THE UNIVERSITY OF CHICAGO

POLYMERIZATION OF THE ETS FAMILY TRANSCRIPTIONAL REPRESSOR YAN: THEORY AND EXPERIMENT

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOPHYSICS

BY

C. MATTHEW HOPE

CHICAGO, ILLINOIS

DECEMBER 2017

For Andrew, who the world knows as a scholar of $Y_1Ba_2Cu_3O_{7-\delta}$ crystals with enhanced flux pinning properties, and who I know as the brother who showed me the value of pursuing your dreams with courage

Table of Contents

List of Figures

Abstract

In order to specify the many cell types of an organism, transcription factors must form transcriptional regulatory complexes at enhancers to control the expression of genes that determine cell fate. These complexes are essential to the spatiotemporal specificity of gene regulation during both development and disease; yet molecular understanding of their exact composition and operation is limited. Both protein-protein and protein-DNA interactions among the transcription factors bound to enhancer DNA will determine the stability and regulatory meaning of a given complex. Here, we use the polymerizing *Drosophila* ETS family transcriptional repressor Yan as a model transcription factor to explore the impact of altering the affinity of protein-protein interactions on the outcome of gene regulation. To do so, we developed a model of Yan binding to enhancer elements at equilibrium, which depends on a minimal number of interactions including polymerization. In the process, we discovered that the model reproduces features of Yan DNA binding *in vivo*, and show that the polymerization of Yan highlights a trade-off between binding affinity and specificity, with implications for Yan as a transcriptional repressor. Furthermore, the model predicted a role for medium-affinity polymerization in maintaining Yan function, with increased polymerization beyond Yan wild type levels resulting in loss of function. To test this, we employed an engineering approach based on structural and sequence-homology to create mutants of Yan with increased self-association affinity over four orders of magnitude. These mutants result in loss of function phenotypes proportional to their strength of polymerization, and result in a dramatic restructuring of the subcellular localization of Yan. Additionally, examination of specific cell fate defects in the *Drosophila* eye revealed a role for regulation of

polymerization in the execution of the various regulatory decisions Yan makes throughout development. Taken together, this work broadens the picture of the implications of Yan polymerization, and establishes a firm role for balanced protein-protein and protein-DNA interactions in the transcriptional regulation of development.

Chapter 1: Introduction and Background

C. Matthew Hope

1.1 Transcription regulatory complexes drive multicellular development

1.1.1 Development is a progressive process of specifying cell fate

Development in multicellular animals is a process of executing a profoundly simple principle — "one cell becomes one organism"— in a manner that belies the complexity of gene regulation required to do so. Although there are many phenomena encompassed in the study of development that have fascinated biologists for more than a century, at its core, it is a process of progressive cellular decision-making. The fundamental problem is this: every cell in an organism has a copy of the same genome, yet every cell must adopt a final cell fate out a myriad number to choose from. Simply put, how can cells 'know' what they should become, and how can one get different outcomes from the same set of instructions?

The solution to this problem relies on gene-specific regulation of the information contained within DNA. Although the plans for every cell type are available to any given cell, the appropriate genes are activated or repressed by intricate control mechanisms. By regulating the process of transcription to turn on expression of one gene but not another, these regulatory mechanisms form the molecular underpinning of how cells make their cell fate decisions. With the ability to differentiate one outcome from another, many decisions can then be linked together to create an iterative process by which cells move from a pluripotent state to progressively more restricted states, until arriving at their terminally differentiated state.

Although regulating transcription allows cells to make critical developmental choices, the precise molecular nature of the complexes that do so remain poorly understood. In order to meaningfully orchestrate development, transcriptional regulatory

complexes have to achieve some simple (yet profound) requirements at the broadest levels. First, they must be able to distinguish different genes in order to regulate them appropriately— activating or repressing all genes indiscriminately will never allow for a specific cell fate choice. Secondly, they must transmit regulatory information from the DNA to RNA polymerase in order to exert some measurable effect on transcription, either positively or negatively. And lastly, transcriptional regulatory complexes must operate with spatiotemporal specificity, to take signaling cues from the entire organism in order to regulate the correct gene in the correct cell at the correct time. Though these principles are broad, decades of research in model organisms have shown their applicability across organisms, tissues, and contexts, including for human development and disease; therefore, greater molecular understanding of the action of transcriptional regulatory complexes remains a key project of molecular biology.

1.1.2 Transcription factors link cell signaling pathways to cis*-regulatory elements*

To regulate transcription during development, there are three broad components that collaborate to transmit the flow of information from outside the cell to the ultimate outcome inside the nucleus. At the center of the action of transcriptional regulatory complexes are sequence-specific transcription factors (TFs), which by definition are proteins that can read specific DNA sequences and interact with other transcriptional regulators in order to exert an effect on transcription. TFs were first discovered in viruses and prokaryotes with the isolation of phage lambda and E. coli *lac* repressors respectively (Ptashne 1967; Gilbert, Walter and Müller-Hill 1966), which embodied the earlier theoretical predictions of Jacob and Monod for elements that can regulate gene expression (Jacob and Monod 1961). Later, TFs were shown to operate in eukaryotes

(Dynan and Tjian 1983), and with the completion of the sequencing of the human genome, it is now clear that TFs are pervasive, with roughly 2000 TFs in humans (Vaquerizas et al. 2009; Lander 2001), of many different families (Luscombe et al. 2000). These proteins are responsible for driving development and play an outsized role in disease (Boyadjiev and Jabs 2001), especially cancer (Furney et al. 2006).

Cell-to-cell signaling pathways play an equally important role in development, for the simple reason that cells require external cues to determine where they are spatiotemporally, either in the context of a tissue or an entire developing organism. Although there are many important signaling pathways, some especially critical ones for development include RTK, Notch, Wnt, BMP, JAK/STAT, Shh, and nuclear receptors. All seven of these pathways are conserved throughout bilaterians (Pires-daSilva and Sommer 2003), with some being conserved across metazoa (Loh et al. 2016). Signaling molecules can be either secreted or membrane-bound, and signaling pathways are used iteratively throughout development. Critically, the information that is external to the cell must be transmitted internally in order to affect a change in transcription. Therefore, ultimately, signaling pathways must impinge upon TFs, with each of these pathways having dedicated, signaling-responsive TFs that mediate signaling information. Although signaling-responsive TFs are not the only TFs utilized during development, their ability to convey information outside of the cell increases their importance in properly regulating gene expression.

The final component in transcriptional regulation are sequences of DNA known as *cis*-regulatory elements, or enhancers. For clarity, the term "*cis*-regulatory elements" includes many types of elements, including enhancers, insulators, and repressive

elements. Enhancers, therefore, are a subset of *cis*-regulatory elements, but the terms are often used interchangeably. These are pieces of DNA that do not code for protein-coding genes, but instead contain clustered TF binding sites that can regulate gene expression of nearby protein-coding genes in a spatiotemporal pattern. Early examples of gene regulation in prokaryotes, such as the aforementioned lac operon (Gilbert, Walter and Müller-Hill 1966) and the discovery of sigma factors (BURGESS et al. 1969), established a paradigm of regulating transcription from promoter proximal regions. Therefore, it came as a surprise when an enhancer within the simian virus SV40 was shown to activate gene expression of mammalian beta-globin genes at a distance of over 10kb (Banerji et al. 1981). It is now clear that long-range regulation of gene expression by enhancers is one of the key hallmarks of eukaryotic transcriptional regulation. Enhancers function by serving as templates for the binding of groups of TFs and then forming loops to bring these complexes into close physical proximity with promoters to activate transcription. The critical importance of enhancers in development was later shown for eve stripe 2 (Small et al. 1992).

Taken together, transcriptional regulation during development is a flow of information, with TFs playing a key role to bridge the receipt of input in the form of cell-to-cell signals to transcriptional output through binding to *cis*-regulatory elements. A nice illustrative example of this paradigm at work comes from R7 photoreceptor specification in the *Drosophila* eye, which is one of the earliest developmental contexts to molecularly describe the linkages among all three components and connect it to a cell fate choice. Briefly, the adult compound eye of Drosophila is composed of 700-800 units called ommatidia, each of which has a stereotypic number of cell types, including eight light-

sensing photoreceptors, four cone cells, and accessory pigment and bristle cells (Tomlinson and Ready 1987). The eye was an important model system for establishing the role of signaling pathways in development because histological sectioning established the repeating structure and cell types of each facet (Waddington and Perry 1960) and clonal analysis established that cell types were not specified by lineage, but instead relied upon inductive signaling (Hotta and Benzer 1970; Ready et al. 1976). Genetic screens for mutants that failed to respond to UV light uncovered a gene called *sevenless* that is required for specifying UV-sensitive R7 photoreceptors in each ommatidium [Tomlinson and Ready 1986 Science]. When molecularly cloned, *sev* was shown to encode a Receptor Tyrosine Kinase (RTK) (Hafen et al. 1987; U Banerjee et al. 1987; Utpal Banerjee et al. 1987), and soon after, the corresponding ligand, *bride of sevenless,* was identified (Reinke and Zipursky 1988), thus establishing the core extracellular components of the signaling pathway that promotes R7 fate. The use of further genetic screens that enhance or suppress the *sevenless* phenotype uncovered other extracellular signaling pathways involved in R7 specification, namely Epidermal Growth Factor Receptor (EGFR) and Notch(Nagaraj and Banerjee 2004), and also lead to the identification of a cascade of intracellular factors that transduce the extracellular signals from RTK signaling (briefly: Ras, Sos, Raf, Drk, Ksr, Mek, and rolled (Rubin et al. 1997)). These components form a conserved RTK signaling cassette, the power of which had been previously appreciated from studies of oncogenic transformation of vertebrate cell lines, but came into focus as a mechanism to transmit extracellular signals from an RTK to the small GTPase Ras, and then to a series of phosophorylation events that

ultimately result in activation of Mitogen Activated Protein Kinase (MAPK) (Rubin et al. 1997).

Upon elucidating this signaling cascade, it quickly became clear that the signals from Sevenless and other pathways had to be transmitted to TFs in order to carry out changes in transcription. Eventually this lead to the identification of two E-26 Specific (ETS) family TFs called Yan and Pointed (Lai and Rubin 1992; O'Neill et al. 1994; Klämbt 1993) that respond to activated MAPK, negatively and positively respectively, and also bind to targets throughout development that are downstream of RTK signaling. One such target was shown to be *prospero*, a homeodomain TF (Chu-Lagraff et al. 1991) required for R7 fate, and subsequent work identified an enhancer at the *prospero* locus that responds to both Yan and Pointed as well as other TFs (Xu et al. 2000), thus completing the linkage from signal to TF to regulatory element. It should be emphasized that the uncovering of this paradigm at play in *Drosophila* R7 development was complemented by parallel discoveries in other fields, notably *C. elegans* and mammalian cell lines, and the key discoveries in these three areas cemented the importance of a common developmental logic for specifying cell fate.

1.1.3 Transcription factors form regulatory complexes to specify cell fate

Since the establishment of the above paradigm for development, there have been great strides in uncovering the commonalities and complexities of transcriptional regulation in multi-cellular animals. For instance, the concept of individual TFs acting upon a target gene has given way to the idea of transcriptional regulatory complexes. These complexes are defined by multiple TFs bound at a *cis*-regulatory element, such as

an enhancer, that regulate gene expression. Importantly, the interactions among TFs are just as important as the interactions between the TFs and the enhancer DNA.

Transcriptional regulatory complexes are a natural extension of the principle of combinatorial control of development. Due to the fact that there are a handful of signaling pathways that specify cell fate, but many cell fates to be specified, using a combination of cell signaling inputs (instead of a one-signal, one-fate paradigm) expands the complexity needed to pattern an entire organism. Therefore, signaling pathways are used iteratively throughout development and their outcomes, in terms of specification, are highly context dependent. A correlated idea is that in order to get tight spatial boundaries to pattern cells from signaling molecules that can diffuse widely throughout a tissue, combinatorial control may allow for specification to occur at the intersection of otherwise broad gradients.

The molecular reflection of combinatorial control by cell signaling pathways is the fact that cell-signaling dependent TFs converge at enhancers to form transcriptional regulatory complexes. In this way, the presence of a TF in the complex represents the input received from its dedicated signaling pathway, and the combination of multiple TFs stabilizes the complex to allow the recruitment of co-factors to effect transcription. This principle was nicely encapsulated by the near-simultaneous discovery of three developmental enhancers in *Drosophila,* the aforementioned *prospero* enhancer, an enhancer of dPax2 (Fu et al. 1997), and the muscle heart enhancer that regulates *eve* in the embryo (Halfon et al. 2000). All three require input from multiple signaling pathways and require the binding of multiple signaling-dependent TFs in transcriptional regulatory complexes to induce target gene expression.

1.2 Protein-protein interactions and protein-DNA interactions control TF behavior

1.2.1 The role of sequence specific and sequence non-specific DNA interactions

The most important role of protein-DNA interactions in TFs is to recognize specific sequences of DNA nucleotides, and this specificity is achieved through a remarkable number of structures and strategies (Rohs et al. 2010). The first structures of TFs to be co-crystallized with DNA—including CAP (McKay and Steitz 1981), lambda repressor (Pabo and Lewis 1982), and lambda Cro (Anderson et al. 1981)— suggested a remarkably simple model of one-to-one interactions between amino acid sidechains and basepairs, although the subsequent explosion in structures of DNA-bound proteins has shown the simple model to not be the case. Instead the complex principles of molecular recognition still hold for DNA-binding proteins, which rely on electrostatic interactions to the negatively charged sugar-phosphate DNA backbone, van der Waals packing, steric recognition, and hydrogen bonding between amino acids and nucleotides to form protein-DNA complexes (Garvie and Wolberger 2001). Within a sequence-specific interaction, there are often multiple modes of recognition at play, which loosely divide into two categories: base-specific contacts which are termed direct readout, and recognition of the shape of the DNA, especially the dimensions of the major and minor grooves which is called indirect readout (Garvie and Wolberger 2001).

Many types of protein folds can recognize DNA in a sequence specific fashion, and although there are too many to describe in detail, a few folds are used extensively by developmental TFs and should be highlighted. One of the most prominent in humans is

the helix-turn-helix (HTH) domain family, which subdivides into classic HTHs such as homeodomains, and winged helix-turn-helix domains (wHTH) which includes ETS domains and Forkhead domains (Hollenhorst et al. 2011). This family is defined by a right-handed three helix bundle, with a characteristic 'recognition' helix that inserts into the major groove of DNA to form base specific contacts. The other helices in the core motif also provide sequence specificity, position the recognition helix, and stabilize the fold of the bundle. The largest family of DNA binding proteins in humans is zinc finger proteins, which include Cys2-His2 zinc fingers and Cys2-Cys2 zinc fingers which bind as individual subunits, as well as Zn2-Cys6 zinc fingers that function as obligate dimers (Razin et al. 2012). All zinc finger proteins are defined by their coordination of a zinc cations, which due to the absence of extensive hydrophobic residues to form the fold, use the binding of zinc to hold together the small domain. In the case of Cys2-His2 and Cys2-Cys2 zinc finger proteins, multiple finger subunits are often chained together within a given protein, with each domain imparting 2-3 basepairs of sequence specificity, although the sequence interactions among domains can be complex in the context of the full length TF. Lastly, nuclear receptors comprise a very large class of eukaryotic TFs, with over 150 members in humans. Nuclear receptors usually function as dimers, although a few have been shown to function as monomers, and frequently the binding of a small molecule signal influences their dimerization and function (Bain et al. 2007).

1.2.2 The transcription factor search problem

Eukaryotic genomes can contain billions of basepairs of DNA, yet to regulate specific genes during development, TFs must discriminate the correct locations to build transcriptional regulatory complexes from the much larger milieu of off-target sites. The

number of non-sequence specific sites to potentially bind will be much larger than the number of sequence-specific sites, and of the sequence-specific sites, only a subset will be functional to regulation of transcription. Surprisingly, TFs are capable of binding their sites to regulate genes in a manner that is faster than would be expected from three dimensional diffusion alone (Otto G. Berg and von Hippel 1985; O G Berg et al. 1981), and further study has shed light upon the precise mechanisms by which TFs search the genome.

Three predominate models of TF search include one-dimensional sliding along DNA, hopping, and looping (also known as intersegmental transfer) (Smith and Matthews 2016). It should be noted that all of these models require both sequence specific binding as well as sequence non-specific binding of a TF to DNA, and are not necessarily mutually exclusive. One-dimensional sliding requires weak interactions between the TF and non-specific DNA, which will mostly be mediated by electrostatic interactions between the negatively charged DNA phosphate backbone and basic sidechains of the TF (Winkler et al. 1993; Viadiu and Aggarwal 2000; Kalodimos 2004; Givaty and Levy 2009). It has been theoretically shown that by reducing the dimensionality of search from three dimensions to one, the association rate at which TFs can apprehend their target sites increases greatly (Halford and Marko 2004). Sliding has been observed in both prokaryotic and eukaryotic TFs— for example, the canonical LacI repressor (Mahmutovic et al. 2015; Elf et al. 2007) as well as the Hox family homeodomains (Vuzman et al. 2010) — and there is mixed evidence if sliding TFs can circumvent obstacles along DNA, such as nucleosomes in eukaryotes, without completely

dissociating (i.e. engaging in 'hopping').(Gorman and Greene 2008; Hedglin and O'Brien 2008; Pluciennik and Modrich 2007).

The two additional modes of TF search are hopping and looping. Hopping involves a TF temporarily dissociating from the DNA, undergoing three dimensional diffusion for a short time, and then re-associating with DNA in a new location, while looping accomplishes a similar feat, but without dissociating first from the DNA. Hopping by TFs can increase the rate of locating a specific binding site (Givaty and Levy 2009), and can provide a mechanism to bypass obstacles during sliding. Looping requires multiple interfaces with DNA in order to form a bridging complex between the two DNA segments and the protein complex itself (Doucleff and Clore 2008) and may increase as a search mechanism in situations where DNA is tightly packed. Due to the need to engage multiple interfaces in DNA recognition, TFs may be particularly well suited to utilize multivalent interactions involving several DNA binding domains to facilitate looping (Schmidt et al. 2014). As mentioned above, these modes of facilitated diffusion are not mutually exclusive, and it is likely that most TFs employ all three in order to recognize their targets in the genome, in varying proportions that are likely to be specific to the TF in question (Guardiani et al. 2014).

1.2.3 Protein-protein interactions as a means to recognize target sites

In addition to protein-DNA interactions, transcriptional regulatory complexes utilize protein-proteins interactions in multiple ways to execute regulation of transcription, and broadly speaking, these roles can be subdivided into three categories: by influencing the affinity or selectivity of DNA recognition, by forming regulatory combinations at enhancers, and by recruiting non-sequence specific transcriptional

regulators to control RNA polymerase. The first of these roles for protein-protein interactions mainly relies on dimerization or multimerization of TFs in recognizing their cognate sites on DNA (Funnell and Crossley 2012).

Many TFs require either homo- or hetero-dimerization in order to recognize high affinity DNA binding sites in eukaryotic *cis*-regulatory regions. Prominent families of TFs that require partners for DNA recognition include the aforementioned nuclear receptor superfamily (Germain and Bourguet 2013) as well as basic helix-loop-helix (bHLH) (Garrell and Campuzano 1991), and basic leucine zipper (bZIP) (Vinson et al. 2006) TFs. Dimerization frequently increases the affinity of a TF-DNA interaction versus a monomer with DNA, because more protein-DNA contacts can be made with two molecules, in addition to cooperativity effects from forming a productive protein-protein interaction. By forming a dimer to bind DNA, it is also possible for protein-protein interactions to induce allosteric changes in the two partner binding domain, resulting in more favorable contacts with DNA and an increase in the affinity of protein-DNA interactions. Notable examples of this include the interaction between RUNX1 and CBFβ and the complex formed by FOS/JUN heterodimers and NFAT (L. Chen et al. 1998).

In addition to modulating protein-DNA binding affinity, protein-protein interactions can also modulate sequence specificity of DNA binding, and therefore influence preferences for forming one transcriptional regulatory complex over another. This occurs notably for bZIP family of proteins, where partner selection between two bZIP TFs can determine high affinity for one sequence versus another. For example, the aforementioned bZIP protein JUN can form heterodimers with either FOS or another bZIP protein ATF2, with the sites recognized by each dimer having different sequences

and biological functions (van Dam and Castellazzi 2001). Nuclear receptors are also notably influenced by partner selection (Wilce et al. 2012) with formation of different dimers determining the sequence preferences for binding.

1.2.4 Combinatorial collectives of transcription factors impart enhancer with spatiotemporal specificity

As discussed above, eukaryotic development relies upon iterative use of a small number of cell-to-cell signaling pathways with dedicated signaling-responsive TFs. Given the relatively small number of TFs and the large number of genes to be regulated over the course of development, the combination of TFs at a given enhancer can impart regulatory specificity (termed "combinatorial control"). Multiple TFs acting in concert at enhancers necessarily requires protein-protein interactions among TFs, and combinatorial control is a key role that protein-protein interactions play in the function of transcriptional regulatory complexes (Davidson and Levine 2008; Spitz and Furlong 2012).

Many of the most well-characterized enhancers in developmental biology require inputs from multiple TFs. Examples from *Drosophila* include the *eve* stripe 2 and muscle heart enhancers (MHE), as well as the sparkling enhancer of *dPax2* and aforementioned enhancer of *prospero* in the eye. These enhancers are united in their requirement for different TF binding sites, and furthermore the order and arrangement of binding sites are necessary for proper function because they facilitate or precluded protein-protein interactions between bound TFs. In the case of the MHE, several signaling-responsive TFs converge at this element, including dTCF downstream of Wg signaling, Mad downstream of Dpp signaling, Yan/Pnt downstream of RTK signaling and Twist which specifies mesodermal fate. (Halfon et al. 2000) Thus, the molecular architecture of this

enhancer is encapsulates the principle of combinatorial control, because the cells that will have expression driven by the MHE will necessarily require the activity of these TFs, and thus only cells at the intersection of these signals will be competent to respond.

The action of TFs at enhancers has lead to several models of enhancer function during development, each with varying levels of stringency with respect to the order and arrangement of TF binding sites. Briefly, the billboard model posits that enhancers sum activating and repressive inputs from TFs with little to no regard as to the arrangement of TF binding, with the possibility of maintaining enhancer function while switching one TF for another, as long as the correct levels of input are held constant (Kulkarni 2003). At another other extreme, the enhanceosome model posits a highly rigid and conserved grammar to enhancers, where shifting any TF binding site results in a loss of activity, with a classic example being the interferon-β enhancer (Panne et al. 2007). While there are examples to support either model, it is likely that most developmental enhancers rely on a logic that splits the difference between these two modes, with some arrangements of binding sites within an enhancer being relatively fragile, while others are flexible (Evans et al. 2012). Thorough dissection of the *sparkling* enhancer provided support for both flexibility and rigidity in terms of binding site arrangement, and a new model termed "transcription factor collectives" posits that this is the predominant mode for developmental enhancers (Junion et al. 2012). Ultimately, the requirements for binding site arrangement are likely to be enhancer- and context-specific.

An important feature with respect to binding site arrangement within enhancers is that strict spacing is likely to be enforced by protein-protein interactions between TFs bound at the sites. Thus, by moving the relative spacing between two binding sites, a

protein-protein interaction between TFs may become more or less favorable, with consequences for the affinity of the complex to the enhancer. This explains the surprising observation that TF binding sites undergo rapid rearrangement through evolution while maintaining the same pattern of gene expression (Ludwig et al. 2005)and suggests that models of enhancer activity that focus on conservation of TF-TF interactions may be more predictive of enhancer function.

1.2.5 Protein-protein interactions allow TFs to regulate RNA polymerase

Ultimately, the information contained in the non-coding genome must read by transcriptional regulatory complexes and be transmitted to RNA polymerase II (PolII) in order to affect the transcription of genes during development. Thus, the final role of protein-protein interactions in transcriptional regulatory complexes is to recruit other co-factors and co-regulators that do not contain sequence specificity in order to stimulate transcription.

One of the key control points for regulation during development is at the level of transcription initiation. Years of structural and genetic characterization of has revealed a conserved set of general transcription factors that are required for transcription from all eukaryotic promoters, including TFIIB, TFIID, TFIIE, TFIIF, and TFIIH, and collectively form the eukaryotic pre-initiation complex (Sainsbury et al. 2015). Additionally, the TFIID contains the TATA binding protein (TBP) as well as TBP-associated factors (TAFs), which play key roles in transcriptional regulation from subsets of promoters. Furthermore, Mediator is a large multi-subunit eukaryotic transcriptional complex that provides specificity to certain types of promoters by physically connecting enhancers and promoters, among many other functions (Jeronimo and Robert 2017). In

contrast to early views of the core promoter as merely a TATA box and adjacent DNA, recent work has shown that promoters and promoter-proximal elements are diverse in terms of the transcriptional complexes they will recruit(Vo Ngoc et al. 2017). The formation of any pre-initiation complex necessarily requires extensive protein-DNA and protein-protein interactions, as well as communication between enhancers and core promoters facilitated by protein-protein interactions. Indeed, it is clear that there is biochemical specificity in terms of recognition between enhancers and promoters, with enhancers displaying a preference for certain types of promoters (Butler and Kadonaga 2001).

In addition to complexes formed at core promoters, another key role for TFs is the recruitment of co-activators or co-repressors. These are regulators that rely on the sequence specificity of TFs to be recruited to specific sites in the genome, and then mediate the activation or repression of transcription through several distinct mechanisms. For co-activators recruited by TFs, a common mechanism is acetylation of histones via recruitment histone acetyltransferase activity (HAT), which in turn establishes a more permissive environment for transcription (Galvani and Thiriet 2015). Prominent co-activators for development that possess HAT-activity include the paralogs CREB-Binding-Protein (CBP) and p300 (Ogryzko et al. 1996), the SAGA complex (L. Wang and YR Dent 2014), and the MYST-family acetyltransferase TIP60 complex(Voss and Thomas 2009). Additionally, some co-activators have roles in directly recruiting or stimulating PolII, such as the aforementioned Mediator complex and TFIID, while CBP and components of the SAGA complex may fulfill this mechanism of activation in addition to their roles in modifying histones (Mannervik 2014). Conversely, many co-

repressors repress transcription by recruiting histone deactylase (HDAC) activity, such as the Groucho/TLE family of co-repressors(Turki-Judeh and Courey 2012) and CtBP (Stankiewicz et al. 2014). The precise composition of complexes recruited to a given enhancer is likely to be context-dependent, and to further complicate the picture, the same TF can recruit diverse co-regulators in a context dependent manner (Qi et al. 2008; Nibu et al. 1998; Fang et al. 2006). Protein-protein interactions between TFs and co-regulators are therefore at the center of the function of transcription regulatory complexes, and are an important foundation to understanding transcriptional regulation during development.

1.2.6 *An emerging role for medium-affinity interactions in development*

The role of high affinity binding sites for TFs at *cis*-regulatory elements has long been known, and frequently the function of these sites has been assessed by creating reporter constructs with mutations that reduce TF binding affinity to the level of a non-specific protein-DNA interaction. Recently, there has been renewed interest in the role of low- to medium-affinity TF binding sites in shaping the function of transcriptional regulatory complexes (Crocker et al. 2015; Ramos and Barolo 2013; Parker et al. 2011; Farley et al. 2015). In essence, this is the idea that strong binding of TFs to an enhancer can create a transcriptional regulatory complex that has reduced spatiotemporal specificity or signaling responsiveness, and that through evolution, medium-affinity TF binding sites are under selection to maintain enhancer function. Support for this hypothesis has been found in various contexts across model organisms for quite some time (Driever et al. 1989; Jiang and Levine 1993; Gaudet 2002). For example, Ramos and colleagues discovered that high affinity Gli binding sites from a *patched* enhancer in

Drosophila drive weaker expression than medium-affinity sites when transplanted into other Hedgehog-signaing responsive elements (Ramos and Barolo 2013). Crocker et al reached a similar conclusion in examining the role of medium affinity Hox sites for proper output of enhancers of *shavenbaby* (Crocker et al. 2015). Farley and colleagues expanded this conclusion into the *Ciona* and highlighted a role for sub-optimization in a high-throughput screen of enhancers. It was shown that spacing between ETS and GATA sites is optimal or affinity is high, but placing both together within one construct drives strong ectopic expression of the construct in a variety of tissues (Farley et al. 2015). Taken all together, this work reemphasizes the role that stability of transcriptional regulatory complexes play in gene regulation. It should also be noted that this work focused on the strength of protein-DNA interactions, while the strength of protein-protein interactions remains unexplored.

1.3 ETS family transcription factors

1.3.1 ETS transcription factors are essential to metazoan development

Transcription factors of the E-Twenty-Six specific (ETS) family, mentioned above, have served as a useful window into understanding TF biology in general, as well as uncovering molecular links between TFs, cell-to-cell signaling, and development. They are one of the most important families of TFs in metazoan development, and are exceptionally well characterized genetically, molecularly, and structurally. Furthermore, ETS TFs play an outsized role in disease processes, most notably in hematological malignancies and prostate cancers.

ETS TFs were first discovered in the avian acute leukemia retrovirus E26, and comparison of the V-*ets* oncogene (Nunn et al. 1983) to its cellular homologs in other

species (Leprince et al. 1983)(Watson et al. 1985) lead to the identification of a novel DNA binding domain (also called ETS) that is diagnostic for all ETS family TFs. ETS TFs are present in all metazoan species that have been sequenced to date (Hart et al. 2000), with 8 members in *Drosophila*, 10 in *C. elegans*, and 28 in humans (Findlay et al. 2013). Of the human ETS TFs, the family can be subdivided further into four groups based upon preferences for sequence-specific DNA binding (Wei et al. 2010). In addition to the ETS DNA binding domain found in all ETS TFs by definition, there are other protein-protein interaction domains frequently found in ETS TF family members, which include the sterile alpha motif (SAM) domain (also called the PNT domain, found in 11 human ETS TFs), the OST domain of GABPA, and the B-box of the TCF ETS sub group, ELK1, ELK3, and ELK4 (Hollenhorst et al. 2011).

The requirements of ETS family factors in mammalian biology and development are diverse, but they play primary roles in epithelial tissues, angiogenesis, and differentiation of blood lineages (Findlay et al. 2013). For example, ETV2, FLI1, and TEL (also known as ETV6) all play critical roles in the biology of haematopoietic stem cells, both during development (ETV2) and in adult haematopoiesis (FLI1 and TEL) (Ciau-Uitz et al. 2013). ETV2 collaborates with VEGF pathway member FLK1 in the yolk sac to form hemangioblasts, the common precursor of blood and endothelial lineages, and ETV2 deficient mice consequently do not form these lineages (Lee et al. 2008). In the related process of adult haematopoiesis, both FLI1 and TEL are required for VEGF-independent and -dependent establishment of haematopoietic stem cells (L. C. Wang et al. 1998). Other roles for ETS family factors in blood development include requirements for ETS1 and ELF4 in the development of cells within the T-cell lineage,

including T-cells, natural killer cells, and natural killer T-cells (Muthusamy et al. 1995; Bories et al. 1995; Lacorazza et al. 2002) and a pervasive role for SPI1 in commitment to both the myeloid and lymphoid lineages, affecting B-cells, T-cells, and many myeloid derived cell-types (McKercher et al. 1996; Scott et al. 1997). All together, 11 of the 27 murine ETS family factors show significant embryonic or post-embryonic lethality in knockout models (Findlay et al. 2013), underscoring their important role in mammalian biology.

Curiously, most ETS family factors are pervasively co-expressed, and despite minimal differences in sequence-specificity between ETS family factors, individual TFs within the ETS family have distinct roles and functions (Hollenhorst et al. 2004). This suggests that for a given context, co-occupancy and overlap of ETS family factors is to be expected, but there are regulatory decision points where a certain family member will take the lead. It has been hypothesized that minimal differences in sequence specificity, along with choice of partner protein interactions are capable of achieving the regulatory specificity needed to execute development (Hollenhorst et al. 2011).

In *Drosophila,* there are eight ETS family factors, the best characterized of which included the aforementioned Yan (also called *anterior open*) and Pointed (*pnt*). Yan and Pnt have roles in embryogenesis in controlling specification of pericardial precursors (Halfon 2000), as well as multiple roles in eye development in controlling photoreceptor differentiation. The other ETS family factors include Ets21C, Ets65A, Ets96B, Ets97D, Ets98B, and Eid74EF (T. Chen et al. 1992). Many of these ETS family factors are not well characterized, although some contexts have been established for a few. For example, Ets21C plays a role in *Drosophila* innate immunity (Chambers et al. 2012), Eid74EF is

ecdysone-responsive and has roles in *Drosophila* polytene chromosome biology (Fletcher and Thummel 1995), and Ets97D has roles in oogenesis and egg chamber development (Gajewski and Schulz 1995). Similar to mammalian ETS family factors, *Drosophila* ETS factors display co-expression in various tissues.

1.3.2 Properties of the ETS DNA binding domain

The ETS DNA binding domain is diagnostic of ETS family TFs, and the ETS domains of many mammalian ETS TFs have been structurally and biochemically well characterized. ETS domains recognize a core of 5'-GGAA-3' or 5'-GGAT-3' sequences, although typically nine base pairs centered on this core motif contribute to binding specificity, as part of a larger 12-15 basepair footprint (Hollenhorst et al. 2011). The ETS domain is a winged helix-turn-helix domain (wHTH) with three core alpha helices (H1, H2, and H3) as well as four beta strands that form an anti-parallel sheet (Beta1-4). Like most members of the HTH family, the ETS domain has a core recognition helix, in this case H3, which forms sequence-specific contacts in the major groove with the two guanine nucleotides in the core sequence via two conserved arginine residues. Sequence-specific contacts outside of the core GGAA/T motif are very rare, suggesting that the majority of ETS domain recognition occurs through indirect readout and shape recognition (Szymczyna and Arrowsmith 2000).

Frequently, mammalian ETS domains contain helices appended to the ETS domain that control DNA binding behavior. ETS1 was the first ETS TF where this was shown for this to be the case, where two helices N-terminal to the ETS domain (HI-1 and -2) and two helices C-terminal to the ETS domain (H4 and H5) are subject to phosphorylation and regulate auto-inhibition (Pufall 2005). In the absence of

phosphorylation, HI-1 is unstructured, permitting association with DNA, but upon phosphorylation, the helix fold against the core ETS domain and auto-inhibits DNA binding approximately 50-fold. It has also been shown that ETV6 contains two similar C-terminal helices that auto-inhibit DNA binding (Coyne et al. 2012), and that the stability of the auto-inhibitory helix contributes directly to increase inhibition of binding, with unfolding permitting recognition of high affinity interactions (De et al. 2016). Recently, the DNA binding domains of ETV1, ETV4, and ETV5 have also been shown to contain helices that function by a similar mechanism (Currie et al. 2017). To date, no such helices have been found in *Drosophila* ETS TFs, and it is intriguing to speculate if this might be a general strategy of mammalian ETS factors to increase specificity and affinity in the much larger mammalian genome.

In addition to regulation of DNA binding within ETS domains, there is also significant selection of partner TFs to recognize compound binding sites via direct protein-protein interactions between the ETS domain and the DNA binding domain of another TF. For example, the aforementioned TCF sub-group of ETS factors, including ELK1, ELK3, and ELK4 can all partner with Serum Response Factor (SRF) TFs via direct interactions with the SRF MADS DNA binding domain (Buchwalter et al. 2004), and in the process reorients residues in the recognition helix of the ETS domain to rewire sequence specificity (Mo et al. 2001). A similar type of interaction occurs in partnership between ETS1 and PAX5, where residues from PAX5 also contact the ETS1 H3 helix to reorient sequence-specific contacts and permit recognition of a non-canonical GGAG core present at an important B-cell promoter (Garvie et al. 2001). Other heterotypic interactions include binding between ETS factor GABPA and GABPB1(Batchelor 1998),

ETS1 and RUNX1(Goetz et al. 2000), and SP1 and IR4 (Escalante et al. 2002). In addition to protein-protein interactions with members of other TF families, ETS factors also display homotypic interactions at the level of ETS to ETS binding. ETS1 can form homodimers via two distinct interfaces (Babayeva et al. 2010, 2012), which counteracts the auto inhibition in the ETS1 monomer and permits binding to tandem repeats on DNA. This strategy to counteract auto-inhibition and recognize specific sites on DNA may be a widespread feature of ETS DNA binding, as ELK1, ETV1, and FEV all display the propensity to form homo-dimers in complex with DNA (Cooper et al. 2014). Given the tendency for TFs to group together at enhancers in order to cooperatively bind their targets, it appears likely that further partner protein interactions will emerge to shed light on how extensively co-expressed ETS factors achieve regulatory specificity *in vivo*.

1.3.3 Properties of the Sterile Alpha Motif (SAM) domain

While the ETS domain is primarily responsible for site recognition, in many ETS family TFs an additional domain called the sterile alpha motif (SAM) domain mediates diverse protein-protein interactions necessary for ETS function. SAM domains are an all alpha-helical fold that mediates many diverse functions across eukaryotes, including both protein-protein, protein-RNA, and protein-lipid interactions (Qiao and Bowie 2005). The SAM domains that are appended to ETS family factors are a subset of the larger SAM fold (sometimes referred to as PNT domains), and do not contain an additional 3_{10} helix (H2') found in the larger SAM family (Mackereth et al. 2004) SAM domains are found in eleven of the 28 mammalian ETS factors, and just Yan and Pnt contain SAM domains in the *Drosophila* ETS family.

One of the primary features of SAM domains in ETS factors is to mediated homotypic, head-to-tail polymerization, in the case of both Yan and TEL (C. A. Kim et al. 2001; Qiao et al. 2004; Zhang et al. 2010). The ability of TEL to polymerize was first noted in characterization of in-frame translocation mutations that are capable of driving cancer(Golub et al. 1996) , including fusions to both tyrosine kinases and transcription factors (De Braekeleer et al. 2012). Structural characterization of the TEL SAM domain revealed that TEL polymerizes via association between two distinct surfaces on the SAM domain: the end-helix (EH) and mid-loop (ML) surfaces (C. A. Kim et al. 2001). These surfaces form a hydrophobic core, with residues making salt bridge interactions that modulate affinity scattered to the periphery of the interface (Cetinbas et al. 2013). All polymerizing SAM domains characterized to date show that the polymer grows as a left handed helix, and visualization of TEL polymers via electron microscopy showed that TEL SAM is capable of forming very large (500 angstroms or more) structures (C. A. Kim et al. 2001; Chongwoo A. Kim et al. 2005). Characterization of Yan demonstrated a similar mechanism of self-association (Qiao et al. 2004; Zhang et al. 2010), albeit with a much reduced affinity between two SAM molecules. The exact roles of polymerization in ETS function are not known. Self-association can mediate cooperative TEL binding to ETS sites *in vitro* (Green et al. 2010), and although it has been hypothesized that it may contribute to transcriptional repression, the exact mechanism of how this may be the case remains to be elucidated.

In addition to polymerization, SAM domains in ETS family factors have been shown to operate as heterotypic protein-protein interaction interfaces, specifically for docking kinases in order to induce phosphorylation. For example, ETS1 and ETS2 both

contain recognition sites on their SAM domains to associate with MAPK and thus induce phosphorylation of themselves (Seidel and Graves 2002). MAPK phosphorylation in turn unfolds a helix adjacent to the SAM domain that serves to recruit the co-activator CBP (Nelson et al. 2010), thus linking RTK signaling activity to transcriptional activation. The P2 isoform of Pnt has been shown to have a similar flexible alpha helix N-terminal to its SAM domain that can be phosphorylated by MAPK *in vitro* (Lau et al. 2012), and the SAM domain also serves as a docking module for MAPK (Qiao et al. 2006). Therefore it has been hypothesized that Pnt might operate through a similar mechanism of MAPK induced recruitment of co-activators.

The SAM domains of ETS family transcription factors are subject to other heterologous protein-protein interactions that regulate their activity. The *Drosophila* protein Mae (also known as Edl) does not contain an ETS DNA binding domain, but contains a SAM domain capable of interacting with both Yan and Pnt. Importantly for its regulation of Yan, the crystal structure of Mae demonstrates that it is capable of binding the EH surface of Yan, but does not have a ML surface that can further polymerize, thus serving as an effective capping protein for Yan polymerization (Qiao et al. 2004). Additionally, the affinity of the Yan-Mae interaction is three orders of magnitude stronger than the Yan-Yan interaction, so that at equilibrium Mae can effectively compete and depolymerize Yan polymers (Qiao et al. 2004). Mae also binds the SAM domain of Pnt, and is thought to block the docking of MAPK and thus prevent the activation of Pnt via the mechanism mentioned above (Qiao et al. 2006). Currently there are no known interactions with Mae-like proteins in mammalian systems that would regulate ETS family factors in a similar manner.

1.4 Yan is a model ETS family TF in Drosophila

1.4.1 Yan plays key roles in Drosophila *embryogenesis and eye development*

The Drosophila ETS family factor Yan (also called anterior open) is an essential transcriptional repressor and regulator of multiple cell fate decision throughout development. Yan fits neatly into the paradigm of TFs operating in development, by both being responsive to cell-to-cell signaling cues from the RTK pathway and collaborating with other TFs at cis-regulatory elements in order to represses transcription. Yan was first identified in the classic screens for regulators of Drosophila embryogenesis (Nüsslein-Volhard et al. 1984) and was later characterized from transpose-mediated screens of regulators of eye development (Lai and Rubin 1992; Tei et al. 1992). Yan plays multiple roles throughout development, with primary roles during embryogenesis and eye development, which was alluded to in Section 1.1 as a paradigm of a single cell fate specification choice, and will now be expanded upon in greater detail.

Eye development in *Drosophila* occurs by patterning an epithelial sheet of cells known as the eye-antennal imaginal disc (Tomlinson and Ready 1987). The earliest cells that will form the eye imaginal disc are specified during embryogenesis, which begin expressing a master regulatory network of TFs known as the Retinal Determination Gene Network (RDGN) (Davis and Rebay 2017). As the animal moves from embryogenesis through the three instars of larval development, the RDGN delimits the eye field, primes cells to respond to later signaling cues, and grows the disc from a few cells to a large tissue. At the transition between larval and pupal stages, the eye discs is comprised of a large number of developmentally equivalent progenitor cells, all of which are competent to adopt the many subsequent cell fates of the eye. Then, a progressive wave of signaling

proceeds from posterior to anterior in the disc, called the morphogenetic furrow (MF), which begins cell fate specification. Since all cells in the disc have the potential to form any particular cell type, cell-to-cell signaling is used iteratively— signaling at the MF specifies initial cell types, which then signal to other uncommitted progenitors, recruit them towards their fate, and then further signal to other sub-types. Thus within one disc, ommatidia are arrange spatially within the disc from youngest and least developed at the MF to oldest and most developed at the posterior most portion of the disc. In this way, all the cells of an ommatidium (i.e. one of the facets of the eye), are recruited, and once specification has been achieved, morphogenesis occurs during pupal stages to realize the functionalities of these cell types (Tomlinson and Ready 1987).

The cells that must form within the core of each ommatidium include eight photoreceptors to sense light, four cone cells to secrete the lens that covers the eye, and two primary pigment cells to make optically-protective pigment. Each ommatidium then shares six secondary pigment cells, three tertiary pigment cells, and three mechano-sensory bristle cells (Kumar 2012). The order listed above follows the order of recruitment from early at the MF through eclosure into adulthood. Furthermore, there are subsets of the earliest recruited cells, photoreceptors numbered R1-8, that are recruited in a specific order. EGFR signaling promotes cells toward a photoreceptor fate during eye development (Malartre 2016), and as a major effector of EGFR, Yan has predominate roles in inhibiting over-recruitment of specific photoreceptor cell types. There are five key specification steps, corresponding to which photoreceptors are being recruited, with some pairs of photoreceptors being developmentally equivalent at the points of

recruitment: namely, R8, R2/R5, R3/R4, R1/R6, and finally R7, which are specified in that order (Tomlinson and Ready 1987).

Yan is best characterized as a regulator of R7 fate in the eye. Yan directly binds and represses an enhancer of the homeodomain TF Prospero, which is necessary for R7 specification(Xu et al. 2000), and loss of Yan function leads to over-specification the number of R7 photoreceptors recruited to each ommatidium (Lai and Rubin 1992; Tei et al. 1992). High levels of MAPK signaling are needed to achieve down regulation of Yan and allow specification to proceed; therefore, input from both EGFR and Sevenless RTKs is necessary to dismantle Yan mediated repression (Freeman 1996). As an important point of crosstalk between the EGFR and Notch signaling pathways, the transcriptional effector of Notch signaling, Suppressor of Hairless, direct activates the transcription of *yan*, and thus helps establish the R7-specific threshold of RTK activation required (Rohrbaugh et al. 2002). The phenotype of classic Yan hypomorphs in the eye also suggests a role in the recruitment of outer photoreceptors (R1-6), as these alleles show ectopic photoreceptor recruitment to these positions (Lai and Rubin 1992). However, specific roles or transcriptional targets for Yan's action at these earlier cell fate decisions are not known.

Yan also plays a well-understood role in repression of Eve expression at stage 11 of embryogenesis. Eve expressing cells will go on to the form pericardial cells, which serve analogous functions to nephrons in vertebrates, in order to filter the hemolymph of the fly from toxins (Mills and King 1965). The regulation of Eve expression is controlled by an enhancer known as the Muscle Heart Enhancer (Halfon et al. 2000) which integrates inputs from multiple signaling pathways, including RTK signaling via Yan and

Pnt. In addition to the MHE, it has been noted that Yan and Pnt are recruited to repressive elements at the Eve locus, and serve to dampen variability in Eve expression and insure fidelity in the Eve expression pattern (Webber, Zhang, Mitchell-Dick, et al. 2013).

1.4.2 Yan operates downstream of the RTK/Ras/MAPK signaling cascade

Yan operate as one of the key downstream effectors of RTK/Ras/MAPK signaling, although it far from the only effector. Additionally, it should be noted that there are several RTKs besides those mentioned here that are used throughout *Drosophila* development, although all RTKs feed into the same pathway leading to the activation of Ras (Sopko and Perrimon 2013). A P-element insertion in the Yan gene was capable of giving both increased R7 photoreceptors and extra outer photoreceptors, and genetic interaction with members of the RTK/Ras/MAPK cascade, including EGFR and Sevenless, placed Yan downstream of Ras signaling (Lai and Rubin 1992). The expression pattern of Yan and sequence homology to mammalian ETS TFs thus suggested it was an important nuclear link between extra-cellular signaling pathways and transcription.

Subsequent characterization of Yan demonstrated that it is a direct target of MAPK phosphorylation (O'Neill et al. 1994), and acts in opposition to another *Drosophila* ETS factor Pnt, mentioned above. Pnt has two well-characterized isoforms, PntP1 and PntP2, that share a common ETS domain (Klämbt 1993), but critically, PntP2 contains a MAPK-responsive SAM domain, while PntP1 is a constitutively active transcriptional activator (O'Neill et al. 1994). PntP2 is also phosphorylated by MAPK (Brunner et al. 1994) and stimulates transcription upon RTK activation. Synthesizing these lines of evidence has lead to a model of Yan and Pnt competition in response to

RTK signaling. In the absence of signaling, Yan represses target gene expression, but upon activating RTK signaling and MAPK, Yan is phosphorylated and downregulated, while PntP2 is activated by MAPK phosphorylation and stimulates transcription from the same target genes. Eventually, transcription transitions to activation from PntP1, which is stimulated by PntP2 itself (Shwartz et al. 2013), in order to allow a sustained transcriptional response in the absence of the original RTK stimulus.

1.4.3 Regulation of Yan nuclear localization controls Yan function

One of the most important aspects of control over Yan function by MAPK has to do with its sub-cellular localization to the nucleus. Yan contains nine canonical MAPK phosphorylation motifs, and under normal circumstances, phosphorylation at these positions results in export of Yan from the nucleus, via a classic importin beta/ Crm1-mediated mechanism (Tootle et al. 2003). Upon exit from the nucleus, numerous PEST motifs make Yan a target for rapid degradation (Price and Lai 1999), and the F-box protein Fbl6 has a further role in ubiquitinating Yan for degradation (Roukens et al. 2008). When the MAPK motifs in Yan are mutated to prevent phosphorylation, the result is Yan nuclear localization even in the presence of strong activation of the RTK pathway, and consequently constitutive Yan-mediated repression (Rebay and Rubin 1995). This phosphorylation defective construct of Yan, called Yan^{Act}, is constitutively active and has severe phenotypes when expressed during development. In addition to the role of MAPK in regulating Yan's nuclear localization and activity, interaction with the SAM domain of Yan also play a role. Mae was shown to promote export from the nucleus, possibly by depolymerizing Yan and making it more amenable to recognition by the export machinery (Tootle et al. 2003). Furthermore, mutations in Yan that restrict it monomers

result in weaker localization to the nucleus, perhaps arising from increased export (Zhang et al. 2010). This is consistent with a model where there is active competition between localization of Yan into transcriptional regulatory complexes, and export of Yan from the nucleus for degradation.

1.4.4 Roles for SAM-mediated polymerization in controlling Yan function

The SAM domain also has roles in shaping the transcriptional and functional activity of Yan. Introduction of the missense mutations A86D and V105R into the two opposing surfaces of the SAM domain restricts Yan to monomers *in vitro* (Qiao et al. 2004), and perturbs the function of Yan *in vivo* (Zhang et al. 2010; Webber, Zhang, Cote, et al. 2013). These mutations were further shown to be relevant as they were able to decrease the fluorescence recovery time of fluorescently-tagged Yan in a Fluoresence Recovery After Photobleaching (FRAP) assay, indicative that Yan can form higher-order polymers in living cells. Additionally, these monomerizing mutations were shown to be necessary for down regulation of Yan target genes in an over-expression context, suggesting that polymerization is necessary to form active transcriptional regulatory complexes . Interestingly, rescue constructs containing YanA86D or YanV105R were shown to weakly rescue Yan null mutations (Webber, Zhang, Cote, et al. 2013), with significant embryonic and larval lethality, but mixed strains that form dimers completely rescue genetic function. Additionally, the chromatin occupancy of Yan monomers compared to wild type is slightly reduced, but the overall pattern of occupancy is largely maintained, suggesting that polymerization may be a partially contributing factor to the occupancy of Yan on DNA. (Webber, Zhang, Cote, et al. 2013) The exact mechanism of how

polymerization may contribute to or mediate Yan's repressive ability is not currently known.

1.5 Scope of dissertation

The core of this dissertation centers upon uncovering the relationships between Yan polymerization, and Yan DNA occupancy, transcriptional repression, and function. The methods utilized to do so fall neatly into theory and experiment. Chapter 2 establishes a model of Yan fractional occupancy from first principles at chemical equilibrium and explores the ramifications of polymerization for occupancy of transcription factors. Chapter 3 tests some of the predictions of the model by engineering mutants of Yan with increased SAM-SAM affinity, called SuperYan. These mutants were characterized in terms of their repressive ability showing a surprising inverse relationship between self-association affinity and repression. In the process, a role for medium-affinity protein-protein interactions among TFs was uncovered, and the model was revisited to reflect a picture of balanced interactions both on and off of DNA. Finally, Chapter 4 discusses the implications and future directions for expanding understanding of Yan's polymerization.

1.6 References

Anderson, W F, D H Ohlendorf, Y Takeda, and B W Matthews. 1981. "Structure of the Cro Repressor from Bacteriophage Lambda and Its Interaction with DNA." *Nature* 290 (5809):754–58.

Babayeva, Nigar D., Oxana I. Baranovskaya, and Tahir H. Tahirov. 2012. "Structural Basis of Ets1 Cooperative Binding to Widely Separated Sites on Promoter DNA." *PLoS ONE* 7 (3).

Babayeva, Nigar D., Phillip J. Wilder, Masaaki Shiina, Koshiki Mino, Michelle Desler, Kazuhiro Ogata, Angie Rizzino, and Tahir H. Tahirov. 2010. "Structural Basis of Ets1 Cooperative Binding to Palindromic Sequences on Stromelysin-1 Promoter DNA." *Cell Cycle* 9 (15):3054–62.

Bain, David L., Aaron F. Heneghan, Keith D. Connaghan-Jones, and Michael T. Miura. 2007. “Nuclear Receptor Structure: Implications for Function.” *Annual Review of Physiology* 69 (1):201–20.

Banerjee, U, P J Renfranz, J A Pollock, and S Benzer. 1987. “Molecular Characterization and Expression of Sevenless, a Gene Involved in Neuronal Pattern Formation in the Drosophila Eye.” *Cell* 49 (2):281–91.

Banerjee, Utpal, Patricia J. Renfranz, David R. Hinton, Bruce A. Rabin, and Seymour Benzer. 1987. “The Sevenless+ Protein Is Expressed Apically in Cell Membranes of Developing Drosophila Retina; It Is Not Restricted to Cell R7.” *Cell* 51 (1):151–58.

Banerji, Julian, Sandro Rusconi, and Walter Schaffner. 1981. “Expression of a ??-Globin Gene Is Enhanced by Remote SV40 DNA Sequences.” *Cell* 27 (2 PART 1):299–308.

Batchelor, A. H. 1998. “The Structure of GABP/: An ETS Domain- Ankyrin Repeat Heterodimer Bound to DNA.” *Science* 279 (5353):1037–41.

Berg, O G, R B Winter, and P H von Hippel. 1981. “Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 1. Models and Theory.” *Biochemistry* 20 (24):6929–48.

Berg, Otto G., and Peter H. von Hippel. 1985. “Diffusion-Controlled Macromolecular Interactions.” *Annual Review of Biophysics and Biophysical Chemistry* 14 (1):131–58.

Bories, J C, D M Willerford, D Grévin, L Davidson, A Camus, P Martin, D Stéhelin, and F W Alt. 1995. “Increased T-Cell Apoptosis and Terminal B-Cell Differentiation Induced by Inactivation of the Ets-1 Proto-Oncogene.” *Nature* 377 (6550):635–38.

Boyadjiev, Sa, and Ew Jabs. 2001. “Online Mendelian Inheritance in Man (OMIM) as a Knowledgebase for Human Developmental Disorders.” *Clinical Genetics* 57 (4):253–66.

Braekeleer, Etienne De, Nathalie Douet-Guilbert, Fr??d??ric Morel, Marie Jos??e Le Bris, Audrey Basinko, and Marc De Braekeleer. 2012. “ETV6 Fusion Genes in Hematological Malignancies: A Review.” *Leukemia Research*.

Brunner, D, K Dücker, N Oellers, E Hafen, H Scholz, and C Klämbt. 1994. “The ETS Domain Protein Pointed-P2 Is a Target of MAP Kinase in the Sevenless Signal Transduction Pathway.” *Nature*.

Buchwalter, Gilles, Christian Gross, and Bohdan Wasylyk. 2004. “Ets Ternary Complex Transcription Factors.” *Gene*.

BURGESS, RICHARD R., ANDREW A. TRAVERS, JOHN J. DUNN, and EKKEHARD K. F. BAUTZ. 1969. “Factor Stimulating Transcription by RNA Polymerase.” *Nature* 221 (5175):43–46.

Butler, J. E.F., and J. T. Kadonaga. 2001. “Enhancer-Promoter Specificity Mediated by DPE or TATA Core Promoter Motifs.” *Genes and Development* 15 (19):2515–19.

Cetinbas, Naniye, Helen Huang-Hobbs, Cristina Tognon, Gabriel Leprivier, Jianghong An, Steven McKinney, Mary Bowden, et al. 2013. “Mutation of the Salt Bridge-Forming Residues in the ETV6-SAM Domain Interface Blocks ETV6-NTRK3-Induced Cellular Transformation.” *Journal of Biological Chemistry* 288 (39):27940–50.

Chambers, Moria C., Kyung Han Song, and David S. Schneider. 2012. “Listeria Monocytogenes Infection Causes Metabolic Shifts in Drosophila Melanogaster.” *PLoS ONE* 7 (12).

Chen, L, J N Glover, P G Hogan, A Rao, and S C Harrison. 1998. “Structure of the DNA-Binding Domains from NFAT, Fos and Jun Bound Specifically to DNA.” *Nature* 392 (6671):42–48.

Chen, T, M Bunting, F D Karim, and C S Thummel. 1992. “Isolation and Characterization of Five Drosophila Genes That Encode an Ets-Related DNA Binding Domain.” *Developmental Biology* 151 (1):176–91.

Chu-Lagraff, Q, D M Wright, L K McNeil, and C Q Doe. 1991. “The Prospero Gene Encodes a Divergent Homeodomain Protein That Controls Neuronal Identity in Drosophila.” *Development (Cambridge, England)* Suppl 2:79–85.

Ciau-Uitz, Aldo, Lu Wang, Roger Patient, and Feng Liu. 2013. “ETS Transcription Factors in Hematopoietic Stem Cell Development.” *Blood Cells, Molecules & Diseases* 51 (4):248–55.

Cooper, Christopher D O, Joseph A Newman, and Opher Gileadi. 2014. “Recent Advances in the Structural Molecular Biology of Ets Transcription Factors: Interactions, Interfaces and Inhibition.” *Biochemical Society Transactions* 42 (1):130–38.

Coyne, H. Jerome, Soumya De, Mark Okon, Sean M. Green, Niraja Bhachech, Barbara J. Graves, and Lawrence P. McIntosh. 2012. “Autoinhibition of ETV6 (TEL) DNA Binding: Appended Helices Sterically Block the ETS Domain.” *Journal of Molecular Biology* 421 (1):67–84.

Crocker, Justin, Namiko Abe, Lucrezia Rinaldi, Alistair P. McGregor, Nicolás Frankel, Shu Wang, Ahmad Alsawadi, et al. 2015. “Low Affinity Binding Site Clusters Confer HOX Specificity and Regulatory Robustness.” *Cell* 160 (1–2):191–203.

Currie, Simon L., Desmond K.W. Lau, Jedediah J. Doane, Frank G. Whitby, Mark Okon, Lawrence P. McIntosh, and Barbara J. Graves. 2017. "Structured and Disordered Regions Cooperatively Mediate DNA-Binding Autoinhibition of ETS Factors ETV1, ETV4 and ETV5." *Nucleic Acids Research* 45 (5):2223–41.

Dam, H van, and M Castellazzi. 2001. "Distinct Roles of Jun : Fos and Jun : ATF Dimers in Oncogenesis." *Oncogene* 20 (19):2453–64.

Davidson, Eric H, and Michael S Levine. 2008. "Properties of Developmental Gene Regulatory Networks." *PNAS* 105 (51):20063–66.

Davis, Trevor L., and Ilaria Rebay. 2017. "Master Regulators in Development: Views from the Drosophila Retinal Determination and Mammalian Pluripotency Gene Networks." *Developmental Biology*.

De, Soumya, Mark Okon, Barbara J. Graves, and Lawrence P. McIntosh. 2016. "Autoinhibition of ETV6 DNA Binding Is Established by the Stability of Its Inhibitory Helix." *Journal of Molecular Biology* 428 (8):1515–30.

Doucleff, Michaeleen, and G Marius Clore. 2008. "Global Jumping and Domain-Specific Intersegment Transfer between DNA Cognate Sites of the Multidomain Transcription Factor Oct-1." *Proceedings of the National Academy of Sciences of the United States of America* 105 (37):13871–76.

Driever, Wolfgang, Gudrun Thoma, and Christiane Nüsslein-Volhard. 1989. "Determination of Spatial Domains of Zygotic Gene Expression in the Drosophila Embryo by the Affinity of Binding Sites for the Bicoid Morphogen." *Nature* 340 (6232):363–67.

Dynan, William S., and Robert Tjian. 1983. "The Promoter-Specific Transcription Factor Sp1 Binds to Upstream Sequences in the SV40 Early Promoter." *Cell* 35 (1):79–87.

Elf, J., G.-W. Li, and X. S. Xie. 2007. "Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell." *Science* 316 (5828):1191–94.

Escalante, Carlos R., Abraham L. Brass, Jagan M.R. Pongubala, Ella Shatova, Leyi Shen, Harinder Singh, and Aneel K. Aggarwal. 2002. "Crystal Structure of PU.1/IRF-4/DNA Ternary Complex." *Molecular Cell* 10 (5):1097–1105.

Evans, Nicole C., Christina I. Swanson, and Scott Barolo. 2012. "Sparkling Insights into Enhancer Structure, Function, and Evolution." *Current Topics in Developmental Biology* 98:97–120.

Fang, Ming, Jiong Li, Timothy Blauwkamp, Chandan Bhambhani, Nathan Campbell, and Ken M Cadigan. 2006. "C-Terminal-Binding Protein Directly Activates and

Represses Wnt Transcriptional Targets in Drosophila." *The EMBO Journal* 25 (12):2735–45.

Farley, E. K., K. M. Olson, W. Zhang, A. J. Brandt, D. S. Rokhsar, and M. S. Levine. 2015. "Suboptimization of Developmental Enhancers." *Science* 350 (6258):325–28.

Findlay, Victoria J., Amanda C. LaRue, David P. Turner, Patricia M. Watson, and Dennis K. Watson. 2013. "Understanding the Role of ETS-Mediated Gene Regulation in Complex Biological Processes." *Advances in Cancer Research* 119:10–61.

Fletcher, J C, and C S Thummel. 1995. "The Drosophila E74 Gene Is Required for the Proper Stage- and Tissue- Specific Transcription of Ecdysone-Regulated Genes at the Onset of Metamorphosis." *Development* 121 (5):1411–21.

Freeman, Matthew. 1996. "Reiterative Use of the EGF Receptor Triggers Differentiation of All Cell Types in the Drosophila Eye." *Cell* 87 (4):651–60.

Fu, W, H Duan, E Frei, M Noll, Andrew P Jarman, Seth A. Johnson, Katharine J. Harmon, et al. 1997. "Shaven and Sparkling Are Mutations in Separate Enhancers of the Drosophila Pax2 Homolog." *Development (Cambridge, England)* 454 (5):1215–18.

Funnell, Alister P W, and Merlin Crossley. 2012. "Homo- and Heterodimerization in Transcriptional Regulation." *Advances in Experimental Medicine and Biology*.

Furney, Simon J, Desmond G Higgins, Christos A Ouzounis, and Núria López-Bigas. 2006. "Structural and Functional Properties of Genes Involved in Human Cancer." *BMC Genomics* 7:3.

Gajewski, K M, and R A Schulz. 1995. "Requirement of the ETS Domain Transcription Factor D-ELG for Egg Chamber Patterning and Development during Drosophila Oogenesis." *Oncogene* 11 (6):1033–40.

Galvani, Angélique, and Christophe Thiriet. 2015. "Nucleosome Dancing at the Tempo of Histone Tail Acetylation." *Genes*.

Garrell, Joan, and Sonsoles Campuzano. 1991. "The Helix-loop-helix Domain: A Common Motif for Bristles, Muscles and Sex." *BioEssays*.

Garvie, Colin W., James Hagman, and Cynthia Wolberger. 2001. "Structural Studies of Ets-1/Pax5 Complex Formation on DNA." *Molecular Cell* 8 (6):1267–76.

Garvie, Colin W., and Cynthia Wolberger. 2001. "Recognition of Specific DNA Sequences." *Molecular Cell*.

Gaudet, J. 2002. "Regulation of Organogenesis by the Caenorhabditis Elegans FoxA

Protein PHA-4." *Science* 295 (5556):821–25.

Germain, Pierre, and William Bourguet. 2013. "Dimerization of Nuclear Receptors." *Methods in Cell Biology* 117:21–41.

Gilbert, Walter and Müller-Hill, B. 1966. "The Isolation of the Lac Repressor." *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* 12 (1):41–43.

Givaty, Ohad, and Yaakov Levy. 2009. "Protein Sliding along DNA: Dynamics and Structural Characterization." *Journal of Molecular Biology* 385 (4):1087–97.

Goetz, Tamara L, T L Gu, Nancy A Speck, and Barbara J Graves. 2000. "Auto-Inhibition of Ets-1 Is Counteracted by DNA Binding Cooperativity with Core-Binding Factor alpha2." *Molecular and Cellular Biology* 20 (1):81–90.

Golub, T R, A Goga, G F Barker, D E Afar, J McLaughlin, S K Bohlander, J D Rowley, O N Witte, and D G Gilliland. 1996. "Oligomerization of the ABL Tyrosine Kinase by the Ets Protein TEL in Human Leukemia." *Molecular and Cellular Biology* 16 (8):4107–16.

Gorman, Jason, and Eric C Greene. 2008. "Visualizing One-Dimensional Diffusion of Proteins along DNA." *Nature Structural & Molecular Biology* 15 (8):768–74.

Green, Sean M., H. Jerome Coyne, Lawrence P. McIntosh, and Barbara J. Graves. 2010. "DNA Binding by the ETS Protein TEL (ETV6) Is Regulated by Autoinhibition and Self-Association." *Journal of Biological Chemistry* 285 (24):18496–504.

Guardiani, Carlo, Massimo Cencini, and Fabio Cecconi. 2014. "Coarse-Grained Modeling of Protein Unspecifically Bound to DNA." *Physical Biology* 11 (2):26003.

Hafen, E, K Basler, J E Edstroem, and G M Rubin. 1987. "Sevenless, a Cell-Specific Homeotic Gene of Drosophila, Encodes a Putative Transmembrane Receptor with a Tyrosine Kinase Domain." *Science* 236 (4797):55–63.

Halfon, Marc S, Ana Carmena, Stephen Gisselbrecht, Charles M Sackerson, F Jiménez, Mary K Baylies, and Alan M Michelson. 2000. "Ras Pathway Specificity Is Determined by the Integration of Multiple Signal-Activated and Tissue-Restricted Transcription Factors." *Cell* 103 (1):63–74.

Halford, Stephen E., and John F. Marko. 2004. "How Do Site-Specific DNA-Binding Proteins Find Their Targets?" *Nucleic Acids Research*.

Hart, Adam H, Roddie Reventar, and Alan Bernstein. 2000. "Genetic Analysis of ETS Genes in C. Elegans." *Oncogene* 19 (55):6400–6408.

Hedglin, Mark, and Patrick J. O'Brien. 2008. "Human Alkyladenine DNA Glycosylase Employs a Processive Search for DNA Damage." *Biochemistry* 47 (44):11434–45.

Hollenhorst, Peter C., David A. Jones, and Barbara J. Graves. 2004. "Expression Profiles Frame the Promoter Specificity Dilemma of the ETS Family of Transcription Factors." *Nucleic Acids Research* 32 (18):5693–5702.

Hollenhorst, Peter C., Lawrence P. McIntosh, and Barbara J. Graves. 2011. "Genomic and Biochemical Insights into the Specificity of ETS Transcription Factors." *Annual Review of Biochemistry* 80 (1):437–71.

Hotta, Y, and S Benzer. 1970. "Genetic Dissection of the Drosophila Nervous System by Means of Mosaics." *Proceedings of the National Academy of Sciences of the United States of America* 67 (3):1156–63.

Jacob, François, and Jacques Monod. 1961. "Genetic Regulatory Mechanisms in the Synthesis of Proteins." *Journal of Molecular Biology*.

Jeronimo, Célia, and François Robert. 2017. "The Mediator Complex: At the Nexus of RNA Polymerase II Transcription." *Trends in Cell Biology*, 2017.

Jiang, Jin, and Michael Levine. 1993. "Binding Affinities and Cooperative Interactions with bHLH Activators Delimit Threshold Responses to the Dorsal Gradient Morphogen." *Cell* 72 (5):741–52.

Junion, Guillaume, Mikhail Spivakov, Charles Girardot, Martina Braun, E. Hilary Gustafson, Ewan Birney, and Eileen E M Furlong. 2012. "A Transcription Factor Collective Defines Cardiac Cell Fate and Reflects Lineage History." *Cell* 148 (3):473–86.

Kalodimos, C. G. 2004. "Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes." *Science* 305 (5682):386–89.

Kim, C. A., M. L. Phillips, W. Kim, M. Gingery, H. H. Tran, M. A. Robinson, S. Faham, and J. U. Bowie. 2001. "Polymerization of the SAM Domain of TEL in Leukemogenesis and Transcriptional Repression." *EMBO Journal* 20 (15):4173–82.

Kim, Chongwoo A., Michael R. Sawaya, Duilio Cascio, Woojae Kim, and James U. Bowie. 2005. "Structural Organization of a Sex-Comb-on-Midleg/polyhomeotic Copolymer." *Journal of Biological Chemistry* 280 (30):27769–75.

Klämbt, C. 1993. "The Drosophila Gene Pointed Encodes Two ETS-like Proteins Which Are Involved in the Development of the Midline Glial Cells." *Development (Cambridge, England)* 117 (1):163–76.

Kulkarni, Meghana M. 2003. "Information Display by Transcriptional Enhancers."

Development 130 (26):6569–75.

Kumar, Justin P. 2012. “Building an Ommatidium One Cell at a Time.” *Developmental Dynamics*.

Lacorazza, H. Daniel, Yasushi Miyazaki, Antonio Di Cristofano, Anthony Deblasio, Cyrus Hedvat, Jin Zhang, Carlos Cordon-Cardo, Shifeng Mao, Pier Paolo Pandolfi, and Stephen D. Nimer. 2002. “The ETS Protein MEF Plays a Critical Role in Perforin Gene Expression and the Development of Natural Killer and NK-T Cells.” *Immunity* 17 (4):437–49.

Lai, Zhi Chun, and Gerald M. Rubin. 1992. “Negative Control of Photoreceptor Development in Drosophila by the Product of the Yan Gene, an ETS Domain Protein.” *Cell* 70 (4):609–20.

Lander, E S. 2001. “Initial Sequencing and Analysis of the Human Genome.” *Nature* 409 (6822):860–921.

Lau, Desmond K.W., Mark Okon, and Lawrence P. McIntosh. 2012. “The PNT Domain from Drosophila Pointed-P2 Contains a Dynamic N-Terminal Helix Preceded by a Disordered Phosphoacceptor Sequence.” *Protein Science* 21 (11):1716–25.

Lee, Dongjun, Changwon Park, Ho Lee, Jesse J. Lugus, Seok Hyung Kim, Elizabeth Arentson, Yun Shin Chung, et al. 2008. “ER71 Acts Downstream of BMP, Notch, and Wnt Signaling in Blood and Vessel Progenitor Specification.” *Cell Stem Cell* 2 (5):497–507.

Leprince, D, A Gegonne, J Coll, C de Taisne, A Schneeberger, C Lagrou, and D Stehelin. 1983. “A Putative Second Cell-Derived Oncogene of the Avian Leukaemia Retrovirus E26.” *Nature* 306 (5941):395–97.

Loh, Kyle M., Renée van Amerongen, and Roel Nusse. 2016. “Generating Cellular Diversity and Spatial Form: Wnt Signaling and the Evolution of Multicellular Animals.” *Developmental Cell*.

Ludwig, Michael Z., Arnar Palsson, Elena Alekseeva, Casey M. Bergman, Janaki Nathan, and Martin Kreitman. 2005. “Functional Evolution of a Cis-Regulatory Module.” *PLoS Biology* 3 (4):0588–98.

Luscombe, Nicholas M, Susan E Austin, Helen M Berman†, and Janet M Thornton. 2000. “An Overview of the Structures of Protein-DNA Complexes.” *Genome Biology* 1 (1):1–1.

Mackereth, Cameron D., Manuela Schärpf, Lisa N. Gentile, Scott E. MacIntosh, Carolyn M. Slupsky, and Lawrence P. McIntosh. 2004. “Diversity in Structure and Function of the Ets Family PNT Domains.” *Journal of Molecular Biology* 342 (4):1249–64.

Mahmutovic, Anel, Otto G. Berg, and Johan Elf. 2015. “What Matters for Lac Repressor Search in Vivo - Sliding, Hopping, Intersegment Transfer, Crowding on DNA or Recognition?” *Nucleic Acids Research* 43 (7):3454–64.

Malartre, Marianne. 2016. “Regulatory Mechanisms of EGFR Signalling during Drosophila Eye Development.” *Cellular and Molecular Life Sciences*.

Mannervik, Mattias. 2014. “Control of Drosophila Embryo Patterning by Transcriptional Co-Regulators.” *Experimental Cell Research*.

McKay, D B, and T A Steitz. 1981. “Structure of Catabolite Gene Activator Protein at 2.9 A Resolution Suggests Binding to Left-Handed B-DNA.” *Nature* 290 (5809):744–49.

McKercher, S R, B E Torbett, K L Anderson, G W Henkel, D J Vestal, H Baribault, M Klemsz, et al. 1996. “Targeted Disruption of the PU.1 Gene Results in Multiple Hematopoietic Abnormalities.” *The EMBO Journal* 15 (20):5647–58.

Mills, R. P., and R. C. King. 1965. “The Pericardial Cells of Drosophila Melanogaster.” *Quarterly Journal of Microscopical Science* s3-106 (1962):261–68.

Mo, Y, W Ho, K Johnston, and R Marmorstein. 2001. “Crystal Structure of a Ternary SAP-1/SRF/c-Fos SRE DNA Complex.” *Journal of Molecular Biology* 314 (3):495–506.

Muthusamy, Natarajan, Kevin Barton, and Jeffrey M Leiden. 1995. “Defective Activation and Survival of T Cells Lacking Ets-1 Transcription Factor.” *Nature*.

Nagaraj, Raghavendra, and Utpal Banerjee. 2004. “The Little R Cell That Could.” *International Journal of Developmental Biology*.

Nelson, M. L., H.-S. Kang, G. M. Lee, A. G. Blaszczak, D. K. W. Lau, L. P. McIntosh, and B. J. Graves. 2010. “Ras Signaling Requires Dynamic Properties of Ets1 for Phosphorylation-Enhanced Binding to Coactivator CBP.” *Proceedings of the National Academy of Sciences* 107 (22):10026–31.

Nibu, Yutaka, Hailan Zhang, Ewa Bajor, Scott Barolo, Stephen Small, and Michael Levine. 1998. “dCtBP Mediates Transcriptional Repression by Knirps, Krüppel and Snail in the Drosophila Embryo.” *The EMBO Journal* 17 (23):7009–20.

Nunn, MF, PH Seeburg, C Moscovici, and PH Duesberg. 1983. “Tripartite Structure of the Avian Erythroblastosis Virus E26 Transforming Gene.” *Nature* 306 (5941):391–95.

Nüsslein-Volhard, C, E Wieschaus, and H Kluding. 1984. “Roux’s Archives of

Developmental Biology Mutations Affecting the Pattern of the Larval Cuticle in Drosophila Melanogaster I. Zygotic Loci on the Second Chromosome." *Roux's Arch Dev Biol* 193:267–82.

O'Neill, Elizabeth M., Ilaria Rebay, Robert Tjian, and Gerald M. Rubin. 1994. "The Activities of Two Ets-Related Transcription Factors Required for Drosophila Eye Development Are Modulated by the Ras/MAPK Pathway." *Cell* 78 (1):137–47.

Ogryzko, Vasily V., R. Louis Schiltz, Valya Russanova, Bruce H. Howard, and Yoshihiro Nakatani. 1996. "The Transcriptional Coactivators p300 and CBP Are Histone Acetyltransferases." *Cell* 87 (5):953–59.

Pabo, C O, and M Lewis. 1982. "The Operator-Binding Domain of Lambda Repressor: Structure and DNA Recognition." *Nature* 298 (5873):443–47.

Panne, D, T Maniatis, and S C Harrison. 2007. "An Atomic Model of the Interferon-Beta Enhanceosome." *Cell* 129 (6):1111–23.

Parker, David S, Michael A White, Andrea I Ramos, Barak A Cohen, Scott Barolo, S. Barolo, J. W. Posakony, et al. 2011. "The Cis-Regulatory Logic of Hedgehog Gradient Responses: Key Roles for Gli Binding Affinity, Competition, and Cooperativity." *Science Signaling* 4 (176):ra38.

Pires-daSilva, André, and Ralf J. Sommer. 2003. "The Evolution of Signalling Pathways in Animal Development." *Nature Reviews Genetics* 4 (1):39–49.

Pluciennik, Anna, and Paul Modrich. 2007. "Protein Roadblocks and Helix Discontinuities Are Barriers to the Initiation of Mismatch Repair." *Proceedings of the National Academy of Sciences of the United States of America* 104 (31):12709–13.

Price, Mitch D., and Zhi Chun Lai. 1999. "The Yan Gene Is Highly Conserved in Drosophila and Its Expression Suggests a Complex Role throughout Development." *Development Genes and Evolution* 209 (4):207–17.

Ptashne, Mark. 1967. "Isolation of the Λ Phage Repressor." *Proceedings of the National Academy of Sciences of the United States of America* 57 (2):306.

Pufall, M. A. 2005. "Variable Control of Ets-1 DNA Binding by Multiple Phosphates in an Unstructured Region." *Science* 309 (5731):142–45.

Qi, Dai, Mattias Bergman, Hitoshi Aihara, Yutaka Nibu, and Mattias Mannervik. 2008. "Drosophila Ebi Mediates Snail-Dependent Transcriptional Repression through HDAC3-Induced Histone Deacetylation." *The EMBO Journal* 2726 (6):898–909.

Qiao, Feng, and James U Bowie. 2005. "The Many Faces of SAM." *Science's STKE :*

Signal Transduction Knowledge Environment 2005 (January):re7.

Qiao, Feng, Bryan Harada, Haiyun Song, Julian Whitelegge, Albert J Courey, and James U Bowie. 2006. “Mae Inhibits Pointed-P2 Transcriptional Activity by Blocking Its MAPK Docking Site.” *The EMBO Journal* 25 (1):70–79.

Qiao, Feng, Haiyun Song, Chongwoo A. Kim, Michael R. Sawaya, Jacob B. Hunter, Mari Gingery, Ilaria Rebay, Albert J. Courey, and James U. Bowie. 2004. “Derepression by Depolymerization: Structural Insights into the Regulation of Yan by Mae.” *Cell* 118 (2):163–73.

Ramos, Andrea I, and Scott Barolo. 2013. “Low-Affinity Transcription Factor Binding Sites Shape Morphogen Responses and Enhancer Evolution.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 368:20130018.

Razin, S. V., V. V. Borunova, O. G. Maksimenko, and O. L. Kantidze. 2012. “Cys2His2 Zinc Finger Protein Family: Classification, Functions, and Major Members.” *Biochemistry (Moscow)* 77 (3):217–26.

Ready, Donald F., Thomas E. Hanson, and Seymour Benzer. 1976. “Development of the Drosophila Retina, a Neurocrystalline Lattice.” *Developmental Biology* 53 (2):217–40.

Rebay, Ilaria, and Gerald M. Rubin. 1995. “Yan Functions as a General Inhibitor of Differentiation and Is Negatively Regulated by Activation of the Ras1/MAPK Pathway.” *Cell* 81 (6):857–66.

Reinke, Rosemary, and S. Lawrence Zipursky. 1988. “Cell-Cell Interaction in the Drosophila Retina: The Bride of Sevenless Gene Is Required in Photoreceptor Cell R8 for R7 Cell Development.” *Cell* 55 (2):321–30.

Rohrbaugh, Margaret, Edward Ramos, Duc Nguyen, Mitch Price, Yu Wen, and Zhi Chun Lai. 2002. “Notch Activation of Yan Expression Is Antagonized by RTK/pointed Signaling in the Drosophila Eye.” *Current Biology* 12 (7):576–81.

Rohs, Remo, Xiangshu Jin, Sean M. West, Rohit Joshi, Barry Honig, and Richard S. Mann. 2010. “Origins of Specificity in Protein-DNA Recognition.” *Annual Review of Biochemistry* 79 (1):233–69.

Roukens, M Guy, Mariam Alloul-Ramdhani, Setareh Moghadasi, Marjolein Op den Brouw, and David a Baker. 2008. “Downregulation of Vertebrate Tel (ETV6) and Drosophila Yan Is Facilitated by an Evolutionarily Conserved Mechanism of F-Box-Mediated Ubiquitination.” *Molecular and Cellular Biology* 28 (13):4394–4406.

Rubin, G. M., H. C. Chang, F. Karim, T. Laverty, N. R. Michaud, D. K. Morrison, I.

Rebay, A. Tang, M. Therrien, and D. A. Wassarman. 1997. “Signal Transduction Downstream from RAS in Drosophila.” In *Cold Spring Harbor Symposia on Quantitative Biology*, 62:347–52.

Sainsbury, Sarah, Carrie Bernecky, and Patrick Cramer. 2015. “Structural Basis of Transcription Initiation by RNA Polymerase II.” *Nature Reviews Molecular Cell Biology* 16 (3):129–43.

Schmidt, Hugo G., Sven Sewitz, Steven S. Andrews, and Karen Lipkow. 2014. “An Integrated Model of Transcription Factor Diffusion Shows the Importance of Intersegmental Transfer and Quaternary Protein Structure for Target Site Finding.” *PLoS ONE* 9 (10).

Scott, Edward W, Robert C Fisher, Marilyn C Olson, Eli W Kehrli, M.Celeste Simon, and Harinder Singh. 1997. “PU.1 Functions in a Cell-Autonomous Manner to Control the Differentiation of Multipotential Lymphoid–Myeloid Progenitors.” *Immunity* 6 (4):437–47.

Seidel, J. J., and B. J. Graves. 2002. “An ERK2 Docking Site in the Pointed Domain Distinguishes a Subset of ETS Transcription Factors.” *Genes and Development* 16 (1):127–37.

Shwartz, Arkadi, Shaul Yogev, Eyal D Schejter, and Ben-zion Shilo. 2013. “Sequential Activation of ETS Proteins Provides a Sustained Transcriptional Response to EGFR Signaling.” *Development* 140:2746–54.

Small, S, A Blair, and M Levine. 1992. “Regulation of Even-Skipped Stripe 2 in the Drosophila Embryo.” *The EMBO Journal* 11 (11):4047–57.

Smith, Ngaio C., and Jacqueline M. Matthews. 2016. “Mechanisms of DNA-Binding Specificity and Functional Gene Regulation by Transcription Factors.” *Current Opinion in Structural Biology*.

Sopko, Richelle, and Norbert Perrimon. 2013. “Receptor Tyrosine Kinases in Drosophila Development.” *Cold Spring Harbor Perspectives in Biology* 5 (6).

Spitz, Francois, and Eileen E M Furlong. 2012. “Transcription Factors: From Enhancer Binding to Developmental Control.” *Nature Reviews Genetics* 13:613–26.

Stankiewicz, Trisha R., Josie J. Gray, Aimee N. Winter, and Daniel A. Linseman. 2014. “C-Terminal Binding Proteins: Central Players in Development and Disease.” *Biomolecular Concepts*.

Szymczyna, Blair R., and Cheryl H. Arrowsmith. 2000. “DNA Binding Specificity Studies of Four ETS Proteins Support an Indirect Read-out Mechanism of Protein-DNA Recognition.” *Journal of Biological Chemistry* 275 (37):28363–70.

Tei, H., I. Nihonmatsu, T. Yokokura, R. Ueda, Y. Sano, T. Okuda, K. Sato, K. Hirata, S. C. Fujita, and D. Yamamoto. 1992. “Pokkuri, a Drosophila Gene Encoding an E-26-Specific (Ets) Domain Protein, Prevents Overproduction of the R7 Photoreceptor.” *Proceedings of the National Academy of Sciences* 89 (15):6856–60.

Tomlinson, Andrew, and Donald F. Ready. 1987. “Cell Fate in the Drosophila Ommatidium.” *Developmental Biology* 123 (1):264–75.

Tootle, Tina L, Philina S Lee, and Ilaria Rebay. 2003. “CRM1-Mediated Nuclear Export and Regulated Activity of the Receptor Tyrosine Kinase Antagonist YAN Require Specific Interactions with MAE.” *Development (Cambridge, England)* 130 (5):845–57.

Turki-Judeh, Wiam, and Albert J. Courey. 2012. “Groucho. A Corepressor with Instructive Roles in Development.” *Current Topics in Developmental Biology* 98:65–96.

Vaquerizas, Juan M., Sarah K. Kummerfeld, Sarah A. Teichmann, and Nicholas M. Luscombe. 2009. “A Census of Human Transcription Factors: Function, Expression and Evolution.” *Nature Reviews Genetics* 10 (4):252–63.

Viadiu, Hector, and Aneel K. Aggarwal. 2000. “Structure of BamHI Bound to Nonspecific DNA: A Model for DNA Sliding.” *Molecular Cell* 5 (5):889–95.

Vinson, Charles, Asha Acharya, and Elizabeth J. Taparowsky. 2006. “Deciphering B-ZIP Transcription Factor Interactions in Vitro and in Vivo.” *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1759 (1–2):4–12.

Vo Ngoc, Long, Yuan-Liang Wang, George A Kassavetis, and James T Kadonaga. 2017. “The Punctilious RNA Polymerase II Core Promoter.” *Genes & Development* 31 (13):1289–1301.

Voss, Anne K., and Tim Thomas. 2009. “MYST Family Histone Acetyltransferases Take Center Stage in Stem Cells and Development.” *BioEssays*.

Vuzman, Dana, Ariel Azia, and Yaakov Levy. 2010. “Searching DNA via a ‘Monkey Bar’ Mechanism: The Significance of Disordered Tails.” *Journal of Molecular Biology* 396 (3):674–84.

Waddington, C. H., and Margaret M. Perry. 1960. “The Ultra-Structure of the Developing Eye of Drosophila.” *Proceedings of the Royal Society B: Biological Sciences* 153 (951):155–78.

Wang, Li Chun, Wojciech Swat, Yuko Fujiwara, Laurie Davidson, Jane Visvader, Frank Kuo, Fred W. Alt, D. Gary Gilliland, Todd R. Golub, and Stuart H. Orkin. 1998.

"The TEL/ETV6 Gene Is Required Specifically for Hematopoiesis in the Bone Marrow." *Genes and Development* 12 (15):2392–2402.

Wang, Li, and Sharon YR Dent. 2014. "Functions of SAGA in Development and Disease." *Epigenomics* 6 (3):329–39.

Watson, D K, M J McWilliams-Smith, M F Nunn, P H Duesberg, S J O'Brien, and T S Papas. 1985. "The Ets Sequence from the Transforming Gene of Avian Erythroblastosis Virus, E26, Has Unique Domains on Human Chromosomes 11 and 21: Both Loci Are Transcriptionally Active." *Proceedings of the National Academy of Sciences of the United States of America* 82 (21):7294–98.

Webber, Jemma L., Jie Zhang, Lauren Cote, Pavithra Vivekanand, Xiaochun Ni, Jie Zhou, Nicolas Nègre, Richard W. Carthew, Kevin P. White, and Ilaria Rebay. 2013. "The Relationship between Long-Range Chromatin Occupancy and Polymerization of the Drosophila Ets Family Transcriptional Repressor Yan." *Genetics* 193 (2):633–49.

Webber, Jemma L., Jie Zhang, Aaron Mitchell-Dick, and Ilaria Rebay. 2013. "3D Chromatin Interactions Organize Yan Chromatin Occupancy and Repression at the Even-Skipped Locus." *Genes and Development* 27 (21):2293–98.

Wei, Gong-Hong, Gwenael Badis, Michael F Berger, Teemu Kivioja, Kimmo Palin, Martin Enge, Martin Bonke, et al. 2010. "Genome-Wide Analysis of ETS-Family DNA-Binding in Vitro and in Vivo." *The EMBO Journal* 29 (13):2147–60.

Wilce, Jackie, Julian Vivian, and Matthew Wilce. 2012. "Oligonucleotide Binding Proteins: The Occurrence of Dimer and Multimer Formation." *Advances in Experimental Medicine and Biology* 747:91–104.

Winkler, F K, D W Banner, C Oefner, D Tsernoglou, R S Brown, S P Heathman, R K Bryan, P D Martin, K Petratos, and K S Wilson. 1993. "The Crystal Structure of EcoRV Endonuclease and of Its Complexes with Cognate and Non-Cognate DNA Fragments." *The EMBO Journal* 12 (5):1781–95.

Xu, Chunyan, Rachele C. Kauffmann, Jianjun Zhang, Susan Kladny, and Richard W. Carthew. 2000. "Overlapping Activators and Repressors Delimit Transcriptional Response to Receptor Tyrosine Kinase Signals in the Drosophila Eye." *Cell* 103 (1):87–97.

Zhang, Jie, Thomas G W Graham, Pavithra Vivekanand, Lauren Cote, Maureen Cetera, and Ilaria Rebay. 2010. "Sterile Alpha Motif Domain-Mediated Self-Association Plays an Essential Role in Modulating the Activity of the Drosophila ETS Family Transcriptional Repressor Yan." *Molecular and Cellular Biology* 30 (5):1158–70.

Chapter 2: DNA occupancy of polymerizing transcription factors: a chemical model of the ETS family factor, Yan

C. Matthew Hope, Ilaria Rebay, John Reinitz

Biophysical Journal 112(1): 180-192, 2017

2.1 Abstract

Transcription factors use both protein-DNA and protein-protein interactions to assemble appropriate complexes to regulate gene expression. While most transcription factors operate as monomers or dimers, a few, including the ETS family repressors, *Drosophila melanogaster* Yan and its human homolog TEL/ETV6, can polymerize. Although polymerization is required for both the normal and oncogenic function of Yan and TEL/ETV6, the mechanisms by which it influences the recruitment, organization, and stability of transcriptional complexes remain poorly understood. Further, a quantitative description of the DNA occupancy of a polymerizing transcription factor is lacking, and has broader applications to the conceptually-related area of polymerizing chromatin regulators. To expand the theoretical basis for understanding how the oligomeric state of a transcriptional regulator influences its chromatin occupancy and function, we leveraged the extensive biochemical characterization of ETS factors to develop a mathematical model of Yan occupancy at chemical equilibrium. We find that spreading condensation from a specific binding site can take place in a path-independent manner given reasonable values of the free energies of specific and non-specific DNA binding and protein-protein cooperativity. Our calculations show that polymerization confers a transcription factor with the unique ability to extend occupancy across DNA regions far from specific binding sites. In contrast, dimerization promotes recruitment to clustered binding sites and maximizes discrimination between specific and non-specific sites. We speculate that the association with non-specific DNA afforded by polymerization may enable regulatory behaviors that are well-suited to transcriptional repressors, but perhaps incompatible with precise activation.

2.2 Introduction

Sequence specific transcription factors (TFs) regulate the gene expression programs that drive multi-cellular animal development. To achieve this, TFs combine protein-DNA and protein-protein interactions to assemble transcriptional regulatory complexes. While protein-DNA interactions enable recognition of and binding to specific DNA sequences, a wide variety of homo- and heterotypic protein-protein interactions modulate DNA binding, promote associations between different TFs (Funnell and Crossley 2012), and recruit co-activators (Weake and Workman 2010) or co-repressors (Perissi et al. 2010). Homo- or heterotypic dimerization is common among TFs, with notable examples including helix-loop-helix proteins (Garrell and Campuzano 1991), nuclear receptors (Germain and Bourguet 2013), and E26 transformation-specific (ETS) family TFs (Li et al. 2000). More broadly, the full spectrum of protein-protein interactions stabilizes occupancy of TF complexes at appropriate *cis*-regulatory elements and increases the combinatorial specificity and complexity for target gene regulation (Evans et al. 2012). Thus, understanding how both protein-DNA and protein-protein interactions determine TF occupancy is necessary for gaining molecular insight into the logic of development (Davidson and Levine 2008).

Notable among the various types of protein-protein interactions, certain classes of transcriptional regulators are capable of open-ended polymerization, forming large homotypic complexes (Qiao et al. 2004). The effects of oligomerization beyond dimerization are not well-characterized from either a functional or a theoretical perspective, and most studies of polymerizing transcriptional regulators have focused on multi-member gene silencing complexes that contain sequence specific activities (Kueng

et al. 2013; Grossniklaus and Paro 2014). However, there is a class of polymerizing TFs that unite both sequence specific DNA binding and homotypic self-association in one molecule—namely members of the ETS family of TFs, *Drosophila* Yan and its human ortholog TEL or ETV6 (Zhang et al. 2010; Kim et al. 2001).

Yan and TEL are both transcriptional repressors that regulate gene expression downstream of Receptor Tyrosine Kinase(RTK)/Ras/Mitogen-activated Protein Kinase(MAPK) signaling (Lai and Rubin 1992; O'Neill et al. 1994; Golub et al. 1994). Both bind DNA in a sequence specific fashion via their ETS DNA binding domain, and both also homotypically polymerize via a sterile alpha motif (SAM) domain (Zhang et al. 2010; Kim et al. 2001; Poirel et al. 1997). SAM-mediated self-association is required for the repressive function of both Yan and TEL, and measurements of diffusion kinetics in cultured *Drosophila* cells together with genetic rescue experiments argue for the presence and importance of SAM-dependent higher order Yan complexes *in vivo* (Zhang et al. 2010; Tognon et al. 2004).

Despite these indications of functional significance, whether the polymerizing ability of TFs like Yan and TEL can produce distinct DNA occupancy, and so confer novel regulatory potential, as compared to that of more conventional monomeric or dimeric TFs is not known. Suggestively, chromatin immunoprecipitation experiments in *Drosophila* have shown that regions of Yan occupancy are quantitatively broader than most regions associated with sequence specific TFs (Webber et al. 2013). Since direct *in vivo* measurements of the specific protein-protein and protein-DNA contacts that produce a particular chromatin binding profile is currently experimentally impossible, we turned

to a modeling approach to gain insight into the mechanisms by which polymerization might drive the assembly of higher order TF complexes on DNA.

The DNA occupancy of TFs has been modeled mathematically by a variety of approaches, most extensively using one-dimensional lattice models (McGhee and von Hippel 1974; Hippel et al. 1974; Johnson et al. 1981; Tsodikov et al. 2001) (reviewed in (Teif and Rippe 2010)). These models describe TF binding to long stretches of DNA at chemical equilibrium, and have investigated questions concerning specific and non-specific binding, cooperative interactions among molecules, and orientation of binding sites. The thermodynamic feasibility of the nucleation and spreading of a sequence specific, polymerizing TF remains uninvestigated. In this work, we present a model of Yan occupancy that recapitulates some of the features of Yan binding across the genome (Webber et al. 2013), and provides a framework to consider the occupancy of polymerizing TFs more broadly. We show that Yan occupancy can spread from specific recruitment sites to more distal sites at equilibrium, and that this behavior depends on both protein-DNA and protein-protein interactions. Additionally, we calculate the phase space for the behavior of systems that have different strengths of interactions, and demonstrate that clustering recruitment sites increases Yan occupancy, even for the same strength of interactions and concentrations. Lastly, we demonstrate quantitative differences in the stability, size and sequence-specificity of transcription complexes formed by polymerizing versus non-polymerizing TFs. Taken together, our results expand the approaches utilized for assessing TF occupancy and highlight distinct roles for polymerization in determining TF occupancy patterns at equilibrium.

2.3 Results

2.3.1 Formulation of the model

We set out to develop a model of Yan binding at chemical equilibrium, utilizing this well-characterized protein to explore the role of polymerization in TF occupancy. Our model conceptualizes the genome as a series of binding sites for Yan, arranged in a one-dimensional lattice, which we refer to as an element (Fig. 2.1A). In contrast to other one-dimensional lattice models of TF binding that consider an infinitely large genome (McGhee and von Hippel 1974; Tsodikov et al. 2001; Teif and Rippe 2010), we calculated occupancy for a discrete element. Unless otherwise noted, the elements we calculate contain one specific binding site for Yan and *n* non-specific binding sites. This reflects our intuition that the majority of contacts a transcription factor makes with DNA are not sequence specific, but instead are governed by weaker electrostatic interactions with the DNA phosphate backbone (Marcovitz and Levy 2011). Thus higher affinity sites, which we term ETS sites, make up a small fraction of the total sites and confer the sequence-specific contacts typical of ETS family transcription factors, while lower affinity non-specific sites make up the majority. Although our approach limits the scope of the binding region considered, it provides arbitrary control over the configuration of binding sites within the element, as well as the calculation of fractional occupancies of interest. Our model also incorporates parameters to represent Yan's protein-DNA and SAM-mediated protein-protein interactions: α, β, and γ (Fig. 2.1A). α represents the free energy value of specific binding to ETS sites, while β represents the free energy value of non-specific Yan binding to low affinity sites. When two Yan molecules are adjacent to one another on an element, γ represents the free energy of a protein-protein interaction

via their SAM domains. As Yan polymerizes, each additional adjacent Yan molecule contributes an additional increment of γ. Using this framework, the fractional occupancy of any microstate(s) of an element can be calculated as a function of four parameters, n, α, β, γ, and the concentration of Yan (Fig. 2.1B; further details in Materials and Methods).

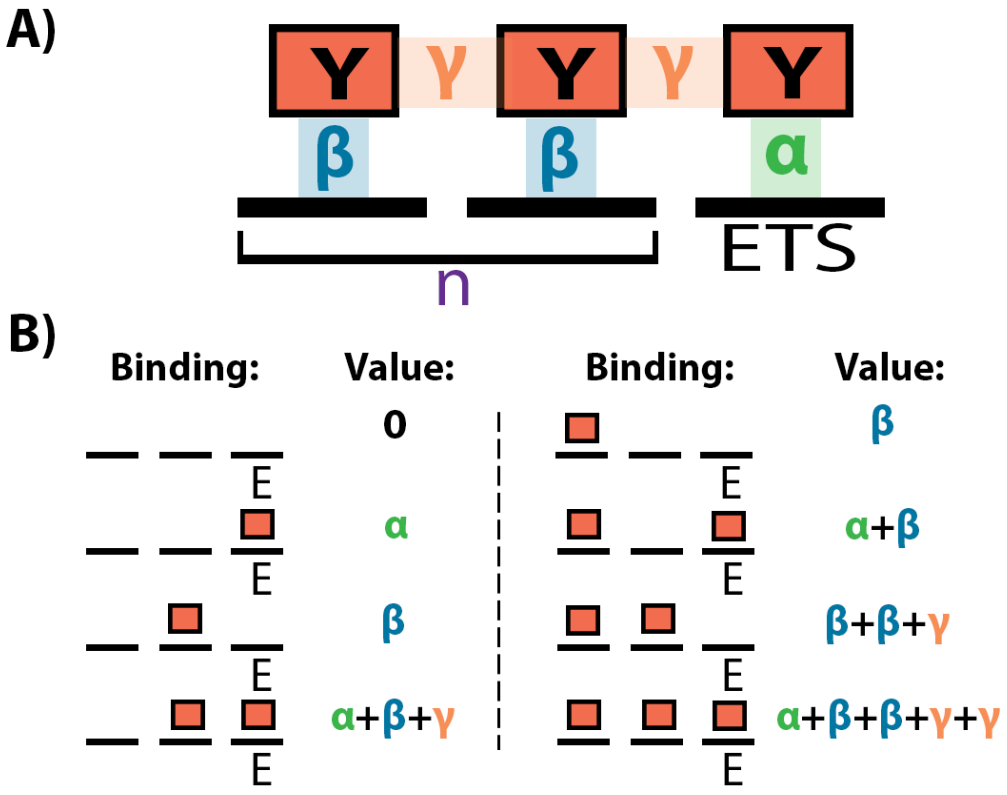


Figure 2.1: Lattice model for Yan binding at equilibrium.

A) Yan molecules (red boxes) occupy discrete sites (bold black lines) at an element of DNA. The class of elements considered by the model contain one high affinity specific binding site (ETS site) and n nonspecific binding sites. Yan molecules participate in different types of interactions depending on their occupancy at the element, and the free energies of these interactions are given by the terms α, β, and γ. α (green) represents the specific DNA binding free energy, β (blue) represents the non-specific DNA binding free energy, and γ (orange) represents the SAM-mediated protein-protein interaction free energy, contingent upon two molecules of Yan adjacent to one another in the lattice. B) Configurations of binding and free energy values for an element where n is equal to 2. The binary representation of the index of configurations (0,1,…,6,7) corresponds to the binding configuration itself. This provides a straightforward means of manipulating and

Figure 2.1 (cont.): scoring many microstates. Each microstate is shown with its free energy in terms of α, β, and γ.

2.3.2 Parameterization and implementation of the model

Calculating the fractional occupancies of interest requires special consideration from a computational perspective. Because the number of microstates grows exponentially as *n* increases, the microstates must be generated and scored in an efficient, structured way. To minimize the complexity of the calculation, we utilized a binary notation scheme for ordering microstates. We assigned each microstate an integer index *k,* where *k* is a value from 0 to 2^{n+1}-1 whose binary representation reflects the molecular configuration of the microstate. Although ordering the calculation in this fashion does not reduce the computational complexity *per se*, the increased efficiency reduces the computing resources required.

To begin calculating fractional occupancies for Yan, we used the literature to parameterize the model with reasonable values of α, β, γ, and Yan concentration. Although there are no biochemical measurements of Yan's specific or non-specific binding to DNA, these affinities have been measured for its human homolog, TEL, enabling us to assign values for α and β (De et al. 2014). Previous work has measured the affinity between two SAM molecules of Yan (Qiao et al. 2004), which we used for γ. These values for α, β, and γ (-9.955, -5.837, and -7.043 kcal/mol, respectively) are referred to as the wild type parameters, α_{wt}, β_{wt}, γ_{wt}. Lastly, there are no published measurements of nuclear Yan concentration, but based on measurements of other transcription factors (Biggin 2011), we set a reasonable range of concentrations from 0.1-100nM.

To determine what size of element would permit exploration of the broad scale features of Yan occupancy without requiring exorbitant amounts of memory to track microstates, we explored Yan binding across a range of element sizes from n=2 to 23; a subset of these fractional occupancy curves are shown in Figure 2.S1. At low values of n (Figure 2.S1A-B), Yan binding curves were shallow across all concentrations calculated. As n increases, the binding profiles sharpened quickly. At elements of size n=19 to 22 (Figure 2.S1F-I), the sharpening slowed significantly, with these elements showing highly similar profiles. We also plotted the concentrations of 50% fractional occupancy for specific sites across the collection of elements from n=2 to 23 (Figure 2.S2). Elements with low values of n (1 to 5) required higher concentrations of Yan to attain 50% fractional occupancy, but as n increased, all positions measured asymptotically approached a value of ~0.4nM. Therefore, we selected 24 site elements (n=23) for all further calculations because the Yan fractional occupancy profile was representative of the larger set of fractional occupancy curves, and the computational cost was deemed reasonable.

2.3.3 Yan occupancy spreads across the element at equilibrium

Genome-wide ChIP studies of Yan occupancy have noted extensive bound regions that are larger than those typically occupied by other *Drosophila* TFs (Webber et al. 2013; Roy et al. 2010). To ask whether our model could recapitulate and explain this feature, we calculated Yan occupancy across an element at equilibrium over a range of concentrations. Microstates were counted if they contained a bound Yan molecule at a given position on the element, starting at the ETS site and moving outward to the most distal site of the element. The results show that Yan fractional occupancy at all positions

increased with concentration and saturated to completion (Figure 2.2A).

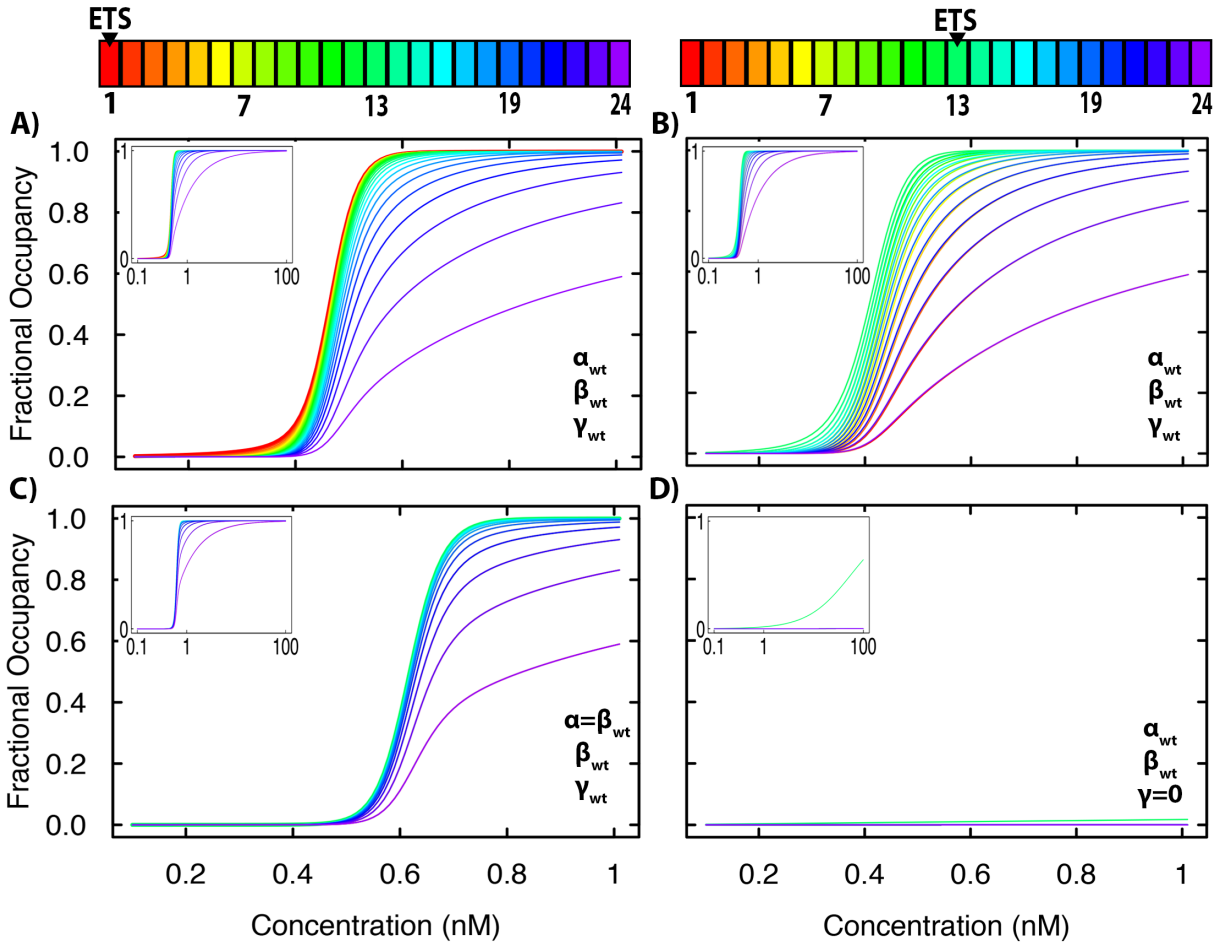


Figure 2.2: Yan occupancy spreads across the element at equilibrium, and depends on both protein-DNA and protein-protein interactions.

All plots show fractional occupancy of positions along a 24 site element as a function of concentration; insets plot the same data with concentration on a log scale. The ETS site is at position 1 in A and B, and at position 13 in C and D, as shown in the keys. A) Site by site occupancy using the wild type values for Yan binding parameters. Occupancy is highest at the ETS site (red) and progressively decreases at sites further from the ETS, suggestive of a spreading profile. B) Site by site occupancy for an element with the ETS site in the interior at position 13, using wild type values for α, β, and γ. C) Site by site occupancy when the specific DNA binding term is set equal to the wild type value for non-specific binding, i.e. an element without any specific binding sites. Occupancy is highest in the center of the element (green) and decreases symmetrically from the center. D) Site by site occupancy when the protein-protein interaction term (γ) is set to 0 kcal/mol. Significant occupancy is only observed at the ETS site for high concentrations (green curve at position 13). Note that identical lines are plotted on top of one another.

Interestingly, we noted a middle regime of concentrations where Yan preferentially occupied the ETS site with progressively less occupancy at sites further from the ETS site. This suggests that Yan can generate a broad binding profile across the element, and can do so at chemical equilibrium without requiring active mechanisms to establish this type of profile. We hypothesized that this could be explained by contiguous chains of Yan molecules nucleating at the ETS site and spreading outward towards distal, non-specific sites. To rule out the possibility that edge effects might be producing this behavior, we repeated the calculation for an element of the same size, but with the ETS site in the middle (Figure 2.2B). Yan occupancy spread outwards bi-directionally and equally from the ETS site and was not affected by the position of the ETS site within the element. Highest occupancy occurred at the ETS site (position 13, light green curve). Because the ETS site is positioned slightly asymmetrically in the center of the element, there was marginally higher fractional occupancy in positions 14 to 24 versus 1 to 12 (seen most clearly from 0.4 - 0.6nM). Based on these results, we conclude that the broad chromatin occupancy profiles of Yan observed *in vivo* could reflect binding at equilibrium.

Next, we assessed the sensitivity of the spreading pattern to changes in the strength of Yan's protein-DNA and protein-protein interactions. First, we explored the impact of Yan's protein-DNA interactions by varying α and β, the specific and non-specific DNA binding parameters. As a control, we set α and β equal to 0 kcal/mol, thus eliminating Yan protein-DNA interactions. No occupancy was detected at all concentrations considered (Figure 2.S3A), confirming that protein-protein interactions alone are insufficient to drive occupancy. To explore the contribution of sequence

specific DNA binding, we set α equal to β_{wt} in order to represent an element with no ETS site, and then recalculated Yan occupancy. In contrast to the preferential occupancy at the ETS site and spreading to the most distal site observed with the wild type parameters (Figure 2.2A), in the absence of a high affinity ETS site, maximal occupancy occurred at the middle positions of the element and then tapered distally in both directions (Figure 2.2C). This suggests that Yan is condensed on DNA, but that binding is not nucleated from the ETS site.

We also examined the role of the protein-protein interaction term γ in driving Yan fractional occupancy. When γ was set to 0 kcal/mol and α and β were held at their wild type values, occupancy at non-specific sites was reduced to nearly zero at all concentrations, and even the ETS site was only half occupied at a Yan concentration of 100nM (Figure 2.2D). The position of the ETS site had no effect on occupancy when the protein-protein interaction term was set to 0, (Figure 2.S3B), validating the essential contribution of SAM-mediated interactions. Additionally, we varied the strength of protein-protein interaction across a wide range from 0 to -12 kcal/mol and saw that this parameter strongly influenced occupancy across the element (Figure 2.S4). At weak values of γ (-5 to -6 kcal/mol, Figure 2.S4A-C) there was almost no occupancy, whereas at strong values of γ (-8 to -9 kcal/mol; Figure 2.S4G-I) the element was completely occupied at every site. In a narrow middle regime of protein-protein affinities (-6.5 to -7.5 kcal/mol; Figure 2.S4D-F), the spreading profile was observed. Fractional occupancy profiles with more extreme values of γ represented a continuation of the corresponding curves in Figure 2.S4 (data not shown).

Taken together these results highlight distinct roles for protein-DNA and protein-protein interactions in influencing Yan occupancy across the element. Both non-specific protein-DNA interactions and protein-protein interactions contribute strongly to Yan fractional occupancy. In contrast, specific DNA binding interactions contribute modestly to the level of occupancy, but impact the shape of the distribution by increasing occupancy at specific binding sites. Therefore, we hypothesize that specific DNA binding may facilitate recruitment to certain regions of the genome, while non-specific DNA binding may allow for spreading into more distal positions.

2.3.4 Nucleated microstates drive Yan fractional occupancy

One mechanism that could produce the observed occupancy profiles is contiguous chains of Yan spreading across the element via SAM-mediated interactions. To test this idea, we investigated which microstates contribute most strongly to the spreading profiles by calculating the fractional occupancy of "self-associated" microstates, which we define as microstates that have at least one contiguous chain of two or more Yan molecules. This definition encompasses dimers, trimers, and so on, up to a chain that spans the entire element. Of the self-associated microstates, the terms “nucleated" versus "non-nucleated" distinguish those in which the chain of Yan molecules includes or excludes the ETS site, respectively. Because nucleated and non-nucleated microstates are mutually exclusive sets, non-nucleated microstates can be calculated by removing nucleated microstates from the set of all self-associated microstates.

To explore further the role of sequence-specific binding interactions in nucleating Yan occupancy, we compared all self-associated, nucleated, and non-nucleated microstates under different conditions for α and β. When compared using the wild type

Yan parameters, nucleated microstates and self-associated microstates contributed almost equally to Yan occupancy, with non-nucleated microstates contributing negligibly across all concentrations (Fig. 2.3A). When α is equal to β_{wt}, meaning there is no contribution from sequence-specific interactions, nucleated microstates made up a much smaller proportion of the self-associated microstates until saturation was reached at ~10nM (Fig. 2.3B). There was also a middle regime of concentrations where non-nucleated microstates were more likely than nucleated microstates (~0.5-0.8nM), indicating that under these conditions, chains of Yan molecules are condensed on the element, but are not nucleated by the ETS site. This confirms that condensation can occur with just non-specific DNA binding, and suggests that removing the relative advantage of specific DNA binding allows microstates that do not encompass the ETS site to predominate. Thus α biases the distribution of bound molecules towards nucleated binding at the ETS site.

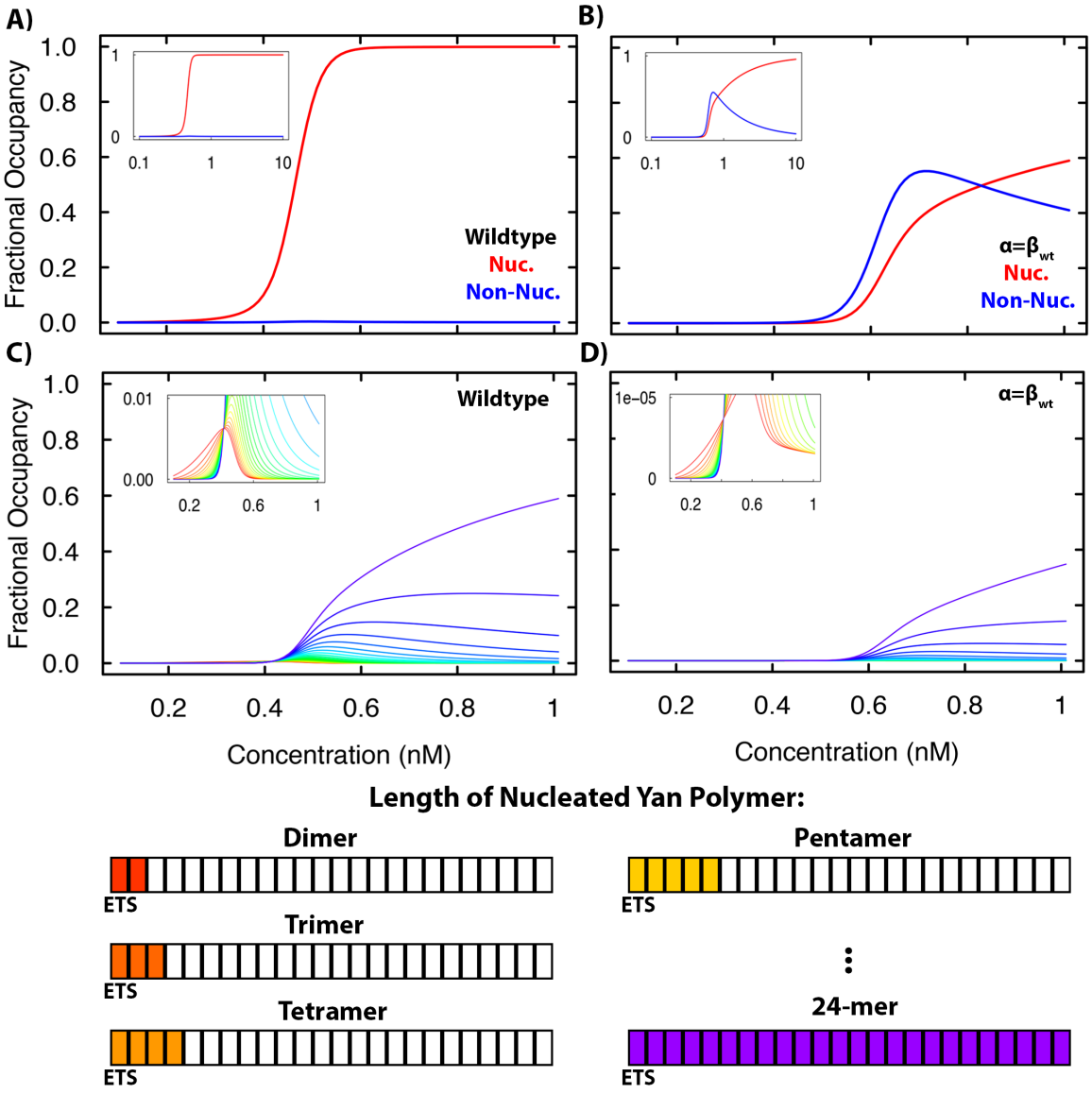


Figure 2.3: **Occupancy of nucleated microstates depends on specific DNA binding of Yan.**

All graphs show fractional occupancy of types of Yan microstates as a function of concentration. Insets show the same data with concentration on a log scale (A and B) or a zoomed view of data at low fractional occupancies on a linear scale (C and D). A) Fractional occupancy of nucleated microstates with two or more self-associated molecules versus non-nucleated microstates with two or more self-associated molecules, calculated for wild type parameters. B) Same as A, but with specific DNA binding term set equal to wild type values for non-specific DNA binding. C and D) Nucleated microstates with exactly *x* self-associated molecules, values of *x* from 2 to 24 are shown. Colors progress through values of x with red representing 2 and purple representing 24 (see key below). C) Fractional occupancy of nucleated microstates with exactly *x* self-

Figure 2.3 (cont.): associated molecules, for the wild type parameters of Yan. D) Same as C, but with specific binding set equal to the wild type value of non-specific binding.

Next we measured the exact distribution of chain lengths for nucleated microstates. As shown in Figure 2.3C, the fractional occupancy of nucleated microstates with exactly *x* molecules was plotted, up to 24 self-associated molecules. At low concentrations (<0.4nM), nucleated chains of 2 molecules were the most likely and chains of 24 molecules the least likely. At ~0.4nM, all curves passed through a single point of equal fractional occupancy, and at higher concentrations, chains of 24 molecules were the most likely nucleated microstates (Fig. 2.3C). When we repeated the calculation with α equal to β_{wt} (Fig. 2.3D), there was weaker fractional occupancy for all chain lengths considered. In addition, the low concentration regime behavior had significantly lower fractional occupancy (compare scales of 0-0.01 and 0-0.00001 in the insets of Fig. 2.3C and 2.3D), and the point where all curves have equal fractional occupancy occurred at the same concentration of Yan. We conclude that chain length within nucleated mircostates follows the broader trend of Yan occupancy with respect to concentration—at low concentration, the number of Yan molecules is limiting, and at a high concentration, large chain lengths predominate. This relationship holds true when α is equal to β_{wt}, even though fractional occupancies are decreased across the board.

2.3.5 Exploration of parameter space

Our initial results suggested distinct roles for protein-DNA and protein-protein interactions in determining Yan binding across the element. To test the relationships between these interactions, we systematically co-varied α, β, and γ and reiterated the calculation over a wide area of parameter space. The fractional occupancies of all nucleated microstates for a given parameter set were calculated and plotted as a spectral

heat map. We fixed Yan concentration at 0.1 nM for all spectral heat maps shown in Figure 2.4, because at this concentration, the broad trends upon changing α, β, and γ are most easily distinguished.

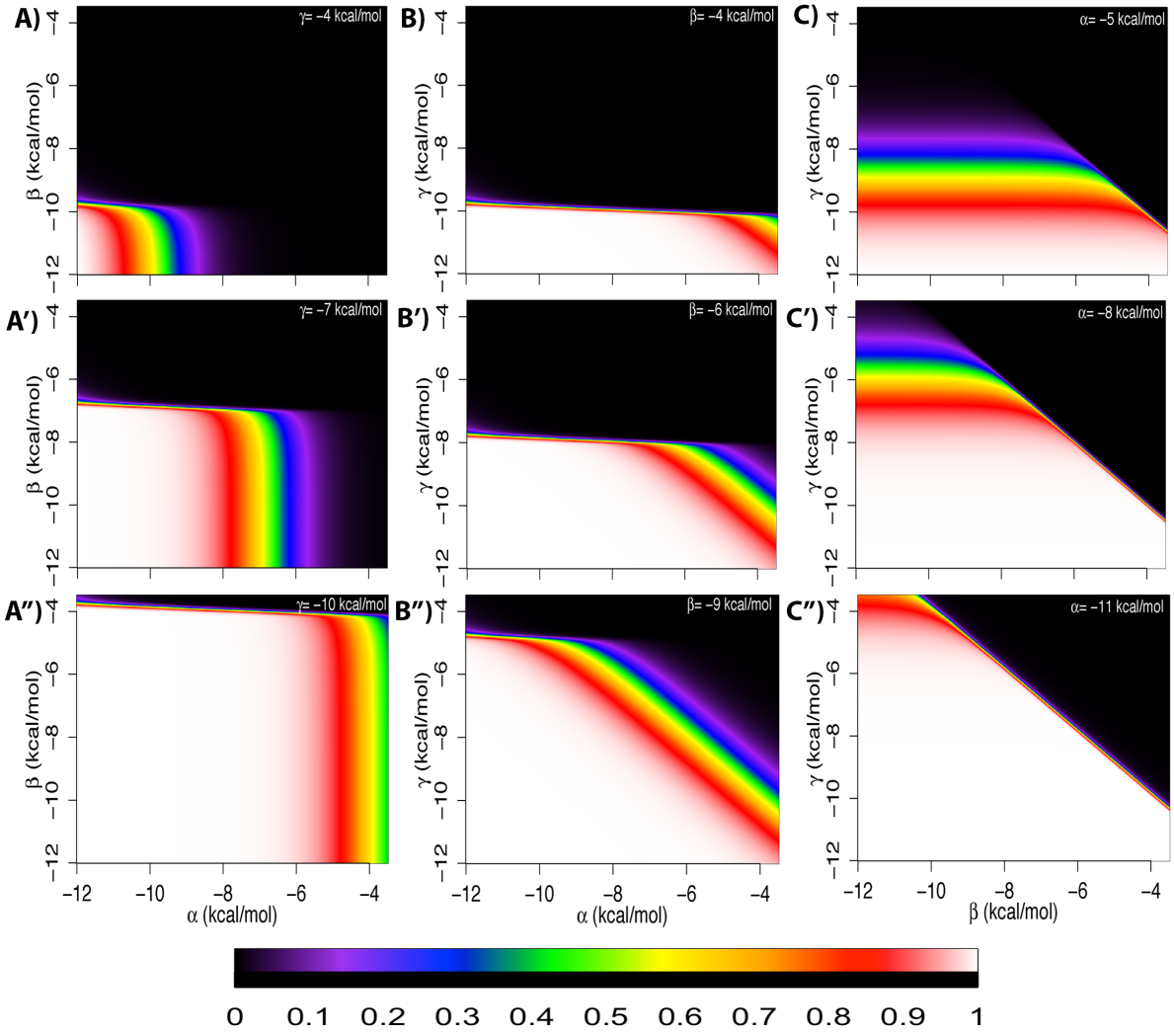


Figure 2.4: Exploration of parameter space for nucleated, self-associated microstates.

All graphs are spectral heat maps plotting fractional occupancy of nucleated microstates with two or more self-associated molecules. Fractional occupancy of 0 is represented in black, and fractional occupancy increases moving through the visible color spectrum, ending with fractional occupancy of 1 represented in white (see scale bar below). Values of the parameters α, β, and γ are plotted along the x or y axes from -3.5kcal/mol to -12.0kcal/mol in 0.125kcal/mol increments. All heat maps shown are calculated at a concentration of Yan of 0.1 nM. A-A") α versus β; increasing values of γ are shown

Figure 2.4 (cont.): descending in the column (-4.0, -7.0, -10.0 kcal/mol, respectively) B-B") α versus γ; increasing values of β are shown descending in the column (-4.0, -6.0, -9.0 kcal/mol, respectively) C-C") β versus γ, increasing values of α are shown descending in the column (-5.0, -8.0, -11.0 kcal/mol, respectively)

We first mapped the relationship between α and β as γ was increased (Fig. 2.4A). At lower values of both α and β, nucleated fractional occupancy was increased. Consistent with non-specific interactions contributing more strongly to occupancy than specific interactions (Fig. 2.2), occupancy increased sharply for small increases in β (e.g. -6.75 kcal/mol to -7 kcal/mol, Fig. 2.4A') as compared to α, which slowly increased occupancy over a wide range of values (e.g. -5 kcal/mol to -9 kcal/mol, Fig. 2.4A'). However, once occupancy was achieved at a given value of α, additional strengthening of β did not change the nucleated fractional occupancy. Thus, without non-specific DNA binding interactions, significant occupancy cannot be achieved; specific DNA binding interactions instead determine the extent that binding occurs at the ETS versus other positions. Increasing the strength of γ resulted in greater nucleated occupancy. Specifically, as γ was increased, the sharp transition in occupancy occurred at weaker values of β, although the total width of the transition in α was unaffected (i.e. the overall shape of the map translocates to the upper right corner, but does not sharpen or change shape; compare Fig. 2.4A to 2.4A' and 2.4A' to 2.4A"). This was also the case if Yan concentration was increased—the sharp transition in β occurred at lower values while the transition in α was unaffected (Figure 2.S5A'-A'').

We next examined the relationship between α and γ, with increasing strength of β (Fig. 2.4B). Nucleated occupancy increased by strengthening all three parameters. Reminiscent of the plots in Fig. 2.4A-A", there was a sharp transition in occupancy for γ (-7.75 kcal/mol to -8.0 kcal/mol, Fig. 2.4B') and a broad, fixed-width transition for α.

However, this transition was along a diagonal slope, suggesting a compensatory trade-off between α and γ. This suggests that γ plays a role in increasing both occupancy at the element and nucleated occupancy, in contrast to β where strengthening the interaction increases occupancy but not nucleated occupancy. Neither the width of the transition in α nor the sharpness of the transition in γ changed as the strength of β increased, and increasing Yan concentration maintained these relationships (Figure 2.S5B-B'').

Lastly, we plotted the relationship between β and γ as α was increased (Figure 2.4C-C''). The nucleated fractional occupancy did not depend on the value of β as evidenced by the horizontal stripes of color, representing a wide range of values of β (analogous to the vertical stripes in Fig. 2.4A-A"). Additionally, the negative slope of the diagonal suggested a compensatory trade-off for β and γ in determining occupancy, meaning that the same occupancy was achieved by strengthening γ and weakening β (and vice versa). Increasing the strength of α translocated the figure towards higher fractional occupancy but kept the same steepness of the transition. At increased concentrations, the shape of the graph translocated, but the sharpness of the transition in α was unchanged (Figure 2.S5C-C''). Taken all together, these results suggest that β and γ are both required for occupancy at the element, and that loss of strength in one interaction can be compensated by increasing the other; α does not have a dominating role in determining occupancy. In contrast, in the context of nucleated occupancy, α plays an important role and has a compensatory relationship with γ, while β is less influential.

2.3.6 Spreading is robust to variation in non-specific binding

In the calculations described above, the non-specific DNA binding affinity term β represented sequence-independent electrostatic binding to the DNA phosphate backbone

but ignored the effects of sequence heterogeneity. To capture the assumption that DNA binding at non-consensus sites will occur with a range of affinities according to differences in nucleotide content, we next asked how sensitive the patterns of Yan occupancy are to variation in non-specific DNA binding. Using the wild type value of β as a minimum threshold, we drew values of β from a half Gaussian distribution with a fixed standard deviation in order to create random non-specific sites with varying affinities, and then recalculated Yan occupancy. Increasing the variation in β by increasing the standard deviation gradually shifted the profile toward greater occupancy at lower concentrations, eventually producing profiles where non-specific sites had greater occupancy than the ETS site (Figure 2.S6A-C). To quantify the point where variation overwhelms the normal spreading profile, we varied the standard deviation of β and calculated occupancy for 100 random elements for each value. At the concentration where the ETS site is at half maximal occupancy, we compared fractional occupancy of all other positions to the ETS site—if the ETS site had greater occupancy than any other position, we considered the spreading profile to be intact. Figure 2.S6D shows the fraction of elements with spreading profiles for various values of the standard deviation. At a value of 0.6, which corresponds to up to a 55-fold increase in affinity in β (-5.837 kcal/mol to -8.287 kcal/mol), the spreading profile was equally likely as not. As a 55-fold increase in affinity approaches that of a sequence specific binding interaction, the calculation provides a complete exploration of the effects of non-specific binding across the plausible affinity range. Therefore we conclude that the spreading pattern is robust with respect to heterogeneity in non-specific DNA binding.

2.3.7 Clustering of multiple ETS sites increase fractional occupancy

Given that TF binding sites tend to be clustered into regulatory elements (Suryamohan and Halfon 2015), the number and spacing between ETS binding sites could play important roles in determining Yan occupancy. We therefore expanded the model to include two specific ETS binding sites separated by an integer number of non-specific binding sites. Using the wild type parameters, we repeatedly calculated fractional occupancy at one ETS site as the second ETS site was moved step-wise from the opposite end of the element towards the first site, and then plotted this as a function of Yan concentration (Figure 2.5). Occupancy at the first ETS site increased as the opposing site was moved closer, with maximal occupancy achieved when the two sites were arranged as a tandem repeat. In all configurations, occupancy at two-site elements was greater than that at a single-site element (Figure 2.5A). Thus clustering transcription binding sites increases their occupancy in a distance dependent manner. To confirm the role of protein-protein interactions in generating this behavior, we repeated the calculation when γ equals 0 kcal/mol, and found that the distance dependent increase in occupancy disappeared (Figure 2.5B). We conclude that the arrangement of binding sites relative to one another within an element can stabilize occupancy in a manner that depends on protein-protein interactions bridging the binding sites.

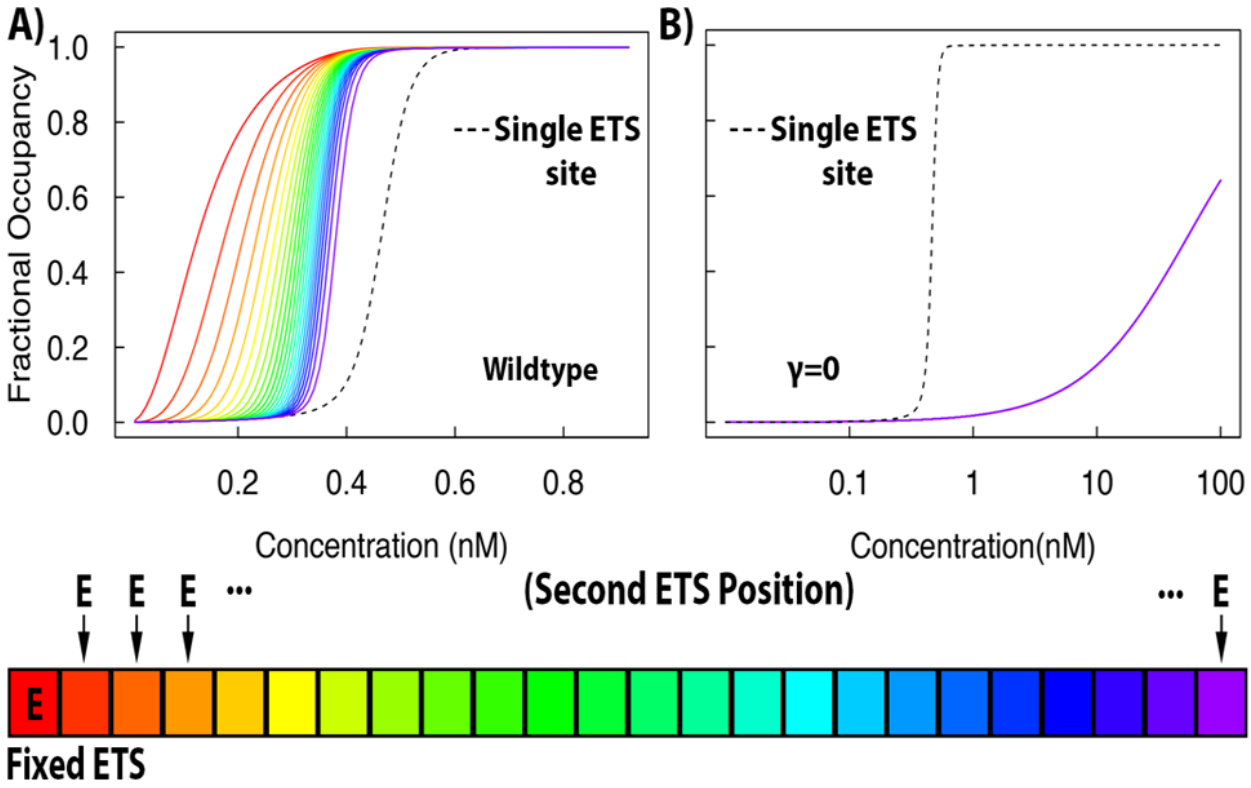


Figure 2.5: Clustering multiple ETS sites within an element increases occupancy at the sites.

One ETS site is held fixed, while another ETS site is moved from the adjacent position (orange) to the most distal position (purple). The plotted fractional occupancies are measured at the fixed ETS site as a function of concentration. Dashed lines denote fractional occupancy at the fixed site without an additional ETS site in the system. A) Fractional occupancy with the wild type parameters of Yan B) Fractional occupancy when the protein-protein interaction term is set to 0 kcal/mol. Note that all lines are identical in fractional occupancy and plotted on top of one another.

We also noted in the wild type parameter calculation that as the two ETS sites were brought together, the slope of the occupancy curves became less sharp (Figure 2.5A). This suggests that site clustering exerts a stronger effect on weakly or partially occupied binding sites as compared to strongly occupied binding sites—for example, fractional occupancy of 0.1 is achieved at ~6-fold lower concentration by clustering sites (0.05 versus 0.35 nM), whereas fractional occupancy of 0.9 is achieved at a ~1.5-fold lower concentration by clustering (0.27 versus 0.42 nM). Therefore, clustering TF

binding sites may preferentially result in partial occupancy of these sites at equilibrium, consistent with an emerging role for the conservation and function of medium to weak affinity TF binding sites in regulatory element function (Ramos and Barolo 2013; Parker et al. 2011; Crocker et al. 2015).

2.3.8 Restricting the extent of Yan polymerization results in preferential occupancy at specific sites

Because many TFs operate as dimers, we asked which features of the Yan occupancy profiles could be explained by dimeric interactions and which required higher order polymerization. When we limited Yan to forming dimeric complexes, occupancy was dramatically reduced for all sites across an element (Figure 2.6A, solid lines) as compared to polymeric Yan (Figure 2.6A, dashed lines; coloring by position across the element is the same as in Figure 2.2). Interestingly, there was strong occupancy at the ETS site and the immediately adjacent non-specific site, and greatly reduced occupancy at all other sites. For example, focusing on the concentration regime around 10 nM, fractional occupancies were ~0.9 for the ETS and adjacent site (red solid lines), but near 0 elsewhere (blue/purple solid lines). These results predict that TFs that typically dimerize will have strikingly different chromatin occupancy profiles than those that polymerize. Specifically, whereas polymers can condense on DNA in the absence of specific binding interactions (Figure 2.2D), dimers will only stably occupy specific DNA binding sites. Additionally, the relative difference between binding at specific versus non-specific sites was much greater for dimers than for polymers. Although the fractional occupancy of a dimer was lower than that of a polymer at a given Yan concentration,

dimers were better able to discriminate between specific binding sites and non-specific sites, and did so over a wide concentration range.

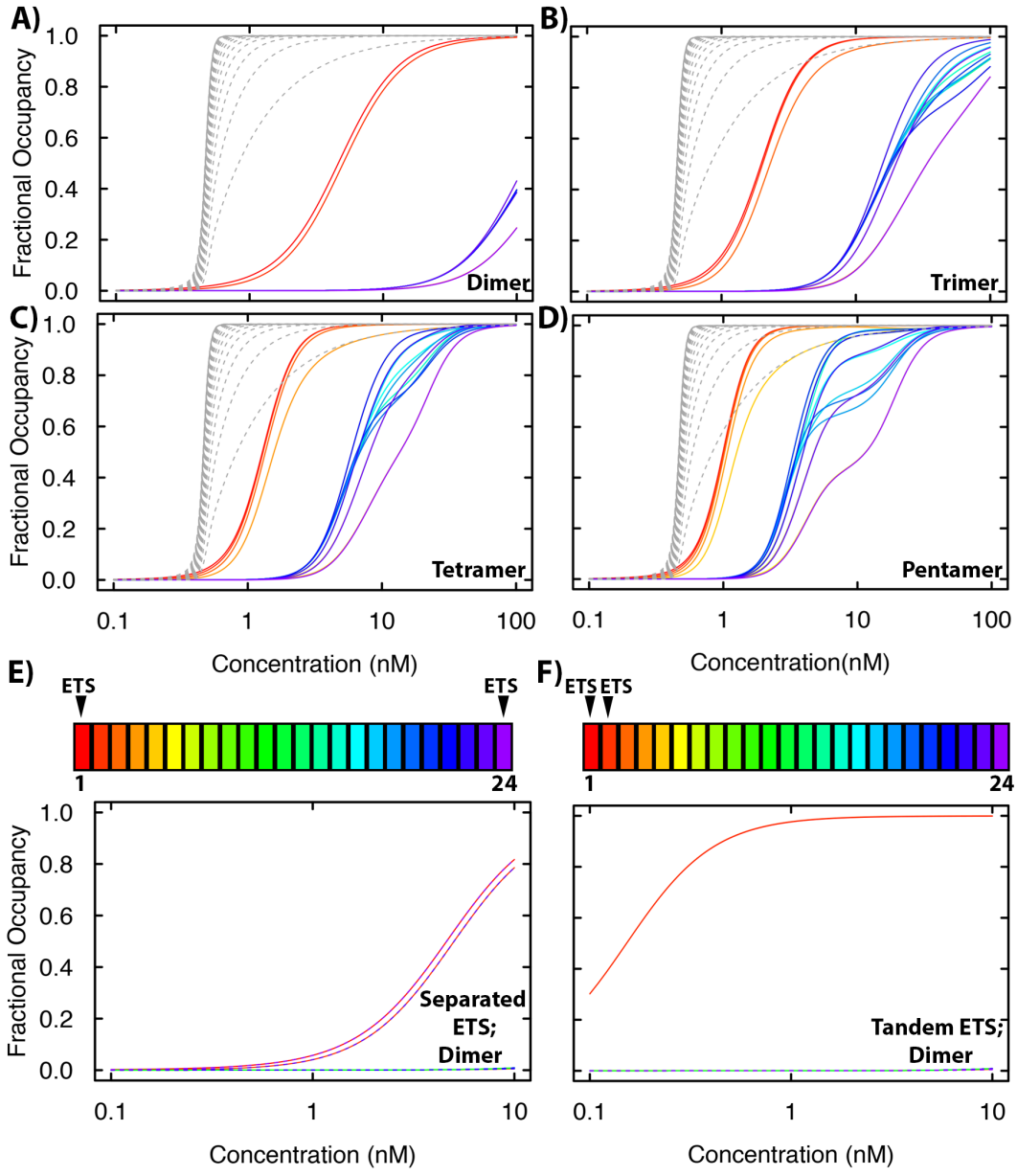


Figure 2.6: Restricting Yan polymerization decreases occupancy at distal sites but maintains occupancy at specific sites, especially tandem ETS sites.

Fractional occupancies are plotted as a function of concentration as in Fig. 2 with red curves representing the ETS site and purple curves representing the most distal site. Wild type data from Fig. 2A is plotted as gray dashed lines. Wild type parameters for α, β, and γ were used for all calculations. A) Yan restricted to dimers B) Yan restricted to trimers

Figure 2.6 (cont.): C) Yan restricted to tetramers D) Yan restricted to pentamers. E and F) Fractional occupancy of Yan restricted to dimers, with multiple ETS sites E) Fractional occupancy at an element with two maximally separated ETS sites. Dashed lines are shown for some positions 13 through 24 to show overlap of curves. F) Fractional occupancy at an element with a tandem pair of ETS sites at positions 1 and 2.

Figures 2.6B-D show the results of extending the calculation to consider trimers, tetramers, and pentamers. Occupancy remained decreased relative to polymers, but increased relative to dimers. Similar to the pattern with dimers, there was strong occupancy at the ETS site and immediately adjacent sites, but far less occupancy at all other sites. The gap in occupancy between the ETS adjacent sites and all other sites decreased as the extent of polymerization increased, suggesting that the ability to discriminate between specific and non-specific sites decreases as polymerization increases. For trimers and above, the fractional occupancy curves at distal sites were not smooth sigmoids, but showed more complicated behavior that arises from edge effects at the end of the element. For example, a trimer cannot fit into two vacant sites near the edge of the region modeled. Such numerical artifacts were not observed in earlier calculations (Figure 2.2B) because polymers fill the element completely with no partial chains at the edge. Taken together, these results suggest that Yan's spreading profile is highly dependent on the extent of polymerization, and by extension, that typical dimerizing TFs will exhibit significantly different DNA occupancy profiles under the same conditions.

Given that recruitment to specific sites was reduced for dimers compared to polymers, we asked if clustering of ETS sites could restore occupancy to the level observed for the polymer. To do so, we compared occupancy of Yan dimers across two elements, one in which the two ETS sites were maximally separated and one in which the two ETS sites

were immediately adjacent. Similar to the single ETS element calculations (Figure 2.6A-D), Yan dimers were recruited equally to the separated ETS sites and sites immediately adjacent to the ETS sites, while occupancy at all other sites was negligible (Figure 2.6E). In contrast, when ETS sites were placed immediately adjacent to one another, occupancy was much higher at both sites and rivaled the recruitment to tandem sites observed with polymers (compare Fig. 2.6F to Fig. 2.5B). We conclude that Yan dimers and polymers can exhibit equivalent recruitment to a tandem pair of ETS sites, but not to single ETS sites.

2.4 Discussion

In this study, we describe a mathematical model of Yan DNA binding that provides a new framework for understanding its *in vivo* occupancy patterns and for exploring the unique behaviors that polymerization can confer to a transcription factor. We find that Yan occupancy spreads across DNA at equilibrium, and that this behavior shows distinct requirements for both protein-DNA and protein-protein interactions. More broadly, our results predict that a polymerizing versus non-polymerizing TF will form different kinds of complexes on a given enhancer element, with respect to both the degree of occupancy and the requirement for sequence specific DNA binding sites. We propose that a key role of polymerization is to recruit and stabilize TF accumulation across regions of an enhancer that lack specific binding sites. In contrast, DNA occupancy of monomeric or dimeric TFs is more tightly restricted to specific sites of recruitment. This result may offer insight into why so many transcription factors have evolved to operate as dimers and further suggests that emphasizing the contributions of tandem versus individual binding sites might improve the success of both bioinformatic predictions and

experimental manipulations of enhancers. We also note that although the model was formulated for Yan, it extends in principle to any TF or combinations of interacting TFs, with the only differences in the precise values of α, β, or γ.

A major conclusion from our calculations is that polymerization-driven TF spreading across DNA can occur at equilibrium. The traditional assumption is that TFs are first recruited to DNA via recognition of high affinity sequence-specific binding sites and then spread via protein-protein interactions. Our model similarly predicts highest occupancy at specific sites with decreasing occupancy distally, but reveals that a precise, step-wise kinetic process is not required to produce such a profile. Instead, our model is path independent, which suggests that spreading of a polymerizing TF occurs because it is thermodynamically most favorable, and that non-polymerizing TFs do not spread primarily because the complexes are thermodynamically less favorable. This finding also bears implications for the spreading of repressive chromatin complexes assembled by either the Polycomb group (PcG) proteins or the Silent Information Regulator (SIR) factors (Kueng et al. 2013; Grossniklaus and Paro 2014). Traditionally, these systems have been thought to function analogously to polymerizing TFs in that they are recruited by specific binding of one protein (to a DNA sequence or histone mark), followed by binding and polymerization of another protein (Grewal and Moazed 2003). However, there is also thought to be a requirement for cycles of action, such as recruiting the histone modification enzyme that installs the original mark, thus resulting in a positive feedback loop (Grewal and Moazed 2003). Our results suggest that, in addition to those kinetic processes, the binding profile of spreading regulators could be substantially influenced by polymerization at equilibrium, depending on the strength of protein-DNA

and protein-protein interactions. Conversely, the occupancy patterns of a polymerizing TF like Yan or TEL will almost certainly be modified by cycles of post-translational modification, recruitment of cofactors and other TFs, and perhaps interactions with PcG or other polymerized repressive complexes. Supporting the notion of heteropolymeric repressive complexes, the polymerizing PcG protein L(3)MBT has been shown to complex directly with TEL1 via a SAM-SAM interaction (Boccuni et al. 2003; Knight et al. 2011).

Based on the thermodynamics of binding at equilibrium, we speculate further that the distinct occupancy patterns calculated for a polymerizing transcription factor produce intrinsically repressive regulatory behavior that is perhaps incompatible with precise activating regulation of gene expression. First, our results suggest that multimerization can be viewed as a trade-off between occupancy and discrimination of proper binding sites; as the extent of polymerization increases, the gap between specific and non-specific occupancy narrows. Because activating TFs may have a greater requirement for binding precision, restricting them to dimers enables the sharpest distinction in occupancy of specific versus non-specific sites. In contrast, repressors may use both active sequence-specific mechanisms and passive sequence-non-specific mechanisms of gene regulation. In the latter scenario, transcription factor polymerization would drive the spreading of oligomeric complexes into distal, non-specific sites on DNA, thereby effectively occluding access to other factors, both repressive and activating. Although a polymerized activator would have a similar effect, because activating complexes are often comprised of collectives of different TFs bound to clustered specific binding sites in enhancers, with the combinatorics of binding important for regulatory specificity (Junion et al. 2012),

general steric interference would seem counterproductive as it would either inappropriately repress gene expression or lead to uncontrolled activation. As noted above in the text, clustering of binding sites produces a greater fold increase in occupancy for lower affinity sites as opposed to higher affinity sites (Figure 2.5A), which may explain the preference for clustered, weak sites in driving transcriptional activation in some developmental enhancers (Ramos and Barolo 2013; Parker et al. 2011; Crocker et al. 2015). Given that introduction of high affinity binding sites has been shown to disrupt the expression patterns of these enhancers, either by reducing levels of expression or inducing ectopic expression, it is reasonable to assume that excess TF occupancy may be deleterious at these elements. Therefore, polymerization might be counterproductive not only for reducing binding site discrimination, but also for globally increasing occupancy beyond threshold levels at sensitive developmental enhancers.

In this context, it is interesting to consider the role that polymerization might play in transcriptional mis-regulation driving oncogenesis (De Braekeleer et al. 2012). Translocations involving the *TEL/ETV6* gene (the human homolog of Yan) are prevalent in human leukemia, and several of these link the N-terminal exons that encode the polymerizing SAM in frame to other TFs. For example, the most common cause of childhood acute lymphoblastic leukemia is a reciprocal translocation that fuses the TEL SAM onto the RUNX family transcription factor AML1 (De Braekeleer et al. 2012). Mechanistically, we hypothesize that conversion of AML1 into a polymerizing TF may permit spreading into regions of chromatin it would not usually occupy. Although the conversion to polymerizing TF could also fundamentally alter AML1 transcriptional

activity, even absent an activity change, altered occupancy on DNA could substantially dysregulate the cell's transcriptional program and lead to oncogenesis.

While translocations involving TEL are thought to generate polymerizing repressive complexes, a handful of oncogenic translocations that target other TFs are hypothesized to generate polymerizing activating complexes. Prime examples are the reciprocal translocations that fuse the polymerizing low-complexity (LC) transactivation domains of the TET-family proteins, Ewing Sarcoma protein (EWS) or Fused in Ewing Sarcoma (FUS), with the DNA-binding domain of the ETS-family members FLI1 or ERG (Kwon et al. 2014; Paronetto 2013). Based on our modeling results, we speculate that the reduced requirement for sequence specific binding associated with polymerization will produce more promiscuous hyper-activating complexes. If correct, this would argue that TF polymerization is compatible with repression but not with precise activation of transcription.

Returning to Yan, our observation that the polymerization state of Yan can facilitate the formation of distinct complexes on DNA has explanatory power for several aspects of Yan biology. First, experiments comparing the extent of polymerization required for full *in vivo* function have concluded that while Yan monomers have extremely limited repressive ability, Yan dimers are almost, but perhaps not quite, as functional as wild type polymerization-capable Yan (Zhang et al. 2010; Webber et al. 2013). Parallel experiments addressing this question for human TEL have similarly concluded that dimerization confers significant, but not complete, function(Tognon et al. 2004). Based on our model, we speculate that the relatively strong recruitment of Yan/TEL dimers to clustered, specific binding sites may be sufficient to execute a

majority of their transcriptional regulatory functions. However at certain enhancers and/or in specific developmental contexts, the broader occupancy conferred by polymerization must confer biologically significant regulation.

Second, polymerization at equilibrium could theoretically explain some features of actual Yan chromatin occupancy patterns. Briefly, a meta-analysis of chromatin occupancy patterns of *Drosophila* TFs concluded that Yan DNA-bound regions are broader than those associated with typical non-polymerizing TFs (Webber et al. 2013). Consistent with our findings that clustering of specific binding sites can increase Yan occupancy, the top 600 Yan-bound peaks showed an enrichment for both number of ETS sites and number of clustered ETS sites (Webber et al. 2013). To consider how our model of polymerization at equilibrium might explain observed peaks, we consider the prominent subset of extensively Yan-bound regions termed high density regions (HDRs). Yan HDRs were measured to be 4kb in length; taking into account that chromatin immunoprecipitation (ChIP) of sonicated DNA fragments will define regions broader than those actually bound by the TF (Landt et al. 2012), the actual length of an average HDR might be closer to 3kb. The ETS domain of TEL (Yan's human homolog) bound to specific DNA has been crystallized, and shown to have a footprint of about 10bp (De et al. 2014), which would translate to 240 base pairs of consecutive footprints for a 24 site element. *In vitro* DNA binding analyses have showed that TEL dimers can bind cooperatively to ETS sites separated by up to 50bp (Green et al. 2010), which if true *in vivo*, could increase the size of bound regions by about a factor of five, permitting occupancy up to 1500bp in length. Our model is agnostic to the actual conformation of DNA, and so if one assumes that short stretches of intervening DNA can be looped out without impacting polymerizing

interactions, then the effective binding range might increase further. Assuming these estimates are reasonably accurate, then a majority of the broad, Yan-binding peaks observed could be explained by polymerization at equilibrium. Additionally, we note that our previous work comparing the genome-wide occupancy profile of Yan monomers to that of wild type Yan showed that the median size of Yan HDRs was reduced by ~20% (Webber et al. 2013). This indicates that polymerization contributes to the Yan occupancy profile, but suggests that additional interactions contribute to the ultimate binding pattern. We speculate that cooperation with other transcriptional regulatory complexes could stabilize Yan monomer binding and lead to an underestimate of the contribution of polymerization at equilibrium to Yan occupancy. Given that Yan operates in a milieu that includes many other TFs, future iterations of our model will incorporate cooperative and competitive interactions with other TFs that could positively or negatively influence polymerization-driven Yan occupancy.

2.5 Materials and Methods

2.5.1 Calculation of fractional occupancy at equilibrium

We represent the genome as a series of discrete, potential binding sites for Yan, arranged in a one-dimensional lattice, which we refer to as an element (Figure 2.1A). A pattern of Yan occupancy across an element is referred to as a microstate. At constant temperature and pressure, the fractional occupancy of a microstate *k* at equilibrium is given by

$$f_k = \frac{\exp\left(-\Delta G_k/\mathrm{RT}\right)[\mathrm{Yan}]^{\mathrm{a(k)}}}{\mathrm{Z}}.$$

In this case, ΔG_k is the change in Gibbs free energy upon Yan binding for microstate *k*, R is the gas constant (1.987204118 x 10^3 kcal/mol degree), T is the absolute temperature (300 degrees K for all calculations), [Yan] is the concentration of free molecules of Yan in the nucleus, a(*k*) is a function representing the number of bound Yan molecules in microstate *k*. The numerator of Equation 1 is a Boltzmann weight for the microstate. The partition function, Z, is the sum of Boltzmann weights for all microstates considered, and is given by

$$Z = \sum_k \exp\left(-\Delta G_k / RT\right)[Yan]^{a(k)}.$$

Binding sites within an element come in two types, specific and non-specific, with non-specific sites greatly outnumbering specific sites, which we refer to as ETS sites. ETS sites have a greater affinity for Yan binding than non-specific sites, and non-specific sites are taken to be of uniform, lower affinity for Yan. Molecules of Yan can either be bound or unbound at each position on the element in a binary fashion, and each molecule of Yan can bind only a single site at one time. Unless otherwise noted in the text, we considered a class of elements characterized by one ETS site in the right-most position, flanked by *n* non-specific binding sites to the left. In this system, the total number of sites is *n* + 1, and the total number of microstates is equal to 2^{n+1}.

To calculate the fractional occupancy of microstates of interest, our model requires five parameters: α, β, γ, [Yan], and *n* (Figure 2.1A). α represents the change in Gibbs free energy of specific binding to ETS sites, while β represents the change in free energy of non-specific Yan binding to lower affinity sites. Given the affinity differences between the two types of binding sites, α will always have a lower ΔG than β. γ

represents the change in free energy of association between two adjacent Yan molecules via their SAM domains. Reflecting Yan's ability to polymerize, each additional adjacent Yan molecule contributes an additional increment of γ. The free energy of Yan binding is fully parameterized by α, β, and γ. Thus, the sum of interactions in a microstate, in terms of α, β, and γ, reflects the exact configuration of Yan molecules and defines the free energy of the microstate, ΔG_k.

The fractional occupancy of all microstates of a given element were calculated, from k=0 to k=2^{n+1}-1. The free energies of all microstates of an element of length n+1 can be calculated iteratively from smaller elements, from n=0 up to the specified size. For the special case where n=0, which corresponds to an element with just one ETS and no non-specific sites, there are only two free energies possible: $\Delta G_{k=0} = 0$ and $\Delta G_{k=1} = \alpha$. The free energy of any microstate in an element of size n+1 (where n does not equal zero) is given by

$$\begin{aligned} \Delta G_{k,n+1} &= \Delta G_{k,n} && \text{iff site } n\text{+1 unbound} \\ &= \Delta G_{k,n} + \beta && \text{iff site } n\text{+1 bound; site } n \text{ unbound} \\ &= \Delta G_{k,n} + \beta + \gamma && \text{iff site } n\text{+1 bound; site } n \text{ bound.} \end{aligned}$$

The calculation was ordered in such a way that the molecular configuration of Yan molecules in a microstate was equivalent to the binary representation of the index of the microstate (Figure 2.1B). For instance, k = 5 represented in binary notation is **1 0 1**, and k = 5 indexes a microstate where two bound Yan molecules are separated by one empty

non-specific binding site. Microstates of interest can be returned by integral arithmetic operations on the index of the microstate. For example, the states where the ETS site is occupied are solutions of the equation k=1 (mod 2), easily expressible in C++ by the elementary remainder operator, "%".

2.5.2 Computation considerations

All calculations were performed using C++ compiled with gcc ver.4.4.7, and are available for download at https://uchicago.box.edu/v/YanPolymerization, or by request. Unless otherwise noted, 24 site elements were chosen for computation. For an element of this size, there are 16,777,216 distinct microstates. For calculation of fractional occupancy by position within an element (Figures 2.2, 2.5, and 2.6), microstates were selected if they contained a bound Yan molecule at the position considered, which was implemented by bit-wise operations on the index of the microstate. For calculation of self-associated microstates (Figures 2.3A-B), microstates were counted if two or more molecules were adjacent to one another in the lattice, which was implemented by tracking adjacent positions with an integer flag. Microstates were counted if the adjacent position is occupied (i.e. flag is raised) and the current position is occupied. Nucleated microstates (Figure 2.3A-B) were counted if the ETS site is occupied and a self-associated ligand is bound at the adjacent site. Thus, the expression "k%3==0" returns the set of microstates of interest. For calculation of exact nucleated chain length in Figures 2.3C and 2.3D, nucleated microstates were counted as before, with the additional caveat that the binding site adjacent to the end of the chain be unoccupied. In this way, nucleated chains of Yan exactly x molecules (but not x+1) can be counted—the only exception being the singular microstate with 24 self-associated molecules. It is important to note that microstates with

chains of self-associated Yan molecules greater than *x* will be counted by this method; however, the measurement of fractional occupancy concerns nucleated microstates, which by definition requires binding at the ETS site, and not the broader set of self-associated microstates. For the calculation of spectral heat maps in Figure 2.4, Boltzmann weights were calculated in log space to prevent overflow and then converted. Each spectral heat map represents 341 by 341 calculations.

In calculations where the extent of Yan polymerization was restricted to dimers, trimers, tetramers, and pentamers (Figure 2.6), the free energy values of microstates were calculated in a different manner than specified in equation 3. Whereas each additional adjacent bound molecule contributes an additional increment of γ in all other calculations, when polymerization is restricted, values of γ are added up to the maximum permitted size of Yan complex. For example, when Yan is restricted to trimers, three adjacent molecules contribute two increments of γ to the free energy of a microstate, but an additional fourth adjacent molecule represents the start of a new trimer chain and does not contribute an additional increment of γ. Computationally, the extent of polymerization in a microstate was determined by a series of integer flags. Microstates were assumed to be making the maximum permitted number of protein-protein interactions allowed for the permitted size of the Yan complex. For example, four adjacent molecules were counted as "Y+Y+Y+Y" for unrestricted Yan polymers, where "+" denotes an interaction. However, when Yan is restricted to dimers, four adjacent molecules were counted as "Y+Y,Y+Y", and not as “Y,Y+Y,Y”, which results in two values of γ added instead of one.

2.6 Acknowledgements

The authors thank Trevor Davis, James Fuller, Nicelio Sanchez-Luege, Jemma Webber, and members of the Reinitz and Rebay labs for helpful comments on the manuscript, and Kenneth Barr, Zhihao Lou, and Nicelio Sanchez-Luege for help in implementing the code. MH and IR are supported by NIH grant R01 GM080372. MH has been supported by NIH grants T32 GM007183 and 2T32 HL007381-36A1. JR is supported by NIH grant 2R01 OD10936.

2.7 Author Contributions

M.H., I.R., and J.R. designed experiments; M.H. and J.R. wrote the code; M.H. performed the experiments; M.H., I.R., and J.R. analyzed data; M.H., I.R., and J.R. wrote the paper. The authors declare no conflicts of interest.

2.8 References

Biggin, Mark D. 2011. “Animal Transcription Networks as Highly Connected, Quantitative Continua.” *Developmental Cell*.

Boccuni, Piernicola, Donal MacGrogan, Joseph M. Scandura, and Stephen D. Nimer. 2003. “The Human L(3)MBT Polycomb Group Protein Is a Transcriptional Repressor and Interacts Physically and Functionally with TEL (ETV6).” *Journal of Biological Chemistry* 278 (17):15412–20.

Braekeleer, Etienne De, Nathalie Douet-Guilbert, Fr??d??ric Morel, Marie Jos??e Le Bris, Audrey Basinko, and Marc De Braekeleer. 2012. “ETV6 Fusion Genes in Hematological Malignancies: A Review.” *Leukemia Research*.

Crocker, Justin, Namiko Abe, Lucrezia Rinaldi, Alistair P. McGregor, Nicolás Frankel, Shu Wang, Ahmad Alsawadi, et al. 2015. “Low Affinity Binding Site Clusters Confer HOX Specificity and Regulatory Robustness.” *Cell* 160 (1–2):191–203.

Davidson, Eric H, and Michael S Levine. 2008. “Properties of Developmental Gene Regulatory Networks.” *PNAS* 105 (51):20063–66.

De, Soumya, Anson C K Chan, H. Jerome Coyne, Niraja Bhachech, Ulrike Hermsdorf, Mark Okon, Michael E P Murphy, Barbara J. Graves, and Lawrence P. McIntosh.

2014. “Steric Mechanism of Auto-Inhibitory Regulation of Specific and Non-Specific Dna Binding by the Ets Transcriptional Repressor ETV6.” *Journal of Molecular Biology* 426 (7):1390–1406.

Evans, Nicole C., Christina I. Swanson, and Scott Barolo. 2012. “Sparkling Insights into Enhancer Structure, Function, and Evolution.” *Current Topics in Developmental Biology* 98:97–120.

Funnell, Alister P W, and Merlin Crossley. 2012. “Homo- and Heterodimerization in Transcriptional Regulation.” *Advances in Experimental Medicine and Biology*.

Garrell, Joan, and Sonsoles Campuzano. 1991. “The Helix-loop-helix Domain: A Common Motif for Bristles, Muscles and Sex.” *BioEssays*.

Germain, Pierre, and William Bourguet. 2013. “Dimerization of Nuclear Receptors.” *Methods in Cell Biology* 117:21–41.

Golub, Todd R., George F. Barker, Michael Lovett, and D. Gary Gilliland. 1994. “Fusion of PDGF Receptor β to a Novel Ets-like Gene, Tel, in Chronic Myelomonocytic Leukemia with t(5;12) Chromosomal Translocation.” *Cell* 77 (2):307–16.

Green, Sean M., H. Jerome Coyne, Lawrence P. McIntosh, and Barbara J. Graves. 2010. “DNA Binding by the ETS Protein TEL (ETV6) Is Regulated by Autoinhibition and Self-Association.” *Journal of Biological Chemistry* 285 (24):18496–504.

Grewal, Shiv I S, and Danesh Moazed. 2003. “Heterochromatin and Epigenetic Control of Gene Expression.” *Science (New York, N.Y.)* 301 (5634):798–802.

Grossniklaus, Ueli, and Renato Paro. 2014. “Transcriptional Silencing by Polycomb-Group Proteins.” *Cold Spring Harbor Perspectives in Biology* 6 (11):1–26.

Hippel, P. H. V., A. Revzin, C. A. Gross, and A. C. Wang. 1974. “Non-Specific DNA Binding of Genome Regulating Proteins as a Biological Control Mechanism: 1. The Lac Operon: Equilibrium Aspects.” *Proceedings of the National Academy of Sciences* 71 (12):4808–12.

Johnson, A D, A R Poteete, G Lauer, R T Sauer, G K Ackers, and M Ptashne. 1981. “Lambda Repressor and Cro - Components of an Efficient Molecular Switch.” *Nature* 294 (5838):217–23.

Junion, Guillaume, Mikhail Spivakov, Charles Girardot, Martina Braun, E. Hilary Gustafson, Ewan Birney, and Eileen E M Furlong. 2012. “A Transcription Factor Collective Defines Cardiac Cell Fate and Reflects Lineage History.” *Cell* 148 (3):473–86.

Kim, C. A., M. L. Phillips, W. Kim, M. Gingery, H. H. Tran, M. A. Robinson, S. Faham,

and J. U. Bowie. 2001. “Polymerization of the SAM Domain of TEL in Leukemogenesis and Transcriptional Repression.” *EMBO Journal* 20 (15):4173–82.

Knight, Mary Jane, Catherine Leettola, Mari Gingery, Hao Li, and James U. Bowie. 2011. “A Human Sterile Alpha Motif Domain Polymerizome.” *Protein Science* 20 (10):1697–1706.

Kueng, Stephanie, Mariano Oppikofer, and Susan M. Gasser. 2013. “SIR Proteins and the Assembly of Silent Chromatin in Budding Yeast.” *Annual Review of Genetics* 47 (1):275–306.

Kwon, Ilmin, Masato Kato, Siheng Xiang, Leeju Wu, Pano Theodoropoulos, Hamid Mirzaei, Tina Han, Shanhai Xie, Jeffry L. Corden, and Steven L. McKnight. 2014. “Phosphorylation-Regulated Binding of RNA Polymerase Ii to Fibrous Polymers of Low-Complexity Domains.” *Cell*.

Lai, Zhi Chun, and Gerald M. Rubin. 1992. “Negative Control of Photoreceptor Development in Drosophila by the Product of the Yan Gene, an ETS Domain Protein.” *Cell* 70 (4):609–20.

Landt, Stephen G., Georgi K. Marinov, Anshul Kundaje, Pouya Kheradpour, Florencia Pauli, Serafim Batzoglou, Bradley E. Bernstein, et al. 2012. “ChIP-Seq Guidelines and Practices of the ENCODE and modENCODE Consortia.” *Genome Research*.

Li, Runzhao, Huiping Pei, and Dennis K Watson. 2000. “Regulation of Ets Function by Protein - Protein Interactions.” *Oncogene* 19:6514–23.

Marcovitz, a., and Y. Levy. 2011. “Frustration in Protein-DNA Binding Influences Conformational Switching and Target Search Kinetics.” *Proceedings of the National Academy of Sciences* 108 (44):17957–62.

McGhee, James D., and Peter H. von Hippel. 1974. “Theoretical Aspects of DNA-Protein Interactions: Co-Operative and Non-Co-Operative Binding of Large Ligands to a One-Dimensional Homogeneous Lattice.” *Journal of Molecular Biology* 86 (2):469–89.

O’Neill, Elizabeth M., Ilaria Rebay, Robert Tjian, and Gerald M. Rubin. 1994. “The Activities of Two Ets-Related Transcription Factors Required for Drosophila Eye Development Are Modulated by the Ras/MAPK Pathway.” *Cell* 78 (1):137–47.

Parker, David S, Michael A White, Andrea I Ramos, Barak A Cohen, Scott Barolo, S. Barolo, J. W. Posakony, et al. 2011. “The Cis-Regulatory Logic of Hedgehog Gradient Responses: Key Roles for Gli Binding Affinity, Competition, and Cooperativity.” *Science Signaling* 4 (176):ra38.

Paronetto, Maria Paola. 2013. “Ewing Sarcoma Protein: A Key Player in Human

Cancer." *International Journal of Cell Biology*.

Perissi, Valentina, Kristen Jepsen, Christopher K. Glass, and Michael G. Rosenfeld. 2010. "Deconstructing Repression: Evolving Models of Co-Repressor Action." *Nature Reviews Genetics* 11 (2):109–23.

Poirel, H, C Oury, C Carron, E Duprez, Y Laabi, A Tsapis, S P Romana, et al. 1997. "The TEL Gene Products: Nuclear Phosphoproteins with DNA Binding Properties." *Oncogene* 14 (3):349–57.

Qiao, Feng, Haiyun Song, Chongwoo A. Kim, Michael R. Sawaya, Jacob B. Hunter, Mari Gingery, Ilaria Rebay, Albert J. Courey, and James U. Bowie. 2004. "Derepression by Depolymerization: Structural Insights into the Regulation of Yan by Mae." *Cell* 118 (2):163–73.

Ramos, Andrea I, and Scott Barolo. 2013. "Low-Affinity Transcription Factor Binding Sites Shape Morphogen Responses and Enhancer Evolution." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 368:20130018.

Roy, S., J. Ernst, P. V. Kharchenko, P. Kheradpour, N. Negre, M. L. Eaton, J. M. Landolin, et al. 2010. "Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE." *Science* 330 (6012):1787–97.

Suryamohan, Kushal, and Marc S. Halfon. 2015. "Identifying Transcriptional Cis-Regulatory Modules in Animal Genomes." *Wiley Interdisciplinary Reviews: Developmental Biology* 4 (2):59–84.

Teif, Vladimir B, and Karsten Rippe. 2010. "Statistical–mechanical Lattice Models for protein–DNA Binding in Chromatin." *Journal of Physics: Condensed Matter* 22 (41):414105.

Tognon, Cristina E, Cameron D Mackereth, Aruna M Somasiri, Lawrence P McIntosh, and Poul H B Sorensen. 2004. "Mutations in the SAM Domain of the ETV6-NTRK3 Chimeric Tyrosine Kinase Block Polymerization and Transformation Activity." *Molecular and Cellular Biology* 24 (11):4636–50.

Tsodikov, Oleg V., Jill A. Holbrook, Irina A. Shkel, and M. Thomas Record. 2001. "Analytic Binding Isotherms Describing Competitive Interactions of a Protein Ligand with Specific and Nonspecific Sites on the Same DNA Oligomer." *Biophysical Journal* 81 (4):1960–69.

Weake, V, and J Workman. 2010. "Inducible Gene Expression: Diverse Regulatory Mechanisms." *Nat Rev Genet* 11 (6):426–37.

Webber, Jemma L., Jie Zhang, Lauren Cote, Pavithra Vivekanand, Xiaochun Ni, Jie

Zhou, Nicolas Nègre, Richard W. Carthew, Kevin P. White, and Ilaria Rebay. 2013. "The Relationship between Long-Range Chromatin Occupancy and Polymerization of the Drosophila Ets Family Transcriptional Repressor Yan." *Genetics* 193 (2):633–49.

Zhang, Jie, Thomas G W Graham, Pavithra Vivekanand, Lauren Cote, Maureen Cetera, and Ilaria Rebay. 2010. "Sterile Alpha Motif Domain-Mediated Self-Association Plays an Essential Role in Modulating the Activity of the Drosophila ETS Family Transcriptional Repressor Yan." *Molecular and Cellular Biology* 30 (5):1158–70.

2.9 Supplementary Figures

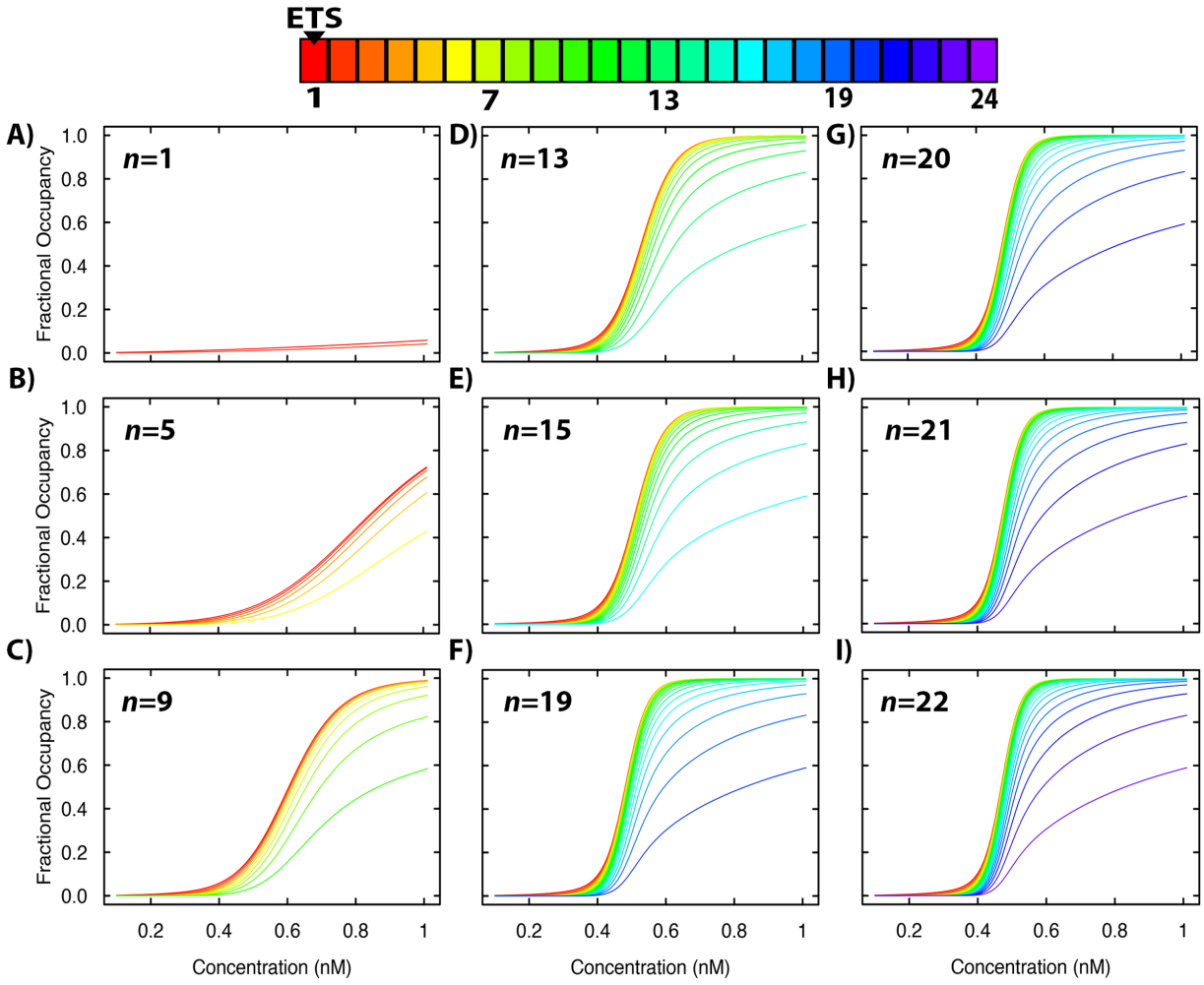


Figure 2.S1: Yan binding profiles for elements of increasing size.

The value of *n* is labeled for each calculation, and Yan fractional occupancy was plotted across the element using the wild type parameters for α, β, and γ. Fractional occupancy is plotted as a function of concentration for all position within the element, as shown in the key.

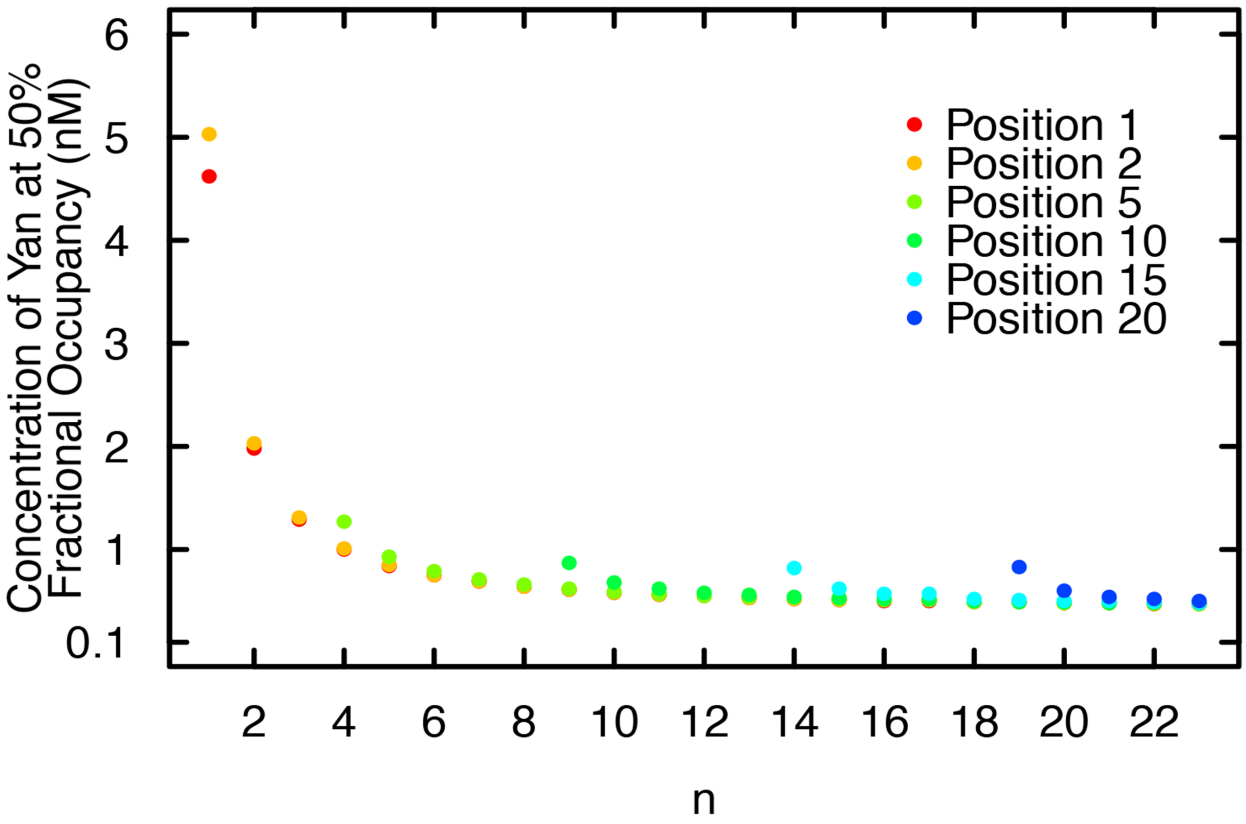


Figure 2.S2: Yan binding profiles converge as *n* increases.

The concentration at which 50% fractional occupancy is achieved is plotted as a function of element size, and displayed by position within the element. If an element is too small to encompass a given position, it is not plotted. Note data points are plotted on top of one another.

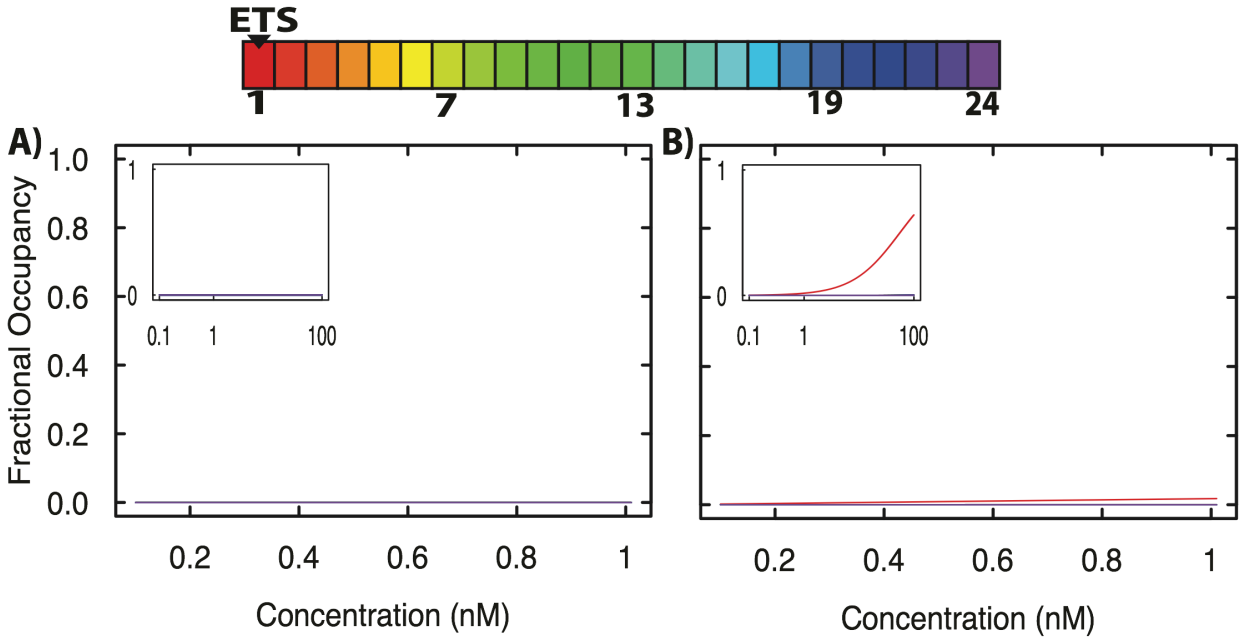


Figure 2.S3: Yan fractional occupancy depends on protein-DNA and protein-protein interactions.

Fractional occupancy is plotted as a function of concentration for all position within the element, as shown in the key. Note that lines are plotted on top of one another. Insets represent log scale concentration. A) Yan binding profile for an element where α and β are set to 0 kcal/mol. B) Yan binding profile for an element where γ is set to 0 kcal/mol.

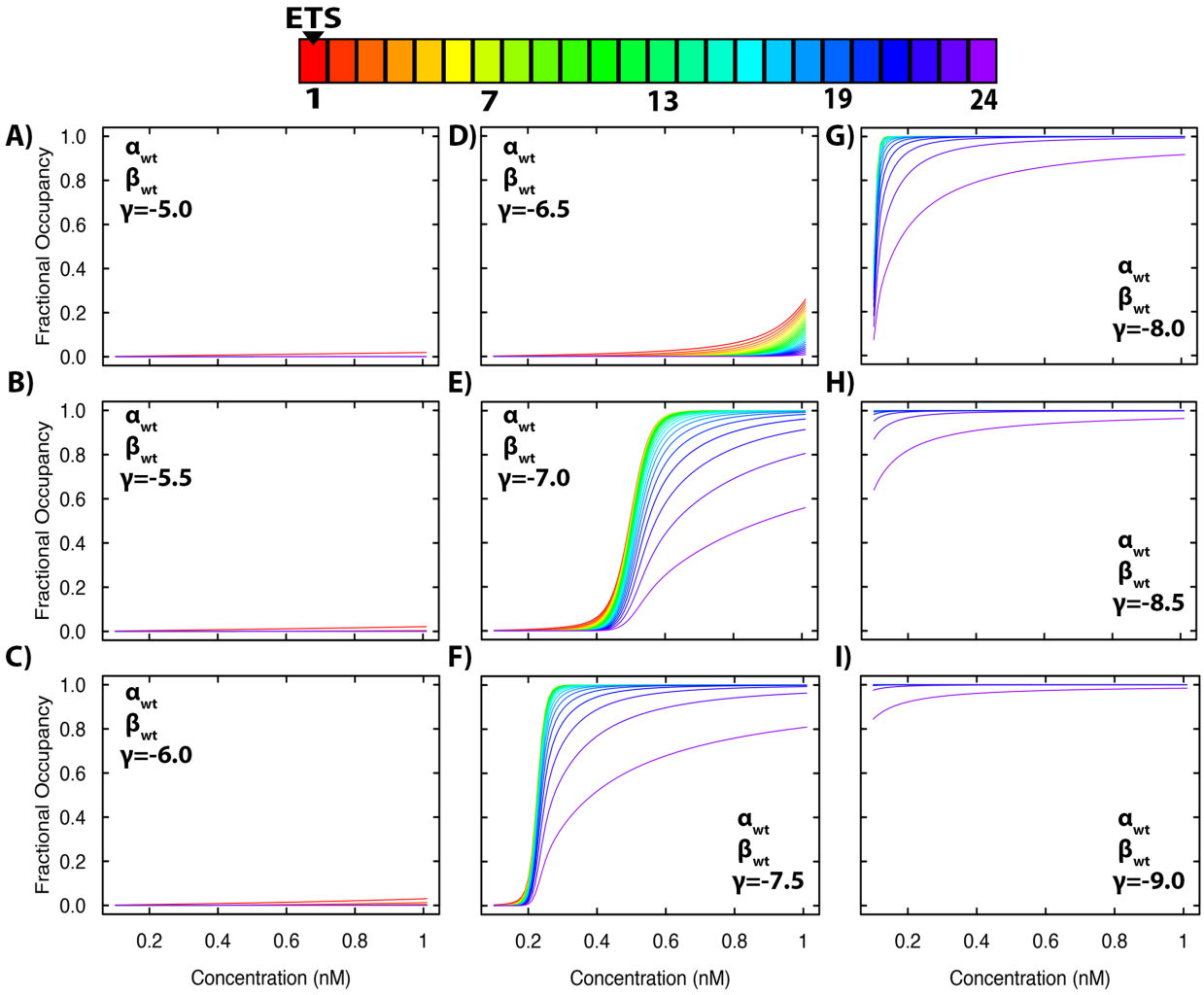


Figure 2.S4: Protein-protein interactions strongly influence Yan occupancy.

Yan occupancy was calculated using the wild type values of α and β, while decreasing γ from -5.0 to -9.0 kcal/mol in 0.5 kcal/mol increments (A-I). Fractional occupancy is plotted as a function of concentration for all position within the element, as shown in the key.

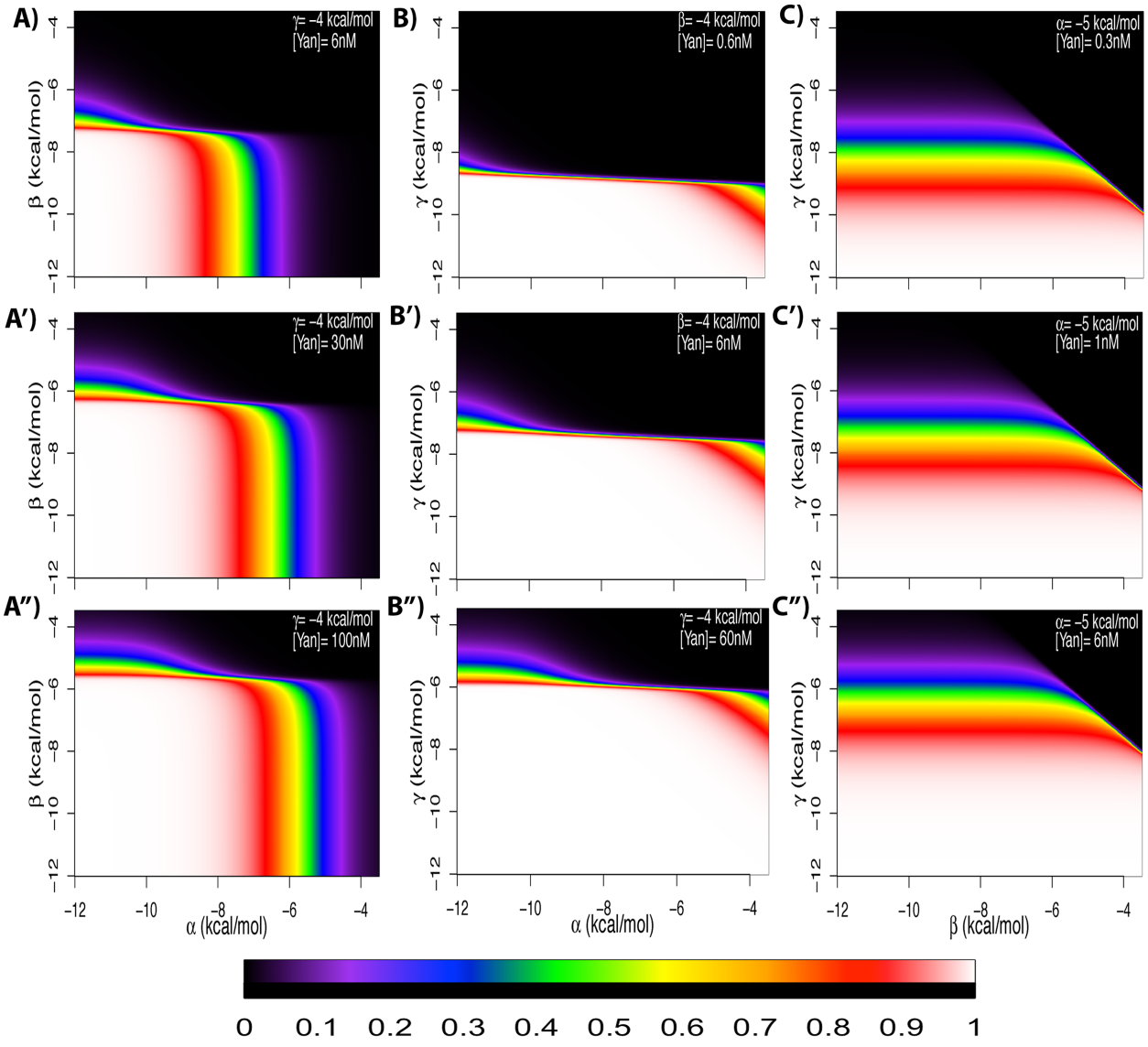


Figure 2.S5: Increasing Yan concentration translocates spectral heat maps of nucleate fractional occupancy.

All nucleated fractional occupancies are plotted on a spectral scale, as shown in the key below. A-A'') α versus β, with [Yan] at 6 nM, 30 nM, and 100 nM. B-B'') α versus γ, with [Yan] at 0.6 nM, 6 nM, and 60 nM. C-C'') β versus γ, with [Yan] at 0.3 nM, 1 nM, and 6 nM.

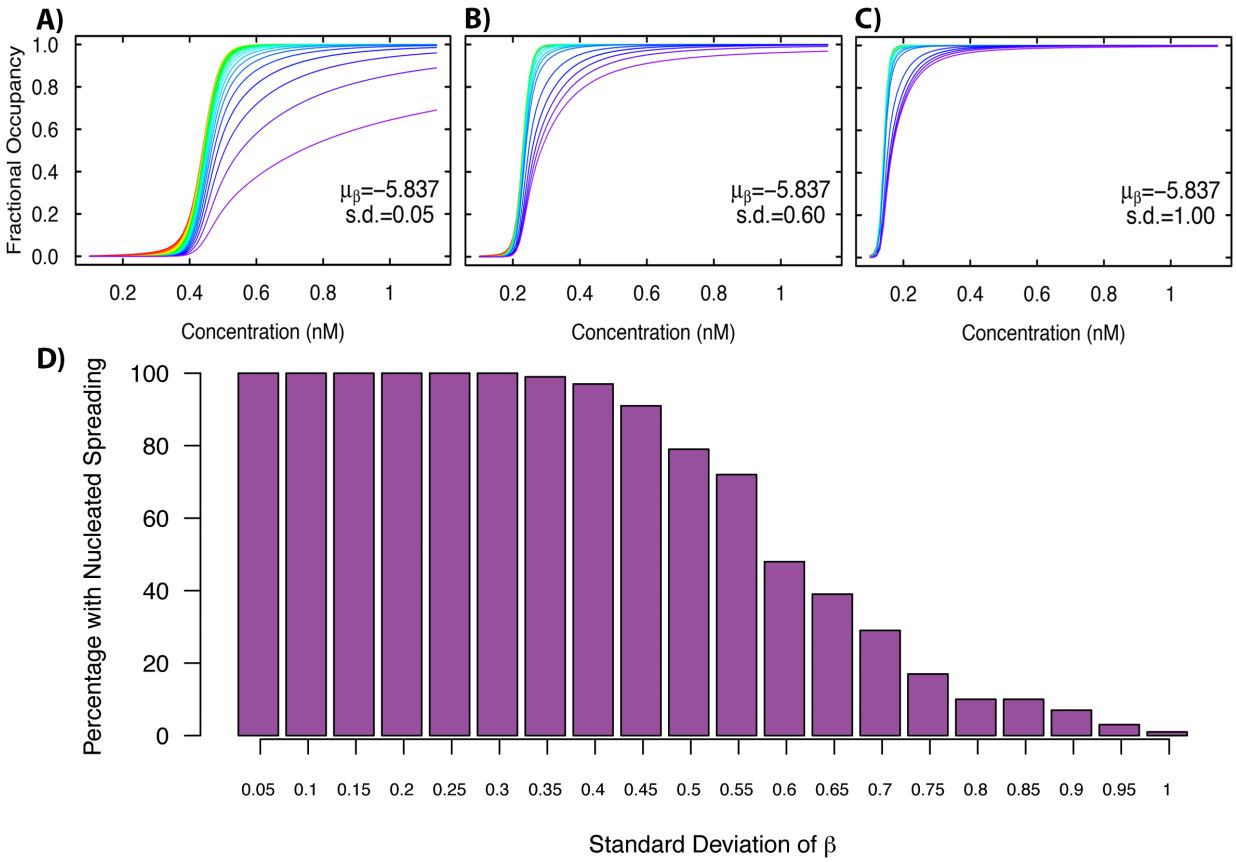


Figure 2.S6: The Yan spreading profile is robust to variation in β.

A-C) Fractional occupancy by position versus Yan concentration, with the ETS site in red, progressing through the visible color spectrum to the most distal site in purple. Values of β for each position were selected from a half Gaussian distribution with mean μ_β and a standard deviation (s.d.). A) Standard deviation of 0.05 B) Standard deviation of 0.6 C) Standard deviation of 1.0. D) Percentage of elements with nucleated spreading profile, as a function of the standard deviation in values of β. Random elements were generated and scored as demonstrating nucleated spreading if the ETS site had higher occupancy than any other site, measured at the concentration of half-maximal occupancy for the ETS. 100 random elements were generated and scored for each standard deviation value.

Chapter 3: Engineering of ETS-family factor Yan reveals a requirement for balanced protein-protein interaction affinity among transcription factors during development

C. Matthew Hope and Ilaria Rebay

3.1 Abstract

During development, precisely tuned DNA binding affinity enables transcription factors to build dynamic complexes that regulate gene expression. Cooperative protein-protein interactions among transcription factors at enhancers are also critical to regulatory output and spatiotemporal specificity. Whether the affinities of these interactions are similarly tuned to a range that balances robustness with regulatory specificity is not known. The Drosophila ETS-family transcription factor Yan, whose repressive ability requires self-association via SAM-SAM interactions, provides a perfect model. Using a structure-based engineering strategy, we increased Yan affinity over four orders of magnitude. This produced a dramatic subcellular redistribution of Yan into punctate structures, reduced repressive output and compromised survival. Examination of photoreceptor specification defects revealed cell-type specific differences in response to altered SAM-driven self-association, suggesting Yan utilizes different polymeric species for specific regulatory decisions. We conclude that protein-protein interactions broadly shape the formation of regulatory complexes, with medium-affinity interactions essential to proper transcriptional output.

3.2 Introduction

Development is a process of progressive cell fate acquisition that requires extraordinary precision in gene expression. Sequence-specific transcription factors (TFs) provide the molecular underpinning of this process by binding to *cis*-regulatory elements, known as enhancers, in order to activate or repress gene expression in spatiotemporal patterns (Davidson and Levine 2008; Spitz and Furlong 2012) and in disease (Smith and Shilatifard 2014). The recruitment and organization of transcriptional regulatory

complexes at enhancers is dictated primarily by two types of interactions: sequence-specific protein-DNA interactions and protein-protein interactions. Protein-DNA interactions allow TFs to translate sequence information in non-coding DNA, namely the number, order, and affinity of TF binding sites within an enhancer, in order to build regulatory complexes at the correct genomic locations. Protein-protein interactions among TFs serve multiple purposes. For example, cooperative interactions can increase fractional occupancy of TFs at enhancers and can promote association with essential cofactors and general regulators of RNA polymerase (Funnell and Crossley 2012). At the molecular level, how the affinities of a particular TF's protein-DNA and protein-protein interactions need to be tuned to ensure recruitment of regulatory complexes that have the correct composition, dynamics, and function needed to produce the rapid, shifting gene expression patterns necessary for development is poorly understood.

While there are no simple rules for predicting the output of a given enhancer, decades of perturbing specific protein-DNA interactions within model enhancers have uncovered several key organizing principles (Evans et al. 2012; Reinitz and Sharp 1995). Of note to our work are recent studies emphasizing the importance of clustered low- and medium-affinity TF binding sites in regulatory elements that control developmental gene expression. Thus in a variety of examples, the introduction of high-affinity TF binding sites produced deleterious effects on enhancer output or spatiotemporal specificity (Ramos and Barolo 2013; Parker et al. 2011; Lorberbaum et al. 2016; Crocker et al. 2015; Cary et al. 2017; Farley et al. 2015) perhaps because the regulatory complexes recruited by TFs anchored by such high-affinity interactions lack the necessary dynamics.

Thus, part of the solution in resolving the tradeoff in protein-DNA affinity versus specificity must be to tune the cooperative TF-TF interactions in a way that achieves sufficient occupancy at target sites. Although equally important as protein-DNA interactions in determining the output from an enhancer, the functions of TF protein-protein interactions at enhancers are more opaque. In *Drosophila,* extensive co-occupancy of TFs at enhancers has been noted(Junion et al. 2012), as well conservation of binding site motifs that preserve cooperative TF binding (Kazemian et al. 2013), yet few studies have modulated TF protein-protein interactions directly to assess the impact on transcriptional regulatory complexes. A recent thermodynamic modeling effort by our group revealed that the extent of homotypic TF-TF interactions may also play a role in the tradeoff between affinity and binding site specificity, with longer polymers showing reduced binding site discrimination (Hope et al. 2017). This raises the question of how modulating TF protein-protein interaction affinity impacts downstream transcriptional regulation. *A priori*, it is possible that higher affinity interactions would stabilize a given complex, thus resulting in stronger recruitment of transcriptional co-regulators and increased transcriptional output, either activating or repressive. Alternatively, and analogous to the importance of medium-affinity protein-DNA interactions, higher affinity protein-protein interactions might compromise regulatory dynamics or target gene specificity, thereby decreasing or dysregulating enhancer output. In either case, the impact of increasing protein-protein interaction affinity for a TF remains unexplored.

The *Drosophila* E-Twenty-Six (ETS) family TF Yan provides a well-characterized system ideally suited to addressing these questions. Yan is a conserved transcriptional repressor that works downstream of Receptor Tyrosine Kinase (RTK)

signaling to regulate cell fate decisions during embryonic and retinal development (Lai and Rubin 1992). Upon RTK activation, Mitogen Activated Protein Kinase (MAPK) phosphorylates Yan (O'Neill et al. 1994). This triggers the dismantling of Yan repressive complexes, allowing its export from the nucleus and degradation in the cytoplasm, which in turn enables other activating TFs and cofactors to turn on previously repressed target genes and initiate the cell fate transition. If MAPK-mediated phosphorylation of Yan is blocked, Yan remains nuclearly localized, constitutively represses transcription and blocks the fate switch (Rebay and Rubin 1995). In contrast, insufficient Yan repressive activity permits inappropriate induction of gene expression programs which leads to ectopic fate specification.

Two key motifs determine the protein-DNA and protein-protein interactions, respectively, that organize Yan's regulatory functions and dynamics: the ETS DNA binding domain and the sterile alpha motif (SAM). The ETS domain binds as a monomer to DNA motifs with a core GGAA/T sequence and also provides a critical nuclear localization sequence. Homotypic SAM-SAM interactions are essential for full *in vivo* function and required to establish active repressive complexes(Webber et al. 2013; Zhang et al. 2010). Thus mutations that block SAM-SAM association and limit Yan to a monomeric state abrogate repressive ability and show strong loss-of-function phenotypes. An intriguing feature of the Yan SAM-SAM interaction is that *in vitro* the isolated domain will form long helical polymers. Although mutations that restrict Yan to dimers fully rescue the lethality and phenotypes associated with *yan* loss (IR, unpublished observation), higher-order Yan complexes have been detected in cells and can influence repressive output at certain target enhancers target genes(Zhang et al. 2010). Thus while

the majority of Yan biology appears to be tuned to utilize lower-order species, how the affinity of the SAM-SAM interaction shapes this distribution to ensure proper dynamics and specificity to regulation of gene expression is not understood.

To address this question, we expanded our previous modeling effort and uncovered a requirement for medium-affinity Yan interactions, based on a minimal model of Yan binding at equilibrium. To test the predictions of the model, we designed a molecular engineering strategy to isolate Yan variants with increased SAM-SAM affinity. By exploiting the sequence- and structure-level evolutionary conservation between Yan and its human orthologue TEL/ETV6, we identified four residues whose manipulation enabled us to tune the affinity of the Yan SAM interaction over four orders of magnitude. Using a series of assays in *Drosophila* cultured cells and in the developing retina, we show that increased SAM affinity impairs Yan's repressive activity and *in vivo* function. At the cellular level, increased SAM affinity promotes and stabilizes Yan self-association, leading to the redistribution of Yan into punctate structures in both the cytoplasm and nucleus. We speculate that these structures lack the necessary dynamics for nucleo-cytoplasmic Yan trafficking and assembly of effective regulatory complexes in the nucleus. Finally, our cellular analysis of highly polymerizing Yan mutants in the eye uncovered an atypical Yan loss of function phenotype, which we speculate is a demonstration of different requirements for Yan polymerization at different transcriptional regulatory complexes throughout development. Taken together, our results suggest a spectacular dynamic range of affinities lies latent within the SAM, and by extension, strong evolutionary pressure to maintain medium-affinity interactions that preserve transcriptional repression and biological function.

3.3 Results

3.3.1 Modeling of Yan DNA occupancy predicts a requirement for medium-affinity polymerization

In order to explore the functional consequences of altering Yan polymerization, we revisited a previously published *in silico* model of Yan occupancy on DNA based on binding at chemical equilibrium (Hope et al. 2017). Briefly, this model relies on calculating Yan fractional occupancy at discrete sites on DNA using four parameters: the strength of sequence-specific DNA binding, non-specific DNA binding, polymerization, and the concentration of Yan. Our results showed that binding at equilibrium was able to produce Yan binding profiles that spread from sequence-specific ETS sites to non-specific sites on DNA, and importantly, that this occurred at parameter values that match previously measured affinities for ETS-family TFs (Qiao et al. 2004; De et al. 2014). However, an import factor that was not taken into account in the original modeling effort was polymerization off of DNA. Given that isolated SAM domains are sufficient to polymerize *in vitro* (Kim et al. 2001; Qiao et al. 2004), it is reasonable to assume that SAM-SAM self-association can occur both on and off DNA, and that this might affect the balance of Yan molecules bound at target sites. Therefore, we updated the model to take this into account and calculated Yan fractional occupancy as a function of polymerization strength.

Intriguingly, the model demonstrated a new behavior with respect to polymerization affinity. Whereas increasing polymerization in the previous model inevitably resulted in increased fractional occupancy on DNA, allowing polymerization off of DNA reveals a clear picture of an affinity-dependent penalty for Yan occupancy.

Figure 1A plots the fractional occupancy at a single ETS site surrounded by many non-specific binding sites, as a function of Yan concentration for the wild type parameter values of Yan. Occupancy was calculated over four orders of magnitude of concentration from 0.1nM to 1000nM, which is a back-of-the-envelope interpretation of the physiological relevant regime of any TF concentration based on measured values in *Drosophila* (Biggin 2011), although the true range of Yan concentration may be more narrow. As expected, fractional occupancy increases with Yan concentration until the site is completely saturated. However, there is a stark penalty for increased polymerization (Figure 1B). Although at lower concentrations fractional occupancy is increased, our model predicts that at higher concentrations of Yan, increasing polymerization affinity will result in decreased fractional occupancy on DNA.

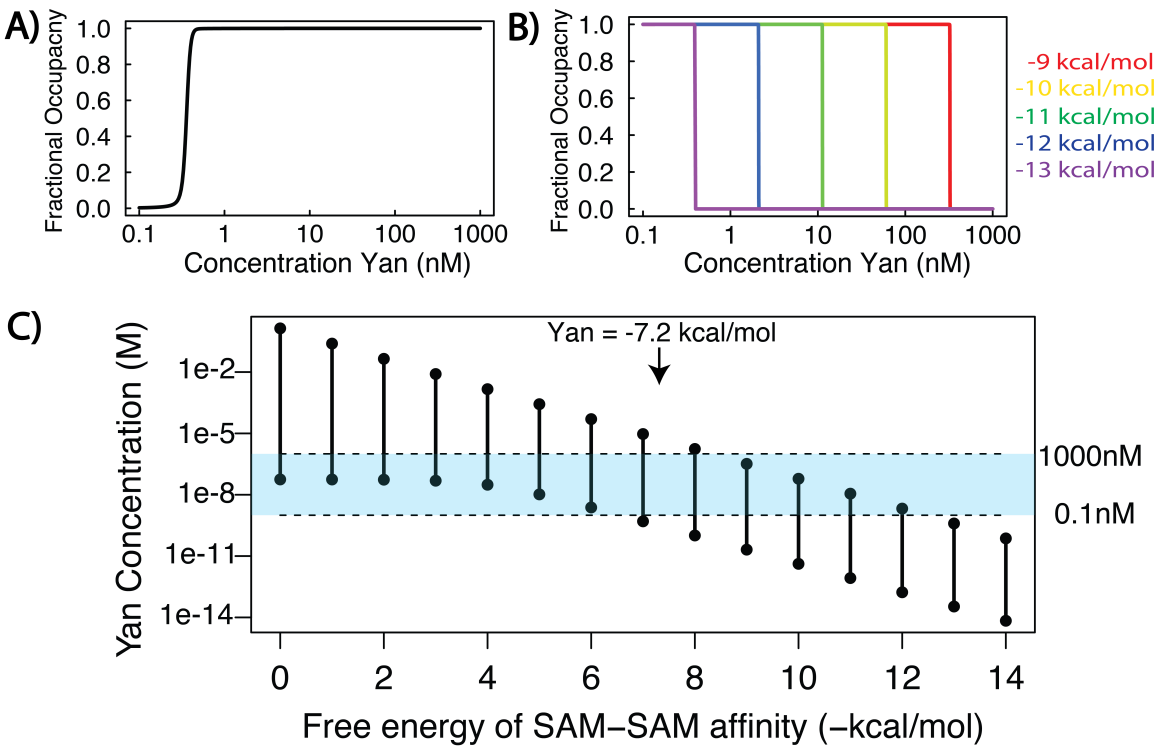


Figure 3.1: Equilibrium modeling shows that increasing polymerization affinity results in decreased occupancy of Yan on DNA.

A) Model of Yan fractional occupancy at a specific binding site as a function of Yan concentration, for wild type parameters of Yan protein-DNA and protein-protein affinity. Polymerization off of DNA is taken into account. B) Fractional occupancy of Yan when SAM-SAM affinity is strengthened, showing progressive loss of fractional occupancy as polymerization affinity increases (red through purple curves). C) Concentration ranges of 50% or greater Yan fractional occupancy, as a function of SAM-SAM affinity. As affinity increases, Yan occupancy moves outside the physiologically relevant range (blue shading).

To fully explore the behavior of Yan with respect to polymerization, we calculated the concentrations where ETS site occupancy will be 50% or greater, as a function of polymerization affinity over a large range (Figure 1C). The results show that for any given strength of polymerization, as expected, there are concentrations where Yan is too limited to appreciably occupy specific DNA sites. However, there is also a regime of high Yan concentration where polymerization off of DNA impinges on the amount of Yan bound to DNA. Importantly, the concentration regime of Yan occupancy shifts as

polymerization increases, relative to the physiological range of Yan concentration (light blue shading, Figure 1C). Thus, at weaker values of polymerization affinity (0 to -5 kcal/mol), there is little to no occupancy at specific binding sites, but there is a middle regime from -6 to -10 kcal/mol of Yan occupancy in the relevant regime of Yan concentration. Furthermore, our model predicts that at extreme values of Yan polymerization(-10 to -14 kcal/mol), occupancy will be once again lost due to overly strong SAM-SAM affinity.

Taken together, these results support a role for medium-affinity protein-protein interactions in driving Yan occupancy. Interestingly, the results predict a relatively wide range of polymerization affinities that can still result in Yan occupancy, followed by a precipitous decline once extreme values of SAM-SAM affinity are reached. Thus to confirm the predictions of the model, we set out to develop a series of Yan mutants that would explore the entirety of the high-affinity regime of Yan polymerization, and in the process, test the relationship between SAM-SAM affinity and transcriptional function.

3.3.2 A screen for Yan mutants that increase SAM-SAM affinity

In order to isolate mutants that increase SAM-SAM affinity, we adopted a strategy of increasing Yan affinity by mutating it toward the sequence of the mammalian TEL SAM domain, whose self-association affinity is three orders of magnitude higher than Yan SAM (Kim et al. 2001; Qiao et al. 2004). We reasoned that the divergent residues between Yan and TEL at the interface between two molecules could be the key to this difference in affinity. Of the 22 residues at the interface between two SAM molecules in the TEL crystal structure, only nine differed between Yan and TEL (Figure 3.2A). We noticed that four of the nine residues clustered to a small region at the periphery of the

SAM-SAM interface, and that three of these residues make salt-bridge contacts that span the interface. Mutation of these residues has been previously shown to modulate TEL's polymerization affinity *in vitro* and to block the transformation ability of TEL oncogenic fusions in cultured cells (Cetinbas et al. 2013). Hypothesizing that facilitating TEL-like interactions including salt-bridges in this clustered region would increase Yan SAM affinity, we made all possible combinations of the four mutations Yan R92K, A93E, G96R, and H97Y.

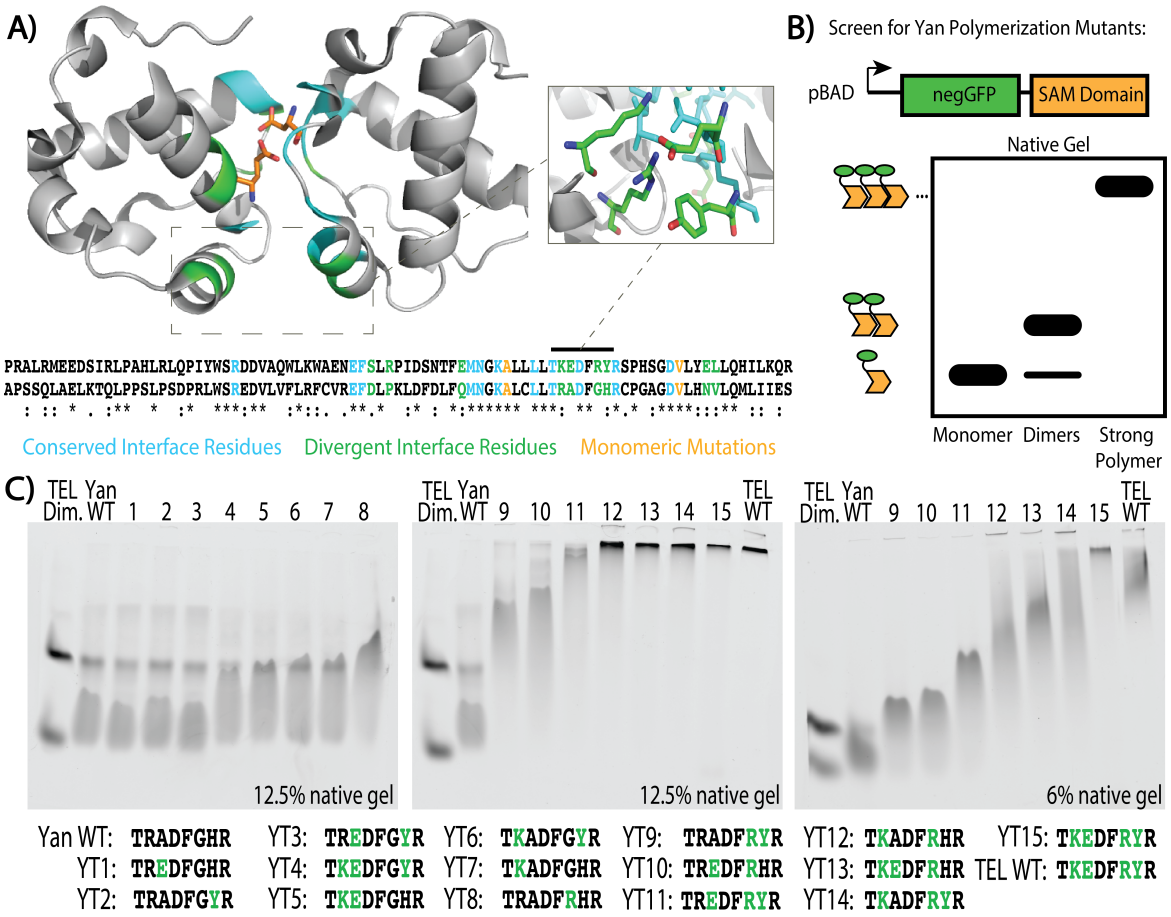


Figure 3.2: A small group of mutations are capable of shifting Yan polymerization to TEL-like affinity *in vitro*.

A) Cartoon representation of the TEL SAM crystal structure (PDB: 1JI7, Kim. et. al.) with the primary sequence alignment of Yan and TEL SAM domains. Residues at the interface are emphasized, with conserved residues colored blue, divergent residues colored green, and previously published monomerizing positions colored orange. Inset,

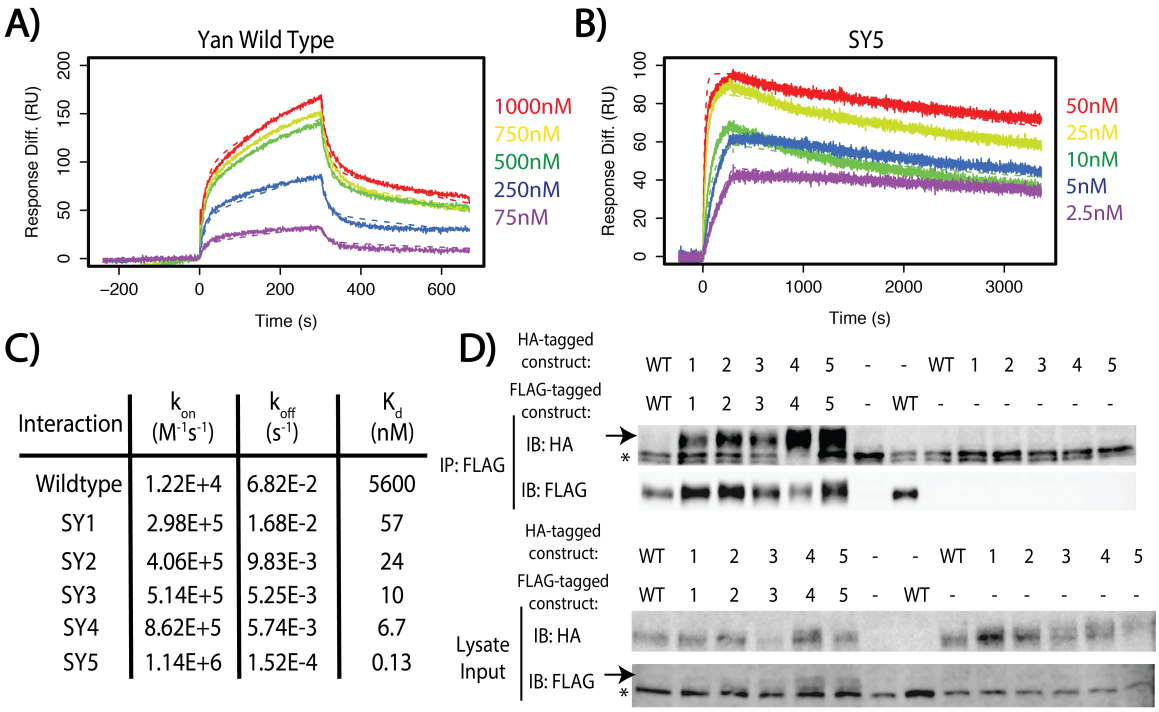
Figure 3.2 (cont.): TEL residues corresponding to YanR92, A93, G96, and H97. B) Schematic of negGFP Native Gel Assay. SAM domains are expressed and purified with a supercharged negGFP tag that allows separation by size on a native gel, and visualization by GFP fluorescence. C) Native gels of all possible combinations of the four mutations R92K, A93E, G96R, and H97Y. Each gel contains a mixture of TEL species to mark monomers and dimers, in addition to Yan wild type. Mutants are arranged in increasing order with identity listed below.

To effectively screen these fifteen Yan-to-TEL (YT) SAM mutants, we adapted a native gel shift assay used previously to examine polymerization of human SAM domains (Knight et al. 2011) (Figure 3.2B). The negGFP gel shift assay centers upon expressing and purifying various SAMs fused to an engineered, super negative GFP tag (Lawrence et al. 2007). The negGFP tag imparts a standard amount of charge per SAM molecule, allowing separation of samples by size of SAM polymer on a native PAGE gel, and visualization of polymeric species via GFP fluorescence. Two controls were included as reference points to mark the lower and upper limits of polymerization: an unequal mixture of two monomeric forms of TEL SAM (TEL A93E and TEL V112E) (Kim et al. 2001) marked the mobility of monomers and dimers, while the wild type TEL SAM demonstrated the limited migration of higher order polymers.

The mobility shifts seen in the negGFP native gel assay identify Yan SAM variants with increased polymerization that matches and exceeds that of TEL SAM (Figure 3.2C). Although two individual mutants, R92K and G96R, modulated mobility on their own (YT7 and YT8 respectively), multiple mutations were required for greater shifts, with simultaneous mutation of all four residues (YT15) surpassing the polymerization extent of TEL SAM. Three of the mutant SAMs, designated YT1-3, migrated with mobilities indistinguishable from that of the wild type Yan SAM domain, which runs primarily as monomers and dimers with some high-order polymer species,

consistent with its low affinity (Qiao et al. 2004). The next five, YT4-8, shifted mobility toward the dimer form, while YT9-11 exhibited a gradation of reduced mobility. YT12-15 initially showed mobility indistinguishable from that of TEL SAM, but analysis on lower percentage gels revealed that the polymerization ability of YT15 exceeded that of TEL SAM. Thus based on their relative mobility shifts, the re-engineered Yan SAMs appear to span more than three orders of magnitude of affinity.

To quantitatively measure affinities, we turned to pair-wise Surface Plasmon Resonance (SPR) spectroscopy. We selected five mutants that appeared to span the entire affinity range: YT8, YT10, YT12, YT13, and YT15, which for simplicity we refer to as SuperYan (SY) 1-5, respectively. Following the published strategy (Qiao et al. 2004) we blocked higher order polymerization by introducing the monomerizing mutations, A86D and V105R, into each SY construct and then measured the affinity between each SY monomer pair; native gels confirmed each protein ran as a monomer (Figure 3.S1). The K_ds measured by SPR followed the rank order established in the native gel shift assay, revealing a four orders of magnitude range in binding affinity between wild type Yan and SY5 (Figure 3.3A-C and Figure 3.S2). Confirming the quality of our measurements, our determination of wild type Yan affinity as 5.6uM is in excellent agreement with the previously measured value of 7uM (Qiao et al. 2004). Additionally, our measurements of 6.7nM for SY4 and 0.13nM for SY5 accurately predict their migration relative to TEL (2nM) on the native gels (Kim et al. 2001)(Figure 3.2C).



Interaction	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})	K_d (nM)
Wildtype	1.22E+4	6.82E-2	5600
SY1	2.98E+5	1.68E-2	57
SY2	4.06E+5	9.83E-3	24
SY3	5.14E+5	5.25E-3	10
SY4	8.62E+5	5.74E-3	6.7
SY5	1.14E+6	1.52E-4	0.13

Figure 3.3: SuperYan mutations quantitatively increase protein affinity *in vitro* and *in vivo*.

A) Representative sensogram of the wild type Yan SAM interaction, via SPR. Concentrations of analyte are color-coded and listed to the right of the sensogram, and the kinetic model is shown as corresponding dashed lines. B) Sensogram of the SY5 interaction, demonstrating stronger binding affinity. C) Table of parameters obtained from kinetic fitting of sensogram for Yan and SY1-5 interactions. D) Co-immunoprecipitation of Yan and SY1-5 constructs. IPs are shown in the top panel, while lysate inputs are shown on the bottom panel. FLAG-tagged Yan constructs were immunoprecipitated and both HA-tagged and FLAG-tagged proteins were immunoblotted. Arrows mark the size of the Yan constructs, while asterisks mark unreduced IgG in the IP and a non-specific FLAG-reactive species in the lysate, respectively.

Finally we confirmed that the increased SAM-SAM affinity of the SY mutations could influence self-association of full-length Yan proteins. Using co-immunoprecipation of FLAG and HA-tagged constructs expressed in *Drosophila* S2 cells, under stringent conditions in which the wild type Yan interaction is not detectable, SY1-5 showed increasingly robust co-immunoprecipation (Figure 3.3D), roughly proportional with their measured affinities (Figure 3.3C). Taken together, SY1-5 represent a tunable suite of Yan

mutants that span a range of SAM-SAM affinities of approximately four orders of magnitude, ideal for assessing how the affinity of a homotypic protein-protein interaction impacts TF function.

3.3.3 Increasing SAM-SAM affinity induces separable defects in Yan repression and sub-cellular localization

With the suite of SuperYan mutants in hand, we next compared the repressive activity of Yan versus SY1-5 in transcriptional reporter assays in transfected cultured cells, using a reporter generated from regions bound by Yan *in vivo* (Webber et al. 2013). While the repressive outputs of SY1-4 were all statistically indistinguishable from that of wild type Yan, SY5, the one with the highest SAM-SAM affinity, showed a significant roughly two-fold reduction in repressive activity (Figure 3.4A). When RTK signaling was stimulated by co-transfecting cells with constitutively active Ras (RasV12), all constructs demonstrated similar relief of Yan-mediated repression (Figure 3.4B), suggesting signaling responsiveness is not compromised in SuperYan mutants. Because the slightly reduced repressive output from SY4 in the absence of RasV12 was on the borderline of significance ($p=.08$), we retested it using a different reporter comprised of just multimerized ETS sites, and found its repressive ability was indeed reduced relative to that of wild type Yan (Figure 3.S3A). This suggests that the defect in repression in the strong SuperYan mutants is a widespread phenomena that is not specific to a given reporter construct. The subcellular localization of all mutants was indistinguishable from that of wild type Yan (Figure 3.S3B), indicating that the reduced repressive output of SY4 and SY5 reflects a difference in activity. Together these results suggest that effective

Yan-mediated repression can tolerate a wide range of SAM affinities, but that when the affinity of the TEL SAM is approached or exceeded, function becomes impaired.

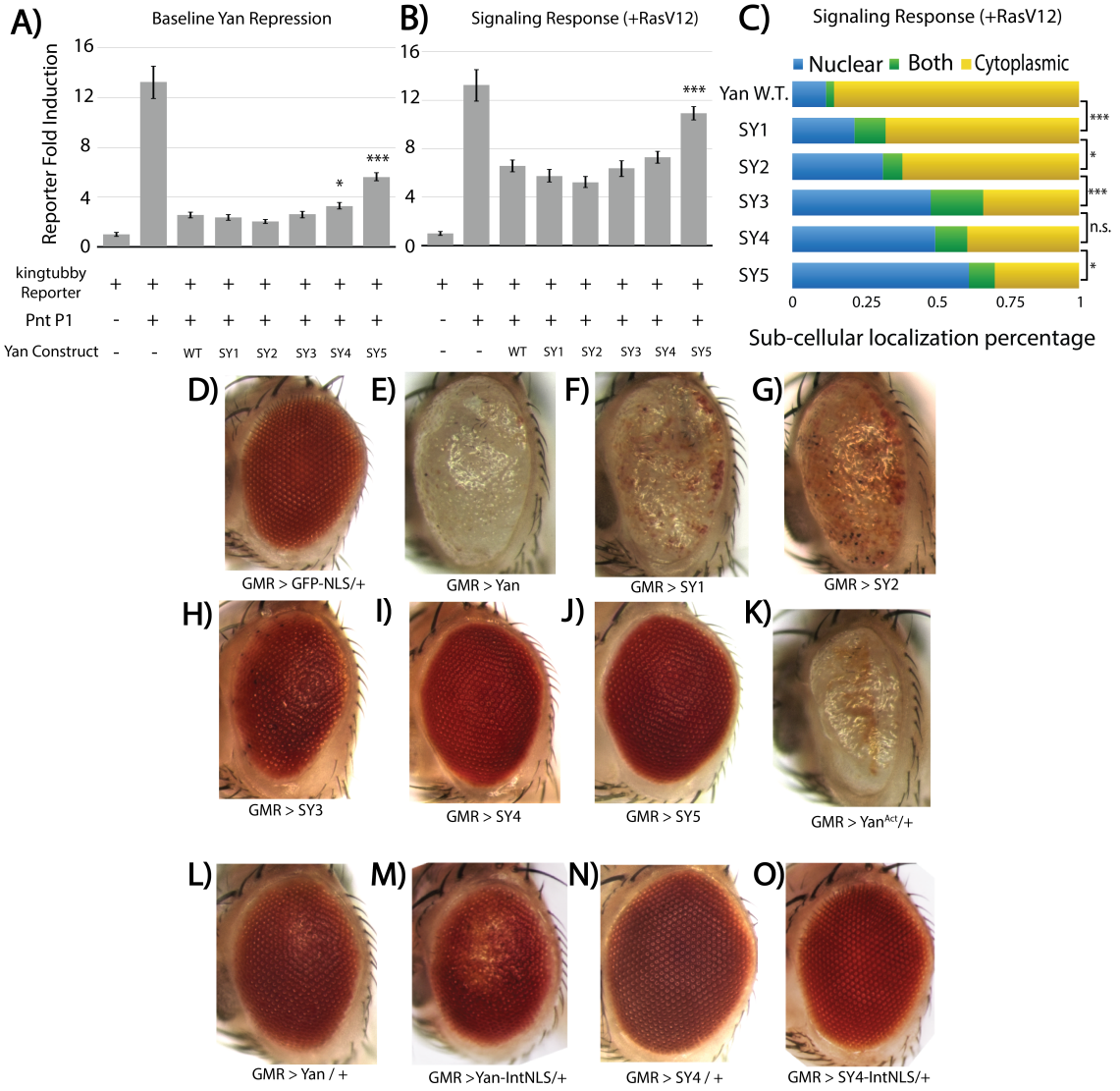


Figure 3.4: Increasing SAM-SAM affinity results in decreased transcriptional repressive activity.

A-B) Luciferase transcription assays of Yan and SY1-5 using a Yan/Pnt responsive transcriptional reporter. The reporter responds to both Pnt and Yan in the expected manner, and Yan's transcriptional repression can be relieved by activating RTK signaling with RasV12. SY4 and SY5 show significant decreases in baseline transcriptional repression. ($p < .1$, and $p < .01$ respectively) Significance was calculated via pair-wise Student T-tests with Bonferroni correction. A) Assay without RasV12 B) Assay with RasV12 C) SuperYan mutants show increasing sub-cellular localization to the nucleus in the presence of RasV12. Genotypes were scored blind and binned into nuclear localization, cytoplasmic localization, or both. Significance was calculated by pairwise

Figure 3.4 (cont.): chi-square tests with Bonferroni correction (* = p < .05 ; *** = p < .001 ; n.s. = not significant). D-O) Over-expression of Yan mutants using the UAS-GAL4 system, driving with GMR-GAL4 at 25C D) GFP-NLS control E-J) Progressive recovery of eye structure and pigment with Yan and SY1-5 K) YanAct, which cannot be exported from the nucleus by RasV12 gives a robust phenotype L-M) YanIntNLS has a strong phenotype in one copy of the transgene, compared to Yan wild type N-O) SY4IntNLS is indistinguishable from SY4, with one transgene copy.

Yan localization and repressive function is negatively regulated by RTK/MAPK signaling (Rebay and Rubin 1995). We therefore tested whether increasing SAM-SAM affinity altered the sub-cellular localization of SuperYan mutants in response to RasV12. Whereas the distribution of wild type Yan localization in S2 cells shifts from predominantly nuclear to predominantly cytoplasmic upon co-transfection of RasV12, as SAM-SAM affinity increased, this response diminished (Figure 3.4C). However in the transcription assay, even at the highest affinities, the attenuation of repression in response to RasV12 was not significantly reduced (Figure 3.4A and Figure 3.S3A). Thus paradoxically, increasing SAM-SAM affinity causes a transcriptional repression defect in the absence of RasV12, but a localization defect in the presence of RasV12. This suggests that affinity driven increases in polymerization produces complexes that are both refractory to nuclear export and incompatible with forming proper transcriptional regulatory complexes at target enhancers.

To test these ideas further, we examined the consequences of overexpressing SY1-5 in the developing *Drosophila* eye, a context where Yan is well-characterized. Consistent with previous findings that modest overexpression of wild type Yan does not strongly perturb development (Rebay and Rubin 1995), two copies of the UAS-SY transgenes were needed to produce strong phenotypes (Figures 3.4E-J and 3.S4A-F). A clear inverse correlation between the severity of the rough eye phenotype and SAM-SAM affinity

emerged. For example, ommatidial patterning and pigmentation are virtually ablated in GMR>SY1 and severely disrupted in SY2 eyes (Figure 3.4E,F), while only very mild patterning defects are detected in GMR>SY4 or SY5 animals (Figure 3.4I,J). However even the strongest SY phenotypes did not match the severity of expressing one copy of a signaling-refractory version of Yan, Yan^{Act}(Figure 3.4K). Qualitative examination of Yan levels and localization in the third instar imaginal disc showed comparable increases in nuclear expression of each SY protein relative to endogenous Yan, and a noticeable increase in overall levels between one and two copies of a given transgene (Figure 3.S4G-T). This suggests that SY mutants are appropriately localized to the nucleus, but that as SAM affinity increases, Yan transcriptional activity becomes compromised.

As an independent test of this interpretation, we added an internal Nuclear Localization Sequence (IntNLS) to both wild type Yan and one of the high affinity mutants, SY4. Previous studies have shown that addition of a strong NLS reduces signal-induced nuclear export and degradation of Yan, thereby stabilizing transcriptional repression and increasing the severity of overexpression-induced defects (Tootle et al. 2003). Consistent with this, NLS addition enhanced the Yan over-expression phenotype (Figure 3.4L-M). In contrast, the eyes of flies expressing SY4-IntNLS were indistinguishable from those expressing unaltered SY4 (Figure 3.4N,O), suggesting that the reduced activity of SY mutants results from defects in repressive ability rather than from altered subcellular localization.

3.3.4 Increased SAM-SAM affinity causes a distinct punctal localization and loss of function

The results described above suggest that increasing the affinity of Yan self-association reduces transcriptional repression capability. However because the assays relied on Yan overexpression, it was possible that the unusually high protein levels promoted complexes that would not be formed under physiological Yan concentrations. We therefore used CRISPR/Cas9-mediated mutagenesis to generate the three strongest mutants (SY3-5) in the *yan* locus and then compared their function and phenotypes relative to wild type and *yan* null backgrounds, in order to make a straightforward assessment of these mutants without endogenous Yan.

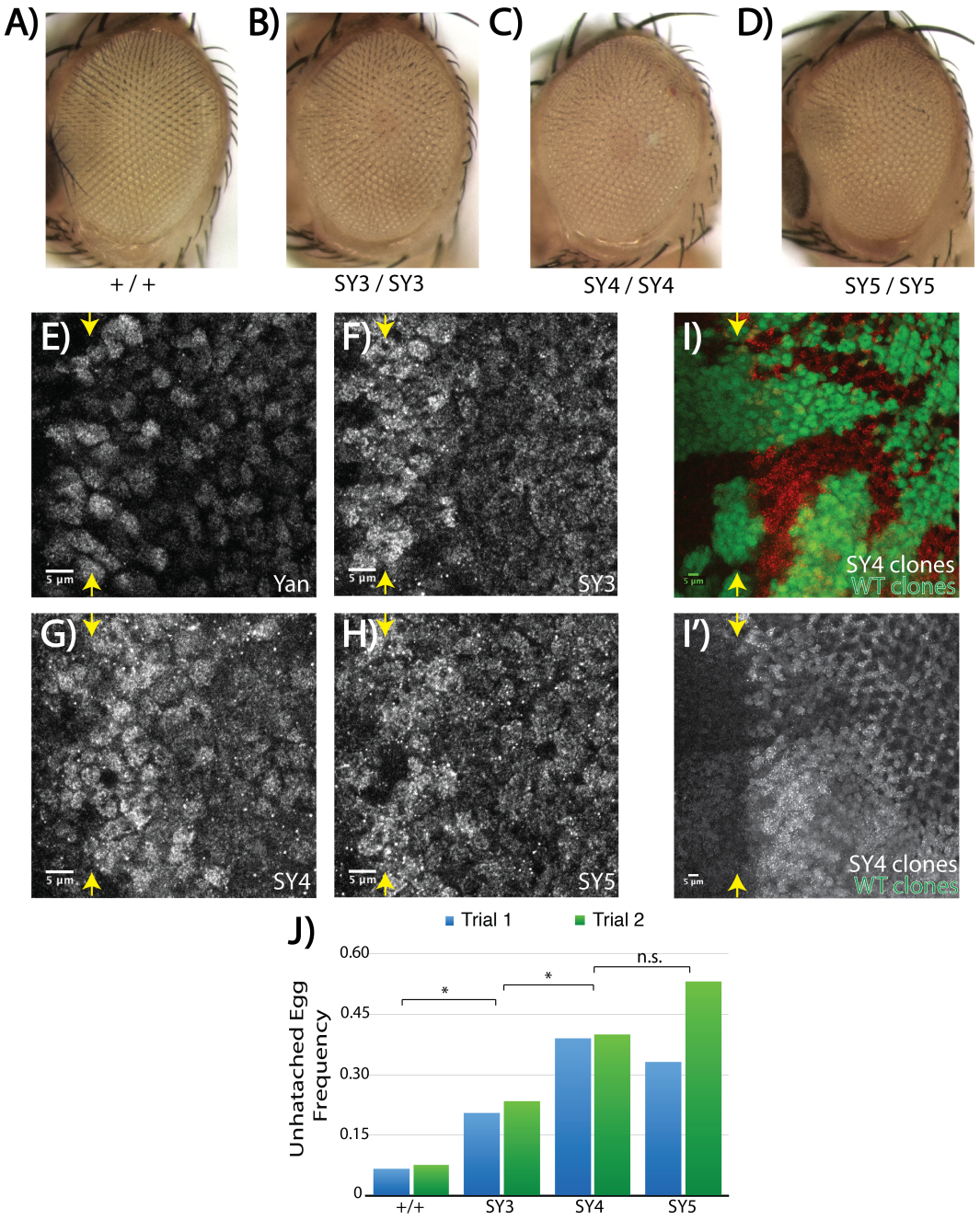


Figure 3.5: Strong SuperYan mutants show atypical photoreceptor recruitment and distinct punctate cellular staining.

A-D) Adult eye images of homozygous mutants of SY3-5, generated via CRISPR/Cas9-mediated mutagenesis. E-H) Immunofluoresence of Yan in wild type and SY3-5 in third instar eye discs. SuperYan discs show increased cytoplasmic localization and increased punctate staining. I-I') Mitotic clones of SY4, generated with the Flp-FRT staining, showing increase in Yan puncta. J) Lethality assays of Yan and SY3-5, measured by failure of eggs to hatch during embryogenesis. Each trial represents n=800 eggs of each genotype, with the exception of SY5, which was n=145 and 161 respectively.

As a primary test of genetic fitness, we measured lethality at the embryonic stage of development, in which Yan plays key roles to repress transcription (Halfon et al. 2000; Rogge et al. 1995). Figure 3.5J shows that lethality relative to wild type increased significantly in the case of SY3, to approximately 20%, and that SY4 and SY5 showed a further elevated lethality of approximately 40%. Strong Yan null mutants have been previously shown to have a 100% lethality rate (Webber et al. 2013). Consistent with the classic Yan phenotype of failure to close the anterior portion of the head cuticle, we frequently observed eggs with an anterior open phenotype, indicative of Yan loss of function. Although we did not measure fertility rates, we were unable to maintain SY5 as a homozygous stock. Additionally, animals that survived embryogenesis eclosed with rough eyes, consistent with Yan's known roles in eye development. Although we noted some in degree of eye roughness within each (data not shown), overall the severity of the phenotypes correlated positively with increased SAM-SAM affinity (Figure 3.5A-D) Thus, in contrast to the trend observed by over-expression, in the context of the endogenous locus, the severity of the loss of function phenotypes are proportional to the increase in SAM-SAM affinity.

Examination of Yan expression during development revealed striking changes to Yan sub-cellular localization, with the formation of Yan puncta in both the nucleus and cytoplasm and an overall increase in the levels of Yan staining (Figure 3.5E-H). The number of puncta formed roughly followed the proportional increase expected from the increased SAM-SAM affinity of the SY3-5 mutants. Mitotic clones of SY4 confirmed the increase in both levels and punctual distribution relative to wild type (Figure 3.5I-I'). Thus, the cellular phenotype of the SuperYan mutants suggests a loss of function by

sequestration mechanism, where puncta formation in the cytoplasm decreases nuclear localization and puncta formation in the nucleus prevents productive binding to target enhancers. We were intrigued as to why puncta were not observed in over-expression experiments in S2 cells and in eye discs, and hypothesized that endogenous Yan is capable of intercalating with SuperYan in order to break up large polymers. In the over-expression of SY3-5 in eye discs, puncta were observed under the strongest over-expression conditions (Fig 3.S4R-T), but not to the extent seen in the homozygous SuperYan alleles, consistent with this hypothesis. As a further test of this hypothesis, we over-expressed Yan and SY5 in the developing wing where there is no endogenous Yan, and observed puncta in the case of SY5, but not for Yan wild type (Figure 3.S5A-B).

If increased affinity leads to a distribution of Yan polymer sizes incompatible with normal function, then disruption of higher-order Yan polymers should suppress SuperYan phenotypes. We tested two different interactions that achieve this *in vitro*. First, monomerizing mutations (V105R or A86D) completely block polymerization (Figure 3.S1), and so if co-expressed with wild type Yan or SY, should cap and disrupt the extent of polymerization. Indeed, transheteroallelic combinations of SY5 and YanV105R or YanA86D had eyes with completely wild type morphology (Figure 3.6D-E and data not shown). Secondly, Mae is a protein that contains a SAM domain with high affinity for Yan, that can both limit *in vitro* polymerization and antagonize Yan function S2 cells and *in vivo* (Qiao et al. 2004; Tootle et al. 2003). Mae proved to be a potent suppressor of the Super Yan phenotype. Thus expression from the UAS-Mae transgene was sufficiently leaky at 25°C that it suppressed the SY5 rough eye phenotype; flies of

the same genotype raised at 20°C had rough eyes qualitatively comparable to those of SY5 alone (Figure 3.6A-C).

Finally, we characterized the specific eye phenotype of the SuperYan mutants at the cellular level, and uncovered an intriguing phenotype. Histological sections revealed over-recruitment of photoreceptors, further consistent with increased SAM affinity reducing Yan function (Figure 3.6G-I). However in contrast to characterized *yan* hypomorphs where extra R7 cells are most frequent (for example in yan^{1} animals, 88% of ommatidia have extra R7s but only 27% have extra outer photoreceptors (Lai and Rubin 1992)), the SY alleles predominantly increased the number of outer photoreceptor rhabdomeres, with ~30% of SY4 or SY5 ommatidia recruiting at least one extra outer rhabdomeres, often near the R3/R4 position (Figure 3.6F). This increase in outer photoreceptors matches exactly the increase in yan^{1}, but the corresponding lack of effect on R7 specification suggests that the increased SAM-SAM affinity somehow uncouples the two loss of function phenotypes associated with classic *yan* hypomorphs. Consistent with the adult sections, SY5 eye imaginal discs showed normal induction of Pros-positive R7 cells (Figure 3.S5C-D) and increased specification of Salm-positive R3/4 cells (Figure 3.S5 E-F).

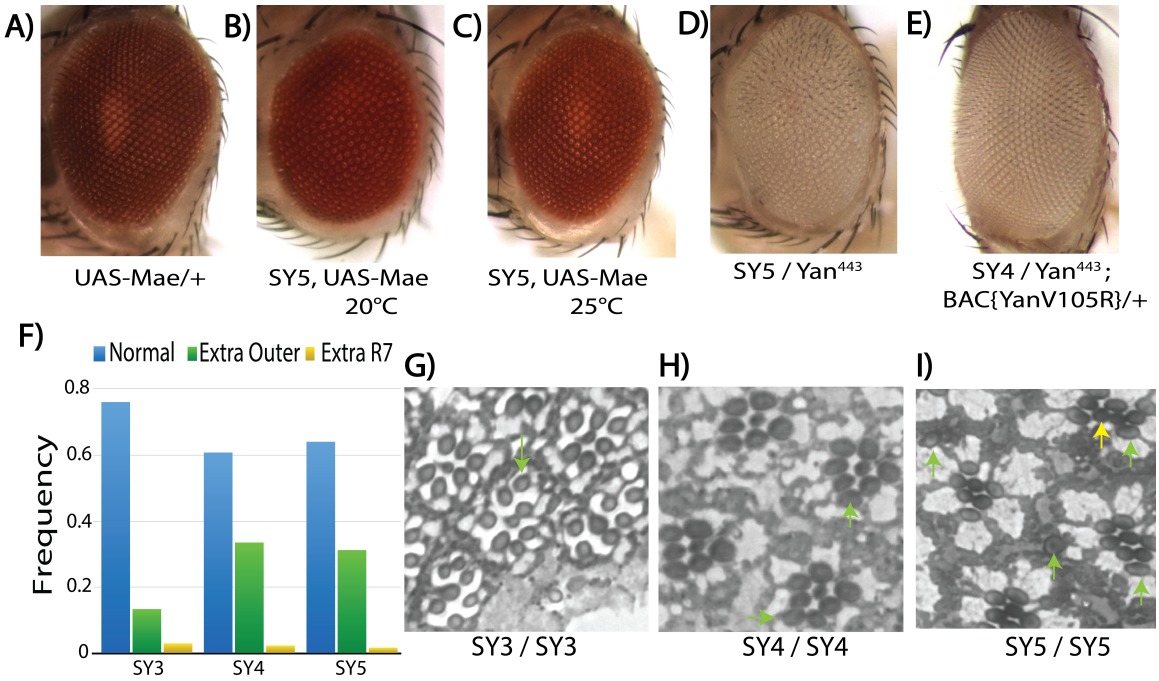


Figure 3.6: SuperYan phenotypes can be rescued by perturbations that shift Yan polymer distribution

A-C) Adult eye images of modification of SY5 phenotype with UAS-Mae, a capping protein for Yan polymers D-E) Adult eye images for rescue of SY4 phenotype with an exogenous copy of Yan monomer, carried on a BAC transgene F) Quantification of photoreceptor recruitment defects, comparing normal ommatidia to ommatidia with extra outer photoreceptors or R7 photoreceptors G-I) Histological staining of SY3-5 mutants, showing predominant increase in extra outer photoreceptors (green arrows) versus R7 photoreceptors (observed rarely, yellow arrows)

Taken together our results support a model where the SuperYan alleles behave as genetic hypomorphs due to excessive self-association, resulting in large punctal aggregates that sequester functional Yan away from its target enhancers. However, the distinct consequences on outer versus R7 photoreceptor recruitment suggest there may be different requirements for or different regulation of polymerization in different cell types.

3.4 Discussion

In order to achieve a clearer picture of how the stability of transcriptional regulatory complexes depends on the strength of protein-protein interactions within the complex, we

followed the predictions of a Yan occupancy model and created a tunable suite of mutants to increase Yan polymerization. Our results demonstrate that the SuperYan mutations increase the affinity of Yan polymerization both *in vitro* and *in vivo*, and that unexpectedly, increasing the strength of polymerization induces a loss of function phenotype that arises from sequestration into large puncta of Yan. Thus, we have uncovered a role for medium-affinity TF interactions in maintaining the function of transcriptional regulatory complexes, which we envision as analogous to the role of medium-affinity protein-DNA interactions at enhancer binding sites. Whereas recent investigations that have explored the role of medium-affinity TF binding sites in enhancers have noted that these sites are required for proper transcriptional output, the requirement for medium-affinity polymerization of Yan arises not just at the level of binding to an enhancer. Instead, the requirement for polymerization at the wild type level of Yan is balanced across all of Yan's roles as a transcription factor— from localization to DNA search to nuclear export upon signal activation— suggesting that polymerization shapes a broad swath of Yan's behavior. To our knowledge, this is the first demonstration of increased protein-protein interaction affinities altering the transcriptional and functional outputs of a TF, and we suspect that this relationship will be applicable to other TFs, including non-polymerizing TFs. Further, our results imply that protein-protein interaction affinity may be a key parameter tuned through evolution in order to balance competing requirements for TFs during development, especially considering the small number of residue changes that resulted in dramatic polymerization shifts we observed.

In addition to the aforementioned roles of polymerization, our exploration of the loss of function phenotype of the SuperYan mutants in the adult eye strongly suggests that increasing polymerization alters the output of Yan's regulation of different target genes. If increasing polymerization evenly depleted the pool of active Yan by sequestration, one would imagine the SuperYan mutants would recreate the known phenotypes of weak Yan hypomorphs. Instead, we observed an uncoupling of the classic Yan hypomorph phenotype, with a strong increase in extra outer photoreceptors, but no increase in R7 photoreceptors. The key marker of R7 identity, Prospero, has previously been shown to be a Yan target gene (Xu et al. 2000), and yet we observed no increase in Prospero staining in SY5 as would be expected for relief of Yan repression. We believe that this hints at different requirements for Yan polymerization at different target genes throughout development, although we can not rule out the possibility that increased MAPK signaling in R7 cells (Doroquez and Rebay 2006) compensates for over-polymerization by more strongly antagonizing Yan polymerization. Although the molecular underpinning of this potential difference in target gene regulation remains to be elucidated, our results imply that altering protein-protein interaction affinity can have effects at the level of transcriptional complex formation, in addition to the broader effects on Yan localization.

Our *in vitro* results cement the role of charged residues at the periphery of the SAM-SAM interface in tuning the affinity of polymerization. Previously, TEL residues K99, E100, and R103 (Yan residues R92, A93, R96) have been shown to contribute to TEL polymerization both *in vitro* and in terms of signaling output of the TEL-NTRK3 oncogenic fusion protein (Cetinbas et al. 2013). Our results confirm their importance and

show that the Yan H97Y is an additional mutation that further modifies the polymerization of the SAM domain, past the affinity of wild type TEL (SY4) to sub-nanomolar affinity (SY5). Interestingly, this suggests that there are residues in TEL that prevent it from achieving the higher polymerization affinity of SY5, and may further argue that although TEL polymerization is very strong, it too is not tuned to maximal affinity. Further exploration of the differences between SY5 and TEL may be able to illuminate the identity of the residues responsible for the difference in affinity. Additionally, we note that, to our knowledge, SY5 is the strongest quantitatively measured SAM-SAM interaction, and engineering of other SAM domain structures outside of the ETS family, such as Polyhomeotic SAM or Sex Comb on Midleg (Scm) SAM (Qiao and Bowie 2005), might be aided by taking advantage of the strategy of placing salt-bridge residues at the periphery of the SAM-SAM interface.

Lastly, the punctual phenotype of the SuperYan mutants has a striking resemblance to dynamic protein aggregates, also known as membrane-less organelles, stress granules, or liquid-like droplets (Kaganovich 2017). The biology of these structures is diverse and still emerging, but it is clear that in some cases, protein-protein interactions can drive the formation of functional aggregates that respond to cellular stimuli. Especially important are the role of low-complexity domains or intrinsically disordered regions, which are frequently enriched in TFs (Staby et al. 2017). Formation of puncta can be observed occasionally in wild type Yan, although not nearly as strikingly as in the SuperYan mutants. Therefore, we speculate that Yan, like many TFs, may endogenously participate in granule-like complexes, and that the SuperYan mutations dramatically shift the propensity to form these complexes. Importantly, we cannot make a determination as

to how dynamic these puncta may be without live cell imaging, which is a key aspect of the biology of granule-like complexes. However, we note that MAPK phosophorylation of Yan may serve as switch on the formation of puncta, and further study on the effect of this phosphorylation on Yan and the SuperYan mutants is warranted.

3.5 Materials and Methods

3.5.1 Molecular biology and cloning

The Yan SAM domain (YanS33-S117) and the TEL SAM domain (TELA40-Q123) were subcloned into the His-A scGFP-30 /pBAD (kindly provided by James Bowie and Catherine Leetola) through PCR with primers MH53+MH54 and MH55+MH56 respectively, followed by digestion with NotI and HindII and low melting agarose (LMA) ligation. Subsequent generation of point mutations YT1-15, TELA93D, TELV112E, YanA86D, YanV105R, and constructs for SPR were carried out with the indicated primers via Quikchange Mutagenesis (Stratagene), followed by sequencing. Of the constructs used for SPR, either a TEV cleavage site or a 6XHis-tag with TEL cleavage site was added to the construct by annealing oligos MH300+MH301 or MH302+303 respectively, and LMA ligating into the relevant YTnegGFP/pBAD construct digested with KpnI and NotI.

To generate constructs for co-immunoprecipitation, a previously generated full length YanWT/pENTR3C construct (Zhang et al. 2010) was used for LR-mediated exchange into pAWF and pAWH expression vectors, per the manufacturer's instructions. The C-terminal stop codon was removed via Quikchange mutagenesis with primers MH308+309. Since the Yan SAM domain exists on an endogenous EcoRI fragment within the wild type Yan sequence, SY1-5 mutations were generated by sub-cloning the

relevant fragments into EcoRI-digested YanWT/pAWF and /pAWH from the pBAD expression constructs.

To generate constructs for S2 cell transcription assays and S2 cell staining, a previously generated untagged YanWT/pMT was used as a substrate for Quikchange with the relevant primers to generate SY1-5/pMT. Yan^{Act}/pMT, RasV12/pMT, and pMT empty were generated previously. FLAG-PntP1/pMT was generated via PCR with MH15+MH50 followed by digestion with KpnI and SalI and LMA ligation into N-terminal FLAG-pMT vector.

To generate constructs for UAS mediated over-expression, the full length Yan sequence was sub cloned into pUASTattb vector using PCR with primers MH40(find thomas Kozak sequence)+MH41 and the YanWT/pMT vector as template, followed by digestion with KpnI and XbaI and LMA ligation. UAS-SY1-5 were cloned identically using the respective pMT constructs as template for PCR. YanWTIntNLS/pMT vector was generated previously(Tootle et al. 2003), and the appropriate primers were used to Quikchange the vector to generate SY4IntNLS/pMT, and then UAS constructs from these vectors were generated as described above.

To generate genomic rescue constructs for CRISPR/Cas9-mediated mutagenesis, primers MH250+MH251 were used to amplify the Yan locus using genomic DNA extracted from the strain w-,y-,vasa>cas9, at positions 2L:2,159,158 - 2L:2,161,166 (Dm Assembly 6). This generated a ~2kb product with KpnI and SacI restriction sites that was sub cloned into pBS-SKII+ using KpnI and SacI digestion followed by LMA ligation. To generate SY3-5, the appropriate Quikchange primers were used on the Yan genomic construct. To generate the gRNA construct used for cleavage, primers MH_+MH_ were

annealed and ligated into BbsI digested pU6-gRNA vector. Note that sequencing of the genomic construct generated from the w-,y-,vasa>cas9 background revealed a silent mutation (YanThr91 ACG->ACA) which was incorporated into the gRNA used to allow efficient cleavage.

3.5.2 Fly strains and genetics

UAS constructs of Yan, SY1-5, YanIntNLS, and SY4IntNLS were generated by injecting into the stock vasa>phiC31;;86Fb-attB (BL#247249). CRISPR/Cas9-mediated mutagenesis was accomplished by injecting into the stock vasa>Cas9. The following lines are also available from the Bloomington stock center: w1118, FRT40A, GMR-GAL4, dpp-GAL4/TM6B, was used as a wild type comparison (BL5905). SY4,FRT40A was generated by recombining the generated SY4 allele with FRT40A. Yan^{443}, and Yan-V105R BAC were generated and used previously (Webber et al. 2013). UAS-Mae constructs were generated previously(Tootle et al. 2003).

3.5.3 negGFP Native Gel Assay

Assay was adapted from (Knight et al. 2011). YTnegGFP/pBAD constructs were transformed into DH10B cells. 200mL expression cultures were grown at 37C for 2.5 hours until cultures had reached OD600 0.6-0.8 and then induced at 18C with 0.2% L-(+)-arabinose (Sigma) for 8 hours. Cells were pelleted, frozen, and then lysed in SAM Native Gel Lysis Buffer (20mM Tris, 1M NaCl, 5mM $MgCl_2$, pH 7.5) with 1mM DTT, 1mg/mL lysozyme, and cOmplete Mini protease inhibitor tablets (1 tablet per 10mLs of buffer; Roche). Lysis proceeded for 30min with rotating at 4C, followed by sonication (Fisher Sonic Dismembrator Model 500, 1 min total time, 10 sec on, 10 sec off, 20% amplitude). Lysates were cleared by spinning at 16,000 x g at 4C for 15min, and then

incubated with HisPur Cobalt Resin (Thermo) in Binding Buffer (50mM sodium phosphate, 300mM NaCl, 10mM imidazole, pH7.4) for 60 minutes, according to the manufacturers instructions. Beads were washed twice with two bed volumes of Binding Buffer, and then eluted in Elution Buffer (Binding Buffer + 150mM imidazole). The fluorescence of YT constructs was measured with a 96-well plate reader (Synergy Neo HST) diluted 1:10 in Elution Buffer, with filters set to an excitation wavelength of 485nm, emission 516nm, and gain set to 35. Measurements were buffer subtracted and constructs were standardized to 375,00 RFU.

To run the native gel assay, 10-well native acrylamide gels were poured at 12.5% and 6%, with a stacker of 3% acrylamide. Samples were loaded with 5X Native Gel Sample Buffer (300mM Tris, 50% glycerol, bromphenol blue, pH 6.8), and gels were run in Native Gel Running Buffer (40mM Tricine, 60mM Tris) at 4C at 22V for ~48 hours (12.5% gels) or ~18 hours (6%). Gels in Figure S1 were run identically, except gradient gels (Mini-PROTEAN TGX Stain-Free, 4-15%, BioRad) were used instead of a single percentage gels, run for ~24 hours. Gels were imaged on a Typhoon excitation 488nm and emission of 516nm.

3.5.4 SPR purification and measurement of affinity

Constructs for SPR were purified in the same manner as for the native gel assay. Following elution, constructs were cleaved with TEV protease overnight at 4C at a A_{280} ratio of construct to TEV of 100:1. Constructs were further purified with an AktaPure FPLC system equipped with a HiLoad 16/600 Superdex75 column, in SAM Native Gel Lysis Buffer. Samples eluted in two peaks consistent with the larger negGFP fragment and the smaller SAM domain, and the peak centered on 80mL of elution volume were

collected and concentrated using Amicon Ultra 15- Ultracel 3K centrifugal concentration filters (Millipore). Samples were flash frozen in liquid nitrogen using 10% glycerol as cryoprotectant. Residual His-binding was removed via another 1 hour incubation with HisPur Cobalt Resin *in situ*, and concentrations were measured via A_{280} in triplicate with a NanoDrop ND-1000 Spectrophotometer. SPR measurements were taken with a ProteonXPR36 machine (BioRad) using a HTE chip. 5mM $NiSO_4$ was used to regenerate the chip, and ligand was used at a concentration of 100nM for each interaction. Sensograms were measured in duplicate, and fit to a two-state kinetic binding model.

3.5.5 Co-immunoprecipitation experiments

Drosophila S2 cells were maintained in Schnieder Insect Media (Sigma) with Insect Media Supplement (Sigma) and peniccilin/streptamycin. Cells were grown confluence and transfected using 250ug/mL DDAB (Sigma). 1500ng of the appropriate Yan construct (in pAWF or pAWH expression vectors) were transfected with 9.0E6 cells in 6mL cultures, with pMT empty used to standardize total DNA transfected. 48 hours after transfection, cells were pelleted for 3min at 4C at 350 x g and resuspended in IP lysis buffer (50mM Tris, 100mM NaCl, 1% NP-40, 2mM EDTA, 2mM EGTA, pH 8.0) with cOmplete Mini protease inhibitor tablets and 0.5mM DTT. Cells were syringe-passed seven times with a 27.5 gauge syringe to induce lysis, and incubated at 4C for 30min. Lysates were cleared by centrifuging at 16,000 x g at 4C for 10min, and incubated with 20uL of Anti-FLAG-M2 Affinity Gel (Sigma) in a 1:1 slurry for 2 hours. Samples were washed three times with 500uL of IP lysis buffer, boiled, and both lysates and IP were run on an SDS PAGE gel. The gel was transferred to a nitrocellulose membrane, blocked with 1% casein for 1 hour, and incubated with 1:3000 mouse anti-HA and 1:1000 rabbit

anti-FLAG overnight at 4C. Blots were washed 5X with TBST and incubated with 1:2000 Alexa 488 and 1:2000 Alexa 600 for two hours, washed 4X with TBST, 1X with TBS and imaged.

3.5.6 Transcription assays

Drosophila S2 cells were transfected with 100ng of *king-tubby* luciferase or 100ng of 6X-ETS luciferase, 100ng of the relevant Yan/pMT construct, 100ng of PntP1/pMTFLAG, 20ng of actin>Renilla luciferase, and if applicable, 5ng of RasV12/pMT using DDAB as described above. Cells were grown in 1.5mL cultures and transfected with 2.25E6 cells per condition, and induced with 700mM $CuSO_4$ at 1:1000 24 hours after transfection. 24 hours after induction, cells were collected by spinning for 2min at 835 x g at 4C, lysed in 170uL Transcription Assay Lysis Buffer (100mM Potassium Phosphate, 0.5% NP-40, pH7.8), and incubated at 4C for 30min. Lysates were cleared by centrifuging for 10 min at 16,000 x g at 4C. Luciferase measurements were made using an Autolumat Plus LB 953, for both firefly luciferase and renilla luciferase using Luciferase Buffer (10mM Mg Acetate, 100mM Tris Acetate, 1mM EDTA, pH 7.8) with 4.5mM ATP (Fisher) and 77uM D-luciferin (Pierce), and Renilla Buffer (25mM sodium pyrophosphate, 10mM Na Acetate, 15mM EDTA, 500mM Na_2SO_4 500mM NaCl, pH 5.0) with 4mM coelenterazine (Promega) in ethanol at 1:1000. Measurements were made in technical triplicates using 50uL per sample to measure biological replicates, using the same tube for both firefly and renilla measurements, with renilla activity as a transfection control. The ratio of Firefly RLU to Renilla RLU was taken as transcriptional activity, and all measurements were normalized to reporter alone.

3.5.7 Drosophila S2 cell staining

Drosophila S2 cells were transfected and induced with 1000ng of the relevant Yan/pMT construct, and 50ng RasV12/pMT if applicable, as above. Cells were settled on poly-L lysine coated slides for one hour, fixed with 4% para-formaldehyde and 0.1% Triton for 10min, washed 5X with PBST, incubated with mouse anti-Yan monoclonal antibody at 1:500 (DSHB 8B12H9) in PBST with 5% normal goat serum (NGS) for 1 hour, washed 5X with PBST, incubated with anti-mouse Cy3 1:2000 and DAPI 1:2000 in PBST with 5% NGS for 1 hour, washed 5X with PBST and mounted with n-propyl gallate mounting medium. Sub-cellular localization was assessed via immunofluoresence with a light microscope.

3.5.8 Generation of transgenic UAS- and CRISPR/Cas9 constructs

Flies were injected broadly in accordance with the procedure outlined in (Fujioka M Dev. Bio. Protocols II). 1ug/uL of Yan/puASTattB constructs were injected, and 1ug/uL of the relevant genomic rescue construct and 50ng/uL of gRNA were injected for Cas9 mutagenesis. UAS constructs were screened for insertion of a mini-w^+ marker, while a preliminary screen for CRISPR-mediated mutagenesis was carried out by crossing to T2B (sev>RasV12)(Rebay et al. 2000), selecting for enhancement of the rough eye phenotype. Candidates of SY3-5 were further screened by genomic DNA extraction and PCR, and confirmed by sequencing.

3.5.9 Staining, Immunofluoresence of Imaginal Discs, Microscopy

White pre-pupal 3rd instar larvae were dissected in S2 cell medium and eye or wing discs were fixed with 4% paraformaldehyde with 0.1% Triton, washed 3X in PBST, blocked for 1 hour in PBST+ 5% NGS, and incubated overnight in primary antibody in PBST +

NGS. Antibodies used include: mouse anti-Yan monoclonal antibody at 1:500 (DSHB 8B12H9), guinea pig anti-salm polyclonal antibody at 1:500 (a gift from Claude Desplan), and mouse anti-Pros monoclonal antibody at 1:50 (MR1A). Discs were washed with PBST 3X and incubated with secondary antibodies for two hours, washed again and mounted. Secondaries used include Cy3 conjugated goat anti-mouse (Jackson Immunoresearch) and anti-guinea pig (Jackson Immunoresearch) and DAPI, all at 1:2000. Confocal images were taken on a Zeiss LSM 880 microscope, using 0.5um to 1.0um slices to step through the desired part of the tissue and then maximally projected. To image adult eyes, adults were decapitated and imaged with a Canon EOS Rebel camera fitted to a Leica dissecting microscope. Photographs were merged using a "Perfect Focus" macro, executed within iSolution Lite software (IMT-Digital).

3.5.10 Modeling of Yan fractional occupancy

Fractional occupancy of Yan was calculated largely in accordance to previously described methods (Hope et al. 2017), and we refer the reader to the methods section of the paper for extensive explanation. The modification of the model to account for polymerization off of DNA takes the form of a multiplier on the effective concentration of free Yan to bind DNA. The fraction of free Yan monomers is given by $1 / \{1 + [Yan]/K_d + ([Yan]/K_d)^2 + ([Yan]/K_d)^3 + \ldots ([Yan]/K_d)^n\}$ where [Yan] is the concentration of total Yan, K_d is the dissociation constant of SAM-SAM affinity, and *n* is the maximum length of polymer. All calculations were performed at $n=50$, as increasing *n* was shown to asymptotically approach the same fractional occupancy curves.

3.6 Acknowledgements

The authors would like to thank James Bowie, Catherine Leetola, and Claude Desplan for providing reagents, John Reinitz for discussion on updating the model, and members of the Rebay lab for critical reading and helpful discussion on the manuscript. MH and IR are supported by NIH grant R01 GM080372. MH has been supported by NIH grants T32 GM007183 and 2T32 HL007381-36A1.

3.7 Author contributions

M.H. and I.R. designed experiments; M.H. and I.R. performed experiments; M.H. wrote the modeling code; M.H. and I.R. analyzed data; M.H. and I.R. wrote the paper. The authors declare no conflicts of interest.

3.8 References

Biggin, Mark D. 2011. “Animal Transcription Networks as Highly Connected, Quantitative Continua.” *Developmental Cell*.

Cary, Gregory A., Alys M. Cheatle Jarvela, Rene D. Francolini, and Veronica F. Hinman. 2017. “Genome-Wide Use of High- and Low-Affinity Tbrain Transcription Factor Binding Sites during Echinoderm Development.” *Proceedings of the National Academy of Sciences* 114 (23):5854–61.

Cetinbas, Naniye, Helen Huang-Hobbs, Cristina Tognon, Gabriel Leprivier, Jianghong An, Steven McKinney, Mary Bowden, et al. 2013. “Mutation of the Salt Bridge-Forming Residues in the ETV6-SAM Domain Interface Blocks ETV6-NTRK3-Induced Cellular Transformation.” *Journal of Biological Chemistry* 288 (39):27940–50.

Crocker, Justin, Namiko Abe, Lucrezia Rinaldi, Alistair P. McGregor, Nicolás Frankel, Shu Wang, Ahmad Alsawadi, et al. 2015. “Low Affinity Binding Site Clusters Confer HOX Specificity and Regulatory Robustness.” *Cell* 160 (1–2):191–203.

Davidson, Eric H, and Michael S Levine. 2008. “Properties of Developmental Gene Regulatory Networks.” *PNAS* 105 (51):20063–66.

De, Soumya, Anson C K Chan, H. Jerome Coyne, Niraja Bhachech, Ulrike Hermsdorf, Mark Okon, Michael E P Murphy, Barbara J. Graves, and Lawrence P. McIntosh.

2014. “Steric Mechanism of Auto-Inhibitory Regulation of Specific and Non-Specific Dna Binding by the Ets Transcriptional Repressor ETV6.” *Journal of Molecular Biology* 426 (7):1390–1406.

Doroquez, David B., and Ilaria Rebay. 2006. “Signal Integration during Development: Mechanisms of EGFR and Notch Pathway Function and Cross-Talk.” *Critical Reviews in Biochemistry and Molecular Biology*.

Evans, Nicole C., Christina I. Swanson, and Scott Barolo. 2012. “Sparkling Insights into Enhancer Structure, Function, and Evolution.” *Current Topics in Developmental Biology* 98:97–120.

Farley, E. K., K. M. Olson, W. Zhang, A. J. Brandt, D. S. Rokhsar, and M. S. Levine. 2015. “Suboptimization of Developmental Enhancers.” *Science* 350 (6258):325–28.

Funnell, Alister P W, and Merlin Crossley. 2012. “Homo- and Heterodimerization in Transcriptional Regulation.” *Advances in Experimental Medicine and Biology*.

Halfon, Marc S, Ana Carmena, Stephen Gisselbrecht, Charles M Sackerson, F Jiménez, Mary K Baylies, and Alan M Michelson. 2000. “Ras Pathway Specificity Is Determined by the Integration of Multiple Signal-Activated and Tissue-Restricted Transcription Factors.” *Cell* 103 (1):63–74.

Hope, C. Matthew, Ilaria Rebay, and John Reinitz. 2017. “DNA Occupancy of Polymerizing Transcription Factors: A Chemical Model of the ETS Family Factor Yan.” *Biophysical Journal* 112 (1):180–92.

Junion, Guillaume, Mikhail Spivakov, Charles Girardot, Martina Braun, E. Hilary Gustafson, Ewan Birney, and Eileen E M Furlong. 2012. “A Transcription Factor Collective Defines Cardiac Cell Fate and Reflects Lineage History.” *Cell* 148 (3):473–86.

Kaganovich, Daniel. 2017. “There Is an Inclusion for That: Material Properties of Protein Granules Provide a Platform for Building Diverse Cellular Functions.” *Trends in Biochemical Sciences*, 2017.

Kazemian, Majid, Hannah Pham, Scot A. Wolfe, Michael H. Brodsky, and Saurabh Sinha. 2013. “Widespread Evidence of Cooperative DNA Binding by Transcription Factors in Drosophila Development.” *Nucleic Acids Research* 41 (17):8237–52.

Kim, C. A., M. L. Phillips, W. Kim, M. Gingery, H. H. Tran, M. A. Robinson, S. Faham, and J. U. Bowie. 2001. “Polymerization of the SAM Domain of TEL in Leukemogenesis and Transcriptional Repression.” *EMBO Journal* 20 (15):4173–82.

Knight, Mary Jane, Catherine Leettola, Mari Gingery, Hao Li, and James U. Bowie. 2011. “A Human Sterile Alpha Motif Domain Polymerizome.” *Protein Science* 20

(10):1697–1706.

Lai, Zhi Chun, and Gerald M. Rubin. 1992. “Negative Control of Photoreceptor Development in Drosophila by the Product of the Yan Gene, an ETS Domain Protein.” *Cell* 70 (4):609–20.

Lawrence, Michael S., Kevin J. Phillips, and David R. Liu. 2007. “Supercharging Proteins Can Impart Unusual Resilience.” *Journal of the American Chemical Society* 129 (33):10110–12.

Lorberbaum, David S., Andrea I. Ramos, Kevin A. Peterson, Brandon S. Carpenter, David S. Parker, Sandip De, Lauren E. Hillers, et al. 2016. “An Ancient yet Flexible Cis-Regulatory Architecture Allows Localized Hedgehog Tuning by patched/Ptch1.” *eLife* 5 (MAY2016).

O’Neill, Elizabeth M., Ilaria Rebay, Robert Tjian, and Gerald M. Rubin. 1994. “The Activities of Two Ets-Related Transcription Factors Required for Drosophila Eye Development Are Modulated by the Ras/MAPK Pathway.” *Cell* 78 (1):137–47.

Parker, David S, Michael A White, Andrea I Ramos, Barak A Cohen, Scott Barolo, S. Barolo, J. W. Posakony, et al. 2011. “The Cis-Regulatory Logic of Hedgehog Gradient Responses: Key Roles for Gli Binding Affinity, Competition, and Cooperativity.” *Science Signaling* 4 (176):ra38.

Qiao, Feng, and James U Bowie. 2005. “The Many Faces of SAM.” *Science’s STKE : Signal Transduction Knowledge Environment* 2005 (January):re7.

Qiao, Feng, Haiyun Song, Chongwoo A. Kim, Michael R. Sawaya, Jacob B. Hunter, Mari Gingery, Ilaria Rebay, Albert J. Courey, and James U. Bowie. 2004. “Derepression by Depolymerization: Structural Insights into the Regulation of Yan by Mae.” *Cell* 118 (2):163–73.

Ramos, Andrea I, and Scott Barolo. 2013. “Low-Affinity Transcription Factor Binding Sites Shape Morphogen Responses and Enhancer Evolution.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 368:20130018.

Rebay, Ilaria, Fangli Chen, Francis Hsiao, Peter A. Kolodziej, Bing H. Kuang, Todd Laverty, Chris Suh, Matthew Voas, Andrina Williams, and Gerald M. Rubin. 2000. “A Genetic Screen for Novel Components of the Ras/mitogen-Activated Protein Kinase Signaling Pathway That Interact with the Yan Gene of Drosophila Identifies Split Ends, a New RNA Recognition Motif-Containing Protein.” *Genetics* 154 (2):695–712.

Rebay, Ilaria, and Gerald M. Rubin. 1995. “Yan Functions as a General Inhibitor of Differentiation and Is Negatively Regulated by Activation of the Ras1/MAPK

Pathway." *Cell* 81 (6):857–66.

Reinitz, John, and David H. Sharp. 1995. "Mechanism of Eve Stripe Formation." *Mechanisms of Development* 49 (1–2):133–58.

Rogge, R, P J Green, J Urano, S Horn-Saban, M Mlodzik, B Z Shilo, V Hartenstein, and U Banerjee. 1995. "The Role of Yan in Mediating the Choice between Cell Division and Differentiation." *Development (Cambridge, England)* 121 (12):3947–58.

Smith, Edwin, and Ali Shilatifard. 2014. "Enhancer Biology and Enhanceropathies." *Nature Structural & Molecular Biology* 21 (3):210–19.

Spitz, Francois, and Eileen E M Furlong. 2012. "Transcription Factors: From Enhancer Binding to Developmental Control." *Nature Reviews Genetics* 13:613–26.

Staby, Lasse, Charlotte O'Shea, Martin Willemoës, Frederik Theisen, Birthe B. Kragelund, and Karen Skriver. 2017. "Eukaryotic Transcription Factors: Paradigms of Protein Intrinsic Disorder." *Biochemical Journal* 474 (15):2509–32.

Tootle, Tina L, Philina S Lee, and Ilaria Rebay. 2003. "CRM1-Mediated Nuclear Export and Regulated Activity of the Receptor Tyrosine Kinase Antagonist YAN Require Specific Interactions with MAE." *Development (Cambridge, England)* 130 (5):845–57.

Webber, Jemma L., Jie Zhang, Lauren Cote, Pavithra Vivekanand, Xiaochun Ni, Jie Zhou, Nicolas Nègre, Richard W. Carthew, Kevin P. White, and Ilaria Rebay. 2013. "The Relationship between Long-Range Chromatin Occupancy and Polymerization of the Drosophila Ets Family Transcriptional Repressor Yan." *Genetics* 193 (2):633–49.

Xu, Chunyan, Rachele C. Kauffmann, Jianjun Zhang, Susan Kladny, and Richard W. Carthew. 2000. "Overlapping Activators and Repressors Delimit Transcriptional Response to Receptor Tyrosine Kinase Signals in the Drosophila Eye." *Cell* 103 (1):87–97.

Zhang, Jie, Thomas G W Graham, Pavithra Vivekanand, Lauren Cote, Maureen Cetera, and Ilaria Rebay. 2010. "Sterile Alpha Motif Domain-Mediated Self-Association Plays an Essential Role in Modulating the Activity of the Drosophila ETS Family Transcriptional Repressor Yan." *Molecular and Cellular Biology* 30 (5):1158–70.

3.9 Supplemental Figures

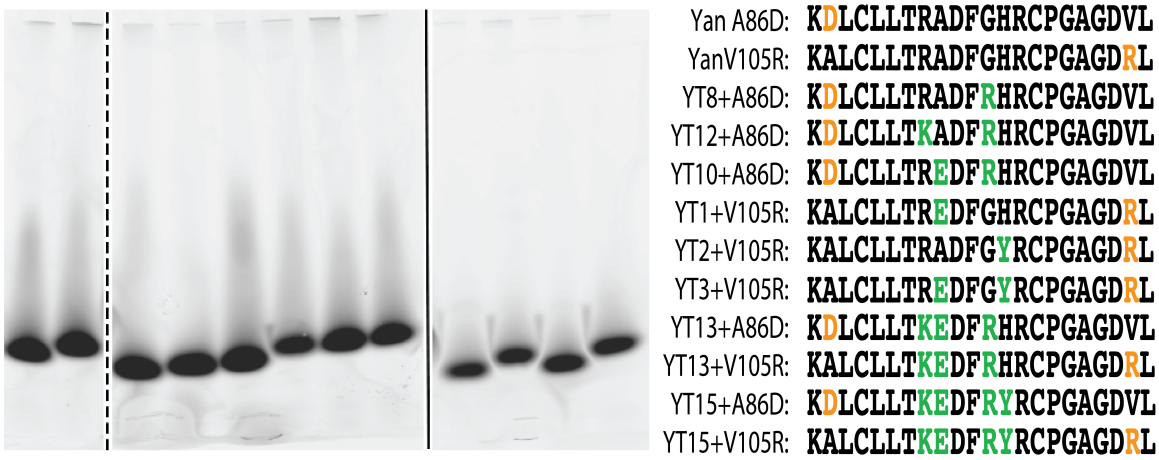


Figure 3.S1: Yan A86D and V105R mutations are capable of monomerizing SuperYan mutations, to allow measurement of affinity.

negGFP native gel assay for various Yan constructs, listed in right-hand panel. All constructs that contain a monomerizing mutation run as monomers. Note that in contrast to single percentage native gels in Figure 1, the gels shown are 4-15% gradient gels. Image shown is taken from three separate gels loaded and run in parallel.

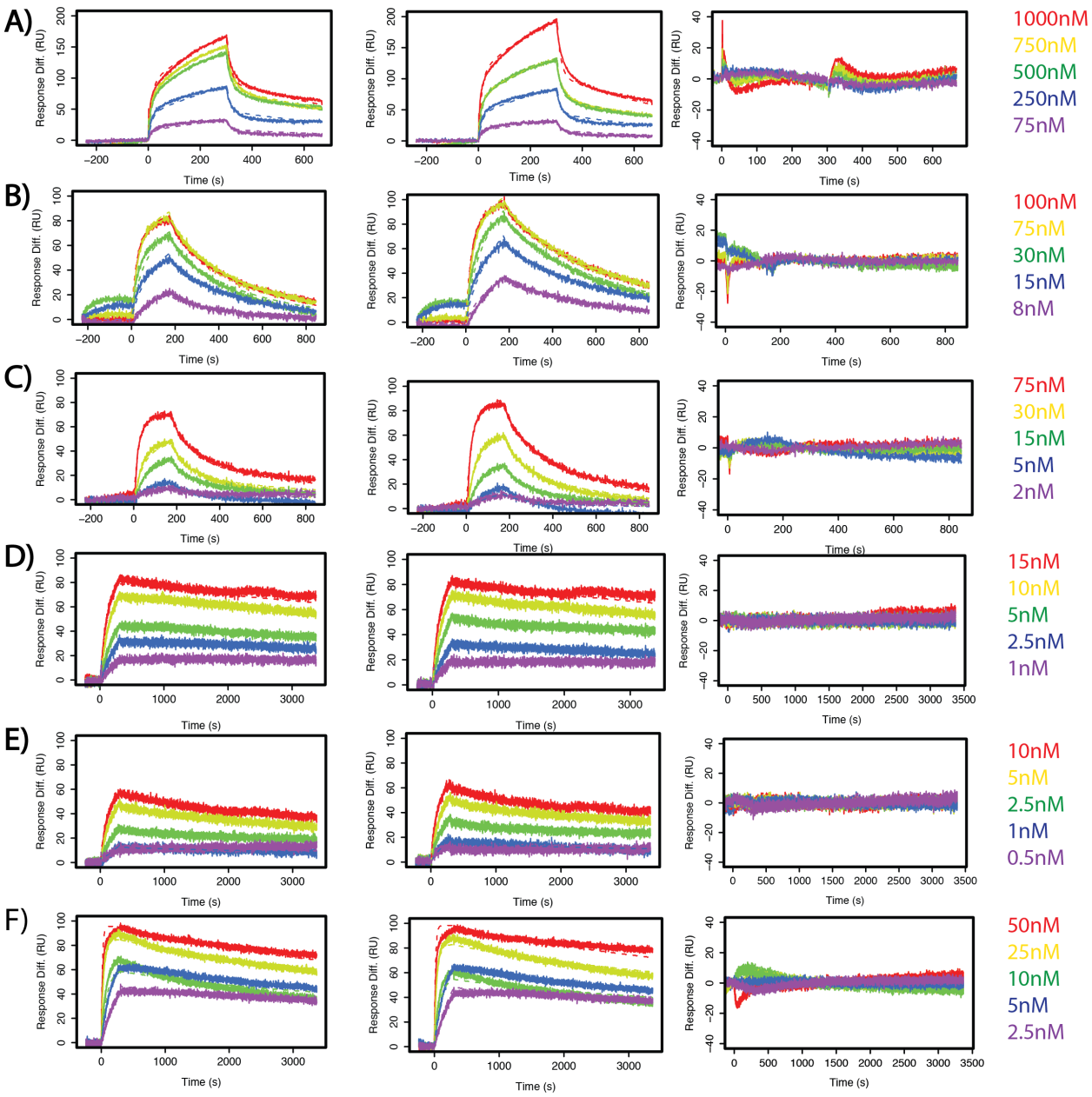


Figure 3.S2: SuperYan mutations increase SAM-SAM affinity via SPR.

A-F) Two sensograms with kinetic model fits and residuals for the measured interactions. A) Yan wild type B) SY1 C) SY2 D) SY3 E) SY4 F) SY5

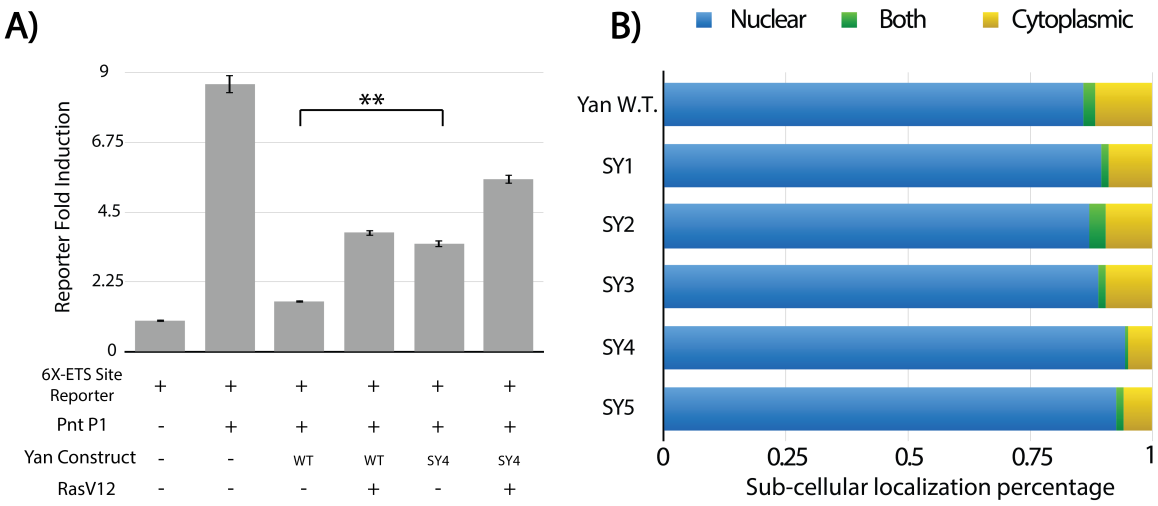


Figure 3.S3: SuperYan mutants show defects in transcriptional repression, and normal recruitment to the nucleus in the absence of signaling.

A) Transcription assay with a multimerized (6X ETS binding site) reporter, with just SY4 alone. Significance was measure by Bonferroni-corrected t-test ($p<.05$) B) Sub-cellular localization of SuperYan mutants in the absence of RasV12. All mutants were similarly localized to the nucleus.

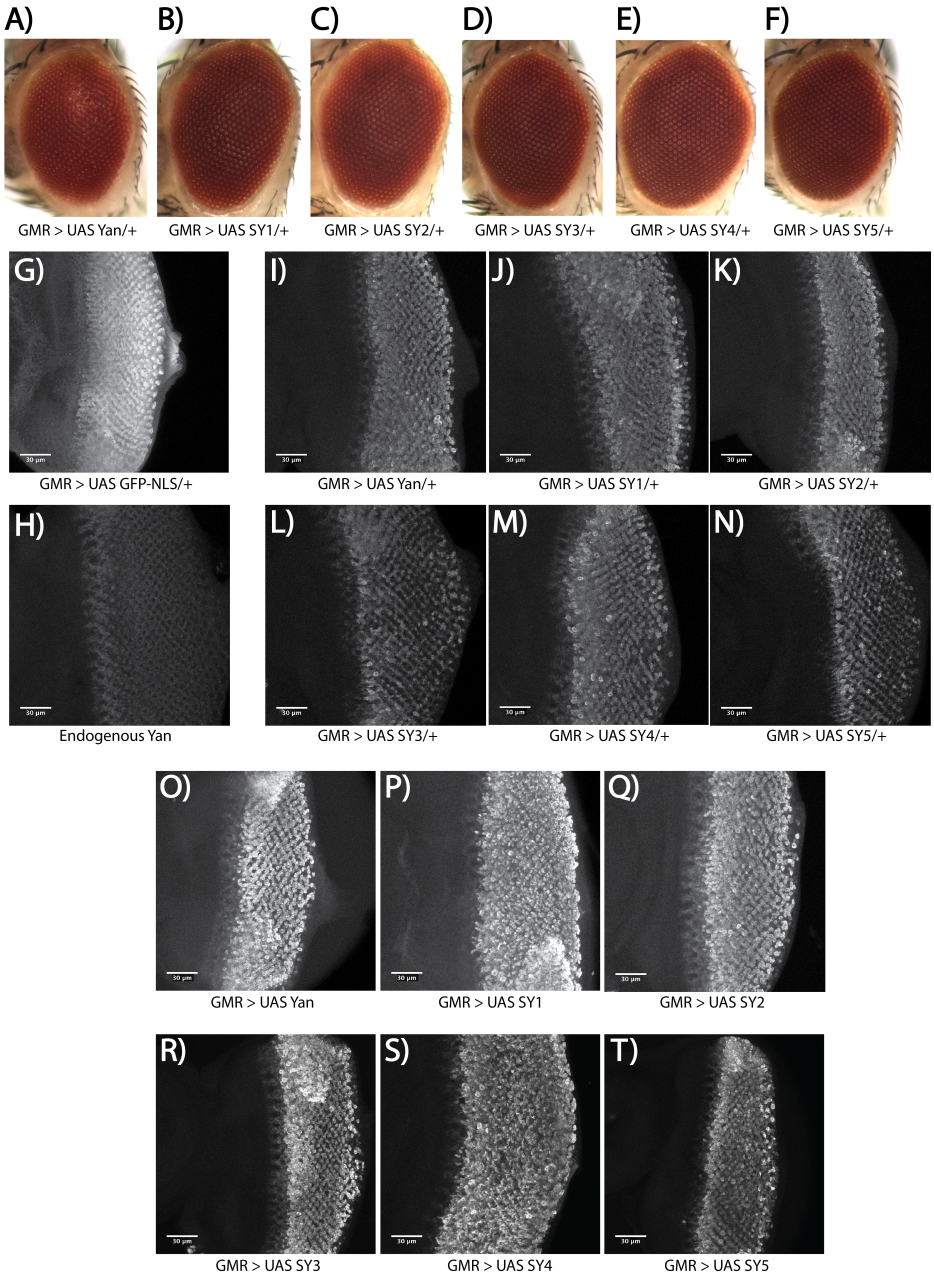


Figure 3.S4: Overexpression of SuperYan constructs shows localization to the nucleus.

A-F) Adult eye phenotype of over expression of Yan and SY1-5 in one copy, driven with GMR-GAL4 at 25C G) Staining of GMR> UAS-GFP-NLS demonstrating pattern of driver H) Endogenous Yan staining in the third larval instar eye disc I-N) Staining of Yan for over expression in one copy of the transgene, for Yan wild type and SY1-5. O-T) Staining of Yan for over expression in two copies of the transgene, for Yan wild type and SY1-5. All confocal microscopy images are maximum intensity projections, taken with the same microscope settings.

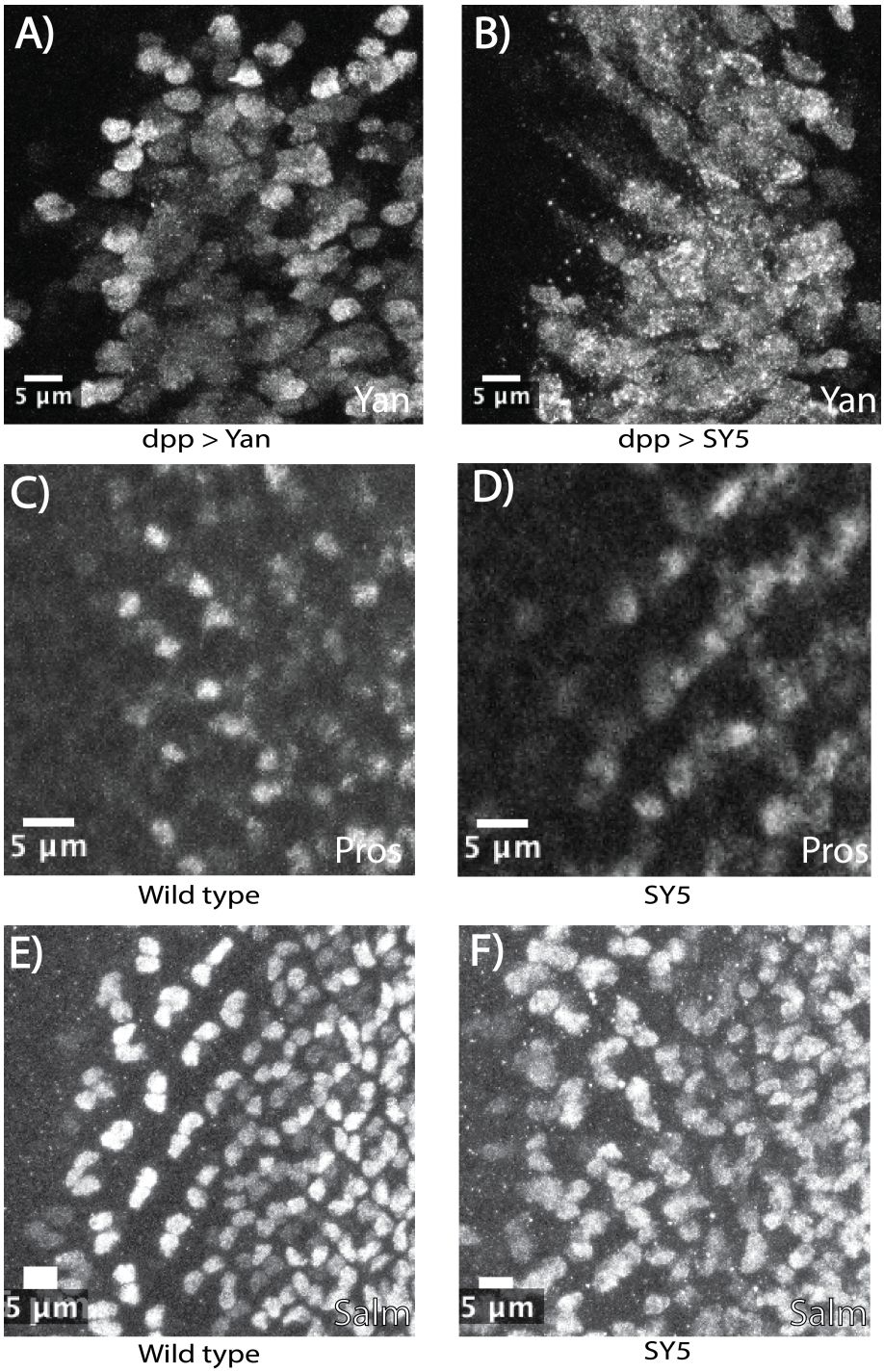


Figure 3.S5: Marker of R3/4 fate, but not R7 fate, is expanded in SY5 genotype; SuperYan puncta are observed in the third instar wing.

A-B) Staining of dpp-GAL4 > UAS-Yan and UAS-SY5 respectively, demonstrating increase in puncta in the wing, a context that lacks endogenous Yan. C-D) Staining of Prospero, demonstrating no expansion of the R7 lineage in SY5 genotype E-F) Staining of Spalt-major (Salm), demonstrating increased numbers of R3/4 cells

Chapter 4: Discussion and Future Directions

C. Matthew Hope

4.1 Strategies for increasing protein-protein interaction affinity in Yan transcriptional regulatory complexes

The SuperYan mutations offer a useful window into studying the effects of increasing the strength of protein-protein interactions in a given transcriptional regulatory complex, and taken together, our results argue that a key consequence of increasing self-association Yan is formation of complexes that are incompatible with proper function. Although increasing self-association revealed a requirement for tuned Yan affinity in complex formation, one major outstanding question is if increased self-association plays a role in transcriptional outcomes for complexes that are already formed. It is possible to envision that increasing the strength of a protein-protein interaction would stabilize a given regulatory complex on DNA once assembled. Mechanistically, this might result in over-recruitment of transcriptional co-effectors and therefore increase the activity of a transcriptional complex. Alternatively, the increased stability of a complex may prevent turnover of the complex during development, and this lack of dynamics may result in normal levels of transcriptional activity but for too long of a time period to properly specify cell fate. The increased affinity of the SuperYan mutations resulted in defects in transcriptional regulatory complex formation in the first place, but I propose testing further roles of increased self-association affinity by designing strategies to utilize the SuperYan mutations while bypassing these defects.

One possibility is exploring weaker polymerization mutants identified in the negGFP native gel screen to look for behaviors that diverge from the stronger, characterized SuperYan alleles. Quantification of the increases in self-association affinity showed that SY1 (the weakest mutant selected for further characterization) increased

self-association roughly 100-fold relative to Yan wild type, indicating that there is a large range of self-association affinities to explore between these two mutants. It is possible that modestly increasing self-association affinity would result in more stable complexes on DNA, but that further increasing self-association affinity drives Yan into puncta. If true, these weaker alleles might represent a gain of function regime that was skipped over by examining the extreme ends of increasing Yan polymerization. The simple test of this hypothesis is to produce UAS transgenes of these mutants in the same manner as the UAS-SY1-5 transgenes, and over express them in the eye using GMR-GAL4. Yan wild type in one copy was shown to produce a very mild rough eye phenotype, and therefore should provide the best comparison to look for enhancement by slightly increasing self-association with these weak mutants. If enhancement is uncovered, then the weak alleles could be similarly characterized by their transcriptional behavior using the transcription assay, and eye phenotype using CRISPR/Cas9 mutagenesis.

It is also possible that increasing self-association affinity modestly also results in driving Yan off of DNA, even for the weakest mutants identified in the screen. Therefore an alternative strategy is to use the monomeric Yan mutations A86D and V105R to create Yan dimers with increased self-association at one interface, thus preventing excessive polymerization beyond dimers while allowing control over protein-protein interaction affinity. Given the fact that wild type dimers are capable of greatly rescuing Yan lethality and eye phenotypes (IR, unpublished observation), this suggests that much of Yan biology may be tuned to utilize smaller order species and therefore it would be reasonable to compare interaction affinity among different dimers. To do this, I propose using the identical mutations explored via pairwise SPR and generating and over-expressing UAS

transgenes, with Yan-A86D and Yan-V105R as the proper comparison. The A86D and V105R mutations are capable of completely monomerizing the SuperYan mutations *in vitro,* therefore I expect that these mutants will not form puncta *in vivo* and will have a greater chance of forming high affinity regulatory complexes on DNA. If this is the case, I expect enhancement of the phenotype observed with over-expression of Yan wild type dimers, which would establish a context for increased self-association resulting in Yan gain of function behavior. Alternatively, it is possible to imagine that the ability to form one high affinity interface would still interfere with Yan localization, even though the effective polymerization distribution is greatly reduced. For example, if Yan import or export can only occur at the level of a Yan monomer, forming high affinity dimers could still result in phenotypes that resemble the SY3-5 alleles.

If it is indeed possible to bypass the loss of function by sequestration behavior of the strongest SuperYan alleles and form stronger transcriptional regulatory complexes on DNA, it will be important to mechanistically distinguish between increased transcriptional activity versus improper complex turnover, as alluded to above. To do so, both the established transcription assay in S2 cells as well as target gene staining in the eye provide facile means of testing increased transcriptional activity, with the expectation that stronger Yan complexes may result in stronger target gene repression and decreased staining of transcriptional targets *in vivo*. Alternatively, a simple test of looking at the turnover of Yan complexes would entail staining for Yan via immunofluoresence in larval eye discs. Due to the fact that the Yan expression pattern is well established (Boisclair Lachance et al. 2014), it is possible that increasing self-association will result in increased Yan staining in the eye progenitor population long after the expression of

Yan would typically be resolved to cone cells, and therefore may be consistent with persistent, hyper-stable complexes. Taken together, these experiment may be able to further delineate roles for self-association in influencing Yan transcriptional activity, and reveal new biology downstream of initial Yan complex formation.

4.2 SuperYan mutant phenotype potentially uncouples Yan regulation of R3/4 and R7 fate

In eye development, EGFR signaling is used iteratively to specify cell fates in the developing retina (Doroquez and Rebay 2006; Freeman 1996) and consequently, Yan and Pnt have multiple functions at different developmental decision points as major signaling effectors of RTK signaling. Consistent with this, the classic phenotype of Yan hypomorphs is a mix of photoreceptor defects(Lai and Rubin 1992), with an increase in both the number of outer photoreceptors and a striking increase in the number of inner R7 photoreceptors. Given that these two populations are recruited at different points during development and have differential signaling requirements (Freeman 1996), it follows that there are at least two separable regulatory decisions that Yan makes during development— one for outer photoreceptors in the R8, R2/5, R3/4 precluster, and a later one for R7. The SuperYan mutants showed an intriguing phenotype of increased outer photoreceptors but a greatly suppressed R7 phenotype (Fig3.4F-H), and further experiments showed that during development, Pros-positive R7 precursors are not expanded, but Salm-positive R3/4 precursors are expanded (Figure3.S5). These results evoke two mechanistic hypotheses to be explored further: 1) That the known role of elevated RTK-signaling in R7 cells (Freeman 1996; Nagaraj and Banerjee 2004) is capable of shifting the SuperYan polymer distribution to rescue function, and 2) That

different regulatory decisions may require different polymerized species of Yan, with the outer photoreceptor decision relying on smaller species and the R7 decision relying on larger species.

In order to explore these hypotheses further, I must first address a caveat. In principle, it is possible that the SuperYan mutations have a novel activity that recruits a new TF or relieves repression from a TF that is not otherwise a target of Yan, resulting in induction of R3/4 fate as opposed to weakening repression on the endogenous Yan target of R3/4 fate. While I cannot rule out this possibility, the behavior of the SuperYan mutants as genetic hypomorphs argues against this ectopic activity hypothesis. Additionally, over expression of the SuperYan mutants did not induce any noticeable change in eye roughness, which might be expected if the activity of the mutants were recruiting novel transcriptional activity. Lastly, the phenotype of the classic Yan hypomorph alleles is a 27% increase in the number of ommatidia with extra outer photoreceptors, which is almost perfectly in line with the phenotype of the SY4-5 mutants (Fig3.4E). Therefore, I interpret the SuperYan phenotype as loss of repressive input into the R3/4 regulatory decision, but no loss of repressive input into the R7 regulatory decision. In essence, the phenotype is evocative of uncoupling these two decisions, with repressive input into the R7 cell fate choice being significantly rescued compared to the simple Yan^{1} hypomorph (Lai and Rubin 1992).

4.3 Higher MAPK signaling activity in R7 photoreceptors may rescue SuperYan phenotype

One explanation as to why R7 repressive input is rescued in the SuperYan mutants has to do with the level of RTK signaling presumptive R7 cells receive, compared to other

photoreceptor subtypes. In order to specify R7 fate, the input of both EGFR and Sevenless RTKs are needed (Freeman 1996), resulting in elevated signaling through the Ras/MAPK signaling cassette. If phosphorylation by MAPK of the SuperYan mutants is elevated in these cells, it is possible that the act of phosphorylation is capable of disrupting the polymerization of Yan, resulting in a shift of polymers back towards a more wild type distribution and rescuing the proper regulation of R7. The cells of the pre-cluster, including R3/4, experience lower EGFR signaling without Sevenless signaling, and therefore the SuperYan mutants would remain sequestered within puncta and unable to regulate the R3/4 decision. This explanation requires that phosphorylation outside of the SAM domain by MAPK is able to affect the affinity of SAM-mediated polymerization, for which there is no currently evidence. However, as I note below, the SuperYan mutants display an uncoupled relationship between nuclear localization and transcriptional repression, therefore it remains a possibility that MAPK phosphorylation can actively depolymerize Yan, or at the least, prevent further Yan polymerization.

In order to test this hypothesis I propose several experiments. First, if the extra RTK signaling activity in R7 precursors is rescuing the polymerization-based defect of the SuperYan mutants, then decreasing RTK signaling in those cells should have the effect of increasing SuperYan polymerization, resulting in less repression, and an increase in the number of R7 cells specified. This would be a paradoxical (and striking) phenotype because the classic effect of decreasing RTK signaling is to abolish R7 fate. Due to the fact that RTK signaling has such pervasive roles in photoreceptor recruitment, to answer this question requires precision in perturbing just R7 cells without compromising the specification of other photoreceptors in the eye and precluding

evaluation of the phenotype. Additionally, I believe that it is likely that strongly decreasing RTK signaling will result in the classic phenotype of abolishing R7 fate, therefore one would use relatively modest perturbations to see if there is a regime where R7 photoreceptors increase in the SuperYan mutant background. The first would be to try modification of the *sevenless* background, which will decrease signaling specifically to the presumptive R7 cells, and evaluating the phenotype by histological sectioning. Additionally, the *sev*-GAL4 driver is sufficiently specific to drive RNAi transgenes against RTK components, and controllable by temperature, thus providing a second method of slightly decreasing signaling. Lastly, I propose using heterozygosity for RTK components, specifically *sev, boss*, *DER,* and *rl* (MAPK) to modify the adult photoreceptor phenotype of SuperYan mutants by histological sectioning, with the caveat that other photoreceptor cell types will be affected.

In addition to genetic perturbations, it would also be important to explore the biochemical aspect of this hypothesis, which is that polymerization can be attenuated by MAPK phosphorylation of full length Yan. It should be noted that the SAM domain of Pnt contains a flexible helix N-terminal to the SAM that is capable of being phosphorylated by MAPK, and increases its flexibility upon phosphorylation (Lau et al. 2012). This same region is found in the human homolog of Pnt, ETS1, and phosphorylation by MAPK potentiates binding of the co-activator CBP (Nelson et al. 2010). Therefore, there is precedent for regions appended to SAM domains of ETS TFs making a difference for transcriptional regulation with respect to MAPK phosphorylation. Furthermore, the most potent single MAPK phosphorylation site in Yan is immediately N-terminal to the SAM domain as well (Rebay and Rubin 1995). One

intriguing possibility is that phosphorylation may potentiate an interaction with a helix outside of the SAM domain and the canonical SAM interface, thereby decreasing Yan polymerization potential by occluding the polymerization interface. Given the success of heteronuclear single quantum coherence (HSQC) NMR in visualizing MAPK phosphorylation of the dynamic helix appended to the Pnt SAM domain (Lau et al. 2012), this might be a useful technique to dissect if a similar MAPK-dependent phenomena is at play in Yan's SAM domain and adjacent regions. I propose purifying Yan and comparing in vitro MAPK phosphorylated Yan with wild type to try to observe if conformational changes are occurring upon phosphorylation. Since full length Yan is likely to be too large for an HSQC NMR experiment, an alternative is to attempt Dynamic Light Scattering (DLS) with Yan in vitro, to look for shifts to smaller populations upon MAPK phosphorylation, with the caveat that adding a negGFP tag to full length Yan may be required to solubilize it for experiments in vitro. If hints of phosphorylation-dependent changes in polymerization affinity are observed, pairwise Surface Plasmon Resonance (SPR) spectroscopy could be employed with full length Yan or large fragments including the SAM domain to try to make a measurement of differences in affinity.

4.4 SuperYan phenotype may reveal polymerization sensitive enhancers of photoreceptor fate

An alternative hypothesis to be explored is that there are different requirements for extent of polymerization for different genes that Yan regulates. This hypothesis presupposes that there are regulatory elements of a gene that controls the R3/4 decision of Yan that requires lower-order species of Yan and regulatory elements controlling the R7 decision that require higher-order species of Yan. To illustrate this, I will contrast the

hypomorphic Yan^1 allele with the SuperYan mutants. In the case of the Yan^1 allele, a decrease in the amount of Yan protein results in preferential loss of higher-polymerized Yan species, simply because more Yan molecules are required to make larger polymers. This results in potent derepression of regulatory elements controlling R7 fate, and a large increase in supernumerary R7 photoreceptors. By contrast, the SuperYan alleles shift the distribution of Yan polymers toward a size that would be needed to properly regulate R7 fate and away from a size that would be need to regulate R3/4 fate, thus preferential favoring proper regulation of inner photoreceptor recruitment, while disfavoring outer photoreceptor regulation. An intriguing prediction of this model would be that alleles that restrict Yan to monomers or dimers, if expressed at proper nuclear concentrations, may rescue the regulation of the R3/4 fate decision, but result in extra R7 photoreceptors, as opposed to hypomorphs which show an increase in both.

The key confirmation of this hypothesis would be identification of target genes (and regulatory regions) of Yan with differential requirements for polymerization. To that end, I propose a candidate approach based on the known biology of R7 and R3/4 fate. Prospero is a previously identified Yan target gene, that was shown to not be expanded in the SuperYan mutants (Figure3.S5). Additionally, Prospero has a well-characterized enhancer that is known to be responsive to Yan and Pnt, among other TFs (Xu et al. 2000; Hayashi et al. 2008). One simple test would be to compare the transcriptional repressive ability of Yan and the SuperYan mutants on the *prospero* enhancer in a transcription assay, and with previously generated LacZ reporters of the enhancer *in vivo* (Xu et al. 2000), with the prediction that baseline repression from SuperYan mutants should be better than at other target genes, such as *king-tubby*. If true, this may indicate

that the structure of the *prospero* enhancer is wired for regulation by higher-order Yan polymers.

Importantly, the Yan target of the R3/4 fate decision would also have to be elucidated. One intriguing candidate is *salm* which is necessary for R3/4 fate early in specification (Mollereau et al. 2001), was expanded in SuperYan mutants(Fig3.S5), and has several strong peaks of Yan binding via chromatin immunoprecipitation and sequencing (ChIP-seq) in the eye (Webber and Rebay, unpublished). I propose cloning potential enhancer regions of *salm* based on Yan binding peaks, and assessing their response in both transcription assays and reporter gene constructs, with the expectation that *salm* expression will be significantly de-repressed in SuperYan mutants. A caveat to this hypothesis is that *salm* has later roles in specifying R7 and R8 identity during pupal stages (Domingos 2004) and must be repressed by Seven-up in R3/4 to allow their specification to proceed. Therefore, an important correlate of this hypothesis is that Seven-up is sufficiently robust to convert an expanded Salm-positive population to an R3/4, and would suggest that loss of *seven-up* should convert the extra outers in a SuperYan background toward R7 and R8 fate. If the targeted approach to uncover polymerization-sensitive enhancers does not yield results, I propose using ChIP-seq to examine Yan bound regions in SuperYan mutants. The specificity of the SuperYan phenotype suggests that cell fate determination will not be globally disrupted, and changes can be compared to wild type Yan to find specific regions of peak loss in SuperYan mutants. Furthermore, the gradient of affinity in SuperYan mutants can be employed to tune the loss of Yan binding to identify the most sensitive regions.

Ultimately, if regulatory regions of Yan that require higher- and lower-order Yan polymers can be identified, a long term project will be to assess the architectural differences in TF binding sites that determine the differential requirements for polymerization. Given the results of the modeling in Chapter 2 in revealing a role for clustering of TF binding sites in increasing Yan fractional occupancy, one hypothesis is that use of tandem ETS sites may drive the requirement for polymerization of higher-order or lower-order species. Relatedly, heterotypic interactions between Yan and other TFs at these enhancers may enforce a requirement for one class of Yan polymers versus another. A simple model would be that tandem ETS sites greatly increase the amount of Yan recruited to an element and drive homotypic, higher-order polymerization, while heterotypic interactions between Yan and another TF would be non-polymerizing by definition, and enforce a requirement for small species like Yan monomers. Therefore, assessing the TF binding site composition of these two classes of enhancers may reveal differences in regulation. If no significant differences are found at the level of TF binding sites, it is also possible that the differential requirement for size of polymers enters at the level of differential co-factor recruitment, with one set of co-repressors recognizing smaller Yan regulatory complexes and another set recognizing larger regulatory complexes. Given that many Yan-interacting co-repressors are already known (Zhang J "Regulation of gene expression by the *Drosophila* ETS family transcriptional repressor Yan" PhD diss, University of Chicago, Chicago, IL (2011)) a simple test would be to use genetics (via heterozygosity or RNAi transgenes) to modify the expression of reporters *in vivo*, and confirm recruitment of co-repressors via chromatin immunoprecipitation. It is tempting to speculate that the TLE-family co-repressor Groucho may play a role as a

regulator of higher-polymerization enhancers, given Groucho's extensive recruitment to highly polymerized Yan bound regions (Webber and Rebay, unpublished), and its propensity to also form oligomers (Turki-Judeh and Courey 2012), therefore I propose examining Groucho as a candidate for polymerization-dependent recruitment.

4.5 Role for SuperYan mutants in affecting Wingless signaling

In addition to differentiating between the roles of Yan in outer photoreceptor and inner photoreceptor recruitment, the SuperYan mutants display occasional severe defects in establishing the eye field (Hope and Rebay, unpublished), which may be consistent with roles of Yan regulation at other points of eye development. Loss of the eye field is observed consistently in a population of SuperYan mutants, and frequently affects one eye field but not another in a given animal. I hypothesize that this regulation has to do with Yan's known inputs into Wingless (Wg) signaling via association with Armadillo (the *Drosophila* homolog of beta-catenin) (Olson et al. 2011), and the loss of the eye field in these rare mutants arrises from aberrant Wg-signaling activity (Baonza and Freeman 2002). Mechanistically, I envision this occurring by the increase in amount of Yan protein in SuperYan mutants binding to and protecting Armadillo from degradation, resulting in increased signaling. To test this hypothesis, I propose trying to modify the occurrence of this rare phenotype by decreasing negative regulators of Wg-signaling in the eye, specifically *axin* (Hamada 1999) which is part of the core machinery that constitutively degrades Armadillo in the absence of Wg-signaling. I predict that heterozygosity for axin may increase the incidence of loss of eye field in SuperYan mutants, and that weak knockdown of axin with RNAi transgenes with different tissue

specific drivers may result in establishing a broad temporal requirement for this phenomenon.

4.6 Evidence for a dual role of MAPK phosphorylation in down-regulating Yan

One intriguing outcome of the assessment of the sub-cellular localization of SuperYan mutants was the observation that SuperYan mutants have proportionally stronger localization to the nucleus in the presence of RasV12, versus Yan wild type (Fig 3.3B). This result is coupled with the fact that the SuperYan mutants appear to be as Ras-responsive as Yan wild type in terms of their transcriptional activity (Figure 3.3A and Figure 3.S3A), and thus a paradox in that stronger Yan localization to the nucleus does not lead to stronger Yan transcriptional activity. While it appears clear that this is partly due to the fact that SuperYan mutants form puncta that are incompatible with transcriptional repression of target genes (Figure 3.4I-L) it is interesting to compare these results to other perturbations that affect the nuclear localization of Yan. For example, Yan^{Act} has eight MAPK phosphorylation sites mutated, cannot be exported from the nucleus, and is a potent repressor of cell fate specification leading to strong developmental defects (Figure 3.3J). Therefore, if both the SuperYan mutants and Yan^{Act} are more localized to nucleus, how do they result in different outcomes with respect to transcription?

An intriguing possibility is that MAPK phosphorylation of Yan may serve multiple roles in the process of dismantling Yan's transcriptional repression. For example, phosphorylation of Yan may preclude it from interacting with its partner co-repressors or alter its DNA binding affinity, therefore evicting it from target regions and into the nucleoplasm. As a second step, recognition of MAPK-phosphorylated Yan by the

Crm1-mediated export machinery (Tootle et al. 2003)then further results in down regulation of Yan via nuclear export and degradation. Therefore, Yan^{Act}, by virtue of being unable to be phosphorylated, may retain its ability to bind to Yan regulatory regions, but the SuperYan mutants uncouple MAPK regulation by being down regulated by the first process, but are too large to be exported by the second process resulting in their nuclear localization.

To test this hypothesis, I propose combining the Yan^{Act} and SuperYan mutations in one construct. If MAPK phosphorylation inhibits complex formation on DNA, the lack of phosphorylation could result in more SuperYan associated with the DNA, therefore rescuing the baseline transcriptional repression of the SuperYan mutants and restoring them to the level of Yan^{Act}. Alternatively, excessive polymerization could result in Yan transcriptional regulatory complexes not being formed in the first place, in which case the SuperYan mutations may be dominant over the Yan^{Act} mutations and further characterization would be required. In addition to disrupting MAPK phosphorylation sites, I propose making phosphomimetic mutations of these sites in the background of the SuperYan mutations. I predict that these mutations may potentially exacerbate the baseline repression defect of SuperYan while maintaining nuclear localization, which would argue that MAPK phosphorylation is playing a role in down regulating Yan complex formation. In these two sets of experiments, transcriptional function could be quickly assessed in both luciferase transcription assay and UAS over-expression transgenes, with the previous results for both the SuperYan mutations and Yan^{Act} serving as comparison points.

In addition to genetics, a biochemical approach might be useful in uncovering a role for MAPK phosphorylation in reducing Yan association to either DNA or partner components of Yan transcriptional regulatory complexes. To test this, I propose trying to assess a role for MAPK phosphorylation in direct binding of Yan to DNA in a electrophoretic mobility shift assay (EMSA) with *in vitro* MAPK-phosphorylated Yan, with the expectation that phosphorylation might decrease Yan's DNA binding affinity.

4.7 Expanding the model of Yan fractional occupancy

In an effort to better understand the occupancy of Yan at transcriptional regulatory elements, and to elucidate the relationship between polymerization and Yan DNA binding, we established a simple model of Yan fractional occupancy at equilibrium (Chapter 2). Intriguingly, our results showed that Yan has an occupancy profile consistent with binding to ETS sites and spreading into non-specific sites on DNA. In addition, the modeling efforts demonstrated roles for both clustering of TF binding sites and extent of polymerization in shaping the Yan occupancy profile. Revisiting the model to allow polymerization off of DNA elucidated a role for high affinity polymerization in decreasing Yan occupancy, and highlighted the importance of tuned polymerization affinity in achieving proper Yan fractional occupancy. A longterm goal of this effort will be to further refine the model to eventually aid in prediction of transcriptional activity from regulatory elements responsive to Yan. Therefore some of the changes necessary to achieve this goal will be discussed below.

Due to the fact that enhancers rarely operate under the control of single TFs, but instead rely on multiple activators and repressors to achieve robust and spatiotemporally specific gene expression, a major future direction of the modeling effort will be to

incorporate competition and binding from other TFs beside Yan. An obvious first candidate for this effort would be Pnt, as co-occupancy and competition between these two ETS family factors has already been established (Graham et al. 2010). Inherent in making proper assessments of Yan and Pnt occupancy at a given enhancer would be high quality position weight matrices (PWMs) of the two factors, and careful measurement of their specific and non-specific DNA binding affinites, via EMSA. Importantly, existing PWMs from FlyFactorSurvey (Zhu et al. 2011) show that Yan and Pnt share similar DNA binding motifs, but with important differences in specificity outside the core GGAA motif. Therefore, transitioning to a nucleotide-based model of binding as opposed to a site-based model will be an important next step in modeling Yan occupancy from sequence.

Another key area to address in order to properly parameterize the model will be to determine the range of function Yan nuclear concentration *in vivo*. Like most TF binding at equilibrium, the behavior of Yan depends strongly on its nuclear concentration, which is especially important due to the tight regulation of Yan import and export. Initial modeling used a broad range of reasonable concentrations for TFs in general, but more narrowly measuring the concentration of Yan will determine if the model captures behavior that is physiologically relevant. To this end, I propose a combination of methods to measuring concentration *in vivo* based on tagging Yan with fluorescent protein tags. Using step-wise photobleaching, it is possible to measure the number of fluorescently-tagged molecules of a protein of interest and correlate that to the absolute fluorescence intensity of a sample (Verdaasdonk et al. 2014). This technique works best for small numbers of molecules, but the advantage is that it can be performed in living tissues

.Therefore, I propose using a relatively weak, cell-type specific driver in Drosophila embryos to measure small numbers of molecules via photobleaching, given the fact that protocols for live imaging are now well established, and correlate that to total fluorescence of Yan tagged in the same manner at its endogenous locus. Furthermore, a key question is what is the size of the polymer distribution *in vivo*? To answer this, I propose a similar technique of Fluoresence Correlation Spectroscopy (FCS) (Verdaasdonk et al. 2014), using a photoactivable fluorophore to stimulate small numbers of Yan molecules and accurately measure their binding *in vivo*. Reasoning that the optically active Yan molecules should be able to interface with optically inactive molecules, FCS may be able to provide a quantitative picture of the dynamics of Yan in living cells, which could then be correlated to a size distribution of polymer.

4.8 Modeling of Yan regulatory elements to predict gene expression activity

While addressing the above issues will provide a more detailed picture of Yan occupancy in the current modeling framework, a long term goal of modeling Yan occupancy will be prediction of the output of regulatory elements regulated by Yan. This is a very difficult problem, since the output of any given regulatory element will typically depend on the action of many TFs including Yan, and thus requires knowledge of the concentration and binding affinities of all the TFs that serve as inputs into the model. However, some groups have made progress in predicting regulatory element activity from DNA sequence and quantitative TF binding information(Janssens et al. 2006; Kim et al. 2013; Barr and Reinitz 2017). One strong candidate element for modeling Yan regulation in an actual *in vivo* setting is the Muscle Heart Enhancer (MHE) (Halfon et al. 2000). The MHE has the advantage that many TF binding sites have been mapped, for Yan and Pnt

as well as other TFs that are known to be important for MHE activation, including dTCF, Mothers against dpp (Mad), Twist, and Tinman. Many of these TFs have PWMs generated by multiple methods (Nitta et al. 2015; Zhu et al. 2011), thus adding another important component to prediction of MHE expression. However, I propose three areas that likely will have to be well-established before modeling expression from a complex element like the MHE can be successful.

First, identifying all the possible TFs that may operate on the MHE will be necessary, so that inputs to the model are not missing. This task is made easier by the fact that there are RNA-seq data sets during this stage of embryogenesis, which narrows the field of candidate TFs to be tested to those that are expressed in the mesodermal precursors that give rise to the Eve-positive preclusters. Secondly, interactions between a given candidate TF and output from the MHE can be tested by using a MHE-GFP reporter, which drives robust expression, and can be easily enhanced or suppressed. Thus heterozygosity or RNAi transgenes could be employed to further validate an interaction of a given TF at the MHE. And lastly, binding at the MHE could be verified via chromatin immunoprecipitation or *in vitro* binding to an MHE EMSA probe. While it possible that all TF inputs to the MHE have been previously identified, searching for additional inputs will preclude developing an incomplete model that may fail to capture the behavior of the MHE.

A second important project will be to identify the relative concentrations of TFs at single cell resolution within the embryo. Fortunately, an atlas of Yan and Pnt gene expression already exists at this resolution(Boisclair Lachance et al. 2014), however it is likely that concentrations will have to be measured for other TFs that serve as inputs into

the model. I propose facilitating this process by tagging TFs using a similar small epitope tag and then staining for various TFs via immunofluoresence. Although this procedure will amplify the signal in a non-linear fashion, the key measurement is relative concentration of a given TF compared to others, which is simplified by using the same staining procedure. Additionally, other groups have shown that deriving quantitative single-cell measurements of TF concentration via immunofluoresence is feasible in the *Drosophila* embryo and produces data that can be used for producing high quality models of gene expression (Pisarev et al. 2009; Surkova et al. 2008).

Lastly, I predict that in a similar fashion to other enhancers, that cooperative protein-protein interactions among TFs will be key to shaping the output of the MHE. Thus although there are no known direct interactions between Yan and other TFs, it reasonable to assume that they may exist, especially given the presence of other TF binding motifs in Yan ChIP datasets (Webber et al. 2013). With just a handful of potential TFs acting at the MHE, I propose probing for protein-protein interactions using co-immunoprecipitation experiments, and assessing direct binding using *in vitro* binding assays. Given that one of the main functions of enhancers in development is to serve as platforms for the association of transcriptional regulatory complexes, I predict that interactions between Yan and other TFs are likely to be both important and predictive in terms of understanding the output of gene regulation during development.

4.7 References

Baonza, Antonio, and Matthew Freeman. 2002. “Control of Drosophila Eye Specification by Wingless Signalling.” *Development (Cambridge, England)* 129 (23):5313–22.

Barr, Kenneth A., and John Reinitz. 2017. “A Sequence Level Model of an Intact Locus Predicts the Location and Function of Nonadditive Enhancers.” *PLoS ONE* 12 (7).

Boisclair Lachance, Jean François, Nicolás Peláez, Justin J. Cassidy, Jemma L. Webber, Ilaria Rebay, and Richard W. Carthew. 2014. "A Comparative Study of Pointed and Yan Expression Reveals New Complexity to the Transcriptional Networks Downstream of Receptor Tyrosine Kinase Signaling." *Developmental Biology* 385 (2):263–78.

Domingos, P. M. 2004. "Spalt Transcription Factors Are Required for R3/R4 Specification and Establishment of Planar Cell Polarity in the Drosophila Eye." *Development* 131 (22):5695–5702.

Doroquez, David B., and Ilaria Rebay. 2006. "Signal Integration during Development: Mechanisms of EGFR and Notch Pathway Function and Cross-Talk." *Critical Reviews in Biochemistry and Molecular Biology*.

Freeman, Matthew. 1996. "Reiterative Use of the EGF Receptor Triggers Differentiation of All Cell Types in the Drosophila Eye." *Cell* 87 (4):651–60.

Graham, T. G. W., S. M. A. Tabei, A. R. Dinner, and I. Rebay. 2010. "Modeling Bistable Cell-Fate Choices in the Drosophila Eye: Qualitative and Quantitative Perspectives." *Development* 137 (14):2265–78.

Halfon, Marc S, Ana Carmena, Stephen Gisselbrecht, Charles M Sackerson, F Jiménez, Mary K Baylies, and Alan M Michelson. 2000. "Ras Pathway Specificity Is Determined by the Integration of Multiple Signal-Activated and Tissue-Restricted Transcription Factors." *Cell* 103 (1):63–74.

Hamada, F. 1999. "Negative Regulation of Wingless Signaling by D-Axin, a Drosophila Homolog of Axin." *Science* 283 (5408):1739–42.

Hayashi, Takashi, Chunyan Xu, and Richard W Carthew. 2008. "Cell-Type-Specific Transcription of Prospero Is Controlled by Combinatorial Signaling in the Drosophila Eye." *Development (Cambridge, England)* 135 (16):2787–96.

Janssens, Hilde, Shuling Hou, Johannes Jaeger, Ah-Ram Kim, Ekaterina Myasnikova, David Sharp, and John Reinitz. 2006. "Quantitative and Predictive Model of Transcriptional Control of the Drosophila Melanogaster Even Skipped Gene." *Nature Genetics* 38 (10):1159–65.

Kim, Ah Ram, Carlos Martinez, John Ionides, Alexandre F. Ramos, Michael Z. Ludwig, Nobuo Ogawa, David H. Sharp, and John Reinitz. 2013. "Rearrangements of 2.5 Kilobases of Noncoding DNA from the Drosophila Even-Skipped Locus Define Predictive Rules of Genomic Cis-Regulatory Logic." *PLoS Genetics* 9 (2).

Lai, Zhi Chun, and Gerald M. Rubin. 1992. "Negative Control of Photoreceptor Development in Drosophila by the Product of the Yan Gene, an ETS Domain Protein." *Cell* 70 (4):609–20.

Lau, Desmond K.W., Mark Okon, and Lawrence P. McIntosh. 2012. “The PNT Domain from Drosophila Pointed-P2 Contains a Dynamic N-Terminal Helix Preceded by a Disordered Phosphoacceptor Sequence.” *Protein Science* 21 (11):1716–25.

Mollereau, Bertrand, Maria Dominguez, Rebecca Webel, Nansi Jo Colley, Benison Keung, Jose F de Celis, and Claude Desplan. 2001. “Two-Step Process for Photoreceptor Formation in Drosophila.” *Nature* 412 (6850):911–13.

Nagaraj, Raghavendra, and Utpal Banerjee. 2004. “The Little R Cell That Could.” *International Journal of Developmental Biology*.

Nelson, M. L., H.-S. Kang, G. M. Lee, A. G. Blaszczak, D. K. W. Lau, L. P. McIntosh, and B. J. Graves. 2010. “Ras Signaling Requires Dynamic Properties of Ets1 for Phosphorylation-Enhanced Binding to Coactivator CBP.” *Proceedings of the National Academy of Sciences* 107 (22):10026–31.

Nitta, Kazuhiro R., Arttu Jolma, Yimeng Yin, Ekaterina Morgunova, Teemu Kivioja, Junaid Akhtar, Korneel Hens, et al. 2015. “Conservation of Transcription Factor Binding Specificities across 600 Million Years of Bilateria Evolution.” *eLife* 2015 (4).

Olson, Emily R, Raluca Pancratov, Sujash S Chatterjee, Binita Changkakoty, Zeeshan Pervaiz, and Ramanuj DasGupta. 2011. “Yan, an ETS-Domain Transcription Factor, Negatively Modulates the Wingless Pathway in the Drosophila Eye.” *EMBO Reports* 12 (10):1047–54.

Pisarev, Andrei, Ekaterina Poustelnikova, Maria Samsonova, and John Reinitz. 2009. “FlyEx, the Quantitative Atlas on Segmentation Gene Expression at Cellular Resolution.” *Nucleic Acids Research* 37 (SUPPL. 1).

Rebay, Ilaria, and Gerald M. Rubin. 1995. “Yan Functions as a General Inhibitor of Differentiation and Is Negatively Regulated by Activation of the Ras1/MAPK Pathway.” *Cell* 81 (6):857–66.

Surkova, Svetlana, Ekaterina Myasnikova, Hilde Janssens, Konstantin N. Kozlov, Anastasia A. Samsonova, John Reinitz, and Maria Samsonova. 2008. “Pipeline for Acquisition of Quantitative Data on Segmentation Gene Expression from Confocal Images.” *Fly*.

Tootle, Tina L, Philina S Lee, and Ilaria Rebay. 2003. “CRM1-Mediated Nuclear Export and Regulated Activity of the Receptor Tyrosine Kinase Antagonist YAN Require Specific Interactions with MAE.” *Development (Cambridge, England)* 130 (5):845–57.

Turki-Judeh, Wiam, and Albert J. Courey. 2012. “Groucho. A Corepressor with

Instructive Roles in Development." *Current Topics in Developmental Biology* 98:65–96.

Verdaasdonk, Jolien Suzanne, Josh Lawrimore, and Kerry Bloom. 2014. "Determining Absolute Protein Numbers by Quantitative Fluorescence Microscopy." *Methods in Cell Biology* 123:347–65.

Webber, Jemma L., Jie Zhang, Lauren Cote, Pavithra Vivekanand, Xiaochun Ni, Jie Zhou, Nicolas Nègre, Richard W. Carthew, Kevin P. White, and Ilaria Rebay. 2013. "The Relationship between Long-Range Chromatin Occupancy and Polymerization of the Drosophila Ets Family Transcriptional Repressor Yan." *Genetics* 193 (2):633–49.

Xu, Chunyan, Rachele C. Kauffmann, Jianjun Zhang, Susan Kladny, and Richard W. Carthew. 2000. "Overlapping Activators and Repressors Delimit Transcriptional Response to Receptor Tyrosine Kinase Signals in the Drosophila Eye." *Cell* 103 (1):87–97.

Zhu, Lihua Julie, Ryan G. Christensen, Majid Kazemian, Christopher J. Hull, Metewo Selase Enuameh, Matthew D. Basciotta, Jessie A. Brasefield, et al. 2011. "FlyFactorSurvey: A Database of Drosophila Transcription Factor Binding Specificities Determined Using the Bacterial One-Hybrid System." *Nucleic Acids Research* 39 (SUPPL. 1).

Appendix I: Polymerization and engineering of the Pnt SAM domain

C. Matthew Hope and Ilaria Rebay

A1.1 Results and Discussion

Yan and Pointed (Pnt) are both RTK-responsive transcription factors, and are comprised of very similar protein domain architectures, with each containing a SAM and an ETS domain. While polymerization of Yan's SAM domain has been shown to be important for full Yan function (Zhang et al. 2010; Webber et al. 2013), the role of polymerization in Pnt has not been well explored. Specifically, it has not been carefully assessed if Pnt is capable of polymerizing. If Pnt is capable of polymerization, then it would be important to understand how polymerization affects its transcriptional function, especially in relation to competition or cooperation with Yan in forming transcriptional regulatory complexes. If Pnt is incapable of polymerization, then it would be equally important to understand biochemically what prevents polymerization and if there is a deleterious impact on fitness conferred by Pnt polymerization. Given the fact that relatively few amino acid changes were capable of wildly altering Yan's self-association affinity, it is possible that a small number of residues in Pnt may be responsible for tuning its self-association affinity, in either a pro- or anti-polymerization fashion.

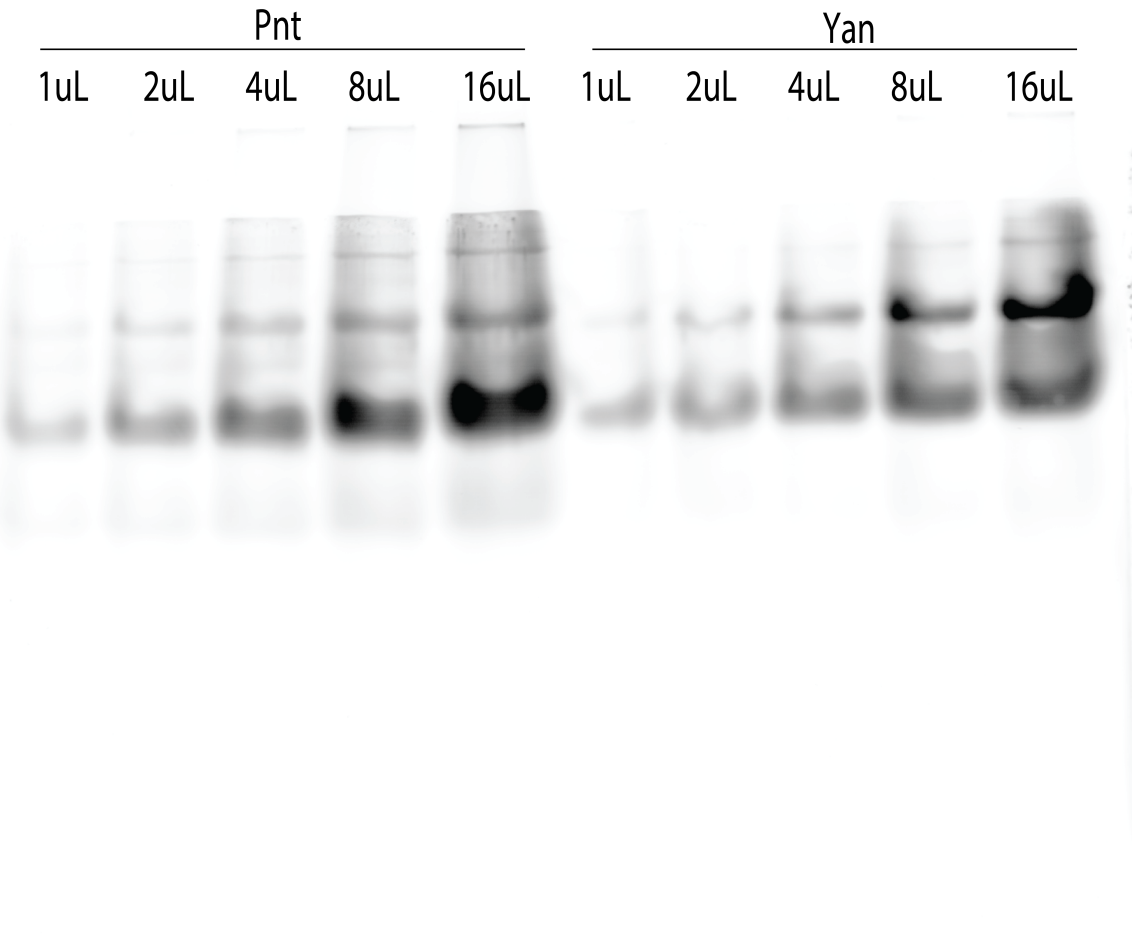


Figure A1.1: Comparison of polymerization abilities of Pnt SAM and Yan SAM.

Pnt SAM and Yan SAM samples were normalized to the fluorescence of Yan SAM, and increasing volumes of each (from 1-16uL) were run on a 12.5% native gel. Yan shows a strong increase in the proportion of dimers relative to Pnt, although Pnt displays monomers, dimers, and higher-order species.

Therefore, we set out to assess polymerization of Pnt *in vitro* to determine whether or not it is capable of self-associating, by using the negGFP Native Gel Assay to assess Pnt polymerization relative to Yan. Both Yan and Pnt SAM domains were purified *in vitro* and assayed in a concentration-matched fashion to permit rough comparison in their interaction affinities (Figure A1.1). We observed that Yan runs as monomers and dimers, with a strong predominate band at the dimer level; in a similar fashion, we observe monomers and dimers for Pnt, with a weaker band at the dimer level. At high

concentrations, both proteins were capable of producing higher mobility bands beyond dimers. Taken together, we conclude that Pnt can form higher order species beyond monomers *in vitro* and does so with a weaker affinity than Yan wild type. Formally, it is possible that the bands we observe are a function of high concentrations of purified protein producing non-specific higher order polymers, as opposed to polymerization through the canonical End Helix (EH) - Mid Loop (ML) surface. Therefore, two key outstanding experiments to verify if Pnt polymerization is real will be to use classic SAM monomerizing mutation to control polymerization in the expected fashion, and ultimately to use electron microscopy to visualize the formation of Pnt filaments. Taken together, we interpret Pnt self-association *in vitro* as the starting point for further investigation into Pnt self-association *in vivo*.

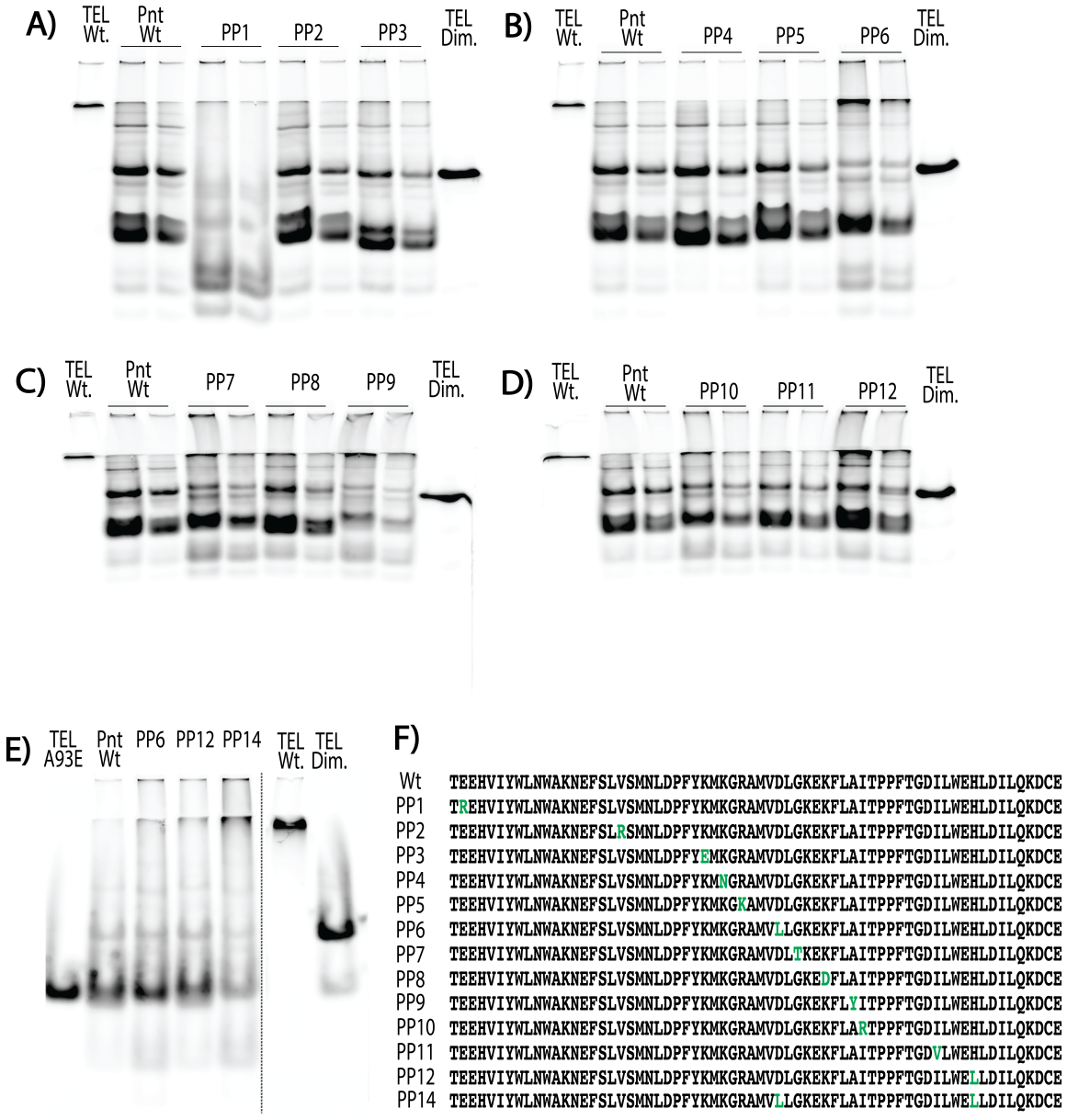


Figure A1.2: Two point mutations are capable of shifting Pnt polymerization *in vitro*.

A-D) negGFP native gels of SAM domains, run on 12.5% native gels. For each Pnt construct, two concentrations are shown (10uL and 5uL) although concentrations are not normalized to GFP fluorescence across constructs. TEL wild type, and mixed TEL dimers in equal proportion are included to mark higher order polymers and dimers, respectively. A) Pnt, PP1, PP2, and PP3 B) Pnt, PP4, PP5, PP6 C) Pnt, PP7, PP8, PP9 D) Pnt, PP10, PP11, PP12 E) negGFP native gel comparing Pnt to PP6, PP12, and a construct with both mutations (PP14). Intervening lanes have been removed for clarity. F) Diagram of primary sequence and mutations of all PolyPnt constructs.

Taking a cue from our strategy of increasing Yan affinity by mutating toward TEL SAM sequences, we next sought to ask the question if the Pnt SAM domain is also capable of adopting a wide range of self-association affinities. Restricting our search to mutations at the interface between TEL SAM molecules in the TEL crystal structure (Kim et al. 2001), we identified 12 residues that are divergent between Pnt and TEL at these positions (see Figure A1.2F). We then made all possible single mutations and ran a negGFP Native Gel Assay to screen for mutations that increase Pnt polymerization. Figures A1.2A-D show the results of the screen, for two concentrations of each mutant, numbered PolyPnt (PP) 1-12. We selected PP6 and PP12 for further characterization because they showed an increase in higher order species (A1.2B Lane 8 and A1.2D Lane 8 respectively). We then compared polymerization of these mutations individually and in combination (PP14) on the same gel in a concentration matched fashion (Figure A1.2E), demonstrating that these two mutations together produce a strong shift in polymerization that runs at the same level of the wild type TEL SAM domain (Lane 5). Taken together, we interpret that a relatively small number of mutations are capable of controlling Pnt polymerization *in vitro*. Given the apparently large shift in affinity we observe, it will be important to confirm that this increase is mediated by canonical polymerization, and not non-specific protein binding. To do so, we suggest using monomerizing mutations on opposite interfaces of Pnt to produce a dimer with high affinity, which would be indicative of canonical polymerization. If indeed polymerization occurs in Pnt and is tunable with a few mutations, then further experiments to explore the functional impact of polymerization in Pnt are warranted.

A1.2 Materials and Methods

Pnt was cloned into the His-negGFP/pBAD vector via PCR with NotI and HindIII restriction sites. Pnt mutants were then generated via Quikchange mutagenesis, and generation of PP14 was accomplished via serial Quikchange mutagenesis. negGFP tagged proteins were expressed and purified as described in Chapter 3, and native gels were run on 12.5% native gels overnight at 22V.

A1.3 References

Kim, C. A., M. L. Phillips, W. Kim, M. Gingery, H. H. Tran, M. A. Robinson, S. Faham, and J. U. Bowie. 2001. "Polymerization of the SAM Domain of TEL in Leukemogenesis and Transcriptional Repression." *EMBO Journal* 20 (15):4173–82.

Webber, Jemma L., Jie Zhang, Lauren Cote, Pavithra Vivekanand, Xiaochun Ni, Jie Zhou, Nicolas Nègre, Richard W. Carthew, Kevin P. White, and Ilaria Rebay. 2013. "The Relationship between Long-Range Chromatin Occupancy and Polymerization of the Drosophila Ets Family Transcriptional Repressor Yan." *Genetics* 193 (2):633–49.

Zhang, Jie, Thomas G W Graham, Pavithra Vivekanand, Lauren Cote, Maureen Cetera, and Ilaria Rebay. 2010. "Sterile Alpha Motif Domain-Mediated Self-Association Plays an Essential Role in Modulating the Activity of the Drosophila ETS Family Transcriptional Repressor Yan." *Molecular and Cellular Biology* 30 (5):1158–70.

Appendix II: Characterization of Domain-swapped ETS factors

C. Matthew Hope and Ilaria Rebay

A2.1 Results and Discussion

Our modeling efforts (Chapter 2) demonstrated that the ability to polymerize drives fundamental differences in the occupancy profile of a transcription factor on DNA. To better understand the functional impacts of altering polymerization in a transcription factor, we created two constructs with swapped SAM domains, PntP2 with Yan's SAM domain (PntYanSAM) and Yan with TEL's SAM domain (YanTelSAM). In the case of PntYanSAM, the logic behind the creation of this construct is to achieve *de novo* introduction of polymerization into a TF that is not otherwise thought to polymerize. Mechanistically, there are at least two possibilities for the effect of introducing polymerization into a non-polymerizing TF: increased self-association could result in greater occupancy of activating complexes on DNA and therefore increasing Pnt's role as a transcriptional activator, or alternatively, polymerization might result in the formation of complexes that are incompatible with activation by spreading Pnt via polymerization across target loci, and therefore polymerization may be sufficient to confer repression. In the case of YanTelSAM, we also sought to understand how increased self-association affinity effects Yan mediated repression, and the domain swap of TEL's SAM domain was a forerunner of the more targeted approach to increase polymerization in the creation of the SuperYan mutants (Chapter 3), which demonstrated this relationship more clearly.

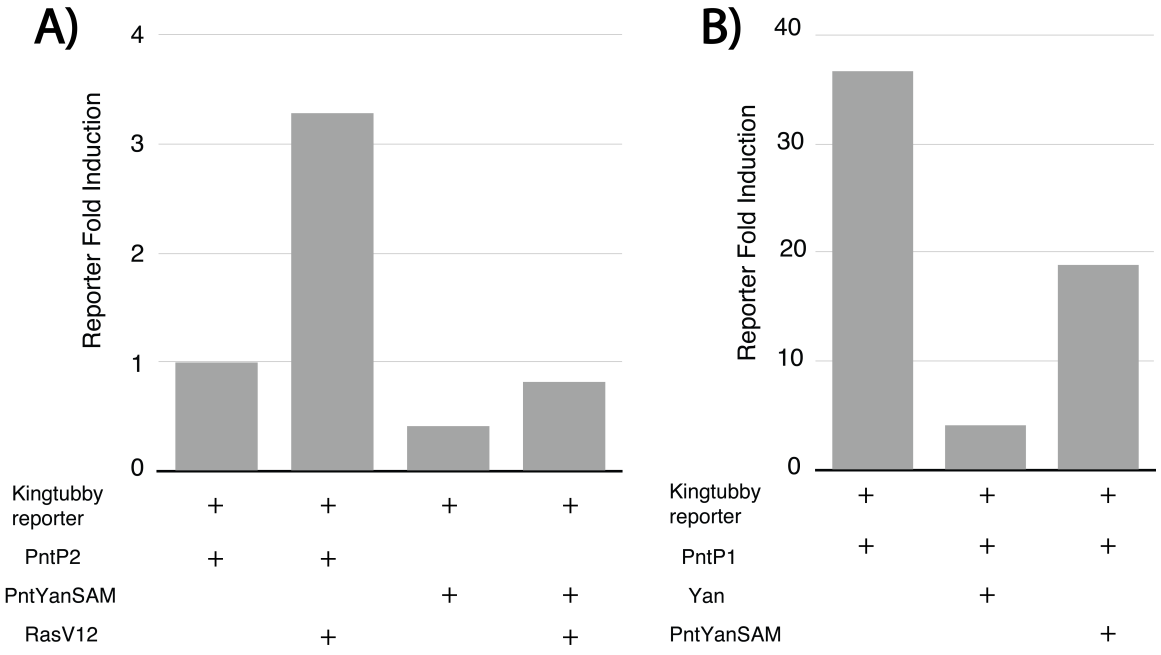


Figure A2.1: PntYanSAM shows decreased ability to activate transcription.

A) Transcription assay comparing PntP2 activation with and without Ras to PntYanSAM. B) Transcription assay comparing Yan and PntYanSAM's ability to repress activation by PntP1. Measurements are the average of three technical replicates, in one biological replicate.

To test the effects of these construct we performed luciferase transcription assays to specifically query effects on transcription at a known target element of both Yan and Pnt (*king tubby)* (Webber et al. 2013). FigureA2.1A shows the comparison for PntYanSAM to PntP2 in order to assess the effect on activation. PntP2 alone drives very little reporter gene expression, but co-transfection with RasV12 increases Pnt's activation, as has been previously shown (O'Neill et al. 1994). If polymerization was able to increase PntYanSAM's activation ability, one would expect activation above the level of PntP2, but instead reporter gene expression is reduced. To test Pnt's ability to act as repressor, we compared it to Yan in terms of ability to repress transcription from PntP1 (Figure A2.1B), where PntYanSAM was able to reduce transcription from the reporter, but not the extent seen with Yan. Taken together, these trends hint at the ability for Yan

SAM mediated polymerization to alter Pnt's behavior, although some key experiments remain to be done. For example, it is known that the Pnt SAM domain contains a docking site for MAPK to bind and phosphorylate Pnt— one possibility is that the whole domain swap removed the ability for RasV12 to stimulate Pnt activation. Therefore it would be helpful to clarify the role of polymerization by introducing the PolyPnt mutations identified in Appendix I, in order to increase polymerization but maintain MAPK binding, and determine if a similar defect is observed. Similarly, if MAPK input is compromised in PntYanSAM then it may be behaving antagonistically to PntP1 at the level of competition for binding, as opposed to establishing active repressive complexes in the manner of Yan. Therefore, it will be important to compare PntYanSAM with PntP2 in a competition with PntP1 to determine if this is the case, or to see in just the Pnt ETS domain is sufficient to reduce the activation of PntP1.

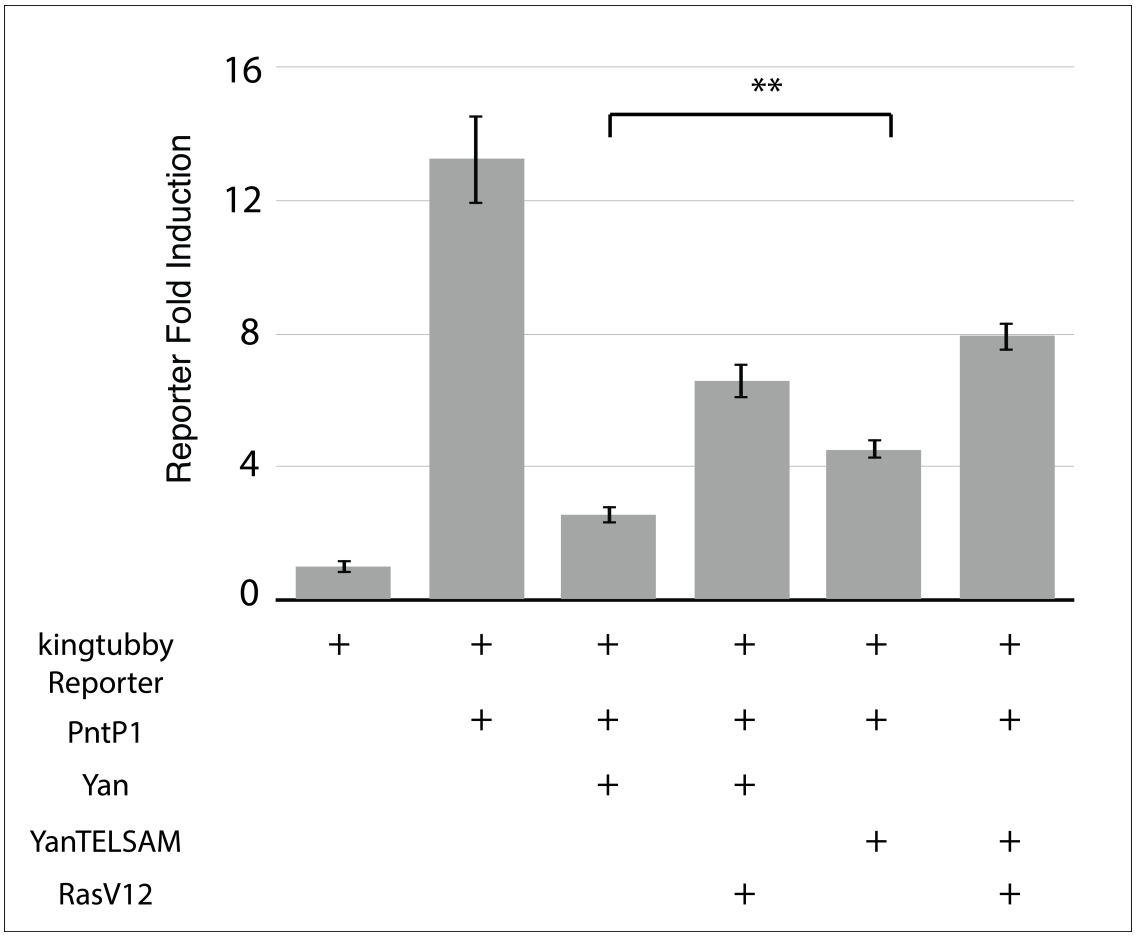


Figure A2.2: YanTelSAM shows elevated baseline of transcriptional repression.

Transcription assay comparing Yan and YanTelSAM's ability to repress the activation of PntP1. Significance represents a pair-wise Student t-test with Bonferroni correction, $p<.01$. Measurements represent the average of three biological replicates.

In the case of YanTelSAM, we observed a similar trend to the behavior of the strongest SuperYan mutants, SY4 and SY5, with respect to transcription. Relative to Yan wild type, YanTelSAM demonstrated an increased baseline of transcriptional repression, but was still responsive to RasV12-mediated relief (FigureA2.2). If YanTelSAM has a similar self-association affinity to that of wild type TEL, then this is consistent with a similar loss of function by sequestration mechanism being at play. Furthermore, YanTelSAM bolsters the argument that Yan self-association affinity is tuned to the wild type level, since two separate methods of increasing self-association in Yan resulted in a

similar mechanistic defect in Yan-mediated repression. Given the fact that similar effects can be exerted by the SuperYan mutants with a much smaller number of amino acid changes, we propose further exploring the SuperYan mutants as a means of assessing the impact of increased polymerization in Yan.

A2.2 Materials and Methods

In order to generate PntYanSAM and YanTelSAM, primers were generated to amplify the Yan and TEL SAM domains, with overlap with the regions just outside the PntP2 and Yan SAM domains, respectively. Thus, Yan SAM (Q35-E116) was inserted between Pnt F166 and E251, and TEL SAM (A40- Q123) was inserted between Yan Q35 and S117. Transcription assays were performed as described in Chapter 3, with N-terminally FLAG tagged PntP1 and PntP2. In Figure A2.1A, 100ng of kingtubby reporter and 20ng of act>Renilla reporter were used in all conditions, while 100ng of PntP2, 100ng of PntYanSAM, and 100ng of RasV12 were used as indicated. In Figure A2.1B, 100ng of kingtubby reporter, 20ng of act>Renilla reporter, and 600ng of PntP1 were used in all conditions, while 100ng of Yan and 100ng of PntYanSAM were used as indicated. In Figure A2.2, 100ng of kingtubby reporter, 20ng of act>Renilla reporter, and 100ng of PntP1 were used in all conditions except reporter alone, and 100ng of Yan, 100ng of YanTelSAM, and 5ng of RasV12 were used as indicated.

A2.3 References

O'Neill, Elizabeth M., Ilaria Rebay, Robert Tjian, and Gerald M. Rubin. 1994. "The Activities of Two Ets-Related Transcription Factors Required for Drosophila Eye Development Are Modulated by the Ras/MAPK Pathway." *Cell* 78 (1):137–47.

Webber, Jemma L., Jie Zhang, Lauren Cote, Pavithra Vivekanand, Xiaochun Ni, Jie Zhou, Nicolas Nègre, Richard W. Carthew, Kevin P. White, and Ilaria Rebay. 2013. "The Relationship between Long-Range Chromatin Occupancy and Polymerization

of the Drosophila Ets Family Transcriptional Repressor Yan." *Genetics* 193 (2):633–49.

Appendix III: Characterizing SuperYan polymerization *in vivo*

C. Matthew Hope and Ilaria Rebay

A3.1 Results and Discussion

One key expectation of increasing Yan self-association affinity is that this increases the proportion of Yan molecules that form long Yan polymers at equilibrium. Due to the fact that Yan polymers are difficult to work with *in vitro*, and that we wish to assess the nature of polymerization *in vivo*, the best established assay for querying the state of Yan polymerization is a Fluoresence Recovery After Photobleaching (FRAP) assay (Zhang et al. 2010). In this assay, molecules of Yan are fluorescently tagged with a mEGFP tag and expressed in transiently transfected S2 cells. A small area of the cell is bleached with a laser and recovery time of fluorescent signal in the bleached area is recorded, with the expectation that molecules in larger complexes should have reduced diffusion rates into the bleached area and therefore longer recovery times. Therefore, we compared one of the SuperYan constructs (SY4) to Yan wild type to determine if we could observe an increase in recovery time.

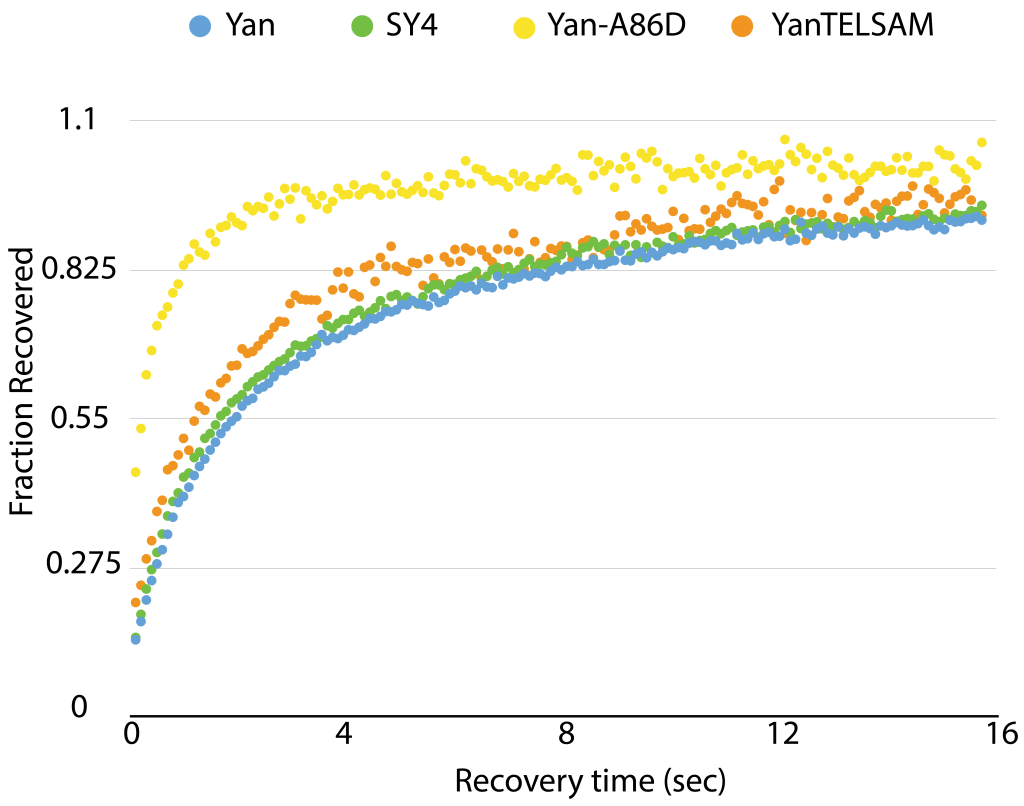


Figure A3.1: SY4 shows similar recovery time to Yan and YanTelSAM via FRAP.

Fraction recovered after photobleaching is plotted as a function of time for Yan, SY4, Yan-A86D, and YanTelSAM. The curves shown represent the average of multiple bleaching events across cells, with n = 39, 40, 12, and 8 for wild type, SY4, A86D, and YanTelSAM respectively.

Surprisingly, Yan wild type and SY4 showed no appreciable difference in recovery time, averaged over many photobleaching events. In order to establish that the assay was working, we included YanA86D which had been previously shown to have a much faster recovery time relative to Yan wild type (Zhang et al. 2010). Consistent with our expectations, we were able to observe much faster recovery of YanA86D, with these mutations being capable of monomerizing Yan. As an additional control, we compared Yan wild type and SY4 to Yan replaced with TEL's SAM domain, and observed similar recovery times for all three constructs. Given the fact that puncta were not observed in S2 cells, and that we hypothesize that endogenous Yan is capable of intercalating and

disrupting large SuperYan polymers, an obvious explanation is that endogenous Yan in S2 cells is exerting an effect on the transiently transfected constructs, resulting in recovery times that appear normal. One obvious test of this is to co-transfect dsRNA against Yan wild type with the fluorescent constructs to see if puncta can be observed when endogenous Yan is knocked down, which might permit better characterization in the FRAP assay.

A3.2 Materials and Methods

A previously generated construct of YanIntNLS-mEGFP/pPL17 (Zhang et al. 2010) was used as a template for Quikchange mutagenesis to generate SY4IntNLS-mEGFP/pPL17. Additionally, YanA86D-2XNLS-mEGFP/pPL17 was also generated previously. Note that the pPL17 vector drives constitutive expression from an *armadillo* promoter, and contains a kanamycin resistance gene. To create YanTELSAM/pPL17, YanTELSAM/pMT was digested with EcoRI to produce a YanTELSAM fragment, and ligated into YanIntNLS-mEGFP/pPL17.

FRAP Experiments

Cells were transfected with the appropriate constructs and imaged two days later to allow expression from the constitutive *armadillo* promoter. 18 by 18mm coverslips were treated with 10mg/mL ConA in PBS and allowed to evaporate. Then cells were settled on the slides for at least one hour in a humid chamber. Immediately before imaging, excess media was drawn off with a Kimwipe and then the slide was placed on an additional microscope slide apparatus (see Thomas Graham's notes on construction). Additional S2 media was added back to the cells to keep them alive during imaging

without risking excess media contaminating the confocal objectives. If necessary, the coverslip with cells can be attached to the slide using nail polish to prevent sliding.

Imaging was carried out with a 63X oil objective with 20X digital zoom. GFP was imaged with excitation at 5% of transmittance, and bleaching was carried out with all lines of an argon laser at 100% transmittance, and then recovery was monitored at maximum speed with no line averaging. To select an area for bleaching, a circle was selected approximately one quarter of the nucleus in size, and applied to all imaged cells. Recordings were taken for approximately 40 frames to allow the fluorescent signal to stabilize before bleaching, and then recovery was allowed to proceed for approximately 160 frames, or about 16 seconds.

To estimate fractional recovery over time, three areas were measured per cell for integrated density in ImageJ imaging software: fluorescence in the bleached spot, fluorescence in a non-bleached area of the nucleus, and background fluorescence outside the cell entirely. Background fluorescence was subtracted from both bleached and unbleached measurements, and then the ratio bleached to unbleached was calculated for all frames of the recording. Lastly, the measurements were normalized to the average of the measurements of the final five frames before bleaching, in order to roughly normalize the recovery measurements to the intensity pre-bleaching. Finally, recovery curves were averaged over many cells to arrive at the final curves for comparison.

A3.3 References

Zhang, Jie, Thomas G W Graham, Pavithra Vivekanand, Lauren Cote, Maureen Cetera, and Ilaria Rebay. 2010. “Sterile Alpha Motif Domain-Mediated Self-Association Plays an Essential Role in Modulating the Activity of the Drosophila ETS Family Transcriptional Repressor Yan.” *Molecular and Cellular Biology* 30 (5):1158–70.

Appendix IV: Implementation of code to calculate Yan occupancy

C. Matthew Hope, Ilaria Rebay, and John Reinitz

Attached to this thesis is a zip file of C++ code used for calculation of Yan fractional occupancy. Please see code and comments for descriptions, and especially the methods section of Chapter 2.

Acknowledgements

I would like to thank all of my teachers who have mentored me over the course of my career, both in the classroom and in labs, at Phoenixville High School, Brandeis University, the Woods Hole Oceanographic Institution, and University of Chicago. They put me on a path long before I could glimpse it for myself, and their patience has provided me with a lifetime of benefit. I especially thank Ilaria Rebay for her endless support and guidance.

I also thank my friends, both in Chicago and across the world, for showing me the joys of living, and allowing me to find myself as a person with their help. Our friends are the few relationships we choose in life, and I feel lucky to have chosen each other.

I would be nowhere without the love and support of every member of my family. I remember especially Walter and Ruth Hope and Mildred Simpson, who set the perpetual example of strength and grace. I also thank my mother Diane, my father Chris, and my brother Andrew, who have lifted me up, cheered me on, and made me so grateful for the indelible bonds of family.

And finally, I would like to thank Ryan, who has been with me every step of the way through graduate school, celebrating the good times, and supporting me in the hard times. In a good relationship, you find out what you want, and in a great relationship, you find out what you are— Ryan is both good and great, and I thank him from the bottom of my heart. For all the time I've spent in lab, I hope to spend tenfold more with you in whatever adventure comes next.