University of South Florida
Digital Commons @ University of South Florida

USF Tampa Graduate Theses and Dissertations | USF Graduate Theses and Dissertations

October 2023

Thermodynamic frustration of TAD2 and PRR contribute to autoinhibition of p53

Emily Gregory
University of South Florida

Follow this and additional works at: https://digitalcommons.usf.edu/etd

Part of the Molecular Biology Commons

Scholar Commons Citation

Gregory, Emily, "Thermodynamic frustration of TAD2 and PRR contribute to autoinhibition of p53" (2023). *USF Tampa Graduate Theses and Dissertations.*
https://digitalcommons.usf.edu/etd/10042

This Dissertation is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.

Thermodynamic frustration of TAD2 and PRR contribute to autoinhibition of p53

by

Emily Gregory

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Molecular Biosciences
College of Arts and Sciences
University of South Florida

Major Professor: Gary Wayne Daughdrill, Ph.D.
Yu Chen, Ph.D.
Kristina Schmidt, Ph.D.
Libin Ye, Ph.D.

Date of Approval:
November 7, 2023

Keywords: intrinsically disordered proteins, cooperativity, effective concentration, fluorescence anisotropy

Copyright © 2023, Emily Gregory

ACKNOWLEDGMENTS

Many thanks to my major professor, Dr. Gary Daughdrill, who has been unfailingly supportive and enthusiastic about this project. I can't thank him enough for the generosity he has displayed in his mentorship. This work is made possible through the funding of his grants: NIH grants R01CA141244, R01GM115556, and Florida Department of Health Project 20B17. I must also thank my lab mates for their support, especially Dr. Wade Borcherds for the data that motivated this work, Dr. Asokan Anbanandam for his work on NMR assignments for the KIX protein, and Ethan Dixon for his assistance with protein purification, as well as my committee members Drs. Yu Chen, Kristina Schmidt, Libin Ye, and, formerly, Younghoon Kee for their guidance.

Additionally, I must thank Dr. Jianfeng Cai for the opportunity to collaborate on the projects elaborated on in Chapter Five as well as several of his students for assisting me with the use of their microplate reader for fluorescence anisotropy assays: Drs. Mi Zhou, Minghui Wang, and graduate students Bo Huang and Yu Yu Win.

Lastly, I must thank my family, especially my mother, for their enduring support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviations	Definitions
CBP	CREB binding protein
CCT	Counterion condensation theory
C_{eff}	Effective concentration
cMyb	Myeloblastosis protein
CREB	cAMP response element binding protein
DBD	DNA binding domain
FA	Fluorescence anisotropy
GST	Glutathione S-transferase
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IS	Ionic strength
ITC	Isothermal titration calorimetry
KIX	Kinase-inducible interacting domain
Lc	Contour length
Lp	Persistence length
Mdm2	Mouse double minute homolog 2
MdmX	Murine double minute X
MLL	Mixed lineage leukemia 1 protein
MSA	Multiple sequence alignments
NCPR	Net charge per residue
NMR	Nuclear magnetic resonance
PIN1	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1
pKID	Phosphorylated kinase inducible domain
pI	End-to-end distance distribution
PRR	Proline rich region
PPII	Polyproline II helix
PTM	Posttranslational modification
REG	Regulatory domain
R_S	Stokes radius
SEC	Size exclusion chromatography
SEOE	Standard error of estimate
SliM	Short linear motif
TAD	Transactivation domain
TAD1	Transactivation domain 1
TAD2	Transactivation domain 2
TET	Tetramerization domain

ABSTRACT

The intrinsically disordered transcription factor and tumor suppressor p53 binds to promoter response element DNA upon cellular stress and activates genes associated with cell cycle arrest, senescence, and apoptosis. Disruption of sequence specific binding to target gene promoters is heavily implicated in human health, where a majority of cancers contain mutations localized to the DNA binding domain (DBD) of p53. P53 DNA binding is regulated by posttranslational modifications, associations with cellular factors, and by an autoinhibitory intramolecular interaction. The autoinhibitory intramolecular interaction occurs when the disordered N-terminal transactivation domain (TAD) interacts with the ordered DBD. Previous work in the Daughdrill lab showed that the second transactivation domain (TAD2) and the proline rich region (PRR) are responsible for inhibition of DNA binding. The goal of this study is to investigate the specific features of TAD2 and PRR that result in inhibition and to gain insight into how these interactions regulate DNA binding.

The Interaction of the disordered TAD2 and PRR with DBD Is multivalent and dynamic. We studied fragments of p53 that included only the DBD and a minimal fragment with maximal inhibition of DNA binding that includes TAD2, PRR and the DBD (ND). We then systematically mutated physicochemical features in TAD2 and PRR to reduce or eliminate inhibition of DNA binding. The TAD2 mutants targeted the negatively charged residues of TAD2, nonpolar residues of TAD2, a conserved motif implicated in p53 transactivation, or a complete deletion of TAD2. PRR mutants were designed to

eliminate chain stiffness due to proline content, potential nonpolar interactions between PRR and DBD, a known pi-cation interaction between PRR residue W91 and DBD residue R174, or to replace the PRR with a flexible linker composed of alternating Gly, Ser, and Thr residues. The effects of these mutations on DNA binding affinity to target and nontarget DNA sequences were measured using fluorescence anisotropy and analytical size exclusion chromatography was used to measure changes in the Stokes radius of p53 ND.

We find TAD2 mutations moderately restore DNA binding to ND, disrupting the intramolecular interaction and increasing the Stokes radius. By analyzing DNA binding under varying salt concentrations using the counterion condensation theory, we find a change in the apparent excess ion release mediated by the charged residues of TAD2, suggesting a mechanism of energetic control over the DNA binding process. We find the PRR is directly involved in autoinhibition but also has a frustrating effect on the interaction between TAD2 and DBD. When TAD2 is deleted, PRR can inhibit DNA binding by a factor of 10 compared to DBD but when TAD2 is present PRR controls its orientation and reduces its ability to inhibit DNA binding.

Analysis of the effective concentration of TAD2 based on PRR suggests autoinhibition is not optimized. Evolutionary analysis suggests the intramolecular interaction is likely present in birds and most mammals, and the frustrated component may have emerged simultaneously.

The results of our experiments define a system where the molecular features of TAD2 and PRR simultaneously compete and cooperate to maintain optimal autoinhibition.

CHAPTER ONE: P53 IS AN INTRINSICALLY DISORDERED TUMOR SUPPRESSOR

General characteristics of intrinsically disordered proteins

Protein function is determined by structure. For intrinsically disordered proteins (IDPs), function does not depend on stable secondary or tertiary structure making primary structure paramount (1). Approximately 30-50% of eukaryotic proteins have intrinsically disordered regions (IDRs) (2). IDRs contribute to cellular complexity and are frequently found in transcription factors (3, 4) (1).

IDRs can vary from short linear motifs that are 3-10 residues long to up greater than 500 residue in length, but they display shared features such as low sequence complexity with a relatively high percentage of charged residues and low percentage of hydrophobic residues (4, 5). Disordered regions often have elevated rates of evolution as they are not as constrained by secondary and tertiary structure as ordered domains (6). The range of conformational flexibility and heterogeneity for IDRs is broad (7), and many IDRs undergo coupled folding and binding to other proteins and DNA (8). Conformational flexibility can allow a single disordered region to assume a wide array of structures depending on the interaction partner, sometimes termed promiscuity (9) with the result that IDRs are often found in interaction hubs and are able to bind multiple partners (10). Similarly, their enrichment in PTM sites, which are exposed for protein-protein interactions, increase their tunability in interactomes (10).

Recent developments have shown that disordered proteins can also bind without folding, remaining dynamic even when tightly and specifically bound (11, 12). IDRs have effects on specificity in binding through a variety of mechanisms that are prevalent in DNA binding proteins (13-15).

P53 is an intrinsically disordered protein involved in health and disease

p53 is the most frequently mutated protein found in cancers (16-18). Figure 1.1A shows the number of missense p53 mutations found in human tumors using data taken from the Catalogue of Somatic Mutations in Cancer (COSMIC) where those residues with the highest mutation frequency are labeled in red (19). A domain map of p53, Figure 1.1B, shows its domains: the N-terminal transactivation domain (TAD), the core DNA binding domain (DBD), the tetramerization domain (TET), and the regulatory domain (REG) (20). The TAD is composed of three subdomains: transactivation domain 1 (TAD1) and transactivation domain 2 (TAD2), and the proline rich region (PRR). About 50% of the p53 protein is disordered: Figure 1.1C shows an IUPRED plot of p53 where residues scored above 0.5 are predicted to be disordered and those below 0.5 are predicted to be disordered, which has been confirmed by experimentation (21-24). Structural characterization of full length p53 is hindered by the high degree of disorder in its N- and C-termini. The structure predictor AlphaFold2 is able to give a low-confidence prediction for disordered regions; however, the prediction for p53 does not capture some known features and the single structure predicted is inherently unable to represent its conformational heterogeneity (25, 26). Experimentally, crystallization has not been successful for visualizing residues N-terminal of P89 (27), and while a recent

study included a cryo-EM image of full-length p53, technical challenges remain and the published structure does not recapitulate some known features of p53 (28).

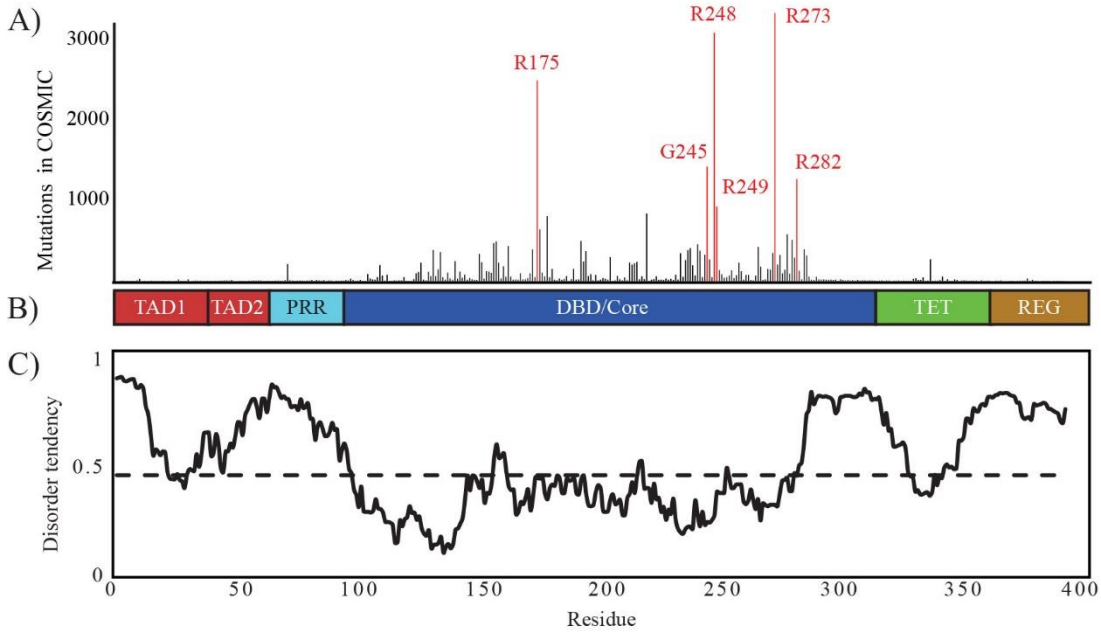


Figure 1.1: Structure of p53. A) Number of missense mutations of p53 found in cancers from the COSMIC database with largest peaks labeled in red, B) Domain map of p53, C) IUPRED plot of p53.

The domain organization of p53 is conserved throughout the animal kingdom, though with the notable absence of the TAD in poriferans, cnidarians, and arthropods (29, 30). This domain organization is partially conserved in the p63 and p73 paralogs, although the TAD is poorly conserved between the three, and p63 and p73 have additional C-terminal domains (31, 32). As is common among DNA binding proteins, the DBD fold is highly conserved across long evolutionary time (33): D. *melanogaster* and human DBD share only 24% sequence similarity and yet bind the same DNA sequence with high affinity (34); C. *elegans* DBD also binds similar DNA sequences as the human protein (35). Slow evolution of the DNA binding function is seen in many transcription factors: FOXO transcription factor paralogs that are millions of years divergent bind

essentially the same DNA sequences as each other; the difference in gene selection between paralogs comes from interactions with external binding factors and from flanking sequences, which are often disordered (36).

DBD is the primary site of cancer-related mutations. Approximately 30% of all cancer-related p53 missense mutations occur at six sites in the DBD referred to as hotspot mutations, which are shown in Figure 1.1A. Mutations at three of the hotspots directly affect DNA binding (R248, R273, R282), and mutations at three others affect the structural stability of the domain (R175, G245, R249), indirectly affecting DNA binding (37, 38). Mutations associated with disease at other sites are also common (39).

The typical location and behavior of p53 is shown in Figure 2.1. Under normal cellular conditions, p53 is constitutively produced and sequestered in the cytoplasm via interaction with its regulators murine double minute X (MdmX), which holds p53 in an inactive state, and mouse double minute homolog 2 (Mdm2), an E3 ubiquitin ligase that targets p53 for ubiquitination and degradation (40, 41). Upon cellular stress, phosphorylation of p53 causes p53 to dissociate from Mdm2 and MdmX after which it migrates to either the nucleus to activate transcriptional responses to stress or to the mitochondria to activate a transcription-independent apoptotic response (18).

While p53 is a hub protein that interacts with many partners, its most studied function is in the activation of gene transcription (42, 43). Shown in Figure 1.3B-C, in response to cellular stress or DNA damage, p53 migrates to the nucleus and binds to promoter response element DNA as a dimer of dimers, an interaction that is partially mediated by the tetramerization domain, recruiting members of the general transcription

machinery and thus initiating the formation of the preinitiation complex. It is notable that under normal conditions in the cell, p53 exists as a monomer, dimer, and tetramer (44).

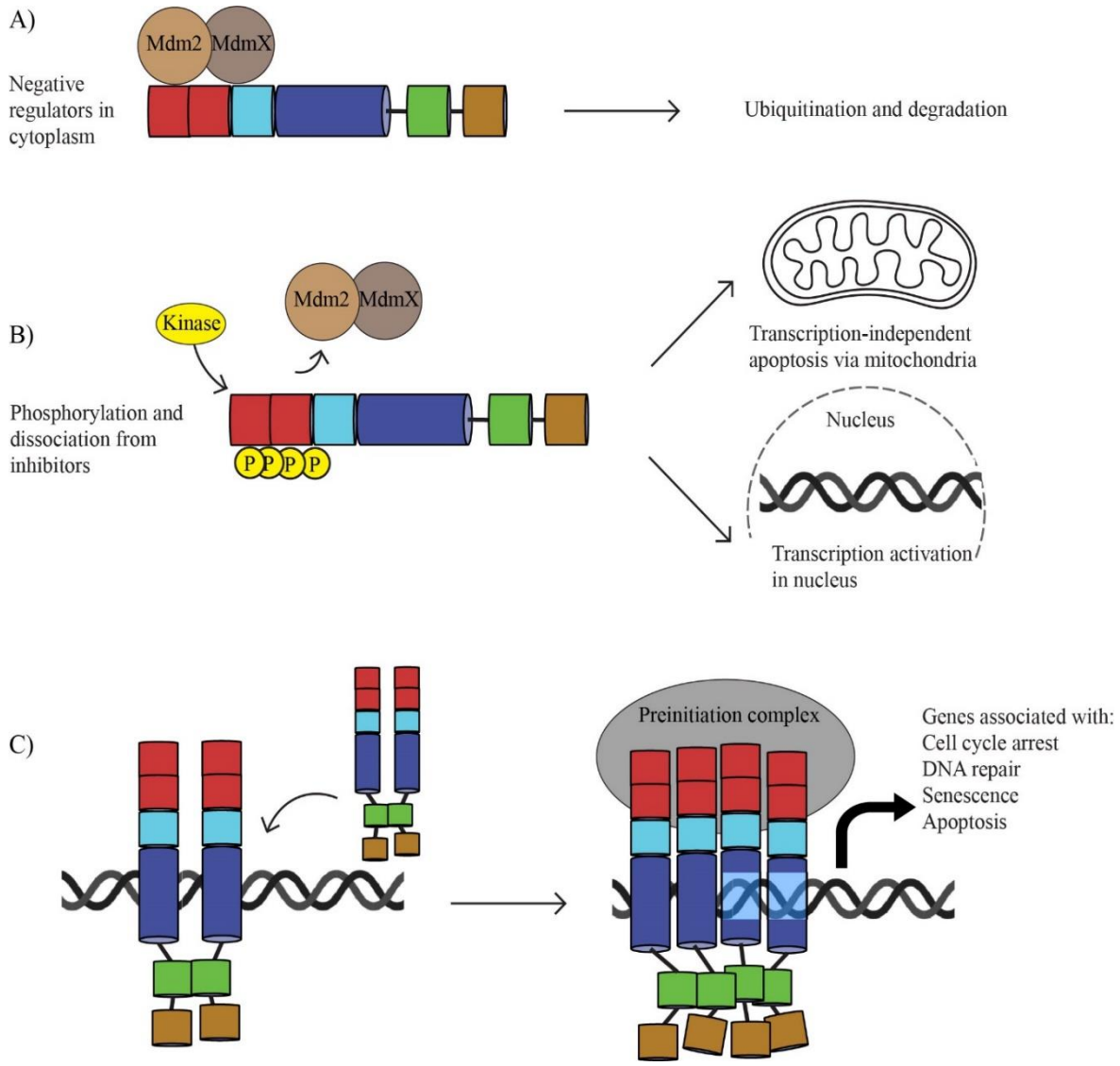


Figure 1.2: p53 in the cell. A) Under unstressed conditions, p53 is sequestered in the cytoplasm with its negative regulators, Mdm2 and MdmX, resulting in ubiquitination and degradation. B) Upon cellular stress, kinases phosphorylate p53 TAD, resulting in the dissociation of Mdm2 and MdmX and migration of p53 to the mitochondria, where it activates apoptosis, or to the nucleus, where it activates gene expression. C) A dimer of p53 binds the promoter response element of a target gene. This dimer recruits a second dimer to form a homotetramer. The homotetramer recruits cofactors and the general transcription machinery to form the preinitiation complex, which leads to transcription of target genes.

P53 as a DNA binding protein

p53 binds DNA as a tetramer at specific sequences that correspond to promoter response element sites (45). The consensus sequence for p53 is composed of two half sites or four quarter sites, defined as RRRCWWGYYY where R is any purine, W is A or T, and Y is any pyrimidine (46, 47). This sequence is highly degenerate: substitutions are common, 35% of promoters bound by p53 have an insertion or deletion (47) and others have multiple copies of the consensus sequence (37) or only three quarter sites (48). A study of nucleotide substitutions in the consensus sequence has quantified the energetic penalty for substitutions at each site, finding that the central C A/T T/A G sequence is the most crucial for high binding affinity in the context of full-length p53 *in vitro* (49). The binding affinity of p53 to various promoter response elements varies by >50-fold *in vitro* with a noticeable trend towards lower affinity for pro-apoptotic promoter response elements (50). Binding affinity has been linked to the ability of p53 to activate transcription (51, 52); however, these studies taken together are not sufficient to explain the behavior of p53 *in vivo*, nor do they propose a structural explanation that relates to changing *in vivo* conditions.

Given the degeneracy of the p53 target sequence, the number of potential binding sites in the human genome has been estimated to be 3,700 to 20,000 (49, 53). Like many DNA binding proteins, p53 binds both specific and nonspecific DNA with relatively high affinity (54, 55). P53 is a DNA-bending transcription factor that is particularly effective at binding cruciform DNA, a common structure formed by target sequences that are palindromes (55), and there is a positive correlation between binding affinity and the flexibility of the DNA sequences immediately flanking half sites

(56). P53 promoter response elements are well conserved between higher mammals, although it is notable that the proapoptotic promoters are the least well conserved – interestingly this divergence in promoter response element sequence is not shared by NRF2 and NF-kB transcription factors (57).

Binding of p53 dimers to DNA is highly cooperative (45). Like many or most transcription factors, it is only active in this role as an oligomer (45, 58, 59); however, unlike most transcription factors, p53 can only activate genes as a homotetramer where four DBDs stably interact with DNA (45). Oligomerization is strengthened by the interaction of tetramerization domains, and this increases DNA binding affinity; however, four DBDs alone tetramerize on target DNA (56, 60). Tetrameric p53 is relatively unstable compared to monomeric DBD, possibly due to an increase in DBD local concentration that results in unfolding and aggregation of DBD (61).

The true number of p53 activation targets is unknown (62). A common estimate is approximately 100, with the best characterized being those associated with cell-cycle arrest (p21 and p16) and apoptosis (BAX, PUMA, Noxa), although p53 also activates genes associated with the DNA damage response (DDB2, XPC), self-regulation (Mdm2), and metabolism (TIGAR) (63, 64). It can also activate a transcription-independent pro-apoptotic response via the mitochondria and regulate many other pathways through protein-protein interactions (65, 66). As a participant in cell fate decisions, regulation of p53 occurs via many mechanisms including subcellular localization, ubiquitination and degradation, other posttranslational modifications, and intramolecular interactions that regulate binding to DNA (67). Transcription-dependent functions of p53 are thought to be the most essential of its activities that contribute to tumor

suppression (68). Accordingly, p53 hotspot mutations associated with tumors interfere with DNA binding by directly affecting DNA-contact residues or residues that support the DNA-binding interface (69, 70); it is also notable that this second, conformational class of hotspot mutants result in a destabilized molecule that is prone to aggregation (71, 72).

An intramolecular interaction with p53 decreases DNA binding affinity

It has long been noted that p53's DNA binding affinity is affected by both the disordered N-terminus and the disordered C-terminus (73); however, the lack of complete structural studies has complicated identification of the mechanism. An intramolecular interaction between the N-terminus and DBD has been partially characterized and shown to decrease DNA binding affinity (74-76). Early investigations by our lab suggested that this interaction increases DNA binding specificity by decreasing affinity for non-target DNA (74). The studies described in Chapters 2-5 use the N-terminus and DBD of p53 (ND, human residues 1 – 312) to interrogate this mechanism.

The autoinhibitory interaction is weak. Attempts to quantify the dissociation constant of the untethered domains, termed *in trans* binding affinity, using isothermal titration calorimetry failed, as did competition experiments. NMR studies suggest the *in trans* dissociation constant is in the range of 1 mM. Despite this, the K_D of binding to consensus DNA for DBD alone versus the ND fragment was reduced by 164-fold (i.e., tighter binding) (60). Additionally, the interaction does not result in coupled folding and binding of TAD2 (74).

Our earlier results also suggested that the TAD2 and PRR domains have some sequence specificity when interacting with DBB (74). The TAD2-DBD interaction is

decreased or abolished at high salt concentrations, suggesting an electrostatic component (76), but chemical shift changes in TAD2 at hydrophobic residues suggest a nonelectrostatic component as well (74). Deletion of PRR increased the interaction of TAD2 and DBD, suggesting a regulatory role for PRR, but the features of the region were not probed in detail (75).

Effective concentration and multivalency in autoinhibition

Protein self-regulation by autoinhibition is a way to achieve equilibrium between open and closed states (77), and IDRs play an important role in autoinhibition of many proteins (78-81). An increasing number of autoinhibitory regions have been characterized in transcription factors where disorder is enriched and frequently plays a role in DNA-binding specificity (3, 13, 14).

Interactions between autoinhibitory disordered and ordered domains are generally weak when not linked together (77). The tethering of domains increases the frequency of their interaction, a feature that can be quantified as effective concentration, which is the concentration of untethered protein that binds with equivalent frequency as the tethered domain and is based on the length and sequence of residues separating the interacting’domains (82). Autoinhibitory IDRs have *in trans* binding affinities in the micromolar to millimolar range (83-87) with longer distances separating the domains correlating with tighter untethered binding. The TAD2 and PRR subdomains are immediately adjacent to DBD, which leads us to propose that the interaction between TAD2, PRR, and DBD has evolved to be submaximal at autoinhibition.

The dynamic, or fuzzy, binding seen in the TAD2-DBD interaction is a feature of many autoinhibitory IDR interactions (81, 88-91). Secondary structure in the bound state was not assessed for the PRR due to the high proline content and low sequence complexity. However, a previous group was able to crystalize the C-terminal residues of PRR (91-94) bound to DBD where W91 makes a single contact with R174 of DBD (PDB **2XWR**, shown in Figure 1.3) (27). The many prolines in PRR may participate in hydrophobic interactions or act as hydrogen acceptors in multiple weak interactions with the surface of DBD (92, 93).

Features of TAD2

TAD2 is a disordered hub region of p53 with two regions of minimal transient secondary structure (94) that engages in coupled folding and binding to many partners such as TAZ2 (95), RPA70N (96), and HMGB1 (97) to form an amphipathic helix, though it can bind in an extended conformation as with TFIIH subunit p62 (98). It is enriched in posttranslational modification sites, notably the phosphorylation sites at S46 and T55, which our lab has investigated in the context of Mdm2 and DNA binding, and which other labs have investigated in the context of the intramolecular binding (75, 99).

The transactivation domain of p53 is required for activation of genes; however, there is evidence that the two transactivation subdomains, TAD1 and TAD2, are semi-independent and may participate in the activation of different genes. Brady et al. found in a study with knock-in mice that TAD1 deficient mice (L25Q/W26S) were unable to initiate cell cycle arrest or apoptosis but only TAD1 and TAD2 deficient mice (L25Q/W26S/F53Q/F54S) were unable to initiate senescence and tumor suppression to a degree similar to p53-null mice, suggesting the two subdomains influence separate

processes (100). Other studies have found that either TAD1 or TAD2 is dispensable for transactivation of a variety of genes (101, 102); however, the mechanism that explains this observation remains unclear.

The sequence of TAD2 is poorly conserved, even among vertebrates, but conservation of sequence properties, like acidic residues, is relatively high (103). TAD2 appears to have emerged as a domain separate from TAD1 following the split between the p53 and p63 paralogs that occurred with a whole genome duplication event in jawless fishes about 600 million years ago (104), implying that the role of TAD2 in self-regulation of p53 emerged after this point as N-terminal autoinhibition has not been noted in p53 paralogs.

Features of PRR

The least conserved subdomain of p53 (105), the PRR is a low complexity sequence composed in humans of 36% proline, 33% alanine, 18% charged residues and 12% other residues. Proline-rich regions are common in globular proteins (92, 106), and like other proline-rich regions, the p53 PRR is relatively stiff with a propensity towards polyproline II helical structure (107-109).

The function of the PRR is poorly understood in comparison to that of TAD2. It has at times been speculated to be merely a positioning linker, a participant in transcriptional activation, a player in stimulation of apoptosis, a potentially important domain in signaling, and a domain that plays an important role in aging by as yet unknown means (110). PRR has a smaller known number of interaction partners than TAD2; it is known to interact with p300 (111), Sin3, Mdm2 (112), Cin85, Pin1, Gas7, and Argbp2 (113), although the functional effects of many of these interactions is unknown.

Given its high proline content and the presence of five PxxP motifs, the PRR has been suspected of engaging in signaling events with SH3 domain partners although interaction with SH3 domains has only been marginally investigated (113, 114). A common human polymorphism at 72R/P affects the binding affinity of partners ASPP1/2 and iASPP, leading to differential expression of pro-apoptotic genes (114, 115). The homozygous 72P allele is associated with greater risk or lower survival rates for some types of cancer; however, the mechanism that causes these outcomes remains unclear (116, 117). The peptide prolyl isomerase PIN1 acts on pT81/P82 (118). It has been noted that a familial mutation at P82 increases a specific cancer risk (119) and decreases Mdm2 association, affecting response to DNA damage (118). Mutational frequency of PRR sites found in tumors from the COSMIC database are noted to be more frequent than those found in TAD1 or TAD2 (39); however, the reason for the relative enrichment of mutations in these sites is speculative.

In vivo PRR deletion studies have shown contradictory results. For mouse p53, deletion of PRR in some studies results in inhibition of cell cycle arrest with minimal changes to apoptosis (120, 121) where another shows inhibition of E1A-dependent apoptosis but no inhibition of cell cycle arrest and no change in transactivation of observed genes (119). *In vivo* PRR deletion in human p53 shows minor inhibition of p21 transactivation with a graded effect based on the length of the deletion and a general decrease in RNA levels of most, but not all, p53-mediated genes observed (122). Alternatively, deletion of PRR was found primarily to result in decreased transactivation of specific genes such as PIG3 (123, 124) or BAX (112). PRR is generally thought to influence apoptosis, but a unified mechanism to explain this has not emerged, with

explanations involving targeting of p53 to specific promoters (115), changes in DNA binding or chromatin remodeling (122, 124), and transcription-independent pathways (125). Structural investigations on PRR may explain some of these disparate results.

In cell studies have also noted the importance of PTMs on PRR functioning. Phosphorylation of T81 by c-Jun N-terminal kinase occurs upon DNA damage and secondarily results in isomerization of P82 by peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (PIN1) (126). Interaction with the prolyl isomerase cyclophilin 18 decreases *in vitro* binding to the Gadd45 promoter and deletion of Cyp18 results in increased apoptosis (127).

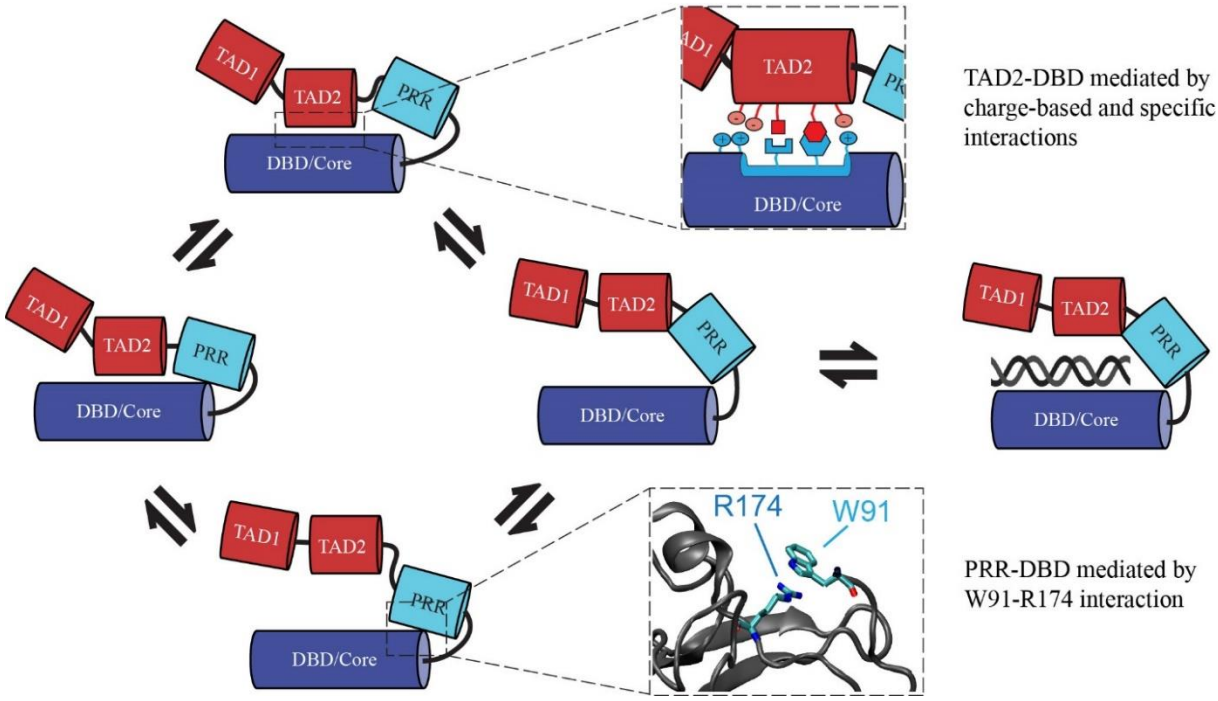


Figure 1.3: Model of the intramolecular interaction and its features. Our model predicts that autoinhibition of DBD is either dominated by the TAD2-DBD interaction, where an inset shows a combination of interaction types involved, or by the PRR-DBD interactions, where the W91-R174 interaction is the primary mediator. Inset here is p53 (91-289) PDB **2XWR**.

Figure 1.3 shows a model in which we propose that the TAD2-DBD and PRR-DBD interactions may be independent or be substates of the autoinhibited protein. Furthermore, we expect the TAD2-DBD and PRR-DBD interactions to be based on different mechanisms. In this model, binding of DNA requires a transition from the closed state to the open state.

Environmental conditions and the counterion condensation theory

The presence of ions in solution inherently affects protein-DNA binding events: positively charged ions that interact with DNA in solution stabilize the helix but also shield it from external electrostatic interactions, and positively charged DNA-binding proteins are similarly shielded by ions to remain soluble (128, 129). Thus, salt concentration can have a large effect on protein-DNA interactions; however, aspects of ion behavior such as the binding specificity and mobility in this context are not well understood (130).

A model of protein-DNA interaction anticipates that increasing salt concentration disrupts charge-based interactions, Figure 1.4A-B, but does not affect hydrogen bonding, Figure 1.4C-D. We recognize that protein-DNA interactions involve more components than these, that the behavior of ions around these molecules is not yet well-defined, and that ion identity and concentration have a multitude of effects on both protein and DNA; however, this illustrates the idea that varying salt concentrations in a protein-DNA binding experiment can allow us to parse the contributions of different types of interactions.

Our model of the intramolecular interaction supposes a combination of electrostatic and hydrophobic interactions contribute to the intramolecular interaction in

p53 (Figure 1.3). We used the counterion condensation theory, a model of ion behavior around DNA, to infer energetic components of the p53-DNA interaction that are altered by weakening specific features of the intramolecular interaction between TAD2 and DBD. Increasing salt concentration generally decreases the strength of protein-DNA interactions, and the counterion condensation theory provides a model to quantify ion release based on changes in DNA binding affinity as ionic strength increases, which discriminates the energetic contribution to binding from ion release.

Dysregulation of DNA binding by p53 is thought to be the major contributor to cancers in which p53 is found to be mutated. Regulation of p53's DNA binding ability is

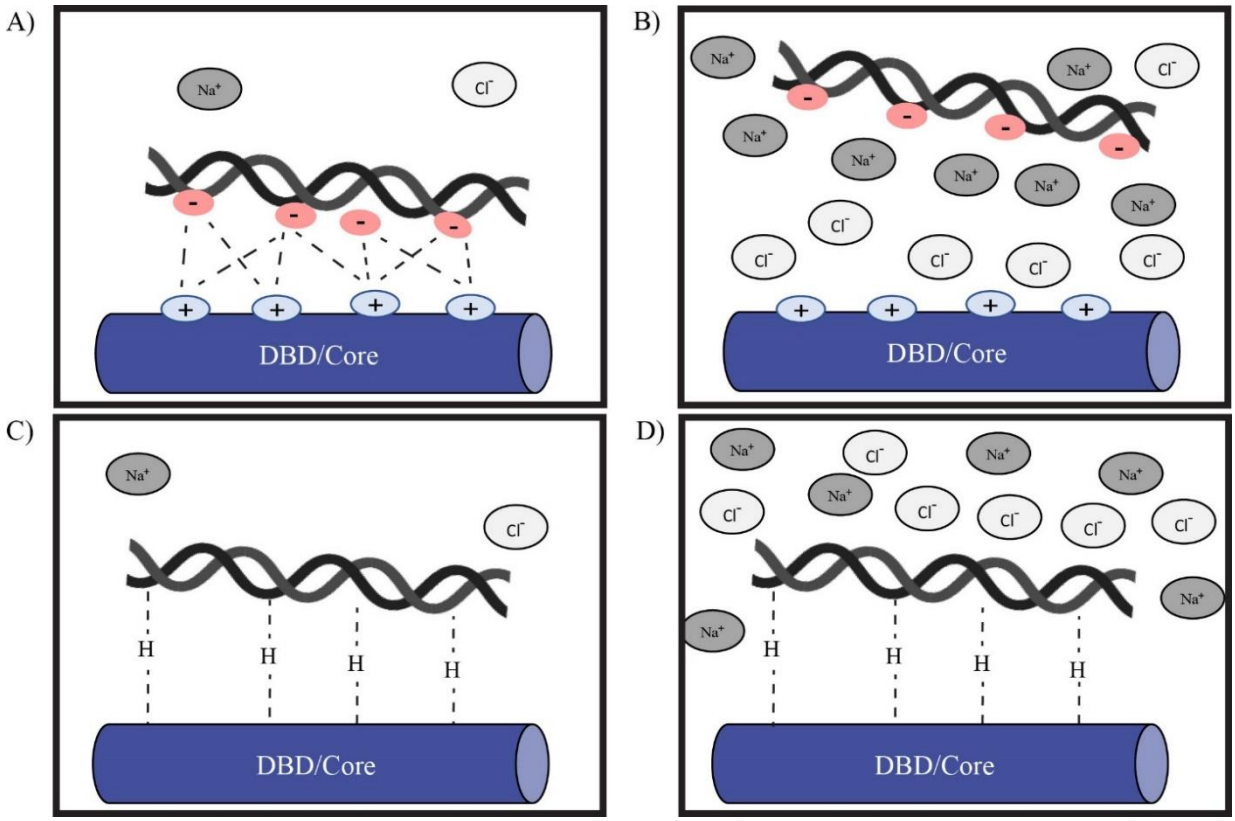


Figure 1.4: Environmental salt conditions affect protein-DNA interactions. A-B) Electrostatic attraction is strong and nonspecific in low-salt conditions and weaker in high-salt conditions. C-D) Hydrogen bonding is specific and is not disrupted by high salt concentration.

comprised of several layers, some of which are poorly understood. This study seeks to characterize an intramolecular interaction that represents a relatively recently discovered layer of regulation. Furthermore, it is clear that understanding the function and structure of p53's disordered regions applies to drug discovery, which is an ongoing challenge (131) and to the question of how p53's behavior contributes to cell fate decisions.

Mimics of intrinsically disordered TADs interact with KIX

The studies described here include NMR investigations of the interaction of various synthetically produced peptide mimics with the ordered Kinase-inducible interacting domain (KIX) of the CREB Binding Protein (CBP). The KIX domain interacts as a coactivator for the disordered TADs of the transcription factors myeloblastosis protein (cMyb), mixed-lineage leukemia 1 protein (MLL), and cAMP response element-binding protein (CREB) (132).

Our collaborators in the lab of Dr. Jianfeng Cai conducted binding affinity assays on synthetic versions of these TADs that utilize either sulfonyl groups or stapling of the peptides to increase their unbound helicity. Our lab conducted ^{1}H-^{15}N HSQC to compare their interaction sites on KIX. Dysfunction and/or mutation of these transcription factors is associated with disease. Thus, these studies are used as a step to assess the use of these compounds as inhibitors of the transcription factors' interactions with KIX.

CHAPTER TWO: SEQUENCE PROPERTIES OF THE TAD2-DBD INTERACTION THAT INHIBITS DNA BINDING

Note to the readers

This chapter is comprised of prior published data, used with the permission of the publishers (see Appendix C) (60).

Rationale

Initial results from our lab suggested that the intramolecular interaction mainly came from the TAD2-DBD interaction with only a minor contribution from the PRR (74). Given the importance of TAD2 for p53 function, we began by investigating the features that contribute to the TAD2-DBD interaction.

TAD2 is a relatively short subdomain (human residues 41-61), Figure 2.1, that is highly negatively charged. We designed mutant versions of the ND WT fragment (human residues 1 – 312) that targeted predicted features of the intramolecular interaction. The ND DE mutant changes all seven acidic residues of TAD2 to alanine, the ND NP mutant changes seven nonpolar residues to alanine, and the ND QS mutant changes only W53/F54 to QS. The W53/F54 sequence is key to TAD2's ability to bind several external partners, and the selection of QS is based on historical precedence (100, 102). Sequences of these mutants, as well as IUPRED predictions and Agadir predictions to assess the likelihood of secondary structure are shown in Figure 2.1. The

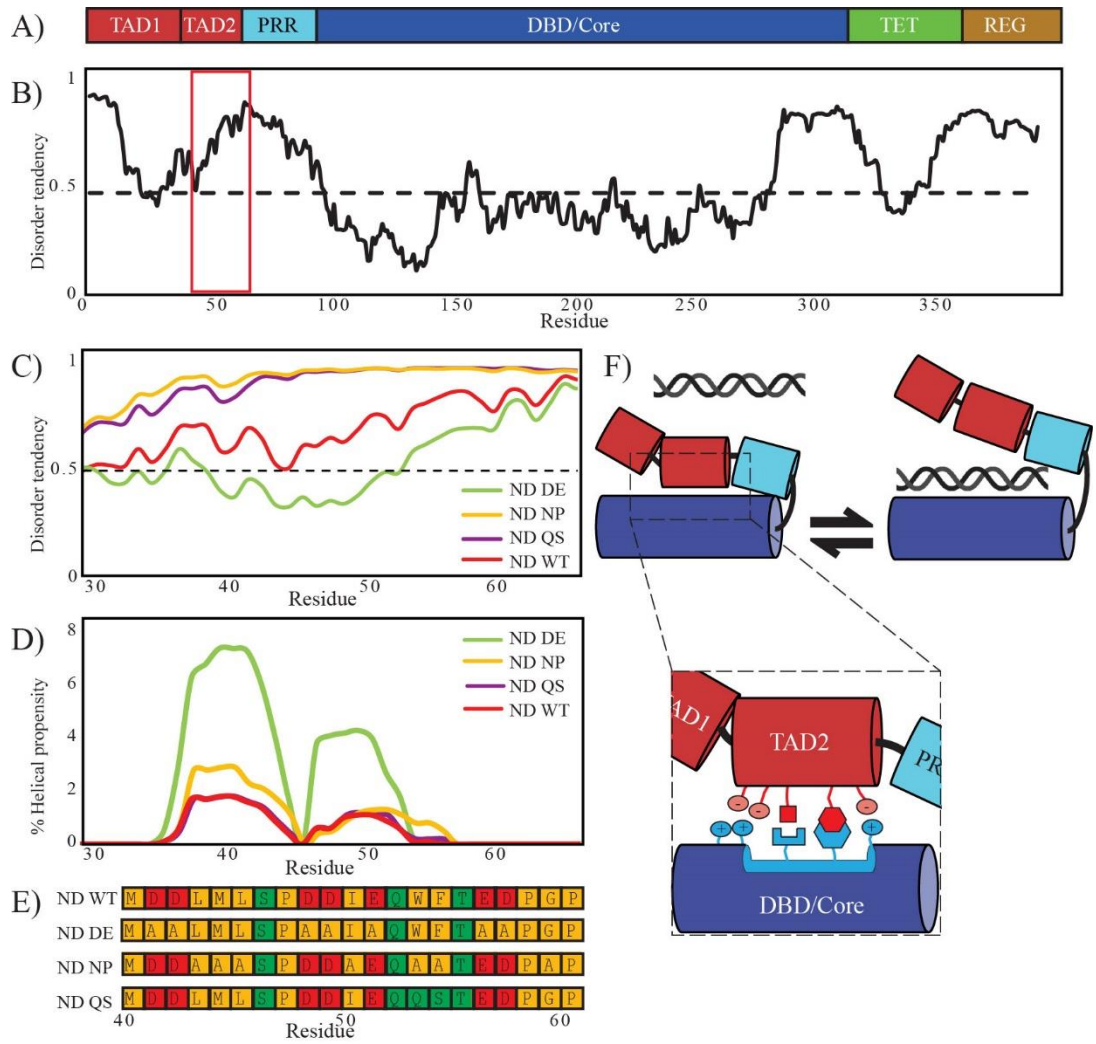


Figure 2.1: TAD2 interacts with DBD via specific and electrostatic interactions. A) Domain map of full-length p53 shows its domains, B) C) IUPRED plot of full length p53, C) IUPRED plot of residues including and flanking TAD2, D) Agadir predictions of WT and mutant TAD2, E) WT and mutant TAD2 sequences, F) model of TAD2-DBD interaction shows a combination of charged and noncharged interactions that must be disrupted for DNA to bind DBD.

combination of features thought to compose the intramolecular interaction is shown in Figure 2.1F.

Due to the weak nature of the TAD2-DBD interaction, direct study of the domains' binding affinity was not feasible. Therefore, we used fluorescence anisotropy to evaluate the effects of TAD2 mutations on the intramolecular interaction using DNA binding as a

proxy, where a decrease in the intramolecular interaction is seen as an increase in DNA binding relative to ND WT. We used ionic strength (IS) to evaluate TAD2 features more finely: in assessing DNA binding affinity from 125-225 mM IS, we expected charge-based features to be strongly affected and nonpolar features to be only weakly affected. Additionally, we used the counterion condensation theory to evaluate the salt-dependent and salt-independent contributions to DNA binding.

We found that our ND TAD2 mutants had greater DNA-binding affinity than ND WT at all IS and for all DNA sequences used. Assessment of the apparent size of the fragments using analytical size exclusion chromatography showed that increased DNA binding affinity of ND mutants correlated with an increase in apparent size, suggesting that the mutations successfully decreased the strength of the intramolecular interaction and resulted in a more open but not unfolded conformation. All ND TAD2 mutants restored DNA binding to a similar degree at physiological ionic strength; however, the mutants' behavior diverged over a range of IS. Mathematical analysis of our binding data suggests that the DBD-DNA interaction is driven by entropy, possibly deriving from ion release, where interaction with nonspecific DNA is less so. The charged residues of TAD2 modulate the release of ions from DBD upon DNA, regulating this energetic component while presumably providing an entropy sink for the actual binding process.

To determine the salt-dependent binding affinity and related inferences about ion release and energetic components of binding, we used the counterion condensation theory, a model of ion behavior in the context of protein-DNA interactions.

The counterion condensation theory

The polyelectrolyte model, sometimes referred to as the oligolysine model and later called the counterion condensation theory (CCT), posits that decreasing protein-DNA binding affinity with increasing salt concentrations can be used to predict ion behavior and the energetic components of the binding event (129, 133, 134).

As shown in Figure 2.2, CCT posits that positively charged ions are condensed nonspecifically on phosphates of the DNA backbone, and the DNA is surrounded by an

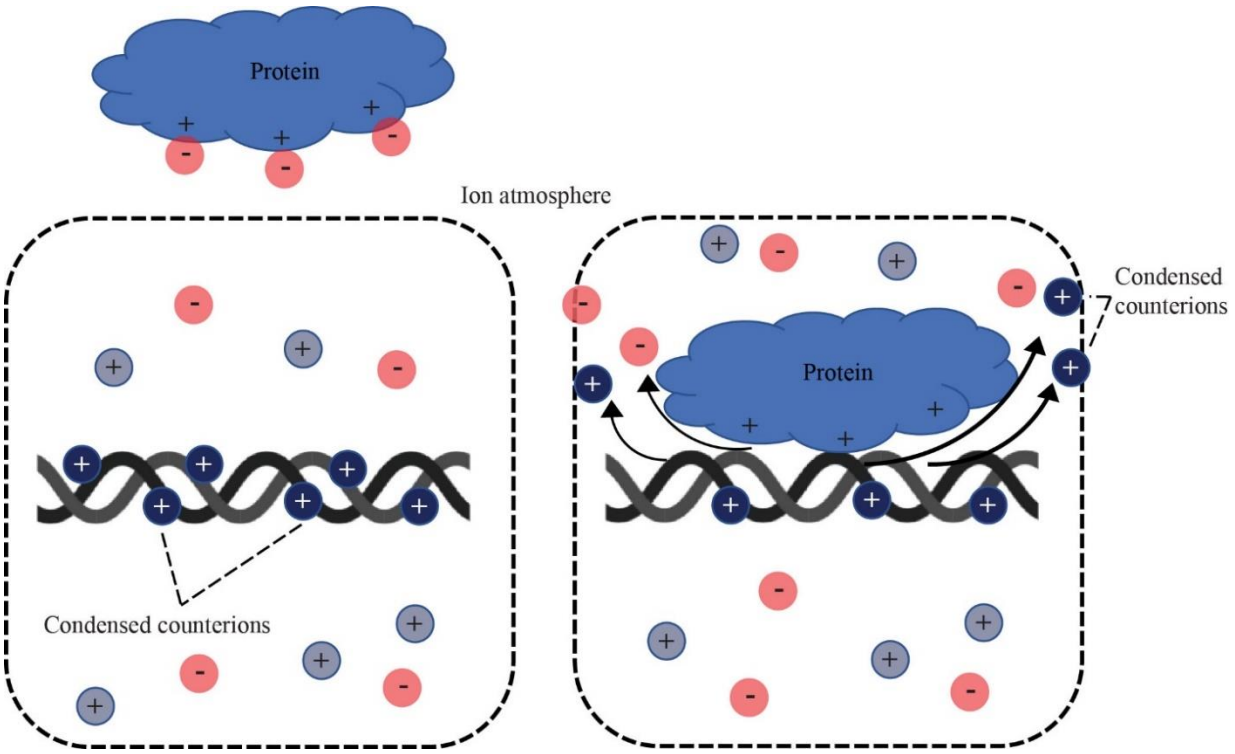


Figure 2.2: The counterion condensation theory predicts ion behavior in protein-DNA interactions. A polyelectrolyte like DNA has negatively charged ions condensed on its surface nonspecifically, and the DNA is surrounded by the ion atmosphere, an environment in which ions behave distinctly from solution. A positively charged protein also binds negatively charged ions. Both protein and DNA must release these bound ions in order to bind each other. The number of condensed counterions that are released from DNA corresponds to the number of phosphate contacts made.

ion atmosphere, whose ions behave distinctly from ions outside of the atmosphere (133). Similarly, a positively charged protein binds negatively charged ions. For a binding event to occur, the ions engaged with charged components must be released before binding to the external partner. For a rod-like DNA segment, each phosphate backbone contact must be deshielded to interact, thus the number of counterions released from DNA corresponds to the number of contacts made.

A double log plot of binding affinity and salt concentration linearizes in a range of salt concentrations that is specific to an individual system. CCT uses the slope of this plot to describe the relationship between the salt-dependent decrease in affinity and the number of ions released from DNA upon binding a protein.

Method scheme

Proteins were produced using a bacterial expression system and purified using affinity chromatography, as described in detail in Chapter 7. High affinity, or specific, DNA binding was assessed with a 20-bp consensus sequence (135) and low affinity, or nonspecific, binding was assessed with a scrambled version of the high affinity consensus sequence. This scrambled sequence is not predicted to give the lowest possible affinity, based on the algorithm studies by Veprintsev and Fersht (49); however, it contains the same GC content as the consensus sequence. DNA sequences for protein constructs were ordered from manufacturers and subcloned into an appropriate vector by our lab, excepting the ND QS construct, which was made using site directed mutagenesis as described in Chapter 7 Methods.

We chose fluorescence anisotropy as our primary binding assay because isothermal titration calorimetry (ITC) had three major problems: 1) enthalpy values at high salt concentrations were greatly diminished as shown in Figure 2.3, 2) DBD binds DNA with a dissociation constant in the low nanomolar range, which is outside the preferred range of ITC, and attempts to lower the C-value resulted in a low signal-to-noise ratio, and 3) the DNA required for ITC would have been prohibitively expensive. Sample ITC heat traces and thermodynamic values are shown in Figure 2.3. and Table 2.1.

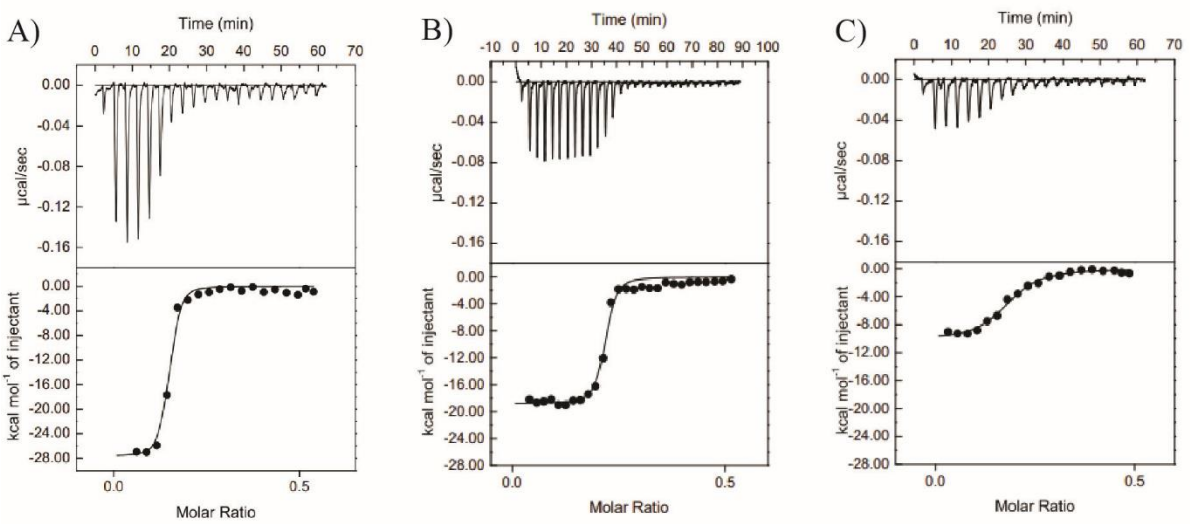


Figure 2.3.: ITC heat curves with increasing salt concentration. DBD with consensus DNA at A) 15 mM IS, B) 85 mM IS, C) 185 mM IS.

Stoichiometric values are in good agreement with the values obtained from FA at all IS, and ΔG values are in good agreement at 85 mM IS; however, values at both extremes deviate from values seen in FA.

Table 2.1: ITC values for DBD binding consensus DNA

[NaCl] (mM)	Total ionic strength	ΔG (kcal/mol)	K_D (nM)	ΔH (kcal/mol)	TΔS kcal/mol)
3	15 mM	-11.41	5.1	-28.82	-17.40
59.74	85 mM	-11.53	2.1	-18.45	-6.91
159.26	185 mM	-10.03	49.3	-11.27	-1.24

FA is a light-based assay that relies on the speed of tumbling molecules differing based on their size (136). A fluorescently tagged small molecule, in this case DNA, emits photons response to stimulation at an appropriate wavelength that is essentially unpolarized due to its tumbling speed. When a larger molecule binds this smaller molecule, the smaller molecule tumbles more slowly than when unbound, and the excited fluorophore emits light that is partially polarized. The use of fluorescence anisotropy has been examined at length in the context of protein-DNA interactions (136) and has been used extensively with p53 and DNA (50, 105, 137).

FA is a resilient assay that can tolerate extreme temperatures and salt concentrations, and, unlike ITC, FA can effectively measure K_D ranges from the nanomolar to millimolar range (136, 138). Our system used a 20-mer DNA molecule tagged with 6-Carboxyfluorescein (6-FAM), an isomer of fluorescein isothiocyanate (FITC) with good quantum yield (0.92 at pH 7.4) (139) whose peak absorbance and emission wavelengths are 495 nm and 520 nm, respectively. Concentrations of tagged DNA were held stable at 10 nM per well with p53 added in one half dilutions starting at ~100-200 μM to 1 nM concentrations. The decrease in polarization of tagged DNA as p53 is added in progressively smaller increments is plotted in logarithmic scale.

Because p53 binds DNA as a dimer of dimers, concentration of p53 is considered as a dimer although it is plotted in concentrations of monomer in Figures 8-9. As shown in Figure 2.4, p53 binding to consensus versus scrambled DNA results in points that are best fit with lines that use different equations. For scrambled DNA, in which the p53 dimer and DNA bind in a 1:1 ratio, we use the traditional equation described at length in Chapter 7 (140). For consensus DNA where the binding of p53 is highly cooperative (45), we use a modified equation where the Hill coefficient, an indication of cooperativity (141), is set to 2. Unlike ITC, FA cannot directly measure thermodynamic parameters. The Van't Hoff equation was used to estimate thermodynamic parameters as described in Chapter 7 (138).

The hydrodynamic radius or Stokes radius (R_S) of a protein defines a protein as a featureless sphere with an average radius (142). In the context of this study, a difference in R_S between two mutants of the same or similar molecular weight represents a difference in their conformations. We utilized analytical size exclusion chromatography (SEC) to determine R_S of p53 fragments. We hypothesized that disruption of the intramolecular interaction must result in the protein occupying an open, or uninhibited state, a greater percentage of the time it travels through an SEC column.

Salt dependent binding affinity linearizes in a double log plot

We conducted binding experiments using fluorescence anisotropy in buffers with IS ranging from 15–225 mM and using two DNA sequences. One is a high affinity sequence taken from a consensus promoter sequence (135), which we refer to as consensus DNA. The other is a scrambled version of this sequence that maintains the same GC content and is used as a representative of nontarget DNA. Figure 2.4 shows

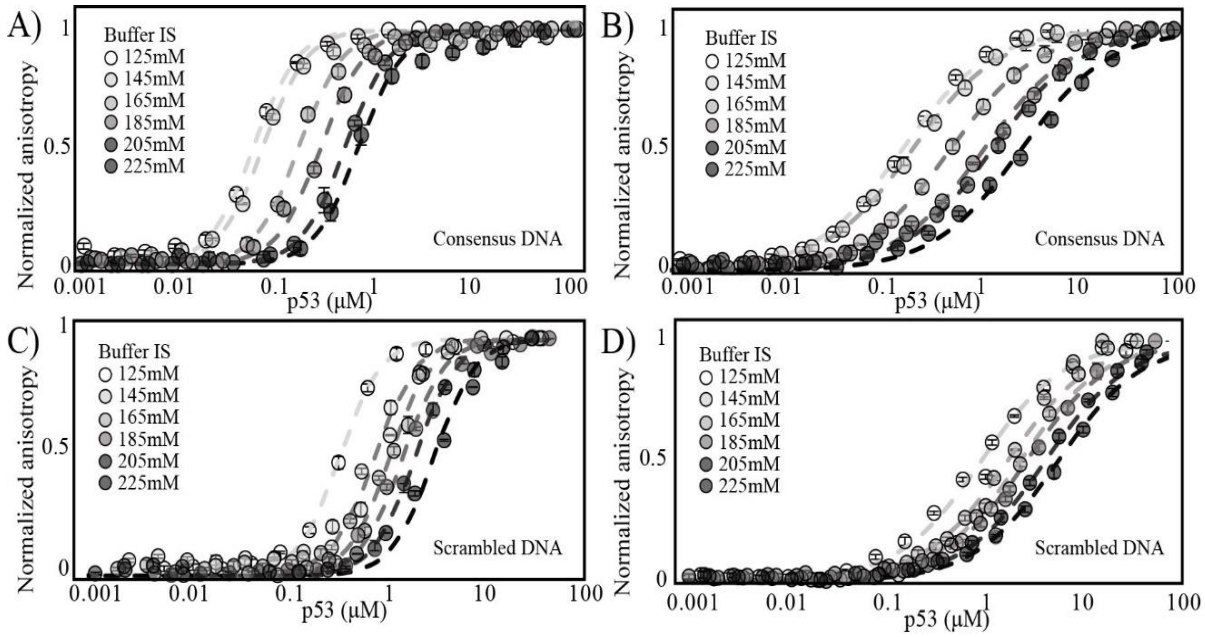


Figure 2.4. DBD and ND WT bind DNA across IS. Fluorescence anisotropy plots show the change in signal from a fluorescently tagged DNA fragment as protein is added: an increase in the concentration of p53 needed to achieve saturation when DNA concentration is kept stable as buffer salt concentration increases. Fluorescence anisotropy plots from 125 – 225 mM IS of (A) DBD bound to consensus DNA; (b) DBD bound to scrambled DNA; C) ND WT bound to consensus DNA; D) ND WT bound to scrambled DNA.

the normalized anisotropy values of fluorescently labeled DNA plotted as a function of DBD or ND WT concentration in buffers with IS ranging from 125 – 225 mM. Dashed lines show the fit to a cooperative binding model in the case of consensus DNA, and to a single-site binding model for scrambled DNA. Both models assume p53 binds DNA as a dimer of dimers (45). As salt concentration increases, binding affinity of p53 to DNA decreases. This is in accordance with observations of p53 specifically (45) and of DNA-binding proteins in general (134, 143). Hill coefficients are approximately 1.8 for p53 binding to consensus DNA and 1 for binding to scrambled DNA. This supports previous studies showing that p53 binds its target DNA in a cooperative manner and nontarget DNA in a noncooperative manner (45). We observed the same trend in cooperativity in

DBD, ND WT and the mutants (Figure A1), but K_D values are 5–200 times larger for ND WT than for DBD (Table A1 and A2). At 125 mM IS the K_D for DBD binding consensus DNA was 0.9 ± 0.07 nM and at 225 mM IS K_D was 104.5 ± 5 nM. For binding to scrambled DNA, K_D ranges from 89.1 ± 5 nM to 1388 ± 44 nM over the same range of IS. These results are in the same range as previously observed binding affinities of DBD to DNA (74, 144). Similar trends are observed for ND WT, for which fluorescence anisotropy curves across a range of IS are shown in Figure 2.4C-D. The K_D for ND WT binding to consensus DNA ranges from 43 ± 3.4 nM to 3861 ± 40 nM and binding to

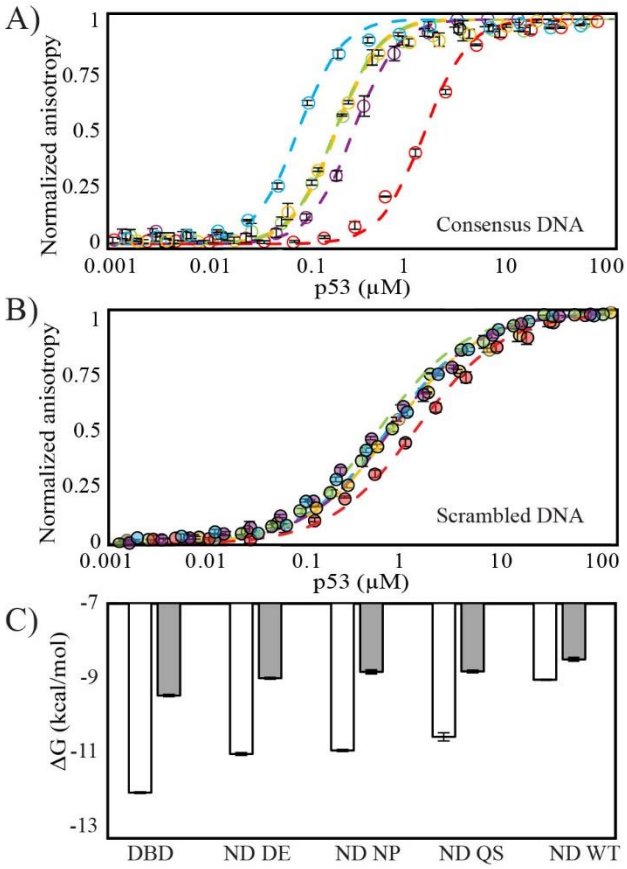


Figure 2.5: **Binding of DBD, ND WT and TAD2 mutants to consensus and scrambled DNA at physiological ionic strength.** Blue is DBD, yellow is ND NP, green is ND DE, purple is ND QS, and red is ND WT. A) Consensus DNA, B) Scrambled DNA, C) ΔG graphs of all binding consensus and scrambled DNA where unfilled and rectangles are consensus and scrambled DNA, respectively.

scrambled DNA ranges from 193 ± 8.2 nM to 3705 ± 230 nM. See Table A1 and A2 for full range of K_D and ΔG values. Error bars in Figure 2.4 represent the standard deviation of three measurements at each IS and the fitting errors presented in Table A3 are the standard error of estimate.

Binding to consensus and scrambled DNA at physiological IS

To determine the contributions of charged and nonpolar interactions between TAD2 and DBD in the autoinhibition of DNA binding we designed three mutants where all aspartic and glutamic acid residues in TAD2 were changed to alanine (ND DE), where all the nonpolar residues from TAD2, including W53 and F54, were changed to alanine (ND NP), and where W53 and F54, were changed to glutamine and serine (ND QS) (See 6E for sequences). The ND QS mutant is based on an early study of p53 in which this mutation inhibited transactivation and apoptosis by inhibiting interactions with multiple domains of CBP/p300 (95, 145, 146). A decrease in the intramolecular interaction should lead to increased DNA binding affinity. Figure 2.5A shows the binding curves of fluorescence anisotropy experiments for DBD, ND WT, and the ND mutants at physiological IS (145 mM). The ND mutants have a binding affinity for consensus DNA that is closer to DBD than ND WT, indicating all the mutants disrupt the intramolecular interaction between TAD2 and DBD. ND DE and ND NP have similar binding affinities to one another for consensus and scrambled DNA, increasing the free energy of binding for consensus DNA relative to ND WT by −1.99 and −1.89 kcal/mol, respectively (Table 2.2). The ND QS mutant has DNA binding affinity between ND NP and ND WT and increases the free energy of consensus DNA binding by −1.49 kcal/mol relative to ND WT.

Table 2.2: ΔΔG values for cooperativity/frustration

Consensus DNA

	DBD	ND DE	ND NP	ND QS	ND WT
DBD	0.00	1.03	1.13	1.49	3.02
ND DE	-1.03	0.00	0.10	0.46	1.99
ND NP	-1.13	-0.10	0.00	0.36	1.89
NDQS	-1.49	-0.46	-0.36	0.00	1.53
ND WT	-3.02	-1.99	-1.89	-1.53	0.00

Scrambled DNA

	DBD	ND DE	ND NP	ND QS	ND WT
DBD	0.00	0.46	0.64	0.63	0.97
ND DE	-0.46	0.00	0.18	0.17	0.51
ND NP	-0.64	-0.18	0.00	-0.01	0.32
NDQS	-0.63	-0.17	0.01	0.00	0.33
ND WT	-0.97	-0.51	-0.32	-0.33	0.00

Notes on Table 2.2: The ΔΔG is the ΔG of the DNA binding of the top of the column protein minus the row header protein. All values are in kcal/mol.

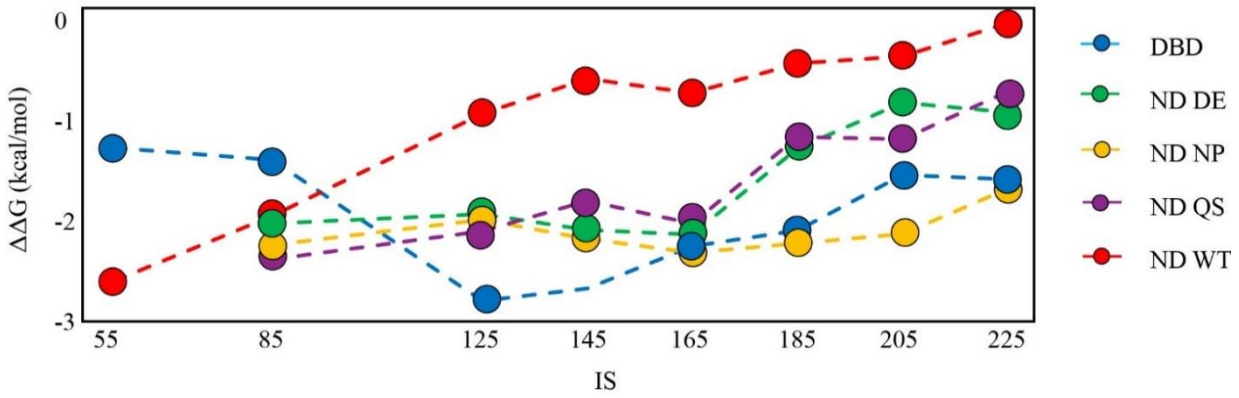


Figure 2.6. Binding specificity of DBD, ND WT, and TAD2 mutants. For each p53 fragment, $\Delta\Delta G = \Delta G_{consensus} - \Delta G_{scrambled}$ at a given IS indicates binding specificity.

Effect of IS on binding specificity of DBD, ND WT, and the ND mutants

Binding specificity is commonly estimated as $\Delta G_{specific}$ — $\Delta G_{nonspecific}$ (147, 148). Figure 2.6 shows the ΔΔG values for DBD and ND WT at 55–225 mM IS, and the ND TAD2 mutants at 85–225 mM IS. Below physiological ionic strength, ND WT has greater specificity than DBD for consensus versus scrambled DNA as evidenced by the larger negative ΔΔG; however, this trend reverses between 85–125 mM IS. Figure 2.6 also shows that at higher IS, ND NP has a similar binding specificity to DBD and the binding specificity for ND DE is closer to ND WT. This is interesting because we expect the nonpolar interactions between TAD2 and DBD to be more specific than the charged interactions and our data shows that removing them increases DNA binding specificity while removing the charged interactions between TAD2 and DBD reduces specificity. We think ND DE has lower binding specificity because the strength of the hydrophobic effect between nonpolar residues in TAD2 and DBD becomes stronger at higher IS (149, 150). In contrast, ΔΔG for ND NP tracks with DBD at higher salt concentrations, indicating that the acidic residues in TAD2 are responsible for inhibiting binding to nonspecific DNA. We expect residues W53 and F54 in TAD2 to play a role in forming specific interactions with DBD but introduction of Q53/S54 reduces DNA binding specificity, suggesting the introduction of these amino acids, and not removal of W53/F54, is driving this effect. The ND WT fragment used in this study lacks the tetramerization domain and only enhances DNA binding specificity at low ionic strength even though it shows strong inhibition of DNA binding and maintains binding cooperativity for specific DNA up to 225 mM IS. As shown in Figure 2.6, the DBD can

bind DNA specifically in the absence of TAD2 and the TET, and Figure 2.5C shows that ND WT inhibits binding to either consensus or scrambled DNA by a similar amount.

In our previous work we showed the intramolecular interaction between TAD2 and DBD in monomeric p53 became intermolecular when the tetramerization domain (TET) was present (74). In a related study, Wright and colleagues showed that adding TAD2 to a p53 fragment containing the DBD and TET enhances DNA binding specificity by inhibiting binding to nonspecific DNA but has no effect on binding to specific DNA (76). The binding studies by Wright and colleagues were conducted at an IS close to 165 mM using similar specific and nonspecific sequences to ours. Using full length p53 with and without TAD2, their K_D ratio for binding was 1 for specific DNA and 5.7 for nonspecific DNA. By comparison our K_D ratios for ND WT and DBD binding to specific and nonspecific DNA are 70 and 5.3, respectively. Taken together these data suggest that inhibition of DNA binding to both specific and nonspecific sequences is driven by the intramolecular interaction between TAD2 and DBD and specificity enhancement depends on this interaction becoming intermolecular when p53 is tetrameric. As mentioned, we think addition of the tetramerization domain reduces the hydrophobic effect between TAD2 and DBD and this could be happening due to differences in the way TAD2 interacts with DBD when the intramolecular interaction becomes intermolecular.

Table 2.3: ΔΔG values for specificity

Ionic strength (mM)	DBD	ND DE	ND NP	ND QS	ND WT
85	-1.34	-1.96	-2.17	-2.32	-0.70
125	-2.72	-1.87	-1.93	-2.04	-2.53
145	-2.60	-2.03	-2.11	-1.74	-1.89
165	-2.19	-2.07	-2.25	-1.96	-0.89
185	-2.03	-1.23	-2.16	-1.11	-0.55
205	-1.49	-0.78	-2.07	-1.14	-0.68
225	-1.53	-0.87	-1.62	-0.67	-0.39

Notes on Table 2.3: Specificity is defined here as $\Delta G_{consensus} - \Delta G_{scrambled}$. All values are in kcal/mol.

Estimating ion release using Counterion Condensation Theory

To assess the sensitivity of the TAD2-DBD interaction to IS, we conducted fluorescence anisotropy binding experiments on ND WT and the ND mutants from 125–225 mM IS. Figure 2.7 shows the linear region of log(K_A) versus log[Salt] plots. Figure 2.7A shows that the binding of consensus DNA to DBD is tighter than to ND WT at every IS and that the presence of TAD2 in ND WT inhibits DNA binding at a level that corresponds to increasing IS by 70–80 mM for DBD. Binding of DBD and ND WT to scrambled DNA (Figure 2.7B) shows a similar trend in affinity where the inhibition of DNA binding by TAD2 corresponds to an increased IS of 40–60 mM for DBD.

Counterion condensation theory proposes that ions are uniformly condensed on DNA at a concentration that is relatively independent of buffer conditions or the type of protein binding. When a positively charged protein binds DNA, a number of counterions equivalent (or fractionally equivalent) to the number of nonspecific ionic contacts made between the protein and DNA backbone are released into solution (133). The oligolysine

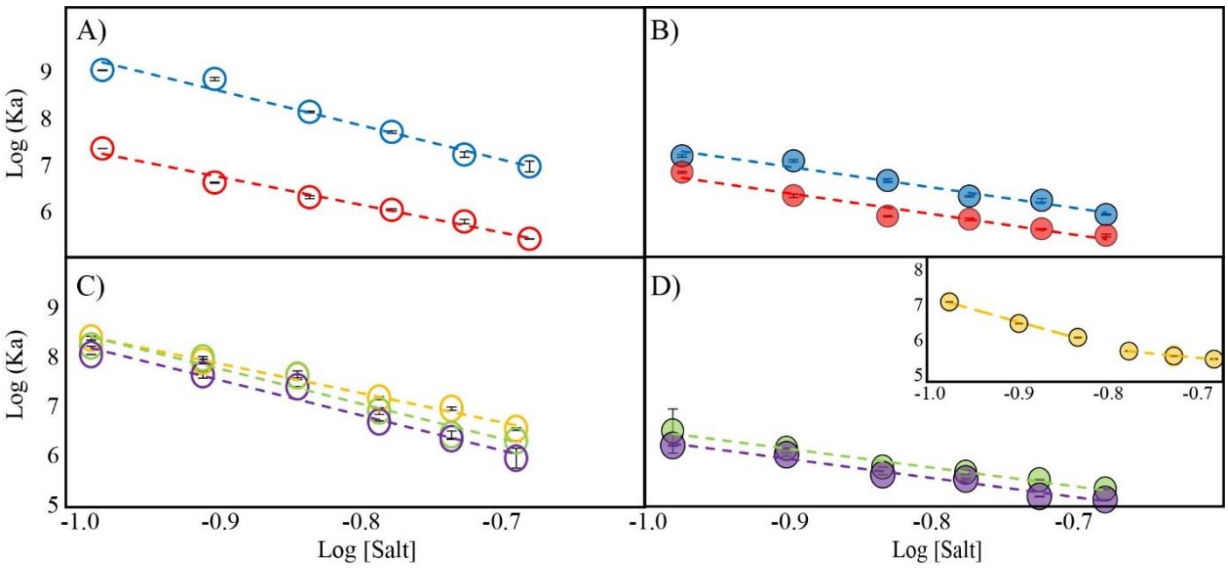


Figure 2.7. **Salt-dependent DNA binding affinity of DBD, ND WT and TAD2 mutants.** Double log plots of fluorescence anisotropy data plotting of Log (K_A) vs Log [Salt] from 125-225mM IS of (a) DBD and ND WT binding to consensus DNA where blue is DBD, red is ND WT, (b) DBD and ND WT binding to scrambled DNA where blue is DBD, red is ND WT (c) ND mutants binding consensus DNA, where green is ND DE, yellow is ND NP, purple is ND QS, (d) ND mutants binding scrambled DNA, where green is ND DE, purple is ND QS. Yellow inset is ND NP, demonstrating two different slopes varying with IS.

model developed by Record and colleagues as an extension of the counterion condensation theory predicts that the observed decrease in DNA binding affinity as salt concentration increases can be used to estimate the number of these nonspecific ionic contacts (129, 151). As described in detail in Chapter 7 Equation 4, the slope (N) of the double log plots Figure 2.7 is proportional to the fractional number of counterions released from the DNA backbone (Ψ), approximately 0.7 per phosphate contact for short oligonucleotides (152), and any excess ions released from the protein (β). According to this theory, a smaller slope corresponds to the release of fewer ions, whether they originate from backbone phosphates or from protein. As shown in Table 2.4, DBD has a larger slope than ND WT when binding consensus DNA, corresponding to greater predicted ion release.

Additional binding data was collected for DBD and ND WT at 15 mM and 55 mM, and for all constructs at 85 mM; however, this data was not included in analysis due to nonlinear behavior at low ionic strength, which is a common occurrence (151). Data is shown in Table A1 and A2.

Table 2.4: Estimated excess ion release in TAD2 mutants

	Consensus DNA		**Scrambled DNA**	
	-Slope	**Excess ion release**	**-Slope**	**Excess ion release**
DBD	7.09	3.9	4.09	0.6
ND DE	7.08	3.6	3.89	0.5
ND NP	5.94	2.4	6.91, 2.36	3.4, 0.0
ND QS	7.16	3.6	3.90	0.4
ND WT	5.99	2.5	4.15	0.7

Crystallographic studies show five DNA backbone contacts made by DBD when bound to the p21 promoter (38, 52). We assume the same number of DNA backbone contacts are made by DBD to consensus DNA because our consensus sequence is similar to the p21 sequence. We also assume ND WT and TAD2 mutants make the same number of contacts as DBD because TAD2 does not interact with DNA (74) or affect binding cooperativity according to the Hill plots in Figure A1. The difference in the slopes between DBD and ND WT when binding consensus DNA corresponds to a difference in the predicted release of excess ions when binding DNA (Table 2.4) where DBD is predicted to release 3.9 excess ions and ND WT is predicted to release 2.5 excess ions. This small difference in ion release corresponds to a difference in salt sensitivity where DBD experiences a 117-fold increase in K_D versus ND WT's 86-fold increase in K_D over this range of IS. We also observe that inhibition of DNA binding is

greater for ND WT as IS decreases, indicating a stronger intramolecular interaction at lower salt concentrations. A similar divergence of salt-dependent binding affinity was seen in a previous study of an autoinhibitory IDR-DBD interaction in which the addition of an acidic domain lowered DNA binding affinity and changed the slope of its double log plot (90). By contrast, ND WT binding to scrambled DNA has a slope similar to that of DBD (Figure 2.7B and Table 2.4). We assume the same number of backbone contacts are made when p53 binds a nontarget sequence as is suggested by structures of low affinity p53-DNA complexes (52). Assuming five backbone contacts, the slopes of ND WT and DBD when binding scrambled DNA correspond to predicted excess ion release of 0.7 and 0.6, respectively.

Figure 2.7C shows ND DE, ND NP, and ND QS bind consensus DNA more tightly than ND WT (also see Table A1 and A2). Slope values for ND DE and ND QS are close to DBD, while ND NP has a slope close to ND WT (Table 2.4). From these results we can make three conclusions: (1) ion release after removal of acidic residues (ND DE) is similar to ion release of DBD, (2) removal of several nonpolar residues in TAD2, including W53 and F54, (ND NP) has no effect on ion release relative to ND WT, and (3) introduction of Q53 and S54, not removal of W53 and F54, is responsible for changes in ion release of ND QS. The first two conclusions were expected and the third suggests the Q53/S54 mutant may do more than interfere with binding to coactivators.

When binding scrambled DNA, the slopes are similar for DBD, ND WT, and the ND DE and ND QS mutants (Figure 2.7B,D). We predict that ND DE and ND QS release 0.5 and 0.4 excess ions, respectively, when binding scrambled DNA, similar to DBD and ND WT. ND NP does not have a single linear slope over the 125–225 mM range when

binding scrambled DNA. Instead, it appears to have a linear portion at 125–165 mM IS with a slope of −6.91 and another linear portion at 185–225 mM IS with a slope of −2.35 as shown in the inset in Figure 2.7D. Slopes and estimated excess ion release from these two states are shown in Table 2.4 to be different from each other and from other p53 fragments. This suggests to us that ND NP binds scrambled DNA in multiple states.

According to the oligolysine model, ΔG of binding can be separated into electrostatic and nonelectrostatic components, where the slopes of the plots in Figure 2.7 multiplied by log[Salt] is the salt-dependent entropy due to ions being released into solution from the phosphate backbone (129, 153). As shown in Figure 2.8A-B and Table 2.5, the Record interpretation of the salt-dependent entropy (Equation 5) predicts salt-

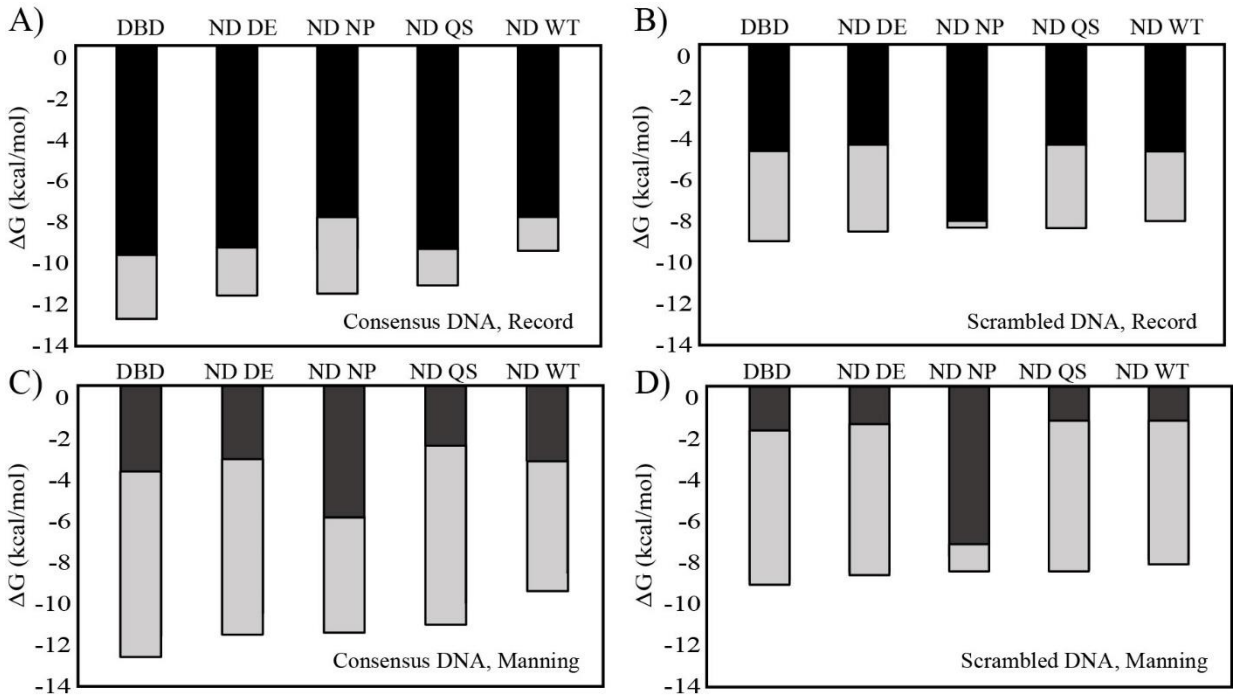


Figure 2.8: Salt-dependent and salt-independent components of Gibbs free energy. At physiological ionic strength, the free energy of binding is shown for DBD, ND WT and TAD2 mutants where ■ is the salt-dependent component and □ is the salt-independent component for A-B) Record's model or C-D) Manning's model.

dependent entropy to be the energetic driver of the p53 fragments binding to consensus DNA, ranging from 68–85% of the total energy. However, in an earlier binding study from our group at an IS of 85 mM using isothermal titration calorimetry we observed a large entropic penalty for DBD binding consensus DNA and a smaller penalty for ND WT and both had a large enthalpy change upon binding (74). Van't Hoff plots using temperature-dependent fluorescence anisotropy data also predict a large enthalpic component of binding (Figure 2.9 and Table 2.6) (52).

Table 2.5: Percentage of Salt-dependent and salt-independent components of Gibbs free energy.

	Consensus DNA			
	Record interpretation		Manning interpretation	
	% Electrostatic	% Nonelectrostatic	% Electrostatic	% Nonelectrostatic
DBD	**79.7**	**20.3**	**69.0**	**31.0**
ND DE	**82.3**	**17.7**	**70.8**	**29.2**
ND NP	**68.0**	**32.0**	**46.6**	**53.4**
ND QS	**85.1**	**14.9**	**74.8**	**25.2**
ND WT	**81.0**	**19.0**	**63.0**	**37.0**
	Scrambled DNA			
DBD	**56.0**	**44.0**	**20.3**	**79.7**
ND DE	**53.9**	**46.1**	**19.4**	**80.6**
ND NP	**97.7**	**2.3**	**84.8**	**15.2**
ND QS	**57.1**	**42.9**	**18.0**	**82.0**
ND WT	**60.6**	**39.4**	**24.5**	**75.5**

Notes on Table: Electrostatic, or salt-dependent, and non-electrostatic, or salt-independent percentages of binding energy are found using either Record's interpretation (Equation 5) or Manning's interpretation (Equation 6) of CCT.

We note here that thermodynamic values derived from a Van't Hoff plot are inferred rather than directly measured. Enthalpic components such as buffer protonation and folding that occurs during binding is not measured here with the result that enthalpic values are always underestimated (154, 155). Van't Hoff analysis is additionally often complicated by unexpected curves at low temperature or a dependence of enthalpy on temperature (156-158). However, we may think of our Van't Hoff data as representing a minimum or approximate enthalpy. The nearly linear plots of DBD and ND bound to scrambled DNA suggests minor heat capacity change (159), which supports predictions of DBD's structure when bound to nontarget DNA (160-162) and the general observation that nonspecific protein-DNA binding exhibits low heat capacity change

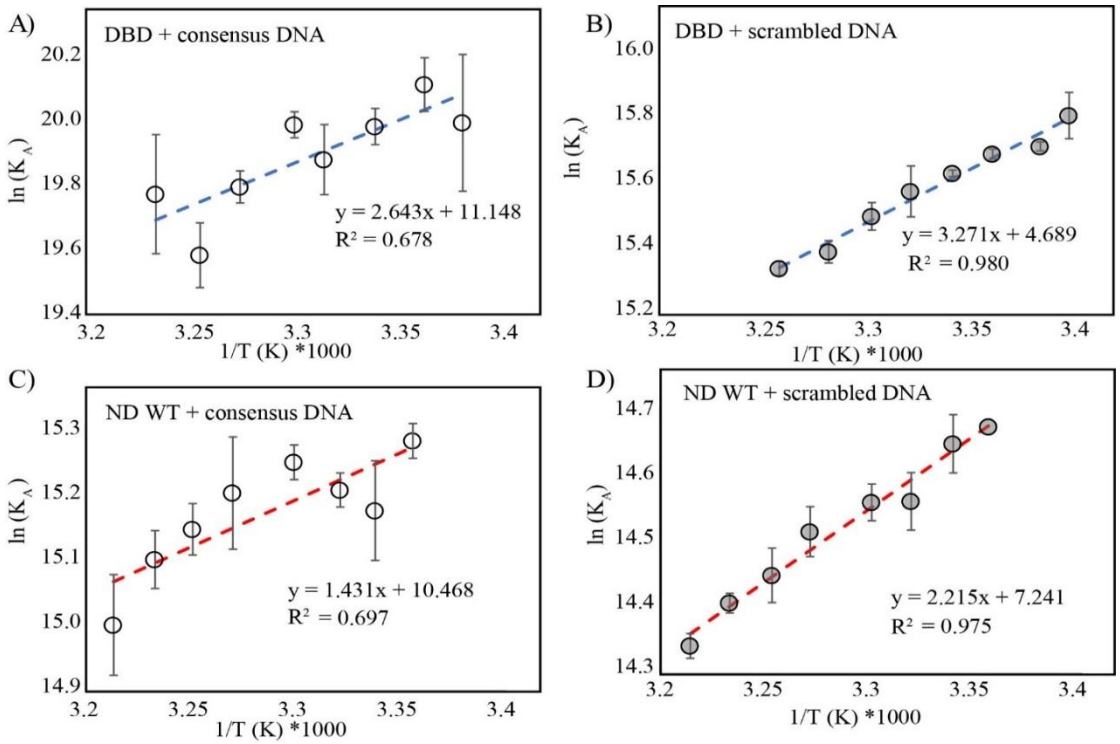


Figure 2.9: **Van't Hoff plots.** A) DBD with consensus DNA, B) ND WT with consensus DNA, C) DBD with scrambled DNA, D) ND WT with scrambled DNA.

(163). The nonlinear plots of DBD and ND bound to consensus DNA suggest a non-negligible heat capacity change upon binding.

Table 2.6: Van't Hoff plot-derived thermodynamics. All values are in kcal/mol and calculated using data collected at 145 mM IS.

Consensus DNA						
	ΔH	**ΔS**	**TΔS**	**ΔH-TΔS**	**Actual ΔG**	**Entropic %**
DBD	-5.252	0.022	6.448	-11.700	-11.838	55
ND WT	-2.843	0.021	6.198	-9.040	-9.047	69
Scrambled DNA						
	ΔH	**ΔS**	**TΔS**	**ΔH-TΔS**	**Actual ΔG**	**Entropic %**
DBD	-6.401	0.008	2.456	-8.856	-8.870	28
ND WT	-4.371	0.012	3.454	-7.825	-7.824	44

This suggests that for p53 the salt-dependent component of binding is not entirely the entropic contribution from ion release. According to the Record model, the salt-dependent and independent contributions to binding free energy for DBD are predicted to be −9.30 kcal/mol and −2.77 kcal/mol, respectively, and for ND WT they are −7.55 kcal/mol and −1.50 kcal/mol, respectively. For all the fragments except ND NP, a smaller contribution for binding to scrambled DNA comes from the salt-dependent component. For DBD, the salt-dependent and independent components of binding to scrambled DNA are −5.14 kcal/mol and −4.32 kcal/mol, respectively, and for ND WT are −5.22 kcal/mol and −3.28 kcal/mol, respectively. An analysis of these components using Manning's model, Equation 6, shown in Figure 2.8C-D and Table 2.5 also predicts that salt-dependent entropy is a larger component of binding to consensus DNA than to scrambled DNA.

Salt-dependent ion release is one of several mechanisms that proteins use to achieve specificity in DNA binding. Studies have characterized systems in which the salt-dependent component of binding is higher for specific than nonspecific DNA binding (164), in which the salt-dependent component is similar for specific and nonspecific DNA binding (13, 165), in which the salt-dependent component is lower for specific than for nonspecific DNA binding (166, 167), in which the salt-dependent component is relatively low for both specific and nonspecific binding (156, 158, 168), and in which the salt-dependent component follows no clear trend between specific and nonspecific DNA binding (169, 170). It appears that our p53 fragments utilize salt-dependent components of the interaction for specific binding to a greater degree than the salt-independent components, and this trend is reversed for nonspecific DNA. Our mutants also follow this trend, with the exception of ND NP, which may switch between two modes depending on the IS.

In summary, using the salt-dependent component of binding, we find that predicted excess ion release upon protein-DNA binding is greater when our p53 fragments binding consensus DNA than scrambled DNA. Whereas excess ion release varies by fragment when binding consensus DNA, it is similar between all fragments when binding scrambled DNA excepting ND NP. This salt-dependent component comprises a variable amount of the free energy of binding for each fragment and generally comprises a greater amount of the free energy of binding for consensus DNA than scrambled DNA.

The TAD2-DBD interaction affects Stokes Radius

Using size exclusion chromatography (SEC) at high IS (410 mM), the elution volumes of p53 fragments were compared to elution volumes of known standards (see methods) to determine their Stokes radii and apparent and actual molecular weights (MW). As shown in Figure 2.10, TAD2 mutants elute at a lower volume than ND WT, which elutes at a lower volume than DBD. As shown in Table 2.7 we find the Stokes radius of DBD to be 2.642 ± 0.040 nm, similar to a previously published Stokes radius of the same DBD fragment using dynamic light scattering (2.74 nm) (171), whereas the radius of ND WT was found to be 3.284 ± 0.051 nm. The change in radius with the tethered TAD is relatively small given that p53 residues 1–93, including TAD1, TAD2, and PRR, has a Stokes radius of 3.2 nm at 5°C (172). ND WT appears to be more

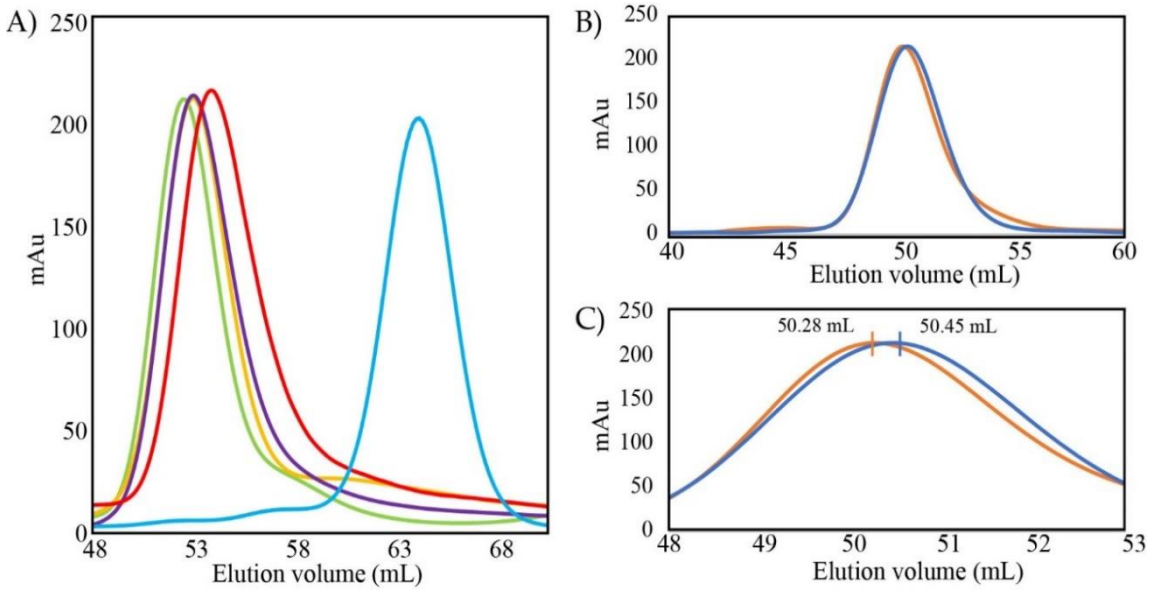


Figure 2.10. Size exclusion chromatography is used to compare p53 fragments. A) Elution profiles of p53 constructs where lower elution volume indicates a larger Stokes radius: — DBD, — ND DE, — ND NP, — ND QS, — ND WT; B) Elution profiles of ND WT at 150 mM IS in orange and 410 mM in dark blue; C) Comparison of elution volumes of ND WT at 150 mM IS in orange and 410 mM in dark blue.

compact than predicted for 93 disordered residues of the TAD attached to 219 ordered residues of the DBD, but ND WT is more expanded than most folded proteins of the same number of residues (Table A4) (173). We expect the intramolecular interaction to cause compaction of the protein as TAD folds over and interacts with DBD. Estimating the Stokes radius of a protein containing both ordered and disordered sections is an ongoing challenge (173, 174). Both DBD and ND WT have an apparent molecular weight greater than their actual molecular weight, as shown in Table 2.7. For DBD this is likely due to a disordered segment near the C-terminus from residues 292–312 (PDB **4HJE**) (52). ND WT and the ND mutants have apparent molecular weights almost twice as large as their actual molecular weights using this technique. Note, these experiments were conducted in 410 mM IS to ensure stability. Shown in Figure 2.10B-C, ND WT's elution volume varies between these two conditions by <0.2 mL, a difference that corresponds to an approximately 0.03 nm difference in Stokes radius and less than 1 kDa difference in apparent molecular weight.

Table 2.7: SEC analysis of DBD, ND WT and ND TAD2 mutants

	Stokes radius (nm)	**Elution volume (mL)**	**Apparent MW (kDa)**	**Actual MW (kDa)**
DBD	2.642 ± 0.040	63.777 ± 0.054	34.603 ± 0.112	24.548
ND DE	3.387 ± 0.053	52.411 ± 0.012	67.825 ± 0.055	34.229
ND NP	3.351 ± 0.053	52.896 ± 0.037	65.913 ± 0.143	34.127
ND QS	3.351 ± 0.053	52.895 ± 0.024	65.915 ± 0.142	34.447
ND WT	3.284 ± 0.051	53.830 ± 0.043	62.624 ± 0.160	34.565

We observe a small decrease in the elution volume of the ND mutants relative to ND WT, but it is larger than the resolution error of the volume measurement (+/−0.02 mL). Small changes in Stokes radii are evidence the mutants do not disrupt the global

structure of ND, which was unexpected given the increase in DNA binding affinity of the mutants relative to ND WT. The smaller elution volumes, and thus slightly larger radii, may be indicative of a less compacted molecule where TAD interacts with DBD less stably.

There are several formulas available to estimate Stokes radius based on sequence. Wilkins and Smith (173) find a good correlation between peptide length and Stokes radius for globular proteins using the equation $R_S = (4.75 \pm 1.11)N^{0.29\pm0.02}$, where N is the number of residues, and R_S is in angstroms. As shown in Table A4, this method predicts an R_S of 1.56 – 3.12 nm for DBD, the range of which overlaps the experimentally determined value of 2.64 nm. The range predicted for ND WT, 1.72 – 3.48 nm, does overlap with the high end of its experimentally determined value, demonstrating that this method is best suited for use with globular proteins. Marsh (174) developed a model to predict R_S based on peptide length as well that takes into account the folded or unfolded nature of the protein, where a folded protein's radius can be estimated by $R_S = 4.92N^{0.285}$, and the radius of an IDP can be estimated by $R_S = 2.49N^{0.509}$, where N is the number of residues. When investigating MeCP2, Ortega-Alarcon et al. noted that the experimentally determined radius for their mixed ordered and disordered system was larger than predicted by any method, suggesting there may be unknown elements that contribute to the radius of a mixed system (175). This challenge exists for p53, as well, where attempts to estimate radius by the methods described or by adding the predicted values of the ordered and disordered segments together gives ranges of radii that often do not overlap with the experimental values, shown in Table A4.

CHAPTER THREE: CONFORMATIONAL PREFERENCES IMPOSED BY W91 AND THE PROLINE RICH REGION REGULATE AUTOINHIBTION OF DNA BINDING FOR THE P53 TUMOR SUPPRESSOR

Rationale

The function of the PRR in the global functioning of p53 is poorly understood, and several deletion and functional studies of the domain have given contradictory results, as described in Chapter 1. Negative effects of full or partial PRR deletions on apoptosis and transactivation seen in cells may be due to changes in interactions with protein partners or changes in the structure and dynamics of PRR (108, 176). We propose a common mechanism to explain the gene expression and cell fate studies that depends on conformational restrictions imposed by the PRR that depend on proline content to increase the persistence length and a known pi-cation interaction between PRR residue W91 and DBA residue 174. As shown in Figure 3.1A-B, the PRR of p53 is a region of low sequence complexity, composed mostly of proline and alanine residues and a small number of aromatic and charged residues.

Structural studies have suggested the PRR is a region that is locally dynamic but globally rigid and it projects TAD1/2 away from DBD. Deletion of the PRR is shown in biophysical assays to increase autoinhibition of DNA binding. Despite this, PRR also meaningfully interacts with DBD. Wells et al. used NMR, small angle x-ray scattering and Flexible Meccano simulations to simulate an ensemble of possible structures for N-

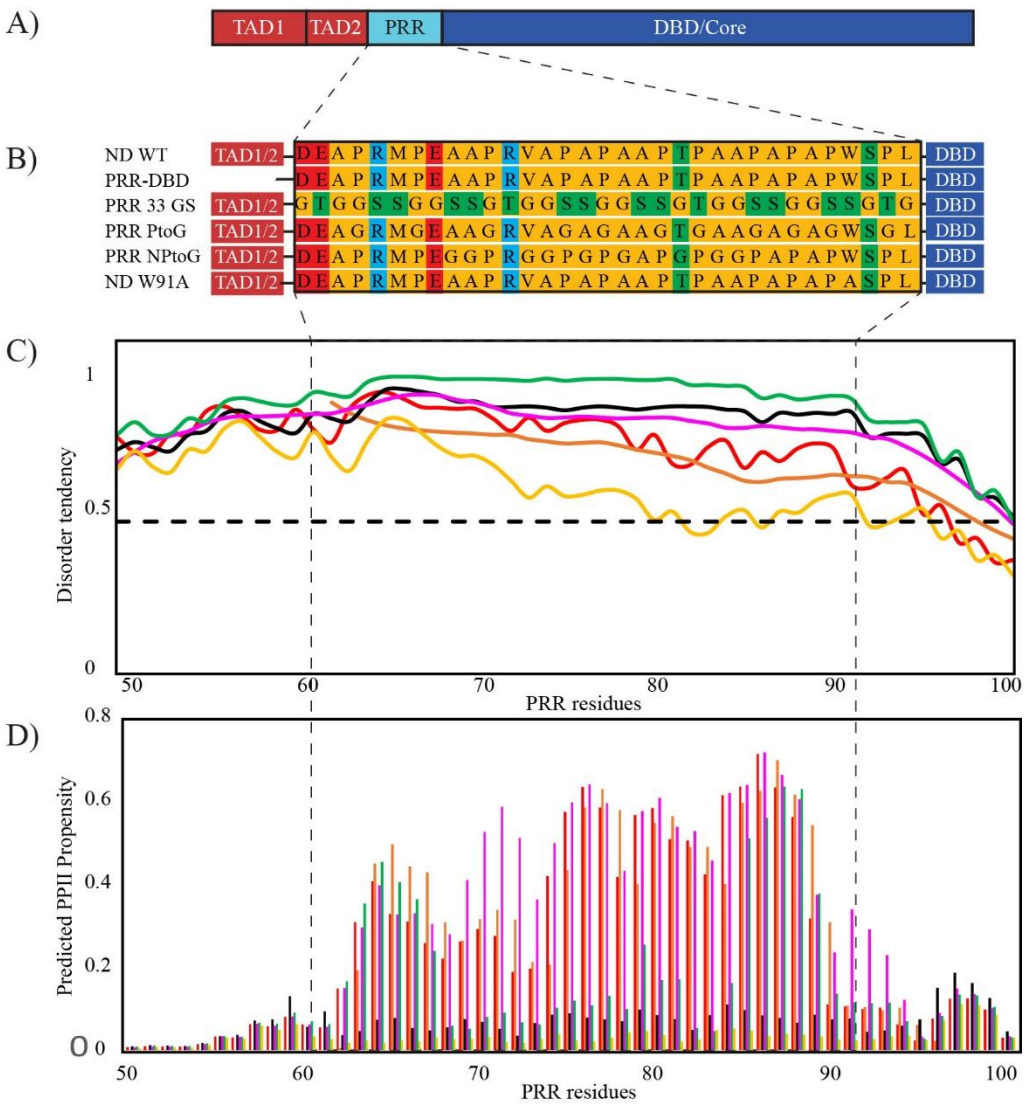


Figure 3.1: PRR mutants described by sequence and features. A) A domain map of ND WT, B) PRR sequences of mutants used in this study. Note, the PRR-DBD fragment has the same PRR sequence as the WT, C) IUPRED plot of PRR mutants, D) Polyproline II helical propensity is predicted by PPIIPRED where — is ND WT, — is PRR-DBD, — is PRR 33 GS, — is PRR PtoG — PRR NPtoG, and — is ND W91A.

and C-terminal disordered regions of p53 and found that TAD residues 60-91 have a high fraction of PPII helix (108). A subset of generated ensembles fit with SAXS data showed TAD2 projecting away from DNA-bound DBD and occupying a large region of space around the tetrameric core. Notably, this study describes the TAD in full length

p53 as more dynamic in the presence of DNA than without, providing early evidence for the intramolecular interaction and a demonstration of the difficulty in modeling it.

Natan et al. crystallized an extended DBD and found an interaction between residues W91 of PRR and R174 of DBD stabilizes DBD, preventing aggregation (shown in Figure 1.3, PDB **2XWR**) (27). It is notable that this study found no significant chemical shift changes in DBD secondary to mutations to the W53/F54 residues of TAD2, which is likely due to the high salt concentrations at which their experiments were conducted. This study found that the folding of the protein at the W91-R174 site increased the thermal stability of DBD and also increased the rate of dissociation of tetrameric DBD; however, the mechanism of this change was not investigated, and the interaction was not considered in the context of DNA interaction.

NMR using labeled methionines showed that an extended p53 DBD including PRR residues 89-93 resulted in a slowed exchange rate for M160 and M169 in DBD (177). The authors describe two sets of cross-peaks that emerge from the extended DBD as open and closed, noting that the open state peaks for M160 and M169 were only 15% as intense as the closed state. Additionally, they noted that the addition of DNA created shifts in M169 similar to those of the “open” conformation, suggesting that W91 affects DNA binding, although the question of how was not pursued further. A W91A mutant decreased the closed state cross-peaks.

A paramagnetic relaxation enhancement study showed the N-terminal region of PRR is in close proximity to the DNA binding pocket of DBD (75), supporting our earlier findings that the PRR interacts with DBD. A deletion mutant of PRR in this study found an increase in autoinhibition of DNA binding, suggesting the PRR-DBD and TAD2-DBD

interactions oppose one another. It is notable, however, that this study replaces the PRR sequence with a 7-residue linker of the PRR and does not investigate specific features of PRR.

The basis of the PRR-DBD interaction remains unclear and its role in frustration of the TAD2-DBD interaction is not fully characterized. Taken together with our lab's recent study, we hypothesized that PRR engages in DNA binding autoinhibition by a complex series of mechanisms based on its sequence and sequence features. The PRR-DBD interaction may be mediated by only the W91-R174 interaction, by multiple weak interactions between prolines and the surface of DBD (178), by an interaction interface formed by the PPII structure of the PPII helix (179), or by specific interactions of projecting R groups that are independent of the PPII helix (74). To address these questions, we generated a series of mutants designed to isolate contributions from features of PRR. Because proline-rich regions and PPII helices tend to engage in multivalent interactions, we expected small scale mutations to have little effect (81, 92). Figure 3.1B shows wildtype and mutant sequences. We generated a mutant that deletes TAD1 and TAD2 entirely to evaluate the autoinhibitory effect of PRR alone on DNA binding (PRR-DBD) as well as a mutant that replaces PRR with a nonreactive linker (PRR 33 GS) (180, 181) of the same number of residues. We also made a mutant that decreases the predicted PPII helical propensity of PRR, assessed via PPIIPred (182), by substituting glycines for all prolines in PRR (PRR PtoG). Glycines were chosen because many of the largest shifts in our previous NMR data came from alanines and because of the high propensity of alanines for PPII helix formation (92, 178). We made a mutant that mutates to glycine the potential nonproline specific

interactions based on those residues that had large chemical shifts in our initial data set (74). These are mostly nonpolar residues, and the mutant is referred to as PRR NptoG. Lastly, we targeted only the W91-R174 interaction with the ND W91A mutant. As with TAD2 mutants, the effect of mutations on the intramolecular interaction is assessed using DNA binding affinity obtained via FA as a proxy. The effect on Stokes radius was assessed with analytical SEC.

We find that the PRR alone participates in autoinhibition of DNA binding in the absence of TAD. Substitution of PRR for a 33 GS linker is thought to yield only autoinhibition provided by TAD2-DBD, yet the total autoinhibition of the PRR-DBD mutant and the PRR 33 GS mutants is greater than that seen in ND WT, suggesting a frustrated set of ensembles. Frustration appears to be mostly mediated by prolines as substitution of prolines for glycines (PRR PtoG) results in greater autoinhibition than is seen in the WT. Elimination of potential specific interactions (PRR NptoG) result in binding affinity intermediate between that of ND WT and PRR PtoG, which may be indicative of an increase in PRR flexibility or of an effect on the interaction of PRR with DBD. The ablation of the W91-R174 interaction yields the largest decrease in DNA binding autoinhibition relative to ND WT of the PRR mutants, suggesting that the modulation of the effective concentration by this interaction is important for function and also that other potential specific interactions in PRR-DBD have little effect on PRR-DBD interaction.

Proline rich regions and polyproline II helices

Proline residues are unique among natural protein amino acids in that their R-group connects to the backbone amine, resulting in a relatively high frequency of *cis* isomers among proline, increased rigidity that extends to nearby residues, and an extended hydrophobic interface (183, 184). Proline rich regions are common in globular proteins (106, 185) and within transcription factors more specifically such as in the erythroid Kruppel-like factor, Smad4, and interferon regulatory factor 3 (186-188).

Regions containing many prolines often adopt the dihedral backbone angles of proline, resulting in PPII helical conformation, which leads to a triangular prism conformation where every third amino acid aligns (189). This conformation creates an interface commonly recognized by SH3 domains, suggesting a role for these regions in responding to signaling (190, 191). The PRR of p53 has been shown to have high polyproline II helical propensity (107-109) and is known to interact with several proteins that contain SH3 domains, although it is not known if the SH3 domains themselves are direct contacts (113). PPII helices can convert to the more compact right-handed polyproline I helices, which are less rigid, based on environmental conditions (192).

PRR mutations affect autoinhibition in p53 DNA binding

Figure 3.2 shows FA binding curves of DBD, ND WT, and PRR mutants interacting with consensus and scrambled DNA with ΔG values at physiological ionic strength (145 mM) in order of decreasing binding affinity to consensus DNA. As with the TAD2 mutants in Chapter 2, all PRR mutants have a Hill coefficient at all ionic strengths

of ~1.8 when binding consensus DNA and a Hill coefficient of 1 when binding scrambled DNA, which can be seen in the different shapes of the fit lines in Figure 3.2A-B.

DBD binds both consensus and scrambled DNA with greater affinity than any other fragment, and binding to consensus DNA varies more between fragments than does binding to scrambled DNA. The ND W91A, PRR-DBD, and PRR 33 GS mutants increase DNA binding affinity relative to ND WT, suggesting a decrease in autoinhibition,

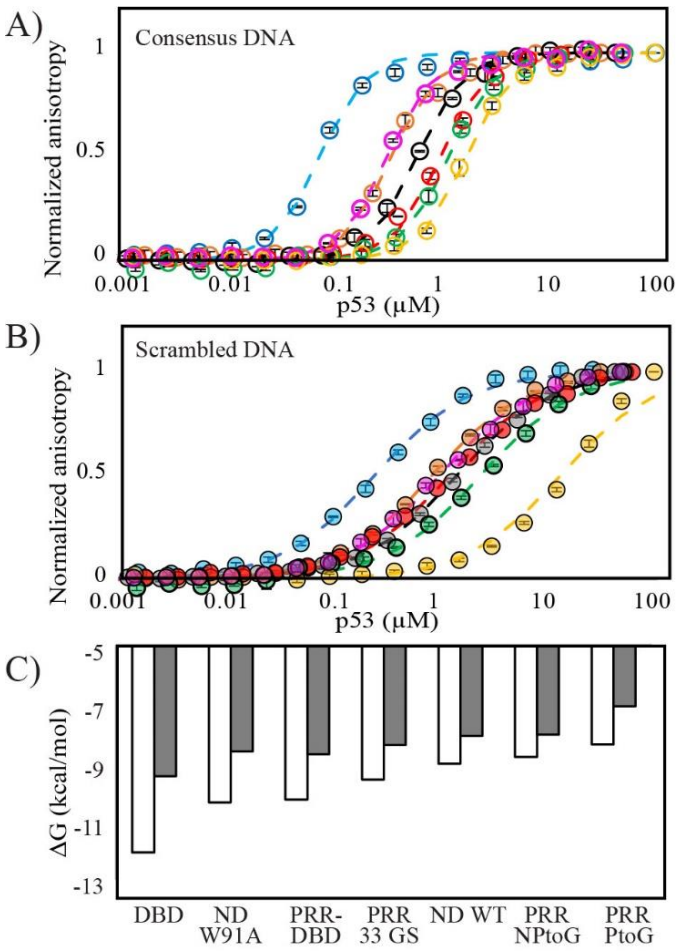


Figure 3.2: Binding of DBD, ND WT, and PRR mutants to consensus and scrambled DNA at physiological ionic strength. Fluorescence anisotropy curves where circles represent data points and dotted lines are fit lines for A) consensus DNA, B) scrambled DNA. C) ΔG values for fragments in kcal/mol where unfilled columns are consensus and grey columns are scrambled DNA. Blue is DBD, orange is PRR-DBD, black is PRR 33 GS, yellow is PRR PtoG green is PRR NPtoG, pink is ND W91A, and red is ND WT.

whereas the PRR PtoG and PRR NptoG mutants decrease DNA binding affinity relative to WT, suggesting an increase in autoinhibition. All mutants were able to saturate the consensus DNA; however, the PRR PtoG mutant, which has the weakest binding of any fragment shown here, did not reach full saturation of scrambled DNA even at a concentration of ~100 μM. Therefore, the ΔG and K_D values of PRR PtoG binding scrambled DNA at 145 mM are low-end estimates, shown in Table A1 and A2. ΔG values for each fragment, Figure 3.2C, shows that the order of decrease in binding affinity is similar for consensus and scrambled DNA.

PRR-DBD and TAD2-DBD interactions are frustrated

Table 3.1 shows the ΔΔG values of fragments relative to each other when binding either consensus or scrambled DNA. The ΔΔG value of any fragment compared to DBD can be thought of as quantified autoinhibition; inversely, the ΔΔG value of a fragment compared to ND WT represents a change in autoinhibition, where a negative value represents an increase in autoinhibition, and a positive value represents a decrease in autoinhibition.

Table 3.1: ΔΔG comparison of fragments at physiological ionic strength
Consensus DNA

	DBD	ND W91A	PRR-DBD	PRR 33 GS	ND WT	PRR NptoG	PRR PtoG
DBD	0	1.70	1.80	2.45	3.02	3.25	3.68
ND W91A	-1.70	0	0.09	0.74	1.32	1.55	1.98
PRR-DBD	-1.80	-0.09	0	0.65	1.22	1.45	1.89
PRR 33 GS	-2.45	-0.74	-0.65	0	0.58	0.81	1.24
ND WT	-3.02	-1.32	-1.22	-0.58	0	0.23	0.66
PRR NptoG	-3.25	-1.55	-1.45	-0.81	-0.23	0	0.43
PRR PtoG	-3.68	-1.98	-1.89	-1.24	-0.66	-0.43	0

Table 3.1 (Continued)

	Scrambled DNA						
	DBD	**ND W91A**	**PRR-DBD**	**PRR 33 GS**	**ND WT**	**PRR NptoG**	**PRR PtoG**
DBD	0	0.84	0.74	0.91	0.97	1.41	2.38
ND W91A	-0.84	0	-0.10	0.07	0.13	0.58	1.54
PRR-DBD	-0.74	0.10	0	0.17	0.22	0.67	1.63
PRR 33 GS	-0.91	-0.07	-0.17	0	0.05	0.50	1.46
ND WT	-0.97	-0.13	-0.22	-0.05	0	0.45	1.41
PRR NptoG	-1.41	-0.58	-0.67	-0.50	-0.45	0	0.96
PRR PtoG	-2.38	-1.54	-1.63	-1.46	-1.41	-0.96	0

Notes on Table 3.1: The ΔΔG is the ΔG of the DNA binding of the top of the column protein minus the row header protein. All values are in kcal/mol at 145 mM IS.

The ΔΔG values of PRR-DBD and PRR 33 GS compared to DBD show that the PRR and TAD2 contribute 1.80 and 2.45 kcal/mol of autoinhibition, respectively. For the PRR-DBD mutant, which deletes TAD1 and TAD2, all autoinhibition must originate in the PRR. The autoinhibition from PRR 33 GS we assume to represent only that of the TAD2-DBD interaction as all sequence features of PRR are substituted with a nonreactive linker while retaining an equal number of residues to separate TAD2 and DBD. The combined autoinhibition of these two mutants then should represent the total autoinhibition. If their activity were only additive, we would expect that when both domains are intact, as in ND WT, the total autoinhibition of consensus DNA binding should be 4.25 kcal/mol. However, the autoinhibitory effect of ND WT (ΔG_{DBD} — $\Delta G_{ND\ WT}$) is only 3.02 kcal/mol, suggesting that there is frustration in this system, where two components within a protein have effects that counteract or compete (193). To compare

this effect in scrambled DNA, the PRR and TAD2 autoinhibitory contributions, described by ΔΔG of PRR-DBD and PRR 33 GS relative to DBD, respectively, are 0.74 and 0.91 kcal/mol for a total expected autoinhibition of 1.65 kcal/mol. Again, the total predicted autoinhibition of these two intact domains is greater than what is seen in ND WT (-0.97 kcal/mol), suggesting frustration. Despite the evidence of frustration, ND WT demonstrates greater autoinhibition than the PRR-DBD or PRR 33 GS mutant alone, suggesting the autoinhibitory effects of the two domains are not mutually exclusive.

Sequence and structure of PRR contribute to frustrated autoinhibition

Two of the PRR mutants shown in Figure 3.2 decrease DNA binding affinity relative to ND WT: PRR PtoG and PRR NptoG. We assume that the change is due to modulation of the intramolecular interaction and, more specifically, to a decrease in the frustration between PRR and TAD2.

PRR PtoG has the greatest autoinhibition of DNA binding among our mutants shown here. This suggests to us that frustration is modulated by the prolines of the PRR. The PPII propensity is predicted to be greatly decreased in this mutant; thus, the increased flexibility of the linker likely allows TAD2 to have greater access to DBD.

PRR NptoG demonstrates autoinhibition that is slightly greater than that seen for ND WT (3.25 kcal versus 3.02 kcal/mol for consensus DNA and -1.41 kcal/mol versus -0.97 kcal/mol for scrambled DNA, relative to DBD). This mutant is also expected to have lower PPII propensity due to its lowered alanine content and thus greater flexibility than the WT sequence despite maintaining high proline content (Figure 3.1D). The increase in autoinhibition with these mutants may be due to: 1) an increase in TAD2

access to DBD due to increased PRR flexibility, 2) tighter binding of the PRR to DBD, or 3) removal of a PRR-DBD interaction site that competes with a TAD2-DBD interaction site. There is a clear correlation between predicted flexibility of the PRR and autoinhibition, where ND WT is more rigid than PRR NptoG, which is more rigid than PRR PtoG, and we suspect that this is the defining feature of increased autoinhibition rather than a change in direct PRR-DBD interaction sites. The total predicted autoinhibition of intact PRR and TAD2, represented by PRR-DBD plus PRR 33 GS, is still greater than that seen in PRR PtoG or PRR NptoG when looking at consensus DNA, suggesting an additional feature that contributes to frustration. By contrast, the autoinhibition of PRR PtoG binding to scrambled DNA is greater than the predicted total, and the meaning of this remains unclear.

The ND W91A mutant results in the greatest decrease in autoinhibition relative to ND WT seen in any of the PRR mutants when binding consensus DNA. Previous studies suggested that the W91-R174 interaction modulated an open and closed state of p53 (27). We hypothesized this open and closed state might affect the TAD2-DBD interaction although R174 is not immediately adjacent to the DNA binding pocket. Our FA data shows the ND W91A mutant decreases DNA binding affinity by -1.70 kcal/mol relative to DBD, a greater restoration of DNA binding than is seen in PRR-DBD for binding consensus DNA, where TAD2 is deleted entirely. Our explanation for this is that the W91-R174 interaction folds the PRR back onto DBD, orienting TAD2 towards DBD even as the rigidity of the PRR decreases TAD2 access to DBD. If this were true, the W91-R174 interaction would increase the ability of TAD2 to bind DBD.

Figure 3.3 shows a model of how PRR mutants affect the flexibility of PRR and the subsequent orientation of TAD2 with K_D values of each fragment binding consensus DNA at 145 mM IS shown. In this model, the ND W91A mutant disrupts the folding back of the PRR towards the DNA binding pocket with the result that TAD2 is now free to sample a much larger volume of space around DBD. The resulting effective concentration is lowered. PRR 33 GS also allows TAD2 to occupy a larger potential volume around DBD because the W91 site is mutated to glycine. However, unlike the ND W91A mutant, the PRR 33 GS mutant is more flexible, and the average occupied space around DBD is smaller due to the shorter average pitch of amino acids in a flexible linker compared to those in a linker in PPII conformation (181, 194).

If we examine the PRR mutant data with that obtained from TAD2 mutants described in Chapter 2, we see the autoinhibition of ND DE and ND NP, which together are expected to equal the total autoinhibition provided by TAD2, adds together for a total of 2.23 kcal/mol. This suggests that these two components come very close to the total autoinhibition of the TAD2-DBD interaction, which is represented here by PRR 33 GS with total autoinhibition of 2.45 kcal/mol.

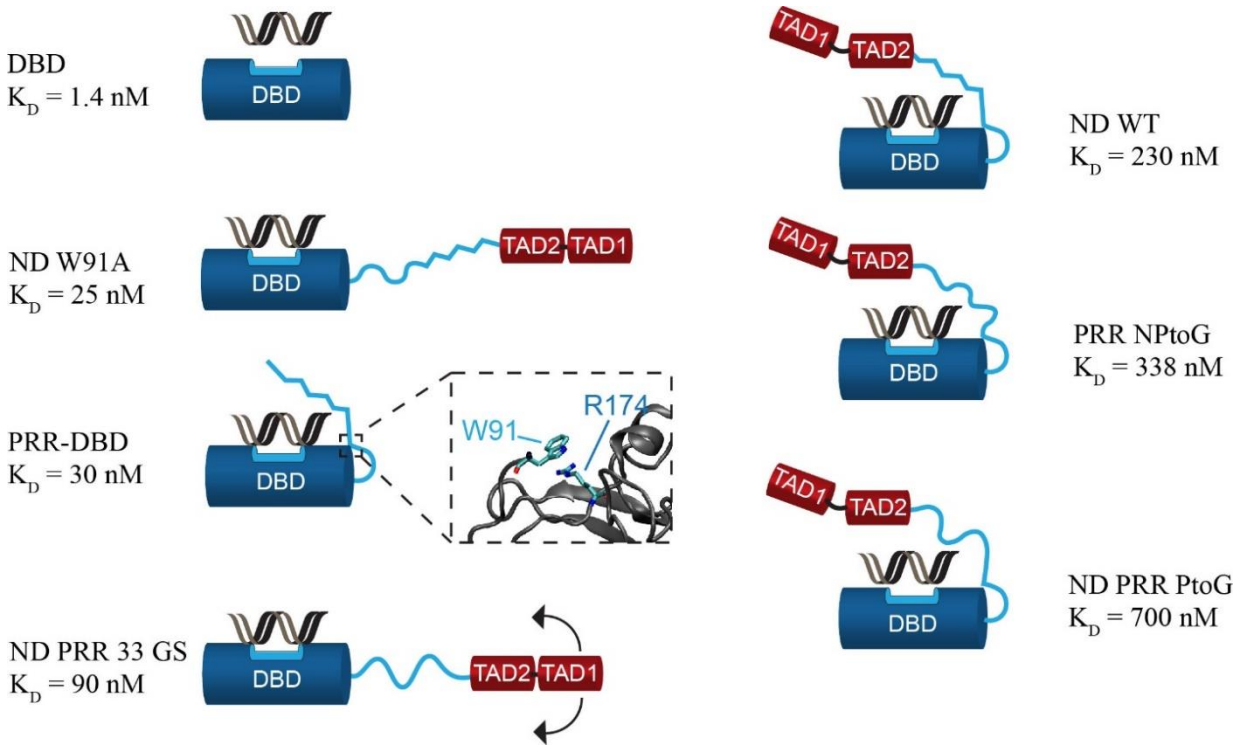


Figure 3.3: PRR mutants affect DBD and TAD2-DBD interaction. Cartoons of PRR, shown in light blue, demonstrate a proposed model where PRR mutants affect the orientation and flexibility of PRR. The location of TAD1 and TAD2 are secondarily affected. K_Ds of each fragment's binding to consensus DNA are shown beside models.

An unexpected aspect of our results, however, comes from comparing TAD2 mutants and PRR mutants. The PRR-DBD mutant autoinhibits consensus DNA binding by 1.80 kcal/mol relative to DBD. The TAD2 mutants ND DE, ND NP, and ND QS autoinhibit consensus DNA binding by 1.13, 1.03, and 1.49 kcal/mol, respectively, relative to DBD. This means fragments of p53 with a partially intact TAD2 inhibit DNA binding less than fragments that lack TAD2 entirely; the addition of a mutated TAD2 unexpectedly decreases autoinhibition. This is only true for the disrupted TAD2-DBD interactions and not for the WT, but it suggests to us that TAD2 may also contribute to the frustration of the PRR-DBD interaction, possibly by allosterically interfering with the W91-R174 interaction.

Taken together, these observations suggest that the closed state of ND, previously visualized as a single collapsed, dynamic state, is likely to be composed of substates in which either the TAD2-DBD interaction or the PRR-DBD interaction is the dominant interaction of the closed state, as shown in Figure 1.2.

PRR mutants affect specificity and apparent ion release

As with TAD2 mutants discussed in Chapter 2, an increase in buffer IS leads to a decrease in DNA binding affinity. The change in binding affinity to consensus and scrambled DNA was measured from 125 mM to 165 mM IS, shown as a double log plot in Figure 3.4. Relatively weak binding was observed for PRR PtoG; binding of PRR PtoG to scrambled DNA did not reach saturation even at a concentration of ~100 μM. For this reason, we did not investigate the high IS binding of any PRR mutants.

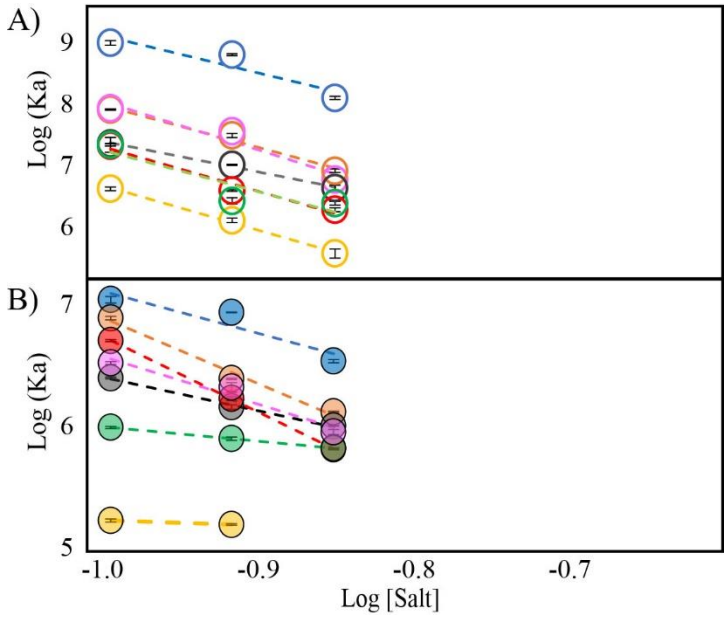


Figure 3.4: Salt-dependent DNA binding affinity of DBD, ND WT, and PRR mutants. Double log plots of fluorescence anisotropy data plotting of Log (K_A) vs Log [Salt] from 125-165mM IS of DBD (blue), ND W91A (pink), PRR-DBD (orange), (gray) PRR 33 GS, PRR PtoG (yellow), PRR NPtoG (green), and ND WT (red) Where A) represents binding to consensus DNA, and B) represents binding to scrambled DNA.

Figure 3.4 shows that DBD demonstrates the tightest binding of any fragment at all IS, and PRR PtoG demonstrates the lowest. For consensus DNA, the binding affinity of each fragment follows the pattern present at 145 mM IS. However, as shown in Table 3.2, the response to salt concentration is slightly varied by fragment, resulting in differing slopes. PRR 33 GS has a smaller slope than the other fragments, suggesting a lack of responsiveness to environmental salt concentration. By contrast the slopes PRR mutants binding to scrambled DNA show divergent behavior, where PRR 33 GS and PRR NPtoG, especially, show less response to salt concentration.

Table 3.2: Estimated excess ion release in PRR mutants

	Consensus DNA		**Scrambled DNA**	
	-Slope, N	**Excess ion release**	**-Slope, N**	**Excess ion release**
DBD	6.02	2.5	3.38	0
ND W91A	7.85	4.4	3.84	0.3
PRR-DBD	6.77	3.3	5.31	1.8
PRR 33 GS	4.94	1.4	2.70	0
ND WT	7.18	3.7	6.08	2.6
PRR NptoG	6.66	3.2	2.87	0
PRR PtoG	7.18	3.7	NA	NA

Notes on Table: due to lack of saturation above 125 mM IS, the slope of ND PRR PtoG binding scrambled DNA could not be evaluated.

The slopes listed here for DBD and ND WT are different for the 125 – 165 mM plots than for the 125 – 225 mM plots in Chapter Two. For DBD, the slope and apparent ion release decreases for both consensus and scrambled DNA, which is likely due to the plateauing of the log(K_A) at low ionic strengths. This plateau does not occur for ND WT at 125 mM, presumably because its binding affinity has a lower threshold not yet

reached. Instead, the narrower range of salt concentrations results in a larger double lot plot slope for ND WT, which indicates a greater apparent ion release that is especially apparent for scrambled DNA. This presents a reverse situation from that described in Chapter 2, where DBD a greater predicted ion release when binding consensus DNA and an equal predicted ion release as ND WT when binding scrambled DNA. From this, we conclude that this narrow range of salt concentrations may not be sufficient for establishing slope for use with the counterion condensation theory. Despite this, comparison of PRR fragments across IS gives us some insight into mutant behavior.

For all fragments shown here, the slope generated from binding consensus DNA is greater than that from scrambled DNA, suggesting a greater component in binding affinity coming from salt-dependent features in binding consensus DNA than in binding scrambled DNA. Mutants that change the flexibility of PRR – PRR 33 GS, PRR NptoG, and PRR PtoG – are predicted to release 3.3, 3.2, and 3.7 excess ions, respectively, when binding consensus DNA. Those that disrupt the W91-R174 interaction – ND W91A and PRR 33 GS – are predicted to release 4.4 and 1.4 excess ions, respectively. A trend is not apparent. However, we note that PRR 33 GS and PRR NptoG appear relatively insensitive to salt concentration, with the smallest slopes for both consensus and scrambled DNA such that they cross over other mutants at 165 mM IS.

The PRR-DBD interaction affects Stokes radius

Analytical SEC was used to evaluate the R_S of PRR mutants. As shown in Figure 3.5, ND WT elutes at the lowest volume and thus has the largest R_S of all fragments shown, DBD elutes at the highest volume and thus has the smallest R_S of all fragments

shown here, and PRR-DBD is intermediate both in molecular size and elution volume. In Chapter 2, mutations to TAD2 resulted in lower elution volumes and larger R_S than ND WT, suggesting an average lesser occupancy of the closed state than seen for ND WT. By contrast, among the PRR mutants, PRR 33 GS, PRR NptoG, PRR PtoG, and ND W91A have lower elution volumes than ND WT. Table 3.3 quantifies the values of R_S and apparent molecular weight.

Increased occupancy of the closed state is supported by the increase in autoinhibition seen in PRR NptoG and PRR PtoG. However, it is notable that while the PRR NptoG mutant is only slightly more autoinhibited than ND WT, its apparent size

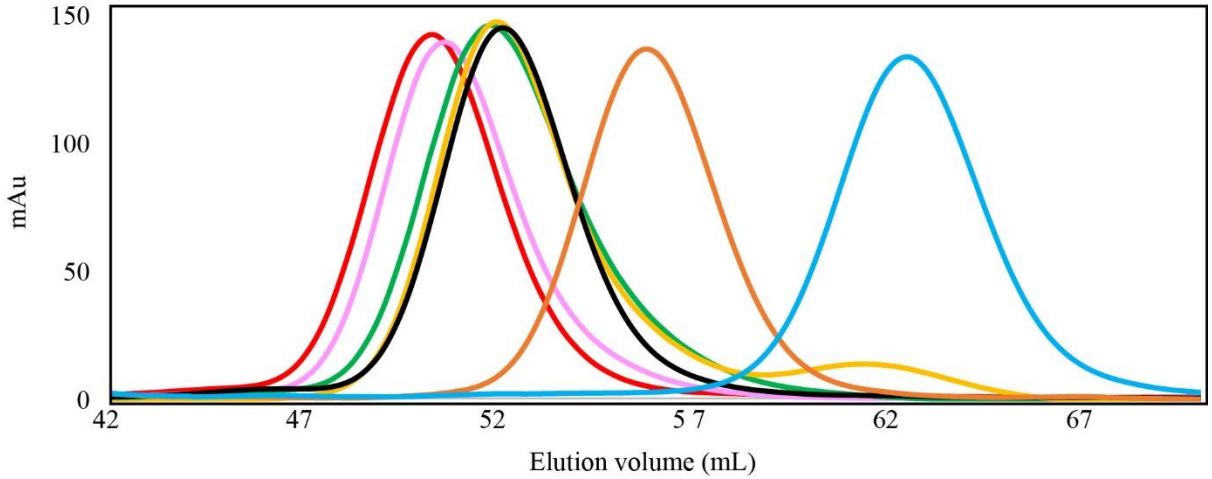


Figure 3.5: Size exclusion chromatography of PRR mutants. Elution profiles of p53 constructs where lower elution volume indicates a larger Stokes radius: Where blue is DBD, orange is PRR-DBD, black is PRR 33GS, yellow is PRR PtoG green is PRR NPtoG, pink is ND W91A, and red is ND WT.

here is very similar to that of PRR PtoG, suggesting that there is not a direct correlation between the closed state occupancy of the apo state under these conditions.

Both PRR 33 GS and ND W91A are less autoinhibited than ND WT and thus are expected to spend less time in the autoinhibited state, yet both appear smaller than ND WT here. It is notable that PRR NptoG, PRR PtoG and PRR 33 GS are grouped closely

together, and all are predicted to significantly change the rigidity of the PRR as shown in Figure 3.1D, making TAD2 more available to bind DBD. It could be that conditions of the SEC experiment favor TAD2-DBD interaction in the apo state. ND W91A, which appears similar in size to ND WT, should have a PRR similar in rigidity to that of ND WT. This suggests to us that the ablation of the W91-R174 interaction seen in the ND W91A mutant favors an orientation of TAD2 that is more accessible to bind DBD, yet the rigidity of the PRR discourages this interaction, resulting in an average closed state occupancy that is only slightly higher than that of ND WT. Taken together, these suggest that the compaction of the molecule when PRR is mutated only mildly correlate with the strength and stability of the TAD2-DBD interaction. Conformational variability of PRR 33 GS and ND W91A may contribute significantly to the ease with which the TAD2-DBD dissociates in the presence of DNA but have only a small effect in the high salt apo state of the size exclusion column.

Table 3.3: SEC analysis of DBD, ND WT, and PRR mutants

	Stokes radius (nm)	**Elution volume (mL)**	**Apparent MW (kDa)**	**Actual MW (kDa)**
DBD	2.612 ± 0.005	62.512 ± 0.069	33.554 ± 0.155	24.548
PRR-DBD	3.082 ± 0.001	55.884 ± 0.011	52.534 ± 0.083	27.798
ND PRR 33 GS	3.375 ± 0.002	52.091 ± 0.157	67.200 ± 0.127	33.643
ND PRR PtoG	3.392 ± 0.001	51.928 ± 0.001	68.114 ± 0.020	34.124
ND PRR NptoG	3.398 ± 0.001	51.860 ± 0.011	68.445 ± 0.046	34.381
ND W91A	3.499 ± 0.006	50.666 ± 0.072	74.118 ± 0.338	34.397
ND WT	3.529 ± 0.020	50.173 ± 0.110	75.853 ± 0.558	34.565

Note on Table 3.3: Elution volumes, apparent MW, and Stokes radii for DBD and ND WT differ from those in Table 2.7 due to the use of a different size exclusion column and set of calibration kit standards.

As described in Chapter 2, R_S values for p53 fragments can be predicted using various systems; however, predictions of a mixed ordered and disordered system are usually larger than experimentally derived radii. Whether this represents a flaw in the prediction system or merely reflects the compaction of the intramolecular interaction is unclear. Supplementary Table A4 shows predicted radii of PRR mutant constructs.

CHAPTER FOUR: AUTOINHIBITION IS MODULATED BY AN EVOLVED LOW EFFECTIVE CONCENTRATION

Rationale

Our work shows that PRR directly inhibits DNA binding by p53 DBD and indirectly reduces inhibition of DNA binding by TAD2. The latter is likely mediated by the rigidity of PRR which was modulated with the mutations described in Chapter 3. However, PRR also serves as a linker that separates TAD2 and DBD to regulate the orientation and effective concentration (C_{eff}) of TAD2.

In this section we investigate the PRR as a linker and use the insertion of inert linkers between either TAD2-PRR or PRR-DBD to assess how changes in C_{eff} and conformational variability affect the intramolecular interaction. We evaluate the C_{eff} of TAD2 based on the length and rigidity of PRR. We also assess the evolutionary conservation of the autoinhibitory mechanism among vertebrates.

PRR as a linker

Disordered linkers are common in transcription factors and serve a variety of functions, such as affecting the orientation of flanking domains and regulating oligomerization and the recruitment of other proteins (26, 105, 195). A recent cryo-EM structure of p53 in complex with the RNA polymerase II complex demonstrates how

TAD1, TAD2, and DBD interact with distinct sites within the complex whose distance is bridged by the PRR, highlighting the importance of adequate spacing between domains for multidomain interactions (196). Linkers that separate domains are often enriched in proline and relatively rigid (92, 197). Despite this, linkers must also maintain a level of flexibility that allows for association of flanking domains when appropriate. Multiple sequence alignments of disordered linkers often show conservation of length, flexibility and sequence composition while showing relatively poor sequence conservation (198, 199), suggesting that linkers can be optimized for specific functions despite low sequence conservation.

Transcription factors often contain a DNA binding domain and a transactivation domain, where the DNA binding domain interacts with DNA and the transactivation domain interacts with cofactors, general transcription factors, or other regulators of gene expression. As such, these domains must be optimally spaced from one another to function independently (195). Thus, the PRR as a linker must balance the optimal spacing of TAD2 and DBD for transactivation and for autoinhibition.

Linker mutants

Glycine and serine-rich linkers are often used in protein design due to their low tendency for secondary structure and biological activity. Threonine can also be included to improve interactions with solvent and decrease interactions with flanking domains (200). We expected a 20-residue linker inserted between residues 93 and 94 of PRR and DBD termed ND P+20, Figure 4.1, would have three effects: 1) to disrupt the W91-R174 interaction, 2) decrease the C_{eff} of TAD2, and 3) increase the conformational

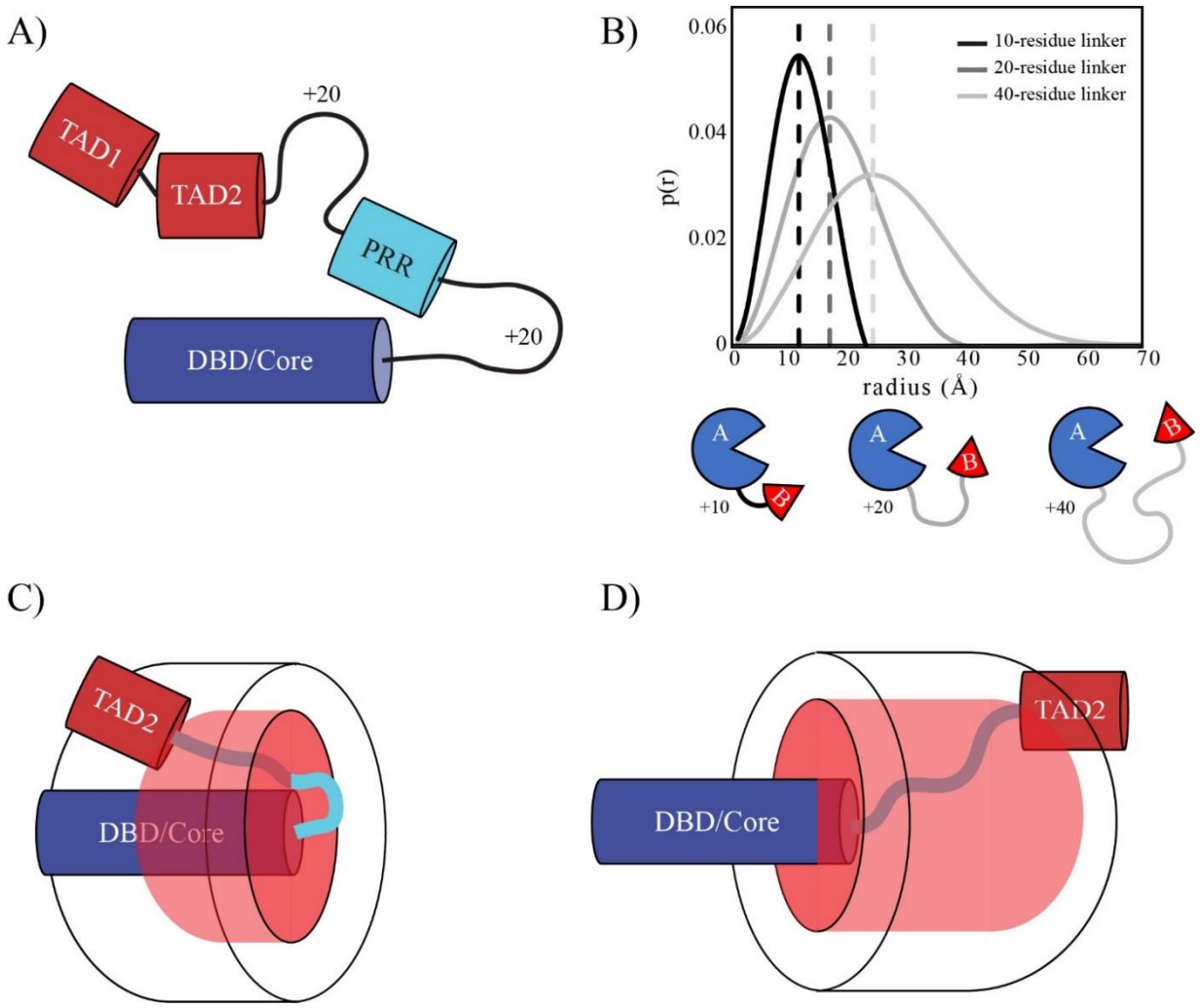


Figure 4.1: Linker mutants and PRR affect global conformation. A) Model of linker insertions between either TAD2 and PRR or between PRR and DBD domains. B) Example of end-to-end distribution of a flexible linker based on the length of linker. C) The space available to be occupied by TAD2 depends on the length and orientation of PRR.

variability of the entire TAD. While disrupting the PRR-DBD interaction is expected to ablate the frustration of the TAD2-DBD interaction, the additional length and flexibility of the TAD was expected to decrease the overall strength of the intramolecular interaction, leading to an increase in DNA binding affinity relative to ND WT. Our experimental results support this model, where ND P+20 bound DNA with an intermediate affinity

between ND WT and DBD. The Stokes radius was approximately the same as ND WT, see Table 4.2 and Figure 4.3.

A linker inserted between residues 60 and 61 of TAD2 and PRR, termed ND T+20, Figure 4.1, was expected to have the following effects: 1) disrupt the frustration provided by the PRR-DBD interaction and 2) decrease the C_{eff}. These two effects have opposite results in the context of DNA binding. A short linker may overcome the frustration of the PRR-DBD interaction, but a long linker is expected to weaken the intramolecular interaction as conformational variety expands. Our results ultimately confirmed that suppression of PRR-based frustration increased the autoinhibitory effect. Stokes radius was similar to that of ND WT and ND P+20.

Wormlike chain model and effective concentration

The wormlike chain model (WLC) is used to predict the end-to-end distance of semiflexible polymers and can approximate the behavior of flexible protein linkers (198, 201, 202). The tethering of two domains by a flexible linker affects their relative orientation. The conservation of such linkers in terms of their length, charge, and flexibility suggests an important role for the linker in maintaining the distance and orientation of the two flanking domains regardless of sequence similarity, as is seen in MdmX (198) and the adenovirus early gene 1A (199).

The WLC produces a probability distribution of the end-to-end distance between a flexible peptide's N- and C-termini, termed end-to-end distance (pI), based on two major variables: the contour length (*Lc*) and persistence length (*Lp*), Figure 4.1B. Contour length is the length of the peptide without bending or stretching and is given as

the number of amino acids multiplied by the average length of a single amino acid, estimated at 3.4 – 4.0 Å per amino acid (203). Here we use the commonly cited figure of 3.8 Å per amino acid (204). Persistence length is a measure of the stiffness of a polymer, where the *Lp* is the smallest distance separating two points where the orientation of one does not affect the orientation of the other. Sequence composition determines *Lp* in protein chains, with random coil chains generally being estimated to have *Lp* of 3-4 Å (204). The *Lp* of PRR, however, is elevated due to the high proline content and has been estimated by FRET and a combination of residual dipolar coupling analysis and flexible Meccano molecular dynamics simulation to be 7 – 8 Å (107, 108). These elements produce the end-to-end distance, which increases slowly with increasing length, see Figure 4.1B.

Our lab has previously shown in ITC studies that neither TAD2 nor PRR is able to inhibit DNA binding by DBD when not tethered to DBD. The weak *in trans* K_D of the TAD2-DBD or PRR-DBD interaction is made physiologically relevant by tethering, where the C_{eff} of TAD is increased. C_{eff} is related to the end-to-end distance, where a short or long end-to-end distance results in higher and lower C_{eff}, respectively. C_{eff} may be optimized or not optimized: a very short linker will frustrate interaction by disallowing appropriate orientation of the two interacting domains as in Mdn1 (205), and a very long linker decreases C_{eff}, often allowing interactions with external binding partners to be competitive as in fatty acid conjugation to small proteins (206), see Figure 4.2.

In our recent review article, we discuss several examples of autoinhibitory IDRs interacting with ordered domains (26). In those examples where the IDR and ordered domain are adjacent or nearly adjacent, C_{eff} is necessarily high, and the trend is for the

in trans binding affinity to be in the high micromolar to low millimolar range. This is seen in the transcription factor MAX with an *in trans* K_D of ~57 μM and an C_{eff} of 198 mM (85). By contrast, autoinhibitory IDRs that are not adjacent to the inhibited ordered domain tend to have *in trans* K_Ds in the micromolar range, as seen with B-Myb (*in* trans K_D of 4.5 μM) (83), Foxo4 (*in trans* K_D of 1.2 μM and C_{eff} of 0.5 mM) (84), UHRF1 (*in trans* KD of 1.6 μM) (86), and MdmX (*in trans* K_D of 8.3 μM and C_{eff} of 2.12 mM) (91).

Polyproline regions are notable for their consistency: studies have frequently used them as molecular rulers where the number of prolines can be used to control the distance separating flanking features (207). We find that the long persistence length of PRR makes it possible for TAD2 to interact with any DBD in the p53 tetramer; however, the C_{eff} is not optimized for any of these in comparison to a more flexible linker of the same number of residues.

Evolutionary evaluation

MSA of TAD1, TAD2 and PRR across vertebrates suggests the intramolecular interaction may have evolved to be relevant in mammals as the sequence of TAD2 is poorly conserved among nonmammalian vertebrates, and PRR only emerges as a stable feature in birds. Whereas major interaction sites in TAD1 are preserved across great evolutionary distance, the short linear motif (SliM) around residues W53/F54 in TAD2 that is suspected to be a major contributor to the intramolecular interaction is conserved only in residue type rather than sequence. The sequence of PRR is also poorly conserved; however, the conservation of features is unexpectedly high, suggesting the emergence of a function in birds. We are not able to detect

coevolutionary coupling between PRR and TAD2. MSA suggests conservation of features in the PRR in mammals and perhaps birds, but not beyond.

Linker insertion mutants recapitulate model of TAD2 and PRR frustration

The insertion of a 20-residue flexible linker between TAD2 and DBD or between PRR and DBD results in divergent effects that agree with model presented in Chapter 3. Because both TAD2 and PRR are intact, changes in autoinhibition are expected to be due to an increase in conformational entropy and changes in the resulting orientation and C_{eff} of TAD2 and PRR. Plots and ΔG values for all fragments shown here were generated using data obtained at 145 mM IS, with data collected at 125 – 165 mM IS, see Table A1 and A2.

ND P+20, which has a 20-residue linker inserted between PRR and DBD, has DNA binding that is intermediate between DBD and ND WT. The decrease in autoinhibition supports a model where the PRR-DBD interaction is disrupted and the TAD2-DBD interaction is lessened by decreasing C_{eff}; the proximity of W91 and R174 is important for their interaction. As with the ND W91A and PRR 33 GS mutants, this data supports the model that the W91-R174 interaction biases TAD2 toward the DBD, increasing its C_{eff} regardless of the high persistence length of PRR.

ND T+20, which has a 20-residue linker inserted between TAD2 and PRR, has weaker DNA binding than ND WT. The increase in autoinhibition agrees with a model where the rigidity of PRR frustrates the TAD2-DBD interaction by controlling TAD2 orientation rather than a model where PRR competes with TAD2 for binding to DBD. A

sufficiently long linker allows TAD2 access to its binding site on DBD while the PRR-DBD interaction remains intact.

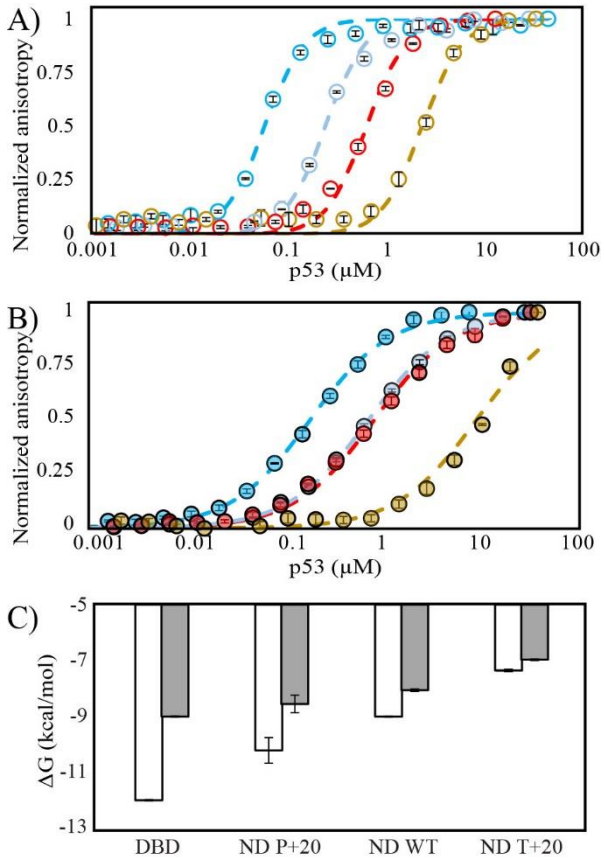


Figure 4.2: Binding of DBD, ND WT and linker mutants to consensus and scrambled DNA at physiological ionic strength. A) Fluorescence anisotropy curves for consensus DNA. B) Fluorescence anisotropy curves for scrambled DNA. C) ΔG values for fragments in kcal/mol.

Table 4.1 shows the quantified autoinhibition or relief of autoinhibition produced by the linker mutants relative to DBD and ND WT. ND P+20 is among the least autoinhibited fragments in this study, with only a 5-fold increase in K_D relative to DBD, Table A1. For comparison, at 145 mM IS the ND DE, ND NP, and ND QS mutants produce approximately 6-, 7-, and 12-fold increases in K_D relative to DBD, respectively. The ND T+20 mutant inhibits DNA binding to a greater degree than any other mutant in this study. As shown in Table 4.1, the quantified autoinhibition of ND T+20 is -4.70

kcal/mol, which is larger than that of the PRR-DBD and PRR 33 GS mutant added together. We suspect that the increase may be due to interactions in other units of DBD within the dimer or tetramer.

Table 4.1: ΔΔG for cooperativity/frustration in linker mutants

Consensus DNA				
	DBD	**ND P+20**	**ND WT**	**ND T+20**
DBD	0.00	1.80	3.02	4.70
ND P+20	-1.80	0.00	1.22	2.89
ND WT	-3.02	-1.22	0.00	1.68
ND T+20	-4.70	-2.89	-1.68	0.00
Scrambled DNA				
	DBD	**ND P+20**	**ND WT**	**ND T+20**
DBD	0.00	0.88	0.97	2.48
ND P+20	-0.88	0.00	0.08	1.60
ND WT	-0.97	-0.08	0.00	1.52
ND T+20	-2.48	-1.60	-1.52	0.00

The N-terminal domain length affects Stokes radius

Analytical SEC was used to evaluate the Stokes radius of the linker mutants. As shown in Figure 4.3, ND T+20 and ND P+20 elute at very similar volumes as ND WT, indicating a similar degree of compaction. Table 4.2 quantifies the R_S and apparent molecular weight. ND T+20 and ND P+20 have a larger molecular weight than ND WT due to the insertion of linkers. ND P+20 is less autoinhibited than ND WT, and ND T+20

is more autoinhibited than ND WT, yet the two are nearly identical to each other and to ND WT in R_S.

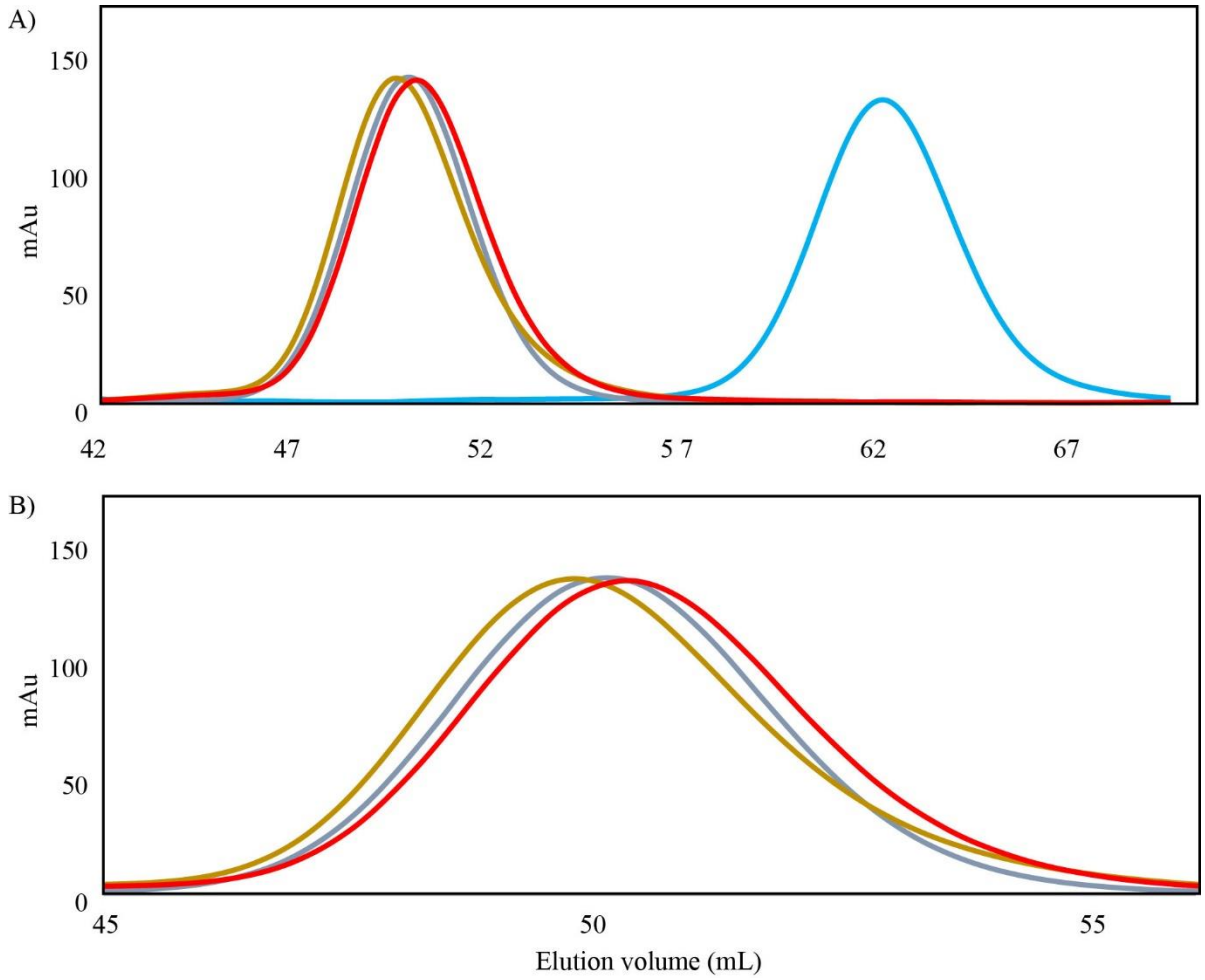


Figure 4.3: Size exclusion chromatography of linker mutants. Elution volumes of proteins from left to right indicate progressively smaller radii. Shown here — is DBD, — is ND WT, — is ND T+20, and — is ND P+20. A) Shows elution volumes from 42 – 69 mL, B) Shows only ND WT and linker mutants.

We assume that the GS linkers do not interact with any part of the protein and do not result in interaction of either TAD2 or PRR with a section of DBD distinct from that seen in ND WT. The consistent Stokes radii and the resulting changes to DNA binding affinity fit our predictions and make a large-scale change in binding pattern unlikely.

Table 4.2: SEC analysis of DBD, ND WT, and PRR mutants

	Stokes radius (nm)	Elution volume (mL)	Apparent Molecular weight (kDa)	Actual Molecular weight (kDa)
DBD	2.612 ± 0.005	62.512 ± 0.069	33.554 ± 0.155	24.548
ND T+20	3.579 ± 0.001	49.752 ± 0.001	78.079 ± 0.001	36.034
ND P+20	3.550 ± 0.035	50.092 ± 0.026	77.028 ± 2.091	36.034
ND WT	3.529 ± 0.020	50.173 ± 0.110	75.853 ± 0.558	34.565

As described in Chapters 2 and 3, predictions for the Stokes radius based on sequence length are not accurate. Predictions for the linker mutants using various estimation methods are shown in Table A4.

Effective concentration of TAD2 based on PRR

The C_{eff} of TAD2 is dramatically affected by the length and sequence of PRR. Figure 4.4A shows the end-to-end distances of two amino acid polymers with the same number of residues but with *Lp* of 3 Å versus 7 Å, where the maximum value, labeled at the dashed line, indicates the distance between the N- and C-termini of the polymer that is most occupied. The distance between the C-terminus of the PRR, residue 93, and the nearest residue of DBD's DNA binding pocket is measured on the crystal structure of DNA-bound p53 within a monomer, dimer, and tetramer, shown in Figure 4.4B and quantified in Table 4.3 (52). Using the actual distance, we can calculate the C_{eff} of TAD2 based on the *Lp* of the linker separating it from a DBD within a monomer, dimer, or tetramer, Figure 4.3C-D.

The distance separating the C-terminal end of PRR, S93, and the nearest residue of DBD that is thought to shift in response to TAD2 in the DNA binding pocket,

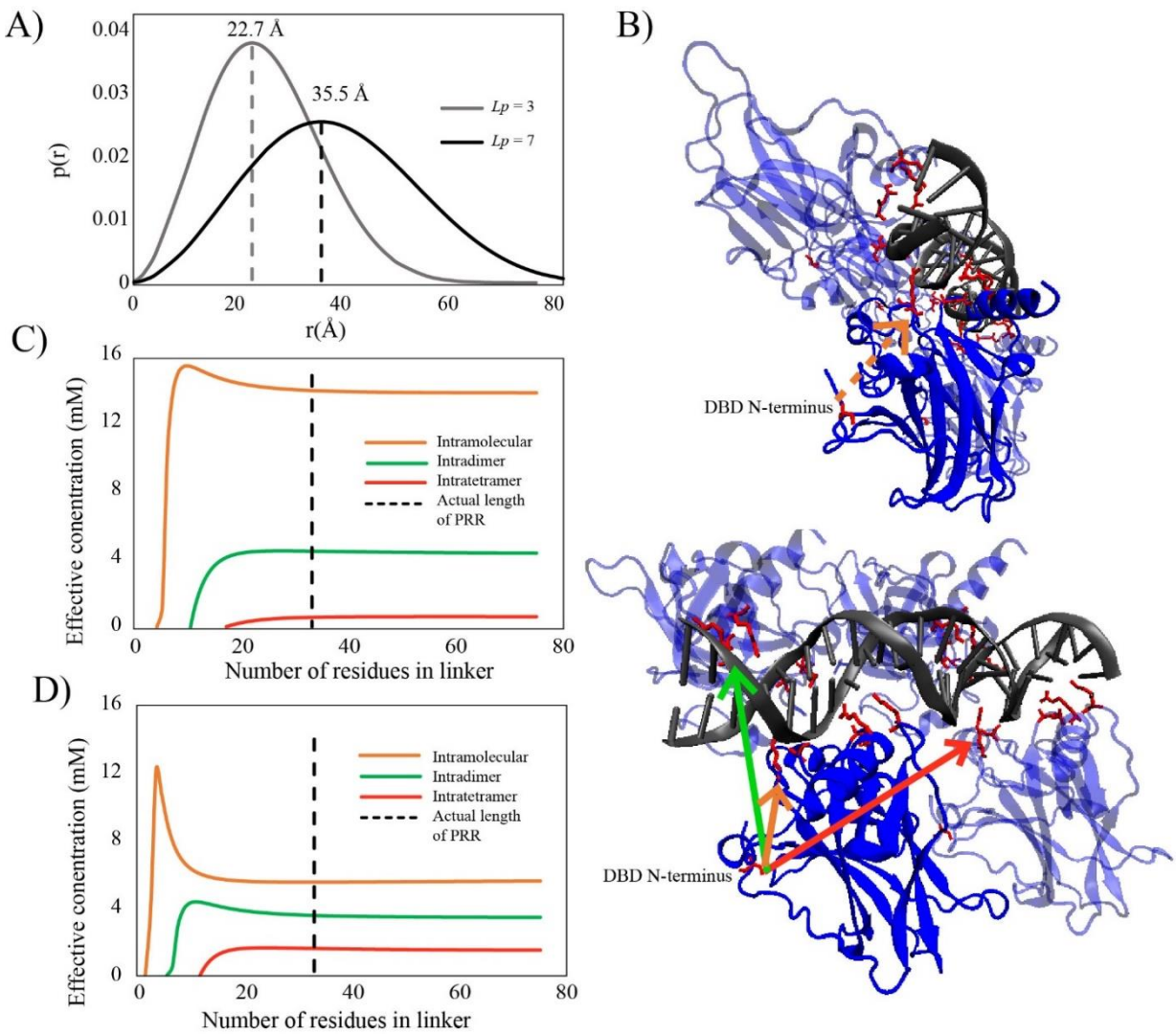


Figure 4.4: Wormlike chain model predicts TAD2 effective concentration. A) The end-to-end distribution of a 33-residue linker if *Lp* = 3 Å and *Lp* = 7 Å with maximum indicated with a dotted line. B) Cartoon model of dimeric and tetrameric DBD where the distance separating the most C-terminal residue of PRR and the nearest DNA binding pocket residue is indicated with arrows (PDB **4HJE**). DNA is shown in black, DBD in blue, and DNA contact residues of DBD are red. C) The effective concentration of TAD2 for DBD for a 33-residue linker if *Lp* = 3 Å depending on the oligomerization state. D) The effective concentration of TAD2 for DBD for a 33-residue linker if *Lp* = 7 Å depending on the oligomerization state.

R248, is 18.05 Å. With an *Lp* of 7 Å, the C_{eff} of this intramolecular interaction is 5.4 mM. The distance separating this most C-terminal residue of PRR in one p53 unit and the DNA binding site of another DBD in the dimer and tetramer is also shown in Table 4.3.

We also calculated these values for a chain with an *Lp* of 3 Å. Figure 4.4C-D plots the C_{eff} versus the length of the PRR and shows that the C_{eff} of TAD2 is highest for a very short linker: 10 residues for a chain with an *Lp* of 3 Å and 4 residues for a chain with an *Lp* of 7 Å. The highest C_{eff} is found for a linker that is less rigid and that interacts intramolecularly. The significantly lower C_{eff} for the more rigid PRR, 5.4 mM compared to 13.9 mM for the flexible linker, suggests that PRR is not optimized for TAD2 autoinhibition of DBD. However, we find the PRR is long enough to bridge all proposed distances regardless of its persistence length.

Table 4.3: Effective concentration based on oligomerization and persistence length

	Distance	C_{eff} if *Lp* = 3 Å (mM)	C_{eff} if *Lp* = 7 Å (mM)
Intramolecular	18.05 Å	13.9	5.4
Intradimer	30.16 Å	4.5	3.5
Intratetramer	43.66 Å	0.6	1.6

The C_{eff} of TAD2 in the ND P+20 and ND T+20 mutants cannot be assessed because the entire linker sequence must be modeled as uniform whereas we expect the inserted GS linker to have a different persistence length than the PRR.

C_{eff} can also be used to assess the occupancy of the autoinhibited state for the apo protein. As described in Chapter 7, the occupancy or fraction bound of TAD2 in the DNA binding pocket of DBD can be calculated using the C_{eff} and the binding affinity of TAD2 for DBD when untethered. We estimate the untethered binding affinity to be 1 mM and use the C_{eff} of the monomeric interaction, 5.4 mM, to find the occupancy of TAD2 in the DNA binding pocket to be 84%, meaning 84% of the apo protein is in a closed, or

autoinhibited state. This figure varies based on the persistence length used and whether the interaction occurs within the monomer, dimer, or tetramer, but the estimate is in agreement with an earlier NMR study in which open and closed states of p53 in the absence of DNA represented 15% and 85% of the intensity of cross-peaks (177).

Evolution and variation of TAD subdomains

Protein domains are subject to evolutionary pressures that depend on their structure and function; accordingly, IDRs tend to evolve more quickly than folded domains due to the relative lack of structural constraints (208), although there are clear exceptions at PTM sites and at SliMs that occur at interaction interfaces (209, 210). Instead, the length or relative frequency of residue types is more preserved than the amino acid sequence in many IDRs. DBD is the most highly conserved region of p53 in both structure and sequence. For example, human and fruit fly DBD share 24% sequence homology. Their solved structures are easily overlaid, and they bind the same DNA sequences with equivalent affinity (34). Other domains within p53, however, are variably conserved. Within the N-terminal TAD, TAD1 is the most highly conserved and PRR has been noted to be the least well conserved subdomain (105). MSA of TAD1, Figure A2, show a recognizable N-terminal sequence, notably with good conservation of phosphorylation sites S15, T18, S20, extending to marsupials with a less conserved C-terminal sequence. Figure 4.5 shows the conservation of amino acid types within each subdomain of TAD that correspond to the MSA of mammals in Figures A2-4. In TAD1, negative charge, polarity, hydrophobicity, and proline content are largely conserved, with the average fractional content of each and its standard deviation shown in Table 4.4.

TAD1 is shorter, but recognizable, in birds as well with a similar percentage of amino acid types.

Table 4.4: Fractional content and variation in conservation of amino acid type in TAD subdomains

Subdomain	Fractional proline content	Fractional charged content	Fractional polar content	Fractional hydrophobic content
TAD1	0.13 ± 0.03	0.19 ± 0.03	0.31 ± 0.04	0.50 ± 0.04
TAD2	0.09 ± 0.05	0.31 ± 0.05	0.17 ± 0.07	0.53 ± 0.06
PRR	0.28 ± 0.06	0.12 ± 0.04	0.16 ± 0.07	0.72 ± 0.08

MSA of TAD2, Figure A3, shows that TAD2 maintains its percentage of charged residues throughout mammals, and aromatic residues at human residues W53/F54 are

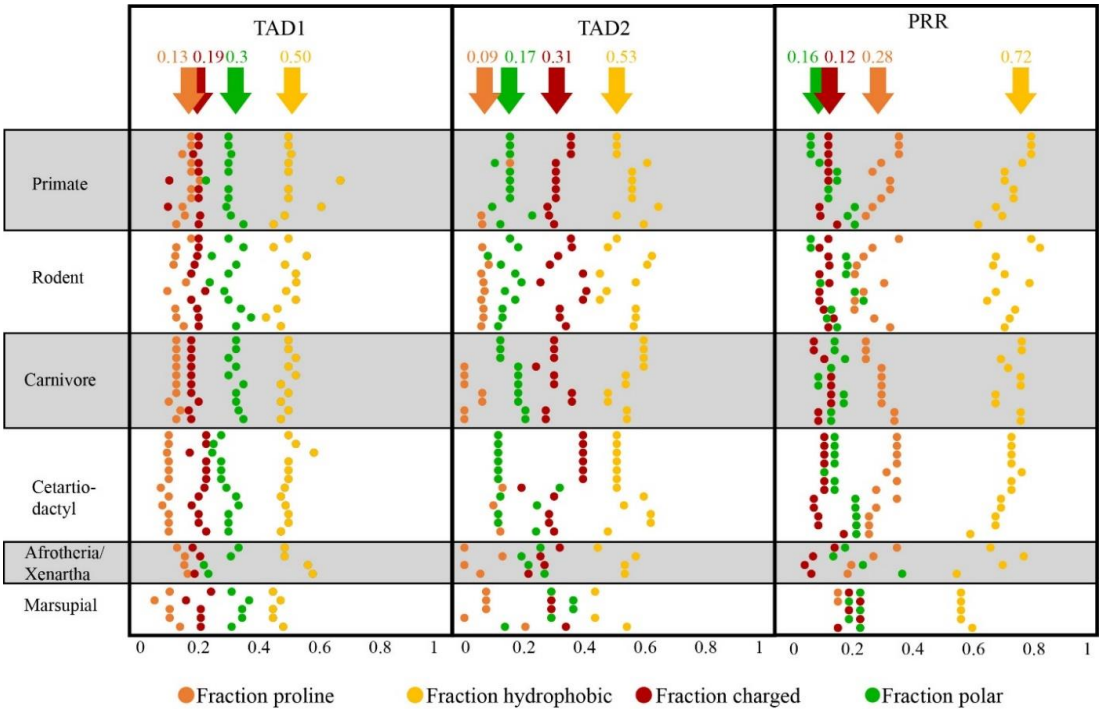


Figure 4.5: Sequence feature conservation of TAD1, TAD2, and PRR among mammalian orders. The average fractional content of each type of residue in the specified domain for all animals listed is shown above each arrow.

maintained in most mammals. Aromatic residues are conserved only in some birds, and the sequence is shorter. The TAD is present in most orders of animals as well as in paralogs p63 and p73 (29); however, it is unclear how the function of TAD2 may have changed over time. Poor sequence conservation does not inherently indicate a lack of conservation of function. Many IDRs involved in autoinhibitory interactions have poor sequence conservation and yet serve similar functions in diverse groups (26, 211, 212). Conservation of amino acid type in TAD2, Figure 4.5, and Table 4.4, shows a slightly greater variability in each category than is seen in TAD1.

MSA of PRR, shown in Figure A5, shows the length of the PRR gradually increases from marsupials through primates. Within individual mammalian orders, PRR is often as well conserved as TAD2 and within some orders is nearly perfectly conserved (110), although its variability across mammals is similar to that seen in TAD2. The sequence of PRR is poorly conserved between groups; however, the sequence composition is surprisingly consistent. It has been noted that PPII regions are often structurally conserved with low sequence conservation (213).

In all mammalian sequences shown in Figure A5, the PRR is comprised of a proline fraction of 15 – 33% with a large percentage of alanines and comparatively low fraction of charged and aromatic residues. This suggests a conserved function of PRR among mammals. Available monotreme sequences are limited and may represent isoforms and so are not included here: echidna T. *aculeatus* has a PRR but exceptionally poor TAD1/2 conservation; the available platypus O. *anatinus* sequence lacks the N-terminus entirely. Among reptiles, the N-terminus sequence shows a wide variation that is difficult to align; however, several birds appear to have PRR

immediately N-terminal to DBD, suggesting the PRR may have emerged in birds and persisted into the mammalian order, as suggested by a previous paper (113). While the PRR is recognizable in many birds, both sequence and feature conservation appear lower than that seen in most mammalian orders, Figure A5.

The observation that interaction interfaces between proteins must coevolve has led to an improvement in the ability to predict protein-protein interactions in recent years (25, 214). This combination of database analysis and coevolutionary analysis has been used to investigate affinity modulation of viral linkers connecting two domains that bind Rb (199), but the method has not yet been fully developed to evaluate coevolution in protein regions that engage in fuzzy interactions, or in cooperative or frustrated effects without directly interacting. This leads us to ask, how might TAD2 and PRR coevolve, and can we predict features of compensatory evolution between these two domains?

In MSA of TAD2 and PRR, it initially appears that the lengthening of PRR corresponds to an increase in the negative charge of TAD2. As shown in Chapter 2, negative charges in TAD2 play an important role in the intramolecular interaction, and the length of PRR has a large effect on the C_{eff} of TAD2. If the intramolecular interaction is conserved and is maintained at a low affinity, then the increases in negative charge of TAD2 must be counteracted by increasing length and/or rigidity of PRR. However, analysis of the net charge per residue (NCPR) (215) of TAD2 versus PRR length or proline content do not yield an obvious correlation, as shown in Figure 4.6A. TAD1,

which is expected to coevolve with PRR to a lesser degree if at all is also shown for comparison.

Similarly, within PRR, an increase in length does not correlate with its proline content or number of charged residues in TAD2 (Figure 4.6B-C). While primate PRR is among the longest, the fractional proline conservation implies it is neither more nor less rigid in other mammals. We expect that a longer PRR would be necessary to separate an increasingly negatively charged TAD2 from DBD; however, the number of negative

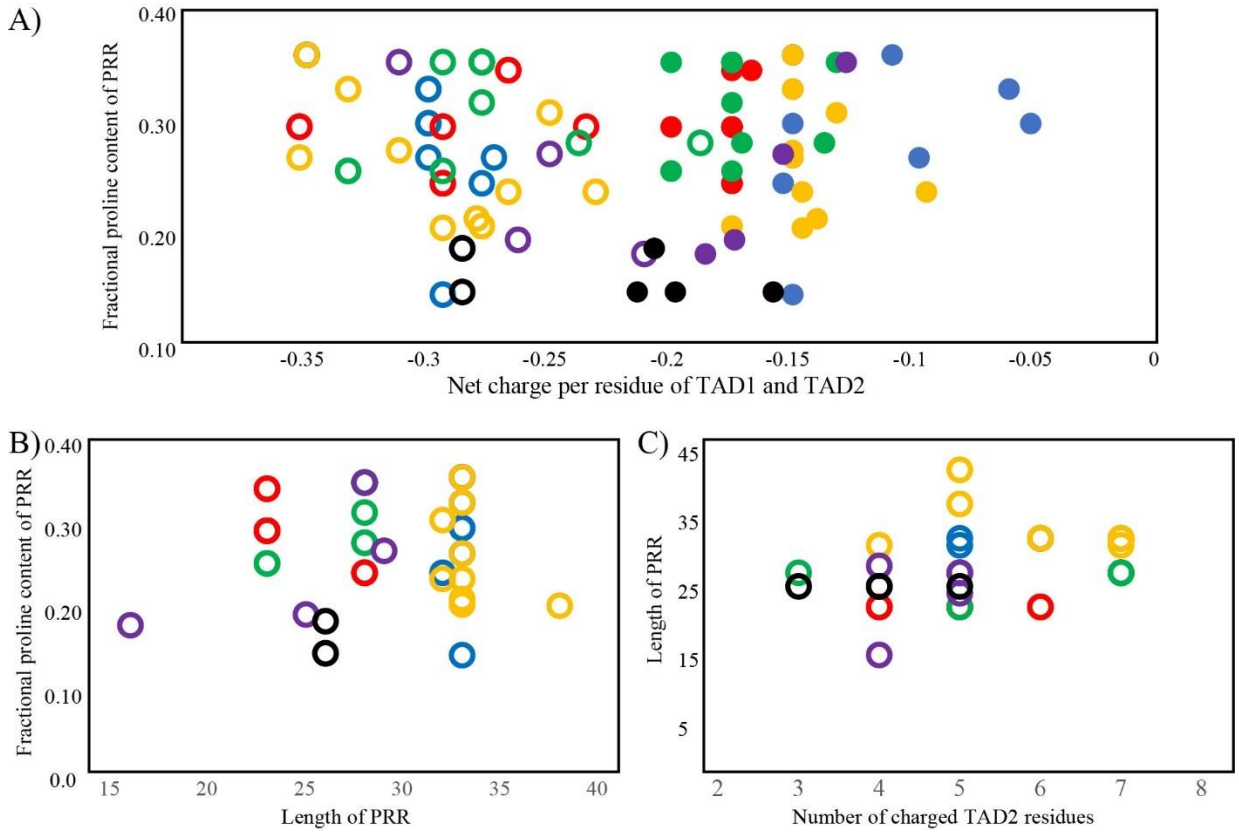


Figure 4.6: Correlation of features of TAD2 and PRR. In all instances, filled circles indicate TAD1 and open circles indicate TAD2, where ○ blue is primates, ○ yellow is rodents, ○ red is carnivores, ○ green is cetartiodactyl, ○ purple is Xenarthra and Afrotheria, and ○ black is marsupial. Correlation of, A) TAD2 NCPR versus PRR length, B) TAD2 NCPR versus PRR proline fractional content, C) length of PRR versus fractional proline content.

charges is usually 5, and while the length of PRR varies between 15 and 44 residues, there is no obvious correlation with PRR length.

We conclude that there is no obvious correlation between the features analyzed. Therefore, while we expect the intramolecular interaction to exist in all or most mammals, the basic features we expect to be important to the frustrated interaction vary by species, implying the intramolecular interaction may have varying autoinhibitory capability.

CHAPTER FIVE: KIX AND MIMIC PEPTIDE BINIDNG

Rationale

The kinase-inducible domain interacting domain (KIX) of the coactivator CREB Binding Protein (CPB) and the homologous p300 interacts with many proteins as an early step in the initiation of gene transcription (216). KIX binds basal transcription machinery and IDRs of transactivation domains of transcription factors such as myeloblastosis protein (cMyb), mixed-lineage leukemia 1 protein (MLL), and cAMP response element-binding protein (CREB) (132). KIX binding partners are frequent drug targets due to their well-known roles in cancer. Three other domains of CBP/p300 – TAZ1, TAZ2, and NCBD – also interact with transactivation domains of transcription factors (146, 217). The bromodomain and lysine acetyltransferase domains of CBP/p300 recognize methylated DNA sequences and modify chromatin in response to cellular factors that regulate transcription (218).

CBP/p300 are among the most universal transcription coactivators and have been found associated with 16,000 promoters and many families of transcription factors (219, 220). Both have similar domain organization and some overlap in their function; however, their activity is distinct to an ill-defined degree (221). In this study, we use KIX from human CBP.

Features of KIX interaction with cMyb, MLL, and CREB

Structurally, KIX is a three-helix bundle with two short 3_{10} helices, labeled G1 and G2 in Figure 5.1A, with two major binding pockets called the cMyb site and the MLL site after their canonical binding partner (PDB **2AGH**) (222). Helices α1 and α3 create a hydrophobic interface, shown interacting with a green ribbon representing the transactivation domain of cMyb in Figure 5.1B-C. A groove formed by the juncture of all

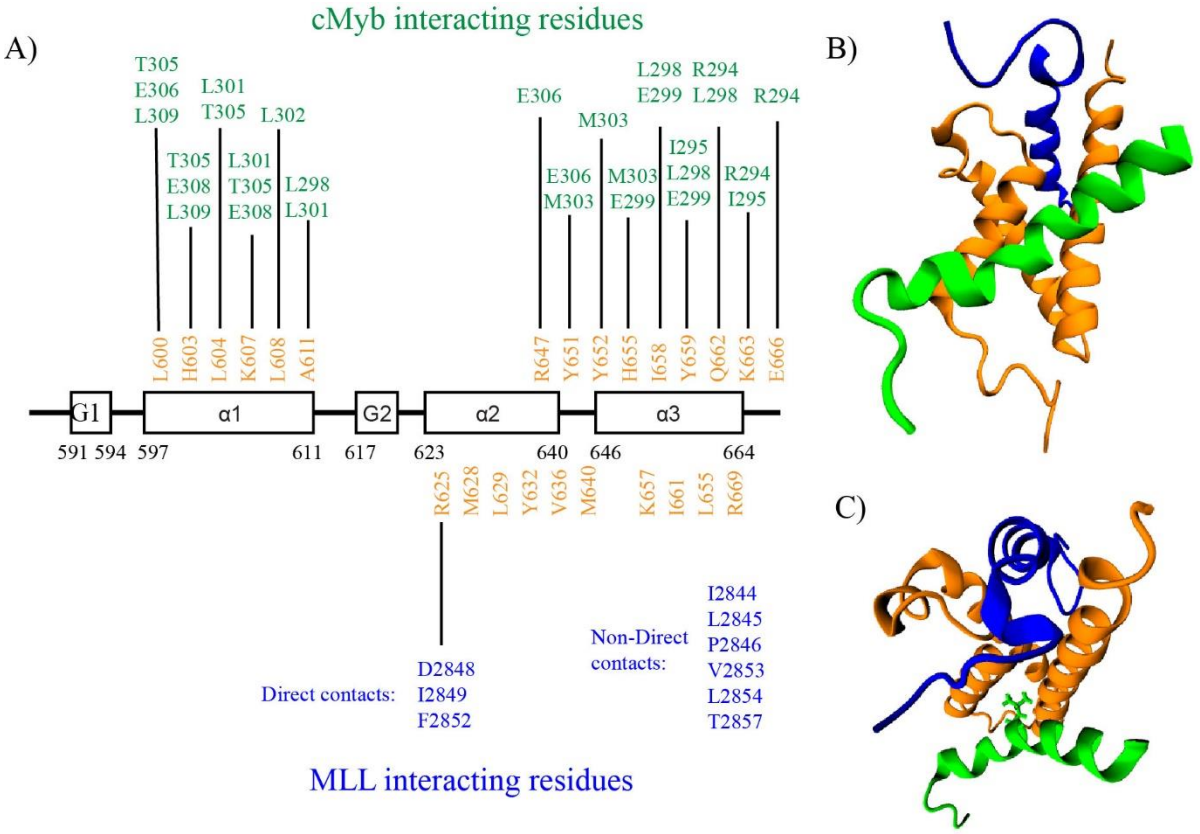


Figure 5.1: Structure of KIX. A) A structural map of the KIX domain shows the layout of alpha helices and 310 helices as rectangles separated by short linkers as black lines. KIX residues, shown in orange, are shown to have direct contacts with cMyb residues, shown in green, and MLL residues, shown in blue. Non-directly contacting residues of MLL are listed, B) Front view of cartoon structures representing the backbones of KIX bound to cMyb, green, and MLL, blue as cartoons. C) Top view of KIX, orange, cMyb, green with Leu 302 side chain shown, and MLL, blue.

three helices creates the MLL binding pocket, and MLL is shown as a blue ribbon in Figure 5.1 interacting with the MLL binding site (PDB **2AGH**) (222).

The two binding sites of KIX are not exclusive. The KIX-interaction sites of cMyb, MLL, and CREB, as well as other KIX-binding proteins, contain a “ΦXXΦΦ” or “LXXLL” motif, where Φ is a hydrophobic residue, and X is any residue (223, 224). The shared binding motif and conformational plasticity of the disordered transactivation domains allow cMyb to bind to the MLL site and vice versa (225). Hence, many binding studies use a relatively low concentration of the transcription factor to avoid oversaturation and interaction with the secondary pocket, as in Poosapati et al. (226) where KIX and cMyb were combined in a 1:0.8 ratio to avoid oversaturation. Despite this apparent competition for binding sites, cooperation of binding to KIX is noted for several combinations of transcription factors, where binding of MLL to the MLL site increases KIX affinity for either cMyb or CREB via an allosteric conformational change in KIX (225, 227).

cMyb, MLL, and CREB engage in a coupled folding and binding event when interacting with KIX that results in an amphipathic helix in which residues are buried in the hydrophobic pockets created within the helical bundle (226, 228-230). These coupled folding and binding events result in a decrease in entropy upon binding, the penalty of which can be decreased by cooperative binding of cMyb and MLL (231) or modulated with strategic mutations (226). As the KIX-interacting transcription factors described here are well known oncogenes, these are targets for drug discovery. Many inhibitors of cMyb binding to KIX have been designed based on mutational studies of cMyb sequences with increased unbound helicity relative to the WT (232). MLL

inhibitors have been similarly designed based on mutational studies (233). Various strategies have also been employed in the design of CREB inhibitors (234).

Structure and function of cMyb WT and mimics

cMyb is known to interact with more than a thousand gene promoters (235) and is especially well known for its role in hematopoiesis where deletion is embryonically lethal (236). Aberrations in cMyb expression are associated with acute myeloid leukemia, as well as other cancers (237-240).

The binding interface of cMyb with KIX, as shown in green text in Figure 5.1A, is in helices α1 and α3. Note, cartoons in Figure 5.1B-C show cMyb residues 291-315. The top view shows the projection of the side chain of one important residue, Leu 302, into the hydrophobic pocket of helices α1 and α3.

The cMyb WT used in this study is a synthetically-made peptide that corresponds to human residues 293 – 309 and which was found to bind to KIX with a K_D of 4.37 μM, similar to that found in a previous study of cMyb WT of a similar length (241). The length of the cMyb TAD used has a significant effect on the binding affinity to KIX. A fragment consisting of residues 275-327 has a K_D of 0.5 μM (226), and a shorter fragment consisting of residues 291-315 demonstrates weaker binding, K_D = 4 μM (232, 241). The binding segment of cMyb, in the strictest sense, is residues 291-310, which pack against the hydrophobic pocket (242), but the sequence outside the primary binding pocket and residues that do not directly contact KIX have a large effect on binding affinity, often by modulation of the entropy of the coupled folding and binding effect (226, 228).

The cMyb mimics and inhibitors used in these studies take two major approaches, both with the same general goal of increasing unbound helicity to reduce the entropic penalty upon binding: 1) introduce alpha-sulphonyl groups to an amine, and 2) staple the backbone of an alpha-sulphonyl group-modified peptide mimic to increase helicity. These are compared to a synthetic cMyb WT sequence.

The cMyb mimics represented in the first subset were conceived of and prepared by Dr. Mi Zhou. These are designated here as cMyb 104-2 and cMyb 109-1; sequences are shown in Figure A5. cMyb 104-2, preserves the direct-contact residues R294, I 295, L298, E299, L301, M303, T305, and L309. Other contact residues are substituted with helix-stabilizing groups. cMyb 104-2 was found to bind KIX with a K_D of 0.93 μM via FA. cMyb 109-1 preserves the direct-contact residues R294, L298, M303, T305, and L309, with extensive backbone modifications to other contact residues. cMyb 109-1 was found to bind with a K_D of 0.89 μM via FA.

The cMyb inhibitors represented in the second subset were conceived of and prepared by Yu Yu Win. These are designated here as LC-A-122-2 and LC-A-122-3, and their structures are shown in Figure A6. These compounds do not directly mimic cMyb WT or preserve direct contact residues; instead, they are predicted to bind a similar site on KIX based on structure alone. LC-A-122-2 and LC-A-122-3 were found to bind KIX with K_Ds of 0.09 μM and 0.235 μM, respectively, via FA.

Structure and function of MLL WT and mimics

MLL, alternatively named Lysine-specific Methyl Transferase 2A (KMT2A), is a transcription factor that includes multiple domains involved in DNA recognition, and it is

thought to be involved in regulation of developmental genes such that disruption of MLL results in abnormal fetal development (243, 244). Rearrangements of MLL are associated with up to 10% of acute myeloid leukemias (245).

Shown in Figure 5.1, the human MLL residues 2838 – 2869 interact with the α2 helix and the C-terminus of the α3 helix. Unlike cMyb, MLL binds KIX through relatively nonspecific hydrophobic interactions, where only R625 of KIX makes direct contact with D2848, I2489, and F2852. Other important MLL residues are shown in Figure 5.1 but are not shown to directly contact KIX residues in the solution structure used here (222).

The cooperative binding of cMyb and MLL to KIX tightens the binding of MLL from a K_D of 2.8 μM to 1.7 μM; conversely, KIX-cMyb alone has a K_D of 10 μM and KIX-MLL:cMyb has K_D of 4 μM (225).

The MLL fragments used in this study include a synthetic MLL WT peptide that spans human residues 2845-2859, and two mimics. The synthetic MLL WT, which does not contain the entire minimal transactivation domain, was found to have a K_D of 3.97 μM via FA, in agreement with an earlier study (225). MLL mimic 6 and MLL mimic 8 were conceived of and prepared by Minghui Wang and correspond roughly to residues 2845 – 2858, where large scale backbone modifications are shown in Figure A7B-C and have reported K_Ds of 0.18 μM and 0.26 μM, respectively, via FA.

Structure and function of CREB WT and mimics

The kinase inducible domain (KID) domain of the transcription factor CREB was the first established binding partner of KIX (246). CREB responds to hormones, growth factors, and a variety of phosphorylation sources to activate genes in a variety of

pathways (247). Overexpression of CREB is implicated in several cancers, notably leukemia (248, 249).

The disordered KID region folds Into two helical regions upon phosphorylation (pKID) by Protein Kinase A (250). As with the other transcription factors here, the coupled folding and binding event follows the induced folding pathway in its interaction with KIX (251). The pKID:KIX interface overlaps with the cMyb:KIX interface where the αB helix of KID contacts the hydrophobic cMyb pocket of the α1 and α3 helices of KIX, and the αA helix of KID contacts the α3 helix of KIX (250, 251).

The CREB mimics represented here were conceived of and prepared by Bo Huang. These are designated here as 78a, 79b, 3_30b, and 133b, although in a recently submitted publication 3_30b is described as S2-18.

Sulfono mimetic cMyb mimics interact with KIX

To compare how the sulfono-mimic peptides bind KIX relative to cMyb WT, NMR titration experiments using ^{15}N-labeled KIX and unlabeled versions of either a synthetic cMyb peptide or the sulfono-mimics were performed. The synthetically produced cMyb peptide was added to KIX in a molar ratio 0.8:1 to avoid interaction with the MLL binding site. The mimics were titrated into KIX to a final molar ratio of 4:1 to ensure saturation of binding. For all the ligands fast exchange was observed and ^{15}N-labeled KIX H-N resonances shifted in proportion to the amount of ligand that was added. Shifts of KIX when bound to synthetic cMyb WT are easily tracked, shown in NMR spectra in Figure 5.2A. Figure 5.2B-C show overlays of 8 spectra each with different molar ratios of the

cMyb mimics added to ^{15}N-labeled KIX. Chemical shifts at each titration point were measured and averaged as described in the methods.

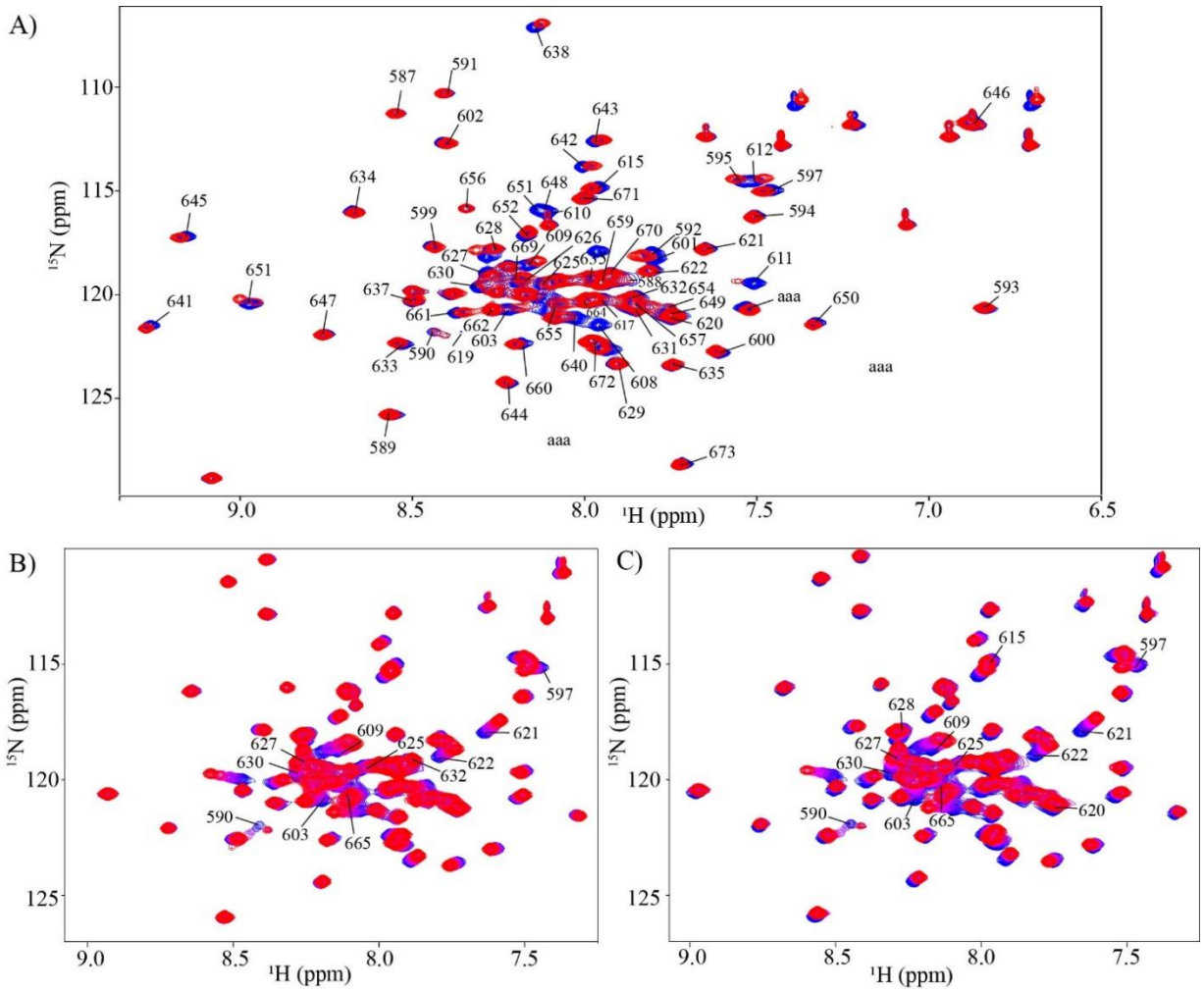


Figure 5.2: NMR spectra of labeled KIX chemical shifts bound to cMyb WT and mimics 104-2 and 109-1. KIX apo (blue) and bound to (red) cMyb where the largest shifts are labeled. A) cMyb WT at a 0.8-fold ratio, B) cMyb 104-2 with titrations up to 4-fold ratio, C) cMyb 109-1 with titrations up to 4-fold ratio.

Figure 5.3 shows the chemical shift changes in KIX upon addition of 0.8-fold or 4-fold cMyb WT or mimics, respectively. Upon binding the synthetic cMyb WT, Figure 5.3A, chemical shift changes in KIX greater than two times the resolution limit of the experiment occur in all three helices, clustered around the canonical cMyb-contact

residues but also in the MLL binding site. The model at right shows shifts greater than the resolution limit as red against a surface model of KIX in orange. These shifts are distributed both in the cMyb binding pocket and across the surface of KIX.

The addition of cMyb mimics 104-2 and 109-1 in four-fold excess of KIX causes chemical shifts throughout the KIX protein, as seen in Figure 5.3B-C. In

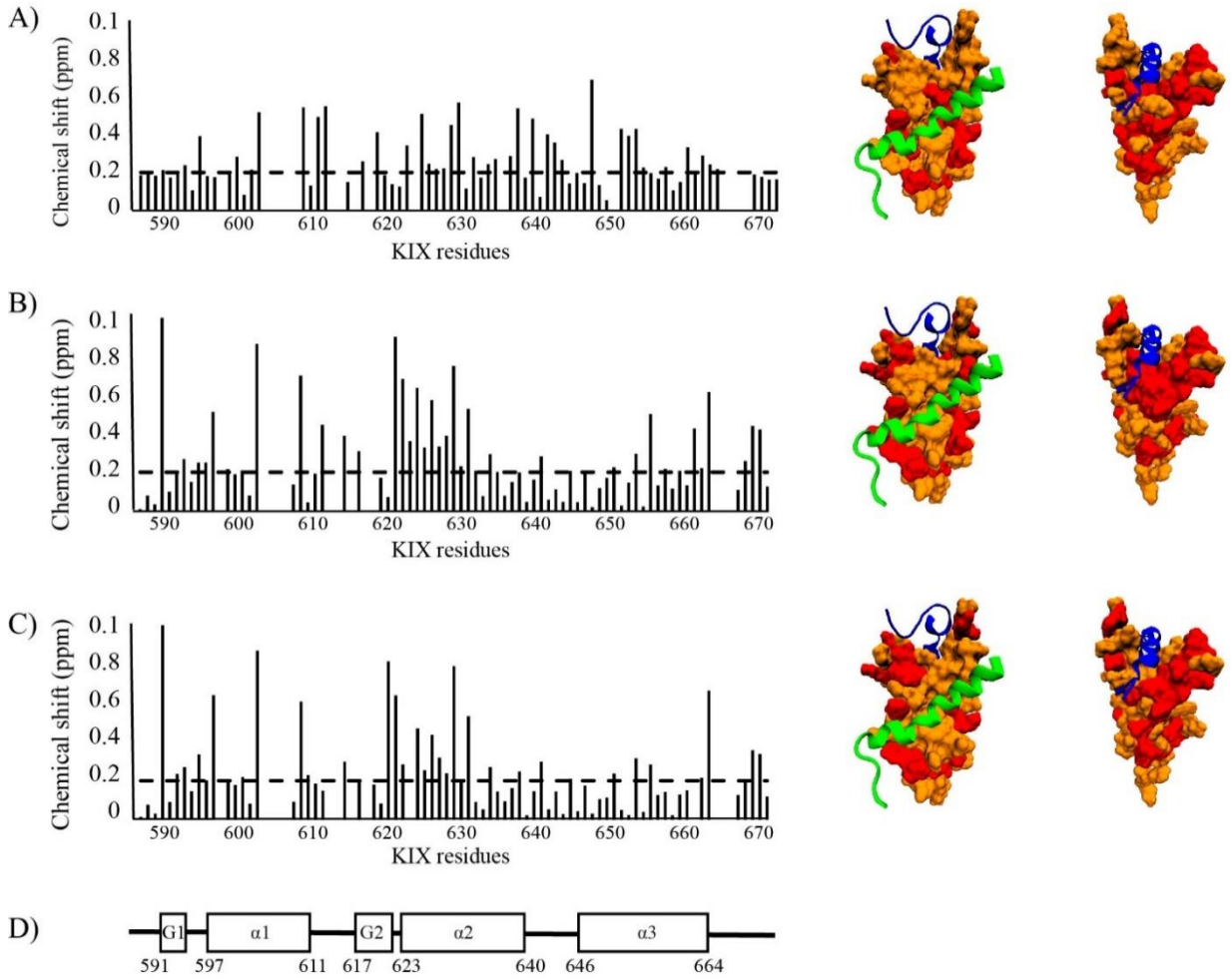


Figure 5.3: Chemical shift changes of KIX bound by cMyb WT, cMyb 104-2, and cMyb 109-1. ^{1}H - ^{15}N HSQC NMR chemical shift changes of cMyb-bound KIX are shown at left and residues most affected on a 3D model at right. Resolution limit is shown as a black dotted line. KIX is orange with largest shifted residues in red. cMyb is shown as a green ribbon. Bar graphs and models are for KIX bound to cMyb constructs at varying ratios: A) synthetic cMyb WT at 1:0.8, B) cMyb mimic 104-2 at 1:4, C) cMyb mimic 109-1 at 1:4, and D) structural map of KIX.

common with cMyb WT, cMyb 104-2 results in large shifts in H603, which directly contacts cMyb WT, as well as in V609, I612, and V630, and cMyb 109-1 results in large shifts in H603 and V630. It is notable that the cMyb mimics here result in shifts more similar each other than to the WT, as seen in large shifts that occur in K590, T597, V609, L621, K622, E627, V630 and L665 in response to both mimics.

Correlation analysis was performed; plots are shown in Figure A8A-C. The Pearson's correlation coefficient compares the chemical shifts of these compounds and shows that mimics 104-2 and 109-1 have a correlation value to cMyb WT of 0.220 and 0.166, respectively. By contrast, the mimics have a correlation coefficient of 0.861 when compared to each other and share several large shifts only between themselves.

Stapled cMyb inhibitors interact with KIX

NMR was conducted on a four-fold excess of stapled cMyb mimics LC-A-122-2 or LC-A-122-3; HSQC spectra of ^{1}H-^{15}N labeled KIX apo and with the cMyb inhibitors are shown in Figure 5.4. LC-A-122-2 resulted in slightly smaller chemical shifts than LC-

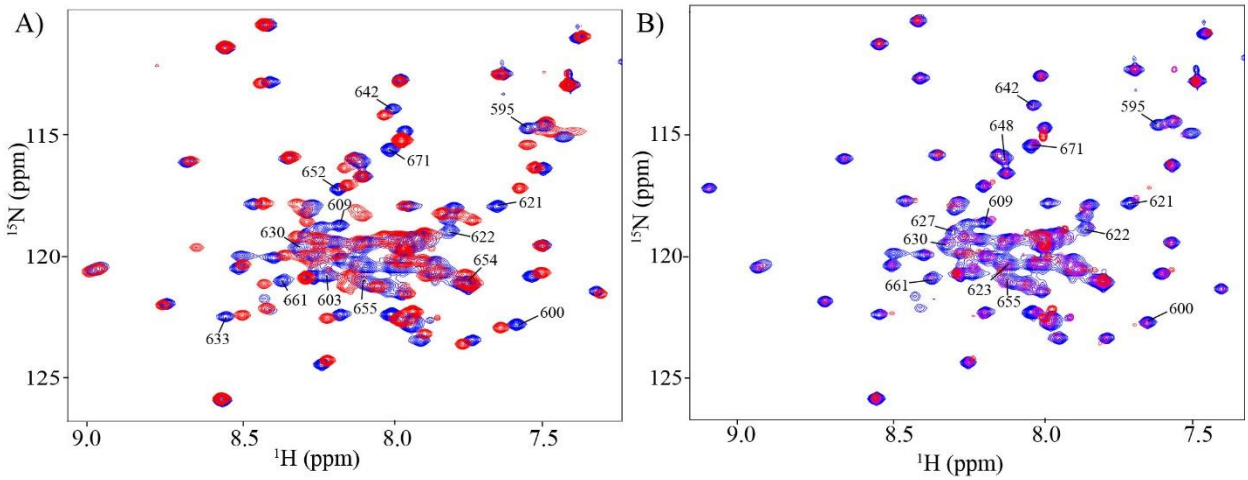


Figure 5.4: NMR spectra of KIX bound to cMyb inhibitors LC-A-122-2 and cMyb mimic LC-A-122-3. KIX apo (blue) bound to four-fold excess of mimic (red), A) LC-A-122-2, or B) LC-A-122-3.

A-122-2. We suspect this is related to the solubility of LC-A-122-2, where full solubilization of the compound in 5% DMSO required approximately three hours in PBS buffer. By contrast, the LC-A-122-3 compound solubilized immediately in the standard NMR buffer (see Methods Chapter 7). The LC-A-122-3 compound resulted in peaks that decrease in intensity as the concentration of compound increases, presumably due to the resonances being in the slow exchange regime at four-fold excess with KIX.

Chemical shift changes in labeled KIX upon addition of compounds are shown in Figure 5.5 with the chemical shift changes greater than the resolution limit mapped onto surface models in red at right. Stapled cMyb inhibitors resulted in shifts primarily in the

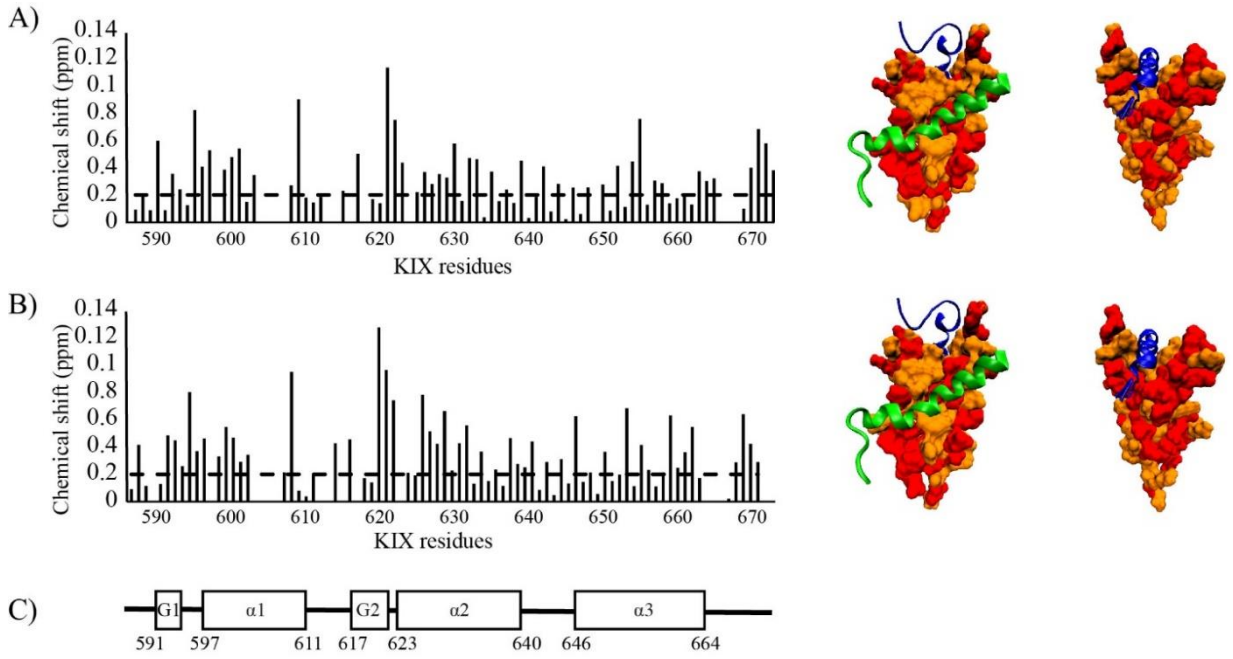


Figure 5.5: Chemical shift changes of KIX bound to cMyb inhibitors LC-A-122-2 and cMyb mimic LC-A-122-3. ^{1}H - ^{15}N HSQC NMR chemical shift changes of inhibitor-bound KIX are shown at left and residues most affected on a 3D model at right. Resolution limit is shown as a black dotted line. KIX is orange with largest shifted residues in red. cMyb is shown as a green ribbon and MLL is shown as a blue ribbon. Bar graphs and models are for KIX bound to compounds in four-fold excess: A) cMyb LC-A122-2, B) cMyb LC-A 122-3, and C) Structural map of KIX.

C-terminus of the α1 helix and N-terminus of the α2 helix of KIX with other smaller shifts dispersed throughout.

In common with chemical shift changes observed in KIX when binding cMyb WT, LC-A-122-2 results in chemical shifts changes at residues V609, V630, E642, H652 and L654. In common with cMyb WT, LC-A-122-3 results in significant shifts at residues L600, V630 and E642. As with the set of cMyb mimics, LC-A-122-2 and LC-A-122-3 share a general profile and are more like each other than like cMyb WT with both mimics resulting in the largest shifts at KIX residues L621 and K622. The Pearson's correlation coefficient of cMyb WT versus LC-A-122-2 is 0.023, cMyb WT versus LC-A-122-3 is 0.077, and LC-A-122-3 versus LC-A-122-2 is 0.837, with plots shown in Figure A8D-F.

MLL mimics interact with KIX

HSQC NMR spectra of 1H-15N labeled KIX apo and in the presence of a four-fold excess of MLL WT, MLL mimic 6 or MLL mimic 8 are shown in Figure 5.6. Figure 5.6A shows large migrations of labeled peaks upon the addition of MLL WT, especially at residues T615, L621, K622, D623, R625, E627, and N628 as was previously observed (225). These unusually large chemical shift changes, up to seventy times greater than the resolution limit and with peak changes up to 1.4 ppm in ^{1}H and up to 4.4 ppm in ^{15}N, in KIX are in or adjacent to the MLL binding pocket of KIX. By contrast, MLL mimic 6 and mimic 8 resulted in chemical shift changes up to four and nine times the resolution limit, respectively. Large shifts in KIX resulting from the addition of MLL mimic 6 and mimic 8 are labeled in Figure 5.6B-C. Titrations were not required to assign

chemical shift changes; however, MLL mimic 8 results the split of V630, L653, and Y659 into two peaks upon four-fold titration, suggesting slow exchange.

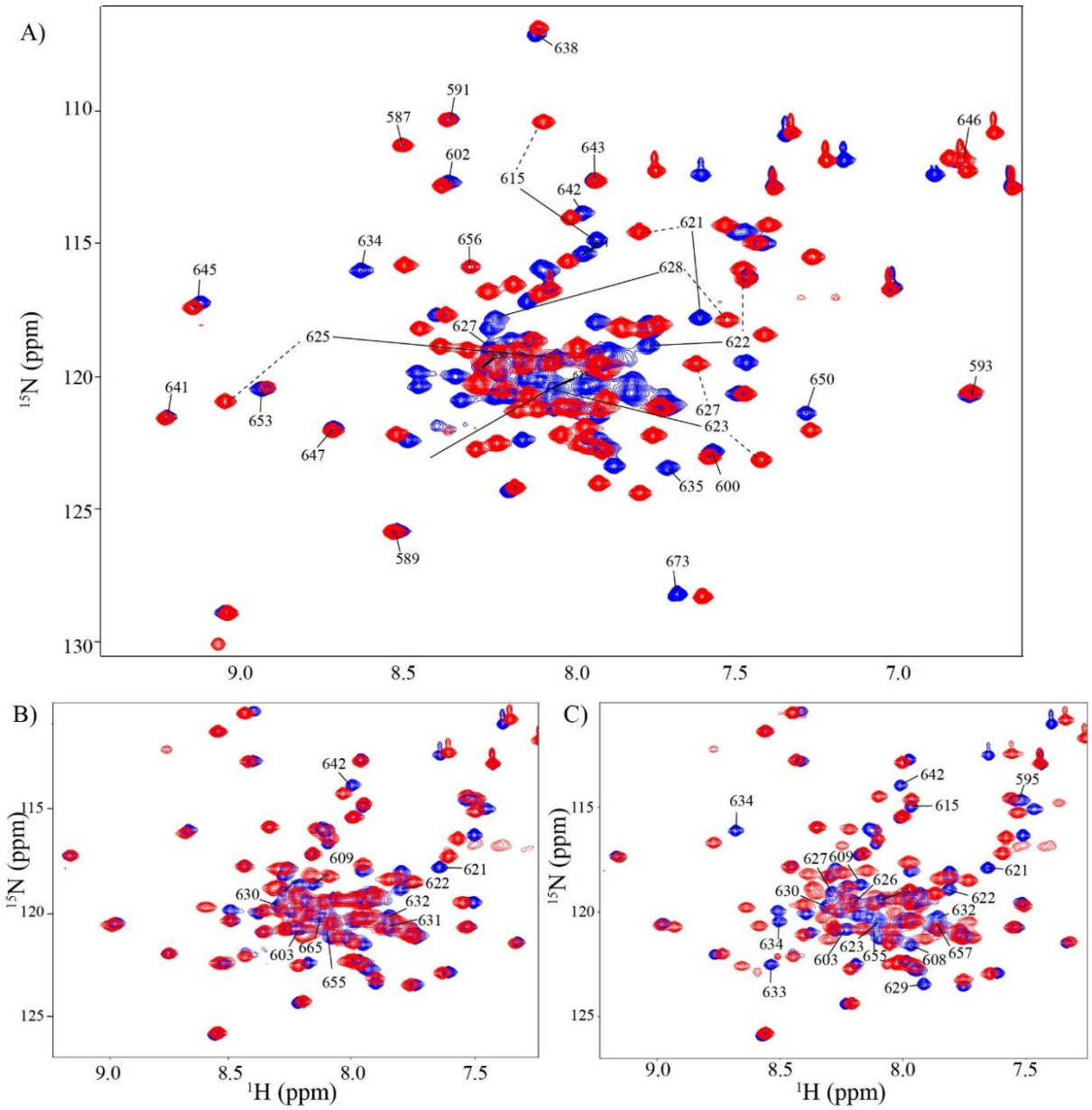


Figure 5.6: NMR spectra of KIX bound to MLL WT, MLL mimic 6 and MLL mimic 8. KIX apo (blue) and bound to four-fold excess (red) of A) MLL WT, B) MLL mimic 6 and, C) MLL mimic 8. Very large shifts are labeled with combination solid and dashed black lines.

The synthetic MLL fragment results in chemical shift changes above the resolution limit over large sections of KIX, as shown in Figure 5.7A. This accords with the noted allosteric effects of MLL: interaction with KIX increases KIX affinity for other transcription factor binding partners that interact with the cMyb binding pocket. Thus, we expect conformational changes to occur at sites in KIX distant from the MLL interaction

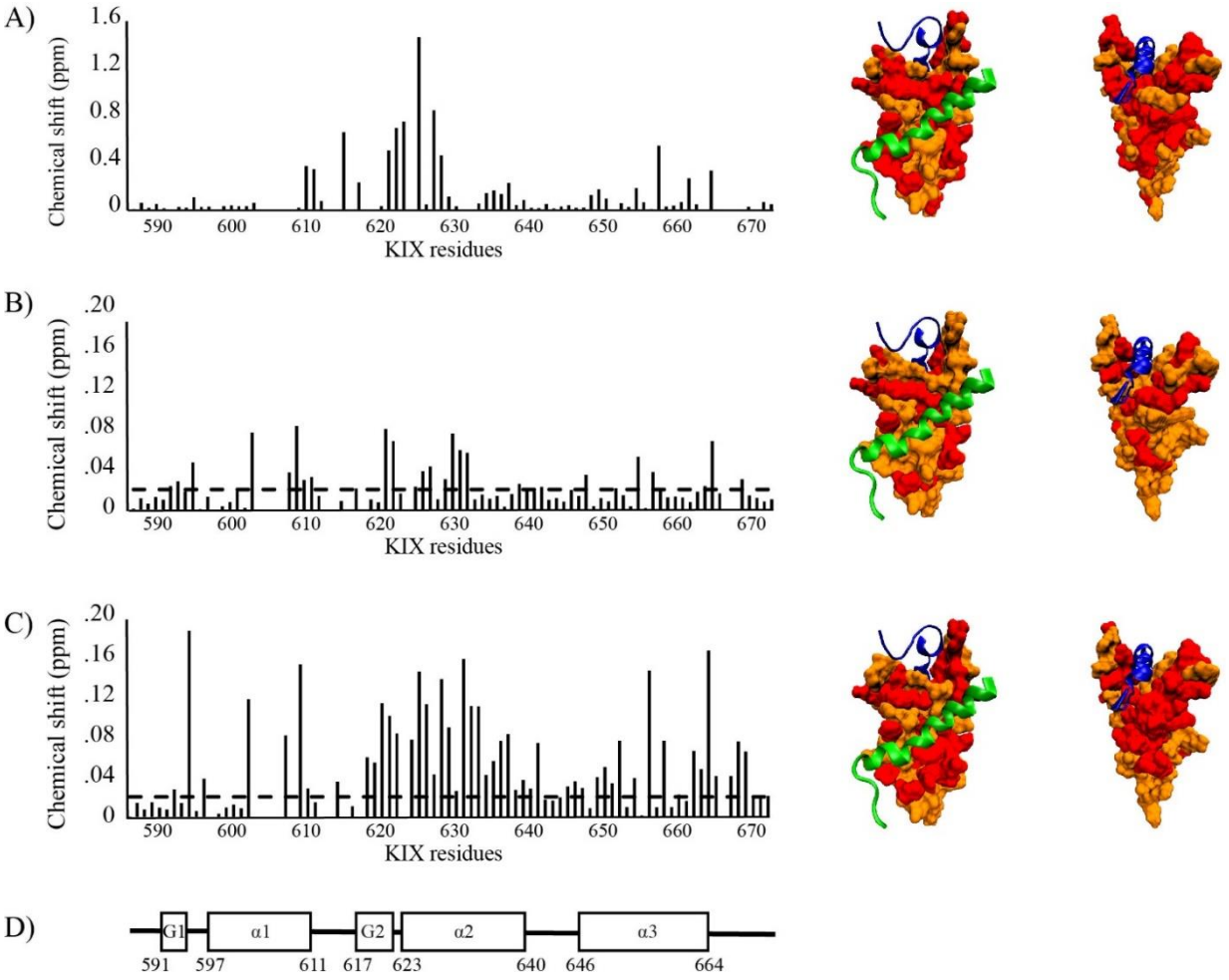


Figure 5.7: Chemical shift changes of KIX bound to MLL WT, MLL mimic 6, and MLL mimic 8. ^{1}H - ^{15}N HSQC NMR chemical shift changes of MLL-bound KIX are shown at left and residues most affected on a 3D model at right. Resolution limit is shown as a black dotted line. KIX is orange with largest shifted residues in red. cMyb is shown as a green ribbon and MLL is shown as a blue ribbon. Bar graphs and models are for KIX bound to MLL compounds at four-fold excess of A) MLL WT, B) MLL mimic 6, and C) MLL mimic 8.

site. The canonical MLL binding site to KIX is approximately residues 625 – 640, comprising the entire α2 helix, and a C-terminal segment of the α3 helix from approximately residues 657 – 669. The synthetic MLL WT fragment used here results in significant shifts throughout both regions, although it is notable that some of the largest shifts – T615, L621, K622, D623 – are adjacent to the pocket.

MLL mimic 6, Figure 5.7B, shares significant shifts with MLL WT at residues E627, V630, Y631, and A632. MLL mimic 6 also results in -terminal α3 helix shifts at A655 and L665, though these do not overlap with the larger shifts of the WT. MLL mimic 8, Figure 5.7C, shares significant shifts with MLL WT at residues 619 – 638 and K657.

As seen in with other compounds, the MLL mimics share significant similarity in the residues affected, including H595, H603, V609, K622, E627, V630, Y631, and A632, several of which are not shared with MLL WT. MLL mimic 8 results in larger shifts than MLL mimic 6. The locations of the shifts between the two, however, are largely overlapped.

CREB mimics interact with KIX

HSQC spectra of ^{1}H-^{15}N labeled KIX apo and in the presence of a four-fold excess of mimics of the KID domain of CREB are shown in Figure 5.8. All compounds resulted in small shifts to KIX residues. The largest shifts are labeled in black.

Figure 5.9 shows the chemical shift changes in KIX bound to CREB mimics with surface models, adapted from PDB file **1KDX** (250), shown at right with the largest resultant chemical shifts of KIX in Figure 5.8. The KID domain is shown as a cyan cartoon. KID binds KIX at the cMyb binding pocket, where the αB helix interacts with the

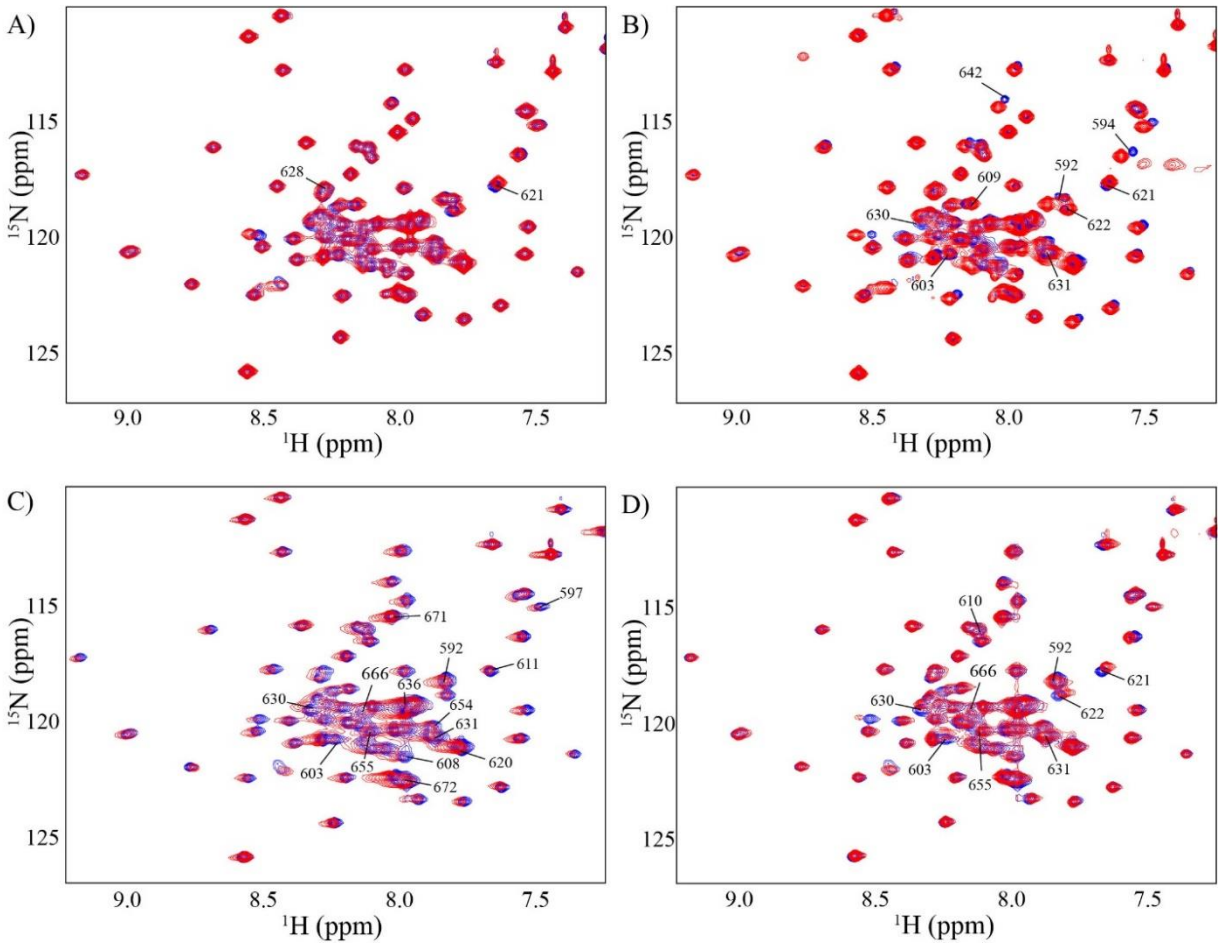


Figure 5.8: NMR spectra of KIX bound to CREB mimics. KIX apo (blue) and bound to four-fold excess (red) of A) compound 78a, B) compound 79b, C) compound 3_30b, and D) compound 2_133b.

hydrophobic pocket created by residues D599, H603, H606, Y650, L653, A654, K657, and I658 with additional lesser interactions with KIX residues S602, and A655. Synthetic CREB mimics, despite resulting in small chemical shifts, share some significant chemical shift with a known interacting residue of KIX. Compound 79b and 2_133b cause a significant shift at H603, also an important residue in cMyb binding. Compound 3_30b affects KIX residues H603, L653, A654, A655, and I658. Compound 78a results

in shifts above the resolution limit only at L621 and N628, which are generally associated with binding of MLL rather than CREB.

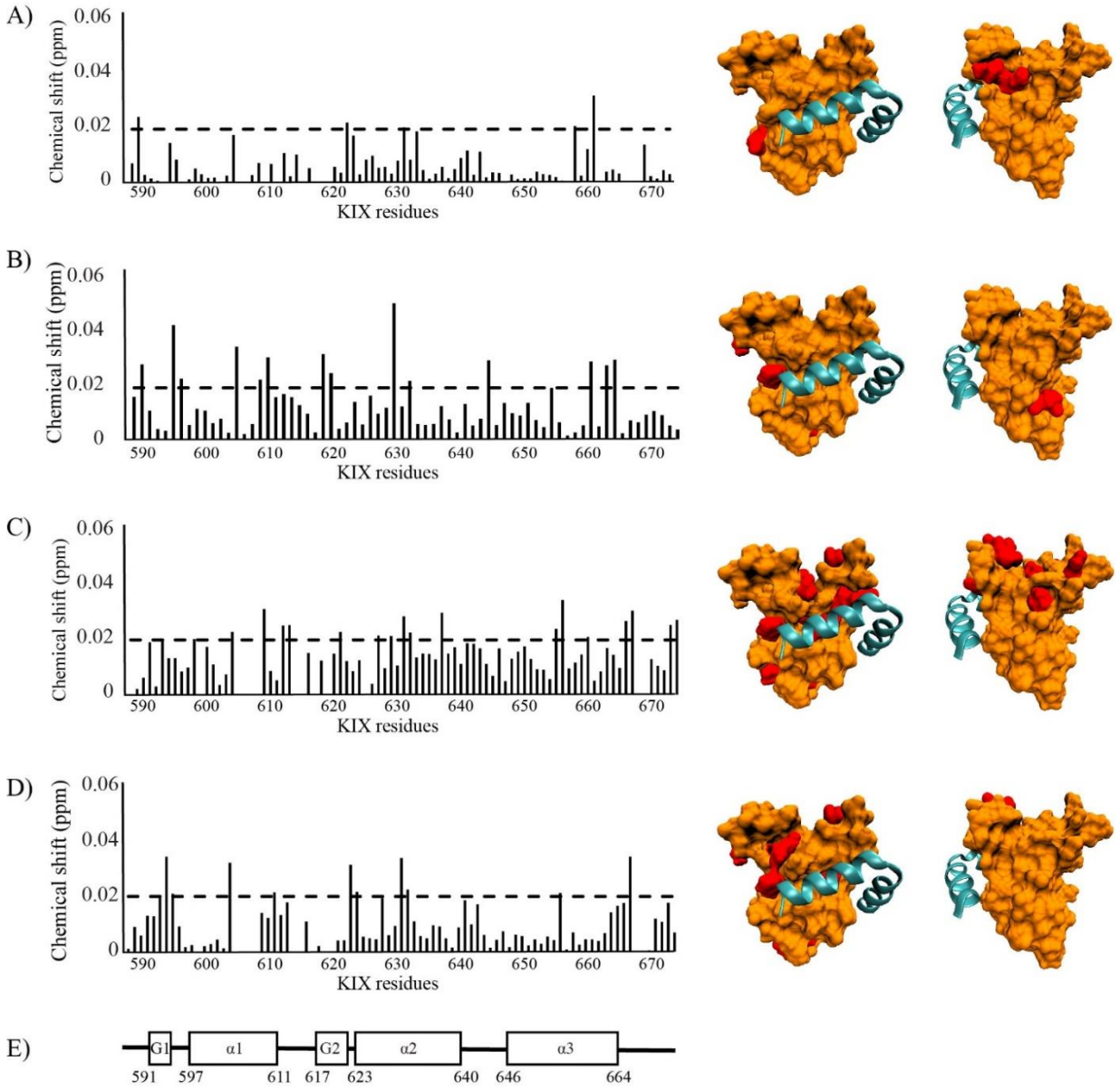


Figure 5.9: Chemical shifts of KIX bound to CREB mimics. ^{1}H - ^{15}N HSQC NMR chemical shift changes of cMyb-bound KIX are shown at left and residues most affected on a 3D model at right. Resolution limit is shown as a black dotted line. KIX is orange with largest shifted residues in red. cMyb is shown as a green ribbon and MLL is shown as a blue ribbon. Bar graphs and models are for KIX bound to cMyb constructs at varying ratios
At right, chemical shift changes of KIX bound to compounds and at right KIX, orange bound to compounds, cyan: A) Stapled peptide 78a, B) Stapled peptide 79b, C) 3_30b, and D) 2-133b, E) structural map of KIX.

The compounds share significant chemical shifts between each other. Residue W592 is significantly shifted in 79b, 3_30b, and 2_133b; L621 is significantly shifted in 78a, 79b, and 2_133b; V630 and A631 are significantly shifted in 79b, 3_30b, and 2_133b; and E666 is shifted in 3_30b and 2_133b.

CHAPTER SIX: DISCUSSION

Note to the readers

Sections of his chapter are comprised of adapted published text, used with the permission of the publishers, (see Appendix C) (60).

In this project we investigated the effects of mutations in TAD2 and PRR that result in changes to the DNA binding affinity of p53, despite neither of these domains interacting with DNA, via modulation of the autoinhibitory intramolecular interaction with DBD. These disordered segments were shown in our lab's previous work to engage in dynamic interactions with DBD with an unclear degree of sequence specificity. Our goal has been to find what features of TAD2 and PRR are responsible for their respective interactions with DBD.

The TA'2-DBD interaction is composed of electrostatic and nonelectrostatic features and modulates ion release

We find that the intramolecular interaction between the TAD2 and DBD domains of p53 is disrupted by mutations targeting multiple types of interactions. Alanine substitutions of TAD2's negatively charged residues, ND DE, increased consensus DNA free energy of binding by −1.99 kcal/mol relative to ND WT, suggesting that electrostatics play a large role in the intramolecular interaction and autoinhibition of DNA binding. Alanine substitutions of nonpolar residues, ND NP, increased DNA free energy

of binding by −1.89 kcal/mol, suggesting a nonelectrostatic component. A targeted substitution of W53/F54 to Q53/S54, ND QS, chosen because of its established ability to disrupt other important TAD2 interactions (95, 145, 146), increases consensus DNA free energy of binding by −1.49 kcal/mol. The sum of the effects of the ND DE and ND NP mutants on the autoinhibition of DNA binding is 1 kcal/mol greater than the effect of ND WT. This indicates the possibility of cooperativity between the acidic, nonpolar, and aromatic residues of TAD2 to inhibit DNA binding.

A previous analysis of transcription factor-DNA complexes using the counterion condensation theory, notably HMG boxes and homeodomains, showed the salt-dependent component of binding was similar for specific and nonspecific DNA and the salt-independent components, attributed to hydrogen bonds and van der Waals interactions, were the drivers of specificity (13). By contrast, our study shows that p53 has a larger salt-dependent component of binding for consensus DNA versus scrambled DNA; according to the counterion condensation theory, this represents a dependency on entropy derived from ion release when p53 binds consensus DNA that is not present when it binds the scrambled DNA sequence. Critiques of the counterion condensation theory have noted that ion release is not the only energetic component of the salt-dependent binding affinity, nor is the salt-dependent component entirely entropic (130, 170, 252, 253). Our data is discussed in the context of entropy derived from predicted ion release; however, our Van’t Hoff data and previous ITC data (74) suggests a large enthalpic component, meaning the difference we see is a combination of ion release and other energetic components that drive specificity.

Our results show how the presence of TAD2 decreases the apparent number of ions released by DBD when binding consensus DNA. We propose that the interactions between the positively charged residues in the DNA binding pocket and the negatively charged residues of TAD2 reduce the need for ionic interactions between those same positive charges of DBD and negatively charged solutes. This conclusion is consistent with the differences in ion release we see between the ND DE and ND NP mutants. The ND DE mutant releases almost the same number of ions as DBD. By eliminating the negative charges of TAD2 we have eliminated the intramolecular screening and now ions from the solute reestablish their positions around the positively charged amino acids of the DBD. The ND NP mutant has the negatively charged residues of TAD2 present, and the ion release is almost identical to that of ND WT. Thus, we show that the differences in ion release between DBD and ND WT are primarily moderated by negatively charged residues in TAD2. We also think the differences in the salt dependence of DNA binding between DBD and ND WT could be relevant for p53 function. Prior to DNA damage TAD1 is primarily responsible for the interaction with MDM2 that leads to p53 degradation (254). However, following DNA damage, posttranslational modifications regulate numerous interactions between TAD2 and other cofactors (146, 255-257). It is reasonable to expect these other interactions will compete with the autoinhibitory function of TAD2, resulting in increased DNA binding.

The fact that TAD2 remains dynamic when bound to DBD implies a potentially multivalent interaction, where if several nearby interactions are possible, the binding affinity of the total segment is increased even if specificity is low and individual binding sites are weak and interchangeable. The dynamic bound state provides the obvious

advantage that, while binding affinity is weak, there is little entropic penalty to binding because TAD2 does not fold. Similarly, the closed conformation of ND, where TAD2 interacts with DBD, should experience a small entropic penalty from the release of TAD2, but may experience a greater penalty from the association of ions from the environment.

One of initial goals in this project was to understand how transactivation might be controlled by autoinhibition. An interpretation of our results gives a possible mechanistic basis for previous observations of p53 mutations in the context of cancer. Mello et al. found that mice with an introduced F53Q/F54S mutation to p53 (equivalent to human W53/F54) had unexpectedly greater survival rates against pancreatic cancer, suggesting a protective role for these mutations (258). A clear structural explanation is that this mutant is less autoinhibited and thus has greater transactivation of some pro-apoptotic genes than the WT.

PRR participates in frustrated autoinhibition of DBD

Our results suggest that there are multiple semi-independent mechanisms for autoinhibition of DNA binding. We find that the PRR 33 GS and PRR-DBD mutants, which are used to represent the autoinhibition originating from the TAD2-DBD and PRR-DBD interactions, respectively, increase consensus free energy of binding by 0.58 and 1.22 kcal/mol. The total autoinhibition when both are intact in ND WT is greater than either alone, meaning their activities are not mutually exclusive. However, the autoinhibition from each subdomain when added together is greater than that seen in ND WT, suggesting frustration.

The PRR appears to have at least two features that affect DNA binding and that act simultaneously: the W91-R174 interaction inhibits DNA binding, and the rigidity of PRR frustrates autoinhibition. The ND W91A mutant increases consensus DNA free energy of binding by -1.32 kcal/mol relative to ND WT which we suppose is due both to disruption of the W91-R174 interaction and to the increased conformational variability of the TAD and resulting decreased C_{eff} of TAD2 when it is not projected towards DBD. The PRR PtoG and PRR NptoG mutants decrease consensus DNA free energy of binding by 0.45 and 1.41 kcal/mol, respectively, relative to ND WT suggesting an increase in autoinhibition when the PRR becomes more flexible and allows TAD2 greater access to the DBD. Note, both these mutants maintain the W91-R174 interaction that orients TAD2 towards DBD.

Our data suggests a model where there are several non-open states of ND with mildly graded levels of autoinhibition, any of which may be stimulated by any number of modifications that affect TAD2 and PRR, such as Pin1-mediated prolyl isomerization at P82 that occurs following T81 phosphorylation (118, 126), Cyp18 prolyl isomerization of P71 (127), or phosphorylation of S46 and T55 (75). The participation of multiple domains in multivalent, allovalent, or frustrated autoinhibition is common (26). In Bruton's tyrosine kinase (BTK), for example, Joseph et al. recently characterized a graded autoinhibitory mechanism controlled by several semi-independent regions of BTK and in which a proline-rich region engaged in inhibition of autoinhibition and acted as a switch between open and closed states (259).

Our data presented here cannot explicitly determine if the decrease in DNA binding affinity of the targeted PRR mutants is due to an increase in the TAD2-DBD

interaction. It could be that the PRR competes with TAD2 for a binding site on DBD or that the PRR mutants interact with DBD strongly. Our SEC analysis suggests a more closed conformation in the PRR PtoG and PRR NptoG mutants. These are not definitively due to a collapse of the TAD onto the DBD but may rather be due to a larger scale conformational shift. However, it is notable that the changes in R_S are relatively small (<3 Å) and are of comparable size to the differences seen in TAD2 mutants in Chapter 2. NMR studies that compare chemical shift changes in DBD for ND WT versus ND PRR PtoG, as an example, could provide the explanation.

The PRR 33 GS mutant is intended to represent an ND variant that completely ablates the sequence specific activities of the PRR, yet equivalently separates TAD2 and DBD. However, the 33-residue GS linker used here will be more flexible than the wild type PRR with the result that the average end-to-end distance will be smaller and the likelihood for intradimer and intratetramer TAD2-DBD interaction will be decreased relative to ND WT. The effect of PRR flexibility on TAD2 is described in detail in Chapter 4. We acknowledge this issue while also noting that the C_{eff} for TAD2-DBD interactions is highest for intramolecular interactions regardless of the substitution of a 33 GS linker for the wild type PRR. A potential alternative for substituting the PRR is to substitute a GS linker whose pI distribution is equivalent to that of the PRR. This is a 54-residue linker if we assume a persistence length of 3 Å. Potential occlusive effects of the GS linker used are not considered here.

Lastly, environmental effects on secondary structure and disorder are not considered in our analysis. PPII helix conformation is likely increased at higher ionic strengths, for example, which may lead to a greater frustrating effect on the TAD2-DBD

interaction (260). The persistence length of PRR in increasing ionic strengths can be assessed in future studies via paramagnetic resonance enhancement or by FRET. Ionic strength likely affects all protein structures and pressures even random coils into more ordered states (261), so it may be that the system is dominated by different autoinhibitory features at different ionic strengths.

TAD2 and PRR provide submaximal autoinhibition

The introduction of a flexible linker between domains reduces the frustration between TAD2 and PRR. The ND T+20 mutant allows the intact TAD2 subdomain to interact with DBD freely while the intact PRR domain also interacts with DBD, decreasing the consensus DNA free energy of binding by 1.68 kcal/mol relative to ND WT. The insertion of a flexible linker that separates PRR and DBD, ND P+20, increases the consensus DNA free energy of binding by -01.22 kcal/mol. The decrease in autoinhibition upon insertion of this linker supports a model where the W91-R174 interaction depends on a high C_{eff}. The remaining autoinhibition for this mutant may originate in TAD2-DBD interactions, PRR-DBD interactions, or a combination of both.

The mutants highlight the relevance of C_{eff} for the TAD-DBD interaction: the proximity of the subdomains to their interaction sites on DBD has a large effect on autoinhibition. Our modeling of the C_{eff} of TAD2 for DBD interactions demonstrates that intramolecular interactions are likely favored, although intradimer and intratetramer interactions are possible. The rigidity of PRR has a large effect on C_{eff} where a flexible linker versus a rigid one yields an C_{eff} for TAD2 of 13.9 mM versus 5.4 mM in the intramolecular interaction. The PRR 33 GS mutant described in Chapter 3 for which the

entire PRR is replaced with a GS linker is equivalent to this model where the C_{eff} of TAD2 is 13.9 mM. Shown in Chapter 3, the autoinhibition provided by $\Delta G_{DBD} - \Delta G_{PRR\ 33\ GS}$ is -2.45 kcal/mol, which is slightly greater than the total autoinhibition from ND DE and ND NP combined. The PRR 33 GS mutant likely produces its effect partly from an increased C_{eff} of TAD2 yet is somewhat decreased by the absence of the orienting W91-R174 interaction.

We find that TAD2 and PRR are present in most mammals with low sequence conservation but high feature conservation: notably TAD2 maintains aromatic residues in a similar position and maintains its high negative charge; PRR has a similar fraction of charged, polar, hydrophobic, and proline residues throughout mammals. Both subdomains are more variable but still recognizable in many birds, suggesting the potential for the emergence of this frustrated system at the time of this group's emergence. There are not currently any functional studies that address PRR evolution.

Cooperation and frustration in autoinhibition vary with ionic strength

As described in Chapters 2 – 4, the ΔΔG value of binding of DBD versus another p53 fragment to consensus or scrambled DNA is a means to quantify autoinhibition. Values for quantified autoinhibition relative to DBD are shown in Table 6.1 for all ionic strengths.

At low ionic strength, binding affinity may not be reliable as ion behavior near DNA is not predictable: DNA may bind ions very tightly so that even at low ionic strength where shielding from buffer ions is low, shielding of protein-DNA interactions by bound ions is equivalent or higher to that at higher ionic strengths as DNA and/or protein

approaches the minimum number of ions bound. Conformational changes with greater hydrophobic burial may also be possible at very low ionic strength. Our data at low ionic strength (85 mM) show ND DE, ND NP, and PRR 33 GS binding consensus DNA more tightly than DBD.

Table 6.1: ΔΔG values across ionic strengths (all in kcal/mol)

Consensus DNA

	85mM	125mM	145mM	165mM	185mM	205mM	225mM
DBD	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ND P+20	NA	-0.94	-1.80	-1.83	NA	NA	NA
ND DE	0.01	-0.97	-1.03	-0.55	-0.95	-0.95	-0.77
ND NP	0.24	-0.74	-1.13	-0.57	-0.60	-0.23	-0.44
ND QS	-0.21	-1.20	-1.49	-0.87	-1.26	-1.06	-1.27
ND W91A	0.00	-1.45	-1.70	-1.82	NA	NA	NA
PRR-DBD	-0.27	-1.48	-1.80	-1.62	NA	NA	NA
PRR 33GS	0.44	-2.24	-2.48	-2.01	NA	NA	NA
ND WT	-0.71	-2.29	-3.02	-2.49	-2.28	-1.95	-2.12
PRR NptoG	-1.08	-2.27	-3.25	-2.34	NA	NA	NA
ND T+20	NA	-2.92	NA	NA	NA	NA	NA
PRR PtoG	-1.84	-3.25	-3.68	-3.46	NA	NA	NA

Scrambled DNA

	85mM	125mM	145mM	165mM	185mM	205mM	225mM
DBD	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ND P+20	NA	-0.54	-0.88	-0.74	NA	NA	NA
ND DE	-0.61	-0.12	-0.46	-0.43	-0.15	-0.24	-0.11
ND NP	-0.59	0.05	-0.64	-0.63	-0.72	-0.80	-0.53
ND QS	-1.19	-0.51	-0.63	-0.65	-0.34	-0.71	-0.41

Table 6.1 (Continued)

ND W91A	-0.98	-0.71	-0.84	-0.79	NA	NA	NA
PRR-DBD	0.46	-0.21	-0.74	-0.57	NA	NA	NA
PRR 33GS	-0.86	-0.88	-1.06	-0.73	NA	NA	NA
ND WT	-1.63	-0.87	-1.38	-1.39	-1.05	-1.08	-0.99
PRR NptoG	-1.57	-1.43	-1.41	-0.98	NA	NA	NA
ND T+20	NA	-0.95	NA	NA	NA	NA	NA
PRR PtoG	-2.75	-2.47	-2.38	-8.92	NA	NA	NA

*Note, each column is the ΔG value of DBD minus the ΔG of the fragment in the row header.

The total inhibition of ND DE and ND NP does not equal the inhibition seen in ND WT or PRR 33 GS at any ionic strength shown here when binding consensus DNA, although this is inconsistent when binding scrambled DNA, suggesting either that their effect remains cooperative regardless of ionic strength or that the inhibition from PRR is present regardless of ionic strength. Similarly, the total autoinhibition from TAD2 + PRR is greater than that of ND WT at all ionic strength, implying that frustration is only mildly salt-dependent.

KIX interacts with binding partner mimics and inhibitors

In collaboration with the lab of Dr. Cai, we conducted NMR experiments to examine the interaction sites of synthetic peptides with the KIX domain of the coactivator CBP. These synthetic peptides utilized sulfonyl groups and stapling to mimic the structure of natural disordered transactivation domains but with increased helical content in an attempt to create an inhibitor to outcompete natural binding partners of KIX.

In general, the mimics tested here cause shifts in KIX both in the expected binding site and outside of it. These effects may be related to the excess of compounds used, generally four-fold, that leads to interactions outside of the primary binding site or to conformational changes in KIX distant from the binding site. Both are reasonable outcomes for mimics of these proteins that agree with the plasticity seen in natural partners of KIX. Given that the purpose of these compounds is to inhibit the interaction of the natural transcription factor partners with KIX, effects on KIX outside of the primary expected binding site are not inherently unwanted. Instead, *in vivo* experiments will be used to determine the effectiveness of these compounds at preventing KIX-partner interactions.

cMyb WT and mimics cause shifts in the primary binding site of KIX despite modifications to several of the residues of the mimics that contact the shifted KIX residues. The mimics also cause shifts in the MLL binding site. Correlation analysis shows that the mimics' effects on KIX are more similar to each other than either is to the cMyb WT peptide. The LCA inhibitors produce shifts with greater magnitude, on average, than the cMyb mimics. These shifts are widespread across the surface of KIX, suggesting either significant nonspecific binding or largescale conformational changes in KIX. Low intensity peaks in the bound state of KIX with the LC-A-122-3 peptide suggest this reaction did not fully reach saturation.

The MLL mimics produce large shifts both within and outside of the MLL binding site. The CREB mimics produce small shifts despite tight binding demonstrated by FA.

Future directions

Transcription factors require oligomerization to achieve target specificity (58). In the case of p53, monomeric DBD binds a degenerate 5-base pair sequence, and the tetramer requires a 20-base pair sequence whose half sites are generally not more than 13 base pairs apart (262). While the number of possible binding sites in the eukaryote nucleus is still immense, tetramerization greatly reduces the number of high affinity sites even as it increases binding affinity through cooperative effects (45). Additionally, it is notable that mutations in the tetramerization domain of p53 diminish the distribution of p53 tetramers in solution and are strongly represented in some cancers (Figure 1.1) (263). However, have tetramerization-deficient mutants displayed decreased specificity? If the TAD-DBD interaction affects oligomerization, we might expect a large effect on specificity.

Our interpretation of previous NMR data suggests that TAD2 interacts with the DNA binding pocket of DBD (74). Thus, we supposed PRR might interact with DBD residues near the C-terminus of PRR, at residue 93, and we noticed the proposed PRR-DBD interaction site overlaps with the canonical DBD dimerization and tetramerization sites, Figure 4.4. Furthermore, an earlier study noted that the presence of the five C-terminal-most residues of PRR (residues 89-93) was enough to increase the subunit exchange rate of tetrameric p53, suggesting that the W91-R174 interaction influence oligomerization of p53.

Crystal structures have reported similar but not identical DBD dimerization site residues when bound to DNA that, being generous in their inclusion, include P177, H178, H179, E 180, R181, R174, M243, and G244 as well as several non-contacting

residues important for stability (38, 264). An alternative dimerization interface in the absence of DNA may also exist that involves residues Y103, G105, S106, Y107, P152, P177, H178, H179, N239, C242, M243, D259, N263, L264, L265, although its functional relevance is not yet clear (265). An alternative organization of the tetramer in the absence of DNA has also been proposed (266). NMR data published by our lab suggests the DNA binding pocket of DBD interacts with TAD2, with smaller chemical shifts between loops 2 and 3 representing possible interaction sites with PRR. These residues overlap with the dimerization site, suggesting a possible mechanism for inhibition of oligomerization.

Despite this, our FA data for all p53 fragments that contain mutations in PRR demonstrate the same Hill coefficient trend as other fragments: ~1.8 – 2 when binding consensus DNA and ~1 when binding scrambled DNA. This suggests that the cooperativity of the p53 dimer upon binding target DNA is not diminished by mutations to the PRR; however, it does not provide information about dissociation rates of p53 from DNA, which may show changes upon mutation of the PRR.

Our data shows a notable decrease in specificity in PRR 33 GS relative to other fragments as ionic strength increases. This may be an artefact of altered sensitivity to ionic strength that originates from any of the domains present. Alternatively, it could be a gain in stability of the tetramer bound to scrambled DNA.

Relatedly, as discussed in Chapter Four, the TAD2-DBD may occur within a monomer, dimer, or tetramer. Our analysis of the C_{eff} suggests that the intramolecular interaction is likely favored within a monomer but that interdimer or intertetramer interactions are possible. NMR studies that measure dynamics of TAD2 may reveal

complex interactions between p53 units in the tetramer and a further layer regulation of DNA binding.

Functional *in vivo* effects of the intramolecular interaction are still unknown. Our analysis of the interaction of TAD2 with DBD suggests the modulation of an energetic component of DNA binding. Previous studies of protein-DNA binding have noted the driving energetic component of the reaction (either entropy or enthalpy) is either defined by the conformation change of the two components upon binding or may be variable based on the DNA sequence (13, 267). Thus, we can imagine the TAD2-DBD interaction to have a functional effect on DBD binding to promoters that could be related to the inherent flexibility of the DNA molecule secondary to the distance between p53 half sites, flanking regions, or base pair identity within the target sequence. However, whether the intramolecular interaction affects promoter selection is still an open question. Promoter occupancy of the various mutants following cellular stress may change based on the nature of the TAD2 or PRR mutation.

Additionally, given the promiscuity of both TAD and DBD, it may be that the intramolecular interaction discourages many other interactions in the cell. *In vivo* studies with fluorescently labeled p53 are expected to show an increase in the time between cellular stress and migration of p53 to the nucleus for those mutants that disrupt the intramolecular interaction, such as with ND P+20. Conversely, those that strengthen the intramolecular interaction may migrate more quickly to the nucleus but transactivate genes more slowly than the WT, such as with ND T+20. These linker mutants would be ideal for such studies due to their intact TAD2 and PRR domains.

CHAPTER SEVEN: METHODS/PROTOCOLS

p53 fragments, plasmid design, and subcloning

The human residue numbers of p53 fragments, as well as their molecular weights, extinction coefficients, and isoelectric points are shown in Table 7.1.

Table 7.1: p53 fragments

Protein fragment	Molecular weight (Da)	Extinction coefficient	Isoelectric point
DBD (94 – 312)	24545	18910	8.83
PRR-DBD (61 – 312)	27798	24410	8.53
ND WT (1 – 312)	34565	35410	5.51
ND DE	34229	35410	6.51
ND NP	34183	29910	5.51
ND QS	34447	29910	5.51
ND PRR 33GS	33643	29910	5.61
ND PRR W91A	34428	29910	5.51
ND PRR PtoG	34124	35410	5.51
ND PRR NptoG	34381	35410	5.51
ND T+20	36769	35410	5.51
ND P+20	36034	35410	5.51

All p53 fragments were ordered from Eurofins Genomics in a pUC vector with a 5’ BamHI cut site (G^GATCC) and 3’ EcoRI cut site (G^AATTC). Synthesized genes were generally lyophilized and resuspended in Tris-EDTA to a concentration of 100 mM. From these, a 100 µL stock of 2 ng/µL was created by diluting with nanopure water. Fragments in pUC vectors were subcloned into a pGEX 6-P2. Shown in Figure 7.1, the

pGEX 6p-2 vector contains a tac promoter site, followed by a GST tag, followed by an HRV3C protease target sequence (LEVLFQ^GP) and a multiple cloning site where a BamHI cut site is 5' to EcoRI cut site. This vector confers resistance to ampicillin and is sequenced using the GEX forward site. Synthesized genes in pUC vectors were digested with BamHI-HF and EcoRI – HF restriction enzymes and subcloned into the pGEX vector that was likewise digested.

As an example, if both plasmids are at a concentration of 250 ng/μL, Table 7.2 demonstrates the amounts of each ingredient for digestion in μL:

Table 7.2: Example DNA digestion for pUC and pGEX vectors for subcloning

	pUC vector with insert				pGEX vector			
	Uncut	BamHI cut	EcoRI cut	BamHI + EcoRI cut	Uncut	BamHI cut	EcoRI cut	BamHI + EcoRI cut
Plasmid DNA	1	1	1	10	1	1	1	10
10X buffer	1	1	1	2	1	1	1	2
BamHI	-	1	-	1	-	1	-	1
EcoRI	-	-	1	1	-	-	1	1
npH_20	8	7	7	6	8	7	7	6
Total	10	10	10	20	10	10	10	20

After digestion, DNA is heated to 65°C for 20 minutes to deactivate the EcoRI enzyme and then allowed to cool to room temperature. An agarose gel is made by the following:100 ml Tris borate EDTA is combined with 1 g low melting temperature agarose (Thermo Fisher, Waltham, MA) and heated to dissolve the agarose. The agarose mixture is allowed to cool to ~50° C and then 1 μL EtBr is mixed in.

Gels are run on a power source at 100 mV for approximately one hour and migration of digested DNA is visualized on a UV illuminator. DNA fragments containing the double cut pGEX vector and the double cut p53 insert from the pUC vector are excised from the gel using a razor and purified from the gel using a Qiagen (Germantown, MD) miniprep gel extraction kit and the associated protocol.

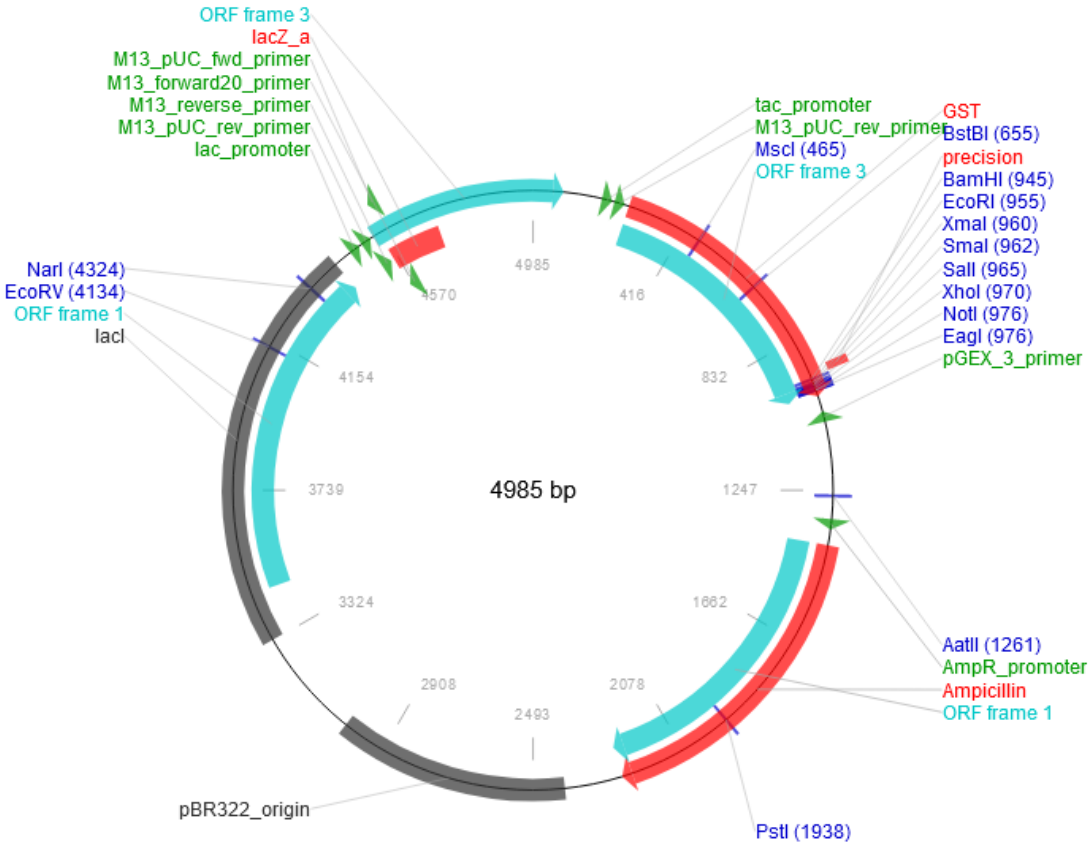


Figure 7.1: Map of pGEX 6p-2 vector. Adapted from Addgene.com. Synthesized genes were transferred from pUC vectors into pGEX 6p-2 vectors using the multiple cloning site (right hand top, shown in blue text).

Purified empty pGEX vector and p53 insert are ligated to form a whole plasmid. Amounts of DNA used for the ligation reaction may be varied, but we have found that a 1:3 molar ratio of plasmid to insert is preferred for this system. Actual ng values are obtained from nebiocalcμLator.com, and all ligation reactions were carried out using 25

ng of pGEX DNA. As an example, if double cut empty pGEX plasmid with a length of ~6000 bp is recovered at 10 ng/μL and the double cut insert with a length of ~950 bp is recovered at 5 ng/μL, ligation is performed using the volumes in μL shown in Table 7.3:

Table 7.3: Example ligation reaction

	Experimental	Positive control	Negative control	Negative control
Empty pGEX vector	2.5	2.5	2.5	2.5
pUC insert	2.37	-	-	-
pGEX insert	-	2.37	-	-
5X Buffer	4	4	4	4
T4 ligase	1	1	1	-
npH_20	10.13	10.13	12.5	13.5
Total	20	20	20	20

The ligation reaction was allowed to proceed at room temperature for one hour. This plasmid DNA is transformed into XL1 blue cells, which are nuclease and recombinase deficient.

All transformations mentioned henceforth follow the same procedure. In a microfuge tube, plasmid DNA was mixed with *E.* coli in a 1:10 ratio and allowed to incubate on ice. Both volume and incubation times vary with the strain: BL21 (DE3 Rosetta) cells use 2:20 μL and incubate 5 minutes; Neb5α cells use 5:50 μL and incubate for 15 minutes; XL1 blue cells use 5:50 μL and incubate for 30 minutes.

Cells were heat shocked in a water bath at 42°C for 30 seconds and then cooled on ice for 2 minutes. Immediately, SOC media without an antibiotic was added at a 5:1 ratio to the volume of cells used, i.e., 250 μL SOC media for 50 μL of cells. Microfuge

tubes were positioned upright in a shaking incubator at a speed of 200 rpm for one hour, after which cells and SOC media were dispensed onto an agar plate containing ampicillin and allowed to grow in a standing incubator at 37°C for approximately sixteen hours. In the case of subcloning, we expect a small number of transformed colonies on the experimental plate and positive control plate and no colonies on either negative control plate.

Post transformation and overnight growth, a single colony of XL1 blue cells is used to inoculate 5 mL of LB+24 media with ampicillin. Cells are allowed to grow overnight and then are processed using a Qiagen miniprep kit and associated protocol. DNA is sequenced by Eurofins Genomics. Correct sequences are subsequently transformed into Neb5α cells for creation of long-term high-quality stock solutions and into BL21 cells for protein expression.

Protein expression

Proteins are produced in BL21 (DE3 Rosetta) *E*. coli. After transformation, plating, and growth on an ampicillin agar plate overnight, a single colony is used to inoculate 25 mL of modified minimal media that is made according to the following 2 L recipe and then filtered for sterility:

<u>Growth media</u>

80 mL “MDMX salts” (see below)

4 mL “O” solution

2 mL “S” solution

400 µL Vitamin B1

4 g D-glucose or 4 g ^{13}C D-glucose

2 g NH_4Cl or ^{15}N NH_4Cl

1 mL ampicillin at 1 mg/mL

QS to 2000 mL in npH_2O and pH 7.32 – 7.36

Recipes for components of growth media are as follows:

<u>MDMX salts</u>

250 mM KH_2PO_4, monobasic

1 M KH_2PO_4, dibasic

625 mM NaCl

QS to 500 mL with npH_2O, pH 7.5

<u>"O" solution</u>

10 mL metal stock solution

264 mM $MgCl_2$

QS to 500 mL

<u>Metal Stock solution</u>

8 mL 12N HCl

5 g $FeCl_2$

184 mg $CaCl_2$

4 mg H_3BO_3

40 mg $MnCl_2$

18 mg $CoCl_2$

4 mg $CuCl_2$

340 mg $ZnCl_2$

605 mg Na_2MoO_4

QS to 100 mL with npH_2O

"S" solution

275 mM K_2SO_4

QS to 100 mL with npH_2O

The inoculated cell culture is allowed to grow for 16 hours at 37°C shaking at 150 rpm. Cell density is measured on a Biorad SmartSpec Plus spectrophotometer. This culture is used to start a 1 L culture with a starting optical density (O.D.) of 0.04. Growth is measured every hour until the O.D. is 0.5, at which point 200 µL of 100 mM $ZnSO_4$ is added for a final concentration of 20 µM, and the culture is moved to a shaking incubator at 15°C. After cooling for 15 minutes, 1 mL of 1 M isopropyl β-D-thiogalactoside (IPTG) is added to induce protein expression and the culture is allowed to grow for 20 hours. Then the culture is pelleted in 1 L increments at 8200 g and stored immediately at -80°C until ready for lysis.

Protein purification

To purify p53 fragments, one liter of pelleted cells was resuspended in 25 mL lysis buffer containing 50 mM Tris (pH 7.4), 500 mM NaCl, 2 mM DTT, 0.02% NaN_3 and a fresh tablet of Pierce EDTA-free protease inhibitor (Thermo Fisher, Waltham, MA). Cells were lysed via French press at approximately 1000 psi and centrifuged at 38000 rcf for one hour.

After centrifugation, the supernatant was passed through a GST Fast-Flow Sepharose column (Cytiva, Marlboro, MA) and eluted with 10 mM reduced glutathione, Figure 7.2A. A sample of the supernatant was collected to be run on a polyacrylamide gel, shown as “S” in Figure 7.2B. The pellet was resuspended in 25 mL of the lysis

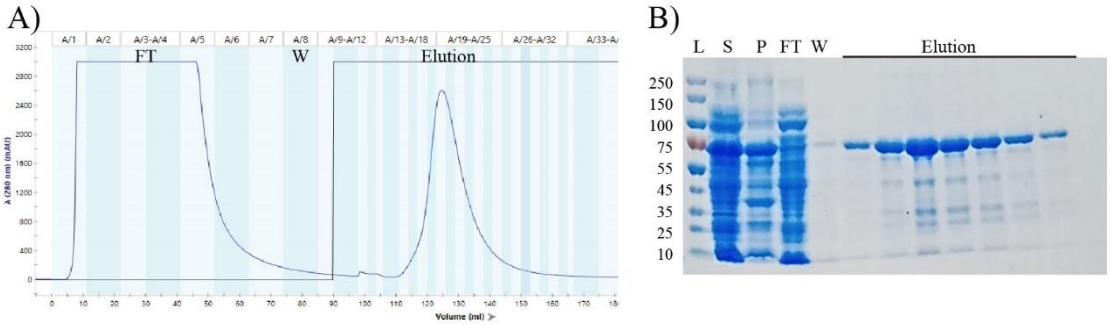


Figure 7.2: Chromatogram and SDS-PAGE gel of GST precleave column purification of ND WT. A) Chromatogram of GST precleave where FT indicates flow through, W indicates wash, Elution indicates the peak that corresponds to the desired protein. B) Polyacrylamide gel where L indicates ladder, S indicates the soluble fraction, P indicates the pellet, W indicates wash, and Elution indicates the peak corresponding to the desired protein.

buffer and ~25 µL of Triton-X-100 and collected for a polyacrylamide gel, shown as “P” in Figure 7.2B. The eluted fractions containing the GST-tagged ND fragments were pooled and incubated with a 1:100 ratio of the HRV3C protease overnight at 4°C to cleave the GST tag. The cleaved GST tag was removed by passing the mixture over a GST column, Figure 7.3. Following separation of p53 and the GST tag, fragments containing the TAD were dialyzed into a no-salt buffer and passed through a Q Sepharose High Performance anion exchange column (Cytiva), eluted in buffer containing 20 mM Tris at a pH of 7 – 8 depending on isoelectric point of the protein, 0 – 1 M NaCl, 2 mM DTT, and 0.02% NaN_3, shown in Figure 7.4. All fragments were analyzed using polyacrylamide gel electrophoresis and protein samples were pooled

and concentrated to 25 – 50 μM and loaded on a 16/600 mm Superdex 75 column (Cytiva,) in a buffer composed of 50 mM NaH_2PO_4 (pH 7), 300 mM NaCl, 1 mM DTT, and 0.02% NaN_3, shown in Figure 7.5. Protein purity was evaluated via SDS-PAGE and concentration assessed using Beer's Law, A = εcl, where A is the absorbance at 280 nm, ε is the extinction coefficient described in Table 1, c is the concentration in moles/L, and l is the pathlength which is 1 cm in this case, using a Nanodrop 1000 Spectrophotometer (Thermo Fisher).

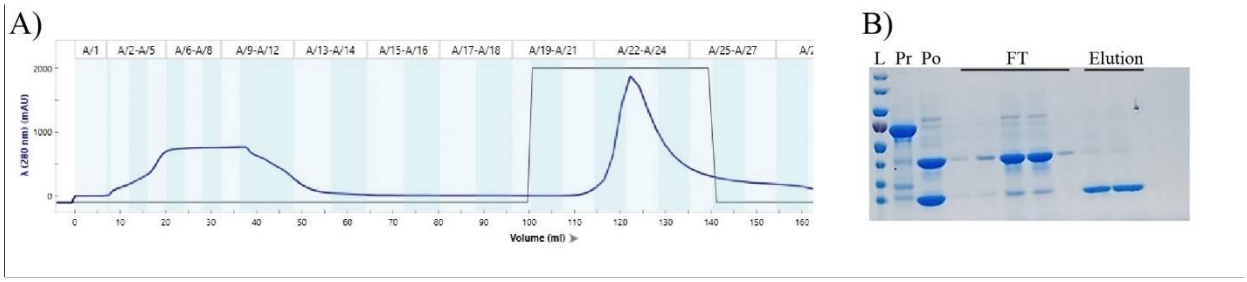


Figure 7.3: Chromatogram and SDS-PAGE gel of GST postcleave column purification of ND WT. A) Chromatogram of GST postcleave after 16 – 24 hours digestion. B) Polyacrylamide gel of GST postcleave fractions where L indicates ladder, Pr indicates precleave, Po indicates postcleave, FT indicates flow through indicates the desired protein, and Elution indicates the cleaved GST tag and other contaminants that bind the GST column.

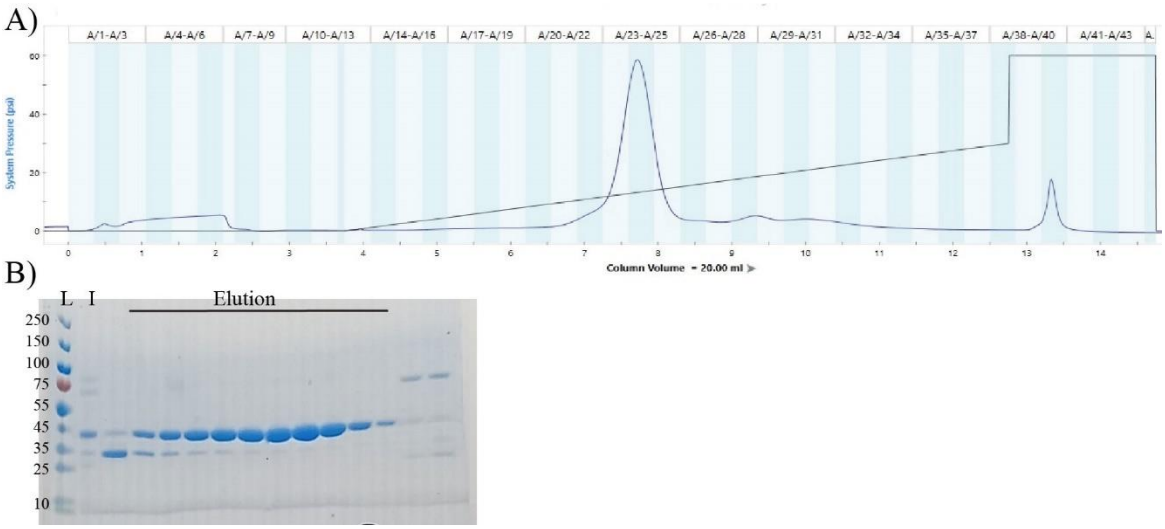


Figure 7.4: Chromatogram and SDS-PAGE gel of Anion exchange column purification of ND WT. Where L indicates ladder, I indicates input, and elution indicates the peak that corresponds to the desired protein.

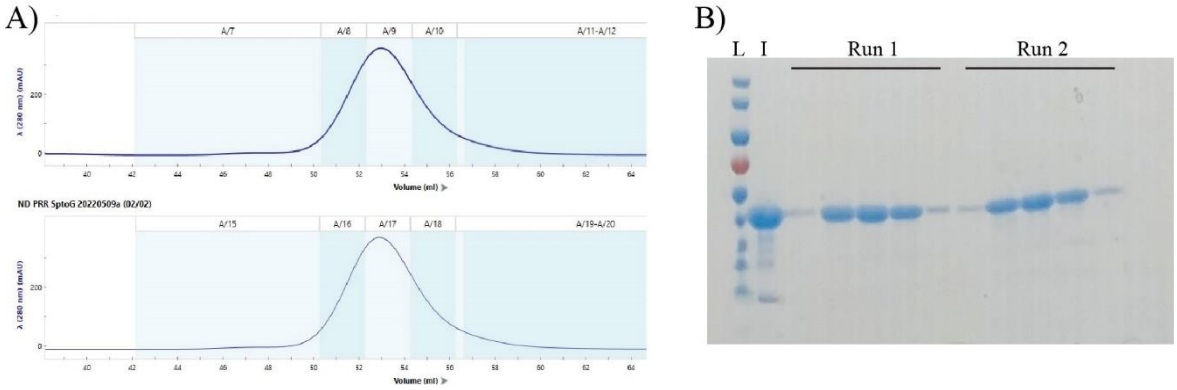


Figure 7.5: Chromatogram and SDS-PAGE gel of SEC column purification of ND WT. A) Two runs on a size exclusion chromatography column. B) Polyacrylamide gel of two runs of SEC where L indicates ladder, I indicates input and Run 1 and Run 2 correspond to the peak of the desired protein elution.

Protein Purification for NMR

The KIX domain of human CBP (residues 586-672) was grown and purified as previously described (226). In brief, the KIX domain of CBP (mouse residues 586 – 672) was purchased from Eurofins (Lancaster, PA) was inserted into a pET28a vector with a six-histidine tag and thrombin cut site N-terminal to the gene using BamHI and EcoRI

cut sites and then transformed into *E*. coli BL21 (DE3) cells. Bacteria was grown at 37°C to an $O.D._{600}$ of 0.6, at which point the cells were transferred to 15°C, induced with 1 mM IPTG, and grown for 22 hours. Cells were pelleted at 8200 g and stored at -80°C. Cells were resuspended in 25 mL of lysis buffer containing one Pierce Protease Inhibitor tablet (Thermo Fisher, Waltham, MA), 50 mM NaH_2PO_4, 300 mM NaCl, 0.02% NaN_3, and 10 mM imidazole, pH 8.0. Cells were lysed using a French press at psi >1000 and lysate was centrifuged at 38000 g. The soluble fraction of lysate was loaded onto a Ni-NTA column (Qiagen, Germantown, MD) and eluted in buffer containing 250 mM imidazole. Protein was then dialyzed overnight into buffer containing 50 mM NaH_2PO_4, 300 mM NaCl, and 0.02% NaN_3, pH 7.0, and the histidine tag was cleaved using a Thrombin CleanCleave Kit (Sigma-Aldrich, Burlington, MA) for two hours at room temperature. Cleaved protein was further purified using a 16/600 mm Superdex 75 size exclusion column (Cytiva, Marlborough, MA). Protein purity was verified using SDS-PAGE and concentration was determined via spectrometry.

Fluorescence anisotropy

Preparation of DNA: HPLC-purified, 6-Carboxyfluorescein (6-FAM) tagged DNA was obtained from IDTDNA (Coralville, IA) as single strands. Double-stranded DNA was annealed by boiling at 95°C for 10 minutes and allowing to cool to room temperature. The sequences used are as follows: consensus 5′ AGACATGCCTAGGACATGCCT and scrambled 5′ TGCCGATCAAAACCGATTCG. Annealing was confirmed using nondenaturing gel electrophoresis. Extinction coefficients for the forward and reverse DNA fragments are 214 and 195, respectively, and the extinction coefficient of 321

UNITS L*mol^{-1}cm^{-1} is used for the annealed sequence. Extinction coefficients were found using https://molbiotools.com/dnacalculator.php or were provided by the manufacturer.

Purified samples of DBD, ND WT, ND DE, ND NP, and ND QS were concentrated to 20 – 200 μM depending on the IS of the buffer and co-dialyzed with DNA twice against a buffer containing 10 mM NaH_2PO_4 (pH 7.4), 100 – 200 mM NaCl, 5 mM DTT, 0.02% NaN_3, and 0.01% Triton-X 100 for a total dilution factor of 1x10^6. 10 nM labeled DNA was aliquoted into Corning™ COSTAR 96-Well Solid Black Polystyrene Microplates (Thermo Fisher) and protein samples were added at increasing concentrations from 1 nM to saturation at 20 – 100 μM for a total volume of 100 μL. Fluorescence was measured using a Synergy H1 microplate reader from Biotek (Winooski, VT) at 25°C, and at 1.5° increments from 21 — 37°C for Van't Hoff analysis. Excitation and emission wavelengths were 485 nm and 528 nm, respectively, with a sample height of 7cm, gain of 50, and shake and delay steps of 30 seconds and 20 seconds, respectively.

Binding affinities were estimated using a cooperative binding model for p53's interaction with consensus DNA as described previously (45) where p53 is evaluated as a dimer:

$$\Delta A = \frac{[A]^2}{K_d + [A]^2} \tag{1}$$

Where ΔA is the normalized anisotropy change, [A] is p53 dimer concentration. Binding affinity to scrambled DNA was calculated using a one-to-one binding model (136):

$$\Delta A = \frac{[A] + [DNA] + K_D - \sqrt{([A] + [DNA] + K_D)^2 - 4[A][DNA]}}{2[DNA]} \quad (2)$$

The Hill coefficient was evaluated using the following equation (136):

$$\Delta A = \frac{[A]^h / {K_D}^h}{1 + [A]^h / {K_D}^h} \quad (3)$$

Where *h* is the Hill coefficient, indicating the cooperativity of the binding event where 1 is a noncooperative event and greater than 1 is a cooperative event. K_D and the Hill coefficient in Equations 1 – 3 were found using the Solver function in Excel. Anisotropy values for individual data points were plotted against the concentration of the p53 dimer and a fit line, generated using the appropriate equation, also plotted against concentration, were compared. The R^2 value of the data points versus the fit line was used to assess how well the fit line predicted individual data points given a K_D value. The solver function of Excel was then used to minimize the R^2 value, resulting in the best fit. For Equation 3, the same method was used, except that both the K_D and Hill coefficient are simultaneously solved while minimizing the R^2 value.

The counterion condensation theory

The counterion condensation theory developed by Record and colleagues expands on the polyelectrolyte theory (133) to estimate ionic contacts and excess ion release for protein-nucleic acid binding (129) using the following relationship:

$$\log(K_A) = \log(K'_A) - N^*\log[\text{Salt}] \quad (4)$$

Where K_A is the association constant, K'_A is the nonelectrostatic component of binding, and $N^*\log[\text{Salt}]$ is the electrostatic component of binding. N is the slope of a double log plot of K_A versus [Salt]. In this theory the electrostatic component of binding refers to the positive entropy associated with ion release (13, 129). It is unclear if this approach can quantitatively discriminate the salt-dependent entropic component of binding from other components, but we think it provides a useful qualitative segregation of components of binding affinity (252, 268). Because of this we refer to these as the salt-dependent and salt-independent components of binding rather than as the electrostatic and nonelectrostatic components. The salt-independent component is inferred from the y-intercept of a $\log(K_A)$ vs log[Salt]. The slope of this plot, N, is further defined as:

$$N = Z\Psi + \beta \quad (5)$$

Where Z is the number of protein-DNA backbone contacts made, Ψ is the fractional number of ions bound by phosphate, 0.7 for short oligonucleotides (152), and β is the number of excess ions released from protein. Our study utilizes only NaCl as the salt. Studies have found that variation of the monovalent cation, which is condensed around and ultimately released from DNA, is unimportant in evaluating ion release (13, 269) although introduction of a divalent cation can have complicated effects on apparent ion release (270). Variation of the anion may affect apparent ion release; however, the change in apparent ion release based on anion identity may reflect on the size of the anion or its relative attraction to water versus the protein side chains and thus varying

the anion is not predicted to reveal additional information about the protein's DNA binding interface (13, 166, 271).

A reevaluation of the theory by Manning and colleagues resulted In the following relationship (153):

$$\log(K_A) = \log(K_0) + \log V + 0.513Z - 0.434 - Z*\log[Salt] \quad (6)$$

Where K_A is the association constant, K_0 is the salt-independent component of binding, V is the reaction volume, and Z represents the number of charged molecules associated with the binding event, which is interchangeable with *N* from Equation 2.

Both these approaches use the section of a double log plot where $\log(K_A)$ versus log[Salt] becomes linear, a range that is uniquely determined for a given protein. In this case, while fluorescence anisotropy was conducted on DBD and ND WT over an IS range of 15 – 225 mM, Supplementary Table 1 (Table S1), the double log plot is linear in the 125 – 225 mM range. Thus, fluorescence anisotropy was only conducted on ND mutants in the 85 – 225 mM range and these were evaluated using the counterion condensation theory from 125 – 225 mM IS.

Experimental enthalpy and entropy estimates were calculated from Van't Hoff plots. These were generated by measuring anisotropy at physiological IS across a range of temperatures as previously described (156). The Gibbs free energy equation can be rearranged to give:

$$Ln(K_A) = \frac{\Delta H}{R} * \frac{1}{T} + \frac{\Delta S}{R} \quad (7)$$

Where ΔH is the change in enthalpy, T is the temperature in Kelvin, ΔS is the change in entropy, and R is the gas constant in kcal/mol. A line is generated from plotting Ln(K_A) versus 1/T, from which the m, slope, represents ΔH/R and the y-intercept represents ΔS/R. Note, this method requires the assumption that ΔH does not change with temperature.

Isothermal titration calorimetry

ITC experiments were performed on a GE MicroCal VP-ITC 200 system instrument. Initial tests used an 85 mM ionic strength buffer composed of 10 mM NaH_2PO_4, pH 6.8, 66 mM NaCl, 0.02% NaN3 and 8 mM BME. Subsequent tests at higher ionic strengths modified the concentration of NaCl only, with ionic strength calculated using the tool provided at https://www.liverpool.ac.uk/pfg/Tools/page23/PepMap.html. Concentrations of p53 DBD in the cell and DNA in the syringe were at concentrations of 5 μM and 12.5 μM, respectively. ND WT was used at a concentration of 50 μM with 125 μM DNA in the syringe. For DBD binding to consensus DNA at 85 mM ionic strength with a K_D of ~1 nM, the c-value = K_A/M is 5,000. Attempts to lower the c-value of DBD with DNA using lower DBD concentrations resulted in noisy ITC data. Stoichiometry typically ranged from 0.19 – 0.25, indicating a 1:4 ratio of protein to DNA. Experiments were performed at 25°C.

Protein and DNA in ITC experiments were co-dialyzed three times for a total dilution factor of $1*10^9$. Experiments were conducted with a 5 μL initial injection of DNA into protein followed by 28 injections of 10 μL for a final molar ratio of 1 to 2.5

protein to DNA. Experiments were conducted in triplicate. Stoichiometry, enthalpy, entropy, and K_D were calculated using MicroCal Origin Software (7.0) using a single site binding model.

Analytical size exclusion chromatography

The Stokes radius of proteins was evaluated using size exclusion chromatography. The process uses the same type of SEC column as described above. Low molecular weight calibration kit standards were obtained from Cytiva and were diluted according to the protocol provided in standard Gel Filtration Buffer and passed through SEC in this same buffer with an ionic strength of 410 mM. Standards were diluted further as described by the kit protocol and run in two batches: One set included ovalbumin and aprotinin and the other included conalbumin, carbonic anhydrase, and ribonuclease, as shown in Figure 2.

Stokes radii (R_H) of the p53 fragments were determined using size exclusion chromatography (SEC). The Cytiva Gel Filtration Calibration Kit LMW was used to generate a calibration curve in a buffer of 50 mM NaH_2PO_4, pH 7.0, 300 mM NaCl, 0.02% NaN_3 using a HiLoad 16/600 mm Superdex 75 column (Cytiva, Marlboro, MA) at 4°C. A high ionic strength buffer was used to reduce binding to the 129nubis129x beads and decrease line broadening. The elution volume of each protein was taken as the average of three injections, each of which contained 0.6 – 0.8 mg/mL of protein. The peak elution volume is used to find the partition coefficient, K_{av}:

$$K_{av} = (V_c - V_o)/(V_t - V_o) \quad (8)$$

Where V_c is the total column volume, V_o is the void volume, and V_t is the elution volume. Kav is plotted against known molecular weights of standards to generate a standard curve, Table 7.4 and Figure 7.6, and the apparent molecular weights of experimental p53 fragments are found using the equation taken from the standard curve.

Stokes radius is found by the following relationship (272):

$$R_s = (0.369 \pm 0.001)*\log(MW) - (0.253 \pm 0.002) \quad (9)$$

Where R_s is the Stokes radius, MW is the apparent molecular weight derived from a graph comparing elution volumes of experimental proteins with calibration kit standards.

Stokes radius tests were conducted on p53 fragments in batches of 3-4 runs at concentrations of 0.6 – 0.8 mg/mL which generally corresponds to 15 – 25 μM. Elution volumes are as described in Chapters Three and Four. We conducted Stokes radius testing on ND WT at 150 mM ionic strength to compare the effect of additional salt on the elution volume. As shown in Figure 2.10, differences in elution volumes for ND WT

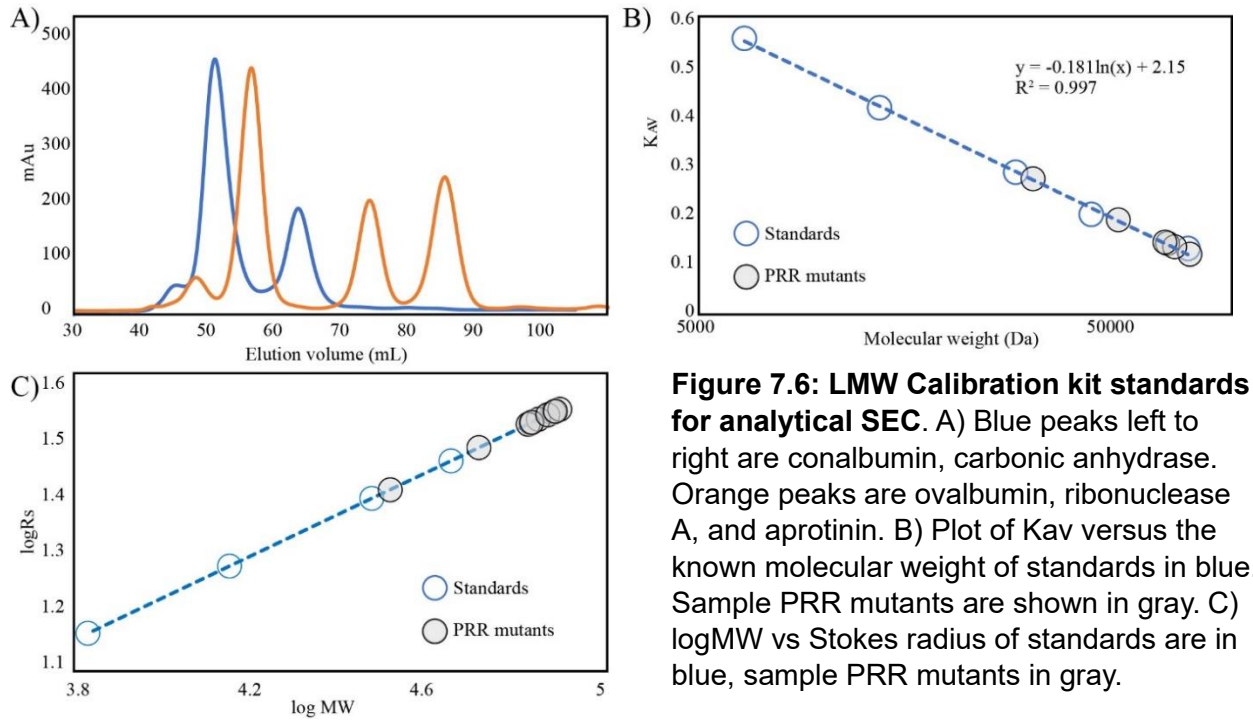


Figure 7.6: LMW Calibration kit standards for analytical SEC. A) Blue peaks left to right are conalbumin, carbonic anhydrase. Orange peaks are ovalbumin, ribonuclease A, and aprotinin. B) Plot of Kav versus the known molecular weight of standards in blue. Sample PRR mutants are shown in gray. C) logMW vs Stokes radius of standards are in blue, sample PRR mutants in gray.

in high and low ionic strength buffers is only 0.17 mL. This suggests to us that the average radius is not greatly changed by ionic strength in this environment.

Table 7.4: Calibration kit standards for analytical SEC

	Elution peak (mL)	K_{av}	Known molecular weight (kDa)
Conalbumin	52.117 ± 0.029	0.127	75.000
Ovalbumin	57.950 ± 0.090	0.198	44.000
Carbonic anhydrase	64.967 ± 0.026	0.285	29.000
Ribonuclease	76.073 ± 0.031	0.419	13.700
Aprotinin	88.020 ± 0.090	0.563	6.500
Dextran Blue	41.750 ± 0.210	-	-

End-to-end distance and Effective concentration calculations

The distribution of end-to-end distances, pl, were determined as previously described (198, 273) and by the following equation:

$$p(r) = 4\pi r^2 (\frac{3}{4\pi L_p L_c})^{\frac{3}{2}} \exp\,(\frac{-3r^2}{4\pi L_p L_c}) \zeta(r, L_p, L_c) \tag{10}$$

Where r is the distance in angstroms separating N- and C-termini, L_p is the persistence length that is either measured experimentally or taken as a standard value from literature as 3 – 4 Å (204), and L_c is the contour length, here using a value of 3.8 Å per residue. The rightmost phrase expands to:

$$\zeta(r, L_p, L_c) = 1 - \left\{\frac{5L_p}{4L_c} - \frac{2r^2}{L_c} + \frac{33r^4}{80L_p{L_c}^3} + \frac{79{L_p}^2}{160{L_c}^2} + \frac{329L_p r^2}{120{L_c}^3} - \frac{6799r^4}{1600{L_c}^4} + \frac{3441r^6}{2800L_p{L_c}^5} - \frac{1089r^8}{28000{L_p}^2{L_c}^6}\right\} \quad (11)$$

End-to-end distance graphs are generated in Excel using Equation 8 with the appropriate *Lc* and *Lp* values as described in Chapter 4 and by using a range of radius values beginning with 1 Å. The maximum value of the distribution – the radius at which the end-to-end distribution is greatest – is found using the Solver function in Excel where the end-to-end distance distribution, pI, is set to a maximum and the radius, r, is solved for. Alternatively, Kjaergaard et al. have developed an online tool to calculate end-to-end distance and effective concentration with variable parameters (82), see https://cloud.chemeslab.org:3939/CeffApp/.

Effective concentration was determined using the following equation:

$$C_{eff} = \frac{p(r_0)}{4\pi r^2} * \frac{10^{27}\text{Å}^3\text{l}^{-1}}{L_0} \quad (12)$$

Where C_{eff} is the effective concentration, L_0 is Avogadro's number, $p(r_o)$ is the pI value at a given radius, and r is the actual distance separating the two interaction sites. The actual distance separating interaction sites here is the distance from the N-terminal-most backbone atom of residue S94 in PDB **4HJE** and the nearest DNA-binding residue of DBD, R248: 18.05 Å. The presence of TAD causes a chemical shift in DBD above the resolution threshold for residue R248, suggesting this residue interacts with TAD (74). Similarly, the distance separating interdimer interaction is traced from residue S94 to S241 of the flanking DBD unit, a distance of 30.16 Å, and the intertetramer distance is traced from S94 to R248, a distance of 43.66 Å. All distances were measured in VMD

(274) using PDB **4HJE**. C_{eff} was calculated in Excel using the parameters described above and confirmed using the calculator from Kjaergaard et a.l (82).

The fraction of bound, *fx Bound*, protein is described by the following:

$$fx\ Bound = \frac{K_{A'}}{1 + K_{A'}} \tag{13}$$

Where $K_{A'}$ is the defined by the following:

$$K_{A'} = K_{A1} * C_{eff} \tag{14}$$

And where K_{A1} is the binding affinity of the untethered TAD to DBD. We estimate this to be 1 mM, and C_{eff} is based on an intramolecular interaction model.

Sequence analysis

Sequences of p53 were obtained from Uniprot (275) and the National Center for Biotechnology Information protein database (276). Sequences were aligned using Geneious Prime 2022.2 (277) using Consensus Alignment and the Blosum62 cost matrix with a gap open penalty of 12, gap extension penalty of 3, and with 2 refinement iterations. Additional multiple sequence alignments were generated using Clustal Omega from EMBL (278).

Disorder tendency was determined using IUPRED (21, 22). Net charge per residue (NCPR) was determined using the CIDER tool (215). PPII helical propensity of PRR mutants was assessed using PPIIPred (182). Helical propensity was estimated using Agadir (279).

NMR

Experiments were performed on ^{15}N-labeled KIX apo or bound to unlabeled synthetic cMyb, MLL, or CREB constructs in a standard NMR buffer composed of 50 mM NaH_2PO_4, 50 mM NaCl, 1 mM EDTA, 5% DMSO, and 0.02% NaN_3 at pH 6.8 or in PBS composed of 10 mM Na_2HPO_4, 1.8 mM KH_2PO4, 2.7 mM KCl, 137 mM NaCl, and 0.02% NaN_3 at pH 7.4 to improve solubility of compounds LC-A-122-2 and 3_30b.

cMyb WT was used at a concentration of 160 μM and cMyb mimics 104-2 and 109-1 were titrated into KIX for final concentrations ranging from 133-800 μM, all against 200 μM samples of ^{15}N-labeled KIX in standard NMR buffer.

For experiments using cMyb LC-A-122-2 and LC-A-122-3, ^{15}N-labeled KIX was used at 260 μM and mimics were used at concentrations of 1040 μM, with titrations at final concentrations ranging from 173 μM to 867 μM being used to confidently establish chemical shift changes for cMyb LC-A-122-3. Experiments were conducted in standard NMR buffer or PBS with plurionic F68 (Thermo Fisher, Waltham, MA) added for a final concentration of 0.05% to improve solubility. For experiments using MLL mimics 6 and 8, ^{15}N-labeled KIX was used at 200 μM and mimics were used at concentrations of 800 μM in standard NMR buffer. For experiments using CREB mimics, ^{15}N-labeled KIX was used at 450 μM and mimics were used at concentrations of 1700 μM in NMR buffer except for mimic 3_30b which used PBS buffer to improve solubility.

All cMyb, MLL, and CREB mimics were resuspended from lyophilized form in 100% DMSO and then diluted in NMR buffer or PBS buffer with KIX for a final concentration of 5% DMSO. Experiments were performed at 25°C on a Varian VNMRS 800-MHz spectrometry with a triple resonance pulse field Z-axis gradient cold probe at

30°C. HSQC used 256 and 1024 points for t1 and t2 dimensions, respectively. Data was processed using NMRFx and analyzed using NMRView (280, 281) with assignments made using previously published data (228). Chemical shifts were calculated as $[((\Delta^{1}H)^{2}+(\Delta^{15}N/5)^{2})/2)]^{1/2}$. Backbone assignments of apo ^{13}C-^{15}N labeled KIX 425 μM were found using 3D HNCO, HNCA, and HNCACB data (282, 283). HNCO experiments used sweep widths of 9689.22 Hz (^{1}H), 2200 Hz (^{15}N), and 2412.096 Hz (^{13}CO) and complex points of 2048 (^{1}H), 60 (^{15}N), and 112 (^{13}CO). The reference for the carbon dimension was 174.095 ppm. HNCA experiments used sweep widths of 9689.22 Hz (^{1}H), 2200 Hz (^{15}N), and 6031.345 Hz ($^{13}C_{\alpha}$) and complex points of 2048 (^{1}H), 80 (^{15}N), and 256 ($^{13}C_{\alpha}$). The reference for the carbon dimension was 565.133 ppm. HNCACB experiments used sweep widths of 9689.22 Hz (^{1}H), 2200 Hz (^{15}N), and 14073.137 Hz ($^{13}C_{\alpha}$) and complex points of 2048 (^{1}H), 80 (^{15}N), and 220 ($^{13}C_{\alpha}$). The reference for the carbon dimension was 46.136 ppm. 3D NMR spectra were processed and analyzed using POKY (284).

Surface models that show chemical shifts >0.02 ppm of KIX residues in response to the addition of compounds were made using VMD 1.9.1 (274, 285) overlaid on a complex of KIX (mouse residues 586-666) with cMyb (mouse residues 291 – 315) and MLL (mouse residues 839 – 869) (PDB **2AGH**) (222) or the KID domain of CREB (rat residues 119-146) (PDB **1KDX)** (250).

LITERATURE CITED

1. Babu MM. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. Biochem Soc Trans. 2016;44(5):1185-200.
2. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol. 2004;337(3):635-45.
3. Guo X, Bulyk ML, Hartemink AJ. Intrinsic disorder within and flanking the DNA-binding domains of human transcription factors. Pac Symp Biocomput. 2012:104-15.
4. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of intrinsically disordered regions and proteins. Chem Rev. 2014;114(13):6589-631.
5. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins. 2001;42(1):38-48.
6. Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and disorder. Curr Opin Struct Biol. 2011;21(3):441-6.
7. Lieutaud P, Ferron F, Uversky AV, Kurgan L, Uversky VN, Longhi S. How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. Intrinsically Disord Proteins. 2016;4(1):e1259708.
8. Wright PE, Dyson HJ. Linking folding and binding. Curr Opin Struct Biol. 2009;19(1):31-8.
9. Chakrabarti P, Chakravarty D. Intrinsically disordered proteins/regions and insight into their biomolecular interactions. Biophys Chem. 2022;283:106769.
10. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. 2015;16(1):18-29.
11. Borgia A, Borgia MB, Bugge K, Kissling VM, Heidarsson PO, Fernandes CB, et al. Extreme disorder in an ultrahigh-affinity protein complex. Nature. 2018;555(7694):61-6.
12. Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem Sci. 2008;33(1):2-8.
13. Privalov PL, Dragan AI, Crane-Robinson C. Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components. Nucleic Acids Research. 2011;39(7):2483-91.
14. Liu Y, Matthews KS, Bondos SE. Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the Drosophila hox protein ultrabithorax. J Biol Chem. 2008;283(30):20874-87.

15. Brodsky S, Jana T, Mittelman K, Chapal M, Kumar DK, Carmi M, et al. Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. Mol Cell. 2020;79(3):459-71 e4.
16. Hainaut P, Hollstein M. p53 and human cancer: The first ten thousand mutations. Adv Cancer Res. 2000;77:81-137.
17. Hupp TR, Meek DW, Midgley CA, Lane DP. Regulation of the Specific DNA-Binding Function of P53. Cell. 1992;71(5):875-86.
18. Kruse JP, Gu W. Modes of p53 Regulation. Cell. 2009;137(4):609-22.
19. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47(D1):D941-D7.
20. Joerger AC, Fersht AR. Structural biology of the tumor suppressor p53. Annu Rev Biochem. 2008;77:557-82.
21. Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics. 2005;21(16):3433-4.
22. Erdos G, Pajkos M, Dosztanyi Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. Nucleic Acids Res. 2021;49(W1):W297-W303.
23. Xue B, Brown CJ, Dunker AK, Uversky VN. Intrinsically disordered regions of p53 family are highly diversified in evolution. Biochim Biophys Acta. 2013;1834(4):725-38.
24. Ayed A, Mulder FA, Yi GS, Lu Y, Kay LE, Arrowsmith CH. Latent and active p53 are identical in conformation. Nat Struct Biol. 2001;8(9):756-60.
25. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9.
26. Fenton M, Gregory, E., Daughdrill, G. Protein disorder and autoinhibition: The role of multivalency and effective concentration. Curr Opin Struct Biol. 2023;83.
27. Natan E, Baloglu C, Pagel K, Freund SM, Morgner N, Robinson CV, et al. Interaction of the p53 DNA-binding domain with its n-terminal extension modulates the stability of the p53 tetramer. J Mol Biol. 2011;409(3):358-68.
28. Solares MJ, Jonaid GM, Luqiu WY, Berry S, Khadela J, Liang Y, et al. High-Resolution Imaging of Human Cancer Proteins Using Microprocessor Materials. Chembiochem. 2022;23(17):e202200310.
29. Nedelcu AM, Tan C. Early diversification and complex evolutionary history of the p53 tumor suppressor gene family. Dev Genes Evol. 2007;217(11-12):801-6.
30. Aberg E, Saccoccia F, Grabherr M, Ore WYJ, Jemth P, Hultqvist G. Evolution of the p53-MDM2 pathway. BMC Evol Biol. 2017;17(1):177.
31. Levine AJ, Tomasini R, McKeon FD, Mak TW, Melino G. The p53 family: guardians of maternal reproduction. Nat Rev Mol Cell Biol. 2011;12(4):259-65.
32. Dotsch V, Bernassola F, Coutandin D, Candi E, Melino G. p63 and p73, the ancestors of p53. Cold Spring Harb Perspect Biol. 2010;2(9):a004887.

33. Nitta KR, Jolma A, Yin Y, Morgunova E, Kivioja T, Akhtar J, et al. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. Elife. 2015;4.
34. Jin S, Martinek S, Joo WS, Wortman JR, Mirkovic N, Sali A, et al. Identification and characterization of a p53 homologue in Drosophila melanogaster. Proc Natl Acad Sci U S A. 2000;97(13):7301-6.
35. Huyen Y, Jeffrey PD, Derry WB, Rothman JH, Pavletich NP, Stavridi ES, et al. Structural differences in the DNA binding domains of human p53 and its C. elegans ortholog Cep-1. Structure. 2004;12(7):1237-43.
36. Spreitzer E, Alderson TR, Bourgeois B, Eggenreich L, Habacher H, Brahmersdorfer G, et al. FOXO transcription factors differ in their dynamics and intra/intermolecular interactions. Curr Res Struct Biol. 2022;4:118-33.
37. Muller PA, Vousden KH. p53 mutations in cancer. Nat Cell Biol. 2013;15(1):2-8.
38. Kitayner M, Rozenberg H, Kessler N, Rabinovich D, Shaulov L, Haran TE, et al. Structural basis of DNA recognition by p53 tetramers. Mol Cell. 2006;22(6):741-53.
39. Hoyos D, Greenbaum B, Levine AJ. The genotypes and phenotypes of missense mutations in the proline domain of the p53 protein. Cell Death Differ. 2022;29(5):938-45.
40. Kubbutat MH, Jones SN, Vousden KH. Regulation of p53 stability by Mdm2. Nature. 1997;387(6630):299-303.
41. Shvarts A, Steegenga WT, Riteco N, van Laar T, Dekker P, Bazuine M, et al. MDMX: a novel p53-binding protein with some functional properties of MDM2. EMBO J. 1996;15(19):5349-57.
42. Collavin L, Lunardi A, Del Sal G. p53-family proteins and their regulators: hubs and spokes in tumor suppression. Cell Death Differ. 2010;17(6):901-11.
43. Vousden KH, Prives C. Blinded by the Light: The Growing Complexity of p53. Cell. 2009;137(3):413-31.
44. Gaglia G, Guan Y, Shah JV, Lahav G. Activation and control of p53 tetramerization in individual living cells. Proc Natl Acad Sci U S A. 2013;110(38):15497-501.
45. Weinberg RL, Veprintsev DB, Fersht AR. Cooperative binding of tetrameric p53 to DNA. J Mol Biol. 2004;341(5):1145-59.
46. McLure KG, Lee PWK. How p53 binds DNA as a tetramer. Embo Journal. 1998;17(12):3342-50.
47. el-Deiry WS, Kern SE, Pietenpol JA, Kinzler KW, Vogelstein B. Definition of a consensus binding site for p53. Nat Genet. 1992;1(1):45-9.
48. Jordan JJ, Menendez D, Inga A, Noureddine M, Bell DA, Resnick MA. Noncanonical DNA motifs as transactivation targets by wild type and mutant p53. PLoS Genet. 2008;4(6):e1000104.
49. Veprintsev DB, Fersht AR. Algorithm for prediction of tumour suppressor p53 affinity for binding sites in DNA. Nucleic Acids Res. 2008;36(5):1589-98.
50. Weinberg RL, Veprintsev DB, Bycroft M, Fersht AR. Comparative binding of p53 to its promoter and DNA recognition elements. J Mol Biol. 2005;348(3):589-96.

51. Chen X, Ko LJ, Jayaraman L, Prives C. p53 levels, functional domains, and DNA damage determine the extent of the apoptotic response of tumor cells. Genes Dev. 1996;10(19):2438-51.
52. Chen Y, Zhang X, Dantas Machado AC, Ding Y, Chen Z, Qin PZ, et al. Structure of p53 binding to the BAX response element reveals DNA unwinding and compression to accommodate base-pair insertion. Nucleic Acids Res. 2013;41(17):8368-76.
53. Sullivan KD, Galbraith MD, Andrysik Z, Espinosa JM. Mechanisms of transcriptional regulation by p53. Cell Death Differ. 2018;25(1):133-43.
54. Arbely E, Natan E, Brandt T, Allen MD, Veprintsev DB, Robinson CV, et al. Acetylation of lysine 120 of p53 endows DNA-binding specificity at effective physiological salt concentration. Proc Natl Acad Sci U S A. 2011;108(20):8251-6.
55. Brazda V, Coufal J. Recognition of Local DNA Structures by p53 Protein. Int J Mol Sci. 2017;18(2).
56. Chene P. The role of tetramerization in p53 function. Oncogene. 2001;20(21):2611-7.
57. Horvath MM, Wang X, Resnick MA, Bell DA. Divergent evolution of human p53 binding sites: cell cycle versus apoptosis. PLoS Genet. 2007;3(7):e127.
58. Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG. Choose your partners: dimerization in eukaryotic transcription factors. Trends Biochem Sci. 2008;33(5):220-9.
59. Funnell AP, Crossley M. Homo- and heterodimerization in transcriptional regulation. Adv Exp Med Biol. 2012;747:105-21.
60. Gregory E, Daughdrill GW. Sequence Properties of An Intramolecular Interaction That Inhibits p53 DNA Binding. Biomolecules. 2022;12(11).
61. Lubin DJ, Butler JS, Loh SN. Folding of tetrameric p53: oligomerization and tumorigenic mutations induce misfolding and loss of function. J Mol Biol. 2010;395(4):705-16.
62. Fischer M. Census and evaluation of p53 target genes. Oncogene. 2017;36(28):3943-56.
63. Hafner A, Bulyk ML, Jambhekar A, Lahav G. The multiple mechanisms that regulate p53 activity and cell fate. Nat Rev Mol Cell Biol. 2019;20(4):199-210.
64. Tsai YY, Su CH, Tarn WY. p53 Activation in Genetic Disorders: Different Routes to the Same Destination. Int J Mol Sci. 2021;22(17).
65. Speidel D. Transcription-independent p53 apoptosis: an alternative route to death. Trends Cell Biol. 2010;20(1):14-24.
66. Fernandez-Fernandez MR, Sot B. The relevance of protein-protein interactions for p53 function: the CPE contribution. Protein Eng Des Sel. 2011;24(1-2):41-51.
67. Beckerman R, Prives C. Transcriptional regulation by p53. Cold Spring Harb Perspect Biol. 2010;2(8):a000935.
68. Kaiser AM, Attardi LD. Deconstructing networks of p53-mediated tumor suppression in vivo. Cell Death Differ. 2018;25(1):93-103.
69. Harms KL, Chen X. The functional domains in p53 family proteins exhibit both common and distinct properties. Cell Death Differ. 2006;13(6):890-7.

70. Cho Y, Gorina S, Jeffrey PD, Pavletich NP. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science. 1994;265(5170):346-55.
71. Li J, Guo M, Chen L, Chen Z, Fu Y, Chen Y. p53 amyloid aggregation in cancer: function, mechanism, and therapy. Exp Hematol Oncol. 2022;11(1):66.
72. Xu J, Reumers J, Couceiro JR, De Smet F, Gallardo R, Rudyak S, et al. Gain of function of mutant p53 by coaggregation with multiple tumor suppressors. Nat Chem Biol. 2011;7(5):285-95.
73. Ko LJ, Prives C. p53: puzzle and paradigm. Genes Dev. 1996;10(9):1054-72.
74. He F, Borcherds W, Song T, Wei X, Das M, Chen L, et al. Interaction between p53 N terminus and core domain regulates specific and nonspecific DNA binding. Proc Natl Acad Sci U S A. 2019;116(18):8859-68.
75. Sun X, Dyson HJ, Wright PE. A phosphorylation-dependent switch in the disordered p53 transactivation domain regulates DNA binding. Proc Natl Acad Sci U S A. 2021;118(1).
76. Krois AS, Dyson HJ, Wright PE. Long-range regulation of p53 DNA binding by its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci U S A. 2018;115(48):E11302-E10.
77. Pufall MA, Graves BJ. Autoinhibitory domains: modular effectors of cellular regulation. Annu Rev Cell Dev Biol. 2002;18:421-62.
78. Trudeau T, Nassar R, Cumberworth A, Wong ET, Woollard G, Gsponer J. Structure and intrinsic disorder in protein autoinhibition. Structure. 2013;21(3):332-41.
79. Gogl G, Kornev AP, Remenyi A, Taylor SS. Disordered Protein Kinase Regions in Regulation of Kinase Domain Cores. Trends Biochem Sci. 2019;44(4):300-11.
80. Ulamec SM, Brockwell DJ, Radford SE. Looking Beyond the Core: The Role of Flanking Regions in the Aggregation of Amyloidogenic Peptides and Proteins. Front Neurosci. 2020;14:611285.
81. Arbesu M, Iruela G, Fuentes H, Teixeira JMC, Pons M. Intramolecular Fuzzy Interactions Involving Intrinsically Disordered Domains. Front Mol Biosci. 2018;5:39.
82. Kjaergaard M, Glavina J, Chemes LB. Predicting the effect of disordered linkers on effective concentrations and avidity with the "Ceff calculator" app. Methods Enzymol. 2021;647:145-71.
83. Wijeratne TU, Guiley KZ, Lee HW, Muller GA, Rubin SM. Cyclin-dependent kinase-mediated phosphorylation and the negative regulatory domain of transcription factor B-Myb modulate its DNA binding. J Biol Chem. 2022:102319.
84. Bourgeois B, Gui T, Hoogeboom D, Hocking HG, Richter G, Spreitzer E, et al. Multiple regulatory intrinsically disordered motifs control FOXO4 transcription factor binding and function. Cell Rep. 2021;36(4):109446.
85. Schutz S, Bergsdorf C, Goretzki B, Lingel A, Renatus M, Gossert AD, et al. The Disordered MAX N-terminus Modulates DNA Binding of the Transcription Factor MYC:MAX. J Mol Biol. 2022;434(22):167833.

86. Fang J, Cheng J, Wang J, Zhang Q, Liu M, Gong R, et al. Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition. Nat Commun. 2016;7:11197.
87. Gong WB, Liang QH, Tong YF, Perrett S, Feng YG. Structural Insight into Chromatin Recognition by Multiple Domains of the Tumor Suppressor RBBP1. Journal of Molecular Biology. 2021;433(21).
88. Desjardins G, Meeker CA, Bhachech N, Currie SL, Okon M, Graves BJ, et al. Synergy of aromatic residues and phosphoserines within the intrinsically disordered DNA-binding inhibitory elements of the Ets-1 transcription factor. P Natl Acad Sci USA. 2014;111(30):11019-24.
89. Kang HS, Sanchez-Rico C, Ebersberger S, Sutandy FXR, Busch A, Welte T, et al. An autoinhibitory intramolecular interaction proof-reads RNA recognition by the essential splicing factor U2AF2. Proc Natl Acad Sci U S A. 2020;117(13):7140-9.
90. Wang X, Greenblatt HM, Bigman LS, Yu B, Pletka CC, Levy Y, et al. Dynamic Autoinhibition of the HMGB1 Protein via Electrostatic Fuzzy Interactions of Intrinsically Disordered Regions. J Mol Biol. 2021;433(18):167122.
91. Fenton M, Borcherds W, Chen L, Anbanandam A, Levy R, Chen J, et al. The MDMX Acidic Domain Uses Allovalency to Bind Both p53 and MDMX. J Mol Biol. 2022;434(22):167844.
92. Williamson MP. The structure and function of proline-rich regions in proteins. Biochem J. 1994;297 (Pt 2):249-60.
93. Zhang L, Kang H, Perez-Aguilar JM, Zhou R. Possible Co-Evolution of Polyglutamine and Polyproline in Huntingtin Protein: Proline-Rich Domain as Transient Folding Chaperone. J Phys Chem Lett. 2022;13(27):6331-41.
94. Lee H, Mok KH, Muhandiram R, Park KH, Suk JE, Kim DH, et al. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. J Biol Chem. 2000;275(38):29426-32.
95. Miller Jenkins LM, Feng H, Durell SR, Tagad HD, Mazur SJ, Tropea JE, et al. Characterization of the p300 Taz2-p53 TAD2 complex and comparison with the p300 Taz2-p53 TAD1 complex. Biochemistry. 2015;54(11):2001-10.
96. Bochkareva E, Kaustov L, Ayed A, Yi GS, Lu Y, Pineda-Lucena A, et al. Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. Proc Natl Acad Sci U S A. 2005;102(43):15412-7.
97. Rowell JP, Simpson KL, Stott K, Watson M, Thomas JO. HMGB1-facilitated p53 DNA binding occurs via HMG-Box/p53 transactivation domain interaction, regulated by the acidic tail. Structure. 2012;20(12):2014-24.
98. Okuda M, Nishimura Y. Extended string binding mode of the phosphorylated transactivation domain of tumor suppressor p53. J Am Chem Soc. 2014;136(40):14143-52.
99. Posttranslational Modification and Protein Disorder Regulate Protein-Protein Interactions and DNA Binding Specificity of p53 [Internet]. 2020. Available from: https://digitalcommons.usf.edu/etd/9545.

100. Brady CA, Jiang D, Mello SS, Johnson TM, Jarvis LA, Kozak MM, et al. Distinct p53 transcriptional programs dictate acute DNA-damage responses and tumor suppression. Cell. 2011;145(4):571-83.
101. Suzuki S, Tsutsumi S, Chen Y, Ozeki C, Okabe A, Kawase T, et al. Identification and characterization of the binding sequences and target genes of p53 lacking the 1st transactivation domain. Cancer Sci. 2020;111(2):451-66.
102. Venot C, Maratrat M, Sierra V, Conseiller E, Debussche L. Definition of a p53 transactivation function-deficient mutant and characterization of two independent p53 transactivation subdomains. Oncogene. 1999;18(14):2405-10.
103. Soussi T, Caron de Fromentel C, May P. Structural aspects of the p53 protein in relation to gene evolution. Oncogene. 1990;5(7):945-52.
104. Straub WE, Weber TA, Schafer B, Candi E, Durst F, Ou HD, et al. The C-terminus of p63 contains multiple regulatory elements with different functions. Cell Death Dis. 2010;1(1):e5.
105. Subekti DRG, Murata A, Itoh Y, Fukuchi S, Takahashi H, Kanbayashi S, et al. The Disordered Linker in p53 Participates in Nonspecific Binding to and One-Dimensional Sliding along DNA Revealed by Single-Molecule Fluorescence Measurements. Biochemistry. 2017;56(32):4134-44.
106. Adzhubei AA, Sternberg MJ. Left-handed polyproline II helices commonly occur in globular proteins. J Mol Biol. 1993;229(2):472-93.
107. Huang F, Rajagopalan S, Settanni G, Marsh RJ, Armoogum DA, Nicolaou N, et al. Multiple conformations of full-length p53 detected with single-molecule fluorescence resonance energy transfer. Proc Natl Acad Sci U S A. 2009;106(49):20758-63.
108. Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, et al. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci U S A. 2008;105(15):5762-7.
109. Perez RB, Tischer A, Auton M, Whitten ST. Alanine and proline content modulate global sensitivity to discrete perturbations in disordered proteins. Proteins. 2014;82(12):3373-84.
110. Bartas M, Brazda V, Volna A, Cerven J, Pecinka P, Zawacka-Pankau JE. The Changes in the p53 Protein across the Animal Kingdom Point to Its Involvement in Longevity. Int J Mol Sci. 2021;22(16).
111. Dornan D, Shimizu H, Burch L, Smith AJ, Hupp TR. The proline repeat domain of p53 binds directly to the transcriptional coactivator p300 and allosterically controls DNA-dependent acetylation of p53. Mol Cell Biol. 2003;23(23):8846-61.
112. Berger M, Vogt Sionov R, Levine AJ, Haupt Y. A role for the polyproline domain of p53 in its regulation by Mdm2. J Biol Chem. 2001;276(6):3785-90.
113. Panni S, Salvioli S, Santonico E, Langone F, Storino F, Altilia S, et al. The adapter protein CD2AP binds to p53 protein in the cytoplasm and can discriminate its polymorphic variants P72R. J Biochem. 2015;157(2):101-11.

114. Bergamaschi D, Samuels Y, Sullivan A, Zvelebil M, Breyssens H, Bisso A, et al. iASPP preferentially binds p53 proline-rich region and modulates apoptotic function of codon 72-polymorphic p53. Nat Genet. 2006;38(10):1133-41.
115. Samuels-Lev Y, O'Connor DJ, Bergamaschi D, Trigiante G, Hsieh JK, Zhong S, et al. ASPP proteins specifically stimulate the apoptotic function of p53. Mol Cell. 2001;8(4):781-94.
116. Tommiska J, Eerola H, Heinonen M, Salonen L, Kaare M, Tallila J, et al. Breast cancer patients with p53 Pro72 homozygous genotype have a poorer survival. Clin Cancer Res. 2005;11(14):5098-103.
117. Fan R, Wu MT, Miller D, Wain JC, Kelsey KT, Wiencke JK, et al. The p53 codon 72 polymorphism and lung cancer risk. Cancer Epidemiol Biomarkers Prev. 2000;9(10):1037-42.
118. Berger M, Stahl N, Del Sal G, Haupt Y. Mutations in proline 82 of p53 impair its activation by Pin1 and Chk2 in response to DNA damage. Molecular and Cellular Biology. 2005;25(13):5380-8.
119. Sakamuro D, Sabbatini P, White E, Prendergast GC. The polyproline region of p53 is required to activate apoptosis but not growth arrest. Oncogene. 1997;15(8):887-98.
120. Ruaro EM, Collavin L, Del Sal G, Haffner R, Oren M, Levine AJ, et al. A proline-rich motif in p53 is required for transactivation-independent growth arrest as induced by Gas1. Proc Natl Acad Sci U S A. 1997;94(9):4675-80.
121. Toledo F, Krummel KA, Lee CJ, Liu CW, Rodewald LW, Tang M, et al. A mouse p53 mutant lacking the proline-rich domain rescues Mdm4 deficiency and provides insight into the Mdm2-Mdm4-p53 regulatory network. Cancer Cell. 2006;9(4):273-85.
122. Zhu J, Jiang J, Zhou W, Zhu K, Chen X. Differential regulation of cellular target genes by p53 devoid of the PXXP motifs with impaired apoptotic activity. Oncogene. 1999;18(12):2149-55.
123. Venot C, Maratrat M, Dureuil C, Conseiller E, Bracco L, Debussche L. The requirement for the p53 proline-rich functional domain for mediation of apoptosis is correlated with specific PIG3 gene transactivation and with transcriptional repression. EMBO J. 1998;17(16):4668-79.
124. Roth J, Koch P, Contente A, Dobbelstein M. Tumor-derived mutations within the DNA-binding domain of p53 that phenotypically resemble the deletion of the proline-rich domain. Oncogene. 2000;19(14):1834-42.
125. Dumont P, Leu JI, Della Pietra AC, 3rd, George DL, Murphy M. The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. Nat Genet. 2003;33(3):357-65.
126. Wolf ER, McAtarsney CP, Bredhold KE, Kline AM, Mayo LD. Mutant and wild-type p53 form complexes with p73 upon phosphorylation by the kinase JNK. Sci Signal. 2018;11(524).

127. Baum N, Schiene-Fischer C, Frost M, Schumann M, Sabapathy K, Ohlenschlager O, et al. The prolyl cis/trans isomerase cyclophilin 18 interacts with the tumor suppressor p53 and modifies its functions in cell cycle regulation and apoptosis. Oncogene. 2009;28(44):3915-25.
128. Singh A, Singh N. Phase diagram of mechanically stretched DNA: The salt effect. Physica A. 2013;392(9):2052-9.
129. Record MT, Jr., Lohman ML, De Haseth P. Ion effects on ligand-nucleic acid interactions. J Mol Biol. 1976;107(2):145-58.
130. Yu B, Pettitt BM, Iwahara J. Dynamics of Ionic Interactions at Protein-Nucleic Acid Interfaces. Acc Chem Res. 2020;53(9):1802-10.
131. Hassin O, Oren M. Drugging p53 in cancer: one protein, many targets. Nat Rev Drug Discov. 2023;22(2):127-44.
132. Huang Y, Gao M, Yang F, Zhang L, Su Z. Deciphering the promiscuous interactions between intrinsically disordered transactivation domains and the KIX domain. Proteins. 2017;85(11):2088-95.
133. Manning GS. Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions .1. Colligative properties. J Chem Phys. 1969;51(3):922-34.
134. Record MT, Jr., Ha JH, Fisher MA. Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. Methods Enzymol. 1991;208:291-343.
135. Wang Y, Schwedes JF, Parks D, Mann K, Tegtmeyer P. Interaction of p53 with its consensus DNA-binding site. Mol Cell Biol. 1995;15(4):2157-65.
136. LiCata VJ, Wowor AJ. Applications of fluorescence anisotropy to the study of protein-DNA interactions. Methods Cell Biol. 2008;84:243-62.
137. Brandt T, Townsley FM, Teufel DP, Freund SM, Veprintsev DB. Molecular basis for modulation of the p53 target selectivity by KLF4. PLoS One. 2012;7(10):e48252.
138. Rossi AM, Taylor CW. Analysis of protein-ligand interactions by fluorescence polarization. Nat Protoc. 2011;6(3):365-87.
139. Povedailo VA, Lysenko IL, Tikhomirov SA, Yakovlev DL, Tsybulsky DA, Kruhlik AS, et al. Fluorescent Properties of Carboxyfluorescein Bifluorophores. J Fluoresc. 2020;30(3):629-35.
140. Heyduk T, Ma Y, Tang H, Ebright RH. Fluorescence anisotropy: rapid, quantitative assay for protein-DNA and protein-protein interaction. Methods Enzymol. 1996;274:492-503.
141. Weiss JN. The Hill equation revisited: uses and misuses. FASEB J. 1997;11(11):835-41.
142. Nygaard M, Kragelund BB, Papaleo E, Lindorff-Larsen K. An Efficient Method for Estimating the Hydrodynamic Radius of Disordered Protein Conformations. Biophys J. 2017;113(3):550-7.
143. Misra VK, Honig B. On the magnitude of the electrostatic contribution to ligand-DNA interactions. Proc Natl Acad Sci U S A. 1995;92(10):4691-5.

144. Beno I, Rosenthal K, Levitine M, Shaulov L, Haran TE. Sequence-dependent cooperative binding of p53 to DNA targets and its relationship to the structural properties of the DNA targets. Nucleic Acids Res. 2011;39(5):1919-32.
145. Zhu J, Zhou W, Jiang J, Chen X. Identification of a novel p53 functional domain that is necessary for mediating apoptosis. J Biol Chem. 1998;273(21):13030-6.
146. Teufel DP, Freund SM, Bycroft M, Fersht AR. Four domains of p300 each bind tightly to a sequence spanning both transactivation subdomains of p53. Proc Natl Acad Sci U S A. 2007;104(17):7009-14.
147. von Hippel PH, Berg OG. On the specificity of DNA-protein interactions. Proc Natl Acad Sci U S A. 1986;83(6):1608-12.
148. Eaton BE, Gold L, Zichi DA. Let's get specific: the relationship between specificity and affinity. Chem Biol. 1995;2(10):633-8.
149. Chen J, Sun Y. Modeling of the salt effects on hydrophobic adsorption equilibrium of protein. J Chromatogr A. 2003;992(1-2):29-40.
150. Zangi R, Hagen M, Berne BJ. Effect of ions on the hydrophobic interaction between two plates. J Am Chem Soc. 2007;129(15):4678-86.
151. Fried MG, Stickle DF. Ion-exchange reactions of proteins during DNA binding. Eur J Biochem. 1993;218(2):469-75.
152. Olmsted MC, Bond JP, Anderson CF, Record MT. Grand-Canonical Monte-Carlo Molecular and Thermodynamic Predictions of Ion Effects on Binding of an Oligocation (L(8+)) to the Center of DNA Oligomers. Biophys J. 1995;68(2):634-47.
153. Fenley MO, Russo C, Manning GS. Theoretical assessment of the oligolysine model for ionic interactions in protein-DNA complexes. J Phys Chem B. 2011;115(32):9864-72.
154. Liu Y, Sturtevant JM. Significant discrepancies between van't Hoff and calorimetric enthalpies. II. Protein Sci. 1995;4(12):2559-61.
155. Datta K, Wowor AJ, Richard AJ, LiCata VJ. Temperature dependence and thermodynamics of Klenow polymerase binding to primed-template DNA. Biophys J. 2006;90(5):1739-51.
156. Zou Y, Bassett H, Walker R, Bishop A, Amin S, Geacintov NE, et al. Hydrophobic forces dominate the thermodynamic characteristics of UvrA-DNA damage interactions. J Mol Biol. 1998;281(1):107-19.
157. Ma H, Zou Y. Thermodynamic characterization of the interaction of mutant UvrB protein with damaged DNA. Biochemistry. 2004;43(14):4206-11.
158. Poon GM, Gross P, Macgregor RB, Jr. The sequence-specific association of the ETS domain of murine PU.1 with DNA exhibits unusual energetics. Biochemistry. 2002;41(7):2361-71.
159. Zhukov A, Karlsson R. Statistical aspects of van't Hoff analysis: a simulation study. J Mol Recognit. 2007;20(5):379-85.
160. Demir O, Ieong PU, Amaro RE. Full-length p53 tetramer bound to DNA and its quaternary dynamics. Oncogene. 2017;36(10):1451-60.

161. Lambrughi M, De Gioia L, Gervasio FL, Lindorff-Larsen K, Nussinov R, Urani C, et al. DNA-binding protects p53 from interactions with cofactors involved in transcription-independent functions. Nucleic Acids Res. 2016;44(19):9096-109.
162. Melero R, Rajagopalan S, Lazaro M, Joerger AC, Brandt T, Veprintsev DB, et al. Electron microscopy studies on the quaternary structure of p53 reveal different binding modes for p53 tetramers in complex with DNA. P Natl Acad Sci USA. 2011;108(2):557-62.
163. Takeda Y, Ross PD, Mudd CP. Thermodynamics of Cro protein-DNA interactions. Proc Natl Acad Sci U S A. 1992;89(17):8180-4.
164. Moraitis MI, Xu H, Matthews KS. Ion concentration and temperature dependence of DNA binding: comparison of PurR and LacI repressor proteins. Biochemistry. 2001;40(27):8109-17.
165. Oda M, Furukawa K, Ogata K, Sarai A, Nakamura H. Thermodynamics of specific and non-specific DNA binding by the c-Myb DNA-binding domain. J Mol Biol. 1998;276(3):571-90.
166. Cravens SL, Hobson M, Stivers JT. Electrostatic properties of complexes along a DNA glycosylase damage search pathway. Biochemistry. 2014;53(48):7680-92.
167. Winter RB, von Hippel PH. Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The Escherichia coli repressor--operator interaction: equilibrium measurements. Biochemistry. 1981;20(24):6948-60.
168. Chakraborty M, Sengupta A, Bhattacharya D, Banerjee S, Chakrabarti A. DNA binding domain of RFX5: interactions with X-box DNA and RFXANK. Biochim Biophys Acta. 2010;1804(10):2016-24.
169. Koblan KS, Ackers GK. Site-specific enthalpic regulation of DNA transcription at bacteriophage lambda OR. Biochemistry. 1992;31(1):57-65.
170. Misra VK, Hecht JL, Sharp KA, Friedman RA, Honig B. Salt effects on protein-DNA interactions. The lambda cI repressor and EcoRI endonuclease. J Mol Biol. 1994;238(2):264-80.
171. Klein C, Georges G, Kunkele KP, Huber R, Engh RA, Hansen S. High thermostability and lack of cooperative DNA binding distinguish the p63 core domain from the homologous tumor suppressor p53. J Biol Chem. 2001;276(40):37390-401.
172. Langridge TD, Tarver MJ, Whitten ST. Temperature effects on the hydrodynamic radius of the intrinsically disordered N-terminal region of the p53 protein. Proteins. 2014;82(4):668-78.
173. Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry. 1999;38(50):16424-31.
174. Marsh JA, Forman-Kay JD. Sequence Determinants of Compaction in Intrinsically Disordered Proteins. Biophys J. 2010;98(10):2383-90.
175. Ortega-Alarcon D, Claveria-Gimeno R, Vega S, Jorge-Torres OC, Esteller M, Abian O, et al. Stabilization Effect of Intrinsically Disordered Regions on Multidomain Proteins: The Case of the Methyl-CpG Protein 2, MeCP2. Biomolecules. 2021;11(8).

176. Toledo F, Lee CJ, Krummel KA, Rodewald LW, Liu CW, Wahl GM. Mouse mutants reveal that putative protein interaction sites in the p53 proline-rich domain are dispensable for tumor suppression. Mol Cell Biol. 2007;27(4):1425-32.
177. Bista M, Freund SM, Fersht AR. Domain-domain interactions in full-length p53 and a specific DNA complex probed by methyl NMR spectroscopy. Proc Natl Acad Sci U S A. 2012;109(39):15752-6.
178. Brown AM, Zondlo NJ. A propensity scale for type II polyproline helices (PPII): aromatic amino acids in proline-rich sequences strongly disfavor PPII due to proline-aromatic interactions. Biochemistry. 2012;51(25):5041-51.
179. Rath A, Davidson AR, Deber CM. The structure of "unstructured" regions in peptides and proteins: role of the polyproline II helix in protein folding and recognition. Biopolymers. 2005;80(2-3):179-85.
180. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev. 2013;65(10):1357-69.
181. van Rosmalen M, Krom M, Merkx M. Tuning the Flexibility of Glycine-Serine Linkers To Allow Rational Design of Multidomain Proteins. Biochemistry. 2017;56(50):6565-74.
182. O'Brien KT, Mooney C, Lopez C, Pollastri G, Shields DC. Prediction of polyproline II secondary structure propensity in proteins. R Soc Open Sci. 2020;7(1):191239.
183. Alderson TR, Lee JH, Charlier C, Ying J, Bax A. Propensity for cis-Proline Formation in Unfolded Proteins. Chembiochem. 2018;19(1):37-42.
184. Kay BK, Williamson MP, Sudol M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB J. 2000;14(2):231-41.
185. Stapley BJ, Creamer TP. A survey of left-handed polyproline II helices. Protein Sci. 1999;8(3):587-95.
186. Bieker JJ, Southwood CM. The erythroid Kruppel-like factor transactivation domain is a critical component for cell-specific inducibility of a beta-globin promoter. Mol Cell Biol. 1995;15(2):852-60.
187. de Caestecker MP, Yahata T, Wang D, Parks WT, Huang S, Hill CS, et al. The Smad4 activation domain (SAD) is a proline-rich, p300-dependent transcriptional activation domain. J Biol Chem. 2000;275(3):2115-22.
188. Lin R, Mamane Y, Hiscott J. Structural and functional analysis of interferon regulatory factor 3: localization of the transactivation and autoinhibitory domains. Mol Cell Biol. 1999;19(4):2465-74.
189. Moradi M, Babin V, Roland C, Darden TA, Sagui C. Conformations and free energy landscapes of polyproline peptides. Proc Natl Acad Sci U S A. 2009;106(49):20746-51.
190. Ball LJ, Kuhne R, Schneider-Mergener J, Oschkinat H. Recognition of proline-rich motifs by protein-protein-interaction domains. Angew Chem Int Ed Engl. 2005;44(19):2852-69.

191. Kurochkina N, Guha U. SH3 domains: modules of protein-protein interactions. Biophys Rev. 2013;5(1):29-39.
192. Tanaka S, Scheraga HA. Theory of the cooperative transition between two ordered conformations of poly(L-proline). III. Molecular theory in the presence of solvent. Macromolecules. 1975;8(4):516-21.
193. Li J, Motlagh HN, Chakuroff C, Thompson EB, Hilser VJ. Thermodynamic dissection of the intrinsically disordered N-terminal domain of human glucocorticoid receptor. J Biol Chem. 2012;287(32):26777-87.
194. Chin AF, Toptygin D, Elam WA, Schrank TP, Hilser VJ. Phosphorylation Increases Persistence Length and End-to-End Distance of a Segment of Tau Protein. Biophys J. 2016;110(2):362-71.
195. Arora PS, Ansari AZ, Best TP, Ptashne M, Dervan PB. Design of artificial transcriptional activators with rigid poly-L-proline linkers. J Am Chem Soc. 2002;124(44):13067-71.
196. Liou SH, Singh SK, Singer RH, Coleman RA, Liu WL. Structure of the p53/RNA polymerase II assembly. Commun Biol. 2021;4(1):397.
197. George RA, Heringa J. An analysis of protein domain linkers: their classification and role in protein folding. Protein Eng. 2002;15(11):871-9.
198. Borcherds W, Becker A, Chen L, Chen J, Chemes LB, Daughdrill GW. Optimal Affinity Enhancement by a Conserved Flexible Linker Controls p53 Mimicry in MdmX. Biophys J. 2017;112(10):2038-42.
199. Gonzalez-Foutel NS, Glavina J, Borcherds WM, Safranchik M, Barrera-Vilarmau S, Sagar A, et al. Conformational buffering underlies functional selection in intrinsically disordered protein regions. Nat Struct Mol Biol. 2022;29(8):781-90.
200. Argos P. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. J Mol Biol. 1990;211(4):943-58.
201. Zhou HX. Loops in proteins can be modeled as worm-like chains. Journal of Physical Chemistry B. 2001;105(29):6763-6.
202. Anderson E, Cole JL. Domain stabilities in protein kinase R (PKR): Evidence for weak interdomain interactions. Biochemistry. 2008;47(17):4887-97.
203. Ainavarapu SR, Brujic J, Huang HH, Wiita AP, Lu H, Li L, et al. Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. Biophys J. 2007;92(1):225-33.
204. Zhou HX. Polymer models of protein stability, folding, and interactions. Biochemistry. 2004;43(8):2141-54.
205. Mickolajczyk KJ, Olinares PDB, Niu Y, Chen N, Warrington SE, Sasaki Y, et al. Long-range intramolecular allostery and regulation in the dynein-like AAA protein Mdn1. Proc Natl Acad Sci U S A. 2020;117(31):18459-69.
206. Cho J, Park J, Tae G, Jin MS, Kwon I. The Minimal Effect of Linker Length for Fatty Acid Conjugation to a Small Protein on the Serum Half-Life Extension. Biomedicines. 2020;8(5).

207. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA. Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. Proc Natl Acad Sci U S A. 2005;102(8):2754-9.
208. Brown CJ, Johnson AK, Daughdrill GW. Comparing Models of Evolution for Ordered and Disordered Proteins. Mol Biol Evol. 2010;27(3):609-21.
209. Glavina J, Palopoli N, Chemes LB. Evolution of SLiM-mediated hijack functions in intrinsically disordered viral proteins. Essays Biochem. 2022;66(7):945-58.
210. Sangster AG, Zarin T, Moses AM. Evolution of short linear motifs and disordered proteins Topic: yeast as model system to study evolution. Curr Opin Genet Dev. 2022;76:101964.
211. Basu U, Mishra N, Farooqui M, Shen J, Johnson LC, Patel SS. The C-terminal tails of the mitochondrial transcription factors Mtf1 and TFB2M are part of an autoinhibitory mechanism that regulates DNA binding. J Biol Chem. 2020;295(20):6823-30.
212. Schweibenz BD, Devarkar SC, Solotchi M, Craig C, Zheng J, Pascal BD, et al. The intrinsically disordered CARDs-Helicase linker in RIG-I is a molecular gate for RNA proofreading. EMBO J. 2022;41(10):e109782.
213. Adzhubei AA, Sternberg MJ. Conservation of polyproline II helices in homologous proteins: implications for structure prediction by model building. Protein Sci. 1994;3(12):2395-410.
214. Coucke A, Uguzzoni G, Oteri F, Cocco S, Monasson R, Weigt M. Direct coevolutionary couplings reflect biophysical residue interactions in proteins. J Chem Phys. 2016;145(17):174102.
215. Holehouse AS, Das RK, Ahad JN, Richardson MO, Pappu RV. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. Biophys J. 2017;112(1):16-21.
216. Vo N, Goodman RH. CREB-binding protein and p300 in transcriptional regulation. J Biol Chem. 2001;276(17):13505-8.
217. Thakur JK, Yadav A, Yadav G. Molecular recognition by the KIX domain and its role in gene regulation. Nucleic Acids Res. 2014;42(4):2112-25.
218. Ogryzko VV, Schiltz RL, Russanova V, Howard BH, Nakatani Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell. 1996;87(5):953-9.
219. Ramos YF, Hestand MS, Verlaan M, Krabbendam E, Ariyurek Y, van Galen M, et al. Genome-wide assessment of differential roles for p300 and CBP in transcription regulation. Nucleic Acids Res. 2010;38(16):5396-408.
220. Goodman RH, Smolik S. CBP/p300 in cell growth, transformation, and development. Genes Dev. 2000;14(13):1553-77.
221. Fauquier L, Azzag K, Parra MAM, Quillien A, Boulet M, Diouf S, et al. CBP and P300 regulate distinct gene networks required for human primary myoblast differentiation and muscle integrity. Sci Rep. 2018;8(1):12629.

222. De Guzman RN, Goto NK, Dyson HJ, Wright PE. Structural basis for cooperative transcription factor binding to the CBP coactivator. J Mol Biol. 2006;355(5):1005-13.
223. Cook PR, Polakowski N, Lemasson I. HTLV-1 HBZ protein deregulates interactions between cellular factors and the KIX domain of p300/CBP. J Mol Biol. 2011;409(3):384-98.
224. Lee CW, Arai M, Martinez-Yamout MA, Dyson HJ, Wright PE. Mapping the interactions of the p53 transactivation domain with the KIX domain of CBP. Biochemistry. 2009;48(10):2115-24.
225. Goto NK, Zor T, Martinez-Yamout M, Dyson HJ, Wright PE. Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. J Biol Chem. 2002;277(45):43168-74.
226. Poosapati A, Gregory E, Borcherds WM, Chemes LB, Daughdrill GW. Uncoupling the Folding and Binding of an Intrinsically Disordered Protein. J Mol Biol. 2018;430(16):2389-402.
227. Ernst P, Wang J, Huang M, Goodman RH, Korsmeyer SJ. MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. Mol Cell Biol. 2001;21(7):2249-58.
228. Poosapati A. Disorder Levels of c-Myb Transactivation Domain Regulate its Binding Affinity to the KIX Domain of CREB Binding Protein. USF Scholar Commons: University of South Florida; 2017.
229. Crabtree MD, Borcherds W, Poosapati A, Shammas SL, Daughdrill GW, Clarke J. Conserved Helix-Flanking Prolines Modulate Intrinsically Disordered Protein:Target Affinity by Altering the Lifetime of the Bound Complex. Biochemistry. 2017;56(18):2379-84.
230. Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature. 2007;447(7147):1021-5.
231. Law SM, Gagnon JK, Mapp AK, Brooks CL, 3rd. Prepaying the entropic cost for allosteric regulation in KIX. Proc Natl Acad Sci U S A. 2014;111(33):12067-72.
232. Suetaka S, Oka Y, Kunihara T, Hayashi Y, Arai M. Rational design of a helical peptide inhibitor targeting c-Myb-KIX interaction. Sci Rep. 2022;12(1):816.
233. Sato N, Suetaka S, Hayashi Y, Arai M. Rational peptide design for inhibition of the KIX-MLL interaction. Sci Rep. 2023;13(1):6330.
234. Sapio L, Salzillo A, Ragone A, Illiano M, Spina A, Naviglio S. Targeting CREB in Cancer Therapy: A Key Candidate or One of Many? An Update. Cancers (Basel). 2020;12(11).
235. Quintana AM, Liu F, O'Rourke JP, Ness SA. Identification and regulation of c-Myb target genes in MCF-7 cells. BMC Cancer. 2011;11:30.
236. Mucenski ML, McLain K, Kier AB, Swerdlow SH, Schreiner CM, Miller TA, et al. A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. Cell. 1991;65(4):677-89.

237. Okada M, Tada M, Kanda N, Masuda M, Mizoguchi H, Kazuma M, et al. c-myb gene analysis in T-cell malignancies with del(6q). Cancer Genet Cytogenet. 1990;48(2):229-36.
238. Torelli G, Venturelli D, Colo A, Zanni C, Selleri L, Moretti L, et al. Expression of c-myb protooncogene and other cell cycle-related genes in normal and neoplastic human colonic mucosa. Cancer Res. 1987;47(20):5266-9.
239. Thompson MA, Flegg R, Westin EH, Ramsay RG. Microsatellite deletions in the c-myb transcriptional attenuator region associated with over-expression in colon tumour cell lines. Oncogene. 1997;14(14):1715-23.
240. Lahortiga I, De Keersmaecker K, Van Vlierberghe P, Graux C, Cauwelier B, Lambert F, et al. Duplication of the MYB oncogene in T cell acute lymphoblastic leukemia. Nat Genet. 2007;39(5):593-5.
241. Arai M, Sugase K, Dyson HJ, Wright PE. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. Proc Natl Acad Sci U S A. 2015;112(31):9614-9.
242. Zor T, De Guzman RN, Dyson HJ, Wright PE. Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. J Mol Biol. 2004;337(3):521-34.
243. Yu BD, Hess JL, Horning SE, Brown GA, Korsmeyer SJ. Altered Hox expression and segmental identity in Mll-mutant mice. Nature. 1995;378(6556):505-8.
244. Yagi H, Deguchi K, Aono A, Tani Y, Kishimoto T, Komori T. Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice. Blood. 1998;92(1):108-17.
245. Muntean AG, Hess JL. The pathogenesis of mixed-lineage leukemia. Annu Rev Pathol. 2012;7:283-301.
246. Chrivia JC, Kwok RP, Lamb N, Hagiwara M, Montminy MR, Goodman RH. Phosphorylated CREB binds specifically to the nuclear protein CBP. Nature. 1993;365(6449):855-9.
247. Shaywitz AJ, Greenberg ME. CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. Annu Rev Biochem. 1999;68:821-61.
248. Shankar DB, Cheng JC, Kinjo K, Federman N, Moore TB, Gill A, et al. The role of CREB as a proto-oncogene in hematopoiesis and in acute myeloid leukemia. Cancer Cell. 2005;7(4):351-62.
249. Steven A, Seliger B. Control of CREB expression in tumors: from molecular mechanisms and signal transduction pathways to therapeutic target. Oncotarget. 2016;7(23):35454-65.
250. Radhakrishnan I, Perez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. Cell. 1997;91(6):741-52.
251. Dahal L, Kwan TOC, Shammas SL, Clarke J. pKID Binds to KIX via an Unstructured Transition State with Nonnative Interactions. Biophys J. 2017;113(12):2713-22.

252. Fogolari F, Elcock AH, Esposito G, Viglino P, Briggs JM, McCammon JA. Electrostatic effects in homeodomain-DNA interactions. J Mol Biol. 1997;267(2):368-81.
253. Sharp KA, Friedman RA, Misra V, Hecht J, Honig B. Salt effects on polyelectrolyte-ligand binding: comparison of Poisson-Boltzmann, and limiting law/counterion binding models. Biopolymers. 1995;36(2):245-62.
254. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, et al. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science. 1996;274(5289):948-53.
255. Ferreon JC, Lee CW, Arai M, Martinez-Yamout MA, Dyson HJ, Wright PE. Cooperative regulation of p53 by modulation of ternary complex formation with CBP/p300 and HDM2. Proc Natl Acad Sci U S A. 2009;106(16):6591-6.
256. Di Lello P, Jenkins LM, Jones TN, Nguyen BD, Hara T, Yamaguchi H, et al. Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53. Mol Cell. 2006;22(6):731-40.
257. Zhao L, Ouyang Y, Li Q, Zhang Z. Modulation of p53 N-terminal transactivation domain 2 conformation ensemble and kinetics by phosphorylation. J Biomol Struct Dyn. 2020;38(9):2613-23.
258. Mello SS, Valente LJ, Raj N, Seoane JA, Flowers BM, McClendon J, et al. A p53 Super-tumor Suppressor Reveals a Tumor Suppressive p53-Ptpn14-Yap Axis in Pancreatic Cancer. Cancer Cell. 2017;32(4):460-73 e6.
259. Joseph RE, Wales TE, Fulton DB, Engen JR, Andreotti AH. Achieving a Graded Immune Response: BTK Adopts a Range of Active/Inactive Conformations Dictated by Multiple Interdomain Contacts. Structure. 2017;25(10):1481-94 e4.
260. Wilcox KG, Kemerer GM, Morozova S. Ionic environment effects on collagen type II persistence length and assembly. J Chem Phys. 2023;158(4):044903.
261. Wilcox KG, Dingle ME, Saha A, Hore MJA, Morozova S. Persistence length of alpha-helical poly-L-lysine. Soft Matter. 2022;18(35):6550-60.
262. Vyas P, Beno I, Xi Z, Stein Y, Golovenko D, Kessler N, et al. Diverse p53/DNA binding modes expand the repertoire of p53 response elements. Proc Natl Acad Sci U S A. 2017;114(40):10624-9.
263. Malkin D, Li FP, Strong LC, Fraumeni JF, Jr., Nelson CE, Kim DH, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science. 1990;250(4985):1233-8.
264. Ho WC, Fitzgerald MX, Marmorstein R. Structure of the p53 core domain dimer bound to DNA. J Biol Chem. 2006;281(29):20494-502.
265. Wei H, Qu L, Dai S, Li Y, Wang H, Feng Y, et al. Structural insight into the molecular mechanism of p53-mediated mitochondrial apoptosis. Nat Commun. 2021;12(1):2280.
266. Wang Y, Rosengarth A, Luecke H. Structure of the human p53 core domain in the absence of DNA. Acta Crystallogr D Biol Crystallogr. 2007;63(Pt 3):276-81.

267. Morgunova E, Yin YM, Das PK, Jolma A, Zhu FJ, Popov A, et al. Two distinct DNA sequences recognized by transcription factors represent enthalpy and entropy optima. Elife. 2018;7.
268. Sharp KA, Friedman RA, Misra V, Hecht J, Honig B. Salt Effects on Polyelectrolyte-Ligand Binding - Comparison of Poisson-Boltzmann, and Limiting Law Counterion Binding Models. Biopolymers. 1995;36(2):245-62.
269. Ha JH, Capp MW, Hohenwalter MD, Baskerville M, Record MT, Jr. Thermodynamic stoichiometries of participation of water, cations and anions in specific and non-specific binding of lac repressor to DNA. Possible thermodynamic origins of the "glutamate effect" on protein-DNA interactions. J Mol Biol. 1992;228(1):252-64.
270. Misra VK, Hecht JL, Yang AS, Honig B. Electrostatic contributions to the binding free energy of the lambdacI repressor to DNA. Biophys J. 1998;75(5):2262-73.
271. Grucza RA, Bradshaw JM, Mitaxov V, Waksman G. Role of electrostatic interactions in SH2 domain recognition: salt-dependence of tyrosyl-phosphorylated peptide binding to the tandem SH2 domain of the Syk kinase and the single SH2 domain of the Src kinase. Biochemistry. 2000;39(33):10072-81.
272. Uversky VN. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. Biochemistry. 1993;32(48):13288-98.
273. Yamakawa H, Stockmayer WH. Statistical Mechanics of Wormlike Chains .2. Excluded Volume Effects. J Chem Phys. 1972;57(7):2843-+.
274. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14(1):33-8, 27-8.
275. UniProt C. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51(D1):D523-D31.
276. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the National Center for Biotechnology Information in 2023. Nucleic Acids Res. 2023;51(D1):D29-D38.
277. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647-9.
278. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 2022;50(W1):W276-9.
279. Munoz V, Serrano L. Elucidating the folding problem of helical peptides using empirical parameters. Nat Struct Biol. 1994;1(6):399-409.
280. Johnson BA. Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol Biol. 2004;278:313-52.
281. Norris M, Fetler B, Marchant J, Johnson BA. NMRFx Processor: a cross-platform NMR data processing program. J Biomol NMR. 2016;65(3-4):205-16.

282. Ikura M, Kay LE, Bax A. A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry. 1990;29(19):4659-67.
283. Muhandiram DR, Kay, L.E. Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. Journal of Magnetic Resonance, Series B. 1994;103(3):203-16.
284. Lee W, Rahimi M, Lee Y, Chiu A. POKY: a software suite for multidimensional NMR and 3D structure calculation of biomolecules. Bioinformatics. 2021;37(18):3041-2.
285. Stone JE, McGreevy R, Isralewitz B, Schulten K. GPU-accelerated analysis and visualization of large structures solved by molecular dynamics flexible fitting. Faraday Discuss. 2014;169:265-83.

APPENDIX A – SUPPLEMENTARY TABLES

Table A1: K_D (nM) of DBD, ND WT and mutants across ionic strengths
Consensus DNA

	15 mM	**55 mM**	**85 mM**	**125 mM**	**145 mM**
DBD	0.018 ± 0.001	0.9 ± 0.01	2.08 ± 0.01	0.90 ± 0.07	1.39 ± 0.01
ND DE	NA	NA	2.1 ± 0.2	4.6 ± 0.3	8.0 ± 0.5
ND NP	NA	NA	1.4 ± 0.08	3.2 ± 0.2	9.4 ± 0.7
ND QS	NA	NA	3.08 ± 0.04	7 ± 1	17 ± 4
PRR-DBD	NA	NA	3.3 ± 0.3	11.9 ± 0.3	29 ± 2
PRR 33GS	NA	NA	1.0 ± 0.1	40 ± 3	87 ± 4
ND WT	0.70 ± 0.08	0.88 ± 0.03	7.0 ± 0.6	43 ± 3	230 ± 7
PRR NptoG	NA	NA	13.0 ± 0.6	41.4 ± 9.8	338 ±25
PRR PtoG	NA	NA	47 ± 2	216 ± 16	703 ± 60
PRR W91A	NA	NA	2.10 ± 0.01	10.5 ± 0.6	25 ± 1
T+20	NA	NA	NA	124 ± 2.6	3902 ± 339
P+20	NA	NA	NA	4.4 ± 0.4	29.4 ± 3.3
	165 mM	**185mM**	**205 mM**	**225 mM**	
DBD	7.1 ± 0.4	19 ± 2	58 ± 15	105 ± 5	
ND DE	17.9 ± 0.7	96 ± 16	290 ± 43	384 ± 10	
ND NP	19 ± 3	52 ± 7	85 ± 6	218 ± 15	
ND QS	31 ± 1	159 ± 9	347 ± 8	860 ± 340	
PRR-DBD	109 ± 7	NA	NA	NA	
PRR 33GS	210 ± 26	NA	NA	NA	
ND WT	473 ± 29	883 ± 49	1570	3713 ± 40	

Table A1 (Continued)

PRR NptoG	371 ± 26	NA	NA	NA
PRR PtoG	2444 ± 420	NA	NA	NA
PRR W91A	153 ± 15	NA	NA	NA
T+20	NA	NA	NA	NA
P+20	155.1 ± 10.3	NA	NA	NA

	Scrambled DNA				
	15 mM	**55 mM**	**85 mM**	**125 mM**	**145 mM**
DBD	0.03 ± 0.004	7.8 ± 1.5	22 ± 2	89 ± 5	113 ± 7
ND DE	NA	NA	25.8 ± 0.2	108 ± 2	245 ± 7
ND NP	NA	NA	55 ± 6	82 ± 2	334 ± 7
ND QS	NA	NA	150 ± 6	213 ± 9	329 ± 25
PRR-DBD	NA	NA	19.2 ± 0.1	126 ± 4	398 ± 6
PRR 33GS	NA	NA	86 ± 3	392 ± 5	678 ± 4
ND WT	1.2 ± 2	69 ± 3	170 ± 10	193 ± 8	578 ± 40
PRR NptoG	NA	NA	287 ± 18	998 ± 21	1234 ± 39
PRR PtoG	NA	NA	2095 ± 84	*5802* ± 178	*6264* ± 102
PRR W91A	NA	NA	106 ± 7	295 ± 7	467 ± 44
T+20	NA	NA	NA	445 ± 37	7479 ± 51
P+20	NA	NA	NA	220 ± 8.8	502 ± 27
	165 mM	**185 mM**	**205 mM**	**225 mM**	
DBD	285 ± 25	590 ± 26	717 ± 62	1388 ± 44	
ND DE	593 ± 58	761 ± 70	1083 ± 50	1676 ± 19	
ND NP	830 ± 27	2006 ± 119	2777 ± 85	3386 ± 196	
ND QS	855 ± 7	1042 ± 58	2371 ± 81	2860 ± 38	
PRR-DBD	746 ± 10	NA	NA	NA	

Table A1 (Continued)

PRR 33GS	970 ± 4	NA	NA	NA
ND WT	1500 ± 46	1743 ± 105	2727 ± 66	3705 ± 230
PRR NptoG	1478 ± 13	NA	NA	NA
PRR PtoG	NA	NA	NA	NA
PRR W91A	1054 ± 39	NA	NA	NA
T+20	NA	NA	NA	NA
P+20	996 ± 38.5	NA	NA	NA

Notes on Table A1: Entries in *italics* did not fully reach saturation; thus values provided here are estimates. Data points for ND mutants that were not collected are represented here as 'NA.'

Table A2: ΔG at varying ionic strength in kcal/mol

	Consensus DNA				
	15 mM	**55 mM**	**85 mM**	**125 mM**	**145 mM**
DBD	-14.65 ± 0.05	-12.42 ± 0.09	-11.83 ± 0.01	-12.33 ± 0.05	-11.11 ± 0.04
ND DE	NA	NA	-11.83 ± 0.01	-11.37 ± 0.07	-11.04 ± 0.04
ND NP	NA	NA	-12.07 ± 0.07	-11.59 ± 0.05	-10.94 ± 0.03
ND QS	NA	NA	-11.60 ± 0.01	-11.14 ± 0.11	-10.58 ± 0.11
PRR-DBD	NA	NA	-11.57 ± 0.05	-10.85 ± 0.02	-10.27 ± 0.05
PRR 33GS	NA	NA	-12.27 ± 0.07	-10.09 ± 0.04	-9.62 ± 0.03
ND WT	-12.16 ± 0.01	-9.76 ± 0.02	-11.12 ± 0.05	-10.04 ± 0.02	-9.05 ± 0.04
PRR NptoG	NA	NA	-10.75 ± 0.03	-10.07 ± 0.38	-8.82 ± 0.04
PRR PtoG	NA	NA	-9.99 ± 0.03	-8.90 ± 0.02	-8.38 ± 0.06
PRR W91A	NA	NA	-11.83 ± 0.02	-10.89 ± 0.04	-10.39 ± 0.02
T+20	NA	NA	NA	-9.43 ± 0.07	-6.76 ± 0.05
P+20	NA	NA	NA	-11.39 ± 0.06	-1026 ± 0.07

Table A2 (Continued)

	165mM	185 mM	205 mM	225 mM
DBD	-11.11 ± 0.04	-10.52 ± 0.08	-9.87 ± 0.15	-9.52 ± 0.01
ND DE	-10.56 ± 0.26	-9.57 ± 0.09	-8.91 ± 0.10	-8.75 ± 0.02
ND NP	-10.54 ± 0.04	-9.93 ± 0.08	-9.64 ± 0.04	-9.08 ± 0.04
ND QS	-10.54 ± 0.01	-9.27 ± 0.11	-8.81 ± 0.03	-8.27 ± 0.24
PRR-DBD	-9.49 ± 0.04	NA	NA	NA
PRR 33GS	-9.10 ± 0.07	NA	NA	NA
ND WT	-8.62 ± 0.03	-8.25 ± 0.06	-7.91 ± 0.02	-7.40 ± 0.27
PRR NptoG	-8.77 ± 0.04	NA	NA	NA
PRR PtoG	-7.65 ± 0.10	NA	NA	NA
PRR W91A	-11.11 ± 0.04	NA	NA	NA
T+20	NA	NA	NA	NA
P+20	-9.28 ± 0.04	NA	NA	NA

	Scrambled DNA				
	15 mM	**55 mM**	**85 mM**	**125 mM**	**145 mM**
DBD	-13.35 ± 0.17	-11.06 ± 0.12	-10.44 ± 0.04	-9.61 ± 0.04	-9.47 ± 0.03
ND DE	NA	NA	-9.88 ± 0.40	-9.50 ± 0.01	-9.01 ± 0.02
ND NP	NA	NA	-12.07 ± 0.07	-11.59 ± 0.05	-10.94 ± 0.03
ND QS	NA	NA	-9.30 ± 0.02	-9.10 ± 0.02	-8.84 ± 0.05
PRR-DBD	NA	NA	-10.95 ± 0.02	-9.41± 0.02	-8.73 ± 0.01
PRR 33GS	NA	NA	-9.63 ± 0.01	-8.74 ± 0.01	-8.41 ± 0.02
ND WT	-12.16 ± 0.01	-9.76 ± 0.02	-9.23 ± 0.04	-9.16 ± 0.03	-8.51 ± 0.04
PRR NptoG	NA	NA	-8.92 ± 0.04	-8.18 ± 0.01	-8.06 ± 0.02
PRR PtoG	NA	NA	-7.75 ± 0.02	*-7.14 ± 0.02*	*-7.08 ± 0.01*
PRR W91A	NA	NA	-9.53 ± 0.04	-9.09 ± 0.01	-8.66 ± 0.06
T+20	NA	NA	NA	-8.66 ± 0.05	*-6.40 ± 0.03*

Table A2 (Continued)

P+20	NA	NA	NA	-9.08 ± 0.02	-8.59 ± 0.03
	165 mM	**185 mM**	**205 mM**	**225 mM**	
DBD	-8.92 ± 0.05	-8.49 ± 0.03	-8.38 ± 0.05	-7.99 ± 0.03	
ND DE	-8.49 ± 0.31	-8.34 ± 0.05	-8.13 ± 0.31	-7.87 ± 0.01	
ND NP	-10.54 ± 0.04	-9.93 ± 0.08	-9.64 ± 0.04	-9.08 ± 0.04	
ND QS	-8.27 ± 0.01	-8.16 ± 0.03	-7.67 ± 0.29	-7.58 ± 0.01	
PRR-DBD	-8.35 ± 0.01	NA	NA	NA	
PRR 33GS	-8.20 ± 0.01	NA	NA	NA	
ND WT	-7.94 ± 0.02	-7.85 ± 0.04	-7.59 ± 0.02	-7.74 ± 0.27	
PRR NptoG	-7.95 ± 0.01	NA	NA	NA	
PRR PtoG	NA	NA	NA	NA	
PRR W91A	-8.16 ± 0.02	NA	NA	NA	
T+20	NA	NA	NA	NA	
P+20	-8.17 ± 0.02	NA	NA	NA	

Notes on Table A2: All values are in kcal/mol. Entries in *italics* did not fully reach saturation; thus values provided here are estimates. Data points for ND mutants that were not collected are represented here as 'NA.'

Table A3: SEOE of DBD, ND WT, and mutants

	Consensus DNA				
	15 mM	**55 mM**	**85 mM**	**125 mM**	**145 mM**
DBD	0.0450	0.0543	0.0368	0.0310	0.0317
ND DE	NA	NA	0.0411	0.0276	0.0332
ND NP	NA	NA	0.0186	0.0402	0.0345
ND QS	NA	NA	0.0351	0.0514	0.0377
PRR-DBD	NA	NA	0.0020	0.0004	0.0005
PRR 33GS	NA	NA	0.0005	0.0022	0.0006
ND WT	0.0852	0.0193	0.0518	0.0254	0.0590

Table A3 (Continued)

PRR NptoG	NA	NA	0.0008	0.0030	0.0010
PRR PtoG	NA	NA	0.1288	0.0003	0.0025
PRR W91A	NA	NA	0.0004	0.0001	0.0005
T+20	NA	NA	NA	0.0019	0.0017
P+20	NA	NA	NA	0.0032	0.0020
	165 mM	**185mM**	**205 mM**	**225 mM**	
DBD	0.0318	0.0263	0.0776	0.0315	
ND DE	0.0186	0.0331	0.0901	0.0820	
ND NP	0.0344	0.0291	0.0520	0.0368	
ND QS	0.0344	0.0291	0.0520	0.0368	
PRR-DBD	0.0001	NA	NA	NA	
PRR 33GS	0.0018	NA	NA	NA	
ND WT	0.0370	0.0395	0.0164	0.0399	
PRR NptoG	0.0020	NA	NA	NA	
PRR PtoG	0.0025	NA	NA	NA	
PRR W91A	0.0007	NA	NA	NA	
T+20	NA	NA	NA	NA	
P+20	0.0026	NA	NA	NA	
			Scrambled DNA		
	15 mM	**55 mM**	**85 mM**	**125 mM**	**145 mM**
DBD	0.0744	0.0180	0.0131	0.0180	0.0173
ND DE	NA	NA	0.0210	0.0306	0.0185
ND NP	NA	NA	0.0224	0.0320	0.0138
ND QS	NA	NA	0.0234	0.0274	0.0161
PRR-DBD	NA	NA	0.0016	0.0004	0.0001
PRR 33GS	NA	NA	0.0006	0.0010	0.0003

Table A3 (Continued)

ND WT	0.0003	0.0198	0.0134	0.0332	0.0189
PRR NptoG	NA	NA	0.0006	0.0010	0.0019
PRR PtoG	NA	NA	0.0004	0.0019	0.0011
PRR W91A	NA	NA	0.0002	0.0010	0.0006
T+20	NA	NA	NA	0.0006	0.0025
P+20	NA	NA	NA	0.0007	0.0100
	165 mM	**185 mM**	**205 mM**	**225 mM**	
DBD	0.0104	0.0243	0.0249	0.0250	
ND DE	0.0168	0.0379	0.0248	0.0344	
ND NP	0.0242	0.0148	0.0206	0.0433	
ND QS	0.0438	0.0178	0.0087	0.0232	
PRR-DBD	0.0001	NA	NA	NA	
PRR 33GS	0.0005	NA	NA	NA	
ND WT	0.0323	0.0201	0.0172	0.0212	
PRR NptoG	0.003	NA	NA	NA	
PRR PtoG	NA	NA	NA	NA	
PRR W91A	0.0003	NA	NA	NA	
T+20	NA	NA	NA	NA	
P+20	0.0001	NA	NA	NA	

Notes on Table A3: Triplicate runs of fluorescence anisotropy were evaluated using standard error of estimate as a measure of data quality using the following:

$$\frac{(Actual - Expected)^2}{dof}$$

Where *Actual* is the observed anisotropy values at a given concentration of protein and DNA, *Expected* is the predicted anisotropy value at these concentrations, and *dof* is degrees of freedom.

Data points for ND mutants that were not collected are represented here as ‘NA.’

Table A4: Predicted R_S based on experimentation versus formulas

	A	B	C	D	E
	Exp. Data	Wilkins & Smith	Marsh ordered	Marsh IDP	Marsh IDP + ordered
DBD	2.71	1.56 – 3.12	2.29	3.87	2.29
	2.74*				
PRR-DBD	3.08	1.62 – 3.26	2.38	4.16	3.76
ND WT	3.28	1.72 – 3.48	2.53	4.63	4.79
ND P+20	3.55	1.75 – 3.54	2.57	4.78	5.05

Stokes radius can be estimated by various methods and compared to experimental values. Values are determined by: A) experimentation. * Denotes experimental data determined using dynamic light scattering. All others were determined using analytical SEC. Note that TAD2 and PRR mutants that have similar (<1 kDa) differences in molecular weight are excluded for clarity. B) Wilkins and Smith's estimate for globular proteins (173), C) Marsh's estimate for ordered proteins (174), D) Marsh's estimate for disordered proteins IDP, E) Marsh's estimate for the ordered DBD and disordered TAD domains added together.

Table A5: Chemical shifts of ^{1}H-^{13}C-^{15}N KIX apo

Residue number	AA	CA	CB	H	N
587	G	45.361		8.558	111.192
588	V	62.285	32.804	7.966	119.393
589	R	56.259	30.417	8.553	125.603
590	K	55.669	32.866	8.447	121.962
591	G	45.351		8.449	110.33
592	W	57.978	28.084	7.855	118.273
593	H	56.752	29.947	6.948	120.287
594	E	58.198	29.593	7.586	116.319
595	H	55.362	31.101	7.54	114.263

Table A5 (Continued)

596	V	62.078	33.347	7.552	120.631
597	T	60.179	71.207	7.495	115.043
598	Q	58.84	28.19	8.98	120.43
599	D	57.087	40.34	8.454	117.659
600	L	58.238	41.702	7.629	122.739
601	R	60.351	30.702	7.819	118.255
602	S	61.672	62.872	8.435	112.619
603	H			8.229	120.843
608	L	58.796	42.65	7.984	121.385
609	V	67.465	31.563	8.168	118.347
610	Q	57.992	28.872	8.104	116.029
611	A	53.985	19.07	7.529	119.356
612	I	62.929	39.302	7.561	114.391
613	F	59.801	38.073	8.252	117.669
615	T	58.731	69.907	7.947	114.609
617	D	51.497	41.938	7.968	120.243
619	A	53.148	18.825	8.338	121.800
620	A	53.253	19.168	7.756	121.14
621	L	56.016	41.816	7.658	117.700
622	K	56.485	32.71	7.815	118.746
623	D			8.102	120.527
625	R	58.774	29.511	8.089	119.233
626	M			8.204	119.50
627	E	59.239	29.737	8.277	117.884
628	N	55.723	38.073	8.266	117.738
629	L	57.986	41.523	7.922	123.185

Table A5 (Continued)

630	V	66.909	31.601	8.395	119.862
631	A	55.454	18.04	7.864	120.492
632	Y	60.972	38.11	7.877	120.087
633	A	55.346	20.503	8.547	122.300
634	K	59.702	32.96	8.686	115.965
635	K	59.308	31.983	7.76	123.314
636	V	65.915	31.804	7.985	119.277
637	E	61.435	28.884	8.516	120.273
638	G	47.603		8.164	107.078
639	D	57.11	40.305	8.197	122.318
640	M	58.793	32.147	8.055	121.047
641	Y	62.082	39.116	9.293	121.337
642	E	57.969	30.242	8.034	114.137
643	S	60.341	64.524	7.988	112.612
644	A	52.921	19.528	8.216	124.161
645	N	53.535	40.275	9.157	117.104
646	S	56.693	65.156	6.899	111.461
647	R	59.062	30.304	8.76	121.85
648	D	57.362	40.404	8.191	117.091
649	E	59.384	30.188	7.775	121.134
650	Y	61.002	38.96	7.36	121.33
651	Y	60.906	38.039	8.164	115.831
652	H	59.631	28.759	8.331	119.057
653	L	57.85	42.018	9.011	120.559
654	L	57.803	41.592	7.789	120.897
655	A	55.447	18.004	8.119	120.908

Table A5 (Continued)

656	E	59.22	29.803	8.351	115.973
657	K	57.993	31.835	7.877	120.424
658	I	66.078	37.944	8.264	119.096
659	Y	60.98	38.156	7.968	119.509
660	K	59.813	32.793	8.206	119.222
661	I	65.109	38.309	8.374	120.79
662	Q	59.522	28.355	8.299	118.921
663	K	58.525	31.929	8.187	119.866
664	E	59.223	29.529	8.102	120.281
665	L	57.665	41.597	8.141	119.975
666	E	58.986	29.618	8.005	120.238
669	R			8.263	119.358
670	R	57.255	30.6	7.936	119.085
671	S	59.003	63.792	8.021	115.371
672	R	56.123	30.629	8.015	122.203
673	L	56.786	43.394	7.75	128.177

Table A6: Chemical shifts of ^{1}H-^{15}N KIX bound to synthetic cMyb WT

Residue number	AA	H	N	Chemical shift change from apo
587	G	8.5502	111.2743	0.018
588	V	7.9514	119.4121	0.019
589	R	8.5682	125.7858	0.018
590	K	8.4329	121.7982	0.021
591	G	8.4103	110.2988	0.017
592	W	7.8317	118.1405	0.020
593	H	6.8396	120.6353	0.024

Table A6 (Continued)

594	E	7.5096	116.2761	0.010
595	H	7.5629	114.4252	0.039
596	V	7.5205	120.7389	0.018
597	T	7.4809	115.0298	0.017
599	D	8.4349	117.7094	0.019
600	L	7.6187	122.7211	0.028
601	R	7.8109	118.3907	0.008
602	S	8.4008	112.7051	0.021
603	H	8.1854	120.6370	0.052
608	L	7.9607	121.4793	0.019
609	V	8.1378	118.3865	0.055
610	Q	8.1082	116.1762	0.013
611	A	7.5542	119.3556	0.050
612	I	7.4803	114.3979	0.056
615	T	7.9798	114.8990	0.015
617	D	7.9741	120.1713	0.026
619	A	8.4066	121.9700	0.042
620	A	7.7496	121.0183	0.018
621	L	7.6532	117.8140	0.014
622	K	7.8145	118.8496	0.012
623	D	8.0691	120.4529	0.034
625	R	8.0715	119.3114	0.022
626	M	8.1801	119.1862	0.051
627	E	8.2673	119.1390	0.025
628	N	8.2642	117.7799	0.022
629	L	7.9023	123.3372	0.022

Table A6 (Continued)

630	V	8.2834	119.4889	0.045
631	A	7.8073	120.6114	0.058
632	Y	7.8680	120.3835	0.011
633	A	8.5430	122.3239	0.028
634	K	8.6672	116.0493	0.017
635	K	7.7449	123.3854	0.024
636	V	7.9880	119.1615	0.027
637	E	8.4954	120.2595	0.029
638	G	8.1251	106.9254	0.054
639	D	8.2027	122.3914	0.017
640	M	7.9922	121.0888	0.049
641	Y	9.2731	121.6253	0.007
642	E	7.9798	113.7969	0.040
643	S	7.9520	112.5484	0.036
644	A	8.2307	124.2059	0.027
645	N	9.1741	117.2512	0.014
646	S	6.8741	111.7950	0.020
647	R	8.7601	121.9362	0.014
648	D	8.0580	115.9771	0.070
649	E	7.7722	121.0394	0.013
650	Y	7.3364	121.4554	0.005
651	Y	8.1336	115.9089	0.016
652	H	8.1637	116.9976	0.043
653	L	8.9543	120.3520	0.040
654	L	7.7527	120.7317	0.043
655	A	8.1085	121.0654	0.023

Table A6 (Continued)

656	E	8.3434	115.8445	0.020
657	K	7.8468	120.5688	0.017
658	I	8.2588	119.8124	0.023
659	Y	7.9853	119.4442	0.010
660	K	8.2794	122.3955	0.015
661	I	8.3515	120.8676	0.034
662	Q	8.2697	120.7420	0.019
663	K	8.1791	119.9538	0.029
664	E	7.9934	120.3027	0.024
665	L	8.1301	120.1263	0.019
669	R	8.2549	119.2387	0.018
670	R	7.9237	119.0744	0.019
671	S	8.0072	115.3793	0.018
672	R	7.9820	122.2535	0.016
673	L	7.7219	128.1726	0.016

Table A7: Chemical shifts of ^{1}H-^{15}N KIX bound to cMyb mimic 104-2

Residue number	AA	H	N	Chemical shift change from apo
587	G	8.5567	111.3940	0.001
588	V	7.9631	119.5825	0.008
589	R	8.5688	125.8943	0.003
590	K	8.5414	122.9096	0.170
591	G	8.4237	110.4346	0.010
592	W	7.8283	118.2333	0.021
593	H	6.8679	120.7106	0.027
594	E	7.5286	116.3599	0.015

Table A7 (Continued)

595	H	7.5188	114.6360	0.025
596	V	7.5222	120.6114	0.025
597	T	7.5168	115.2138	0.051
599	D	8.4290	117.7961	0.022
600	L	7.6368	122.9426	0.019
601	R	7.7910	118.2763	0.021
602	S	8.4253	112.7800	0.008
603	H	8.2577	120.2663	0.087
608	L	7.9554	121.5579	0.014
609	V	8.1140	118.4094	0.070
610	Q	8.1143	116.2231	0.004
611	A	7.5312	119.6485	0.019
612	I	7.4933	114.9030	0.045
615	T	7.9864	115.1858	0.039
617	D	7.9551	120.1628	0.031
619	A	8.3708	121.7956	0.017
620	A	7.7623	121.1150	0.007
621	L	7.6041	117.3875	0.090
622	K	7.7613	118.6372	0.068
623	D	8.1288	120.6978	0.036
625	R	8.0269	119.2756	0.064
626	M	8.2371	119.5119	0.033
627	E	8.2929	118.6622	0.057
628	N	8.3029	118.0322	0.033
629	L	7.8905	123.2538	0.039
630	V	8.3167	119.2039	0.075

Table A7 (Continued)

631	A	7.8412	120.7455	0.023
632	Y	7.9155	120.2973	0.053
633	A	8.5189	122.4970	0.021
634	K	8.6884	116.1364	0.008
635	K	7.7786	123.6609	0.029
636	V	7.9957	119.3485	0.020
637	E	8.5061	120.3916	0.008
638	G	8.1681	107.2395	0.015
639	D	8.2098	122.5614	0.020
640	M	8.0376	121.2828	0.005
641	Y	9.2825	121.6029	0.016
642	E	8.0308	114.1166	0.028
643	S	7.9781	112.7722	0.006
644	A	8.2233	124.3434	0.011
645	N	9.1599	117.3282	0.005
646	S	6.8895	111.8063	0.021
647	R	8.7676	122.0432	0.004
648	D	8.1445	116.1264	0.020
649	E	7.7757	121.1422	0.002
650	Y	7.3245	121.5153	0.012
651	Y	8.1436	116.1300	0.017
652	H	8.1628	117.1768	0.023
653	L	8.9776	120.5572	0.003
654	L	7.7897	120.7970	0.015
655	A	8.1393	120.8466	0.030
656	E	8.3496	115.9600	0.002

Table A7 (Continued)

657	K	7.8965	120.8026	0.050
658	I	8.2692	120.0403	0.013
659	Y	7.9718	119.4259	0.022
660	K	8.2796	122.5816	0.011
661	I	8.3963	120.9389	0.021
662	Q	8.2877	120.8805	0.013
663	K	8.2243	120.0225	0.043
664	E	7.9891	120.3300	0.022
665	L	8.1829	119.9675	0.062
669	R	8.2660	119.2858	0.011
670	R	7.9118	119.0807	0.026
671	S	7.9834	115.2509	0.044
672	R	7.9458	122.3125	0.042
673	L	7.7130	128.3054	0.012

Table A8: Chemical shifts of ^{1}H-^{15}N KIX bound to cMyb mimic 109-1

Residue number	AA	1H	15N	Chemical shift change from apo
587	G	8.5507	111.2725	0.001
588	V	7.9579	119.4577	0.008
589	R	8.5629	125.7705	0.003
590	K	8.5274	122.3849	0.125
591	G	8.4186	110.3155	0.009
592	W	7.8263	118.1260	0.023
593	H	6.8641	120.5657	0.027
594	E	7.5244	116.2573	0.015
595	H	7.5080	114.5684	0.033

Table A8 (Continued)

596	V	7.5191	120.5459	0.020
597	T	7.5186	115.1493	0.064
599	D	8.4246	117.6739	0.020
600	L	7.6274	122.8049	0.018
601	R	7.7867	118.1640	0.022
602	S	8.4189	112.6667	0.008
603	H	8.2320	120.1218	0.087
608	L	7.9583	121.4184	0.009
609	V	8.1178	118.3299	0.061
610	Q	8.1169	116.0240	0.023
611	A	7.5253	119.5097	0.018
612	I	7.5056	114.6106	0.015
615	T	7.9809	114.9632	0.030
617	D	7.9547	120.1209	0.020
619	A	8.3614	121.6782	0.018
620	A	7.7585	121.0026	0.008
621	L	7.6060	117.3373	0.081
622	K	7.7651	118.5300	0.064
623	D	8.1172	120.5152	0.028
625	R	8.0378	119.1890	0.047
626	M	8.2201	119.3834	0.025
627	E	8.2850	118.6449	0.044
628	N	8.2929	117.9190	0.032
629	L	7.8986	123.2010	0.024
630	V	8.3060	119.0449	0.079
631	A	7.8355	120.5838	0.021

Table A8 (Continued)

632	Y	7.9111	120.1927	0.053
633	A	8.5274	122.3849	0.009
634	K	8.6791	116.0191	0.005
635	K	7.7680	123.5384	0.027
636	V	7.9760	119.2722	0.014
637	E	8.4986	120.2738	0.009
638	G	8.1628	107.1460	0.016
639	D	8.2049	122.4335	0.025
640	M	8.0291	121.1615	0.002
641	Y	9.2746	121.4744	0.015
642	E	8.0270	114.0012	0.030
643	S	7.9686	112.6443	0.005
644	A	8.2135	124.2240	0.014
645	N	9.1534	117.2067	0.003
646	S	6.8805	111.6804	0.021
647	R	8.7593	121.9133	0.004
648	D	8.1333	116.0111	0.018
649	E	7.7700	121.1202	0.003
650	Y	7.3210	121.3866	0.011
651	Y	8.1333	116.0111	0.011
652	H	8.1575	117.0492	0.024
653	L	8.9697	120.4525	0.005
654	L	7.7809	120.9414	0.002
655	A	8.1353	120.7633	0.031
656	E	8.3432	115.8349	0.004
657	K	7.8858	120.6664	0.028

Table A8 (Continued)

658	I	8.2658	119.8869	0.013
659	Y	7.9789	119.3814	0.014
660	K	8.2736	122.4236	0.002
661	I	8.3793	120.8197	0.013
662	Q	8.2818	120.7598	0.015
663	K			0.000
664	E	7.9827	120.2170	0.022
665	L	8.1887	119.9270	0.066
669	R	8.2646	119.1741	0.013
670	R	7.9111	118.9849	0.020
671	S	7.9831	115.1939	0.036
672	R	7.9495	122.2131	0.034
673	L	7.7094	128.1820	0.012

Table A9: Chemical shifts of ^{1}H-^{15}N KIX bound to LC-A-122-2

Residue number	AA	1H	15N	Chemical shift change from apo
587	G	8.551	111.259	0.009
588	V	7.961	119.235	0.021
589	R	8.565	125.809	0.009
590	K	8.388	122.001	0.059
591	G	8.417	110.322	0.009
592	W	7.829	118.097	0.035
593	H	6.852	120.563	0.024
594	E	7.520	116.228	0.012
595	H	7.479	114.706	0.082

Table A9 (Continued)

596	V	7.511	120.592	0.040
597	T	7.403	114.750	0.052
599	D	8.422	117.727	0.038
600	L	7.642	122.844	0.047
601	R	7.781	118.142	0.054
602	S	8.422	112.729	0.015
603	H	8.206	120.510	0.034
608	L	7.948	121.429	0.027
609	V	8.102	118.296	0.089
610	Q	8.120	116.119	0.018
611	A	7.524	119.511	0.015
612	I	7.493	114.413	0.019
615	T	7.965	114.918	0.023
617	D	7.950	120.048	0.050
619	A	8.403	122.038	0.017
620	A	7.753	120.991	0.014
621	L	7.590	117.213	0.112
622	K	7.758	118.528	0.074
623	D	8.121	120.601	0.043
625	R	8.112	119.346	0.022
626	M	8.157	119.406	0.037
627	E	8.312	118.934	0.028
628	N	8.296	117.931	0.035
629	L	7.924	123.193	0.033
630	V	8.358	119.779	0.057
631	A	7.821	120.569	0.016

Table A9 (Continued)

632	Y	7.915	120.205	0.047
633	A	8.505	122.364	0.046
634	K	8.676	116.007	0.004
635	K	7.765	123.601	0.037
636	V	7.983	119.202	0.016
637	E	8.496	120.248	0.024
638	G	8.164	107.148	0.014
639	D	8.216	122.476	0.045
640	M	8.029	121.184	0.003
641	Y	9.277	121.457	0.019
642	E	8.033	114.031	0.041
643	S	7.974	112.671	0.008
644	A	8.217	124.230	0.028
645	N	9.158	117.242	0.002
646	S	6.878	111.696	0.025
647	R	8.750	121.907	0.006
648	D	8.136	115.984	0.025
649	E			
650	Y	7.309	121.436	0.027
651	Y	8.138	115.852	0.008
652	H	8.148	116.977	0.041
653	L	8.973	120.504	0.011
654	L	7.785	120.597	0.044
655	A	8.137	121.390	0.075
656	E	8.339	115.807	0.013
657	K	7.882	120.649	0.030

Table A9 (Continued)

658	I	8.235	119.902	0.028
659	Y	7.961	119.547	0.014
660	K	8.290	122.566	0.018
661	I	8.384	120.918	0.019
662	Q	8.278	120.732	0.013
663	K	8.195	119.825	0.037
664	E	7.994	120.165	0.030
665	L	8.126	120.381	0.032
669	R	8.242	119.233	0.010
670	R	7.891	118.921	0.040
671	S	7.968	115.114	0.068
672	R	7.962	122.564	0.057
673	L	7.714	128.271	0.038

Table A10: Chemical shifts of ^{1}H-^{15}N KIX bound to LC-A-122-3

Residue number	AA	1H	15N	Chemical shift change from apo
587	G	8.551	111.259	0.009
588	V	7.961	119.235	0.021
589	R	8.565	125.809	0.009
590	K	8.388	122.001	0.059
591	G	8.417	110.322	0.009
592	W	7.829	118.097	0.035
593	H	6.852	120.563	0.024
594	E	7.520	116.228	0.012
595	H	7.479	114.706	0.082
596	V	7.511	120.592	0.040

Table A10 (Continued)

597	T	7.403	114.750	0.052
599	D	8.422	117.727	0.038
600	L	7.642	122.844	0.047
601	R	7.781	118.142	0.054
602	S	8.422	112.729	0.015
603	H	8.206	120.510	0.034
608	L	7.948	121.429	0.027
609	V	8.102	118.296	0.089
610	Q	8.120	116.119	0.018
611	A	7.524	119.511	0.015
612	I	7.493	114.413	0.019
615	T	7.965	114.918	0.023
617	D	7.950	120.048	0.050
619	A	8.403	122.038	0.017
620	A	7.753	120.991	0.014
621	L	7.590	117.213	0.112
622	K	7.758	118.528	0.074
623	D	8.121	120.601	0.043
625	R	8.112	119.346	0.022
626	M	8.157	119.406	0.037
627	E	8.312	118.934	0.028
628	N	8.296	117.931	0.035
629	L	7.924	123.193	0.033
630	V	8.358	119.779	0.057
631	A	7.821	120.569	0.016
632	Y	7.915	120.205	0.047

Table A10 (Continued)

633	A	8.505	122.364	0.046
634	K	8.676	116.007	0.004
635	K	7.765	123.601	0.037
636	V	7.983	119.202	0.016
637	E	8.496	120.248	0.024
638	G	8.164	107.148	0.014
639	D	8.216	122.476	0.045
640	M	8.029	121.184	0.003
641	Y	9.277	121.457	0.019
642	E	8.033	114.031	0.041
643	S	7.974	112.671	0.008
644	A	8.217	124.230	0.028
645	N	9.158	117.242	0.002
646	S	6.878	111.696	0.025
647	R	8.750	121.907	0.006
648	D	8.136	115.984	0.025
649	E			
650	Y	7.309	121.436	0.027
651	Y	8.138	115.852	0.008
652	H	8.148	116.977	0.041
653	L	8.973	120.504	0.011
654	L	7.785	120.597	0.044
655	A	8.137	121.390	0.075
656	E	8.339	115.807	0.013
657	K	7.882	120.649	0.030
658	I	8.235	119.902	0.028

Table A10 (Continued)

659	Y	7.961	119.547	0.014
660	K	8.290	122.566	0.018
661	I	8.384	120.918	0.019
662	Q	8.278	120.732	0.013
663	K	8.195	119.825	0.037
664	E	7.994	120.165	0.030
665	L	8.126	120.381	0.032
669	R	8.242	119.233	0.010
670	R	7.891	118.921	0.040
671	S	7.968	115.114	0.068
672	R	7.962	122.564	0.057
673	L	7.714	128.271	0.038

Table A11: Chemical shifts of ^{1}H-^{15}N KIX bound to synthetic MLL WT

Residue number	AA	1H	15N	Chemical shift change from apo
587	G	8.5559	111.4076	0.0013
588	V	7.9434	119.9087	0.0536
589	R	8.5827	125.9274	0.0118
590	K	8.4169	122.1558	0.0448
591	G	8.4200	110.4439	0.0077
592	W			
593	H	6.83832	120.6748	0.0193
594	E	7.51706	116.4348	0.0144
595	H	7.44843	114.4048	0.1010
596	V	7.5203	120.761	0.0182

Table A11 (Continued)

597	T	7.48787	115.0588	0.0218
599	D	8.42369	117.783	0.0000
600	L	7.62893	123.1349	0.0269
601	R	7.82722	118.3823	0.0326
602	S	8.43635	112.9201	0.0241
603	H	8.17584	120.8238	0.0261
608	L	8.00316	122.0068	0.0675
609	V	8.15865	118.7346	0.0139
610	Q	7.89239	118.2773	0.3628
611	A	7.78276	118.125	0.3369
612	I	7.58049	114.4094	0.0697
615	T	8.12939	110.5319	0.6462
617	D	8.14628	121.3259	0.2257
619	A			
620	A	7.7767	121.2469	0.0260
621	L	7.8395	114.6625	0.4933
622	K	7.3161	115.6061	0.6840
623	D	7.4716	123.2459	0.7377
625	R	9.1108	130.1581	
626	M	9.5316	119.2620	1.4443
627	E	8.1781	119.2915	0.0400
628	N	7.4614	118.5340	0.8318
629	L	7.8272	118.3823	0.4519
630	V	8.3037	119.5820	0.0246
631	A			
632	Y			

Table A11 (Continued)

633	A	8.5733	122.2810	0.0490
634	K	8.5490	115.9126	0.1348
635	K	7.8375	124.4929	0.1581
636	V	8.1042	119.5702	0.1296
637	E	8.4414	118.9558	0.2203
638	G	8.1446	107.0110	0.0349
639	D	8.2687	122.6077	0.0781
640	M	8.0449	121.2995	0.0119
641	Y	9.2710	121.6685	0.0090
642	E	8.0475	114.1389	0.0443
643	S	7.9707	112.7488	0.0061
644	A	8.2161	124.2745	0.0233
645	N	9.1879	117.5088	0.0345
646	S	6.8674	111.9725	0.0133
647	R	8.7671	122.1392	0.0158
648	D	8.2190	116.6331	0.1189
649	E	7.9388	120.9273	0.1668
650	Y	7.3226	122.1251	0.0910
651	Y			
652	H	8.1408	117.0247	0.0515
653	L	8.9638	120.4938	0.0181
654	L	7.9568	121.3254	0.1778
655	A	8.0735	121.2349	0.0583
656	E	8.3500	115.9699	0.0009
657	K	8.3580	119.1162	0.5341
658	I	8.2777	119.8721	0.0228

Table A11 (Continued)

659	Y	7.9625	119.6479	0.0313
660	K	8.3333	122.8285	0.0619
661	I			
662	Q	8.2535	120.6161	0.0422
663	K	8.1929	119.6910	0.0655
664	E	8.3226	120.3606	0.3243
665	L			
669	R	8.2602	119.2104	0.0206
670	R			
671	S	8.0567	115.7730	0.0611
672	R	8.0170	122.5843	0.0436
673	L	7.6490	128.3814	0.0768

Table A12: Chemical shifts of ^{1}H-^{15}N KIX bound to MLL mimic 6

Residue number	AA	1H	15N	Chemical shift change from apo
587	G	8.5609	111.3900	0.001
588	V	7.9695	119.6084	0.011
589	R	8.5624	125.8100	0.006
590	K	8.4479	122.1493	0.013
591	G	8.4501	110.5183	0.010
592	W	7.8614	118.4709	0.023
593	H	6.9431	120.5012	0.028
594	E	7.5797	116.5162	0.018
595	H	7.5007	114.6347	0.047
596	V	7.5445	120.7696	0.000

Table A12 (Continued)

597	T	7.5075	115.2143	0.013
599	D	8.4523	117.8300	0.003
600	L	7.6431	122.9431	0.007
601	R	7.8271	118.3978	0.018
602	S	8.4364	112.8100	0.002
603	H	8.2297	120.2700	0.076
608	L	8.0085	121.5319	0.037
609	V	8.1052	118.3260	0.083
610	Q	8.1183	116.3988	0.029
611	A	7.5585	119.5825	0.032
612	I	7.5410	114.4744	0.013
615	T	7.9636	114.8542	0.009
617	D	7.9531	120.3800	0.021
619	A	8.3739	121.7841	0.010
620	A	7.7746	121.1289	0.007
621	L	7.6184	117.3864	0.080
622	K	7.7627	118.6229	0.068
623	D	8.0990	120.5900	0.016
625	R	8.0697	119.4198	0.023
626	M	8.2380	119.2668	0.038
627	E	8.3159	118.8161	0.043
628	N	8.2878	117.9647	0.010
629	L	7.9146	123.2296	0.030
630	V	8.3354	118.9028	0.075
631	A	7.8154	120.6951	0.059
632	Y	7.9296	120.2940	0.056

Table A12 (Continued)

633	A	8.5548	122.4917	0.010
634	K	8.6926	116.2462	0.015
635	K	7.7758	123.5364	0.011
636	V	7.9903	119.4339	0.013
637	E	8.5131	120.4238	0.003
638	G	8.1819	107.2789	0.015
639	D	8.2319	122.6167	0.026
640	M	8.0688	121.2582	0.020
641	Y	9.3064	121.6205	0.018
642	E	8.0492	114.3607	0.022
643	S	7.9781	112.7937	0.010
644	A	8.2124	124.3100	0.011
645	N	9.1589	117.3304	0.008
646	S	6.8921	111.7601	0.019
647	R	8.7586	122.0256	0.013
648	D	8.1187	116.3875	0.034
649	E	7.7743	121.1550	0.003
650	Y	7.3436	121.5362	0.011
651	Y	8.1610	116.0436	0.008
652	H	8.1717	117.2565	0.021
653	L	8.9908	120.6601	0.014
654	L	7.7743	120.8133	0.003
655	A	8.1849	120.9531	0.052
656	E	8.3496	115.9500	0.002
657	K	7.8360	120.5007	0.037
658	I	8.3006	119.9894	0.022

Table A12 (Continued)

659	Y	8.0012	119.4550	0.012
660	K	8.2779	122.5750	0.013
661	I	8.3722	120.9872	0.012
662	Q	8.2884	120.8540	0.008
663	K	8.2044	120.0279	0.018
664	E	7.9891	120.3268	0.023
665	L	8.2046	120.0288	0.068
666	R	7.9947	120.2360	0.016
669	R	8.2386	119.2696	0.030
670	R	7.9196	119.1600	0.014
671	S	8.0055	115.4747	0.012
672	R	8.0116	122.3802	0.008
673	L	7.7453	128.3677	0.010

Table A13: Chemical shifts of ^{1}H-^{15}N KIX bound to MLL mimic 8

Residue number	AA	1H	15N	Chemical shift change from apo
587	G	8.5612	111.3778	0.000
588	V	7.9698	119.6220	0.014
589	R	8.5604	125.7935	0.007
590	K	8.4457	122.1271	0.015
591	G	8.4491	110.5003	0.009
592	W	7.8457	118.4009	0.007
593	H	6.9443	120.5434	0.027
594	E	7.5804	116.4267	0.014
595	H	7.3648	114.7942	0.185

Table A13 (Continued)

596	V	7.5505	120.7344	0.006
597	T	7.5328	115.2605	0.038
599	D	8.4589	117.8370	0.003
600	L	7.6454	122.9277	0.009
601	R	7.8139	118.3124	0.012
602	S	8.4348	112.7557	0.008
603	H	8.3209	120.2773	0.117
608	L	7.9081	121.2862	0.081
609	V	8.1472	118.0189	0.095
610	Q	8.2469	116.8184	
611	A	7.5079	119.6965	0.028
612	I	7.5620	114.5515	0.015
615	T	7.9651	114.6403	0.035
617	D	7.9698	120.4631	0.011
619	A	8.4027	121.4844	0.059
620	A	7.7184	121.2429	0.054
621	L	7.5940	117.2077	0.113
622	K	7.7327	118.5118	0.101
623	D	8.0480	120.8907	0.083
625	R	8.0165	119.3934	0.077
626	M	8.3534	119.1352	0.144
627	E	8.3035	118.2980	0.112
628	N	8.2476	118.1368	0.042
629	L	7.9731	122.4984	0.137
630	V	8.3693	118.8515	0.089
631	A	7.8612	120.8451	0.026

Table A13 (Continued)

632	Y	8.0262	120.0211	0.157
633	A	8.6558	122.5755	0.110
634	K	8.7692	116.6914	0.110
635	K	7.7561	123.2633	0.042
636	V	7.9752	118.9822	0.056
637	E	8.5848	120.6650	0.076
638	G	8.2478	107.4262	0.082
639	D	8.2129	122.7239	0.027
640	M	8.0568	121.5216	0.037
641	Y	9.3166	121.4940	0.028
642	E	8.0996	114.4922	0.074
643	S	8.0020	112.8659	0.017
644	A	8.2099	124.3824	0.016
645	N	9.1530	117.3865	0.019
646	S	6.8881	111.6713	0.030
647	R	8.7366	122.0067	0.035
648	D	8.1037	116.0429	0.029
649	E	7.7689	121.1397	0.009
650	Y	7.3203	121.4201	0.040
651	Y	8.2178	116.0467	0.050
652	H	8.1607	117.2046	0.033
653	L	8.9289	120.7170	0.076
654	L	7.7681	120.8140	0.010
655	A	8.1637	120.8190	0.039
656	E	8.3500	115.9381	0.001
657	K	7.7292	120.4493	0.146

Table A13 (Continued)

658	I	8.2887	119.9718	0.010
659	Y	8.0626	119.4005	0.076
660	K	8.2880	122.6022	0.010
661	I	8.4041	121.0793	0.022
662	Q	8.2853	120.7438	0.016
663	K	8.2162	119.6865	0.066
664	E	7.9592	120.4046	0.048
665	L	8.2286	119.1912	0.166
666	R	8.0314	119.9725	0.041
669	R	8.2224	118.9085	0.075
670	R	7.8699	119.1217	0.065
671	S	8.0029	115.4214	0.020
672	R	8.0245	122.4225	0.018
673	L	7.7575	128.4079	0.021

Table A14: Chemical shifts of ^{1}H-^{15}N KIX bound to CREB mimic 78a

Residue number	AA	1H	15N	Chemical shift change from apo
587	G	8.560	111.374	0.0003
588	V	7.967	119.529	0.0033
589	R	8.564	125.808	0.0028
590	K	8.440	122.111	0.0088
591	G	8.443	110.472	0.0061
592	W	7.848	118.390	0.0115
593	H	6.927	120.573	0.0180
594	E	7.571	116.447	0.0104
595	H	7.540	114.570	0.0050

Table A14 (Continued)

596	V	7.546	120.759	0.0010
597	T	7.503	115.194	0.0086
599	D	8.454	117.813	0.0019
600	L	7.635	122.955	0.0014
601	R	7.808	118.391	0.0033
602	S	8.436	112.806	0.0030
603	H	8.231	120.668	0.0170
608	L	7.983	121.580	0.0021
609	V	8.160	118.583	0.0145
610	Q	8.115	116.316	0.0074
611	A	7.534	119.562	0.0033
612	I	7.530	114.646	0.0024
615	T	7.959	114.916	0.0049
617	D	7.964	120.359	0.0064
619	A	8.374	121.809	0.0047
620	A	7.771	121.117	0.0037
621	L	7.644	117.668	0.0228
622	K	7.798	118.791	0.0194
623	D	8.110	120.531	0.0037
625	R	8.084	119.419	0.0077
626	M	8.224	119.441	0.0067
627	E	8.296	119.040	0.0057
628	N	8.290	118.100	0.0289
629	L	7.916	123.359	0.0094
630	V	8.333	119.276	0.0198
631	A	7.875	120.690	0.0014

Table A14 (Continued)

632	Y	7.884	120.299	0.0097
633	A	8.538	122.500	0.0058
634	K	8.690	116.145	0.0017
635	K	7.771	123.558	0.0037
636	V	7.993	119.375	0.0032
637	E	8.511	120.428	0.0017
638	G	8.171	107.281	0.0062
639	D	8.218	122.553	0.0071
640	M	8.056	121.267	0.0015
641	Y	9.295	121.586	0.0033
642	E	8.040	114.266	0.0116
643	S	7.986	112.788	0.0015
644	A	8.221	124.318	0.0046
645	N	9.168	117.321	0.0044
646	S	6.888	111.815	0.0079
647	R	8.768	122.045	0.0036
648	D	8.124	116.222	0.0010
649	E	7.792	121.158	0.0053
650	Y	7.353	121.518	0.0015
651	Y	8.167	116.075	0.0029
652	H	8.184	117.287	0.0034
653	L	8.999	120.660	0.0034
654	L	7.775	120.797	0.0167
655	A	8.145	120.978	0.0059
656	E	8.349	115.938	0.0005
657	K	7.853	120.655	0.0064

Table A14 (Continued)

658	I	8.281	119.978	0.0028
659	Y	7.962	119.296	0.0064
660	K	8.283	122.539	0.0012
661	I	8.381	120.995	0.0040
662	Q	8.287	120.868	0.0046
663	K	8.198	120.079	0.0066
664	E	8.003	120.414	0.0046
665	L	8.158	120.131	0.0115
669	R	8.005	120.160	0.0138
670	R	8.269	119.305	0.0023
671	S	7.930	119.190	0.0028
672	R	8.013	115.505	0.0015
673	L	8.009	122.388	0.0058

Table A15: Chemical shifts of ^{1}H-^{15}N KIX bound to CREB mimic 79b

Residue number	AA	1H	15N	Chemical shift change from apo
587	G	8.561	111.370	0.0062
588	V	7.968	119.536	0.0022
589	R	8.560	125.787	0.0130
590	K	8.450	122.124	0.0109
591	G	8.452	110.510	0.0090
592	W	7.862	118.497	0.0242
593	H	6.955	120.484	0.0387
594	E	7.589	116.501	0.0206
595	H	7.542	114.495	0.0139

Table A15 (Continued)

596	V	7.545	120.783	0.0085
597	T	7.511	115.243	0.0129
599	D	8.454	117.828	0.0088
600	L	7.636	122.996	0.0075
601	R	7.807	118.411	0.0111
602	S	8.439	112.810	0.0012
603	H	8.224	120.577	0.0308
608	L	7.986	121.576	0.0030
609	V	8.148	118.544	0.0319
610	Q	8.112	116.163	0.0127
611	A	7.536	119.561	0.0014
612	I	7.535	114.581	0.0162
615	T	7.945	114.841	0.0178
617	D	7.963	120.416	0.0137
619	A	8.374	121.812	0.0071
620	A	7.773	121.128	0.0027
621	L	7.641	117.644	0.0316
622	K	7.794	118.780	0.0285
623	D	8.114	120.549	0.0054
625	R	8.080	119.401	0.0168
626	M	8.222	119.422	0.0052
627	E	8.299	118.995	0.0134
628	N	8.281	118.000	0.0109
629	L	7.914	123.372	0.0161
630	V	8.334	119.117	0.0407
631	A	7.854	120.640	0.0298

Table A15 (Continued)

632	Y	7.892	120.288	0.0119
633	A	8.541	122.497	0.0090
634	K	8.686	116.140	0.0090
635	K	7.774	123.579	0.0038
636	V	7.998	119.390	0.0030
637	E	8.509	120.419	0.0113
638	G	8.171	107.283	0.0013
639	D	8.223	122.575	0.0073
640	M	8.062	121.265	0.0010
641	Y	9.302	121.567	0.0054
642	E	8.045	114.378	0.0230
643	S	7.987	112.810	0.0069
644	A	8.215	124.301	0.0167
645	N	9.164	117.310	0.0128
646	S	6.882	111.799	0.0142
647	R	8.763	122.049	0.0122
648	D	8.115	116.276	0.0147
649	E	7.774	121.139	0.0014
650	Y	7.355	121.540	0.0070
651	Y	8.169	116.065	0.0015
652	H	8.184	117.277	0.0101
653	L	9.009	120.742	0.0079
654	L	7.780	120.877	0.0040
655	A	8.151	121.072	0.0110
656	E	8.347	115.931	0.0082
657	K	7.873	120.683	0.0082

Table A15 (Continued)

658	I	8.288	119.978	0.0030
659	Y	7.970	119.421	0.0042
660	K	8.285	122.553	0.0075
661	I	8.381	120.995	0.0093
662	Q	8.285	120.850	0.0042
663	K	8.198	120.074	0.0091
664	E	8.004	120.410	0.0089
665	L	8.170	120.130	0.0161
669	R	8.006	120.263	0.0072
670	R	8.264	119.295	0.0123
671	S	7.925	119.161	0.0145
672	R	8.010	115.483	0.0127
673	L	8.007	122.362	0.0069

Table A16: Chemical shifts of ^{1}H-^{15}N KIX bound to CREB mimic 3_30b

Residue number	AA	1H	15N	Chemical shift change from apo
587	G	8.573	111.3032	0.0001
588	V	7.975	119.6206	0.0021
589	R	8.57365	125.7791	0.0061
590	K	8.42426	122.1126	0.0191
591	G	8.44293	110.3874	0.0029
592	W	7.85054	118.2444	0.0200
593	H	6.9059	120.58	0.0132
594	E	7.5671	116.35	0.0133
595	H	7.54932	114.4552	0.0084

Table A16 (Continued)

596	V	7.5732	120.72	0.0098
597	T	7.50918	115.1017	0.0203
599	D	8.48659	117.7562	0.0174
600	L	7.6402	122.74	0.0110
601	R	7.83195	118.339	0.0035
602	S	8.44209	112.7308	0.0074
603	H	8.25456	120.596	0.0229
608	L	8.02342	121.4871	0.0313
609	V	8.19912	118.5564	0.0086
610	Q	8.122435	116.272	0.0052
611	A	7.56292	119.5127	0.0252
612	I	7.5857	114.64	0.0254
615	T	7.99236	114.8304	0.0152
617	D	7.994	120.2983	0.0124
619	A	8.394918	121.8075	0.0149
620	A	7.76201	121.0739	0.0229
621	L	7.68083	117.7575	0.0122
622	K	7.8394	118.83	0.0086
623	D	8.1289	120.4141	0.0126
625	R	8.1089	119.3599	0.0039
626	M	8.23937	119.3373	0.0216
627	E	8.31368	118.95	0.0097
628	N	8.3063	117.8342	0.0214
629	L	7.94427	123.282	0.0106
630	V	8.34735	119.3117	0.0286
631	A	7.894	120.6953	0.0227

Table A16 (Continued)

632	Y	7.901	120.2228	0.0136
633	A	8.577	122.4	0.0150
634	K	8.715	116.06	0.0148
635	K	7.7838	123.41	0.0129
636	V	7.995	119.5214	0.0299
637	E	8.546	120.43	0.0150
638	G	8.1918	107.17	0.0173
639	D	8.216	122.3974	0.0112
640	M	8.076	121.1847	0.0188
641	Y	9.309	121.4984	0.0186
642	E	8.050	114.0129	0.0169
643	S	8.010	112.67	0.0112
644	A	8.254	124.32	0.0068
645	N	9.192	117.2944	0.0169
646	S	6.893	111.7989	0.0049
647	R	8.762	121.9521	0.0131
648	D	8.148	116.052	0.0157
649	E	7.798	121.0523	0.0176
650	Y	7.377	121.36	0.0130
651	Y	8.210	117.1786	0.0091
652	H	9.011	120.5422	0.0056
653	L	7.823	121.0665	0.0240
654	L	8.142	121.1166	0.0346
655	A	8.359	115.8386	0.0095
656	E	7.902	120.5181	0.0117
657	K	8.297	119.8382	0.0146

Table A16 (Continued)

658	I	7.960	119.515	0.0211
659	Y	8.166	115.9774	0.0093
660	K	8.293	122.427	0.0050
661	I	8.399	120.9302	0.0085
662	Q	8.306	120.7544	0.0170
663	K	8.210	119.9817	0.0145
664	E	8.027	120.294	0.0098
665	L	8.180	120.0531	0.0269
669	R	8.026	120.1422	0.0307
670	R	8.290	119.2411	0.0130
671	S	7.959	119.2207	0.0105
672	R	8.043	115.49	0.0089
673	L	8.047	122.3389	0.0254

Note: This CREB mimic is also called (also called S2-18).

Table A17: : Chemical shifts of ^{1}H-^{15}N KIX bound to CREB mimic 2_133b

Residue number	SS	1H	15N	Chemical shift change from apo
587	G	8.574	111.303	0.0010
588	V	7.979	119.670	0.0089
589	R	8.576	125.762	0.0058
590	K	8.449	122.085	0.0130
591	G	8.453	110.408	0.0128
592	W	7.850	118.247	0.0203
593	H	6.925	120.507	0.0344
594	E	7.574	116.376	0.0210
595	H	7.550	114.451	0.0091

Table A17 (Continued)

596	V	7.565	120.698	0.0017
597	T	7.488	115.048	0.0024
599	D	8.471	117.738	0.0020
600	L	7.635	122.819	0.0028
601	R	7.832	118.350	0.0044
602	S	8.437	112.694	0.0013
603	H	8.246	120.513	0.0322
608	L	8.006	121.476	0.0140
609	V	8.183	118.514	0.0122
610	Q	8.118	116.123	0.0215
611	A	7.550	119.515	0.0132
612	I	7.568	114.671	0.0178
615	T	7.985	114.841	0.0110
617	D	7.980	120.292	0.0021
619	A	8.384	121.788	0.0041
620	A	7.743	121.054	0.0041
621	L	7.658	117.638	0.0314
622	K	7.814	118.749	0.0216
623	D	8.111	120.394	0.0054
625	R	8.108	119.367	0.0048
626	M	8.221	119.344	0.0045
627	E	8.301	118.885	0.0194
628	N	8.282	117.775	0.0060
629	L	7.928	123.289	0.0094
630	V	8.336	119.247	0.0339
631	A	7.881	120.787	0.0225

Table A17 (Continued)

632	Y	7.899	120.213	0.0110
633	A	8.566	122.384	0.0057
634	K	8.696	116.029	0.0049
635	K	7.779	123.452	0.0095
636	V	7.985	119.345	0.0091
637	E	8.529	120.420	0.0049
638	G	8.174	107.149	0.0015
639	D	8.212	122.416	0.0088
640	M	8.076	121.182	0.0186
641	Y	9.300	121.466	0.0099
642	E	8.040	114.083	0.0171
643	S	8.004	112.679	0.0061
644	A	8.246	124.310	0.0014
645	N	9.174	117.229	0.0043
646	S	6.902	111.771	0.0074
647	R	8.773	121.918	0.0016
648	D	8.136	116.092	0.0062
649	E	7.786	121.043	0.0057
650	Y	7.366	121.386	0.0023
651	Y	8.165	115.938	0.0045
652	H	8.200	117.164	0.0029
653	L	9.002	120.557	0.0055
654	L	7.804	121.056	0.0042
655	A	8.129	121.110	0.0212
656	E	8.368	115.871	0.0009
657	K	7.886	120.587	0.0070

Table A17 (Continued)

658	I	8.283	119.849	0.0026
659	Y	7.985	119.515	0.0045
660	K	8.292	122.394	0.0045
661	I	8.388	120.928	0.0039
662	Q	8.296	120.724	0.0067
663	K	8.192	119.919	0.0142
664	E	8.012	120.274	0.0166
665	L	8.165	120.007	0.0176
669	R	8.017	120.120	0.0344
670	R	8.277	119.229	0.0005
671	S	7.961	119.223	0.0120
672	R	8.044	115.507	0.0108
673	L	8.039	122.336	0.0176

APPENDIX B – SUPPLEMENTARY FIGURES

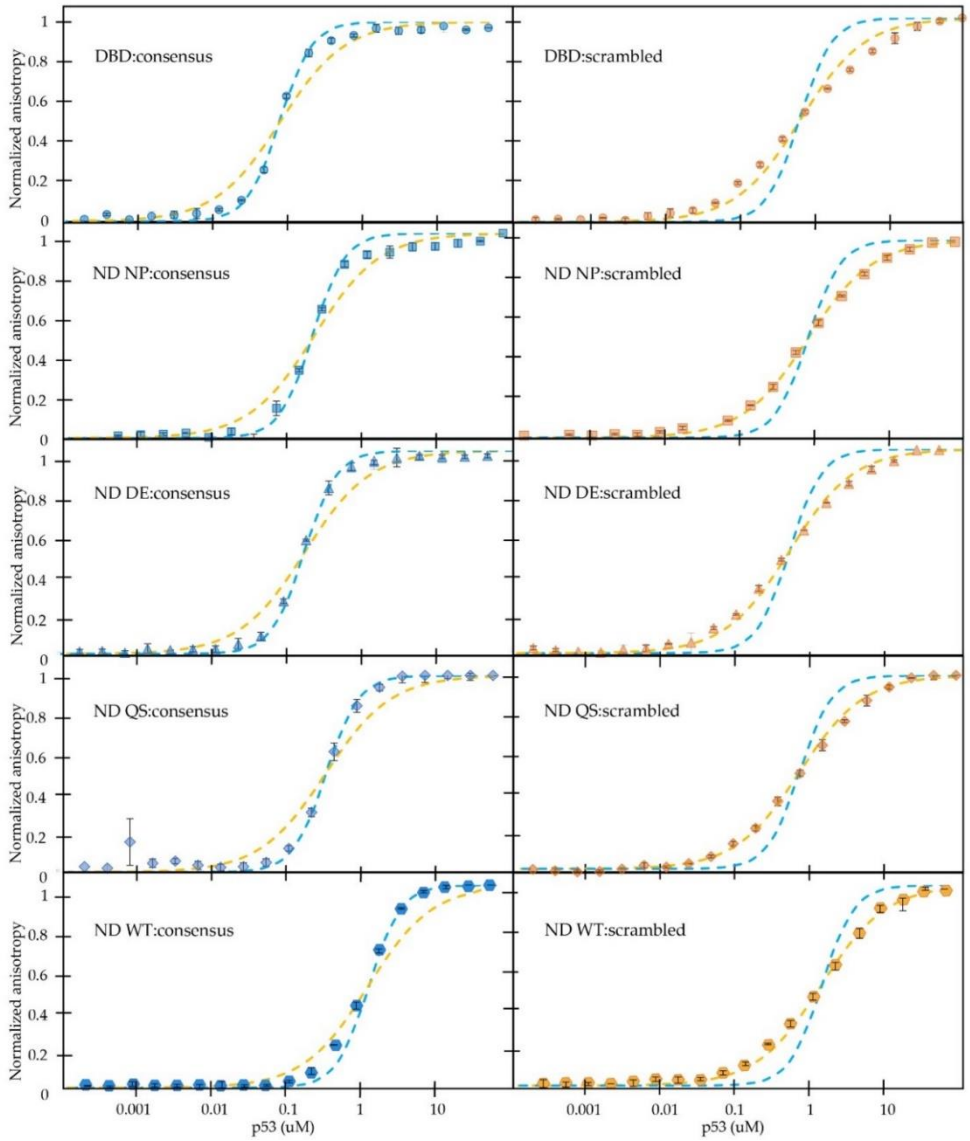


Figure A1: Hill coefficients of TAD2 mutants. Fluorescence anisotropy data points are graphed with fit lines that use a Hill coefficient of one (orange dotted line) or two (blue dotted line). In all cases, data for binding consensus DNA matches a fit line with a Hill coefficient of two, and binding for scrambled DNA matches a fit line with a Hill coefficient of one. This is also true for PRR mutants and across all ionic strengths.

Primates		Uniprot or NCBI ID
H sapiens	----------MEEPQSDPSVEPPLSQETFSDLWKLLPE----NNVLSPLPSQAM----	P04637
P troglodytes	----------MEEPQSDPSVEPPLSQETFSDLWKLLPE----NNVLSPLPSQAM----	H2QC53
G gorilla	MEPCISSQTAFRVTAMEEPQSDPSVEPPLSQETFSDLWKLLPE----NNVLSPLPSQAM----	A0A2I2Y7Z8
P abelii	----------MEEPQSDPSVEPPLSQETFSDLWKLLPE----NNVLSPLPSQAV----	XP_002827020
N leucogynes	----------MEEPQSDPSVEPPLSQETFSDLWKLLPE----NNVLSPLPSQAM----	G1RF61
P 203nubis	----------MEEPQSDPSIEPPLSQETFSDLWKLLPE----NNVLSPLPSQAV	A0A096P2R8
M fuscata	----------MEEPQSDPSIEPPLSQETFSDLWKLLPE----NNVLSPLPSQAV----	P61260
C aethiops	----------MEEPQSDPSIEPPLSQETFSDLWKLLPE----NNVLSPLPSQAV----	P13481
C jacchus	----------MEEPQSDLSIEPPLSCGVWWV---PGPLTALFTHLQSS SLSQPV---	A0A2R8MYD9
C syrichta	----------MEEPQSDLSIE-PLSQETFSDLWKLLPE----NNVLSPSLSPPV----	A0A1U7U5H4
O garnettii	----------MEETQSDLTIEPPLSQETFSDLWKLLPE----NNVLSSSLSPPV----	H0XGB0
Rodents		
T belangeri	----------MEEPQSDPSVEPPLSQETFSDLWKLLPE----NNVLSPLPSQAM----	Q9TTA1
O cuniculus	----------MEESQSDLSLEPPLSQETFSDLWKLLPE----NNLLTTSLNPPV----	Q95330
M musculus	----------MTAMEESQSDISLELPLSQETFSGLWKLLPP----EDILPSPHCM------	P02340
R norvegicus	----------MEDSQSDMSIELPLSQETFSCLWKLLPP----DDILPTTATGSPNSM-	P10361
I tridecemlineatus	----------MEEPQSDLSIEPPLSQETFSDLWNLLPE----NNVLSPVLSPPM----	I3N5N2
M unguiculatus	----------MEEPQSDLSIEPPLSQETFSDLWKLLPP----KNLLSALEPM------	Q920Y0
N galili	----------MEEQQSDLSIEPPLSQETFSDLWKLLPE---QNNVLSTPLSPNSM---	A0A0K1TP12
M marmota	----------MEEPQSDLSIEPPLSQETFSDLWNLLPE----NNVLSPVLSPPM----	A0A8CA0W6
C griseus	----------MEEPQSDLSIELPLSQETFSDLWKLLPP----NNVLSTLPSSDSI---	O09185
C canadensis	----------MEESQSELSIEPPLSQETFSDFWKLLPE----NNLLSTSSSPPM----	A0A250YHC8
C lanigera	----------MEEPQSDLSIEPPLSQETFSDLWKLLPE----NNVLSNSLPSPM----	A0A8C2YMX2
Carnivores		
L canadensis	----------MQEPPLELTIEPPLSQETFSELWNLLPE----NNVLSSELSSAM---	A0A667I4D8
P leon	----------MQEPPLELTIEPPLSQETFSELWNLLPE----NNVLSSELSSAM---	A0A8C8Y413
F sylvestris	----------MQEPPLELTIEPPLSQETFSELWNLLPE----NNVLSSELSSAM----	P41685
S suricatta	----------MQEPPLELNIEPPLSQETFSELWNLLPE----NNVLSAELSSGV----	XP_029783014
Z californianus	----------MQDPQSELTIDPPLSQETFSELWNLLPE----NNVLSAELSPAV----	A0A6J2FP67
A melanoleuca	----------MQDPQSELTIDPPLSQETFSELWNLLPE----NNVLSSELSPGV----	D2HPX0
U maritimus	----------MQDPQSELTIDPPLSQETFSELWNLLPE----NNVLSSELSPGV----	A0A384BVC2
V vulpes	----------MQEPQSELNIDPPLSQETFSELWNLLPE----NNVLSSELCPAV----	A0A3Q7TIN0
C lupus fam.	----------MEESQSELNIDPPLSQETFSELWNLLPE----NNVLSSELCPAV----	Q29537
M putorius	----------MQDPQSELTIDPPLSQETFSELWNLLPE----NNVLSPAV--------	M3YC88
E lutra kenyoni	----------MQDPQSELTIDPPLSQETFSELWNLLPE----NNVLSSELSPAV----	A0A2Y9L4Q2

Certartiodactyl

O orca	----------MEESQAELGVEPPLSQETFSDLWKLLPE----NNLLSSELSPAV----	XP_004266944
T truncata	----------MEELQAELGVEPPLSQETFSDLWKLLPE----NNLLSSELSPAV----	A0A2U4C2U9
L obliquidens	--MCPLGHTASQGAVMEELQAELGVEPPLSQETFSDLWKLLPE----NNLLSSELSPAV----	XP_026986729
D leucas	----------MEESQAELGVEPPLSQETFSDLWKLLPE----NNLLSSELSPAV----	Q8SPZ3
B mysticetus	----------MEESQAELGVEPPLSQETFSDLWKLLPE----NNLLSSELSPAV----	QZI95188
B davidsoni	----------MEESQAELGVEPPLSQESFSDLWKLLPE----NNLLSSELSPAV----	A0A383YUA7
S scrofa	----------MEESQSELGVEPPLSQETFSDLWKLLPE----NNLLSSELSLAAV---	Q9TUB2
B bubalis	----------MEESQAELNVEPPLSQETFSDLWNLLPE----NNLLSSELSAPV----	F6MDM8
O virginiana	----MKGHTTFQGATMEESQAELGMEPPLSQETFSDLWNLLPE----NNLLSSELSSPV----	A0A6J0VGP1
O aries	----------MEESQAELGVEPPLSQETFSDLWNLLPE----NNLLSSELSAPV----	P51664
C hircus	----------MEESQAELGVEPPLSQETFSDLWNLLPE----NNLLSSELSAPV----	A0A452G0A3
V pacos	----------MEESQSELGVEPPLSQETFSDLWKLLPE----NNLLSSEFSPAV----	A0A6I9IU13
B taurus	----------MEESQAELNVEPPLSQETFSDLWNLLPE----NNLLSSELSAPV----	P67939
E caballus	----------MEETQTELGIEPPLSQETFSDLWKLLPE----NNVLSPDLSPAV----	E9NME8

Xenarthra/Afrotheria

D novemcinctus	----------MEEPPSDLSIEAPLSQETFSDLWKLLPQ----NNVLSPSLSTM-----	XP_023447566
C didactylus	----------MEEPQSDLSIEPPLSQETFSDLWKLLPE----NNVLSPSPSAV-----	XP_037665509
T manatus	----------MEEPQSDLSTELPLSQETFSYLWELLPE----KPVLSPSLPPAVEVVE	A0A2Y9R5B9
O afer afer	----------MEEPQSDLCTDLPLSQETFSYLWELLPE----NPVLSPSIPPAVEVA-	A0A8B7AVU1

Marsupials

S harrisii	----------MEESLSD-LEPPLSQETFSDLWKLQWN----N-LP----------	G3WS63
T vulpecula	----------MEESLSD-LEPPLSQETFSDIWNLQWN----N-LP----------	XP_036622872
P cinereus	----------MEESLSD-LEPPLSQETFSDIWNLQWN----N-MP----------	A0A6P5IIX8
V ursinus	----------MEESLSD-LEPPLSQETFSDIWNLQWN----N-MP----------	A0A4X2JX83
N eugenii	----------MEESLSD-LEPPLSQETFPDLWNLQWN----N-L----------	A0A1W5W828

Birds

G gallus	----------MAEEMEPPLEPTEVFMDLWSMLPYSMQQLPL----------	NP_990595
A platyrhynchos	----------MAEELEPPLEPPEIFLELWNMLPDNMHSLSP----------	XP_038027745
C ustulatus	----------MAEDLEPPLEG-EGFLDWWKTLPDNIGSI----------	XP_032940146.1
C canorus	----------MAEELEPPLGG-EGFLDHWNMLPDNINSF----------	XP_053908747.1
A gentilis	----------MA----------	XP_049648596.1

Figure A2: Multiple sequence alignment of TAD1. MSA of mammals grouped into major orders and ordered by their evolutionary distance from humans; birds are shown at bottom.

Primates		Uniprot or NCBI ID
H sapiens	---DDLMLSPDD-------IEQWFTEDPGP-------	P04637
P troglodytes	---DDLMLSPDD-------IEQWFTEDPGP-------	H2QC53
G gorilla	---DDLMLSPDD-------IEQWFTEDPGP-------	A0A2I2Y7Z8
P abelii	---DDLLLSPDD-------IAQWFIEDPGP-------	XP_002827020
N leucogynes	---DDLMLSPED-------IAQWFTEDPGP-------	G1RF61
P 205nubis	---DDLMLSPDD-------LAQWLTEDPGP-------	A0A096P2R8
M fuscata	---DDLMLSPDD-------LAQWLTEDPGP-------	P61260
C aethiops	---DDLMLSPDD-------LAQWLTEDPGP-------	P13481
C jacchus	---DDLMLSPDD-----IDIAQWLSQDPVP-------	A0A2R8MYD9
C syrichta	---DDLILSTED-------IANWFSEGP---------	A0A1U7U5H4
O garnettii	---DDLMLSPD-------IVNWFDEGP--------	H0XGB0
Rodents		
T belangeri	---DDLMLSPDD-------IEQWFTEDPGP-------	Q9TTA1
O cuniculus	---DDL-LSAED-------VANWLNEDP---------	Q95330
M musculus	---DDLLL-PQD-------VEEFFEGPSEALRVSGAP	P02340
R norvegicus	---EDLFL-PQD-------VAELLEGPEEALQVS-AP	P10361
I tridecemlineatus	---DDLLLSSED-------VENWFDKGP---------	I3N5N2
M unguiculatus	---EDLLLPQD--------VTSWLGDA----------	Q920Y0
N galili	---EDLLLSPED-------VANWLDD-P---------	A0A0K1TP12
M marmota	---DDLLLSSED-------VENWFDKGP---------	A0A8C6A0W6
C griseus	---EELFLSEN--------VAGWLEDPG---------	O09185
C canadensis	---DDLLFPDE--------VANWLEGQ----------	A0A250YHC8
C lanigera	---DDLLLCPED-------VVNWLEENP---------	A0A8C2YMX2
Carnivores		
L canadensis	---NELPLS-ED-------VANWLDEAP---------	A0A667I4D8
P leon	---NELPLS-ED-------VANWLDEAP---------	A0A8C8Y413
F sylvestris	---NELPLS-ED-------VANWLDEAP---------	P41685
S suricatta	---NELPLS-ED-------VTHWLDEAP---------	XP_029783014
Z californianus	---DELLLS-EG-------VATWLDAGS---------	A0A6J2FP67
A melanoleuca	---DELLLS-EG-------VVNWMDEGS---------	D2HPX0
U maritimus	---DELLLS-EG-------VVNWMDEGS---------	A0A384BVC2
V vulpes	---DELLL-PES-------VVNWLDEDS---------	A0A3Q7TIN0
C lupus fam.	---DELLL-PES-------VVNWLDEDS---------	Q29537
M putorius	---DELL-S-E--------GVNWLGEGS---------	M3YC88
E lutra kenyoni	---DELL-S-E--------GVNWLGEGS---------	A0A2Y9L4Q2

Certartiodactyl		
O orca	---DDLLLSPED-------VANWLDERP---------	XP_004266944
T truncata	---DDLLLSPED-------VANWLDERP---------	A0A2U4C2U9
L obliquidens	---DDLLLSPED-------VANWLDERP---------	XP_026986729
D leucas	---DDLLLSPED-------VANWLDERP---------	Q8SPZ3
B mysticetus	---DDLLLSPED-------VANWLDERP---------	QZI95188
B davidsoni	---DDLLLSPED-------VANWLDERP---------	A0A383YUA7
S scrofa	---NDLLLS-P--------VTNWLDENP---------	Q9TUB2
B bubalis	---DDLLPYT-D-------VATWLDECP---------	F6MDM8
O virginiana	SDLLSPVMTSED-------VANWLDECP---------	A0A6J0VGP1
O aries	---DDLLPYSED-------VVTWLDECP---------	P51664
C hircus	---DDLLPYSED-------VVTWLDECP---------	A0A452G0A3
V pacos	---DELLLSPE--------VTNWLDENP---------	A0A6I9IU13
B taurus	---DDLLPYT-D-------VATWLDECP---------	P67939
E caballus	---NNLLLSPD--------VVNWLDEGP---------	E9NME8
Xenarthra/Afrotheria		
D novemcinctus	---EDLLLS-DD-------VTSWFEGQ----------	XP_023447566
C didactylus	---DDLFL-PED-------VASWLNNP----------	XP_037665509
T manatus	---DDLLLT-ED-------AATWLESQVGA-------	A0A2Y9R5B9
O afer afer	---DNLLF-PEV-------TANWLENEVGT-------	A0A8B7AVU1
Marsupials		
S harrisii	---DDLSN-IDDIQFASPDSINWLENEEEN-------	G3WS63
T vulpecula	---DDLSN-IDNIQFASPDNMSWLGNEEES-------	XP_036622872
P cinereus	---DDLSN-IDDIQFASPDDINWLGNEEEN-------	A0A6P5IIX8
V ursinus	---DDLSN-IDDIQFASADDINWLRNEEEN-------	A0A4X2JX83
N eugenii	---DDLSN-IDDKYFAPPDSGTWLGNEEES-------	A0A1W5W828
Birds		
G gallus	---PE----------D---HSNWQELSPLE-------	NP_990595
A platyrhynchos	---PD----------D---PLAVQDLCPLE-------	XP_038027745
C ustulatus	---PE----------E---TLEWEDLTTLG-------	XP_032940146.1
C canorus	---LD----------D---PPEWQELSPLG-------	XP_053908747.1
A gentilis	--------------G---AEPPSGSPCGPP-------	XP_049648596.1

Figure A3: Multiple sequence alignment of TAD2. MSA of mammals grouped into major orders and ordered by their evolutionary distance from humans; birds are shown at bottom.

Primates		Uniprot or NCBI ID
H sapiens	DEAPRM----------PEAAPPVAPAPAAPTPAAPAPAPSWPL	P04637
P troglodytes	DEAPRM----------PEAAPPVAPAPAAPTPAAPAPAPSWPL	H2QC53
G gorilla	DEAPRM----------PEAAPPVAPAPAAPTPAAPAPAPSWPL	A0A2I2Y7Z8
P abelii	DEAPRM----------SEAASPVGPAPAAPIPAAPAPAPSWPL	XP_002827020
N leucogynes	HEAPRM----------SEAAPPMAPASAAPTLAAPAPSPSWPL	G1RF61
P 207nubis	DEAPRM----------SEATPPMAPTPAAPTPAAPAPAPSWPL	A0A096P2R8
M fuscata	DEAPRM----------SEAAPPMAPTPAAPTPAAPAPAPSWPL	P61260
C aethiops	DEAPRM----------SEAAPHMAPTPAAPTPAAPAPAPSWPL	P13481
C jacchus	DEAPTV----------SEAPPAMAQAPAAPTLVAPTPAPSWPL	A0A2R8MYD9
C syrichta	DEALR----------TAPAPVAPTPAASTQAAPAPGTPWPL	A0A1U7U5H4
O garnettii	DEALRT----------LEDPAPVASTTAALTPAASAPVTGWPL	H0XGB0
Rodents		
T belangeri	DEAPRM----------PEAAPPVAPAPAAPTPAAPAPAPSWPL	Q9TTA1
O cuniculus	EEGLRV----------PAAPAPEAPAPAAPALAAPAPATSWPL	Q95330
M musculus	----------AAQDPVTETPGPVAPAPATPWPL	P02340
R norvegicus	----------AAQEPGTEAPAPVAPASATPWPL	P10361
I tridecemlineatus	DEALQM----------SAAPAPKAPTPAASTLAAPAPATSWPL	I3N5N2
M unguiculatus	DEALPV----------CTAPAEGPAPEAPAPAAPAPPASWPL	Q920Y0
N galili	DEALQV----------PAAAITGDPVTETSAPVAPPPATPWPL	A0A0K1TP12
M marmota	DEALQM----------SAAPAPKAPTPAASTLAAPSPATSWPL	A0A8C6A0W6
C griseus	EALQG-----SAAAAAPAAPAAEDPVAETPAPVASAPATPWPL	O09185
C canadensis	DETLQILAAPVSKAPAPEVPAPEVPVPEVPAPAAPAPVTSWPL	A0A250YHC8
C lanigera	DEDVQM----------SAAPVPEPPTPAAPAPAAPPPATSWPL	A0A8C2YMX2
Carnivores		
L canadensis	DDASGM----------SAVPAPAAPAPATPAPAISWPL	A0A667I4D8
P leon	DDASGM----------SAVPAPAAPAPATPAPAISWPL	A0A8C8Y413
F sylvestris	DDASGM----------SAVPAPAAPAPATPAPAISWPL	P41685
S suricatta	DDASRM----------SAAPAPAAPAATTPAPAISWPL	XP_029783014
Z californianus	DDAPRM----------PAAPAP-----AGPGPTTSWPL	A0A6J2FP67
A melanoleuca	DDTPRM----------PVAPAP-----AAPGPAISWPL	D2HPX0
U maritimus	DDTPRM----------PVAPAP-----AAPGPAISWPL	A0A384BVC2
V vulpes	DDAPRM----------PATSAP-----TAPGPAPSWPL	A0A3Q7TIN0
C lupus fam.	DDAPRM----------PATSAP-----TAPGPAPSWPL	Q29537
M putorius	NDAPRM----------PATPAP-----AAPGPAPSWPL	M3YC88
E lutra kenyoni	NDAPRM----------PATPAP-----AAPGPAPSWPL	A0A2Y9L4Q2

Certartiodactyl

O orca	DEAPQM----------PEPPAPAAPTPAAPAPATSWPL	XP_004266944
T truncata	DEAPQM----------PEPPAPAAPTPAAPAPATSWPL	A0A2U4C2U9
L obliquidens	DEAPQM----------PEPPAPAAPTPAAPAPATSWPL	XP_026986729
D leucas	DEAPQM----------PEPPAPAAPTPAAPAPATSWPL	Q8SPZ3
B mysticetus	DEAPQM----------PEPPALAAPAPAAPAPATSWPL	QZI95188
B davidsoni	DEAPQM----------PEPTAPAAPAPAAPAPATSWPL	A0A383YUA7
S scrofa	DDASRV----------PAPPAATAPAPAAPAPATSWPL	Q9TUB2
B bubalis	NEVPQM----------PEPSAPAAPPPATPAPATSWPL	F6MDM8
O virginiana	NEAPQM----------AEPLAPAALPPATPTPATSWPL	A0A6J0VGP1
O aries	NEAPQM----------PEPPAQAAL-----APATSWPL	P51664
C hircus	NEAPQM----------PEPPAQAAL-----APATSWPL	A0A452G0A3
V pacos	DEAPRM----------QEPPVP-----TASAPATSWPL	A0A6I9IU13
B taurus	NEAPQM----------PEPSAPAAPPPATPAPATSWPL	P67939
E caballus	NEAPQM----------PEPSAPAAPPPATPAPATSWPL	E9NME8

Xenarthra/Afrotheria

D novemcinctus	DEPLRT----------PEAPAPT-----TPAPAAPTPATPWPL	XP_023447566
C didactylus	EEPVG----------TPAAAALATTPVPAAPAPATPWPL	XP_037665509
T manatus	---QEI----------SAGPAPATPTPVALIPATSWTL	A0A2Y9R5B9
O afer afer	----------QGISEAPTPATSWTLP	A0A8B7AVU1

Marsupials

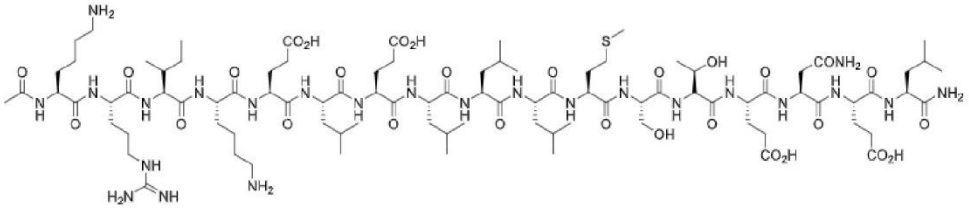
S harrisii	----------PGFQVSLPPVNLPT	G3WS63
T vulpecula	----------PGLPVSAPPVNLPT	XP_036622872
P cinereus	----------PGLQVPAPPVNLPT	A0A6P5IIX8
V ursinus	----------PGFQVPAPPVNLPT	A0A4X2JX83
N eugenii	----------LSVQVPVPPVSLPT	A0A1W5W828

Birds

G gallus	PSD-------PPPPPPPPPLPLAAAAPPPLNPPTPPRAAPS	NP_990595
A platyrhynchos	PS-------EPPPG------PPPSTEPPPAAPPEPPRASPS	XP_038027745
C ustulatus	VS-------PEGP----------PPEPPRQVEPPAPPPA	XP_032940146.1
C canorus	P-----------PPEPPPAVLVPPPPSLPPS	XP_053908747.1
A gentilis	PSD-------PPPPPPPPPLPLAAAAPPPLNPPTPPRAAPS	XP_049648596.1

Figure A4: Multiple sequence alignment of PRR. MSA of mammals grouped into major orders and ordered by their evolutionary distance from humans; birds are shown at bottom.

A) cMyb WT

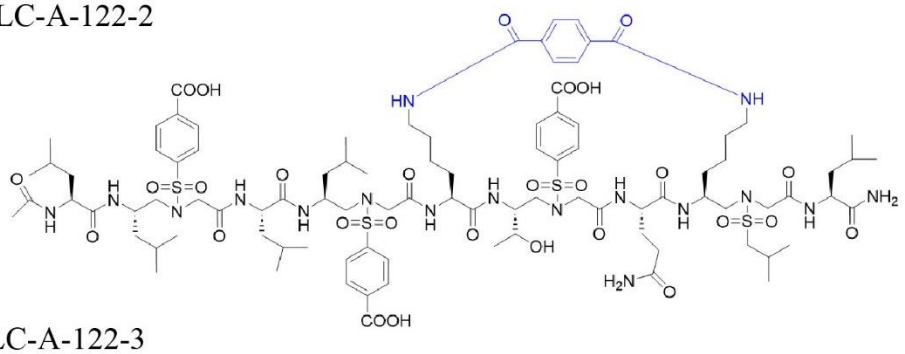


B) cMyb 104-2

C) cMyb 109-1

Figure A5: Synthetic cMyb WT and mimic 104-2 and 109-1 structures. A) cMyb WT, B) cMyb 104-2, and C) cMyb 109-1.

A) LC-A-122-2



B) LC-A-122-3

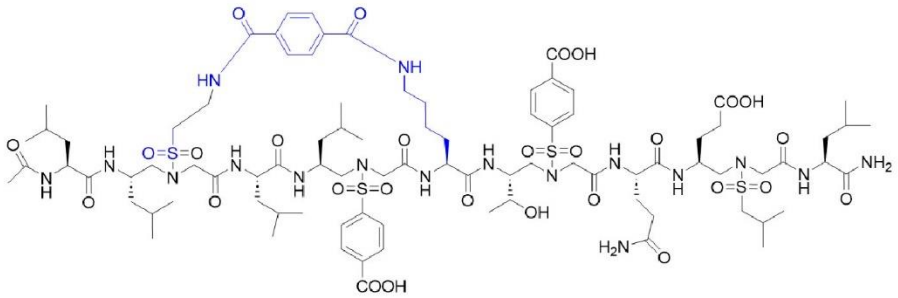


Figure A6: cMyb inhibitors LC-A-122-2 and LC-A-122-3 structures. A) cMyb inhibitor LC-A-122-2 and B) cMyb inhibitor LC-A-122-3.

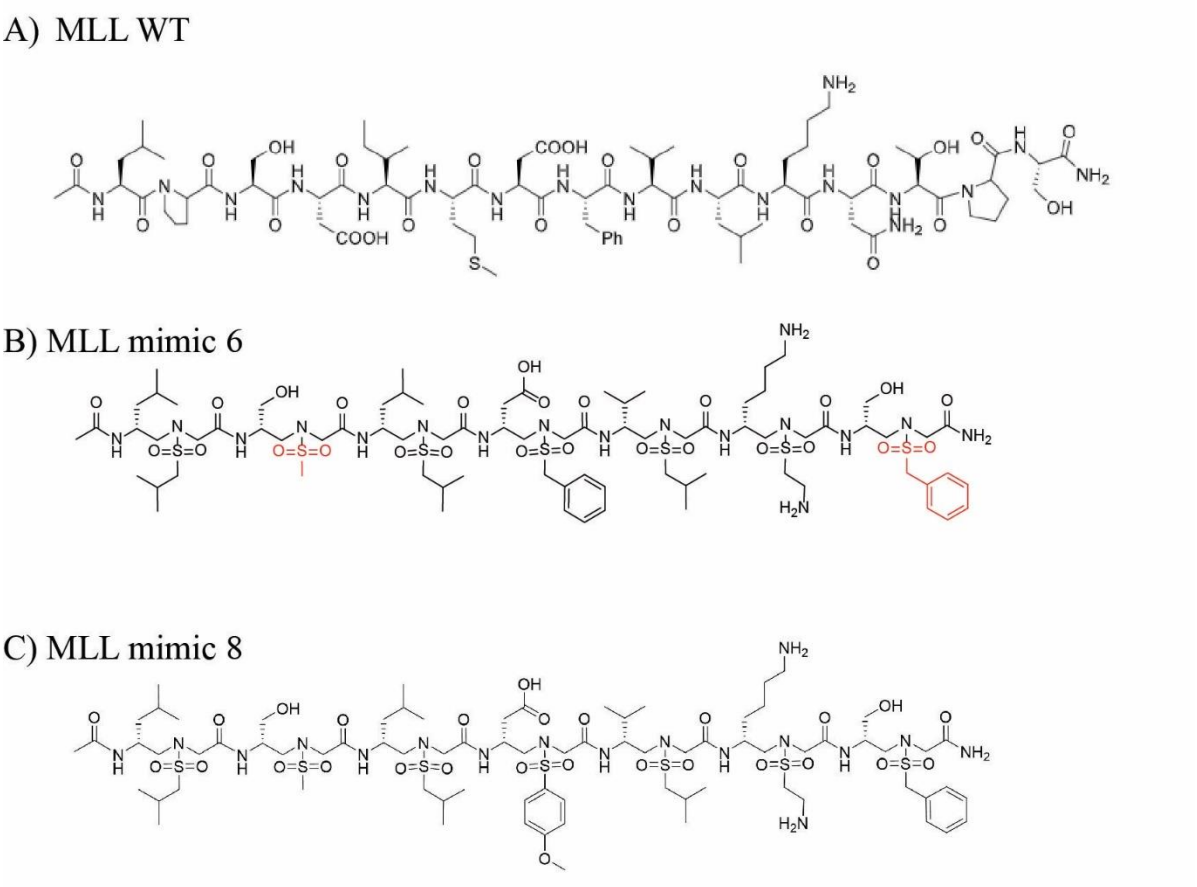


Figure A7: Synthetic MLL WT and MLL mimic 6 and 8 structures. A) MLL WT, B) MLL mimic 6, and C) MLL mimic 8.

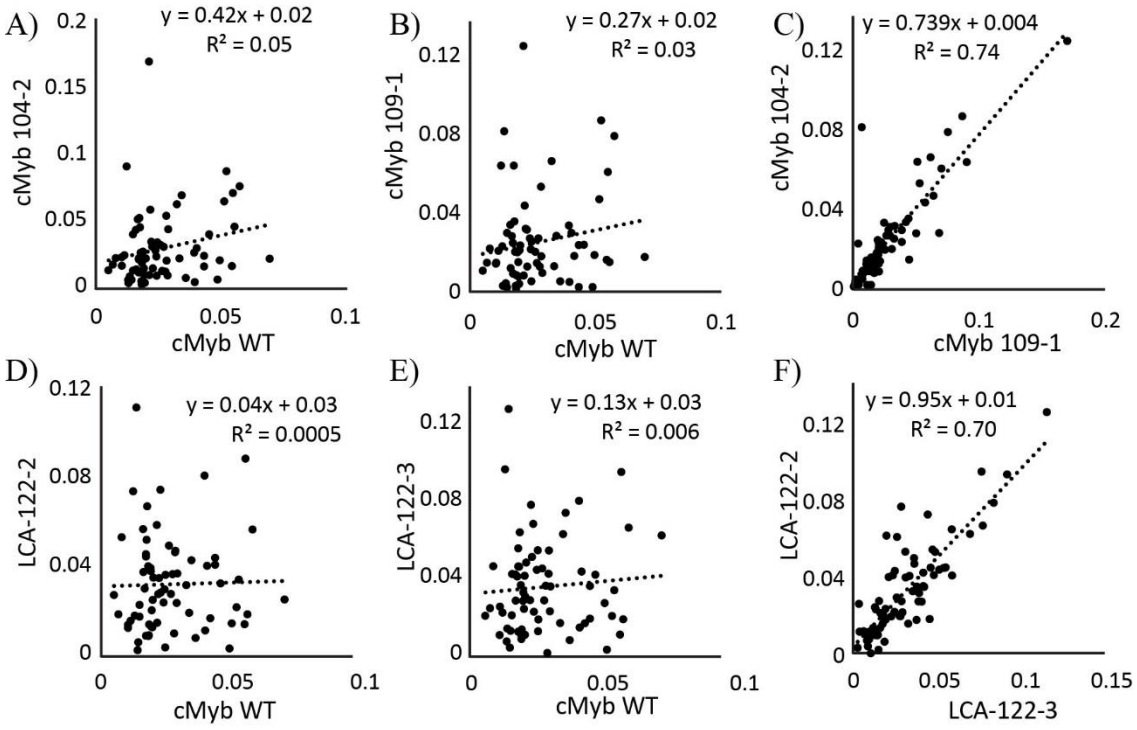


Figure A8: Correlation plots of cMyb WT and cMyb mimics. A) cMyb WT and cMyb 104-2, B) cMyb WT and cMyb 109-1, C) cMyb 109-1 and cMyb 104-2, D) cMyb WT and LC-A-122-2, E) cMyb WT and LC-A-122-3, and F) LC-A-122-3 and LC-A-122-2.

APPENDIX C – COPYRIGHT PERMISSIONS

The following is permission for the article in Chapters Two and Six:

https://www.mdpi.com/authors/rights

MDPI

Journals Topics Information Author Services Initiatives About

Search for Articles: Title / Keyword Author / Affiliation / Email All Journals All Article Types

Copyrights

Copyright and Licensing

For all articles published in MDPI journals, copyright is retained by the authors. Articles are licensed under an open access Creative Commons CC BY 4.0 license, meaning that anyone may download and read the paper for free. In addition, the article may be reused and quoted provided that the original published version is cited. These conditions allow for maximum use and exposure of the work, while ensuring that the authors receive proper credit.

In exceptional circumstances articles may be licensed differently. If you have specific condition (such as one linked to funding) that does not allow this license, please mention this to the editorial office of the journal at submission. Exceptions will be granted at the discretion of the publisher.