THE STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF ETS TRANSCRIPTION FACTORS WITH RELEVANCE TO PROSTATE CANCER

by

Simon Logan Currie

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Oncological Sciences

The University of Utah

August 2016

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Simon Logan Currie**

has been approved by the following supervisory committee members:

Barbara J. Graves	, Chair	**4/29/2016** Date Approved
Donald E. Ayer	, Member	**4/29/2016** Date Approved
Christopher P. Hill	, Member	**4/29/2016** Date Approved
Katherine S. Ullman	, Member	**4/29/2016** Date Approved
Dennis R. Winge	, Member	**4/292016** Date Approved

and by **Bradley Cairns** , Chair/Dean of the Department/College/School of **Oncological Sciences**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Appropriate regulation of gene expression is important for the development and homeostasis of multicellular organisms. DNA sequence-specific transcription factors play a central role in regulating the first step of gene expression, transcription. The aberrant expression of transcription factors is a common mechanism for the initiation and progression of many human cancers. The ETS family of transcription factors consists of twenty-eight human proteins that contain a conserved DNA-binding domain, termed the ETS domain. ETS factors have varied roles in organismal development and disease etiology. For example, ETS proteins from the ERG and ETV1/4/5 subfamilies are overexpressed in the majority of prostate cancers and contribute to cancer initiation and progression. In stark contrast, EHF and SPDEF are two ETS factors present in normal prostate tissue that have been characterized as tumor suppressors whose genes are often deleted during cancer progression. The phenotypic dichotomy displayed between these subclasses of ETS factors suggests that the understanding of the molecular mechanisms that underlie transcription factors' roles in normal and disease settings may provide additional opportunities for therapeutic intervention.

Here we describe the DNA-binding autoinhibition of ETS factors ETV1,

ETV4, and ETV5. An intrinsically disordered region and an α-helix cooperate to inhibit DNA-binding by altering the positioning of the DNA-recognition α-helix of the ETS domain. These inhibitory elements are distinct from those that have been previously described for other ETS factors. We also characterize the interaction of Mediator subunit 25 (MED25) with the transcriptional activation and DNA-binding domains of ETV4. The inhibitory α-helix of ETV4 provides a unique interaction surface for MED25, as compared to other ETS domains, and interaction with MED25 activates the DNA-binding of ETV4. We also demonstrate the differential ability of ETS factors to bind to DNA with JUN-FOS at composite DNA binding sites. These distinct intra- and intermolecular interactions distinguish ETS oncoproteins and tumor suppressors in prostate cancer and may, in part, underlie their phenotypic differences. Finally, we present an assay for ETS-DNA interactions that is amenable to high-throughput screening for small molecule inhibitors. This assay could be further modified to incorporate any of the previously described partnerships.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my mentor, Barbara Graves. Her time, patience, persistence, curiosity, feigned interest in college football, and her relentless commitment to teaching have been formative in my development as a scientist. I would also like to acknowledge the members of my committee for their scientific input and encouragement. I thank the many collaborators who contributed to the science in this dissertation: Desmond Lau, Lawrence McIntosh, and Jack Skalicky with NMR spectroscopy; Frank Whitby and Chris Hill with X-ray crystallography; Steve Warner, Alexis Mollard, Kelly Doyle, Xiao-Hui Liu, Hari Vankayalapati, and Dave Bearss with screening for small-molecule inhibitors. I also want to thank overlapping members of the Graves lab, and in particular Jed Doane, Katie Evans, and Bethany Madison. Their intelligence, curiosity, and drive have contributed tremendously to my development as a scientist.

I wish to express my love and admiration to my family. Their support and encouragement of my pursuits has provided me with sustenance. Their accompaniment in the wild lands of Utah has made life sweet. I would also like to express my gratitude for the many friends that I have been fortunate to spend time with in Utah. In particular, I would like to thank Dennis and Lauri Winge. Without their friendship and support, this work would not have been possible.

CHAPTER 1

INTRODUCTION

Gene-specific regulation of transcription

Appropriate spatiotemporal control of gene expression is necessary for normal development and homeostasis[1]. Conversely, misregulation of key genes is causal for many diseases, including cancer[2]. Transcription is the first step of gene expression and is a highly coordinated process. DNA sequence-specific transcription factors bind to cognate sites in DNA and modulate the recruitment of RNA polymerase II (Pol II), general transcription factors (GTFs), the Mediator complex, chromatin-modifying or –remodeling complexes, and other transcription factors. A network of molecular interactions dictates the unique composition of proteins at any given gene, and this diverse composition is integrated into a graded ability to recruit and activate Pol II for processive transcription[3,4] (**Fig. 1.1a**).

Typical DNA sequence-specific transcription factors have a discrete DNA-binding domain (DBD) that recognizes related DNA binding sites with varying affinty[5], and an activation domain(s) (AD) that interacts with transcriptional coactivators[2,6]. DBDs have diverse structural composition, such as homeodomains, zinc fingers, helix-loop-helix, or basic region leucine zippers, and read both specific DNA sequences and the general shape of DNA through contact with the exposed base on a nucleotide or the phosphate backbone of DNA, respectively[7]. Multiple transcription factors can coordinately control the transcription of specific genes by modifying protein-DNA contacts and/or by adding new protein-protein contacts[8-15]. The largest structurally characterized example is that of the *IFN-B* enhancer[9,10] (**Fig. 1.1c**). Seven distinct transcription

factors bind to a ~ 50 nucleotide segment of DNA. Interestingly, despite the dense binding of transcription factors on the *IFN-B* enhancer, there is a paucity of protein-protein interactions between the factors. Additionally, mutagenesis of the protein contacts demonstrates that they are not required for cooperative DNA-binding in this example. Rather, this suggests that the cooperative binding of multiple transcription factors at the *IFN-B* enhancer is facilitated through DNA. Other examples have more explicitly demonstrated the role of DNA sequences in influencing cooperative protein binding[16], as well as protein conformation and regulatory activity[17]. Thus, both the DNA sequences of transcription factor binding sites and the spacing between these binding sites influence the binding of multiple transcription factors at regulatory regions.

Sequence-specific transcription factors utilize an AD(s) to interact with GTFs and transcriptional coactivators, such as subunits of the Mediator complex, to recruit and activate Pol II. ADs are disordered in isolation but often become more α-helical in the presence of a coactivator target[18-22]. ADs can interact with multiple distinct coactivators[23-27], and the disordered nature of ADs is postulated to be important for this promiscuous recognition[28]. For example, p53 AD - coactivator structures demonstrate that the specific coactivator target influences the coactivator-bound structure of the AD[29-33] (**Fig. 1.1d**). Furthermore, distinct p53 AD mutants differentially affect gene-specific transcription[34], which suggests that there is variable requirement for individual p53 – coactivator interactions at particular p53 target genes. In sum, the transcription factors at a regulatory region affect the binding of other transcription factors and recruit of transcriptional

coactivators and GTFs to influence transcriptional output from nearby genes.

ETS transcription factors

The ETS family of transcription factors consists of 28 genes in humans, defined by the obligate presence of the ETS domain, a winged helix-turn-helix DNA-binding domain[35] (**Fig. 1.2**). The ETS domain is conserved across human ETS factors; ~ 15% and 45% of sequences are absolutely or functionally conserved, respectively. This conservation is more striking from a structural standpoint as the structures of ETS domains from different factors align with typical root-mean-square-deviation values of ~ 1 Å[36]. A subset of ETS factors also contain a pointed (PNT) domain which facilitates protein-protein interactions[35]. Diverse sequences and structures that flank the ETS domain in different ETS factors contribute to autoinhibition (discussed below) and/or facilitate protein-protein interactions[35,37]. Outside of these structured domains, ETS factors primarily consist of intrinsically disordered regions (IDRs), which are enriched in transcription factors in general[38,39].

Sequences that reside outside of the ETS domain and inhibit the DNA-binding of the ETS domain have been observed in most ETS subfamilies[40-46] and have been structurally described for ETS1[47,48], ETV6[37,49,50], and ERG[51] (**Fig. 1.3**). In ETS1, the most thoroughly characterized example of DNA-binding autoinhibition in the ETS family, four α-helices that flank the ETS domain impart a slight ~ 2-fold level of inhibition[52,53]. An IDR, termed the serine-rich region (SRR), increases this inhibition to ~ 30-fold and phosphorylation of serines within

the SRR further increases inhibition another ~1000-fold[47,48]. Cooperative DNA-binding with either PAX5 or RUNX1 ablates the autoinhibition of ETS1 by disrupting the inhibitory module[13-15,54]. Therefore, ETS1 autoinhibition serves as a model for the integration of posttranslational modifications and protein-protein interactions in the regulation of DNA-binding affinity. This regulation provides a route for transcriptional regulation by an individual ETS factor, as discussed below. ETV6 and ERG, the other two ETS factors for which autoinhibition has been structurally characterized, have distinct mechanisms of autoinhibition, implying that the cellular regulation of autoinhibition for these factors will also be distinct.

Some ETS factors have been described as master regulators of cell identity. For example, SPI1, also known as PU.1, dictates differentiation of hematopoietic progenitors along the myeloid lineage[55], and ETV2 is sufficient for the conversion of fibroblasts into endothelial cells[56]. Most cell-types express multiple ETS factors and conversely, most ETS factors are expressed in multiple cell-types[57]. This co-expression of ETS factors results in many ETS DNA-binding sites in the genome being redundantly occupied by multiple ETS proteins[58]. However, the broad array of phenotypes produced from genetic disruptions of individual ETS factors[35], as well as the examples of ETS factors serving as master regulators of cell identity, suggest that some genes are regulated by a single ETS factor.

The determinants of redundant regulation by multiple ETS factors versus specific regulation by an individual ETS factor have been described for ETS1 in

T-cells. Whereas redundantly regulated sites contain near-consensus ETS-binding DNA sequences, ETS1-specific sites possess variant DNA sequences that disproportionately weaken the binding of other ETS factors such as ELF1[59]. Additionally, only ETS1-specific sites are part of an ETS-RUNX composite DNA-binding site[58,59]. RUNX1 and ETS1 cooperatively bind to DNA by relieving the autoinhibition of the other factor[54,58,60]. In summary, variant ETS sites and cooperative binding with other transcription factors establish the specific regulation of genomic sites by individual ETS factors.

Phenotypic diversity of ETS factors in prostate cancer

ETS transcription factors exhibit an intriguing phenotypic dichotomy in the context of prostate cancer. EHF and SPDEF, two ETS factors that are highly expressed in normal prostate[57], are commonly down-regulated or deleted during disease progression[61-67]. These ETS proteins are of prognostic value in prostate cancer as patients with lower levels display poorer overall and biochemical recurrence-free survival[61,63], and patients with higher levels of SPDEF have prolonged response to androgen deprivation therapy[67]. In contrast, ETS factors from the ERG and ETV1/4/5 subfamilies are overexpressed in the majority of prostate cancer patients[68]. Chromosomal rearrangements resulting in prostate-specific or constitutively-expressed promoters controlling the transcription of *ERG* or *ETV1/4/5* genes are the most common cause for the overexpression of these genes[53,68-70], although other cellular mechanisms have been described[71]. ERG or ETV1/4/5 factors are sufficient to generate prostatic intraepithelial

neoplasia, an early stage of prostate cancer[72-74]. Additional genetic events, such as the deletion of *PTEN*, are frequently observed in prostate cancer patients with *ERG* or *ETV1/4/5* rearrangements[68,75], and these multiple genetic events synergize to generate prostate cancer in mouse models[72,73,76-78]. Interestingly, ETV1 and ETV4 mouse models exhibit more aggressive and metastatic forms of disease as compared to ERG mouse models, although it is debated as to whether this difference faithfully reflects the human disease[68,76-78].

Although it is clear that the aberrant expression of *ERG* or *ETV1/4/5* genes contributes to prostate cancer progression, the description of molecular mechanisms that facilitate this phenotypic response are incomplete. Due to the central role of the androgen receptor (AR) in the etiology and treatment of prostate cancer[79,80], the interplay of ETS factors and AR has been a focus of many studies. ERG and ETV1 physically interact with AR and act as pioneer factors to expand AR's binding to new genomic loci[74,76,77,81]. Whereas ETV1 and AR synergize to drive higher transcriptional output[74,77], ERG appears to dampen the transcriptional affect of AR alone[77,81]. This difference may be due, at least in part, to ETV1 driving prostate cell-autonomous production of the androgen hormone. ETS factors normally expressed in the prostate, such as SPDEF and ELF3, also interact with and perturb the transcriptional activity of AR[82,83]. Therefore, multiple ETS factors are capable of influencing the location and transcriptional activity of AR, and the overexpression of ERG or ETV1, may drive prostate cancer by disrupting the normal ETS-AR balance in normal prostate cells by altering the location and activity of AR.

Activator protein 1 (AP1) is a heterodimeric transcription factor composed of JUN and FOS subunits. ETS – AP1 is a transcription factor partnership that has more clearly illustrated the opposite roles of ETS factors in prostate cancer. ERG, ETV1, and ETV4 drive the transcription of genes that are near composite ETS-AP1 sites, whereas SPDEF and EHF repress the transcription of these genes[64,84]. These ETS-AP1 composite sites are near genes that are important for cell migration, such as matrix metalloproteases and the extracellular matrix remodeler urokinase plasminogen activator[84,85]. Concordantly, overexpression of ERG, ETV1, or ETV4 is sufficient for the increased migration of normal prostate cells. There is also specificity from the AP1 side as the JUN proteins cJUN, JUNB, and JUND differentially regulate transcription at ETS-AP1 composite sites[86]. While it is clear that the tumor suppressor ETS factors, EHF and SPDEF, as well as oncogenic ETS factors, ERG, ETV1, and ETV4, regulate ETS-AP1 composite sites in an opposing manner, the molecular basis for this distinction is unclear.

Targeting transcription factors for inhibition

The deviant expression or activity of transcription factors is characteristic of many human cancers, making misregulated transcription factors desirable therapeutic targets[2,87,88]. However, several intrinsic characteristics of transcription factors make therapeutic intervention difficult. With the exception of nuclear hormone receptors[80,89], transcription factors do not possess highly concave ligand pockets that have served as energetically favorable targets in proteins

such as kinases[90] or chromatin modifying enzymes[91,92]. Rather, the DBD of transcription factors presents a broad surface for extensive DNA contact and is highly conserved throughout transcription factor families (**Fig. 1.2**), making potent, yet selective inhibition of this interface a challenge. Transcription factors make numerous protein-protein contacts with transcriptional cofactors[29] (**Fig. 1.1c**), therefore prior knowledge of which, if any, of these interactions is necessary for the disease state is required. Additionally, much like protein-DNA interactions, protein-protein interfaces typically form broad interfaces that are difficult to selectively inhibit. Despite these challenges, progress has been made in inhibiting transcription factors through modulation of expression or protein stability and through blocking protein-protein interfaces.

Inhibitors of the bromodomain and extraterminal (BET) protein BRD4, a transcriptional coactivator, competitively compete with acetylated transcription factors and histone tails to bind with BRD4[93,94]. These inhibitors are active in many forms of hematological and solid cancers[95-97]. Originally, the mode-of-action for BRD4 inhibitors was thought to be the reduced transcription of the oncogenic transcription factor MYC[95,97]. In many cancers, down-regulation of MYC is, at least in part, responsible for the effect of BRD4 inhibition. However, in other cancer subtypes, BRD4 inhibition appears to be completely independent of the transcription and expression of MYC[98]. Therefore, BRD4 inhibition serves as an example of decreasing the expression of a transcription factor and/or of blocking a transcriptionally important protein-protein interface, depending on the disease context.

ERG, as discussed above, is overexpressed in the majority of prostate cancer patients[53,68]. The inhibition of a deubiquitinase enzyme, ubiquitin-specific peptidase 9 X-linked (USP9X), that deubiquinates ERG results in ERG degradation and inhibits growth of ERG-positive tumors in mouse xenograft models of prostate cancer[99]. Conjugation of a phthalimide-derived molecule to JQ1, one of the BRD4 inhibitors discussed above, increases BRD4 protein degradation through recruitment of the cereblon (CRBN) E3-ubiquitin ligase, and is efficacious in a mouse model of leukemia[100]. A distinct phthalimide-derived molecule has anticancer activity against multiple myeloma cell lines and patient cells by specifically decreasing the protein levels of the Ikaros family zinc finger transcription factors 1 and 3 through a similar mechanism[101]. In the opposite direction, inhibition of the E3-ubiquitin ligases MDM2 or MDM4 increases the stability of the tumor suppressor p53 and improves survival in mouse models of lymphomas, sarcomas, and liver carcinomas[102-104]. Therefore, modulating the expression of oncogenic or tumor suppressor transcription factors appears to be a tractable method for the targeted treatment of many cancers.

Other strategies have focused on the inhibition of protein-protein interfaces that are crucial for disease phenotype. An inhibitor that blocks the interaction between transcription factors CBFβ and RUNX1 delays leukemia progression in mice[105]. CBFβ/SMMHC chromosomal rearrangements in acute myeloid leukemia result in an aberrant trifold symmetry in the CBFβ-RUNX1 interaction, and the inhibitor mimics this additional symmetry in order to specifically inhibit CBFβ/SMMHC - RUNX1 interactions over wild-type CBFβ –

RUNX1 interactions. Disruption of ELF3 with MED23 decreased the expression of the HER2 oncoprotein and specifically inhibited the cell-growth of HER2-expressing breast cancer cell lines[106-108]. Therefore, the inhibition of protein-protein interactions required for disease progression is another viable strategy for targeting transcription factors.

Summary of research chapters

The common misregulation of transcription factor expression and activity in cancers makes these factors attractive, yet difficult, therapeutic targets. The challenges of, as well as previous successes in, inhibiting transcription factors indicate that elucidation of the mechanisms by which these proteins contribute to cancer can be leveraged into generating more selective inhibitors. Chapter 2 describes the development of an assay that is amenable to high-throughput screening for small molecule inhibitors for the interaction of ETS1 and DNA. The lead compound from this screen was nonspecific in inhibiting ETS and other transcription factors, but the screen is robust in differentiating between positive and negative hits and is transferrable to other ETS factors with roles in cancer progression. Chapter 3 details the characterization of DNA-binding autoinhibition in the ETV1/4/5 subfamily of transcription factors. An intrinsically disordered region and an α-helix cooperatively inhibit the DNA binding of the ETS domain of ETV4 by modulating the positioning of the DNA-recognition α-helix. Acetylation of lysine residues within the intrinsically disordered region ablates autoinhibition, suggesting that this is a route for the *in vivo* regulation of ETV4 DNA binding.

The involvement of activation and DNA-binding domains of ETV4 in the interaction with MED25 is identified in Chapter 4. Divergent sequences within the ETS domain, and a secondary structural element that is specific to the ETV1/4/5 subfamily form the DNA-binding domain interaction surface of the ETV4-MED25 interaction, thereby dictating the specificity of MED25 for ETV4 as compared to other ETS factors. Chapter 5 reveals that cJUN-FOS cooperates with, or antagonizes, the DNA-binding of ERG and EHF, respectively. This distinction supports the differential regulation of ETS-AP-1-controlled genes by these factors, and may underlie their distinctive phenotypes in the context of prostate cancer. A summary and future directions are presented in Chapter 6.

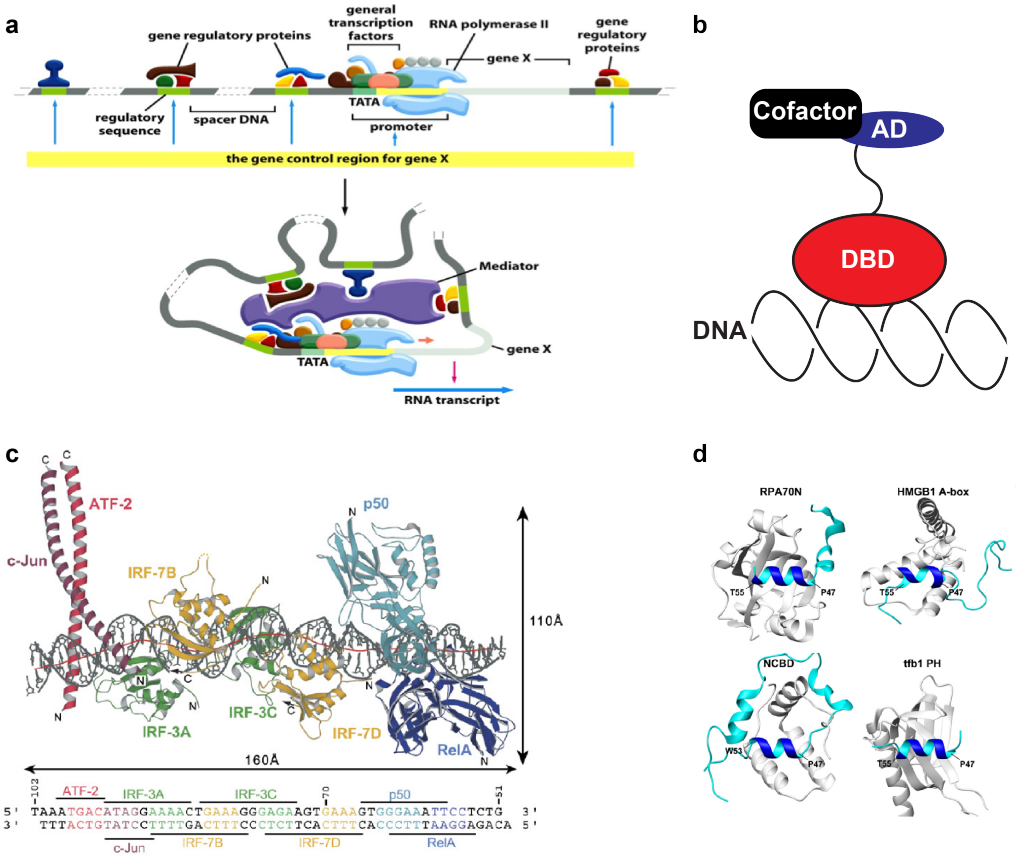


Figure 1.1 Eukaryotic transcription preinitiation complex. (**a**) Cartoon depicting eukaryotic transcription preinitiation complex including RNA polymerase II (Pol II), general transcription factors, gene regulatory proteins or transcription factors, and the Mediator complex. Image from Alberts *et al.*, 2014. (**b**) Cartoon depiction of a sequence-specific transcription factor illustrating two minimal domains; a DNA-binding domain, DBD, red, and an activation domain, AD, blue. (**c**) The structure of *IFN-B* enhancer[10] is an example of an enhanceosome, or a region of DNA where multiple transcription factors bind to modulate Pol II recruitment and activity at a gene. Image from Panne *et al.*, 2007. (**d**) The AD of p53 has unique secondary structural characteristics depending on the individual coactivator. Image from Okuda *et al.*, 2014.

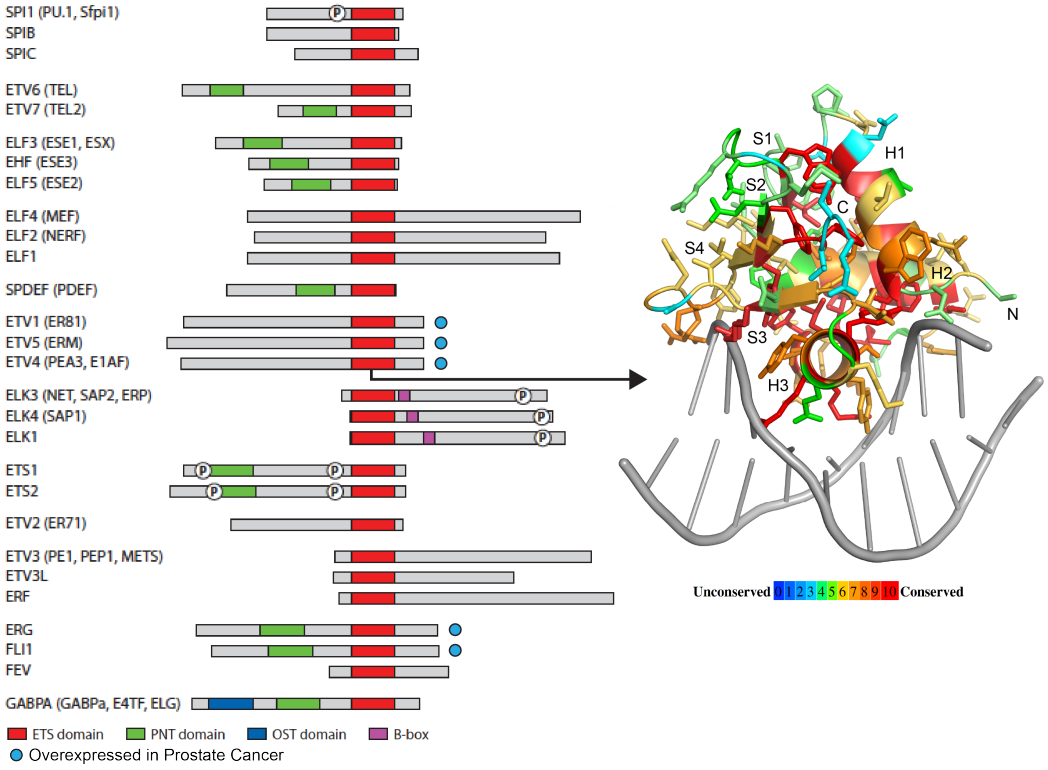


Figure 1.2 ETS family of transcription factors. Left, ETS transcription factors are depicted as rectangles with red, green, blue, and purple boxes depicting ETS, PNT, OST, and B-box domains, respectively. Light blue circles denote ETS factors that are overexpressed in prostate cancer. Right, structure of ETV4 ETS domain with the protein backbone in cartoon format and amino acid side chains in stick format. Individual amino acids are colored according to their conservation in all human ETS domains. N and C refer to the N-terminus and the C-terminus of the ETS domain, and H1, H2, H3 and S1, S2, S3, S4 refer to the α-helices and β-strands in order from N- to C-terminus. Rectangle depiction of ETS factors modified from Hollenhorst *et al*., 2011.

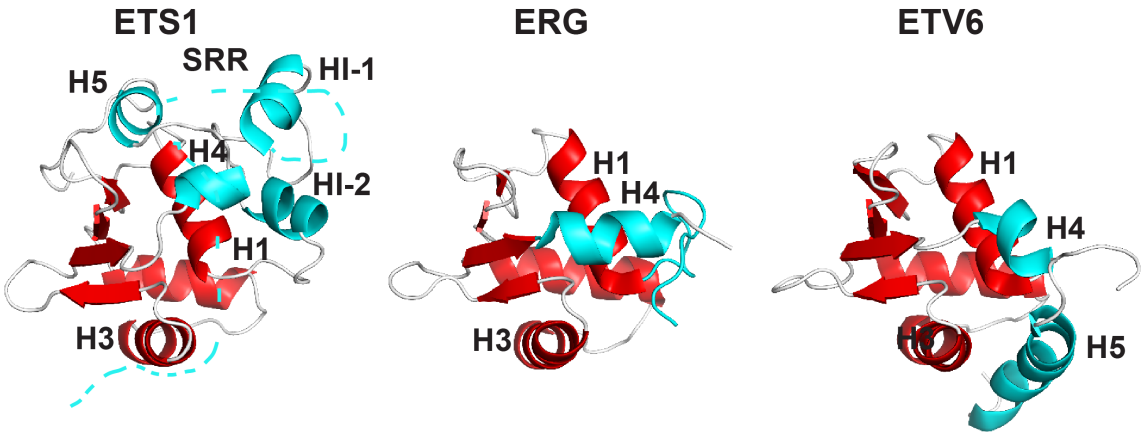


Figure 1.3 Structurally characterized examples of DNA-binding autoinhibition in ETS factors. ETS1[47,48,53], ERG[51], and ETV6[37,49,50] are examples of ETS factors with inhibitory sequences, cyan, that have been structurally characterized. The conserved ETS domain is colored red. The serine-rich region (SRR) inhibitory element in ETS1 is represented as a dotted line as it is intrinsically disordered and does not take on a fixed position or structure while inhibiting the ETS domain.

References

1. Lee, T.I. & Young, R.A. Transcription of eukaryotic protein-coding genes. *Annu Rev Genet* **34**, 77-137 (2000).

2. Bhagwat, A.S. & Vakoc, C.R. Targeting transcription factors in cancer. *Trends Cancer* **1**, 53-65 (2015).

3. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**, 129-43 (2015).

4. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K. & Walter, P. (ed.) *Molecular Biology of the Cell*, 1464 (Garland Science, New York, NY, USA, 2014).

5. Jolma, A. et al. DNA-binding specificities of human transcription factors. *Cell* **152**, 327-39 (2013).

6. Poss, Z.C., Ebmeier, C.C. & Taatjes, D.J. The Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol* **48**, 575-608 (2013).

7. Slattery, M. et al. Absence of a simple code: how transcription factors read the genome. *Trends Biochem Sci* **39**, 381-99 (2014).

8. Chen, L., Glover, J.N., Hogan, P.G., Rao, A. & Harrison, S.C. Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature* **392**, 42-8 (1998).

9. Panne, D., Maniatis, T. & Harrison, S.C. Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer. *EMBO J* **23**, 4384-93 (2004).

10. Panne, D., Maniatis, T. & Harrison, S.C. An atomic model of the interferon-beta enhanceosome. *Cell* **129**, 1111-23 (2007).

11. Mo, Y., Ho, W., Johnston, K. & Marmorstein, R. Crystal structure of a ternary SAP-1/SRF/c-fos SRE DNA complex. *J Mol Biol* **314**, 495-506 (2001).

12. Garvie, C.W., Hagman, J. & Wolberger, C. Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell* **8**, 1267-76 (2001).

13. Garvie, C.W., Pufall, M.A., Graves, B.J. & Wolberger, C. Structural analysis of the autoinhibition of Ets-1 and its role in protein partnerships. *J Biol Chem* **277**, 45529-36 (2002).

14. Shrivastava, T. et al. Structural basis of Ets1 activation by Runx1. *Leukemia* **28**, 2040-8 (2014).

15. Shiina, M. et al. A novel allosteric mechanism on protein-DNA interactions underlying the phosphorylation-dependent regulation of Ets1 target gene expressions. *J Mol Biol* **427**, 1655-69 (2015).

16. Kim, S. et al. Probing allostery through DNA. *Science* **339**, 816-9 (2013).

17. Meijsing, S.H. et al. DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* **324**, 407-10 (2009).

18. Dyson, H.J. & Wright, P.E. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **12**, 54-60 (2002).

19. Landrieu, I. et al. Characterization of ERM transactivation domain binding to the ACID/PTOV domain of the Mediator subunit MED25. *Nucleic Acids Res* **43**, 7110-21 (2015).

20. Vojnic, E. et al. Structure and VP16 binding of the Mediator Med25 activator interaction domain. *Nat Struct Mol Biol* **18**, 404-9 (2011).

21. Milbradt, A.G. et al. Structure of the VP16 transactivator target in the Mediator. *Nat Struct Mol Biol* **18**, 410-5 (2011).

22. Brzovic, P.S. et al. The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. *Mol Cell* **44**, 942-53 (2011).

23. Yang, F. et al. An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature* **442**, 700-4 (2006).

24. Galbraith, M.D. et al. ERK phosphorylation of MED14 in promoter complexes during mitogen-induced gene activation by Elk-1. *Nucleic Acids Res* **41**, 10241-53 (2013).

25. Defossez, P.A., Baert, J.L., Monnot, M. & de Launoit, Y. The ETS family member ERM contains an alpha-helical acidic activation domain that contacts TAFII60. *Nucleic Acids Res* **25**, 4455-63 (1997).

26. Verger, A. et al. The Mediator complex subunit MED25 is targeted by the N-terminal transactivation domain of the PEA3 group members. *Nucleic Acids Res* **41**, 4847-59 (2013).

27. Ito, M. et al. Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol Cell* **3**, 361-70 (1999).

28. Berlow, R.B., Dyson, H.J. & Wright, P.E. Functional advantages of dynamic protein disorder. *FEBS Lett* **589**, 2433-40 (2015).

29. Okuda, M. & Nishimura, Y. Extended string binding mode of the phosphorylated transactivation domain of tumor suppressor p53. *J Am Chem Soc* **136**, 14143-52 (2014).

30. Lee, C.W., Martinez-Yamout, M.A., Dyson, H.J. & Wright, P.E. Structure of the p53 transactivation domain in complex with the nuclear receptor coactivator binding domain of CREB binding protein. *Biochemistry* **49**, 9964-71 (2010).

31. Di Lello, P. et al. Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53. *Mol Cell* **22**, 731-40 (2006).

32. Rowell, J.P., Simpson, K.L., Stott, K., Watson, M. & Thomas, J.O. HMGB1-facilitated p53 DNA binding occurs via HMG-Box/p53 transactivation domain interaction, regulated by the acidic tail. *Structure* **20**, 2014-24 (2012).

33. Bochkareva, E. et al. Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. *Proc Natl Acad Sci U S A* **102**, 15412-7 (2005).

34. Brady, C.A. et al. Distinct p53 transcriptional programs dictate acute DNA-damage responses and tumor suppression. *Cell* **145**, 571-83 (2011).

35. Hollenhorst, P.C., McIntosh, L.P. & Graves, B.J. Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu Rev Biochem* **80**, 437-71 (2011).

36. Cooper, C.D., Newman, J.A., Aitkenhead, H., Allerston, C.K. & Gileadi, O. 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. *J Biol Chem* **290**, 13692-709 (2015).

37. Coyne, H.J., 3rd et al. Autoinhibition of ETV6 (TEL) DNA binding: appended helices sterically block the ETS domain. *J Mol Biol* **421**, 67-84 (2012).

38. Liu, J. et al. Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873-88 (2006).

39. Wright, P.E. & Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* **16**, 18-29 (2015).

40. Janknecht, R., Zinck, R., Ernst, W.H. & Nordheim, A. Functional dissection of the transcription factor Elk-1. *Oncogene* **9**, 1273-8 (1994).

41. Oettgen, P. et al. Characterization of ESE-2, a novel ESE-1-related Ets transcription factor that is restricted to glandular epithelium and differentiated keratinocytes. *J Biol Chem* **274**, 29439-52 (1999).

42. Greenall, A., Willingham, N., Cheung, E., Boam, D.S. & Sharrocks, A.D. DNA binding by the ETS-domain transcription factor PEA3 is regulated by intramolecular and intermolecular protein.protein interactions. *J Biol Chem* **276**, 16207-15 (2001).

43. Maira, S.M., Wurtz, J.M. & Wasylyk, B. Net (ERP/SAP2) one of the Ras-inducible TCFs, has a novel inhibitory domain with resemblance to the helix-loop-helix motif. *EMBO J* **15**, 5849-65 (1996).

44. Bojovic, B.B. & Hassell, J.A. The PEA3 Ets transcription factor comprises multiple domains that regulate transactivation and DNA binding. *J Biol Chem* **276**, 4509-21 (2001).

45. Stinson, J. et al. Regulation of TCF ETS-domain transcription factors by helix-loop-helix motifs. *Nucleic Acids Res* **31**, 4717-28 (2003).

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

47. Pufall, M.A. et al. Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* **309**, 142-5 (2005).

48. Lee, G.M. et al. The affinity of Ets-1 for DNA is modulated by phosphorylation through transient interactions of an unstructured region. *J Mol Biol* **382**, 1014-30 (2008).

49. Green, S.M., Coyne, H.J., 3rd, McIntosh, L.P. & Graves, B.J. DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J Biol Chem* **285**, 18496-504 (2010).

50. De, S. et al. Steric mechanism of auto-inhibitory regulation of specific and non-specific DNA binding by the ETS transcriptional repressor ETV6. *J Mol Biol* **426**, 1390-406 (2014).

51. Regan, M.C. et al. Structural and dynamic studies of the transcription factor ERG reveal DNA binding is allosterically autoinhibited. *Proc Natl Acad Sci U S A* **110**, 13374-9 (2013).

52. Jonsen, M.D., Petersen, J.M., Xu, Q.P. & Graves, B.J. Characterization of the cooperative function of inhibitory sequences in Ets-1. *Mol Cell Biol* **16**, 2065-73 (1996).

53. Tomlins, S.A. et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-8 (2005).

54. Goetz, T.L., Gu, T.L., Speck, N.A. & Graves, B.J. Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. *Mol Cell Biol* **20**, 81-90 (2000).

55. Nerlov, C. & Graf, T. PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes Dev* **12**, 2403-12 (1998).

56. Morita, R. et al. ETS transcription factor ETV2 directly converts human fibroblasts into functional endothelial cells. *Proc Natl Acad Sci U S A* **112**, 160-5 (2015).

57. Hollenhorst, P.C., Jones, D.A. & Graves, B.J. Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. *Nucleic Acids Res* **32**, 5693-702 (2004).

58. Hollenhorst, P.C., Shah, A.A., Hopkins, C. & Graves, B.J. Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. *Genes Dev* **21**, 1882-94 (2007).

59. Hollenhorst, P.C. et al. DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genet* **5**, e1000778 (2009).

60. Gu, T.L., Goetz, T.L., Graves, B.J. & Speck, N.A. Auto-inhibition and partner proteins, core-binding factor beta (CBFbeta) and Ets-1, modulate DNA binding by CBFalpha2 (AML1). *Mol Cell Biol* **20**, 91-103 (2000).

61. Cheng, X.H. et al. SPDEF inhibits prostate carcinogenesis by disrupting a positive feedback loop in regulation of the Foxm1 oncogene. *PLoS Genet* **10**, e1004656 (2014).

62. Cangemi, R. et al. Reduced expression and tumor suppressor function of the ETS transcription factor ESE-3 in prostate cancer. *Oncogene* **27**, 2877-85 (2008).

63. Albino, D. et al. ESE3/EHF controls epithelial cell differentiation and its loss leads to prostate tumors with mesenchymal and stem-like features. *Cancer Res* **72**, 2889-900 (2012).

64. Tugores, A. et al. The epithelium-specific ETS protein EHF/ESE-3 is a context-dependent transcriptional repressor downstream of MAPK signaling cascades. *J Biol Chem* **276**, 20397-406 (2001).

65. Johnson, T.R. et al. Loss of PDEF, a prostate-derived Ets factor is associated with aggressive phenotype of prostate cancer: regulation of MMP 9 by PDEF. *Mol Cancer* **9**, 148 (2010).

66. Steffan, J.J., Koul, S., Meacham, R.B. & Koul, H.K. The transcription factor SPDEF suppresses prostate tumor metastasis. *J Biol Chem* **287**, 29968-78 (2012).

67. Haller, A.C. et al. High SPDEF may identify patients who will have a prolonged response to androgen deprivation therapy. *Prostate* **74**, 509-19 (2014).

68. Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215-28 (2015).

69. Tomlins, S.A. et al. TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. *Cancer Res* **66**, 3396-400 (2006).

70. Helgeson, B.E. et al. Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer. *Cancer Res* **68**, 73-80 (2008).

71. Vitari, A.C. et al. COP1 is a tumour suppressor that causes degradation of ETS transcription factors. *Nature* **474**, 403-6 (2011).

72. King, J.C. et al. Cooperativity of TMPRSS2-ERG with PI3-kinase pathway activation in prostate oncogenesis. *Nat Genet* **41**, 524-6 (2009).

73. Zong, Y. et al. ETS family transcription factors collaborate with alternative signaling pathways to induce carcinoma from adult murine prostate cells. *Proc Natl Acad Sci U S A* **106**, 12465-70 (2009).

74. Shin, S. et al. Induction of prostatic intraepithelial neoplasia and modulation of androgen receptor by ETS variant 1/ETS-related protein 81. *Cancer Res* **69**, 8102-10 (2009).

75. Taylor, B.S. et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11-22 (2010).

76. Chen, Y. et al. ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nat Med* **19**, 1023-9 (2013).

77. Baena, E. et al. ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. *Genes Dev* **27**, 683-98 (2013).

78. Aytes, A. et al. ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. *Proc Natl Acad Sci U S A* **110**, E3506-15 (2013).

79. Chen, C.D. et al. Molecular determinants of resistance to antiandrogen therapy. *Nat Med* **10**, 33-9 (2004).

80. Tran, C. et al. Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* **324**, 787-90 (2009).

81. Yu, J. et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* **17**, 443-54 (2010).

82. Oettgen, P. et al. PDEF, a novel prostate epithelium-specific ets transcription factor, interacts with the androgen receptor and activates prostate-specific antigen gene expression. *J Biol Chem* **275**, 1216-25 (2000).

83. Shatnawi, A. et al. ELF3 is a repressor of androgen receptor action in prostate cancer cells. *Oncogene* **33**, 862-71 (2014).

84. Hollenhorst, P.C. et al. Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes Dev* **25**, 2147-57 (2011).

85. Hollenhorst, P.C., Paul, L., Ferris, M.W. & Graves, B.J. The ETS gene ETV4 is required for anchorage-independent growth and a cell proliferation gene expression program in PC3 prostate cells. *Genes Cancer* **1**, 1044-1052 (2011).

86. Selvaraj, N., Budka, J.A., Ferris, M.W., Plotnik, J.P. & Hollenhorst, P.C. Extracellular signal-regulated kinase signaling regulates the opposing roles of JUN family transcription factors at ETS/AP-1 sites and in cell migration. *Mol Cell Biol* **35**, 88-100 (2015).

87. Darnell, J.E., Jr. Transcription factors as targets for cancer therapy. *Nat Rev Cancer* **2**, 740-9 (2002).

88. Koehler, A.N. A complex task? Direct modulation of transcription factors with small molecules. *Curr Opin Chem Biol* **14**, 331-40 (2010).

89. Rathkopf, D.E. et al. Phase I study of ARN-509, a novel antiandrogen, in the treatment of castration-resistant prostate cancer. *J Clin Oncol* **31**, 3525-30 (2013).

90. Mollard, A. et al. Design, synthesis and biological evaluation of a series of novel Axl kinase inhibitors. *ACS Med Chem Lett* **2**, 907-912 (2011).

91. Sorna, V. et al. High-throughput virtual screening identifies novel N'-(1-phenylethylidene)-benzohydrazides as potent, specific, and reversible LSD1 inhibitors. *J Med Chem* **56**, 9496-508 (2013).

92. Kruidenier, L. et al. A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* **488**, 404-8 (2012).

93. Nicodeme, E. et al. Suppression of inflammation by a synthetic histone mimic. *Nature* **468**, 1119-23 (2010).

94. Filippakopoulos, P. et al. Selective inhibition of BET bromodomains. *Nature* **468**, 1067-73 (2010).

95. Delmore, J.E. et al. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904-17 (2011).

96. Dawson, M.A. et al. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* **478**, 529-33 (2011).

97. Mertz, J.A. et al. Targeting MYC dependence in cancer by inhibiting BET bromodomains. *Proc Natl Acad Sci U S A* **108**, 16669-74 (2011).

98. Shi, J. & Vakoc, C.R. The mechanisms behind the therapeutic activity of BET bromodomain inhibition. *Mol Cell* **54**, 728-36 (2014).

99. Wang, S. et al. Ablation of the oncogenic transcription factor ERG by deubiquitinase inhibition in prostate cancer. *Proc Natl Acad Sci U S A* **111**, 4251-6 (2014).

100. Winter, G.E. et al. Drug development. Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science* **348**, 1376-81 (2015).

101. Lu, G. et al. The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science* **343**, 305-9 (2014).

102. Xue, W. et al. Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* **445**, 656-60 (2007).

103. Ventura, A. et al. Restoration of p53 function leads to tumour regression in vivo. *Nature* **445**, 661-5 (2007).

104. Zhang, Q., Zeng, S.X. & Lu, H. Targeting p53-MDM2-MDMX loop for cancer therapy. *Subcell Biochem* **85**, 281-319 (2014).

105. Illendula, A. et al. Chemical biology. A small-molecule inhibitor of the aberrant transcription factor CBFbeta-SMMHC delays leukemia in mice. *Science* **347**, 779-84 (2015).

106. Asada, S. et al. External control of Her2 expression and cancer cell growth by targeting a Ras-linked coactivator. *Proc Natl Acad Sci U S A* **99**, 12747-52 (2002).

107. Asada, S., Choi, Y. & Uesugi, M. A gene-expression inhibitor that targets an alpha-helix-mediated protein interaction. *J Am Chem Soc* **125**, 4992-3 (2003).

108. Shimogawa, H. et al. A wrench-shaped synthetic molecule that modulates a transcription factor-coactivator interaction. *J Am Chem Soc* **126**, 3461-71 (2004).

CHAPTER 2

DEVELOPMENT OF A HIGH-THROUGHPUT SCREENING ASSAY FOR INHIBITORS OF ETS TRANSCRIPTION FACTORS

This research is in preparation for submission to the Journal of Biomolecular Screening. Simon L Currie, Stephen L Warner, Hariprasad Vankayalapati, David J Bearss and Barbara J Graves. Development of a High-Throughput Screening Assay for Inhibitors of ETS Transcription Factors.

Abstract

ETS transcription factors from the ERG and ETV1/4/5 subfamilies are overexpressed in the majority of prostate cancer patients and contribute to disease progression. Here, we developed an *in vitro* assay for ETS transcription factors binding to DNA that is amenable to high-throughput screening. Using ETS1 as a model for ETS transcription factors, we applied these assays to screen 110 compounds that were derived from a high-throughput virtual screen. We find that lower affinity DNA binding sites, similar to those which ERG and ETV1 bind to in prostate cells, allow for higher inhibition from many of these test compounds. Additionally, we demonstrate that these assays are robust for the ETS transcription factors that are overexpressed in prostate cancer, such as ERG, ETV1, and ETV5.

Introduction

DNA sequence-specific transcription factors influence RNA polymerase activity in a gene-specific manner and are among the major factors that regulate normal development and define cellular fate. Transcription factors are often misregulated in human cancers, with the most abundant examples being the downregulation of the p53 tumor suppressor and upregulation of the C-MYC oncoprotein[1]. From this perspective, transcription factors are highly desirable therapeutic targets. Yet, with the exception of steroid hormone receptors, transcription factors are difficult therapeutic targets due to the lack of highly concave ligand-binding surfaces. However, there are some successful examples

of modulating the DNA occupancy of transcription factors through the inhibition of protein-protein interfaces[2,3].

The ETS family of transcription factors contains 28 genes in humans that possess a conserved ETS DNA-binding domain (**Fig. 2.1a**). Factors of the ERG (ERG, FLI1, FEV) and PEA3 (ETV1, ETV4, ETV5) subfamilies are involved in chromosomal rearrangements that result in the overexpression of these proteins in the majority of prostate cancer patients[4]. Preclinical modeling of prostate cancer suggests that the overexpression of ERG, ETV1, or ETV4 contributes to further disease progression[5,6], indicating that these transcription factors are desirable therapeutic targets.

Here we have designed *in vitro* DNA-binding assays for ETS transcription factors that are amenable to high-throughput screening. We piloted these assays using ETS1 and a library of 110 compounds derived from high-throughput virtual screening. Furthermore, we demonstrate that using lower affinity ETS DNA binding sites, similar to those bound by ERG and ETV1 in prostate cells, raises the efficacy of inhibitors of ETS–DNA interactions. Lastly, we establish that these *in vitro* assays can be used with the prostate-cancer relevant transcription factors ERG, ETV1, and ETV5.

Results

ETS1 DN279 (amino acids 279 – 441) was used to pilot *in vitro* assays that could be utilized for high-throughput screening of potential small molecule inhibitors of ETS-DNA interaction. This fragment has robust expression in a

recombinant system and contains the same affinity for DNA as full-length ETS1[7]. The ETS domains of ETS1, ERG, and ETV1 are conserved from an amino-acid sequence and structural perspective (**Fig. 2.1a** and **Fig. 2.2a**). Therefore, ETS1 serves as a good model for the DNA-binding of these other ETS factors and inhibitors that prevent ETS1 from binding to DNA would likely also inhibit ERG and ETV1.

ETS1 DN279 was expressed in *E. coli* and thoroughly purified using a Ni^{2+} affinity column, a cation exchange column, and a size exclusion column (**Fig. 2.2b**). Using electrophoretic mobility shift assay (EMSA) we measured the binding of ETS1 DN279 to DNA with a consensus ETS site (5'-CCGGAAGT-3'), termed SC1 (Selected Clone 1)[8]. The K_D of 0.4 nM is in agreement with previous measurements for this fragment (**Fig. 2.2c**)[9]. The yield of ETS1 DN279 was generally around five milligrams of purified protein per liter of bacterial culture, which provided plenty of protein for this study and could be efficiently scaled up to provide enough protein for a high-throughput screen.

We next optimized screening conditions with the validated ETS1 DN279 for two potential high-throughput assays: fluorescence polarization, and ALPHAScreen. The fluorescence polarization assay utilized a fluorescein-tagged SC1 DNA and measures the change in the polarization of fluorescently emitted light when the DNA is free in solution versus when the DNA is bound by a transcription factor. The ALPHAScreen assay brings beads that engage in fluorescence resonance energy transfer (FRET) into proximity through conjugation to a transcription factor and its recognition DNA site using Ni^{2+}-His_6

and streptavidin-biotin interactions, respectively. First, titration of either DNA demonstrated that using 5 nM of fluorescein-tagged DNA for fluorescence polarization or 10 nM of biotin-tagged DNA for ALPHAScreen minimized the amount of DNA while still retaining a large signal in these assays with ETS1 DN279. Using these amounts of DNA, titration of ETS1 DN279 showed a dose-dependent response in these two assays with a concentration of around 30-50 nM generating maximum signal (**Fig. 2.1b,c**). Based on these titrations, 10 nm concentrations of ETS1 DN279 were used in the fluorescence polarization and ALPHAScreen assays, respectively, for compound screening studies. The maximum signal and the baseline were used to calculate a Z' factor for these assays. The fluorescence polarization assay had a Z' factor of 0.4 and the ALPHAScreen assay had a Z' factor of 0.7. Z' factors above 0.5 are considered to be excellent assays for high-throughput screening purposes[10]. Whereas the ETS1 ALPHAScreen assay already clears this guideline, the ETS1 fluorescence polarization assay is close and could likely be optimized to achieve Z' factors over 0.5.

Computer modeling was utilized to enrich for likely bioactive compounds in the limited number of compounds to be screened using these newly established *in vitro* assays. Briefly, PocketFinder (ICM) and SiteMap (Schrödinger) were used to define ligand-binding pockets in the ETS domain of ETS1 (**Fig. 2.3**). Sequential rounds of virtual screening using one of these defined ligand-binding pockets, ETS1 site 1, culled a starting library of 13 million compounds down to 110 compounds to be tested in the *in vitro* ETS1 DNA binding assays. In addition

to the predicted strength of interaction with ETS1, these compounds were also filtered to optimize chemical diversity and enrich for compounds with favorable physicochemical properties.

A constant concentration of protein and DNA, as indicated above, was used to test the inhibition of each of the 110 compounds that resulted from virtual screening. These compounds were tested at a single concentration of 60 mM and each compound or control was measured in quadruplicates. Using three standard deviations above the baseline (3-SD) as a cutoff, only two compounds in the fluorescence polarization assay and four compounds in the ALPHAScreen assay significantly inhibited the ETS1 DN279 – DNA interaction. Only one of these compounds significantly inhibited this interaction in both assays (**Fig. 2.4a**).

To further investigate these compounds, as well as some additional compounds that were close to the 3-SD cutoff, we utilized the 'TruHits' false positive screen in ALPHAScreen. In this assay, a small molecule that covalently-conjugates biotin and His_6 together is used in lieu of the biomolecules of interest, in this case ETS1 DN279 and SC1 DNA. Compounds that inhibit the false positive assay must do so through a manner inherent to the assay itself such as by absorbing light in the donor or emission wavelengths or by disrupting the streptavidin-biotin or His_6 - Ni^{2+} interactions that conjugate the biomolecules to the ALPHA beads. All of the compounds that strongly inhibited the ALPHAScreen assay also strongly inhibited this false positive assay (**Fig 2.4b**). Only two compounds that had weak to moderate inhibition of the ALPHAScreen assay displayed differential preference for inhibiting the ALPHAScreen assay more

robustly than the false positive assay.

With very few, if any, actual hits from our first round of *in vitro* screening, we next considered potential adjustments to our assays. One potential challenge with this screen is that strength of the ETS1 DN279 – SC1 DNA (K_D = 0.4 nM) interaction might conceal the discovery of lead compounds with relatively lower affinity for ETS1 DN279, which then could be further optimized for inhibition. To address this, we switched from SC1 (5'-GCCGGAAGTG-3'), the highest affinity DNA sequence for ETS1, to a weaker ETS1 binding site, SC13 (5'-ACAGGATATC-3') [8]. By EMSA, ETS1 DN279 bound to SC13 with K_D of 3 nM. This roughly 10-fold weaker interaction is consistent with the difference observed between SC1 and SC13 DNA with other ETS1 truncations[8].

We rescreened the 110 compounds against ETS1 and SC13 DNA. Eighteen of these compounds inhibited the ETS1-SC13 interaction above the 3-SD cutoff, as compared to only four for the ETS1-SC1 interaction (**Fig. 2.5a**). While many of these compounds still inhibited the 'TruHits' false positive assay, 12/20 compounds showed more inhibition in the ETS1-SC13 assay than the false positive assay (**Fig. 2.5b**), compared to only 2/12 compounds that showed more inhibition in the ETS1-SC1 assay than the false positive assay (**Fig. 2.4b**). Therefore, using the weaker interaction of ETS1 with SC13 DNA appears to enable more compounds to disrupt this interaction. Additionally, a significant part of the inhibition observed in the ALPHAScreen assays for most of these compounds appears to come from "off-target" effects in the assay, besides interrupting the ETS-DNA interaction. However, as several of these compounds

display stronger inhibition of the ETS1-SC13 assay than the false positive assay, these compounds may inhibit the ETS1-DNA interaction in addition to the ALPHAScreen assay in general.

In both the ETS1-SC1 and ETS1-SC13 screens, the same compound, CIT-0312, displayed the largest differential between inhibition of ETS1 DN279-DNA assays and inhibition of the false positive assay. Therefore, this compound displayed the most inhibition of the ETS – DNA interaction, as opposed to inhibiting other components of the ALPHAScreen assay. Additionally, CIT-0312 more robustly inhibited the ETS-SC13 interaction (73%) than the ETS-SC1 interaction (36%), as would be expected given the relatively weaker affinity of the ETS-SC13 interaction. This compound inhibited ETS1 DN279 – SC13 DNA interaction in the ALPHAScreen assay with an IC_{50} of 8.0 ± 0.3 mM (mean ± standard deviation). To confirm this inhibition, we tested CIT-0312 using EMSAs. In this orthogonal assay, CIT-0312 inhibited the ETS1 DN279 – SC1 DNA with an IC_{50} of 27 ± 5 mM (mean ± standard deviation) (**Fig. 2.6**). Further investigation demonstrated that this compound lacked specificity as it similarly inhibited cJUN-FOS and FOXA1 transcription factors from binding to their cognate DNA recognition sites. Therefore, this particular compound must be inhibiting the DNA binding of ETS1, as well as other transcription factors, through a nonspecific mechanism that is distinct from the prediction of our *in silico* modeling (**Fig. 2.3**).

Within the ETS family of transcription factors, ERG and ETV1/4/5 subfamily proteins are overexpressed in a number of cancers, including prostate

cancer[4], making therapeutic inhibition of these proteins desirable. To establish that the screening assays used here for ETS1 are suitable for these proteins we expressed and purified full length, His_6-tagged ERG, ETV1, and ETV5. Titrations of these proteins with 10 nM of biotin-tagged SC1 DNA and streptavidin donor and nickel chelate acceptor beads established that these proteins similarly generate robust ALPHAScreen signal, with a maximum signal observed around 20-70 nM, depending on the individual protein (**Fig. 2.7**). Each of these interactions had Z' factors over 0.5 (ERG, 0.8; ETV1, 0.6; ETV5, 0.8), suggesting that they would be suitable for high-throughput screening.

Discussion

In summary, we have established that two *in vitro* assays, ALPHAScreen and fluorescence polarization, are suitable for high-throughput screening of potential small molecule inhibitors of ETS1–DNA interactions. Using weaker affinity DNA, such as SC13, was advantageous for more readily identifying lead compounds from our screens. Interestingly, these nonconsensus DNA sites may be more desirable biologically as well as they more closely resemble the ERG and ETV1 DNA-binding sites that are relevant in prostate cancer[5,6,11]. In contrast, consensus ETS sites, such as SC1, are redundantly regulated by multiple ETS factors and control the expression of housekeeping genes[12]. Lastly, we have demonstrated that ETS factors with high clinical relevance, such as ERG and ETV1, can be used in these screening assays.

Methods

DNA constructs

Human cDNAs corresponding to full-length ETV1, ETV5, and ERG were cloned into the bacterial expression vector pET28 (Novagen) using standard sequence- and ligation-independent cloning strategies and the following oligos:

ETV1 fwd: 5'-CTGGTGCCGCGCGGCAGCCATATGATGGATGGATTTTATGACCAGCAA-3'

ETV1 rev: 5'-GTGCTCGAGTGCGGCCGCAAGCTTTTAATACACGTAGCCTTCGTTGTA-3'

ETV5 fwd: 5'-CTGGTGCCGCGCGGCAGCCATATGATGGACGGGTTTTATGATCAGCAA-3'

ETV5 rev: 5'-GTGCTCGAGTGCGGCCGCAAGCTTTTAGTAAGCAAAGCCTTCGGCATA-3'

ERG fwd: 5'-CTGGTGCCGCGCGGCAGCCATATGATGGCCAGCACTATTAAGGAAGCC-3'

ERG rev: 5'-GTGCTCGAGTGCGGCCGCAAGCTTTTAGTAGTAAGTGCCCAGATGAGA-3'

ETS1 DN279 construct in pET28 was cloned as previously described [7].

Expression and purification of proteins

All proteins were produced in *Escherichia coli* (IDE3) cells. ETS1 DN279 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-b-D-thiogalactopyranoside (IPTG), and grown at 30 °C for ~ 3 hours.

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 40k rpm for at least 30 minutes 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 and dialyzed overnight at 4 °C into 25 mM Tris pH 7.9, 10% glycerol (v:v), 1mM EDTA, 50 mM KCl, and 1 mM dithiothreitol (DTT). After centrifugation at 40k rpm for 30 minutes at 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) and eluted fractions were analyzed by SDS-PAGE for purified ETS proteins. The final, purified protein was then concentrated on a 10-kDa molecular weight cut-off (MWCO) Centricon device, snap-frozen in liquid nitrogen, and stored at -80°C in single-use aliquots for subsequent *in vitro* studies.

Full-length ERG, ETV1, and ETV5 generally expressed more efficiently in the insoluble fraction using IPTG induction as described above. Harvested cells were resuspended as described above, sonicated and centrifuged at 15k rpm for 15 minutes 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 hour at 4 °C, the sample was centrifuged for 40k rpm for at least 30 minutes at 4 °C. The soluble fraction was loaded onto a Ni^{2+} NTA 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 to 500 mM imidazole, the remaining purification steps, 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 full-length proteins compared to ETS1 DN279.

Protein concentrations were measured using averages from the following two methods after ensuring that the concentrations from each method were in agreement with one another (within ~ 2 fold).

Protein concentrations were determined by measuring the absorbance at 595 nm of 20 uL of protein combined with 1 mL of Protein Assay Dye Reagent (diluted 1:5 in deionized water)(Bio-Rad) and comparing to a bovine serum albumin standard curve. Molecular weights for each ETS protein were calculated using the Peptide Property Calculator (Northwestern).

Additionally, absorbance at a wavelength of 280 nm was measured on samples of protein mixed with 6M Guanidine HCl (Thermo Scientific) at a 1:1 ratio and compared to a blank. Protein concentrations were determined using Beer's Law ($Abs280nm = \varepsilon * l * c$) with extinction coefficients for each protein

calculated using Peptide Property Calculator (Northwestern).

Electrophoretic mobility shift assays (EMSAs)

DNA-binding assays of ETS factors utilized duplexed 27-bp oligonucleotides 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 consensus ETS binding site motif. 0.2 nanomoles of each of these oligonucleotides, as measured by absorbance at 260 nM on a NanoDrop 1000 (Thermo Scientific), were labeled with [g-^{32}P] ATP (Perkin Elmer) using T4 polynucleotide kinase (Thermo Scientific) at 37° C for ~ 30-60 minutes. After purification over a Bio-Spin® 6 chromatography column (Bio-Rad), the combined oligonucleotides were incubated at 100 °C for ~ 5 minutes, and then cooled to room temperature over 1-2 hours. For binding reactions, the DNA concentration was diluted to 1 x 10^{-11} M and held constant, whereas protein concentrations ranged ~ 6 orders of magnitude with the exact concentration range dependent on the K_D of the particular protein fragment. Protein concentration was determined after thawing each aliquot of protein, as described above. The binding reactions were incubated for 3 hours at 4° C in a buffer containing 25 mM Tris pH 7.9, 0.1 mM EDTA, 60 mM KCl, 6 mM $MgCl_2$, 200 mg/mL BSA, 10 mM DTT, 2.5 ng/mL poly(dIdC), and 10% (v:v) glycerol and then resolved on an 8% (w:v) native polyacrylamide gel at 4 °C. 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 $[P]_t$ 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; SynergySoftware). 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.

Computational methods

Computational methods were used as previously described[13]. All computational studies used PDB ID 2NNY[14] for the structural coordinates of ETS1. PocketFinder (ICM) and SiteMap (Schrödinger) were used to define ligand-binding sites. Out of the three ETS1 protein and one DNA ligand-binding sites that were defined by PocketFinder and SiteMap, only ETS1 site 1 (**Fig. 2.3**) was used for docking studies.

The compound database was prepared using Ligprep 2.1.23 of the Schrödinger Suite and ICM's inbuilt preparation of three-dimensional ligands. A small molecule ligand library of 13 million compounds was docked against ETS1 using Glide High Throughput Virtual Screen. The top ~ 15% ranked compounds were then redocked with the relatively more computationally expensive Glide standard precision scoring. The top ~ 0.5% ranked were then subjected to further virtual screening using Glide extra-precision and ICM docking and scoring methods.

The final compounds that were identified for *in vitro* screening were the

top ranking compounds from this final round of virtual screening that also met certain physicochemical criteria, such as solubility > 50 mg/mL, permeability > 50 nmol/s, and polar surface area < 120 $Å^2$ as determined by QikProp. In addition to these rankings, redundant compounds were removed using ICM Molcart to improve the chemical diversity of the final set of compounds. Visual inspection of the docking results was used to evaluate binding mode, position, and orientation. In sum, this process resulted in 110 compounds that were purchased and screened using *in vitro* ETS1 DNA-binding assays.

Fluorescence polarization

Fluorescence polarization reactions were performed in the same buffer as described above for EMSAs. Reactions were carried out in 20 mL volumes in black 384 well plates (Corning). The protein, DNA, and compound were incubated for 30 minutes at room temperature, protected from light. Timecourse studies demonstrated that less than 5 minutes were required for the protein-DNA reaction to reach equilibrium; however, we went with a longer incubation time to encourage compound – protein interactions, potentially with significantly lower affinity and kinetics, to also reach equilibrium. Reactions containing up to 5% DMSO showed no influence on the DNA-protein interaction. Plates were then analyzed on an Envision 2104 Multilabel Reader (Perkin Elmer). To calculate percent inhibition, the signal (mp) for each compound was compared to positive (10 nM protein, 5 nM DNA, 0 mM compound; set to 0% inhibition) and negative (0 nM protein, 5 nM DNA, 0 mM compound; set to 100% inhibition) controls.

ALPHAScreen

ALPHAScreen reactions were performed in the same buffer as described above for EMSAs except without 10% glycerol as this caused aggregation of the ALPHA beads. Reactions were carried out in 25 mL volumes in 384 well white OptiPlate-384 HB plates (Perkin Elmer). ALPHAScreen was performed according to manufacturer's recommendations. Briefly, protein, compound, and DNA were incubated at room temperature for 60 minutes, protected from light. Nickel chelate acceptor beads were then added followed by another 60-minute incubation at room temperature, protected from light. Then streptavidin donor beads were added followed by another 60-minute incubation at room temperature. Plates were then analyzed on an Envision 2104 Multilabel Reader (Perkin Elmer). To calculate percent inhibition, the signal (cps) for each compound was then compared to positive (10 nM protein, 10 nM DNA, 0 mM compound; set to 0% inhibition) and negative (0 nM protein, 10 nM DNA, 0 mM compound; set to 100% inhibition) controls.

Comparison of fluorescence polarization and ALPHAScreen assays

In order to compare the assay performance between fluorescence polarization and ALPHAScreen assays for ETS1, the following equation was used to calculate a Z' factor (μ and σ are mean and standard deviation and c+ and c- are positive and negative controls) [10]:

$$Z' = 1 - \frac{3\sigma_{c+} + 3\sigma_{c-}}{|\mu_{c+} - \mu_{c-}|}$$

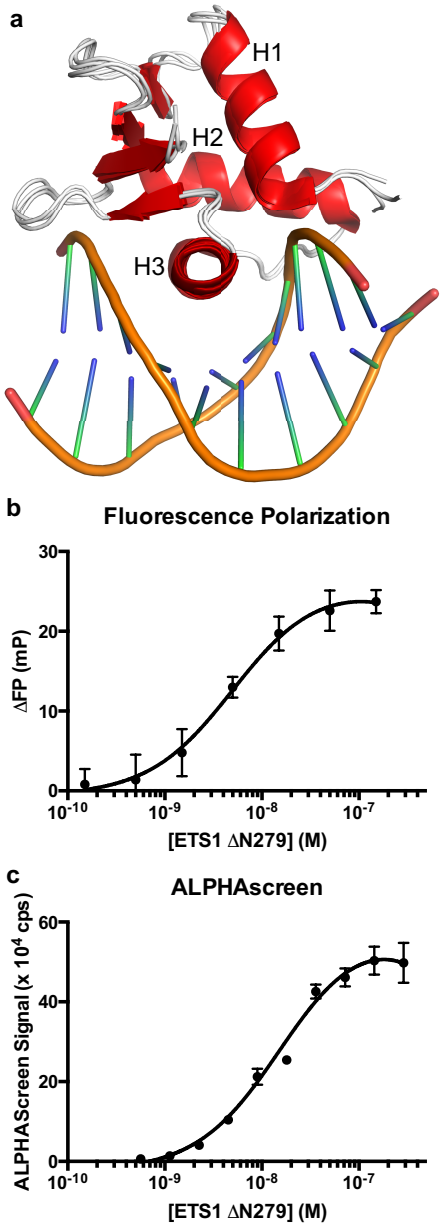


Figure 2.1 ETS1-DNA interaction in fluorescence polarization and ALPHAScreen assays. (**a**) Structural alignment of ETS domains from ETS1 (PDB: 2NNY), ERG (4IRI), and ETV1 (4BNC) bound to DNA. H1, H2, and H3 indicate the order of the α-helices in the ETS domain from N-terminus to C-terminus, according to previous nomenclature. (**b**) Fluorescence polarization assay with a titration of ETS1 DN279 and 5 nM of 3' fluorescein-labeled SC1 DNA. (**c**) ALPHAScreen assay with a titration of ETS1 DN279 and 10 nM of 5' biotin-labeled SC1 DNA.

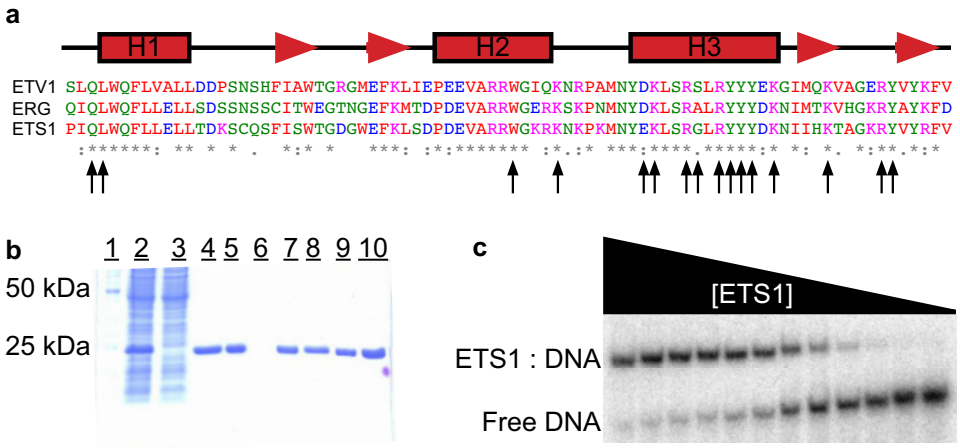


Figure 2.2 Purification and validation of ETS1 protein. (**a**) Sequence alignment of ETS domain from ETV1, ERG, and ETS1. Coloring of amino acid sequence according to Clustal Omega. Rectangles and arrows above sequence alignment refer to α-helices and β-strands, respectively. Arrows below the sequences indicate amino acids that make contact with DNA (Hydrogen bonds, salt bridges, or hydrophobic contacts). (**b**) Coomassie-stained gel of purification of ETS1 DN279. Lanes: (1) marker, (2) soluble fraction, (3) Ni^{2+} affinity flow-through, (4) Ni^{2+} affinity elution, (5) heparin load, (6) heparin flow-through, (7) heparin elution, (8) size-exclusion load, (9) size exclusion peak, (10) size-exclusion peak, 5x quantity in lane 9. (**c**) Electromobility shift assay with ETS1 DN279 and ^{32}P-labelled SC1 DNA. The measured K_D of 0.38 nM is in agreement with literature values (0.44 ± 0.04 nM)[9].

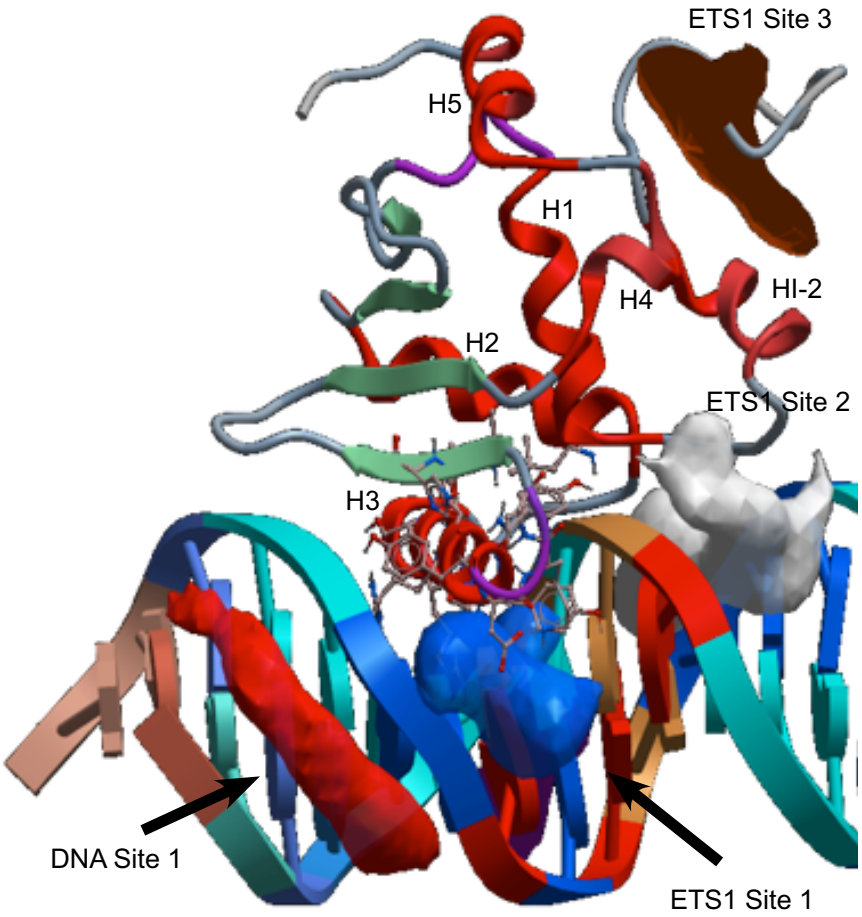


Figure 2.3 Computational modeling of potential inhibitor binding sites on ETS1-DNA interaction. PocketFinder (ICM) and SiteMap (Schrödinger) were used to define the ligand-binding site for docking studies. Only ETS1 site 1 was used for the virtual-screening of compounds in this report. The cartoon of ETS1-DNA complex is from X-ray crystallography-based structure (2NNY.pdb)[14].

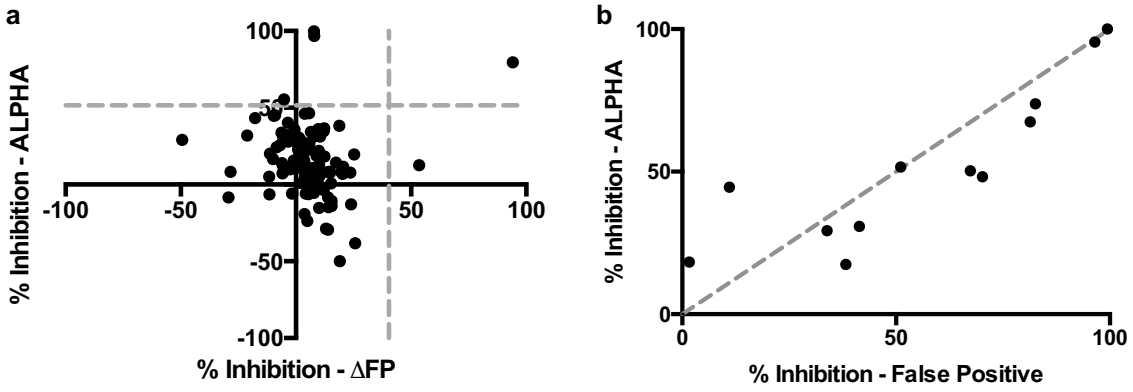


Figure 2.4 *In vitro* screen for inhibitors of ETS1 – SC1 DNA interaction by fluorescence polarization and ALPHAScreen. (**a**) 110 compounds identified from virtual screening were assayed for inhibition of ETS1 – SC1 DNA interaction using fluorescence polarization (DFP) and ALPHAScreen (ALPHA). Percent inhibition was calculated with reference to positive (protein and DNA, no compound) and negative (DNA only, no protein or compound) controls. Dotted gray lines indicate three standard deviations separation from the baseline for each assay. (**b**) Counterscreen of the top hits from ALPHAScreen assay using the TruHits false positive kit. Dotted gray line indicates where the percent inhibition of the ALPHAScreen assay and the false positive assay are equal.

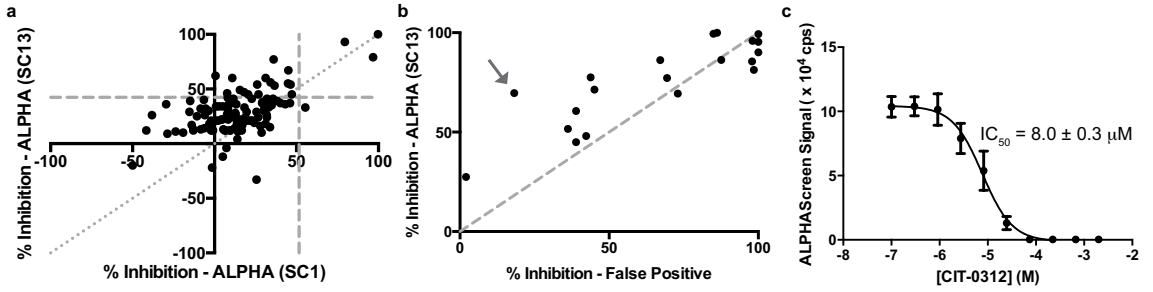


Figure 2.5 Screen for inhibitors of ETS1 – SC13 DNA interaction using ALPHAscreen. (**a**) Comparison of inhibition efficiency of compounds against ETS1 and SC1 DNA (x-axis) and ETS1 and SC13 DNA (y-axis). SC13 is a lower-affinity ETS binding site. Horizontal and vertical dotted gray lines indicate three standard deviations separation from the baseline for each screen. Diagonal, finely dotted line indicates where the inhibition against both of these DNA sequences is equal. (**b**) Counterscreen of the top hits from ETS1-SC13 assay using the TruHits false positive kit. Dotted gray line indicates where the percent inhibition of the ALPHAScreen assay and the false positive assay are equal. Arrow indicates the compound with the largest differential of inhibition of ETS-DNA assay compared to false positive assay, and was used for further studies. (**c**) Representative titration of compound CIT-0312 using ALPHAScreen assay with ETS1 DN279 and SC13 DNA. Indicated IC_{50} of 8.0 ± 0.3 mM (mean ± standard deviation) for this compound was calculated from three replicate experiments.

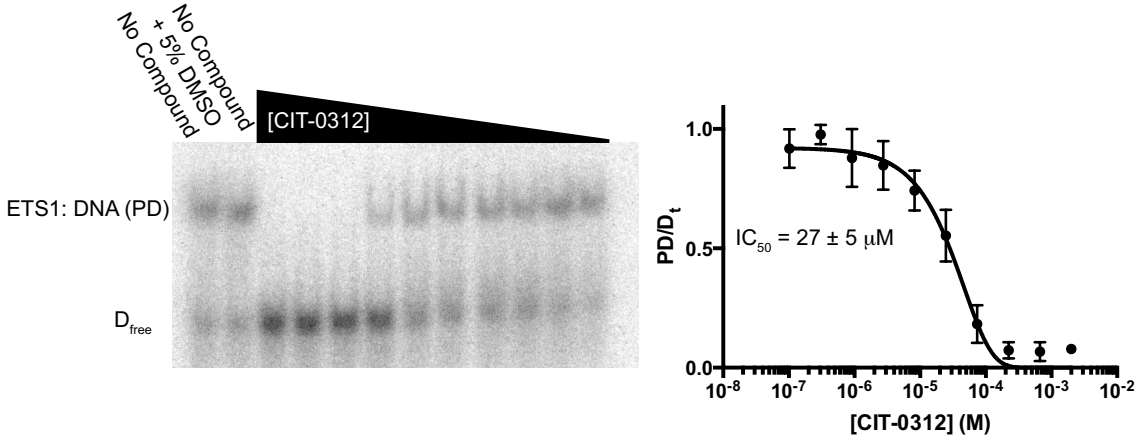


Figure 2.6 EMSA confirmation of compound inhibition. Left, representative titration of CIT-0312 against ETS1 DN279 and SC1 DNA using EMSA. Right, plot of ETS1 DN279 bound to DNA versus total DNA (PD/D_t) against CIT-0312 concentration. Indicated IC_{50} of 27± 5 mM (mean ± standard deviation) for this compound was calculated from three replicate experiments.

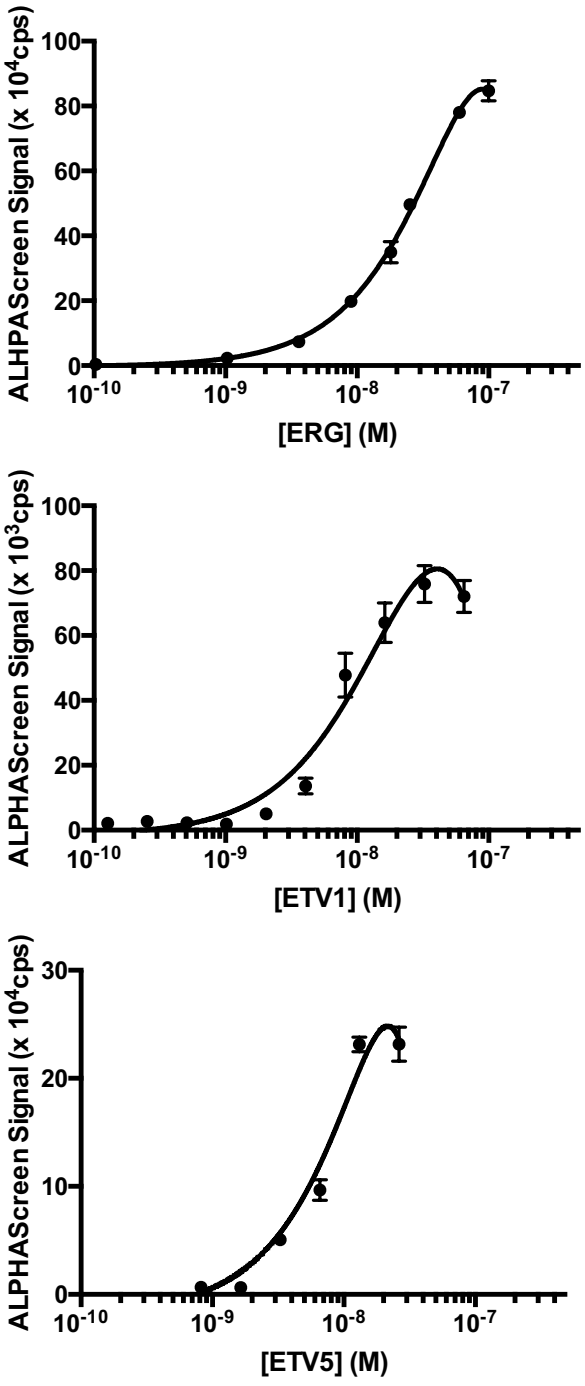


Figure 2.7 ERG-, ETV1-, and ETV5-DNA interactions in the ALPHAScreen assay. ALPHAScreen assay with a titration of ERG (top), ETV1 (middle), or ETV5 (bottom) and 10 nM 5' biotin-labeled SC1 DNA.

References

1. Bhagwat, A.S. & Vakoc, C.R. Targeting transcription factors in cancer. *Trends Cancer* **1**, 53-65 (2015).

2. Illendula, A. et al. Chemical biology. A small-molecule inhibitor of the aberrant transcription factor CBFbeta-SMMHC delays leukemia in mice. *Science* **347**, 779-84 (2015).

3. Hammoudeh, D.I., Follis, A.V., Prochownik, E.V. & Metallo, S.J. Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *J Am Chem Soc* **131**, 7390-401 (2009).

4. Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215-28 (2015).

5. Baena, E. et al. ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. *Genes Dev* **27**, 683-98 (2013).

6. Chen, Y. et al. ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nat Med* **19**, 1023-9 (2013).

7. Pufall, M.A. et al. Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* **309**, 142-5 (2005).

8. Nye, J.A., Petersen, J.M., Gunther, C.V., Jonsen, M.D. & Graves, B.J. Interaction of murine ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. *Genes Dev* **6**, 975-90 (1992).

9. Lee, G.M. et al. The affinity of Ets-1 for DNA is modulated by phosphorylation through transient interactions of an unstructured region. *J Mol Biol* **382**, 1014-30 (2008).

10. Zhang, J.H., Chung, T.D. & Oldenburg, K.R. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* **4**, 67-73 (1999).

11. Hollenhorst, P.C. et al. Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes Dev* **25**, 2147-57 (2011).

12. Hollenhorst, P.C. et al. DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genet* **5**, e1000778 (2009).

13. Sorna, V. et al. High-throughput virtual screening identifies novel N'-(1-phenylethylidene)-benzohydrazides as potent, specific, and reversible LSD1 inhibitors. *J Med Chem* **56**, 9496-508 (2013).

14. Lamber, E.P. et al. Regulation of the transcription factor Ets-1 by DNA-mediated homo-dimerization. *EMBO J* **27**, 2006-17 (2008).

CHAPTER 3

STRUCTURED AND DISORDERED REGIONS COOPERATIVELY MEDIATE DNA-BINDING AUTOINHIBITION OF ETS FACTORS ETV1, ETV4, AND ETV5

This research has been submitted for publication at eLife. Simon L Currie, Desmond K Lau, Jedediah J Doane, Frank G Whitby, Mark Okon, Lawrence P McIntosh and Barbara J Graves. Structured and disordered regions cooperatively mediate DNA-binding autoinhibition of ETS factors ETV1, ETV4, and ETV5.

Abstract

The ETS transcription factors ETV1, ETV4, and ETV5 are often overexpressed in prostate cancer. Here we described the molecular basis of DNA-binding autoinhibition of these factors. 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 ETS domain and the C-terminal inhibitory domain to mediate autoinhibition. Our studies reveal a distinctive mechanism for DNA-binding autoinhibition in the ETV1/4/5 subfamily of ETS proteins and suggest routes for their regulation through cellular pathways and therapeutic interventions.

Introduction

Autoinhibition occurs in diverse proteins and allows for spatiotemporal modulation of activity in response to various inputs such as signaling pathways and macromolecular partnerships[1]. This self-dampening activity can influence the equilibria between the active and inactive states of proteins by serving as the integration point for posttranslational modifications and protein interactions. Alternative intramolecular and intermolecular interactions are often the key attribute for an autoinhibitory element[2,3]. Notably, both structured elements and

intrinsically disordered regions (IDR) can be effective inhibitory elements[4]. However, conformational disorder allows for distinct and adaptable recognition of both intramolecular surfaces and diverse interacting proteins[5]. IDR function in autoinhibition adds to the growing evidence for significant roles for disorder in protein function, thereby regulating cellular processes, including transcription[6].

The *ETS* gene family, which encodes 28 human transcription factors (reviewed in[7]), provides a model system to expand our understanding of the role of IDRs in autoinhibition. Autoinhibition of the conserved DNA binding domain, termed the ETS domain, is reported for multiple family members. In the case of ETS1, the most thoroughly characterized example, a serine-rich IDR inhibits DNA binding through transient phosphorylation-enhanced interactions with the structured ETS domain and flanking N- and C-terminal inhibitory α-helices[8,9]. However, distinct modes of autoinhibition, involving IDRs and appended helices, have been reported for members of different ETS subfamilies[10-13] that lack the serine-rich IDR and flanking α-helices of ETS1. Autoinhibition of a particular ETS factor is also regulated by a distinct set of posttranslational modifications[8,14] and protein-protein interactions[15,16]. Unique regulation corresponding to divergent modes of autoinhibition has been posited as one mechanism to account for specific gene regulation by individual ETS factors[7,17].

The involvement of the *ETS* genes of the ERG and ETV1/4/5 (also known as PEA3) subfamilies in prostate cancer motivated our interest in a better understanding of autoinhibition of these ETS factors. Chromosomal rearrangements involving ERG and ETV1/4/5 subfamilies are observed in the

majority of prostate cancer tumors[9,18,19]. 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[20]. In addition, ETV1 and ETV4 mediate PI3-kinase and Ras signaling pathways, resulting in aggressive and metastatic disease phenotypes[21,22]. Although DNA-binding autoinhibition has been reported within the ETV1/4/5 subfamily[11,16,23,24], detailed characterization is lacking. A mechanistic understanding of the autoinhibition of these factors may provide insights into their roles in prostate cancer progression and 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 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 is an IDR that transiently interacts with the ETS domain and the C-terminal inhibitory domain. Acetylation of the N-terminal inhibitory domain relieves autoinhibition, likely through disruption of its intramolecular interaction with the ETS domain and C-terminal inhibitory domain. From these findings, we propose a model for DNA-binding autoinhibition in the ETV1/4/5 subfamily that evokes a conformational equilibrium modulated by interplay of structured and disordered sequences.

Results

DNA binding by ETV1, ETV4, and ETV5 is autoinhibited

We initially sought to determine the magnitude of autoinhibition in the ERG and ETV1/4/5 subfamilies of ETS factors. Towards this aim, we measured the DNA binding affinities (K_D values), of the full-length proteins and nearly-minimal DNA-binding domains (DBD) for ERG, FLI1, ETV1, ETV4, and ETV5 (**Fig. 3.1a**). Robust autoinhibition was observed in ETV1, ETV4, and ETV5, with the full-length proteins displaying ~10- to 30-fold weaker binding than their minimal DBDs (**Fig. 3.1b-d** and **Table 3.1**). These levels of autoinhibition are comparable to those previously reported for ETS1[8] and ETV6[10]. In contrast, ERG and its subfamily member FLI1 displayed a modest 2- to 3-fold autoinhibition, as also previously reported[11]. Interestingly, the K_D values cluster in a pattern that reflects their subfamily phylogenetic classifications (**Fig. 3.1e**)[25]. 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 further mechanistic studies.

We chose ETV4 as a model factor to further investigate autoinhibition in the ETV1/4/5 subfamily. Partial proteolysis aided the design of truncation boundaries for mapping inhibitory elements (**Fig. 3.2a-c**). We found that the predominant trypsin-resistant fragment, spanning amino acids 165-484, retained comparable levels of autoinhibition to full-length ETV4 (**Fig. 3.3a** 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 synergize to

yield higher than additive levels of inhibition (**Fig. 3.3b**). Hereafter, these regions will be denoted as the NID (N-terminal inhibitory domain) and the CID (C-terminal inhibitory domain), whereas the nearly-minimal DBD will be denoted as an uninhibited species. Based on the autoinhibition mechanisms of other ETS proteins[10,26], we hypothesized that the ETV1/4/5 NID and CID function through direct interactions with the ETS domain (or possibly with each other) to cooperatively inhibit DNA binding (**Fig. 3.3c**).

An α-helix in the CID mediates autoinhibition

To elucidate further the mechanism(s) of autoinhibition by the NID and CID, we undertook crystallographic studies on members of the ETV1/4/5 subfamily. Structures for inhibited fragments of ETV1 and ETV4 (1.4 and 1.1 Å resolution data, respectively), which contain both the ETS domain and the CID (**Fig. 3.4a-b** and **Table 3.3**) as mapped in ETV4 (**Fig. 3.3**), were very similar with a root mean square deviation (RMSD) of 0.16 Å for alignment of their ETS domains. The CID includes an α-helix, termed H4, that 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 Trp344 from H1, Ile407 from the loop between H3 and S3, and Phe420 from S4 **(Fig. 3.4c)**. 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, had no effect on DNA binding (**Fig. 3.4d**). 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 H4.

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 (**Fig. 3.1**) were predicted to have a shorter or possibly unfolded helix H4. As with ETV4 (**Fig. 3.3a**), loss of these homologous residues in ETV1 and ETV5 also activated DNA binding (**Fig. 3.5a**). 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 (1.8 Å resolution; **Table 3.3**). 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 (**Fig. 3.5b**). This alternate position is accommodated in the crystal by intermolecular contacts between the truncated H4 and the ETS domain of a neighboring molecule (**Fig. 3.6**). 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 **Fig. 3.4b** and **3.5b**). 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 alternate position of truncated H4 is potentially a consequence of crystallization, we propose that this repositioning indicates an intrinsic mobility of the CID, an idea pursued further by NMR studies below.

A CID connection to the DNA-recognition helix H3
mediates autoinhibition

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 (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 (**Fig. 3.7a**). 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 3 and 4. 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 (**Fig. 3.5c**). Further comparison with the structure of ETV4 in complex with DNA[27] demonstrated that in the DNA-bound form, H3 of ETV4 is also shifted to a similar position as observed for uninhibited ETV5 (**Fig. 3.5c** and **Fig. 3.7c**). 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 we

observed in the uninhibited species.

Having observed the activating phenotype of the ETV4 mutant L430A (**Fig. 3.4d**) and the variable positioning of the DNA-recognition helix H3 in our crystal structures (**Fig. 3.5c**), we hypothesized that Leu430 inhibits the ETS domain by modulating H3 positioning through an interaction with Ile407 in the H3-S3 loop. We tested this postulate by mutating Ile407 and Leu430 each to alanine separately and in combination. The ETV4 mutant I407A had an almost 5-fold reduction in DNA-binding affinity compared to the wild-type protein, but importantly, this mutation also abrogated the activating nature of L430A in the double mutant I407A/L430A (**Fig. 3.5d**). We conclude that the Leu430-Ile407 interaction is required for CID-mediated autoinhibition and propose that CID-mediated autoinhibition functions by restricting the accessible conformations of H3.

Dynamic features of CID autoinhibition mechanism
detected by NMR

To further investigate the mechanism of autoinhibition, we utilized NMR spectroscopy to compare uninhibited and CID-inhibited species. ^{15}N-HSQC spectra were analyzed for two ETV4 fragments, which displayed the same affinities for DNA as the uninhibited and CID-inhibited species discussed above (**Fig. 3.8a-c**). Based on mainchain chemical shifts, the two proteins in solution contained truncated or full-length H4, as observed by crystallography. Spectral differences demonstrated that amides in the loop between H1 and S1, the C-

terminal end of H3, and H4 were most affected by the presence of the full-length H4 (**Fig. 3.8d,e**). The amino acids in H3 that were perturbed 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.

In additional NMR experiments, the dynamics of uninhibited ETV4 were analyzed using amide hydrogen exchange (HX). 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 (**Fig. 3.8f,g**). In contrast, residues preceding the ETS domain and in loop regions had lower protection factors, which is indicative of conformational flexibility. In addition, many residues, and especially those in H3 and H4, exhibited HX that was too fast to detect by $^1H/^2H$ exchange yet too slow to be measured by $^1H/^1H$ magnetization transfer. This implies that the protection factors of these residues are also low, in the approximate range of 50 to 1000. Thus, the DNA-recognition α-helix H3, and the inhibitory CID helix H4 are conformationally dynamic and readily sample partially unfolded states detectable by HX. Similar behavior is observed with the DNA-recognition and inhibitory helices of ETS1[8,14] and ETV6[12,13]. We conclude that the CID autoinhibitory mechanism requires an equilibrium involving dynamic interactions between helices H4 and H3.

An intrinsically disordered NID connects to CID and

helix H3

Attempts to characterize the NID structurally by crystallization of larger fragments of ETV4 were unsuccessful, potentially reflecting the predicted disordered nature of the NID (**Fig. 3.2d**). Consistent with this prediction, the ^{15}N-HSQC spectrum of the isolated ETV4 NID, amino acids 165-336, displayed limited $^{1}H^{N}$ chemical shift dispersion, (**Fig. 3.9a)**. An analysis of the assigned mainchain ^{1}H, ^{13}C, and ^{15}N chemical shifts confirmed that the NID predominantly samples random coil conformations (**Fig. 3.9b**). Circular dichroism added additional evidence for the overall disordered character of the NID (**Fig. 3.10**).

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[28]. Therefore, we asked whether the NID is still disordered while making inhibitory contacts with the ETS domain. Using intein technology, we ligated the ^{15}N-labeled NID, residues 165-336, to an unlabeled ETV4 fragment spanning the ETS domain and the CID, amino acids 337-436. The NID spectrum retained limited $^{1}H^{N}$ chemical shift dispersion, indicating the lack of any persistent induced secondary structure when covalently linked "in cis" to a fragment spanning the ETS domain and CID (**Fig. 3.9a,c**). This comparison indicates that the NID remains disordered while transiently interacting with the ETS domain and/or CID of ETV4.

With a better understanding of the separated NID and the CID, we next investigated the basis of cooperative inhibition imparted by these two inhibitory

domains. To map the possible interaction of the NID with the ETS domain and the CID of ETV4, we compared the ^{15}N-HSQC spectra of an ETV4 fragment containing the ETS domain and the CID only, amino acids 337-436, with an ETV4 fragment also containing the NID, amino acids 165-436 (**Fig. 3.11a**). As described above, for spectral simplification, an unlabeled NID was added by intein technology to a ^{15}N-labeled fragment containing the ETS domain and CID. In addition to the expected changes at the N-terminus of H1, the ligated NID weakly perturbed amides in H2, the C-terminal region of H3 and the surface-exposed face of the CID (**Fig. 3.11b,c**). These chemical shift perturbations indicate that the NID directly or indirectly affects the structural environment of amides in these regions of the ETS domain and the CID.

To further characterize the NID regions that are responsible for autoinhibition, we investigated the activities of ETV4 variants bearing acetylated lysines. Multiple lysine acetylation sites have been described for ETV4[29,30] with some reported to activate DNA binding[31]. Two known sites of acetylation, Lys226 and Lys260, reside within the NID. Acetylation of either Lys226 or Lys260, independently, resulted in a decrease of DNA binding autoinhibition by 2.8 or 1.6 fold, respectively (**Fig. 3.12**). We propose that acetylation of these lysines relieves autoinhibition by disrupting intramolecular interactions between the NID and the ETS domain and/or CID of ETV4.

Discussion

Cooperative autoinhibition of DNA binding

Here, we have observed that the ETV1/4/5 subfamily of ETS factors include regions N- and C-terminal to the ETS domain that act cooperatively to inhibit DNA binding. Based on the following evidence, we propose that CID-mediated inhibition functions by restricting the binding-competent conformation of the DNA-recognition helix H3. The C-terminal region of the DNA-recognition helix H3 adopts a position in uninhibited ETV5 that is distinct from that in CID-inhibited ETV1 and ETV4, but is very similar to that of the DNA-bound form of ETV4[27]. NMR-spectroscopy confirms that the CID influences H3, consistent with a model of conformationally-induced inhibition. Additionally, Leu430 and Ile407 are required for CID-mediated inhibition and establish a direct link between H4 and H3 for mediating inhibition.

In contrast to the structured CID, the NID is predominantly intrinsically disordered and interacts only transiently with the ETS domain and CID. Most notably, these interactions, which could be direct or propagated through the protein, localize on the C-terminal region of H3, as well as on H4. Therefore, the NID may cooperate directly with the CID by reinforcing its inhibitory positioning, and/or indirectly through interaction with helix H3. Acetylation of the NID activates DNA binding, likely through disruption of the intramolecular interactions between the NID and the ETS domain and/or the CID. In conclusion, we submit that apo-forms of ETV1, ETV4, and ETV5 are in a dynamic equilibrium between conformations of H3 that are competent and incompetent for binding to DNA

(**Fig. 3.13a**). Potential mechanisms for the relief of autoinhibition for these proteins include co-localization with the acetyltransferase p300[29], alternative splicing or use of a translational start site that would remove the NID[4,32,33], and protein-protein interactions that would disrupt or reinforce the intramolecular interactions between the NID and/or the CID and the ETS domain[16,31].

Autoinhibition in ETS family of transcription factors

The characterization of autoinhibition in the ETV1/4/5 subfamily adds to the diversity of structural elements utilized in inhibiting DNA binding by ETS factors (**Fig. 3.13b**). Despite this diversity, common themes of autoinhibition in ETS factors include the low stability of the inhibitory helices and integration of structured and IDR inhibitory elements.

In the cases of ETS1 and ETV6, inhibitory α-helices unfold upon binding DNA[13,34], whereas DNA-bound structures of ETV1/4/5 subfamily members[27] demonstrate that H4 remains folded. However, in uninhibited ETV5, the truncated α-helix H4 shows a drastic repositioning away from the ETS domain, forming intermolecular interactions with a neighboring molecule within the crystal lattice. Along with the low HX protection factors of H4, these structural data suggested that H4 is flexible and our functional data demonstrated a connection between H3 and H4 influences DNA binding.

Regulation of internal molecular motion is important for transcription factor binding to specific DNA sequences[35] and a dynamic, active state has been proposed as a requirement for transcription factor recognition of cognate DNA

sites[36]. The inhibitory sequences of ETS1, ETV6, and ERG have been shown to dampen the internal dynamics of the ETS domain[8,11,12]. Therefore, despite the structural divergence of ETS inhibitory elements, they may share a convergent mechanism of restricting the internal dynamics of the ETS domain. The conformational changes in inhibited versus uninhibited ETV1/4/5 factors suggests that the dampening of motions may also contribute to autoinhibition in this subfamily, but additional experiments are required to confirm this hypothesis.

The modulation of autoinhibition by both structured and disordered regions is a shared feature all structurally-characterized ETS factors[8,10,11]. The cooperation of structured and disordered inhibitory elements in the ETV1/4/5 subfamily factors is most similar to that of ETS1. Four α-helices flanking the ETS domain of ETS1 provide a slight level of inhibition[26], and this autoinhibition is reinforced by an IDR, termed the serine-rich region (SRR)[8,14]. 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 helices of ETS1 and its ETS domain. Furthermore, tyrosine and phenylalanine residues, amino acids that are usually depleted within IDRs[37], are present in the SRR of ETS1[38] and in the NID of ETV1, ETV4, and ETV5. In ETS1, these aromatic residues reside in a repeating pattern of Ser-(Tyr/Phe)-Asp repeats that is not observed in the NID of ETV1, ETV4, and ETV5.

Autoinhibition via IDRs as a molecular target

In light of the challenge of designing an inhibitor targeted at a conserved DNA interface, the presence of heterogeneous inhibitory elements amongst ETS factors provides some hope for new strategies. IDRs often form intermolecular interactions with structured domains, and progress has been made in targeting inhibitors to the structured components of these interactions[39,40]. However, targeting the IDR directly is difficult using structure-based or rational inhibitor design due to their conformational heterogeneity and a lack of binding pockets suitable for energetically favorable interactions with a small molecule inhibitor[41]. Nevertheless, recent reports have described high-throughput screening approaches that successfully identified inhibitors that modulate protein activity through direct interaction with IDRs. Examples include the transcription factor c-MYC, the protein tyrosine phosphatase PTP1B, the TAF2 subunit of the general transcription factor TFIID, and ETV1[42-45]. Although the exact epitope(s) on ETV1 targeted by the inhibitor was not identified structurally, the small molecule inhibited cell migration in a mechanism dependent on the acetylation of Lys33 and Lys116. These lysines are distinct from ETV4 Lys226 and Lys260, which are involved in ETV4 autoinhibition, but also reside within IDRs. These examples of small molecule – IDR interactions demonstrate the feasibility of targeting IDRs and their function. The use of a small molecule to inhibit DNA binding by reinforcing the interaction between the ETS domain and the disordered NID of ETV1/4/5 subfamily proteins would be a novel molecular therapeutic approach.

Methods

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[46]. Point mutations were introduced into the ETV4 plasmid using the QuikChange site-directed mutagenesis protocol (Stratagene). For acetylation studies, codons encoding 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[47].

Expression and purification of proteins

All proteins were produced in *Escherichia coli* (lDE3) 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-b-D-thiogalactopyranoside (IPTG), and grown at 30 °C for ~ 3 hours. 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 40k rpm in a 45 Ti rotor (Beckman) for at least 30 minutes 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 40k rpm for 30 minutes at 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 protein was 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 “auto-induction” protocol[48]. Briefly, bacteria in 250 mL of autoinduction media were grown in 4 L flasks at 37 °C to an OD_{600} ~ 0.6 – 1. The temperature was then reduced to 30 °C and cultures were grown for another ~ 12 – 24 hours. Final OD_{600} values were typically ~ 6 – 12, indicating robust autoinduction. Harvested cells were

resuspended as described above, sonicated, and centrifuged at 15k rpm in a JA-17 rotor (Beckman) for 15 minutes 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 hour at 4 °C, the sample was centrifuged for 40k rpm for at least 30 minutes 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.

Acetylated full-length ETV4 proteins were expressed according to a published protocol[47]. Briefly, expression was induced with IPTG, as described above, but in the presence of 10 mM acetyllysine and 20 mM nicotinamide and a plasmid encoding an amber tRNA that has been mutated in order to be charged with 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 Na_2HPO_4, 500 mM NaCl, 10 mM imidazole, 6M 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 minutes of sonication. The cell lysate was spun down by centrifuging at 25,000 x g for 1 hour 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 Na_2HPO_4, 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 Na_2HPO_4, 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).

Expressed protein ligation and purification

The DNA sequence encoding ETV4 ETS domain and CID, amino acids 337-436, was sub-cloned into bacterial expression vector pEM5B (kind gift from Dr. Pierre Barraud, Université Paris Descartes) between XhoI and BamHI 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[49]. The protein construct was expressed in M9 media, purified under denaturing conditions, and refolded as described above. Protein was concentrated to 0.3 mM as measured by absorbance at 280 nm (predicted ε_{280} 57995 $M^{-1}cm^{-1}$) and stored in the inactive reaction buffer (50 mM HEPES, 200 mM NaCl, 0.1 mM TCEP, pH 7).

The DNA sequence encoding ETV4 NID, amino acids 165-336, was sub-cloned into pEM9B (kind gift from Dr. Pierre Barraud) between NdeI and SapI 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, cells were harvested and resuspended in native buffer (50 mM Na_2HPO_4, 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 under native conditions 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, as measured by absorbance at 280 nm (predicted ε_{280}

63760 $M^{-1}cm^{-1}$), and stored in the inactive reaction buffer (50 mM HEPES, 200 mM NaCl, 0.1 mM TCEP, pH 7).

Purified protein samples containing ^{15}N-labeled ETV4 ETS domain and CID, amino acids 337-436, and unlabeled ETV4 NID, amino acids 165-336, 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. 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 (20 mM Na_2HPO_4, 200 mM NaCl, 2 mM DTT, 0.1 mM EDTA, pH 6.5). For the ligation reaction using ^{15}N-labeled ETV4 NID, amino acids 165-336, and unlabeled ETV4 ETS domain and CID, amino acids 337-436, 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.

Electrophoretic mobility shift assays (EMSAs)

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 consensus 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 [g-^{32}P] ATP using T4 polynucleotide kinase at 37 °C for ~ 30 – 60 minutes. After purification over a Bio-Spin 6 chromatography column (Bio-Rad), the oligonucleotides were incubated at 100 °C for ~ 5 minutes, and then cooled to room temperature over 1 – 2 hours. The DNA for EMSAs was diluted to 1 x 10^{-12} M and held constant, whereas protein concentrations ranged ~ 6 orders of magnitude with the exact concentrations dependent on 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 mg) 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 minutes 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 mg/mL BSA, 10 mM DTT, 2.5 ng/mL 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 $[P]_t$ 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.

Partial proteolysis

For tryptic digestion studies, 20 mL ETV4 (FL) at 20 mM 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 minutes 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.

Crystallization and structure determination

Purified 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, amino acids 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 ammonium sulfate at pH 6.5 and 20 °C. CID-Inhibited ETV4, amino acids 337-441, was crystallized

against a reservoir of 1 M di-ammonium phosphate and 0.1 M sodium acetate at pH 4.5 and 20 °C. Uninhibited ETV5, amino acids 364-457, was crystallized against a reservoir of 0.2 M di-ammonium pH 5.0 and 20% PEG 3350 at pH 5.0 and 4 °C.

Crystals were immersed briefly in mother liquor containing 20% glycerol, and then cryocooled 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[50]. Phases were determined by molecular replacement with Phaser-MR[51] using the ETS domain of ETS1 (1MD0.pdb) as a search model. Models were built with COOT[52] and refined with PHENIX[53]. PyMOL (Schrödinger, LLC) was used to render molecular structure figures.

Model geometries were analyzed by MolProbity[54] within PHENIX. For ETV1, 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, 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, 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.

The coordinate files have been deposited to the RCSB under accession codes 5ILS, 5ILU, and 5ILV.

Circular dichroism spectroscopy

Frozen ETV4 NID, amino acids 165-336, aliquots were thawed, dialyzed overnight into 20 mM Na_2HPO_4, 50 mM NaCl, pH 7.9, and diluted to 25 mM concentration. CD spectra were recorded at 4 °C over 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[55].

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[56] and Sparky[57]. Signals from mainchain and sidechain ^{1}H, ^{13}C, and ^{15}N nuclei were assigned by standard multidimensional heteronuclear correlation experiments, including ^{15}N-HSQC, HNCO, HN(CA)CO, CBCA(CO)NH, and HNCACB[58]. Amide $^{1}H/^{2}H$, after transfer into ~ 99% D_2O NMR sample buffer via a spin column, and CLEANEX-PM $^{1}H/^{1}H$ hydrogen exchange (HX) measurements were recorded using 850 MHz NMR spectrometer and analyzed as described previously[12,59].

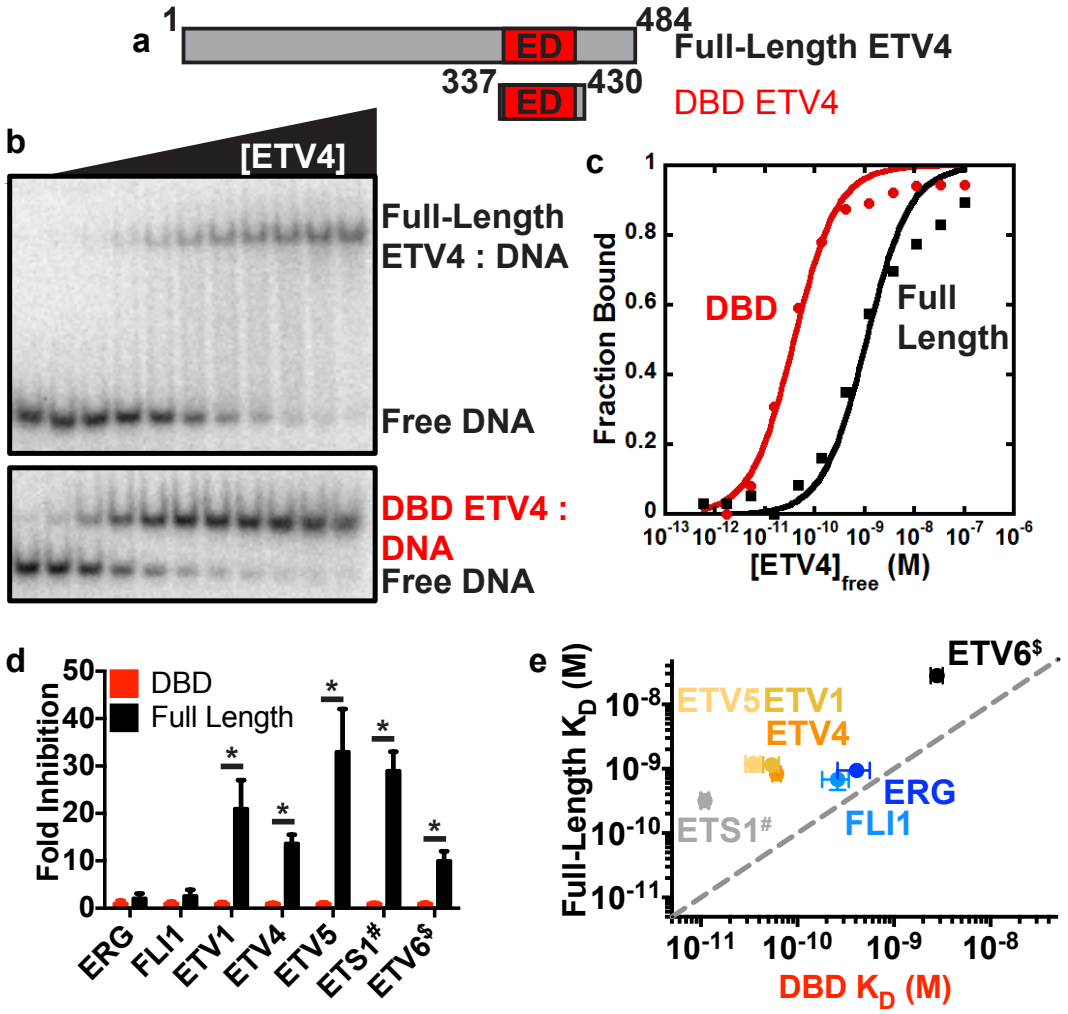


Figure 3.1 Autoinhibition in the ERG and ETV1/4/5 subfamilies. (**a**) Schematic of full-length protein, FL, and nearly minimal DNA-binding domain, DBD, for ETV4. Based on the sequences of all ETS factors, the conserved ETS domain, ED, is noted in red. (**b, c**) Representative examples of EMSA gels and binding isotherms for ETV4 FL or DBD with a double-stranded DNA duplex containing a core ETS binding site, 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[8,#] and ETV6[10,$] data are included for comparison. Mean and standard error of the mean from at least three replicates are plotted; "*" indicates $p < 0.01$. See Table 3.1 for K_D values and number of replicates for each protein. (**e**) K_D values of FL versus DBD for each of the ETS factors tested. The dotted line on the diagonal represents no autoinhibition [i.e., K_D (FL) = K_D (DBD)].

Table 3.1
Equilibrium dissociation constants and fold-inhibition values for ETS transcription factors

ETS Factor	Fragment	K_D (x 10^{-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[8].
[e] Data included for comparison from reference[10].

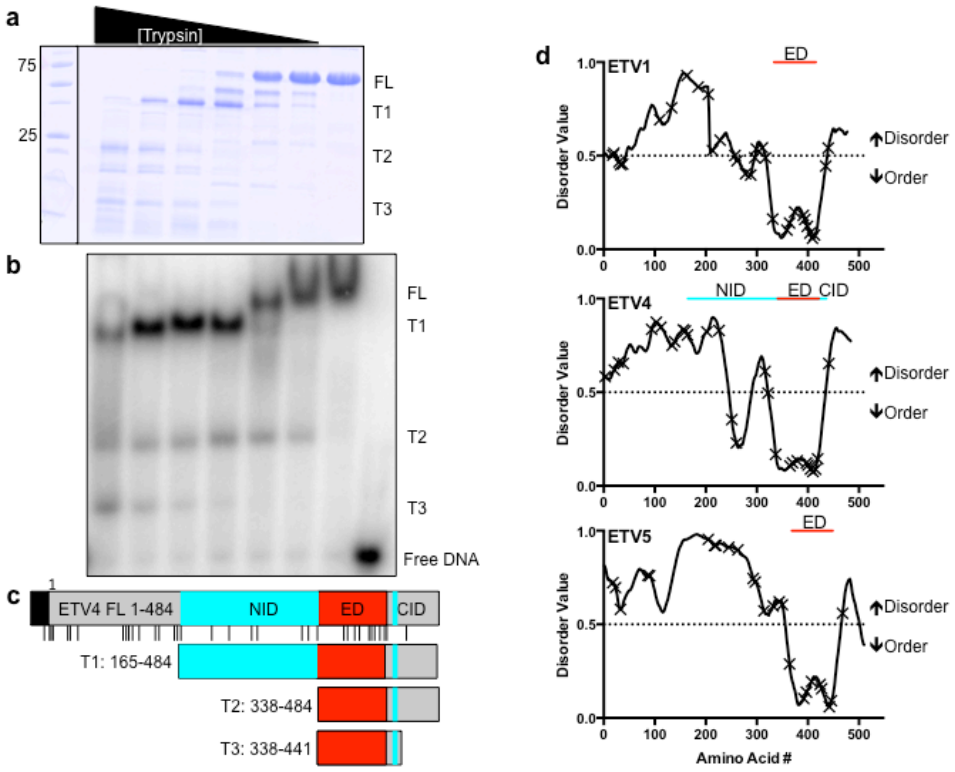


Figure 3.2 ETV4 165-484 is a trypsin-resistant fragment. (**a**) SDS-PAGE gel of partial trypsin proteolysis of ETV4. The left-most lane contains protein molecular weight standards, and the next seven lanes show products from digestion with 450, 150, 45, 15, 4.5, 1.5, and 0 ng of trypsin. A representative 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 arbitrarily 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. The ETS domain (ED) is noted in red, and N-terminal inhibitory domain (NID) and C-terminal inhibitory domain (CID), as identified for ETV4 (Fig. 3.3), 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[60], 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 (Fig. 3.3).

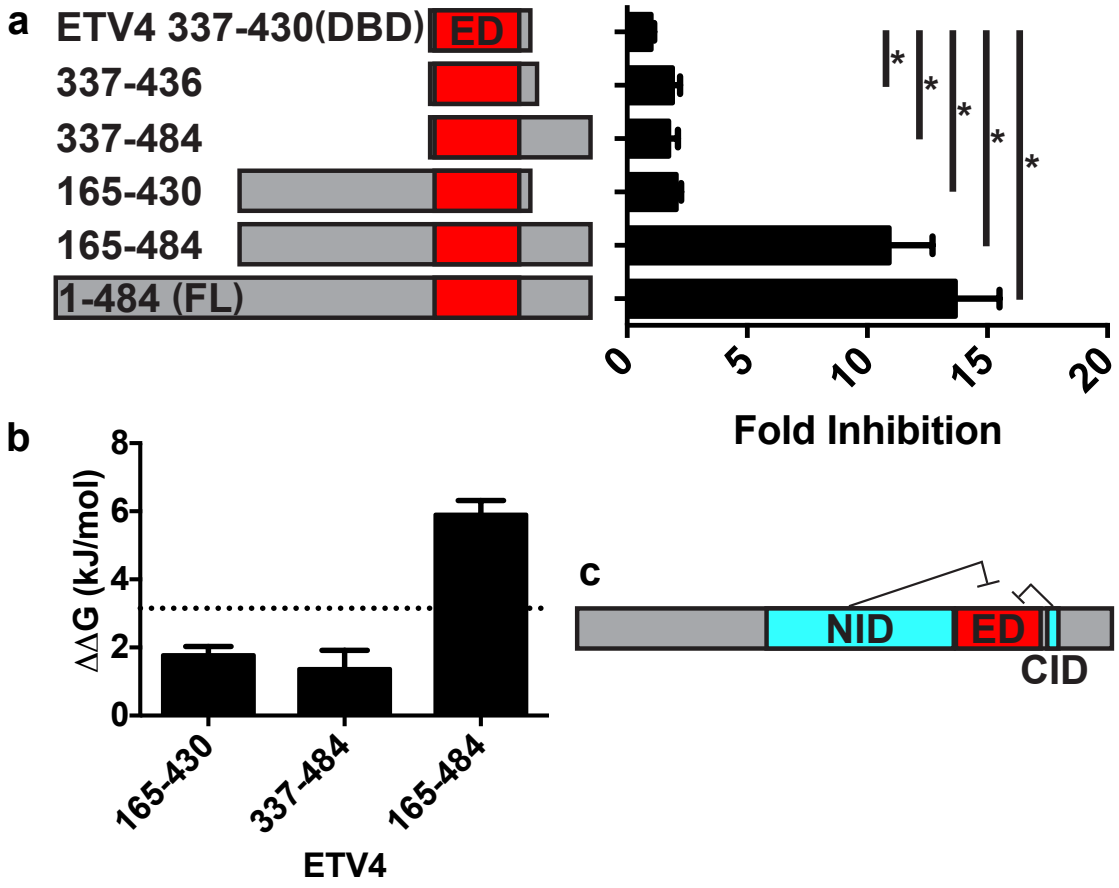


Figure 3.3 NID and CID cooperate to inhibit ETV4 DNA binding. (**a**) Fold inhibition of the ETV4 fragments with mean and standard error of the mean displayed. Fold inhibition calculated as K_D (fragment) / K_D (DBD). "*" Indicates $p < 0.01$. See Table 3.2 for K_D values and number of replicates for each protein. (**b**) DDG = RT ln (K_D ETV4 inhibited fragment / K_D ETV4 337-430) measured for fragments containing the NID, amino acids 165-430, the CID, 337-484, or both, 165-484. The dotted line indicates the sum of the DDG values for 165-430 and 337-484. (**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 and fold-inhibition values for ETV4 fragments

ETV4 Fragment	K_D (x10^{-11} M)[a]	Fold Inhibition[a,b]	n	*p*[c]
337-430 (DBD)	6.1 ± 0.6	1.0 ± 0.1	25	-
337-436	12 ± 2	1.9 ± 0.3	23	0.009
337-484	11 ± 2	1.7 ± 0.4	4	0.04
165-430	12.5 ± 0.3	2.1 ± 0.2	3	0.03
165-484	66 ± 9	11 ± 2	18	3 x 10^{-7}
1-484 (FL)	83 ± 8	14 ± 2	35	4 x 10^{-7}

[a] Mean and standard error of the mean are given for K_D and fold-inhibition values.
[b] ETV4 (DBD) 337-430, 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.

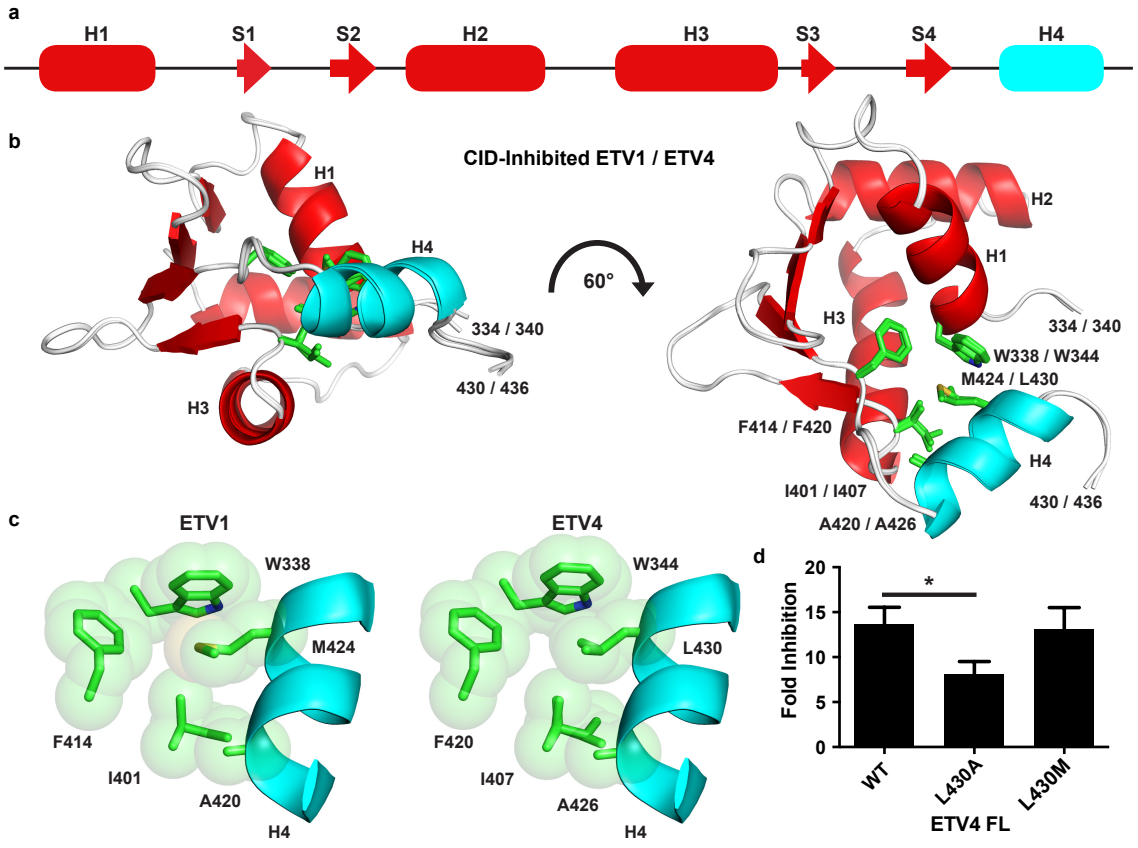


Figure 3.4 CID inhibits DNA binding through hydrophobic contacts between α-helix H4 and the ETS domain**.** (**a**) Schematic of ETS domain, H1-H3 and S1-S4, and α-helix H4 of ETV1, ETV4, and ETV5. ETS domain, red; inhibitory elements, cyan; α-helices, cylinders; β-strands, arrows. (**b**) Cartoon representations of the aligned structures for the ETS domain and CID of ETV1 and ETV4, amino acids 332-435 and 337-441, respectively. Displayed in stick format are Ala426 and Leu430 from α-helix H4 in ETV4, and the analogous amino acids Ala420/Met424 from ETV1, as well as the conserved amino acids in the ETS domain that form a hydrophobic cluster. Numbering for amino acids and endpoints denoted as: ETV1/ETV4. See Table 3.4 for homologous residues and numbering for ETV1 and 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. (**d**) Fold Inhibition of ETV4 FL in its wild-type form, WT (n=35), or with point mutations Leu430Ala (n=11) or Leu430Met (n=3). "*" Indicates $p < 0.05$.

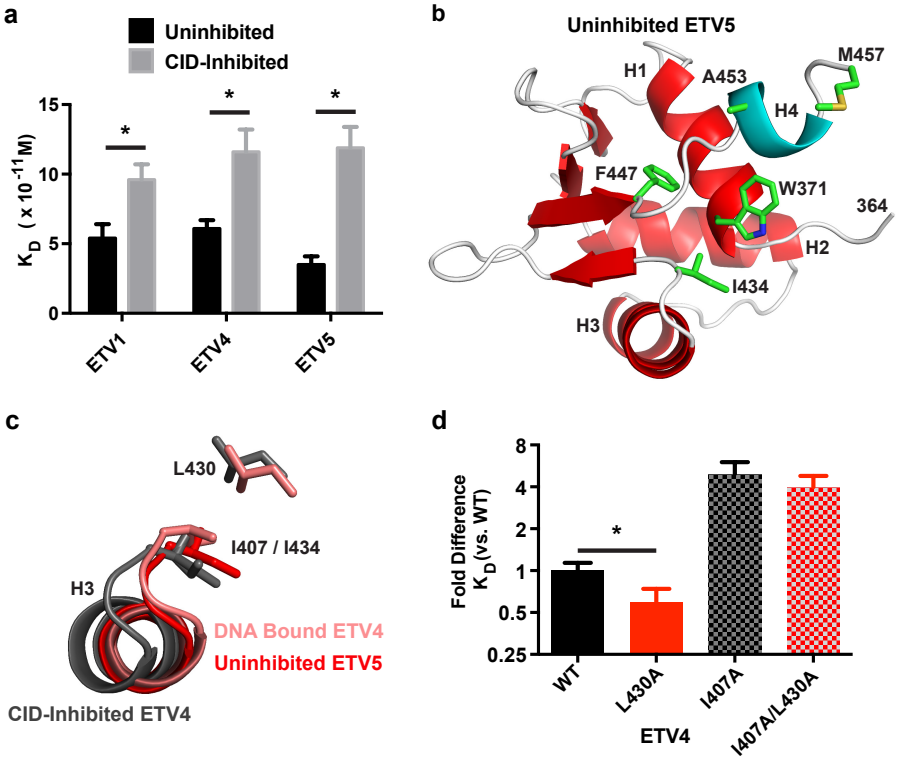


Figure 3.5 Connection between CID and DNA-recognition α-helix H3 mediates autoinhibition. (**a**) Equilibrium dissociation constant, K_D, values for uninhibited and CID-inhibited ETV1, amino acids 332-425 (n=6) and 332-430 (n=7), respectively, ETV4, 337-430 (n=25) and 337-436 (n=23), and ETV5, 364-457 (n=4) and 364–463 (n=7). "*" Indicates $p < 0.05$. (**b**) Crystal structure of uninhibited ETV5, amino acids 364-457, showing the truncated H4 and the same selected sidechains as in Figure 3.4. (**c**) H3 positioning from CID-inhibited ETV4, gray; uninhibited ETV5, red; and ETV4 bound to DNA, pink (4UUV.pdb)[27]. Structures were aligned to DNA-bound ETV4 across the entire protein sequence (See Fig. 3.6). 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. See Table 3.4 for homologous residues and numbering for ETV4 and ETV5. (**d**) Comparison of K_D values for ETV4 FL in its wild-type form, WT (n=35), or with point mutations Leu430Ala (n=11), Ile407Ala (n=4), or both Ile407Ala and Leu430Ala (n=4). "*" Indicates $p < 0.05$. Fold difference for K_D values are relative to ETV4 FL.

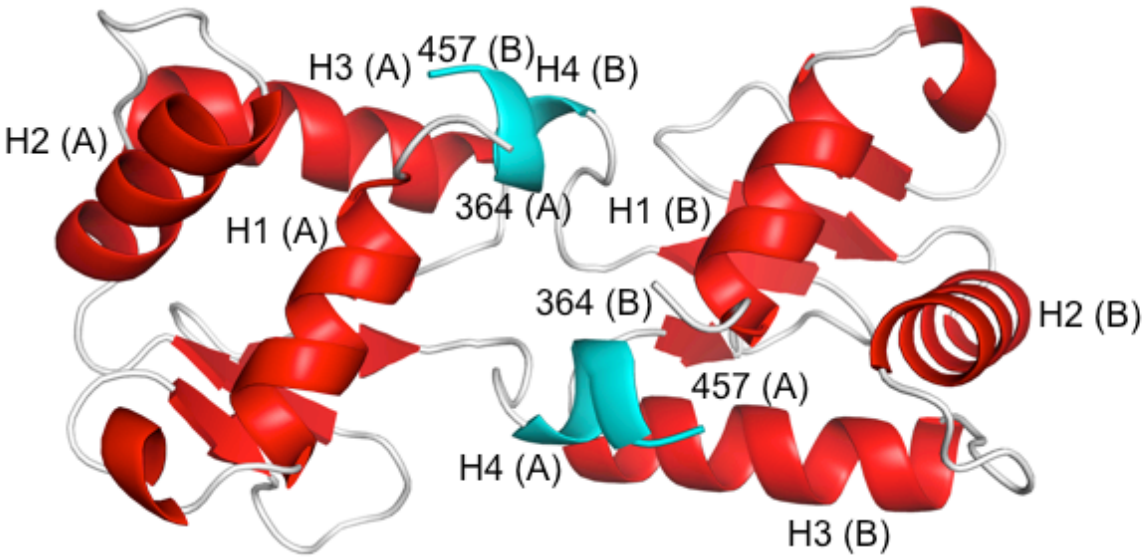


Figure 3.6 Crystal packing of uninhibited ETV5. (A) and (B) distinguish the two molecules of uninhibited ETV5, amino acids 364-457. The contacts between (A) and (B) may affect the position of truncated α-helix H4 (cyan) as compared to the position in solution or in the intact H4 in inhibited ETV5.

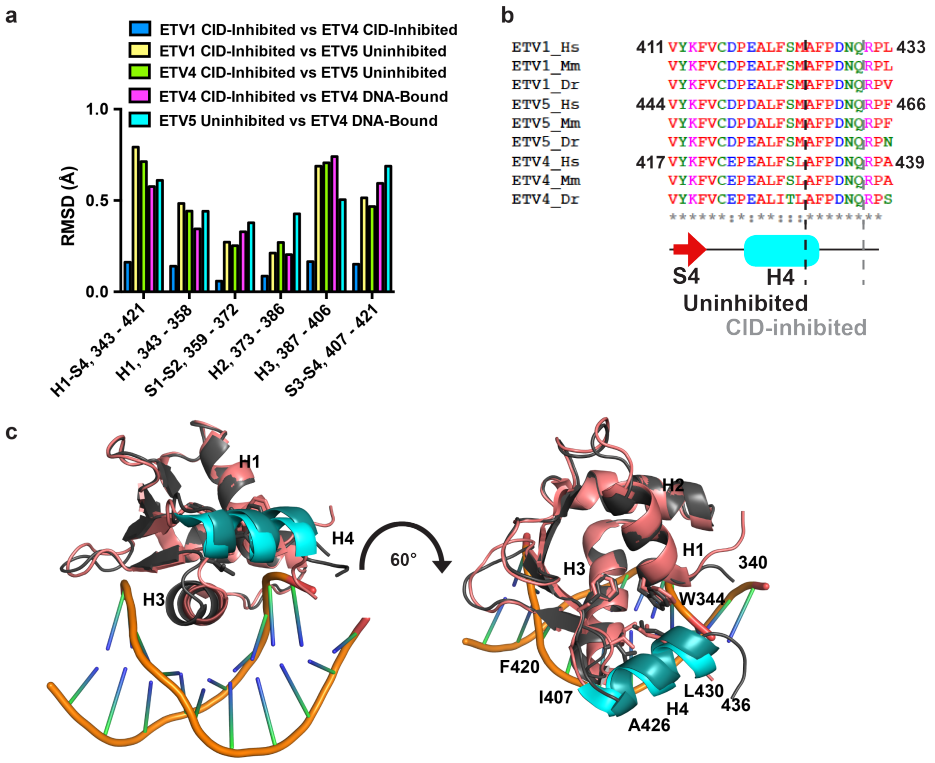


Figure 3.7 Structural comparison of CID-inhibited ETV1 and ETV4 with uninhibited ETV5. (**a**) Root mean square deviations were calculated for backbone atoms to compare the crystal structures of uninhibited ETV5, amino acids 364-457, with CID-inhibited ETV1, 331-435, and ETV4, 337-441, and DNA-bound ETV4, 337-441 (4UUV.pdb)[27]. Secondary structural elements are defined as in Figure 3.4 and the numbering refers only 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 relatively lower for the uninhibited ETV5 / 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[61]. The red arrow and cyan cylinder indicate β-strand S4 of the ETS domain and H4, respectively. The two vertical dashed lines, black and gray, identify truncation endpoints that cause activation or retain CID inhibition, respectively. (**c**) CID-inhibited ETV4 in its apo (this study) and DNA-bound forms (4UUV.pdb)[27] were aligned based on the entire protein sequence. ETS domain and inhibitory residues are colored gray and dark teal, respectively, for the apo ETV4 and pink and cyan, respectively, for the DNA-bound ETV4. Selected side chains are displayed in stick format as in Figure 3.4. Comparison with the apo form demonstrates that there are subtle shifts of backbone atoms in the C-terminus of H3, as well as H4.

Table 3.3
Data collection and refinement statistics

	CID-Inhibited ETV1 332-435	CID-Inhibited ETV4 337-441	Uninhibited ETV5 364-457
Data Collection			
Crystal	BETV108	CETV402	AEAV501
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)
Resolution (Å)	55.00 - 1.40	45.00 - 1.10	30.00 - 1.80
Resolution (Å) (high-resolution shell)	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)	100.0 (100.0)	99.2 (97.3)
<I/σI>	16 (1.9)	5 (0.9)	9 (1.0)
Mosaicity (°)	0.3	0.2	0.09
R(pim)	0.018 (0.243)	0.020 (0.676)	0.039 (0.363)
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) (high-resolution shell)	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} (high-resolution shell)	0.157 (0.217)	0.181 (0.361)	0.186 (0.247)
R_{free} (high-resolution shell)	0.178 (0.237)	0.201 (0.388)	0.234 (0.285)
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

Values in parentheses are for highest-resolution shell. One crystal was used to measure the data for each structure.

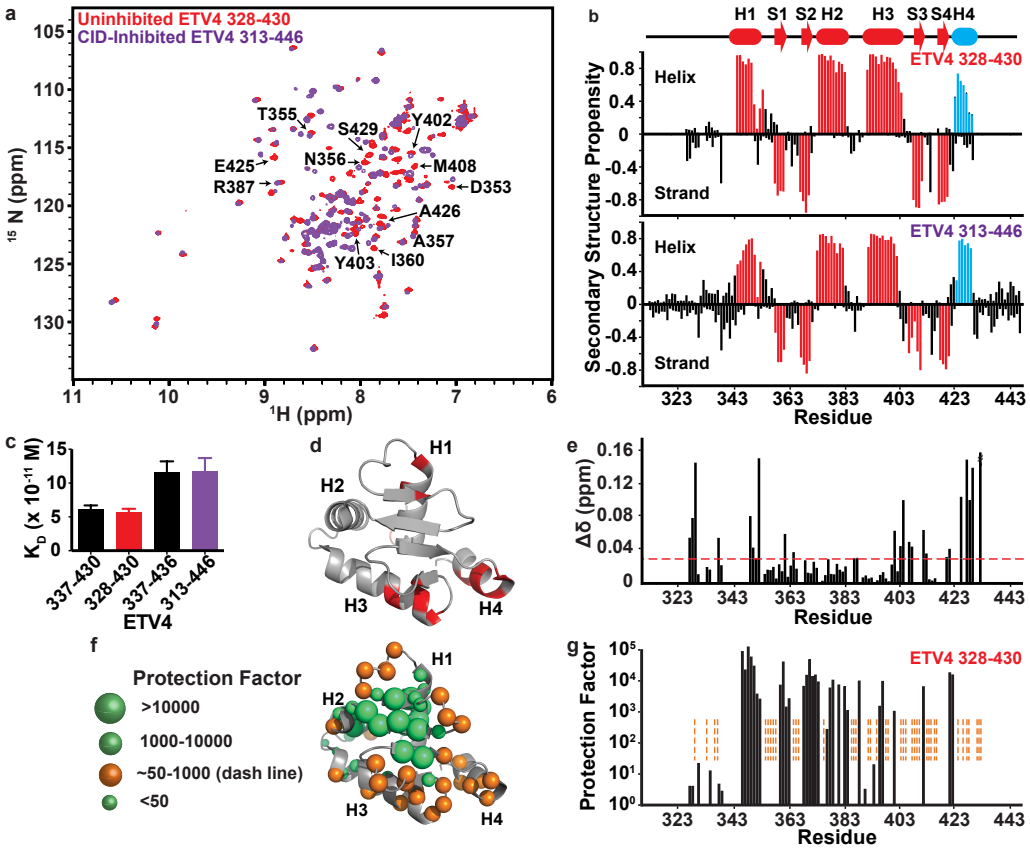


Figure 3.8 The CID perturbs the dynamic DNA-recognition α-helix H3. (**a**) Overlaid ^{15}N-HSQC spectra of uninhibited ETV4, amino acids 328-430, red, and CID-inhibited ETV4, 313-446, purple. (**b**) Secondary structure propensities for the two ETV4 fragments calculated from $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO chemical shifts[62]. Helix, strand (shown as negative values), and coil (not shown) propensities sum to 1. Colored histogram bars identify amides in helices or stands of the ETS domain, red, and CID, cyan, as observed in the X-ray crystal structure of ETV4, top cartoon. (**c**) K_D values for the ETV4 fragments used for NMR spectroscopy studies, red (n=4) or purple (n=4), compared to those used for X-ray crystallography, black (n=25 and n=23 for ETV4 337-430 and ETV4 337-436, respectively). (**d,e**) 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 mapped onto the crystal structure of ETV4 337-436 and plotted as a histogram. Perturbed residues with $\Delta\delta > 0.025$ ppm, horizontal dashed line, are highlighted in red on the structure. (**f,g**) Amide HX protection factors of uninhibited ETV4 328-430, are mapped onto the crystal structure of ETV4 337-441, left, and plotted as a histogram, right. Green spheres indicate protection factors $\gtrsim$ 1000, determined from $^{1}H/^{2}H$ exchange, and $\lesssim$50 obtained from $^{1}H/^{1}H$ CLEANEX measurements. The orange spheres and dashed histogram lines identify amides with resolved, assigned ^{15}N-HSQC signals that exchanged too fast or too slow to be quantitated by these two approaches, respectively, and thus have protection factors in the range of ~ 50 to 1000. Missing values correspond to prolines or residues with unassigned or overlapped NMR signals.

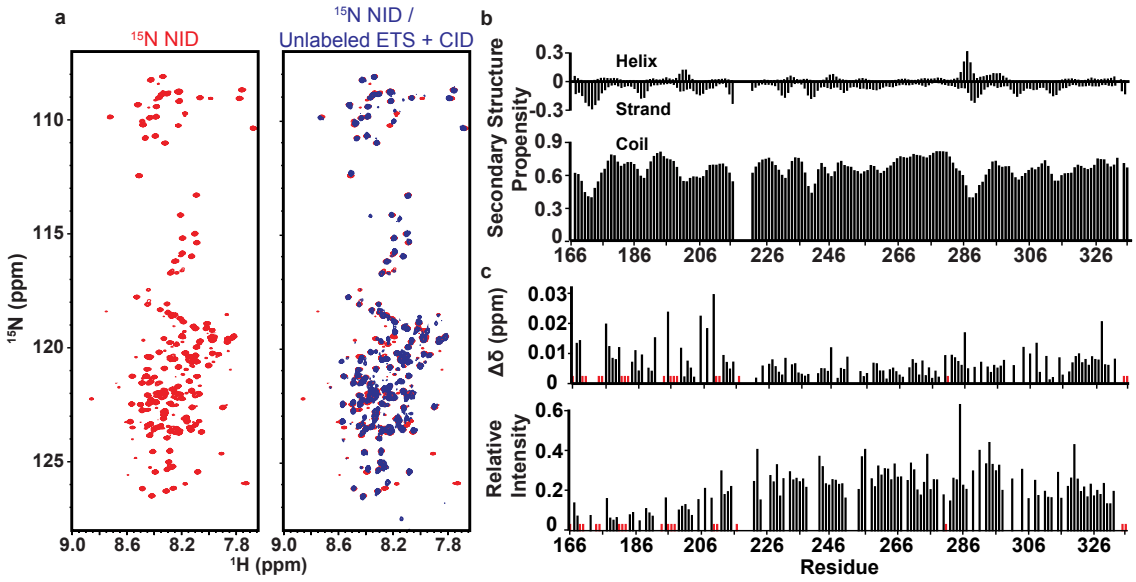


Figure 3.9 The NID is intrinsically disordered. (**a**) Left, the ^{15}N-HSQC spectrum of the ETV4 NID, amino acids 165-336, red. Right, the overlapped spectra of ^{15}N-labled ETV4 NID alone, red, and ligated to the unlabeled ETS domain and CID, amino acids 337-436, blue. (**b**) Secondary structure propensities for a-helical and β-strand, top, and random coil conformations, bottom, calculated from mainchain chemical shifts ($^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO) of the NID with the algorithm MICS [62]. (**c**) Comparison of ^{15}N-HSQC amide chemical shifts ($\Delta\delta = [(\Delta\delta_H)^2 + (0.2\Delta\delta_N)^2]^{½}$), top, and peak intensities, bottom, for the NID from the two spectra in (a). 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. Missing histogram bars correspond to unassigned amides in the isolated NID and prolines.

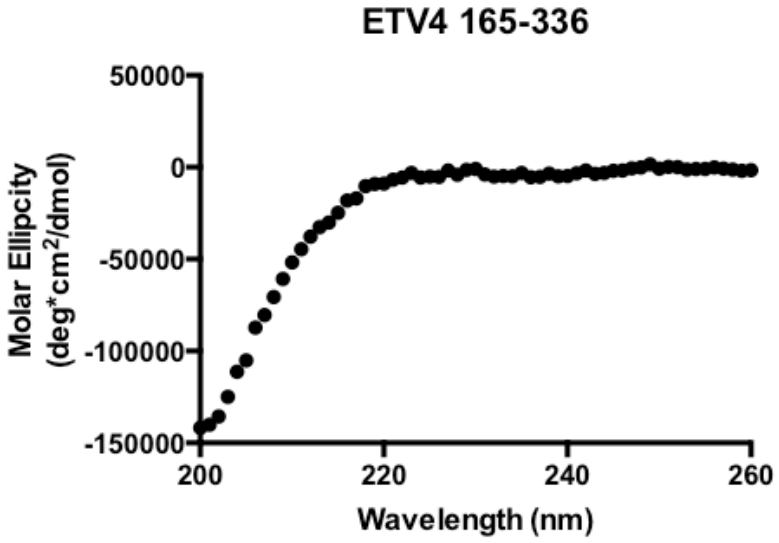


Figure 3.10 Circular dichroism of the NID. Circular dichroism spectrum of the ETV4 NID, amino acids 165-336, at 4 °C and pH 7.9, is indicative of a random coil conformation. Three scans were collected in series and averaged after visually verifying their consistency.

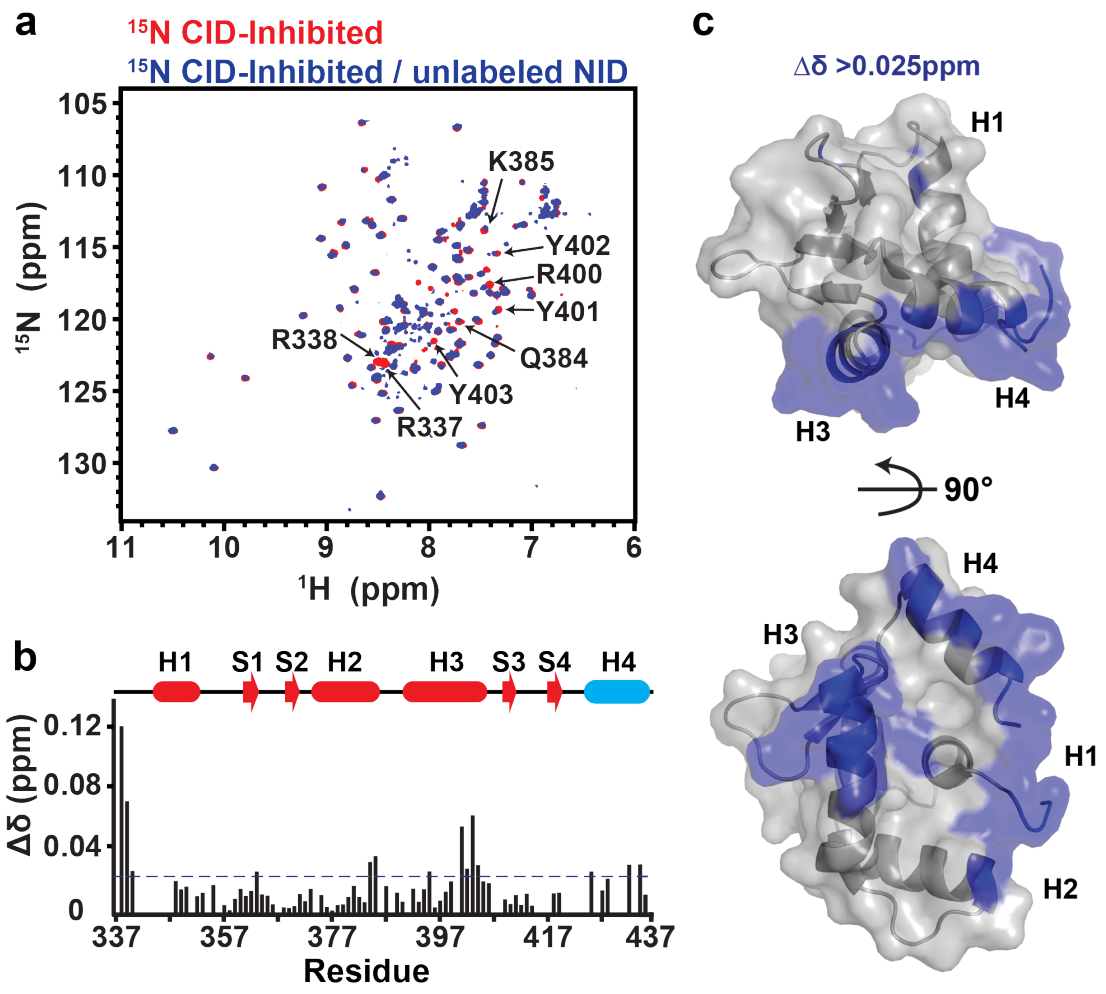


Figure 3.11 The NID perturbs the DNA-recognition α-helix H3 and the CID. (**a**) Overlaid ^{15}N-HSQC spectra of ETV4 ETS domain and CID alone, amino acids 337-436, red, and with the unlabeled NID, amino acids 165-336, added via intein ligation, blue. Selected peaks are labeled with corresponding residues. (**b, c**) The amide chemical shift perturbations, ($Dd = [(Dd_H)^2 + (0.2Dd_N)^2]^{½}$), resulting from the ligated NID are displayed in histogram format and mapped onto the structure of ETV4 ETS domain and CID. Blue, Dd > 0.025 ppm; gray, Dd < 0.025 ppm, prolines, and residues with unassigned NMR signals

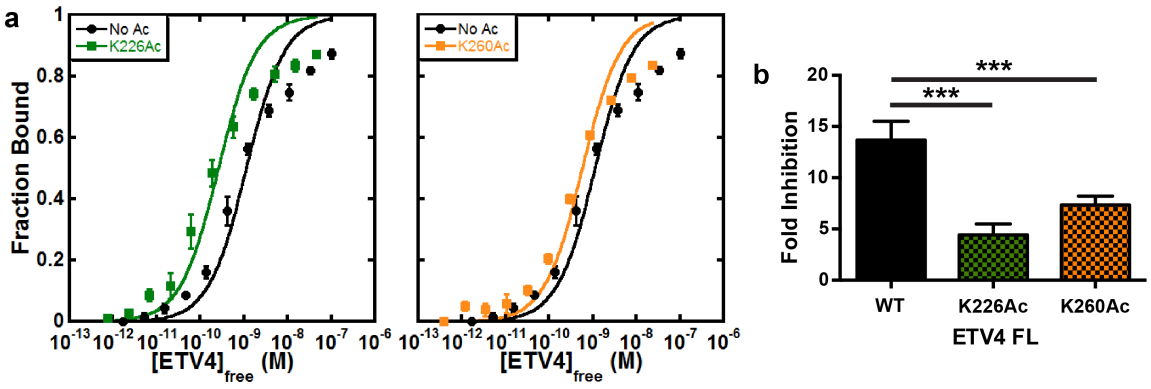


Figure 3.12 Acetylation relieves NID-dependent autoinhibition. (**a**) Binding isotherms for ETV4 (FL), unacetylated, black, and acetylated at Lys226, green, or Lys260, orange. Data points and error bars correspond to the mean and standard error of the mean from four replicates. (**b**) Quantification of fold inhibition comparing ETV4 FL with no acetylation with ETV4 K226Ac and K260Ac. The DNA binding of ETV4 Lys226Ac (K_D, 30 ± 6 x 10^{-11} M) and ETV4 Lys260Ac (K_D, 51 ± 3 x 10^{-11} M) was inhibited 4 ± 1 fold and 7 ± 1 fold, respectively, whereas, ETV4 (K_D, 83 ± 8 x 10^{-11} M) with no acetylation was inhibited 14 ± 2 fold. All fold inhibition values are relative to uninhibited ETV4 337-430 (K_D, 6.1 ± 0.6 x 10^{-11} M). "***" Indicates $p < 0.001$.

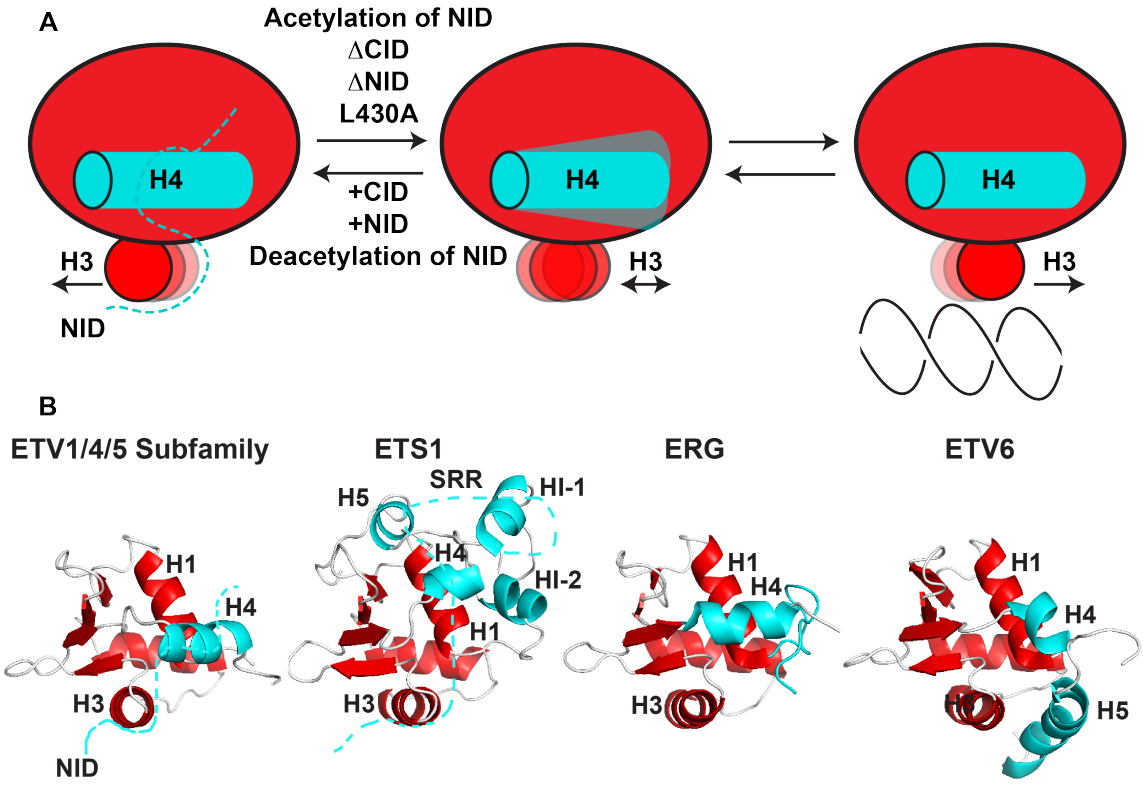


Figure 3.13 Autoinhibition in ETS family of transcription factors (ETS domain, red; inhibitory elements, cyan). (**a**) Model for autoinhibition in ETV1/4/5 subfamily illustrating a hypothetical equilibrium between apo forms that are inactive (left) and active (center) for binding DNA, as well as the DNA-bound form (right). Parameters that influence this equilibrium are listed. Dotted cyan line refers to the disordered NID. (**b**) Examples of structurally-characterized autoinhibited ETS factors: ETV1/4/5 subfamily (this study), ETS1[9,14,15], ERG[11], and ETV6[10,12,13]. Dotted cyan line for ETS1 refers to the disordered serine-rich region (SRR).

Table 3.4
Numbering of homologous amino acids for ETV1, ETV4, and ETV5

ETV1	ETV4	ETV5
Lys228	Lys226	Lys263
Lys257	Lys260	Lys293
Gly333	Gly339	Gly366
Trp338	Trp344	Trp371
Leu344	Leu350	Leu377
Arg394	Arg400	Arg427
Tyr395	Tyr401	Tyr428
Tyr396	Tyr402	Tyr429
Tyr397	Tyr403	Tyr430
Lys399	Lys405	Lys432
Ile401	Ile407	Ile434
Lys404	Lys410	Lys437
Phe414	Phe420	Phe447
Asp417	Glu423	Asp450
Ala420	Ala426	Ala453
Phe422	Phe428	Phe447
Ser423	Ser429	Ser456
Met424	Leu430	Met457
Phe426	Phe432	Phe459

References

1. Pufall, M.A. & Graves, B.J. Autoinhibitory domains: modular effectors of cellular regulation. *Annu Rev Cell Dev Biol* **18**, 421-62 (2002).

2. Morreale, A. et al. Structure of Cdc42 bound to the GTPase binding domain of PAK. *Nat Struct Biol* **7**, 384-8 (2000).

3. Kim, A.S., Kakalis, L.T., Abdul-Manan, N., Liu, G.A. & Rosen, M.K. Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein. *Nature* **404**, 151-8 (2000).

4. Trudeau, T. et al. Structure and intrinsic disorder in protein autoinhibition. *Structure* **21**, 332-41 (2013).

5. Wright, P.E. & Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* **16**, 18-29 (2015).

6. Liu, J. et al. Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873-88 (2006).

7. Hollenhorst, P.C., McIntosh, L.P. & Graves, B.J. Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu Rev Biochem* **80**, 437-71 (2011).

8. Pufall, M.A. et al. Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* **309**, 142-5 (2005).

9. Tomlins, S.A. et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-8 (2005).

10. Green, S.M., Coyne, H.J., 3rd, McIntosh, L.P. & Graves, B.J. DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J Biol Chem* **285**, 18496-504 (2010).

11. Regan, M.C. et al. Structural and dynamic studies of the transcription factor ERG reveal DNA binding is allosterically autoinhibited. *Proc Natl Acad Sci U S A* **110**, 13374-9 (2013).

12. Coyne, H.J., 3rd et al. Autoinhibition of ETV6 (TEL) DNA binding: appended helices sterically block the ETS domain. *J Mol Biol* **421**, 67-84 (2012).

13. De, S. et al. Steric mechanism of auto-inhibitory regulation of specific and non-specific DNA binding by the ETS transcriptional repressor ETV6. *J Mol Biol* **426**, 1390-406 (2014).

14. Lee, G.M. et al. The affinity of Ets-1 for DNA is modulated by phosphorylation through transient interactions of an unstructured region. *J Mol Biol* **382**, 1014-30 (2008).

15. Garvie, C.W., Pufall, M.A., Graves, B.J. & Wolberger, C. Structural analysis of the autoinhibition of Ets-1 and its role in protein partnerships. *J Biol Chem* **277**, 45529-36 (2002).

16. Greenall, A., Willingham, N., Cheung, E., Boam, D.S. & Sharrocks, A.D. DNA binding by the ETS-domain transcription factor PEA3 is regulated by intramolecular and intermolecular protein.protein interactions. *J Biol Chem* **276**, 16207-15 (2001).

17. Hollenhorst, P.C., Shah, A.A., Hopkins, C. & Graves, B.J. Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. *Genes Dev* **21**, 1882-94 (2007).

18. Tomlins, S.A. et al. TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. *Cancer Res* **66**, 3396-400 (2006).

19. Helgeson, B.E. et al. Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer. *Cancer Res* **68**, 73-80 (2008).

20. Tomlins, S.A. et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* **448**, 595-9 (2007).

21. Baena, E. et al. ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. *Genes Dev* **27**, 683-98 (2013).

22. Aytes, A. et al. ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. *Proc Natl Acad Sci U S A* **110**, E3506-15 (2013).

23. Bojovic, B.B. & Hassell, J.A. The PEA3 Ets transcription factor comprises multiple domains that regulate transactivation and DNA binding. *J Biol Chem* **276**, 4509-21 (2001).

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

25. Hollenhorst, P.C. et al. Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes Dev* **25**, 2147-57 (2011).

26. Jonsen, M.D., Petersen, J.M., Xu, Q.P. & Graves, B.J. Characterization of the cooperative function of inhibitory sequences in Ets-1. *Mol Cell Biol* **16**, 2065-73 (1996).

27. Cooper, C.D., Newman, J.A., Aitkenhead, H., Allerston, C.K. & Gileadi, O. 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. *J Biol Chem* **290**, 13692-709 (2015).

28. Dyson, H.J. & Wright, P.E. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **12**, 54-60 (2002).

29. Guo, B., Panagiotaki, N., Warwood, S. & Sharrocks, A.D. Dynamic modification of the ETS transcription factor PEA3 by sumoylation and p300-mediated acetylation. *Nucleic Acids Res* **39**, 6403-13 (2011).

30. Guo, B. & Sharrocks, A.D. Extracellular signal-regulated kinase mitogen-activated protein kinase signaling initiates a dynamic interplay between sumoylation and ubiquitination to regulate the activity of the transcriptional activator PEA3. *Mol Cell Biol* **29**, 3204-18 (2009).

31. Kim, H.J., Kim, S.H., Yu, E.J., Seo, W.Y. & Kim, J.H. A positive role of DBC1 in PEA3-mediated progression of estrogen receptor-negative breast cancer. *Oncogene* (2014).

32. Coutte, L. et al. Characterization of the human and mouse ETV1/ER81 transcription factor genes: role of the two alternatively spliced isoforms in the human. *Oncogene* **18**, 6278-86 (1999).

33. Ingolia, N.T., Lareau, L.F. & Weissman, J.S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789-802 (2011).

34. Petersen, J.M. et al. Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an alpha helix. *Science* **269**, 1866-9 (1995).

35. Tzeng, S.R. & Kalodimos, C.G. Dynamic activation of an allosteric regulatory protein. *Nature* **462**, 368-72 (2009).

36. Tzeng, S.R. & Kalodimos, C.G. Protein activity regulation by conformational entropy. *Nature* **488**, 236-40 (2012).

37. Williams, R.M. et al. The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac Symp Biocomput*, 89-100 (2001).

38. Desjardins, G. et al. Synergy of aromatic residues and phosphoserines within the intrinsically disordered DNA-binding inhibitory elements of the Ets-1 transcription factor. *Proc Natl Acad Sci U S A* **111**, 11019-24 (2014).

39. Lao, B.B. et al. Rational design of topographical helix mimics as potent inhibitors of protein-protein interactions. *J Am Chem Soc* **136**, 7877-88 (2014).

40. Procko, E. et al. A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. *Cell* **157**, 1644-56 (2014).

41. Zhang, Y., Cao, H. & Liu, Z. Binding cavities and druggability of intrinsically disordered proteins. *Protein Sci* **24**, 688-705 (2015).

42. Hammoudeh, D.I., Follis, A.V., Prochownik, E.V. & Metallo, S.J. Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *J Am Chem Soc* **131**, 7390-401 (2009).

43. Krishnan, N. et al. Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. *Nat Chem Biol* **10**, 558-66 (2014).

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

45. Pop, M.S. et al. A small molecule that binds and inhibits the ETV1 transcription factor oncoprotein. *Mol Cancer Ther* **13**, 1492-502 (2014).

46. Li, M.Z. & Elledge, S.J. SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol Biol* **852**, 51-9 (2012).

47. Neumann, H., Peak-Chew, S.Y. & Chin, J.W. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat Chem Biol* **4**, 232-4 (2008).

48. Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207-34 (2005).

49. Barraud, P. & Allain, F.H. 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. *J Biomol NMR* **55**, 119-38 (2013).

50. Otwinowski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326 (1997).

51. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).

52. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).

53. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).

54. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).

55. Greenfield, N.J. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**, 2876-90 (2006).

56. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93 (1995).

57. Goddard, T.D. & Kneller, D.G. Sparky, 3rd ed. (1999).

58. Sattler, M., Schleucher, J., Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Mag Reson Spect* **34**, 93-158 (1999).

59. Hwang, T.L., van Zijl, P.C. & Mori, S. 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. *J Biomol NMR* **11**, 221-6 (1998).

60. Radivojac, P., Obradovic, Z., Brown, C.J. & Dunker, A.K. Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pac Symp Biocomput*, 216-27 (2003).

61. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).

62. Shen, Y. & Bax, A. Identification of helix capping and b-turn motifs from NMR chemical shifts. *J Biomol NMR* **52**, 211-32 (2012).

CHAPTER 4

THE ACTIVATION AND DNA-BINDING DOMAINS OF ETV4 INTERACT WITH MEDIATOR SUBUNIT 25

This research is in preparation for submission to a peer-reviewed journal. Simon L Currie, Jedediah J Doane, Kathryn S Evans, Kathleen A Clark and Barbara J Graves. The activation and DNA-binding domains of ETV4 interact with Mediator subunit 25.

Abstract

The activation domains of DNA sequence-specific transcription factors recruit the Mediator complex through interactions with individual subunits. Previously, it was demonstrated that the N-terminal activation domain of ETV5 interacts with Mediator subunit 25 (MED25). We establish in this report that both the N-terminal activation domain and the DNA-binding domain of ETV4 interact with MED25. The interactions of each of these ETV4 domains with MED25 display distinct kinetics and combined, they contribute to a higher-affinity interaction of full-length ETV4 with MED25. Interaction with MED25 is selective for the ETV1/4/5 subfamily of ETS transcription factors as ETS1 and EHF do not appreciably bind to MED25. This selectivity arises from divergent amino acids within the ETS domain and distinct flanking sequences outside of, but proximal to, the ETS domain. Our findings are the first example of an ETS DNA-binding domain interacting with a Mediator subunit and demonstrate that both activation and DNA-binding domains can contribute to ETS transcription factor – Mediator interactions.

Introduction

Sequence-specific DNA-binding transcription factors regulate eukaryotic transcription through interactions with general transcription factors, coactivators, and chromatin remodelers in order to recruit and affect the activity of RNA polymerase II (Pol II)[1]. The Mediator complex is a critical transcriptional coactivator that serves as a primary conduit for transmitting regulatory signals

from specific transcription factors to Pol II[2]. The 26 subunits of mammalian Mediator (not including the CDK8 kinase module) form distinct modules termed the head, middle, and tail modules. A reconstituted complex comprised of 15 subunits from the head and middle modules represents the minimal functional, or "core"[1], complex required for the general coactivator function of Mediator[3]. In contrast, the presence of and requirement for other subunits of Mediator, primarily those that compose the tail module, is more variable and gene-specific[4-11]. As individual transcription factors recruit the Mediator complex through distinct Mediator subunits[5,7,12,13], the simplest model to explain gene-specificity is that non-core Mediator subunits are only required for the transcription of the genes that they are directly recruited to via interactions with sequence-specific transcription factors[2]. This model accurately depicts many Mediator subunit-regulated genes[4-6,9-11], although more complex mechanisms involving transcription factor partnerships or recruitment of transcriptional repressors have been described[8,14-16].

Transcription factors primarily utilize activation domains (ADs) to recruit Mediator subunits. ADs are often short peptide sequences that are disordered in isolation, but form amphipathic α-helices when binding to their Mediator subunit targets[17-21]. Less frequently DNA-binding domains (DBD) of transcription factors have been implicated in interactions with Mediator subunits[22,23].

ETV1, ETV4, and ETV5 form a subgroup of the ETS (E26-transformation specific) family of transcription factors. This subgroup is aberrantly overexpressed in a subset of prostate cancers[24-26], which promotes PI3-kinase

and RAS signaling pathways resulting in an aggressive and metastatic disease phenotype[27,28]. Previously it was demonstrated that ETV5 interacts with the activator interacting domain (ACID) of Mediator subunit 25 (MED25) through its N-terminal AD[20,29]. Due to chromosomal rearrangements, prostate cancers frequently harbor truncations of ETV1 that lack the AD, suggesting that the AD is dispensable for ETV1's function in prostate cancer[30,31]. Previously characterized interactions between VP16 and MED25 demonstrated that the two ADs of VP16 bind to separate faces of the MED25 ACID domain[17,18]. The N-terminal ADs of ETV5 and VP16 interact with same surface on MED25, therefore we hypothesized that ETV1/4/5 subfamily factors contain an additional MED25-binding site outside of the AD that can interact with MED25 in the absence of the AD.

Here we demonstrate that both the N-terminal AD and the DBD of ETV4 interact with the ACID domain of MED25. The kinetics of each of these interactions are distinct and combined these interactions contribute to a higher-affinity binding of full-length ETV4 with MED25. Using NMR spectroscopy, we detected partially overlapping, yet distinct, faces of the ACID domain interact with these two regions of ETV4. Reciprocal NMR experiments determined that the ACID domain of MED25 perturbs a broad surface on the DNA-binding domain of ETV4. Mutagenesis of residues in the AD or in the DBD of ETV4 both weaken the affinity of the interaction with MED25, confirming the contribution of both domains in this interaction. This is the first reported interaction between a DBD in an ETS factor and a Mediator subunit and provides a rationale for selective

interaction of individual Mediator subunits with the divergent DBDs of ETS factors.

Results

The activation and ETS domains of ETV4 bind to MED25

We used biolayer interferometry to measure the interaction between MED25 ACID (amino acids 391-553) and truncations of ETV4 (**Fig. 4.1**). The N-terminal AD, amino acids 43-84, bound to MED25 with a K_D of 680 ± 60 nM (**Fig. 4.1a** and **Table 4.1**). This value is comparable to previous measurements of the interaction between the AD of ETV5 and MED25 by fluorescence polarization (K_D, 580 ± 20 nM) and by isothermal calorimetry (K_D, 540 ± 40 nM) [20,29]. Intriguingly, full-length ETV4 (amino acids 1-484) bound to MED25 with an approximately 100-fold higher affinity (K_D, 6 ± 1 nM). Correspondingly, ETV4 165-484, which lacks the AD, also bound to MED25 with a slightly higher affinity than that of the AD (K_D, 300 ± 80 nM). Interestingly, the kinetics of the N-terminal and C-terminal portions of ETV4 interacting with MED25 differed (**Fig. 4.1b** and **Fig. 4.2**). Whereas the AD had association and dissociation rate constants (k_a and k_d, respectively) that reflected relatively fast kinetics of interaction, ETV4 165-484 had k_a and k_d values, indicating that the interaction with this fragment and Mediator is defined by relatively slow kinetics. The interaction between full-length ETV4 and MED25 had a comparable k_a to the AD – MED25 interaction and comparable k_d to the ETV4 165-484 – MED25 interaction. Therefore, we

concluded that both the N-terminal AD and a C-terminal component of ETV4 contribute to the higher-affinity binding of full-length ETV4 with MED25.

To pinpoint the region(s) within ETV4 165-484 that are responsible for interaction, with MED25 we tested further truncations. ETV4 165-436 interacts with MED25 with a K_D that is indistinguishable from ETV4 165-484 (**Fig. 4.1a**), although the kinetics of these interactions are slightly different (**Fig. 4.1b**). ETV4 165-336 only measurably interacted with MED25 at concentrations greater than 50 mM, which precluded a full titration to determine the exact K_D of this interaction. An estimate based on only the highest concentration suggests that ETV4 165-336 interacts with MED25 with an approximate K_D of ~ 5 mM (**Fig. 4.2f**). Unfortunately, ETV4 337-436 nonspecifically bound to the streptavidin sensors in the absence of MED25 (**Fig. 4.2g**), which precluded measuring the interaction strength between this truncation and MED25. Despite our inability to measure this interaction, we interpreted the weak interaction of ETV4 165-336 - MED25 (~ 5,000,000 nM), in comparison with the ETV4 165-436 - MED25 interaction (290 ± 70 nM), to indicate that the C-terminal contribution in binding with MED25 requires the DBD of ETV4, amino acids 337-436.

Next we tested the specificity of the interaction with MED25 amongst ETS transcription factors. The N-terminal activation domain is conserved between ETV1, ETV4, and ETV5, but is not detected by sequence similarity in other ETS proteins. However, there is conservation amongst other ETS factors for the highly conserved C-terminal component, which spans the ETS domain. Unlike ETV4, ETS1 and EHF, two ETS factors outside the ETV1/4/5 subfamily, do not

measurably interact with MED25 (**Fig. 4.3**). This suggests that the specific C-terminal component(s) of ETV4 responsible for interaction with MED25 are not widely conserved across ETS proteins.

ETV4 ETS domain and AD perturb distinct surfaces of MED25

To obtain amino acid-resolution of the interactions between ETV4 and MED25, we utilized NMR spectroscopy and a previously characterized ^{15}N-HSQC spectrum of the ACID domain of MED25[17,18]. We titrated ^{15}N-labeled MED25 ACID with either the AD, amino acids 43-84, or the DBD, amino acids 337-436, of ETV4. The addition of the AD of ETV4 resulted in widespread changes in the ^{15}N-HSQC spectrum of MED25 ACID (**Fig. 4.4a**) in the form of chemical shift perturbations (**Fig. 4.4b**) and changes in the relative intensities of peaks (**Fig. 4.4c**). In comparison, the addition of the DBD resulted in subtler changes in the spectrum of MED25 ACID (**Fig. 4.5**). We mapped the strongest perturbations from both ETV4 titrations (**Fig. 4.4b,c** and **Fig. 4.5b,c**) onto the structure of MED25 ACID (**Fig. 4.6**). The ACID domain is a seven-stranded β-barrel with four α-helices (**Fig. 4.6a**). Two of the α-helices, H2 and H4, are oriented parallel to the lengthwise edge of the β-barrel. The ETV4 AD and ETS domain perturb distinct, yet partially overlapping, subsets of surface-exposed residues on MED25 ACID. The overlapping perturbations are centered on H2 and H4 (**Fig. 4.6b,c**). The distinct portions of the AD- and DBD-perturbed surfaces reside on opposing sides of the β-barrel, with the AD-perturbed residues

on the S3/S5 side and the DBD-perturbed residues on the S4/S6/S7 side. Therefore, we conclude that the DBD and the AD of ETV4 interact with distinct, yet overlapping surfaces on MED25 ACID.

MED25 ACID activates the DNA-binding of ETV4
through interaction with divergent residues

Previously, studies structurally-characterized the interaction between the ETV5 AD and MED25, demonstrating that the AD becomes more helical in character in the MED25-bound state and that the hydrophobic residues in the amphipathic α-helix of the AD are critical for this interaction[20]. Based on the robust sequence conservation between the AD of ETV1, ETV4, and ETV5 (**Fig. 4.7**), we surmised that the ETV4 AD would interact with MED25 in a conserved fashion. Therefore, we focused on characterizing the residues that are important for the DBD interaction with MED25.

We titrated ^{15}N-labeled ETV4 DBD, amino acids 337-436, with unlabeled MED25 ACID (**Fig. 4.8a**). MED25 perturbed a broad surface on the ETV4 DBD with residues from all four α-helices and the β-sheet being influenced (**Fig. 4.8c,d**). Because of the observed selectivity for MED25 interacting with ETV4 but not ETS1 or EHF, we examined the conservation of the residues perturbed by MED25. In general for ETS transcription factors, the interior core of the ETS domain and amino acids that form the DNA-binding interface were most highly conserved (**Fig. 4.9**). In contrast, exterior facing amino acids that do not form the DNA-binding interface were more divergent. The amino acids that were

perturbed by MED25 also reflected this trend (**Fig. 4.9b**). In addition, the peaks with the largest chemical shift perturbation and change in relative peak intensities, Ser429 and Glu425, respectively, reside just outside of the ETS domain in helix H4, a structural element that is unique to the ETV1/4/5 subfamily. Therefore, we reasoned that surface-exposed, divergent amino acids in the ETS domain and H4 are critical for the interaction between ETV4 and MED25.

To test the functional importance of individual amino acids, we mutated select residues to alanine in the context of full-length ETV4 and used biolayer interferometry to analyze the affect of these mutations on the binding affinity of MED25 ACID for ETV4. Mutations from the AD, F54A, and the DBD, S429A, resulted in 10- and 30-fold increases, respectively, in K_D values (**Fig. 4.10a** and **Table 4.2**). Interestingly, the AD mutation F54A had more of an influence on the k_a of the interaction whereas the DBD mutation S429A had more of an influence on the k_d of the interaction (**Fig. 4.10b** and **Table 4.2**). These mutants support the observation from ETV4 truncations (**Fig. 4.1**) that the AD and the DBD rate constants reflect relatively fast and slow kinetics, respectively, for these domains' interactions with MED25.

We previously demonstrated that helix H4 inhibits the DNA-binding of the ETS domain of ETV4 (Chapter 3). Therefore, we next tested whether interaction with MED25 modulates the DNA-binding affinity of ETV4. Addition of MED25 results in a more prominent ETV4:DNA EMSA band (**Fig. 5.11a**). As a control, equivalent amounts of MED25 had no effect on DNA in the absence of ETV4. MED25 also slightly increases the affinity of ETV4, but not ETS1, for DNA (**Fig.**

5.11b). While this activation is slight (1.4 ± 0.2 fold), it is approaching statistical significance (p = 0.1 in a heteroscedastic t-test) and similar in magnitude to the level of autoinhibition conferred by H4 (1.9 ± 0.3 fold) (Chapter 3). Therefore, this data suggests that interaction with MED25 activates the DNA-binding of ETV4 through relief of H4-mediated autoinhibiton of the ETS domain.

Discussion

ETV4 AD and DBD contribute to MED25 interaction

Here, we have observed a bipartite interaction between full-length ETV4 and the ACID domain of MED25. Both the AD and the DBD of ETV4 are capable of independent interaction with MED25 ACID. The higher-affinity interaction of full-length ETV4 and MED25 reflects the fast association rate of the AD-MED25 interaction and the slow dissociation rate of the DBD-MED25 interaction. In support of distinct contributions from the AD and DBD, these domains interact with distinct, yet overlapping regions on the surface of MED25 ACID. Interaction with MED25 is not conserved in the ETS transcription factors ETS1 and EHF. While the AD of ETV1/4/5 is specific to that subfamily, the presence of a conserved ETS domain in all ETS transcription factors would seemingly confound the specificity of interaction amongst ETS factors with MED25. However, we found that MED25 perturbs the b-sheet of the ETS domain, and α-helix H4 that resides just C-terminal of the ETS domain. The β-sheet is poorly conserved amongst ETS factors, and H4 is specific to the ETV1/4/5 subfamily in terms of its sequence and relative positioning relative to the ETS domain.

Therefore, we suggest that these two components create a unique interface for MED25 that is distinct from most, if not all, ETS factors outside of the ETV1/4/5 subfamily. The point mutation of serine 429 to alanine drastically diminishes the interaction of ETV4 with MED25, reinforcing the importance of H4 to this interaction. Lastly, we find that the interaction of MED25 strengthens the binding of ETV4 to DNA, likely through ablation of the autoinhibition that H4 imparts on the ETS domain. In summary, we have found that the interaction with MED25 involves both the AD and DBD and activates the DNA binding of ETV4.

The ETV4 AD-MED25 interaction is consistent with the previously reported characterizations of the ETV5 AD-MED25 interaction[20,29] based on the functional importance of conserved hydrophobic residues in both ADs, and the similar interaction surface with MED25 for both ADs. In contrast, we found that the DBD of ETV4 also interacts with MED25. All ETV4 amino acids implicated by NMR in the DBD-MED25 interaction are conserved in ETV1 and ETV5. In addition, ETV1, ETV4, and ETV5 have structurally-conserved DBDs[32]. Therefore, it is highly likely that the DBD-MED25 interaction observed here for ETV4 is also conserved in ETV1 and ETV5.

Mediator interactions with other ETS transcription factors

The TCF subfamily of ETS factors, ELK1[7,9,33], ELK3[9], and ELK4[9], as well as ELF3[19] interact with Mediator subunit 23 (MED23). The transcriptional activity of ELK1 was largely ablated in the absence of MED23, whereas ELK3 and ELK4

demonstrated a more modest dependence on MED23[9]. This finding suggested that ELK3 and ELK4 are capable of interacting with the Mediator complex even in the absence of MED23. However, recent work suggests that ELK1 can also associate with the Mediator complex even in the absence of MED23[33]. The redundancy of ELK1, ELK3, and ELK4 interacting with multiple subunits of the Mediator complex is a common feature that has been noted for the ADs of other transcription factors[34-37].

The ADs of ETV1/4/5 and TCF subfamilies differ in sequence (**Fig. 4.7**). However, due to the flexible nature of ADs, functional binding to common protein targets is observed even in the case of significant sequence divergence[21]. For example, the ADs of ATF6 and VP16, which also bind to MED25 ACID[17,18,38], show no sequence conservation with ETV1/4/5 ADs. Further investigation is warranted to investigate whether ETV1/4/5 and TCF ADs have overlapping Mediator subunit targets. Additionally, the above studies on TCF factors utilized only the AD of these proteins. Further analysis of TCF, or other ETS factors, with Mediator subunits should also examine the possible contribution of the DBDs to these interactions.

ETV4-MED25 interaction and prostate cancer

ETV1/4/5 subfamily factors are often overexpressed in prostate cancer, and in a subset of these instances, the N-terminus of ETV1, including the AD, is truncated due to chromosomal rearrangements[30,31]. This truncation has been shown to increase the stability of ETV1, as two of the main E3 ubiquitin ligase

recognition sites are located in the N-terminus and are lost due to the truncation[39,40]. Here we have demonstrated that even with the loss of the AD, ETV1/4/5 subfamily factors can still interact with MED25 through their DBD. The requirement of interaction with MED25 for ETV1/4/5 subfamily factors' role in prostate cancer remains to be established. Disruption of ELF3-MED23 interaction with a small molecule inhibitor effectively decreased the expression of the receptor tyrosine kinase ERBB2 and selectively killed breast cancer cell lines[19,41]. Therefore, perturbation of transcription factor–mediator interfaces are tractable potential therapeutic targets[42]. Future studies will establish whether the ETV1/4/5 – MED25 interface could be a molecular target with therapeutic potential.

Methods

Protein expression and purification

Full-length ETS factors and and truncated ETV4 fragments were cloned into the pET28 (Novagen) bacterial expression vector using sequence- and ligation-independent cloning (SLIC)[43]. MED25 ACID (amino acids 391-553) was cloned into a vector with N-terminal GST, avitag, and HIS_6 tags for biolayer interferometry and pET28 for NMR spectroscopy.

All proteins were expressed in (λDE3) *Escherichia coli* cells. MED25 ACID and ETV4 1-164, 43-84, and 337-436 expressed into the soluble fraction, and were grown in 1L cultures of Luria broth (LB) 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 hours. For MED25 ACID protein used in BLI, 1mL of 50mM biotin was

added at the induction point. Cells were centrifuged at 6,000 RPM in a JLA 8.1 rotor (Beckmann), resuspended in a buffer containing 25 mM Tris, pH 7.9, 200mM NaCl, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 2 mM 2-mercaptoethanol (BME), and 1 mM phenylmethanesulfonylfluoride (PMSF), and snap-frozen in liquid nitrogen. After 3-5 freeze-thaw cycles, cells were lysed by sonication and ultracentrifuged at 40k RPM in a Beckman Ti45 rotor. The soluble fraction was then loaded onto a Ni^{2+} affinity column (GE) and eluted over 20 column volumes of a 5-500mM imidazole gradient. For MED25 ACID used in BLI, protein eluted from the Ni^{2+} column was loaded onto a GST affinity column and eluted with the same buffer with 15mM glutathione. Elutions were then dialyzed overnight into 25 mM Tris pH 7.9, 10% glycerol (v:v), 1 mM EDTA, 50 mM KCl and 1 mM dithiothreitol (DTT). After ultracentrifugation as previously described, proteins were purified over a SP-sepharose cation exchange column (GE) (ETV4 337-436 and MED25 391-553) or Q-sepharose anion exchange column (GE) (ETV4 1-164 and 43-84) using a 50 – 1000 mM linear gradient of KCl. Proteins were then further purified over 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. Purified proteins were concentrated on a 10-kDa molecular weight cutoff (MWCO) centricon device, snap-frozen in liquid nitrogen, and stored in single-use aliquots at -80° C for subsequent NMR or biolayer interferometry studies.

Full-length ETV4 and ETV4 165-436, 165-336, and 165-484 expressed into the insoluble fraction using an autoinduction protocol[44]. Briefly, bacteria in

250 mL of autoinduction media were grown in 4 L flasks at 37 °C to an OD_{600} ~ 0.6 – 1. The temperature was then reduced to 30 °C and cultures were grown for another ~ 12 – 24 hours. Final OD_{600} values were typically ~ 6 – 12, indicating robust autoinduction. Harvested cells were resuspended as described above, sonicated and centrifuged at 15k rpm in a JA-17 rotor (Beckman) for 15 minutes 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 hour at 4 °C, the sample was centrifuged for 40k rpm for at least 30 minutes 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. A Q-sepharose anion-exchange column was used due to differing isoelectric points of the desired proteins.

Bio-layer interferometry

Data were collected using an Octet Red96 instrument (ForteBio) and processed with the intrument's software. 500 nM of biotinylated MED25 protein was immobilized using high-precision streptavidin sensors (ForteBio) for ~ 100 seconds to get a response of ~ 1.5 nm. Interaction experiments were conducted

using 150 mM NaCl, 25 mM Tris pH 7.9, 1 mM Tris(2-carboxyethl)phosphine (TCEP)(Sigma), 5 ug/mL bovine serum albumin (Sigma) and 0.5% (v:v) Tween20 (Sigma). Biosensors were dipped in various concentrations of the analyte of interest to measure association, and transferred back to buffer wells for monitoring dissociation. For quantitative analysis, six titration points of ETV4, with exact concentrations varying dependent on the affinity of the interaction between that ETV4 truncation/mutant and MED25, were fit using a global (full) analysis. Kinetic constants were determined from the mathematical fit of a 1:1 binding model. Mean and standard deviation of K_D, k_a, and k_d values from at least three independent experimental replicates are displayed in figures and tables.

NMR spectroscopy

^{1}H-^{15}N HSQC measurements were recorded on a 500MHz Varian Inova spectrometer at 25°C in NMR buffer (20mM Sodium Phosphate, pH 6.5, 200mM NaCl, 2mM Dithiothreitol (DTT), 1mM EDTA acid, 10% D_2O). Assignments for MED25 ACID[17,45] and ETV4 337-436 (Chapter 3) were transferred from previous work and titration data processed and analyzed using Sparky[46] (UCSF).

Electrophoretic mobility shift assays (EMSAs)

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 consensus 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 [g-^{32}P] ATP using T4 polynucleotide kinase at 37 °C for ~ 30 – 60 minutes. After purification over a Bio-Spin 6 chromatography column (Bio-Rad), the oligonucleotides were incubated at 100 °C for ~ 5 minutes, and then cooled to room temperature over 1 – 2 hours. The DNA for EMSAs was diluted to 1 x 10^{-12} M and held constant, whereas ETV4 and ETS1 concentrations ranged from 1 x 10^{-7} to ~1 x 10^{-12}M. Protein concentrations were determined after thawing each aliquot of protein, using the Protein Assay Dye Reagent. Equivalent starting amounts (0.2 mg) 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 minutes 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 mg/mL BSA, 10 mM DTT, 100 ng/mL poly(dIdC), and 10% (v:v) glycerol, and then resolved on an 3.5 or 5% (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 $[P]_t$ 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.

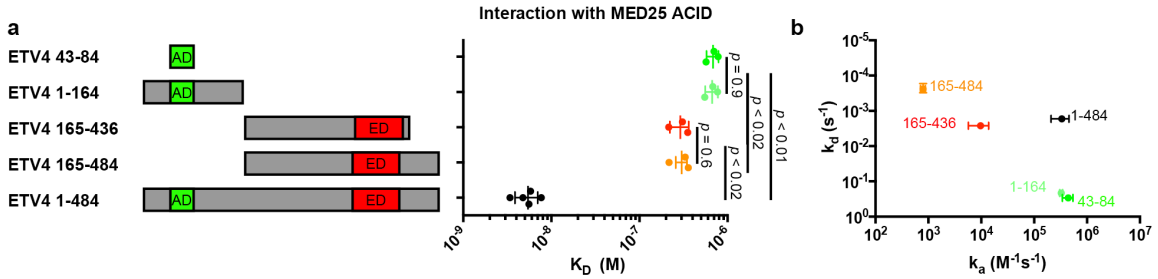


Figure 4.1 N- and C-terminal regions of ETV4 interact with MED25. (**a**) Equilibrium dissociation constants (K_D) of ETV4 truncations with MED25 as measured by biolayer interferometry. Values from three individual experiments are displayed as well as the mean and standard deviation. Two-tailed, heteroscedastic t-tests were used to calculate p-values between different fragments. The activation domain and ETS domain are abbreviated as AD and ED, respectively. (**b**) Comparison of association (k_a) and dissociation (k_d) rate constants for the interactions between different ETV4 truncations and MED25. The mean and standard deviation calculated from three individual experiments are displayed.

Table 4.1 Interaction of ETV4 truncations with MED25

ETV4	K_D (x 10^{-9} M)[a]	k_a (x 10^3 $M^{-1}s^{-1}$)[a]	k_d (x 10^{-3} s^{-1})[a]
43-84	680 ± 60	400 ± 100	290 ± 30
1-164	670 ± 60	330 ± 20	210 ± 20
165-436	290 ± 70	10 ± 4	2.7 ± 0.4
165-484	300 ± 80	0.80 ± 0.08	0.2 ± 0.1
1-484	6 ± 2	300 ± 100	1.7 ± 0.2

[a]Mean and standard deviation from at least three individual experiments.

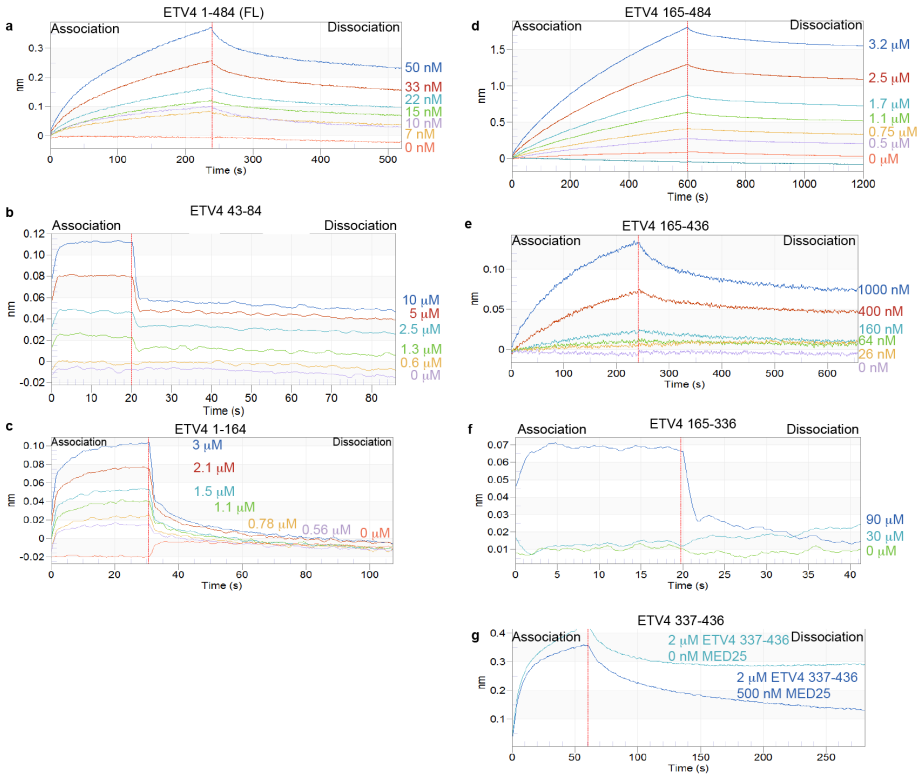


Figure 4.2 Representative example of each interaction between ETV4 truncations and MED25. Biolayer interferometry graphs for the interactions between MED25 and ETV4 1-484 (full length) (**a**), 165-484 (**b**), 165-436 (**c**), 43-84 (N-terminal transactivation domain) (**d**), 1-164 (**e**), 165-336 (**f**), and 337-436 (**g**). The left side of each graph is the association between MED25 and ETV4, and the right side is the dissociation. Concentrations of ETV4 truncations are displayed on the right side of each graph. The interaction affinity could not be measured for ETV4 337-436 (g) as there was too much nonspecific interaction between this truncation and the streptavidin biolayer tip.

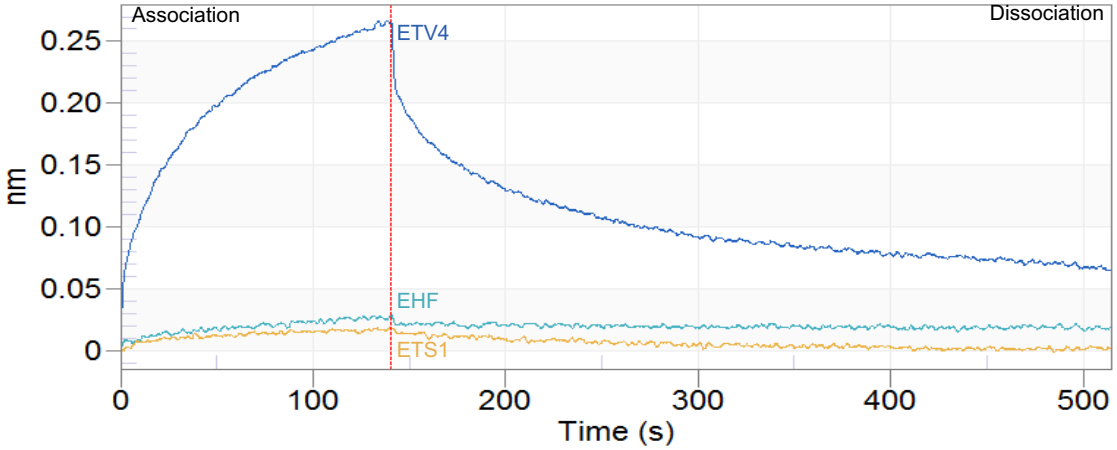


Figure 4.3 MED25 does not interact with ETS1 or EHF. Biolayer interferometry graph showing the interaction between a single concentration (4 mM) of ETV4, EHF, or ETS1 with MED25. The lack of association between EHF or ETS1 with MED25 at this concentration suggests that these ETS factors either do not interact with MED25 or do so with an affinity that is substantially lower than that of ETV4.

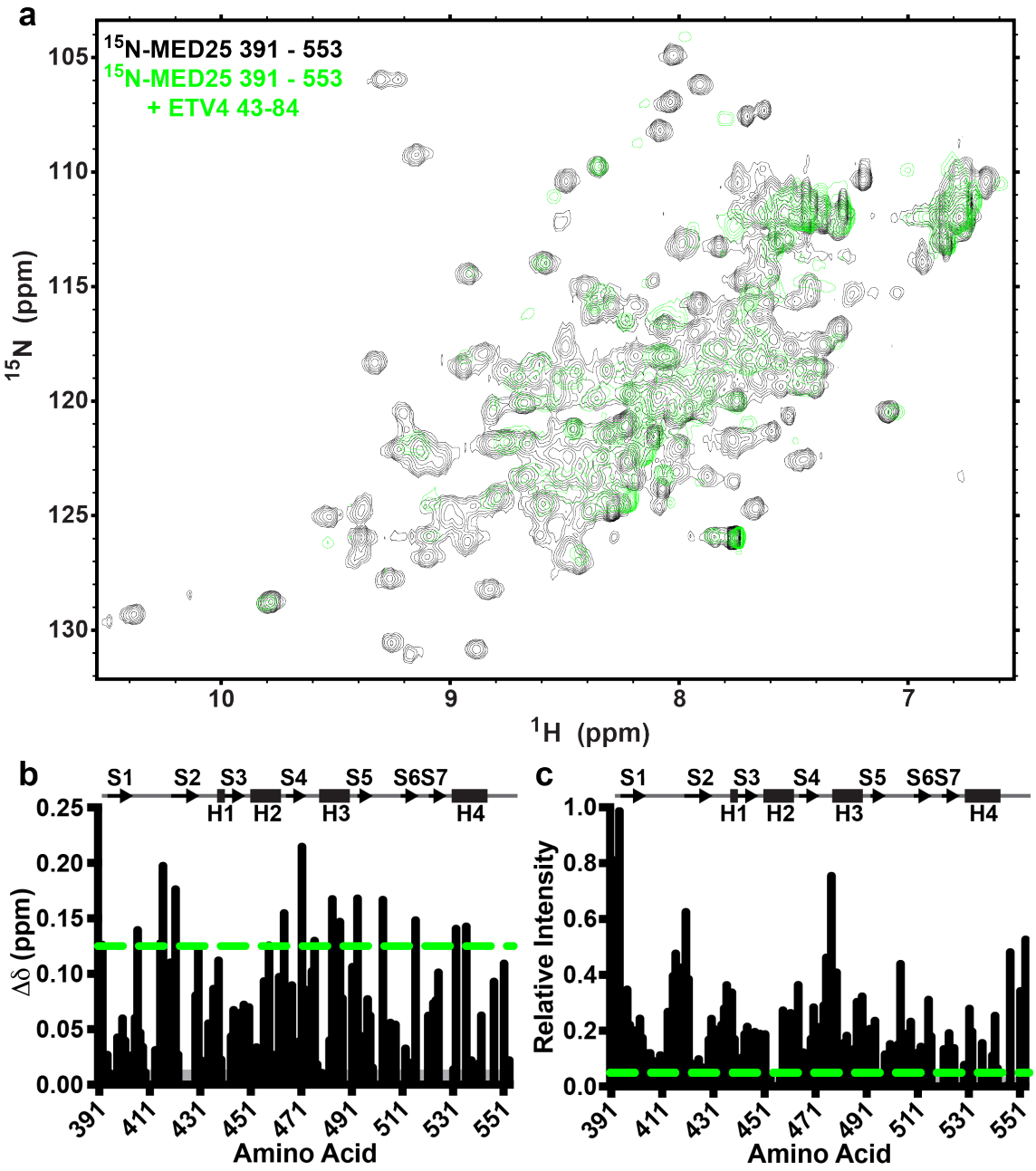


Figure 4.4 ETV4 43-84 perturbation of MED25 391-553. (**a**) Overlaid NMR spectra of ^{15}N-MED25 391-553 alone (black) and with 1.2 molar equivalents of unlabeled ETV4 43-84 (blue). (**b,c**) Dd (ppm) and relative intensity of NMR peaks comparing spectra of MED25 391-553 alone and with ETV4 43-84. Schematic of secondary structure of MED25; arrows and rectangles represent β-strands and α-helices, respectively.

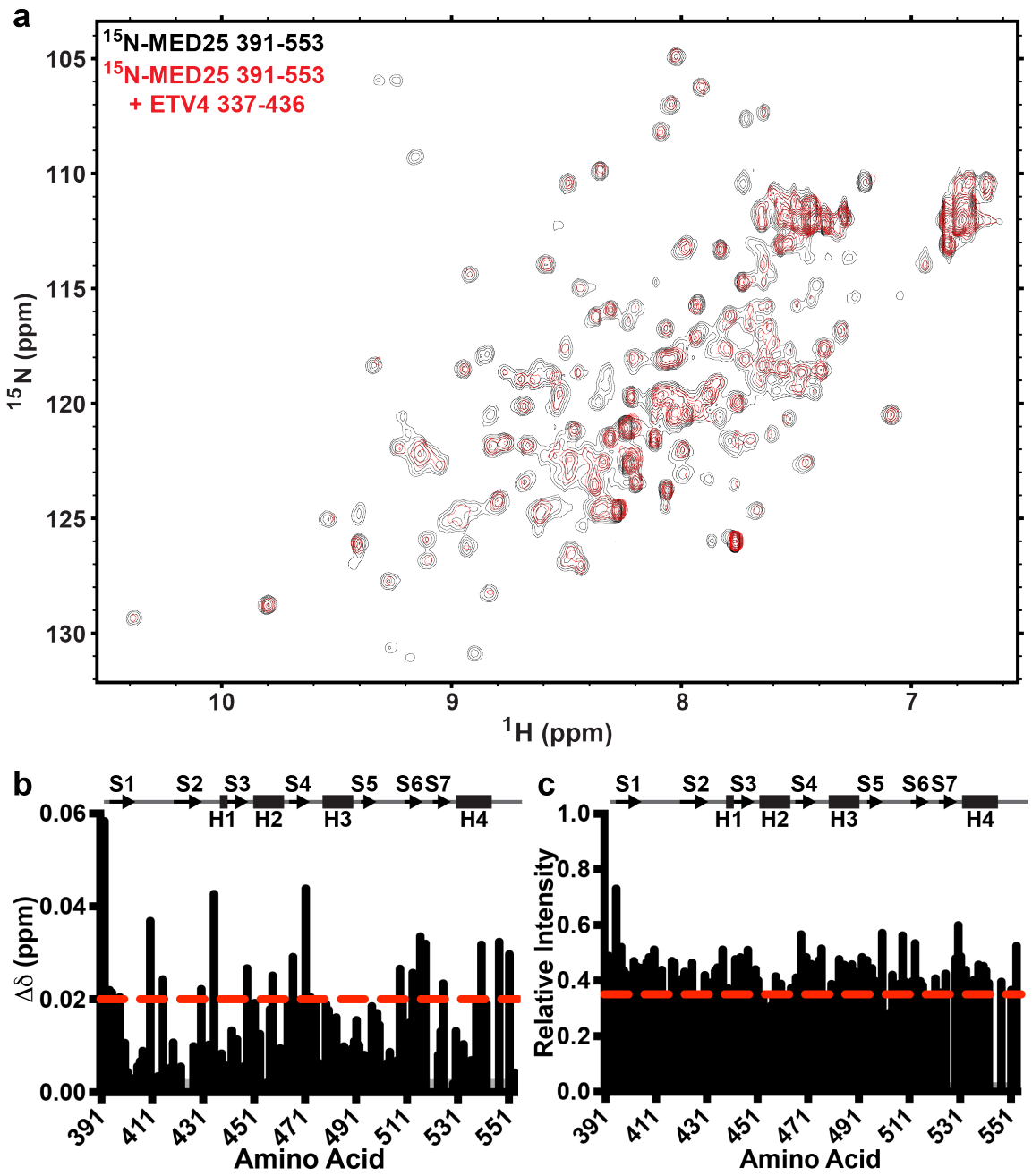


Figure 4.5 ETV4 337-436 perturbation of MED25 391-553. (**a**) Overlaid NMR spectra of ^{15}N-MED25 391-553 alone (black) and with 1.2 molar equivalents of unlabeled ETV4 337-436 (red). (**b,c**) Dd (ppm) and relative intensity of NMR peaks comparing spectra of MED25 391-553 alone and with ETV4 43-84. Schematic of secondary structure of MED25; arrows and rectangles represent β-strands and α-helices, respectively.

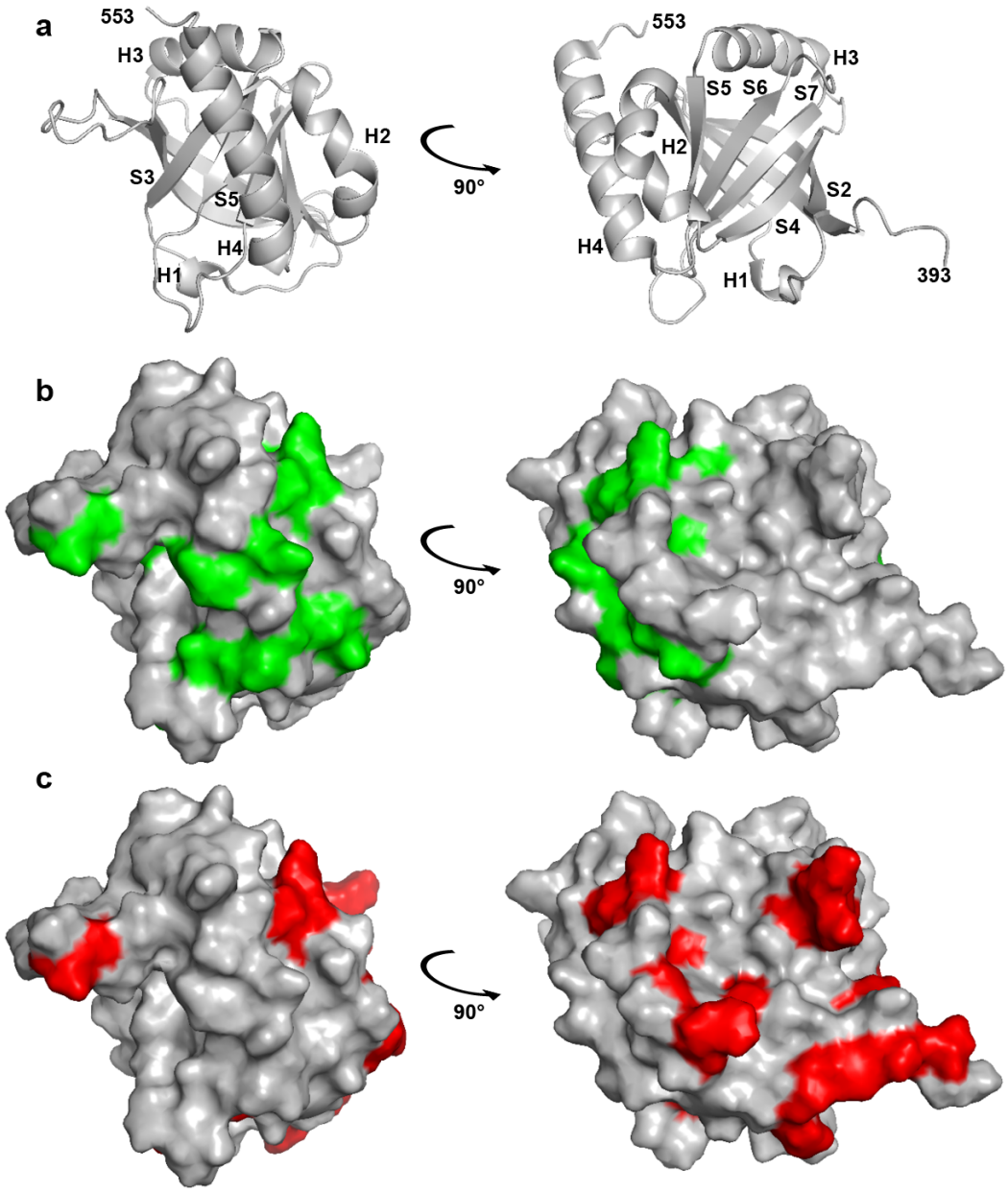


Figure 4.6 ETV4 43-84 and 337-436 perturb distinct, but partially overlapping interfaces on MED25 391-553. (**a**) Cartoon depiction of MED25 ACID with two different views. (**b**) Spectral changes upon addition of ETV4 43-84, Dd > 0.125 ppm or relative peak intensity < 0.05 (Fig 2b,c), are mapped onto the structure of MED25 391-553 (blue). MED25 391-553 oriented as in (a), but shown in surface format. (**c**) Spectral changes upon addition of ETV4 337-436, Dd > 0.02 ppm or relative peak intensity < 0.35 (Fig 3b,c), are mapped onto the structure of MED25 391-553 (red). MED25 ACID oriented as in (a), but shown in surface format.

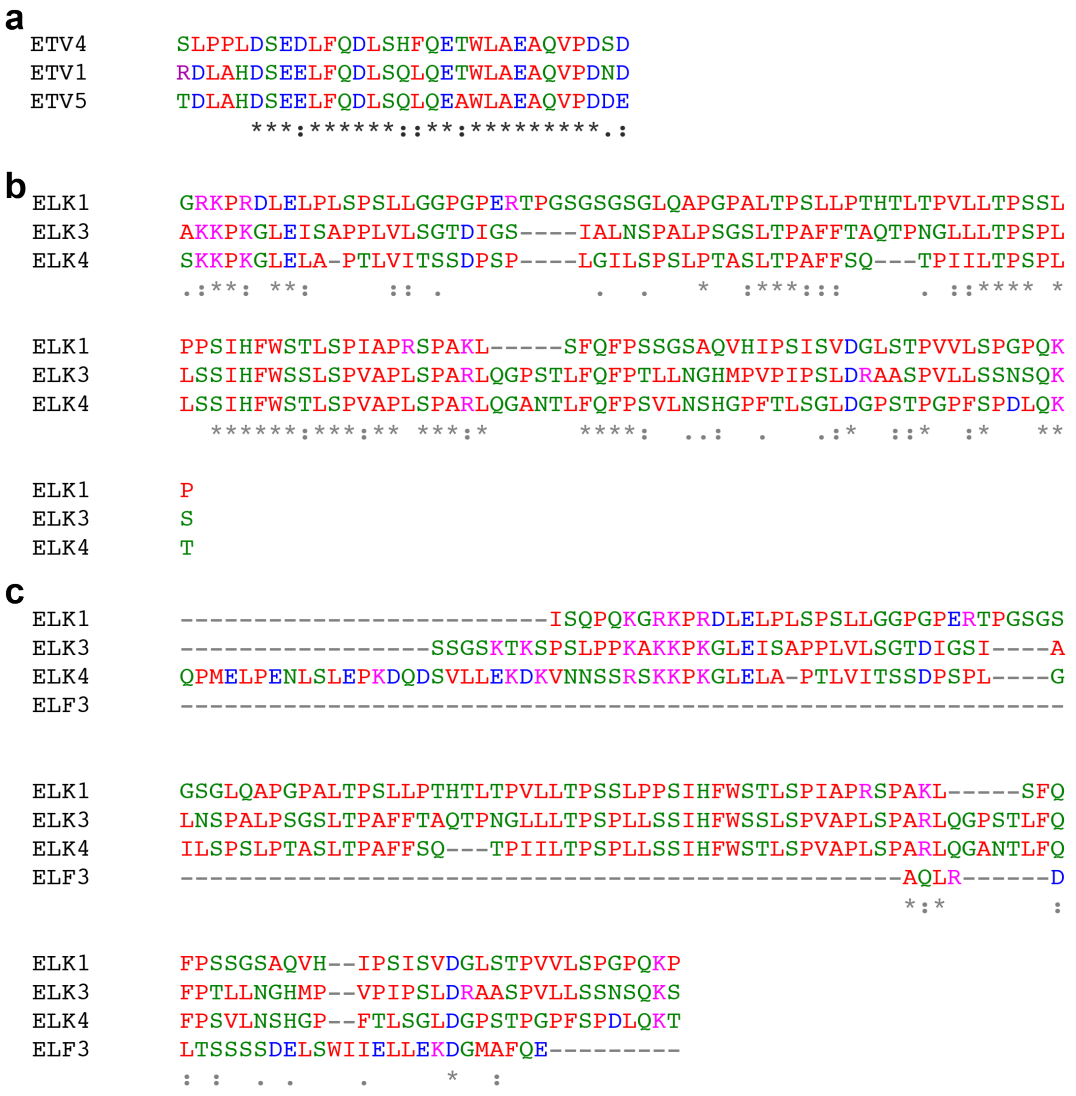
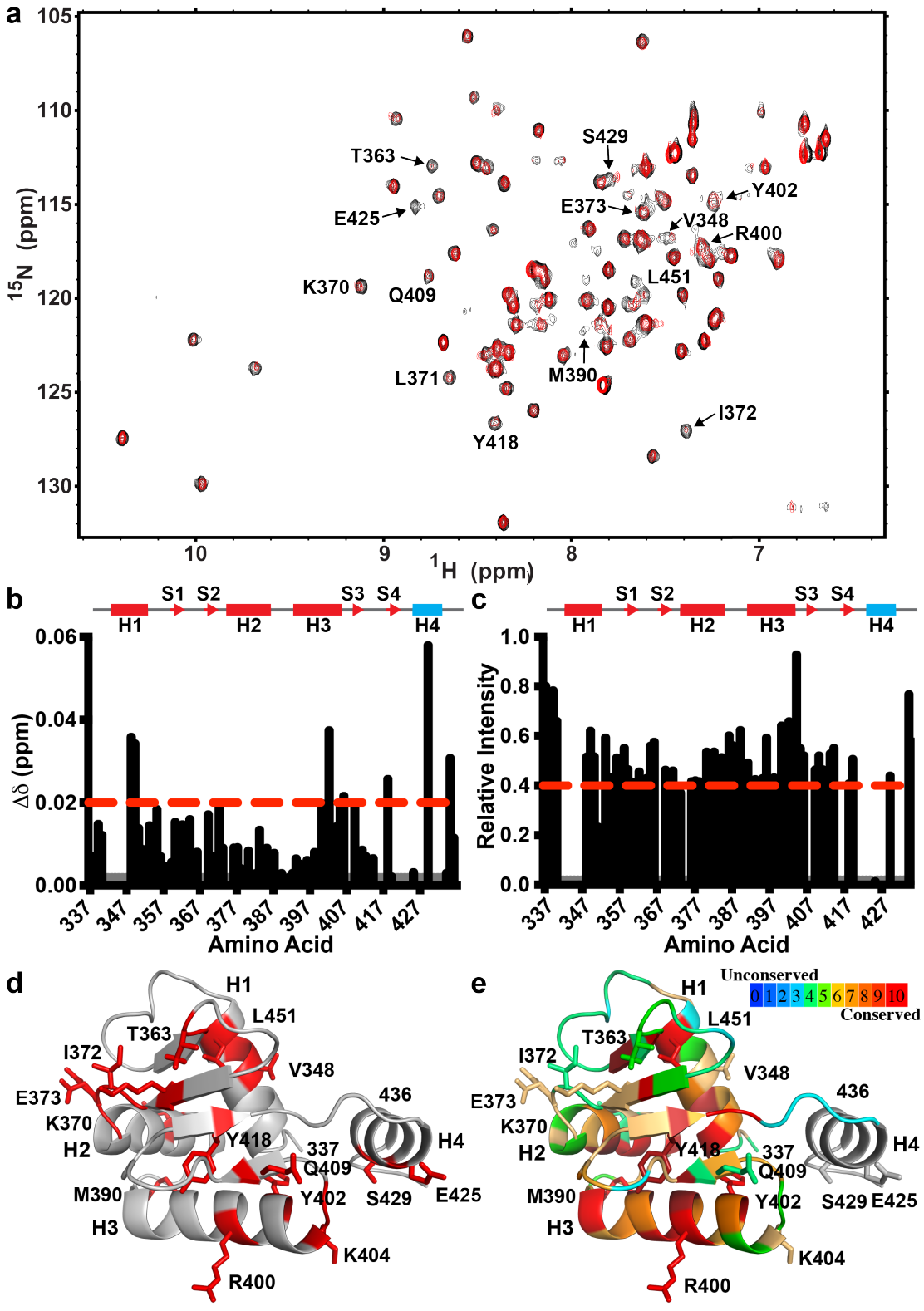


Figure 4.7 Sequence alignments for activation domains of select ETS factors. (**a**) Alignment of the N-terminal AD from ETV1, ETV4, and ETV5. (**b**) Alignment of the AD from ELK1, ELK3, and ELK4. (**c**) Alignment as in (b), but with the AD from ELF3 also aligned. ELK1, ELK3, ELK4, and ELF3 are all capable of interacting with MED23.

Figure 4.8 MED25 interacts with a divergent interface on the DBD of ETV4. (**a**) Overlaid spectra of 15N-ETV4 ED, amino acids 337-436, alone (black), and with 1.2 molar equivalents of unlabeled MED25 ACID (red). (**b,c**) Dd (ppm) and relative intensity of NMR peaks comparing spectra of ETV4 337-436 alone and with MED25 391-553. Schematic above histograms refers to the secondary structure of ETV4 DBD. Rectangles and arrows are α-helices and β-strands, red is ETS domain, cyan is divergent element specific to ETV1/4/5 subfamily. (**d**) Spectral changes upon addition of MED25 391-553; Dd > 0.02 ppm or relative peak intensity < 0.4, are mapped onto the structure of ETV4 337-436 in cartoon format. Perturbed residues are indicated by red coloring and side-chains are displayed in stick format. (**e**) Structure of ETV4 337-436 as in (d), but coloring refers to conservation of amino acids in the ETS domain. H4 is colored gray because is not structurally conserved in ETS proteins outside of the ETV1/4/5 subfamily. See Figure 4.9 for further details on sequence conservation.


a
105
110
115
120
125
130
15N (ppm)
10
9
8
7
1H (ppm)
T363
S429
Y402
E373
E425
V348
R400
L451
K370
Q409
L371
M390
I372
Y418
b
S1 S2 S3 S4
H1 H2 H3 H4
0.06
0.04
0.02
0.00
Δδ (ppm)
337 347 357 367 377 387 397 407 417 427
Amino Acid
c
S1 S2 S3 S4
H1 H2 H3 H4
1.0
0.8
0.6
0.4
0.2
0.0
Relative Intensity
337 347 357 367 377 387 397 407 417 427
Amino Acid
d
H1
L451
T363
I372
V348
E373
436
K370
H2
Y418
337
H4
Q409
M390
Y402
S429
E425
H3
K404
R400
e
Unconserved
0 1 2 3 4 5 6 7 8 9 10
Conserved
H1
L451
T363
I372
V348
E373
436
K370
H2
Y418
337
H4
Q409
M390
Y402
S429
E425
H3
K404
R400

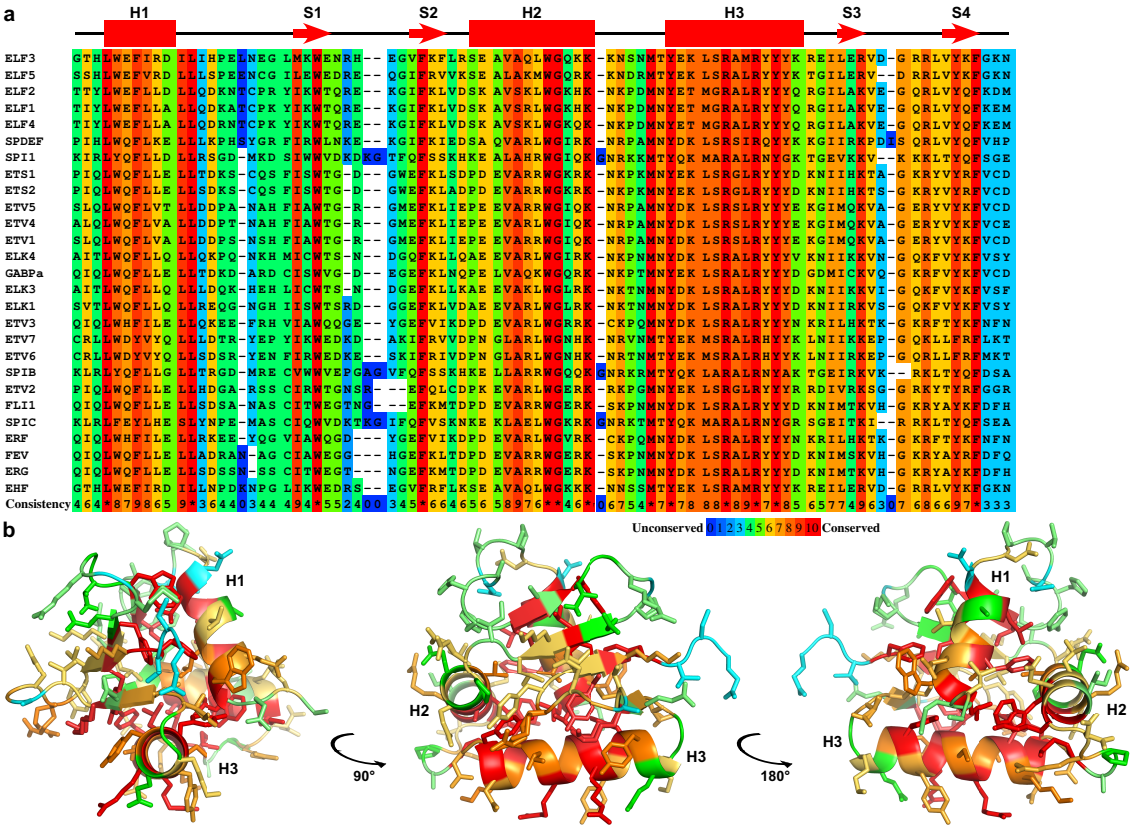


Figure 4.9 Conservation of ETS domain. (**a**) Sequence alignment of ETS domain using Praline multiple sequence alignment tool[47]. Schematic above sequence alignment represents secondary structure elements of the ETS domain; rectangles and arrows represent α-helices and β-strands, respectively. (**b**) ETS domain from crystal structure of ETV4 shown in three different views and colored according to (a). Note that the highest conservation occurs in α-helices as well as core-facing amino acids and amino acids that are critical for DNA-binding (H3). In contrast, loops and the β-sheet are relatively less conserved.

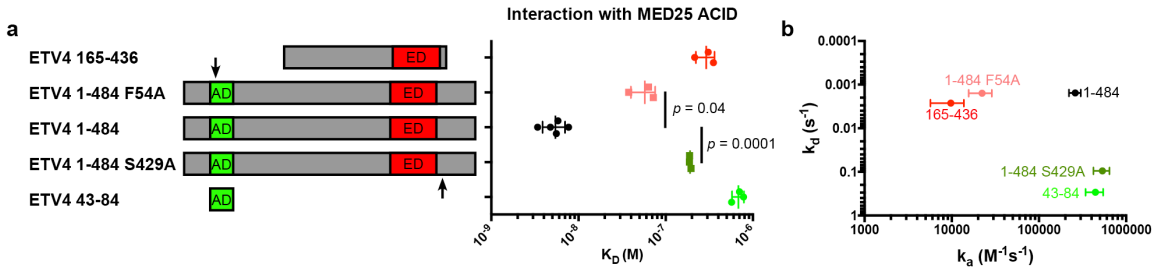


Figure 4.10 ETV4 point mutations ablate interaction with MED25. (**a**) Equilibrium dissociation constants (K_D) of ETV4 point mutants F54A and S429A with MED25 as measured by biolayer interferometry. Values from three individual experiments are displayed as well as the mean and standard deviation. Two-tailed, heteroscedastic t-tests were used to calculate p-values between different fragments. The activation domain and ETS domain are abbreviated as AD and ED, respectively. (**b**) Comparison of association (k_a) and dissociation (k_d) rate constants for the interactions between ETV4 point mutants and MED25. The mean and standard deviation calculated from three individual experiments are displayed. Values for ETV4 1-484, 43-84, and 165-436 (Fig. 4.1) are included for reference.

Table 4.2 Interaction of ETV4 point mutants with MED25

ETV4	K_D (x 10^{-9} M)[a]	k_a (x 10^{3} $M^{-1}s^{-1}$)[a]	k_d (x 10^{-3} s^{-1})[a]
1-484	6 ± 1	300 ± 100	1.7 ± 0.2
1-484 F54A	60 ± 10	22 ± 7	1.2 ± 0.3
1-484 S249A	190 ± 35	500 ± 100	90 ± 10

[a]Mean and standard deviation from at least three individual experiments.

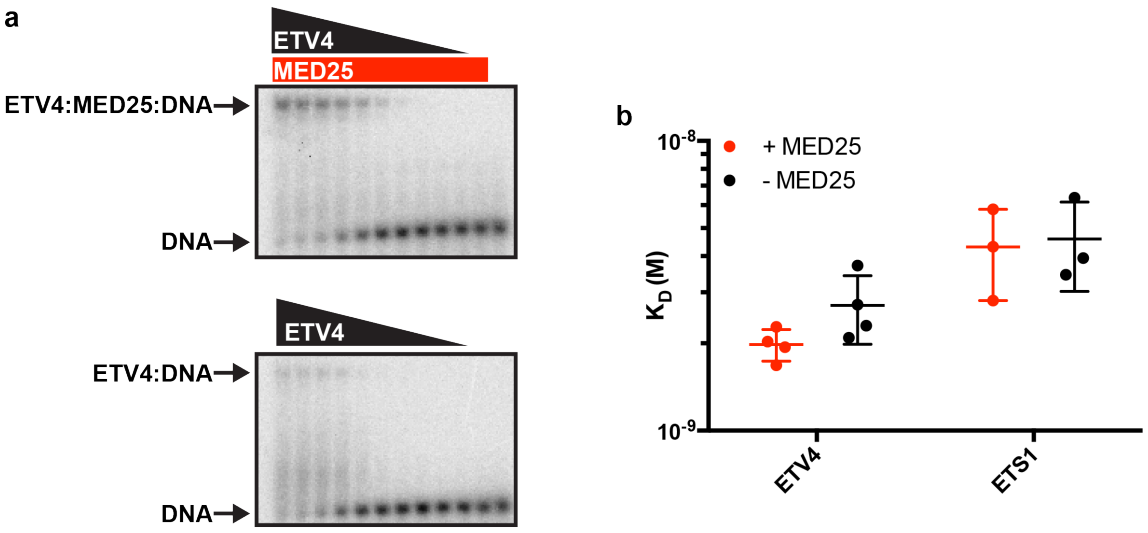


Figure 4.11 Interaction with MED25 activates the DNA binding of ETV4. (**a**) EMSAs with titrations of ETV4 with (top) or without (bottom) MED25 ACID domain. (**b**) K_D values of ETV4 and ETS1 for DNA with (red) or without (black) MED25. Each dot represents an individual replicate and lines refer to mean and standard deviation.

References

1. Plaschka, C., Nozawa, K. & Cramer, P. Mediator architecture and RNA polymerase II interaction. *J Mol Biol* **428**, 2569-74 (2016).

2. Allen, B.L. & Taatjes, D.J. The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* **16**, 155-66 (2015).

3. Cevher, M.A. et al. Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nat Struct Mol Biol* **21**, 1028-34 (2014).

4. Ge, K. et al. Transcription coactivator TRAP220 is required for PPAR gamma 2-stimulated adipogenesis. *Nature* **417**, 563-7 (2002).

5. Zhang, X. et al. MED1/TRAP220 exists predominantly in a TRAP/ Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription. *Mol Cell* **19**, 89-100 (2005).

6. Kim, T.W. et al. MED16 and MED23 of Mediator are coactivators of lipopolysaccharide- and heat-shock-induced transcriptional activators. *Proc Natl Acad Sci U S A* **101**, 12153-8 (2004).

7. Stevens, J.L. et al. Transcription control by E1A and MAP kinase pathway via Sur2 mediator subunit. *Science* **296**, 755-8 (2002).

8. van Essen, D., Engist, B., Natoli, G. & Saccani, S. Two modes of transcriptional activation at native promoters by NF-kappaB p65. *PLoS Biol* **7**, e73 (2009).

9. Balamotis, M.A. et al. Complexity in transcription control at the activation domain-mediator interface. *Sci Signal* **2**, ra20 (2009).

10. Rana, R., Surapureddi, S., Kam, W., Ferguson, S. & Goldstein, J.A. Med25 is required for RNA polymerase II recruitment to specific promoters, thus regulating xenobiotic and lipid metabolism in human liver. *Mol Cell Biol* **31**, 466-81 (2011).

11. Chen, W., Rogatsky, I. & Garabedian, M.J. MED14 and MED1 differentially regulate target-specific gene activation by the glucocorticoid receptor. *Mol Endocrinol* **20**, 560-72 (2006).

12. Park, J.M. et al. In vivo requirement of activator-specific binding targets of mediator. *Mol Cell Biol* **20**, 8709-19 (2000).

13. Boyer, T.G., Martin, M.E., Lees, E., Ricciardi, R.P. & Berk, A.J. Mammalian Srb/Mediator complex is targeted by adenovirus E1A protein. *Nature* **399**, 276-9 (1999).

14. Ding, N. et al. MED19 and MED26 are synergistic functional targets of the RE1 silencing transcription factor in epigenetic silencing of neuronal gene expression. *J Biol Chem* **284**, 2648-56 (2009).

15. Ding, N. et al. Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. *Mol Cell* **31**, 347-59 (2008).

16. Griffiths, S.J. et al. A systematic analysis of host factors reveals a Med23-interferon-lambda regulatory axis against herpes simplex virus type 1 replication. *PLoS Pathog* **9**, e1003514 (2013).

17. Milbradt, A.G. et al. Structure of the VP16 transactivator target in the Mediator. *Nat Struct Mol Biol* **18**, 410-5 (2011).

18. Vojnic, E. et al. Structure and VP16 binding of the Mediator Med25 activator interaction domain. *Nat Struct Mol Biol* **18**, 404-9 (2011).

19. Asada, S. et al. External control of Her2 expression and cancer cell growth by targeting a Ras-linked coactivator. *Proc Natl Acad Sci U S A* **99**, 12747-52 (2002).

20. Landrieu, I. et al. Characterization of ERM transactivation domain binding to the ACID/PTOV domain of the Mediator subunit MED25. *Nucleic Acids Res* **43**, 7110-21 (2015).

21. Warfield, L., Tuttle, L.M., Pacheco, D., Klevit, R.E. & Hahn, S. A sequence-specific transcription activator motif and powerful synthetic variants that bind Mediator using a fuzzy protein interface. *Proc Natl Acad Sci U S A* **111**, E3506-13 (2014).

22. Gordon, D.F. et al. MED220/thyroid receptor-associated protein 220 functions as a transcriptional coactivator with Pit-1 and GATA-2 on the thyrotropin-beta promoter in thyrotropes. *Mol Endocrinol* **20**, 1073-89 (2006).

23. Mo, X., Kowenz-Leutz, E., Xu, H. & Leutz, A. Ras induces mediator complex exchange on C/EBP beta. *Mol Cell* **13**, 241-50 (2004).

24. Helgeson, B.E. et al. Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer. *Cancer Res* **68**, 73-80 (2008).

25. Tomlins, S.A. et al. TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. *Cancer Res* **66**, 3396-400 (2006).

26. Tomlins, S.A. et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-8 (2005).

27. Aytes, A. et al. ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. *Proc Natl Acad Sci U S A* **110**, E3506-15 (2013).

28. Baena, E. et al. ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. *Genes Dev* **27**, 683-98 (2013).

29. Verger, A. et al. The Mediator complex subunit MED25 is targeted by the N-terminal transactivation domain of the PEA3 group members. *Nucleic Acids Res* **41**, 4847-59 (2013).

30. Clark, J.P. & Cooper, C.S. ETS gene fusions in prostate cancer. *Nat Rev Urol* **6**, 429-39 (2009).

31. Kumar-Sinha, C., Tomlins, S.A. & Chinnaiyan, A.M. Recurrent gene fusions in prostate cancer. *Nat Rev Cancer* **8**, 497-511 (2008).

32. Cooper, C.D., Newman, J.A., Aitkenhead, H., Allerston, C.K. & Gileadi, O. 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. *J Biol Chem* **290**, 13692-709 (2015).

33. Galbraith, M.D. et al. ERK phosphorylation of MED14 in promoter complexes during mitogen-induced gene activation by Elk-1. *Nucleic Acids Res* **41**, 10241-53 (2013).

34. Ito, M. et al. Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol Cell* **3**, 361-70 (1999).

35. Hittelman, A.B., Burakov, D., Iniguez-Lluhi, J.A., Freedman, L.P. & Garabedian, M.J. Differential regulation of glucocorticoid receptor transcriptional activation via AF-1-associated proteins. *EMBO J* **18**, 5380-8 (1999).

36. Zhou, R. et al. SOX9 interacts with a component of the human thyroid hormone receptor-associated protein complex. *Nucleic Acids Res* **30**, 3245-52 (2002).

37. Nakamura, Y. et al. Wwp2 is essential for palatogenesis mediated by the interaction between Sox9 and mediator subunit 25. *Nat Commun* **2**, 251 (2011).

38. Sela, D. et al. Role for human mediator subunit MED25 in recruitment of mediator to promoters by endoplasmic reticulum stress-responsive transcription factor ATF6alpha. *J Biol Chem* **288**, 26179-87 (2013).

39. Vitari, A.C. et al. COP1 is a tumour suppressor that causes degradation of ETS transcription factors. *Nature* **474**, 403-6 (2011).

40. Baert, J.L. et al. The E3 ubiquitin ligase complex component COP1 regulates PEA3 group member stability and transcriptional activity. *Oncogene* **29**, 1810-20 (2010).

41. Shimogawa, H. et al. A wrench-shaped synthetic molecule that modulates a transcription factor-coactivator interaction. *J Am Chem Soc* **126**, 3461-71 (2004).

42. Phillips, A.J. & Taatjes, D.J. Small molecule probes to target the human Mediator complex. *Isr J Chem* **53**, 588-595 (2013).

43. Li, M.Z. & Elledge, S.J. SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol Biol* **852**, 51-9 (2012).

44. Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207-34 (2005).

45. Eletsky, A. et al. Solution NMR structure of MED25(391-543) comprising the activator-interacting domain (ACID) of human mediator subunit 25. *J Struct Funct Genomics* **12**, 159-66 (2011).

46. Goddard, T.D. & Kneller, D.G. Sparky, 3rd ed. (1999).

47. Simossis, V.A. & Heringa, J. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res* **33**, W289-94 (2005).

CHAPTER 5

ETS AND AP1 FACTOR INTERACTIONS AT COMPOSITE DNA BINDING SITES PROVIDE A BASIS FOR ALTERNATIVE ROLES IN PROSTATE CANCER

This research is in preparation for submission to a peer-reviewed journal. Simon L Currie, Bethany Madison, Kathleen A Clark, Peter C Hollenhorst and Barbara J Graves. ETS and AP1 factor interactions at composite DNA binding sites provide a basis for alternative roles in prostate cancer.

Abstract

ETS transcription factors from the ERG and ETV1/4/5 subfamilies are overexpressed in the majority of prostate cancers and drive cancer progression, in part, by activating the expression of cell migration genes that are regulated by composite ETS and AP1 DNA binding sites. In contrast, EHF and SPDEF, ETS factors that are normally expressed in the prostate, have been characterized as tumor suppressors of prostate cancer and transcriptionally repress ETS-AP1 regulated loci. Here we demonstrate that JUN-FOS displays negative cooperativity with EHF and SPDEF at closely spaced ETS-AP1 DNA-binding sites. In contrast, ERG and ETV4 are capable of binding to DNA with JUN-FOS without mutual interaction or with positive cooperativity, respectively. Furthermore, ERG binds to the nonoptimal ETS sequence characteristic of ETS-AP1 composite sites with ~14-fold lower affinity than a consensus high affinity ETS sequence, whereas EHF affinity differs by a more modest ~3-fold. We develop a model of competing ETS factors with EHF binding to ETS-AP1 composite sites in isolation whereas ERG binds to these sites with JUN-FOS. These findings provide evidence for the differential transcriptional regulation observed among ETS factors at ETS-AP1 composite sites.

Introduction

Multiple transcription factor (TF) binding sites in enhancers and promoters allow for the combinatorial control of gene transcription by integrating multiple inputs, such as signaling pathways, into a single transcriptional output[1,2]. TFs are

classified into families based on DNA-binding domain structure[3]. TFs from an individual family recognize similar DNA sequences[4,5] and have overlapping expression profiles[6], often resulting in a competition between multiple TFs for the same binding site at many genomic loci[7]. A neighboring, distinct TF site can influence this competition for DNA-binding by selecting for, or against, one of the competing TFs through cooperative or antagonistic DNA-binding, respectively.

The ETS (E26-transformation specific) family of transcription factors are defined by their DNA-binding domain, a winged helix-turn-helix structure termed the ETS domain[8,9]. Transcription factor partnerships with other transcription factors have been described for several ETS factors[10-15], and these partnerships are specific for an individual ETS factor or subfamily of ETS factors. Therefore, composite DNA sites that contain an ETS site in combination with another TF binding site are distinctly regulated by a single or subfamily of ETS factors[7,16].

ETS transcription factors from the ERG and ETV1/4/5 subfamilies are overexpressed in the majority of prostate cancers[17] and contribute to the development of prostate cancer in mouse models of the disease[18-20]. In contrast, the ETS factors EHF and SPDEF, which are expressed in normal prostate tissue, have been characterized as tumor suppressors in the context of prostate cancer[21,22]. Evidence suggests that this phenotypic dichotomy of ETS factors in the prostate may derive, in part, from differential regulation of composite ETS and activator protein 1 (AP1) DNA-binding sites. Genome-wide analyses demonstrated that ERG and ETV1/4/5 factors bind to ETS-AP1 DNA-binding sites and activate the transcription of ETS-AP1 regulated genes that are

important for processes such as cell migration[23]. In contrast, EHF and SPDEF have been shown to transcriptionally repress particular ETS–AP1 regulated genes[23,24]. ERG and ETV1/4/5 factors have been shown to cooperatively bind to DNA with JUN-FOS[25], suggesting that differences in cooperative DNA-binding may distinguish between ETS factors at ETS-AP1 sites. It is currently unknown how EHF or SPDEF affect JUN-FOS activity at ETS-AP1 tandem DNA sites.

Here we demonstrate a differential ability of ETS factors to bind to DNA with JUN-FOS. Using the previously characterized uridine phosphorylase gene promoter (UPP), we find that ETS factors can either display negative cooperativity (EHF and SPDEF), or positive cooperativity (ERG and FLI1) with JUN-FOS on ETS-AP1 composite DNA sites. In addition, ETV1 and ETV4 bind together with JUN-FOS and have no influence on the affinity of DNA-binding. Interestingly, we find that EHF has a stronger affinity for the nonoptimal –AGGAA- ETS motif that occurs within ETS-AP1 motifs than ERG. These differences in affinity map to the ETS domain and are likely due to sequence divergence of amino acids that contact DNA. Furthermore, we discover that the features of EHF required for DNA-binding antagonism with JUN-FOS reside outside of the ETS domain. In summary, this study establishes the differences in binding to ETS-AP1 consensus sites and in cooperation with JUN-FOS between distinct ETS factors and provides a biochemical basis for alternative functionality of ETS factors in prostate cancer.

Results

JUN-FOS differentially affects ETS factor binding to composite ETS-AP1 sites

The promoter for the uridine phosphorylase gene (UPP) contains a composite ETS-AP1 DNA motif. We used electrophoretic mobility shift assays (EMSAs) to determine the equilibrium dissociation constants (K_D) of ETS factors for the UPP promoter in the presence or absence of saturating JUN-FOS. We titrated the ETS factors EHF, SPDEF, ETV1, ETV4, ERG, and FLI1 with the UPP promoter alone, and with JUN-FOS bound to DNA (**Fig. 5.1**). EHF and SPDEF had relatively tighter binding (lower K_D values) for the UPP DNA compared to the other ETS factors, with values varying up to ~10-fold between ETS factors. Additionally, JUN-FOS strongly deterred the binding of EHF (~30-fold) and SPDEF (~20-fold) (**Figs. 5.1** and **5.2**). In comparison, ETV1 and ETV4 bound with about the same affinity with or without JUN-FOS, and ERG (~5-fold) and FLI1 (~20-fold) bound more strongly with JUN-FOS, as previously suggested[25]. These results demonstrate that ETS factors respond differently to the presence of JUN-FOS in a composite ETS-AP1 DNA site with ERG and FLI1, ETV1 and ETV4, and EHF and SPDEF displaying cooperative, noncooperative, and anti-cooperative DNA-binding with JUN-FOS, respectively.

DNA determinants of ETS-AP1 DNA binding

We were surprised by the substantial difference in the affinity of ETS factors for the UPP DNA in the absence of JUN-FOS, so we next probed the

basis of this variance. The consensus high-affinity DNA site for all of these ETS factors is -CCGGAAGT-, with the exception that SPDEF shows an equal preference for A and T at the sixth nucleotide (-CCGGAAGT- and –CCGGATGT-)[4,26]. There is only a single C to A difference between the consensus high-affinity ETS site and the ETS site observed in composite ETS-AP1 sites (-C*A*GGAAGT-), and this change occurs outside of the -GGAA- core of the ETS binding site[23]. While EHF and SPDEF belong to a different DNA-binding subclass of ETS factors than the other ETS factors tested here, previous findings suggest that all of these ETS factors have a similar preference for C over A at this position[4]. EHF and ERG both bound to consensus high-affinity ETS sites (-CCGGAAGT-) with similar K_D values of 1.0 ± 0.1 and 0.7 ± 0.2 nM, respectively (**Fig. 5.3**). Incorporating the single C to A change (-C*A*GGAAGT-) into the ETS site resulted in decreased affinity of all ETS factors for the DNA. However, the difference was only ~ 3-fold for EHF (1.0 ± 0.1 versus 2.9 ± 0.5 nM), whereas it was ~ 14-fold for ERG (0.7 ± 0.2 versus 10 ± 5 nM). Furthermore, both EHF and ERG bound to an ETS-AP1 composite site with similar affinity as the high affinity ETS site with the C to A change. Therefore, while the single C to A change decreases affinity for all ETS factors tested here, it has a stronger effect on ERG. Therefore, in the absence of JUN-FOS, EHF binds with higher affinity to ETS-AP1 sites than ERG.

Although all ETS-AP1 composite sites have a nonoptimal ETS-binding sequence, ETS-AP1 composite sites display variable spacing between the ETS and AP1 binding sequences[23]. Following previous nomenclature, the most common spacing is +6 base-pairs (bp), which corresponds to –

AGGAAGTGAXTCA- with the core ETS and AP1 binding sequences underlined, and X referring to any nucleotide (**Fig. 5.4**). However, other spacing is frequently observed for ETS-AP1 composite sites. This is in contrast to ETS-RUNX composite DNA sites that have a singular dominant spacing between ETS and RUNX binding sequences[16].

We next investigated how the variable spacing of ETS and AP1 sequences influences the binding of JUN-FOS and ETS factors to these composite sites. With 6 bp spacing between ETS and AP1 DNA binding sites, we again observed that JUN-FOS strongly deters the binding of EHF (**Fig. 5.5a-b**). However, 11, 16, or 22 bp spacing accommodates the binding of both EHF and JUN-FOS, and EHF binds to both free DNA and DNA with JUN-FOS bound with similar K_D values. In contrast, ERG can bind to DNA with JUN-FOS on all of the DNA spacing options that were tested (**Fig. 5.5c-d**). These data suggest that the difference between EHF and ERG in binding to DNA with JUN-FOS is limited to composite sites with less than 11 bp between ETS and AP1 DNA-binding motifs.

Given the strong antagonism displayed by JUN-FOS towards EHF binding at a composite ETS-AP1 DNA site with 6-bp spacing, we next investigated the reciprocality of this relationship. The presence of EHF had no detectable impact on JUN-FOS binding to composite sites with the more distal 11, 16, and 22 bp DNA spacing. Furthermore, we observed slower mobility bands corresponding to co-occupancy of EHF and JUN-FOS on the DNA (**Figs. 5.6** and **5.7**). At composite sites with 6-bp spacing of DNA, EHF antagonized JUN-FOS by only a modest 2-fold, compared to the ~100-fold decrease that was observed for EHF

binding. Slower mobility bands are not observed, as was the case with the reciprocal experiment (**Fig. 5.5 a-b**), indicating that JUN-FOS efficiently replaces EHF on the DNA. Interestingly, these results suggest nonreciprocal cooperativity in which JUN-FOS robustly deterred EHF from binding to DNA at the 6-bp composite sites whereas EHF only weakly deterred JUN-FOS binding. Previous studies have demonstrated that the dissociation rate constant for JUN-FOS is relatively smaller (slower dissociation rates) compared to ETS factors[25]. Therefore, while the equilibrium dissociation constants of JUN-FOS and EHF are relatively similar, this difference in the kinetics of interaction with DNA may underlie the nonreciprocal nature of antagonizing DNA binding between EHF and JUN-FOS.

DNA-binding domains (DBDs) of EHF and JUN-FOS
are not sufficient for antagonism of DNA binding

The close proximity required for antagonistic DNA binding between EHF and JUN-FOS suggested that the features required for JUN-FOS antagonizing the DNA binding of EHF might be located near the (DBDs) of these factors. We devised a rough structural model of the DBDs of these factors using DNA-bound structures of the ETS domain from ERG[27] (4IRI.pdb) or ELF3[28] (3JTG.pdb), the closest paralog of EHF with a solved structure for the ETS domain, and the bZIP domains of JUN-FOS[29] (1FOS.pdb) (**Fig. 5.8**). These structures were aligned on a modeled 6-bp composite motif. This model suggests that the ETS domain from both ERG and EHF can bind DNA with JUN-FOS without any steric hindrance.

However, the inability of full-length EHF to bind to this site indicates two possibilities. Either our model does not accurately reflect the DNA-bound state of ETS and JUN-FOS factors, or a region outside of the ETS domain of EHF is necessary for the antagonism of EHF and JUN-FOS.

To test the whether the DBDs of EHF and JUN-FOS are sufficient for DNA-binding antagonism, we expressed and purified fragments corresponding to the ETS domain of EHF (amino acids 203-300) and the bZIP domains of JUN (amino acids 250-319) and FOS (amino acids 131-203). After purification we combined JUN and FOS DBDs at a 1:2 molar ratio and this heterodimer bound to DNA with much stronger affinity (K_D = 1.2 nM) than JUN (K_D = 40 nM) or FOS homodimers (K_D too high to be determined) (**Fig. 5.9**). EHF and JUN-FOS DBDs bound to the 6-bp composite ETS – AP1 site with similar affinity as full-length EHF and JUN-FOS, respectively (**Fig. 5.10**). Interestingly, the DBD of EHF is able to bind to the 6-bp composite ETS – AP1 DNA site with the DBD of JUN-FOS (**Fig. 5.11a**). In contrast, full-length EHF displaces JUN-FOS DBD (**Fig. 5.11b**), as previously observed with full-length JUN-FOS. We conclude that the ETS domain of EHF can bind to DNA with the bZIP domains of JUN-FOS, indicating that a region outside of the ETS domain of EHF is required for antagonizing activity of JUN-FOS. There is a band of minor intensity with slower mobility than the full-length EHF:DNA band. This band might correspond to full-length EHF and JUN-FOS DBD co-occupying a single DNA duplex. In this case, components outside of both DBDs of both ETS and AP1 factors contribute to DNA-binding antagonism.

Discussion

Here we have shown that ETS factors have contrasting abilities to bind to composite ETS-AP1 sites. The two components that contribute to differential binding by ETS factors to ETS-AP1 sites are (i) differing affinities between ETS factors for the modified ETS site that is characteristic of ETS-AP1 composite motifs and (ii) distinct capabilities between ETS factors for binding to these composite motifs with JUN-FOS. Whereas EHF strongly binds to ETS-AP1 composite sites, its binding is antagonistic to simultaneous JUN-FOS binding. In contrast, ERG weakly binds to ETS-AP1 composite sites, but can bind to DNA in conjunction with JUN-FOS. The difference in affinity for the ETS-AP1 composite motif maps to the ETS domains of ERG and EHF, thus implicating divergent sequences that contact DNA (**Fig. 5.12**). However, the ETS domain of EHF is not sufficient for the anticooperative DNA-binding with JUN-FOS. Thus, we hypothesize that the region just N terminal of the ETS domain is responsible for this effect based on two lines of evidence. First, we observe anticooperative binding between EHF and JUN-FOS with 6-bp spacing but not with 11-bp or greater spacing between the two sequences, indicating that the region of EHF responsible for antagonism is proximally located, at least in three-dimensional space, to the ETS domain. Second, our structural model demonstrates that with the 6-bp spacing the N terminus of the ETS domain is oriented towards, and in close proximity to (~ 3 Å) JUN-FOS. This hypothesis is also consistent with observations of a fragment of SPI1 containing the ETS domain with an additional 15 amino acids N-terminal and 5 amino acids C-terminal that also anti-

cooperatively binds to DNA with JUN-FOS[25]. The sequences just N-terminal of the ETS domain are not conserved between SPI1 and EHF; however, both factors contain stretches of positive residues, KKK for SPI1 and KKH for EHF, that are evolutionarily conserved for each factor (**Fig. 5.12**). We hypothesize that these positive residues may result in charge repulsion with the highly basic DBDs of JUN and FOS. Further experiments that would disrupt these residues or structural studies that would define the interaction will be necessary to test this hypothesis. In summary, differences sequences within and outside of the ETS domain for EHF and ERG underlie the differential ability of these factors to bind to ETS-AP1 motifs and cooperatively bind to DNA with JUN-FOS.

The RAS/RAF/MEK/ERK (RAS-MAPK) signaling pathway amplifies the transcription output of ERG and ETV1/4/5 factors, as well as JUN-FOS, but not EHF or SPDEF[30-32]. Transcriptional activation in response to RAS-MAPK signaling is a common feature of many ETS factors, including several that are expressed in normal prostate tissue such as ETS1 and GABPA. Interestingly, among other ETS factors that are transcriptionally activated by RAS-MAPK signaling, all, except SPIB, belong to the same DNA-binding subclass as ERG and ETV4[4]. This suggests that, similarly to ERG, these RAS-MAPK-responsive ETS factors would also bind more poorly to the –AGGAA- ETS motif in ETS-AP1 consensus sites. Therefore, ETS sites within the composite ETS-AP1 motifs appear “suboptimized”[33] in that they favor binding by the ETS factors, such as EHF and SPDEF, which do not amplify transcriptional output in response to RAS-MAPK signaling. Additionally, EHF and SPDEF compete with the other factors at

that site, ERG, ETV4, and JUN-FOS, which are capable of amplifying transcriptional output in response to RAS-MAPK signaling. This competition likely dampens the transcription of ETS-AP1 regulated genes in the normal prostate[23] (**Fig. 5.13**). However, common genetic lesions in prostate cancer, such as the loss of EHF[34] or SPDEF[22], and/or the overexpression of ERG or ETV1/4/5 factors[17], would perturb this hypothesized competition resulting in the overexpression genes that are important for cellular migration, a hallmark of the invasive stage of oncogenesis promoted by ETS factors[23].

Methods

Expression plasmids

Human ETS and JUN-FOS 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[35].

Expression and purification of proteins

All proteins were produced in *Escherichia coli* (IDE3) cells. Minimal DNA-binding domains for ETS and JUN-FOS proteins 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 0.5 mM isopropyl-b-D-thiogalactopyranoside (IPTG), and grown at 30 °C for ~ 3 hours.

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) (5 mM BME for JUN-FOS DBDs), and 1 mM phenylmethanesulfonylfluoride (PMSF). Cells were lysed by sonication and centrifuged at 40k rpm in a 45 Ti rotor (Beckman) for at least 30 minutes 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 ETS 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). 10 mM DTT and 10% glycerol (v:v) was added to JUN and FOS DBDs which were then flash frozen in single-use aliquots for EMSAs. After centrifugation at 40k rpm for 30 minutes at 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 protein was then concentrated on a 3-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 and JUN-FOS factors generally expressed more efficiently in the insoluble fraction (with the exception of EHF which was expressed and

purified as described above) using either an IPTG expression protocol as described above or an “auto-induction” protocol[36]. Briefly, bacteria in 250 mL of autoinduction media were grown in 4 L flasks at 37 °C to an OD_{600} ~ 0.6 – 1. The temperature was then reduced to 30 °C and cultures were grown for another ~ 12 – 24 hours. Final OD_{600} values were typically ~ 6 – 12, indicating robust autoinduction. Harvested cells were resuspended as described above, sonicated and centrifuged at 15k rpm in a JA-17 rotor (Beckman) for 15 minutes 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 hour at 4 °C, the sample was centrifuged for 40k rpm for at least 30 minutes 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. HIS_6-tagged FOS and untagged JUN were combined prior to Ni^{2+} column purification to enrich for a molar excess of FOS. JUN-FOS heterodimers were then purified with size-exclusion column but not ion-exchange column. Thrombin cleavage of His_6-tags was not used on FL ETS and JUN-FOS factors due to internal cleavage sites for some proteins. Also, either a Q-sepharose anion-exchange column or a SP-

sepharose cation-exchange column was used depending on the isoelectric point of the protein.

Electrophoretic mobility shift assays (EMSAs)

DNA-binding assays utilized the following duplexed oligonucleotides:

UPP promoter

5’-TAGG**GGAA**A**TGACTCA**TTCA-3’

5’-TGAATGAGTCATTTCCCCTA-3’

High Affinity ETS consensus

5’-TCGACGGCCAAGCC**GGAA**GTGAGTGCC-3

5’-TCGAGGCACTCACTTCCGGCTTGGCCG-3’;

High Affinity ETS consensus C2A

5’-TCGACGGCCAAGCA**GGAA**GTGAGTGCC-3

5’-TCGAGGCACTCACTTCCTGCTTGGCCG-3’;

ETS-AP1 6 bp spacing

5’-TCGACGGCCAAGCA**GGAA**G**TGACTCA**GCCCGATCG-3’

5’-TCGACGATCGGGCTGAGTCACTTCCTGCTTGGCCG-3’;

ETS-AP1 11 bp spacing

5’-TCGACGGCCAAGCA**GGAA**GTAAAG**TGACTCA**GCCCGATCG-3’

5’-TCGACGATCGGGCTGAGTCACTTTACTTCCTGCTTGGCCG-3’

ETS-AP1 16 bp spacing

5’-TCGACGGCCAAGCA**GGAA**GTACGTACAAG**TGACTCA**GCCCGATCG-3’

5’-TCGACGATCGGGCTGAGTCACTTGTACGTACTTCCTGCTTGGCCG-3’

ETS-AP1 22 bp spacing

5’-

TCGACGGCCAAGCA**GGAA**GTACGTACGTACGTAAG**TGACTCA**GCCCGATC

G-3’

5’-

TCGACGATCGGGCTGAGTCACTTACGTACGTACGTACTTCCTGCTTGGCCG

3’

Boldface **GGAA** and **TGACTCA** indicate the ETS and AP1 binding site motifs, respectively. Each of these oligonucleotides, at 2 μM as measured by absorbance at 260 nM on a NanoDrop 1000 (Thermo Scientific), were labeled with [g-^{32}P] ATP using T4 polynucleotide kinase at 37 °C for ~ 30 – 60 minutes. After purification over a Bio-Spin 6 chromatography column (Bio-Rad), the oligonucleotides were incubated at 100 °C for ~ 5 minutes, and then cooled to room temperature over 1 – 2 hours. The DNA for EMSAs was diluted to 1 x 10^{-12} M and held constant, whereas protein concentrations ranged ~ 6 orders of magnitude with the exact concentrations dependent on 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 mg) 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 minutes 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 mg/mL BSA, 10 mM DTT, 2.5 ng/mL 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 $[P]_t$ 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.

Structural modeling

Structural modeling was performed using PyMol (Schrödinger, LLC). DNA-bound structures of JUN-FOS[29] (1FOS.pdb), ERG[27] (4IRI.pdb), or ELF3[28] (3JTG.pdb) were aligned using DNA sequences that would overlap in a composite ETS-AP1 motif with 6 bp DNA spacing: 4IRI/B nucleotides 9-12 and 3JTG/C nucleotides 111-114 were overlapped with 1FOS/C nucleotides 6-9.

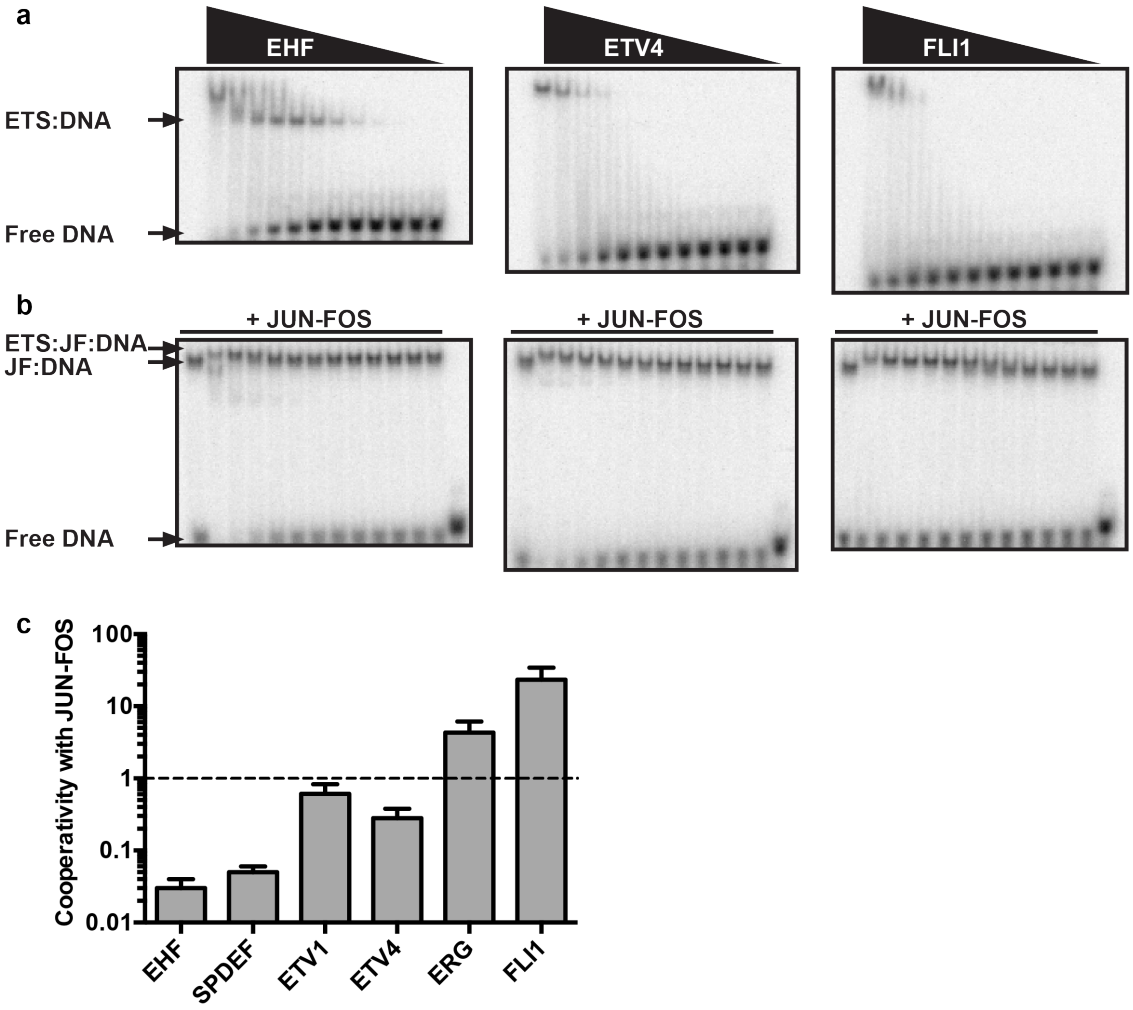


Figure 5.1 JUN-FOS differentially influences the DNA binding of ETS factors. (**a**) Representative EMSAs for EHF (left), ETV4 (middle), and FLI1 (right), to the UPP promoter DNA duplex. Note that higher band for EHF corresponds to two EHF molecules bound to the DNA duplex, as previously observed[25]. (**b**) As in (a), but with JUN-FOS bound to DNA duplex. (**c**) Quantification of cooperativity with JUN-FOS, where cooperativity = K_D (ETS only) / K_D (ETS + JUN-FOS). See Fig. 5.2 for quantification.

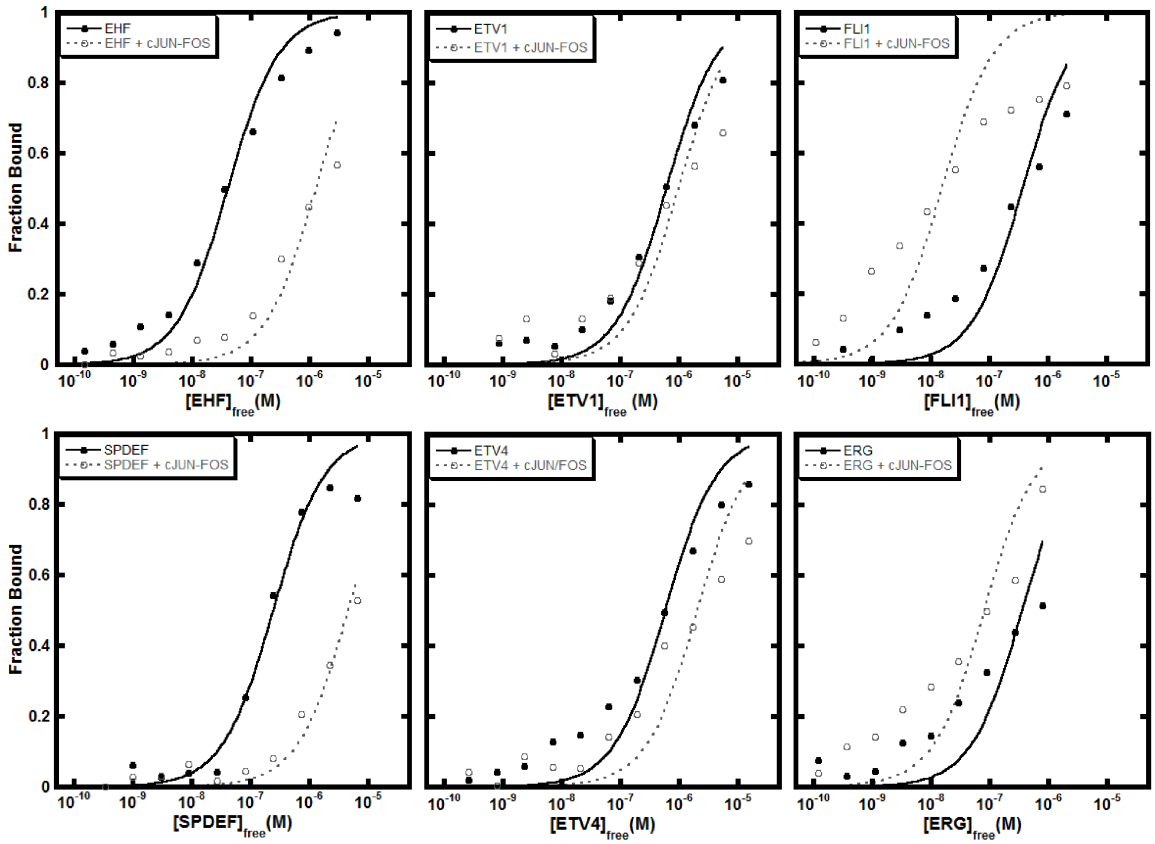


Figure 5.2 JUN-FOS differentially influences the DNA binding of ETS factors. Binding isotherms for ETS factors binding to DNA bearing the sequence of the UPP promoter in the absence (black) or presence of JUN-FOS (gray). Each data point is the mean from two replicates. See Methods for details.

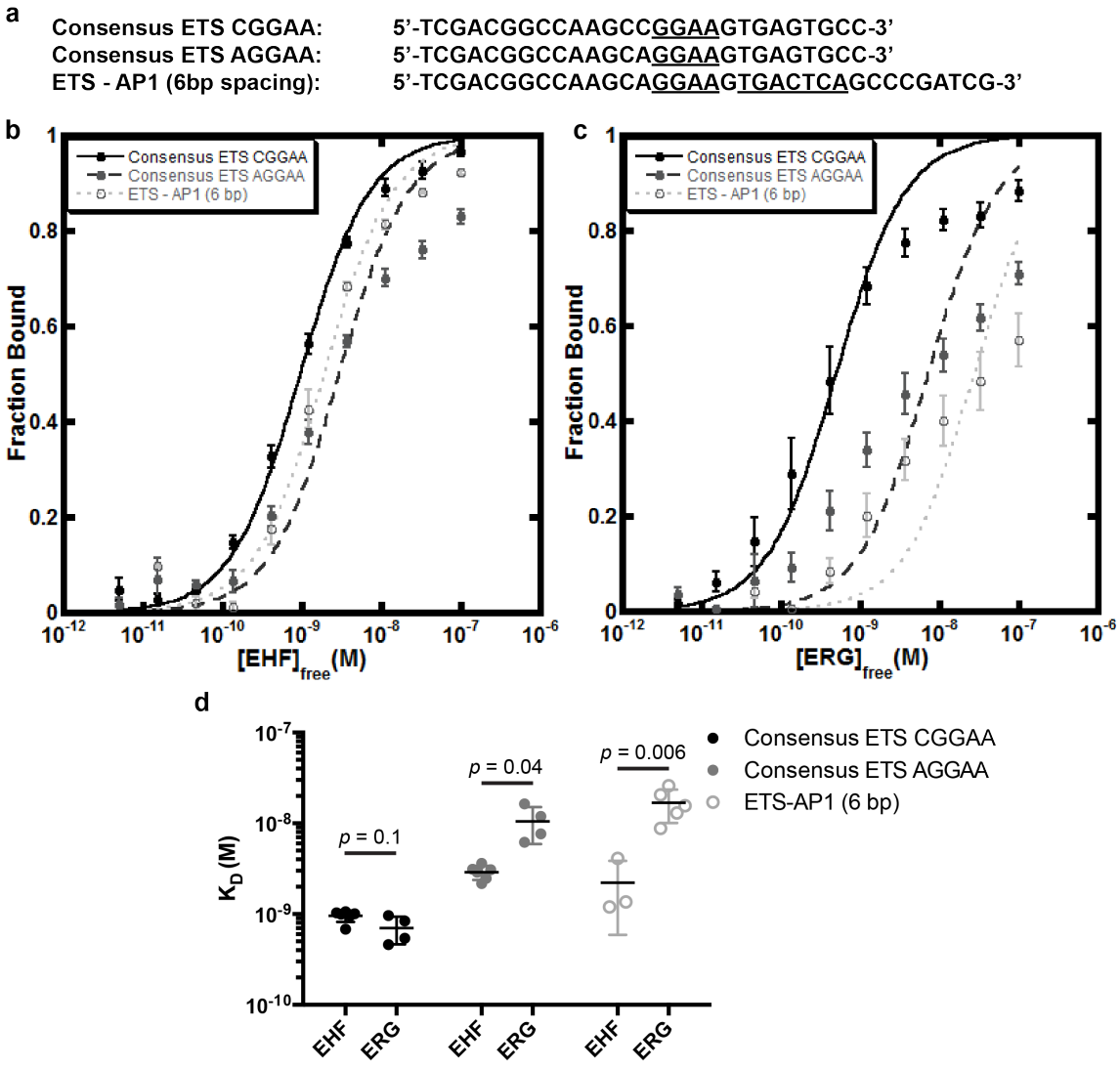


Figure 5.3 Single-nucleotide change in ETS binding sequence differentially affects the DNA binding of ETS factors. (**a**) Top strand of DNA duplexes used for measuring K_D values with EHF or ERG. ETS and AP1 DNA-binding sites are underlined. Binding isotherms for EHF (**b**) and ERG (**c**) with the DNA duplex containing a consensus DNA sequence for these factors (black), a single nucleotide change (dark gray) that occurs at ETS-AP1 composite motifs, or an ETS-AP1 composite site (light gray) with a DNA spacing of 6 bp between ETS and AP1 binding sequences. Mean and standard error of the mean are displayed from at least three replicates. (**d**) Comparison of K_D values for EHF and ERG for all three DNA duplexes. Each dot depicts a single replicate and lines depict the mean and standard deviation for each group of replicates. Two-tailed heteroscedastic t-tests were used to calculate *p* values.

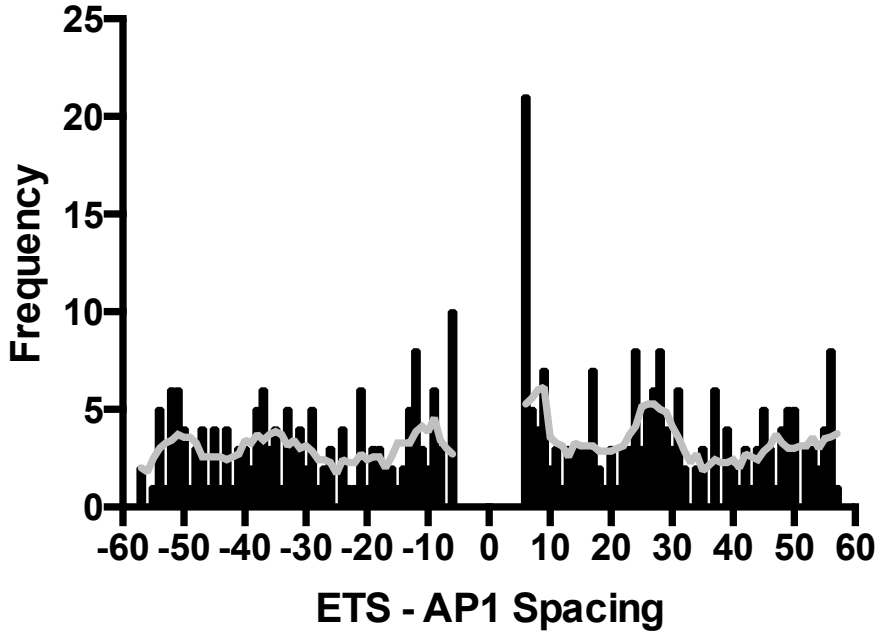


Figure 5.4 Spacing of *in vivo* ETS and AP1 binding sequences. Frequency of occurrence for each given spacing between ETS and JF binding sequences. Data derived from MEME analysis of ERG occupied genomic regions[23]. Gray line is the frequency smoothed over a 6bp window. Data are modified from Hollenhorst et al., 2011.

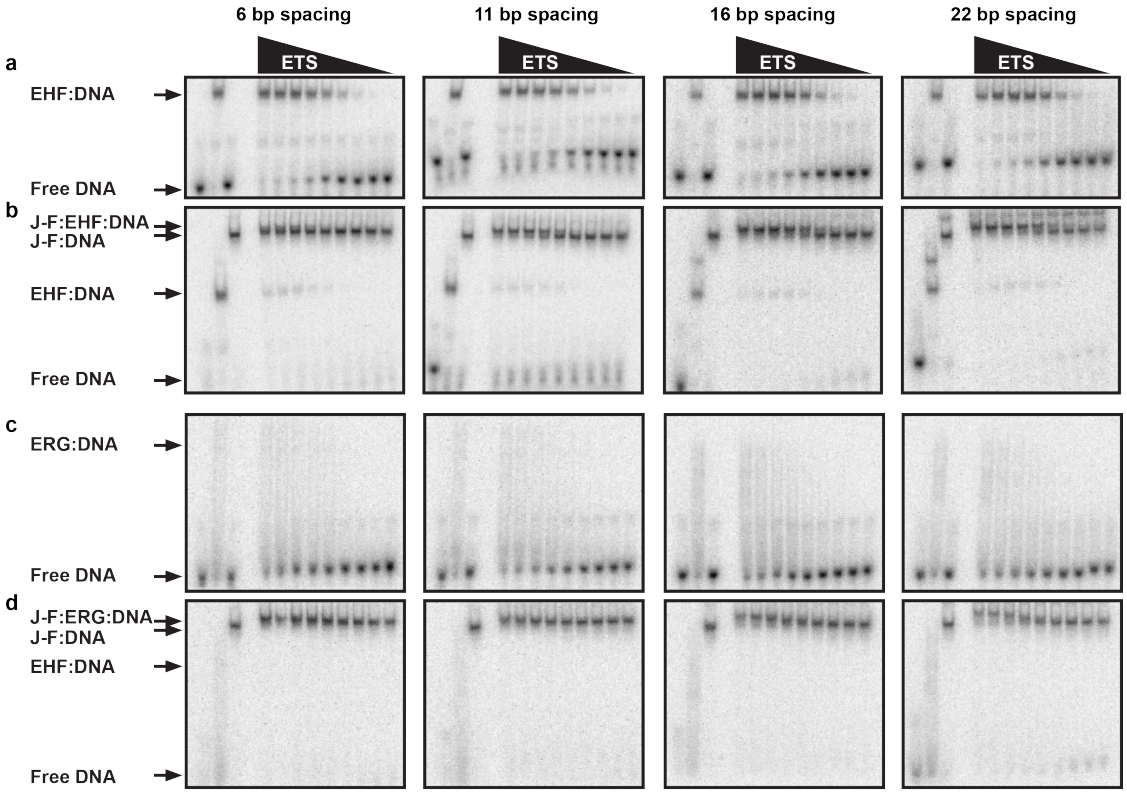


Figure 5.5 Spacing dependence of JUN-FOS (J-F) antagonism of EHF binding. Representative EMSAs. The variable bp spacing at top refers to DNA length between ETS and AP1 DNA binding sites. First three lanes of each EMSA are controls with DNA only, ETS:DNA, and JUN-FOS:DNA, respectively. Triangle denotes a titration with an increasing amount of ETS factor. EHF alone (**a**) and with JUN-FOS bound to DNA (**b**). ERG alone (**c**) and with JUN-FOS (**d**).

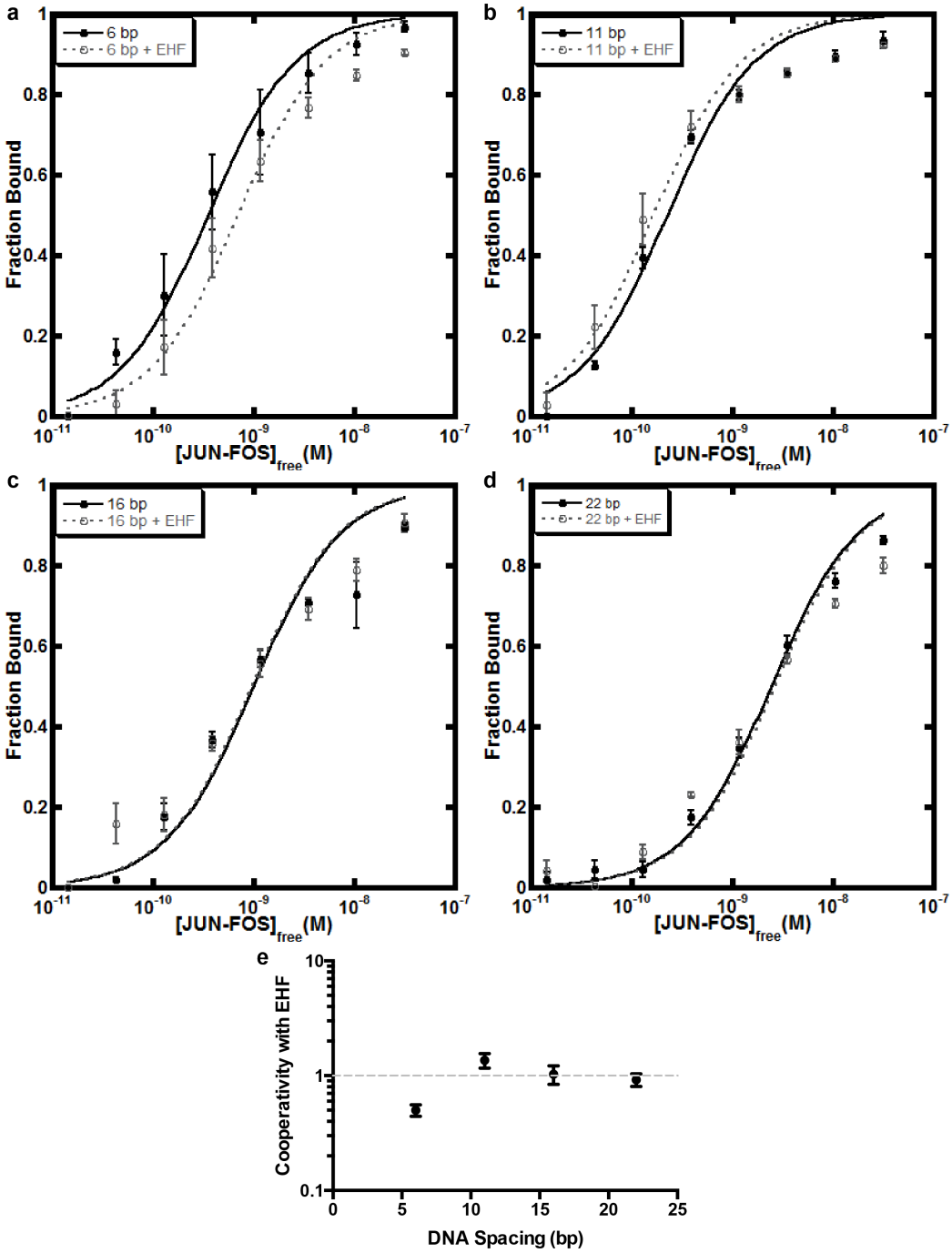


Figure 5.6 Spacing dependence of EHF antagonism of JUN-FOS binding. Binding isotherms for JUN-FOS alone (black) or with EHF (gray) for DNA duplexes with spacing of 6 (**a**), 11 (**b**), 16 (**c**), or 22 bp (**d**) between ETS and JUN-FOS DNA-binding sites. See Fig. 5.7 for representative EMSAs. (**e**) Cooperativity of JUN-FOS DNA-binding with EHF as a function of DNA spacing. Cooperativity = K_D (JUN-FOS) / K_D (JUN-FOS with EHF). Mean and standard error of the mean are displayed for three replicates.

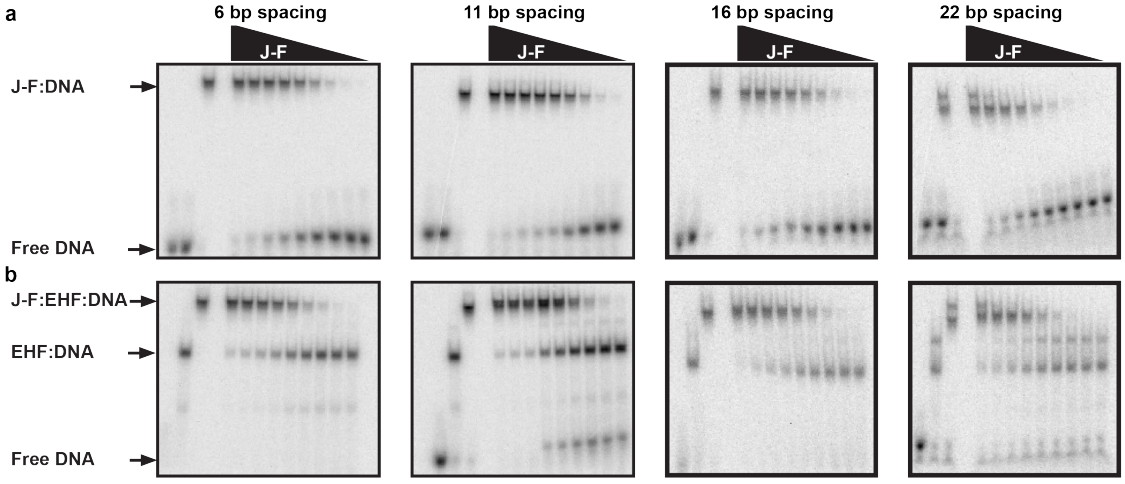


Figure 5.7 EHF effects on JUN-FOS binding. Representative EMSAs with DNA sequence with spacing between ETS and AP1 sites as displayed above EMSAs. First three lanes of each EMSA are controls with DNA only, ETS:DNA, and JUN-FOS:DNA, respectively. Triangle denotes a titration with an increasing amount of JUN-FOS. (**a**) EMSAs for JUN-FOS. (**b**) EMSAs for JUN-FOS with EHF bound to DNA. See Fig. 5.6 for quantification.

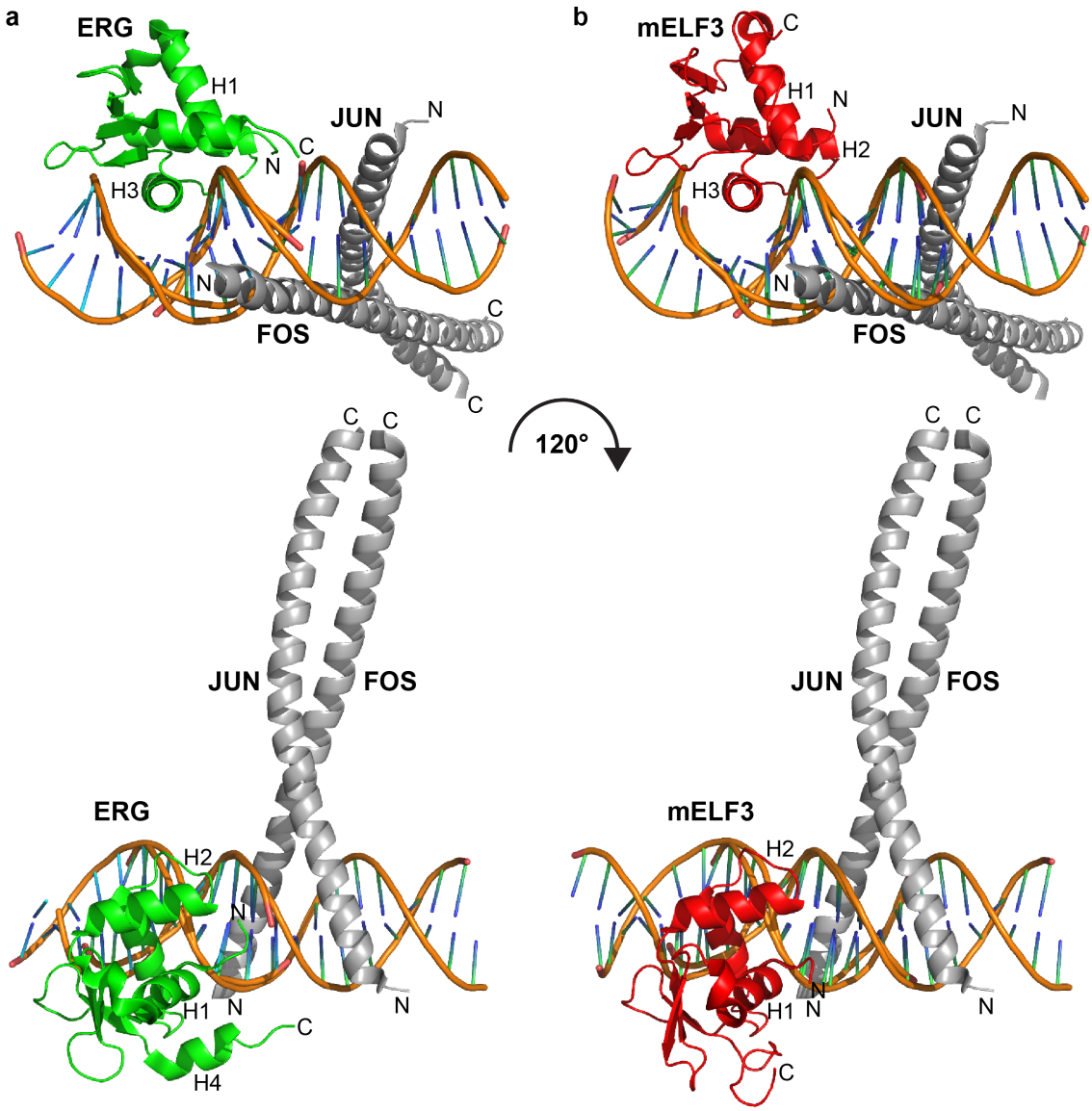


Figure 5.8 Structural model for ETS and AP1 factors bound to DNA. (**a**) Structures of ERG (4IRI.pdb)[27] and JUN-FOS (1FOS.pdb)[29] bound to DNA were aligned to assemble a composite DNA site and ternary complex. See Methods for details. (**b**) As in (a) but with mouse ELF3 (3JTG.pdb)[28], a closely related paralog of EHF. The JUN-FOS heterodimer can bind to consensus palindromic DNA sequences in either orientation[29], so the relative positioning of JUN and FOS relative to ETS factors in this model is arbitrary. The alternative orientations may be in equilibrium between both orientations in complexes in solution.

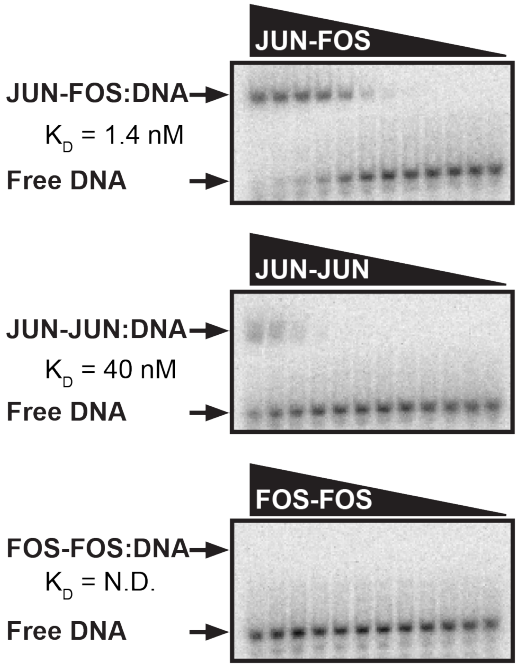


Figure 5.9 DNA binding affinities of JUN-FOS heterodimer, JUN or FOS homodimers. Representative EMSAs. Equivalent titrations of JUN-FOS heterodimer (mixed at a 1:2 JUN:FOS molar ratio, top) and JUN (middle) or FOS (bottom) homodimers. Dimers were assembled with DBDs of JUN (amino acids 250-319) and FOS (amino acids 131-203). K_D values were calculated as in Fig. 5.2.

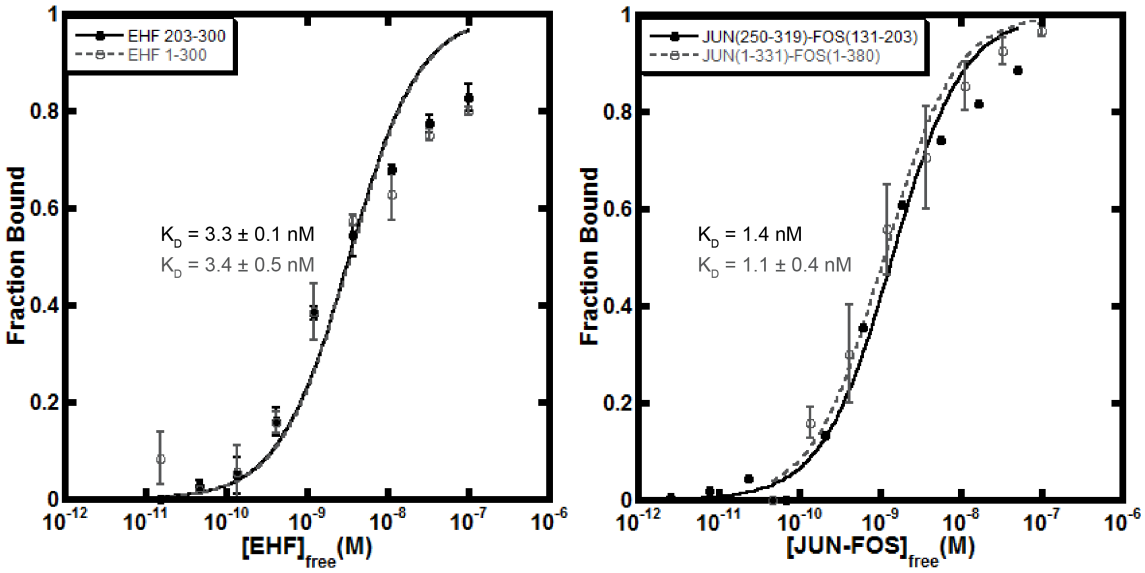


Figure 5.10 DBDs and full-length EHF and JUN-FOS bind to ETS-AP1 composite site with similar affinity. Binding isotherms for EHF full length (gray) and DBD (black), left, and JUN-FOS full length (gray) and DBD (black), right.

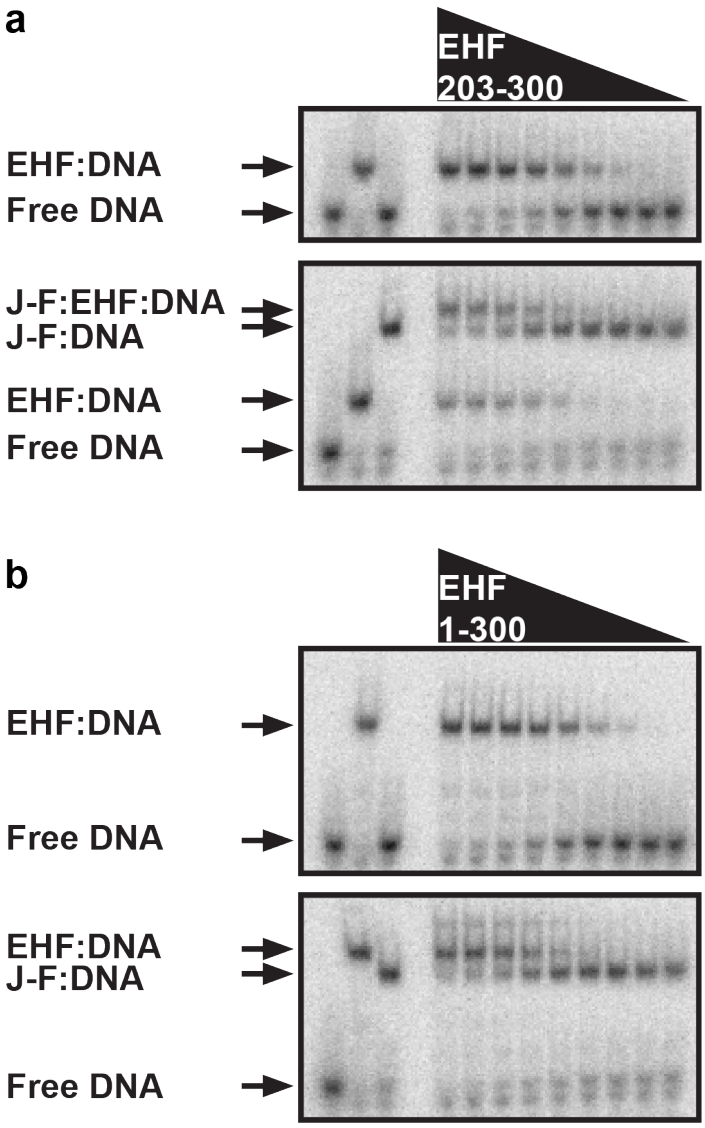


Figure 5.11 The DBDs of EHF and JUN-FOS can simultaneously bind ETS – AP1 composite DNA sites with 6-bp spacing. Representative EMSAs with the first three lanes of each EMSA being DNA only, EHF:DNA, and JUN-FOS:DNA controls, respectively. Triangle denotes a titration with an increasing amount of EHF. (**a**) ETS domain of EHF, amino acids 203-300, was titrated in absence (top), or presence of the JUN-FOS DBDs (bottom). (**b**) Full-length EHF, amino acids 1-300 binding analyzed as in (a).

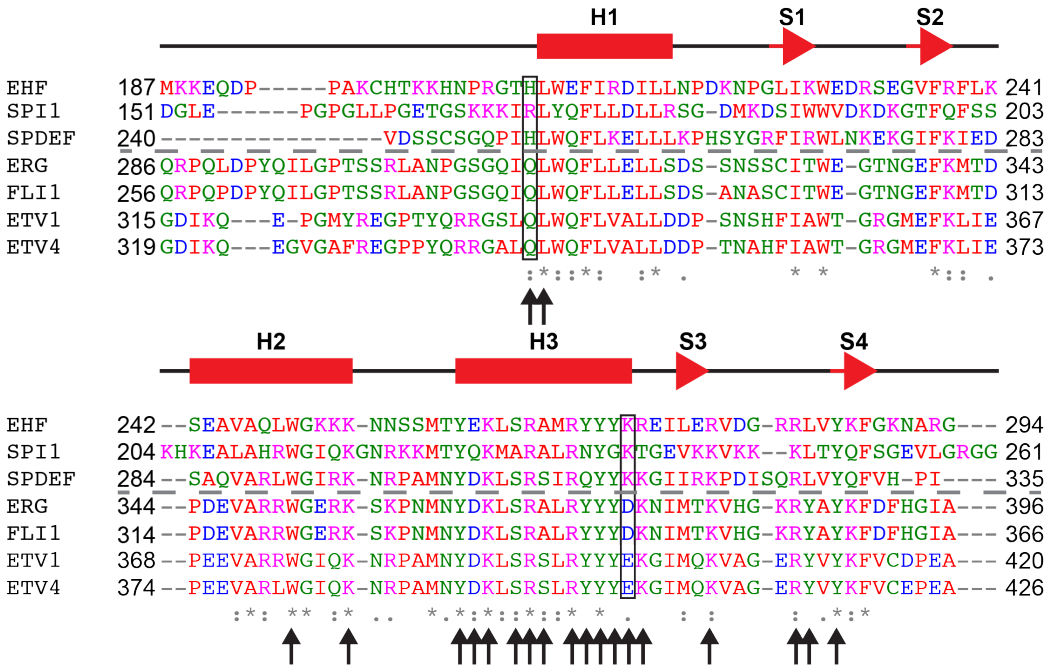


Figure 5.12 Sequence conservation of ETS domain and flanking regions for selected ETS factors. Sequences are aligned for ETS factors that do (below dotted line) or do not (above dotted line) display anticooperativity in binding with JUN-FOS to DNA with 6 bp spacing between factor binding sites using Clustal Omega[37]. Amino acids are colored according to physicochemical properties. Red rectangles and arrows above a sequence represent α-helices and β-strands, respectively. Arrows below the sequence denote amino acids that contact DNA in the DNA-bound structures of SPI1[8], SPDEF[26], ERG[27], or ETV1[38] Amino acids with clear differences between ETS factors that do or do not display anti-cooperativity in binding to DNA with JUN-FOS are circled.

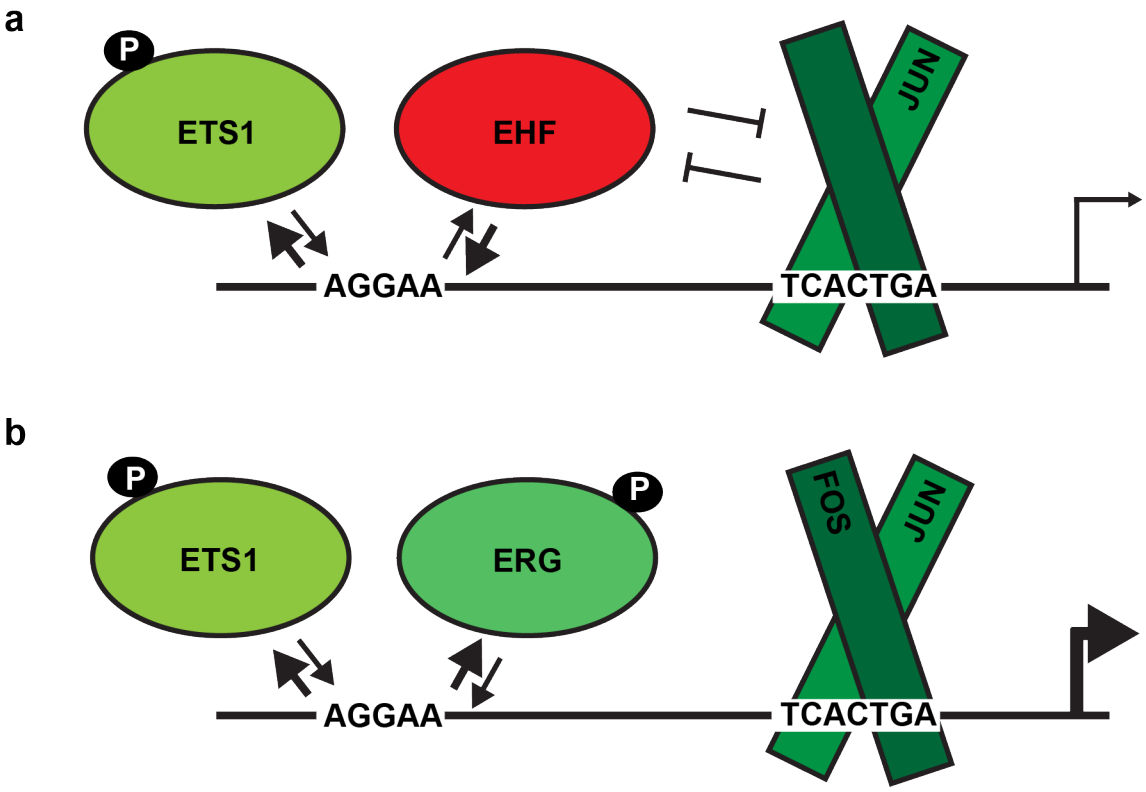


Figure 5.13 Model for EHF/SPDEF repression of ETS-AP1 regulated genes. (**a**) EHF / SPDEF, denoted here as EHF for simplicity, compete with other ETS factors for binding to ETS-AP1 binding sites in normal prostate cells. EHF binds more strongly to these sites than the other RAS-MAPK responsive ETS factors, depicted here as phosphorylated ETS1, and antagonizes the DNA-binding of JUN-FOS, in order to repress the transcriptional activation of ETS-AP1 regulated genes. (**b**) The addition/expression of ERG or ETV1/4/5 subfamily factors and/or the loss of EHF/SPDEF in prostate cancer promotes the expression of ETS-AP1 regulated genes by perturbing the balance between RAS-MAPK responsive factors that can bind to DNA with JUN-FOS and RAS-MAPK irresponsive factors that antagonize JUN-FOS DNA binding. The spacing between ETS and JUN-FOS binding sequences is exaggerated and not to scale in order to incorporate multiple ETS proteins in competition for the ETS site.

References

1. Lee, T.I. & Young, R.A. Transcription of eukaryotic protein-coding genes. *Annu Rev Genet* **34**, 77-137 (2000).

2. Panne, D. The enhanceosome. *Curr Opin Struct Biol* **18**, 236-42 (2008).

3. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. & Luscombe, N.M. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**, 252-63 (2009).

4. Wei, G.H. et al. Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J* **29**, 2147-60 (2010).

5. Jolma, A. et al. DNA-binding specificities of human transcription factors. *Cell* **152**, 327-39 (2013).

6. Hollenhorst, P.C., Jones, D.A. & Graves, B.J. Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. *Nucleic Acids Res* **32**, 5693-702 (2004).

7. Hollenhorst, P.C., Shah, A.A., Hopkins, C. & Graves, B.J. Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. *Genes Dev* **21**, 1882-94 (2007).

8. Kodandapani, R. et al. A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. *Nature* **380**, 456-60 (1996).

9. Donaldson, L.W., Petersen, J.M., Graves, B.J. & McIntosh, L.P. Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. *EMBO J* **15**, 125-34 (1996).

10. Batchelor, A.H., Piper, D.E., de la Brousse, F.C., McKnight, S.L. & Wolberger, C. The structure of GABPalpha/beta: an ETS domain- ankyrin repeat heterodimer bound to DNA. *Science* **279**, 1037-41 (1998).

11. Hassler, M. & Richmond, T.J. The B-box dominates SAP-1-SRF interactions in the structure of the ternary complex. *EMBO J* **20**, 3018-28 (2001).

12. Garvie, C.W., Hagman, J. & Wolberger, C. Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell* **8**, 1267-76 (2001).

13. Escalante, C.R. et al. Crystal structure of PU.1/IRF-4/DNA ternary complex. *Mol Cell* **10**, 1097-105 (2002).

14. Shrivastava, T. et al. Structural basis of Ets1 activation by Runx1. *Leukemia* **28**, 2040-8 (2014).

15. Shiina, M. et al. A novel allosteric mechanism on protein-DNA interactions underlying the phosphorylation-dependent regulation of Ets1 target gene expressions. *J Mol Biol* **427**, 1655-69 (2015).

16. Hollenhorst, P.C. et al. DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genet* **5**, e1000778 (2009).

17. Tomlins, S.A. et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* **448**, 595-9 (2007).

18. Baena, E. et al. ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. *Genes Dev* **27**, 683-98 (2013).

19. Chen, Y. et al. ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nat Med* **19**, 1023-9 (2013).

20. Aytes, A. et al. ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. *Proc Natl Acad Sci U S A* **110**, E3506-15 (2013).

21. Cangemi, R. et al. Reduced expression and tumor suppressor function of the ETS transcription factor ESE-3 in prostate cancer. *Oncogene* **27**, 2877-85 (2008).

22. Cheng, X.H. et al. SPDEF inhibits prostate carcinogenesis by disrupting a positive feedback loop in regulation of the Foxm1 oncogene. *PLoS Genet* **10**, e1004656 (2014).

23. Hollenhorst, P.C. et al. Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes Dev* **25**, 2147-57 (2011).

24. Tugores, A. et al. The epithelium-specific ETS protein EHF/ESE-3 is a context-dependent transcriptional repressor downstream of MAPK signaling cascades. *J Biol Chem* **276**, 20397-406 (2001).

25. Kim, S., Denny, C.T. & Wisdom, R. Cooperative DNA binding with AP-1 proteins is required for transformation by EWS-Ets fusion proteins. *Mol Cell Biol* **26**, 2467-78 (2006).

26. Wang, Y. et al. Analysis of the 2.0 A crystal structure of the protein-DNA complex of the human PDEF Ets domain bound to the prostate specific antigen regulatory site. *Biochemistry* **44**, 7095-106 (2005).

27. Regan, M.C. et al. Structural and dynamic studies of the transcription factor ERG reveal DNA binding is allosterically autoinhibited. *Proc Natl Acad Sci U S A* **110**, 13374-9 (2013).

28. Agarkar, V.B., Babayeva, N.D., Wilder, P.J., Rizzino, A. & Tahirov, T.H. Crystal structure of mouse Elf3 C-terminal DNA-binding domain in complex with type II TGF-beta receptor promoter DNA. *J Mol Biol* **397**, 278-89 (2010).

29. Glover, J.N. & Harrison, S.C. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* **373**, 257-61 (1995).

30. Hibi, M., Lin, A., Smeal, T., Minden, A. & Karin, M. Identification of an oncoprotein- and UV-responsive protein kinase that binds and potentiates the c-Jun activation domain. *Genes Dev* **7**, 2135-48 (1993).

31. Janknecht, R. Analysis of the ERK-stimulated ETS transcription factor ER81. *Mol Cell Biol* **16**, 1550-6 (1996).

32. Selvaraj, N., Kedage, V. & Hollenhorst, P.C. Comparison of MAPK specificity across the ETS transcription factor family identifies a high-affinity ERK interaction required for ERG function in prostate cells. *Cell Commun Signal* **13**, 12 (2015).

33. Farley, E.K. et al. Suboptimization of developmental enhancers. *Science* **350**, 325-8 (2015).

34. Albino, D. et al. ESE3/EHF controls epithelial cell differentiation and its loss leads to prostate tumors with mesenchymal and stem-like features. *Cancer Res* **72**, 2889-900 (2012).

35. Li, M.Z. & Elledge, S.J. SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol Biol* **852**, 51-9 (2012).

36. Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207-34 (2005).

37. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).

38. Cooper, C.D., Newman, J.A., Aitkenhead, H., Allerston, C.K. & Gileadi, O. 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. *J Biol Chem* **290**, 13692-709 (2015).

CHAPTER 6

SUMMARY AND FUTURE DIRECTIONS

Summary

Here we have described distinct intra- and intermolecular protein-protein interactions that distinguish between ETS transcription factors that serve as oncoproteins and tumor suppressors in the context of prostate cancer. The DNA-binding of ETV1, ETV4, and ETV5 is autoinhibited through proposed intramolecular interactions between two inhibitory regions and the ETS domain (Chapter 3). Less autoinhibition is observed for ERG and FLI1, and no autoinhibition is observed for EHF (Chapter 5). Mediator subunit 25 interacts with the activation domain and DNA-binding domain of ETV4, but does not interact with EHF (Chapter 4). EHF, but not ERG or ETV4, antagonizes the DNA binding of JUN-FOS (Chapter 5). Future studies will determine if these distinctions underlie the differential roles observed for ETS factors in prostate cancer.

In analyzing the ETS factors that act as oncoproteins in prostate cancer, we determined that ETV1, ETV4, and ETV5 have higher magnitudes of DNA-binding autoinhibition than did ERG or FLI1 (Chapter 3). Work published during our investigations confirmed this lower magnitude of autoinhibition for ERG[1]. Previous studies had qualitatively observed DNA-binding autoinhibition in the ETS subfamily of ETV1, ETV4, and ETV5, but mechanistic detail of this inhibition was not interrogated[2-4]. We found that an intrinsically disordered region and an α-helix cooperatively inhibit the ETS domain by perturbing the DNA-recognition α-helix. Acetylation of lysine residues within the disordered region abrogated autoinhibition, providing one potential cellular mechanism for the regulation of DNA binding by ETV1, ETV4, and ETV5. The inhibitory domains are distinct from

those previously described for ETS1[5,6] or ETV6[7,8]. Partnerships with the transcription factors RUNX1[9] and PAX5[10] relieve the autoinhibition of ETS1. Therefore, distinct protein partnerships may relieve the autoinhibition of the ETV1/4/5 subclass of ETS factors.

We analyzed whether the interaction with MED25 relieves the DNA-binding autoinhibition of ETV4 (Chapter 4). Previous studies established that the N-terminal activation domain (AD) of ETV5 was sufficient for interaction with the activator-interacting domain (ACID) of MED25[11]. However, we surmised and subsequently discovered a secondary binding site consisting of the ETS domain and inhibitory α-helix of ETV4. Our findings have some similarity to the previously characterized bipartite interaction between the viral protein VP16 and MED25[12,13]. Mutation of residues within the inhibitory α-helix strongly ablated the interaction of ETV4 with MED25. The inhibitory α-helix is specific to the ETV1/4/5 subfamily of ETS factors, providing a rationale for the specificity of MED25 interacting with ETV4 but not EHF or ETS1. Interaction with MED25 activated the DNA binding of ETV4, likely by disrupting the interaction between the inhibitory α-helix and the ETS domain of ETV4.

A previous study demonstrated that ERG, ETV1, and ETV4 bind to composite ETS-AP1 DNA binding sites when overexpressed in prostate cells[14]. Therefore, we asked whether binding nearby with JUN-FOS relieves ETV4 autoinhibition (Chapter 5). While we did not find evidence for modulation of ETV4 autoinhibition, we did observe that ERG and ETV4 bind to ETS-AP1 sites with JUN-FOS. In contrast, EHF and SPDEF antagonize JUN-FOS binding to DNA.

Additionally, the nonoptimal ETS motif that occurs in the composite ETS-AP1 DNA sites differentially disfavors ERG binding compared to EHF. These two findings suggest alternative scenarios whereby either EHF or SPDEF bind in isolation or ERG/ETV4 and JUN-FOS bind jointly to composite ETS-AP1 sites. This model is in agreement with the transcriptionally repressive effect of SPDEF and EHF on genes regulated by composite ETS-AP1 sites[14,15].

We generated assays for ETS1 binding to DNA that are amenable to high-throughput screening for small-molecule inhibitors of this interaction (Chapter 2). While we performed our limited small-molecule screens with ETS1, we also demonstrated that other ETS factors, such as ERG and ETV1/4/5 subfamily factors, could be utilized in these assays. Using a nonoptimal ETS motif in the DNA sequence increased the inhibition of our lead compounds in these assays. Interestingly, the nonoptimal ETS motif used in the screening assay is similar to the ETS motif in ETS-AP1 composite sites. Therefore, in addition to potentiating inhibitors in this assay, the use of nonoptimal ETS motifs such as the ETS-AP1 composite site may also represent a more desirable molecular target.

Future directions

Further definition and affinity maturation of the
minimal inhibitory element in ETV4 NID

We broadly defined the N-terminal inhibitory domain (NID) as amino acids 165-336 in ETV4 and demonstrated that acetylation of two lysine residues within the NID as relieves the autoinhibition of ETV4. However, we have not yet defined

the amino acids within the NID that confer inhibition on the ETS domain. Definition of a minimal inhibitory unit would further our understanding of mechanism of NID-mediated autoinhibition in ETV4.

Additional experiments not described in Chapter 3 have contributed to our understanding of NID-mediated autoinhibition of ETV4. Further truncations suggested that amino acids 165-202 and 288-336 are dispensable for the NID's inhibitory function narrowing the essential region to 203-288 (data not shown). Both lysine residues that relieve autoinihbition upon acetylation, 226 and 260, reside within this region. However, these studies are difficult to perform due to the unpredictable stability of truncated proteins and the relatively low level of magnitude of autoinhibition. For example, the dramatic instability of ETV4 246-436 precludes accurate measurement of this fragment's affinity for DNA. We would predict from our truncation studies that the complete ablation of NID-mediated autoinhibition would result in a ~6-fold activation. We observe three- and 2-fold activation from the acetylation of lysines 226 and 260, respectively, suggesting that multiple regions within the NID contribute to the autoinhibition of ETV4 (**Fig. 6.1a**). These regions show some resemblance to one another in their conservation of hydrophobic amino acids around the acetylated lysine residues. One method that may potentially simplify the interrogation of the molecular interactions that mediate ETV4 autoinhibition would be to artificially add single or multiple copies of the conserved region around lysine 226 to the ETS domain of ETV4 (**Fig. 6.1b**). If this minimal region around lysine 226 truly encompasses the single 'unit' of autoinhibition within the NID, we would expect to see higher levels

of autoinhibition with more repeating copies of the unit[16]. This would then provide a more robust assay to test the functional importance of individual amino acids in the NID-ETS domain interaction. Charge-reversal mutations in the NID (lysines 226 and 260 to glutamic acid) and the ETS domain (glutamic acids 404, 423, and 425) do not ablate autoinhibition. Therefore, my current hypothesis is that aromatic residues surrounding lysines 226 and 260 interact with tyrosines 401 and 403 of the DNA-recognition α-helix in the ETS domain through π-π stacking interactions. These tyrosines are conserved in most (23/28) human ETS factors, but are both not conserved in the SPI1 subfamily. Mutation of these tyrosines to asparagine and glycine, the corresponding amino acids in SPI1, would test their role in DNA-binding autoinhibition.

Definition of the minimal "unit" of the NID would allow for engineering enhanced affinity and specificity of the NID for the ETS domain of ETV4 - a process termed affinity maturation[17]. The development of an inhibitory peptide with increased affinity and specificity for the ETS domain of ETV4 could serve as a tool to understand the consequences of dampening the function of ETV4 in prostate cancer and could serve as a basis for designing a small-molecule inhibitor[18]. We estimate that the separated NID (amino acids 165-336) and CID-inhibited ETV4 fragment (amino acids 337-436) interact in *trans* with an equilibrium dissociation constant, K_D, of ~ 40 mM based on biolayer interferometry experiments (data not shown). Truncation to a minimal inhibitory unit of ~20 amino acids would allow for saturating mutagenesis with either phage[19] or yeast[20] surface display using interaction with the CID-inhibited

fragment of ETV4 as the selection (**Fig. 6.1c**). While there is a high likelihood of substantially increasing the affinity of the NID-ETS interaction of ETV4 with this method, the more challenging task would be to specifically increase the affinity of the NID interaction with the ETS domain of ETV4, while generating or retaining specificity compared to other ETS domains. For example, the ETS domain of EHF could be used as a control to search for matured peptides that bind with high affinity to ETV4, but not to EHF. Such a peptide would meet the goal of development of a tool for studying the specific biological roles of ETV4. Additionally, structural understanding of the interaction between the inhibitory peptide and the ETS domain of ETV4 would guide future site-directed chemical screens to search for a specific small-molecule inhibitor[21].

Determining the importance of the ETV4 - MED25
interaction in prostate cancer

Previously, ETV4 has been shown to promote the metastasis of prostate cancer in response to activated PI3-kinase and Ras signaling in a mouse model of prostate cancer[22]. The overexpression of ETV4 in normal prostate cell lines is also sufficient to induce anchorage-independent growth[22,23] and cell migration[14]. Therefore, both simple cell line models and whole organism mouse models exist and can be used to test the importance of the ETV4-MED25 interaction in prostate cancer. For example, wild-type ETV4 could be compared to variant ETV4 bearing point mutations (e.g., F54A, S429A, and F54A/S429A) that are deficient for interaction with MED25 for their abilities to promote transformation or

migration of normal prostate cells. The activation domain of ETV1/4/5 subfamily members also interacts with other general transcriptional factors[24] and coactivators[25], so it would be important to perform the reciprocal experiments testing wild-type MED25 or mutants[26] that are deficient for interaction with ETV4 in these functional assays. MED25 is expressed in normal prostate. Therefore, two alternative approaches may provide a better readout than the overexpression of this protein. The overexpression of only the ACID domain of MED25 could "squelch" the activity of MED25 when ETV4 is also expressed. Wild-type and mutant fragments could be compared. Alternatively, expression of MED25 could be reduced by RNAi or mutations that interfere with ETV4 interaction could be introduced by CRISPR technology in PC3 cells, a prostate cancer cell line that expresses ETV4[27]. If these assays suggest that the MED25-ETV4 interaction is important for transformation or migration of prostate cells, these mutant proteins could be engineered into previously established mouse models of prostate cancer[22,28,29] in order to study the importance of this interaction *in vivo*.

Recently, the disparate distribution of transcription factors and Mediator complexes at enhancers was described[30]. Approximately 40% of Mediator complexes were concentrated at less than 3% of the enhancers, and these so-called "super enhancers" regulated highly expressed genes that define the cellular identity. In several human cancers, the super enhancers appear to be redistributed by oncogenic transcription factors, including ETS factors[31,32]. Therefore, it would be interesting to analyze the distribution of Mediator complexes in prostate cells before and after the expression of ETV4 using

MED25 as a proxy for Mediator complexes. Chromatin immunoprecipitation (ChIP) of MED25 and ETV4, individually, followed by next-generation sequencing (ChIP-seq) would allow for the identification of co-occupied genomic targets and determination of whether the expression of ETV4 changes the distribution of MED25. These regions could then be overlapped with enhancers as defined by the presence of the coactivator protein p300 and the histone modifications H3K4me1 and K3K27ac[30,33].

ETS – Mediator interactions

We propose two possible models for the ETV4 – MED25 interaction perturbing Mediator complexes in prostate cells. In the first model, ETV4, through interaction with MED25, perturbs the distribution of Mediator complexes and recruits Mediator to new genomic loci in prostate cells. In the second model, the interaction of ETV4 with Mediator through MED25 results in a conformational change that activates the Mediator complex by influencing the recruitment of RNA polymerase II and other cofactors. These models are not necessarily mutually exclusive, but we will refer to them as the distribution model and the activation model, respectively.

ETV1/4/5 factors interacting with MED25 add to the short list of characterized ETS - Mediator subunit interactions; ELK1, ELK3, ELK4, and ELF3 interact with MED23[34-36]. As these ETS factors are expressed in the normal prostate, their interaction with MED23 may help define the distribution of Mediator complexes in the normal prostate. Interestingly, ELF3 belongs to the

same ETS subgroup as EHF, a tumor suppressor in the prostate[37]. Gene disruption of MED23 and MED25 results in distinct defects in smooth muscle cell differentiation[38] and chondrocyte differentiation[39], respectively, suggesting that these subunits are required for the transcription of a distinct subset of genes. Establishing the genes that ELK1, ELK3, ELK4, or ELF3 co-regulate with MED23 in normal prostate cells would allow us to determine if the overexpression of ETV4 alters the transcription of, for example, ELF3-MED23 co-regulated genes. The aforementioned distribution model would predict that the overexpression of ETV4 would repress the transcription of ELF3-MED23 co-occupied genes by recruiting Mediator complex away from these loci. ChIP-seq experiments, as described above, for ELF3 and MED23 to determine co-occupied regions would be performed before and after ETV4 overexpression in prostate cells. Similar occupancy of MED23 at ELF3-MED23 co-occupied sites before and after ETV4 overexpression would disfavor the distribution model whereas loss of MED23 at these sites would favor the distribution model.

Interaction with ADs from distinct transcription factors imparts different conformational rearrangements of Mediator[40-42] and the recruitment of distinct cofactors to the Mediator complex[43]. Therefore, ETV4-MED25 interaction may modulate Mediator activity in a manner that is distinct from ELF3-MED23 interactions. Hybrid ETV4 – ELF3 constructs could be used to test this possibility. For example, the AD of ETV4 could be swapped into ELF3. Then the transcriptional output from ELF3-MED23 co-regulated genes could be compared to wild type ELF3. Similar levels of transcriptional output would favor a pure

distribution model, whereas activation of transcription with the ELF3-ETV4(AD) hybrid would support differential activation of Mediator from these transcription factors. As we determined that the inhibitory α-helix flanking the ETS domain of ETV4 also contributes to the binding of MED25, the contribution of the DBDs of these factors would need to be considered in these hybrid experiments.

One potential complication of this line of investigation is the capability of many transcription factor ADs to interact with multiple subunits of the Mediator complex[44-47]. This appears to be the case for the ADs of ELK1, ELK3, and ELK4 as each of these factors has a distinct dependence on MED23 for transcriptional activity[35] and are capable of associating with the Mediator complex in the absence of MED23[48]. Determining all of the Mediator subunit targets of ETS factors in normal prostate and prostate cancer would be necessary in defining the differential recruitment of Mediator complexes by these factors. For example, as a first pass, a pull-down or an immunoprecipitation with ETV4 followed by western blot analysis of individual Mediator subunits would be performed in cells expressing MED25, or with MED25 knocked out. If the results indicate that ETV4 is still capable of interacting with Mediator complexes then mass spectrometry analysis of these MED25-null complexes could be utilized to identify putative ETV4-targets amongst other Mediator subunits[43,49]. Nuclear extract from cells without MED25 and without the individual putative Mediator subunits could then be used to reanalyze ETV4 binding to modified Mediator complexes. Parallel experiments would be performed with the other ETS factors of interest.

Inhibition of specific AD interactions with small peptides

The activation domain of ETV1/4/5 is disordered in isolation, but forms an amphipathic α-helix when interacting with transcriptional coactivators. The bulky hydrophobic groups of these ADs are functionally important for coactivator interactions. This structural character is representative of many transcription factor ADs[50]. The dogma for this structural character is that the disordered, flexible nature of ADs allow for interaction with multiple distinct protein surfaces[51,52]. If this dogma were valid, then we would hypothesize that reinforcing the bound state of an AD with any single cofactor would inhibit the transcriptional activity of that transcription factor. Currently, the absence of solved structures for the AD of ETV1/4/5 factors bound to transcriptional cofactors limits the use of these proteins for this approach. In contrast, the AD of p53 is an ideal model system to test this hypothesis due to the wealth of solved structures with numerous cofactors[53-57] and the documented biological effects of mutations within the AD[58]. In addition to furthering our understanding of the chameleon-like nature of transcription factor ADs, these studies would also provide the basis to generate protein domains that specifically inhibit individual AD-cofactor interactions.

Multiple structural conformations in the interaction with distinct coactivators has been established for the AD of p53[53-57]. Intriguingly, the mutation of distinct hydrophobic amino acids within the AD of p53 selectively ablates transcription of p53 target genes. The L25Q/W26S double mutant ablates the transcription of genes that are important for the response to DNA damage

and the W53S/F54Q double mutant ablates the transcription of genes that mediate tumor suppression[58]. The simplest model to describe this distinction is that p53 interacts with different coactivators at the genes that control these two different cellular functions. The wealth of p53 AD – coactivator structures that exist allows for “domain grafting”, a computational approach that incorporates amino acids that are known hot-spots for an interaction into a new protein domain that favors interaction with the specific coactivator of interest[59,60]. For instance, protein domains could be grafted on to W53 and F54, two hot-spot residues for the interaction of the p53 AD with the nuclear receptor coactivator binding domain (NCBD) of the CREB-binding protein (CBP)[61](**Fig. 6.2**). Alternatively, other structures that interact with this same region of p53 AD, such as the RPA70 subunit of replication protein A[54], the A box of high-mobility group B1[55], or the pleckstrin homology domain from the TFIIH subunit p62[56,62] could be used as the starting model. Suitable grafted domains could be further optimized with phage or yeast surface display, as described above (**Fig. 6.1c**), and checked for affinity against other p53 coactivators. The resulting protein domains from this particular strategy would be predicted to specifically disrupt the transcription of p53 target genes that are important for tumor suppression[58]. However, these domains would serve as tools to probe the requirement of specific cofactors at p53-regulated genes. Additionally, this general strategy could be used to develop *de novo* protein domains that inhibit p53 gain of function mutants that drive cancer development[63].

The structural partnerships and biological knowledge of ETV1/4/5

activation domains are currently underdeveloped to take the same exact approach, although the maturation of the AD or of H4, the inhibitory α-helix, using phage or yeast surface display (**Fig. 6.1c**) could be utilized to generate higher-affinity binders for MED25. In order to utilize the approach outlined above, structures of ETV4 bound to cofactors, such as MED25, would need to be determined.

RAS-MAPK signaling and ETS factors

We have described an interesting distinction between ETS factors in their differential affinity for nonoptimal ETS motifs that occur within ETS-AP1 sites and in their differential ability to bind with JUN-FOS at these composite sites. These differences lead us to a model where EHF/SPDEF – type ETS factors bind to ETS-AP1 composite sites by themselves, and ERG/ETV4 – type ETS factors bind to ETS-AP1 composite sites with JUN-FOS. This model is consistent with previous descriptions of EHF and SPDEF as transcriptional repressors and ERG, ETV1, and ETV4 as transcriptional activators of ETS-AP1 composite sites[14,15]. Next, we will test whether this model aptly describes ETS-AP1 partnerships in cells.

First, we will explore the preferences of ETS factor binding with JUN-FOS by genomic occupancy experiments. Genome-wide analyses of ERG, ETV1, and ETV4 following the overexpression of these factors in normal prostate cells have been performed, and these factors occupy regions of the genome that are enriched with ETS-AP1 composite sites[14]. EHF and SPDEF binding to select

ETS-AP1 composite sites in normal prostate cells has been described, but not defined in a genome-wide manner[14,15]. ChIP-seq experiments for EHF/SPDEF in normal prostate cells would first describe the prevalence of EHF/SPDEF binding to ETS-AP1 composite sites in the normal prostate. Next, overexpression of ERG in conjunction with ChIP experiments for JUN or FOS would determine whether the occupancy of JUN-FOS is affected by specific ETS factors, as predicted by our *in vitro* experiments.

Next, we will test whether the DNA-binding intrinsic to the ETS domain, the antagonism of JUN-FOS binding nearby, or both features, are important for EHF and ERG in cells. The differential binding to nonoptimal ETS DNA sites and the differential binding with JUN-FOS appear to be separate features of ERG and EHF: the difference in DNA affinity resides within the ETS domain of both factors; whereas, the antagonism of JUN-FOS binding occurs outside of the ETS domain of EHF. We could design mutations that will interconvert each, or both, of these differential properties between ERG and EHF, then test in the genomic occupancy experiments.

In the future, we would plan to move closer to studying expression of ETS-AP1 regulated genes, either by surveying changes in RNA expression of genes associated with genomic-occupancy or by use of cell migration as a proxy.

Development of assays to screen for inhibitors of ETS partnerships

Transcription factors remain difficult drug targets. Small ideal drug-like molecules are a fraction of the size of the large and broad surfaces that are

typical of transcription factor – DNA interactions[64]. Further understanding of the mechanisms through which transcription factors operate provides alternative target sites for the therapeutic disruption of transcription factor function. If either the ETV4 - MED25 interaction or the ERG - JUN-FOS partnership appears to be functionally important for prostate cancer, as determined by the experiments described above, this would motivate the development of inhibitors of these partnerships. The assays described in Chapter 2 could be modified to incorporate either of these components (**Fig. 6.3**). For example, ETV4 would be added to fluorescein-tagged DNA to increase the fluorescence polarization of the DNA. MED25 would interact with ETV4 and further increase the size of the DNA-bound complex and therefore the fluorescence polarization signal. Efficient inhibitors of the ETV4 - MED25 interaction would reduce the fluorescence polarization signal back to level of the ETV4 – DNA complex. A conceptually similar assay could be designed for ERG and JUN-FOS binding to an ETS-AP1 composite DNA site. These modified assays could then be used to perform high-throughput screens for inhibitors of these interactions.

Finale

The development of small molecule inhibitors for transcription factors against the macromolecular contacts that these proteins form is a daunting challenge. However, the disease relevance of ETS factors, as well as many other transcription factors, necessitates the development of novel strategies for the therapeutic perturbation of these proteins' functions[65,66]. Here, I have described

the DNA-binding autoinhibition of ETV4 and its interaction with MED25, both features that appear to be unique to the ETV1/4/5 subclass of ETS factors. Additionally, I have demonstrated distinct partnerships with JUN-FOS transcription factors between different subclasses of ETS factors. I propose that these findings provide a framework for novel routes to specifically perturb the transcriptional activity of individual subclasses of ETS factors.

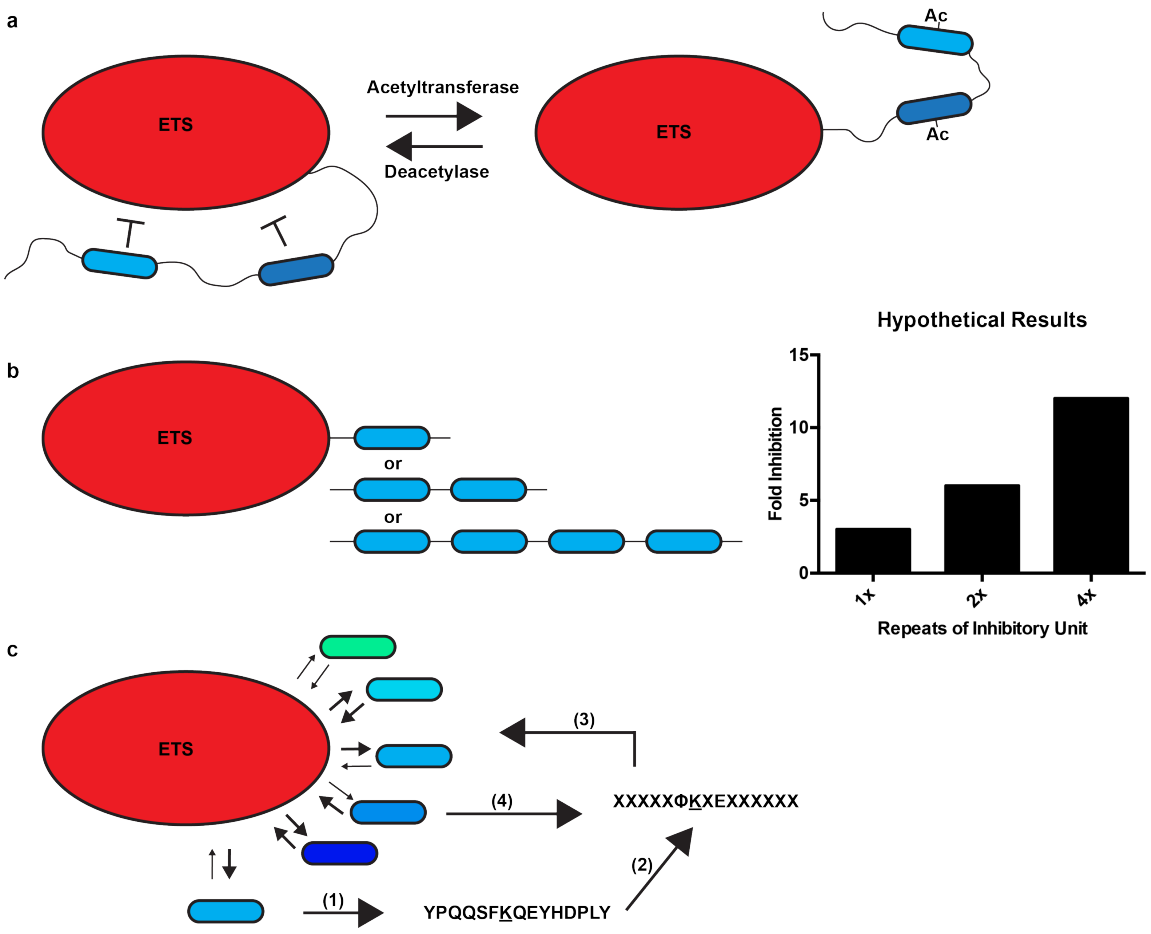


Figure 6.1. Further investigation of the ETV4 N-terminal inhibitory domain (NID). (**a**) Model for NID-mediated autoinhibition. Two separate regions of the NID (cyan ovals) inhibit DNA-binding by the ETS domain. Acetylation of lysine residues within these regions disrupts autoinhibition. (**b**) To facilitate further analysis of autoinhibition within the NID, I propose attaching single or multiple copies of the conserved region surrounding lysine 226. I would expect that the more copies that are present, the higher the level of DNA-binding inhibition would be (right). This more robust level of autoinhibition would enable easier dissection of the inhibitory region through point mutagenesis. (**c**) Definition of the minimal inhibitory unit would allow for affinity maturation of the interaction between this inhibitory unit and the ETS domain of ETV4. (1,2) The conserved region surrounding lysine 226 would be randomly mutagenized. In this case, I designed an experimental setup that would conserve the flanking sequences (ΦKXE) that allow for dual acetylation and sumolyation of lysine 226. (3) Interaction with the ETS domain of ETV4 would be used to select inhibitory sequences with higher affinity from the mutagenized library. (4) Subsequent rounds of mutagenesis are used while fixing residues that favor ETV4 interaction. A control ETS domain, from EHF for example, could be used to select for specificity as well as affinity.

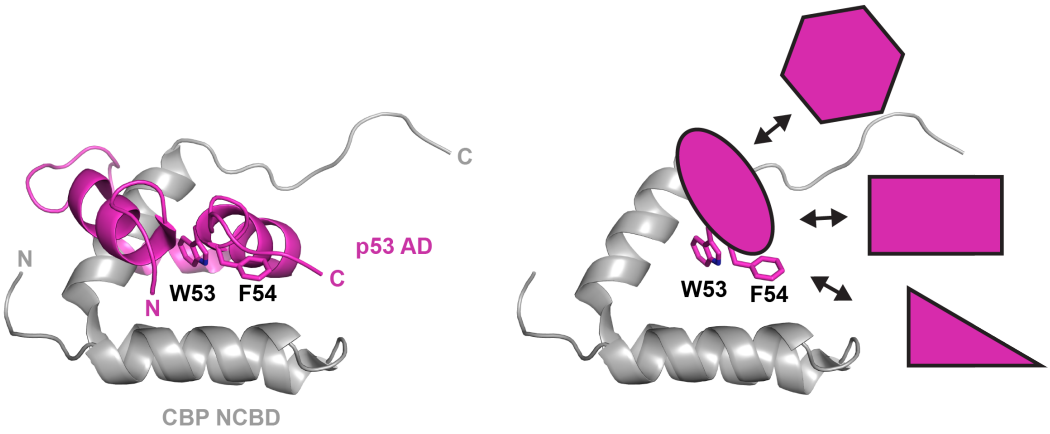


Figure 6.2. Domain grafting as a method for increasing the specificity of an activation domain for a particular coactivator target**.** Left, the structure of the activation domain of p53 (magenta) bound to the NCBD domain of CBP (gray)[53] (2L14.pdb). W53 and F54 of the p53 AD are labeled and shown in stick format. These residues are critical for interaction with CBP NCBD. Right, after optimizing the orientation of the side chains of W53 and F54 for interaction with CBP NCBD, the incorporation of these amino acids into a library of protein domains (magenta polygons) is performed in silico to predict which domain(s) would favor interaction with CBP NCBD, in a process termed domain grafting. The resulting domain grafts could be analyzed for specific interaction with CBP NCBD, compared to other targets of p53 AD, and further optimized using the affinity maturation process described that was previously described (Fig. 6.1c).

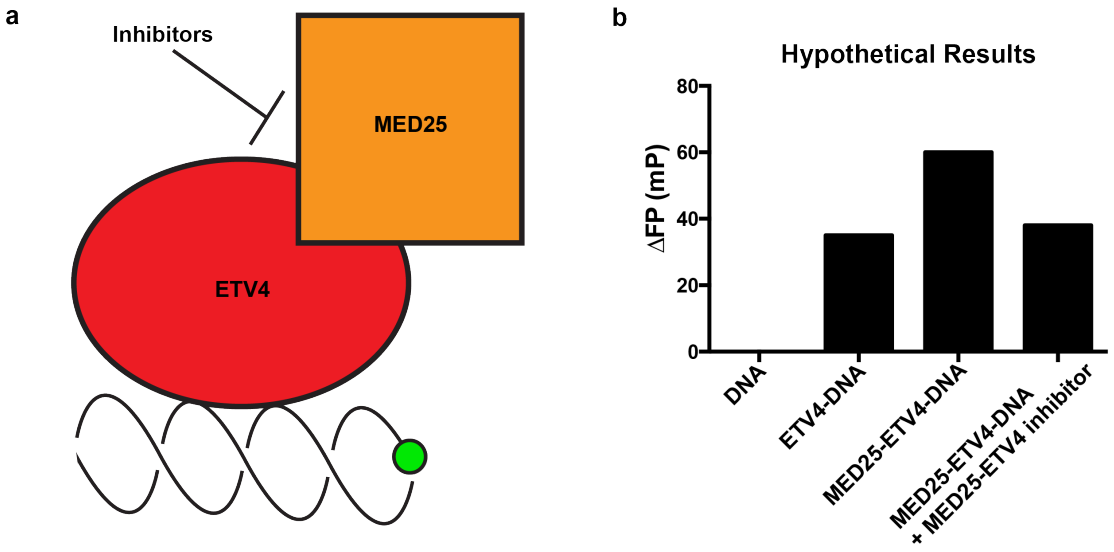


Figure 6.3 Hypothetical assay to screen for inhibitors of ETV4 – MED25 interaction. (**a**) Schematic of fluorescence polarization assay with fluorescein-labeled DNA, ETV4, and MED25. (**b**) Predicted results for an inhibitor of ETV4-MED25 interaction. Addition of ETV4 increases the fluorescence polarization of the DNA (ETV4-DNA), and interaction between MED25 and ETV4 further increases the fluorescence polarization of the DNA (MED25-ETV4-DNA). Efficient inhibition of the MED25-ETV4 interaction would return the fluorescence polarization signal to near ETV4-DNA levels.

References

1. Regan, M.C. et al. Structural and dynamic studies of the transcription factor ERG reveal DNA binding is allosterically autoinhibited. *Proc Natl Acad Sci U S A* **110**, 13374-9 (2013).

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

3. Greenall, A., Willingham, N., Cheung, E., Boam, D.S. & Sharrocks, A.D. DNA binding by the ETS-domain transcription factor PEA3 is regulated by intramolecular and intermolecular protein.protein interactions. *J Biol Chem* **276**, 16207-15 (2001).

4. Bojovic, B.B. & Hassell, J.A. The PEA3 Ets transcription factor comprises multiple domains that regulate transactivation and DNA binding. *J Biol Chem* **276**, 4509-21 (2001).

5. Tomlins, S.A. et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-8 (2005).

6. Pufall, M.A. et al. Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* **309**, 142-5 (2005).

7. Green, S.M., Coyne, H.J., 3rd, McIntosh, L.P. & Graves, B.J. DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J Biol Chem* **285**, 18496-504 (2010).

8. Coyne, H.J., 3rd et al. Autoinhibition of ETV6 (TEL) DNA binding: appended helices sterically block the ETS domain. *J Mol Biol* **421**, 67-84 (2012).

9. Goetz, T.L., Gu, T.L., Speck, N.A. & Graves, B.J. Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. *Mol Cell Biol* **20**, 81-90 (2000).

10. Garvie, C.W., Pufall, M.A., Graves, B.J. & Wolberger, C. Structural analysis of the autoinhibition of Ets-1 and its role in protein partnerships. *J Biol Chem* **277**, 45529-36 (2002).

11. Verger, A. et al. The Mediator complex subunit MED25 is targeted by the N-terminal transactivation domain of the PEA3 group members. *Nucleic Acids Res* **41**, 4847-59 (2013).

12. Milbradt, A.G. et al. Structure of the VP16 transactivator target in the Mediator. *Nat Struct Mol Biol* **18**, 410-5 (2011).

13. Vojnic, E. et al. Structure and VP16 binding of the Mediator Med25 activator interaction domain. *Nat Struct Mol Biol* **18**, 404-9 (2011).

14. Hollenhorst, P.C. et al. Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes Dev* **25**, 2147-57 (2011).

15. Tugores, A. et al. The epithelium-specific ETS protein EHF/ESE-3 is a context-dependent transcriptional repressor downstream of MAPK signaling cascades. *J Biol Chem* **276**, 20397-406 (2001).

16. Desjardins, G. et al. Synergy of aromatic residues and phosphoserines within the intrinsically disordered DNA-binding inhibitory elements of the Ets-1 transcription factor. *Proc Natl Acad Sci U S A* **111**, 11019-24 (2014).

17. Whitehead, T.A., Baker, D. & Fleishman, S.J. Computational design of novel protein binders and experimental affinity maturation. *Methods Enzymol* **523**, 1-19 (2013).

18. Arkin, M.R., Tang, Y. & Wells, J.A. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem Biol* **21**, 1102-14 (2014).

19. Huang, H. & Sidhu, S.S. Studying binding specificities of peptide recognition modules by high-throughput phage display selections. *Methods Mol Biol* **781**, 87-97 (2011).

20. Procko, E. et al. A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. *Cell* **157**, 1644-56 (2014).

21. Rettenmaier, T.J. et al. A small-molecule mimic of a peptide docking motif inhibits the protein kinase PDK1. *Proc Natl Acad Sci U S A* **111**, 18590-5 (2014).

22. Aytes, A. et al. ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. *Proc Natl Acad Sci U S A* **110**, E3506-15 (2013).

23. Hollenhorst, P.C., Paul, L., Ferris, M.W. & Graves, B.J. The ETS gene ETV4 is required for anchorage-independent growth and a cell proliferation gene expression program in PC3 prostate cells. *Genes Cancer* **1**, 1044-1052 (2011).

24. Defossez, P.A., Baert, J.L., Monnot, M. & de Launoit, Y. The ETS family member ERM contains an alpha-helical acidic activation domain that contacts TAFII60. *Nucleic Acids Res* **25**, 4455-63 (1997).

25. Goel, A. & Janknecht, R. Acetylation-mediated transcriptional activation of the ETS protein ER81 by p300, P/CAF, and HER2/Neu. *Mol Cell Biol* **23**, 6243-54 (2003).

26. Landrieu, I. et al. Characterization of ERM transactivation domain binding to the ACID/PTOV domain of the Mediator subunit MED25. *Nucleic Acids Res* **43**, 7110-21 (2015).

27. Hollenhorst, P.C., Jones, D.A. & Graves, B.J. Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. *Nucleic Acids Res* **32**, 5693-702 (2004).

28. Baena, E. et al. ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. *Genes Dev* **27**, 683-98 (2013).

29. Chen, Y. et al. ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nat Med* **19**, 1023-9 (2013).

30. Whyte, W.A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-19 (2013).

31. Loven, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-34 (2013).

32. Yang, H. et al. ETS family transcriptional regulators drive chromatin dynamics and malignancy in squamous cell carcinomas. *Elife* **4**(2015).

33. Khan, A. & Zhang, X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res* **44**, D164-71 (2016).

34. Stevens, J.L. et al. Transcription control by E1A and MAP kinase pathway via Sur2 mediator subunit. *Science* **296**, 755-8 (2002).

35. Balamotis, M.A. et al. Complexity in transcription control at the activation domain-mediator interface. *Sci Signal* **2**, ra20 (2009).

36. Asada, S. et al. External control of Her2 expression and cancer cell growth by targeting a Ras-linked coactivator. *Proc Natl Acad Sci U S A* **99**, 12747-52 (2002).

37. Cangemi, R. et al. Reduced expression and tumor suppressor function of the ETS transcription factor ESE-3 in prostate cancer. *Oncogene* **27**, 2877-85 (2008).

38. Yin, J.W. et al. Mediator MED23 plays opposing roles in directing smooth muscle cell and adipocyte differentiation. *Genes Dev* **26**, 2192-205 (2012).

39. Nakamura, Y. et al. Wwp2 is essential for palatogenesis mediated by the interaction between Sox9 and mediator subunit 25. *Nat Commun* **2**, 251 (2011).

40. Taatjes, D.J., Naar, A.M., Andel, F., 3rd, Nogales, E. & Tjian, R. Structure, function, and activator-induced conformations of the CRSP coactivator. *Science* **295**, 1058-62 (2002).

41. Taatjes, D.J. & Tjian, R. Structure and function of CRSP/Med2; a promoter-selective transcriptional coactivator complex. *Mol Cell* **14**, 675-83 (2004).

42. Meyer, K.D., Lin, S.C., Bernecky, C., Gao, Y. & Taatjes, D.J. p53 activates transcription by directing structural shifts in Mediator. *Nat Struct Mol Biol* **17**, 753-60 (2010).

43. Ebmeier, C.C. & Taatjes, D.J. Activator-Mediator binding regulates Mediator-cofactor interactions. *Proc Natl Acad Sci U S A* **107**, 11283-8 (2010).

44. Ito, M. et al. Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol Cell* **3**, 361-70 (1999).

45. Hittelman, A.B., Burakov, D., Iniguez-Lluhi, J.A., Freedman, L.P. & Garabedian, M.J. Differential regulation of glucocorticoid receptor transcriptional activation via AF-1-associated proteins. *EMBO J* **18**, 5380-8 (1999).

46. Lau, J.F., Nusinzon, I., Burakov, D., Freedman, L.P. & Horvath, C.M. Role of metazoan mediator proteins in interferon-responsive transcription. *Mol Cell Biol* **23**, 620-8 (2003).

47. Chen, W., Rogatsky, I. & Garabedian, M.J. MED14 and MED1 differentially regulate target-specific gene activation by the glucocorticoid receptor. *Mol Endocrinol* **20**, 560-72 (2006).

48. Galbraith, M.D. et al. ERK phosphorylation of MED14 in promoter complexes during mitogen-induced gene activation by Elk-1. *Nucleic Acids Res* **41**, 10241-53 (2013).

49. Sela, D. et al. Role for human mediator subunit MED25 in recruitment of mediator to promoters by endoplasmic reticulum stress-responsive transcription factor ATF6alpha. *J Biol Chem* **288**, 26179-87 (2013).

50. Warfield, L., Tuttle, L.M., Pacheco, D., Klevit, R.E. & Hahn, S. A sequence-specific transcription activator motif and powerful synthetic

variants that bind Mediator using a fuzzy protein interface. *Proc Natl Acad Sci U S A* **111**, E3506-13 (2014).

51. Dyson, H.J. & Wright, P.E. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **12**, 54-60 (2002).

52. Wright, P.E. & Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* **16**, 18-29 (2015).

53. Lee, C.W., Martinez-Yamout, M.A., Dyson, H.J. & Wright, P.E. Structure of the p53 transactivation domain in complex with the nuclear receptor coactivator binding domain of CREB binding protein. *Biochemistry* **49**, 9964-71 (2010).

54. Bochkareva, E. et al. Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. *Proc Natl Acad Sci U S A* **102**, 15412-7 (2005).

55. Rowell, J.P., Simpson, K.L., Stott, K., Watson, M. & Thomas, J.O. HMGB1-facilitated p53 DNA binding occurs via HMG-Box/p53 transactivation domain interaction, regulated by the acidic tail. *Structure* **20**, 2014-24 (2012).

56. Okuda, M. & Nishimura, Y. Extended string binding mode of the phosphorylated transactivation domain of tumor suppressor p53. *J Am Chem Soc* **136**, 14143-52 (2014).

57. Krois, A.S., Ferreon, J.C., Martinez-Yamout, M.A., Dyson, H.J. & Wright, P.E. Recognition of the disordered p53 transactivation domain by the transcriptional adapter zinc finger domains of CREB-binding protein. *Proc Natl Acad Sci U S A* **113**, E1853-62 (2016).

58. Brady, C.A. et al. Distinct p53 transcriptional programs dictate acute DNA-damage responses and tumor suppression. *Cell* **145**, 571-83 (2011).

59. Fleishman, S.J. et al. Hotspot-centric de novo design of protein binders. *J Mol Biol* **413**, 1047-62 (2011).

60. Fleishman, S.J. et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816-21 (2011).

61. Arai, M., Ferreon, J.C. & Wright, P.E. Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *J Am Chem Soc* **134**, 3792-803 (2012).

62. Di Lello, P. et al. Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53. *Mol Cell* **22**, 731-40 (2006).

63. Zhu, J. et al. Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature* **525**, 206-11 (2015).

64. Koehler, A.N. A complex task? Direct modulation of transcription factors with small molecules. *Curr Opin Chem Biol* **14**, 331-40 (2010).

65. Wang, S. et al. Ablation of the oncogenic transcription factor ERG by deubiquitinase inhibition in prostate cancer. *Proc Natl Acad Sci U S A* **111**, 4251-6 (2014).

66. Illendula, A. et al. Chemical biology. A small-molecule inhibitor of the aberrant transcription factor CBFbeta-SMMHC delays leukemia in mice. *Science* **347**, 779-84 (2015).