November 2017

Disorder Levels of c-Myb Transactivation Domain Regulate its Binding Affinity to the KIX Domain of CREB Binding Protein

Anusha Poosapati
University of South Florida, anushap@mail.usf.edu

Disorder Levels of c-Myb Transactivation Domain Regulate its Binding Affinity to the KIX Domain of CREB Binding Protein

by

Anusha Poosapati

A thesis submitted in partial fulfilment
of the requirements for the degree of
Master of Science
with a concentration in Cell and Molecular Biology
Department of Cell Biology, Microbiology and Molecular Biology
College of Arts and Sciences
University of South Florida

Major Professor: Gary Wayne Daughdrill, Ph.D.
Kristina Schmidt, Ph.D.
Sameer Varma, Ph.D.

Date of Approval:
November 3, 2017

Keywords: Coupled folding and binding, Intrinsically disordered proteins, Nuclear Magnetic Resonance Spectroscopy, transient helicity

ACKNOWLEDGMENTS

I would like to acknowledge my principal investigator, Dr. Gary Wayne Daughdrill, for accepting me in his lab. This work would not have been possible without his continual guidance, support and patience. I would like to thank my committee members, Dr. Kristina H. Schmidt and Dr. Sameer Varma for their guidance, feedback and valuable inputs throughout the work. I would also like to thank Wade M. Borcherds, a postdoctoral researcher in our lab for training me, and for his immense help and support throughout my time in the program. A special thank you to him for carrying out the NMR experiments. I would like to thank NMR core facility and Centre for Drug Discovery and Innovation for maintaining the magnets. Lastly, I would like to thank the Department of Cell Biology, Microbiology and Molecular Biology for accepting me into the program and giving me a chance to begin my research career.

TABLE OF CONTENTS

List of Tables

List of Figures

ABSTRACT

Intrinsically disordered proteins (IDPs) do not form stable tertiary structures like their ordered partners. They exist as heterogeneous ensembles that fluctuate over a time scale. Intrinsically disordered regions and proteins are found across different phyla and exert crucial biological functions. They exhibit transient secondary structures in their free state and become folded upon binding to their protein partners via a mechanism called coupled folding and binding. Some IDPs form alpha helices when bound to their protein partners. We observed a set of cancer associated IDPs where the helical binding segments of IDPs are flanked by prolines on both the sides. Helix-breaking prolines are frequently found in IDPs flanking the binding segment and are evolutionarily conserved across phyla. Two studies have shown that helix flanking prolines modulate the function of IDPs by regulating the levels of disorder in their free state and in turn regulating the binding affinities to their partners. We aimed to study if this is a common phenomenon in IDPs that exhibit similar pattern in the conservation of helix flanking prolines. We chose to test the hypothesis in c-Myb-KIX, IDP-target system, in which the disordered protein exhibits high residual helicity levels in its free state.

c-Myb is a hematopoietic regulator that plays a crucial role in cancer by binding to the KIX domain of CBP. Studying the functional regulation of c-Myb by modulating the disorder levels in c-Myb and in IDPs in general provides a better understanding of the way IDPs function and can be used in therapeutic strategies as IDPs are known

to be involved in regulating various cellular processes and diseases. To study the effect of conserved helix flanking prolines on the residual helicity levels of c-Myb and its binding affinity to the KIX domain of CBP, we mutated the prolines to alanines. Mutating prolines to alanines increased the helicity levels of c-Myb in its free state. This small increase in the helicity levels of a highly helical c-Myb showed almost no effect on the binding affinity between cMyb and KIX. We hypothesized that there is a helical threshold for coupled folding and binding beyond which helicity levels of the free state IDP have no effect on its binding to their ordered protein partner. To test this hypothesis, we mutated solvent exposed amino acid residues in c-Myb that reduce its overall helicity and studied its effect on the binding affinity between c-Myb and KIX. Over a broad range of reduction in helicity levels of the free state did not show an effect on the binding affinity but beyond a certain level, decrease in helicity levels showed pronounced effects on the binding affinity between c-Myb and KIX.

CHAPTER 1

INTRODUCTION

1.1 Intrinsically Disordered Proteins

Most proteins adopt compact folded structures to exert cellular functions. They fold into stable tertiary and quaternary structures [1, 2] positioning the group of amino acid residues that are involved in the catalysis of various chemical reactions on the solvent-exposed side. In the past two decades, a set of proteins that do not fall into this category has been receiving recognition. These proteins, called the intrinsically disordered proteins (IDPs), are highly dynamic and do not form stable tertiary structures [3-5], but nevertheless are important for the functioning of the cell [3, 6-20]. They form specific heterogeneous conformational ensembles that fluctuate over time [21, 22].

IDPs are highly prevalent in nature and found across kingdoms especially with eukaryotes exhibiting a higher percentage of IDPs than algae and archaea [6, 23-25]. More than one-third of the eukaryotic proteins contain regions with disorder of over 30 residues or higher in length [6, 7, 11, 26]. Eukaryotic transcription factors [27] and disease-associated proteins are rich in intrinsic disordered regions (IDRs) [9, 28-30]. IDPs regulate several biological processes like transcription, translation, cell signalling, cell proliferation, etc. All these processes are regulated through post translational modifications [31, 32], alternative splicing [33, 34], insertions and deletions [35], and alternative open reading frames [36]. Post translational

modifications [31] and either small molecule ligand [31, 37] or macromolecular protein binding induces IDRs or IDPs to become structured, or structured domains to become unstructured [38]. Disorder to order transition is known to be involved in the regulation of many cellular processes [39].

1.1.1 Structure and Composition of IDPs – Ordered v/s Disordered Proteins

Composition of amino acid residues varies greatly between disordered and ordered proteins [4, 6, 40-43]. IDPs exhibit a low sequence complexity with a low fraction of hydrophobic residues and a high fraction of polar-charged residues [5, 44-47]. Ordered proteins are rich in non-polar and hydrophobic residues that collapse into a compact hydrophobic core thus stabilising their structure. The depletion of hydrophobic residues in IDRs prevents the formation of a hydrophobic core, catering to its flexibility. The presence of a high fraction of charged residues in IDPs allows repulsion of the neighbouring amino acid residues, further preventing and destabilizing the formation of a compact structure thus facilitating in the maintenance of their flexibility. Acidic and basic amino acids and structure-breaking amino acid residues like prolines and glycines are abundantly found in IDRs [5, 48]. These structure-preventing amino acids further extend the IDRs allowing multiple points of attachment when encountering their binding partners. Disordered regions are also depleted of aromatic amino acids. They exhibit high sequence variability including multiple insertions and deletions [35, 49]. Long insertions and deletions usually are associated with disordered regions [35, 50]. The flexibility of disordered regions allows binding to multiple protein partners [13, 41, 51, 52], which is a common feature of IDPs.

IDPs, with constant fluctuations in their conformational ensembles and high repulsions between neighbouring residues, do not maintain an equilibrium state for the positions of their atoms and bond angles thus rendering flexible backbone dihedral angles, phi and psi [7, 53]. All proteins maintain some level of flexibility in aqueous solutions. Despite this fact, ordered proteins, including non-folded structures like loops, are comprised of ensembles with all its members having nearly identical Ramachandran angles. Contrary to that, disordered proteins are comprised of ensembles with individual members possessing unique Ramachandran angles [45]. Disordered regions exist in a broad range of the structural spectrum from completely unfolded to partially folded to structures with near natively folded conformations, i.e., collapsed, extended or intermediates between collapsed and extended forms [11, 45, 54, 55]. As previously described, IDPs are comprised of a high fraction of charged residues. Disordered proteins with a significant net charge exist as extended forms; the higher the fraction of like charges, the higher the extension in the disordered region [3, 56, 57]. IDPs that possess a net neutral charge along with a few hydrophilic amino acid residues exist in collapsed forms, forming random structures but not defined secondary structures. Such proteins fall under the pre-molten globule class of IDPs [3, 22]. Disordered proteins with a significant fraction of hydrophobic residues i.e., IDRs that possess structure-promoting sequence patterns, but insufficient folding energy exist as collapsed forms with stable secondary structures but unstable or no tertiary interactions. This class of IDPs is termed molten globules [3, 58].

While ordered proteins are classified into different groups based on their secondary structures [1, 2], the broad range of structural differences in disordered regions concomitant with their failure to fold into complex structural levels, makes it difficult to classify IDPs into distinct groups [45]. Despite the difficulty, attempts have

been made to classify disordered proteins into various classes based on their structure, sequence, biophysical and functional features [59]. They are further grouped into different classes based on their evolution and interactions with other proteins [59].

IDPs can be recognized and studied by various techniques such as to Circular Dichroism (CD), Nuclear Magnetic Resonance (NMR) Spectroscopy, Optical Rotatory Dispersion (ORD), X-ray Crystallography [6, 11, 60]. Among all the available techniques, NMR is well suited for studying IDPs [61]. Disordered regions can be distinguished from ordered regions by a far U.V spectrum from CD analysis, and low chemical shift dispersions and negative nuclear overhauser effects (NOEs) from NMR analysis.

1.1.2 Functions of Intrinsic Disordered Proteins and Intrinsic Disordered Regions

Disordered proteins and regions are functional in all their structural forms i.e., pre-molten and molten globules, and extended forms. IDPs are vital in regulating numerous biological processes by binding to other proteins [15, 16, 62], nucleic acids [34, 63-66] and are also involved in ribosomal assembly [39, 67, 68], allosteric regulation [69-73] and chaperone activity [8, 38, 74, 75]. Post-translational modifications like methylation, acetylation, phosphorylation and ubiquitination are known to be linked to disordered proteins and regions [15]. Apart from participating in molecular recognition and assembly, as in the case of ribosomal assembly, disordered regions play a pivotal role in complex protein-protein interaction networks acting as network hubs [13, 76-78]. In the current section some of the functional roles of IDPs are described in detail.

1.1.2.1 Solubility enhancement. Protein solubility is dependent upon the interactions of amino acid residues with water. Ordered proteins are collapsed into a compact hydrophobic core exposing the polar, charged residues to the solvent-exposed side. Charged residues interact with water leading to the solubility of protein in aqueous solutions [59]. There are some hydrophobic residues on the solvent-exposed side that interact with other proteins. These residues combined with the denaturation of folded regions lead to the exposure of a significant fraction of hydrophobic residues that are not in favour of interacting with water, sometimes leading to aggregation and precipitation. Disordered proteins are not folded into compact structures and are rich in polar, charged residues. Charged residues favourably interact with water, aiding in the solubility of IDRs. Compact globular structures are susceptible to denaturation by heat; disordered regions failing to fold into stable tertiary structures are resistant to heat denaturation. One such example is the cyclin dependent kinase inhibitor protein, p21, that plays a key role in apoptosis. It was found to be stable and soluble for a wide range of temperatures ranging from 5-90 ºC [52]. In fact, early analysis and purification of disordered proteins was done by denaturation of proteins by heat [79]. Therefore, disordered proteins pose an advantage in laboratory protein expression and purification, over ordered proteins [80]. Studies have been performed where addition of disordered proteins as tags to naturally poorly-soluble proteins was shown to enhance the solubility of the protein [81, 82].

1.1.2.2 Chaperones. Chaperones are proteins that assist in folding of other proteins and RNA [83-87]. Intrinsic disordered regions are abundantly found in chaperones; half of the RNA chaperone sequences and one-third of protein chaperone sequences are disordered [87, 88]. Disordered regions are well adapted to be

chaperones. Their lack of definite structure allows them to bind to multiple binding partners and, therefore, are suitable for binding to a wide range of protein and RNA molecules. Interactions involving disordered regions are faster when compared to those involving only ordered regions. This is because the unfolded structure of IDRs allows easy access to multiple amino acid residues, making recognition by their binding partners easier. IDPs constantly fluctuate between different conformational ensembles, which allows numerous non-specific interactions between IDPs and their binding partners, ultimately increasing the life-time of the bound complex [89, 90]. This quick interaction and prolonged association time with disordered chaperones gives sufficient time for the misfolded protein to unfold and refold correctly. Fast interaction between disordered chaperones and misfolded proteins also helps in preventing the formation of aggregates and consequent precipitation, by enhancing the solubility of the misfolded protein [59].

1.1.2.3 Post-translational modifications. IDPs are associated with an enormous number of post-translational modifications, as many as a million [8, 16, 24, 91]. This enormity emphasizes the importance of intrinsically disordered proteins in cellular signalling and regulation. The flexibility of disordered regions facilitates easy access to modifying enzymes such as kinases and acetylases for recognition and binding. Flexibility of disordered regions also enables transient interactions between IDRs and modifying enzymes [92, 93]. Upon binding of the modifying enzyme, disordered proteins lose their conformational entropy, in turn leading to a reduction in the overall free energy of binding, which results in a weak interaction. Processes involving post-translational modifications require binding of the protein to more than one macromolecule i.e., the modifying enzyme and the receptor molecule. Disordered proteins are well suited for this because unstructured regions can fold into different

conformations required for different targets [8]. Disordered regions are known to be highly enriched in phosphorylation sites [93, 94].

1.1.3 Coupled Folding and Binding

Coupled folding and binding involves the binding and folding of an intrinsically disordered protein to their protein partner to exert important cellular functions. IDPs exhibit transient secondary structures in free state but fold upon binding to their protein partners the process of which is termed as coupled folding and binding. Two mechanisms namely conformational selection and induced fit are believed to be linked to coupled folding and binding. The mechanism by which IDPs bind to their targets has been debated for decades but evidences have shown that IDPs bind to their partners by either of the extreme mechanisms or both together.

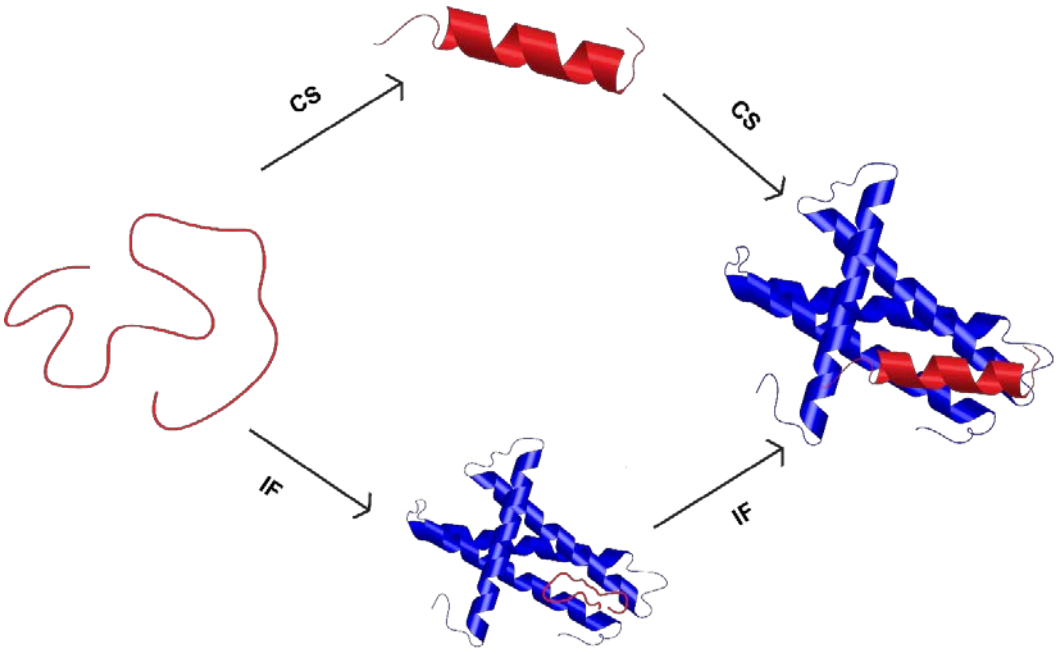


Figure 1: Coupled folding and binding. Mechanism of coupled folding and binding showing induced fit and conformational selection mechanisms. IF denotes induced fit and CS denotes conformational selection.

In the conformational selection mechanism, disordered protein which fluctuates between different conformations adopts a conformation that is close to the IDP in its bound form which is then followed by binding to its protein partner (**Figure 2**) i.e., a conformation is selected for binding and hence the name, conformational selection. In the induced fit mechanism, disordered protein makes a couple of weak interactions with its ordered protein which is followed by its folding and subsequent tight interaction with its protein partner.

1.1.4 Evolution of IDPs

IDRs show higher mutation rates than ordered proteins [49] which is expected since these proteins unlike ordered proteins don't possess a compact hydrophobic core that needs to be preserved. Despite high mutation rates, certain amino acid residues like tyrosine, tryptophan, leucine, phenylalanine and proline are highly conserved in IDRs [6, 42, 44].

1.2 Myeloblastosis Protein, c-Myb

c-Myb (**My**elo**b**lastosis protein) is a hematopoietic transcription factor that plays a crucial role in regulating cellular differentiation and cell proliferation [95-99]. It decides the fate of the cell i.e., decides whether the cell proceeds with continuous proliferation or cell cycle arrest is induced and the cell starts differentiating. c-Myb is highly conserved and is found not only in insects and vertebrates but also in some plants and invertebrates [99-101]. Reduction or loss of c-Myb function is linked to embryonic lethality and most human leukemias, colorectal, breast and some epithelial

cancers [96, 99, 102-108]. c-Myb is abundantly expressed in hematopoietic cells but is also seen in other types of cells [109-112].

1.2.1 History of c-Myb

c-Myb belongs to the myb family of transcription factors, which, like c-Myb, are highly conserved and act as transcriptional coactivators. The other members of the family include A-Myb and B-Myb which are structurally similar to c-Myb with minute differences which are discussed in the next section [113]. Myb gene, termed v-Myb, was originally discovered as an oncogene in chicken leukemia viruses, Avian Myeloblastosis Virus and E26 virus [114]. It was later revealed that v-Myb is a truncated version of c-Myb that was transduced by the viruses from chicken cells [114, 115]. c-Myb is the normal cellular counterpart of v-Myb and hence the name c-Myb which means cellular Myb.

1.2.2 Structure of c-Myb

The proto-oncogene c-Myb is translated into two protein forms. The major translated form is a 75 kDa residue protein that acts a hematopoietic transcriptional activator. It consists of a conserved N-terminal DNA binding domain, an acidic transcriptional activation domain and a C-terminal specificity and negative regulatory domain (**Figure 2**). The C-terminal domain consists of various motifs with specific functions. It includes a “FAETL” domain that is responsible for its oncogenic activity [116], an “EVES” domain that is responsible for intramolecular and intermolecular interactions during negative regulation of the protein [117]. It also consists of a “TPTPF” domain whose function is unknown. Poly-proline regions are also seen in the C-terminal domain and are speculated to be involved in conformational changes [95].

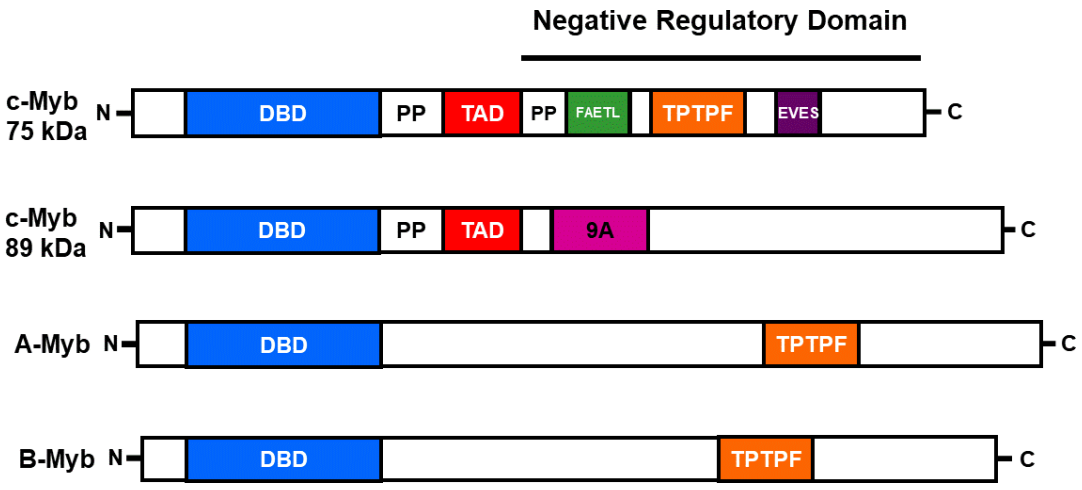


Figure 2. Schematic diagram of members of the Myb protein family. It shows the DNA binding domain (DBD), transactivation domain (TAD) and negative regulatory domain that has various amino acid residue motifs. Blocks with PP represent the poly-proline rich regions. Box with 9A in c-Myb 89 kDa schematic diagram represents the protein encoded from alternative splicing.

The minor translated form of c-Myb arises from alternative splicing of the proto-oncogene which is seen in hematopoietic cells of birds, mice and humans [118-120]. It is an 89 kDa protein which comes from an alternative splice form that results in the addition of a 363-bps fragment between exons 9 and 10. This region was given the name exon 9A. A-Myb and B-Myb are structurally similar to c-Myb though they have specific functions (**Figure 2**). DNA binding domain and the "TPTPF" domains are seen in A-Myb and B-Myb, too. v-Myb also shows structural similarities with c-Myb. Although, DNA binding domain and the transactivation domain from c-Myb are conserved in v-Myb, several truncations are seen in the N- and C-terminal regions. v-Myb from AMV virus also has eleven point-mutations throughout the length of the protein that are shown to contribute to its oncogenic activity [121-123]. v-Myb from E26 virus is a complex form of v-Myb where an oncogenic peptide is added to the C-terminus region. Both v-Myb proteins show truncations in the C-terminus region and

are responsible for oncogenicity thus implying that the C-terminus region of c-Myb has tumour suppressing activity [124, 125]. c-Myb exhibits disorder levels throughout its structure as shown in **figure 3.**

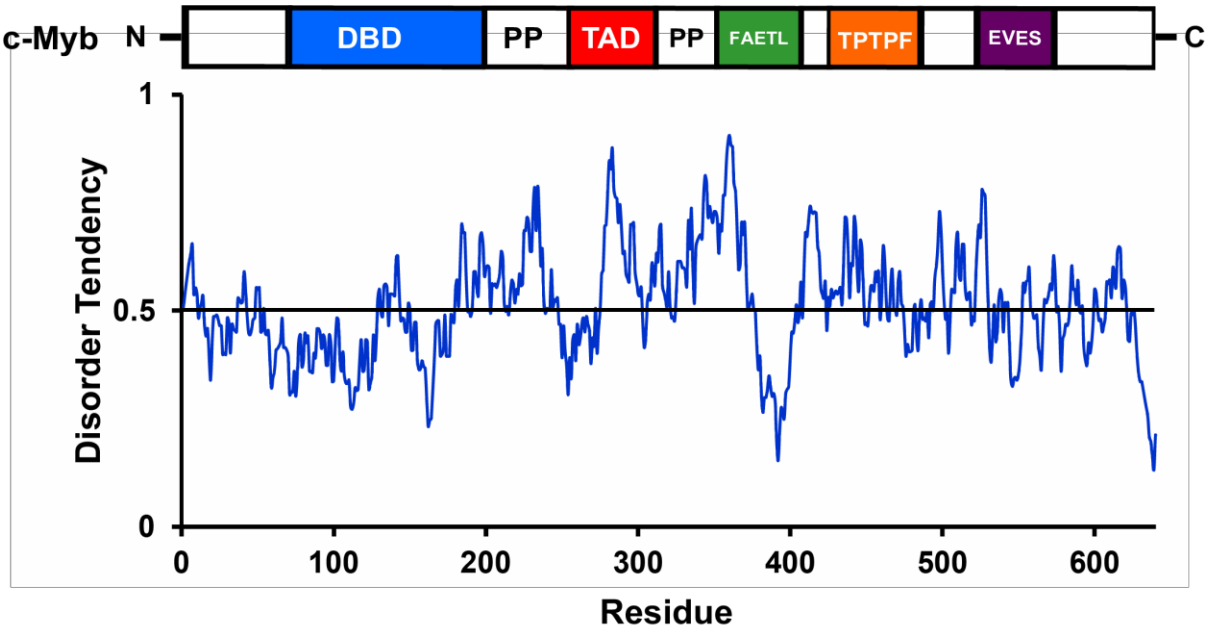


Figure 3: Disorder plot of full length c-Myb. Figure shows the disorder tendency plot of full length c-Myb obtained from the IUPRED predictor aligned with the schematic diagram of c-Myb. Values above 0.5 suggest disorder level tendencies.

1.2.3 Functions of c-Myb

As mentioned previously, c-Myb is a DNA binding protein that acts as a transcriptional activator and plays a key role in cell proliferation and differentiation. Functions of each domain in the protein are described below. The N-terminal DNA binding domain (DBD) binds specifically to the consensus sequence PyAACT/GG that is known as the MYB binding site (MBS) [126]. It consists of three tandem sequence repeats R1, R2 and R3 of approximately 50 amino acid residues long, which are highly conserved [99, 127, 128]. All these three repeats consist of tryptophans separated by

18-19 amino acid residues and are highly conserved among all Myb proteins and across species [99, 113, 128]. R2 and R3 repeats contain two helix-turn-helix motifs and are required for the binding of c-Myb to DNA, and R1 is responsible for the stabilization of DNA-R2R3 complex. The hydrophobicity of conserved aromatic tryptophan residues in the tandem repeats is essential for maintaining the stability of helix-turn-helix motifs [129]. Transactivation domain of c-Myb binds to several proteins and transactivates the expression of multiple genes thus regulating several cellular processes including but not limited to cell cycle regulation and cell signalling. This study is focussed on its binding to the CREB-Binding Protein (CBP), a histone acetyl transferase. The mechanism of this interaction is described in detail in the later section of this chapter. The C-terminal domain is responsible for negative regulation of the protein. Studies involving deletion of C-terminal sequences in c-Myb and v-Myb, which lacks the C-terminal region, show an increase in the transcriptional activity of c-Myb [124, 130].

1.2.3.1 Role in hematopoiesis and adult brain neurogenesis. c-Myb regulates cell proliferation and differentiation in hematopoietic progenitor cells and neural progenitor cells [131]. Early studies have linked c-Myb's role to hematopoiesis [132-135]. c-Myb knockout mice die of anoxia during embryogenesis at day 15. A loss in almost all blood cell lineages and a reduced growth of fetal liver is seen in these mice which implies that c-Myb is essential in the early stages of development [136]. Expression of c-Myb is generally restricted to progenitor cells and is downregulated as cells begin differentiating [135, 137]. Overexpression of c-Myb prevents differentiation of myeloid and erythroid lineages [138, 139]. An exception to this is seen in B and T cell lymphocyte differentiation, which require the presence of c-Myb at almost all stages although its expression is downregulated as cells differentiate. T cell

differentiation pathway requires expression of c-Myb at all stages except for the early stages i.e., differentiation of DN1 (double negative 1) and DN2 cells [140]. Differentiation of DN3, DN4 and DP (double positive) cells require c-Myb. Likewise, B-cell differentiation pathway requires the expression of c-Myb to differentiate from pro-B cells to pre-B cells to B cells [141].

In adult vertebrates, new neural cells are continuously produced and differentiated throughout the lifetime of an individual. Conditional knock out of c-Myb in mice brain, caused defects in adult brain neurogenesis and was linked to decreased expression of crucial neurogenesis genes Pax6 and sox2. Knockout mice show defects in neural progenitor cells leading to reduced or no production of new neural cells, enlarged ventricle spaces and defects in the circulation of cerebrospinal fluid [131].

1.2.3.2 Cell cycle regulation. Activity of c-Myb is highest during late G1 phase and S phase. Destruction of c-Myb RNA arrests cell cycle at G1 phase which implies that c-Myb is essential for cell cycle progression. c-Myb is known to regulate a number of genes that are linked to cell growth, like stem cell growth factor c-kit [142], fibroblast growth factor FGF2 [143] and DNA topoisomerase IIα [144]. c-Myb also regulates the expression of bcl-2 which is responsible for apoptosis [145].

1.2.4 Role in Cancer

c-Myb is linked to leukemia, breast, and colon cancers but is also known to be related to other cancers like melanoma, oesophageal and pancreatic cancer [131]. c-Myb is required for the proliferation and survival of chronic myeloid leukemia (CML) and acute myeloid leukemia (AML) cells, and its over-expression is observed in both cancers [146, 147]. Duplication and translocation of c-myb gene is observed in T cell

acute lymphoblastic leukemia [148, 149]. Over expression of c-Myb is also seen in more than 80% of the cases in colorectal cancers [150-152]. Mutations in the first intron of c-myb, which is linked to its regulation by transcriptional elongation (see 'regulation of c-myb' section), are found in primary colorectal cancers [105]. In other studies, performed using primary colorectal cancer genomic DNA, hypermutations in the poly T tract and the stem loop were also observed [106].

1.2.5 Regulation of c-Myb

c-Myb has a half-life of approximately thirty minutes [131]. Activity of c-Myb is regulated by altering its protein levels through various post translational modifications including phosphorylation, acetylation, sumoylation and ubiquitylation. Mature c-Myb RNA is restricted to hematopoietic cells and intestinal epithelial tissues although differentiation in hematopoietic cells results in the reduction of c-Myb levels.

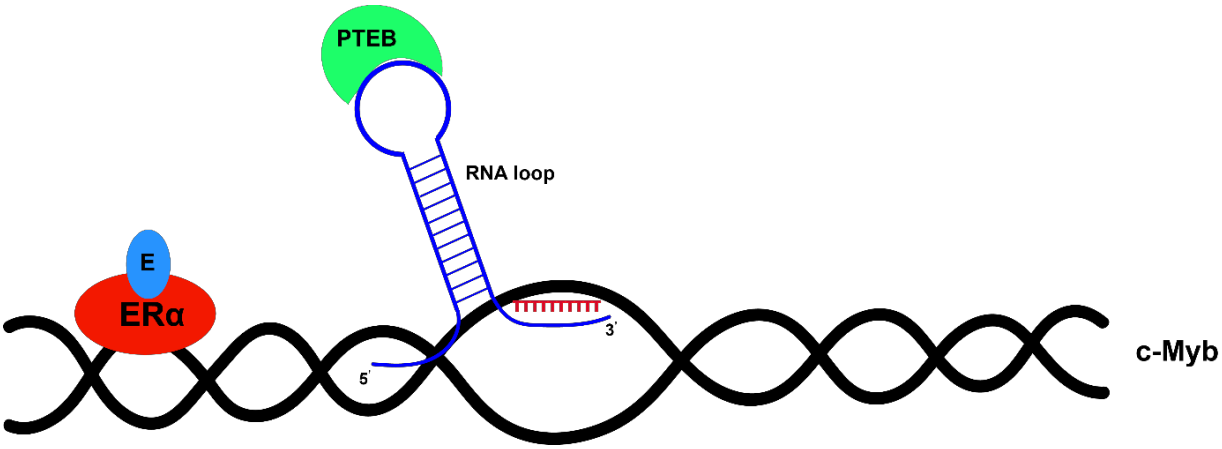


Figure 4. Model representing c-Myb regulation through elongation control. Figure shows the poly T tract present in c-Myb gene. It also shows the RNA stem loop that binds to elongation control proteins abrogating their access to DNA binding. ERα binds upstream of the poly T tract and induces transcription.

Expression of c-Myb is tightly regulated by blocking transcriptional elongation in the first intron [105-107, 153-156]. Region in the human c-myb gene that is responsible for the attenuation of its expression consists of a poly T tract of 19-23 thymidine residues that binds to an RNA stem loop eventually blocking its transcription (**Figure 4**). Poly T tract is preceded by a region which encodes an RNA stem loop that interacts with elongation control proteins like positive elongation transcription factor B (PTEB) [105, 106]. In colon cancers, a reduction in the size of poly T tract and/or mutation in the region encoding the RNA stem loop can occur, leading to an increased expression of c-Myb and cellular proliferation. In breast cancers, oestrogen receptor-α interacts with the gene when bound to oestradiol which relieves elongation block.

1.2.6 c-Myb Target Genes and Binding Partners

c-Myb target genes include genes that are responsible for proliferation such as *CCNA1, CCNB1, CCNE1, MYC, KIT, BCL2*. Target genes linked to differentiation include *MIM1, CD4, PTCRA, ELA2* among others. c-Myb also targets a few housekeeping genes like *GSTM1* and *MAT2A*.

The most important protein-protein interaction for c-Myb is binding to the CREB binding protein (CBP a.k.a. p300) through its transactivation (TAD) domain. c-Myb binds to the KIX domain of CBP/p300 via coupled folding and binding [157, 158]. Interaction between CBP and c-Myb is described in detail in the later sections of the chapter. c-Myb also binds to another co-activator, p100, through its DNA binding domain. Binding of c-Myb DNA binding domain to p100 is believed to abrogate the intramolecular interaction between DNA binding domain and the EVES motif in C-terminal regulatory domain by competing with it. Phosphorylation of S528 from the EVES motif reduces c-Myb activity [117]. It is speculated that phosphorylation of S528

induces intramolecular binding and thereby c-Myb transactivation is inhibited. p100 disrupts this binding and relieves the inhibition. The C-terminal regulatory domain of c-Myb binds to a variety of protein partners. The FAETL motif, which lies in the leucine zipper, is responsible for transactivation and transforming ability. Mutation of FAETL motif and leucine zipper results in increased transcriptional activity revealing that it is involved in negative regulatory interactions [116, 159, 160].

1.2.7 CREB Binding Protein (CBP), c-Myb'S Binding Partner

CREB (cAMP response element binding protein) binding protein (CBP) is a transcriptional co-activator that binds to several proteins including general transcription factors and DNA in a sequence specific manner [161, 162]. CBP binds to a broad range of proteins that are involved in crucial cellular processes like cell growth, differentiation, DNA repair and apoptosis [163, 164]. CBP exhibits an inherent acetyl transferase activity with which it acetylates not only the histone tails of chromosomes but also some proteins and transcription factors like c-Myb [165].

A schematic diagram of CBP, with all its functional domains is shown below (**Figure 5**). CBP is a 2441 amino acid residue protein that includes the following domains: 1) A nuclear receptor interaction domain which binds to the DNA and promotes transcription of adjacent genes. 2) A cysteine- and histidine-rich domain called the TAZ1 domain that binds to proteins such as p53, MDM2 and E1A. 3) KIX domain that binds to CREB, p53, c-Myb, STAT1 and BRCA1. 4) A bromo-domain. 5) A second cysteine-histidine (CH2) rich domain. 6) A histone acetyl transferase activity (HAT) domain. 7) A zinc finger. 8) TAZ2 domain that binds to proteins such as p53, TFIIB, E2F, E1A and GATA1. 9) A Nuclear Co-Activator Binding Domain [164, 166, 167].

Figure 5: Schematic diagram of mammalian CBP. Above picture represents all the functional domains of CBP. Domains with known structures are shown in color. Abbreviations are given to each domain, which are explained in the text.

1.2.8 Interaction Between c-Myb TAD and the KIX Domain of CBP

The transactivation domain of c-Myb binds to the KIX domain of CBP via coupled folding and binding. This interaction results in the activation and transcription of many genes that are linked to biological functions. c-Myb transactivation domain exhibits disorder levels in its free state. An IUPRED plot of c-Myb TAD disorder is shown in **Figure 6**.

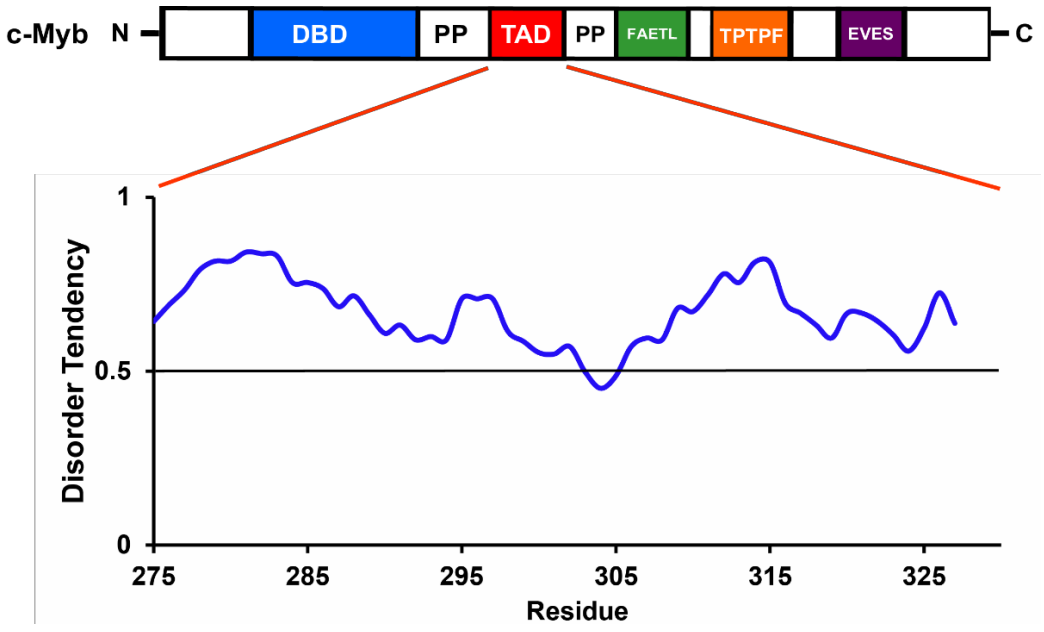


Figure 6: Disorder tendency of c-Myb TAD as predicted using IUPRED. Disorder tendency plot of c-Myb TAD obtained from the IUPRED predictor. Values above 0.5 suggest disorder levels. Schematic diagram of full length c-Myb is placed on top of the plot.

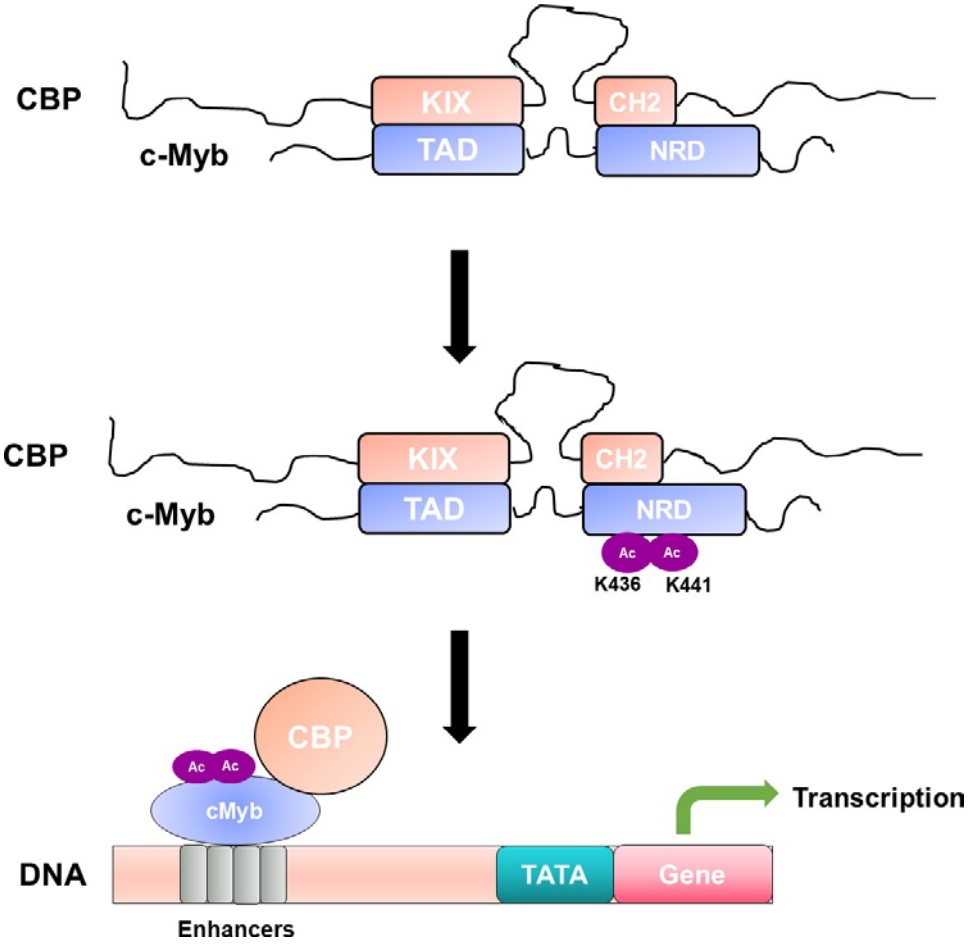


Figure 7: Figure showing mechanism of c-Myb activation by CBP. KIX domain of CBP binds to the transactivation domain of c-Myb while CH2 domain of CBP binds to the negative regulatory domain (NRD). Upon binding of CH2 domain to the NRD of c-Myb, the acetyl transferase activity domain in CBP acetylates lysine residues at positions 436 and 441. Acetylated c-Myb binds to the enhancers of DNA and promotes transcription of downstream genes.

CBP binds to the transactivation and regulatory domains of c-Myb through its KIX and CH2 domains respectively (**Figure 7**). When the CH2 domain of CBP binds to the regulatory domain of c-Myb, the histone acetyltransferase activity region in CBP acetylates lysine 437 and lysine 448 residues of c-Myb that are present in the region containing TPTPF motif. Acetylated c-Myb is translocated in-to the nucleus and activates transcription.

1.3 Specific Aims

Intrinsically disordered proteins do not form tertiary structures but exhibit transient secondary structures in aqueous solutions. They are involved in various protein-protein interactions and become folded upon binding to their ordered proteins via a process called coupled folding and binding. Many IDPs, like the transactivation domain of c-Myb exhibit transient helical structures in free state that resemble their bound structures. Conserved prolines are found flanking the helical binding segments of some cancer associated IDPs. We hypothesized that the levels of transient helicity in the free state disordered proteins regulate the binding affinity of IDPs to their protein partners. Previous studies have shown that increasing the transient helicity levels of p53, a tumour suppressor protein, and MLL, mixed lineage leukaemia protein, in their free state affects their binding affinities to MDM2 and KIX respectively. It was observed that conserved helix flanking prolines regulate transient helicity levels in both these systems. Do conserved helix flanking prolines regulate transient helicity levels of disordered proteins and in turn the binding affinities to their protein partners in all IDPs? We aimed to study the effect of changing transient helicity in c-Myb TAD, which exhibits high levels of transient helicity in its free state, on its binding affinity to the KIX domain of CBP. To test our hypothesis, we designed the following aims.

Aim 1 (Chapter Two) - Study the effect of conserved helix flanking prolines on the transient helicity levels of c-Myb TAD and its binding affinity to the KIX domain of CBP.

Aim 2 (Chapter Three) – Study the effect of changing transient helicity levels of c-Myb TAD on its binding affinity to the KIX domain of CBP by mutating solvent exposed sites in c-Myb TAD.

CHAPTER 2

EFFECT OF CONSERVED HELIX FLANKING PROLINES ON THE STUCTURE AND BINDING AFFINITY OF IDPs TO THEIR TARGETS

Note to the readers: This chapter contains published data from Biochemistry, 2017, 15: 2379-2384, and has been reproduced with permission.

2.1 Rationale: Conservation of Prolines in c-Myb TAD and Other IDPs

As discussed previously, a common phenomenon observed in IDPs is that they fold upon binding to their protein partners via coupled folding and binding [168]. Some of these proteins exhibit transient helical structure in aqueous solution and form amphipathic alpha helices upon binding to their ordered protein partners. Proline residues are frequently found flanking the helical binding regions in these IDPs (**Figure 8**) [5, 169]. If prolines, known helix breakers, are commonly seen flanking the helical binding segments in disordered proteins, and are conserved across species, it was speculated that they play a crucial role in biological function and that they might be doing so by regulating the levels of disorder in IDPs. In this study, the effect of conserved helix flanking prolines on the structure and function of IDPs is tested in the transactivation domains of three IDPs. **Figure 8** shows the alignments of transactivation domains of three IDPs, namely p53, c-Myb and MLL, that form alpha helices when bound to their respective protein partners. The rest of the chapter is focussed on studies conducted using the human sequences of transactivation

domains all the IDPs shown in the **figure 8**. The influence of conserved helix flanking prolines in these IDPs and their effects on the transient helicity levels and the binding affinities to their ordered protein partners is described in this chapter.

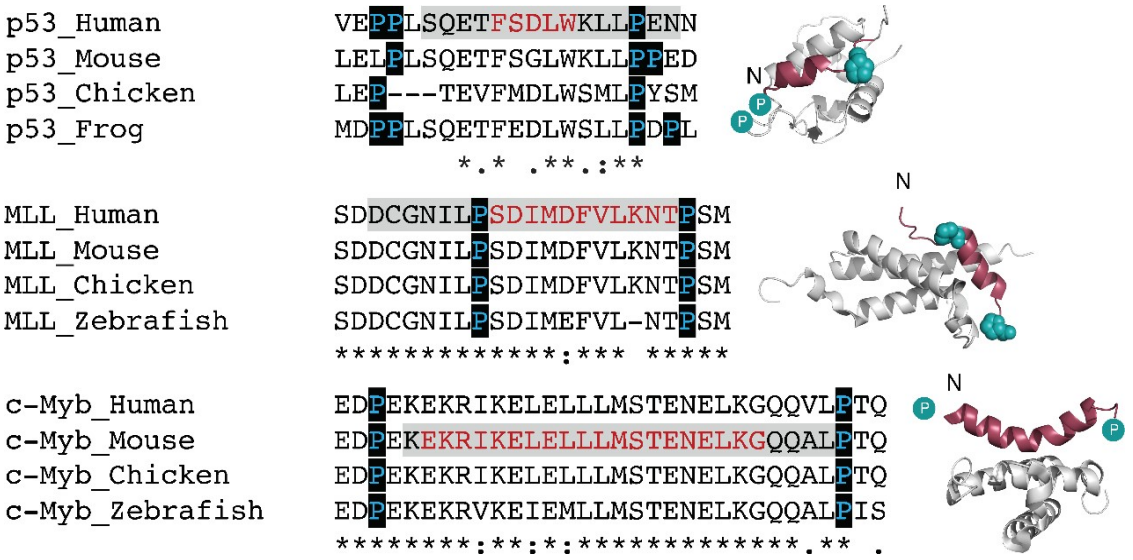


Figure 8: Alignment of IDPs showing conserved helix flanking prolines. Above picture shows conserved helix flanking prolines in three IDPs in various species (human, mouse, chicken, frog and zebrafish) adapted from previously published material [170]. These IDPs form alpha helices when bound to their protein partners. Bound complexes taken from PDB are shown to the right: IDPs in grey and ordered partners in red. Starting from the top, p53:MDM2 (PDB ID 1YCR), MLL:CBP KIX (PDB ID 2LXS) and c-Myb:CBP KIX (PDB ID 1SB0). Cyan spheres represent helix flanking prolines. Prolines that are not present in the PDB structures are represented as cyan spheres. N represents the N-terminal end of IDPs. Alignments shown in the middle panel were obtained using Clustal Omega. Helix flanking prolines are highlighted in blue in a black shaded area and regions that form alpha helices upon binding are shown in red. Grey shaded area represents the region used in the structure. Asterisks denote regions that are fully conserved. Colons indicate regions that are conserved with strong similar amino acid properties and period denotes sequences conserved with weak similar amino acid properties. This figure is taken from “Conserved Helix Flanking Prolines Modulate Intrinsically Disordered Protein:Target Affinity by Altering the Lifetime of the Bound Complex,” by Crabtree MD et.al., 2017, *Biochemistry*, 56, p. 2379. CC-BY.

2.2 Evidence of Conserved Helix Flanking Prolines Regulating the Binding Affinity Between IDP and Target

Previously published work shows the effect of conserved helix flanking prolines in p53 and MLL on their binding affinity to their respective protein interaction partners [170]. Structural and functional roles of conserved helix flanking prolines in these IDPs is studied by mutating the N- and C-terminal prolines to alanines. Structural effects were analysed by nuclear magnetic resonance spectroscopy, an excellent tool to study IDPs. The immediate effect of removal of prolines on the function of IDPs, the binding affinity to their protein partners, is studied using isothermal titration calorimetry and binding kinetics were measured using stopped-flow fluorescence techniques.

2.2.1 Effect of Conserved Helix Flanking Prolines on the Residual Structure of IDPs

NMR experiments were performed on uniformly ^{15}N and ^{13}C labelled MLL and p53. Proteins were purified as described in [170] and dialysed in NMR buffer. HSQC and three dimensional HNCACB and HNCO experiments were performed at 25 ºC and assignments were made as described in chapter 5. When the C-terminal proline was mutated in p53, the peak residual helicity of the resultant mutant in free state was increased approximately by 40% (**Table 1**). N-terminal proline mutations did not show a significant change in the residual helicity of p53. In MLL, the C-terminal proline mutation had no effect on the residual helicity in the free state while the N-terminal proline showed an increase in the helicity of the binding region by approximately 1.4-fold (**Figure 9**). The increase in the helicity for MLL was uniform across the length of the helix but the ppm change observed was within the digital resolution of the NMR experiments.

Peptide	Helicity (NMR)	Helicity (CD)	K_d Equilibrium (nM)	K_d Kinetics (nM)	k_{on} ($x10^6$ M^{-1} s^{-1})	k_{off} (s^{-1})
p53 WT	3	n.d	240 ± 60	176 ± 8	25 ± 1	4.40 ± 0.03
p53 P27A	6	n.d	25 ± 3	n.d	30.7 ± 0.7	fast, 0.37 ± 0.01 slow, 0.072 ± 0.005
p53 P12/13A	3	n.d	220 ± 30	n.d	22 ± 2	n.d
p53 P12/13/27A	6	n.d	17 ± 6	n.d	32.4 ± 0.9	n.d
MLL WT	2	13	600 ± 60	530 ± 30	24 ± 1	12.4 ± 0.3
MLL P21A	n.d	14	700 ± 120	560 ± 15	22.0 ± 0.6	12.22 ± 0.08
MLL P9/21A	3	16	17000 ± 2000	9000 ± 6000	20 ± 14	200 ± 16
MLL L8A	n.d	12	35000 ± 4000	n.d	n.d	170 ± 14

Table 1: Helicity and binding data of WT and mutant p53 and MLL. Table shows the average helicity of the peptides derived from CD and NMR analyses. It also shows the dissociation rate constants, k_{on} and k_{off} rates of WT and mutant peptides obtained from the interaction. Binding affinities of complexes obtained from kinetic and equilibrium studies are listed in the table. This table is adapted from "Conserved Helix Flanking Prolines Modulate Intrinsically Disordered Protein:Target Affinity by Altering the Lifetime of the Bound Complex," by Crabtree MD et.al., 2017, *Biochemistry*, 56, p. 2379. CC-BY.

2.2.2 Influence of Helix Flanking Prolines on the Binding Affinity of IDP:Target

Binding affinities between IDPs and their partners were measured using equilibrium binding experiments and stopped flow kinetics. The proline to alanine mutants of p53 and MLL showing a change in residual helicity have a change in the binding affinity to their ordered partners but the sign of the changes is different.

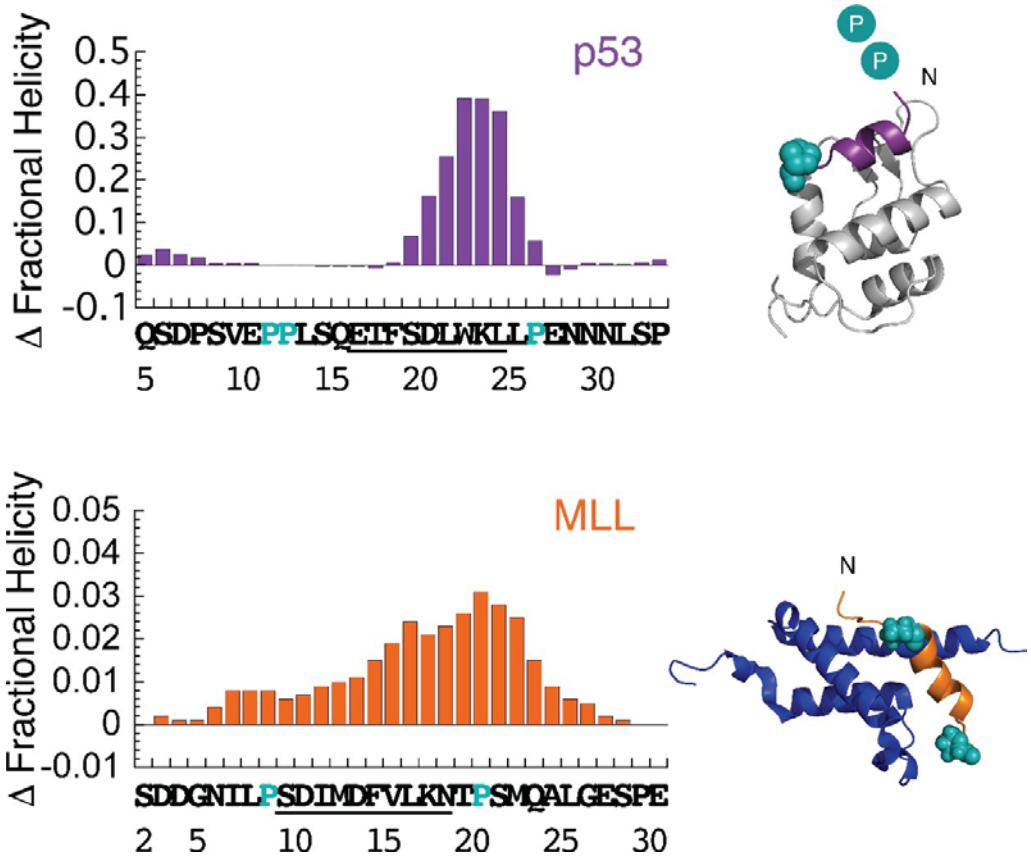


Figure 9: Fractional helicity plots of MLL and p53 showing the differences in residual helicity between WT and mutants. Figure in the top panel shows the change in residual helicity between p53 P27A and p53 WT while the bottom panel shows the change in residual helicity between MLL P9/21A and MLL WT. In the top figure, only the segment of the peptide showing a change in the residual helicity is shown. This figure is taken from "Conserved Helix Flanking Prolines Modulate Intrinsically Disordered Protein:Target Affinity by Altering the Lifetime of the Bound Complex," by Crabtree MD et.al., 2017, *Biochemistry*, 56, p. 2379. CC-BY. Reprinted with permission.

In p53 P27A, an increase in the residual helicity resulted in an increase in the binding affinity by 10-fold (**Figure 10, Table 1**). Conversely, in MLL, an increase in the residual helicity of P9/P21A showed a 25-fold reduction in the binding affinity between MLL and KIX. Data obtained from these experiments suggests that prolines alter the binding affinity of IDPs to their targets by regulating the residual helicity levels of IDPs in their free state, but it also implies that an increase in the residual helicity of IDPs doesn't always enhance their binding affinities.

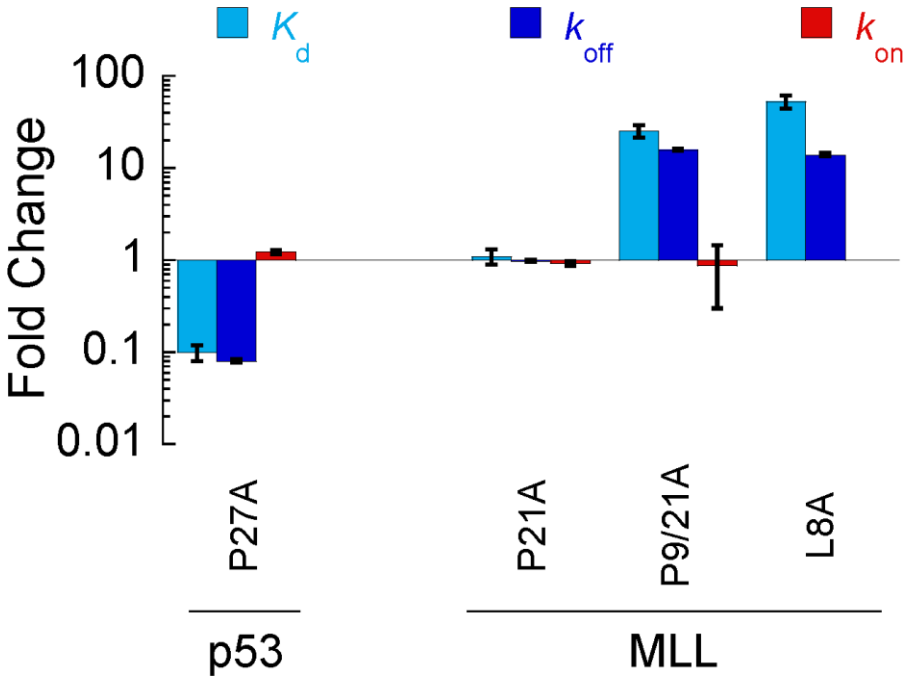


Figure 10: Fold change in the binding affinities, and association and dissociation rate constants of IDPs to their targets. Error bars for association and dissociation rate constants represent the error in the fit except for MLL double mutant where error bars denote the standard deviation of the mean (n=7). Error bars for Kd represent the standard deviation of the mean where n=3. This figure is taken from "Conserved Helix Flanking Prolines Modulate Intrinsically Disordered Protein:Target Affinity by Altering the Lifetime of the Bound Complex," by Crabtree MD et.al., 2017, *Biochemistry*, 56, p. 2379. CC-BY. Reprinted with permission.

A 1.4-fold increase in helicity levels in MLL led to 25-fold reduction in binding affinity. Is this change in binding affinity coming entirely from the change in the amount of helical content in free state MLL or by disrupting the binding interface between MLL and KIX? To answer this question, we aimed to mutate the leucine residue, which is immediately N-terminal to the proline at position 9, that sits in the hydrophobic pocket of KIX. This mutation (L8A) showed a 53-fold reduction in the binding affinity of MLL and KIX (**Figure 10**).

Prolines altering binding affinities in a different manner in each IDP emphasizes the need to study the mechanism behind regulation of binding affinities by levels of transient residual helicity. Stopped-flow fluorescence experiments were performed to analyse the kinetics of IDP:target binding.

2.2.3 Mechanism of Regulation

Binding kinetics experiments revealed a change in the dissociation rate constant in both the IDP-target systems. In p53 P27A mutant, association rate constant showed less than 1.2-fold reduction when compared to WT. Two dissociation rate constants were seen with P27A and the fastest of the two rates showed a 12-fold reduction when compared to wild type (**Figure 11**). In MLL P9/21A too, the association rate constant was unaffected while there was a 15.8-fold increase in the dissociation rate constant. The change in affinity in MLL L8A and KIX interaction corresponds to a 13.9-fold increase in the dissociation rate constant with no apparent changes in the association rate constant (**Figure 11, Table 1**). These studies suggest that the conserved helix flanking prolines regulate the binding affinities between IDPs and their targets, not by enhancing the rate of complex formation, but by altering the stability of the complex through dissociation rates.

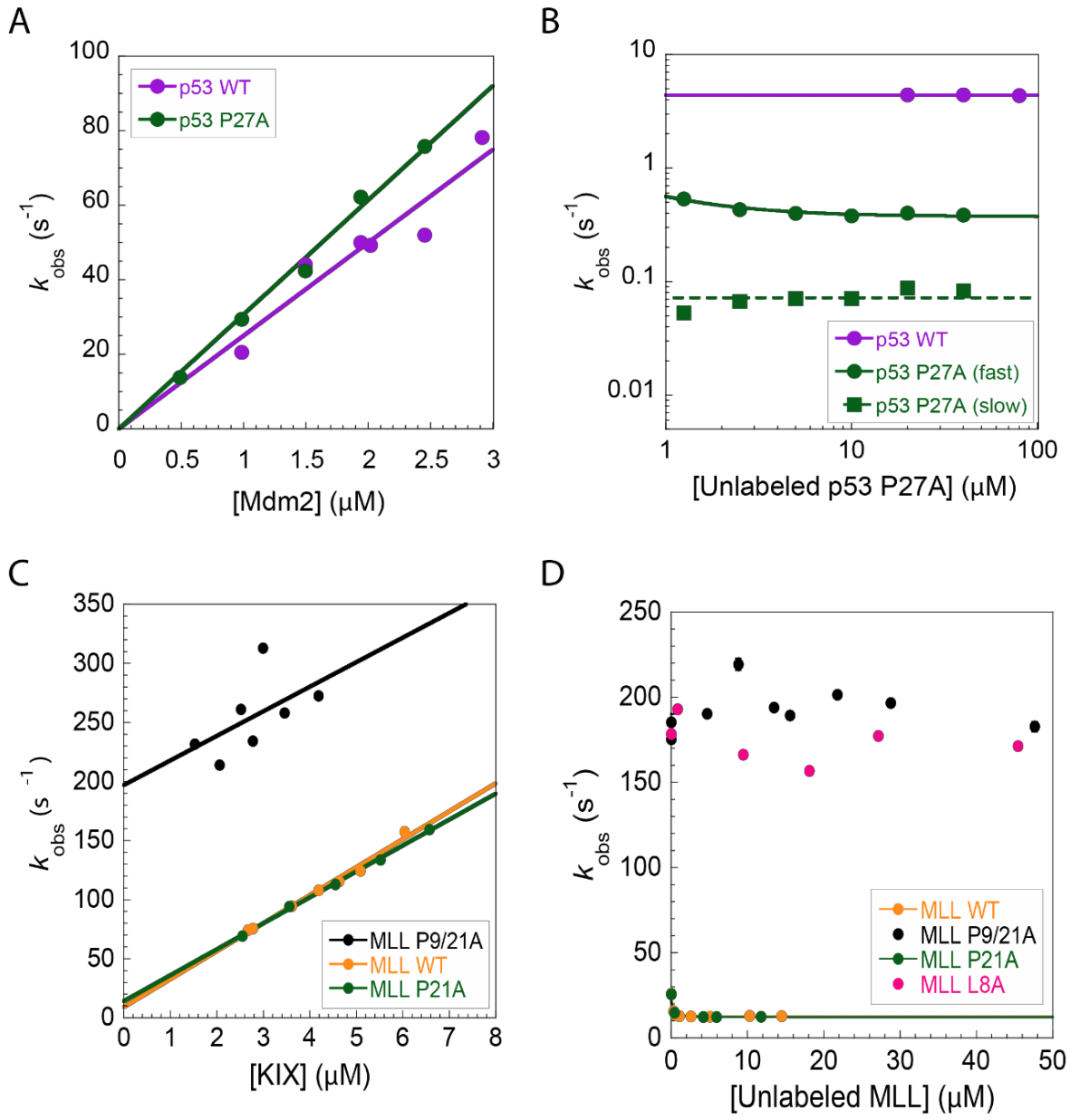


Figure 11: Binding kinetics studies of p53:MDM2 and MLL:KIX. A and C show observed association rate constants observed under pseudo-first order conditions of p53 and MDM2 respectively with an excess of MDM2 and KIX respectively. Association rate constant, k_{on} is obtained from a gradient of the straight-line fit. B and D show observed dissociation rate constants of p53 and MLL obtained from competition studies. This figure is taken from "Conserved Helix Flanking Prolines Modulate Intrinsically Disordered Protein:Target Affinity by Altering the Lifetime of the Bound Complex," by Crabtree MD et.al., 2017, *Biochemistry*, 56, p. 2379. CC-BY. Reprinted with permission.

2.3 Influence of helix flanking prolines on the binding of c-Myb to KIX

N- and C-terminal prolines in the transactivation domain of c-Myb at positions 289 and 316 were mutated to alanines, as described in section 5.6, resulting in two mutants, c-Myb P289A and c-Myb P289A P316A. Detailed description of the expression and purification of proteins is outlined in sections 5.1 and 5.2. Concentration of proteins was calculated using their extinction coefficients and the absorbance of U.V light at 280 nm. Protein samples were dialysed into suitable buffers before analysis according to the protocols described in chapter 5. Circular dichroism and nuclear magnetic resonance spectroscopy were used to determine the percentage of alpha helical content in mutant and WT c-Myb TAD peptides.

2.3.1 Prolines influence the transient helicity of c-Myb

Circular dichroism (CD) is a rapid technique that can be used to study the secondary structures of peptides and proteins. Secondary structures can be distinguished from each other by their unique shapes in a far U.V spectrum. Alpha helices show characteristic dips in molar ellipticity of circularly polarized light at 222 nm and 208 nm and peak at 196 nm. The magnitude of the dip at 222 nm is proportional to mean helicity of the peptide. Detailed description of the procedure for collecting and analysing CD spectra is outlined in section 5.3.

CD experiments were performed in a far U.V. range (190-260 nm) using Jasco J-815 spectropolarimeter at 25 ºC in a 0.1 mm path length cell. Proteins were dialysed by a million-fold in three steps and the experiments were conducted at a concentration of 1mg/ml. Mutation of prolines to alanines showed an increase in helicity in both the mutants. (**Figure 12**). Close observation of the curves reveals a greater dip in the mutants when compared to wild type indicating higher helicity levels. Molar residual

ellipticity and average helicity of the peptides were calculated as described in [171, 172] (**Table 2**). Mutating the N-terminal proline at position 289 to alanine increased the mean helicity by 1.28-fold. Mutating both the N- and C-terminal prolines increased the mean helicity by 1.35-fold when compared to WT.

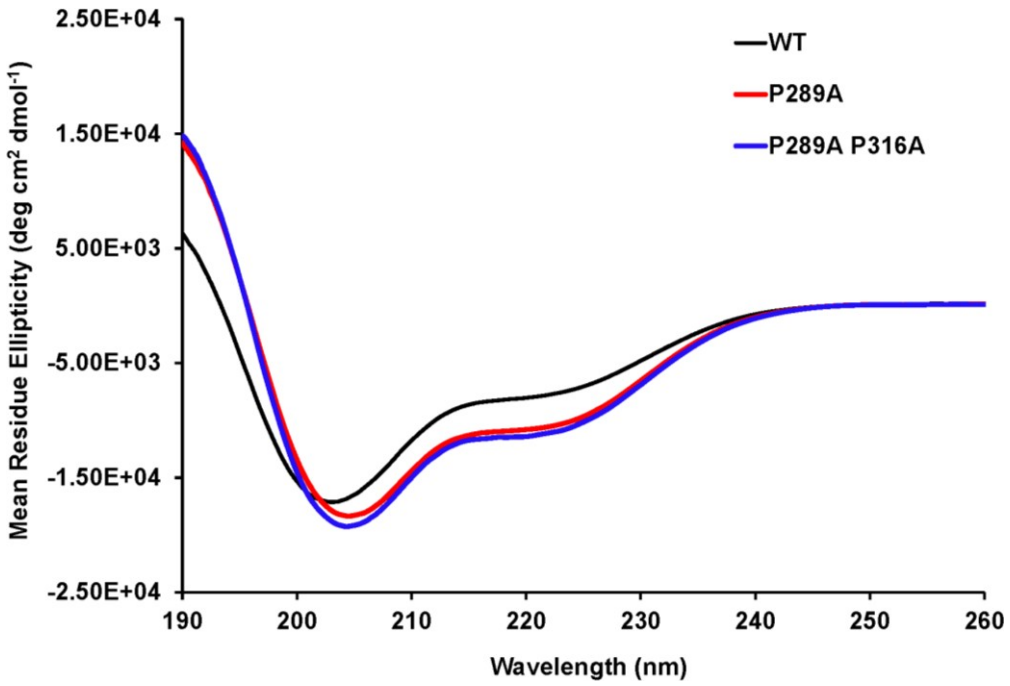


Figure 12: CD plot of WT and proline mutants of c-Myb. Molar residual ellipticities of all the peptides were plotted against wavelength. Black curve represents the molar ellipticities of WT c-Myb while red and blue represent the same for P289A and P289A P316A respectively. Molar ellipticities were obtained from the mean of the two CD accumulations collected using the same sample.

Although, CD is advantageous over other techniques in obtaining a quick measurement, it gives only a global measurement of the helicity levels of peptide. Residue specific information such as the residues involved in helix and the region of the peptide where the mutations caused changes cannot be determined from CD analysis. This emphasizes the need to perform NMR analysis, a technique which gives residue specific information of secondary structures.

2.3.2 Residue specific analysis of the transient helicity levels of c-Myb and the proline mutants

NMR experiments were performed for c-Myb WT, P289A and P289A P316A to obtain information on helicity at the residual level. NMR analysis gives information on the residues specifically affected by the mutation and the mechanism of increased helicity i.e., whether the increased helicity resulted from an increase in the amount of helix per residue or an elongation in the length of the helix. HSQC and 3D HNCACB and HNCO experiments were conducted on all three peptides at 25 ºC as described in section 5.4.

Peptide	% Helicity (CD)
WT	20.81
P289A	26.75
P289A P316A	28.22

Table 2: Percentage helicity of c-Myb WT and proline mutants. Above table shows the percentage helicity of c-Myb WT and mutants calculated from molar residual ellipticity values at 222 nm. % Helicity was calculated using molar ellipticities obtained from two CD accumulations collected using the same sample.

As seen in CD analysis, both the N-terminal and C-terminal proline mutants show an increase in the helicity level. Some residues could not be assigned which are indicated in the chemical shift tables in Appendix B. **Figure 13** shows alpha carbon secondary chemical shifts of WT and mutant c-Myb along with the δ2D curves. Secondary chemical shifts are an accurate measure to determine the secondary structures of peptides and proteins. Information from alpha carbon chemical shifts can be used to measure the helicity of peptides. Secondary chemical shifts were calculated

by subtracting the chemical shifts of random coils from the chemical shift data obtained from assignments. Secondary chemical shifts are used to calculate residual helicity percentages using δ2D analysis developed by Vendruscolo [173]. Mean helicity of each peptide was obtained from the sum of helicities of all the residues in the peptide. NMR analysis showed a 1.23-fold increase in the mean helicity when the N-terminal proline is mutated to alanine. When both the prolines were mutated to alanines, there was a 1.28-fold increase in the mean helicity (**Table 3**). Mean helicity values obtained from CD and NMR analysis complement each other (**Table 3**).

Peptide	% Helicity (CD)	% Helicity (NMR)
WT	20.81	21.8
P289A	26.75	26.9
P289A P316A	28.22	28

Table 3: Average helicity of c-Myb WT and proline mutants obtained from CD and NMR analysis. Helicity values obtained from CD are an average of two accumulations collected using the same sample.

Changes in helicity were observed at and around the sites of mutation in both the mutants. The helix is extended on the N- and C-terminal regions when respective prolines were mutated to alanines. NMR analysis also reveals that the region in WT c-Myb TAD that forms an alpha helix is highly helical in its free state i.e., most of the residues involved in the helix have a helical propensity greater than 75% (**Figure 13**). It must be noted that, though we see an increase in the mean helical content of the peptide, the amount of increased helical content at the residue level is less than two-fold.

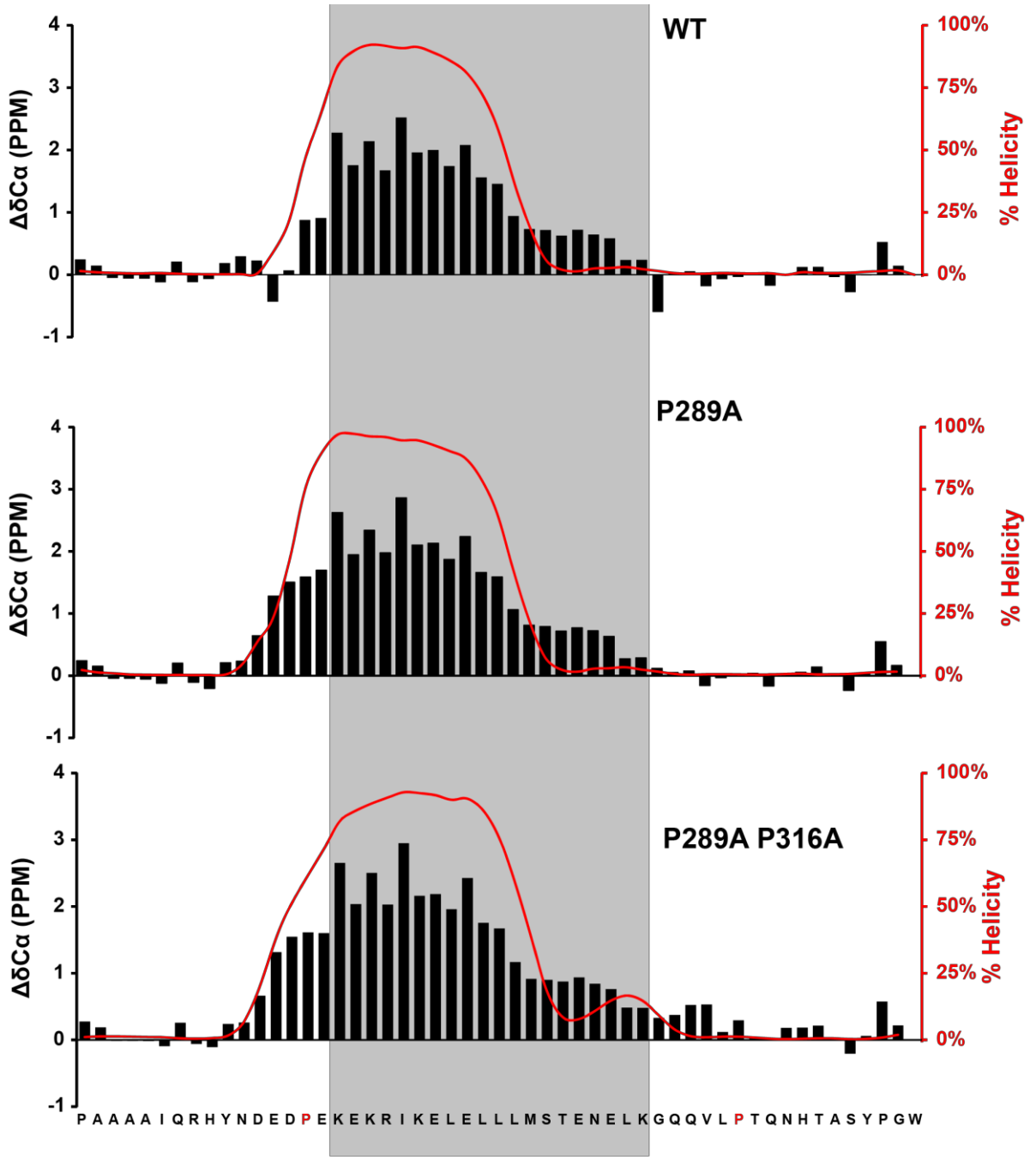


Figure 13: Alpha carbon secondary chemical shifts and δ2d analysis of c-Myb WT and proline mutants. Figure shows the α carbon secondary chemical shift deviations (ΔδCα, black bars) and the residual helical percentages (red smoothed lines) for WT and mutants of c-Myb TAD. X-axis shows the sequence used in all our studies. Prolines used in mutations are shown in red. Grey shaded region represents the region that binds to the KIX domain of CBP. Some assignments couldn't be made which is mentioned in Appendix B. Missing data points in residual helicity percentages are joined as a smooth curve.

2.3.3 Characterisation of Binding Affinity Between c-Myb and the KIX Domain of CBP

Isothermal titration calorimetry (ITC) was used to study the effect of increased helicity in c-Myb mutants. Data was collected with an excess of c-Myb at 25 ºC on a set of three samples prepared from the same culture preparation and average of the three runs is calculated (**Table 4**). WT and mutant c-Myb peptides were used at a concentration of 0.15 mM and KIX was used at a concentration of 15 µM. ITC is an accurate way of measuring binding affinity as it is extremely sensitive to a very small change in heat. It measures the heat utilized or released during a chemical reaction from which the enthalpy and association constants of the reaction can be deduced. Twenty-nine small injections of concentrated c-Myb were titrated into KIX in our experiments. Heat traces of all the injections were then integrated and a curve was plotted using a single-site binding model. Titration curves from one of the three runs of each peptide are shown in **Figure 15**.

The change in residual helicity in both the proline mutants showed almost no effect on their binding affinities to the KIX domain of CBP. Mutating N-terminal proline to alanine showed a 1.24-fold increase in the dissociation constant implying that the increase in helicity resulted in a reduction in the binding affinity between c-Myb TAD and the KIX domain of CBP (**Figure 15**). Mutating both the N- and C-terminal prolines to alanines resulted in a 1.06-fold increase in the dissociation constant which also implies that the increased helicity led to a reduction in binding affinity. But, these small changes in binding affinity (**Table 4**) fall under the error level of the instrument.

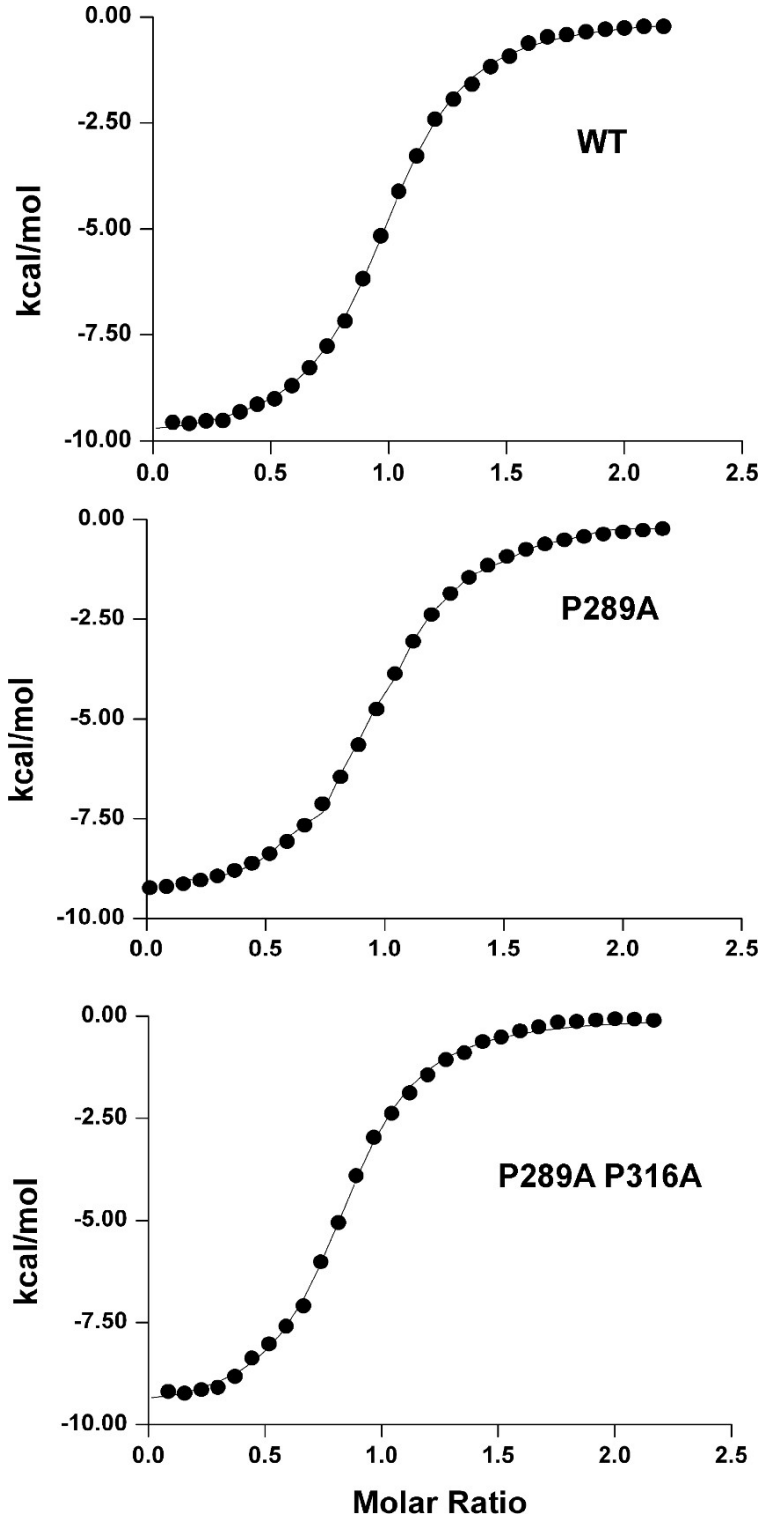


Figure 15: ITC curves of c-Myb WT and proline mutants. Y-axis denotes the heat of injection in Kcal/mol. Each solid circle on the plot represents each injection. X-axis represents the ratio of protein to ligand at the time each injection. Curves from one of the three runs with each peptide is shown.

Peptide	% Helicity (CD)	% Helicity (NMR)	K_d (μM)
WT	20.81	21.8	0.5 ± 0.03
P289A	26.75	26.9	0.62 ± 0.02
P289A P316A	28.22	28	0.53 ± 0.02

Table 4: Average helicity and binding affinity of c-Myb WT and proline mutants. Combined table showing mean helicity of peptides obtained from CD and NMR experiments. K_d is calculated from the association constant obtained from ITC. Error represents the standard deviation of mean from three ITC runs.

CHAPTER 3

CHARACTERISING THE RELATIONSHIP BETWEEN DISORDER LEVELS AND BINDING AFFINITY IN c-Myb TAD

3.1 Rationale

Mutations of conserved helix flanking prolines did not show any effect on the binding affinity of c-Myb to KIX but showed an effect in the binding affinities of p53 and MLL to their respective protein partners. Do levels of disorder in the free state of IDP regulate binding in all IDPs? Does c-Myb TAD fall under such a category? Is there a helical threshold for energetic penalty in coupled folding and binding? To answer these questions, we developed a model of c-Myb TAD – KIX where we designed mutations in c-Myb TAD with levels of transient helicity that are less than WT with the goal of reducing the transient helicity to a point where we observe a reduction in binding affinity. Since these mutations were made in the region of c-Myb helix that binds to KIX, care was taken to ensure we did not mutate any residues that directly contact KIX. Binding experiments were performed on the mutants to study the relationship between disorder levels of c-Myb TAD and its binding affinity to the KIX domain of CBP.

3.2 Examining the Binding Interface Between c-Myb and KIX

To accurately measure the effect of disorder levels on the binding affinity, it is important to ensure that the mutants created are not largely different in their structures

from WT. Mutations must be created in such a way that the overall charge and structure of the peptide is preserved. To account for the change in binding affinity entirely to the change in disorder levels in free state, it is also crucial to ensure that the binding interface is not altered. To satisfy all the above-mentioned conditions, the structure of the complex between c-Myb TAD and the KIX domain of CBP was examined and sites that are not involved in the binding were chosen for mutagenesis (**Figure 16**). Solution structure of the KIX domain of CBP bound to mouse c-Myb TAD that is deposited in Protein Data Bank is taken as a reference for analysis [174]. Eight residues that don't interact with KIX residues in the binding were found in the helical region of the transactivation domain of c-Myb. From the eight available sites, four sites were selected based on the reductions in helicities predicted by the Agadir software which is described in detail in the next section. Among these four residues, two of them are on the N-terminal side of the helix (E292, K293), one residue is closer to the N-terminal region (K296) and one residue is right in the middle of the helix (L300) and is adjacent to the leucine that sits in the hydrophobic pocket of KIX.

3.3 Designing Mutations in Solvent Exposed Sites

Residues at positions 292 (E), 293 (K), 296 (K) and 300 (L) were selected for mutations. Change in helicity was predicted with several conservative amino acid substitutions in the four locations using Agadir software. Agadir is an algorithm used to predict alpha helical content in peptides. pH and temperature were set to our experimental conditions and Agadir was run with several amino acid substitutions. From the Agadir data obtained, mutations with significant change in helicity were picked and mutants were made using site directed mutagenesis as described in section 5.6.

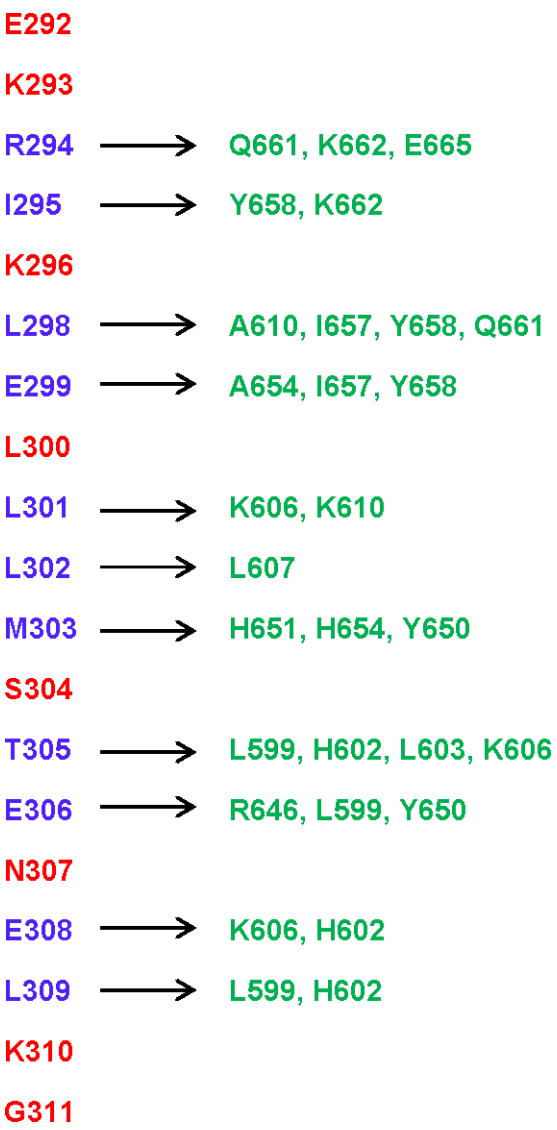


Figure 16: Amino acid interactions between c-Myb TAD and the KIX domain of CBP. Residues in the left, highlighted in blue represent amino acid residues of c-Myb TAD that form an alpha helix when bound to KIX. Residues to the right, highlighted in green are KIX residues that are involved in binding. Arrows indicate the residues interacting with each other. Type of interaction is not shown in the picture. Residues highlighted in red are the residues in the helical region that do not participate in binding.

As the residual helical content of c-Myb TAD in the helical region is already high in its free state, only mutations with significant predicted reduction in helicity were chosen for experimental studies. Mutating the residues with a goal of helicity enhancement is not advantageous since these mutations might not enhance the

helicity of already highly helical c-Myb TAD to a significant extent that can show an impact on the binding affinity between c-Myb TAD and KIX. Despite this fact, one mutation that is predicted to enhances the helicity of free state c-Myb TAD was used for analysis which is described below. Mutations, E292D, K293H, K296H and L300V, were chosen for our studies based on the predicted helical propensities by Agadir software (**Table 5**). Leucine at position 300 was also mutated with proline and glycine, known helix breakers, to obtain a mutant with promising reduction in helicity. Leucine at position 300, being in the middle of the helix should show a greater impact on modulating the residual helicity of free state IDP. Apart from all these mutations, the G311A mutant was also made. This glycine to alanine mutation, as predicted, should enhance the helicity of c-Myb TAD in its free state. Glycine at position 311, being closer to the binding segment when compared to the proline to alanine mutation at position 316, is a good candidate to study the effect of enhanced helicity on binding affinity.

Peptide	% Helicity (Agadir)
G311A	19.29
WT	19.23
E292D	13.45
K293H	11.01
K296H	7.94
L300V	15.48
L300G	8.12
L300P	4.74

Table 5: Percentage helical prediction of single amino acid substitutions in c-Myb TAD using Agadir. Agadir calculates the mean percentage helicity from the residual helical propensities that it predicts. Amino acids substitutions were also picked based on the amount of reduction/change in helicity observed in the helical region of predictions. Percentage helicity of WT as predicted by Agadir is also shown.

3.4 Modulating Levels of Transient Helicity in c-Myb TAD

3.4.1 Analysis of Alpha Helical Content by Circular Dichroism

Circular dichroism was first used to study and analyse the structures of mutants as it is economical, and results can be obtained in a short period of time. NMR spectroscopy was then performed on the mutants that showed a high change in helicity levels to investigate the residual helicity changes in each one of them. Circular dichroism was performed in the same way as proline mutants. Detailed outline of the procedure is described in section 5.3. Mutants were expressed and purified in the same way as WT c-Myb although the induction times for each of the mutants varied which is outlined in section 5.2. From circular dichroism analysis, it was observed that mutants show changes in helicity levels as predicted by Agadir, but in the changes in mean helicity of mutants are not as pronounced as Agadir predicted (**Table 6**).

Peptide	% Helicity (Agadir)	% Helicity (CD)
G311A	19.29	24.31
WT	19.23	20.81
E292D	13.45	20.53
K293H	11.01	16.28
K296H	7.94	19.38
L300V	15.48	17.39
L300G	8.12	14.21
L300P	4.74	7.77

Table 6: Differences in helicity percentage levels as predicted by Agadir and as experimentally observed by CD. Percentage helicity of observed and predicted WT and mutant c-Myb TADs are shown in the table. Although, Agadir over-estimated the reduction in helicity, it follows the general trend as experimental values, as shown in the table.

Change in helicity level of E292D mutant was almost none while L300P and L300G showed substantial reductions in the levels of transient helicity (**Figure 17 and Figure 18**). L300V also revealed a strong reduction in helicity levels, although it is not as pronounced as the other leucine mutants. Reduction of helicity in K293H and K296H is not in compliance with the rest of mutants and we think this could be due to the isomerization of histidine residues at the pH used. Further analysis should be done on these mutants at varying pH levels to study the pH dependence.

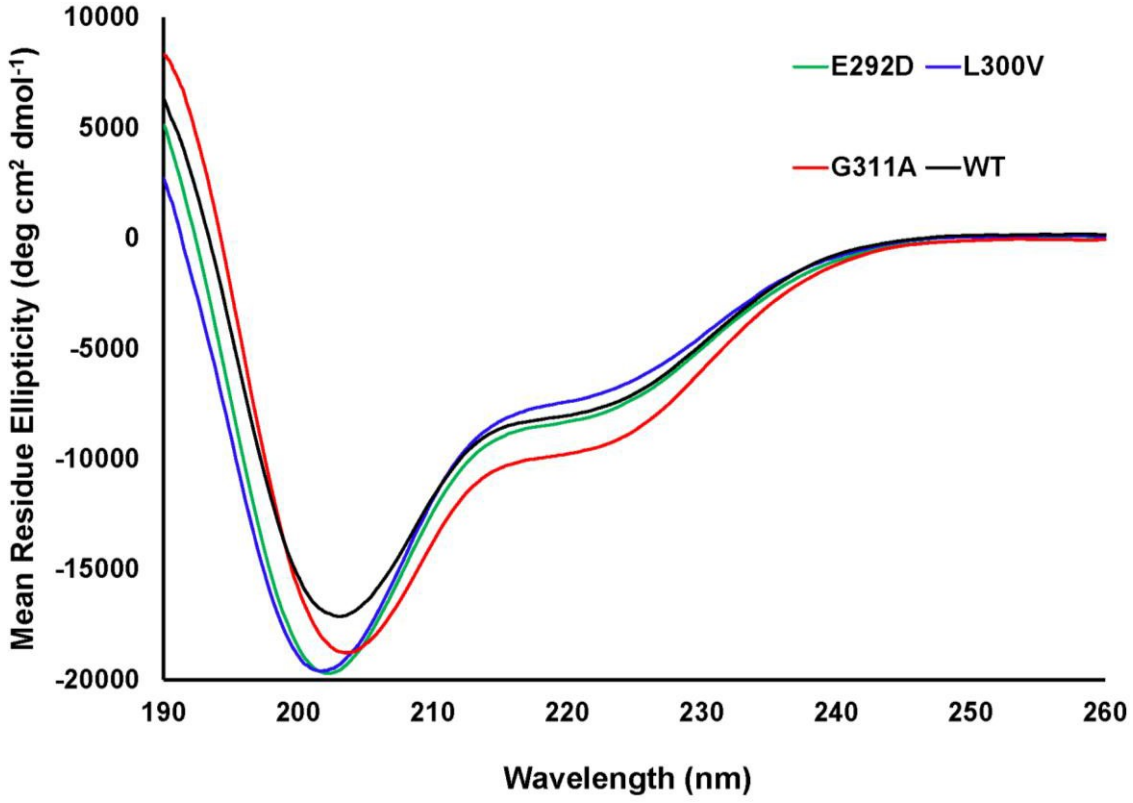


Figure 17: CD plot of c-Myb mutants and WT. Curve in black denotes the molar ellipticity of WT c-Myb while red, green and blue represent molar residue ellipticity of mutants in the decreasing order of helicity levels i.e., G311A, E292D and L300V respectively. Data for K293H and K296H are not shown in the plot.

Based on the NMR and CD data obtained for proline mutants described in chapter 2, information obtained from CD can be relied upon. Plots are separated in two graphs to better notice the changes in curves that are close (**Figure 17 and Figure 18**). Figure 17 shows a dip in the G311A curve at 222 nm when compared to WT indicating that the helicity level is increased. Calculating the mean helicities from molar residual ellipticities revealed a 1.19-fold increase in the helicity level for G311A. E292D and L300V shows reductions in the dip at 222 nm indicating a reduction in the helicity levels. Mena helicity calculations revealed a 1.19-fold reduction in the helicity in L300Vmutant while there is almost no change in the helicity level in E292D. K293H and K296H showed 1.26- and 1.07-fold reductions in the helicity levels respectively when compared to WT(**Table 6**). Curves in **figure 17** and values from **table 6** suggest that these conservative mutations have affected the residual helicity levels but not at a level where we can see an impact on their binding affinities to the KIX domain of CBP.

L300P and L300G curves from **figure 18** show a reduction in the dip at 222 nm when compared to WT thus implying that the levels of transient helicity were reduced. In the case of L300P, the curve is almost flat at 222 nm. Careful observation of the curves reveals a shift in the dips of L300P and L300G from 208 nm to the left, shifting towards the characteristic curve of a random coil. This suggests that the transient helicity levels in these mutants were disrupted on a greater level. From the mean helicity level calculations, we noticed a 1.46-fold and 2.67-fold reductions in the helicity levels of L300G and L300P respectively. NMR analysis was performed on the mutants that showed outstanding reduction in helicity levels i.e., L300P and L300G to characterize the changes in helicity at residual level.

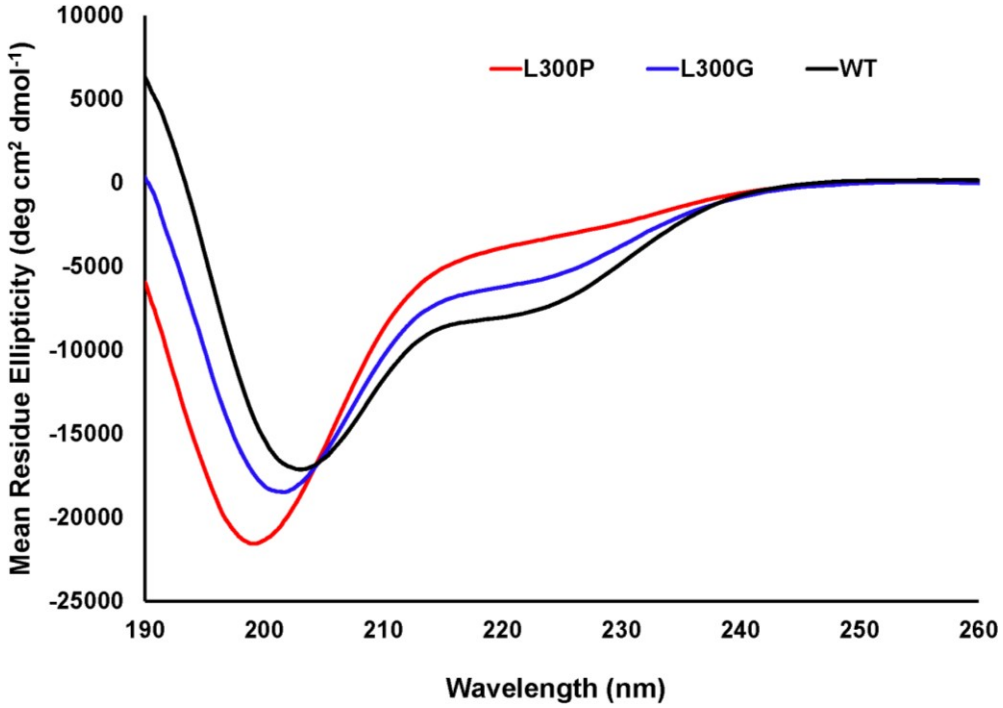


Figure 18: CD plot of WT c-Myb and leucine mutants. Red curve indicates L300P while blue curve represents L300G and black curve represents WT molar residual ellipticity data.

3.4.2 Characterizing the Alpha Helical Content at Residual Level

NMR analysis showed similar reductions in the mean helicity levels as seen with CD (**Figure 19**). E292D shows a 1.18-fold reduction in the mean helicity while L300G and L300P showed 1.57- and 3.96-fold reductions in the mean helicity levels respectively. **Table 7** shows a comparison of mean helicity data obtained from Agadir and experimental analysis. Examining the helicity curves from **figure 19** reveals a reduction in helicity at and around the site of mutation, at least until three residues on both the sides of mutation. Leucine at position 300, being in the middle of the helix showed pronounced reductions in the levels of helicity at and around the site of mutations when mutated to proline and glycine. In L300G, the helix was disrupted until seven residues N-terminal to the site of mutation and a couple of residues on the C-

terminal side of the mutation. A 25% reduction in the peak helicity was observed. In L300P, helix was disrupted entirely around the site of the mutation, until three residues on both the sides of the mutation which includes the leucine residue that sits in the hydrophobic pocket of KIX. It also reduced the helicity levels until seven residues N-terminal to the site of mutation reducing the peak helicity by almost half.

Peptide	% Helicity (Agadir)	% Helicity (CD)	% Helicity (NMR)
G311A	19.29	24.31	n.d
WT	19.23	20.81	21.8
E292D	13.45	20.53	19.2
L300V	15.48	17.39	n.d
L300G	8.12	14.21	13.8
L300P	4.74	7.77	5.5

Table 7: Comparison of mean percentage helicities of WT and mutant c-Myb obtained from Agadir, CD and NMR. Table shows the mean percentage helicity values of WT and mutant c-Myb peptides as predicted from Agadir and experimentally verified using CD and NMR analysis. n.d represents not determined.

3.5 Characterization of binding affinity of mutant c-Myb to the KIX domain of CBP

Binding affinities were measured on the mutants created except for the lysine to histidine mutants. Experimental conditions were kept the same as proline mutations. Since the leucine mutations showed a greater level of reduction in helicity, we expected a weak binding between c-Myb TAD and KIX with these mutants. Hence, the concentrations of proteins used for binding experiments was increased when compared to wild type to obtain a decent signal.

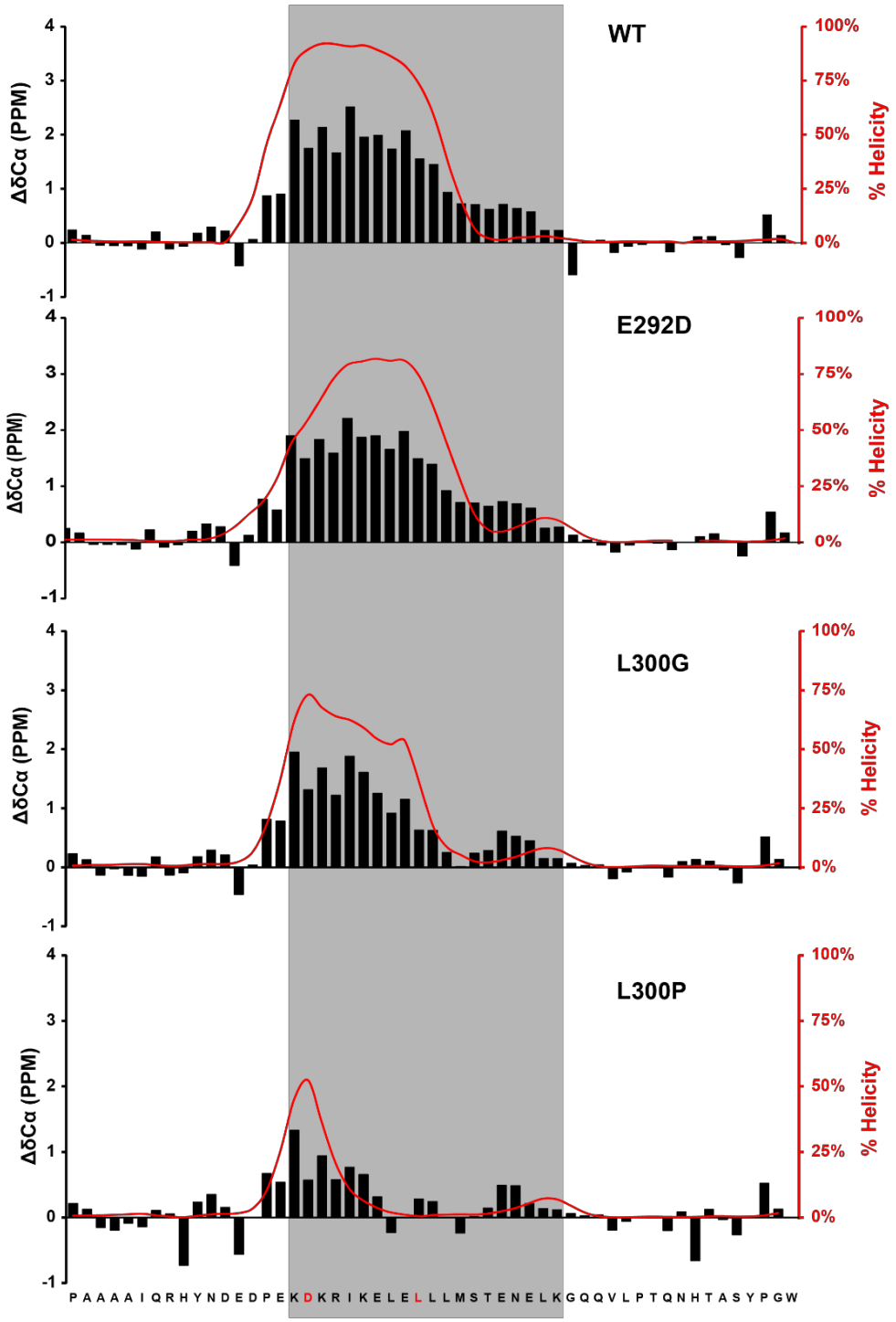


Figure 19: Alpha carbon secondary chemical shifts and δ2D plots of Leucine mutants and E29D. Figure shows the α carbon secondary chemical shift deviations (ΔδCα, black bars) and the residual helical percentages (red smoothed lines) for WT and mutants of c-Myb TAD. X-axis shows the sequence used in all our studies. Sites of mutations are shown in red. Grey shaded region represents the region that binds to the KIX domain of CBP. Some assignments couldn't be made which is mentioned in Appendix B. Missing data points in residual helicity percentages are joined as a smooth curve.

L300G binding was measured at a five-fold increase in the concentrations of protein and ligand while L300P was measured at a ten-fold increase in the concentrations. There was a 1.26- and 1.64-fold increase in the dissociation constants when E292D and L300V were titrated into KIX respectively implying the binding affinity is reduced in both the cases (**Figure 20 and Table 8**). We noticed a 1.08-fold reduction in the dissociation constant when G311A was titrated into KIX which suggests that the increased helicity in G311A resulted in an increased binding affinity between c-Myb TAD and the KIX domain of CBP. Introducing glycine at position 300 further weakened the binding affinity which was explained by a 4.5-fold increase in the dissociation constant suggesting a weak binding between c-Myb TAD and KIX. Binding affinity between L300P and KIX was very small, marked by a 600-fold reduction in the binding affinity, implying that the binding between c-Myb TAD and KIX was almost entirely abolished when a proline was introduced in the middle of the helix (**Figure 20**). L300P binding was repeated a couple of times with a sample size of three and the results were consistent.

Peptide	% Helicity (CD)	% Helicity (NMR)	K_d (μM)
G311A	24.31	n.d	0.46 ± 0.03
WT	20.81	21.8	0.5 ± 0.03
E292D	20.53	19.2	0.63 ± 0.02
L300V	17.39	n.d	0.82 ± 0.05
L300G	14.21	13.8	2.24 ± 0.02
L300P	7.77	5.5	290 ± 101

Table 8: Comparison of helicity and binding affinities of WT and mutant c-Mybs. Observed helicity percentages obtained from both techniques are shown in the table. n.d denotes not determined.

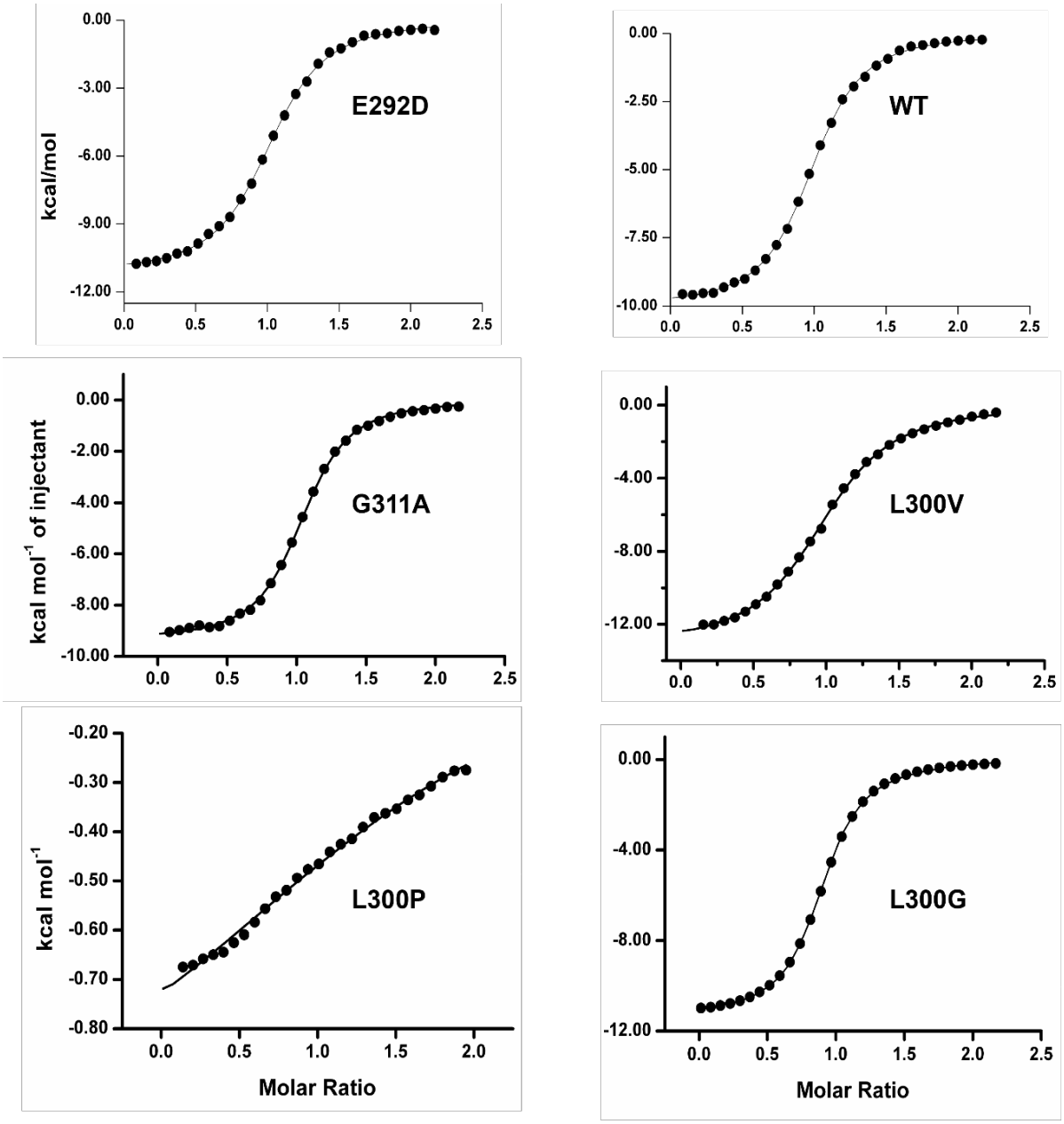


Figure 20: Binding curves of c-Myb conservative and leucine mutants. WT curve is shown for a comparison. All the curves were fit using a single site binding model. Curves represent data from a single run. Each sample was run with three runs.

Mean association constants measured directly by the instrument were used to calculate dissociation constants shown in **Table 8**. Standard deviation was obtained from dissociation constant values from three sample runs. L300P shows a high error

in the binding affinity but catering to the low/no binding affinity as seen in the curve, and the error of the instrument, such high errors are expected. ΔG values were calculated from association constants obtained from the experiment using the formula:

$$\Delta G = -RT\ln K_a$$

where, R is the gas constant, T is the temperature, in degrees Kelvin, at which the experiment was performed, and K_a is the association constant. **Table 9** shows the list of all c-Myb mutants and WT along with their helicities, binding affinities and ΔG values.

Peptide	% Helicity (CD)	% Helicity (NMR)	K_d (μM)	ΔG Kcal/mol
P289A P316A	28.22	28	0.53 ± 0.02	-8.56
P289A	26.75	26.9	0.62 ± 0.02	-8.46
G311A	24.31	n.d	0.46 ± 0.03	-8.64
WT	20.81	21.8	0.5 ± 0.03	-8.6
E292D	20.53	19.2	0.63 ± 0.02	-8.46
L300V	17.39	n.d	0.82 ± 0.05	-8.3
L300G	14.21	13.8	2.24 ± 0.02	-7.7
L300P	7.77	5.5	290 ± 101	-5

Table 9: Comparison of helicities and binding affinities of WT and all mutant c-Mybs. Observed helicity percentages obtained from both techniques are shown in the table. n.d denotes not determined. Dissociation constants were calculated as a mean from association constants obtained from three sample runs. Error represents the standard deviation of the mean. ΔG values were calculated from association constants.

CHAPTER 4

DISCUSSION

4.1 Effect of Prolines on Residual Helicity and Binding Affinity in p53:MDM2 and MLL:KIX systems

The effect of prolines on the residual helicity of free state IDPs was studied using nuclear magnetic resonance spectroscopy. Mutating helix flanking prolines to alanines resulted in an increase in the residual helicity in both systems. We observed a 1.5-fold increase in the mean helicity in MLL while there was a two-fold increase in the mean helicity in p53, when the helix flanking prolines were mutated to alanines. The effect of increased helicity on the binding affinity to their ordered partners was studied using equilibrium binding experiments and stopped-flow kinetics. An increase in the helicity of free state IDP is expected to result in an increase in the binding affinity of the bound complex. As expected, we noticed a 10-fold increase in the binding affinity of p53 and MDM2 when the C-terminal proline is mutated. A one percent increase in the helicity of free state MLL, that resulted from mutating N-terminal proline to alanine, led to a 25-fold reduction in the binding affinity between MLL and KIX, opposite of what we expect.

The mechanism behind changes in binding affinity in p53:MDM2 and KIX:MLL complexes was studied using stopped-flow kinetics. A ten-fold reduction in the dissociation rate constant was observed when the C-terminal proline was mutated to alanine in p53. We noticed a 15.8-fold increase in the dissociate rate constant when

N-terminal proline is mutated to alanine in MLL. Kinetic studies revealed that the changes in binding affinity are caused by changes in dissociation rate constants but not association rate constants implying that the influence of conserved helix flanking prolines is not on the rate of complex formation but on the stability of the bound complex which indicates that in both these cases, the mechanism of binding is induced-fit.

A one-percent increase in helicity in MLL corresponding to a 25-fold reduction in the binding affinity between MLL and KIX did not seem real. This huge reduction in binding affinity might be explained by the loss of proline side chain interactions with KIX caused by the mutation. We questioned if the huge reduction in binding affinity is coming solely from the loss of proline side chain interaction or from the disruption of binding interface. Examining the bound structure of MLL and KIX revealed a leucine, immediately N-terminal to the proline, that packs into a hydrophobic pocket in KIX (**Figure 8**). Mutating this leucine to alanine (L8A) showed a 53-fold reduction in the binding affinity between MLL and KIX (**Table 1**). A 13.9-fold increase in the dissociation rate constant was observed in MLL L8A – KIX complex which we think is majorly contributing to the 15.8-fold reduction in the dissociation rate constant in P9/21A mutation. These results indicate that the 25-fold reduction in binding affinity seen in P9/21A mutation is predominantly coming from the disruption of the interaction between leucine and KIX. Prolines are known to influence the conformation of residues preceding them. This suggests that proline to alanine mutation in MLL disrupted its binding affinity to KIX not by altering the amount the helix but by changing the direction of the polypeptide.

4.2 Effect of Conserved Helix Flanking Prolines in the c-Myb Transactivation Domain on the Binding Affinity Between c-Myb and the KIX Domain of CBP

Effect of conserved helix flanking prolines on the helicity of the free state c-Myb was studied using circular dichroism and nuclear magnetic resonance spectroscopy. Mutating N- and C-terminal prolines resulted in a 1.2- and 1.3-fold increase in the helicity levels as observed by circular dichroism. NMR spectroscopy revealed changes in helicity levels at and around the sites of mutations. Mean helicity levels from CD and NMR spectroscopy complement each other. NMR spectroscopy revealed that the free state of c-Myb is highly helical i.e., the residual helicities of most of the amino acid residues in the helical region range between 80 – 90%, the peak residual helicity being 92.1%. Increase in helicity levels showed almost no effect on the binding affinity between c-Myb TAD and the KIX domain of CBP. An increase in the helicity levels not showing any effect on the binding affinity between an already highly helical c-Myb and its ordered protein KIX might indicate that there is a helical threshold level beyond which any changes in the helicity levels do not influence the binding affinity between IDP and its ordered protein partner. If such a threshold level exists, we think that WT c-Myb TAD is above this threshold level while WT p53 is below this apparent threshold level.

4.3 Effect of solvent exposed amino acid substitutions in c-Myb TAD on its Binding Affinity to the KIX Domain of CBP

Solvent exposed amino acid sites in c-Myb TAD were utilized for mutations to reduce its overall helicity. Conservatives mutations in the solvent exposed sites were made at positions 292, 293, 296 and 300 resulting in mutations E292D, K293H, K296H and L300V. All these conservative mutations showed slight reductions in the helicity

levels which in turn showed very small reductions in the binding affinity between cMyb TAD and KIX. CD analysis revealed almost no change in the helicity levels with E292D mutation while L300V mutation showed a 1.2-fold reduction in helicity levels. Although, lysine to histidine mutations at positions 293 and 296 showed reduction in the helicity levels when compared to WT, they do not follow the same trend with Agadir predictions as our other mutations. Preliminary studies indicated that this might be resulting from ionization of histidine at our experimental pH levels. Binding experiments were not performed on these mutants. E292D mutation showed a 1.26-fold reduction in the binding affinity while L300V mutation showed a 1.3-fold reduction.

Conservative mutations showed a broad range in the reduction of helicity but did not show a significant reduction in the binding affinity between c-Myb TAD and KIX. We aimed to reduce the helical content further to go below the helical threshold we hypothesized by introducing helix-breaking amino acid residues in the middle of the helix. To achieve this, leucine at position 300 was mutated to either glycine or proline. A 1.5-fold reduction in the helicity level was observed from CD and NMR studies when leucine at position 300 was mutated to glycine. Mutating leucine at position 300 to proline resulted in a 2.6-fold and 3.96-fold reductions in the helicity levels obtained from CD and NMR studies, respectively. NMR studies revealed a reduction in the helicity levels at and around the site of mutation in L300G while in L300P mutation the helix is completely disrupted at the site of mutation along with reducing the helical content on the N-terminal side of the mutation. Isothermal titration calorimetry experiments showed a four-fold reduction in the binding affinity between c-Myb L300G and KIX when compared to WT-KIX complex binding affinity. A very weak binding, marked by a 600-fold reduction in binding affinity, was observed between L300P and KIX. Introduction of helix-breaking amino acid residues like

proline and glycine might also change the inherent structure of the peptide. Hence, it is important to make sure that the weak binding affinities in L300P and L300G mutations are not arising from a disrupted binding interface between c-Myb TAD and KIX but solely by changes in the residual helicity levels of free state c-Myb. Chemical shift mapping experiments must be performed with KIX bound to WT and mutant c-Myb to determine whether the binding interface between c-Myb TAD and KIX is disrupted by introducing proline/glycine.

4.4 Conclusion

We think that the helix flanking prolines regulate the binding affinity of IDPs to their ordered proteins by regulating the disorder levels of IDPs in their free state, but the mechanism of regulation might differ in each IDP. In the case of p53, conserved helix flanking prolines regulate the binding affinity between p53 and MDM2 by tuning the amount of helical content in the region that forms helix. This is explained by an increase in the amount of helix caused by proline to alanine mutations. In the case of MLL, conserved helix flanking prolines regulate the binding affinity between MLL and KIX by holding the direction of polypeptide such that the leucine preceding the N-terminal proline packs in the hydrophobic pocket of KIX. Conserved helix-flanking prolines in an already highly helical cMyb TAD do not have any effect on its binding affinity to the KIX domain of CBP. Reducing the helicity of cMyb TAD over a broad range did not show a high reduction in the binding affinity but introducing either a proline or glycine showed pronounced effects on the binding affinity between c-Myb TAD and KIX (**Figure 22**). Our results indicate that there might be a helical threshold level for energetic penalty in coupled and binding and we think introducing proline/glycine in c-Myb TAD reduced its helicity levels to below this apparent helical

threshold level. Helix-coil-transition theory further explains our results. Formation of the first H-bond in an alpha helix requires the immobilisation of four amino acid residues but the formation of subsequent H-bond interactions require the immobilisation of only one amino acid residue per each H-bond interaction. We think introducing proline/glycine in the middle of the helix in c-Myb TAD disrupted the formation of the first H-bond destabilizing the helix and thus resulting in a weak binding affinity.

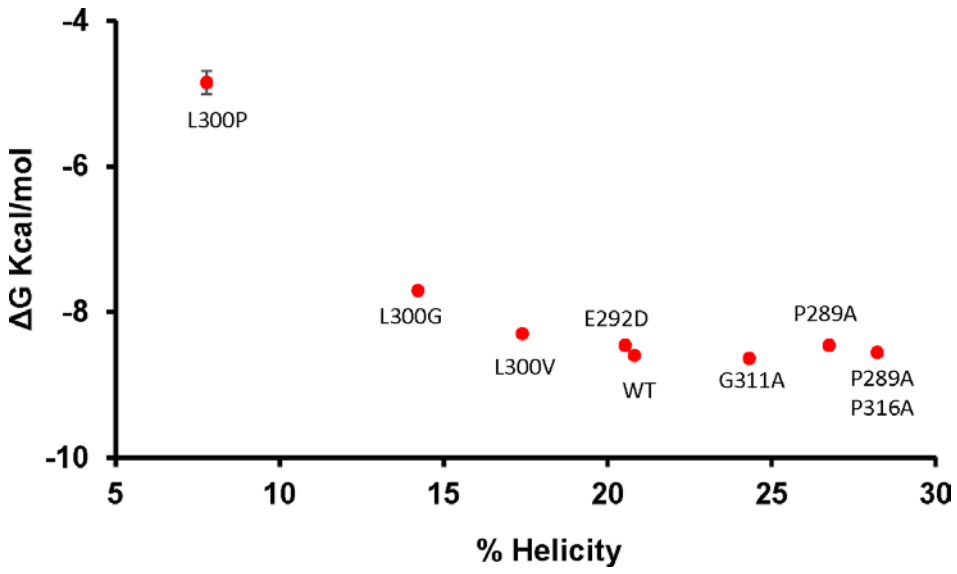


Figure 21: Correlation plot of % Helicity and binding affinity of WT and mutant c-Myb peptides. % Helicity values obtained from CD and binding affinities from ITC are drawn as a plot to show the relation between levels of disorder and binding affinity between c-Myb and KIX.

CHAPTER 5

MATERIALS AND METHODS

5.1 KIX Sample Preparation and Purification Scheme

5.1.1 Synthesis and Sub-Cloning

The human KIX gene (residues 586-672), synthesized from Eurofins, was cloned into pET-28a vector. The vector is selected for kanamycin resistance and consists of T7 promoter which is followed by a six-histidine tag and a thrombin cleavage site (**Figure 1**). KIX plasmid was inserted downstream of T7 promoter in the multiple cloning site between BamHI and XhoI restriction sites.

5.1.2 Transformations

KIX plasmid was transformed into BL21 (DE3) competent *E. coli* cells using the following protocol. BL21 (DE3) cells contain the gene for T7 RNA polymerase under the control of lacUV5 promoter. Therefore, T7 polymerase can be induced by lactose or Isopropyl β-D-1-thiogalactopyranoside (IPTG) allowing maximal expression of the genes cloned downstream of T7 promoter.

1. Gently mix 10 µL of BL21(DE3) cells and 1uL of 1 ng/µL KIX plasmid in a centrifuge tube.
2. Incubate on ice for 10 minutes.
3. Heat shock at 42 ºC for 30 seconds.

4. Incubate on ice for 2 minutes.
5. Add 100 µL of SOC and incubate in the shaker incubator at 37 ºC and 200 RPM for 1 hour.
6. Plate cells on a kanamycin resistant plate and incubate at 37 ºC overnight.
7. For a negative control, take 10 µL of cells and repeat steps 2 through 6.
8. Store the plates in the refrigerator the next morning until use.

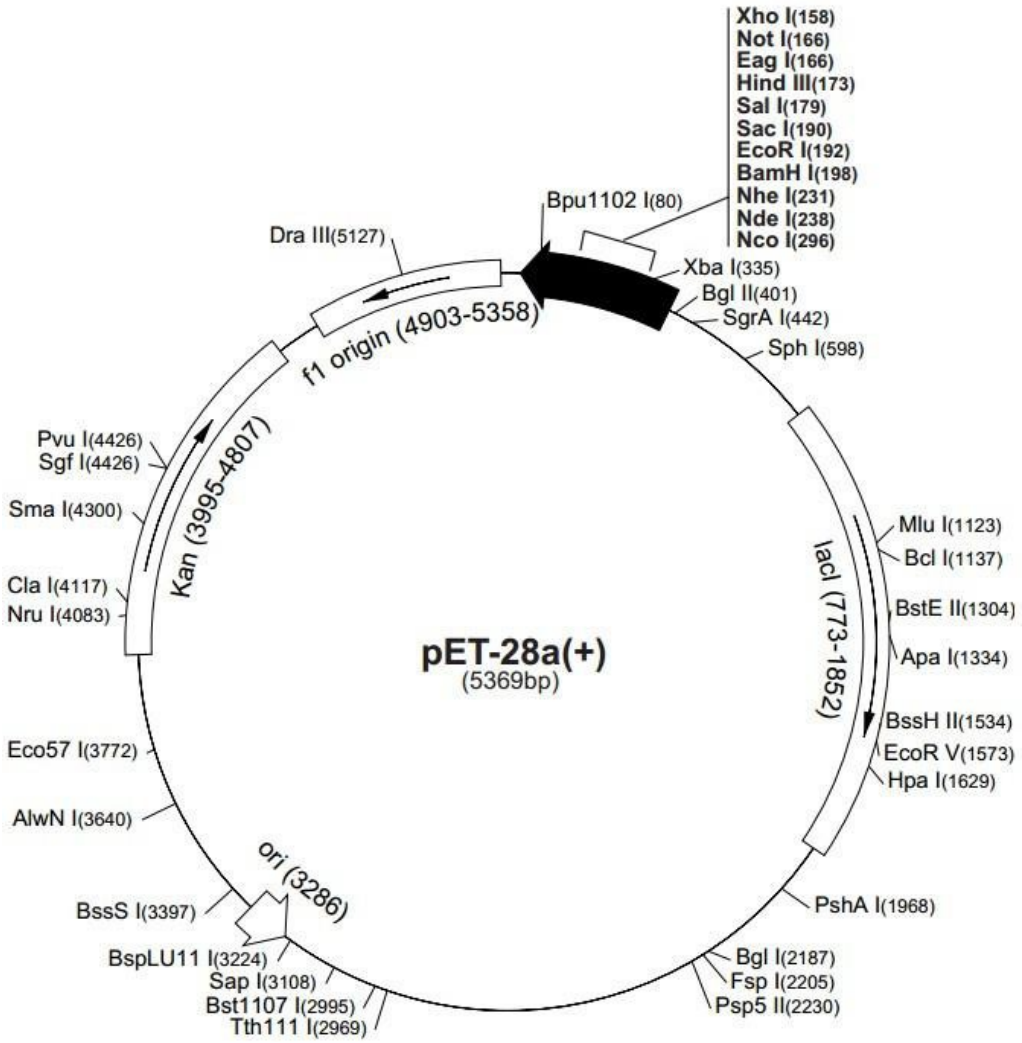


Figure 22: pET-28a Vector map. The gene was codon optimized for *E. coli* and inserted between Xho I and BamH I sites. Image is taken from Novagen.

5.1.3 Expression

Transformed *E. coli* colonies containing KIX plasmid that are not more than seven days old were grown in minimal medium (M9 medium) for expression and subsequent lysis. Protein was usually expressed in 2 L of M9 medium using the following recipe.

Recipe for making 2 L of M9 medium: To 1700 mL of water add

1. 200 mL of 10X M9 salts

   To make 2 L of 10X M9 salts add:

   a. Na2HPO4 120g

   b. KH2PO4 60g

   c. NaCl 10g

   d. Do NOT pH
2. 4 mL of 1 M $MgSO_4$
3. 4 mL of 50 mM $CaCl_2$
4. 2 mL of 0.01 M $FeCl_3$
5. 400 µL of 5 mg/mL vitamin B1
6. 20 % D-Glucose
7. pH the solution to 7.3 – 7.5.
8. Bring it up to 2 L and filter sterilize into two 1 L bottles. Take 2 mL of this media in a 15-mL falcon tube and label it as – N – Kan + cells (Control 1)
9. Add 1 g of ammonium chloride to each 1 L media and shake. Take 2 mL of this media in a falcon tube and label it as + N – Kan + cells (Control 2)

10. Add 500 µL of 60 mg/L of kanamycin to each 1 L media and shake. Take 2 mL of this media in a falcon tube and label it as + N + Kan – cells (Control 3)

A 50 mL of starter culture (along with controls) was inoculated with colonies from fresh transformations and allowed to grow at 37 ºC with constant aeration for 15 – 17 hours and was then used to start the main culture at an O.D of 0.04 measured at a wavelength of 600 nm. Cells were allowed to grow at 37 ºC until they reached an O.D of 0.6 and induced with 1 mM IPTG at 15 ºC. Culture was incubated for protein expression at 15 ºC for 22 hours (induction time tested for maximum protein expression level; **Figure 23**). Cells were then harvested by centrifugation at 7168 RCF for five minutes and stored at -80 ºC until purification. Samples of cells were collected pre- and post-induction for SDS-PAGE analysis of protein expression.

5.1.4 KIX Purification Scheme

Pellets from 1 L culture were usually purified at once. Harvested cells were resuspended in 25 mL of nickel lysis/load buffer (recipe at the end of section 5.1.3.1) with one Pierce™ Protease Inhibitor tablet dissolved. The resuspended cells were lysed using French pressure cell press and centrifuged at 38000 RCF for one hour. Cell lysate was then run through a nickel column containing 35 mL of Ni-NTA resin on a Fast Protein Liquid Chromatography (FPLC) system – either Biologic Duo flow or NGC systems, both purchased from BioRad. The six-histidine tag sticks to nickel column. Purification procedure of the protein through nickel column is described below in detail.

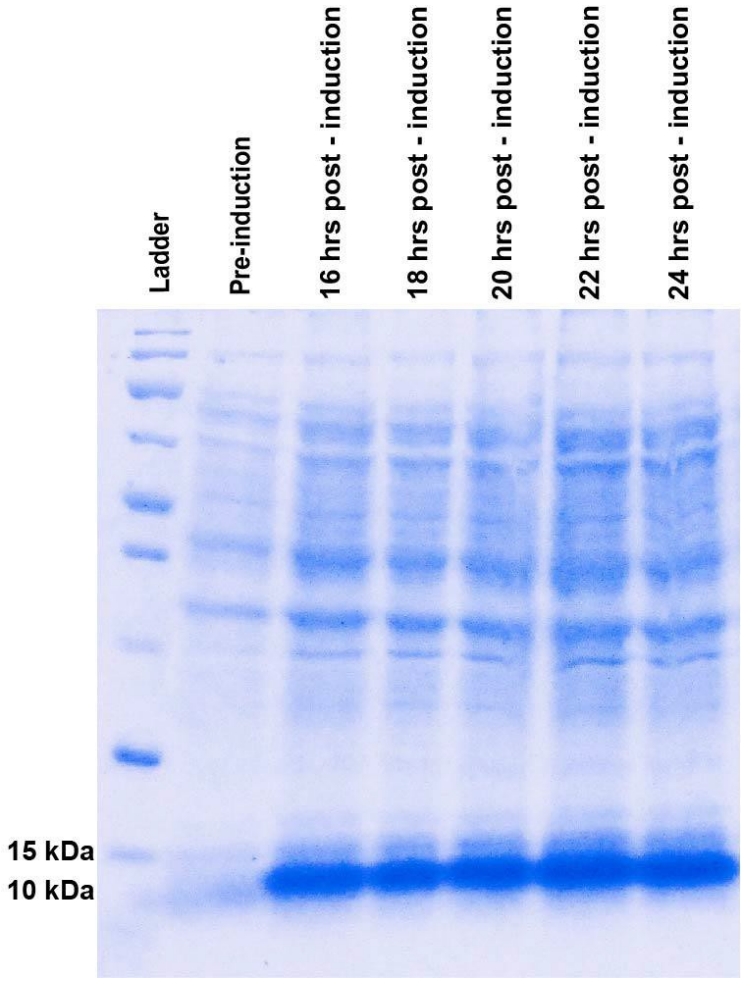


Figure 23: KIX expression gel. The above figure shows expression levels of KIX 16-24 hours post-induction. Culture was grown in 1 L M9 medium and induced at 15 ºC with 1 mM IPTG. Gel was run on a 6-16 % SDS gel at 195 V for 45 minutes.

5.1.4.1 Purification through nickel column. Nickel column was initially equilibrated with one column volume of nickel lysis/load buffer and then the column was charged with cell lysate. Column was then washed with one column volume of nickel lysis buffer during which most of the *E. coli* proteins pass through the column while KIX protein that is attached to six histidine tag sticks to the nickel column. Column is then washed with two column volumes of nickel wash buffer (a mixture of 85% of nickel lysis buffer and 15 % of nickel elution buffer mixed by the FPLC system during the purification run according to a protocol pre-defined by the user). During this

step, any proteins that are bound non-specifically to the column are washed off. Column is then washed with two column volumes of nickel elution buffer. In this step, the high concentration of imidazole in the elution buffer elutes the protein (**Figure 24 and Figure 25**). Buffers were run through the column at a flow rate of 3 mL/min throughout the purification procedure while lysate was loaded onto the column at a flow rate of 2 mL/min. Fractions from the flow through, wash and elute steps were always analysed by SDS-PAGE and the clean elute fractions were collected for subsequent purification. Eluted KIX protein along with the six-histidine tag and the thrombin cleavage site results in a 12 kDa peptide. Protein was concentrated and dialysed in gel filtration buffer for cleaving. KIX was found to precipitate at a concentration of 220 µM in elution buffer. It was therefore ensured that the protein was never concentrated above 200 µM in this step. Protein concentrations are determined as described in section 5.8. Nickel buffers and gel filtration buffer recipes are detailed below.

a. Ni lysis/load buffer
    i. 50 mM sodium phosphate monobasic monohydrate (NaH_2PO_4)
    ii. 300 mM sodium chloride (NaCl)
    iii. 10 mM imidazole
    iv. 0.02 % sodium azide (NaN_3)
    v. pH 8

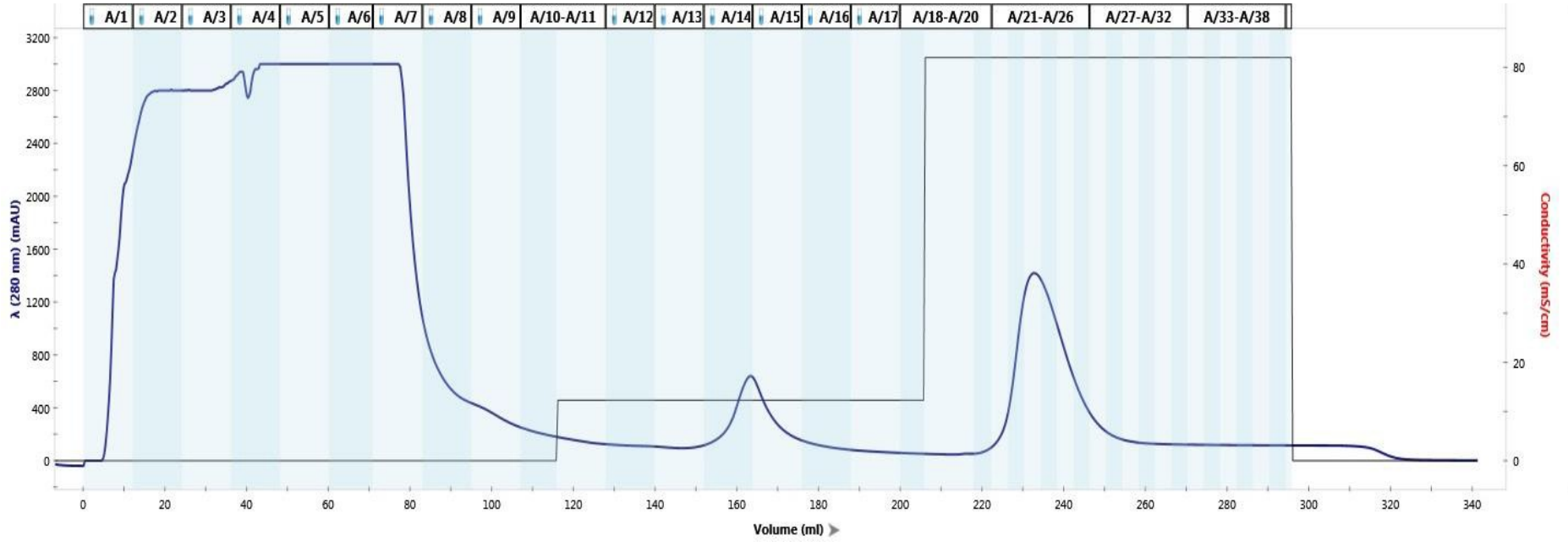


Figure 24: Pre-cleaved KIX purification by nickel column. Above figure shows the chromatogram of 1L KIX purified on a nickel column by BioRad NGC FPLC system. Solid blue line shows U.V absorbance throughout the run. Solid black line represents the percentage of buffer B running at every phase in the run. Fractions 1-9 represent the flow through peak, fractions 13-15 represent the wash peak and fractions 21-27 represent the elution peak.

b. Ni Elution buffer

   i. 50 mM sodium phosphate monobasic monohydrate (NaH_2PO_4)

   ii. 300 mM sodium chloride (NaCl)

   iii. 250 mM imidazole

   iv. 0.02 % sodium azide (NaN_3)

   v. pH 8

c. Gel Filtration Buffer (GFB)

   i. 50 mM sodium phosphate monobasic monohydrate (NaH_2PO_4)

   ii. 300 mM sodium Chloride (NaCl)

   iii. 1 mM EDTA

   iv. 0.02 % sodium azide (NaN_3)

   v. pH 7

5.1.4.2 Thrombin cleavage. Six histidine tag was cleaved from the protein using thrombin cleave kit purchased from Sigma Aldrich. Removal of tag was achieved by incubating the protein, having been dialysed in GFB, with thrombin beads for two hours. A cleave test was performed initially to determine the minimum time required for complete cleaving. Protein samples were collected before incubating with thrombin beads and periodically during incubation of protein with the beads. Samples were analysed on a SDS gel (**Figure 26**) and it was seen that two hours of incubation is sufficient for complete cleaving. Two hours of cleaving was used for subsequent protein preparations. Upon cleaving, two residues are left over from the thrombin cleave site namely Glycine and Serine. Thrombin cleavage results in a 10 kDa protein which was always analysed on a SDS gel to ensure complete cleaving. Cleaved protein was concentrated to 4-6 mL to run on size exclusion column to remove any impurities left over from nickel column purification.

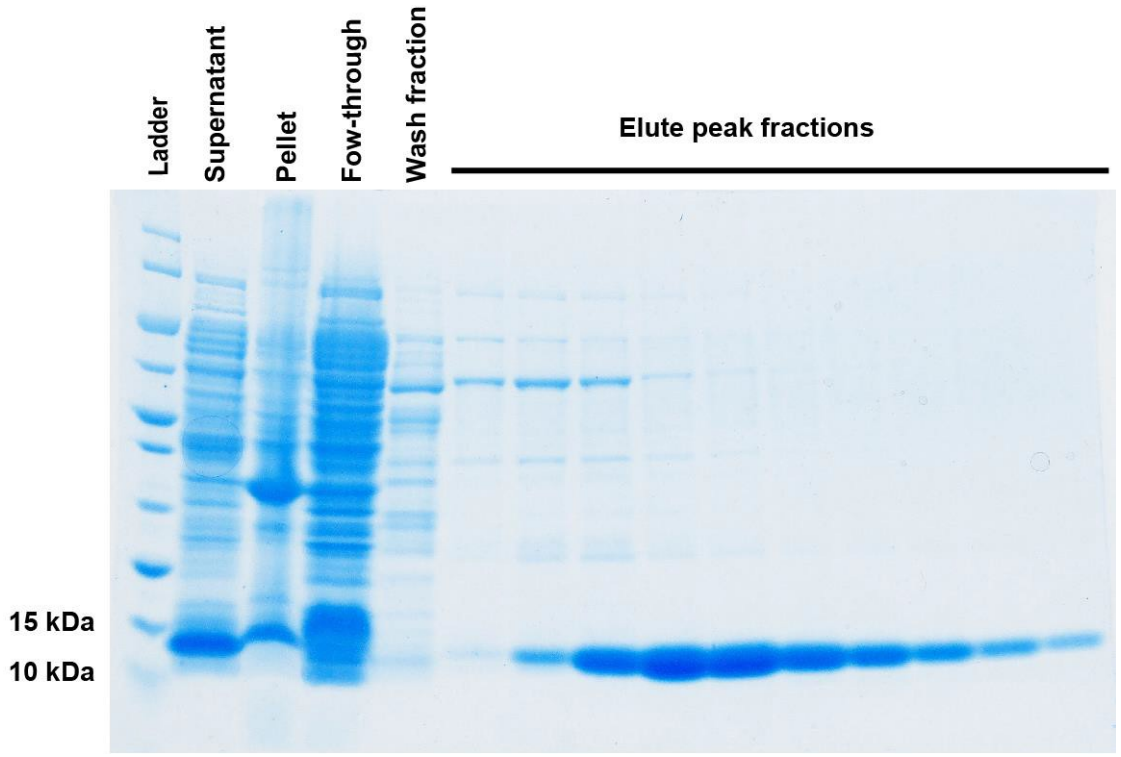


Figure 25: SDS-PAGE gel of KIX purification through nickel column. Above gel shows the samples taken from cell lysate after lysis and centrifugation (Supernatant and pellet). It also shows samples taken from the flow through, wash and elute peak fractions obtained after purification through nickel column.

5.1.4.3 Protein purification through size exclusion column. Size exclusion column separates proteins based on the size of the protein. Multiple injections with small volumes are usually run to ensure good resolution. 2 mL injections at 1 mL/min flow rate was shown to work best with our proteins. All size exclusion columns were run on a GE HiLoad XK 16/60 Superdex 75 pg column with a resin volume of 120 mL. Size exclusion purification scheme is described below (**Figure 26 and Figure 27**).

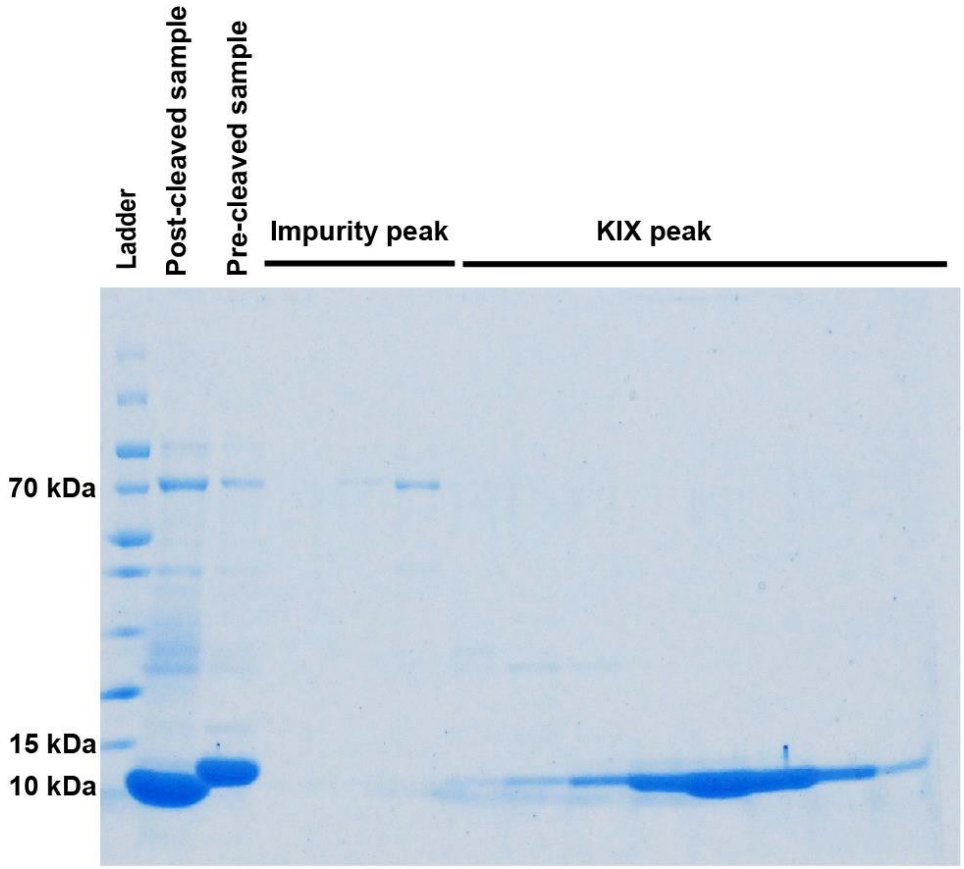


Figure 26: SDS-PAGE gel of KIX cleavage and purification through SEC. Above figure shows the pre-cleaved and concentrated post-cleaved samples of KIX. It also shows the fractions from the impurity peak and KIX peak obtained after purification of KIX through size exclusion chromatography.

Column is initially equilibrated with 1.25 column volumes of GFB and 2 mL of protein is loaded onto the column. Protein was then eluted with 1.5 column volumes of GFB separating the impurities from our protein of interest. Column is re-equilibrated with 0.25 column volumes of GFB before loading the next 2 mL of the protein onto the column. Fractions from the protein and impurity peak were always analysed by SDS gel (**Figure 26**) and clean fractions are pooled, concentrated and stored at 4 ºC until further use. Cleaned protein is generally immediately dialysed with NMR, CD or ITC buffers for subsequent analysis.

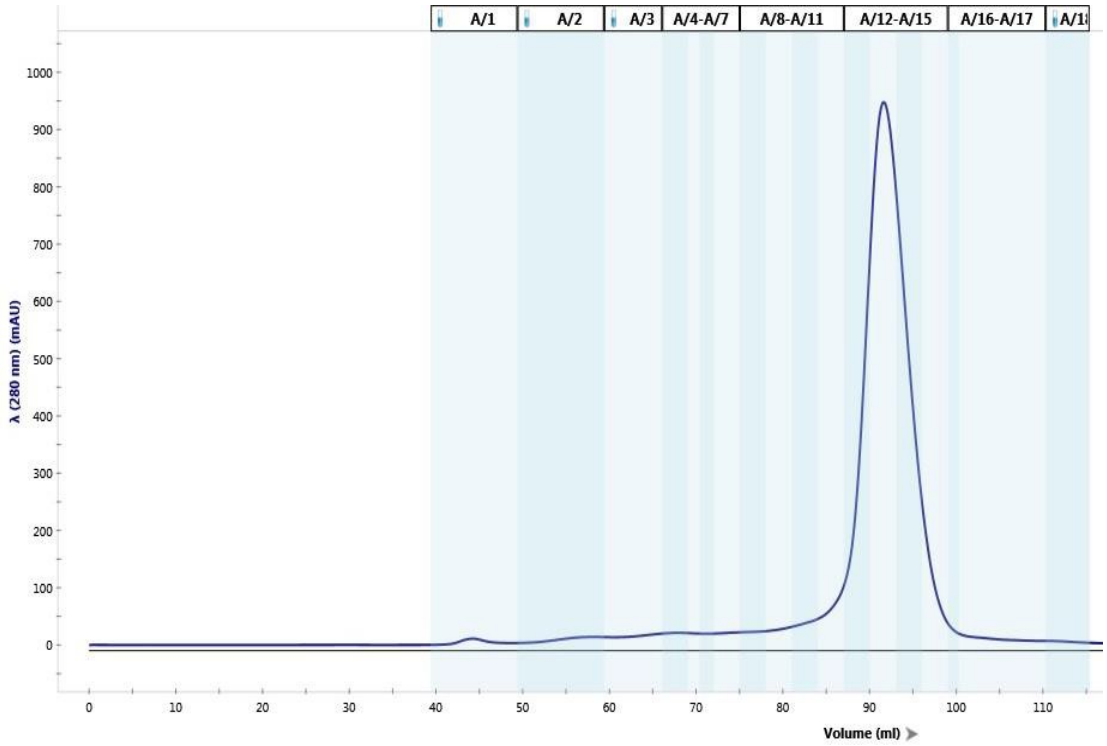


Figure 27: Chromatogram of KIX purified through size exclusion column. Above figure shows a single run KIX purified through size exclusion column on BioRad NGC system. Peak under fractions 1-3 represents the impurity and peak under fractions 10-16 represent the pure KIX protein. Image has been cropped to show just the peak portion. Equilibration of the column with GFB before the run and in between multiple runs are not shown in the above figure.

5.2 c-Myb Sample Preparation and Purification

5.2.1 Cloning

The human c-Myb gene (residues 275-327) sub-cloned into a modified pRSET-A vector was a kind gift from Dr. Jane Clarke (University of Cambridge). Details of modifications in the vector are described below. The vector is selected for ampicillin resistance and consists of a T7 promoter which is followed by a six-histidine tag (**Figure 28**). The Xpress Epitope EK region was removed. An insert consisting of GB1

tag, which helps to increase the solubility of the protein, followed by thrombin cleavage site and c-Myb gene was inserted between BamHI and EcoRI sites.

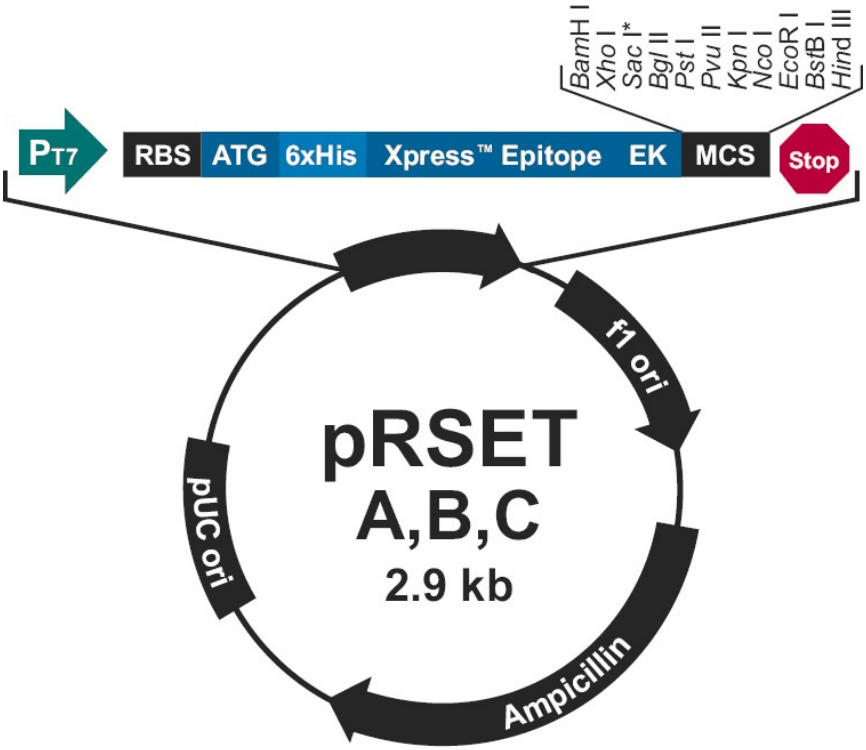


Figure 28: Vector map of pRSET A, B and C vectors. pRSET A, B and C are similar to each other. They have different open reading frames in the gene coding the N-terminal peptide and their multiple cloning sites vary a bit. Image is taken from Thermo Fisher Scientific.

5.2.2 Transformations

c-Myb plasmid was transformed into C41 (DE3) competent *E. coli* cells using the following protocol. C41 (DE3) cells contain the gene for T7 RNA polymerase under the control of lacUV5 promoter just like the BL21 (DE3) cells. Therefore, T7 polymerase can be induced by lactose or isopropyl β-D-1-thiogalactopyranoside (IPTG) allowing maximal expression of the genes cloned downstream of the T7 promoter.

1. Gently mix 10 µL of C41(DE3) cells and 1uL of 1 ng/µL c-Myb WT/mutant plasmid in a centrifuge tube.
2. Incubate on ice for thirty minutes.
3. Heat shock at 42 ºC for 30 seconds.
4. Incubate on ice for 2 minutes.
5. Add 100 µL of SOC and incubate in the shaker incubator at 37 ºC and 200 RPM for 1 hour.
6. Plate cells on an ampicillin resistant plate and incubate at 37 ºC overnight.
7. For a negative control, take 10 µL of cells and repeat steps 2 through 6.
8. Store the plates in the refrigerator the next morning until use.

5.2.3 Expression

Transformed *E. coli* colonies, containing c-Myb WT or mutant plasmid, that were not more than seven days old were grown in M9 medium for expression and subsequent lysis. Protein was usually expressed in 2 L of M9 medium using the same recipe that's described for the expression of KIX protein with slight variations. M9 medium and controls were made up of antibiotic ampicillin instead of kanamycin. Starter cultures along with controls were inoculated with colonies from fresh transformations, and allowed to grow at 37 ºC with constant aeration for 15 – 17 hours and then used to start the main culture at an O.D of 0.04 measured at a wavelength of 600 nm. Cells were allowed to grow at 37 ºC until they reached an O.D of 0.6 and were then induced with 1 mM IPTG at 37 ºC. Induction times for WT and mutants varied between 3-6 hrs. Expression tests were conducted and induction times required for maximum protein expression levels were selected for the WT and each mutant **(Table 10)**. Following protein expression, cells were harvested by centrifugation at

7168 RCF for five minutes and stored at -80 ºC until later purification. Samples of cells were collected pre-and post-induction for SDS-PAGE analysis of protein expression.

c-Myb peptide	Duration of induction representing maximum protein expression levels (in hrs)
WT	3
P289A	3
P289A P316A	3
G311A	5
E292D	4
K293H	5
K296H	6
L300V	6
L300P	4
L300G	4

Table 10: Protein induction times for c-Myb WT and its mutants. Table showing maximum protein expression levels for c-Myb WT and each mutant. c-Myb WT and all the mutants were induced at 37 ºC and an O.D of 0.6. Expression tests were conducted to determine the times to maximum protein expression levels after induction with IPTG.

5.2.4 c-Myb Purification Scheme

Pellets from 2L culture were usually purified at once. Harvested cells were resuspended in 25 mL of nickel lysis/load buffer, lysed and centrifuged as described in section 5.1.4. Purification procedure of the protein through nickel column is described below in detail.

5.2.4.1 Purification through nickel column. Purification of cell lysate through nickel column was done as described in section 5.1.4.1 (**Figure 29**). Fractions from

the flow through, wash and elute steps were always analysed by SDS-PAGE (**Figure 30**) and the clean eluate fractions were collected for subsequent purification. Eluted c-Myb protein along with the six-histidine tag and the GB1 tag that includes thrombin cleavage site results in a 15 kDa residue. Protein was concentrated and dialysed in nickel load buffer for cleaving.

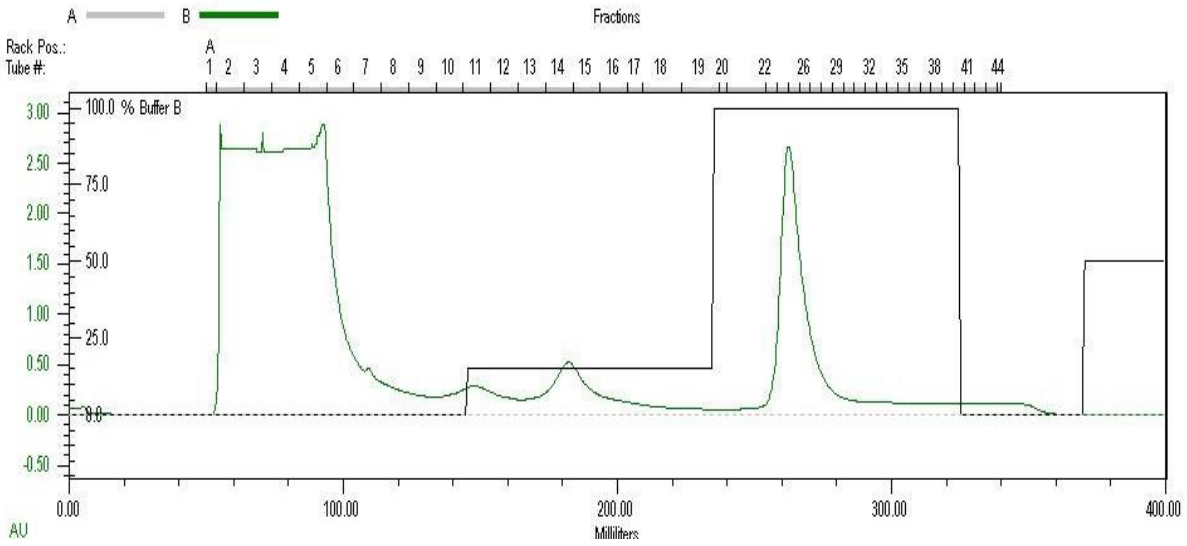


Figure 29: Nickel column chromatogram of c-Myb. Above figure shows the chromatogram of 2L c-Myb culture run on BioRad Biologic DuoFlow system. Green solid line indicates U.V absorbance and the black solid line indicates the percentage of buffer B running through the column at different stages of chromatography. Peak under fractions 1-8 represents the flow through peak, peak under fractions 10-16 represents the wash peak and the peak under fractions 22-28 represents elution peak.

5.2.4.2 Thrombin cleavage. Six histidine tag along with GB1 tag was cleaved from the protein using thrombin cleave kit purchased from Sigma Aldrich. Removal of tag was achieved by incubating the protein, that is dialysed in nickel load buffer, with thrombin beads from the kit. A cleave test was performed on WT and all mutant c-Myb proteins and it was seen that 3 hours of incubation with thrombin beads is sufficient for complete cleaving. Protein samples were collected before incubating with thrombin

beads and periodically during incubation of protein with the beads. Upon cleaving, two residues are left over from the thrombin cleave site namely glycine and serine. Thrombin cleavage results in a 6 kDa protein and 9 kDa tag (histidine tag and GB1 tag together) which were always analysed on a SDS gel (**Figure 31**) to ensure complete cleaving. Cleaved protein was run on a Nickel column again to separate the 9 kDa tag from the 6 kDa protein.

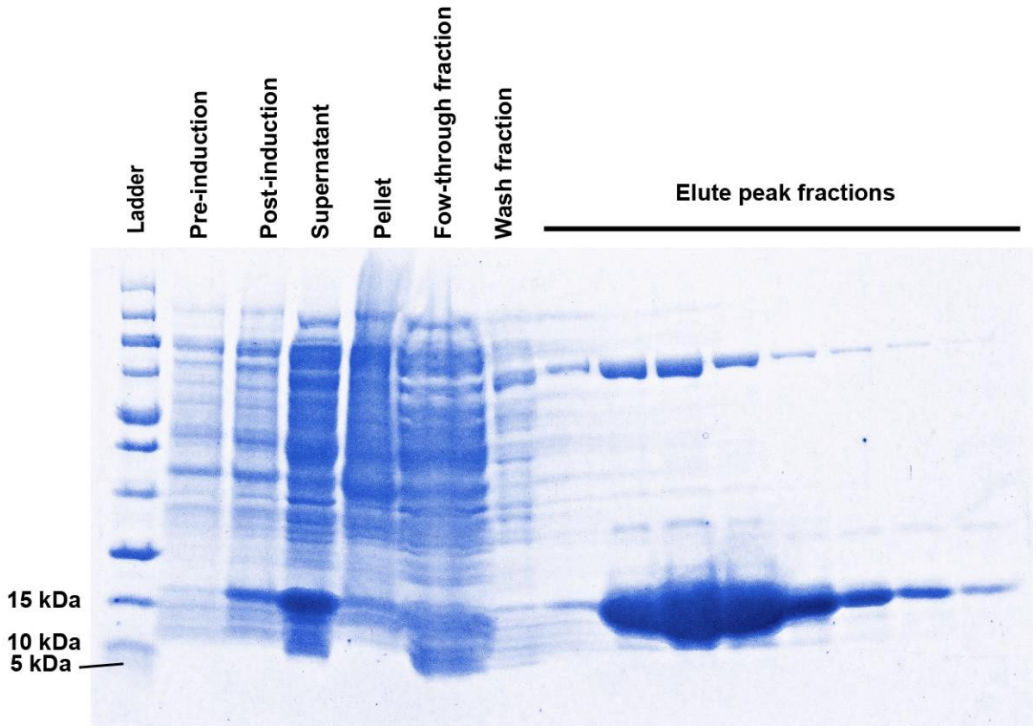


Figure 30: SDS-PAGE gel of pre-cleaved c-Myb. Above figure shows a 2L culture purification of c-Myb mutant L300P. It shows c-Myb at various steps from pre-induction to purification after nickel column. Pre- and post-inductions samples are also shown on the same gel. Flow through, wash and elute peak fractions represent fractions obtained through nickel column purification.

5.2.4.3 Protein purification through post-cleave nickel column. Nickel column was run with the same parameters described in section 5.1.4.1. In this purification step, cleaved c-Myb protein that is not attached to the six-histidine tag is collected in the flow through (F.T) fractions while the GB1 tag that is attached to the

six-histidine tag is collected in the elution fractions (**Figure 32**). Fractions from the F.T, wash and elute were run on a SDS gel for analysis (**Figure 31**). Clean flow through fractions were collected, concentrated to 4-6 mL and run on size exclusion column.

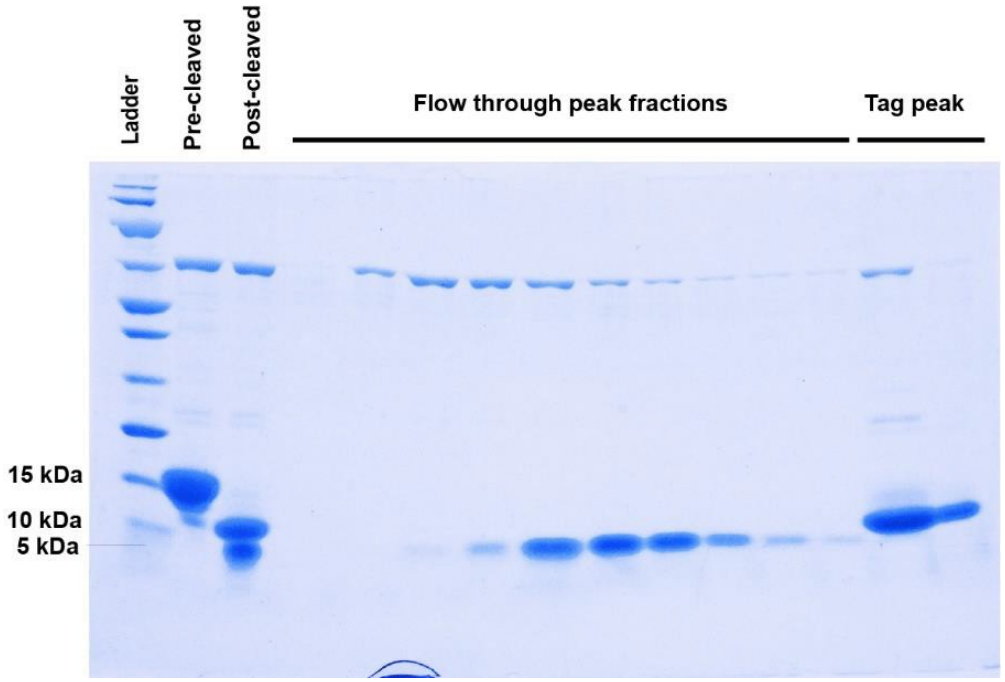


Figure 31: SDS-PAGE gel of pre- and post-cleaved c-Myb and post-cleaved purification fractions. Concentrated pre- and post-cleaved samples of c-Myb are shown in the above picture. Above figure also shows F.T and elute peak fractions obtained from purification through nickel column after cleaving using thrombin beads.

5.2.4.4 Protein purification through size exclusion column. Size exclusion column purification (**Figure 33**) was done as described in section 5.1.4.3. Fractions from the protein and impurity peaks were analysed by SDS gel electrophoresis (**Figure 34**) and clean fractions were pooled, concentrated and stored in the fridge until further use. Cleaned protein is generally immediately dialysed with NMR, CD or ITC buffers for subsequent analysis.

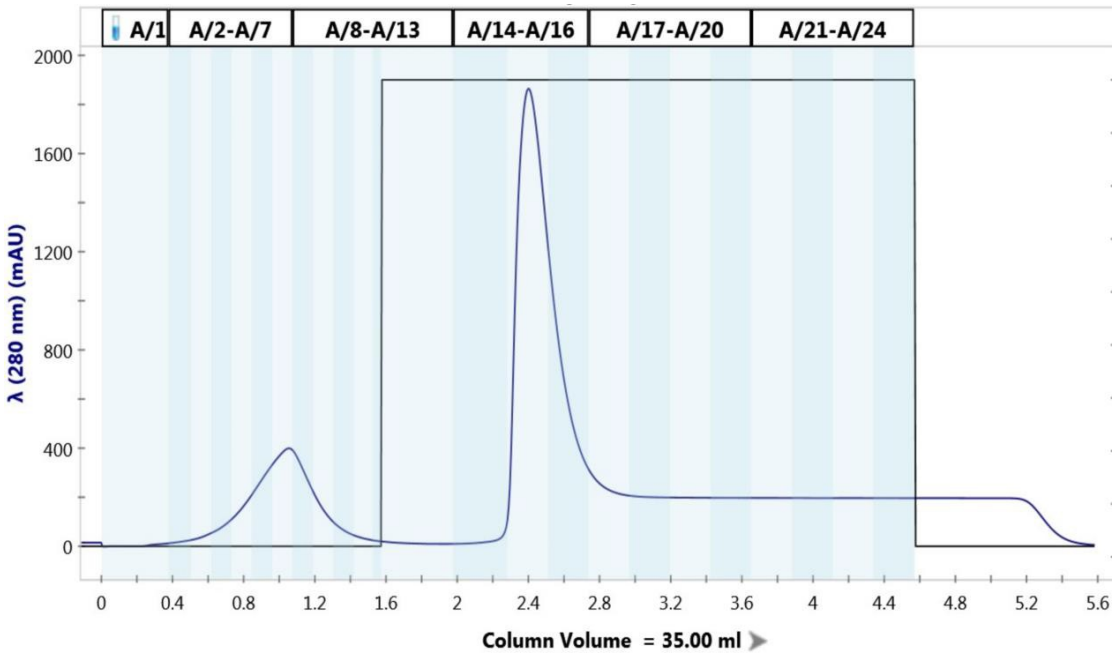


Figure 32: Chromatogram of c-Myb purified through post-cleave nickel column. Above picture shows the chromatogram of a 2L culture of c-Myb P289A P316A purified through nickel column after cleaving with thrombin beads. Peak under fractions 1-12 represent the flow through peak that contains cleaved protein and peak under fractions 15-17 represent the tag peak that contains GB1 tag attached to six-histidine tag.

5.3 Circular Dichroism

Circular Dichroism (CD) can be used to study the secondary structure of proteins and their folding properties [175]. CD is defined as the difference in the absorption of left handed and right handed circularly polarized light by an asymmetric molecule, in this case our protein of interest. CD spectra were measured on a JASCO J-815 spectropolarimeter at 25 ºC and 10 ºC. Proteins were dialysed by million-fold with CD buffer (refer to the end of the section) prior to measurement. CD spectra for all the proteins were measured in the far-UV range (190 – 260 nm). Protein solution of 1 mg/mL was placed in a 0.1 mm path length cell and the spectra were acquired under constant purging with nitrogen at a bandwidth of 1 nm and a scanning speed of 10 nm/min. Two spectra were collected for sample and averaged.

CD buffer

i. 50 mM sodium phosphate monobasic monohydrate (NaH_2PO_4)

ii. 50 mM sodium chloride (NaCl)

iii. pH 6.8

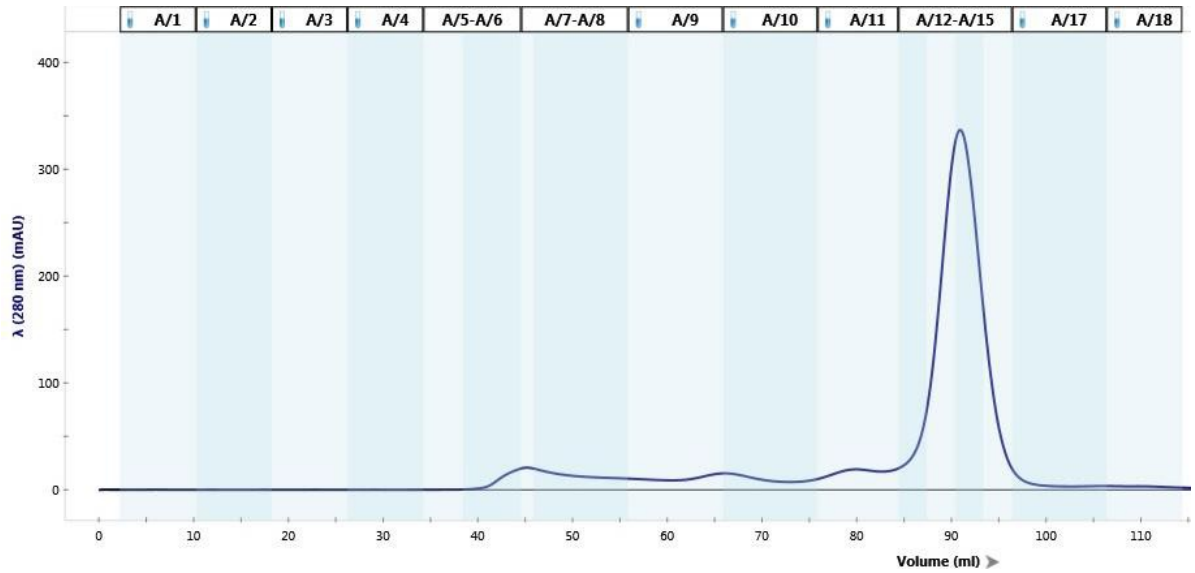


Figure 33: Chromatogram of c-Myb purified through size exclusion column. Above figure shows a single run of c-Myb L300P purified through size exclusion column on BioRad NGC system. Peaks under fractions 6-10 represent impurity peaks and peak under fractions 12-16 represent the pure c-Myb protein. Image has been cropped to show just the peak portion. Equilibration of the column with GFB before the run and in between multiple runs are not shown in the above figure.

5.4 Nuclear Magnetic Resonance (NMR) Spectroscopy

All NMR experiments were performed using 800 MHz Varian spectrometer. Uniformly labelled ^{15}N and ^{13}C proteins were made as described in the above sections. HSQC and HNCACB experiments were performed at 25 ºC or 10 ºC at a concentration range of 0.2-0.5 mM in 50 mM sodium phosphate, 50 mM sodium chloride, 1mM EDTA and 0.02% sodium azide with 10% D_2O at pH 6.8. All NMR spectra were processed using NMRFx processor [176] and analysed using NMRViewJ [177]. Random coil

chemical shift data was collected from the ncIDP database to calculate the secondary chemical shift deviations. Residual helicity of each peptide is calculated from the obtained secondary chemical shift deviation data by using the δ2D program developed by Vendruscolo [173]. Proteins were dialysed by million-fold in the following buffer before analysis.

NMR buffer

i. 50 mM sodium phosphate monobasic monohydrate (NaH_2PO_4)

ii. 50 mM sodium chloride (NaCl)

iii. 0.02 % sodium azide

iv. pH 6.8

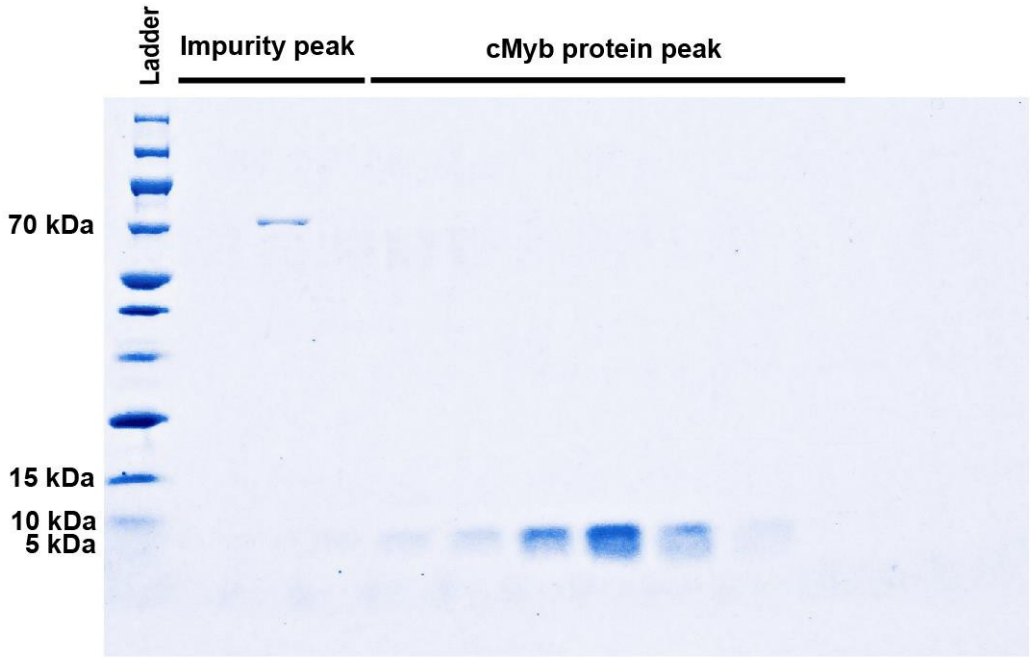


Figure 34: SDS-PAGE gel of c-Myb purified through size exclusion. Above figure shows samples taken from the fractions of impurity peak and the pure protein peak obtained after size exclusion column.

5.5 Isothermal Titration Calorimetry (ITC)

All ITC experiments were conducted at 25 ºC using MicroCal VP-ITC isothermal titration calorimeter in 50 mM sodium phosphate, 150 mM sodium chloride, 1 mM EDTA and 0.02 % sodium azide at pH 6.8. KIX was loaded into the cell at a concentration of 15 μM and the injection syringe was loaded with either mutant or WT c-Myb at a concentration of 150 μM. A single ITC run is performed with 29 injections each with a volume of 10 μL injected at a flow rate of 0.5 μL/sec. Three such runs were performed with every protein and the average binding affinity values with standard deviations were reported. Experimental data obtained from ITC was analysed using the Origin70 ITC software. Data from 29 injections in each run was integrated and was fit using a single site binding model. Proteins were dialysed by million-fold in the following buffer before analysis.

ITC buffer

i. 50 mM sodium phosphate monobasic monohydrate (NaH_2PO_4)
ii. 150 mM sodium chloride (NaCl)
iii. 0.02 % sodium azide
iv. pH 6.8

5.6 Site Directed Mutagenesis

Site directed mutagenesis was performed using the Stratagene QuickChange kit. QuickChange kit can be used to make single amino acid substitutions and insertions and deletions. Mutagenesis is performed using a supercoiled dsDNA (plasmid DNA) and oligonucleotide forward and reverse primers containing the desired mutation. Primers that are complimentary to the opposite strands of the plasmid DNA

are extended by PCR. Primers for the desired mutation were designed using PrimerX software and purchased from Operon. The following protocol was used.

5.6.1 Setting up the Polymerase Chain Reaction

1. Prepare a control reaction of 25 µL by adding:
    i. 2.5 µL of 10 X reaction buffer
    ii. 2 µL of pWhitescript 4.5 kb control plasmid (5 ng/µL)
    iii. 0.625 µL of oligonucleotide control primer 1 (100 ng/µL)]
    iv. 0.625 µL of oligonucleotide control primer 2 (100 ng/µL)]
    v. 0.500 µL of dNTP mix
    vi. Bring it up to a final volume of 25 µL with double distilled water.
2. Prepare sample reactions with a range of template plasmid concentrations such as 2.5, 5, 10 and 25 ng. To each tube of sample reactions, add
    i. 2.5 µL of 10 X reaction buffer
    ii. Template plasmid at the above-mentioned concentrations (X µL of 2.5 / 5 / 10 / 25 ng/µL)
    iii. 0.500 µL (125 ng/µL) of forward primer
    iv. 0.500 µL (125 ng/µL) of reverse primer
    v. 0.500 µL of dNTP mix
    vi. Bring it up to 25 µL with ddH_2O
3. Add 0.5 µL of *PfuTurbo* DNA polymerase to the control and sample reactions

5.6.2 Polymerase Chain Reaction

In this step, plasmid with the target site for mutation is denatured and annealed with the oligonucleotides containing the desired mutation. Thermocycler was set up as shown in **table 11**. Tubes are stored in the refrigerator until the next step.

Step	Temperature	Time	No. of Cycles
Initial denaturation	**95 ºC**	**30 seconds**	**1**
Denaturation, Annealing and Extension	**95 ºC**	**30 seconds**	**18**
	55 ºC	**1 minute**	
	68 ºC	**1 min/kb of plasmid**	
Final extension	**68 ºC**	**5 minutes**	**1**
Hold	**4 ºC**	**-**	**-**

Table 11: Steps of polymerase chain reaction. Table showing the steps of polymerase chain reaction with temperatures, duration of each cycle and the no of cycles used for each step.

5.6.3 Dpn I Digestion

PCR results in nicked circular strands of both parental DNA and DNA with desired mutation. Parental DNA from the mixture is digested using Dpn I which is specific to methylated DNA. DNA from *E. coli* is dam methylated and is therefore susceptible to Dpn I digestion. Dpn I digestion is done as follows.

i. Add 0.5 µL of Dpn I enzyme (10 U/µL).

ii. Blip spin microfuge tubes.

iii. Incubate at 37 ºC for 2 hours with agitation.

iv. Store the tubes in refrigerator until next step.

5.6.4 Transformations

Circular, nicked double-stranded mutated DNA is transformed using supercompetent XL1-Blue *E. coli* cells. After transformation, nicks in the mutated plasmid are repaired by XL1-Blue cells. Transformations were done using the following protocol.

1. Add 25 µL of XL1-Blue cells to a sterile labelled micro centrifuge tube.
2. Add 0.5 µL of Dpn I digested sample to respective tubes and mix gently.
3. Set up a positive control reaction with XL1-Blue cells and pUC plasmid.
4. Set up a mutagenesis control reaction with XL1-Blue cells and pWhitescript mutagenesis control plasmid.
5. Incubate on ice for 30 minutes.
6. Heat shock at 42 ºC for 30 seconds.
7. Incubate on ice for 2 minutes.
8. Add 125 µL of NZY+ broth to each tube and incubate for 1 hour 37 ºC in a shaker incubator at 200 RPM.
9. Plate 250 µL of pWhitescript mutagenesis control on LB gar plates with appropriate antibiotic containing 80 µg/mL X-gal and 20 mM IPTG.
10. Plate 5 µL of transformed pUC plasmid cells. 200 µL of NZY+ broth can be added to the plates for spreading.
11. Plate entire transformation reaction of Dpn I digested plasmid.
12. Incubate the plates at 37 ºC for more than 16 hours.

There should be 50-800 mutant control colonies with 80% being blue.

There should be >250 colonies of the transformation control with >98% being blue.

There should be 10-1000 sample colonies.

5.6.5 Minipreps

Minipreparation of plasmid DNA obtained from the above step was made from three to four isolated colonies and sent for sequencing to ensure that the right amino acid substitution was made.

5.6.5.1 Overnight cultures. Four to five single colonies were picked from a freshly streaked plate to inoculate 5 mL of LB medium supplemented with appropriate antibiotic. Tubes with the cultures were incubated for 14-16 hours at 37 ºC on a shaker incubator with aeration. 15 mL falcon tubes were used for overnight cultures to attain proper aeration. Minipreps were done the following morning using GeneJET Plasmid Miniprep kit from Thermo Scientific according to the protocol described below.

5.6.5.2 Purification of plasmid DNA.

1. Harvest the cultures by centrifugation at 6800 x g for five minutes.
2. Discard the supernatant and resuspend the pellet in 250 µL of the resuspension solution pre-mixed with RNAse A. Resuspend the pellet by pipetting up and down until no clumps are seen.
3. Transfer the cell lysate to a microcentrifuge tube and add 250 µL of the lysis solution. Mix thoroughly by inverting the tube up and down until the solution becomes viscous and slightly clear. This step shouldn't exceed more than five minutes.
4. Add 250 µL of neutralisation buffer and mix thoroughly by inverting the tube up and down 4-6 times. The solution becomes cloudy.
5. Centrifuge for 5 minutes at 12000 x g to pellet the cell debris.
6. Transfer the supernatant to the GeneJET spin column provided by the kit.
7. Centrifuge for 1 minute at 12000 x g and discard the flow through.

8. Add 500 µL of wash buffer and centrifuge for 1 minute at 12000 x g. Discard the flow through.

9. Repeat the above step and centrifuge for an additional one minute to remove any remaining wash solution.

10. Transfer the spin column to a sterilized 1.5 mL centrifuge tube. Add 20-50 µL of elution buffer to the center of the spin column and incubate for two minutes at room temperature.

11. Centrifuge for two minutes at 12000 x g to elute the plasmid DNA.

12. Store plasmids in the refrigerator.

5.6.6 Determination of Plasmid DNA Purity and Sequencing

Purity of the plasmid DNA was checked by U.V absorbance using ND-1000 Nanodrop. Plasmid DNA purity can be estimated by measuring the ratios of absorbances at various wavelengths i.e., 260 nm / 280 nm and 260 nm / 230 nm. The 260 nm / 280 nm ratio estimates the ratio of the amounts of nucleic acids and protein present in the sample. Ratio greater than 1.8 indicates low protein contamination. 260 nm / 280 nm ratio estimates the contamination of organic solvents and chaotropic salts. A ratio of greater than 2.0 indicates a pure sample. Samples with desirable purity were sent to Eurofins MWG Operon for sequencing to ensure that the plasmid DNA has the right sequence.

5.7 Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis (SDS-PAGE)

The separation of macromolecules under the influence of an electric field is called electrophoresis. SDS-PAGE is the most common type of electrophoresis used

to separate proteins. SDS is an anionic detergent that denatures proteins i.e., breaks down the complex tertiary structure of proteins making the polypeptide linear. SDS has a net negative charge and binds to linear polypeptide chains uniformly i.e., the charge of the polypeptide is proportional to its molecular weight thus allowing polypeptides to be separated solely by differences in molecular weight. Masking of polypeptides with SDS results in an overall negative charge on polypeptides attracting them towards anode in an electric field.

Samples were generally run on a 6 % to 16 % gradient SDS gel at 195 V for 45 minutes. Samples with complexes of KIX bound to mutant and WT were run on a 10 % to 20 % gradient SDS gel at 195 V for 45 minutes. Molecular weights of cleaved KIX and cleaved c-Myb are 9 kDa and 6 kDa respectively and thus a 10 % to 20 % gradient gel allows a better resolution in the separation of both the molecules. 40 % acrylamide solution with 29:1 acrylamide/bisacrylamide composition was used to make all gels.

5.8 Determination of Protein Concentration

Concentration of proteins is determined by measuring the U.V absorbance of sample at 280 nm using NanoDrop ND-1000 Spectrophotometer. Concentration of the sample is calculated from the absorbance value at 280 nm using the Beer's Law Equation i.e., Absorbance = e*L*c where 'e' is the molar extinction coefficient, 'L' is the path length of the cell holder and 'c' is the concentration of the solution. Molar extinction coefficients were calculated using the ProtParam tool provided by ExPAsy: SIB Bioinformatics Resource Portal.

REFERENCES

1. Murzin, A.G., et al., *SCOP: a structural classification of proteins database for the investigation of sequences and structures.* J Mol Biol, 1995. **247**(4): p. 536-40.
2. Orengo, C.A., et al., *CATH--a hierarchic classification of protein domain structures.* Structure, 1997. **5**(8): p. 1093-108.
3. Oldfield, C.J. and A.K. Dunker, *Intrinsically disordered proteins and intrinsically disordered protein regions.* Annu Rev Biochem, 2014. **83**: p. 553-84.
4. Uversky, V.N. and A.K. Dunker, *Understanding protein non-folding.* Biochim Biophys Acta, 2010. **1804**(6): p. 1231-64.
5. Theillet, F.X., et al., *The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins.* Intrinsically Disord Proteins, 2013. **1**(1): p. e24360.
6. Dunker, A.K., et al., *Intrinsically disordered protein.* J Mol Graph Model, 2001. **19**(1): p. 26-59.
7. Dunker, A.K., et al., *Intrinsic disorder and protein function.* Biochemistry, 2002. **41**(21): p. 6573-82.
8. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions.* Nat Rev Mol Cell Biol, 2005. **6**(3): p. 197-208.
9. Iakoucheva, L.M., et al., *Intrinsic disorder in cell-signaling and cancer-associated proteins.* J Mol Biol, 2002. **323**(3): p. 573-84.

10. Tompa, P., *Intrinsically unstructured proteins.* Trends Biochem Sci, 2002. **27**(10): p. 527-33.
11. Uversky, V.N., *Natively unfolded proteins: a point where biology waits for physics.* Protein Sci, 2002. **11**(4): p. 739-56.
12. Wright, P.E. and H.J. Dyson, *Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm.* J Mol Biol, 1999. **293**(2): p. 321-31.
13. Dunker, A.K., et al., *Flexible nets. The roles of intrinsic disorder in protein interaction networks.* FEBS J, 2005. **272**(20): p. 5129-48.
14. Uversky, V.N., C.J. Oldfield, and A.K. Dunker, *Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling.* J Mol Recognit, 2005. **18**(5): p. 343-84.
15. Xie, H., et al., *Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions.* J Proteome Res, 2007. **6**(5): p. 1882-98.
16. Vucetic, S., et al., *Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions.* J Proteome Res, 2007. **6**(5): p. 1899-916.
17. Liu, J., J.R. Faeder, and C.J. Camacho, *Toward a quantitative theory of intrinsically disordered proteins and their function.* Proc Natl Acad Sci U S A, 2009. **106**(47): p. 19819-23.
18. Kim, P.M., et al., *The role of disorder in interaction networks: a structural analysis.* Mol Syst Biol, 2008. **4**: p. 179.
19. Oldfield, C.J., et al., *Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners.* BMC Genomics, 2008. **9 Suppl 1**: p. S1.

20. Wright, P.E. and H.J. Dyson, *Linking folding and binding.* Curr Opin Struct Biol, 2009. **19**(1): p. 31-8.

21. Ptitsyn, O.B., V.E. Bychkova, and V.N. Uversky, *Kinetic and equilibrium folding intermediates.* Philos Trans R Soc Lond B Biol Sci, 1995. **348**(1323): p. 35-41.

22. Uversky, V.N. and O.B. Ptitsyn, *Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature.* J Mol Biol, 1996. **255**(1): p. 215-28.

23. Dunker, A.K., et al., *Intrinsic protein disorder in complete genomes.* Genome Inform Ser Workshop Genome Inform, 2000. **11**: p. 161-71.

24. Ward, J.J., et al., *Prediction and functional analysis of native disorder in proteins from the three kingdoms of life.* J Mol Biol, 2004. **337**(3): p. 635-45.

25. Xue, B., A.K. Dunker, and V.N. Uversky, *Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life.* J Biomol Struct Dyn, 2012. **30**(2): p. 137-49.

26. Babu, M.M., et al., *Intrinsically disordered proteins: regulation and disease.* Curr Opin Struct Biol, 2011. **21**(3): p. 432-40.

27. Liu, J., et al., *Intrinsic disorder in transcription factors.* Biochemistry, 2006. **45**(22): p. 6873-88.

28. Uversky, V.N., C.J. Oldfield, and A.K. Dunker, *Intrinsically disordered proteins in human diseases: introducing the D2 concept.* Annu Rev Biophys, 2008. **37**: p. 215-46.

29. Vacic, V. and L.M. Iakoucheva, *Disease mutations in disordered regions--exception to the rule?* Mol Biosyst, 2012. **8**(1): p. 27-32.

30. Vacic, V., et al., *Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder.* PLoS Comput Biol, 2012. **8**(10): p. e1002709.

31. Johnson, L.N., *Glycogen phosphorylase: control by phosphorylation and allosteric effectors.* FASEB J, 1992. **6**(6): p. 2274-82.

32. Gao, J. and D. Xu, *Correlation between posttranslational modification and intrinsic disorder in protein.* Pac Symp Biocomput, 2012: p. 94-103.

33. Romero, P.R., et al., *Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms.* Proc Natl Acad Sci U S A, 2006. **103**(22): p. 8390-5.

34. Liu, Y., K.S. Matthews, and S.E. Bondos, *Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the Drosophila hox protein ultrabithorax.* J Biol Chem, 2008. **283**(30): p. 20874-87.

35. Light, S., et al., *Long indels are disordered: a study of disorder and indels in homologous eukaryotic proteins.* Biochim Biophys Acta, 2013. **1834**(5): p. 890-7.

36. Kovacs, E., et al., *Dual coding in alternative reading frames correlates with intrinsic protein disorder.* Proc Natl Acad Sci U S A, 2010. **107**(12): p. 5429-34.

37. Uversky, V.N., et al., *Rigidity of human alpha-fetoprotein tertiary structure is under ligand control.* Biochemistry, 1997. **36**(44): p. 13638-45.

38. Reichmann, D. and U. Jakob, *The roles of conditional disorder in redox proteins.* Curr Opin Struct Biol, 2013. **23**(3): p. 436-42.

39. Timsit, Y., et al., *The role of disordered ribosomal protein extensions in the early steps of eubacterial 50 S ribosomal subunit assembly.* Int J Mol Sci, 2009. **10**(3): p. 817-34.

40. Uversky, V.N., J.R. Gillespie, and A.L. Fink, *Why are "natively unfolded" proteins unstructured under physiologic conditions?* Proteins, 2000. **41**(3): p. 415-27.

41. Dunker, A.K., et al., *Protein disorder and the evolution of molecular recognition: theory, predictions and observations.* Pac Symp Biocomput, 1998: p. 473-84.

42. Radivojac, P., et al., *Intrinsic disorder and functional proteomics.* Biophys J, 2007. **92**(5): p. 1439-56.

43. Vacic, V., et al., *Composition Profiler: a tool for discovery and visualization of amino acid composition differences.* BMC Bioinformatics, 2007. **8**: p. 211.

44. Romero, P., et al., *Sequence complexity of disordered protein.* Proteins, 2001. **42**(1): p. 38-48.

45. Vucetic, S., et al., *Flavors of protein disorder.* Proteins, 2003. **52**(4): p. 573-84.

46. Xie, Q., et al., *The Sequence Attribute Method for Determining Relationships Between Sequence and Protein Disorder.* Genome Inform Ser Workshop Genome Inform, 1998. **9**: p. 193-200.

47. Dunker, A.K., C.J. Brown, and Z. Obradovic, *Identification and functions of usefully disordered proteins.* Adv Protein Chem, 2002. **62**: p. 25-49.

48. Campen, A., et al., *TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder.* Protein Pept Lett, 2008. **15**(9): p. 956-63.

49. Brown, C.J., et al., *Evolution and disorder.* Curr Opin Struct Biol, 2011. **21**(3): p. 441-6.

50. Fukuchi, S., et al., *Intrinsically disordered loops inserted into the structural domains of human proteins.* J Mol Biol, 2006. **355**(4): p. 845-57.

51. Hsu, W.L., et al., *Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding.* Protein Sci, 2013. **22**(3): p. 258-73.

52. Kriwacki, R.W., et al., *Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity.* Proc Natl Acad Sci U S A, 1996. **93**(21): p. 11504-9.

53. Dunker, A.K. and R.W. Kriwacki, *The orderly chaos of proteins.* Sci Am, 2011. **304**(4): p. 68-73.

54. Ptitsyn, O.B. and V.N. Uversky, *The molten globule is a third thermodynamical state of protein molecules.* FEBS Lett, 1994. **341**(1): p. 15-8.

55. Dunker, A.K. and Z. Obradovic, *The protein trinity--linking function and disorder.* Nat Biotechnol, 2001. **19**(9): p. 805-6.

56. Mao, A.H., et al., *Net charge per residue modulates conformational ensembles of intrinsically disordered proteins.* Proc Natl Acad Sci U S A, 2010. **107**(18): p. 8183-8.

57. Das, R.K. and R.V. Pappu, *Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues.* Proc Natl Acad Sci U S A, 2013. **110**(33): p. 13392-7.

58. Uversky, V.N., et al., *Ligand-free form of human alpha-fetoprotein: evidence for the molten globule state.* FEBS Lett, 1997. **410**(2-3): p. 280-4.

59. van der Lee, R., et al., *Classification of intrinsically disordered regions and proteins.* Chem Rev, 2014. **114**(13): p. 6589-631.

60. Hansen, J.C., et al., *Intrinsic protein disorder, amino acid composition, and histone terminal domains.* J Biol Chem, 2006. **281**(4): p. 1853-6.

61. Jensen, M.R., R.W. Ruigrok, and M. Blackledge, *Describing intrinsically disordered proteins at atomic resolution by NMR.* Curr Opin Struct Biol, 2013. **23**(3): p. 426-35.

62. Xie, H., et al., *Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins.* J Proteome Res, 2007. **6**(5): p. 1917-32.

63. Gehring, W.J., et al., *Homeodomain-DNA recognition.* Cell, 1994. **78**(2): p. 211-23.

64. Holmes, K.C., *Flexibility in tobacco mosaic virus.* Ciba Found Symp, 1983. **93**: p. 116-38.

65. Spolar, R.S. and M.T. Record, Jr., *Coupling of local folding to site-specific binding of proteins to DNA.* Science, 1994. **263**(5148): p. 777-84.

66. Fink, A.L., *Natively unfolded proteins.* Curr Opin Struct Biol, 2005. **15**(1): p. 35-41.

67. Peng, Z., et al., *A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome.* Cell Mol Life Sci, 2014. **71**(8): p. 1477-504.

68. Moore, P.B., *How should we think about the ribosome?* Annu Rev Biophys, 2012. **41**: p. 1-19.

69. Hilser, V.J., *Biochemistry. An ensemble view of allostery.* Science, 2010. **327**(5966): p. 653-4.

70. Beckett, D., *Regulating transcription regulators via allostery and flexibility.* Proc Natl Acad Sci U S A, 2009. **106**(52): p. 22035-6.

71. Motlagh, H.N. and V.J. Hilser, *Agonism/antagonism switching in allosteric ensembles.* Proc Natl Acad Sci U S A, 2012. **109**(11): p. 4134-9.

72. Kumar, R. and I.J. McEwan, *Allosteric modulators of steroid hormone receptors: structural dynamics and gene regulation.* Endocr Rev, 2012. **33**(2): p. 271-99.

73. Ferreon, A.C., et al., *Modulation of allostery by protein intrinsic disorder.* Nature, 2013. **498**(7454): p. 390-4.

74. Tompa, P. and P. Csermely, *The role of structural disorder in the function of RNA and protein chaperones.* FASEB J, 2004. **18**(11): p. 1169-75.

75. Tompa, P., *The interplay between structure and function in intrinsically unstructured proteins.* FEBS Lett, 2005. **579**(15): p. 3346-54.

76. Patil, A. and H. Nakamura, *Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks.* FEBS Lett, 2006. **580**(8): p. 2041-5.

77. Haynes, C., et al., *Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes.* PLoS Comput Biol, 2006. **2**(8): p. e100.

78. Dosztanyi, Z., et al., *Disorder and sequence repeats in hub proteins and their implications for network evolution.* J Proteome Res, 2006. **5**(11): p. 2985-95.

79. Hengst, L., et al., *A cell cycle-regulated inhibitor of cyclin-dependent kinases.* Proc Natl Acad Sci U S A, 1994. **91**(12): p. 5291-5.

80. Singh, J., et al., *The use of Group 3 LEA proteins as fusion partners in facilitating recombinant expression of recalcitrant proteins in E. coli.* Protein Expr Purif, 2009. **67**(1): p. 15-22.

81. Oldfield, C.J., et al., *Utilization of protein intrinsic disorder knowledge in structural proteomics.* Biochim Biophys Acta, 2013. **1834**(2): p. 487-98.

82. Santner, A.A., et al., *Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression.* Biochemistry, 2012. **51**(37): p. 7250-62.

83. Herschlag, D., *RNA chaperones and the RNA folding problem.* J Biol Chem, 1995. **270**(36): p. 20871-4.

84. Semrad, K., *Proteins with RNA chaperone activity: a world of diverse proteins with a common task-impediment of RNA misfolding.* Biochem Res Int, 2011. **2011**: p. 532908.

85. Hendrick, J.P. and F.U. Hartl, *The role of molecular chaperones in protein folding.* FASEB J, 1995. **9**(15): p. 1559-69.

86. Beissinger, M. and J. Buchner, *How chaperones fold proteins.* Biol Chem, 1998. **379**(3): p. 245-59.

87. Kim, Y.E., et al., *Molecular chaperone functions in protein folding and proteostasis.* Annu Rev Biochem, 2013. **82**: p. 323-55.

88. Ivanyi-Nagy, R., et al., *Disordered RNA chaperone proteins: from functions to disease.* Cell Mol Life Sci, 2005. **62**(13): p. 1409-17.

89. Pontius, B.W., *Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association.* Trends Biochem Sci, 1993. **18**(5): p. 181-6.

90. Rogers, J.M., A. Steward, and J. Clarke, *Folding and binding of an intrinsically disordered protein: fast, but not 'diffusion-limited'.* J Am Chem Soc, 2013. **135**(4): p. 1415-22.

91. Uversky, V.N., *Intrinsic disorder-based protein interactions and their modulators.* Curr Pharm Des, 2013. **19**(23): p. 4191-213.

92. Diella, F., et al., *Understanding eukaryotic linear motifs and their role in cell signaling and regulation.* Front Biosci, 2008. **13**: p. 6580-603.

93. Galea, C.A., et al., *Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits.* Biochemistry, 2008. **47**(29): p. 7598-609.

94. Gsponer, J. and M.M. Babu, *The rules of disorder or why disorder rules.* Prog Biophys Mol Biol, 2009. **99**(2-3): p. 94-103.

95. George, O.L. and S.A. Ness, *Situational awareness: regulation of the myb transcription factor in differentiation, the cell cycle and oncogenesis.* Cancers (Basel), 2014. **6**(4): p. 2049-71.

96. Zhou, Y. and S.A. Ness, *Myb proteins: angels and demons in normal and transformed cells.* Front Biosci (Landmark Ed), 2011. **16**: p. 1109-31.

97. Lipsick, J.S., et al., *Functional evolution of the Myb oncogene family.* Blood Cells Mol Dis, 2001. **27**(2): p. 456-8.

98. Ness, S.A., *The Myb oncoprotein: regulating a regulator.* Biochim Biophys Acta, 1996. **1288**(3): p. F123-39.

99. Lipsick, J.S., *One billion years of Myb.* Oncogene, 1996. **13**(2): p. 223-35.

100. Du, H., et al., *Biochemical and molecular characterization of plant MYB transcription factor family.* Biochemistry (Mosc), 2009. **74**(1): p. 1-11.

101. Feller, A., et al., *Evolutionary and comparative analysis of MYB and bHLH plant transcription factors.* Plant J, 2011. **66**(1): p. 94-116.

102. Lavu, S. and E.P. Reddy, *Structural organization and nucleotide sequence of mouse c-myb oncogene: activation in ABPL tumors is due to viral integration in an intron which results in the deletion of the 5' coding sequences.* Nucleic Acids Res, 1986. **14**(13): p. 5309-20.

103. Weinstein, Y., et al., *Truncation of the c-myb gene by a retroviral integration in an interleukin 3-dependent myeloid leukemia cell line.* Proc Natl Acad Sci U S A, 1986. **83**(14): p. 5010-4.

104. Shen-Ong, G.L., et al., *Two modes of c-myb activation in virus-induced mouse myeloid tumors.* Mol Cell Biol, 1986. **6**(2): p. 380-92.

105. Hugo, H., et al., *Mutations in the MYB intron I regulatory sequence increase transcription in colon cancers.* Genes Chromosomes Cancer, 2006. **45**(12): p. 1143-54.

106. Thompson, M.A., et al., *Microsatellite deletions in the c-myb transcriptional attenuator region associated with over-expression in colon tumour cell lines.* Oncogene, 1997. **14**(14): p. 1715-23.

107. Drabsch, Y., et al., *Mechanism of and requirement for estrogen-regulated MYB expression in estrogen-receptor-positive breast cancer cells.* Proc Natl Acad Sci U S A, 2007. **104**(34): p. 13762-7.

108. Guerin, M., et al., *Strong association between c-myb and oestrogen-receptor expression in human breast cancer.* Oncogene, 1990. **5**(1): p. 131-5.

109. Kastan, M.B., K.D. Stone, and C.I. Civin, *Nuclear oncoprotein expression as a function of lineage, differentiation stage, and proliferative status of normal human hematopoietic cells.* Blood, 1989. **74**(5): p. 1517-24.

110. Sitzmann, J., K. Noben-Trauth, and K.H. Klempnauer, *Expression of mouse c-myb during embryonic development.* Oncogene, 1995. **11**(11): p. 2273-9.

111. Reilly, C.F., et al., *Heparin prevents vascular smooth muscle cell progression through the G1 phase of the cell cycle.* J Biol Chem, 1989. **264**(12): p. 6990-5.

112. Rosenthal, M.A., et al., *Colonic expression of c-myb is initiated in utero and continues throughout adult life.* Cell Growth Differ, 1996. **7**(7): p. 961-7.

113. Nomura, N., et al., *Isolation of human cDNA clones of myb-related genes, A-myb and B-myb.* Nucleic Acids Res, 1988. **16**(23): p. 11075-89.

114. Baluda, M.A. and E.P. Reddy, *Anatomy of an integrated avian myeloblastosis provirus: structure and function.* Oncogene, 1994. **9**(10): p. 2761-74.

115. Oh, I.H. and E.P. Reddy, *The myb gene family in cell growth, differentiation and apoptosis.* Oncogene, 1999. **18**(19): p. 3017-33.

116. Fu, S.L. and J.S. Lipsick, *FAETL motif required for leukemic transformation by v-Myb.* J Virol, 1996. **70**(8): p. 5600-10.

117. Dash, A.B., F.C. Orrico, and S.A. Ness, *The EVES motif mediates both intermolecular and intramolecular regulation of c-Myb.* Genes Dev, 1996. **10**(15): p. 1858-69.

118. Dudek, H. and E.P. Reddy, *Identification of two translational products for c-myb.* Oncogene, 1989. **4**(9): p. 1061-6.

119. Dudek, H. and E.P. Reddy, *Murine myeloid leukemias with aberrant myb loci show heterogeneous expression of novel myb proteins.* Oncogene, 1989. **4**(12): p. 1489-95.

120. Dasgupta, P., et al., *Myb protein binds to human immunodeficiency virus 1 long terminal repeat (LTR) sequences and transactivates LTR-mediated transcription.* Proc Natl Acad Sci U S A, 1990. **87**(20): p. 8090-4.

121. Dini, P.W., J.T. Eltman, and J.S. Lipsick, *Mutations in the DNA-binding and transcriptional activation domains of v-Myb cooperate in transformation.* J Virol, 1995. **69**(4): p. 2515-24.

122. Dini, P.W. and J.S. Lipsick, *Oncogenic truncation of the first repeat of c-Myb decreases DNA binding in vitro and in vivo.* Mol Cell Biol, 1993. **13**(12): p. 7334-48.

123. Lei, W., et al., *Positive and negative determinants of target gene specificity in myb transcription factors.* J Biol Chem, 2004. **279**(28): p. 29519-27.

124. Gonda, T.J., C. Buckmaster, and R.G. Ramsay, *Activation of c-myb by carboxy-terminal truncation: relationship to transformation of murine haemopoietic cells in vitro.* EMBO J, 1989. **8**(6): p. 1777-83.

125. Gonda, T.J., et al., *Generation of altered transcripts by retroviral insertion within the c-myb gene in two murine monocytic leukemias.* J Virol, 1987. **61**(9): p. 2754-63.

126. Biedenkapp, H., et al., *Viral myb oncogene encodes a sequence-specific DNA-binding activity.* Nature, 1988. **335**(6193): p. 835-7.

127. Miglarese, M.R., et al., *Differential regulation of c-Myb-induced transcription activation by a phosphorylation site in the negative regulatory domain.* J Biol Chem, 1996. **271**(37): p. 22697-705.

128. Mettus, R.V., et al., *Murine A-myb: evidence for differential splicing and tissue-specific expression.* Oncogene, 1994. **9**(10): p. 3077-86.

129. Saikumar, P., R. Murali, and E.P. Reddy, *Role of tryptophan repeats and flanking amino acids in Myb-DNA interactions.* Proc Natl Acad Sci U S A, 1990. **87**(21): p. 8452-6.

130. Sakura, H., et al., *Delineation of three functional domains of the transcriptional activator encoded by the c-myb protooncogene.* Proc Natl Acad Sci U S A, 1989. **86**(15): p. 5758-62.

131. Ramsay, R.G. and T.J. Gonda, *MYB function in normal and cancer cells.* Nat Rev Cancer, 2008. **8**(7): p. 523-34.

132. Lipsick, J.S. and M.A. Baluda, *The myb oncogene.* Gene Amplif Anal, 1986. **4**: p. 73-98.

133. Thompson, M.A. and R.G. Ramsay, *Myb: an old oncoprotein with new roles.* Bioessays, 1995. **17**(4): p. 341-50.

134. Sheiness, D. and M. Gardinier, *Expression of a proto-oncogene (proto-myb) in hemopoietic tissues of mice.* Mol Cell Biol, 1984. **4**(7): p. 1206-12.

135. Westin, E.H., et al., *Differential expression of the amv gene in human hematopoietic cells.* Proc Natl Acad Sci U S A, 1982. **79**(7): p. 2194-8.

136. Mucenski, M.L., et al., *A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis.* Cell, 1991. **65**(4): p. 677-89.

137. Gonda, T.J. and D. Metcalf, *Expression of myb, myc and fos proto-oncogenes during the differentiation of a murine myeloid leukaemia.* Nature, 1984. **310**(5974): p. 249-51.

138. Clarke, M.F., et al., *Constitutive expression of a c-myb cDNA blocks Friend murine erythroleukemia cell differentiation.* Mol Cell Biol, 1988. **8**(2): p. 884-92.

139. Yanagisawa, H., et al., *Constitutive expression of exogenous c-myb gene causes maturation block in monocyte-macrophage differentiation.* Biochim Biophys Acta, 1991. **1088**(3): p. 380-4.

140. Allen, R.D., 3rd, T.P. Bender, and G. Siu, *c-Myb is essential for early T cell development.* Genes Dev, 1999. **13**(9): p. 1073-8.

141. Thomas, M.D., et al., *c-Myb is critical for B cell development and maintenance of follicular B cells.* Immunity, 2005. **23**(3): p. 275-86.

142. Hogg, A., et al., *Inactivation of a c-Myb/estrogen receptor fusion protein in transformed primary cells leads to granulocyte/macrophage differentiation and down regulation of c-kit but not c-myc or cdc2.* Oncogene, 1997. **15**(24): p. 2885-98.

143. Miglarese, M.R., R. Halaban, and N.W. Gibson, *Regulation of fibroblast growth factor 2 expression in melanoma cells by the c-MYB proto-oncoprotein.* Cell Growth Differ, 1997. **8**(11): p. 1199-210.

144. Brandt, T.L., et al., *c-Myb trans-activates the human DNA topoisomerase IIalpha gene promoter.* J Biol Chem, 1997. **272**(10): p. 6278-84.

145. Salomoni, P., et al., *Resistance to apoptosis in CTLL-2 cells constitutively expressing c-Myb is associated with induction of BCL-2 expression and Myb-dependent regulation of bcl-2 promoter activity.* Proc Natl Acad Sci U S A, 1997. **94**(7): p. 3296-301.

146. Anfossi, G., A.M. Gewirtz, and B. Calabretta, *An oligomer complementary to c-myb-encoded mRNA inhibits proliferation of human myeloid leukemia cell lines.* Proc Natl Acad Sci U S A, 1989. **86**(9): p. 3379-83.

147. Calabretta, B., et al., *Normal and leukemic hematopoietic cells manifest differential sensitivity to inhibitory effects of c-myb antisense oligodeoxynucleotides: an in vitro study relevant to bone marrow purging.* Proc Natl Acad Sci U S A, 1991. **88**(6): p. 2351-5.

148. Lahortiga, I., et al., *Duplication of the MYB oncogene in T cell acute lymphoblastic leukemia.* Nat Genet, 2007. **39**(5): p. 593-5.

149. Clappier, E., et al., *The C-MYB locus is involved in chromosomal translocation and genomic duplications in human T-cell acute leukemia (T-ALL), the translocation defining a new T-ALL subtype in very young children.* Blood, 2007. **110**(4): p. 1251-61.

150. Ramsay, R.G., et al., *Myb expression is higher in malignant human colonic carcinoma and premalignant adenomatous polyps than in normal mucosa.* Cell Growth Differ, 1992. **3**(10): p. 723-30.

151. Melani, C., et al., *Inhibition of proliferation by c-myb antisense oligodeoxynucleotides in colon adenocarcinoma cell lines that express c-myb.* Cancer Res, 1991. **51**(11): p. 2897-901.

152. Torelli, G., 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): p. 5266-9.

153. Bender, T.P., C.B. Thompson, and W.M. Kuehl, *Differential expression of c-myb mRNA in murine B lymphomas by a block to transcription elongation.* Science, 1987. **237**(4821): p. 1473-6.

154. Watson, R.J., *A transcriptional arrest mechanism involved in controlling constitutive levels of mouse c-myb mRNA.* Oncogene, 1988. **2**(3): p. 267-72.

155. Dooley, S., et al., *c-myb intron I protein binding and association with transcriptional activity in leukemic cells.* Leuk Res, 1996. **20**(5): p. 429-39.

156. Wang, D.M., et al., *Functional analysis of carboxy-terminal deletion mutants of c-Myb.* J Virol, 1999. **73**(7): p. 5875-86.

157. Dai, P., et al., *CBP as a transcriptional coactivator of c-Myb.* Genes Dev, 1996. **10**(5): p. 528-40.

158. Oelgeschlager, M., et al., *Interaction of the co-activator CBP with Myb proteins: effects on Myb-specific transactivation and on the cooperativity with NF-M.* EMBO J, 1996. **15**(11): p. 2771-80.

159. Kanei-Ishii, C., et al., *Transactivation and transformation by Myb are negatively regulated by a leucine-zipper structure.* Proc Natl Acad Sci U S A, 1992. **89**(7): p. 3088-92.

160. Nomura, T., et al., *Negative autoregulation of c-Myb activity by homodimer formation through the leucine zipper.* J Biol Chem, 1993. **268**(29): p. 21914-23.

161. Karamouzis, M.V., P.A. Konstantinopoulos, and A.G. Papavassiliou, *Roles of CREB-binding protein (CBP)/p300 in respiratory epithelium tumorigenesis.* Cell Res, 2007. **17**(4): p. 324-32.

162. Janknecht, R., *The versatile functions of the transcriptional coactivators p300 and CBP and their roles in disease.* Histol Histopathol, 2002. **17**(2): p. 657-68.

163. Giles, R.H., D.J. Peters, and M.H. Breuning, *Conjunction dysfunction: CBP/p300 in human disease.* Trends Genet, 1998. **14**(5): p. 178-83.

164. Goodman, R.H. and S. Smolik, *CBP/p300 in cell growth, transformation, and development.* Genes Dev, 2000. **14**(13): p. 1553-77.

165. Kalkhoven, E., *CBP and p300: HATs for different occasions.* Biochem Pharmacol, 2004. **68**(6): p. 1145-55.

166. Krois, A.S., et al., *Recognition of the disordered p53 transactivation domain by the transcriptional adapter zinc finger domains of CREB-binding protein.* Proc Natl Acad Sci U S A, 2016. **113**(13): p. E1853-62.

167. Lin, C.H., et al., *A small domain of CBP/p300 binds diverse proteins: solution structure and functional studies.* Mol Cell, 2001. **8**(3): p. 581-90.

168. Dyson, H.J. and P.E. Wright, *Coupling of folding and binding for unstructured proteins.* Curr Opin Struct Biol, 2002. **12**(1): p. 54-60.

169. Lee, C., et al., *Contribution of proline to the pre-structuring tendency of transient helical secondary structure elements in intrinsically disordered proteins.* Biochim Biophys Acta, 2014. **1840**(3): p. 993-1003.

170. Crabtree, M.D., et al., *Conserved Helix-Flanking Prolines Modulate Intrinsically Disordered Protein:Target Affinity by Altering the Lifetime of the Bound Complex.* Biochemistry, 2017. **56**(18): p. 2379-2384.

171. Munoz, V. and L. Serrano, *Elucidating the folding problem of helical peptides using empirical parameters.* Nat Struct Biol, 1994. **1**(6): p. 399-409.

172. Munoz, V. and L. Serrano, *Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence.* J Mol Biol, 1995. **245**(3): p. 297-308.

173. Camilloni, C., et al., *Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts.* Biochemistry, 2012. **51**(11): p. 2224-31.

174. Zor, T., et al., *Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb.* J Mol Biol, 2004. **337**(3): p. 521-34.

175. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure.* Nat Protoc, 2006. **1**(6): p. 2876-90.

176. Norris, M., et al., *NMRFx Processor: a cross-platform NMR data processing program.* J Biomol NMR, 2016. **65**(3-4): p. 205-16.

177. Johnson, B.A. and R.A. Blevins, *NMR View: A computer program for the visualization and analysis of NMR data.* J Biomol NMR, 1994. **4**(5): p. 603-14.

APPENDICES

Appendix A: Chemical Shifts Tables of Mutant and WT c-Myb

Table A1. c-Myb WT Chemical Shifts

Residue	CO	CB	CA	N	HN
P	176.251648	31.9837093	63.395771		
A	177.64209	19.1026306	52.5764198	124.217903	8.31011009
A	177.64209	19.1777191	52.5184784	122.844193	8.10978031
A	177.74147	19.1612301	52.5113792	122.844193	8.10978031
A	177.805115	19.1250191	52.5675011	123.112343	8.15645981
I	176.801651	38.706871	61.1877708	119.833977	7.94857979
Q	175.567673	29.4332008	55.702919	124.410347	8.31521034
R	175.62117	30.9693203	55.8497391	122.57386	8.25009251
H	174.663742	30.5723801	55.7702293	120.528023	8.36273956
Y	175.344482	39.0033913	57.9470406	121.819237	8.18362999
N	174.801315	39.0571404	53.1883507	120.481697	8.51933956
D	176.118149	41.1490517	54.7804184	120.7649	8.25166988
E	175.969894	30.3205891	56.4079094	120.092346	8.23711014
D		41.5247612	52.1508217	122.293633	8.33296585
P	178.363571	32.2467117	64.2790527		
E	177.673996	29.4087296	57.4385681	118.486214	8.37530994
K	177.834854	32.5905991	58.7876511	120.894569	7.70233011
E	177.952209	29.5658493	58.4550095	118.676239	8.13712025
K	178.074249	32.6411285	58.4596901	120.362953	7.86625004
R	178.074249	30.3211994	57.8246994		
I	177.802597	38.0331001	63.6509285	120.535843	7.98291016
K	178.444061	32.227169	58.2094116	121.750847	7.99701977
E	178.49115	29.7106991	58.6601295	119.747589	8.1718502
L	178.672684	41.9338112	57.1037292	120.840157	7.97055006
E	178.379105	29.8735104	58.6294098	119.796982	8.13099957
L	178.87825	41.967659	56.8112984	119.997749	7.84325981
L	178.659241	42.0739403	56.6159401	121.14859	7.85596991
L	178.146713	41.97369	56.2192993	120.654701	8.00646973

Residue	CO	CB	CA	N	HN
M	176.803207	32.8791695	56.2530594	118.894569	7.95686007
S	175.288483	63.7126007	58.9786491	115.932518	8.11594963
T	175.126801	69.59095	62.6361504	115.724358	8.15273952
E	176.834793	29.9831009	57.4510384	122.273293	8.33024025
N	175.752365	38.9301987	53.9741096	118.611122	8.29827023
E	176.902603	30.1086502	57.3457603	121.053787	8.26537991
L	177.734085	41.9289589	55.524559	122.033363	8.14163017
K	177.36174	32.8578796	56.7266006	121.216743	8.10945034
G	174.247452		44.7326813	109.141586	8.29001999
Q	175.903793	29.5503597	55.7700005	119.459084	8.12030029
Q	175.646729	29.5394001	55.8062096	121.496887	8.36600018
V	175.813919	32.7392807	62.1061516	121.976791	8.1723299
L		41.8241615	52.8375282	127.59005	8.33372974
P	177.071472	32.0555611	63.1059494		
T	174.625351	69.59095	61.85	113.777023	8.07275963
Q		29.9831009	55.7450104	122.25383	8.32050991
N					
H	175.275528	30.30937	56.2547493		
T	174.16095	69.7615509	61.9369698	115.542633	8.08308029
A	177.365311	19.3240604	52.5523987	126.207626	8.26251984
S	173.492157	63.7662392	58.0805702	114.608559	8.05494976
Y		38.4071312	55.4808807	123.029411	7.98932981
P	176.993896	31.7978401	63.4661713		
G	172.969543		45.2038498	108.569702	7.71261978
W		29.9346294	58.4260292	125.815163	7.44950008

Table A2. c-Myb P289A Chemical Shifts

Residue	CO	CB	CA	N	HN
P	176.6	31.94	63.15	138.91	
A	177.56	18.97	52.43	124.47	8.39
A	177.67	18.98	52.57	123.26	8.19
A	177.67	18.98	52.57	123.26	8.19
A	177.93	19.06	52.63	123.1	8.17
I	176.41	38.43	61.31	119.79	8.04
Q	175.84	29.44	55.49	124.36	8.47
R	175.99	30.67	55.97	122.76	8.32

Residue	CO	CB	CA	N	HN
H	174.25	29.7	55.84	120.45	8.47
Y	175.19	38.64	57.76	121.77	8.14
N	174.34	38.89	52.89	121.02	8.31
D	176.43	40.95	54.55	120.81	8.33
E	176.33	30.12	56.84	121.08	8.38
D	176.14	40.99	54.32	121.43	8.34
A	177.97	18.99	52.82	124.49	8.22
E	176.56	30.1	56.53	121.41	8.54
K	176.34	32.88	56.49	122.21	8.39
E	176.49	30.15	56.7	122.15	8.42
K	176.66	32.78	56.44	121.97	8.26
R	176.15	30.68	56.15	122.43	8.33
I	176.13	38.55	61.13	122.49	8.16
K	176.68	32.88	56.25	125.46	8.46
E	176.4	30.07	56.66	121.68	8.39
L	177.5	42.16	55.36	123	8.25
E	176.4	30.03	56.55	121.21	8.32
L	177.39	42.04	55.25	122.71	8.18
L	177.31	42.02	55.16	122.31	8.11
L	177.65	41.86	55.28	122.57	8.13
M	176.35	32.63	55.52	120.78	8.31
S	174.79	63.71	58.26	117.42	8.46
T	174.49	69.71	62.01	116.22	8.26
E	176.15	30.14	56.73	123.23	8.49
N	175.29	38.74	53.33	119.79	8.53
E	176.46	29.97	56.76	121.07	8.35
L	177.49	42.12	55.29	123.17	8.21
K	177.11	32.78	56.49	122.19	8.33
G	173.96	9999	45.33	110.01	8.46
Q	175.96	29.51	55.74	120.14	8.24
Q	175.69	29.49	55.75	121.88	8.43
V	176.01	32.56	62.29	121.71	8.18
L	174.97	41.64	52.91	127.46	8.32
P	177.1	32.02	63.14	137.52	9999
T	174.53	69.65	61.85	114.51	8.32
Q	175.53	29.42	55.92	122.69	8.45
N	175	38.77	53.11	120.14	8.46
H	174.97	29.6	56.13	119.7	8.48
T	174.06	69.72	61.81	115.78	8.13
A	177.81	19.12	52.59	126.53	8.37

Residue	CO	CB	CA	N	HN
S	174.46	63.65	58.36	114.91	8.21
Y	173.28	38.17	55.49	123.6	8.23
P	176.91	32.09	62.94	139.34	9999
G	173.9	9999	45.06	109.08	8.63
W					

Table A3. c-Myb P289A P316A Chemical Shifts

Residue	CO	CB	CA	N	HN
P	176.6	31.94	63.15	138.91	
A	177.56	18.97	52.43	124.47	8.39
A	177.67	18.98	52.57	123.26	8.19
A	177.67	18.98	52.57	123.26	8.19
A	177.93	19.06	52.63	123.1	8.17
I	176.41	38.43	61.31	119.79	8.04
Q	175.84	29.44	55.49	124.36	8.47
R	175.99	30.67	55.97	122.76	8.32
H	174.25	29.7	55.84	120.45	8.47
Y	175.19	38.64	57.76	121.77	8.14
N	174.34	38.89	52.89	121.02	8.31
D	176.43	40.95	54.55	120.81	8.33
E	176.33	30.12	56.84	121.08	8.38
D	176.14	40.99	54.32	121.43	8.34
A	177.97	18.99	52.82	124.49	8.22
E	176.67	30.12	56.67	120.2	8.34
K	176.69	32.87	56.51	122.48	8.38
E	176.49	30.15	56.7	122.15	8.42
K	176.57	32.85	56.32	122.51	8.36
R	176.15	30.68	56.15	122.43	8.33
I	176.13	38.55	61.13	122.49	8.16
K	176.68	32.88	56.25	125.46	8.46
E	176.4	30.07	56.66	121.68	8.39
L	177.5	42.16	55.36	123	8.25
E	176.4	30.03	56.55	121.21	8.32
L	177.39	42.04	55.25	122.71	8.18
L	177.31	42.02	55.16	122.31	8.11
L	177.65	41.86	55.28	122.57	8.13

Residue	CO	CB	CA	N	HN
M	176.35	32.63	55.52	120.78	8.31
S	174.79	63.71	58.26	117.42	8.46
T	174.49	69.71	62.01	116.22	8.26
E	176.15	30.14	56.73	123.23	8.49
N	175.29	38.74	53.33	119.79	8.53
E	176.46	29.97	56.76	121.07	8.35
L	177.49	42.12	55.29	123.17	8.21
K	177.11	32.78	56.49	122.19	8.33
G	173.96	9999	45.33	110.01	8.46
Q	175.96	29.51	55.74	120.14	8.24
Q	175.69	29.49	55.75	121.88	8.43
V	176.01	32.56	62.29	121.71	8.18
L	177.02	42.08	55.15	126.38	8.29
A	177.98	19.04	52.56	124.82	8.3
T	174.64	69.66	61.99	113.3	8.12
Q	175.53	29.42	55.92	122.69	8.45
N	175	38.77	53.11	120.14	8.46
H	174.97	29.6	56.13	119.7	8.48
T	174.06	69.72	61.81	115.78	8.13
A	177.81	19.12	52.59	126.53	8.37
S	174.46	63.65	58.36	114.91	8.21
Y	173.28	38.17	55.49	123.6	8.23
P	176.91	32.09	62.94	139.34	9999
G	173.9		45.06	109.08	8.63

Table A4. c-Myb E292D Chemical Shifts

Residue	CO	CB	CA	N	HN
P	176.6	31.94	63.15	138.91	
A	177.56	18.97	52.43	124.47	8.39
A	177.67	18.98	52.57	123.26	8.19
A	177.67	18.98	52.57	123.26	8.19
A	177.93	19.06	52.63	123.1	8.17
I	176.41	38.43	61.31	119.79	8.04
Q	175.84	29.44	55.49	124.36	8.47
R	175.99	30.67	55.97	122.76	8.32
H	174.25	29.7	55.84	120.45	8.47

Residue	CO	CB	CA	N	HN
Y	175.19	38.64	57.76	121.77	8.14
N	174.34	38.89	52.89	121.02	8.31
D	176.43	40.95	54.55	120.81	8.33
E	176.33	30.12	56.84	121.08	8.38
D	174.08	40.54	52.08	122.5	8.36
P	177.1	31.96	63.4	137.19	9999
E	176.56	30.1	56.53	121.41	8.54
K	176.34	32.88	56.49	122.21	8.39
D	176.35	41.02	54.38	121.59	8.34
K	176.66	32.78	56.44	121.97	8.26
R	176.15	30.68	56.15	122.43	8.33
I	176.13	38.55	61.13	122.49	8.16
K	176.68	32.88	56.25	125.46	8.46
E	176.4	30.07	56.66	121.68	8.39
L	177.5	42.16	55.36	123	8.25
E	176.4	30.03	56.55	121.21	8.32
L	177.39	42.04	55.25	122.71	8.18
L	177.31	42.02	55.16	122.31	8.11
L	177.65	41.86	55.28	122.57	8.13
M	176.35	32.63	55.52	120.78	8.31
S	174.79	63.71	58.26	117.42	8.46
T	174.49	69.71	62.01	116.22	8.26
E	176.15	30.14	56.73	123.23	8.49
N	175.29	38.74	53.33	119.79	8.53
E	176.46	29.97	56.76	121.07	8.35
L	177.49	42.12	55.29	123.17	8.21
K	177.11	32.78	56.49	122.19	8.33
G	173.96	9999	45.33	110.01	8.46
Q	175.96	29.51	55.74	120.14	8.24
Q	175.69	29.49	55.75	121.88	8.43
V	176.01	32.56	62.29	121.71	8.18
L	174.97	41.64	52.91	127.46	8.32
P	177.1	32.02	63.14	137.52	9999
T	174.53	69.65	61.85	114.51	8.32
Q	175.53	29.42	55.92	122.69	8.45
N	175	38.77	53.11	120.14	8.46
H	174.97	29.6	56.13	119.7	8.48
T	174.06	69.72	61.81	115.78	8.13
A	177.81	19.12	52.59	126.53	8.37

Residue	CO	CB	CA	N	HN
S	174.46	63.65	58.36	114.91	8.21
Y	173.28	38.17	55.49	123.6	8.23
P	176.91	32.09	62.94	139.34	9999
G	173.9		45.06	109.08	8.63

Table A5. c-Myb L300G Chemical Shifts

Residue	CO	CB	CA	N	HN
P	176.6	31.94	63.15	138.91	
A	177.56	18.97	52.43	124.47	8.39
A	177.67	18.98	52.57	123.26	8.19
A	177.67	18.98	52.57	123.26	8.19
A	177.93	19.06	52.63	123.1	8.17
I	176.41	38.43	61.31	119.79	8.04
Q	175.84	29.44	55.49	124.36	8.47
R	175.99	30.67	55.97	122.76	8.32
H	174.25	29.7	55.84	120.45	8.47
Y	175.19	38.64	57.76	121.77	8.14
N	174.34	38.89	52.89	121.02	8.31
D	176.43	40.95	54.55	120.81	8.33
E	176.33	30.12	56.84	121.08	8.38
D	174.08	40.54	52.08	122.5	8.36
P	177.1	31.96	63.4	137.19	9999
E	176.56	30.1	56.53	121.41	8.54
K	176.69	32.87	56.51	122.48	8.38
E	176.49	30.15	56.7	122.15	8.42
K	176.57	32.85	56.32	122.51	8.36
R	176.15	30.68	56.15	122.43	8.33
I	176.13	38.55	61.13	122.49	8.16
K	176.68	32.88	56.25	125.46	8.46
E	176.4	30.07	56.66	121.68	8.39
L	177.5	42.16	55.36	123	8.25
E	177.01	30.08	56.73	121.6	8.41
G	173.99	9999	45.22	109.61	8.39
L	177.43	42.16	55.08	121.48	8
L	177.65	41.86	55.28	122.57	8.13

Residue	CO	CB	CA	N	HN
M	176.35	32.63	55.52	120.78	8.31
S	174.79	63.71	58.26	117.42	8.46
T	174.49	69.71	62.01	116.22	8.26
E	176.15	30.14	56.73	123.23	8.49
N	175.29	38.74	53.33	119.79	8.53
E	176.46	29.97	56.76	121.07	8.35
L	177.49	42.12	55.29	123.17	8.21
K	177.11	32.78	56.49	122.19	8.33
G	173.96	9999	45.33	110.01	8.46
Q	175.96	29.51	55.74	120.14	8.24
Q	175.69	29.49	55.75	121.88	8.43
V	176.01	32.56	62.29	121.71	8.18
L	174.97	41.64	52.91	127.46	8.32
P	177.1	32.02	63.14	137.52	9999
T	174.53	69.65	61.85	114.51	8.32
Q	175.53	29.42	55.92	122.69	8.45
N	175	38.77	53.11	120.14	8.46
H	174.97	29.6	56.13	119.7	8.48
T	174.06	69.72	61.81	115.78	8.13
A	177.81	19.12	52.59	126.53	8.37
S	174.46	63.65	58.36	114.91	8.21
Y	173.28	38.17	55.49	123.6	8.23
P	176.91	32.09	62.94	139.34	9999
G	173.9		45.06	109.08	8.63

Table A6. c-Myb L300P Chemical Shifts

Residue	CO	CB	CA	N	HN
P	176.6	31.94	63.15	138.91	
A	177.56	18.97	52.43	124.47	8.39
A	177.67	18.98	52.57	123.26	8.19
A	177.67	18.98	52.57	123.26	8.19
A	177.93	19.06	52.63	123.1	8.17
I	176.41	38.43	61.31	119.79	8.04
Q	175.84	29.44	55.49	124.36	8.47
R	175.99	30.67	55.97	122.76	8.32

Residue	CO	CB	CA	N	HN
H	174.25	29.7	55.84	120.45	8.47
Y	175.19	38.64	57.76	121.77	8.14
N	174.34	38.89	52.89	121.02	8.31
D	176.43	40.95	54.55	120.81	8.33
E	176.33	30.12	56.84	121.08	8.38
D	174.08	40.54	52.08	122.5	8.36
P	177.1	31.96	63.4	137.19	9999
E	176.56	30.1	56.53	121.41	8.54
K	176.69	32.87	56.51	122.48	8.38
E	176.49	30.15	56.7	122.15	8.42
K	176.57	32.85	56.32	122.51	8.36
R	176.15	30.68	56.15	122.43	8.33
I	176.13	38.55	61.13	122.49	8.16
K	176.68	32.88	56.25	125.46	8.46
E	176.4	30.07	56.66	121.68	8.39
L	177.5	42.16	55.36	123	8.25
E	174.14	29.66	54.31	122.68	8.38
P	177.1	32.02	63.14	137.52	9999
L	177.39	42.01	55.1	122.05	8.29
L	177.65	41.86	55.28	122.57	8.13
M	176.35	32.63	55.52	120.78	8.31
S	174.79	63.71	58.26	117.42	8.46
T	174.49	69.71	62.01	116.22	8.26
E	176.15	30.14	56.73	123.23	8.49
N	175.29	38.74	53.33	119.79	8.53
E	176.46	29.97	56.76	121.07	8.35
L	177.49	42.12	55.29	123.17	8.21
K	177.11	32.78	56.49	122.19	8.33
G	173.96	9999	45.33	110.01	8.46
Q	175.96	29.51	55.74	120.14	8.24
Q	175.69	29.49	55.75	121.88	8.43
V	176.01	32.56	62.29	121.71	8.18
L	174.97	41.64	52.91	127.46	8.32
P	177.1	32.02	63.14	137.52	9999
T	174.53	69.65	61.85	114.51	8.32
Q	175.53	29.42	55.92	122.69	8.45
N	175	38.77	53.11	120.14	8.46
H	174.97	29.6	56.13	119.7	8.48
T	174.06	69.72	61.81	115.78	8.13

Residue	CO	CB	CA	N	HN
A	177.81	19.12	52.59	126.53	8.37
S	174.46	63.65	58.36	114.91	8.21
Y	173.28	38.17	55.49	123.6	8.23
P	176.91	32.09	62.94	139.34	9999
G	173.9		45.06	109.08	8.63

Appendix B: Agadir Prediction Plots

Agadir predictions in the segment of c-Myb TAD that forms an alpha helix was plotted for WT and mutants

Figure B1. c-Myb WT Agadir Plot

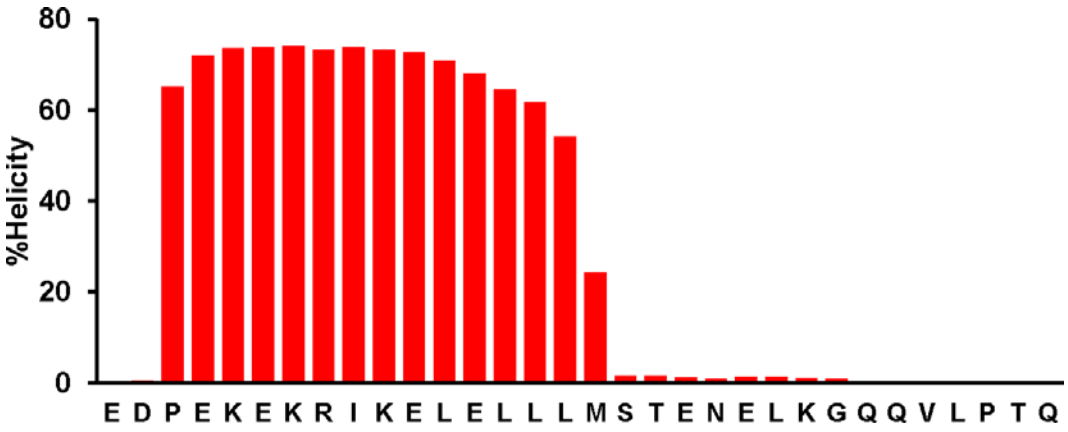


Figure B2. c-Myb E292D Agadir Plot

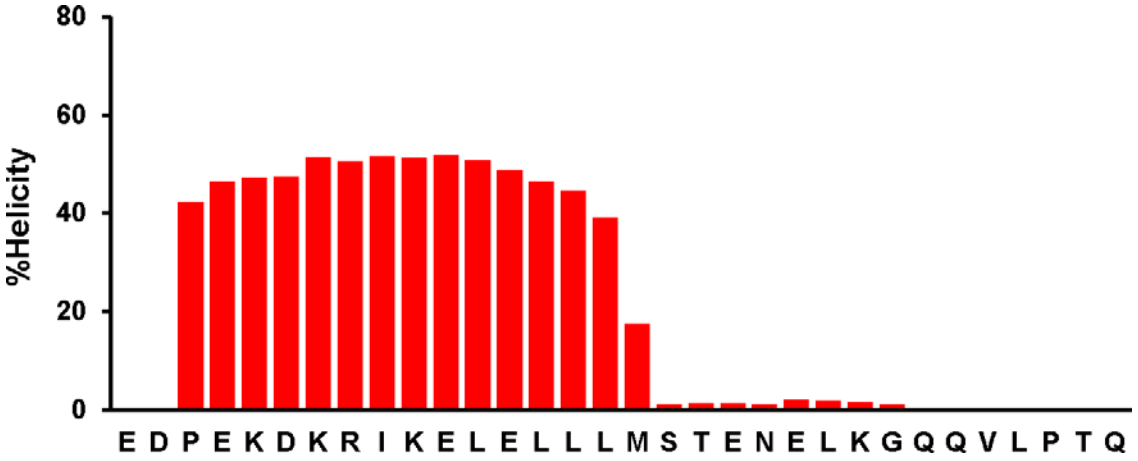


Figure B3. c-Myb K293H Agadir Plot

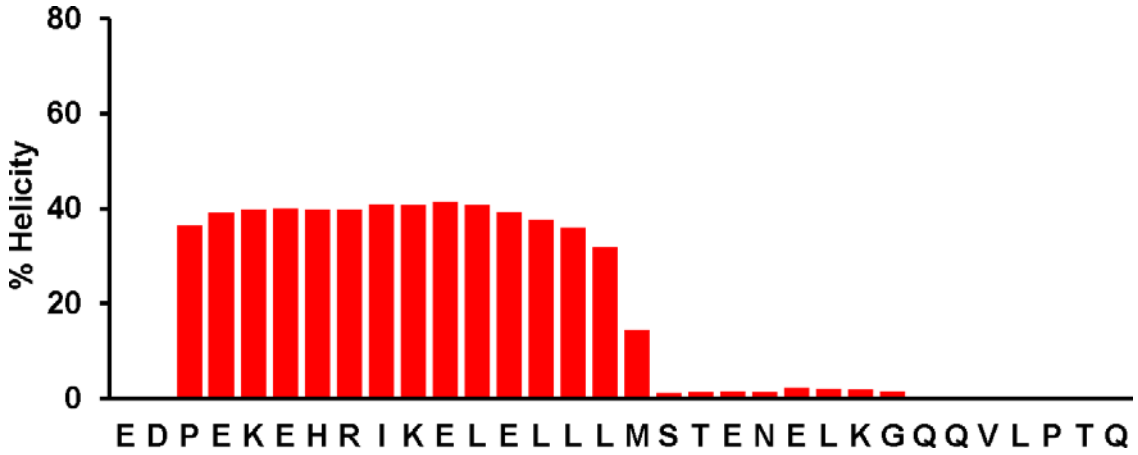


Figure B4. c-Myb K296H Agadir Plot

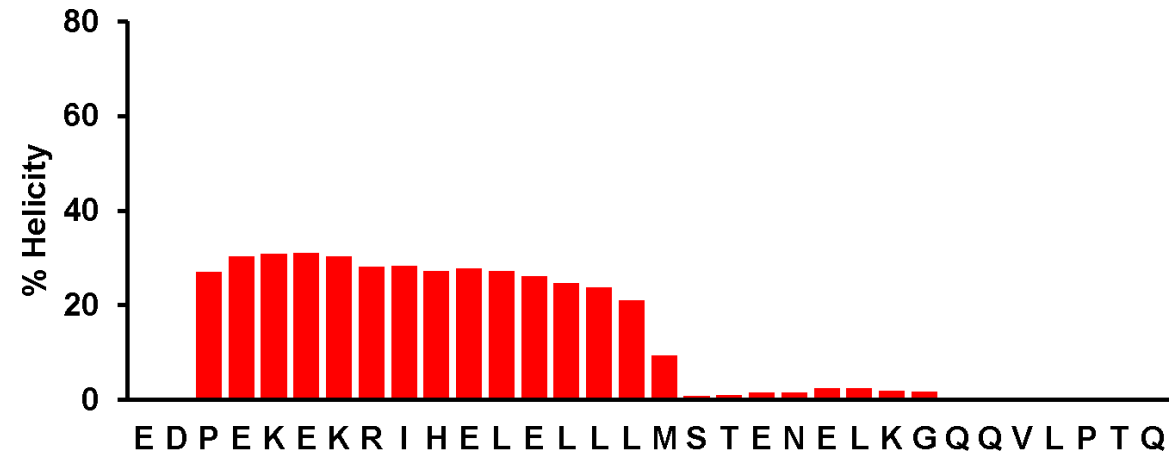


Figure B5. c-Myb L300V Agadir Plot

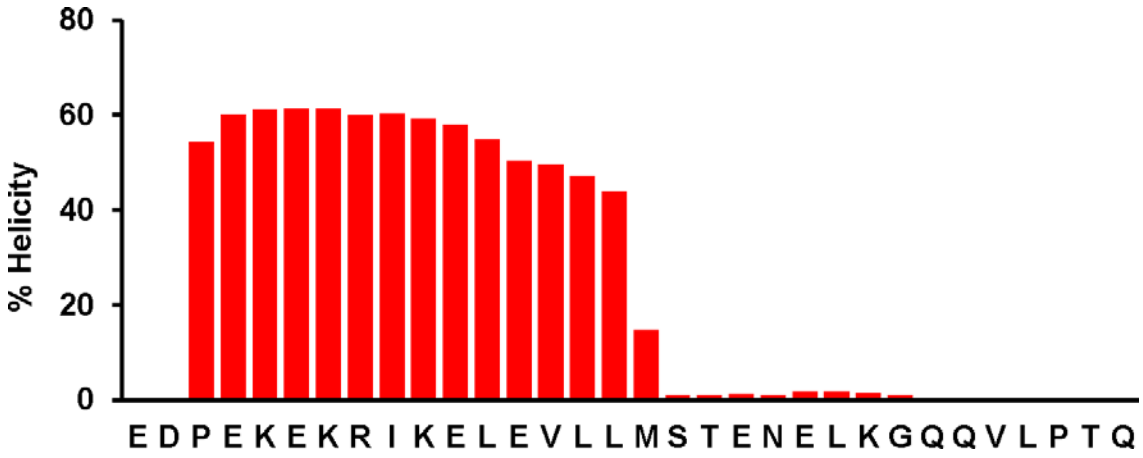


Figure B6. c-Myb L300G Agadir Plot

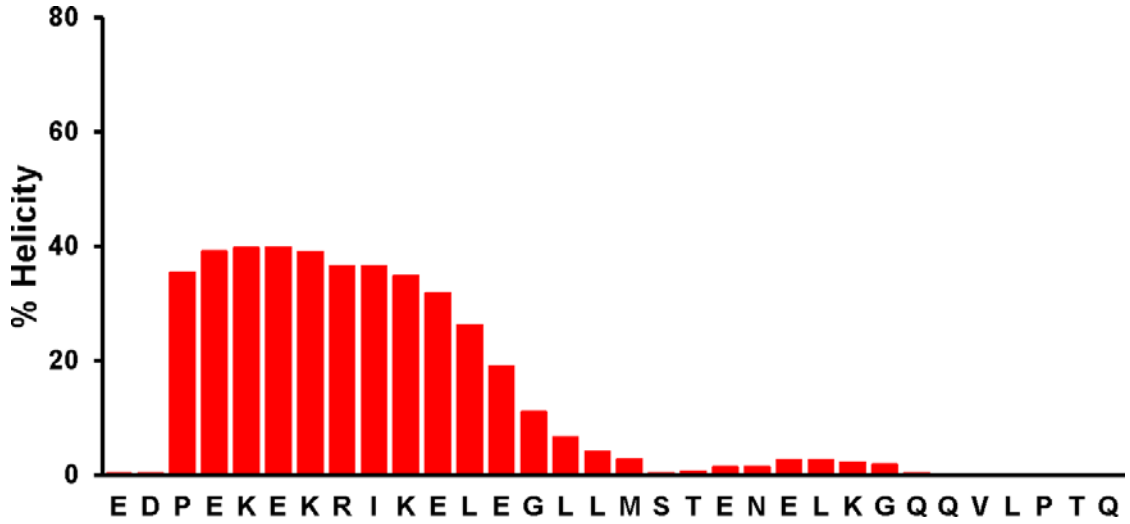
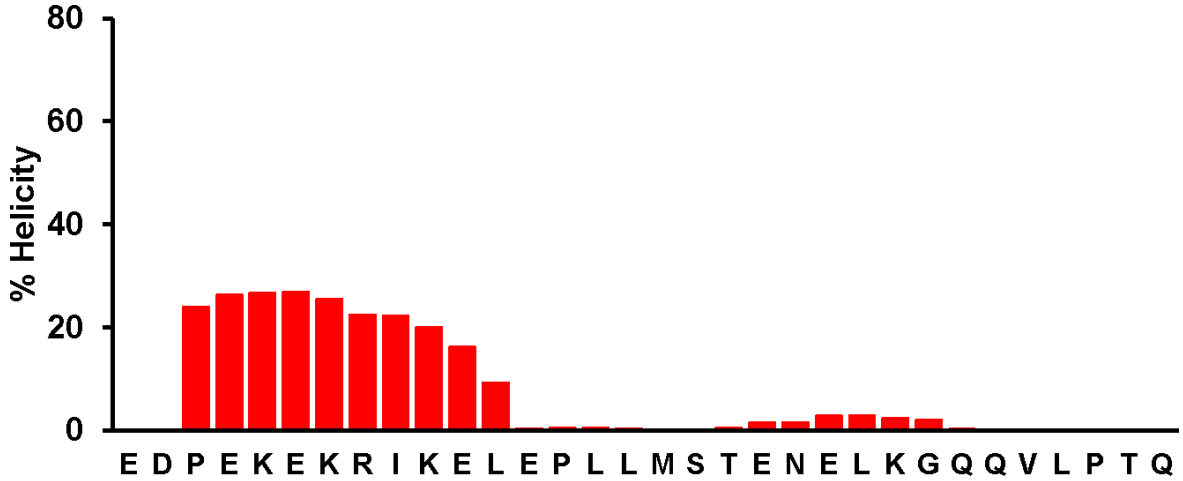


Figure B7. c-Myb L300P Agadir Plot



APPENDIX C: COPYRIGHT INFORMATION

Note. From “Conserved Helix Flanking Prolines Modulate Intrinsically Disordered Protein:Target Affinity by Altering the Lifetime of the Bound Complex,” by Crabtree MD et.al., 2017, *Biochemistry*, 56, p. 2379. CC-BY. Reprinted with permission.

Figure C1: License for use of published material

Conserved Helix-Flanking Prolines Modulate Intrinsically Disordered Protein:Target Affinity by Altering the Lifetime of the Bound Complex

Michael D. Crabtree[†], Wade Borcherds[‡§], Anusha Poosapati[‡§], Sarah L. Shammas[*†∥], Gary W. Daughdrill[*‡§], and Jane Clarke[*†]

[†] Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, U.K.

[‡] Department of Cell Biology, Microbiology and Molecular Biology, University of South Florida, Tampa, Florida 33620, United States

[§] Florida Center for Drug Discovery and Innovation, University of South Florida, Tampa, Florida 33612, United States

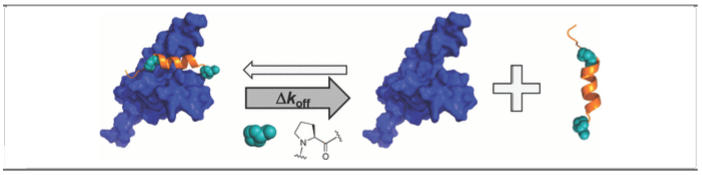
Biochemistry, **2017**, *56* (18), pp 2379–2384

DOI: 10.1021/acs.biochem.7b00179

Publication Date (Web): April 20, 2017

*E-mail: sarah.shammas@bioch.ox.ac.uk., *E-mail: gdaughdrill@usf.edu., *E-mail: jc162@cam.ac.uk.

Abstract



Appropriate integration of cellular signals requires a delicate balance of ligand–target binding affinities. Increasing the level of residual structure in intrinsically disordered proteins (IDPs), which are overrepresented in these cellular processes, has been shown previously to enhance binding affinities and alter cellular function. Conserved proline residues are commonly found flanking regions of IDPs that become helical upon interacting with a partner protein. Here, we mutate these helix-flanking prolines in p53 and MLL and find opposite effects on binding affinity upon an increase in free IDP helicity. In both cases, changes in affinity were due to alterations in dissociation, not association, rate constants, which is inconsistent with conformational selection mechanisms. We conclude that, contrary to previous suggestions, helix-flanking prolines do not regulate affinity by modulating the rate of complex formation. Instead, they influence binding affinities by controlling the lifetime of the bound complex.

Intrinsically disordered proteins (IDPs) or intrinsically disordered regions of proteins (IDRs) lack a well-defined three-dimensional fold and exist as ensembles of conformations with variable levels of transient structure.(1) A subset of IDRs fold upon binding to their partner macromolecule.(2) For IDRs that form α-helices upon coupled folding and binding, increasing the residual helicity of the free IDR has been shown to increase the affinity for the partner protein.(3-5) Conserved proline residues are commonly found flanking transiently helical binding segments of IDRs(6, 7) (Figure