kern 2007). Indeed, ETS domains have fast motions in the range of ps-ns and slower motions in the order of seconds to hours, and these motions are dampened upon autoinhibition (Lee et al. 2005; Green et al. 2010; Regan et al. 2013). Here we demonstrated that the ETS domains of ETV4 and PU.1 have similar dynamics features as Ets1 and ETV6. In particular, the DNA recognition helix H3 is flexible, as evidenced by modest PF's ~ 1000, whereas the stable core of the ETS domain is composed of helix H1 and the β-strands. Although we performed MD simulations to ~900 ns (long by computational standards, but still relatively short on the timescale of protein motions), these did not show significant backbone fluctuation on helix H3. This suggests that the motions of helix H3, detected by amide HX, are relatively slow (>μs). This is not surprising because slower domain motions on the μs – ms timescale are likely most biologically relevant, because they are close to the timescales on which fundamental processes such as docking, protein folding, and allosteric transitions occur (Akke 2002). In contrast, substantial conformational fluctuations were calculated for the "turn" and "wing" loops. This is consistent with the sub-ns timescale motions of these regions of the ETS domains detected via amide ^{15}N relaxation measurements.

The flexibility of a DNA binding domain is proposed to be important in facilitating the search for specific target sequence by allowing to facile interactions with non-specific DNAs (Kalodimos, Biris, et al. 2004). The ability to interact with both cognate and non-specific DNA has been shown for Ets1 and ETV6 to occur via the same canonical ETS domain interface (Desjardins et al. 2016; De et al. 2014). However, nonspecific DNA binding results from dynamic electrostatic interactions, whereas specific DNA binding is dependent upon well-defined hydrogen bonding interactions between the protein sidechains and the DNA bases. A similar behavior is likely to be true for PU.1 and ETV4 given their comparable overall dynamics.

Consistent with protection factors and order parameters from NMR relaxation measurements, MD simulations reveal that the backbone of the terminal helices and loop regions sample multiple conformations. The "turn" between helix H2 and H3 of PU.1 fluctuates significantly, in agreement with the reduced S^2 order parameters of residues forming this loop region. The flexibility of the "turn" has been suggested in other helix-turn-helix motifs, such as the *lac*R repressor, to be required to allow correct positioning of the recognition helix in the major groove of the DNA while also providing flanking contacts to the phosphodiester backbone (Kalodimos, Boelens, et al. 2004). Here, we propose that flexibility of the "turn" is dictated by its size, glycine/proline content, and charge state such that it can help to differentiate binding sites through indirect read-out of the variable residues flanking the $^{5'}$GGA(A/T)$^{3'}$ core (Figure 4-16). This also suggests an important link between dynamics and DNA-binding by the ETS domain.

An emerging theme of molecular hydration and targeting epigenetically modified DNA is linked to the unique affinity and specificity of the ETS domains (Poon 2012; Poon & Macgregor 2004). Detailed investigation of the crystal structures of the ETS domain of Ets1 and PU.1 in complex with DNA revealed that PU.1 utilized more water mediated contacts. As such, PU.1 is more sensitive to osmotic pressure and was determined to have different kinetics to DNA binding compared to Ets1 (Poon & Macgregor 2003; Poon & Macgregor 2004; Poon 2012). How this unique physical chemistry is linked to dynamics

remains to be determined, and may provide new insights into the paradigm of structure-dynamics-function of the ETS domain.

Through dynamical network analysis, it highlights the communication pathway between different parts of the ETS domain. The β-strands serve as a hub to relay dynamical information as indicated by the connectedness to other communities. In particular, motion of helix H3 is relayed via the strand S3 and S4 and mutating the conserved phenylalanine located near the end of strand S4 destabilizes the ETV4 ETS domain. Although speculative, altering the dynamics of ETS domain allosterically can attenuate the motions of the DNA recognition helix H3, likely via the β-sheet that lead to alternative specificity of the ETS domain. Overall, our data support the notion that these motions are conserved features of all ETS proteins.
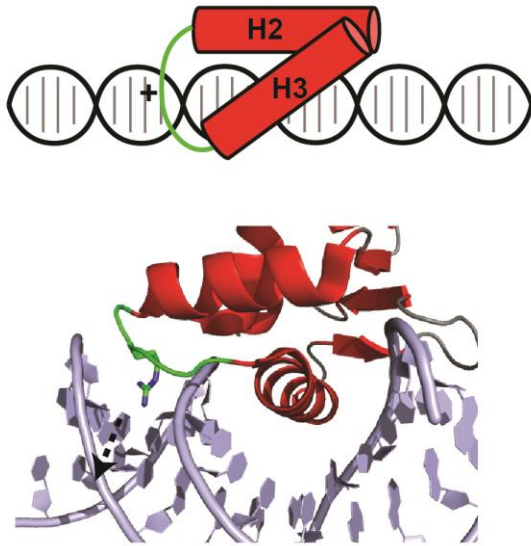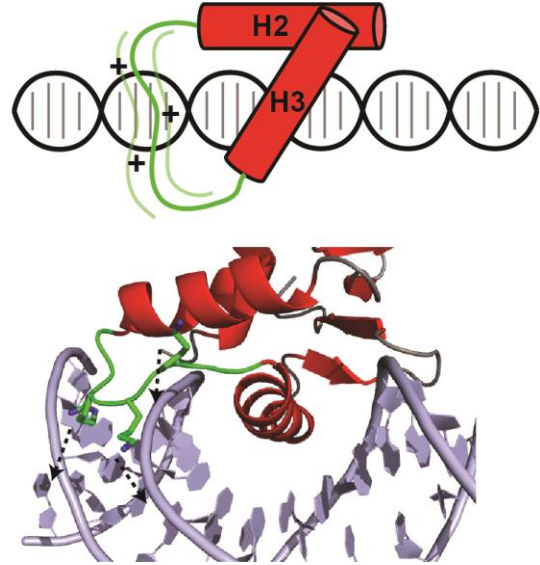
Figure 4-16 Electrostatic and conformational freedom model explaining the flexibility of the "turn" between helix H2 and H3 of ETS domain.

The "turn" of ETV6 has a single arginine, whereas PU.1 has one arginine and two lysines plus an extra glycine. These factors provided the contact needed between the "turn" and the phosphodiester backbone as seen in the crystal structures (ETV6: 4MGH.pdb, PU.1: 1PUE.pdb). MD simulations and NMR relaxations revealed that, in the absence of DNA, the "turn" of PU.1 fluctuates more than the one of ETV6. This provides a possible link between dynamics and DNA-binding.

4.4. Materials and methods

4.4.1. Expression plasmids and protein purification

The sequences of the ETS factors used in this study are shown in Figure 4-1. Ets1[301-440] and ETV6[335-426] have been described previously (Desjardins et al. 2014; Coyne et al. 2012; Lee et al. 2005; Lee et al. 2008; De et al. 2014). The cDNAs encoding the uninhibited ETS domains of human ETV4 (residues 337-436) and mouse ERG (residues 307-407) were cloned into the bacterial expression vector pET28 (Novagen) using PCR amplification and restriction sites ligation. The gene encoding the mouse PU.1 ETS domain (residues 161-272) was provided by Gregory M. K. Poon (Washington State University) (Wang et al. 2014) in a pQE-60 plasmid and subsequently cloned into the pET28MHL expression vector.

Samples of the ETV4 ETS domains were expressed and purified as described in Chapter 3. A similar protocol was used to purify ERG ETS domain. The uninhibited ETS domains of Ets1 (residues 301-440) and ETV6 (residues 335-426) were expressed and purified using published protocols (Desjardins et al. 2014) (De et al. 2014).

For the PU.1 ETS domain, the plasmid containing this protein was transformed into *E. coli* BL21(λDE3). Cultures were inoculated in 1 L M9 media with 1 g of (^{15}N, 99%)-NH$_4$Cl (and for ^{13}C labeling, 3 g of (^{13}C$_6$, 99%) D-glucose) and grown at 37 °C until the OD$_{600}$ reached 0.6. A final concentration of 1 mM IPTG was added to induce protein expression, followed by 4 hrs of growth at 37 °C. Cells were harvested and resuspended in 20 mM sodium phosphate, 0.5M NaCl, 4 M guanidinium-HCl, 20 mM imidazole, pH 7.4, and 0.2x Protease Inhibitor Cocktail (Roche). The cells were then lysed by 30 minutes of sonication (50% duty cycle max) and cell homogenization with EmulsilFlex-C5 (5 passages) at 4 °C. After centrifugation, the supernatants were loaded by passing twice onto a Ni^{2+} affinity column (GE Healthcare Life Science). After washing the column with washing buffer (20 mM sodium phosphate, 0.5M NaCl, 4 M guanidinium-HCL, 60 mM imidazole, pH 7.4), the

His$_6$-tagged protein was eluted with the washing buffer plus 1 M imidazole. The purified denatured PU.1[167-272] was refolded by overnight dialysis in refolding buffer (20 mM sodium phosphate, 150 mM NaCl, and 0.1 mM EDTA, pH 7.5). After dialysis, the sample was spun at 4000 g to pellet any precipitate, which was then discarded. TEV protease (20 µL of 200 µM) was added to the supernatant (what volume), followed by dialysis in refolding buffer, plus 1 mM TCEP as required to maintain TEV activity. Cleavage of the His$_6$-tag was verified by mass spectrometry (MALDI-TOF) and SDS-PAGE. Cleaved PU.1[167-272] was concentrated to a volume < 2 mL and further purified using Superdex S75 gel filtration with a running buffer consisting of 20 mM potassium phosphate, 150 mM KCl and 50 µM EDTA, pH 5.5. Eluted fractions were analyzed by SDS-PAGE and dialyzed into NMR buffer (20 mM potassium phosphate, 150 mM KCl and 50 µM EDTA, pH 5.5). The final, purified protein samples were then concentrated on a 3-kDa MWCO Centricon device, snap-frozen with liquid nitrogen, and stored at -80 °C for future use.

4.4.2. Circular dichroism spectroscopy

Circular dichroism (CD) spectra were measured from 190 nm to 260 nm using a JASCO-J-810 spectropolarimeter. Data were recorded at 25 °C using a 100 nm/min scan rate, 100 mdeg sensitivity and 0.1 s response time. The protein samples (10 µM, 0.1 cm path length cell) were in the following buffers: ETV4[337-436] (20 mM sodium phosphate, 200 mM NaCl, 2 mM DTT, 0.1 mM EDTA, pH 6.5), Ets1[301-440] (20 mM MES, 50 mM NaCl, 2 mM DTT, pH 6.5), ETV6[335-426] (20 mM sodium phosphate, 50 mM NaCl, pH 6.5), PU.1[167-272] (20 mM potassium phosphate, 150 mM KCl and 50 µM EDTA, pH 5.5), and ERG[307-407] (20 mM sodium phosphate, 200 mM NaCl, 2 mM DTT, 0.1 mM EDTA, pH 6.5). These buffers were chosen to match those used for NMR spectroscopic studies of the various ETS domains. CD signals were converted to mean residue ellipticity using the equation

$$\theta MR = (\frac{\theta}{10})/(conc \; x \; l \; x \; n)$$

Where Θ represents the observed signal, conc. represents the protein concentration, l represents the path length, and n represents the number of residues.

Thermal denaturation curves measured by monitoring the CD signal at 222 nm as the protein samples were heated from 20 °C to 95 °C at 0.5 °C/min. Thermodynamic parameters were obtained by fitting the resulting thermal denaturation curves to the linear extrapolation model. This assumes a two-state unfolding transition without a significant temperature dependence of ΔH° and ΔS° (e.g., $\Delta C_p = 0$) model using the following equations (Greenfield 2009; O'Shea et al. 1992):

$$\Delta G° = \Delta H° - T\Delta S°$$

$$\Delta G° = -RT\ln K$$

$$\alpha = K/(1 + K)$$

$$\theta_T = \theta_{fT} + \alpha(\theta_{uT} - \theta_{fT})$$

$$\theta_{fT} = \theta_f^° + m_f T$$

$$\theta_{uT} = \theta_u^° + m_u T$$

In these equations, θ_T is the measured CD signal as a function of temperature T, ΔG°, ΔH° and ΔS° are the free energy, enthalpy and entropy changes of unfolding, respectively. T_m is the midpoint unfolding temperature, K is the equilibrium unfolding constant, α is the fraction unfolded, $\theta_{f,T}$ and $\theta_{u,T}$ are the CD signals of the folded and unfolded states as a function of temperature, $\theta_f°$ is the CD signal of the fully folded form extrapolated linearly to T = 0, m_f is the temperature dependence of the CD signal of the folded protein, $\theta_u°$ is the extrapolated CD signal of the unfolded form at T = 0, and m_u is the temperature dependence of the CD signal of the unfolded protein. Combining these equations generates the expression below of θ_T versus T to be fitted in GraphPad Prism and thereby yield ΔH° and ΔS° (from which $T_m = \Delta H°/\Delta S°$):

$$\theta_T = \left(\theta_f^° + m_f T\right) + \left[\theta_u^° + m_u T - \theta_f^° - m_f T\right]\left[e^{\left(-\frac{\Delta H}{RT}\right)+\left(\frac{\Delta S}{R}\right)}/\left(1 + e^{\left(-\frac{\Delta H}{RT}\right)+\left(\frac{\Delta S}{R}\right)}\right)\right]$$

4.4.3. PU.1[167-272] structure determination by NMR spectroscopy

Standard heteronuclear scalar correlation NMR experiments were used to collected to assign signals from the ^{1}H, ^{13}C, ^{15}N nuclei in the backbone and side chains of uniformly ^{13}C/^{15}N-labeled PU.1[167-272] (Sattler et al. 1999). NOE-derived distance restraints were

obtained from a simultaneous three-dimensional ^{1}H-^{15}N/^{13}C-^{1}H NOESY-HSQC spectrum encompassing the aliphatics, aromatic, and amide regions (t_{mix} = 110 ms) (Pascal et al. 1994; Zwahlen et al. 1998). The data were recorded with a cryoprobe-equipped Bruker Avance 850 MHz spectrometer, processed with NMRpipe (Delaglio et al. 1995) and analyzed with NMRFAM-Sparky (Lee et al. 2015). The NMR-derived structure ensembles of PU.1$^{167-272}$ were calculated using CYANA 3.0 (Güntert 2004) and Ponderosa (W. Lee et al. 2011) using inputs including chemical shift assignments, dihedral angle restraints from TALOS+ (Shen et al. 2009), and unassigned NOESY cross-peaks. Structure calculations combined with automated NOESY spectra assignments were performed in seven iterative steps each yielding 100 structures. The final 20 lowest-energy structures were further refined with NMRe using explicit solvent and molecular dynamics simulations (Ryu et al. 2015). The chemical shifts and structural coordinates of PU.1$^{167-272}$ have been deposited in the BioMagResBank and the RSCB Protein Data Bank under accession codes 30303 and 5W3G, respectively.

4.4.4. ^{15}N relaxation experiments

Unless otherwise specified, all experiments were performed at 25 ºC in 95% sample buffer with 5% D_2O lock solvent, 1% NaN$_3$ (w/v) and 1% Protease Cocktail Inhibitor (Roche; 1 tablet dissolved in 1 mL). ETV4$^{337-436}$ was prepared and analyzed as described in Chapter 3. PU.1$^{167-272}$ was 0.18-0.35 mM in buffer containing 20 mM potassium phosphate, 150 mM KCl and 50 μM EDTA, pH 5.5.. Amide ^{15}N T_1, T_2 and NOE relaxation experiments (Farrow et al. 1994) were collected on ^{15}N-labled PU.1$^{167-272}$ and ^{15}N-labled ETV4$^{328-430}$ using a Bruker Avance 500 MHz and 600 MHz spectrometer, respectively. All spectra were processed and analyzed using Topspin, NMRPipe (Delaglio et al. 1995) and Sparky (Lee et al. 2015). The peak intensities were fit to a single exponential decay with Sparky to obtain T_1 and T_2 lifetimes. The heteronuclear ^{1}H-^{15}N NOE values were determined by taking the ratio of corresponding peak heights, acquired with and without ^{1}H saturation (Coyne et al. 2012). The Model-free (Lipari & Szabo 1982) global correlation

time and S^2 order parameters were calculated from T_1, T_2, NOE data using Tensor2 (Dosset et al. 2000).

4.4.5. Amide hydrogen exchange experiments

Protium-deuterium HX experiments for PU.1[167-272] were conducted at 20 °C with data recorded on a Bruker Avance 500 MHz spectrometer. Using a swinging bucket table top centrifuge, protonated ^{15}N-labeled protein was rapidly exchanged through a Sephadex G-25 spin column (~ 10 mL) equilibrated with the sample buffer, prepared in 99% D_2O and lacking NaN_3 and protease inhibitors. A series of ^{15}N-HSQC spectra were recorded with progressively longer time intervals spanning 7 days. Initially, three ^{15}N HSQC spectra were recorded with 2 scans/FID (roughly 10 mins per HSQC). Thereafter, 40 ^{15}N HSQC spectra were recorded with 6 scans/FID (roughly 30 mins per HSQC) without any intervening delays. Subsequently, spectra were recorded intermittently for six days after the transfer. The sample was removed from the magnet after 6 days and stored at room temperature (20 °C). It was intermittently returned to the spectrometer and ^{15}N HSQC spectra collected for 7 days. After completion of the experiment, the sample pH* (uncorrected pH meter reading) was measured as 5.54. The resulting data were fit with Matlab to the exponential decay function $I_t = (I_0 - I_\infty)e^{-(kex)t} + I_\infty$ where I_t is the intensity of a given amide $^1N^H$-^{15}N signal as a function of time t, I_0 is the initial intensity, I_∞ is the base line intensity extrapolated to infinite time, and k_{ex} is the exchange rate constant. The I_∞ values of most residues were allowed to float, reflecting the ~10% residual HDO in the NMR sample. For the few most slowly exchanging residues (e.g., with significant intensities after 13 days), their baseline intensities were constrained to 10% of the starting signal, making the exponential decay function $I_t = 0.9I_0e^{-(kex)t} + 0.1I_0$.

Amide protium-protium HX rates for ^{15}N-labled PU.1[167-272] were measured by the CLEANEX method in sample buffer at pH values of 5.80, 6.48, 7.49 and 8.54. In each case, with exception of pH 5.80 (Bruker Avance 500 MHz), 8 spectra were collected on a

Bruker Avance 850 MHz spectrometer with exchange delays of 10, 20, 30, 40, 50, 60, 80 and 100 ms. The exponential growth function used to analyze CLEANEX experiments is $\frac{I_t}{I_{ref}} = (\frac{k_{ex}}{k_{ex}+R_1})(1 - e^{-(k_{ex}+R_1)t})$ where I_t is the amide $^1H^N$-^{15}N signal intensity at a given time, I_{ref} is the intensity of the reference spectrum collected with a 12 sec recycle delay to ensure complete water relaxation, k_{ex} is the exchange rate constant, and R_1 is the apparent relaxation rate of water. A correction of 1/0.7 was used to scale the results of the CLEANEX fits to account for the estimated steady-state magnetization of water (~ 0.7).

The protection factor (PF) for each residue with a measurable k_{ex} under one or more experimental conditions was calculated as a ratio of the predicted k_{pred} to the observed k_{ex}. The former were calculated using the program Sphere (Zhang 1995) based on an unstructured polypeptide with an identical sequence to the PU.1[167-272] fragment and corrected for temperature, pH and solvent (^1H or ^2H). In the cases where PFs were obtained with two or more conditions, the higher PF was reported.

4.4.6. Molecular dynamics simulations

The structures used as a starting point for the MD simulations were taken from the PDB as follows; Ets1[331-440] (PDB: 1R36), ETV4[337-434] (PDB: 4CO8), ETV6[335-426] (PDB: 2MD5), PU.1[171-258] (PDB: 1PUE). We picked these structures because they were readily available with the closest boundaries to our studies. For the NMR derived structural ensembles (Ets1, ETV6), the lowest energy structure of each was used. The structure of Ets1 was truncated to the core Ets domain. (HI1/HI2 deleted) and in the case of PU.1, the DNA was removed. PROPKA3.1 (Olsson et al. 2011) was used to determine protonation states at pH = 7. All proteins were solvated in a cuboid explicit TIP3P water box and Na+ and Cl- ions were added to neutralize the systems. The final number of atoms was 18218, 19717, 18895 and 18886 respectively while the final box dimensions were 15 $Å^2$. After 5000 steps of steepest descent energy minimization of the solvent with the protein coordinates fixed and an additional 10000 steps for all atoms including the protein, the systems were heated

to 300 K over 50 ps and 1 ns of equilibration was performed. The production run of a total of 910 ns of Langevin dynamics for all setups was performed with an integration step size of 2 fs in the modified AMBER ff14SB all-atom force field using the PREMD module in AMBER14 (Cossio et al. 2012). The isobaric isothermal ensemble was used at 300 K and a pressure of 1 atm with periodic boundary conditions and the long range electrostatic interactions were accounted for using the particle-mesh Ewald sum (Essmann et al. 1995). A cutoff of 10 Å was used for long-range non-bonded interactions. The SHAKE algorithm was used to constrain bonds connecting to hydrogens (Lambrakos et al. 1989). Backbone (N, Cα, CO) RMSD time courses were calculated from the trajectories aligned to the core structures (helix H1 to H4) using CPPTRAJ (Roe & Cheatham 2013), a module of AMBER14. This module was further used to calculate the per-residue backbone (N, CA, CO) RMS fluctuations and they were mapped onto the structures and visualized using Pymol (DeLano 2002). Normalized covariance (correlation) analysis was performed using CARMA (Glykos 2006; Sethi et al. 2009). The community network analysis was performed as described previously (Sethi et al. 2009; Bui & Gsponer 2014) and communities and critical nodes were mapped onto the structure and visualized in VMD. CNA describes residues as nodes and connections between them as edges. An edge connecting two nodes is present if heavy atoms of the corresponding residues are within 5 Å of each other in 75 % of the analyzed trajectory. The weight of an edge is based on its betweenness or how often the connection is used for a shortest path connecting any two atoms through bonds. This betweenness is calculated for all edges and is subsequently utilized in the Girvan–Newman algorithm to determine communities of highly interconnected residues. Critical pathways are defined as the highest betweenness edges connecting two communities. Communities were schematically visualized using CYTOSCAPE (Smoot et al. 2011).

Chapter 5.   Conclusion and future studies

Eukaryotic transcription factors up- and down-regulate the expression of the target genes in a tightly controlled manner. The ETS family is a set of transcription factors that utilize the conserved ETS domain to recognize a consensus core $^{5'}$GGAA/T$^{3'}$ in the promoter and enhancer regions of target genes. In humans, there are 28 ETS transcription factor paralogs, and about one-third also contains the PNT domain. This domain mediates protein-protein and protein-ligand interaction to further aid transcriptional activity. Targeting the activities of these domains and linkers can provide a distinct route to regulate the specificities of various ETS factors in response to different signaling events.

The well characterized Ets1 highlights several important routes to specificity. First, the PNT domain of Ets1 is regulated by the dynamic helix H0 in a phosphorylation dependent manner. MAPK phosphorylates Thr38 and Ser41 of Ets1 PNT domain, shifting the conformational equilibrium of a dynamic helix towards the open state. This further exposes the binding interface for the TAZ1 domain of the transcriptional co-activator CBP, thereby facilitating Ets1-targeted gene expression. Second, the DNA-binding ability of Ets1 ETS domain is attenuated by both the inhibitory helices and a serine-rich region. The binding affinity for DNA is reduced by ~ 2 fold in the presence of the inhibitory helices, and increasing levels of multisite CaMKII-dependent phosphorylation of the serine-rich region greatly diminishes the binding affinity for DNA. Building upon these striking observations, the goal of my thesis was to elucidate the structural and dynamic mechanisms underlying autoinhibition and regulation of different members of the ETS family.

5.1. Ras-MAPK signaling through the PNT domain

In chapter 2, I investigated the structural and dynamic effects of phosphorylation on the Pointed-P2 PNT domain using NMR spectroscopy. Pointed-P2 is a *Drosophila* ortholog

of human Ets1 and it is also involved in Ras/MAP kinase-regulated gene expression. Pointed-P2 contains a DNA-binding ETS domain as well as a PNT domain that serves as a docking module to enhance phosphorylation of an adjacent phosphoacceptor threonine by the ERK2 MAP kinase Rolled. Similar to the Ets1 PNT domain, the Pointed-P2 PNT domain also has two helices appended N-terminal to the core helical bundle (a SAM domain). Using NMR chemical shift, ^{15}N relaxation, and amide hydrogen exchange experiments, I demonstrated that the N-terminal helix H0 of the Pointed-P2 PNT domain is dynamic, whereas the helical bundle forms the rigid core. Through *in vitro* phosphorylation using ERK2 kinase, I identified three phosphoacceptor sites N-terminal to helix H0. Of these, only Thr151 corresponds to a MAP kinase consensus phospho-acceptor sequence (Pro-x-Ser/Thr-Pro). Phosphorylation of these residues has a minor effect on secondary structure and dynamics on the Pointed-P2 PNT domain. More importantly, using NMR-monitored titrations, I demonstrated that the phosphoacceptor sites, as well as the region the of the core helical bundle, serve as a kinase docking site. Based on the homology model derived from the Ets1 PNT domain, I concluded that the dynamic helix H0 must be displaced to allow both the docking of the kinase and binding to the Mae PNT domain, a protein partner that negatively regulates Pointed-P2.

5.2. Future studies of Pointed-P2

Much effort has been put into understanding the molecular basis for the regulation of Pointed-P2 and its role in the *Drosophila* sevenless signaling cascade. The high resolution structure of the Pointed-P2 PNT domain with the Rolled kinase or additional components of the transcriptional machinery remains to be obtained. This could be done using a combination of X-ray crystallography and NMR spectroscopy, with NMR spectroscopy focusing on the dynamics of the proteins. In addition to my work, one could further dissect the dynamics of the PNT domain by performing HDX or relaxation dispersion to look at slow motions and conformational exchange. This is especially important given the fact that helix H0 serves as a docking module to position the phosphoaccetors for the kinase, and the helix H0 of Ets1 PNT domain can adopt different conformations upon phosphorylation. In addition, the complex interplay between the

transcriptional repressor Yan, the transcriptional activator Pointed-P2, and the Pointed-P2 negative regulator Mae in a phosphorylation-dependent manner needs to be further characterized to better uncover the molecular basis of gene regulation.

5.3. Expanding the autoinhibition repertoire

An important question that remains to be more fully answered in the ETS family is how can a set of structurally similar proteins that bind to highly conserved DNA sequences with similar affinity regulate distinct biological functions. Part of the answer lies within the autoinhibition mechanism that some of the ETS proteins utilize to regulate their DNA-binding ability. Autoinhibition in the ETS domain is well exemplified by Ets1 and ETV6, in which inhibitory sequences form α-helices attenuate the DNA-binding affinity of their ETS domains allosterically or sterically, respectively.

To this end, my goals in chapter 3 were to expand the mechanistic repertoire of autoinhibition of ETS transcription factors. Furthermore, ETV1/4/5 are of great medical interest given the close link with prostate cancer. Chromosomal translocations that place the *ETV1/4/5* genes under control of the *TMPRSS2* promoter (an androgen-responsive, prostate-specific gene) result in their overexpression to drive aberrant gene expression. This project was done in close collaboration with Dr. Barbara Graves' group at the University of Utah. With the aid of limited proteolysis, we identified and expressed various autoinhibited ETV1/4/5 fragments. Through X-ray crystallography and NMR spectroscopy, we identified the inhibitory elements located both N- and C-terminal to the ETS domain that cooperate to repress DNA binding. The CID is an α-helix that packs against the ETS domain and perturbs the positioning of the DNA-recognition helix. The NID is intrinsically disordered and transiently interacts with the CID and the DNA-recognition helix, particularly to Tyr401 and Tyr403, to mediate autoinhibition. In addition, acetylation of Lys226 and Lys260 within the N-terminal inhibitory sequences activates DNA binding. Collectively, these studies uncovered that ETV1/4/5 DNA-binding autoinhibition utilizes both structural and intrinsically disordered elements in an acetylation-dependent manner to regulate gene expression.

5.4. Future studies of ETV1/4/5

In spite of extensive structural studies of ETV1/4/5, the physical and thermodynamic mechanisms that drive the autoinhibition remain to be addressed. In particular, the transient interactions between the intrinsically disordered sequences (NID) and the ETS domain are poorly defined. The NID, having ~ 180 amino acids, is almost double the length of the ETS domain of ETV4 with roughly equal distribution of polar and hydrophobic amino acids. The use of protein ligation technology to covalently link the NID to the ETS domain coarsely defined the interacting region. However, the amino acids that are responsible for the interaction with the ETS domain and the CID are still not well mapped. One could certainly use paramagnetic relaxation enhancement (PRE) to probe the precise interface by labeling the cysteines within the NID. There are four native cysteines within the NID, Cys192, Cys217, Cys275, and Cys311. Site-directed mutagenesis could be used to mutate the NID so that only a single cysteine remains for paramagnetic spin labeling. Subsequently, alanine-scanning mutagenesis could be used to test the role of inhibitory residues that appear key to autoinhibition.

Finally, definitive evidence for the autoinhibition of ETV1/4/5 *in vivo* is lacking. Inspired by Ets1 autoinhibition by the SRR, phosphorylation of the NID could be part of the mechanism to reinforce autoinhibition. Indeed, ETV1/4/5 can be phosphorylated by PKA (Baert et al. 2002) and conserved serines located within the NID can be found in the ETV1/4/5 sub-family. Understanding autoinhibition may help to explain the transcriptional regulation of ETV1/4/5 and inspire therapeutic strategies to offset the activities of these proteins in cancerous prostate cells.

5.5. Dynamic properties of ETS domain

Throughout the course of studying ETS domain autoinhibition, it is striking that structural elements of the ETS domain become unfolded upon DNA binding. Also the autoinhibitory elements of both Ets1 and ETV6 dampen motions of their ETS domains and stabilize their

flexible DNA-recognition helix H3 (Pufall et al. 2005; Lee et al. 2008; Coyne et al. 2012). This leads us to believe that ETS domains exist in a conformational equilibria between flexible active and rigid inactive states, and that the structural plasticity of the ETS domain is required for searching for their specific DNA sequence.

Accordingly, I selected four minimal ETS domains (ETV4, PU.1, Ets1, ETV6) that each recognize a slightly different consensus DNA that differ on the flanking nucleotides outside of the core $^{5'}$GGAA/T$^{3'}$ (Wei et al. 2010) to further investigate the relationship between ETS dynamics and DNA-binding. Thermal denaturation and HX revealed that PU.1 and ETV4 are less stable than Ets1, whereas ETV6 has an unusually high Tm value due to a rather broad thermal unfolding transition. Our NMR-derived PU.1 structure shows that a dynamic helix H4 exists in solution but the functional role of this helix remains to be explored. A comparison of the protection factors of these ETS domains indicate a conserved stable core, composed of helix H1 and strand S1 and S2. Interestingly, the DNA-recognition helix H3 and to a certain extent helix H2, are relatively less protected, indicative of local conformational transitions. Finally, we used molecular dynamics simulations to help provide a unified description of the motions of ETS domains. We discovered that the β-strands (S1-S4) serve as a central hub that relays information across different part of the ETS domain. Furthermore, there are distinct flexibility differences of the "turns" and "wings" of different ETS domains that might link to indirect readout of DNA sequences. I also hypothesized that a critical residue, Phe420, is responsible for stabilizing ETV4 and therefore mutation should change its dynamics and stability. In conclusion, these studies bridged the gap between ETS structure, dynamics and function to better understand how ETS domain interact with DNA.

5.6. Future studies of ETS domain dynamics

A full understanding of the protein structure and dynamics can provide great insight into their function. As an example, Dr. Soumya De engineered helix H5 of ETV6 to make it more stable and thereby more autoinhibitory (De et al. 2016). Although still at early stage, altering the dynamics of ETS domains can potentially change interactions with specific

and non-specific DNA at both the thermodynamic and kinetic levels (Desjardins et al. 2016; De et al. 2014). With the aid of MD simulations, one could design mutants of ETS domains that have the potential to (de)stabilize the protein. In parallel, the specificities of these proteins could be examined using unbiased *in vitro* SELEX approaches in combination with high-throughput next-gen DNA sequencing (Zykovich et al. 2009). Furthermore, understanding the protein structure in detail could aid the design of drugs to target specific interfaces. Indeed, in collaboration with the groups of Dr. Michael Cox, Dr. Paul Rennie and Dr. Artem Cherkasov at the BC Prostate Center, we identified small molecules that disrupt DNA binding by ERG. These potential lead compounds may inspire therapeutics to regulate the *in vivo* activities of ETS factors linked to cancers.

Bibliography

Abramczyk, O. et al., 2007. Expanding the repertoire of an ERK2 recruitment site: Cysteine footprinting identifies the D-recruitment site as a mediator of Ets-1 binding. *Biochemistry*, 46(32), pp.9174–9186.

Adams, P.D. et al., 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), pp.213–221.

Akke, M., 2002. NMR methods for characterizing microsecond to millisecond dynamics in recognition and catalysis. *Current Opinion in Structural Biology*, 12(5), pp.642–647.

Andrew, C.D. et al., 2002. Effect of phosphorylation on alpha-helix stability as a function of position. *Biochemistry*, 41(6), pp.1897–1905.

Anthis, N.J. & Clore, G.M., 2015. Visualizing transient dark states by NMR spectroscopy. *Quarterly Reviews of Biophysics.*, 1, pp.35–116.

Augustijn, K.D. et al., 2002. Structural characterization of the PIT-1/ETS-1 interaction: PIT-1 phosphorylation regulates PIT-1/ETS-1 binding. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20), pp.12657–12662.

Aytes, A. et al., 2013. ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 110(37), pp.E3506-15.

Baena, E. et al., 2013. ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. *Genes and Development*, 27(6), pp.683–698.

Baert, J.-L. et al., 2007. The 26S proteasome system degrades the ERM transcription factor and regulates its transcription-enhancing activity. *Oncogene*, 26(3), pp.415–24.

Baert, J.L. et al., 2002. ERM transactivation is up-regulated by the repression of DNA binding after the PKA phosphorylation of a consensus site at the edge of the ETS domain. *Journal of Biological Chemistry*, 277(2), pp.1002–1012.

Bai, Y. et al., 1993. Primary Structure Effects on Peptide Group Hydrogen Exchange. *Proteins*, 17(1), pp.75–86.

Baker, D.A. et al., 2001. Mae mediates MAP kinase phosphorylation of Ets transcription factors in Drosophila. *Nature*, 411(6835), pp.330–334.

Barraud, P. & Allain, F.H.T., 2013. Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: How the relative orientation between RRMs influences the nucleic acid binding topology. *Journal of Biomolecular NMR*, 55(1), pp.119–138.

Battiste, J.L. & Wagner, G., 2000. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry*, 39(18), pp.5355–5365.

van den Bedem, H. & Fraser, J.S., 2015. Integrative, dynamic structural biology at atomic resolution—it's about time. *Nature Methods*, 12(4), pp.307–318.

Bienkiewicz, E.A. & Lumb, K.J., 1999. Random-coil chemical shifts of phosphorylated amino acids. *Journal of Biomolecular NMR*, 15(3), pp.203–206.

Bohlander, S.K., 2005. ETV6: A versatile player in leukemogenesis. *Seminars in Cancer Biology*, 15(3), pp.162–174.

Bojović, B.B. & Hassell, J.A., 2001. The PEA3 Ets Transcription Factor Comprises Multiple Domains That Regulate Transactivation and DNA Binding. *Journal of Biological Chemistry*, 276(6), pp.4509–4521.

Boros, J. et al., 2009. Elucidation of the ELK1 target gene network reveals a role in the coordinate regulation of core components of the gene regulation machinery. *Genome Research*, 19(11), pp.1963–1973.

Bosc, D.G., Goueli, B.S. & Janknecht, R., 2001. HER2/Neu-mediated activation of the ETS transcription factor ER81 and its target gene MMP-1. *Oncogene*, 20, pp.6215–6224.

De Braekeleer, E. et al., 2012. ETV6 fusion genes in hematological malignancies: A review. *Leukemia Research*, 36(8), pp.945–961.

Brivanlou, A.H. & Darnell, J.E., 2002. Signal transduction and the control of gene expression. *Science (New York, N.Y.)*, 295(5556), pp.813–8.

Brown, L.A. et al., 1998. Molecular characterization of the zebrafish PEA3 ETS-domain transcription factor. *Oncogene*, 17(1), pp.93–104.

Brunner, D. et al., 1994. The ETS domain protein pointed-P2 is a target of MAP kinase in the sevenless signal transduction pathway. *Nature*, 370, pp.386–389.

Bui, J.M. & Gsponer, J., 2014. Phosphorylation of an intrinsically disordered segment in Ets1 shifts conformational sampling toward binding-competent substates. *Structure*, 22(8), pp.1196–1203.

Cai, C. et al., 2007. ETV1 is a novel androgen receptor-regulated gene that mediates prostate cancer cell invasion. *Molecular Endocrinology (Baltimore, Md.)*, 21(8), pp.1835–1846.

Callaway, K. et al., 2010. Phosphorylation of the transcription factor ets-1 by ERK2: Rapid dissociation of ADP and phospho-ets-1. *Biochemistry*, 49(17), pp.3619–3630.

Callaway, K.A. et al., 2006. Properties and regulation of a transiently assembled ERK2-Ets-1 signaling complex. *Biochemistry*, 45(46), pp.13719–13733.

Camilloni, C. et al., 2012. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry*, 51(11), pp.2224–2231.

Campbell, E. et al., 2016. The role of protein dynamics in the evolution of new enzyme function. *Nature chemical biology*, 12(September), pp.944–950.

Chakrabarti, S.R. & Nucifora, G., 1999. The leukemia-associated gene TEL encodes a transcription repressor which associates with SMRT and mSin3A. *Biochemical and Biophysical Research Communications*, 264, pp.871–877.

Chen, V.B. et al., 2010. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(1), pp.12–21.

Chen, Y. et al., 2013. ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nature Medicine*, 19(8), pp.1023–9.

Chi, P. et al., 2010. ETV1 is a lineage survival factor that cooperates with KIT in gastrointestinal stromal tumours. *Nature*, 467(7317), pp.849–53.

Cooper, C.D.O. et al., 2015. Structures of the Ets protein DNA-binding domains of transcription factors Etv1, Etv4, Etv5, and Fev: Determinants of DNA binding and redox regulation by disulfide bond formation. *Journal of Biological Chemistry*, 290(22), pp.13692–13709.

Cossio, M.L.T. et al., 2012. *Amber14 Reference manual*,

Coyne, H.J. et al., 2012. Autoinhibition of ETV6 (TEL) DNA binding: Appended helices sterically block the ETS domain. *Journal of Molecular Biology*, 421(1), pp.67–84.

Currie, S.L. et al., 2017. Structured and disordered regions cooperatively mediate DNA-binding autoinhibition of ETS factors ETV1, ETV4 and ETV5. *Nucleic Acids Research*, pp.1–19.

De, S. et al., 2016. Autoinhibition of ETV6 DNA binding is established by the stability of its inhibitory helix. *Journal of Molecular Biology*, 428(8), pp.1515–1530.

De, S. et al., 2014. Steric mechanism of auto-inhibitory regulation of specific and non-specific dna binding by the ets transcriptional repressor ETV6. *Journal of Molecular Biology*, 426(7), pp.1390–1406.

Debes, J.D. et al., 2003. p300 in Prostate Cancer Proliferation and Progression. *Cancer Research*, 63(22), pp.7638–7640.

Defossez, P. a et al., 1997. The ETS family member ERM contains an alpha-helical acidic activation domain that contacts TAFII60. *Nucleic acids research*, 25(22), pp.4455–4463.

Degerny, C. et al., 2005. SUMO modification of the Ets-related transcription factor ERM inhibits its transcriptional activity. *Journal of Biological Chemistry*, 280(26), pp.24330–24338.

Delaglio, F. et al., 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3), pp.277–293.

DeLano, W.L., 2002. The PyMOL Molecular Graphics System, Version 1.1. *Schrödinger LLC*, p.http://www.pymol.org.

Delano, W.L. & Bromberg, S., 2004. *PyMOL User's Guide*,

Desjardins, G. et al., 2016. Conformational Dynamics and the Binding of Specific and Nonspecific DNA by the Autoinhibited Transcription Factor Ets-1. *Biochemistry*, 55(29), pp.4105–4118.

Desjardins, G. et al., 2014. Synergy of aromatic residues and phosphoserines within the intrinsically disordered DNA-binding inhibitory elements of the Ets-1 transcription factor. *Proceedings of the National Academy of Sciences of the United States of America*, 111(30), pp.11019–11024.

Djebali, S. et al., 2012. Landscape of transcription in human cells. *Nature*, 489(7414), pp.101–8.

Dosset, P. et al., 2000. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *Journal of Biomolecular NMR*, 16(1), pp.23–28.

Dunker, A.K. et al., 2002. Intrinsic Disorder and Protein Function. *Biochemistry*, 41(21), pp.6573–6582.

Dyson, H.J. & Wright, P.E., 2002. Coupling of folding and binding for unstructured proteins. *Current Opinion in Structural Biology*, 12(1), pp.54–60.

Dyson, H.J. & Wright, P.E., 2005. Intrinsically unstructured proteins and their functions. *Nature reviews. Molecular Cell Biology*, 6(3), pp.197–208.

Emsley, P. et al., 2010. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4), pp.486–501.

Escalante, C.R., Brass, A.L., et al., 2002. Crystal structure of PU.1/IRF-4/DNA ternary complex. *Molecular Cell*, 10(5), pp.1097–1105.

Escalante, C.R., Shen, L., et al., 2002. Crystallization and characterization of PU.1/IRF-4/DNA ternary complex. *Journal of Structural Biology*, 139(1), pp.55–59.

Essmann, U. et al., 1995. A smooth particle mesh Ewald method. *Journal of Chemical Physics*, 103(1995), pp.8577–8593.

Farrow, N. et al., 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*, 33(19), pp.5984–6003.

Firlej, V. et al., 2005. Pea3 transcription factor cooperates with USF-1 in regulation of the murine bax transcription without binding to an Ets-binding site. *Journal of Biological Chemistry*, 280(2), pp.887–898.

Fischer, E., 1894. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dtsch. Chem. Ges.*, 27(3), pp.2985–2993.

Foulds, C.E. et al., 2004. Ras/mitogen-activated protein kinase signaling activates Ets-1

and Ets-2 by CBP/p300 recruitment. *Molecular and Cellular Biology*, 24(24), pp.10954–64.

Freiburger, L. et al., 2015. Efficient segmental isotope labeling of multi-domain proteins using Sortase A. *Journal of Biomolecular NMR*, 63(1), pp.1–8.

Fuxreiter, M., 2012. Fuzziness: linking regulation to protein dynamics. *Mol. BioSyst.*, 8(1), pp.168–177.

Fuxreiter, M., Simon, I. & Bondos, S., 2011. Dynamic protein-DNA recognition: Beyond what can be seen. *Trends in Biochemical Sciences*, 36(8), pp.415–423.

Garvie, C.W. et al., 2002. Structural analysis of the autoinhibition of Ets-1 and its role in protein partnerships. *Journal of Biological Chemistry*, 277(47), pp.45529–45536.

Garvie, C.W., Hagman, J. & Wolberger, C., 2001. Structural studies of Ets-1/Pax5 complex formation on DNA. *Molecular Cell*, 8(6), pp.1267–1276.

Gasteiger, E. et al., 2005. *The Proteomics Protocols Handbook*,

Girdwood, D. et al., 2003. p300 transcriptional repression is mediated by SUMO modification. *Molecular Cell*, 11(4), pp.1043–1054.

Glykos, N.M., 2006. Carma: a molecular dynamics analysis program. *Journal of Computational Chemistry*, 27(14), pp.1765–8.

Gocke, C.B., Yu, H. & Kang, J., 2005. Systematic identification and analysis of mammalian small ubiquitin-like modifier substrates. *Journal of Biological Chemistry*, 280(6), pp.5004–5012.

Goel, A. & Janknecht, R., 2003. Acetylation-mediated transcriptional activation of the ETS protein ER81 by p300, P/CAF, and HER2/Neu. *Molecular and Cellular Biology*, 23(17), pp.6243–6254.

Goetz, T.L. et al., 2000. Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. *Molecular and Cellular Biology*, 20(1), pp.81–90.

Gonzalez, F.A., Raden, D.L. & Davis, R.J., 1991. Identification of substrate recognition determinants for human ERK1 and ERK2 protein kinases. *Journal of Biological Chemistry*, 266(33), pp.22159–22163.

Green, S.M. et al., 2010. DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *Journal of Biological Chemistry*, 285(24),

pp.18496–18504.

Greenall, A. et al., 2001. DNA Binding by the ETS-domain Transcription Factor PEA3 is Regulated by Intramolecular and Intermolecular Protein·Protein Interactions. *Journal of Biological Chemistry*, 276(19), pp.16207–16215.

Greenfield, N.J., 2009. Using circular dichroism collected as a funcion of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature Protocols*, 1(6), pp.2527–2535.

Greenfield, N.J., 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, 1(6), pp.2876–2890.

Güntert, P., 2004. Automated NMR structure calculation with CYANA. *Methods in Molecular Biology (Clifton, N.J.)*, 278(1980), pp.353–378.

Guo, B. et al., 2011. Dynamic modification of the ETS transcription factor PEA3 by sumoylation and p300-mediated acetylation. *Nucleic Acids Research*, 39(15), pp.6303–6313.

Gutman, A. & Wasylyk, B., 1991. Nuclear targets for transcription regulation by oncogenes. *Trends in Genetics : TIG*, 7(2), pp.49–54.

De Guzman, R.N. et al., 2005. CBP/p300 TAZ1 domain forms a structured scaffold for ligand binding. *Biochemistry*, 44(2), pp.490–497.

Hafsa, N.E. & Wishart, D.S., 2014. CSI 2.0: A significantly improved version of the Chemical Shift Index. *Journal of Biomolecular NMR*, 60(2–3), pp.131–146.

Hahn, S.L. et al., 1997. Modulation of ETS-1 transcriptional activity by huUBC9, a ubiquitin- conjugating enzyme [published erratum appears in Oncogene 1998 Feb 5;16(5):691]. *Oncogene*, 15(12), pp.1489–95.

Hammoudeh, D.I. et al., 2009. Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *Journal of the American Chemical Society*, 131(21), pp.7390–7401.

Heery, D. et al., 1997. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature*, 387, pp.733–736.

Helgeson, B.E. et al., 2008. Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer. *Cancer Research*, 68(1), pp.73–80.

Henzler-Wildman, K. & Kern, D., 2007. Dynamic personalities of proteins. *Nature*,

450(7172), pp.964–972.

Hiebert, S.W. et al., 1996. The t(12;21) translocation converts AML-1B from an activator to a repressor of transcription. *Molecular and cellular biology*, 16(4), pp.1349–1355.

Hollenhorst, P.C. et al., 2009. DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genetics*, 5(12).

Hollenhorst, P.C. et al., 2007. Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. *Genes and Development*, 21(15), pp.1882–1894.

Hollenhorst, P.C., Ferris, M.W., et al., 2011. Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes and Development*, 25(20), pp.2147–2157.

Hollenhorst, P.C., McIntosh, L.P. & Graves, B.J., 2011. Genomic and biochemical insights into the specificity of ETS transcription factors. *Annual Review of Biochemistry*, 80(1), pp.437–471.

Huang, W. & Waknitz, M., 2009. ETS gene fusions and prostate cancer. *American Journal of Translational Research*, 1(4), pp.341–351.

Hwang, T.L., van Zijl, P.C.M. & Mori, S., 1998. Accurate quantitation of water-amide proton exchange rates using the Phase-Modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme RID B-8680-2008. *Journal of Biomolecular NMR*, 11(2), pp.221–226.

Isharwal, S. et al., 2008. p300 (Histone acetyltransferase) biomarker predicts prostate cancer biochemical recurrence and correlates with changes in epithelia nuclear size and shape. *Prostate*, 68(10), pp.1097–1104.

Jané-Valbuena, J. et al., 2010. An oncogenic role for ETV1 in melanoma. *Cancer Research*, 70(5), pp.2075–2084.

Ji, Z. et al., 2007. Regulation of the Ets-1 transcription factor by sumoylation and ubiquitinylation. *Oncogene*, 26(3), pp.395–406.

Jia, X. et al., 1999. Backbone dynamics of a short PU.1 ETS domain. *Journal of Molecular Biology*, 292(5), pp.1083–93.

Johansson, H. et al., 2015. Specific and Nonspecific Interactions in Ultraweak Protein − Protein Associations Revealed by Solvent Paramagnetic Relaxation

Enhancements.

Jonsen, M.D. et al., 1996. Characterization of the cooperative function of inhibitory sequences in Ets-1. *Molecular and cellular biology*, 16(5), pp.2065–2073.

Juven-Gershon, T. & Kadonaga, J.T., 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental Biology*, 339(2), pp.225–229.

Kalodimos, C.G., Biris, N., et al., 2004. Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes. *Science*, 305(5682), pp.386–389.

Kalodimos, C.G., Boelens, R. & Kaptein, R., 2004. Toward an integrated model of protein-DNA recognition as inferred from NMR studies on the Lac repressor system. *Chemical Reviews*, 104(8), pp.3567–3586.

Karplus, M. & Kuriyan, J., 2005. Molecular dynamics and protein function. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 102(19), pp.6679–6685.

Kim, A.S. et al., 2000. Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein. *Nature*, 404(6774), pp.151–158.

Kim, C.A. et al., 2001. Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *EMBO Journal*, 20(15), pp.4173–4182.

Kim, H.J. et al., 2014. A positive role of DBC1 in PEA3-mediated progression of estrogen receptor-negative breast cancer. *Oncogene*, (October), pp.1–9.

Klämbt, C., 1993. The Drosophila gene pointed encodes two ETS-like proteins which are involved in the development of the midline glial cells. *Development (Cambridge, England)*, 117(1), pp.163–176.

Kleckner, I.R. & Foster, M.P., 2011. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1814(8), pp.942–968.

Kodandapani, R. et al., 1996. A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. *Nature*, 380(6573), pp.456–460.

Krishnan, N. et al., 2014. Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. *Nature Chemical Biology*, 10(7), pp.558–566.

Kriwacki, R.W. et al., 1996. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 93(21), pp.11504–9.

Kumar, J.P. et al., 2004. CREB binding protein functions during successive stages of eye development in Drosophila. *Genetics*, 168(2), pp.877–893.

Laget, M.P. et al., 1996. Two functionally distinct domains responsible for transactivation by the Ets family member ERM. *Oncogene*, 12(6), pp.1325–36.

Lambrakos, S.G. et al., 1989. A modified shake algorithm for maintaining rigid bonds in molecular dynamics simulations of large molecules. *Journal of Computational Physics*, 85(2), pp.473–486.

Lau, D. et al., 2016. Design and Selection of IFI16-PAAD Mutants with Improved dsDNA Destabilization Properties. *Journal of Proteomics & Bioinformatics*, 9(11), pp.255–263.

Laudet, V. et al., 1999. Molecular phylogeny of the ETS gene family. *Oncogene*, 18(6), pp.1351–9.

de Launoit, Y. et al., 1997. Structure-function relationships of the PEA3 group of Ets-related transcription factors. *Biochemical and Molecular Medicine*, 61, pp.127–135.

de Launoit, Y. et al., 2006. The Ets transcription factors of the PEA3 group: Transcriptional regulators in metastasis. *Biochimica et Biophysica Acta - Reviews on Cancer*, 1766(1), pp.79–87.

Lee, G.M. et al., 2008. The Affinity of Ets-1 for DNA is Modulated by Phosphorylation Through Transient Interactions of an Unstructured Region. *Journal of Molecular Biology*, 382(4), pp.1014–1030.

Lee, G.M. et al., 2005. The structural and dynamic basis of Ets-1 DNA binding autoinhibition. *Journal of Biological Chemistry*, 280(8), pp.7088–7099.

Lee, R. Van Der et al., 2014. Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews*, 114, pp.6589–6631.

Lee, S. et al., 2011. A model of a MAPK-Substrate complex in an active conformation: A computational and experimental approach. *PLoS ONE*, 6(4).

Lee, W. et al., 2011. PONDEROSA, an automated 3D-NOESY peak picking program,

enables automated protein structure determination. *Bioinformatics*, 27(12), pp.1727–1728.

Lee, W., Tonelli, M. & Markley, J.L., 2015. NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, 31(8), pp.1325–1327.

Lens, Z. et al., 2010. Solution structure of the N-terminal transactivation domain of ERM modified by SUMO-1. *Biochemical and Biophysical Research Communications*, 399(1), pp.104–110.

Levine, M. & Tjian, R., 2003. Transcription regulation and animal diversity. *Nature*, 424(6945), pp.147–151.

Li, M., Pascual, G. & Glass, C.K., 2000. Peroxisome proliferator-activated receptor gamma-dependent repression of the inducible nitric oxide synthase gene. *Molecular and Cellular Biology*, 20(13), pp.4699–707.

Li, M.Z. & Elledge, S.J., 2012. SLIC: A method for sequence- and ligation-independent cloning. *Methods in Molecular Biology*, 852, pp.51–59.

Li, Q.J. et al., 2003. MAP kinase phosphorylation-dependent activation of Elk-1 leads to activation of the co-activator p300. *EMBO Journal*, 22(2), pp.281–291.

Li, R., Pei, H. & Watson, D.K., 2000. Regulation of Ets function by protein - protein interactions. *Oncogene*, 19, pp.6514–6523.

Li, R. & Woodward, C., 1999. The hydrogen exchange core and protein folding. *Protein science : a publication of the Protein Society*, 8(8), pp.1571–1590.

Lin, J.H. et al., 1998. Functionally related motor neuron pool and muscle sensory afferent subtypes defined by coordinate ETS gene expression. *Cell*, 95(3), pp.393–407.

Lipari, G. & Szabo, A., 1982. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules {II}. {A}nalysis of experimental results. *J. Am. Chem. Soc.*, 104(17), pp.4559–4570.

Lise, S. & Jones, D.T., 2005. Sequence patterns associated with disordered regions in proteins. *Proteins: Structure, Function and Genetics*, 58(1), pp.144–150.

Liu, Y., Borchert, G.L. & Phang, J.M., 2004. Polyoma Enhancer Activator 3, an Ets Transcription Factor, Mediates the Induction of Cyclooxygenase-2 by Nitric Oxide in Colorectal Cancer Cells. *Journal of Biological Chemistry*, 279(18), pp.18694–

18700.

Liu, Y. & Kuhlman, B., 2006. RosettaDesign server for protein design. *Nucleic Acids Research*, 34, pp.235–238.

Lyskov, S. et al., 2013. Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE). *PLoS ONE*, 8(5).

Macauley, M.S. et al., 2006. Beads-on-a-string, characterization of Ets-1 sumoylated within its flexible N-terminal sequence. *Journal of Biological Chemistry*, 281(7), pp.4164–4172.

Mackereth, C.D. et al., 2004. Diversity in structure and function of the Ets family PNT domains. *Journal of Molecular Biology*, 342(4), pp.1249–1264.

Malhotra, P. & Udgaonkar, J.B., 2016. How cooperative are protein folding and unfolding transitions? *Protein Science*, 25(11), pp.1924–1941.

Malik, S. & Roeder, R.G., 2010. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature reviews. Genetics*, 11(11), pp.761–72.

Mao, A.H. et al., 2010. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 107(18), pp.8183–8.

Massie, C.E. et al., 2007. New androgen receptor genomic targets show an interaction with the ETS1 transcription factor. *EMBO Rep*, 8(9), pp.871–878.

McCoy, A.J. et al., 2007. Phaser crystallographic software. *Journal of Applied Crystallography*, 40(4), pp.658–674.

McIntosh, L.P. et al., 2009. Detection and assignment of phosphoserine and phosphothreonine residues by 13C-31P spin-echo difference NMR spectroscopy. *Journal of Biomolecular NMR*, 43(1), pp.31–37.

Meruelo, A.D. & Bowie, J.U., 2009. Identifying polymer-forming SAM domains. *Proteins: Structure, Function and Bioinformatics*, 74(1), pp.1–5.

Morreale, a et al., 2000. Structure of Cdc42 bound to the GTPase binding domain of PAK. *Nature Structural Biology*, 7(5), pp.384–388.

Müller-Späth, S. et al., 2010. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences*

of the United States of America, 107(33), pp.14609–14614.

Munde, M. et al., 2014. Structure-dependent inhibition of the ETS-family transcription factor PU.1 by novel heterocyclic diamidines. *Nucleic Acids Research*, 42(2), pp.1379–1390.

Nakae, K. et al., 1995. ERM, a PEA3 subfamily of Ets transcription factors, can cooperate with c-Jun. *Journal of Biological Chemistry*, 270(40), pp.23795–23800.

Nelson, M.L. et al., 2010. Ras signaling requires dynamic properties of Ets1 for phosphorylation-enhanced binding to coactivator CBP. *Proceedings of the National Academy of Sciences of the United States of America*, 107(22), pp.10026–31.

Neumann, H., Peak-Chew, S.Y. & Chin, J.W., 2008. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nature Chemical Biology*, 4(4), pp.232–234.

Nunn, M.F. et al., 1983. Tripartite structure of the avian erythroblastosis virus E26 transforming gene. *Nature*, 306, pp.391–395.

O'Hagan, R.C. et al., 1996. The activity of the Ets transcription factor PEA3 is regulated by two distinct MAPK cascades. *Oncogene*, 13(6), pp.1323–33.

O'Neill, E.M. et al., 1994. The activities of two Ets-related transcription factors required for Drosophila eye development are modulated by the Ras/MAPK pathway. *Cell*, 78(1), pp.137–147.

O'Shea, E.K., Rutkowski, R. & Kim, P.S., 1992. Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. *Cell*, 68(4), pp.699–708.

Oates, M.E. et al., 2013. D2P2: Database of disordered protein predictions. *Nucleic Acids Research*, 41(D1), pp.508–516.

Olsson, M.H.M. et al., 2011. PROPKA3: Consistent treatment of internal and surface residues in empirical p K a predictions. *Journal of Chemical Theory and Computation*, 7(2), pp.525–537.

Otwinowski, Z. & Minor, W., 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, 276, pp.307–326.

Pan, Y. et al., 2010. Mechanisms of transcription factor selectivity. *Trends in Genetics*, 26(2), pp.75–83.

Panne, D., Maniatis, T. & Harrison, S.C., 2007. An Atomic Model of the Interferon- b Enhanceosome. *Cell*, 129, pp.1111–1123.

Pascal, S.M. et al., 1994. Simultaneous Acquisition of 15N- and 13C-Edited NOE Spectra of Proteins Dissovled in H2O. *Journal of Magnetic Resonance Series B*, 103, pp.197–201.

Petersen, J.M. et al., 1995. Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an alpha helix. *Science (New York, N.Y.)*, 269(5232), pp.1866–1869.

Pintacuda, G. & Otting, G., 2002. Identification of protein surfaces by NMR measurements with a paramagnetic Gd(III) chelate. *Journal of the American Chemical Society*, 124(3), pp.372–373.

Piserchio, A. et al., 2011. Solution NMR insights into docking interactions involving inactive ERK2. *Biochemistry*, 50(18), pp.3660–3672.

Pogenberg, V. et al., 2014. Design of a bZip transcription factor with homo/heterodimer-induced DNA-binding preference. *Structure*, 22(3), pp.466–477.

Ponting, C.P., 1995. SAM: A novel motif in yeast sterile and drosophila polyhomeotic proteins. *Protein Science*, 4(9), pp.1928–1930.

Poon, G.M.K., 2012. Sequence discrimination by DNA-binding domain of ETS family transcription factor PU.1 is linked to specific hydration of protein-DNA interface. *Journal of Biological Chemistry*, 287(22), pp.18297–18307.

Poon, G.M.K. & Macgregor, R.B., 2004. A thermodynamic basis of DNA sequence selectivity by the ETS domain of murine PU.1. *Journal of Molecular Biology*, 335(1), pp.113–127.

Poon, G.M.K. & Macgregor, R.B., 2003. Base coupling in sequence-specific site recognition by the ETS domain of murine PU.1. *Journal of Molecular Biology*, 328(4), pp.805–819.

Pop, M.S. et al., 2014. A small molecule that binds and inhibits the ETV1 transcription factor oncoprotein. *Molecular Cancer Therapeutics*, 13(6), pp.1492–502.

Pufall, M. a et al., 2005. Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science (New York, N.Y.)*, 309(July), pp.142–145.

Pufall, M. a & Graves, B.J., 2002. Autoinhibitory domains: modular effectors of cellular regulation. *Annual Review of Cell and Developmental Biology*, 18, pp.421–62.

Qiao, F. et al., 2004. Derepression by depolymerization: Structural insights into the

regulation of Yan by Mae. *Cell*, 118(2), pp.163–173.

Qiao, F. et al., 2006. Mae inhibits Pointed-P2 transcriptional activity by blocking its MAPK docking site. *The EMBO journal*, 25(1), pp.70–79.

Qiao, F. & Bowie, J.U., 2005. The many faces of SAM. *Science's STKE : signal transduction knowledge environment*, 2005(January), p.re7.

Radhakrishnan, I. et al., 1997. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell*, 91, pp.741–752.

Radivojac, P. et al., 2003. Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 227, pp.216–227.

Raes, J. et al., 2007. Protein function space: viewing the limits or limited by our view? *Current Opinion in Structural Biology*, 17(3), pp.362–369.

Rainey, M.A. et al., 2005. Proximity-induced catalysis by the protein kinase ERK2. *Journal of the American Chemical Society*, 127(30), pp.10494–10495.

Rebay, I. & Rubin, G.M., 1995. Yan functions as a general inhibitor of differentiation and is negatively regulated by activation of the Ras1/MAPK pathway. *Cell*, 81(6), pp.857–866.

Regan, M.C. et al., 2013. Structural and dynamic studies of the transcription factor ERG reveal DNA binding is allosterically autoinhibited. *Proc Natl Acad Sci U S A*, 110(33), pp.13374–13379.

Rezsohazy, R. et al., 2015. Cellular and molecular insights into Hox protein action. *Development*, 142(7), pp.1212–1227.

Roe, D.R. & Cheatham, T.E., 2013. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, 9(7), pp.3084–3095.

Rohs, R. et al., 2010. Origins of specificity in protein - DNA recognition. *Annual Review of Biochemistry*, 79, pp.233–69.

Romero, P. et al., 2001. Sequence complexity of disordered protein. *Proteins: Structure, Function and Genetics*, 42(1), pp.38–48.

Ryu, H. et al., 2015. NMRe: A web server for NMR protein structure refinement with

high-quality structure validation scores. *Bioinformatics*, 32(4), pp.611–613.

Sanchez, R. & Zhou, M.-M., 2009. The role of human bromodomains in chromatin biology and gene transcription. *Current Opinion in Drug Discovery & Development*, 12(5), pp.659–65.

Sattler, M., Schleucher, J. & Griesinger, C., 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 34(2), pp.93–158.

Schneikert, J. et al., 1996. Androgen receptor-Ets protein interaction is a novel mechanism for steroid hormone-mediated down-modulation of matrix metalloproteinase expression. *Journal of Biological Chemistry*, 271(39), pp.23907–23913.

Seidel, J.J. & Graves, B.J., 2002. An ERK2 docking site in the Pointed domain distinguishes a subset of ETS transcription factors. *Genes and Development*, 16(1), pp.127–137.

Sethi, A. et al., 2009. Dynamical networks in tRNA:protein complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16), pp.6620–5.

Shen, Y. et al., 2009. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR*, 44(4), pp.213–223.

Shen, Y. & Bax, A., 2012. Identification of helix capping and beta-turn motifs from NMR chemical shifts. *Journal of Biomolecular NMR*, 52(3), pp.211–232.

Shiina, M. et al., 2014. A novel allosteric mechanism on protein-DNA interactions underlying the phosphorylation-dependent regulation of Ets1 target gene expressions. *Journal of Molecular Biology*, 427(8), pp.1655–1669.

Shrivastava, T. et al., 2014. Structural basis of Ets1 activation by Runx1. *Leukemia*, 21(November 2013), pp.1–9.

Sievers, F. et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), p.539.

Sikorski, T.W. & Buratowski, S., 2009. The basal initiation machinery: beyond the general transcription factors. *Current Opinion in Cell Biology*, 21(3), pp.344–351.

Smoot, M.E. et al., 2011. Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics*, 27(3), pp.431–432.

Song, H. et al., 2005. Antagonistic regulation of Yan nuclear export by Mae and Crm1 may increase the stringency of the Ras response. *Genes and Development*, 19(15), pp.1767–1772.

Songyang, Z. et al., 1996. A Structural Basis for Substrate Specificities of Protein Ser/Thr Kinases: Primary Sequence Preference of Casein Kinases I and II, NIMA, Phosphorylase Kinase, Calmodulin- Dependent Kinase II, CDK5, and Erk1. *Molecular and Cellular Biology*, 16(11), pp.6486–6493.

Studier, F.W., 2005. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 41(1), pp.207–234.

Taatjes, D.J., 2010. The human Mediator complex: A versatile, genome-wide regulator of transcription. *Trends in Biochemical Sciences*, 35(6), pp.315–322.

Tajul-Arifin, K. et al., 2003. Identification and analysis of chromodomain-containing proteins encoded in the mouse transcriptome. *Genome Research*, 13(6 B), pp.1416–1429.

Takahashi, A. et al., 2005. E1AF degradation by a ubiquitin-proteasome pathway. *Biochemical and Biophysical Research Communications*, 327(2), pp.575–580.

Tavernelli, I., Cotesta, S. & Di Iorio, E.E., 2003. Protein Dynamics, Thermal Stability, and Free-Energy Landscapes: A Molecular Dynamics Investigation. *Biophysical Journal*, 85(4), pp.2641–2649.

Taylor, J.M. et al., 1997. A role for the ETS domain transcription factor PEA3 in myogenic differentiation. *Molecular and Cellular Biology*, 17(9), pp.5550–5558.

Tomlins, S.A. et al., 2006. TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. *Cancer Research*, 66(7), pp.3396–3400.

Tomlins, S. a et al., 2007. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature*, 448(7153), pp.595–9.

Tomlins, S. a et al., 2005. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science (New York, N.Y.)*, 310(5748), pp.644–648.

Tompa, P. & Fuxreiter, M., 2008. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends in Biochemical Sciences*, 33(1),

pp.2–8.

Tootle, T.L., Lee, P.S. & Rebay, I., 2003. CRM1-mediated nuclear export and regulated activity of the Receptor Tyrosine Kinase antagonist YAN require specific interactions with MAE. *Development (Cambridge, England)*, 130(5), pp.845–857.

Tootle, T.L. & Rebay, I., 2005. Post-translational modifications influence transcription factor activity: A view from the ETS superfamily. *BioEssays*, 27(3), pp.285–298.

Trudeau, T. et al., 2013. Structure and intrinsic disorder in protein autoinhibition. *Structure*, 21(3), pp.332–341.

Uversky, V., Gillespie, J. & Fink, A., 2000. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, 41(3), pp.415–427.

Vaquerizas, J. et al., 2009. A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics*, 10(4), pp.252–263.

Vivekanand, P. & Rebay, I., 2006. Intersection of signal transduction pathways and development. *Annual Review of Genetics*, 40, pp.139–57.

Vivekanand, P., Tootle, T.L. & Rebay, I., 2004. MAE, a dual regulator of the EGFR signaling pathway, is a target of the Ets transcription factors PNT and YAN. *Mechanisms of Development*, 121(12), pp.1469–1479.

Wang, S. et al., 2014. Mechanistic heterogeneity in site recognition by the structurally homologous DNA-binding domains of the ETS family transcription factors Ets-1 and PU.1. *Journal of Biological Chemistry*, 289(31), pp.21605–21616.

Wasylyk, C. et al., 1997. Conserved mechanisms of Ras regulation of evolutionary related transcription factors, Ets1 and Pointed P2. *Oncogene*, 14(November 2015), pp.899–913.

Weathers, E.A. et al., 2004. Reduced amino acid alphabet is sufficient to accurately recognize intrinsically disordered protein. *FEBS Letters*, 576(3), pp.348–352.

Wei, G.-H. et al., 2010. Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *The EMBO Journal*, 29(13), pp.2147–2160.

Williams, R.M. et al., 2001. The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pp.89–100.

Wilson, N.K. et al., 2010. Combinatorial transcriptional control in blood stem/progenitor

cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*, 7(4), pp.532–544.

Woodward, C. & Li, R., 1998. Is the slow exchange core the protein folding core? *Trends in Biochemical Sciences*, 23(10), p.379.

Wright, P.E. & Dyson, H.J., 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nature reviews. Molecular Cell Biology*, 16(1), pp.18–29.

Wright, P.E. & Dyson, H.J., 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*, 293(2), pp.321–31.

Wright, P.E. & Dyson, H.J., 2009. Linking folding and binding. *Current Opinion in Structural Biology*, 19(1), pp.31–38.

Wu, J. & Janknecht, R., 2002. Regulation of the ETS transcription factor ER81 by the 90-kDa ribosomal S6 kinase 1 and protein kinase A. *The Journal of Biological Chemistry*, 277(45), pp.42669–79.

Xin, J.H. et al., 1992. Molecular cloning and characterization of PEA3 , a new member of the Ets oncogene family that is differentially expressed in mouse embryonic cells. *Genes and Development*, 6, pp.481–496.

Yang, B.S. et al., 1996. Ras-mediated phosphorylation of a conserved threonine residue enhances the transactivation activities of c-Ets1 and c-Ets2. *Molecular and Cellular Biology*, 16(2), pp.538–47.

Yang, C. et al., 1998. A role for CREB binding protein and p300 transcriptional coactivators in Ets-1 transactivation functions. *Molecular and Cellular Biology*, 18(4), pp.2218–2229.

Yang, S.H. et al., 1999. The mechanism of phosphorylation-inducible activation of the ETS-domain transcription factor Elk-1. *EMBO Journal*, 18(20), pp.5666–5674.

Zhang, Y., Cao, H. & Liu, Z., 2015. Binding cavities and druggability of intrinsically disordered proteins. *Protein Science*, 24(5), pp.688–705.

Zhang, Z. et al., 2015. Chemical perturbation of an intrinsically disordered region of TFIID distinguishes two modes of transcription initiation. *eLife*, 4(AUGUST2015).

Zhang, Z., 1995. *Protein and peptide structure and interactions studied by hydrogen exchange and NMR*. University of Pennsylvania.

Zwahlen, C. et al., 1998. An NMR experiment for measuring methyl-methyl NOEs in 13C-labeled proteins with high resolution. *Journal of the American Chemical Society*, 120(30), pp.7617–7625.

Zykovich, A., Korf, I. & Segal, D.J., 2009. Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. *Nucleic acids research*, 37(22), p.e151.

Appendices

During the course of my Ph.D studies (2012 - 2017), I was involved in many other projects, not directly related to the focus of this thesis. In 2016, I was able to revisit and publish my M.Sc. work at Dr. Frederic Pio's laboratory with a focus on engineering of a protein to enhance stability and function. **Desmond K.W. Lau**, Kush Dalal, Benjamin Hon, Frederic Pio. (2016) *"Design and selection of IFI16-PAAD mutants with improved dsDNA destabilization properties."* **J. Proteomics & Bioinformatics.** 9: 255-263.