Utilisation of NMR to study N-myc TAD dynamics and interactions

Ewa Barbara Rejnowicz

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Faculty of Biological Sciences, School of Molecular and Cellular Biology

August 2022

Intellectual property and publication statement

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Acknowledgements

Firstly, I would like to thank my supervisor Richard Bayliss for providing me with the opportunity to work on this project. His guidance and scientific insight, as well as genuine care for me as a person, were paramount in these last 4.5 years.

This work would not exist without the support of Eoin Leen. He was the best teacher, supporting me with kind words when I needed them and encouraging me to do better, when I was not giving it my best. There were many stimulating discussions that expanded my knowledge and helped me grow as a scientist, and many non-science related discussions that made me laugh. I would also like to thank Matt Batchelor for tireless effort of teaching me NMR principles and for the humour he brought when it became clear that my NMR aptitude was fairly limited. I cannot also underestimate his in-depth tutorials in Adobe Illustrator. I would also like to thank Jenny Miles for her help in setting up many experiments and for her encouraging and kind presence. To all the members of Bayliss group: Syed Ahanger, Vanda Gunning, Selena Burgess, Mark Richards, Iosifina Sampson, Jaimie Pitts, Liam Maddison and Amelia Lesiuk, thank you for your support, which was not just limited to advising me on science, but also lifted my spirits through many difficult moments. A big thanks also to Arnout Kalverda and Theo Karamanos for their help in NMR experimental set-up and data processing.

My first wobbly steps on my PhD journey were taken in Newcastle, and for that I would like to thank Jane Endicott, Richard Heath and Martin Noble for their patience and guidance. I would also like to thank Claire Jennings and Q-Tip for their moral support and many moments of laughter, and the entire Newcastle team for their support.

This work would not exist without the presence of my parents who made me the person I am today. Thank you for your unconditional love and support throughout the years. None of this would be possible without their faith that I can achieve anything in life. To my most amazing brothers for their mix of support, humour, and tough love, which only brothers can provide. I hope this puts to rest the ongoing family discussion of ‘are you ever going to stop studying?’ –‘No, I might do another PhD.’ And lastly to my best friend, my life partner and my rock, for his unfaltering love, support, long walks in beautiful Yorkshire which got me through the toughest of times.

Table of Contents

Table of figures

Abbreviations

2D	Two dimensional
3D	Three dimensional
bHLH LZ	Basic helix-loop-helix leucine zipper
bp	Base pair
CD	Circular dichroism
Cdk	Cyclin-dependent kinase
CPS	Chemical shift perturbation
cryo-EM	Cryogenic electron microscopy
CSI	Chemical shift index
CTD	C-terminal domain
DM	Double mutant
FID	Free induction decay
FT	Fourier transform
GFP	Green fluorescent protein
GSK	Glycogen synthase kinase
Gst	Glutathione – S - transferase
HSQC	Heteronuclear single quantum coherence
HTRF	Homogenous time resolved fluorescence
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IMAC	Immobilised metal affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
KD	Kinase domain
LB	Lysogeny broth
MB	Myc box
NB	Neuroblastoma
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOEsy	Nuclear Overhauser effect spectroscopy
NUS	Non-uniform sampling
O.D.	Optical density
PBD	Polo box domain
PCR	Polymerase chain reaction
PDB	Protein data bank
PIC	Pre-initiation complex
Pol II	RNA polymerase II
PP2A	Protein phosphatase 2A
pTEFb	Positive transcription elongation factor b
PTM	Post-translational modification
r.f	Radio frequency

RT	Room temperature
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SF9	Spodoptera frugiperda 9
sm-FRET	Single molecule Föster resonance energy transfer
TAD	Transactivation domain
TAF1	TATA-binding protein-associated factor 1
TBP	TATA-binding protein
TF	Transcription factor
TFE	2,2,2-fluoroethanol
TROSY	Transverse relaxation optimised spectroscopy
TRRAP	Transformation/transcription domain associated protein
WT	Wild type

Abstract

The myc family of proteins (c-, N- and L-myc) are transcription factors (TFs) responsible for maintaining the proliferative program in cells. 28% of tumours exploit these properties through deregulation of myc in various ways, which aids cancer through faster growth rates. Structural and biochemical knowledge of c-myc is lacking, with even less studies performed on N-myc, which is one of the main drivers of paediatric tumours. N-myc tumourigenic functions are mediated via intrinsically disordered transactivation domain (TAD), which makes N-myc TAD a suitable candidate for nuclear magnetic resonance (NMR) studies. NMR collects in-solution data on per-residue basis and can handle a dynamic ensemble of structures. Firstly, N-myc TAD backbone assignment was obtained, with the spectral resonant frequencies assigned to individual residues within primary amino acid sequence of N-myc TAD. Chemical shift index analysis, computer modelling and relaxation data revealed that N-myc TAD has two regions with helical propensity: one region within residues W^{77}–W^{88} and the second between A^{122}–E^{132}. Myc is rapidly degraded following mitogen signalling, through a hierarchy of post-translational modifications (PTMs). This work recreated these PTMs *in vitro* using NMR and confirmed that phosphorylation of T^{58} and S^{62} enhances N-myc binding to the Fbw7–Skp1 complex, the substrate-recognition element of E3 ubiquitin ligase, which mediates myc degradation. By employing various biochemical methods, putative partners of N-myc TAD were interrogated, which revealed that N-myc and c-myc do not have the same interactome. For example, N-myc TAD was found to not interact with TAF1-TBP or pTEFb. N-myc TAD interacts with Plk1 through its polo-box domain in the capacity of a binding partner. The novel data presented in this work underline the necessity for tailored research towards N-myc, as N-myc and c-myc might differ in their modality of interactions with binding partners, despite a degree of conservation between the two proteins. This fundamental research will aid our understanding of the mechanism by which the TAD and other intrinsically disordered protein regions utilise transient/latent structures to bind a wide range of partners, despite lacking a stable 3D structure.

1 Introduction

1.1 Myc family of transcription factors

1.1.1 Brief history of myc discovery

Myc was first discovered in the late 70s, when it was demonstrated that three avian tumour-causing retroviruses have an additional element, apart from the commonly seen *pol*, *gag* and *env* viral genes. This gene, following its sequencing in 80's was named *v-myc* (Sheiness and Bishop, 1979; Roussel et al., 1979; Vennstrom et al., 1982; Nair and Burley, 2003). The discovery that this virus can be carcinogenic fed into discussion at the time, whether carcinogenesis is brought about by genetics, infectious agents or environmental factors. Soon it was discovered that humans also carry an orthologue of *v-myc*, which was named *c-myc. v-myc* was most likely acquired by viruses from their hosts during the course of their evolution (Roussel et al., 1979; Bister, 1986; Conacci-Sorrell et al., 2014). Comparative sequence analysis of C-terminal regions between c-myc and other known oncogenes, such as fos, jun and yeast regulatory protein, GCN4, revealed that c-myc is a transcription factor (TF) – a protein characterised by its ability to recognise and bind specific DNA sequences and control transcription of genes (Vogt et al., 1987; Landschulz et al., 1988a). Further research has identified an obligatory partner of c-myc named myc associated factor X (Max) which is necessary for binding to DNA sequences (Blackwood and Eisenman, 1991). The employment of the southern blotting technique identified *n-myc* and *l-myc* genes within human genome, as highly amplified *c-myc* gene homologues in neuroblastoma cell lines and small cell lung cancer cell lines, respectively (Schwab et al., 1983; Nau et al., 1985; Kohl et al., 1983). Collectively, c-myc and N-myc will be referred to as myc, with L-myc omitted in this work due to there being very little research available on this protein.

1.1.2 Overview of structure of myc proteins

c-myc and N-myc are mostly intrinsically disordered, i.e. they lack a stable tertiary structure and are able to assume multiple conformations within short time scales (Wright and Dyson, 2015). At the N-terminus myc proteins are characterised by the presence of five conserved stretches of ~20 residues known as Myc boxes (MBs) 0, I, II, IIIa, IIIb and IV (Figure 1.1.A). Myc proteins also possess a nuclear localisation signal (NLS) which spans residues K^{348} – K^{349} – I^{350} – K^{351} in N-myc. Another common motif is the PEST sequence, which is rich in proline, glutamic acid, serine and threonine residues. In c-myc this domain occupies residues T^{226} – K^{270} (Gregory et al., 2003b), however, this region is poorly conserved between the two paralogues and thus establishment of the exact position of this domain in N-myc, whose sequence is abundant in these residues, is challenging. The role of this sequence is not fully understood, albeit it is commonly seen in proteins with a short half-life such as myc, thus it has been speculated that it mediates degradation of proteins (Rogers et al., 1986). The C-terminal region of myc proteins comprises of DNA binding domain (DBD) and it heterodimerises with Max.

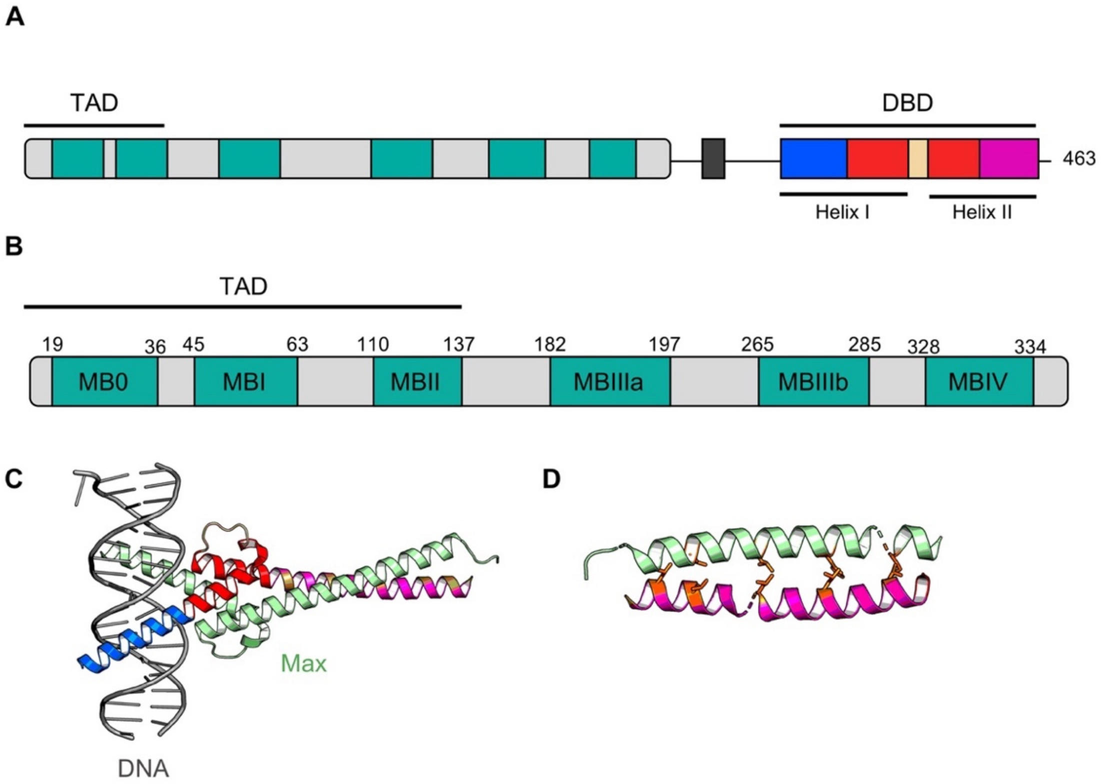


Figure 1.1. The organisation of myc domains.
A. Full-length N-myc. Green boxes at the N-terminus signify MBs I – IV, the dark grey box is the nuclear localisation signal (NLS), helix I is composed of the basic region (blue), loop region (wheat), and helix II contains a leucine zipper (pink). Light grey represents a sequence of N-myc which is not conserved in c-myc and has unknown function. Colours correspond to the structure in C. **B.** The positioning of MBs (green) with N-terminal portion of N-myc protein and their relation to TAD. **C.** Crystal structure of c-myc DBD (multicolour) with Max (light green) and DNA (grey). The colour scheme corresponds to the elements listed in A. **D**. Leucine zipper between c-myc (pink) and Max (light green). Leucines and other hydrophobic residues participating in the leucine zipper are highlighted in orange. (Nair and Burley, 2003). The boundaries for the conserved regions and domains presented in A and B are not to scale and are the estimates.

1.1.2.1 Transactivation domain

The N-terminal portion, spanning MB0 – MBII (Figure 1.1.B), is a transactivation domain (TAD), an essential domain of transcription factors which regulates transcription through the recruitment of other transcription factors, various regulators and transcriptional machinery. N-myc's TAD is localised to the first N-terminal 137 amino acids and contains three MBs (MB0, MBI and MBII). This region is critical for myc regulation of transcription, its stability and transforming abilities (Kato et al., 1990). Truncation of this region causes myc to lose its ability to transform cells, as evidenced by the impaired ability of MycS – a naturally occurring c-myc version lacking first 100 amino acids, to induce transformation (Xiao et al., 1998). Although MB0 is a site of multiple myc interactions, including Aurora A, Bin and Pin1, amongst others, very little is known about its structure and functional significance (Andresen et al., 2012; Helander et al., 2015; Richards et al., 2016; McMahon et al., 1998). MBI is an important site for regulating myc stability as it contains a phospho-degron sequence, which when phosphorylated in sequential fashion, signals for myc degradation (Sears et al., 2000; Yeh et al., 2004; Henriksson et al., 1993; Pulverer et al., 1994; Lutterbach and Hann, 1994). Lastly, MBII, similar to MB0 is a site of important interactions and it directly interacts with Transformation/transcription domain-associated protein (TRRAP complexes; McMahon et al., 1998). Apart from the MBs, there is an important region of interaction with Aurora A, which falls between MBI and MBII (Richards et al., 2016). This region is partially conserved in c-myc and it controls the interaction with TATA binding protein (TBP) and TBP-associated factor 1 (TAF1; Wei et al., 2019).

1.1.2.2 DNA-binding domain

Through its DBD, c-myc heterodimerises with its obligatory partner, Max, and recognises canonical E-box sequences (5′-CACGTG-3′) which are abundant in enhancer regions of the human genome (one per every 28 nucleotides; Figure 1.1 C; Vogt et al., 1987; Landschulz et al., 1988b; Blackwood and Eisenman, 1991; Prendergast et al., 1991; Nair and Burley, 2003; Desbarats et al., 1996; Chang et al., 2015). The basic helix-loop-helix leucine zipper (bHLH LZ) motif is commonly seen in other TFs and is composed of two helices joined by the linker (Figure 1.1 C). The N-terminal helix (helix I), which binds DNA, is rich in basic residues, which allows myc to interact with acidic DNA backbone (Brown et al., 1985; Landschulz et al., 1988a; Nair and Burley, 2003). The second helix (helix II) interacts with Max in the form of two intertwined helices associated through a leucine zipper (pink; Figure 1.1.C).

This is a common motif seen in TFs, with a heptad repeat of leucines which aids dimerisation through the formation of a hydrophobic core (Nair and Burley, 2003). While the crystal structure is of a c-myc:Max dimer, there is evidence suggesting that N-myc uses the same mechanism of heterodimerisation and E-box recognition. The bHLH LZ region of c- and N-myc is highly homologous, displaying nearly 50% sequence identity, indicating that the same interactions regulate the two myc proteins. Secondly, site-directed mutagenesis of the heptad-repeat leucine residues in the leucine zipper abrogates N-myc transforming abilities (Nakajima et al., 1989).

The half-life of Max in cells is relatively long (about 24 h) whereas c-myc is characterised by a short half-life of around ~20 min, indicating that myc's life span is rate-limiting in Max-myc binding to DNA (Ayer et al., 1993; Hurlin et al., 1997; Blackwood et al., 1992a; Blackwood et al., 1992b; Hann and Eisenman, 1984). In the absence of c-myc, Max can heterodimerise with other proteins that belong to the bHLH LZ family, namely Mnt, Mad, Mga and Miz-1, and they share a lot of characteristics with myc, with DBDs, short half-lives and poor homodimerisation properties (Brubaker et al., 2000; Ayer et al., 1995). Lastly, Max can homodimerise, forming a weak dimer, which readily dissociates in the presence of other myc partners (Nair and Burley, 2003; Ayer et al., 1993).

1.1.3 The evolutionary origins of myc proteins

Myc was initially discovered in retroviruses (v-myc), however, Southern blotting later revealed that the sequence was evolutionarily acquired from their host (Roussel et al., 1979; Vennstrom et al., 1982; Bister, 1986). Evidence suggests that myc is present in primitive metazoans, such as diploblastic cnidarian and species closely related to metazoans - choanoflagellate, and is not present in fungi or plants, indicating that myc appeared in the genome following the divergence from these kingdoms (Mahani et al., 2013). The C-terminal bHLH LZ, present in primitive species, has the ability to recognise E-boxes, suggesting that this part of the protein has evolved first (Young et al., 2011; King et al., 2008; Mahani et al., 2013). The N-terminal region of myc proteins is much more versatile, less conserved and appears later in the evolutionary pathway (Mahani et al., 2013; Young et al., 2011). MBI is well conserved across mammals, insects and other distant metazoans and even primitive marine species, such as Cnidaria, but not in species like *C.owczarzaki* which is a primitive unicellular eukaryote (Mahani et al., 2013; Young et al., 2011). MBII and MBIIIb are also well-conserved between species, appearing in vertebrae, insects, and cnidarian (Young et al., 2011). MBIV appears to be a vertebrae-specific motif, present in Zebrafish, but absent in lower complexity species such as insects and tunicatas (Cowling and Cole, 2006; Mahani et al., 2013). Other MBs were not investigated. To summarise, myc originates in primitive metazoans, as a simple DNA binding protein, acquired the N-terminal MB-containing region during later evolution, and is well-established in higher metazoans.

1.2 Myc degradation pathway

1.2.1 Skp-Cullin-F-box proteasomal system

Myc is targeted for proteasomal degradation by Skp-Cullin-F-box (SCF^{Fbw7}), a complex composed of several proteins, each working in concert to recognise the phosphodegron substrate, activate and conjugate ubiquitin for ligation and catalyse the ubiquitination of a recognised substrate (Weissman, 2001). There are three main components of this system: E1 (ubiquitin-activating enzyme), E2 (ubiquitin-conjugating enzyme) which generates ubiquitin chains, and E3 which is a ubiquitin ligase, responsible for adding ubiquitin chains onto a recognised substrate (Zhou et al., 2015; Hershko and Ciechanover, 1998). In addition, the complex features scaffolding proteins, which form bridges between different enzymes and substrates and a ubiquitin-substrate recognition element (Jia et al., 2011; Zheng et al., 2002).

The largest component of SCF^{Fwb7} is Cullin 1 protein, which is the main scaffolding protein within the complex (grey; Figure 1.2 A; Hao et al., 2007). It contains three rigid five-helix cullin repeats acting as a ‘stalk’. Its C-terminus forms a globular Cullin-homology domain with a single β sheet, used to bind Ring box 1 (Zheng et al., 2002; Hao et al., 2007). Rbx1 is a small, ~100 amino acid globular zinc-binding protein with Ring-H2 domain and its role is to recruit E2 conjugating enzyme (Petroski and Deshaies, 2005; Hao et al., 2007). The exact E2 conjugating enzyme that ubiquitinates Myc is not known, however RING-Cullin complexes often recruit cdc34, UBC4/5 or UBC12, and so it is likely to be determined by tissue type and physiological factors (Petroski and Deshaies, 2005; Hao et al., 2007).

At the N-terminus, Cullin 1 interacts with Skp1 through its first cullin repeat (grey; Figure 1.2 A). Skp1 is 163 amino acids folded into eight helices and three β-turns forming a β-sheet and it interacts with F-box and WD repeat domain containing 7

(Fbw7) protein through its two C-terminal helices (Schulman et al., 2000; Hao et al., 2007). Fbw7 possesses a WD40 repeat domain, which is rich in positively charged residues, that mediate interaction with multiple phospho-motifs (Figure 1.2 C; Schulman et al., 2000).

The gap between the substrate recognition part of the complex and the ubiquitin-conjugating parts of the complex is estimated to be 50 Å (Petroski and Deshaies, 2005). Once the substrate is positioned in the proximity, activated ubiquitin chains on the E2 enzyme are conjugated onto the side chain of a specific lysine residue through an isopeptide bond, with K^{52} suggested for both c-myc and N-myc (Zheng et al., 2002; Jaenicke et al., 2016).

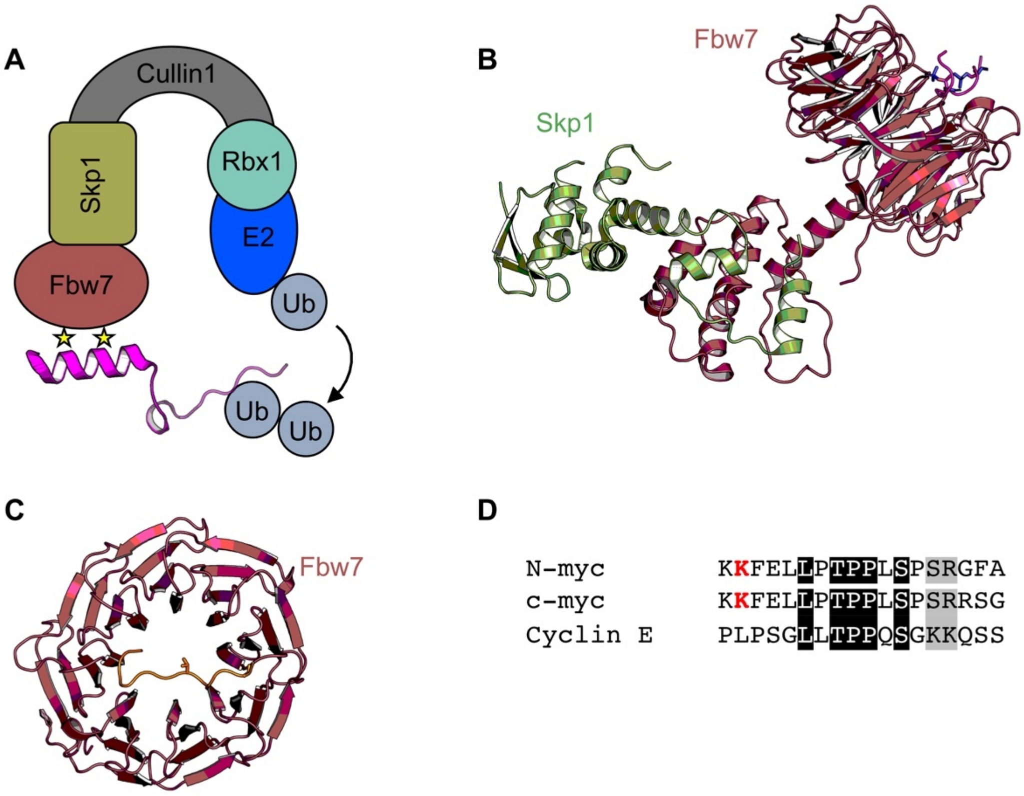


Figure 1.2. SCF^{Fbw7} complex.
A. Schematic representation of SCF^{Fwb7} complex. Myc is marked in pink and stars symbol the phosphorylation sites. Cullin1 is the major scaffolding protein, which at its N-terminus interacts with Skp1 (green). Skp1 then recruits Fbw7 (red) which is responsible for recognizing the ubiquitin substrate. At the C-terminus, Cullin1 interacts with Rbx1 (light green), the main role of which is to recruit E2 ubiquitin conjugating enzyme (blue) and adds ubiquitin (light blue) onto the Fbw7-anchored substrate. **B.** Crystal structure of Fbw7-Skp1 complex. Colour scheme is as per panel A (Hao et al., 2007). **C.** Front view of the Fbw7 WD40 domain which recognises the substrate (in this case the phosphodegron of cyclin E, marked in orange). **D.**

Sequence alignment of N-myc, c-myc and cyclin E phosphodegrons. Conserved residues are highlighted in black and residues with the similar chemistry are highlighted in grey. The putative ubiquitination site (K^{52}) is highlighted in red (Hao et al., 2007; Petroski and Deshaies, 2005).

1.2.2 Regulation of N-myc protein stability through PTMs

Substrates of SCF^{Fbw7} E3 ligase are recognised by a sequence known as a phosphodegron, which is recognised by Fbw7, and must thus first be phosphorylated. In both, c- and N-myc this sequence spans residues $P^{57}TPPLSP^{63}$ (Figure 1.2 D). For recognition by Fbw7, phosphorylation is required on both T^{58} and S^{62} or just T^{58} alone, with reports conflicting on this issue (Sears et al., 2000; Yeh et al., 2004; Hao et al., 2007).

The first modification is phosphorylation of S^{62} (p-S^{62}; Figure 1.3; Sears et al., 2000) and it is often referred to in the literature as stabilising phosphorylation, as an S^{62A} mutant is characterised by a shorter than normal half-life and the lack of transforming ability (Sears et al., 2000; Yeh et al., 2004; Henriksson et al., 1993; Pulverer et al., 1994). As this site has a sequence p-[S/T] – [P], (where p stands for phosphorylated residue) proline-directed kinases extracellular signal-regulated kinase 1 (ERK1) and the cyclin-dependent kinase 1 – cyclin B complex (Cdk1:cyclin B) have been proposed to phosphorylate c-myc and N-myc *in vivo*, respectively (Sears et al., 2000; Sjostrom et al., 2005). Cdk2:cyclin A, JNK or Cdk5:cyclin G_1 have also been shown to phosphorylate S^{62} in c-myc *in vitro* (Helander et al., 2015; Seo et al., 2008; Sjostrom et al., 2005). The selectivity of kinases towards their substrates is often dictated by factors other than just the consensus site, thus their *in vivo* abilities to phosphorylate this site might be restricted by expression profiles, subcellular localisations, signalling and different patterns of gene expression, amongst others.

Priming phosphorylation of S^{62} is a prerequisite for phosphorylation of T^{58} by the glycogen synthase kinase 3 (GSK3) family of kinases, with both α and β isoforms being suggested in the literature as the kinase for this site (Figure 1.3; Sears et al., 2000; Cohen and Frame, 2001; Pulverer et al., 1994; Lutterbach and Hann, 1994; Gregory et al., 2003b). GSK3, unlike most other kinases, lacks an activatory ('priming') phosphorylation residue within its activation loop. For example, in Plk1 this residue is T^{210} and it requires a phosphorylation by Aurora A to achieve its full catalytic activity (Macůrek et al., 2008; Seki et al., 2008). The structural basis for GSK3 priming has been resolved by obtaining a crystal structure, which features a HEPES molecule (pink, Figure 1.3) with its sulphonate group coordinated by R^{96}, R^{180}, K^{205} and V^{214} residues (yellow, Figure 1.3). In physiological conditions, this molecule sits in place of the phosphorylated +4 residue of a substrate and is coordinated by the same residues (Dajani et al., 2001; Bax et al., 2001).

GSK3 kinase recognises the consensus site [S/T] – [X] – [X] – [X] – p-[S/T] (where X stands for any amino acid type; Sutherland, 2011; Doble and Woodgett, 2003) and both c- and N-Myc fit this consensus site with a P^{57}TPPLpSP^{63} phosphodegron. The expression of kinase-dead GSK3β leads to c-myc accumulation (Sears et al., 2000). GSK3β is still able to coimmunoprecipitate with c-Myc^{S62A} better than with c-myc^{WT}, suggesting that T^{62} phosphorylation is not critical for GSK3β binding *per se* but helps the kinase to assume its active conformation (Gregory et al., 2003b).

Subsequent steps of the degradation pathway, occurring post T^{58} phosphorylation are less well-defined and disputed. Peptidyl prolyl *cis/trans* isomerase NIMA-interacting (Pin1) isomerises the peptidyl-prolyl bond between S^{62} and P^{63} from the *cis* to the *trans* isoform, however, the role of Pin1 is controversial (Yeh et al., 2004; Helander et al., 2015; Sears et al., 2000; Welcker et al., 2022). The PTMs that are required for Pin1 recognition of myc substrate are not very clear. Studies of fibroblasts have indicated that T^{58} phosphorylation is critical for the interaction between c-myc and Pin1 (Yeh et al., 2004). This is in opposition to *in vitro* studies performed between c-myc$_{1\text{-}88}$ and Pin1, where T^{58} phosphorylation was dispensable for binding (Helander et al., 2015). In addition, the role of pS^{62} is not clear. Biolayer interferometry data suggest that phosphorylation of this site reduces the binding of Pin1 to c-myc$_{1\text{-}88}$, although Pin1 is a phosphorylation-directed isomerase (Yeh et al., 2004; Helander et al., 2015; Lu and Zhou, 2007). In addition, the catalytic context of this enzyme is not fully known, with conflicting reports indicating that it catalyses *cis*-to-*trans*, *trans*-to-*cis* or both types of isomerisations (Helander et al., 2015; Zhou et al., 2000; Guito et al., 2014).

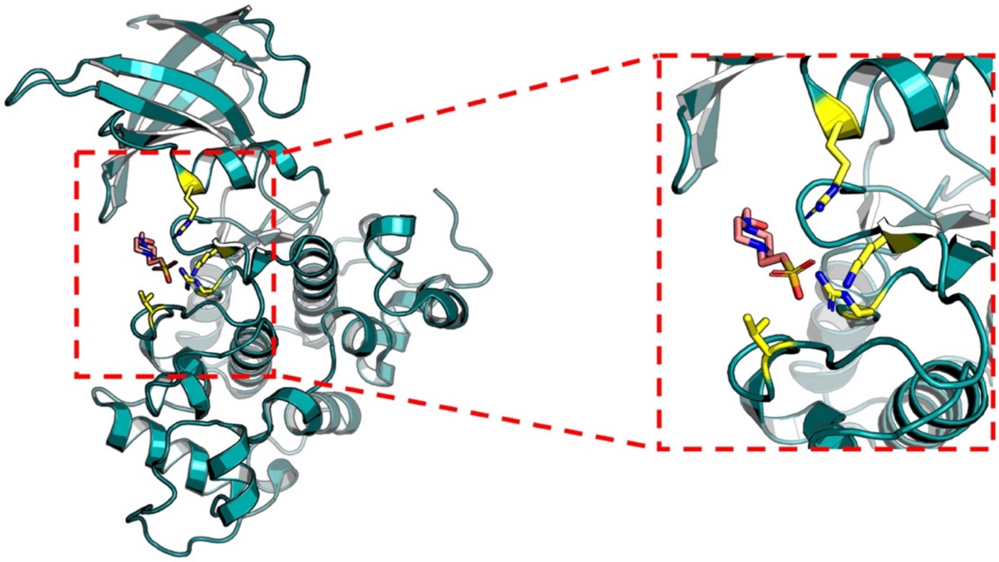


Figure 1.3. Structure of GSK3β.
GSK3β requires a priming phosphorylation at position +4. In myc protein, this site corresponds to the S^{62} residue. The crystal structure of the kinase revealed that GSK3β lacks an activatory phosphorylation site. Instead, this site is filled with the phospho-group at the +4 position of the substrate. In the crystal structure above, R^{96}, R^{180}, K^{205} and V^{211} (yellow) coordinate a HEPES molecule (pink), which possesses a sulfonate group, mimicking the phospho-residue (Dajani et al., 2001).

Peptidyl-prolyl bond isomerisation primes c-myc for the dephosphorylation of S^{62} by protein phosphatase 2A (PP2A; Figure 1.4). PP2A is a proline-directed phosphatase that has been reported to prefer a *trans*- conformation of the proceeding peptide bond, however 90 – 95% of the prolyl-peptidyl bonds within proteins are already in this isoform (Yeh et al., 2004). Mono-phosphorylated p-T^{58} Myc is then directed for ubiquitin conjugation through SCF^{Fwb7} (Weissman, 2001).

The recently published crystal structure of Fbw7 recognising a di-phosphorylated c-myc_{47-66} peptide, suggests that both Pin1 and PP2A are dispensable in the c-myc degradation pathway (Welcker et al., 2022). Similar to the recognition of cyclin E, Fbw7 binds at two phospho-sites within each of the phosphodegrons, spaced apart by four residues. Fbw7 also dimerises and recognises the second c-myc phosphodegron localised to residues c-$myc_{235-265}$, which is poorly conserved in N-myc (Welcker et al., 2022). This will be further explored in Chapter 5.

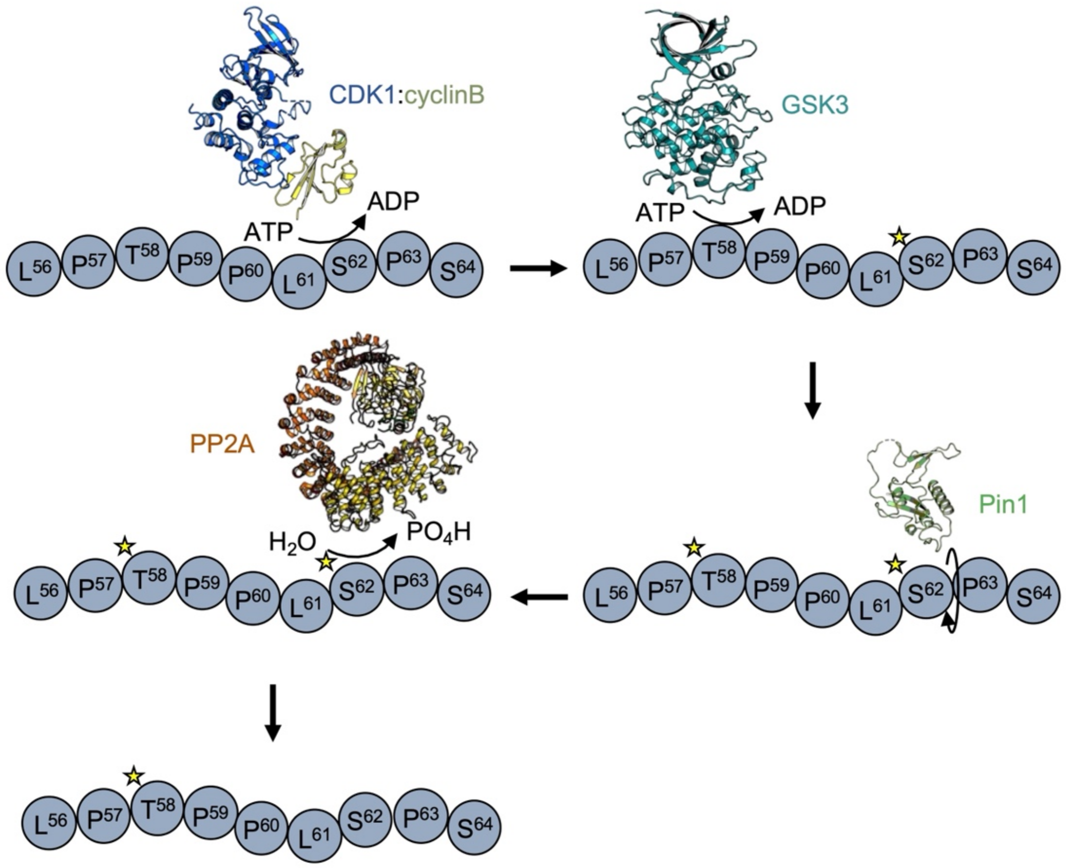


Figure 1.4. Myc degradation pathway.
S^{62} is phosphorylated by CDK1:cyclin B (N-myc) or ERK (c-myc). This is a priming event for the phosphorylation of T^{58} by the GSK3 family of kinases. Pin1 isomerises the peptidyl-prolyl bond from the *cis* to *trans* isoform on position S^{62}-P^{63}, which allows PP2A to dephosphorylate S^{62} (Sears et al., 2000; Yeh et al., 2004; Welcker et al., 2022; Helander et al., 2015).

1.2.3 The role of S^{62} phosphorylation

Data from multiple sources agree that phosphorylation of T^{58} is a necessary step for Fbw7 binding and that abrogation of phosphorylation on this site abolishes the interaction and prolongs myc protein half-life (Sears et al., 2000; Welcker et al., 2022; Yeh et al., 2004). The role of p-S^{62} is less clear with the modification being known to both stabilise myc and propel myc protein on its path to degradation (Sears et al., 2000; Welcker et al., 2022; Yeh et al., 2004).

The presence of p-S^{62}phosphorylation greatly enhances the affinity of c-myc to Fbw7 by 150-fold according to competition assays, performed with either mono- (p-T^{58}) or di-phosphorylated c-myc peptides (Welcker et al., 2022). This is in agreement with data from Yada and colleagues, which demonstrates that di-phosphorylation is critical for the binding of Fbw7, and their substitutions to alanine impair c-myc interaction with Fbw7 *in vitro* (Yada et al., 2004). The S^{62A} mutation prolongs c-myc's half-life in U2OS cells, however, this mutation would also abrogate T^{58} phosphorylation, thus it was impossible to tease apart the individual role of these two modifications (Welcker et al., 2022). This is in contrast to the data demonstrated by Sears, which indicates that S^{62A} has a half-life similar to c-myc^{WT}, most likely due to the lack of S^{62} phosphorylation (Sears et al., 2000). This mutant would not be phosphorylated on T^{58} which most likely causes its enhanced stability. Phosphorylation of S^{62} is also necessary for c-myc degradation as co-transfection with either S^{62A} or T^{58A} S^{62A} c-myc prevents Fbw7-mediated degradation (Welcker et al., 2004). S^{62} mutation is not a mutation hot spot in cancers, as observed with T^{58A} mutations, indicating that the presence of phosphorylation of S^{62} is necessary for cell transforming ability (Tate et al., 2019). A picture that emerges is that S^{62} has a nuanced role in both myc stability and transforming abilities.

1.3 Roles of myc in cells

1.3.1 Myc's role in differentiation and proliferation

Myc is a pleiotropic TF that regulates multiple functions within cells. An exhaustive review of all its functions is beyond the scope of this work. To name a few, myc is involved in metabolism, angiogenesis, apoptosis, cell cycle, amongst many others (Meyer and Penn, 2008; Dang, 2012). Myc is able to promote seemingly opposite processes, such as differentiation and proliferation, indicating that its functions are context dependent.

The seminal work conducted in 2006 by Takashi and Yamanaka cemented the critical role c-myc plays in proliferation of cells. c-myc was one of four factors, in addition to Klf4, Oct4 and Sox2, necessary for de-differentiation of fibroblasts into induced pluripotent stem cells (Takahashi and Yamanaka, 2006). Depletion of c-myc in de-differentiating fibroblasts arrested the process, which was rescued by c-myc^{WT} or a c-myc^{T58A} mutant (Takahashi and Yamanaka, 2006). Fibroblasts induced with Klf4, Oct4 and Sox2, but not c-myc were still able to reprogram, albeit the phenotype took longer to manifest, indicating lower rates of cell division (Wernig et al., 2008). This indicates that c-myc brings a specific role to the stem cell-like phenotype through promotion of cell proliferation and cell division, rather than through reprogramming *per se* (Wernig et al., 2008; Cowling et al., 2007; Jaenicke et al., 2016).

On the other hand, there have been studies showing that overexpression of c-myc can lead to an increase in cell mass, without driving proliferation. This effect was observed in B lymphocytes and the P493-6 B-cell line, as well as in *Drosophila* d-myc (Iritani and Eisenman, 1999; Schuhmacher et al., 1999; Johnston et al., 1999). The mechanism through which myc triggers biomass accumulation, without propelling cell cycle progression is unclear.

N-myc has an indispensable role in the early embryogenesis in mice. Complete and global N-myc knock-out causes murine embryos to develop normally until 10.5 days old, followed by sudden embryo death on the subsequent day (Sawai et al., 1993; Stanton et al., 1992). Prior to their death, they displayed normal physiology and development, suggesting the importance of N-myc in a post-blastocyst stage. Murine embryos also displayed a similar phenotype with a global knock-out of c-myc, with normal development up until day 10.5 of gestation when embryonic lethality ensued (Davis et al., 1993). This embryonically lethal phenotype in mice lacking c-myc can be rescued with N-myc, but not *vice versa* (Malynn et al., 2000). c-myc role in proliferation has long been known (Takahashi and Yamanaka, 2006; Schmidt, 1999). N-myc has been less well investigated, however data indicate that it is both important for proliferation and differentiation in neuronal progenitor cells and murine lung progenitor cells (Knoepfler et al., 2002; Okubo et al., 2005). In lung cell progenitor cells, N-myc is essential in maintaining pluripotency of epithelial cells and its overexpression inhibits differentiation (Okubo et al., 2005). This indicated that the level of N-myc protein determines whether a cell will proliferate or undergo differentiation. This demonstrates a functional difference between N- and c-myc, where c-myc acts as a propeller of the cell cycle, regardless of its expression levels, whereas N-myc can regulate both proliferation and differentiation, depending on its cellular concentration.

During gestation, both proliferation and differentiation play important roles and are finely regulated. It has not been assessed whether myc knock-out caused the sudden pause in proliferation which induces lethality. It is also plausible that despite the similar phenotype (death on day 10.5 of gestation), c- and N-myc are causing embryonic lethality through different mechanisms. c-myc expression is more global and is a necessary factor for proliferation (Takahashi and Yamanaka, 2006). N-myc expression is more restricted and N-myc has been known to drive differentiation of neural tissue.

1.3.2 Myc in normal cells

1.3.2.1 Control of global transcription levels

Myc proteins as general TFs are thought to regulate around ~15% of all genes in humans, whether it is gene repression or gene activation (Zeller et al., 2006). Mycs are devoid of any catalytic activity, rather they act as a versatile platform that assembles transcriptional machinery. The methods used to measure the levels of transcription, often rely on RNA read-outs and these suffer from two major issues. Firstly, enhanced levels of mRNA do not necessarily translate to an increase in protein levels. Secondly, these methods usually standardise pre-myc levels of mRNA, disregarding subtle differences in basal transcription levels between different genes (Patange et al., 2022). Models have been developed that attempt to describe how myc regulates transcription globally within a cell, however, the research is technically challenging. It is beyond the scope of this review to describe the competing theories; however, the two most prominent ones are briefly outlined.

The first model states that myc is an amplifier of global transcription and it recruits transcriptional machinery to all active loci, both within tumours and healthy cells (Kress et al., 2015; Chandriani et al., 2009). This leads to the amplification of a specific cell signature when myc levels are high (Figure 1.5; Kress et al., 2015). Recent studies by Patange and colleagues, revealed that myc increases the dwell time of general TFs within active loci, which leads to an increase in expression levels (Patange et al., 2022). Myc has been known to globally increase the levels of Serine 2 phosphorylation of Polymerase II C-terminal domain (CTD), a marker of global transcriptional activation (Lu et al., 2015).

The second model, dubbed ‘selective’, states that myc has its own signature expression pattern, which leads to the activation of other TFs (Figure 1.5 B). Meta-analysis of myc expression patterns showed very little overlap in myc-driven gene expression across multiple tissues (Rahl and Young, 2014; Kim et al., 2006; Chandriani et al., 2009). Proponents of this model argue that this is not a direct result of myc activity. Instead myc acts as a master TF acting on its downstream TF targets, which then execute their specific programs (Leone et al., 2001).

Myc not only binds to canonical E-box sequences but can also target sub-optimal E-box sequences, usually when it reaches higher-than-normal physiological levels (Allevato et al., 2017). This phenomenon can be explained by both of amplifier and selective models. In the ‘amplifier’ model, occupation of non-canonical E-boxes is dependent on whether these loci are activated within this specific tissue. In the ‘selective’ model, occupation of low-affinity sites at super-optimal c-myc concentrations, leads to activation of downstream TFs that are usually not regulated by c-myc at physiological levels. In addition, c-myc can recruit histone acetyltransferases (HATs) which activate various gene loci through relaxation of chromatin, again in support of both of the models (Neri et al., 2012).

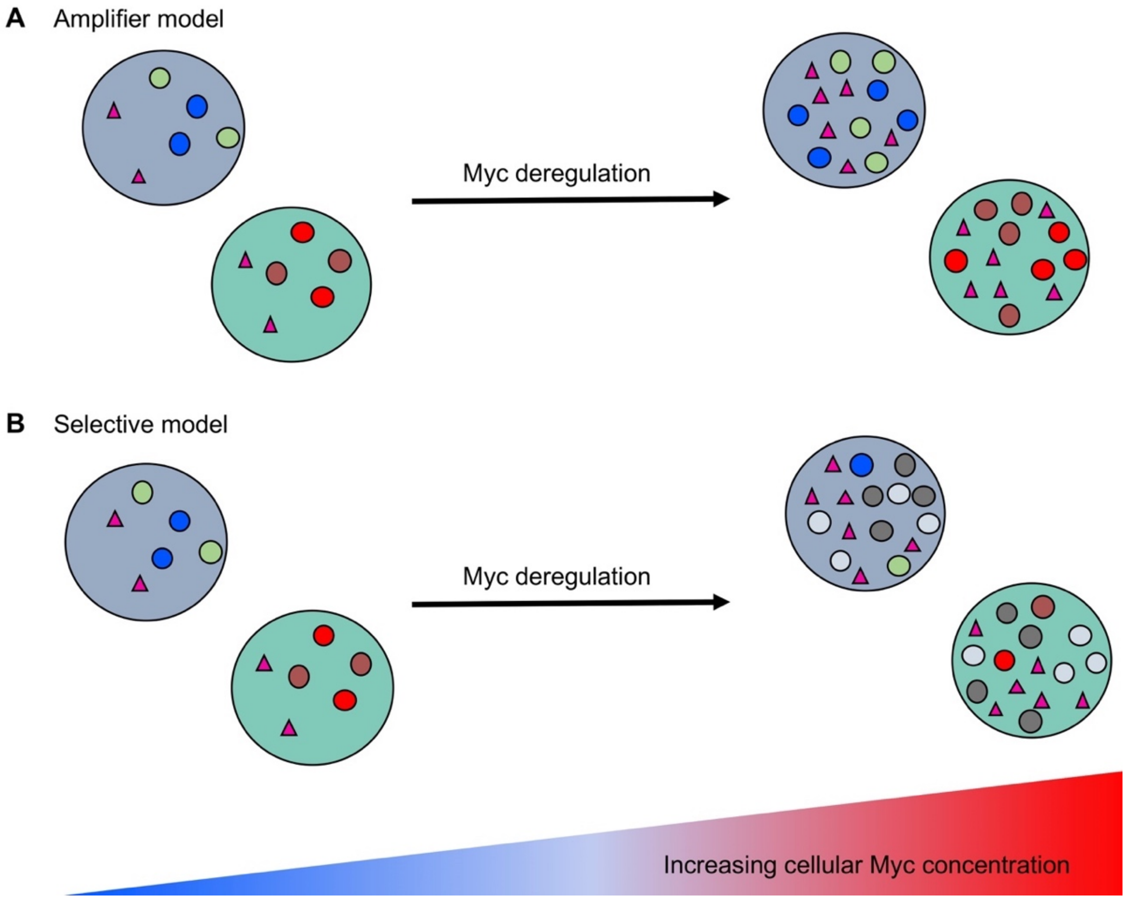


Figure 1.5. Models of myc transcriptional regulation.
A. At normal physiological conditions, myc (pink triangle) regulates the transcription of cell-specific gene signature (different coloured circles). As myc concentration increases due to its deregulation, myc starts amplifying the gene signature specific for a given cell. **B.** At low concentrations, myc promotes the transcription of cell-specific genes. As its levels rise past normative, myc triggers the expression of its specific gene signature. Adapted from Kress et al., 2015.

1.3.2.2 Transcriptional repression by myc

Myc regulates a plethora of antagonistic processes; cell proliferation, apoptosis and differentiation, to name a few (Kalkat et al., 2018). For example, in the presence of mitogenic signals, myc can suppress apoptosis by activating the transcription of genes that downregulate this process. A more direct mechanism is association with myc interacting zinc finger 1 (Miz1) which has the ability to heterodimerise with Myc and represses the transcription of cell-cycle inhibitors (Staller et al., 2001; Peukert et al., 1997). For example, c-myc displaces p300, another partner of Miz1, and represses the transcription of $p15^{Ink4b}$ (Staller et al., 2001). In a similar mechanism, Myc-Miz1 can also abrogate the expression of $p21^{Cip1}$ (Wu et al., 2003). The exact nature of this repressive mechanism is not known (Herold et al., 2002).

1.3.3 Mechanism of myc involvement in transcription

1.3.3.1 Histone modification

The initial stages of transcription require significant chromatin modifications leading to the formation of open chromatin, which is characterised by less coiling and is partially devoid of nucleosomes. This important step allows the accessibility of transcriptional machinery to promoter regions. Histones are mainly disassembled by a change in their PTM signatures. In particular acetylation of lysines masks their positive charge and causes dissociation from the negatively charged backbone of DNA (Venkatesh and Workman, 2015). Both N- and c-myc expression have been correlated with hyperacetylation of histone 3 and histone 4, which is a hallmark of euchromatin (Knoepfler et al., 2006). c-myc recruits TRansformation/tRanscription domain-Associated Protein (TRRAP; McMahon et al., 1998; Saleh et al., 1998; Cowling and Cole, 2006), a large scaffolding protein for multiple histone acetylase (HATs) complexes, modifying the acetylation pattern on histones, leading to transcriptional activation (Saleh et al., 1998; Martinez et al., 2001). TRRAP participates in a few HATs, SPT/ADA/GCN5/Acetyltransferase (STAGA) HAT and TAT-interacting protein 60 kDa (TIP60), GCN5, SPT3, ADA and TAF5L, amongst others (Martinez et al., 2001; Bouchard et al., 2001; Frank et al., 2001). The abrogation of c-myc interacting with TRRAP causes c-myc to lose its ability to transform cells, albeit c-myc still retains the ability to regulate basal gene expression (McMahon et al., 1998). c-myc interacts with WD40-repeat containing protein 5 (WDR5) through its MBIIIa, which is also conserved in N-myc. WDR5 forms part of multiple chromatin regulatory protein complexes, such as H3K4 methyltransferases which are responsible for methylation of H3K4, an epigenetic modification associated with the transcriptional activation (Thomas et al., 2015; Lu et al., 2018). WDR5 is a core component of mixed-lineage leukaemia 1-5 (MLL1-5) H3K4 methyltransferase, thus it is possible that through the interaction of c-myc, WDR5 is recruited to the promoter regions to enhance the activity of MLL1-5 (Lu et al., 2018).

1.3.3.2 Transcription initiation

Transcription initiation is characterised by recognition of core promoters by the transcriptional machinery and is often triggered and aided by TFs, including myc. The pre-initiation complex (PIC) for RNA polymerase II (Pol II) is a mega-Dalton complex that consists of Pol II and general TFs A-H (TFIIA – H) that assemble within the promoter region (Sainsbury et al., 2015). One of the first events is the recognition of a TATA box sequence by TATA-binding protein (TBP), which is located ~30 bps upstream of the transcription start site (TSS). TBP, together with TATA-associated factors (TAFs) forms part of TFIID and is an essential component of transcription *in vitro* (Nogales et al., 2017; Sainsbury et al., 2015). TBP through its saddle-like structure binds DNA and bends it by 90°, necessary for DNA strand melting. c-myc binds to the TBP-TAF1 complex through a short helical segment spanning residues Q^{98} – V^{111}, in a way that prohibits DNA binding (Figure 1.6 B; Wei et al., 2019). TBP is known to bind DNA through hydrophobic interactions with high affinity and it is thought that c-myc aids in the loading of TBP onto DNA (Wei et al., 2019).

1.3.3.3 Elongation of transcription

The elongation stage of transcription is characterised by release of Pol II from a promotor proximal region of the gene into the gene body. Pol II becomes stalled ~50 bps downstream of the TSS in a regulatory process known as promoter proximal pausing (Gilmour and Lis, 1986). It is stalled by negative elongation factor (NELF) and 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB)-sensitivity inducing factor (DSIF). Positive transcription elongation factor B (pTEFb) is a complex of cyclin-dependent kinase 9 (Cdk9) and cyclin T. It phosphorylates NELF and DSIF, causing the dissociation of the former and change in the conformation of the latter to become elongation-permissive (Cramer, 2019) In addition, pTEFb phosphorylates S^2 of the heptad repeats of CTD, which further relieves the inhibitory mechanism and Pol II moves into the elongation stage of transcription (Cramer, 2019).

The role of c-myc in relieving promoter proximal pausing has been well documented using cell and systems biology approaches. However, the structural and molecular biology mechanisms underpinning this are less well-defined. c-myc interacts with cyclin T through MBI and potentially MBII and recruits pTEFb to Pol II (Kanazawa et al., 2003; Rahl et al., 2010; Eberhardy and Farnham, 2002) and a functional knock out of pTEFb through genetic manipulation or pharmacological intervention abrogates the expression of c-myc-regulated genes (Eberhardy and Farnham, 2002; Gargano et al., 2007; Kanazawa et al., 2003). Inhibition of c-myc transactivation activity by dissociation of c-myc from Max causes a marked reduction in Serine 2 phosphorylation, but not Serine 5, indicating that c-myc recruits pTEFb to Pol II and alleviates promoter proximal pause (Figure 1.6 B; Rahl et al., 2010). This suggests that c-myc is responsible for the recruitment of pTEFb. However, whether this is done via a direct interaction between pTEFb and myc or through an intermediary remains to be determined. c-myc also recruits Spt5, a subunit of DSIF. Although initially involved in promoter proximal pausing, DSIF has also emerged as an important factor that enhances Pol II processivity and prevents backtracking of Pol II on DNA (Figure 1.6 B; Shetty et al., 2017).

c-myc has also been shown to interact with Pol II-associated factor 1 (PAF1) complex. In humans, the PAF1 complex consists of Paf1, Cdc73, Leo1, Ctr9 and Rtf1 and its role in regulating transcription is broad and dependent on the physiological state of the cell, and ranges from activating of transcription to inhibition of transcription (Zhu et al., 2005; Jaenicke et al., 2016). PAF1 recruitment usually marks the transition from proximal pause to productive elongation (Jaenicke et al., 2016). c-myc interacts with Cdc73 through MBI in a phosphorylation-dependent manner and is thought to load PAF1 onto pTEFb, promoting phosphorylation of S^2 and transition into a productive elongation (Jaenicke et al., 2016). A similar role has been established for *Drosophila* d-myc and Leo1, however, in this case, the interaction is not restricted to conserved MBs but takes place over a broad central region within myc (Gerlach Jennifer et al., 2017). The role of this interaction is unknown; however it is speculated that it serves to recruit entire PAF1 complex to transcription sites and the knock-out of this interaction leads to a modest decrease in expression of myc-regulated genes (Gerlach Jennifer et al., 2017; Ehara et al., 2017; Baluapuri et al., 2019).

Myc also recruits topoisomerases 1 and 2 (TOP1 and TOP2, respectively), which are the enzymes that uncoil supercoiled DNA. This is especially important during transcription bursts, where heavy Pol II load causes unwinding of DNA strands and introduces coils downstream of the transcription bubble. Topoisomerases nick a strand of DNA and reduce the coiling, preventing DNA transcription stalling and potential DNA strand breaks (Figure 1.6 B; Das et al., 2022).

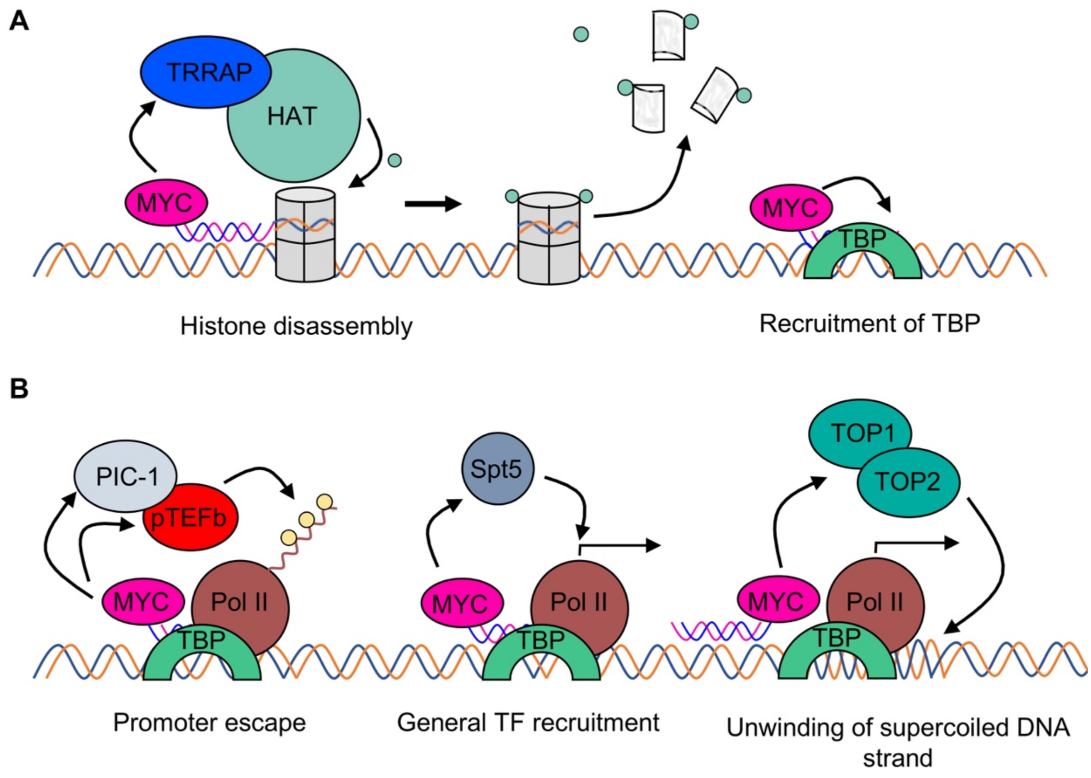


Figure 1.6. Myc involvement in transcription.
A. Myc participates in disassembly of histones, through the recruitment of TRRAP, which in turn recruits various HATs. These enzymes acetylate lysine residues of histones, resulting in dissociation from DNA. c-myc has also been shown to interact with the TBP-TAF1 protein complex, in a manner that prevents TBP from recognising TATA boxes. **B.** Myc is involved in multiple stages of transcription through the recruitment of various factors. Myc directly interacts with TBP protein which is an essential co-factor for Pol II recruitment, Myc also recruits pTEFb and PIC-1 during promoter proximal pause, which allows Pol II to move into productive elongation. c-myc also recruits Spt5 which forms a part of DSIF, however following pTEFb-mediated phosphorylation of CTD, is an essential factor that enhances Pol II processivity and prevents back-tracking. Lastly, c-myc and N-myc are able to recruit TOP1 and TOP2 to active transcription bubbles to prevent super-coiling downstream of the transcription start sites (Das et al., 2022; McMahon et al., 1998; Gerlach Jennifer et al., 2017; Ehara et al., 2017; Baluapuri et al., 2019; Rahl et al., 2010).

1.4 Myc and cancer

1.4.1 Myc deregulation mechanisms

In normal tissues myc is tightly regulated protein with very short half-life of ~20 min (Hann and Eisenman, 1984; Gregory et al., 2003b; Dani et al., 1984). Tumour cells can evolve multiple strategies to bypass the short half-life of myc and maximise its proliferative capabilities (Figure 1.7).

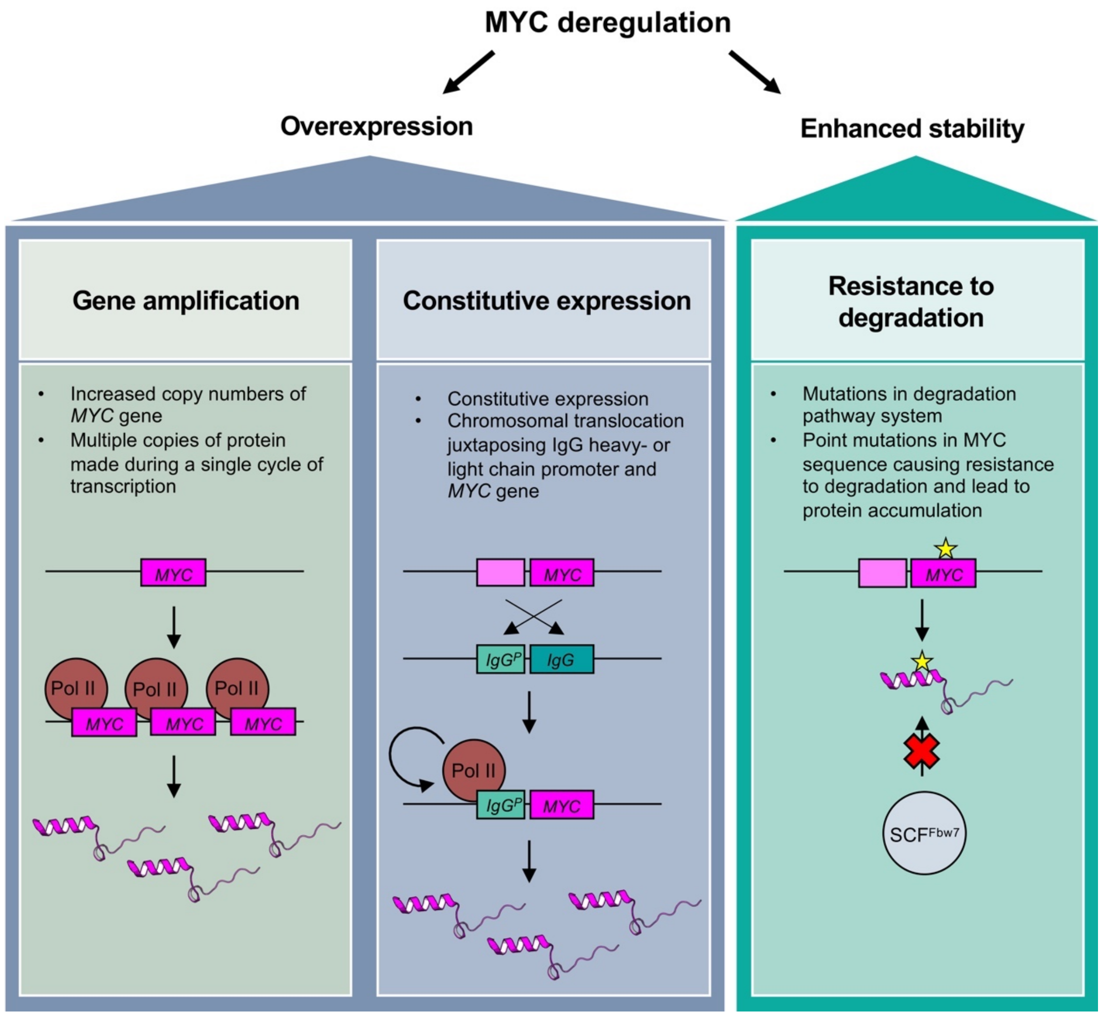


Figure 1.7. Mechanisms of myc deregulation.
Overexpression is the most common mechanism and can happen either through gene amplification or through translocation of the myc gene to active promotes, e.g., heavy immunoglobulin chain promoter. This usually happens in B-cell type malignancies. Myc can also be deregulated by enhanced protein stability. Mutations can be acquired in the proteasomal degradation pathway, but also within the myc sequence itself. A hotspot for mutations in haemotopoietic cancers is T^{58}, frequently mutated to alanine (Bahram et al., 2000; Pugh et al., 2013; Tate et al., 2019).

1.4.1.1 Copy number gains

The amplification of a gene results in multiple copies of a protein being made during a single transcription cycle. This is usually the result of genome doubling or tandem duplication (Kalkat et al., 2018; Burrell et al., 2013). Increased copy numbers of myc are the most common mechanism of myc deregulation.

c-myc gene amplification was first identified in B-cell lymphomas, where it is seen in 8-20% of patients (Figure 1.7; Xia and Zhang, 2020). An in-depth analysis of cancer type *vs.* myc amplifications is beyond the scope of this work, however, an increase in myc copy numbers has been seen in breast, ovarian, prostate and lung cancers, medulloblastoma and NB (Kalkat et al., 2018). In breast cancers, c-myc amplification and/or protein overexpression is detected in ~70% of untreated high-grade breast cancer samples (Blancato et al., 2004). This speaks to the pervasiveness of this mechanism in one of the most common cancers and suggests an urgency to research pharmacological interventions. More so, myc amplification is a marker of poor prognosis, regardless of cancer type, however, myc amplification does not always correlate with an increased level of the protein (Kalkat et al., 2018; Chrzan et al., 2001). Gene amplification of N-myc, as typically seen in NB, can vary between 3 and 300 times, often leading to abnormally high levels of the protein within cells, and is correlated with low survival rate (Seeger et al., 1985; Brodeur et al., 1984).

1.4.1.2 Constitutive expression

Constitutive expression of proteins results in a constant supply of new protein that is not degraded rapidly enough, assuming that physiological degradation mechanisms are intact. In haematopoietic cancers, the most common mechanism driving this phenotypic effect is chromosomal translocation. This results in the transposition of the c-myc gene to the proximity of, for example, heavy or light chain immunoglobulin promoter loci, which results in a constitutive expression (Küppers, 2005; Küppers and Dalla-Favera, 2001). One of the most studied examples of this mechanism is in Burkitt's lymphoma patients where the chromosomal translocation occurs between loci of the c-myc coding region (chromosome 8) and a locus of an enhancer for immunoglobulin heavy chain (Tzankov et al., 2014). The resulting abnormal karyotype (t(8;14)(q24;q23)) is characterised by abnormally high transcriptional output, normally reserved for this highly expressed antibody chain (Tzankov et al., 2014).

1.4.1.3 Resistance to degradation

Temporal regulation of PTMs in the myc phosphodegron region causes its rapid turnover to ensure that myc activation is timed, adequate and in response to proliferative signalling. Resistance to degradations can occur indirectly through dysregulation of proteasomal machinery, or more rarely, through mutations within the phosphodegron sequence (Figure 1.7). In comparison to other mechanisms of myc deregulation, this is seen relatively rarely, perhaps speaking to its reduced effectiveness in sustaining myc signalling. An analysis of cell lines derived from Burkitt's lymphoma, AIDS-related lymphoma and B-cell acute lymphoma has indicated that T^{58A} substitution is the most common mutation in both c- and N-myc

(Bahram et al., 2000; Pugh et al., 2013). T^{58} phosphorylation is a critical step in myc degradation and is critical for the myc turnover (Welcker et al., 2004; Sears et al., 2000).

Another point mutation seen in roughly ~1.7% of N-myc amplified NB cases is the N-myc^{P44L} substitution. Proline 44 is positioned after T^{43}, raising a possibility of abrogating activity of proline-directed kinase against T^{43}, however, no kinase that phosphorylates this site has been identified. There have been reports that N-myc^{P44L} have higher expression levels than WT N-myc, however, this has not been confirmed in any other studies and the functional significance of this site remains a mystery (Pugh et al., 2013).

1.4.2 Myc as a cancer driver

Central nervous system-associated tumours are one of the most prevalent paediatric cancers after leukaemias, and overall are characterised by relatively good survival rates (Siegel et al., 2012; Northcott et al., 2012). The price of these good outcomes comes in the form of aggressive treatment which includes risky surgery, intense chemotherapy and prolonged radiation, which often leaves the paediatric patient with severe treatment-induced disabilities, including retarded physiological and psycho-social development, amongst others (Siegel et al., 2012; Northcott et al., 2012).

N-myc was first discovered in NB – a neuroectodermal-origin tumour, most commonly seen in children (Schwab et al., 1983; Kohl et al., 1983). NB broad pathophysiology ranges from spontaneous regression in pre-18 months old children to a severely malignant phenotype with poor prognosis. N-myc overexpression is estimated to occur in ~20% of all NB cases and is associated with poor cellular differentiation and poor survival outcomes, thus N-myc is the most reliable prognostic marker of this disease (Petroni et al., 2011; Irwin and Park, 2015).

Other types of central nervous system (CNS) tumours with N-myc involvement are medulloblastoma and glioblastoma. In medulloblastoma, N-myc is overexpressed in the most common subtype of this cancer and similarly to NB, its amplification is linked to poor survival outcomes (Northcott et al., 2012). Glioblastoma is another N-myc-driven cancer, however N-myc deregulation in this tumour occurs indirectly, mainly through mutation in histone H3 (glycine is substituted to valine on position 34), which loosens the stringent control over N-myc expression, leading to its overexpression

(Bjerke et al., 2013). N-myc deregulation has also been seen in other CNS tumours, such as astrocytoma and retinoblastoma (Garson et al., 1985; Lee et al., 1984).

N-myc can be upregulated in other types of cancer, such as rhabdomyosarcoma, breast cancer, prostate cancer and small cell lung cancer (SCLC; Jung et al., 2005; Rickman et al., 2018). In one study of rhabdomyosarcoma, N-myc gene amplification was seen in all 15 studied cases of this rare childhood cancer (Williamson et al., 2005). In SCLC, N-myc amplifications were the most significant deregulations, following L-myc amplifications (Peifer et al., 2015). In triple-negative breast cancer, N-myc amplification is seen in ~10% of cases and carries the most dismal prognosis for patients, with limited targeted therapy options, high recurrence rates, and resistance to chemotherapy (Schafer et al., 2020). Even more peculiar is the case of neuroendocrine prostate cancer (NEPC), arising as a secondary tumour due to de-differentiation of prostate adenocarcinoma in response to androgen-deprivation therapy (Beltran et al., 2011). N-myc is overexpressed in ~40% of NEPC cases, owing to gene amplification, and it contributes to the development of neuroendocrine features in these cells, leading to poor clinical outcomes (Beltran et al., 2014). Although limited, these studies show a correlation between N-myc deregulation and poor clinical outcomes and highlight the necessity for the development of pharmacological interventions.

1.5 Intrinsically disordered proteins

1.5.1 Structure-function paradigm

The second half of the 20th century brought on the development of structural biology as a powerful tool in biological research. The use of X-ray crystallography, and more recently cryo-EM, as methods for macromolecular structure determination, solidified the idea that protein function is dependent on its 3D structure, only marginally changing upon binding of a partner (Uversky et al., 2000; Schlessinger et al., 2011; Wright and Dyson, 2015; Babu, 2016). However, as both X-ray crystallography and cryo-EM select for proteins with stable structures and little heterogeneity, the number and the importance of intrinsically disordered proteins (IDPs), which lack a stable tertiary structure and thus are not amendable for classical structural studies, have been underestimated.

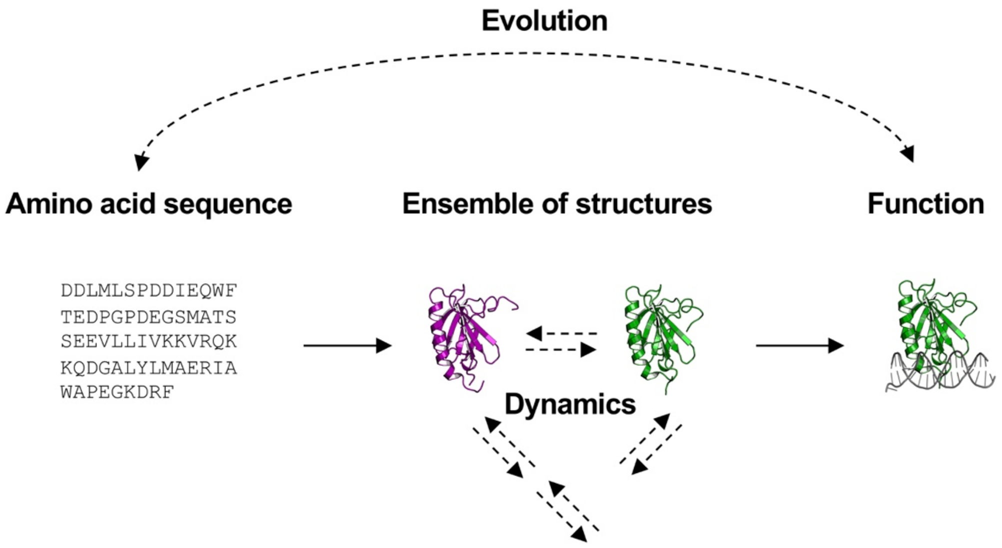


Figure 1.8. Revised structure-function paradigm.
Amino acid sequence determines the presence of secondary and tertiary structures, or lack of thereof, in a protein. Even well-folded globular proteins have multiple conformational states with dynamics determining interconversion between them. A specific function is determined by a single conformation from the ensemble of possible conformations. The function also feeds back into primary amino acid sequence through the process of evolution. Structures taken from (accession code: 2RUK; Okuda and Nishimura, 2014).

Advances in the analysis of the human genome which started in the mid-90's have revealed a large proportion of proteins that defy this paradigm as they do not possess temporally stable 3D structure, but instead are flexible and assume a diverse ensemble of conformations (Chebaro et al., 2015). This has given a rise to the idea that proteins, instead of fitting into discreet conformational categories, fall on a continuum, displaying degrees of plasticity modulated by various factors (Figure 1.8; Chebaro et al., 2015). The amino acid sequence determines both an overall fold and dynamics and the conversion between different states results in protein function (Figure 1.8). A protein can be globular with minimal conformational plasticity, can be a molten globule lacking a 3D shape but retaining secondary structures connected by intrinsically disordered regions (IDRs). Alternatively, a protein can be an IDP, with elements of stable or transient secondary structures (Figure 1.8; Dyson and Wright, 2005). This additional layer of complexity arises from observations that some proteins can fold upon binding or possess domains with various degrees of order/disorder in them.

1.5.2 Features of IDPs

The structural and functional flexibility of IDPs is context-dependent, which makes them particularly beneficial in processes of cell signalling, translation, transcription and other cell cycle processes, where the rapid assembly of diverse protein complexes is essential in response to intra- and extracellular signalling (Wright and Dyson, 2015; Iakoucheva et al., 2002; Galea et al., 2008). Not surprisingly, IDPs are abundant in classes of proteins such as TFs (myc, p53), nucleic acid binding proteins, chromatin binding proteins, amongst many others which integrate multiple and nuanced signalling to generate physiological outcomes (Deiana et al., 2019; Wright and Dyson, 2015).

IDPs have high degrees of local and global motions and can sample a range of transient conformations. This conformational plasticity allows them to interact with multiple binding partners within short time scales (interaction promiscuity) and for a finely tuned regulation of cellular processes (Dyson and Wright, 2005). These interactions tend to be high specificity, with high dissociation constants and relatively low affinity and with an ability to fold into secondary structures upon binding, through small recognition elements (Tompa et al., 2015; Wright and Dyson, 2015; Uversky, 2015; Uversky, 2013; Iakoucheva et al., 2002). Although IDPs lack the ability to catalyse biological reactions, which usually require the presence of a stable hydrophobic active site, the importance of IDPs in health and disease has become progressively clear in recent decades (Wright and Dyson, 2015; Uversky, 2015; Uversky, 2013; Iakoucheva et al., 2002). Unable to assume self-inhibitory conformations, IDPs are often regulated either through PTMs which increase the possible number of binding partners even more (Dyson and Wright, 2005). This also provides a fast and reversible way for potential interactions, shifting equilibrium between different conformations, as a signal for stabilisation, localisation or degradation, amongst others (Gibbs et al., 2017).

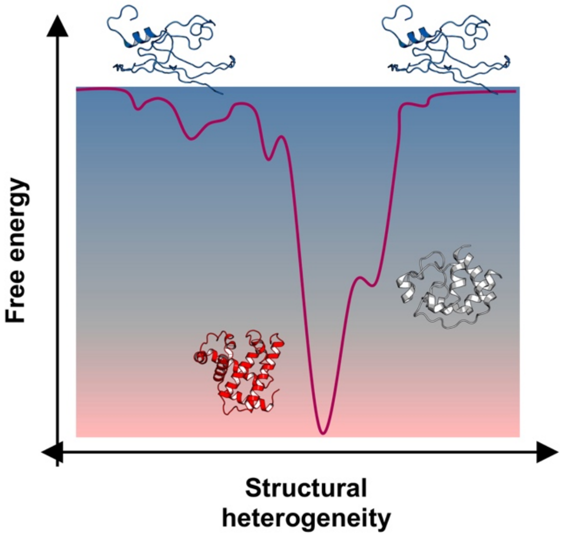


Figure 1.9. Energy landscape of conformational states.
Folded proteins only have one or a small number of energetically favourable states, which drives their folding. This is mainly driven by the presence of a hydrophobic core, which needs to be shielded from the aqueous environment to minimise free energy. IDPs are rich in polar and charged residues and can assume many conformational states, each with low energy. Molten globules are intermediate states between folded and unfolded states characterised by the presence of secondary structures but lacking tight 3D organisation of tertiary structures (Dyson and Wright, 2005). Adapted from (Chong and Ham, 2013).

To achieve their conformational plasticity, IDPs have a lower frequency of bulky hydrophobic amino acids than globular proteins (which drive protein folding through entropically-favourable burying of hydrophobic side chains) and higher net charge driven by a large number of polar and charged residues (Wright and Dyson, 2015; Tompa et al., 2015; Gibbs et al., 2017; Lins and Brasseur, 1995; Uversky, 2015). This allows IDPs and intrinsically disordered regions (IDRs) free movement in the aqueous cellular environment. High proline content is also a common occurrence in IDPs, as the cyclic side chain and a lack of protonated backbone amide, make secondary structures unfavourable.

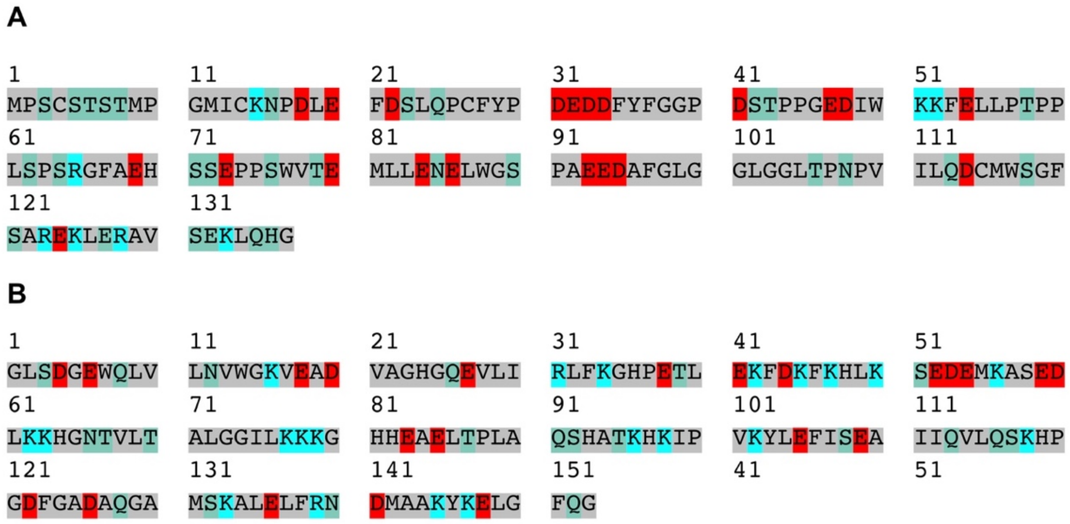
The proportion of IDPs within genomes is positively correlated with the complexity of an organism, and in eukaryotes ~30% of all proteins are either IDPs or possess a significant level of disorder (Dyson and Wright, 2005; Dunker et al., 2001). IDPs can cater to many binding partners, owing to their conformational plasticity and ability to form ‘fuzzy’ complexes, which can be loosely defined as a disorder on binding (Tompa and Fuxreiter, 2008). IDPs can take many interconverting conformations on the surface of a binding partner and form transient secondary structures upon binding and in general are classified as ‘dynamic’, i.e., undergoing rapid changes in a short time-frame on a proteins surface (Gibbs et al., 2017; Wright and Dyson, 2015; Jensen et al., 2013). From an energy landscape point of view, folded proteins usually only have one global energetically favourable state which helps drive the formation of an ordered state. The energy landscape for IDPs is remarkably flat, therefore they can assume many low-energy conformations, most being equally thermodynamically favourable, even in the context of protein binding (Figure 1.9; Papoian, 2008).

1.5.3 N-myc as an IDP

An analysis of the primary amino acid sequence of N-myc$_{1\text{-}137}$ (hereafter referred to as the N-myc TAD) reveals a typical IDP amino acid composition (Figure 1.10). The sequence is abundant in order-disfavouring amino acids. N-myc TAD possesses 17 prolines within its 137 amino acids and this introduces a certain level of rigidity (Campen et al., 2008). The chemical structure of proline disfavours α-helices for several reasons. Firstly, a lack of backbone amide proton prevents proline from participating in H-bond network, necessary to maintain α-helices (Li et al., 1996b). Covalent incorporation of a pyrrolidine cyclic ring side chain into its backbone causes the ϕ angle of the proline to be fixed at a value of -60° resulting in the lack of the required flexibility to maintain α-helix (Li et al., 1996b; Li et al., 1996a). Glycine is the smallest of amino acids and when incorporated into a polypeptide chain is flexible to the point that it is unable to hold the rigidity of secondary structures (Imai and Mitaku, 2005). In addition, it lacks a side chain and instead features two Hα protons. Although classified as hydrophobic, due to the lack of side chain, the hydrophobicity of glycine is more akin to those of polar residues, such as asparagine and glutamine (Zhu et al., 2016).

Further analysis of the amino acid sequence of N-myc TAD reveals the relatively equal number of hydrophobic residues and residues that are either polar or charged (Figure 1.10 A), suggesting that there are not enough bulky hydrophobic residues that could drive the formation of the hydrophobic core. Comparison with amino acid sequence of a protein of similar size which is folded, such as pig myoglobin, reveals a stark contrast in amino acid composition (Figure 1.10 B). Myoglobin is a globular protein composed of ten α-helices (Krzywda et al., 1998). It has a high alanine content which is an α-helix-promoting residue (9.2% of overall sequence vs. 3.6% in N-myc TAD). Myoglobin sequence possesses 4 prolines and 15 glycine residues. The content of polar residues is much lower in myoglobin in comparison to N-myc TAD, and there is a drastic difference between the isoelectric points of the proteins (4.16 *vs.* 6.83, respectively). Myoglobin has more +ve charged residues which serve to form

H-bond lattice within α-helices (Krzywda et al., 1998). In N-myc TAD, the overall charge is not balanced due to high content of acidic aspartate and glutamate with low content of basic lysines and arginines and allow N-myc TAD to interact with aqueous environment (Figure 1.10 A, green).


A

1 MPSCSTSTMP 11 GMICKNPDLE 21 FDSLQPCFYP 31 DEDDFYFGGP 41 DSTPPGEDIW 51 KKFELLPTPP
61 LSPSRGFAEH 71 SSEPPSWVTE 81 MLLENELWGS 91 PAEEDAFGLG 101 GLGGLTPNPV 111 ILQDCMWSGF
121 SAREKLERAV 131 SEKLQHG

B

1 GLSDGEWQLV 11 LNVWGKVEAD 21 VAGHGQEVLI 31 RLFKGHPETL 41 EKFDKFKHLK 51 SEDEMKASED
61 LKKHGNTVLT 71 ALGGILKKKG 81 HHEAELTPLA 91 QSHATKHKIP 101 VKYLEFISEA 111 IIQVLQSKHP
121 GDFGADAQGA 131 MSKALELFRN 141 DMAAKYKELG 151 FQG 41 51

Figure 1.10. Comparison of amino acid sequence of N-myc TAD and myoglobin. **A.** Sequence of N-myc TAD. **B.** Sequence of pig myoglobin. Amino acids that are hydrophobic are highlighted in grey, including glycines and prolines. Residues that are +ve charged (arginine, lysine) are highlighted in blue, residues that are -ve charged (glutamic acid and aspartic acid) are highlighted in red and polar residues are highlighted in green.

1.5.4 Structural methods to study IDPs

The lack of structure makes IDPs challenging candidates for structural studies, as they do not produce crystals for analysis by X-ray crystallography, due to high dynamics of intra- and intermolecular movements (Uversky, 2002; Dunker et al., 2001). Crystal structures that do feature IDPs with a binding partner, only present one of many possible conformations of that IDP. Cryo-EM is also a challenging technique for use with IDPs since it requires a high degree of homogeneity of the sample to allow for averaging different particle views into a 3D map of the protein (Thompson et al., 2016). The most appropriate and informative technique to study IDPs is NMR spectroscopy. NMR, in contrast to cryo-EM and X-ray crystallography is a ‘bottom-up’ approach where, instead of obtaining a model of the entire structure, information is collected for each residue separately in form of chemical shifts and chemical shift perturbations, relaxation rates and spatial restraints and from these the ensemble of tertiary structures can be calculated.

1.6 NMR as a tool to study IDPs

1.6.1 Basic principles of NMR

NMR, as a spectroscopic method, is the study of transitions between energy states of spins. Spin is an intrinsic property of every nucleus and can be defined as the intrinsic angular momentum of a nucleus, i.e., rotation of particle around its own axis (Levitt, 2013). The nuclei motion induced by spin causes nucleus to also possess a magnetic moment and interact with a magnetic field. When placed in strong uniform magnetic field, nuclei behave like mini bar magnets, aligning their magnetic moment along (+*Z*) or against (-*Z*) the direction of B^0 (Figure 1.11; Levitt, 2013). This causes small and detectable differences in energy states between these two orientations for nuclei that have spin of ½, such as ^{1}H, ^{15}N and ^{13}C (Figure 1.12; Levitt, 2013). Through the influence of spin, these magnetic moments do not remain stationary, instead rotate around the *Z* axis in a motion called Larmor precession, a variable characteristic to nucleus type. The Larmor frequency (ω^0) is expressed as

$$\omega^0 = -\gamma B^0,$$

where $-\gamma$ is the gyromagnetic ratio, specific to each nucleus type, and B^0 is a sum of magnetic fields generated by both the magnet and local environment (Levitt, 2013).

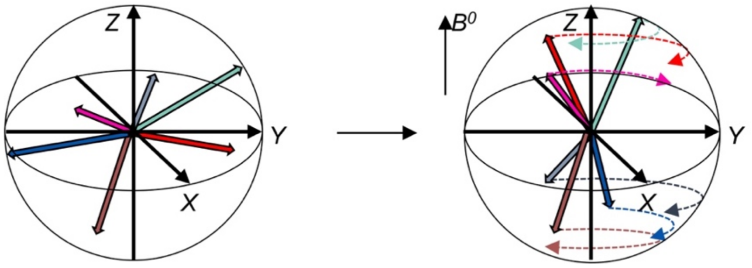


Figure 1.11. Schematic representation of spins behaviour in the absence and presence of an external magnetic field B^0.
In the absence of magnetic field (left), the distribution is isotropic, i.e., all dimensions and directions are equally possible. The application of constant magnetic field B^0, causes spins to align either along *Z* axis or against it. As the lower energy state is in the +*Z* direction, slightly more spins occupy this direction than in the -*Z* direction. Due to their magnetic moments, spins precess around the *Z* axis with a frequency called the Larmor frequency. At this state, spins are in a thermal equilibrium (Levitt, 2013).

This state of thermal equilibrium can be perturbed by the application of radio-frequency (r.f.) pulses that are in resonance with the Larmor frequencies. This causes spins to align and precess within the transverse or *XY* plane perpendicular to the *Z* axis (Figure 1.12; transverse plane).

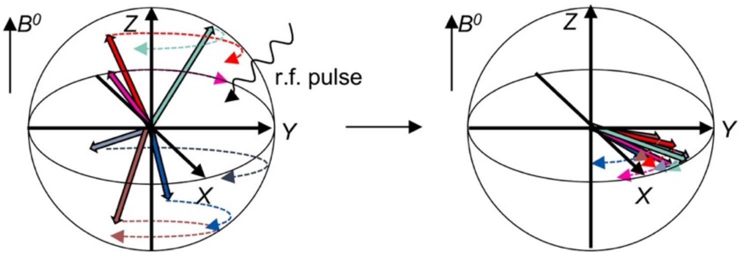


Figure 1.12. Schematic representation of spins perturbed by an r.f. pulse
R.f. pulse is applied in resonance with precessing magnetisation vectors (left panel, curly arrow). This disrupts thermal equilibrium and forces the magnetisation vector onto the *XY* plane (right panel). Initially spins process in the *XY* dimension in coherence with each other (Levitt, 2013).

NMR experiments use r.f. pulses, which vary in length and strength to perturb the equilibrium state of different nuclei and to manipulate couplings between nuclei (the magnetisation of one nucleus causes changes in magnetisation of nearby nuclei). Coupling allows for the observed nucleus to be impacted by neighbouring nuclei and thus provides information about the environment of that nucleus (Levitt, 2013).

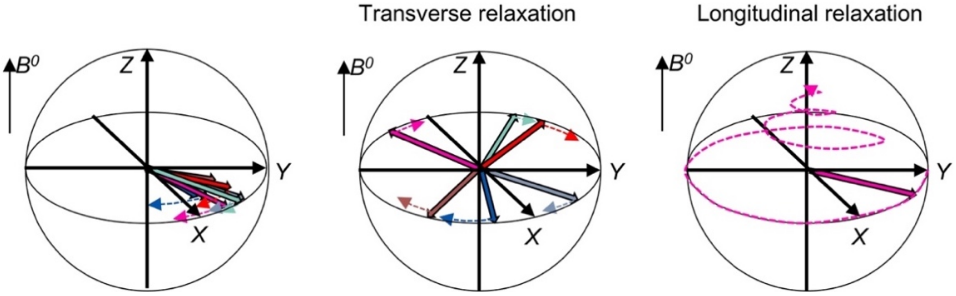


Figure 1.13. Relaxation.
Following the perturbation, net magnetisation preceeses in the *XY* plane in a coherent way (left panel; rotations of individual magnetic vectors are synchronised). After some time, the magnetisation vectors lose their coherence, and the bulk magnetisation vector decreases in the *XY* direction (middle panel). This process of losing coherence and magnetisation in the *XY* plane is called transverse relaxation. At the same time, magnetisation vectors return to their thermal equilibrium along the *Z* axis in a process called longitudinal relaxation (right panel). For clarity, only a single magnetisation vector is shown to show the trajectory of its return to the equilibrium state (Levitt, 2013).

If no further r.f pulses are applied, bulk magnetisation eventually returns to the state of thermal equilibrium (alignment along the *Z* axis) in the process called longitudinal relaxation (Figure 1.13, right panel). Relaxation is also driven by the loss of coherence and loss of bulk magnetisation in the *XY* plane, a process called transverse relaxation

(Figure 1.13; middle panel). Transverse relaxation, described using the time T_2, is, by necessity, always the same as or shorter than longitudinal relaxation (T_1) and in proteins in ranges between 10s and 100s of ms. Longitudinal relaxation typically occurs within 100s of ms to 1 s, but it is determined by factors such as protein size and rotational correlation constant (τ_c,), defined as the time needed for a molecule to rotate by one radian.

NMR experiments are carried out by applying a series of r.f. pulses, tuned to various NMR-active nuclei, which manipulate the positioning of magnetisation vectors and coherences between magnetisation vectors arising from the same type of nuclei (homonuclear) or different types of nuclei (heteronuclear). Pulses are repeated many times, with free induction decays (FID) collected following each pulse sequence. T_1 provides a measure of the length of time required for the return of magnetisation vectors to the equilibrium and thus it determines how soon the next set of pulses can be applied. The FID signal (Figure 1.14) detected along the *X* or *Y* axis, is a function that encodes the measured frequencies and has an intensity envelope that decays over time with a rate R_2 ($R_2 = \frac{1}{T_2}$). The FID is Fourier transformed into the frequency domain and a spectrum is generated with peaks at the measured frequencies (Figure 1.14). The spectrum can be from one- or multi-dimensional, with the same or different nucleus types serving as resonant dimensions or various nuclei used as dimensions. Peak intensity (its height) in the spectrum is inversely related to the width of the (Lorentzian) peak, which can be described using:

$$\Delta v = \frac{R_2}{\pi},$$

where Δv is the linewidth at half the maximum intensity of the peak and R_2 is the transverse relaxation rate for the nucleus. Thus, the intensity of the peak is directly correlated with T_2, which is a critical parameter in protein NMR. The faster the transverse relaxation, the longer the R_2 and the broader the peak, which makes interpretation of spectra challenging, especially if there is peak overlap from nuclei resonating at similar frequencies. There has been a lot of focus on developing techniques that optimise T_2, which will be further discussed in Chapter 3.

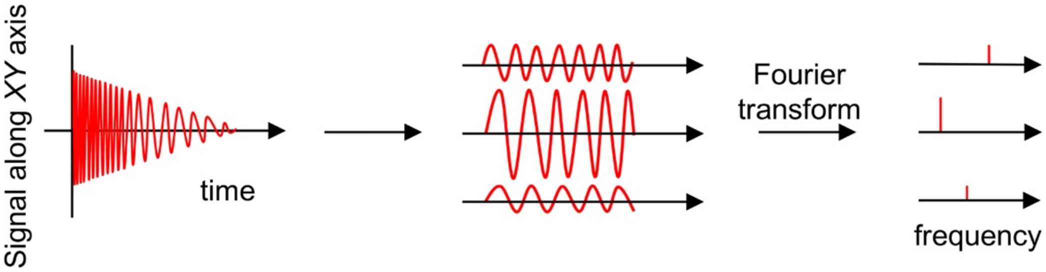


Figure 1.14. The processing of an FID in NMR experiments.
The detected signal in NMR is the free-induction decay (FID) which is a dissipating, oscillating signal in the *XY* plane (undergoing transverse relaxation). The individual components of FID are disentangled and converted into the frequency domain via Fourier transformation

1.6.2 NMR observables

The peaks seen on spectra are the resonant frequencies of observed nuclei. These resonant frequencies, which are measured as a chemical shift (δ), are the result of the local and global magnetic 'shielding' experienced by a nucleus. For instance, the electrons surrounding a nucleus shield it from the magnetic field. δ is an arbitrary unit which, for the purposes of eliminating differences between spectrometers of different magnetic strengths, is calibrated with respect to resonances within a 'standard' inert compound such as tetramethylsilane (TMS) or sodium trimethylsilylpropanesulfonate (DSS). In contrast to other spectroscopic methods, NMR utilises low energies to manipulate spin energy levels. The differences in energy levels between spins mostly average out and only 1 in million spins contribute to the signal observed, thus the units used are parts-per-million (ppm; Levitt, 2013).

1.6.3 ^{1}H-^{15}N HSQC

^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) experiment is one of the most basic experiments used in protein NMR and it serves as a foundation for most of protein NMR experiments. In a ^{1}H-^{15}N HSQC experiment, magnetisation is transferred from an amide proton onto the nitrogen and back onto the proton for detection, which results in a 2D heteronuclear spectrum correlating ^{1}H and ^{15}N resonances (Figure 1.15). The ease of interpretation stems from the fact that most amino acids have only one proton-nitrogen pair, thus producing a single peak on the spectrum. A few exceptions to this rule are the side chain amines of asparagine and glutamine which are usually visible as a pair of peaks with the same ^{15}N frequency and different ^{1}H frequencies. Tryptophan has an NH group positioned within its indole side chain ($H\varepsilon_1$-$N\varepsilon_1$) that also resonates at downfield ^{15}N positions (i.e., higher chemical shift in both ^{1}H and ^{15}N). Arginine and lysine also possess an extra nitrogen – proton pairs in their side chain (e.g. Arg Hε-Nε) that can be observed upfield in ^{15}N depending on conditions (Higman, 2012).

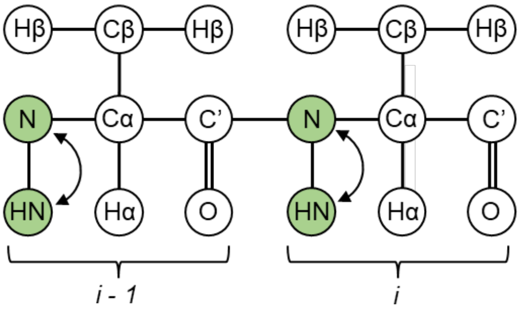


Figure 1.15. Schematic representation of the ^{1}H-^{15}N HSQC experiment. Magnetisation is generated initially on amide proton and then transferred onto the directly bonded nitrogen. The chemical shift is evolved on the nitrogen and the magnetisation is transferred again onto amide proton for detection. Generated spectrum has ^{1}H and ^{15}N as dimensions (Higman, 2012).

1.6.4 Chemical exchange

Chemical exchange is an interconversion between two or more states that causes a change in the chemical environment of nuclei studied (Vallurupalli, 2009). The interconversion can take the form of changes between conformations, between bound and free protein or between monomeric and dimeric/oligomeric protein states, amongst many others. NMR can capture the resonant frequency of two states (δ_A and $\delta_{B,}$) and all transitional states in-between. When interconversion follows slow chemical exchange, the rate of exchange (k_{ex}) is slower than the difference in chemical shift caused by the changed chemical environment ($\Delta\delta$) and occurs on longer time-scale than data acquisition in NMR experiment ($k_{ex} << \Delta\delta$; bottom panel, Figure 1.16). This leads to two observable peaks with the intensities representing the population of nuclei in each of the states (Higman, 2012). As the rate of chemical exchange shifts towards an intermediate regime ($k_{ex} \approx \Delta\delta$), observable nuclei will undergo chemical exchange multiple times during NMR experiment, and thus whilst the data is being acquired some nuclei populations will have δ_A or $\delta_{B.}$ but also chemical shifts of the transitional states between A and B. This results in the broadening of the Lorentzian peak, as multiple chemical shifts contribute to the overall signal acquired (middle panel; Figure 1.16). This effect is even more pronounced in a fast chemical exchange regime ($k_{ex} >> \Delta\delta$), where nuclei exchange multiple times between the two states whilst data is being acquired. In this case, only a single peak will be observed, which is a population-weighted average of the δ_A or δ_B chemical shifts (top panel; Figure 1.17; Higman, 2012).

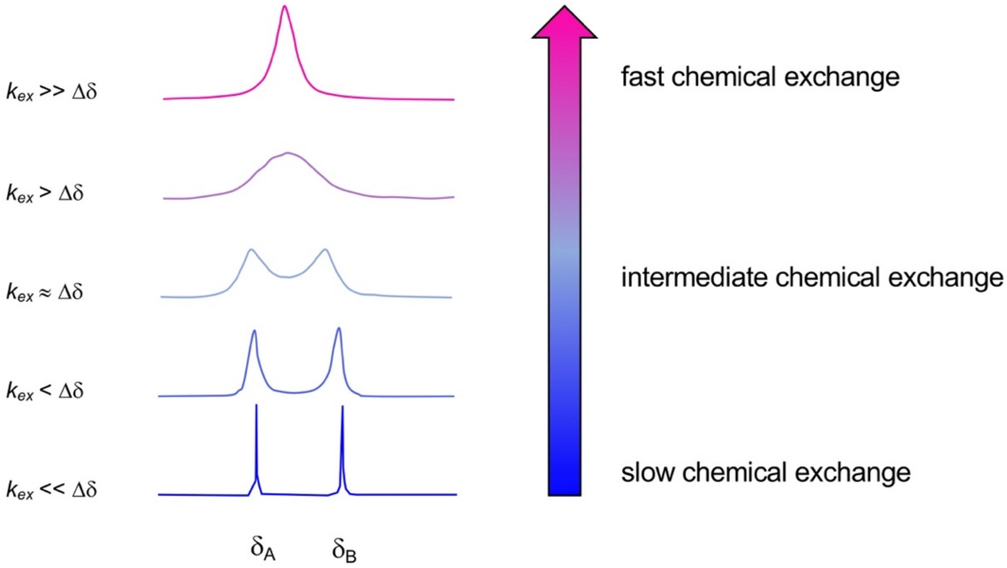


Figure 1.16. Chemical exchange and NMR.
NMR spectroscopy measures the frequency of resonating nuclei. If there is a change in the chemical environment of nuclei it will cause a change in their resonating frequencies. In slow chemical exchange ($k_{ex} << \Delta\delta$; bottom), the exchange occurs at a slower rate than the difference in frequencies ($\Delta\delta$) between chemical shifts of two states (δ_A or δ_B). This results in the observation of two separate peaks, each representing the population of nuclei resonating at δ_A or δ_B. As the chemical exchange changes towards intermediate ($k_{ex} \approx \Delta\delta$; middle), nuclei interchange between δ_A or δ_B at the same time as the NMR experiment is being acquired. The chemical shift of these two states is neither δ_A nor δ_B, albeit it is a population-weighted δ of all states between A and B. Once $k_{ex} >> \Delta\delta$, the interchange between A and B states occurs at a much faster scale and multiple times within the duration of the experiment. This results in a single peak which is population-weighted average of δ_A or δ_B (top panel). Adapted from (Elster, 2021).

1.6.5 Challenges in studying IDPs

Solution NMR is a way of studying IDPs in an aqueous environment, without the need to obtain an array of homogenous conformations. Structure generation requires a collection of various types of NMR experiments, each providing a specific set of information. There are several drawbacks using NMR. The size of the protein impacts the rotational correlation constant τ_c, which in turn determines R_2 relaxation and thus the sensitivity of the experiment. In large proteins τ_c is high due to their slow tumbling (Williamson, 2009). This shortens their R_2 as, as their slow global motions prolong their interaction with other spins, allowing them to cross-relax with other spins nearby more effectively (Williamson, 2009). This results in low peak intensities and large Δv, evidenced by broad linewidths in spectra. Inversely, smaller molecules and molecules with fast global dynamics, such as IDPs, experience slow relaxation rates which in return translates to sharper linewidths in spectra. An indirect way to enhance the strength of the signal is to increase the number of scans in each experiment to improve the signal-to-noise ratio, however, this is restricted by T_1, which increases as a protein size increases.

Larger proteins also contain more residues and thus the number of peaks present in the spectrum increases, potentially leading to peak overlap. If a protein is folded, then peaks in a ^{1}H-^{15}N HSQC spectrum are typically well dispersed, since residues experience distinct chemical environments. In IDPs, each residue samples many local chemical environments on rapid timescales leading to homogenisation of their chemical environment due to the motional averaging (Zhang et al., 2003; Bermel et al., 2012; Kjaergaard and Poulsen, 2011). This is realised in poor dispersion, particularly in the ^{1}H dimension, causing spectral overlap. This may be amplified by low sequence complexity in an IDP or the presence of common motifs within the sequence, which causes similar local chemical environments for some of the residues, complicating the assignment process (Figure 1.17).

The intrinsic flexibility of IDPs also causes their susceptibility to proteasomal degradation, due to the lack of structural shielding of protease-sensitive sites. This results in peak shifting over the course of data collection, usually combined with a decrease in peak intensity or disappearances of the peaks (Figure 1.17). The flexibility of the IDP also contributes to more efficient proton exchange with the solvent, as almost all amides are exposed to the solvent, leading to magnetisation dissipation and peak broadening (Williamson, 2009). Solvent exchange is increased at higher temperatures. In addition, some IDPs may contain regions with a propensity to form transient secondary structures, which are subjected to chemical exchange between structured and unstructured conformers. If the timescale of these exchanges is in an intermediate exchange regime, this leads to broadening of the peak and low intensity (Li and Palmer, 2010).

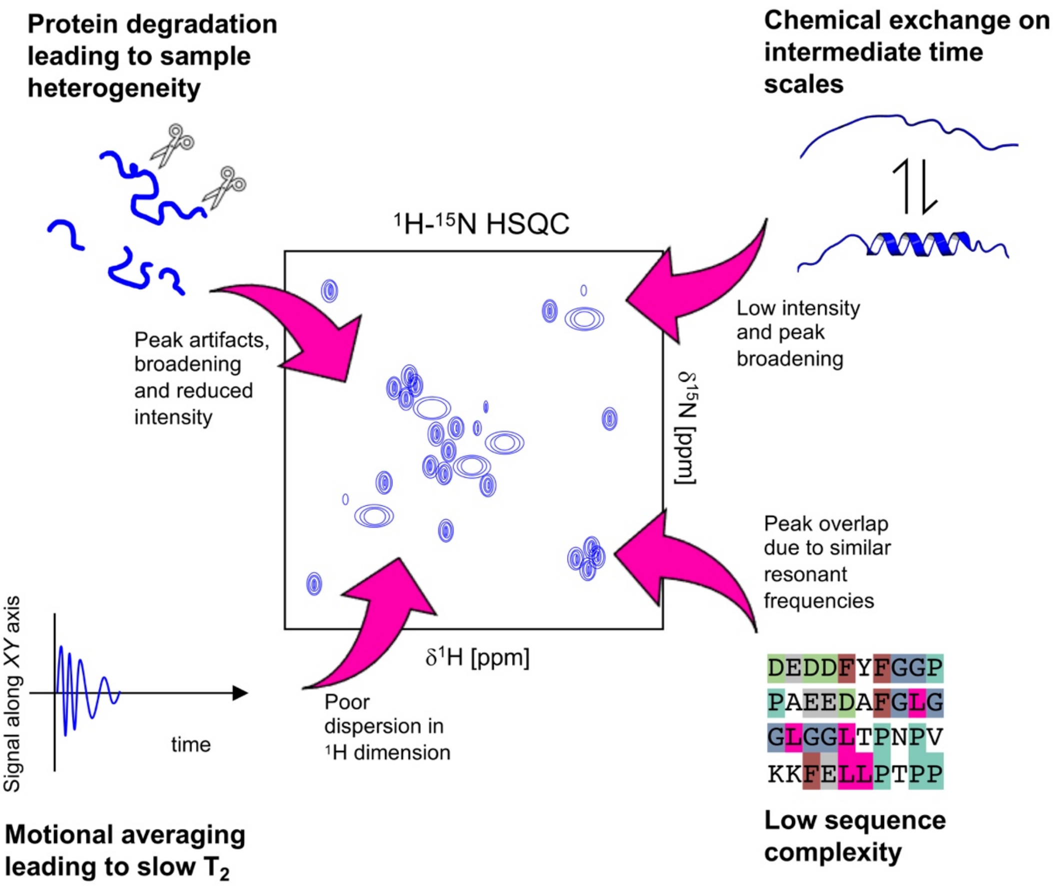


Figure 1.17. NMR challenges when interrogating an IDP.
The middle represents a typical IDP ^{1}H-^{15}N HSQC spectrum with blue peaks in the middle. Poor ^{1}H dispersion is caused by motional averaging, meaning that residues experience a homogenous average of all sampled environments with similar chemical shifts. IDPs also have low sequence complexity with repeating motifs that further the problem of low dispersion. NMR experiments can last from hours to days, usually at ambient temperatures, during which IDPs can undergo proteolytic cleavage, thus changing resonant frequencies of residues as scans are being acquired. IDPs are highly dynamic but can contain regions that exchange between different conformations and transient secondary structures. This can lead to peak broadening and low signal intensity, especially if it occurs on an intermediate time scale.

1.6.6 Advantages of IDPs in NMR

Despite many properties that make IDPs challenging candidates for NMR-based interrogation, they also possess properties that can be exploited to enhance observable spectra.

IDPs lack tertiary structures which makes it possible to optimise the length of the construct without compromising a structure, as it is absent in IDPs. This can improve the interpretability of the spectra in a few ways. Firstly, through the reduction of T_2, which enhances the sharpness of observable peaks. Secondly, truncation results in the reduction of the number of peaks which makes the process of assignment easier as it eliminates multiple potential alternative connectivities between peaks. A well-designed construct can also limit the number of proteolytic sites, expanding the longevity of the sample in the spectrometer. The lack of tertiary structure makes IDPs less sensitive to extreme pH or temperature and allows to manipulate these two parameters. Lower pH and lower temperatures can slow down the rate of proton amide exchange and can be used to enhance the intensity of observable peaks (Zerbe and Bader, no date).

To summarise, NMR is the most appropriate tool to study the dynamic nature of IDP, as it only requires a protein in solution of appropriate concentration and relatively high purity, instead of obtaining a homogenous conformational arrangement. However, the experimental set-up requires optimisation and protein isotope labelling which might be time- and protein-consuming. However, once the technical limitations are overcome, NMR opens a door to a wide range of data that can be obtained, ranging from dynamics and structural studies to interrogation of interactions.

1.7 Introduction to thesis

Despite the prevalent role of myc in cancer, detailed structural, biochemical and biophysical knowledge of this family of proteins is lacking. Due to the lack of studies on N-myc specifically, a lot of assumptions are extrapolated from research performed on c-myc, which due to differential expression patterns, physiological triggers, regulatory mechanisms and a different panel of interactors, is not always an appropriate assumption on N-myc functional and structural mechanisms. This project aims to address some of these lacking data. Previously, our group has published an X-ray crystal structure of an N-myc TAD fragment bound to Aurora A kinase which revealed an important myc interaction sequence that is not contained within conserved MBs (Richards et al., 2016). This served to solidify the idea that not all mechanisms are conserved between N- and c-myc and thus dedicated research is required for each protein.

1.7.1 Structural studies of the whole N-myc TAD

NMR is employed as the main tool to study the dynamics and structural propensity of the N-myc TAD. Chapter 3 describes in detail multiple approaches undertaken in the assignment of the N-myc TAD backbone and an analysis of N-myc TAD in the context of its intrinsically disordered structure. These approaches range from standard ^{1}H-detected triple resonance spectra collection, utilising spatial correlations between nearby nuclei, utilisation of ^{13}C- and ^{15}N-detected spectra, as well as further truncating N-myc TAD to reduce spectral complexity and improve spectral quality.

Chapter 4 focuses on the structural and dynamics characteristics of the N-myc TAD. H^N, N, C', Cα and Cβ chemical shift index was utilised to determine the structural propensity of the TAD. This aspect was analysed in further detail using relaxation experiments to determine both global and per-residue dynamics of the entire protein. This is put in context with structural predictions, which utilise chemical shifts and PBD entries (CS-Rosetta), as well as theoretical prediction models, such as AlphaFold. Chapter 4 also interrogates the two populations of peaks corresponding to residues forming α-helix with complex with Aurora A (Richards et al., 2016). Circular dichroism (CD) experiments were also employed to investigate the secondary structure of W^{77}-W^{88} region.

1.7.2 Interaction studies of N-myc TAD

To further probe the functional characteristics of N-myc TAD, various biochemical interaction studies were performed to (a) test whether N- and c-myc have similar interaction profiles and (b) test the interaction of N-myc with its putative partners. We focus solely on recombinant proteins to simplify the system used and to be able to capture and characterise specific interactions. We interrogate N-myc's TAD in the context of its binding partners and potential kinases that might target some of its sites.

1.7.3 N-myc as a kinase substrate

N-myc TAD is a phosphoprotein and, as an IDP, is subjected to PTMs which modify its functionality. Chapter 5 is dedicated to interrogating a degradation pathway in N-myc, using NMR to monitor PTMs. We recreated the degradation steps *in vitro* using ERK1, and two GSK3 isoforms to obtain di-phosphorylated N-myc. Then we interrogate less well-investigated steps of the pathway, which is the peptidyl-prolyl bond isomerisation between p-S^{62}-P^{63} and dephosphorylation of p-S^{62} residue.

Selected kinases, such as Aurora A and Plk1 have also been implicated to regulate N-myc stability in an indirect fashion and these were also tested to see if they target N-myc as a substrate.

1.7.4 N-myc as a binding partner

Chapter 6 explores the TAD of N-myc in the context of established and putative interactions, utilising recombinant proteins. Many binding partners for c- and N-myc have been established using *in vivo* assays involving immuno-precipitation and immuno-detection approaches. This limits the characterisation only to species that are probed but potentially omits a number of proteins that contribute to the interaction. This leads to the overestimation of the number of direct interactions that myc proteins have. We employ a range of biophysical and biochemical techniques, including NMR, isothermal titration calorimetry (ITC), pull-down assays and homogenous time-resolved fluorescence assays to characterise these interactions both qualitatively and quantitatively, with a focus on recombinant proteins.

2 Methods

Table 2.1. Buffers and reagents used in the experiments

Name	Composition
Analytical SEC buffer	1x TBS, 2 mM β-mercaptoethanol (BME), 5% v/v glycerol
Base buffer	100 mM KH_2PO_4, 10 mM Tris, 300 mM NaCl, 2 mM BME, pH 7.5
Buffer B	1x TBS, 2 mM tris(2-carboxyethyl)phosphine (TCEP), pH 6.9
Circular dichroism (CD) buffer	10 mM NaCl, 10 mM sodium phosphate (Na_2HPO_4/NaH_2PO_4), pH 7.4
Cdk9 lysis buffer	300 mM NaCl, 50 mM Tris-HCl pH 7.5, 2 mM dithiothreitol (DTT)
Cdk9 size-exclusion buffer	300 mM NaCl, 50 mM Tris-HCl pH 7.5, 2 mM DTT, 10% v/v glycerol
Dialysis buffer	TBS, 2 mM BME, pH 7.4
Homogeneous Time Resolved Fluorescence (HTRF) buffer	PBS, 0.005% v/v Tween-20
Glutathione resin pull-down buffer	300 mM NaCl, 50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 2 BME, 0.02% v/v Tween-20
Kinase assay buffer	50 mM HEPES pH 7.5, 500 mM NaCl, 2 mM $MgCl_2$, 10% v/v glycerol, 5 mM BME
Lysis buffer	1x TBS (Melford), 2 mM TCEP, 0.01% v/v NP-40
Luria Bertani (LB) medium	1% w/v tryptone, 1% w/v NaCl, 0.5% w/v yeast extract
M9 Minimal Media (MM)	0.194% w/v $^{15}NH_4Cl$, 0.4% w/v ^{13}C glucose, 1% v/v 100x MEM Eagle vitamin mix (Lonza Group AG), 1% v/v 1000x micronutrient mix, 2 μM $MgSO_4$, 0.2 μM $CaCl_2$, 0.01 μM $FeSO_4 \bullet 7H_2O$, 50 mM Na_2HPO_4, 25 mM KH_2PO_4, 20 mM NaCl
Micronutrient solution (x1000 stock)	1 μM $ZnSO_4$, 400 μM H_3BO_3, 30 μM $CoCl_2$, 10 μM $CuSO_4$, 80 μM $MnCl_2$, 2.4 μM $ZnSO_4$, 3 μM $(NH_4)_2MoO_4$
NMR buffer	1x TBS, 2 mM TCEP, pH 6.9
Phosphorylation buffer	150 mM NaCl, 25 mM Tris pH 7.5, 0.25 mM TCEP, 10 mM $MgCl_2$, 0.5 mM ATP
PBS - Phosphate Buffered Saline 10x solution (Melford)	1.37 M NaCl, 27 mM KCl, 119 mM phosphate buffer, pH 7.4

PBST - Phosphate Buffer Saline + Tween	PBS, 0.1% v/v Tween-20
Plk1$_{13-345}$ T210V ion exchange buffer	25 mM HEPES pH 7.5, 100 mM NaCl, 2 mM $MgCl_2$, 5% v/v glycerol, 2 mM BME
Plk1$_{13-345}$ T210V kinase buffer	50 mM HEPES pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 5% v/v glycerol, 5 mM BME
Plk1$_{13-345}$ T210V dialysis buffer	50 mM HEPES pH 7.5, 500 mM NaCl, 2 mM $MgCl_2$, 10% v/v glycerol, 2 mM BME
Plk1$_{13-345}$ T210V SEC buffer	20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 10% v/v glycerol, 2 mM BME
Ponceau S stain	0.1% w/v Ponceau S, 5% v/v acetic acid, MilliQ water
Protease inhibitor buffer (100x stock)	4x c0mplete^{TM} Mini EDTA-free Protease Inhibitor Cocktail tablets (Roche) dissolved in 2 mL of NMR buffer
Cdk9:cyclin T_1 pull-down buffer	50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$, 2 mM β-mercaptoethanol, 0.02% v/v Tween-20
Running buffer	NuPAGE® MES SDS Running Buffer (20X), MilliQ water
6x SDS-PAGE loading buffer	0.375 M Tris-HCl pH 6.8, 60% v/v glycerol, 12% SDS, 0.6 M DTT, 0.06% w/v Bromophenol Blue
TAF1-TBP pull-down buffer	20 mM HEPES pH 7.2, 150 mM NaCl, 1 mM TCEP, 10% v/v glycerol, 2 mM $MgCl_2$, 0.02% v/v Tween
TAF1-TBP buffer B	20 mM Tris pH 7.5, 10% v/v glycerol, 100 mM NaCl, 5 mM $MgCl_2$, 2 mM BME
TAF1-TBP SEC buffer	20 mM Tris pH 7.5, 5% v/v glycerol, 100 mM NaCl, 5 mM $MgCl_2$, 2 mM BME
TBS – Tris Buffered Saline **10x Solution (Melford)**	250 mM Tris base, 27 mM KCl, 1.37 M NaCl, pH 7.4
Tris-acetate-EDTA (TAE) buffer	40 mM Tris base pH 8.3, 20 mM acetic acid, 1 mM EDTA
Wash buffer	150 mM NaCl, 25 mM Tris pH 7.5, 0.1% v/v NP-40
Western blot transfer buffer	25 mM Tris pH 7.5, 192 mM glycine, 10% v/v methanol
Western blot blocking buffer	5% (w/v) powdered milk dissolved in PBST

All of the buffers and reagents listed were prepared using Milli-Q water, purified to 18.2 Ω by Milli-Q® Reference Water Purification System. Buffers were degassed using 0.2 μM filtered using Nalgene™ Rapid-Flow™ Sterile Disposable Bottle Top Filters (ThermoFisher Scientific) and a vacuum pump unless otherwise specified.

Table 2.2. Plasmids and antibiotic concentration used

Plasmid	**Antibiotic resistance gene**	**Concentration**
pETM6T1	Kanamycin	50 μg/mL
pET30TEV	Kanamycin	50 μg/mL
pACEBAC1	Gentomycin	50 μg/mL
pACYC	Chloramphenicol	35 μg/mL
pCDF	Spectinomycin	100 μg/mL

Table 2.3. Primers used in construct cloning

N-Myc$_{64-137}$	Forward: 5′ CAT CAT CCA TGG GCA GCC GTG GCT TCG CGG AGC AC 3′ Reverse: 5′ CAT CAT GAA TTC TCA GCC GTG CTG CAG CTT CTC GCT 3′
N-myc$_{64-137}$ L82G L87G	Forward: 5′ ACG GAGATG CTG GGT GAG AAC GAG GGG TGG GGC AGC CC 3′ Reverse: 5′ GGG CTG CCC CAC CCC TCG TTC TCA CCC AGC ATC TCC GT 3′

Table 2.4. List of antibodies used in Western blotting

Epitope	Dilution	Type of antibody	Manufacturer
Goat anti-Gst (#27-4577-01)	1:1000	Primary	Sigma Aldrich
Mouse anti-FLAG (#F1804)	1:1000	Primary	Sigma Aldrich
Donkey anti-goat IgG HRP-conjugated (#sc-2020)	1:5000	Secondary	Santa Cruz
Sheep ECL™ anti-mouse IgG HRP-conjugated (#NA931V)	1:5000	Secondary	GE Healthcare

Table 2.5. Peptide sequences used in experimental procedures

Peptide	Sequence
N-myc$_{28-89}$	Ac-YPDEDDFYFGGPDSTPPGEDIWKKFELLPTPP LSPSRGFAEHSSEPPSWVTEMLLENELWGS-NH$_2$
N-myc$_{28-89}$ pT^{58}	Ac-YPDEDDFYFGGPDSTPPGEDIWKKFELLPpTPP LSPSRGFAEHSSEPPSWVTEMLLENELWGS-NH$_2$
N-myc$_{28-89}$ p-S^{62}	Ac-YPDEDDFYFGGPDSTPPGEDIWKKFELLPTPP LpSPSRGFAEHSSEPPSWVTEMLLENELWGS-NH$_2$
N-myc$_{28-89}$ p-T^{58} p-S^{62}	Ac-YPDEDDFYFGGPDSTPPGEDIWKKFELLPpTPP LpSPSRGFAEHSSEPPSWVTEMLLENELWGS-NH$_2$
N-myc$_{62-89}$	Ac- AEHSSEPPSWVTEMLLENELWGS-NH$_2$
N-myc$_{73-89}$	Biot-EPPSWVTEMLLENELWG-NH$_2$
N-myc$_{73-89}$ L82G L87G	Biot-EPPSWVTEMGLENEGWG-NH$_2$
N-myc$_{1-25}$	Biot-MPSCSTSTMPGMICKNPDLEFDSLQ-NH$_2$
N-myc$_{26-50}$	Biot-PCFYPDEDDFYFGGPDSTPPGEDIW-NH$_2$
N-myc$_{51-75}$	Biot-KKFELLPTPPLSPSRGFAEHSSEPP-NH$_2$
N-myc$_{76-100}$	Biot-SWVTEMLLENELWGSPAEEDAFGLG-NH$_2$
N-myc$_{101-125}$	Biot-GLGGLTPNPVILQDCMWSGFSAREK-NH$_2$
N-myc$_{117-141}$	Biot-WSGFSAREKLERAVSEKLQHGRGPP-NH$_2$

Ac: acetylation, p: phosphorylation, biot: biotynylation, underlines residues indicate either post-translational modifications or mutated residues

Table 2.6. Amino acid sequence of protein constructs used experimental procedures

Construct	Sequence
N-myc TAD	GAMMPSCSTSTMPGMICKNPDLEFDSLQPCFYPDEDDFYFGGPDSTPPGED IWKKFELLPTPPLSPSRGFAEHSSEPPSWVTEMLLENELWGSPAEEDAFGL GGLGGLTPNPVILQDCMWSGFSAREKLERAVSEKLQHG
N-myc$_{64-137}$	GAMGSRGFAEHSSEPPSWVTEML**L**ENE**L**WGSPAEEDAFGLGGLGGLTPNPV ILQDCMWSGFSAREKLERAVSEKLQHG
N-myc$_{64-137}$ L82G L87G	GAMGSRGFAEHSSEPPSWVTEML**G**ENE**G**WGSPAEEDAFGLGGLGGLTPNPV ILQDCMWSGFSAREKLERAVSEKLQHG
N-myc TAD S7A	GAMMPSCST**A**TMPGMICKNPDLEFDSLQPCFYPDEDDFYFGGPDSTPPGED IWKKFELLPTPPLSPSRGFAEHSSEPPSWVTEMLLENELWGSPAEEDAFGL GGLGGLTPNPVILQDCMWSGFSAREKLERAVSEKLQHG
TAF1-TBP	MHHHHHHSSGLVPRGSGMKETAAAKFEENLYFQGAMADIGSMSSGENLYFQ GSHMGKTNLANEDEAYEAIFGGEFGSLEIGSYIGGDEARNSKDYTEHLPDA VDFEDEDELADDDDDLPGGGSGGGSGGGSMSGIVPTLQNIVATVTLGCRLD LKTVALHARNAEYNPKRFAAVIMRIREPKTTALIFASGKMVVTGAKSEDDS KLASRKYARIIQKIGFAAKFTDFKIQNIVGSCDVKFPIRLEGLAFSHGTFS SYEPELFPGLIYRMVKPKIVLLIFVSGKIVLTGAKQREEIYQAFEAIYPVL SEFRKM
Aurora A$_{122-403}$ C290A C393A	GAMESKKRQWALEDFEIGRPLGKGKFGNVYLAREKQSKFILALKVLFKAQL EKAGVEHQLRREVEIQSHLRHPNILRLYGYFHDATRVYLILEYAPLGTVYR ELQKLSKFDEQRTATYITELANALSYCHSKRVIHRDIKPENLLLGSAGELK IADFGWSVHAPSSRRTTL**A**GTLDYLPPEMIEGRMHDEKVDLWSLGVLCYEF LVGKPPFEANTYQETYKRISRVEFTFPDFVTEGARDLISRLLKHNPSQRPM LREVLEHPWITANSSKPSN**A**QNKESASKQS
Plk1$_{326-603}$ H538A K540M	GAMAPSSLDPSNRKPLTVLNKGLENPLPERPREKEEPVVRETGEVVDCHLS DMLQQLHSVNASKPSERGLVRQEEAEDPACIPIFWVSKWVDYSDKYGLGYQ LCDNSVGVLFNDSTRLILYNDGDSLQYIERDGTESYLTVSSHPNSLMKKIT LLKYFRNYMSEHLLKAGANITPREGDELARLPYLRTWFRTRSAIILHLSNG SVQINFFQD**A**T**M**LILCPLMAAVTYIDEKRDFRTYRLSLLEEYGCCKELASR LRYARTMVDKLLSSRSASNRLKAS
Plk1$_{13-345}$ T210V	LEAPADPGKAGVPGVAAPGAPAAAPPAKEIPEVLVDPRSRRRYVRGRFLGK GGFAKCFEISDADTKEVFAGKIVPKSLLLKPHQREKMSMEISIHRSLAHQH VVGFHGFFEDNDFVFVVLELCRRRSLLELHKRRKALTEPEARYYLRQIVLG CQYLHRNRVIHRDLKLGNLFLNEDLEVKIGDFGLATKVEYDGERKK**V**LCGT PNYIAPEVLSKKGHSFEVDVWSIGCIMYTLLVGKPPFETSCLKETYLRIKK NEYSIPKHINPVAASLIQKMLQTDPTARPTINELLNDEFFTSGYIPARLPI TCLTIPPRFSIAPSSLDPSNRKPLTVLNK
Plk1$_{3-326}$ T210D	GAMAAVTAGKLARAPADPGKAGVPGVAAPGAPAAAPPAKEIPEVLVDPRSR RRYVRGRFLGKGGFAKCFEISDADTKEVFAGKIVPKSLLLKPHQREKMSME ISIHRSLAHQHVVGFHGFFEDNDFVFVVLELCRRRSLLELHKRRKALTEPE ARYYLRQIVLGCQYLHRNRVIHRDLKLGNLFLNEDLEVKIGDFGLATKVEY DGERKK**D**LCGTPNYIAPEVLSKKGHSFEVDVWSIGCIMYTLLVGKPPFETS CLKETYLRIKKNEYSIPKHINPVAASLIQKMLQTDPTARPTINELLNDEFF TSGYIPARLPITCLTIPPRFS
Gst-Cdk9:cyclin T$_1$	MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLE FPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDI RYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDF MLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQ GWQATFGGGDHPPKSDLEVLFQGPAKQYDSVECPFCDEVSKYEKLAKIGQG TFGEVFKARHRKTGQKVALKKVLMENEKEGFPITALREIKILQLLKHENVV NLIEICRTKASPYNRCKGSIYLVFDFCEHDLAGLLSNVLVKFTLSEIKRVM QMLLNGLYYIHRNKILHRDMKAANVLITRDGVLKLADFGLARAFSLAKNSQ PNRYTNRVVTLWYRPPELLLGERDYGPPIDLWGAGCIMAEMWTRSPIMQGN

	TEQHQLALISQLCGSITPEVWPNVDNYELYEKLELVKGQKRKVKDRLKAYVRDPYALDLIDKLLVLDPAQRIDSDDALNHDFFWSDPMPSDLKGMLST
	MEGERKNNNKRWYFTREQLENSPSRRFGVDPDKELSYRQQAANLLQDMGQRLNVSQLTINTAIVYMHRFYMIQSFTQFPGNSVAPAALFLAAKVEEQPKKLEHVIKVAHTCLHPQESLPDTRSEAYLQQVQDLVILESIILQTLGFELTIDHPHTHVVKCTQLVRASKDLAQTSYFMATNSLHLTTFSLQYTPPVVACVCIHLACKWSNWEIPVSTDGKHWWEYVDATVTLELLDELTHEFLQILEKTPNRLKRIWNWR

2.1 DNA cloning techniques

2.1.1 Bacterial transformation

E. coli DH5α were used for plasmid amplification and BL21-CodonPlus®(DE3)-RIL (Agilent Technologies) were used for protein expression unless otherwise specified. The DH5α strain is characterised by high plasmid stability, they are devoid of endonuclease activity that targets exogenous plasmids (Hanahan, 1983). Frozen chemically competent DH5α bacteria glycerol stocks were thawed on ice and 1 µL of the plasmid of interest (Table 2.2) was added to 50 µL of bacteria and incubated on ice for 15 mins. Bacteria were then heat shocked at 42 °C for 45 s, followed by a recovery on ice for further 2 mins. 200 µL of sterile LB media was supplemented to the bacteria and the transformation reaction was placed at 37 °C for 1 hr, shaking at 200 rpm to allow antibiotic resistance gene expression. 100 µL of transformed bacteria were plated out on agar plates with an appropriate combination and concentration of antibiotics (Table 2.2). Agar plates were incubated overnight at 37 °C.

2.1.2 Isolation of plasmid DNA from *E. coli*

A single colony of *E. coli* DH5α bacteria carrying the plasmid of interest were inoculated in a sterile 50 mL Falcon tube with 5 mL of autoclaved LB media, in addition to plasmid-specific antibiotics (Table 2.2). Cultures were grown at 37 °C overnight in an orbital incubator set for 200 rpm. Bacteria were first pelleted by centrifugation at 2200 x *g* for 10 mins at RT and the LB was decanted. Plasmids were extracted from bacteria cells using QIAprep Spin Miniprep Kit (QIAgen Inc.) and the manufacturer's instructions. with UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen™) was used to elute the DNA.

2.1.3 Estimation of plasmid concentration

To test the amount of the plasmid extracted, DeNovix DS-11 spectrophotometer, set to 'ds DNA' mode was utilised. This setting measures the absorbance at 260 nm, but also informs of RNA and carbohydrate impurities (230 nm absorption) and protein impurities (280 nm absorption) in form of 260/230 ratio and 260/280 ratio, respectively. >1.0µL of extracted plasmid was loaded onto the sample surface and the absorbance at 260 nm was measured. The plasmid stocks were kept frozen at -20 °C.

2.1.4 Polymerase chain reaction

To generate N-myc$_{64-137}$, polymerase chain reaction (PCR) was used. The reaction contained 2.5 μL of 10 μM forward and reverse primers (Table 2.3), 1 μL of petM6T1 N-myc TAD, 12.5 μL of Q5® High-Fidelity 2X Master Mix (New England BioLabs®, Inc.) and topped up to 25 μL with UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen™). All components were gently mixed and placed in a thermal cycler with the following settings:

1. Activate the polymerase by heating to 98 °C for 2 mins.
2. Denaturing at 98 °C for 5 secs
3. Annealing at 72 °C for 2 mins
4. Extending at 72 °C for 2 mins,

with steps 2 – 4 repeated 30 times. The reaction was then held at 4 °C on an 'infinite' setting. PCR reactions were analysed through DNA electrophoresis.

2.1.5 Site-directed mutagenesis

To generate N-myc TAD$^{L82G\ L87G}$, mutagenesis was performed using the Stratagene QuikChange protocol with primer sequences listed in Table 2.3. The total volume of the reaction was 50 μL, with 2.5 μL of 10 μM primers (0.5 μM), 1 μL of petM6T1 N-myc$_{64-137}$ template, 1.5% v/v DMSO and 5 μL of x10 Pfu Reaction buffer (Promega), 50 μM of each dNTPs and 1 μL of Pfu DNA polymerase (Promega) added as a final reagent and topped up with with UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen™). The PCR reactions were performed in Techne Prime Full size thermal cycler with the following settings:

1. Activate the polymerase by heating to 95 °C for 30 secs
2. Denaturing for at 95°C for 30 secs
3. Annealing for at 55 °C for 1 min
4. Extending for at 68 °C for 5 mins

with the last three steps repeated 18 times. The PCR reaction was tested through DNA agarose gel electrophoresis to establish predicted size of the construct (Chapter 2.1.8).

2.1.6 Restriction DNA digest

All of the restriction digest reactions were carried out as a multiple of 20 μL final reaction volume, by mixing 2 μL of UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen™), 2 μL of 10x CutSmart® buffer (New England BioLabs®, Inc.), 14 μL of extracted plasmids and 1μL of respective enzymes, added to the reaction as last reagents and the reaction were incubated 37 °C for ~ 1 hr. The restriction enzymes were sourced from New England Biolabs®, Inc. and, when available, in High-Fidelity form.

2.1.7 DNA electrophoresis

2.1.7.1 Preparation of 1% w/w agarose gel

1 g of powdered UltraPure agarose (Promega) was dissolved in 100 mL of 1x TAE buffer (Table 2.1), briefly mixed and dissolved for ~2 min in a microwave until agarose powder was completely dissolved. Following a brief cooling period on a bench, SYBR® Safe DNA gel stain (Invitrogen Inc.) was added to agarose in dilution of 1:10 000 and mixed thoroughly by swirling. The agarose was poured into a gel cast, with a well cone inserted to mould wells for sample loading. When the gel was firm, the cone was taken out, and the gel was placed in a DNA agarose gel tank and submerged in 1x TAE buffer (Table 2.1). Samples were prepared by mixing 5 parts of the sample with 1 part of 6x DNA loading dye (New England BioLabs®, Inc.) and loaded carefully into the wells. To monitor the electrophoresis, either 1 kb DNA Ladder Plus or 100 bp DNA Ladder (New England BioLabs®, Inc.) was used, depending on the size of DNA investigated. The electrophoresis was set to run at constant voltage of 120 V for ~1 hr or until the blue dye in the sample reached the bottom of a gel. DNA was visualised on TruBlue™ Blue Light Transilluminator box (Edvotek®) which excites SYBR® Safe DNA gel stain in the blue spectrum of 509 nm and imaged using G:Box XT4 and GeneSnap Image Acquisition Software (SYNGENE), set to 'Transilluminator' with exposure times in the millisecond range.

2.1.7.2 DNA band purification following DNA electrophoresis

Bands of interest were identified using the TruBlue™ Blue Light Transilluminator box and then were excised using a sterile scalpel. The gel was placed in a sterile 1.5 mL Eppendorf tube and the DNA was extracted using QIAquick Gel Extraction Kit (QIAgen) kit and protocol. with UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen™) was used to elute the DNA.

2.1.8 Ligation reaction

The ligation reactions were performed in accordance with T4 ligation kit (New England BioLabs). Vector and insert were mixed in 3:1 molar ratio, with 2 μL of 10x Buffer for T4 DNA ligase (New England BioLabs) and 1 μL of T4 DNA ligase (New England BioLabs). The reaction was topped up to 20 μL with UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen™). Negative control was set-up by substituting the insert with UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen™). The ligation was incubated at RT for a minimum of 1 hr and a maximum overnight. Following the ligation reaction, 2 μL of the ligation reaction was transformed in *E. coli* DH5α strain as described in chapter 2.1.2.

2.1.9 Plasmid amplification and sequencing

If the reaction plate yielded a noticeably higher number of colonies than the negative control, few individual colonies were selected and inoculated in 10 mL of sterile LB medium, amplified overnight and plasmids were extracted as per section 2.1.2. The successful ligation reaction was validated by sending 15 μL of extracted plasmid for automated Sanger Sequencing using TubeSeq service provided by Eurofins Genomics. T7 and T7 term primers were used for sequencing of petM6T1 and petTEV30 vectors.

2.2 Insect cell techniques

2.2.1 Insect cells

Insect cells are used for production of proteins that require extensive folding and significant PTMs. They can also be used when recombinant protein is toxic to bacteria cells. Protein expression using insect cells requires a generation of bacmid, which was done according to the Grenoble manual (Berger et al., 2019). The virus was then used to infect *Spodoptera frugiperda* (Sf9) cells and to induce protein production.

2.2.2 Bacmid generation

pABEBAC1 plasmids containing Cdk9$_{1\text{-}330}$ and cyclin T1$_{1\text{-}259}$, under the control of the same promoter, were transformed into Multibac expression system (courtesy of Richard Heath). The bacmid was generated in the following way. 10 µg of pABEBAC1 plasmid was transformed into 100 µL of DH10MultiBac^{TM} cells via electroporation, followed by recovery in 900 µL of SOC media for 4 hr at 37 °C and 220 rpm. DH10MultiBac^{TM} were then plated on agar containing 7 µL/mL gentamycin, 10 µL/mL tetracyclin, 50 µL/mL kanamycin, 100 µL/mL 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal), 40 µg/mL isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated at 37 °C for 24 hr. White colonies were selected for further incubation in 10 mL of LB containing same set of antibiotics as in the SOC media, at 37 °C and 200 rpm overnight in an orbital shaker. The bacmids were extracted using QIAprep Spin Miniprep Kit (QIAgen) and the manufacturer's protocol with the following changes. Following the addition of 350 mL of neutralising Buffer N3 (Chapter 2.1.2), the centrifugation proceeded for 15 mins at 16 000 x *g*, to remove cell debris. DNA was then precipitated by addition of 700 µL of isopropanol, then 200 µL of 70% v/v ethanol, with supernatant being aspirated gently following each step. 50 µL of 70% v/v ethanol was then added to the pellet and the pellet was frozen at -20 °C, ready for Sf9 cell transfection.

2.2.3 Insect cell transfection

4 hr prior to the transfection, Sf9 cells, cultured in Sf-900™ II SFM media containing 1x Antibiotic-Antimycotic (Gibco^{TM}), were diluted to 0.33x 10^6/mL cells in a final volume of 6 mL in 25 mL Corning™ Cell Culture Treated Flasks (Sigma-Aldrich) and placed for 4 hr at 28 °C, for adhesion. In a sterile hood, all ethanol was removed from extracted bacmid and left to air dry. 36 µL of sterile H_2O was added to the bacmid pellet and mixed thoroughly. 300 µL of Sf-900™ II SFM media was added then to each bacmid and left to incubate for 10 min. Transfection reagent mix was prepared with 10 µL of X-treme GENE™ HP DNA Transfection Reagent (Sigma Aldrich) diluted 10-fold with Sf-900™ II SFM media. 105 µL of the transfection reagent mix was added to each bacmid and incubated for another 10 min. Subsequently, the bacmid reagent was pipetted gently into pre-diluted SF9 cells. After 24 hr, the media containing the transfection reagent mix was replaced with Sf-900™ II SFM media. Cells were grown at 28 °C for five days with no agitation, shielded from the light.

2.2.4 Virus propagation, amplification and protein expression

250 µL of Sf-900™ II SFM media (Gibco^{TM}) from transfected SF9 cells, was added to 50 mL of freshly amplified 1.7x 10^6/mL SF9 cells, which were then grown for a further three days at 28 °C at 120 rpm. The scaled bacmid-containing media was harvested by centrifugation 250 *x g* for 10 min. 30 mL of the bacmid-containing supernatant (P_1) was then used to infect 300 mL of 1.7 x10^6/mL SF9 cells and incubated in the dark for three days. The 300 mL cell suspension was centrifuged at 700 x *g* for 30 min. The supernatant was decanted, and pelleted cells were washed with 150 mL of chilled

PBS, followed by another spin. The pellets were either frozen at -80 °C until needed or lysed for protein purification.

2.3 General protein techniques

2.3.1 Protein expression in *E.coli*

BL21-CodonPlus®(DE3)-RIL Competent Cells (Agilent Technologies) were used for protein expression in bacteria. This strain has been modified to also include extra plasmids encoding genes for rare prokaryotic codons, such as *argU*, *ileY*, and *leuW* tRNA genes (BL21-CodonPlus Competent Cells Instruction Manual, Agilent Technologies). Plasmid encoding a protein of interest were transformed into BL21-CodonPlus®(DE3)-RIL Competent Cells as described in chapter 2.1.1, with the inclusion of 35 µg/mL chloramphenicol to agar plates and in subsequent steps. Single colonies were grown overnight at 200 rpm at 37 °C and used as started cultures for large scale expression.

10 mL of starter culture was used to inoculate 1 L of LB broth (containing same selection of antibiotics). Cultures were grown at 200 rpm at 37 °C until they reached O.D of ~0.6, at which point the temperature was reduced to 20 °C and cultures were induced with 0.6 mM IPTG followed by overnight incubation. The following morning pellets were harvested at 4000 rpm (JLA 8.100, Beckman Coulter) for 20 mins at 4 °C. Pellets were either frozen at -80 °C or further processed.

2.3.2 Expression of isotope-labelled protein for NMR

2.3.3 Preparation of M9 minimal media

All laboratory glassware and equipment used for preparation of M9 minimal media (MM) were autoclaved or sterilised with ethanol prior to media preparation. 250 mL of minimal media was prepared for every 1L of LB (Table 2.1). 0.01 g/mL of ^{13}C glucose and 0.002 g/mL of $^{15}NH_4Cl$ were dissolved first in 10 mL of autoclaved MilliQ water by vigorous vortexing. This was supplemented with 250 μL of 1000x micronutrient solution (courtesy of Mark Richards), 2 μM $MgSO_4$, 0.2 μM $CaCl_2$, 0.01 mM $FeSO_4 \cdot 7H_2O$, 1x BME vitamin solution (BME vitamins 100x solution, Sigma-Aldrich). This solution was filtered through 0.2 μm Minisart® Syringe Filter Polyethersulfone membrane filter (Sartorius) and was topped up to 250 mL with sterile MilliQ water. MM was sealed to avoid airborne contamination.

2.3.3.1 ^{15}N- and ^{13}C- isotope labelling of proteins

Initial stages of isotope-labelled protein expression were the same as for non-labelled protein, up until reaching the O.D. of 0.6 (described in chapter 2.3.1). Cultures were then pelleted at 2500 x *g* at 20 °C for 20 min. LB media was decanted and the pellets were resuspended in ~500 mL of autoclaved PBS buffer (Table 2.1) to remove residual LB. Bacterial cultures were centrifuged again as described above, PBS was decanted, and the pellets were resuspended in 250 mL of MM, transferred to autoclaved 2.5 L Erlenmeyer flasks and incubated at 20 °C for further 2 hrs at 200 rpm. Following this incubation period, protein expression was induced with 0.6 mM IPTG as a final concentration. Bacteria were then returned to the incubator for overnight growth and expression and were harvested on the following day as described in section 2.3.1.

2.3.4 Protein purification

All buffers were prepared a day prior to purification, degassed and chilled overnight. All purification steps were performed on ice or in the cold room, unless otherwise specified. Each step of protein purification described in this chapter was followed by analysis through SDS-PAGE, as described in Chapter 2.5.2.

2.3.4.1 Purification of N-myc TAD and N-myc$_{64-137}$

Standard N-myc TAD and N-myc$_{64-137}$ constructs used was cloned into pETM6T1 plasmid which contains a NusA solubility tag (Table 2.6). Pellets were either harvested immediately prior to the purification or thawed in cold water following -80 °C storage. Bacterial pellets were resuspended using buffer B (Table 2.1) supplemented with 2 c0mplete^{TM} Mini EDTA-free Protease Inhibitor Cocktail tablets (Roche) and 10% w/v lysozyme. Samples were sonicated on slushy ice at 10 s - on/ 20 s - off pulses at 40% amplitude for 3 mins using Sonics VibraCell^{TM} sonicator. The lysate was then clarified using pre-chilled high-speed centrifuge (Avanti J-26XP, Beckman Coulter JA-25.50 rotor) set for 50 000 *x g* for 1 hr at 4 °C. ~2 mL bed volume of slurry cobalt resin (HIS-Select® Cobalt Affinity Gel, Sigma; hereafter referred to as cobalt gel resin) was applied to onto Glass Econo-Pac® Chromatography Columns gravity flow column (BioRAD), washed with 50 mL of Milli-Q water and equilibrated with 50 mL buffer B (Table 2.1). The resin was drained and clarified lysate was added onto the capped column and incubated on a rotatory shaker for 30 mins at 4 °C. Flow-through was applied twice to maximise protein binding. Elution was done in batch with increasing imidazole concentrations of 0 mM, 5 mM, 10 mM, 100 mM, 150 mM and 500 mM, added to buffer B. Both N-myc constructs (TAD and 64-137) eluted over a wide range of imidazole, between 5 mM – 150 mM, and these fractions were pooled together and concentrated to the volume of >50 mL, using 3 kDa or 10 kDa Vivaspin Turbo 15 centrifugal concentrators (Sartorius), for N-myc TAD and N-myc$_{64-137}$, respectively. 1:50 mg ratio of TEV-Nia protease-to-protein (courtesy of Amelia Lesiuk) to the substrate was added to remove the 6xHis-NusA tag. The concentrated protein was loaded onto pre-soaked SnakeSkin® Dialysis Tubing 3,500 MWCO (ThermoFisher Scientific) and dialysed in 4 L of buffer B (Table 2.1) at 4 °C overnight.

Dialysed protein was loaded onto HiTrap Q FF 5 mL anion exchange column (GE Healthcare), pre-equilibrated with Buffer B (Table 2.1) and eluted using Äkta Pure system with NaCl gradient increasing to 500 mM, applied over 20 mins at 3 mL min^{-1} flow. N-myc TAD and N-myc$_{64-137}$ eluted at flow-through and low salt concentrations. Fractions were resolved through SDS-PAGE gel and fractions containing N-myc protein were applied onto Glass Econo-Pac® Chromatography Columns gravity flow column (BioRAD), with 2 mL bed volume of 50% slurry cobalt gel resin and incubated on a rotatory shaker for 30 mins at 4 °C. Column was washed with buffer B with either no imidazole or supplemented with 5 mM, 10 mM and 150 mM imidazole, and the proteins eluted at flow-through, 0 mM and 5 mM imidazole concentrations. Fractions containing N-myc protein were pooled together and volume was reduced to > 5 mL using 5 kDa Vivaspin Turbo 15 centrifugal concentrators or 1 kDa Vivaspin® centrifugal concentrators (Sartorius) for N-myc TAD and N-myc$_{64-137}$, respectively. Proteins were subjected to size-exclusion chromatography (SEC) using Superdex 16/60 75pg column (GE Healthcare), pre-equilibrated with buffer B, and purified using Äkta Pure system with 5 mL loop and 1 mL min^{-1} flow rate. The eluted proteins were concentrated again post-SEC as described above to achieve an appropriate final concentration. Proteins were aliquoted, snap frozen in liquid nitrogen and stored at -80 °C until further processing.

2.3.4.2 N-myc TAD expression into inclusion bodies

For the purposes of obtaining N-myc TAD (Table 2.6) which is less prone to proteolytic degradation, N-myc was cloned into a pETTEV30 vector (courtesy of Mark Richards) which contains a TEV-cleavable 6xHis tag but lacks the NusA solubility tag. The protein was purified from inclusion bodies using the protocol of Eoin Leen. In brief, bacterial pellets were resuspended in base buffer (Table 2.1) supplemented with 10% w/v of lysozyme and 1 μg/mL final concentration of DNaseI. Cells were sonicated on slushy ice for 2 mins 30 secs with 15 s - on/ 15 s - off pulses at 40% amplitude and the lysate for clarified at 30 000 *x g* for 30 mins at 4 °C. Pellets were resuspended using 30 mL of 1 M urea buffer made fresh (1 M urea dissolved in base buffer, Table 2.1) and sonicated for 2 mins with 15 s - on/ 15 s - off pulses at 40% amplitude and spun for 20 mins at 30 000 x *g* at 4 °C. This was repeated three more times, each time increasing the concentration of urea to 2 M, 4 M and 6 M, respectively. Fractions were resolved on SDS-PAGE (Chapter 2.4.2) and the solution containing N-myc TAD was diluted 4 fold with dialysis buffer (Table 2.1, with TCEP replaced by 2mM BME) to the total volume of 80 mL and the lysate was incubated with 2 mL bed volume of pre-washed and pre-equilibrated cobalt gel resin for 50 mins on a rotatory shaker at 4 °C. Protein was eluted with base buffer supplemented with increasing imidazole concentrations (0 mM, 5 mM, 10 mM, 100 mM, 150 mM and 500 mM). Fractions containing N-myc TAD were loaded onto pre-soaked SnakeSkin® Dialysis Tubing 3,500 MWCO (ThermoFisher Scientific) with 1:50 mg of TEV-Nia protease-to-protein ratio (courtesy of Amelia Lesiuk) to protein and dialysed overnight in 4 L of pre-chilled dialysis buffer (Table 2.1).

Dialysed protein was subjected to subtractive affinity chromatography, using 25 mL volume of elutions with 0 mM, 5 mM, 10 mM and 150 mM imidazole dissolved in dialysis buffer. N-myc TAD was present in flow-through fractions, which were pooled and concentrated on 5 kDa Vivaspin Turbo 15 (Sartorius) centrifugal concentrators to >5 mL volume. The final purification step was performed on SEC Superdex 16/60 75pg column (GE Healthcare) as described in chapter 2.3.4.1. The peak fractions were pooled, concentrated, snap-frozen in liquid nitrogen and stored at -80 °C.

2.3.5 Purification of N-myc partners

2.3.5.1 Cdk9:cyclin T_1

All protein variants of Cdk9:cyclin-T_1 (Table 2.6) were expressed in, and purified from insect cells. Cdk9 was N-terminally tagged with glutathione S-transferase (Gst) followed by 3c PreScission Protease site. The pellets were resuspended in ~30 μL of Cdk9 lysis buffer (Table 2.1) supplemented with one c0mplete^{TM} Mini EDTA-free Protease Inhibitor Cocktail tablet (Roche) per ~30 mL of lysis buffer. Cells were sonicated with 10 s - on/ 20 s - off pulses at 40% amplitude for 2 mins total sonication time. Lysates were clarified at 50 000 *x g* for 45 mins at 4 °C. Protein was bound to 1.5 mL bed volume of freshly equilibrated Gst-sepharose 4B (GE Healthcare) resin for 2 hrs at 4 °C in Econo-Pac® Chromatography Columns gravity flow column (BioRAD). Flow-through was applied twice to maximise protein binding. The protein was then washed three times with 10 bed resin volumes Cdk9 SEC buffer (Table 2.1) and eluted with 15 mL of 10 mM glutathione pH 8.0 and 20 mM glutathione pH 9.0. Fractions containing Cdk9:cyclin T_1 were concentrated below 5 mL volume using Millipore Amicon Ultra-15 Centrifugal Filter Units and purified further using Superdex 16/60 200 column (GE Healthcare), pre-equilibrated with Cdk9 SEC buffer, and eluted using Äkta Pure system with 5 mL loop and 1 mL min^{-1} flow rate with 1 mL fractions. Peak fractions were concentrated as described above using 10 kDa Amicon Ultra-15 Centrifugal Filters and protein was aliquoted into appropriate volume fractions, snap frozen in liquid nitrogen and stored at -80 °C.

2.3.5.2 Plk1$_{13\text{-}345}$ T210V

Plk1$_{13\text{-}345}$ T210V was expressed as a fusion protein with 6x His tagged monomeric ultra-stable GFP (muGFP) at its N-terminus (Table 2.6), with TEV-Nia site between the two proteins. This construct was expressed in insect cells (courtesy of Eoin Leen). The harvested pellet was resuspended in Plk1$_{13\text{-}345}$ T210V kinase buffer (Table 2.1) and sonicated for 10 s - on/ 20 - s off for 5 mins at 40% amplitude. The lysate was clarified at 40 000 *x g* for 20 mins at 4 °C and it was mixed with 5 mL pre-equilibrated HIS-Select® Cobalt Affinity Gel (Sigma) in Glass Econo-Pac® Chromatography Columns gravity flow column (BioRAD) and rotated for 50 mins at 4 °C with agitation. Recombinant protein was eluted using Plk1$_{13\text{-}345}$ T210V kinase buffer (Table 2.1), supplemented with increasing concentration of imidazole (0 mM, 5 mM, 10 mM, 100 mM, 150 mM and 500 mM). Fractions containing muGFP-Plk1$_{13\text{-}345}$ T210V were pooled together, supplemented with 1:50 mg of TEV-NIa protease-to-protein rato and dialyses overnight using 4 L of pre-chilled PLK1$_{13\text{-}345}$ T210V dialysis buffer (Table 2.1). On the following day, the protein was dialysed again against 4 L of pre-chilled Plk1$_{13\text{-}345}$ T210V ion exchange buffer (Table 2.1) for 2 hrs at 4 °C. The dialysed protein was loaded onto HiTrap Q-FF 5 mL anion exchange column (GE Healthcare) and eluted at 2 mL min^{-1} over 30 mins with 100 to 500 mM NaCl gradient. Peak fractions were pooled together and applied to a HisTrap HP 5 mL column (GE Healthcare), pre-equilibrated with Plk1$_{13\text{-}345}$ T210V kinase buffer. Plk1$_{13\text{-}345}$ T210V eluted at low imidazole concentrations. These were concentrated with a 10 kDa Vivaspin Turbo 15 (Sartorius) to > 5 mL and this was further purified on pre-equilibrated on Superdex 16/60 75pg column (GE Healthcare) as described above, using Plk1$_{13\text{-}345}$ T210V SEC buffer. Peak fractions were snap frozen at ~200 μM final concentration and stored at -80 °C.

2.3.5.3 TAF1 - TBP

TAF1-TBP protein construct is a fusion protein between $TAF1_{6-71}$ and TBP_{61-240} with a triple GSSS linking the two proteins (Table 2.6). TAF1 was also 6x His-tagged for the ease of purification. Bacterial pellets were thawed as described above and were resuspended in 30 mL of chilled TAF1-TBP buffer B (Table 2.1), supplemented with 10% w/v lysozyme, 0.1% v/v NP-40 and 1 μg/mL DNase I final concentration. The pellets were sonicated for 5s - on/ 10 s - off for 3 mins at 40% amplitude. The lysate was clarified by centrifugation at 50 000 *x g* for 1 hr at 4 °C, and the supernatant was applied onto clean Glass Econo-Pac® Chromatography Columns gravity flow column (Bio-RAD), with 2 mL bed volume cobalt gel resin, pre-equilibrated with TAF1-TBP buffer B, and incubated on a rotatory shaker for 30 mins at 4 °C. Flow-through was re-applied to the column twice to maximise protein binding. Elution was done in batch with TAF1-TBP buffer B supplemented with an increasing imidazole gradient of 0 mM, 5 mM, 10 mM, 100 mM, 150 mM and 500 mM. The protein eluted over a wide range of imidazole concentrations, between 5 mM – 150 mM, and these fractions were pooled together and concentrated to the volume of >50 mL, using 10 kDa Vivaspin Turbo 15 centrifugal concentrators (Sartorius). The concentrated protein was dialysed overnight using SnakeSkin® dialysis Tubing 3,500 MWCO (ThermoFisher Scientific) against 4 L of TAF1-TBP buffer B (Table 2.1). Dialysed protein was subjected to Q-FF anion exchange 5 mL column (GE Healthcare) and eluted using an increasing salt gradient from 150 mM to 500 mM NaCl, with 2 mL min^{-1} flow rate over the course of 30 mins. TAF1-TBP eluted in flow-through and the low salt concentration fractions and these were pooled together, and the volume was reduced below 5 mL using 10 kDa Vivaspin Turbo 15 (Sartorius) centrifugal filters. The protein was loaded onto Superdex 16/60 75pg column (GE Healthcare) and purified using Äkta Pure system as described in Chapter 2.4.1. Peak fractions containing TAF1-TBP were concentrated using 10 kDa Vivaspin Turbo 15 (Sartorius), snap frozen in liquid nitrogen and stored at -80 °C.

2.3.6 Protein analysis tools

2.3.6.1 Determination of protein concentration using A_{280}

Protein concentration was measured at 280 nm wavelength, using DeNovix DS-11 spectrophotometer set for ‘Protein A280’. The instrument was blanked with 1.2 μL of matching buffer. Various protein dilutions were prepared to obtain a reading within the instrument reading range. The theoretical extinction coefficient was calculated using ProtParam web server (Swiss Prot group) and the final protein concentration was established using Beer-Lambert law:

$$A = \varepsilon c l,$$

Where A is the absorbance in g L^{-1}, l is the length of the path measured at 10 mm, ε is the molar extinction coefficient and c is the concentration in units related to ε.

2.3.6.2 Protein analysis through Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis (SDS – PAGE)

All of the samples subjected to SDS-PAGE analysis were diluted 1:2 with 6x SDS loading buffer (Table 2.1), boiled at 90 °C for ~2 min and briefly centrifuged using a bench-top centrifuge unless otherwise specified. 19 or 24 µL volume samples were loaded onto pre-cast 10- or 15-wells NuPAGE Bis-Tris 10% pre-cast SDS-PAGE gels (Invitrogen), respectively with the gels generously submerged in running buffer (Table 2.1). The gels were run for ~45 mins at a constant voltage of 200 V, or until the appropriate separation was achieved, as observed by the migration of the protein ladder (SeeBlue Plus2 Pre-stained standard, Invitrogen). Protein bands were visualised by staining with InstantBlue® Coomassie Protein Stain (Expedeon) overnight, to ensure adequate visualisation, and then destained in deionised water and gentle agitation.

2.3.6.3 Western blotting

If SDS-PAGE gel was subjected to Western blotting, then no staining procedure was performed. Instead, the gel was briefly washed in Western blot transfer buffer (Table 2.1) to wash off the running buffer. The gel was sandwiched in the following order starting from a negative electrode plate: two sponges, two layers of a blotting paper, SDS-PAGE gel, 0.2 µm Nitrocellulose Blotting Membrane (Amersham™ Protran™ 0.2 µm NC, GE Healthcare), two blotting papers, four sponges and a positive electrode plate. All air bubbles were removed and all components were generously soaked in transfer buffer. Transfer took place over 1 hr at 30 V at room temperature. Alternatively, slow transfer was used where proteins were transferred for 16 hrs at 4 °C at 16 V. Complete transfer was confirmed by brief staining of the membrane with Ponceau stain (courtesy of Amelia Lesiuk, Table 2.1) for 5 mins. The membrane was then blocked for 20 mins in Western blot blocking buffer (Table 2.1). The primary antibody incubation was either performed overnight at 4 °C, or for 1 hr at RT, with 1° antibody added directly to the blocking buffer (Table 2.1) at the appropriate concentration (Table 2.3). The membrane was washed for 20 mins at RT in PBST

(Table 2.1), with the buffer changed every ~5 mins. 2° antibody incubation was also done at RT for 1 hr, with the antibody dissolved in the blocking buffer at the appropriate concentration (Table 2.3). The membrane was then subjected to the same washing procedure as above. The signal was developed using Pierce™ ECL Western Blotting Substrate (ThermoFisher Scientific) as per protocol. In brief, one part of Detection Reagent 1 was mixed with one part of Detection Reagent 2, briefly mixed and directly applied onto NC membrane. The reaction was allowed to proceed for ~1 min. The excess of the reagent was removed and the membrane was enclosed in a clear plastic envelope and the signal was detected using G:Box XT4 and GeneSnap Image Acquisition Software (SYNGENE), using 30 sec, 1 min, 2.5 mins and 5 mins exposure times, depending on strength of the signal.

2.3.6.4 Circular dichroism (CD)

CD was performed on Chirascan Plus (Applied Photophysics) spectrophotometer using 0.1 cm^{-1} length quartz cuvette path. The measurements were collected at duplicates, every 2 nm at 20 °C. The peptides were diluted from their lyophilised form into CD buffer (10 mM NaCl, 10 mM NaH_2PO_4) at 100 μM final concentration and the pH was adjusted to ~7.0 with 0.01 M NaOH. Both 15% v/v and 30% v/v TFE buffers were dissolved in CD buffer initially to the concentration of 30.75% v/v to achieve the final concentration of 30% v/v with 0.1 mg/mL of peptide as a final concentration. The mean residue molar ellipticity was calculated using formula:

$$\Delta\varepsilon = \frac{[\theta]}{10\, x\, l\, x\, M}$$

, where θ is raw ellipticity is milidegrees, *l* is a path length which is 0.1 cm^{-1} and M is concentrations expressed in molarity. The graphs were plotted using Excel software.

2.3.6.5 Mass spectrometry

To confirm the molecular weight of purified proteins, 20 μM protein samples were submitted to University of Leeds Biomolecular Mass Spectrometry centre for intact mass analysis through time-of-flight mass spectrometry (courtesy of Rachel George). The expected protein size was obtained by computing in primary amino acid sequence into Prot Param (Expasy) web server (Swiss Prot group).

2.4 Protein – protein interaction assays

This chapter describes the experimental techniques employed to interrogate N-myc TAD interactions with various binding partners, described in this work.

2.4.1 Homogenous Time Resolved Fluorescence (HTRF) assay

HTRF assay is a Förster resonance energy transfer (FRET)-based assay which detects the fluorescence emitted between donor (Terbium-FLAG) and the acceptor (Streptavidin-XL655), based on the proximity of components of the complex.

Biotinylated N-myc peptides phosphorylated at different positions (Table 2.5) were prepared in 12 different concentrations in 96-well plate using 1:2 serial dilutions. 40 µL of 8000 nM N-myc peptides were first diluted 1:2 with HTRF buffer (Table 2.1) to the final volume of 60 µL. Then 20 µL of Cdk9:cyclin T_1 or cyclin T_1 was added to each of these wells in the final concentration of 60 nM. These were incubated for 1 hr at RT and 10 µL of each of the wells were transferred to a black opaque 384-well plate in duplicates. 100 µL of 1:1000 of Terbium-FLAG (Anti-FLAG® (M2 clone)-Tb, Cisbo) was mixed with freshly-thawed dye Streptavidin-XL655 (Cisbo) at the final concentration of 250 nM and 10 µL of this reaction was added to each of the wells containing N-myc peptides and Cdk9:cyclin T_1. The reactions were incubated for 1 hr at RT, shielded from the light. The plate was read using a PHERAstar FS plate reader (BMG LABTECH) and MARS software, set for detection of fluorescence emission for a donor at 620 nm and fluorescence emission for an acceptor at 665 nm and the HTRF ratio was calculated using the following equation.

$$HTRF\ ratio = \frac{acceptor\ emission}{donor\ emission}.$$

HTRF ratio was plotted as a function of N-myc peptide concentration using Excel.

2.4.2 Kinase assay with SDS – PAGE as a detection method

Various N-Myc$_{28\text{-}89}$ peptides (phosphorylation at various positions (Table 2.5) were diluted in phosphorylation buffer (Table 2.1) to the final concentration of 10 µM in 100 µL final volume. 0.5 µM of the following kinases were added separately into each of the peptides: Cdk9:cyclin T_1, Cdk1:cyclin B, Cdk2:cyclin A and Cdk4:cyclin D_3 (courtesy of Richard Heath), followed by incubation at 37 °C for 2 hrs. The reactions were terminated by incubation at 4 °C. 15 µL of each of the reactions was mixed with 5 µL of 4x Laemmli sample buffer, (BioRAD) buffer and samples were analysed on SDS-PAGE gel as described in chapter 2.3.6.2.

2.4.3 Kinase assay using ADPsensor^{TM} Universal Kinase Activity Assay Kit

This assay detects fluorescence when ADP is generated and interacts with two sensors provided in the kit. The reactions were performed in 384-well black opaque plates in the total volume of 20 µL. 4 µL of ADPsensor™ kinase buffer (provided in the kit) and 4 µL of the enzyme, in the final concentration of 200 nM in 20 µL, were mixed and aliquoted into wells. The remaining part of the reaction consisted of 400 ng/mL of the substrate, 100 µM ATP, 0.8 µL of ADP sensor I and sensor II (pre-made), 1.6 µL of probe (pre-made), and 4 µL of ADPsensor™ kinase assay buffer. MilliQ water was used to bring the final reaction volume to 20 µL. The blank was also included and consisted of all reagents apart from the enzyme, which was substituted with MilliQ water. The fluorescence was read on Tecan Infinite 200 Pro plate reader set to 535 nm excitation wavelength and 590 nm emission wavelength, set for 50 cycles. The experiment was performed three times. Excel was used to calculate the average emission signal of each cycle with the blank signal subtracted. The emission at 535 nm was plotted as a function of cycle numbers using Excel.

2.4.4 Pull-down assay with biotin–binding resin

Each reaction was performed in 450 µL total volume. 80 µL of Strep-tactin® Sepharose® 50% suspension resin (IBA Lifesciences) was thoroughly washed and equilibrated with appropriate pull-down buffers, depending on interrogated complex (described in detail in the Results chapters). Biotinylated N-myc peptides (Table 2.5), which were used as bait, were added to equilibrated resin at final concentration of 10 µM and incubated for 1 hr at 4 °C, with gentle agitation. To control for non-specific binding, the prey was incubated with the resin as well, with no addition of bait. The resin was thoroughly washed. First, the reactions were centrifuged for 2 mins at 16,000 x *g*. The supernatant was then aspirated carefully to avoid the removal of resin beads (the washing procedure). This was repeated two more times. 15 µM final concentration of bait was then added to the reactions, apart from the negative control, and incubated further for 2 hrs at 4 °C, with gentle agitation and 10% of reaction volumes was saved for input analysis. The reactions were subjected four times to the washing procedure. In the final step, the maximum amount of supernatant was aspirated, and the reaction was eluted by the addition of 80 µL of 6x SDS-PAGE loading buffer and resolved on SDS-PAGE as described in Chapter 2.3.6.2.

2.4.5 Pull-down assay with glutathione resin

50 µL of Glutathione Sepharose™ 4B (GE Healthcare) was equilibrated with Cdk9:cyclin T_1 pull-down buffer (Table 2.1) and incubated for 2 hrs at 4 °C with agitation with 50 µg Gst3C Cdk9:cyclin T_1 (Gst-pTEFb ; courtesy of Richard Heath). Non-specific binding of Gst-pTEFb was reduced by washing the resin three times with

500 µL of pull-down buffer, each time followed by centrifugation at 16,000 x *g* for 2 mins with pull-down buffer aspirated gently to avoid resin loss. 50 µg of various N-myc peptides were added to the reactions and the volume was brought up to 150 µL with pull-down buffer. 20% of reaction volume was taken for the input analysis. The peptides were incubated with Gst-pTEFb for 1 hour at 4 °C with gentle agitation, followed by washing with pull-down buffer as described above. With the final wash, maximum volume of supernatant was taken out and the reaction was topped up to 120 µL with 4x Laemmli Sample Buffer (BioRad). 30 µL of the reaction was taken for pull-down analysis and both input and pull-down were analysed on SDS-PAGE.

2.4.6 Isothermal titration calorimetry (ITC)

To minimise the heat of dilution signals, all proteins interrogated were dialysed overnight using GeBAFlex dialysers of varying molecular weight cut-offs (MWCO) to suit the molecular weight of proteins and using a buffer suitable for ITC and protein stability. The dialysis buffers matched the buffers of titrants. Specific concentrations used in ITC experiments are listed in the Results chapter, however concentrations aimed were 15 µM for titrand and 150 – 200 µM for titrant, depending on protein loss during the dialysis. The programme was set to 2 µL volume injections, every 2 mins, for 45 mins at 25 °C and 750 rpm, using MicroCal ITC_{200} Calorimeter (Malvern). The heat changes were recorded as a thermogram and the area under each peak was integrated using ORIGIN software and Wiseman plot was generated. Manual corrections of baselines were performed when necessary, using the same software.

2.4.7 Analytical size – exclusion chromatography (analytical SEC)

Analytical SEC is a method of testing complex formation without the need for affinity tags, which can obscure or impact interaction sites. To interrogate N-myc TAD's ability to form a complex with Fbw7-Skp1, N-myc TAD was first phosphorylated either at S^{62} or at T^{58} S^{62} positions a day prior, and these modifications were monitored using NMR, as described in 2.5.5. Unphosphorylated N-myc TAD, unphosphorylated N-myc TAD^{S7A}, N-myc $\text{TAD}^{\text{p-S62}}$, and N-myc $\text{TAD}^{\text{p-T58 p-S62}}$ were initially run individually to establish their specific elution volumes. Identical proteins were incubated with pre-formed Fbw7-Skp1 complex in equimolar amounts of the final concentration of 86.9 µM in the total volume of 100 µL, for two hours prior to analytical SEC, with gentle agitation at 4 °C. These reactions were then manually loaded onto Superose™ 12 10/300 GL which was pre-equilibrated with two column volumes of analytical SEC buffer (Table 2.1), with the loop volume of 56 µL and the flow rate of 0.5 mL min^{-1}.

Collected fractions were 0.2 mL in volume. Peak fractions were analysed by SDS-PAGE to test complex formation.

The elution volumes of proteins interrogated were compared against the molecular weight of standard (BioRAD), a composition of proteins of known molecular weights. These were plotted in Excel as molecular weight in kDa *vs.* K_{AV} parameter and the regression line was plotted. The molecular weights of interrogated complexes were then calculated using the regression line equation.

2.5 NMR methodologies

2.5.1 NMR experimental set-up

^{1}H-^{15}N HSQC was the most used NMR experiment in this work and it formed a foundation of all other experiments, including N-myc TAD backbone assignment. The experiments were collected using 120 μM of uniformly isotope-labelled N-myc TAD topped up with NMR buffer (Table 2.1) to either 500 μL total volume (5 mm NMR tubes) or to 320 μL in 3 mm Shigemi® tubes. The sample was supplemented with in-house made x100 stock of cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail tablets (Roche; Table 2.1) and 5% v/v D_2O for lock. Bubbles were removed from the sample by hand centrifuge adapted for NMR tubes.

Typical ^{1}H-^{15}N HSQC was collected using *hsqcetfpf3gpsi* pulse program from Bruker library, using 320 t_1 points and 1028 t_2 points. The sweep width was set to 11.9961 ppm in the ^{1}H dimension (indirect) and 22.0001 ppm in the ^{15}N dimension (direct). Most experiments were acquired at 283 K (10 °C). Tuning, shimming, locking and manual optimisation of 90° pulses was done utilising Bruker TopSpin v3.1 software package. The number of scans was typically set to 4 and was adjusted accordingly, depending on the signal strength. All experiments were collected using States-TPPI acquisition mode. NMR data were processed using NMRpipe (Delaglio et al., 1995) and analysed with CCPNmr Analysis version 2.5.0 (Vranken et al., 2005).

The assignment was performed using ^{1}H-detected 3D spectra (HNCO, HNcaCO, HNCA, HNcoCA, HNCACB and HNcoCACB) which were recorded on Bruker 750 MH Oxford NMR magnet equipped with a 5 mm TCI cryoprobe, with ^{1}H-^{15}N HSQC collected following every experiment to monitor sample changes. The assignment of N-myc$_{64\text{-}137}$ was done using the same experimental set-up, however, the data were collected using non-uniform sampling (NUS) with 25% sampling regiment. NUS data were processed using NMRbox platform using NMRPipe (Maciejewski et al., 2017; Delaglio et al., 1995). Data was reconstructed using NMRDraw using following functions: solvent filter SO, cosine bell, zero fill and Fourier Transform. Phasing was adjusted for each dimension separately and *XY*, *XZ* and *YZ* projections were checked. NUS reconstruction was performed using NMRpipe (Delaglio et al., 1995). ^{15}N- and ^{13}C-detected experiments (CON, CaN, CaCO, CanCO) were collected on a 950 MHz Bruker Ascend Aeon^{TM} NMR spectrometer equipped with a 5 mm TXO cryoprobe, collected at 10 °C, with 768, 72 and 72 real and imaginary points, respectively. The partial assignment was transferred from ^{1}H-^{15}N HSQC of N-myc TAD and HNCO, HNCA and HNcoCA experiments using CCPNmr Analysis version 2.5.0 (Vranken et al., 2005).

NOESY experiments (NOESY ^{15}N-HSQC and ^{15}N HSQC NOESY ^{15}N HSQC) were collected on Bruker 750 MH Oxford NMR magnet equipped with a 5 mm TCI cryoprobe. Similarly, to ^{1}H-^{15}N HSQC, N-myc TAD concentration was kept at 120 μM, with 5% D_2O used for lock purposes and in-house made x100 stock of cOmplete^{TM} Mini EDTA-free Protease Inhibitor Cocktail tablets (Roche) to minimise proteolytic degradation. The spectra were recorded at 10 °C, processed using NMRpipe (Delaglio et al., 1995) and analysed with CCPNmr Analysis version 2.5.0 (Vranken et al., 2005).

The Z-Z exchange experiment (hsqcexf3gpwgph pulse modified by Theo Karamanos) was collected on Bruker 750 MH Oxford NMR magnet equipped with a 5 mm TCI cryoprobe using 300 μM N-myc$_{64\text{-}137}$ labelled with ^{15}N. The mixing time was set to 0.5 ms. The experiment was collected at 10 °C. Tuning, shimming and locking was performed as described above. The data was processed using Bruker TopSpin v3.1 software package.

2.5.2 Secondary chemical shifts

$^{13}C\alpha$ secondary chemical shifts is defined as a difference between experimental and theoretical chemical shifts for random coil (Schwarzinger et al., 2001). The experimental chemical shifts were obtained from Ccp Nmr Analysis package v.2.5.0 (Vranken et al., 2005) following NMR backbone assignment. The reference random coil values for $^{13}C\alpha$ N-myc protein sequence were calculated using ‘Random Coil Chemical Shift calculator for IDPs’ web server (Department of Biology, n.d.) and were adjusted for 10 °C and pH of 6.9. $^{13}C\alpha$ secondary chemical shift was calculated using Excel software.

2.5.3 Prediction of an overall structure

CS-Rosetta web server was utilised to predict the overall structure o N-myc TAD, using chemical shift of $^{1}H^{N}$, ^{15}N, ^{13}CO, $^{13}C\alpha$ and $^{13}C\beta$ from the backbone assignment (Shen et al., 2008). The files were loaded onto the CS-Rosetta web server in STAR format (Shen et al., 2008). Overall, there were 3000 structures generated, with low convergence scores, indicating low confidence of the accuracy of the structures. The best model was selected based on its lowest root mean squared deviation (RMSD) *vs.* energy.

2.5.4 Relaxation experimental set-up

The R_1, R_2 and ^{1}H-^{15}N-nOe were collected on 600 MHz Oxford Instruments, equipped with QCI-P-cryoprobe (5mm). The data were collected at 20 °C with N-myc TAD final concentration of 150 μM. In all experiments, the pseudo-3D pulses were utilised with flip-back pulses for water suppression, shorter CPMG cycle and interleaved FIDs to account for sample changes. ^{1}H-^{15}N-nOe was collected using *15N*_HETNOE customised programme from Bruker library. Spectral widths were set to 11.9777 ppm and 22.0 ppm in ^{1}H and ^{15}N dimensions respectively. The number of scans was set to 32, preceded by 64 dummy scans. The recycle delay was set for 5 sec to allow the magnetisation to return to thermal equilibrium before subsequent pulsing.

T_1 and T_2 experiments were run using 15N_T1_HSQC and 15N_T2_HSQC pulse program, respectively, using 2048, 320 and 11 points set in F_3, F_2 and F_1 dimensions. The number of scans was set to 8 for both experiments, with 32 dummy scans for the collection of T_1 and 64 dummy scans for the collection of T_2. Ten delay times were selected from the set of 0, 9, 19, 19, 29, 39, 59, 59, 79, 119 and 159 ms and 2, 4, 6, 6, 8 10, 12, 14, 14, 16, 20 and 24 ms for T_1 and T_2 experiments, respectively. The recycle delay for both experiments was set for 2.5 s.

The data was processed using NMR Pipe (Delaglio et al., 1995) and analysed with the Ccp Nmr Analysis package (v2.5; Vranken et al., 2005). The backbone assignment of N-myc TAD was transferred to ^{1}H-^{15}N-nOe, T_1 and T_2 spectra, with overlapping peaks and ambiguous peaks omitted in the data analysis. PINT v 2.1.0 was used to calculate R_1 and R_2 relaxation rates by integrating peak volumes and fitting the exponential decay rate functions (Ahlner et al., 2013). Standard errors were calculated using PINT software and the graphs were plotted using Excel.

2.5.5 Phosphorylation studies

NMR was selected as the main method of studying N-myc TAD phosphorylation, where ^{1}H-^{15}N HSQC are collected prior- and post- addition of kinases (Selenko et al., 2008). Data was acquired on 600 MHz Oxford NMR spectrometer, equipped with QCI-P-cryoprobe (5 mm). ^{15}N-labelled N-myc TAD was added to a 5 mm standard NMR tube at a final concentration of ~120 μM. The reaction was supplemented with 5% v/v D_2O, in-house made x1 cOmplete^{TM} Mini EDTA-free Protease Inhibitor Cocktail tablets (Roche; Table 2.1), 4 mM $MgCl_2$ and 1 mM ATP and topped up to 520 μL final volume with NMR buffer (Table 2.1). NMR parameters were set up manually as described in chapter 2.5.1 and the reference spectrum was collected. 0.3 μM final concentration of kinase was added to the reaction tube, spun briefly and the sample was returned to the spectrometer and subsequent ^{1}H-^{15}N HSQCs were acquired. The data was processed using NMR Pipe (Delaglio et al., 1995) and the peaks were integrated with CCPNmr Analysis version 2.5.0 (Vranken et al., 2005). Chemical shift perturbations (CSPs) were calculated using the formula:

$$CSP = \sqrt{(1H\delta_{REF} - 1\mathrm{H}\delta_{OBS})^2 + ((15N\delta_{REF} \ - \ 15N\delta_{OBS})/6.5)^{2\frac{1}{2}}},$$

where δ_{REF} and δ_{OBS} is chemical shift of the reference and altered frequencies. The graphs were analysed and plotted using Excel, with error bars signifying standard deviation (SD), calculated from three biological repeats.

3 N-myc TAD backbone assignment

3.1 Introduction

The unique benefits of NMR stem from its ability to provide a plethora of structural and dynamic data collected in-solution at residue and sub-residue resolution. However, a meaningful interpretation of that data requires an assignment of the peaks in the NMR spectra. In particular, peaks observed in the ^{1}H-^{15}N HSQC spectrum are assigned to specific residues within the amino acid sequence (Figure 3.1). This is achieved by collecting a series of ^{1}H-detected 3D spectra, which correlate the HN peak to ^{13}C', ^{13}Cα and ^{13}Cβ resonances for *i* and *i* –1 residues, thus allowing us to sequentially connect peaks from neighbouring residues. Additionally, some ^{13}Cα and ^{13}Cβ chemical shifts are specific to the amino acid type, and thereby allow the identification of specific residues within a sequence. For example, alanine is the only residue for which the ^{13}Cβ resonates at ~18 ppm, making it an easily identifiable residue in ^{13}Cβ-detected experiments.

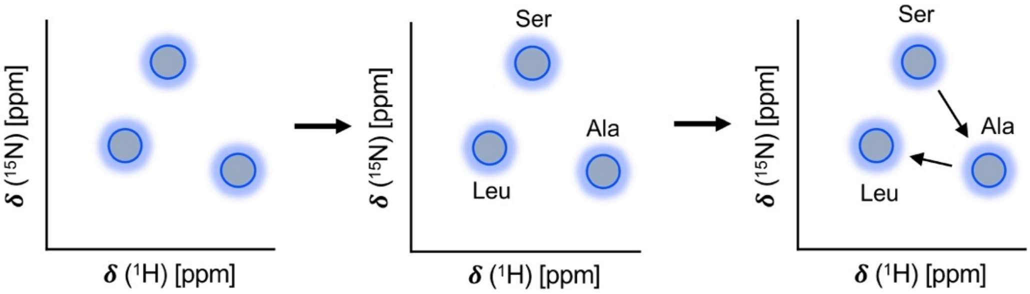


Figure 3.1. The process of NMR assignment.
Peaks of unknown amino acid identity and unknown sequential order (left) are analysed for their ^{13}C', ^{13}Cα and ^{13}Cβ chemical shifts to establish their possible amino acid type (middle panel). Looking at correlated resonances between *i* and *i –1* peaks allows the sequential order of these residues to be established (right).

So far, NMR data available on myc paralogues encompasses the assignment of c-Myc$_{1\text{-}88}$ in its *apo* state and when interacting with Bin (Andresen et al., 2012). Whereas this region contains MB0 and MBI, which are conserved across myc paralogues, there are only around 20 residues within c-Myc$_{1\text{-}88}$ that exactly match the sequence of N-myc, therefore it was impossible to utilise that assignment beyond a few short stretches of residues. In this study, several approaches were taken to achieve the full assignment of N-myc TAD. A suite of standard ^{1}H-detected 3D resonance experiments was collected (HNCO, HNcaCO, HNcoCA, HNCA, HNcocaCB and HNcaCB; (Grzesiek and Bax, 1992). Note, the lower-case nomenclature signifies spins where magnetisation is evolved but not detected. This approach allowed the assignment of ~60% of the entire protein. However, due to the disordered nature of

N-myc TAD, collected spectra were characterised by peak overlap and crowding. There were also instances of peaks characterised by broad linewidths, which is in contrast to sharp and narrow peaks that usually characterise IDPs. The region particularly afflicted by this was the region of documented interaction with Aurora A within S^{76}-W^{88} (Richards et al., 2016). Peak broadening within IDPs is a characteristic of the regions which experience a degree of structure or undergo an intermediate chemical exchange due to some other processes. This will be explored further in Chapter 3.

IDPs are characterised by sensitivity to proteolytic degradation. To shorten the data acquisition, all experiments had BEST pulse sequence. Poor dispersion in the ^{1}H dimension was addressed by also collecting ^{13}C- or ^{15}N- detected experiments. These experiments also allowed resonances for all 17 proline residues present within the sequence to be assigned, which due to their lack of an amide proton, are undetectable through ^{1}H-detected spectra. Due to the failure of the correlation of sequential resonances through *J*-couplings, through-space dipole-dipole interactions were utilised as means to aid the backbone assignment. Lastly, peak crowding had to be addressed by reducing the complexity of spectra through protein truncation. The suite of ^{1}H-detected 3D resonance experiments was then repeated on the truncated N-myc TAD (N-myc$_{64\text{-}137}$), but this time the data was collected through non-uniform sampling (NUS) which aims to improve both the sensitivity and the resolution of experiments.

This chapter describes in detail each of the approaches taken, the rationale behind them, and the outcome. The challenge in the assignment of the N-myc TAD is reflected in the number of different approaches that were taken, that went beyond the routine methods used for assigning an equivalent-sized globular protein.

3.2 Results

3.2.1 Generating N-myc recombinant proteins

A significant amount of time was dedicated to optimising the expression system for N-myc TAD and included screening of various expression systems, temperature, media and solubility and affinity tags.

Multiple *E. coli* and insect cell constructs were made, some of which were test-expressed and some of which were scaled up for protein production. The pET-3d system was the vector of choice for bacterial expression due to the ease of cloning and high copy number (Agilent Technologies, n.a.). Bacterial constructs were expressed in Rosetta^{TM} (DE3) Competent cells (Novagen) for the purposes of protein expression, however other strains were also used, including ArcticExpress (DE3) and SoluBL21TM Competent cells. Bacteria cultures were grown in LB media with either 25 ºC or 18 ºC induction temperature. Alternatively, auto-induction media (AIM) was used, which does not require constant monitoring of O.D. AIM contains glucose and lactose, with the former being utilised first in the initial growth phase and then the metabolism is switched to lactose, which causes the induction of transcription of T7 polymerase and subsequent recombinant protein expression (Rosano and Ceccarelli, 2014).

The baculovirus–insect cell expression system is a eukaryotic system that utilises cells extracted from moth species *Spodoptera frugiperda* cells (SF9; Smith et al., 1983). As eukaryotic cells, their protein translation and post-translational machinery is much more akin to mammalian than bacterial systems and can be used to stabilise recombinant proteins with chaperones or through PTMs. However, this system was labour-intensive with little enhancement to recombinant N-myc TAD stability or quantity produced (data not shown).

Various N-myc protein lengths were tested in bacterial expression systems, based on the structural data available for myc paralogues. The fragment N-myc$_{28\text{-}89}$ has been crystallised with Aurora A, however, the only region visible in the crystal structure is N-myc$_{61\text{-}89}$ (Richards et al., 2016). Both 28-89 and 61-89 fragments were used in the expression studies, with the rationale that these would be more stable constructs for recombinant protein expression, than the full-length TAD fragment. N-myc$_{1\text{-}89}$ was also tested in protein expression trials, as it is the fragment equivalent to c-myc$_{1\text{-}88}$, which has previously been characterised by NMR (Andresen et al., 2012; Helander et al., 2015). All of the truncated TAD fragments were sensitive to proteolytic degradation.

Solubility and affinity protein tags are useful tools that can aid expression and purification and limit proteolysis at the N- and C-terminus. Streptavidin-binding protein (SBP), glutathione-binding protein (Gst) and 6x His were tested, with tags added either at the N- or C- terminus. Apart from their usefulness in the purification process, they did very little to reduce the level of proteolytic degradation or enhance the amount of protein made. N-myc TAD expression was also tested using a PelB sequence that directs protein into the periplasmic space which is devoid of intracellular proteases, however this approach did not yield any detectable protein (Yoon et al., 2010).

None of the conditions tested were successful to produce the soluble protein of sufficient quantities and purity. N-myc, regardless of the expression system, length of construct or solubility tag used, was either insoluble or expressed poorly, in addition to substantial degradation during purification. In the end, pETM6T1 6x His-NusA-TEV-N-myc$_{1\text{-}137}$ was used for large-scale expression in BL21-CodonPlus (DE3) RIL Chemically Competent cells (Agilent; courtesy of Mark Richards) and purification was performed according to a protocol from Eoin Leen. The yield of the construct was modest and averaged around 0.5 mg per 1 L of bacterial cultures, however with multiple rounds of purification it was sufficient to generate enough protein for biochemical characterisation

3.2.2 Selection of NMR conditions

Apart from interaction studies, most NMR experiments were performed at 10 °C. Temperature titrations indicated that higher temperatures lead to a decrease in peak intensities and peak broadening, most likely due to the enhancement of amide proton exchange with bulk water (Figure 3.2; Uversky, 2009). At higher temperatures the rate of this exchange increases, leading to the greater signal loss (Uversky, 2009).

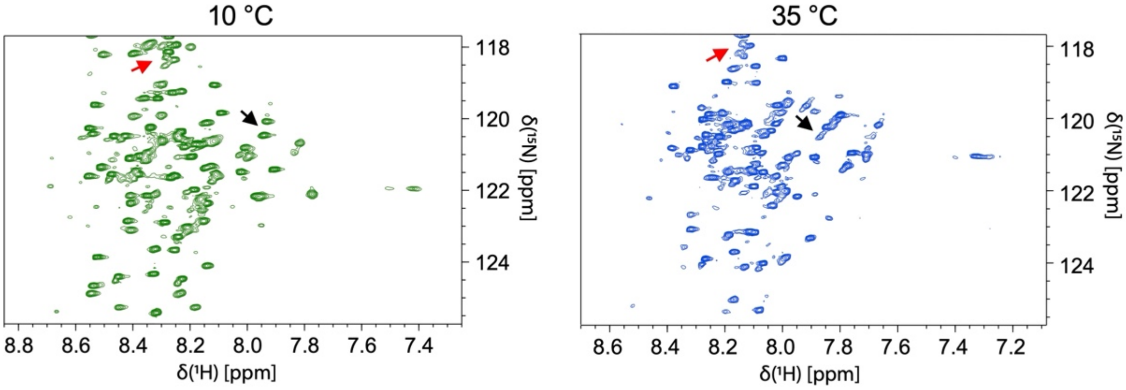


Figure 3.2. ^{1}H-^{15}N HSQC spectra of N-myc TAD.
^{1}H-^{15}N HSQC spectrum at 10 °C (green) and 35 °C (blue). The data was collected on a 600 MHz actively shielded Oxford Instruments NMR magnet equipped with a QCI-P-cryoprobe (5 mm). Only the most crowded central part of the spectra are shown. Increase in temperature causes loss of peak intensities (*e.g.* red arrows) and peak broadening (*e.g.* black arrows). Arrows of the same colour indicate the peaks of the same residues on the 10 °C and 35 °C spectra.

For the purposes of most experiments, a simple-composition buffer was selected, containing a reducing agent to prevent aggregation via cysteine residues and Tris-buffer saline (TBS) with pH adjusted to 6.9. Tris possesses an NH_2 group, which can become protonated depending on pH, and gave ^1H resonances at 3.5 ppm. However this did not significantly impact the ^1H-detected spectra collected since they appear upfield from backbone amide proton resonant frequencies of interest. The pH (6.9) was selected as a mid-point of what is considered physiological pH, ranging between 6.5 and 7.5. Interaction studies included binding partners which tend to be more sensitive to pH due to their globular nature, thus it restricted use of lower pH. However, the pH was still low enough to limit solvent exchange of backbone amide protons and thus avoid loss in signal intensity (Kukic et al., 2013).

3.2.3 NMR backbone assignment

3.2.3.1 ^1H-detected triple resonance experiments

The initial approach in the assignment of N-myc TAD was the collection of ^1H-detected 3D experiments: HNCO, HNcaCO, HNcoCA, HNCA, HNcaCB, HNcoCACB. These utilise through-bond *J* couplings for magnetisation transfer and correlate resonances of *i* and *i –1* residues which allows sequential linking to be established (Kay et al., 1990). They were selected as a first choice for a few reasons. Each experiment correlates the resonances of three different nuclei (thus three dimensions), which allows ^1H–^{15}N correlation peaks that overlap to be resolved using one of three ^{13}C frequencies. ^1H is used for both magnetisation and detection which provides high sensitivity due to its high ^1H γ ratio. After each ^1H-detected 3D experiment a ^1H-^{15}N HSQC spectrum was recorded to monitor the sample integrity, with the initial ^1H-^{15}N HSQC used as a reference spectrum.

The HNCO and HNcaCO pair of experiments is used to assign the backbone carbonyl carbon (C’) and is one of the most commonly utilised tools in protein backbone assignment. In the HNCO experiment, ^1H_i and $^{15}N_i$ are correlated to the $^{13}C'_{i-1}$ resonance through *J*-couplings (Figure 3.3 A; Kay et al., 1990). HNcaCO experiment correlates H_i, N_i and C'_i, however it also correlates H_i, N_i and C'_{i-1}, which overlaps with the peak generated by HNCO (^1H_i,$^{15}N_i$, $^{13}C'_i$; Figure 3.3 B). This provides a means to link C’ resonances of sequential residues (Kay et al., 1990; Clubb et al., 1992). The two experiments that correlate ^1H and ^{15}N to $^{13}C\alpha$ are HNCA and HNcoCA. Both HNcoCA and HNCA experiments provide inter-residual correlations of ^1H_i and $^{15}N_i$ to the preceding $^{13}C\alpha_{i-1}$ (Figure 3.3 C and D; Kay et al., 1990; Grzesiek and Bax, 1992;

Farmer et al., 1992). In the HNCA experiment there is also an additional, stronger intra-residual correlation between $^{1}H_i$ and $^{15}N_i$ to $^{13}C\alpha_i$ (Figure 3.3 D; Bax and Ikura, 1991). The final set of experiments correlates $^{13}C\beta$ with ^{1}H and ^{15}N. In the HNcocaCB experiment $^{1}H_i$ and $^{15}N_i$ are correlated with $^{13}C\beta_{i-1}$ (Figure 3.3 E; Grzesiek and Bax, 1992) . To provide a sequential link, this experiment is complemented with HNcaCB, where again two peaks are often observed, one correlating $^{1}H_i$ and $^{15}N_i$ with $^{13}C\beta_{i-1}$ thus overlapping with the same peak in the HNcocaCB (Figure 3.3 E). The additional intra-residual link is characterised by correlated $^{1}H_i$, $^{15}N_i$ and $^{13}C\beta_i$ chemical shifts (Figure 3.3 F; Grzesiek and Bax, 1992). Each two-experiment set described above allows for the establishment of sequential links by matching ^{13}C resonances and using the intra-residue correlations. Identification of amino acid 'marker' residues in the sequence is achieved using $^{13}C\alpha$ and $^{13}C\beta$ resonances that occur at specific frequencies.

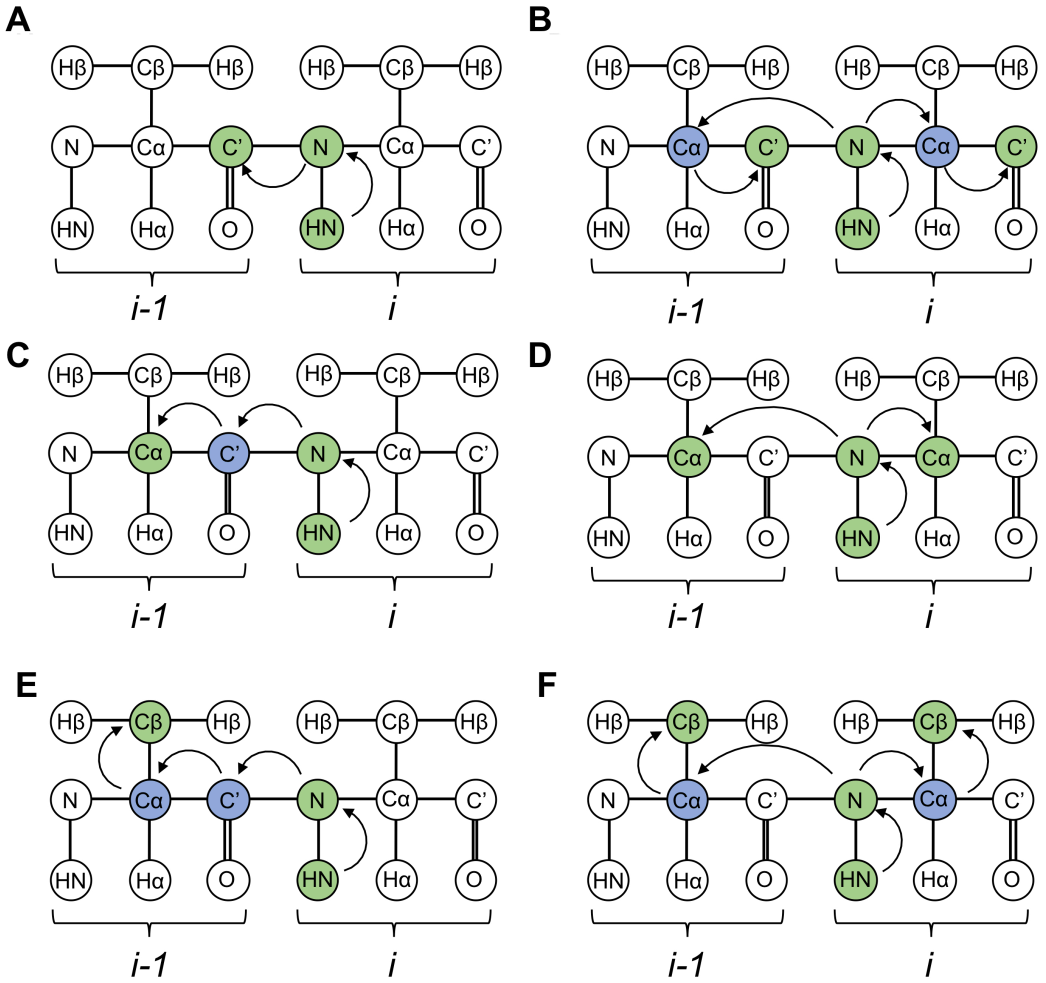


Figure 3.3. Schematic representation of magnetisation transfer in ^{1}H-detected 3D experiments.
A. In the HNCO experiment, magnetisation is transferred from $^1H^N$ onto the ^{15}N and then onto ^{13}C’ on the *i* –1 residue, resulting in a single peak. **B.** In the HNcaCO experiment two peaks are generated as magnetisation from $^1H^N$ is transferred onto ^{15}N and onto two neighbouring $^{13}C\alpha$ from *i* and *i* –1 residues. **C.** In the HNcoCA experiment magnetisation is transferred from $^1H^N$ to ^{15}N and then onto neighbouring $^{13}C\alpha$ of the preceding residue. **D.** In the HNCA magnetisation is transferred from $^1H^N$ onto the ^{15}N and then onto neighbouring $^{13}C\alpha$, resulting in the generation of two peaks, with correlations for $^1H^N_i$, $^{15}N_i$ and $^{13}C\alpha_{i-1}$ and $^1H^N_i$, $^{15}N_i$ and $^{13}C\alpha_i$ resonant frequencies. **E.** In the HNcocaCB experiment magnetisation is transferred from $^1H^N$ to N and then onto the preceding $^{13}C\alpha$ and then onto $^{13}C\beta$ of that residue, generating a single peak with resonant frequencies for $^1H^N_i$, $^{15}N_i$ and $^{13}C\beta_{i-1}$. **F.** In the HNcaCB, magnetisation is transferred from $^1H^N$ onto ^{15}N and onto two neighbouring $^{13}C\alpha$ residues but chemical shift is not developed. Then magnetisation is transferred onto $^{13}C\beta$, resulting in two peaks with resonant frequencies of $^1H^N_i$, $^{15}N_i$ and $^{13}C\beta_{i-1}$ and $^1H^N_i$, $^{15}N_i$ and $^{13}C\beta_i$. Nuclei where chemical shifts are developed are highlighted in green and nuclei where chemical shifts are transferred but not recorded are highlighted in blue. Adapted from (Higman, 2012).

The N-Myc TAD displays a typical IDP ^{1}H-^{15}N HSQC spectrum, characterised by narrow dispersion in the ^{1}H dimension and high chemical shift degeneracy (Figure 3.4; Kosol et al., 2013; Chhabra et al., 2018). Some peaks are characterised by narrow line-widths and some are characterised by broad line-widths, with varying degree of overlap, localised in the centre of the spectrum (pink rectangle; Figure 3.4). For clarity, peaks from asparagine and glutamine side chain NH_2 groups were omitted from the ^{1}H-^{15}N HSQC spectrum. These groups appear as a pair of peaks per residue, displaying the same ^{15}N resonance and different ^{1}H shifts.

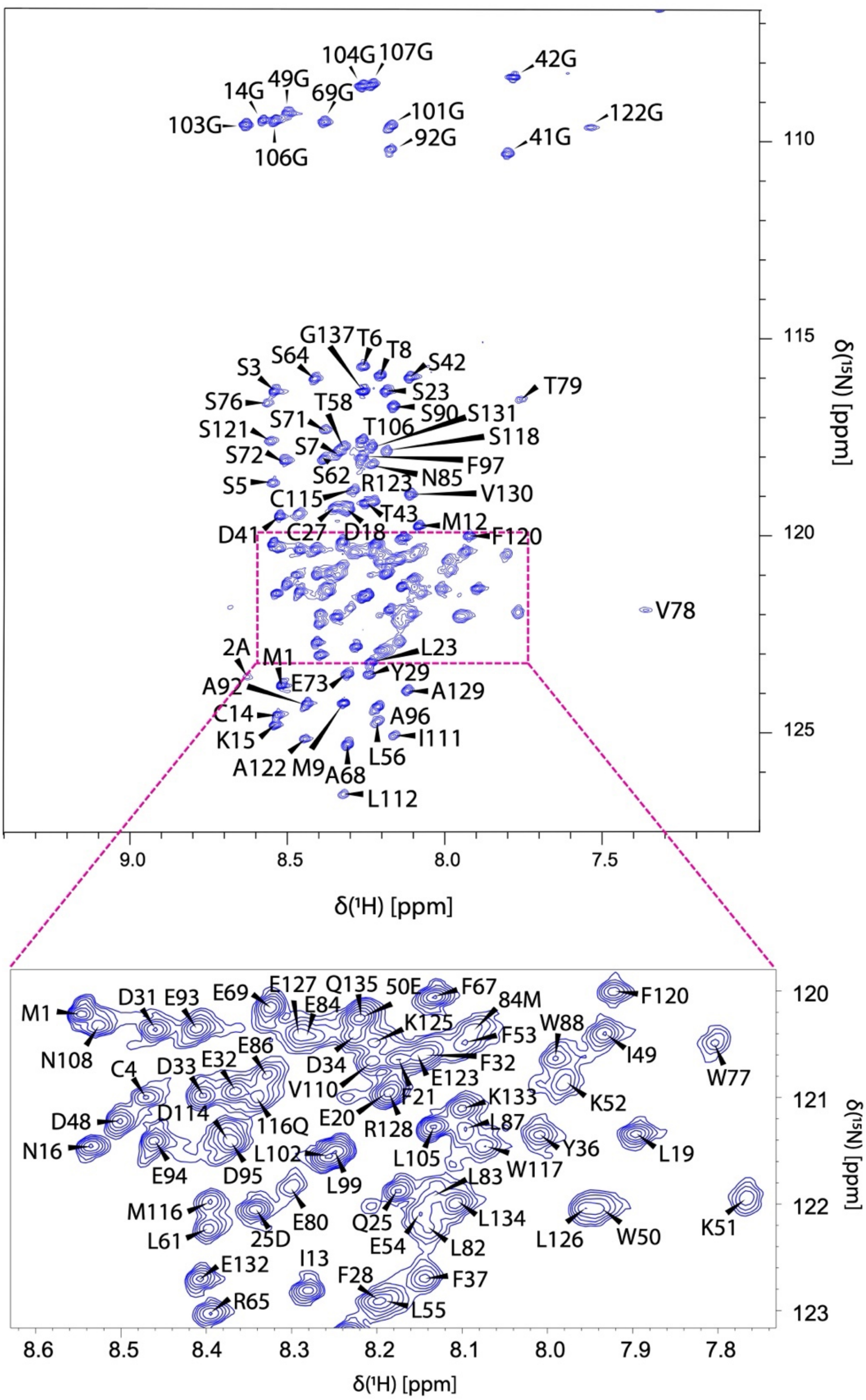


Figure 3.4. ^{1}H-^{15}N HSQC spectrum of N-myc TAD.
Top panel shows completely assigned ^{1}H-^{15}N HSQC spectrum. Arginine, glutamine and tryptophan side chain resonances were omitted for clarity. The centre of the spectrum (pink rectangle) is characterised by a high degree of peak overlap (especially resonating with high ^{1}H degeneracy). This region is magnified in the panel at the bottom.

The utilisation of ^{1}H-detected 3D experiments enabled the assignment of ~60% of the N-myc TAD sequence. A few regions however were particularly challenging to assign and had a high level of ambiguity. Regions D^{18} – F^{37} and S^{76} – W^{88} were characterised by peak broadening and peak overlap and most of them were positioned within regions of high ^{1}H degeneracy within the ^{1}H-^{15}N HSQC spectrum (Figure 3.4). Low primary amino acid sequence complexity, with multiple repeating patterns, such as TS or PT made it difficult to unambiguously identify and assign subsequent residues. There were also multiple prolines present and three PP motifs which break sequential linking, due to their lack of amide protons.

3.2.3.2 ^{13}C- and ^{15}N-detected experiments

To remedy some of the issues encountered with using the suite of standard ^{1}H-detected 3D experiments, ^{15}N- and ^{13}C- detected experiments were employed, which can offer much better dispersion of peaks. Historically, these experiments were not feasible due to the low γ ratio of these nuclei, causing poor sensitivity and low signal-to-noise ratio (Takeuchi et al., 2010; Chhabra et al., 2018). The advent of cryo-probes allowed these nuclei to be used for detection by minimising noise generated from electronic circuits within the spectrometer. ^{15}N and ^{13}C have slower transverse relaxation rates, which also enhances the intensity of the signal detected (Takeuchi et al., 2010; Chhabra et al., 2018). We collected 2D spectra using three ^{13}C-detected experiments: CON, haCACO and haCAnCO and one ^{15}N-detected experiment, haCAN (Takeuchi et al., 2010). All of these experiments correlate resonant frequencies of nuclei within or between *i*-1 and *i* residues, thus allowing multiple strings of connectivity to be established between residues. haCAN and CON experiments correlate heteronuclei by pairing the amide nitrogen with either ^{13}Cα (haCAN; $C\alpha_i$ and N_i) or ^{13}C' (CON; C_{i-1} and N_i) resonances (Figure 3.5 A and B; Takeuchi et al., 2010) . In addition, haCAN also correlates intra-residue: ^{13}Cα with the amide nitrogen of the following residue, thus providing a sequential link ($C\alpha_{i-1}$ and N_i; Figure 3.5 B).

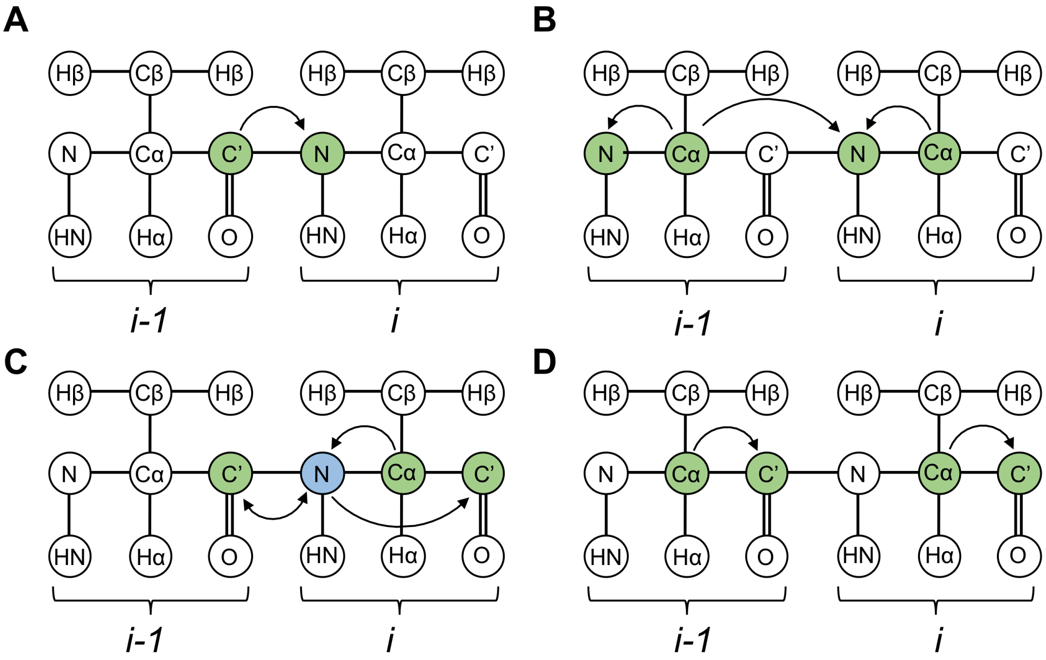


Figure 3.5. Schematic representation of ^{15}N- and ^{13}C-detected experiments. **A.** In the CON experiment magnetisation is generated on $^{13}C'_{i\text{-}1}$ and transferred onto $^{15}N_i$ providing an inter-residue link. **B.** In the haCAN experiment magnetisation is transferred from $^{13}C\alpha$ onto the amide nitrogen on the same or the following residue, providing intra- and inter-residue correlations. **C.** In the haCAnCO experiment, magnetisationis initiated on Hα (but the chemical shift is not recorded) transferred to $^{13}C\alpha$, then onto neighbouring amide nitrogen nuclei and then transferred onto ^{13}C’ of the preceding residue for the detection, providing an inter-residual link. The magnetisation is also transferred to the $^{15}N_{i+1}$ for the detection of $^{13}C'_i$. **D.** haCaCO provides inter-residual link by transferring magnetisation between ^{13}C’ and a neighbouring $^{13}C\alpha$ atom. Adapted from (Higman, 2012).

To provide extra C’–Cα resonance correlations we also collected haCAnCO and haCOCA. In the 2D haCOCA experiment, the transfer of magnetisation starts at $^{13}C'_i$ and finishes on intra-residue $C\alpha_i$ (Bertini et al., 2004). In 2D haCAnCO magnetisation is transferred between $^{13}C\alpha_i$ onto $^{15}N_i$ and then onto $^{13}C'_{i\text{-}1}$. A sequential link is also provided via a second peak that correlates $^{13}C\alpha_i$ with $^{13}C'_{i\text{-}1}$ via $^{15}N_i$ (Bertini et al., 2004; Sahu et al., 2014).

Figure 3.6 illustrates the sequential linking of D^{98} – A^{100} using all four spectra. The CON spectrum was selected as a reference, owing to its excellent resolution and sensitivity (Figure 3.6 orange). In the CON spectrum, peaks correlate the ^{13}C’ frequency of the i – 1 residue and the amide ^{15}N frequency of the i residue (Figure 3.6 orange). CON peaks share the resonant frequencies of ^{15}N nuclei for both major and minor peaks within the haCAN spectrum (green, Figure 3.6). The major peak in haCAN experiment correlates $^{15}N_i$ and $^{13}C\alpha_i$ (green, Figure 3.6). The smaller haCAN peak correlates (between residues) the resonant frequencies of $^{13}C\alpha_{i-1}$ and $^{15}N_i$ (Figure 3.6 green). CON peaks also share C' resonances with the haCAnCO spectrum, which is characterised by two peaks, one major and one minor, with resonant frequencies of $^{13}C'_i$ and $^{13}C\alpha_{i+1}$ and $^{13}C'_i$ and $^{13}C\alpha$, respectively (Figure 3.6 pink). Lastly, haCAnCO minor peak has the same resonant frequency of haCACO ($^{13}C'_i$ and $^{13}C\alpha_{i+1}$; Figure 3.6 blue).

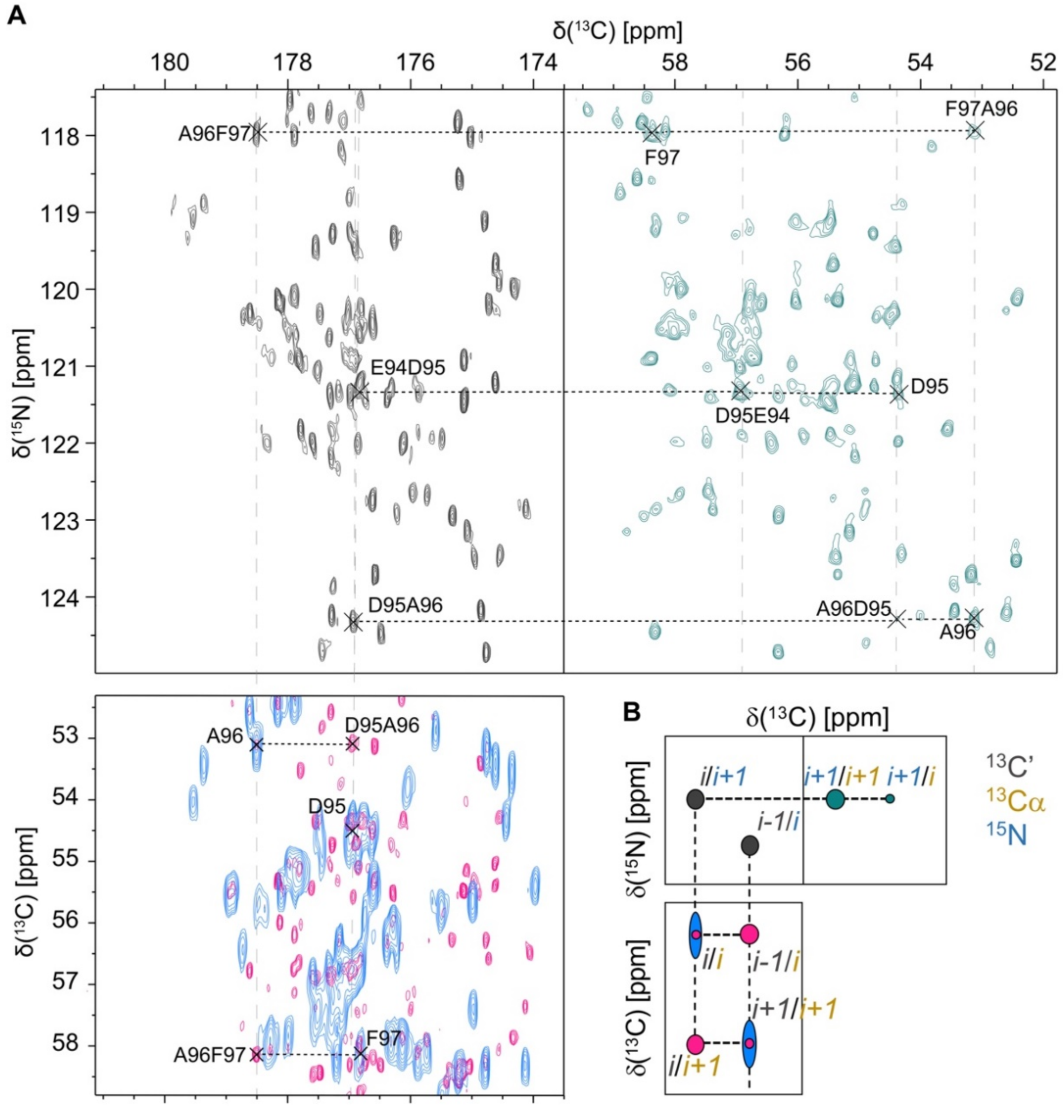


Figure 3.6. Sequential linking using CON, haCAN, haCAnCO and haCACO spectra of N-myc TAD.

CON peaks (graphite) correlate the $^{13}C'$ resonant frequency of residue $i-1$ with the amide ^{15}N resonant frequency of residue i. The haCAN spectrum (green) has two peaks, one major with resonant frequencies of $^{15}N_i$, $^{13}C\alpha_i$ and a minor peak with resonance frequencies of $^{15}N_i$, $^{13}C\alpha_{i-1}$. The CON shares a $^{13}C'$ resonance with the haCACO spectrum (blue) which exhibits only one peak with a frequency of $^{13}C\alpha_i$ and $^{13}C'_i$. This peak overlaps with one peak from the haCAnCO spectrum (pink) that has the same resonant frequencies. The other haCAnCO peak (pink) provides extra correlation of $^{13}C'_{i-1}$ with $^{13}C\alpha_i$ which can be used with the haCAN and CON spectra.

B. Schematic representation of the resonant correlation between CON (orange), haCAN (green), haCACO (blue) and haCAnCO (pink). ^{13}C' frequencies are marked in grey, ^{13}Cα frequencies are marked in yellow and ^{15}N frequencies are in blue.

The utilisation of the ^{15}N- and ^{13}C- detected experiments corroborated the assignment obtained through the collection and the analysis of ^{1}H-detected 3D experiments. In addition, the ^{15}N, ^{13}C' and ^{13}Cα resonant frequencies of all 17 proline residues were also assigned. These occupy a characteristic region of the CON spectrum, downfield in the ^{15}N dimension (132–142 ppm) and C' between 178 - 170 (Figure 3.7 B). Glycine residues (not shown) were also assigned, however as they occupy an upfield region in ^{15}N (107–111 ppm) they were omitted in Figure 3.7 for clarity.

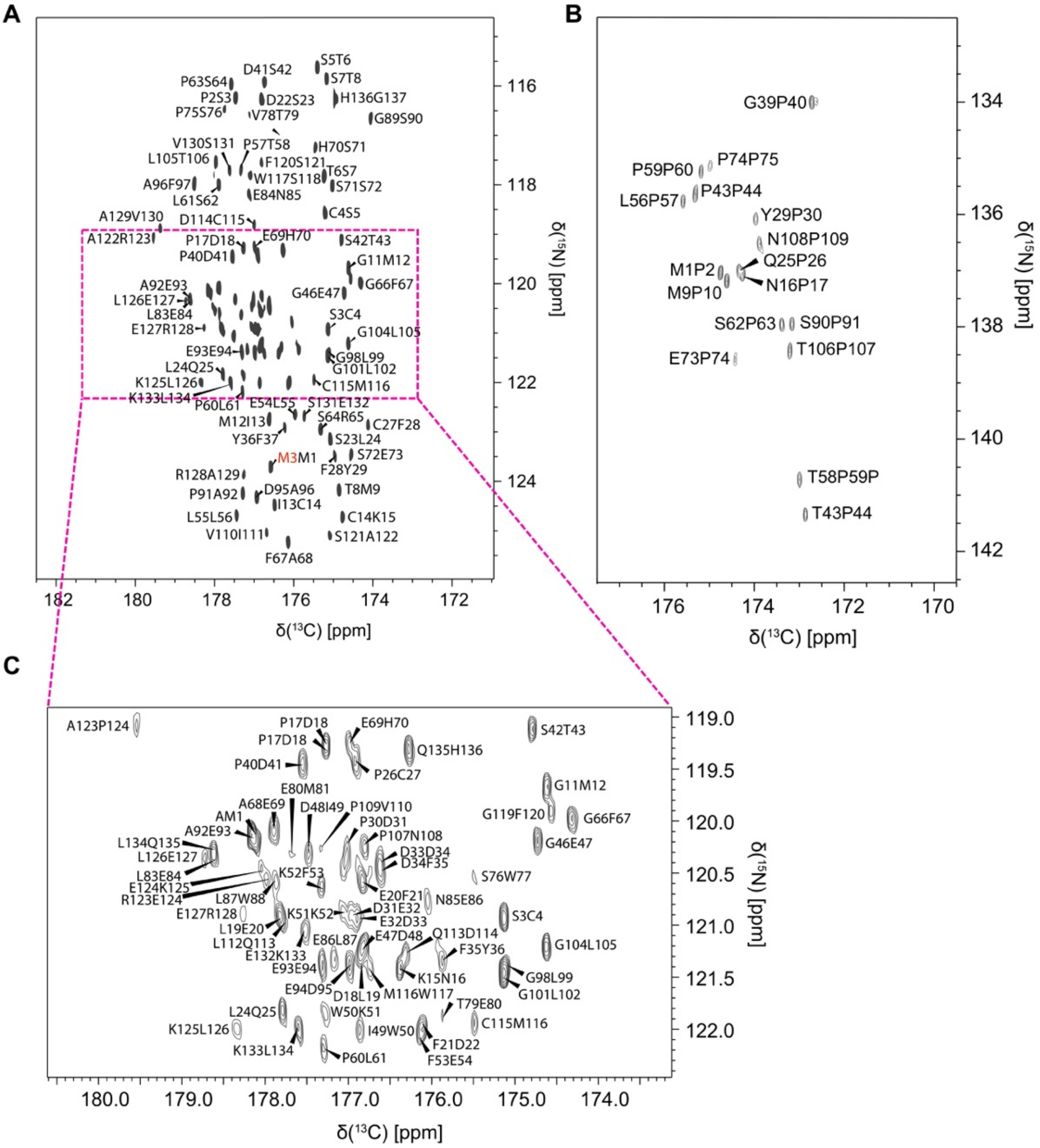


Figure 3.7. CON spectrum of N-myc TAD.
A. The main region of the CON spectrum with resonant frequencies of $^{13}C_{i-1}$ and $^{15}N_i$. Residues labelled in red text are the artefacts of cloning and the TEV-NIa protease site. The glycine region, residing upfield in the ^{15}N dimension, is omitted for clarity. **B**. Region representing peaks resonating at $^{13}C_{i-1}$ and $^{15}N^{Pro}{}_i$ frequency of prolines. **C**. Enlarged overlap region highlighted in red rectangle in A.

The drawbacks of these experiments are issues with their resolution and sensitivity. The haCACO spectrum is characterised by low resolution with broad peaks (blue, Figure 3.6). haCAN and haCAnCO spectra are both affected by sensitivity issues, especially with minor peaks which are often not seen (pink and green; Figure 3.6). Despite their enhanced peak dispersion, ^{15}N- and ^{13}C-detected experiments were still plagued by similar issues as the ^{1}H-detected 3D experiments, *i.e*. peak overlap and peak crowding. Some peaks were later shown to be characterised by low intensity (E^{80}) or were broadened beyond detection or overlapped with other peaks (L^{82} - L^{83} – M^{84}), thereby still preventing the complete assignment of this region of importance.

3.2.3.3 Collection of local constraints through NOEsy

The establishment of sequential linking based on *J*-couplings has proven challenging for some residues in the more crowded regions of the ^{1}H-^{15}N HSQC spectrum, therefore nuclear Overhauser effect (nOe) spectroscopy (NOEsy) was utilised. nOe is the effect whereby spins experience cross-relaxation through dipole-dipole interactions due to the perturbation of spins nearby with radiofrequency pulses (Overhauser, 1953). This effect has an r^6 dependence on distance and is restricted to a maximum distance of about 5 Å, thus it provides information on conformational restraints (Overhauser, 1953; Williamson, 2009; Keeler, 2004). In random coils, the distances between adjacent backbone amide ^{1}H do not exceed the length of 5 Å and thus allow the user to identify sequential residues (Zerbe and Bader, no date). The spectrum generated in a heteronuclear 3D NOESY ^{15}N HSQC experiment is characterised by the presence of diagonal peaks that correspond to the ^{1}H-^{15}N HSQC. Cross-peaks appear as a result of magnetisation transfer onto adjacent backbone amide protons (or side chain protons) and are thus can used to identify neighbouring spins.

Initially, 3D NOESY ^{15}N HSQC was collected, where homonuclear NOESY pulse sequence was applied, in which the magnetisation is transferred between all protons, and then finished with ^{1}H-^{15}N HSQC pulse element that transfers magnetisation onto neighbouring nitrogen and then onto its amide proton (Higman, 2012; Marion et al., 1989; Zuiderweg and Fesik, 1989). This results in a 3D spectrum in which cross-peaks are characterised by resonances of (^1H_i, $^{15}N_{i+1}$, $^1H_{i+1}$ or ^1H_i, $^{15}N_{i-1}$, $^1H_{i-1}$) and the diagonal represents ^{1}H-^{15}N HSQC. In the case of the N-myc TAD, the majority of amide-to-amide cross-peaks overlapped with the diagonal, or the cross-peaks were not detectable at all, thus it was impossible to resolve without ambiguities (Figure 3.8 A). To remedy this problem, ^{15}N HSQC NOESY ^{15}N HSQC was collected which utilises two ^{15}N dimensions for better peak dispersion. The cross peaks have resonant frequencies of $^{15}N_{i+1}$, $^{15}N_i$, ^1H_i. or $^{15}N_{i-1}$, $^{15}N_i$, ^1H_i. The initial ^{1}H-^{15}N HSQC pulse generates magnetisation on the amide proton with the F_1 dimension, determined by amide frequency. This is followed by a fixed mixing time in which ^{1}H-^{1}H nOes are generated through cross-relaxation. The second ^{1}H-^{15}N HSQC pulse element is then applied in which indirect frequencies are detected with F_2 as a nitrogen frequency and F_3 as a direct proton dimension (Frenkiel et al., 1990). The diagonal peak arises from magnetisation that remains on the same nucleus during the mixing time and will have a frequency of $^{15}N_i$, $^{15}N_i$, ^1H_i thereby corresponding to ^{1}H-^{15}N HSQC (Frenkiel et al., 1990; Diercks et al., 1999). As cross peaks are determined by two nitrogen frequencies, and IDPs have good dispersion in the ^{15}N dimension, this spectrum offers better resolution and eliminated some of the overlap of cross peaks with diagonal peaks (orange vs blue; Figure 3.8 A and B). With the aid of this spectrum, the resonant frequencies of residues S^{76} to W^{88} were provisionally identified. With through-space correlations, L^{82} and L^{83} were more easily distinguishable, and these were tentatively assigned (Figure 3.8 B). Nevertheless, despite the better dispersion, there were still regions that remained ambiguous, such as D^{31} – F^{38}. Most of these peaks occur in the crowded region of ^{1}H-^{15}N HSQC, with a high degree of peak overlap which made the assignment of this region challenging.

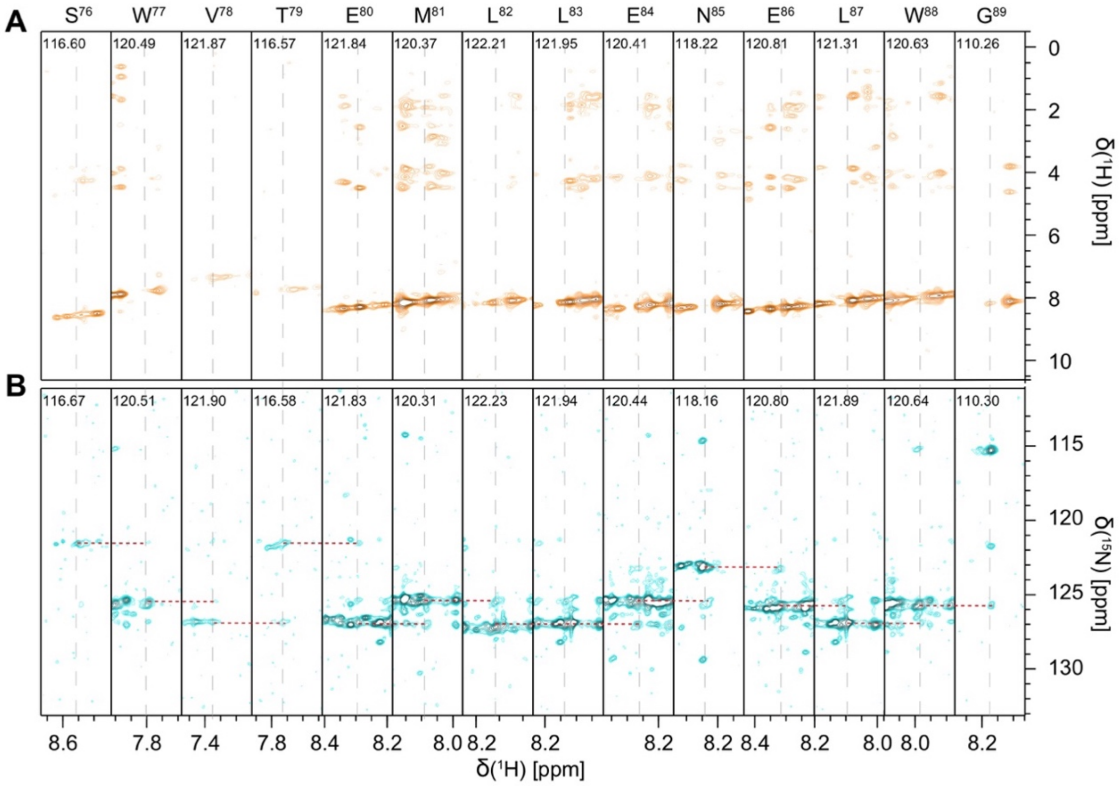


Figure 3.8. NOESY – type spectra of N-myc TAD.
A. NOESY-^{15}N HSQC of residues S^{75} - G^{89}. Strong peaks at the bottom of the spectrum lie on the diagonal and correspond to peaks in ^{1}H-^{15}N HSQC. The peaks resonating at 0–6 ppm in ^{1}H dimension are Hα and Hβ peaks of preceding residues. **B.** ^{15}N HSQC NOESY ^{15}N HSQC spectrum. The diagonal peaks are well-defined peaks at the bottom of the spectrum with resonant frequencies of ^{15}N$_{i}$, ^{15}N$_{i}$, ^{1}H, corresponding to the ^{1}H-^{15}N HSQC spectrum. Red dotted lines join cross-peaks of preceding residue with the diagonal. The number at the top is ^{15}N resonances. Both spectra were set at the same frequencies of ^{1}H$_{i}$ and ^{15}N$_{i}$, for the comparison of the peak intensities.

3.2.3.4 The assignment of N-myc$_{64\text{-}137}$

The N-myc TAD includes 137 residues, with three extra residues at the N-terminus that come as a result of construct cloning and TEV protease cleavage. This makes N-myc TAD ~17 kDa in size which for NMR purposes is a large protein with a high number of peaks in ^{1}H-^{15}N HSQC spectrum. N-myc TAD sequence is also characterised by high number of amino acids that have similar Cα and Cβ resonant frequencies, thus finding sequential links is challenging. For example, glutamine and glutamate resonate at ~57 ppm and ~29 ppm for ^{13}Cα and ^{13}Cβ, respectively. Phenylalanine and tyrosine both resonate at ~58 ppm and ~39 ppm for ^{13}Cα and ^{13}Cβ, respectively. Methionine and lysine have very similar ^{13}Cβ resonant frequencies at ~32 ppm. Leucine ^{13}Cβ nuclei resonate at ~42 ppm which is a quite distinctive resonant frequency, however, a problem arises due to the high content of this amino acid in N-myc TAD (14 residues altogether).

To reduce the number of possible matches in sequential links, N-myc TAD was truncated to N-myc$_{64\text{-}137}$. The shorter construct has more favourable dynamics with faster rotational tumbling rates (τ_c), longer transverse relaxation rates, causing a decrease in linewidths and resulting in sharper peaks (Foster et al., 2007). To minimise the problems of peak broadening and poor signal-to-noise ratio, NUS was also employed as a means of data collection. A 2D experiment is a series of 1D experiments repeated, with each repetition differing by an increase in increment time (Δt_1) in which magnetisation is transferred between heteronuclei (Figure 3.9 A). The resolution of a spectrum is positively correlated to the number of increments collected, as well as a number of scans (Delaglio et al., 2017; Barna et al., 1987). NUS omits collection of data for certain t_1 increments, however as the longest t_1 is still sampled, the resolution is preserved. In addition, more scans can be accommodated in a given instrument time, which also can positively impact the resolution (Figure 3.9 B; Barna et al., 1987). The sampling schedule can also be set in a way to sample more points at the initial stages of FID in t_1 and t_2 dimensions, where the signal-to-noise ratio is the highest (Barna et al., 1987) contributing to the sensitivity of the experiment.

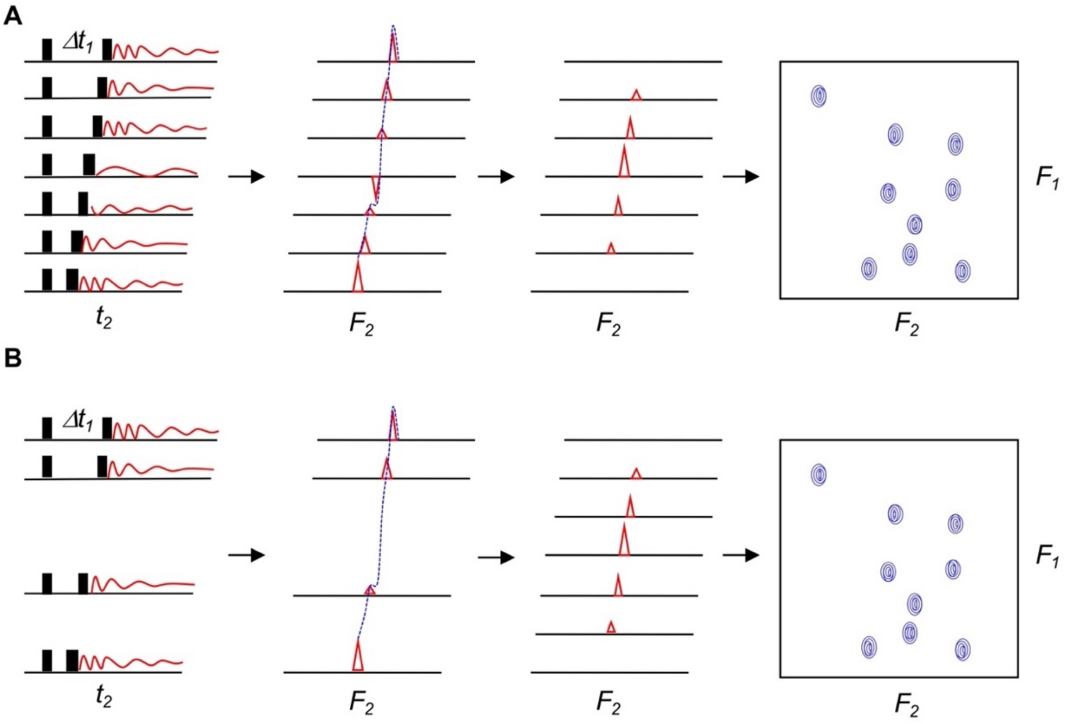


Figure 3.9. Schematic representation of different sampling approaches.
A. In uniformly sampled data every Δt_1 is collected. The resolution is dependent on the collection of the longest Δt_1. **B.** In non-uniform sampled (NUS) certain Δt_1 are

skipped, however the longest *Δt₁* is still collected to preserve the resolution. Enhanced sensitivity stems from the collection of data where signal-to-noise is the highest. In addition, if the acquisition time is shortened, this allows for an increasing number of scans in a given instrument time (Facey, 2016).

NUS sampled ^{1}H-detected 3D experiments were repeated, with uniformly sampled ^{1}H-^{15}N HSQC recorded following every ^{1}H-detected 3D experiment. A 25% sampling schedule was selected, and the data was reconstructed using NmrPipe (Delaglio et al., 1995). The combination of the shorter construct with fewer peaks, with more favourable dynamics, together with NUS data collection generated spectra of enhanced sensitivity and resolution in comparison to N-myc TAD spectra (Figure 3.10 A). This approach resolved the peaks that were overlapping in the uniformly sampled data, either by reducing the total number of peaks or by enhancing the signal-to-noise ratio and narrowing the linewidths of the existing peaks (Figure 3.10 A).

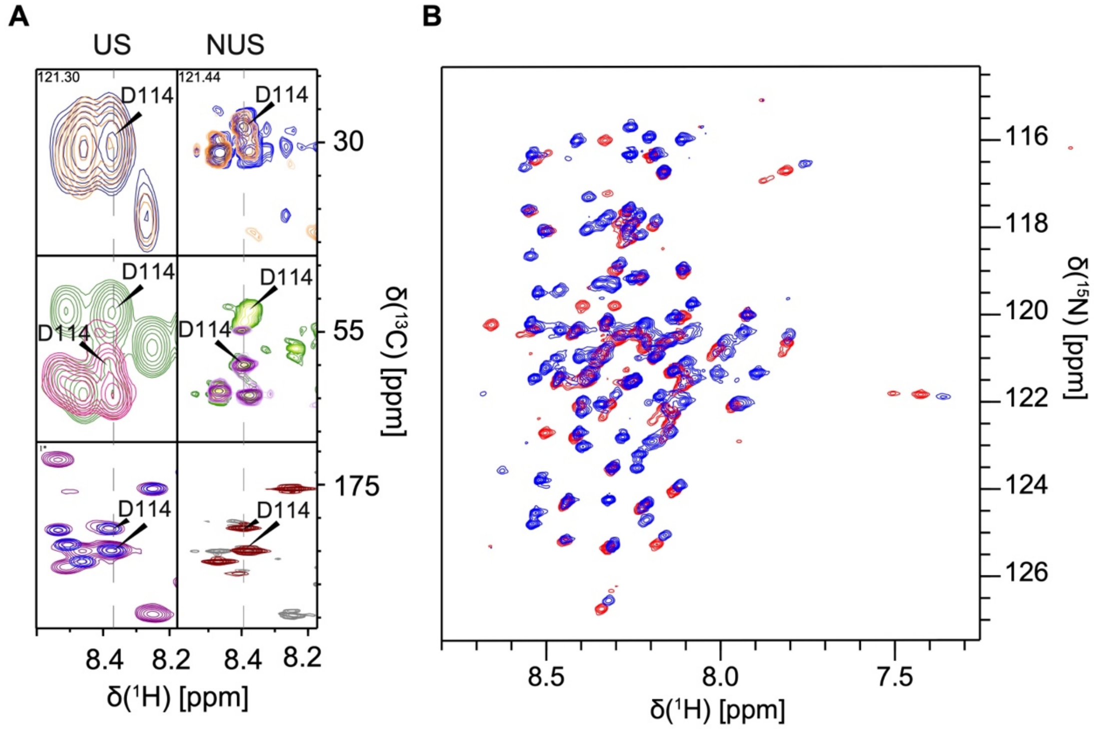


Figure 3.10. Uniform sampling *vs.* non-uniform sampling.
A. The comparison of D^{114} peaks of uniformly sampled N-myc TAD(UN; left) and non-uniformly sampled N-myc$_{64-137}$ (NUS; right) in ^{1}H-detected 3D spectra. Looking from left to right: HNCO is in blue and brown, HNcaCO is in purple and gray, HNCO is in green and lime green, HNcoCA is in pink and violet, HNcaCB is in orange and HNcocaCB is in blue. The numbers in the top squares indicate ^{15}N resonant frequencies. **B.** Overlay of ^{1}H-^{15}N HSQC of N-myc TAD (blue) and N-myc$_{64-137}$ (red). Glycine peaks, and tryptophan, asparagine and glutamine side chain peaks are omitted for clarity.

Comparison of the ^{1}H-^{15}N HSQC spectrum of N-myc TAD (blue) and the ^{1}H-^{15}N HSQC spectrum of N-myc$_{64-137}$ (red) indicated that the peaks overlap very well, apart from the first few N-terminal residues which in the shorter construct have a different sequence to the full-length TAD (Figure 3.10 B). Small uniform differences in peak resonant frequencies can be explained by using different batches of NMR buffer. Regions of increased overlap and peak broadening in N-myc$_{64\text{-}137}$ were easily assigned, and this assignment was transferred onto overlapping peaks in the full-length N-myc$_{1\text{-}137}$ construct (Figure 3.10). As the number of resonances in ^{1}H-^{15}N HSQC was reduced by half, it eliminated some ambiguities in possible sequential assignment. For example, L^{82} and L^{83} were now completely resolved as separate peaks and the assignment from NOESY was confirmed to be correct. NUS reconstruction of data can sometimes result in spectral artifacts; however careful analysis of the overlaid spectra did not reveal any outliers.

3.2.3.5 The assignment of Hα and Hβ of N-myc$_{64\text{-}137}$

Backbone assignment is usually limited to the ^{1}H^N, NH, C', Cα and Cβ. However, the assignment of ^{1}Hα and ^{1}Hβ can be later used in secondary structure propensity calculations.

The assignment of ^{1}Hα and ^{1}Hβ of N-myc$_{1\text{-}137}$ was performed using a HBHAcoNH experiment which transfers magnetisation from ^{1}Hα and ^{1}Hβ onto their respective ^{13}C. From ^{13}Cα the magnetisation is then transferred onto ^{13}C' of the proceeding residue, and then onto ^{15}N and ^{1}H^N of this residue (Figure 3.11; Grzesiek and Bax, 1992). Because the HBHAcoNH experiment does not provide a means for sequential linking, the resonant frequencies observed in HBHAcoNH have to be correlated back to the ^{1}H-^{15}N HSQC for full ^{1}Hα and ^{1}Hβ assignment. Initially, the HBHAcoNH experiment was sampled in a uniform fashion, following the collection of ^{1}H-detected 3D experiments on N-myc TAD (Grzesiek and Bax, 1992; Muhandiram and Kay, 1994). The downside of this experiment is that it takes several hours to record, resulting in protein that was degraded, apparent from observation of peak shifting, broadening and ultimately signal loss (data not shown). Observations from the N-myc$_{64\text{-}137}$ assignment indicated that NUS significantly improves the resolution and shortens the data collection time, thus minimising the effect of proteolytic degradation. A 25% sampling regimen was employed, with ^{1}H-^{15}N HSQC spectra collected prior to and after the HBHAcoNH.

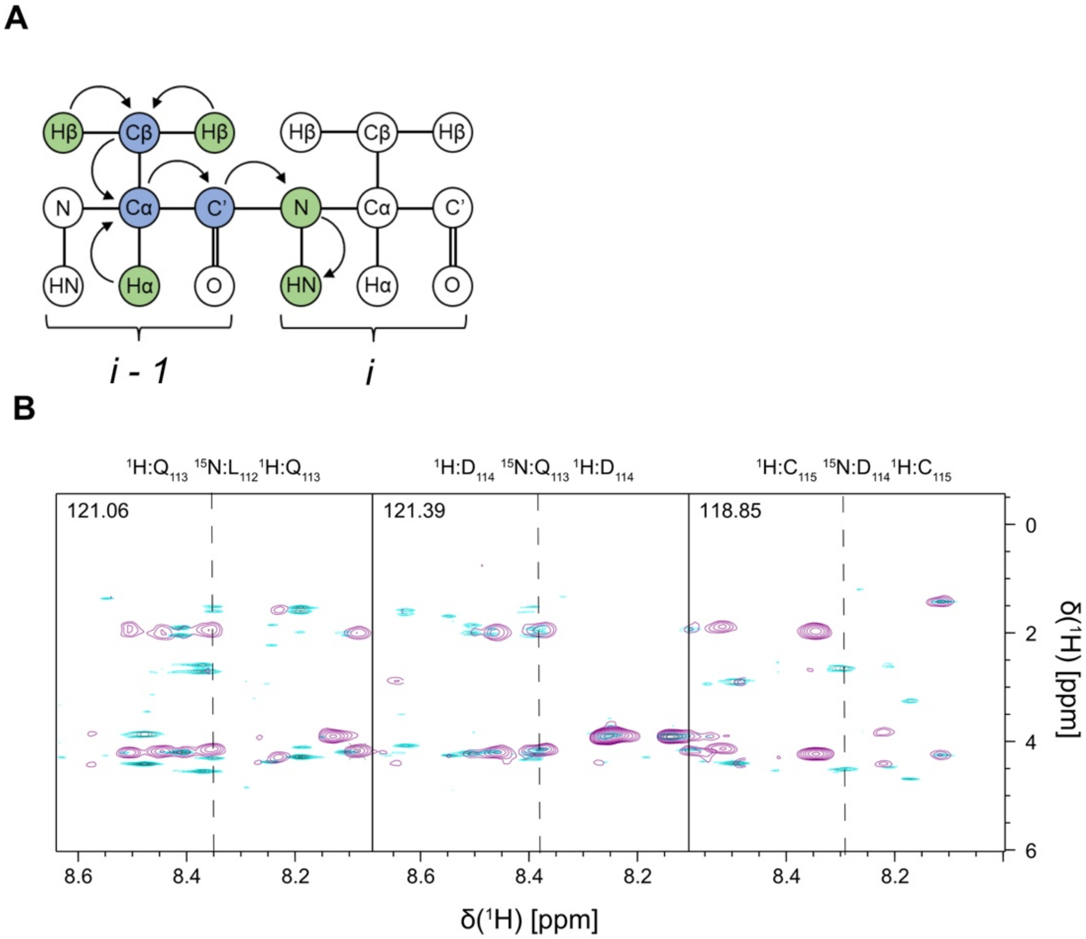


Figure 3.11. HBHAcoNH of N-myc TAD.
A. In HBHAcoNH 3D experiment the magnetisation is transferred from $^1H\alpha$ and $^1H\beta$ onto their respective carbons ($^{13}C\alpha$ and $^{13}C\beta$, respectively). Then the magnetisation is transferred onto $^{13}C'$ and then onto ^{15}N and ^1H of the following residue (Grzesiek and Bax, 1992). **B.** Amide strips of HBHAcoNH experiments showing Q^{113} – D^{114}– C^{115} assignment. Purple is uniformly sampled data and blue is NUS-sampled data. Numbers at the top left corner of each amide strip indicate ^{15}N resonant frequencies. (Higman, 2012).

The utilisation of NUS resulted in peaks with high intensity and narrow linewidths, in comparison to the uniformly sampled data (compare purple and blue, Figure 3.11 B). Using this approach all of the $^1H\alpha$ and $^1H\beta$ were assigned. Due to the differences in side chains, amino acids have characteristic numbers and chemical shift ranges for $^1H\alpha$ and $^1H\beta$ resonances, e.g. glycine has two Hα and no Hβ, whereas alanine has one Hα and three Hβ, all with the same resonant frequency as they experience the same chemical environment. All other naturally occurring amino acids are characterised by a single Hα, however, their numbers of Hβ may vary. Isoleucine and threonine have one Hβ whereas many other residue types have two. Similar to other

^{1}H-detected 3D experiments, proline ^{1}Hα and ^{1}Hβ resonances can be accessed via the following residue, apart from the N-terminal proline in PP motifs. The assignments of ^{1}Hα and ^{1}Hβ resonances granted yet another means to confirm that the assignment obtained was correct through the patterns shown based on amino acid type.

3.3 Discussion

3.3.1 c-myc$_{1-88}$ backbone assignment

The ^{1}H-^{15}N HSQC spectrum of the N-myc TAD reflects the IDP nature of this protein, with low ^{1}H dispersion, low sequence complexity and an abundance of prolines (Dyson and Wright, 2005). Usually, IDPs have fast rotational correlation times, resulting in slow R_2, which is realised in the form of sharp peaks and this was apparent in the N-myc TAD ^{1}H-^{15}N HSQC spectrum (Gibbs et al., 2017). However, there were a few regions within N-myc TAD that were characterised by broader peaks and low signal intensity (*e.g.* W^{77} – M^{84}) indicating regions of less disorder. This will be explored further in Chapter 4.

The related protein c-myc$_{1-88}$ has previously been assigned, and a comparative analysis of c-myc and N-myc ^{1}H-^{15}N HSQC spectra was performed (Andresen et al., 2012). However, due to the majority of this sequence lying outside the bounds of the conserved MBs, there were only 14 residues in total (highlighted in red, Figure 3.12 A) that are characterised by the same chemical environment and thus resonate at the same frequency (red crosses, Figure 3.12 B). These residues overlaid almost perfectly, with slight differences that could be attributed to different lengths of the constructs and different experimental conditions. Residues in MB0 were not conserved, but share similar chemical properties, thus they did not overlay on ^{1}H-^{15}N HSQC. Similar to the ^{1}H-^{15}N HSQC spectrum of N-myc TAD, the spectrum of c-myc$_{1-88}$ was also characterised by heterogeneity in the intensities of peaks, a feature that was attributed by the authors to structural heterogeneity and chemical exchange.

(Andresen et al., 2012).

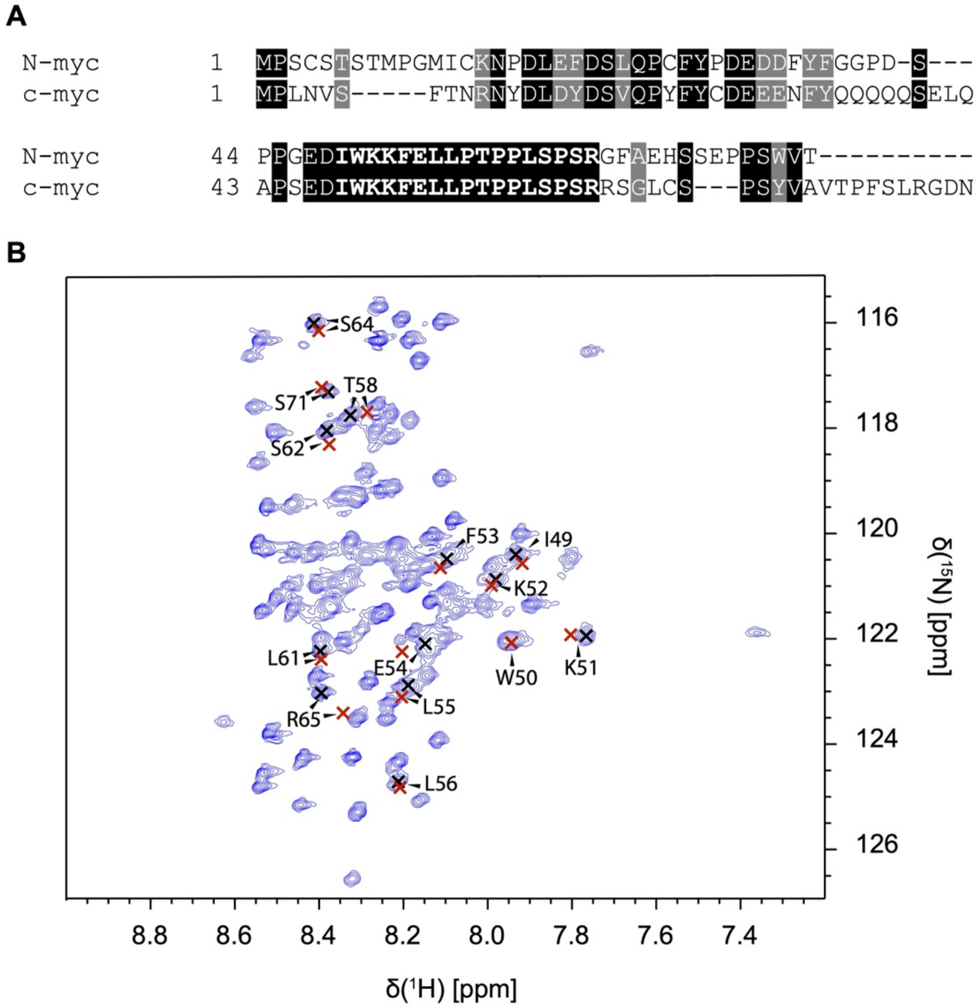


Figure 3.12. Comparison of N-myc TAD and c-myc$_{1-88}$ assignment.
A. Sequence alignment of TAD domains of N-myc$_{1-80}$ (top) and c-myc$_{1-88}$ (bottom). Residues highlighted share similar chemical properties (grey) or are conserved (black). The stretch of residues in bold font are fully conserved and their assignment was compared between the two paralogues. **B**. Overlay of the ^{1}H-^{15}N HSQC spectra of N-myc TAD (blue) and conserved residues in c-myc$_{1-88}$ (red crosses) which were highlighted in A. The black crosses are the N-myc TAD equivalent residues. The glycine region, and arginine, glutamine and tryptophan side chain were omitted for the clarity.

The assignment of c-myc$_{1-88}$ also required multiple approaches, including NUS data sampling, the utilisation of multiple ^{1}H-detectred 3D experiments (HNCA, HNCO, HBHAcoNH, CBCAcoNH) and through-space ^{15}N-NOESY-HSQC, indicating similar issues encountered during the backbone assignment. Despite its smaller size, and more favourable dynamics, the authors were not able to assign all residues (Andresen et al., 2012). They also utilised ^{15}N- and ^{13}C-correlated TOCSY and ^{13}C-NOESY-HSQC. This was complemented with high concentration of c-myc$_{1-88}$ sample (600 μM) and lower temperatures to limit amide proton exchange with bulk water.

To achieve the full assignment, many approaches were taken that utilised the latest developments in the field of NMR. There has been an advancement in the field of optimising NMR acquisition parameters through targeting the longitudinal and transverse relaxation enhancement techniques, either by adjustment of pulsing and/or data collection. In this work we utilised Band-selective Excitation Short-Transient (BEST) technique. BEST-based pulse sequences significantly shorten the acquisition times, which can be a critical parameter in proteolysis-prone proteins such as IDPs (Schanda et al., 2006). The main parameter determining the total length of the experiment is the rapidness of subsequent pulses, which is determined by the length of the time it takes for the magnetisation along the Z axis (usually 67%) before the next set of pulsing, which is determined by T_1. In IDPs T_1 is negatively impacted by the exposure of amide protons to bulk water, which causes slower magnetisation recovery not only on amide protons but also on all surrounding protons in water. BEST pulse sequences selectively target amide protons within the excited spectral window without excitation of the remaining protons, thus amide protons relax faster through cross-relaxation with aliphatic protons and water, whose magnetisation has been held at equilibrium (Schanda et al., 2006; Solyom, 2017). We decided to use this approach in our data collection to minimise the time our sample spends in the spectrometer during subsequent experiments.

The advent of cryoprobes reduced the intrinsic noise from the spectrometer's electronic circuits, making it possible to utilise nuclei with lower gyromagnetic ratios, *i.e.* ^{15}N and ^{13}C, for the purpose of detection (Takeuchi et al., 2010; Chhabra et al., 2018). Utilisation of ^{15}N- and ^{13}C- detected experiments can remedy some of the issues associated with intrinsic disorder. They offer much better dispersion in both dimensions, and thus are able to better resolve peaks that overlap in the ^{1}H dimension. Secondly, with no dependence on H–N correlations, they record resonances directly for proline residues, thus allowing this important and frequently occurring amino acid to be assigned. ^{15}N- and ^{13}C- detected experiments also allow for the data to be acquired at higher temperatures when needed, which in the case of ^{1}H-detected experiments for IDPS can cause line broadening due to the enhancement of proton exchange with bulk water (Gibbs and Kriwacki, 2018).

The greatest obstacle in ^{13}C- and ^{15}N- detected experiments is their intrinsic low sensitivity in comparison to ^{1}H-detected experiments. This can be overcome by using a more concentrated sample, however IDPs are notorious for expressing at lower yields. Use of H^N as the initial source of polarisation, has advantages in terms of enhances sensitivity due to high gyromagnetic ratio of this nucleus. The developing field of ^{13}C- and ^{15}N- detected NMR has sought to harvest this advantage and incorporated ^{1}Hα as sources of polarisation, thus providing sensitivity without signal loss (Felli and Pierattelli, 2022).

In this work, low sensitivity was particularly pronounced for the minor peaks of the haCAN experiment (^{13}Cα_{i},^{15}N$_{i+1}$). Low resolution in the haCACO spectrum was also problematic, however this is the result of having a restricted number of collected points to prevent the evolution of the Cα–Cβ coupling (Felli and Pierattelli, 2022). Faced with these challenges it was imperative that four different experiments were employed, which provided ample correlations across nuclei within and between adjacent residues. The utilisation of ^{13}C- and ^{15}N- detected experiments had a moderate contribution to the assignment of N-myc TAD, apart from the assignment of prolines.

The major impact on the full N-myc TAD assignment was the collection of ^{1}H-detected 3D experiments for a truncated fragment with NUS. A generous NUS regime of 25% was sufficient to generate high-resolution data that resolved overlapped peaks in the centre of ^{1}H dimension, without the introduction of artefacts. Reducing the number of residues through protein truncation also simplified overcrowded regions of the spectra.

3.3.2 Conclusions

Here, we present the full backbone and partial side-chain assignment of N-myc TAD. Prolines were assigned to the extent that it was experimentally possible with C', N, Cα, Cβ, Hα and Hβ assigned. The process required multiple experimental approaches, utilising both through-bond and through-space correlations. It became apparent that each of the approaches taken contributed to the progression of the assignment, however none of them alone were able to eliminate all the problems arising from the inherent nature of the IDP and only through the combination of approaches was the full assignment completed. The assignment presented many of the hallmark difficulties of investigating IDPs through NMR and may serve as an exemplar for approaching the backbone assignment of other IDPs.

4 Structural and dynamics studies of N-myc TAD

4.1 Introduction

Thus far, there has been very little structural information available regarding myc paralogues, despite the devastating prevalence of their deregulation in cancer (Kalkat et al., 2018). Most published articles interrogating the oncogenic properties of myc paralogues are *in vivo* studies of cells or animal models, and thus we lack a good understanding of the structural and functional mechanisms that N-myc protein participates in.

The collection of structural data for IDPs has been plagued with challenges. IDPs are prone to proteolytic cleavage, as they lack the hydrophobic cores that could protect sensitive sites from proteolysis (Wright and Dyson, 2009). By their very nature, they are dynamic even on short timescales, thus obtaining a crystal with an isotropic conformational arrangement is nearly impossible. To partially address these difficulties, minimal fragments of IDPs have been crystallised with larger binding partners. There are three available structures that feature an N-terminal peptide of either c-myc or N-myc in complex with a larger folded protein (Figure 4.1).

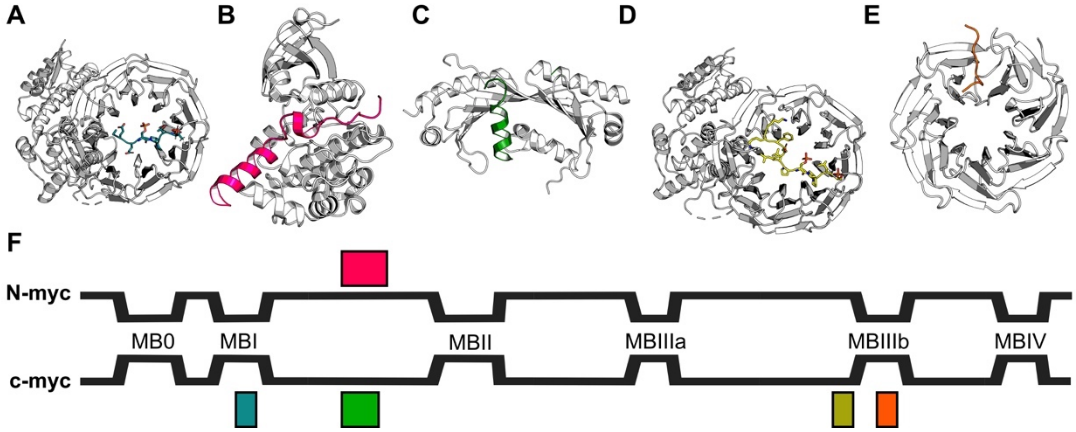


Figure 4.1. Available crystal structures of N-terminal region of myc protein
Crystal structures of **A.** c-myc$_{51\text{-}61}$, **B.** N-myc$_{61\text{-}89}$ (pink), **C.** c-myc$_{96\text{-}125}$ (green), **D.** c-myc$_{241\text{-}248}$ (yellow) and **E.** c-myc$_{260-267}$ (orange). Binding partners are grayed out. **F.** The colour-coded myc fragments were mapped onto a schematic representation of the N-terminal regions of myc proteins (black). Lines converging indicates regions of conservation between c- and N-myc (MBs) whereas non converging lines indicate lack of homology in amino acid sequence between the two proteins (Welcker et al., 2022; Richards et al., 2016; Wei et al., 2019; Thomas et al., 2015).

The first structure features the phosphodegron sequence of c-myc (residues 51-61) in complex with Fbw7-Skp1 (teal; Figure 4.1; Welcker et al., 2022). Two phosphosites, p-T^{58} and p-S^{62} forming polar contacts with Fbw7 WD40 domain (Welcker et al., 2022). The second structure (Figure 4.1 B) features N-myc$_{61-89}$ in complex with Aurora A and this structure will be described in more details below. The third structure is of a small partially-conserved segment of c-myc$_{95-111}$ (A^{95}DQLEMVTELLGGDMV^{111}) interacting with TBP and TAF1, two essential cofactors of the PIC (Figure 4.1 C; accession code: 6E16; Wei et al., 2019). c-myc interacts with these proteins in a fashion that excludes DNA binding, which raises questions regarding the biological significance of this interaction. The fourth structure (Figure 4.1 D) is that of c-myc (H^{241}EEp-TPPp-T^{248}) with Fbw7-Skp1. This is the second phosphodegron within c-myc sequence and it is positioned N-terminally of a conserved MBIIIb and it is not entirely dissimilar from the equivalent N-myc sequence (S^{254}GEDTLSDS^{262}), raising a possibility that the mechanism of regulating myc protein stability is similar between c- and N-myc and both interact with Fbw7 is a similar fashion (Welcker et al., 2022). The fifth structure features a seven-residue MBIIIa peptide (E^{260}EIDVVSV^{267}) of c-Myc binding to a shallow hydrophobic pocket of the WD40 domain of WDR5, part of multiple complexes involved in chromatin modification (Figure 4.1 E; Thomas et al., 2015). There is an additional structural study through NMR of c-myc with Pin1, however it does not provide the structure of the complex, it maps the interaction sites on c-myc sequence and provides the dynamics data of the interaction (Helander et al., 2015).

The only available crystal structure containing N-myc in the PDB is that of N-myc$_{28-89}$ with Aurora A$_{122-403}$ C290A C393A (hereafter referred to as Aurora A; Figure 4.2 A ; Richards et al., 2016). Although the N-myc construct used for crystallisation spans residues 28-89, only residues 62-89 are visible in the crystal structure, implicating that residues 18 – 61 might be intrinsically disordered. N-myc interacts with the N-lobe of Aurora A, then drapes itself on the top of the cleft between the N- and C-lobe of the kinase. There is a small helix seen from A^{68} to S^{71}, and a larger helix composed of residues S^{77} to W^{88}, which sits on top of the C-lobe of Aurora A (Figure 4.2 B). In this structure Aurora A shows hallmarks of an activated kinase, with the salt bridge between K^{162} of β3-sheet region and E^{181} within α-helix is formed which positions αC helix in an ‘in’ conformation, towards from the active site (Roskoski Jr, 2016) with the regulatory and catalytic spines assembled and the DFG motif in an ‘in’ conformation (Modi and Dunbrack, 2019; Kornev et al., 2006; Palmieri and Rastelli, 2013).

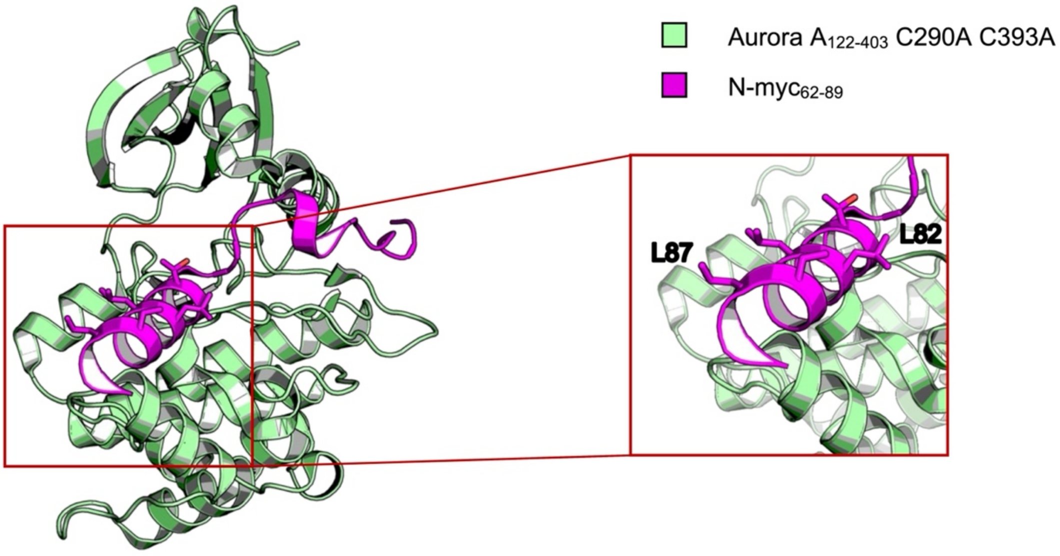


Figure 4.2. Crystal structure of N-myc$_{62-89}$ with Aurora A.
X-ray crystal structure of N-myc (magenta) with Aurora A (light green; accession code: 5G1X; Richards et al., 2016). The red rectangle indicates a close-up of residues participating in the S^{77} - $W^{88}\alpha$-helix with their side chains of the residues that point away from the Aurora A surface.

This ‘divide-and-conquer’ approach of obtaining structural data using small fragments of myc protein does provide some structural insight into myc paralogues, however this ensemble-averaged view fails to address the conformational range of myc in the context of meaningful length of protein fragment. NMR is the only technique geared towards studying conformational ensembles of IDPs, without the necessity for a conformationally homogenous protein sample, and provides in-solution residue and sub-residue resolution.

This chapter covers the work associated with establishing the structure of N-myc TAD. As an IDP, the scope for obtaining a classical structure is limited, however NMR was employed instead to establish individual residues dynamics and structural propensity. In this part of the work, the secondary structure propensity of N-myc TAD was interrogated through $^{13}C\alpha$ secondary chemical shifts and chemical shift index (CSI). This was compared against secondary structure predictions based on primary amino acid sequence. NMR relaxation data that was collected informs of local motions of individual residues in comparison to overall global motions of a protein and are thus a useful probe of major motions that govern dynamics of specific residues. Finally, two structure modelling approaches were utilised. AlphaFold and CS-Rosetta were used to confirm the predictions of secondary structure propensities. The helicity of S^{76} – G^{89} region was further probed using CD spectroscopy.

This chapter also explores the origin of two distinct chemical environments experienced by S^{76} – G^{89} and observed in N-myc$_{64-137}$ ^{1}H-^{15}N HSQC spectrum, realised in the presence of major and minor population peaks. Three possible mechanisms were interrogated, using a variety of tools: Z-Z chemical exchange experiment, R_2 relaxation data, intact mass spectroscopy, amongst others.

This work establishes IDP nature of N-myc TAD, however we have also established regions of α-helical propensity within residues S^{76} – G^{89} and previously unreported A^{125}–A^{132}. Notably, both regions have been documented to be involved in interactions with important binding partners.

4.2 Results

4.2.1 Secondary chemical shift

The observable parameter of NMR is chemical shift, which is a sensitive probe of chemical environment and thus is reflective of differences in average structural geometries. NMR backbone assignment of the N-myc TAD provided chemical shifts for H^N, N, C', Cα, Cβ, Hα and Hβ nuclei for almost all N-myc TAD residues. Secondary chemical shift index (CSI) is an easy method of predicting secondary structure propensity that, unlike many NMR experiments, does not require advanced processing and collection of experimental data above what is necessary for the backbone assignment. ^{13}Cα chemical shifts are particularly useful in detecting secondary structures, as these shifts are heavily influenced by backbone dihedral angle propensities (Buck, 1998; Schwarzinger et al., 2001).

^{13}Cα secondary shifts for N-myc TAD were obtained from the backbone assignment and the predicted random coil secondary shifts were calculated using Random coil chemical shifts for intrinsically disordered proteins from the University of Copenhagen (Department of Biology, n.d.). Secondary chemical shift data for N-myc TAD (Figure 4.3 A) indicates that the N-terminal portion of N-myc TAD is a random coil, with erratic values between adjacent residues. There are two regions that reflect helical propensity and these are W^{77} – E^{86} and A^{122} – E^{132}, with consistent positive values of secondary chemical shifts. As part of the backbone assignment was aided by collecting spectra for the truncated N-myc$_{64-137}$ construct, a ^{13}Cα secondary shift analysis was also performed for this smaller construct. The pattern of the ^{13}Cα secondary shifts is very similar to that of the full-length construct with consecutive positive secondary shift values shown between V^{78}–N^{85} and A^{122}–E^{132} (Figure 4.3.B).

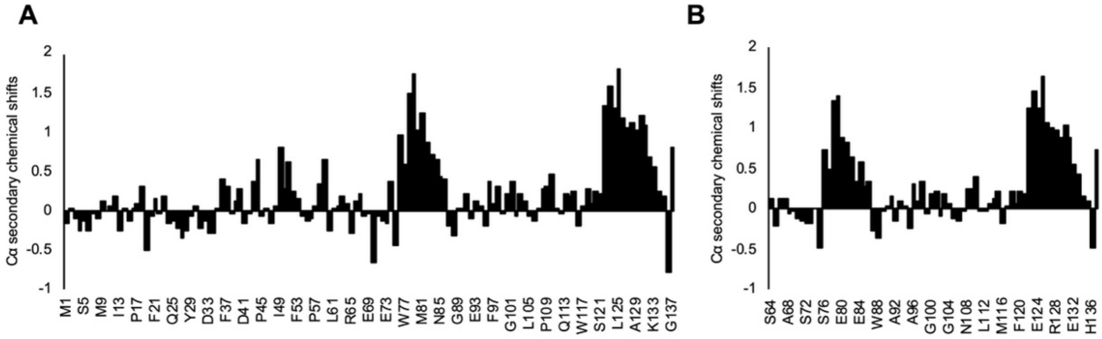


Figure 4.3. ^{13}Cα secondary shifts for N-myc TAD
A. ^{13}Cα secondary shifts for N-myc TAD were plotted against N-myc TAD sequence, were calculated using reference coil values generated by 'Random Coil Chemical Shift calculator for IDPs' (Department of Biology, University of Copenhagen, n.a), and a difference between the observed chemical shift and the reference chemical shift was calculated. **B.** ^{13}Cα secondary shifts for N-myc$_{64\text{-}137}$ plotted against N-myc$_{64\text{-}137}$ sequence, calculated as in A. Positive values indicate α-helices, negative values indicate β-sheets and values nearing 0 or variation in consecutive values indicate a lack of structure.

4.2.2 Prediction of secondary structures from primary amino acid sequence

It is possible to predict the secondary structures from the analysis of the primary sequence alone, without any experimental data. Servers such as JPred4 (Drozdetskiy et al., 2015) and PSIPRED v 3.2 (Jones, 2022) interrogate sequential strings of amino acids against structures already deposited in PDB. The predictions based on a primary amino acid sequence of N-myc TAD agree on the presence of α-helix at the C-terminus (R^{123}–K^{133} and A^{122}–L^{134}, for Jpred4 and PSIPRED, respectively). Jpred4 does not predict any other secondary structures, in contrast to PSIPRED, which places α-helices at residues I^{49}–F^{53}, P^{75}–L^{82}, A^{93}–F^{97} and P^{109}–M^{116} (Figure 4.3). P^{75}–L^{82} and P^{109}–M^{116} α-helices have N-terminal proline, which is commonly seen in other structures. The remaining residues are predicted to be a random coil.

4.2.3 Chemical shift index

Secondary chemical shifts can be transformed into a chemical shift index (CSI) by assigning an integer value of +1, 0 or -1 to chemical shifts for downfield, no shift, or upfield deflection from random coil values, respectively (Wishart et al., 1992). ^{13}Cα are characterised by +1 chemical shifts in α-helices and -1 chemical shifts in β-strands, whereas ^{13}Cβ displays the opposite trend (Higman, 2012). This analysis is extended towards H^{N}, N, C' and Hα chemical shifts also. CSI is commonly used in conjunction with predictors of secondary structures based only on amino acid sequence. We employed CSI 3.0 and Torsion Angle Likeliness Obtained from Shift and Sequence (TALOS-N; Shen and Bax, 2013; Hafsa et al., 2015) to test N-myc

TAD secondary structures. TALOS-N calculates torsion angles from CSI of H_N, N, C', Cα, Cβ, Hα and Hβ by looking at strings of polypeptides matching the sequence of interest and the previously deposited structures (Shen and Bax, 2013). It then generates Ramachandran plots, together with the probability of a given structure (random coil, α-helix or β-sheet; Shen et al., 2009a).

CSI 3.0 utilises TALOS-N in its prediction calculations (Hafsa et al., 2015). Both CSI 3.0 and TALOS-N predict the majority of N-myc TAD to be a random coil. CSI 3.0 only predicts two α-helical regions, spanning residues T^{79}–L^{83} and A^{122}–S^{131}, whereas TALOS-N places two α-helices at W^{77}–L^{83} and A^{122}–E^{132}. In addition, TALOS-N predicts four β-turns, at M^1, M^{11}–C^{14}, F^{28}–Y^{29} and at Y^{36}, however without consecutive strings of similar predictions it is unlikely that these are accurate estimations of secondary structure (Figure 4.4 B). N-myc TAD within this region is prone to proteolytic degradation, causing changes to resonant frequencies in observed HNCA peaks. This could result in abnormal CSI values and the prediction of single-residue β-sheets. Calculated psi (ψ) and phi (φ) angles for the predicted helical regions show typical tight clustering between -150° to -20° and -100° to -45 of Ramanchandran plot, indicating torsion angles are in agreement with α-helix (Hovmöller et al., 2002; Chong and Ham, 2013).

TALOS-N also calculates Random Coil Index (RCI) S^2 order parameter which is a measure of the rigidity of individual residues compared the overall tumbling of a molecule. Residues with high S^2 order parameter are characterised by lower frequency of local motions than the overall tumbling of a molecule, indicating a presence of secondary structure (Berjanskii and Wishart, 2007). RCI S^2 varies across the molecule (Figure 4.4 B), indicating that despite N-myc TAD classification as an IDP, it possesses a lot of variety in its internal motions. Higher (RCI) S^2 values are seen for the predicted helices in W^{77}–L^{83} and A^{122}–E^{132}, but also residues L^{55} – S^{62}, which is an area of N-myc phosphodegron. This region is particularly rich in prolines, which might explain the reduced level of fast internal motions for these residues.

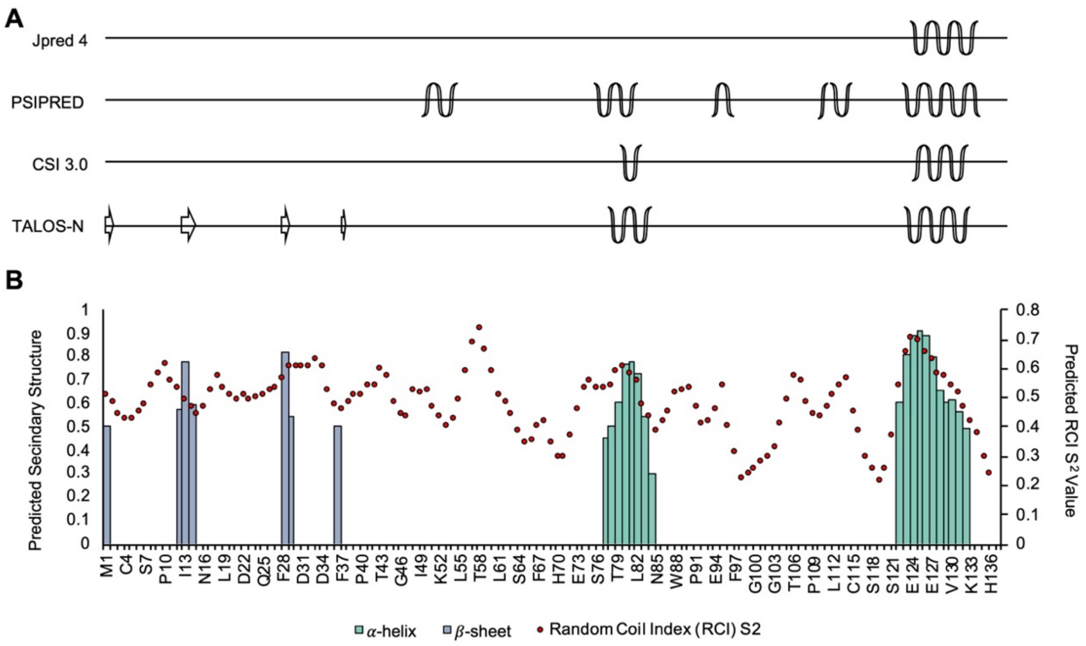


Figure 4.4. Secondary structure propensity of the N-Myc TAD based on primary sequence analysis and CSI predictions from experiment.
A. Secondary structure propensities were analyses based on the primary sequence (Jpred 4 and PSEPRED) or based on primary amino acid sequence and the experimentally-derived CSI values obtained through the NMR assignment. Arrows indicate the presence of predicted β-sheets/turns and ribbons indicate the presence of predicted α-helices. **B** TALOS-N predicted structures with blue bars corresponding to β-sheets, green bars corresponding to α-helices and the regions that lack the bars indicate random coil or unstructured elements. Red circles plotted are the predicted S^2 Random Coil Index.

4.2.4 Relaxation measurements on the N-myc TAD

To investigate further the behaviour of the N-myc TAD in the context of local and global structure, a set of three relaxation measurements, ^{1}H-^{15}N-nOe, R_1 and R_2, was employed to measure the dynamics on ps–ns timescale to probe local dynamics of the protein. ^{1}H-^{15}N-nOe values inform whether individual residue dynamics is determined by overall rotational of the molecule (τ_C) and are thus ‘fixed’ or rigid in nature, or whether their dynamics is faster than τ_C, thus indicating a degree of flexibility (Kay et al., 1989). Negative nOe values indicate large amplitude motions on short timescales which are seen at the N- and C- termini of N-myc TAD (Figure 4.5) as these regions experience the highest degree of freedom in their motions (Yao et al., 2001). Regions with restricted local movements result in increasingly positive nOe values as their rotational motions approach τ_C. Most of the N-myc TAD residues are characterised by ^{1}H-^{15}N-nOe values in the range 0.2–0.4, indicating that most of the TAD is flexible in nature on the picosecond to nanosecond time scale (Figure 4.5). The average nOe values of N-myc TAD, excluding M^1, P^2 and G^{137} is 0.18, but within the sequence are stretches of residues which are characterised by higher nOe values (~0.4). These regions: W^{50}–F^{53}, W^{88}–S^{90} and S^{118}–S^{121} are likely to experience more restrictive motions of the N–H vectors (Figure 4.5; Estrada et al., 2011). Some residues were excluded from the analysis either due to peak overlap or peak broadening where it was impossible to confidently follow peak intensities for independent residues.

R_1 and R_2 measurements were performed to probe longitudinal and transverse relaxation rates, respectively, which depend on the level of internal motions, as well as other factors, and can be related to the structural propensity of a residue. Both R_1 and R_2 measurements report ps – ns motions but R_2 is also impacted by μs – ms timescales (Hellman et al., 2014; Batchelor et al., 2019). R_1 values remain consistent for residues throughout the length of N-myc TAD and average 1.6 s^{-1}, except for the termini of the protein (Figure 4.5), indicating that at ps-ns time scale all residues are similar in their dynamics. R_2 values increase with slower tumbling and are impacted by μs – ms time-scales dynamic processes, which signifies more ‘global’ conformational motions such as protein folding, oligomerisation, or chemical exchange between different conformers (Kovermann et al., 2016). For N-myc TAD R_2 range between 2.0 s^{-1} and 8.0 s^{-1} and average at 4.9 s^{-1}, however they vary across the sequence. The regions of increased R_2 values are W^{50}–F^{53}, L^{87} and G^{89} and I^{111}–E^{132} and this was observed in the regions of latent helicity (Figure 4.5).

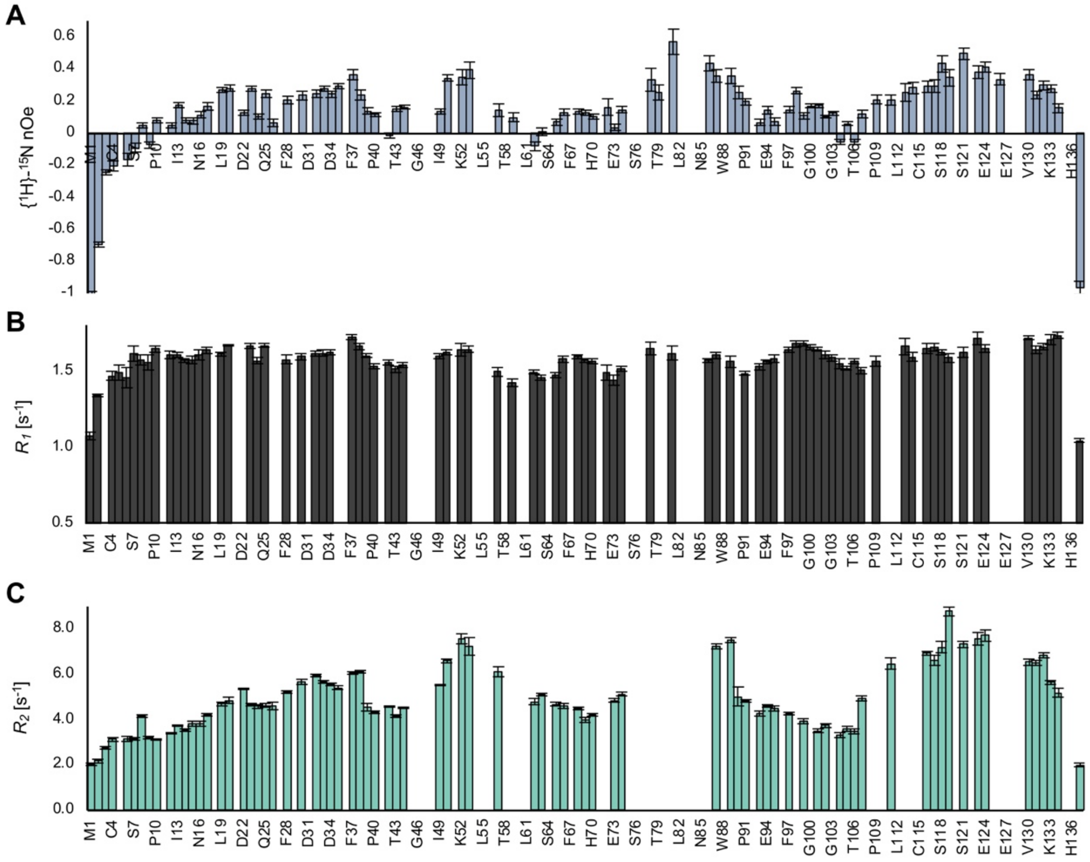


Figure 4.5. Relaxation measurements for the N-myc TAD domain.
^{1}H-^{15}N nOe (**A**), R_1 (**B**) and R_2 (**C**) of N-myc TAD. The first three residues in the protein sequence are artifacts of cloning and TEV-Nia protease cleavage and are omitted from this analysis for clarity. Missing residues are either prolines, not observable in ^{1}H-detected experiments, or from overlapping peaks which were omitted in the data analysis due to ambiguities in peak intensity measurements.

4.2.5 Spectral density mapping

Whereas ^{1}H-^{15}N-nOe, R_1 and R_2 are valuable source of dynamics information, they are largely descriptive and difficult to relate to any parameter of dynamic metric. Relaxation data can be further processed to obtain the read-out of behaviour of individual vectors at specific frequencies. Model-free analysis, which is a first type of analysis, assumes that a molecule tumbles isotropically with a single correlation time. Due to intrinsically disordered nature of N-myc TAD, its vectors are anisotropic and their motion is independent of one another. Spectral density mapping is used instead, which looks at each ^{1}H-^{15}N vector individually and calculates the probability of its motions at a particular frequency (ω; Farrow et al., 1995). Spectral density mapping is described by three equations:

$$R_1 = 3AJ(\omega_N) + AJ(\omega_H - \omega_N) + 6AJ(\omega_H - \omega_N) + BJ(\omega_N) \tag{1}$$

$$R_2 = 2AJ(0) + \frac{3A}{2}J(\omega_N) + \frac{A}{2}J(\omega_H - \omega_N) + 3AJ(\omega_H) + 3AJ(\omega_H + \omega_N) + \frac{2B}{3}J(0) + \frac{B}{2}J(\omega_N) \quad (2)$$

$$\sigma = -AJ(\omega_H - \omega_N) + 6AJ(\omega_H + \omega_N) \quad (3)$$

Where *J* is a spectral density function, ω_N is a ^{15}N frequency, ω_H is ^{1}H frequency. Ideally, to solve spectral density functions, ^{1}H-^{15}N-nOe, R_1 and R_2 are collected at two ^{1}H frequencies. However, if the data is collected at only one frequency, reduced spectral density function can be used instead using few assumptions. As ω_H exceeds ω_N significantly, it can be assumed that $J(\omega_H+\omega_N) \approx J(\omega_H-\omega_N) \approx J(\omega_H)$ are similar and the frequencies that need to be calculated are $J(\omega_N)$, $J(\omega_H)$ and $J(0)$. $J(0)$ describes the probability of motions at ps–ns time scales, whereas $J(\omega_H)$ is more sensitive to the slower motions (μs–ms) and $J(\omega_H\text{-}\omega N)$ is sensitive to all time-scale motions (Hellman et al., 2014). The analysis of frequencies for N-myc TAD is shown in Figure 4.6. The full analysis of this data is beyond the scope of this work, however few observations can be made. Residues with helical propensity (W^{77} – G^{89} and P^{109} – G^{137}), which are highlighted in red in Figure 4.6, are characterised by motions at slower frequencies (μs) which could correspond to chemical exchange. There is also another region (Figure 4.6, green rectangle) that is characterised by ps-ns timescale of motions (calculated by scripts generated by Theodoros Karamanos).

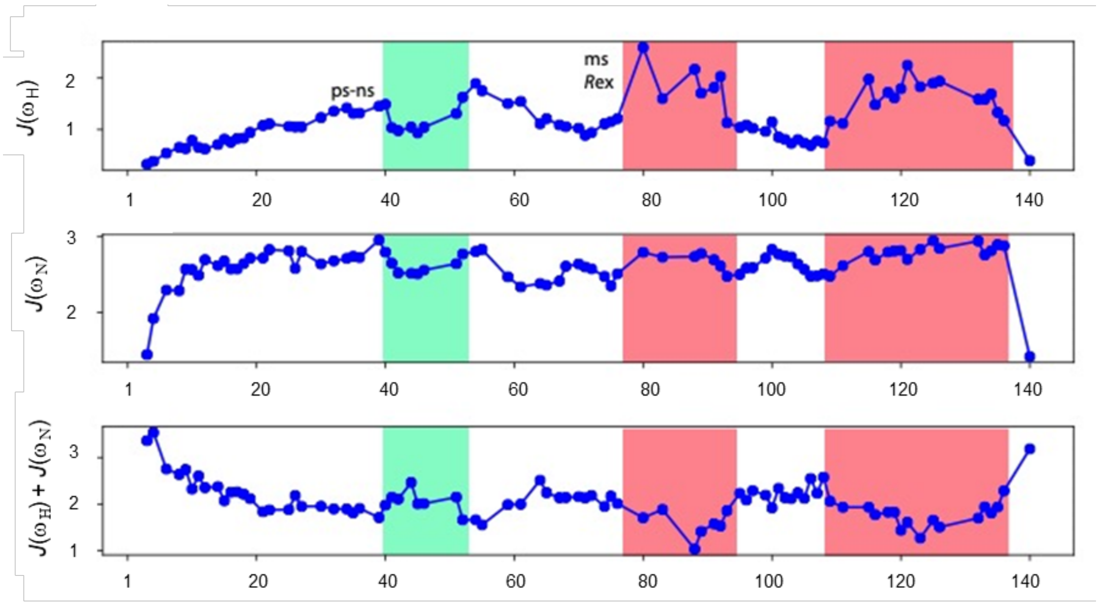


Figure 4.6. Reduced spectral density mapping for N-myc TAD.
Reduced spectral density mapping is the analysis applied to relaxation data, when it was collected only at one spectrometer frequency. It aims to quantitatively establish the probability of motion of ps-ns timescale plotted as a function of residue number. Time-scales of detected motions are indicated in green (ps-ns) or red (ms). R_{ex} refers to relaxation. Courtesy of Theodoros Karamanos.

4.2.6 Modelling of N-myc TAD structure

We utilised two structure-prediction tools to model the overall structure of the N-myc TAD. AlphaFold is one of the latest developments which uses deep machine learning to model protein structures (Jumper et al., 2021). The structure generated shows the N-Myc TAD as mainly disordered, however due to the limited number of N-myc (and c-myc) PDB entries, the confidence levels of predictions are low or very low (Figure 4.7 A). The 'pLDDT' value shown is a derivative of lDDT – Cα metric (local Distance Difference Test), which measures how well local atomic interactions are recreated in the model in comparison to the reference protein (Mariani et al., 2013). AlphaFold indicates that all predictions with pLDDT < 50 have ribbon-like appearance and are implied to be intrinsically disordered (Jumper et al., 2021; Varadi et al., 2022). The regions of predicted α-helices are: D^{48}–E^{54} (70 > pLDDT > 50), S^{76}–E^{86} (70 > pLDDT > 50 for most of the residues); P^{109}– L^{112} (70 > pLDDT > 50) and A^{122}– Q^{135} (90 > pLDDT > 70; Figure 4.7 A).

To incorporate the experimental data into structure modelling, we utilised the chemical shift-Rosetta (CS-Rosetta) server, which predicts the overall structure based on $^{1}H^{N}$, ^{15}N, $^{13}C'$, $^{13}C\alpha$, $^{13}C\beta$, $^{1}H\alpha$ and $^{1}H\beta$ assignment to generate a *de novo* model (Lange Oliver et al., 2012; Shen et al., 2010; Shen et al., 2009b; Shen et al., 2008). In tandem, the depository is searched for similar polypeptides and the lowest energy models are selected and refined. Meshing this two-prong approach results in the generation of accurate 3D structures (Shen et al., 2008). CS-Rosetta has predicted 3000 structures for N-Myc TAD based on its CSI and derived torsion angles generated by TALOS-N, however the predictions are characterised by low convergence and structures score high energy with high mean root square deviations (Cα-RMSD; Figure 4.7 C), indicating a large discrepancy between predicted structures. The best model, characterised by 10.2863 RMSD-Cα against the lowest energy, predicts two α-helices, in positions E^{83}–L^{86} and A^{125}–A^{132} (Figure 4.7 B) and one single-residue β-turn surrounding position W^{120} (Figure 4.7 B).

The predicted structures of AlphaFold and CS-Rosetta are in agreement with secondary chemical shift, CSI and relaxation data which all predict helical propensity in regions of W^{77} – G^{89} and A^{122}–A^{132} and assigns the majority of N-myc TAD as unstructured.

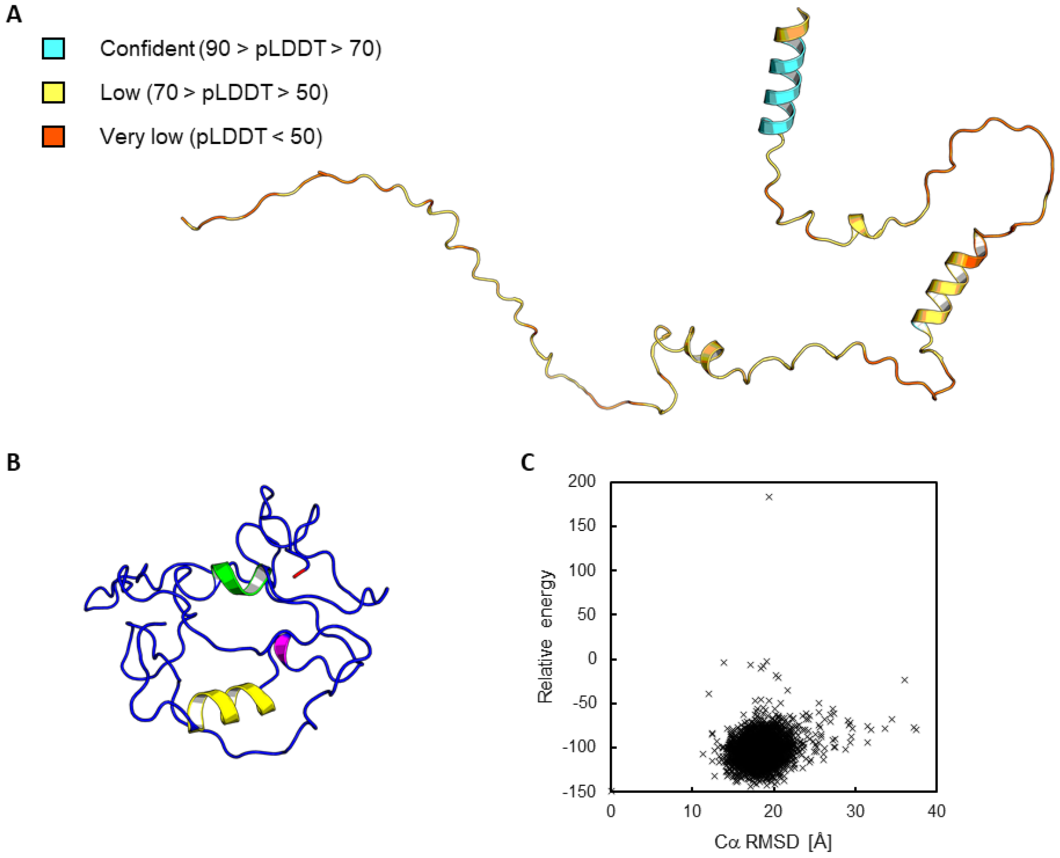


Figure 4.7. N-myc TAD structure predictions.
A. Structure prediction using AlphaFold. Four helices predicted by AlphaFold are D^{48}–E^{54}; S^{76}–E^{86}; P^{109}–L^{112} and A^{122}–Q^{135}. The colours of the structure reflect the structure prediction confidence levels measured as local Distance Difference Test (pLDDT). Full-length N-myc protein was used as an input however, everything apart from N-myc TAD was omitted for clarity. **B.** Structure prediction using the CS-Rosetta web server. The N-terminus is highlighted in red, α-helix E^{83}–L^{86} in green, β-turn around W^{120} in magenta and α-helix A^{122}–A^{132} in yellow. **C.** Each crosshair represents one of 3000 structure predictions generated by CS-Rosetta, scored against their relative energy and Cα-RMSD. (Lange Oliver et al., 2012; Shen et al., 2010; Shen et al., 2009b; Shen et al., 2008; Jumper et al., 2021; Varadi et al., 2022).

4.2.7 Helical propensity of the N-myc$_{75\text{-}89}$ fragment

The crystal structure of N-myc$_{28\text{-}89}$ with Aurora A and NMR experimental data collected in this work on N-myc TAD indicates helical propensity within residues W^{77} – W^{88}. To interrogate whether this helix formation is in response to Aurora A binding or whether this region is helical in *apo* state, synthetic peptides of N-myc$_{73\text{-}89}$ were designed. Apart from a wild type N-myc$_{73\text{-}89}$ peptide (Figure 4.8 A), two mutations were also introduced into this region in position L^{82} and L^{87} (N-myc$_{73\text{-}89}$ DM). Leucines promote helix formation, therefore to reduce helix propensity of this region, these were mutated to glycines, which are known helix-breakers (Imai and Mitaku, 2005; Chakrabartty et al., 1994). In addition, L^{82} and L^{87} point away from the surface of

Aurora A in the crystal structure, thus allowing to decouple interaction-promoted helicity from intrinsic helicity of the peptide (Figure 4.1 B and C; Richards et al., 2016).

The helicity of N-myc$_{73\text{-}89}$ and N-myc$_{73\text{-}89}$ DM was assessed using far-UV circular dichroism (CD) with or without the addition of 2,2,2-fluoroethanol (TFE), a known helix promoter (Cammers-Goodwin et al., 1996). In the absence of TFE, N-myc$_{73\text{-}89}$ WT and N-myc$_{73\text{-}89}$ DM display a typical far-UV absorbance of a random coil, with characteristic single minimum in the mean molar residue ellipticity (MRE) at 200 nm (light blue and ligh green traces; Figure 4.8 A; Greenfield, 2006). N-myc$_{73\text{-}89}$ WT ellipticity scores overall slightly more negative values within the region 197 – 228 nm. In the presence of 30% v/v TFE, differences between the helical propensity of these two peptides becomes more apparent. N-myc$_{73\text{-}89}$ WT absorbs positively at 193 nm and negatively at 208 nm and 222 nm, which is characteristic of helical propensity (dark blue trace, Figure 4.8 A; Woody, 1995). The introduced substitution of two Leu to Gly in N-myc$_{73\text{-}89}$ DM is detrimental for the helical propensity of this region. The far-UV CD profile still resembles that typical for an α-helix, however the MRE values at 208/222 nm are far less negative in comparison to N-myc$_{73\text{-}89}$ WT (dark green; Figure 4.8 B). To interrogate the impact of TFE on N-myc$_{73\text{-}89}$ WT helical propensity, the peptide was diluted into CD buffer containing 15% TFE. The far-UV CD profile indicates an intermediate MRE profile between α-helical and random coil profile in the presence of 15% v/v TFE (Figure 4.8 C). The data indicates that N-myc$_{73\text{-}89}$ WT has much higher helical propensity than N-myc$_{73\text{-}89}$ DM, which is only moderately helical at 30% v/v TFE. Helix formation of N-myc$_{73\text{-}89}$ WT depends on the presence of TFE and positively correlates with the concentration of this compound, which is in agreement with NMR data which predicts a region of increased helical propensity within W^{77} – G^{89} region.

A

$\text{Ac-E}^{73}\text{PPSWVTEMLLENELWGS}^{90} - \text{NH}_2$

$\text{Ac-E}^{73}\text{PPSWVTEM}\underline{\text{G}}\text{LENE}\underline{\text{G}}\text{WGS}^{90} - \text{NH}_2$

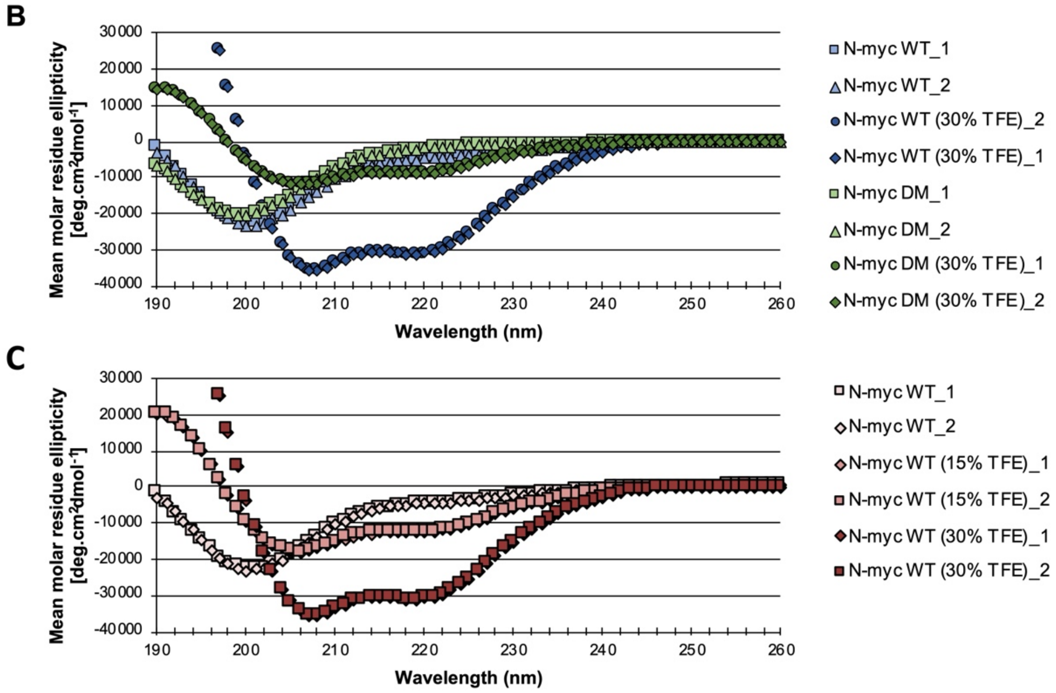


Figure 4.8. Far-UV circular dichroism (CD) of N-myc$_{73\text{-}89}$ WT and N-myc$_{73\text{-}89}$ DM peptides
A. Two N-myc$_{73\text{-}90}$ peptides used for secondary structure investigation through CD. The top peptide is N-myc$_{73\text{-}90}$ WT and the bottom peptide is a double mutant referred to as N-myc$_{73\text{-}90}$ DM. Mutated residues are underlined. **B.** The comparison of N-myc$_{73\text{-}89}$ WT and N-myc$_{73\text{-}89}$ DM mean molar residue ellipticity (MRE) in the presence of 30% v/v TFE (dark blue and dark green traces, respectively) and in the absence of TFE (light blue and light green traces, respectively). **C**. MRE of N-myc$_{73\text{-}89}$ WT in increasing concentrations of TFE. Trace with no TFE present in beige, trace with 15% v/v TFE is pink and the trace with 30% TFE is in dark red. All data was collected at 20° C in CD buffer (Table 2.1).

4.2.8 Interrogation of divergent chemical environments of S^{76} – G^{89}

The assignment performed on N-myc$_{64\text{-}137}$ resulted in collection of ^{1}H-^{15}N HSQC spectrum with better signal-to-noise ratio and revealed two distinct chemical environments for the residues that are involved in the α-helix formation when bound to Aurora A (Figure 4.9 A, blue letters). Residues S^{76} - G^{89} are characterised by two peaks, one larger (major population) and one smaller (minor population). This stretch is preceded by a ProPro motif and is capped at position 91 with another Pro (Figure 4.9 B).

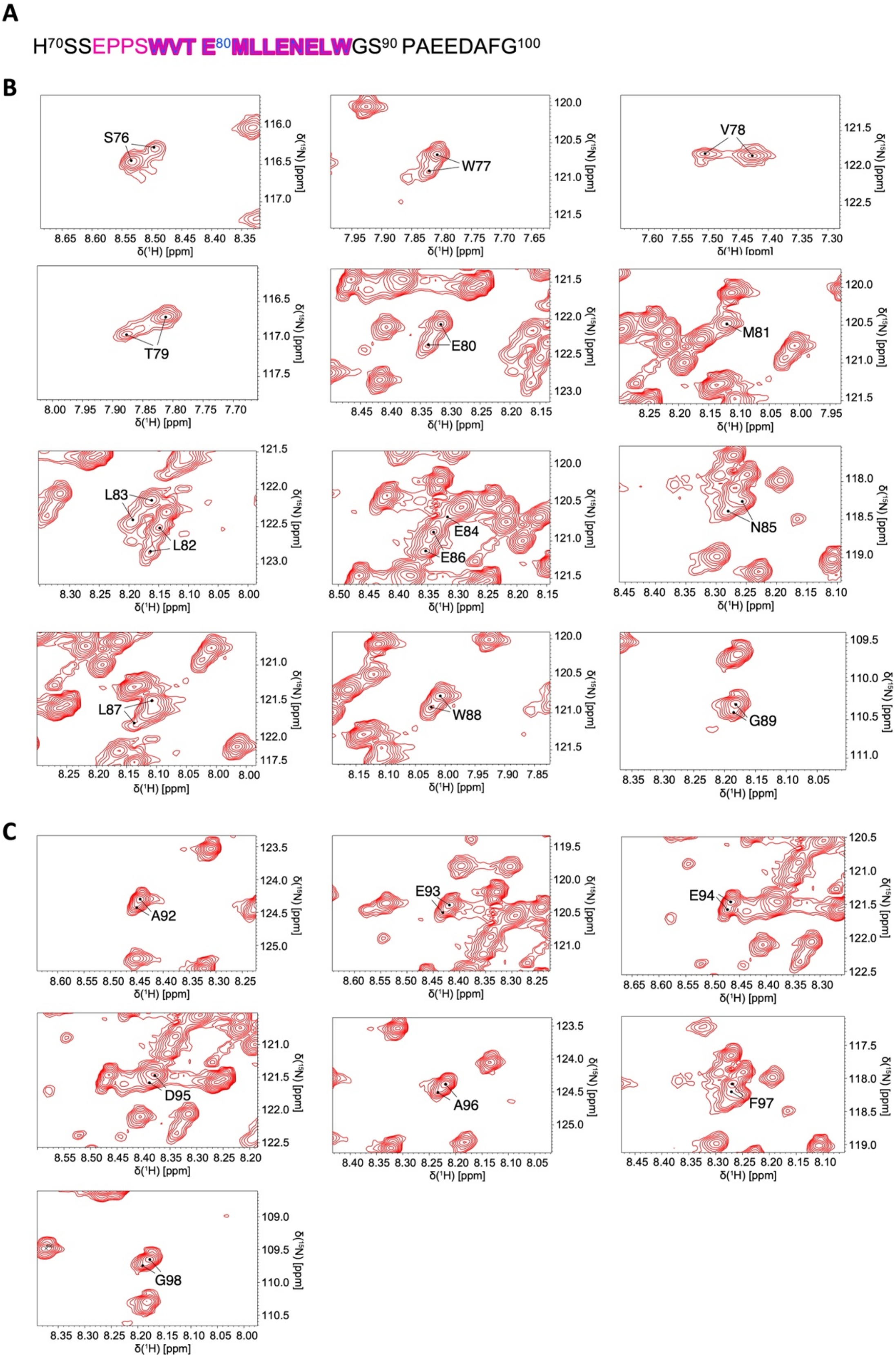


Figure 4.9. Major and minor population of peaks.

A. Primary amino acid sequence of a fragment of N-myc TAD seen in the crystal structure with Aurora A. Blue letters correspond to residues forming a helix and pink letters correspond to the residues seen in the crystal structure as a random coil

(Richards et al., 2016). **B.** Panel showing regions of the N-myc$_{64\text{-}137}$ ^{1}H–^{15}N HSQC spectrum and H–N correlations for residues participating in the formation of a helix when bound with Aurora A. Major and minor peaks are highlighted. **C.** Panel showing regions of the N-myc$_{64\text{-}137}$ ^{1}H–^{15}N HSQC spectrum from residues C-terminal from P^{91}. Major and minor peaks are highlighted.

In NUS ^{1}H-detected triple detected spectra, it was possible to sequentially link minor population peaks (Figure 4.11).

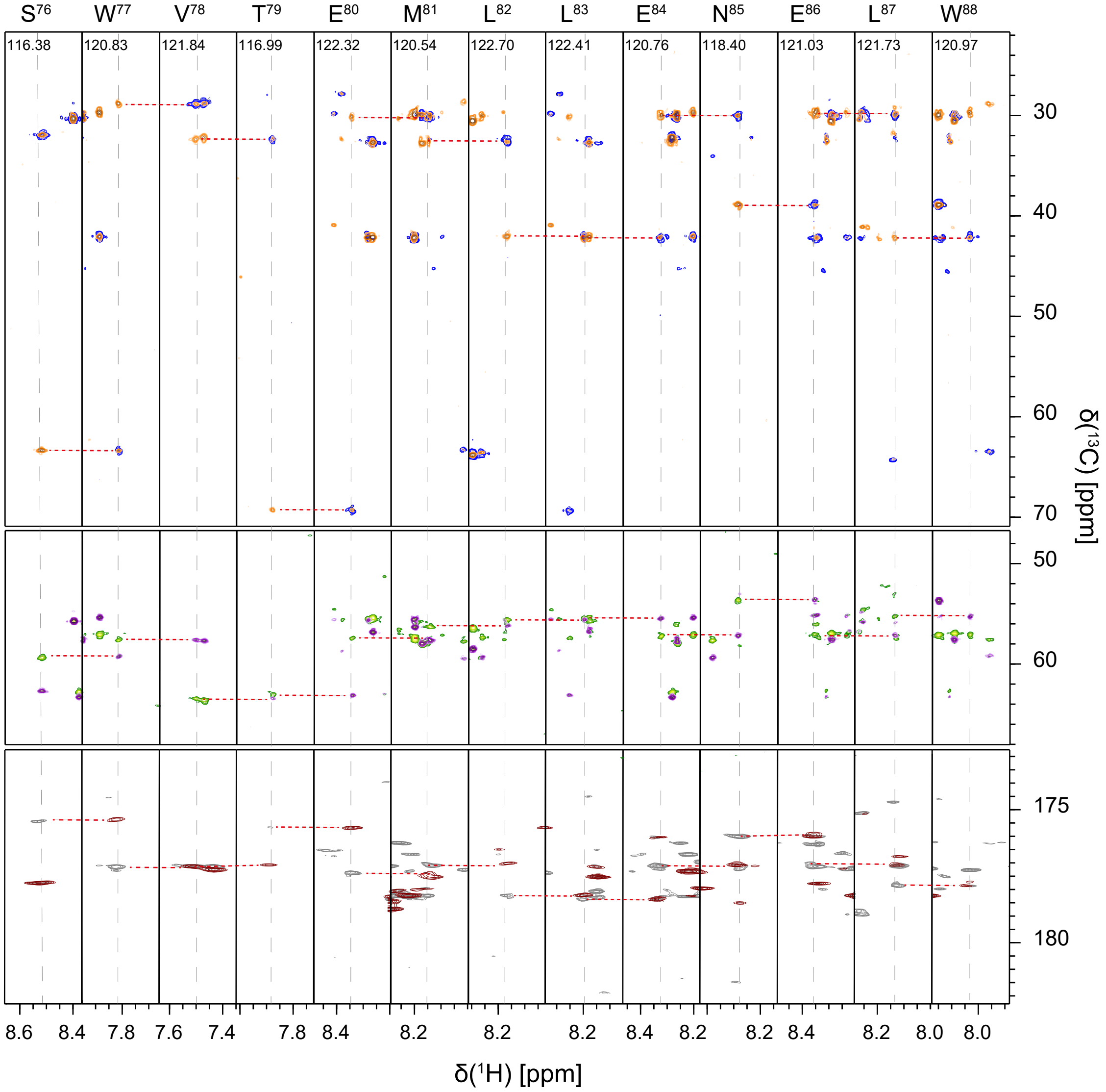


Figure 4.10. Sequential linking of minor population peaks of C’, Cα and Cβ. H(N)–C plane strips through H–N correlations for minor population peaks of residues S^{76}–W^{88}. A series of ^{1}H-detected triple resonance experiments recorded on a 750 MHz Bruker spectrometer were used in the assignment: HNCO (brown), HNcaCO (gray), HNCA (green), HNcoCA (purple), HNcocaCB (orange) and HNcaCB (navy). Data were collected using NUS with 25% sampling. ^{15}N frequencies for the H(N)–C plane are shown at the top of each strip.

We calculated the relative amount of major and minor protein conformers by quantifying ratio of height of major-to-minor H-N correlation peaks for residues within S^{76} – S^{90} region, using CcpNMR Analysis v2.5 software (Vranken et al., 2005). Residues M^{81}, E^{84}, L^{87}, G^{89} and S^{90} were omitted in the quantification due to peak overlap. The quantified ratio varied between residues and was between 1.7 – 3.0 (Figure 4.11 A). To test if this effect is concentration-dependant, ^{1}H-^{15}N HSQC was collected again with doubled concentration of N-myc$_{64-137}$ (300 μM) and the major-to-minor H-N correlation peaks were calculated again. The height ratio of these peaks was smaller in comparison to 150 μM concentration, but only moderately (Figure 4.11 A). If the effect resulting in major and minor peak population was concentration-dependent, the ratio of peak heights would change by half, most likely pushing the equilibrium between these two states towards minor population peak population.

To control for the change in height across entire construct, peaks unrelated to this region were also quantified in the same fashion (A^{92}–G^{98}, N^{108}, V^{110}, C^{115} and W^{117}; Figure 4.11 B). These peaks are also characterised by minor and major population, as they lie within the proximity of proline residue. Most of the peaks involved in the formation of the helix experience a slight decrease in the ratio for 300 μM *vs.* 150 μM concentration, whereas the peaks outside the helix region experience slight increase in the ratio, i.e. their major peak heights increase with an increase in the concentration (Figure 4.11 B).

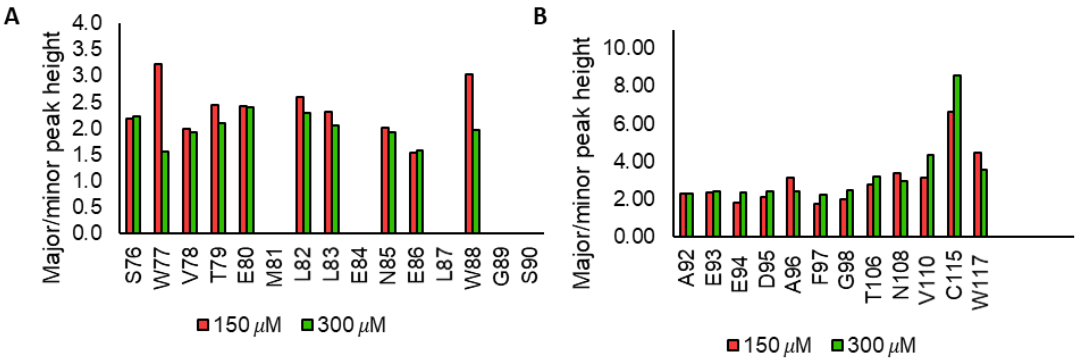


Figure 4.11. Major-to-minor peak height ratio of N-myc$_{64-137}$ construct.
A. Quantification of peak height ratio for the residues in the S^{75}–W^{88} region. **B.** Quantification of peak heights for selected residues of the N-myc$_{64-137}$ construct that are not involved in helix formation on binding to Aurora A. Red bars correspond to 150 μM and green bars correspond to 300 μM.

The ratio of major-to-minor peak fell within the same range of 2:1 for all of the residues at both concentrations, indicating that this effect is not concentration-dependent (Figure 4.11 A and B), thus the process driving these two distinct chemical environments is not concentration-dependent either. This excluded the possibility of a chemical equilibrium between major population which is a random coil and minor population which is a α-helix, being likely. We also quantified $^{13}C\alpha$ secondary chemical shifts for minor peak population, and they did not indicate an increase in secondary shift value, implicating that α-helix formation is not the cause of the difference in chemical environment.

To interrogate this observation further, chemical exchange was probed using Z-Z exchange experiment was used to measure the exchange of transverse relaxation between major and minor peaks as a function of time. Indication of chemical exchange within the set mixing time would show as two cross-peaks on ^{1}H-^{15}N-HSQC spectrum, with one cross-peak labelled with F_1 of major and F_2 of minor frequencies and the second cross peak would be labelled with F_1 of minor and F_2 of major frequency (Figure 4.12 A; Vallurupalli, 2009; Mittermaier and Meneses, 2013). Good signal-to-noise ratio was ensured by using 300 μM N-myc$_{64\text{-}137}$ sample, however no cross-peaks were detected for S^{76}-G^{89} residues, or A^{92} – W^{117} residues (Figure 4.12 B). The long mixing time (0.5 s) used shows a lack of interconversion between major and minor populations, indicating that they belong to a separate species or that the exchange rate occurs at much slower timescale.

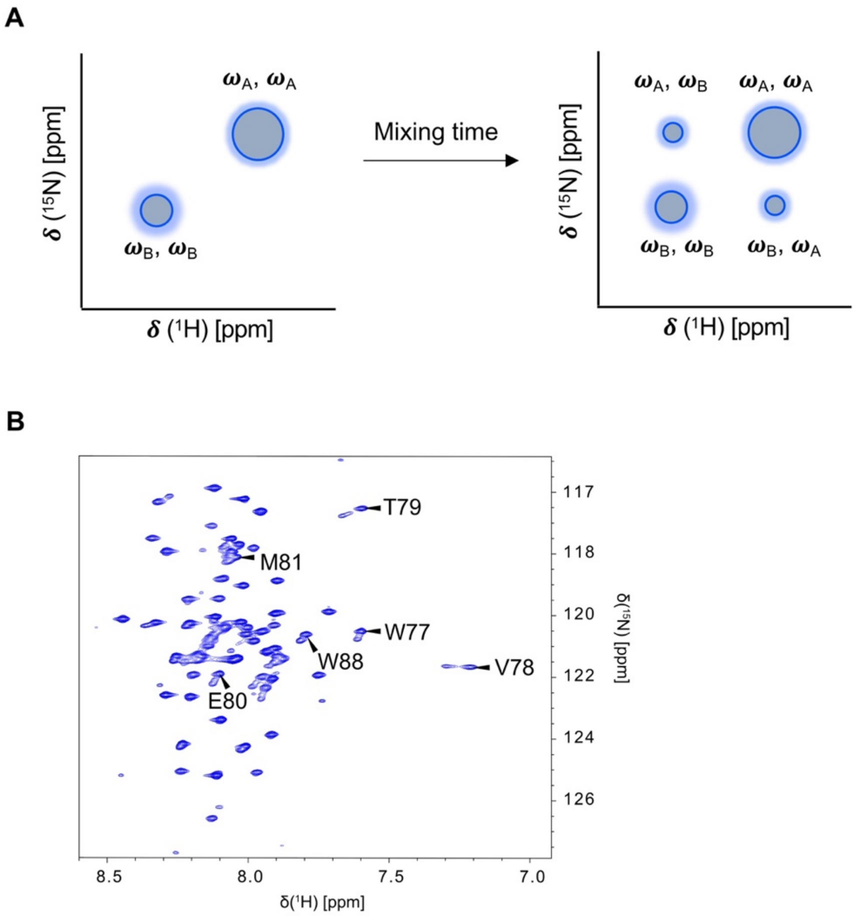


Figure 4.12. Z-Z exchange experiment of N-myc TAD.
A. Z-Z experiment was set up with 0.5 ms mixing time using 300 μM N-myc$_{64\text{-}137}$. The experiment was collected on 750 MHz Oxford NMR magnet, equipped in TCI-cryoprobe (5 mm). The presence of any chemical exchange would be realised in the presence of two extra cross peaks for the highlighted residues labelled with F_1 of major and F_2 of minor peak frequency and the second cross peak would be labelled with F_1 of minor and F_2 of major peak frequency. **B.** Z-Z experiment performed on 300 μM N-myc$_{64\text{-}137}$ sample. Residues experiencing peak splitting are highlighted.

To further investigate the differences in dynamics of these two populations, R_2 relaxation measurement was performed with the ten delays randomly selected between 50 ms – 100 ms. The peaks of minor populations relaxed 50–60% slower than the peaks of major population, as indicating by the smaller relaxation rates (Figure 4.13). M^{81}, E^{84}, N^{85}, E^{85} and L^{86} were excluded from the analysis due to peak overlap.

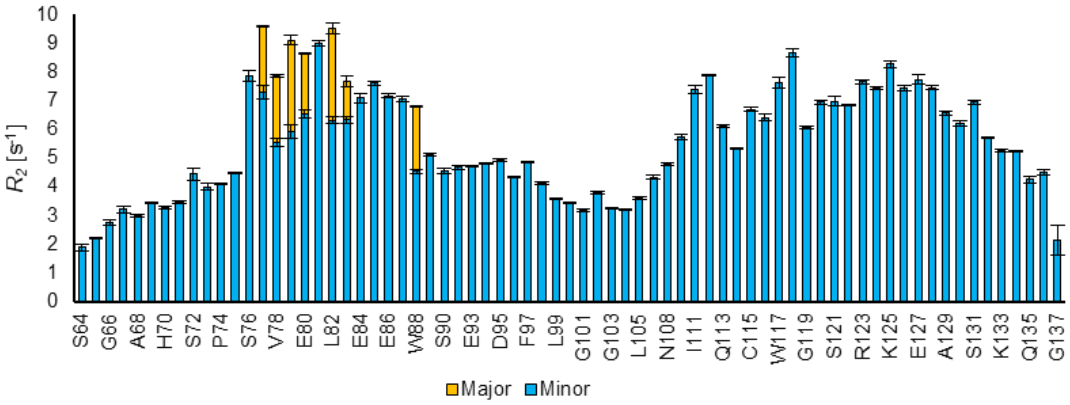


Figure 4.13. *R_2* of major and minor peaks of N-myc$_{64\text{-}137}$.
R_2 was measured on 300 mM N-myc$_{64\text{-}137}$ sample using ten delays selected between 50 ms – 100 ms. Major peaks R_2 rates are shown in blue and minor peaks R_2 rates are indicated in yellow.

To analyse the composition of the protein sample used in NMR, intact mass spectrometry analysis was performed on 20 μM of N-myc$_{64\text{-}137}$, with the sample incubated at RT for 24 hrs without the presence of protease inhibitors, to mimic the degradation occurring within NMR spectrometer during the collection of ^{1}H-detected triple resonance experiments. Intact mass spectrometry analysis indicates N-myc$_{64\text{-}137}$ as a single species of an expected weight of 8485.00 Da as predicted (Figure 4.14).

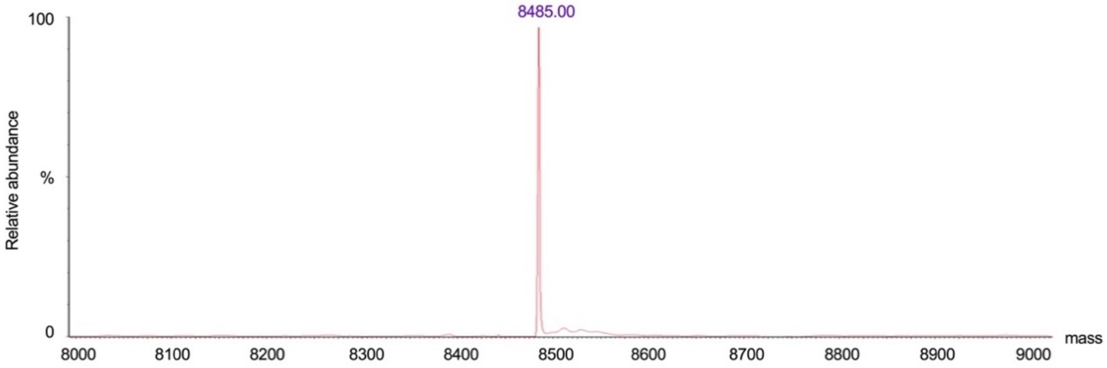


Figure 4.14. Intact mass spectroscopy analysis of N-myc$_{64\text{-}137}$.
The analysis was obtained following the incubation of thawed protein at room temperature, without the presence of protease inhibitors, to replicate the conditions during NMR experiment. The spectrum shows a single species which correlated with the expected molecular weight of 8485.00 Da.

4.3 Discussion

4.3.1 Helical propensity

Although obtaining structural data for IDPs is technically challenging, the employment of NMR allows for assessment of secondary structural propensities and the overall dynamics of a protein. We documented the presence of two transient/latent α-helices within positions S^{76}–W^{88} and A^{122}–E^{132} in otherwise unstructured N-myc TAD protein, using secondary chemical shift, CSI, NMR relaxation data, computer modelling and CD data. Establishing the exact boundaries of these regions is difficult. Each technique employed (secondary chemical shift, CSI, relaxation data, computer modelling) interrogates different aspects relating to structure and dynamics which results in slightly different positioning of these structures within the protein sequence.

N-myc$_{75\text{-}89}$ WT is unstructured when interrogated by far-UV CD, however upon addition of α-helix-promoting agent (TFE) the latent/potential helicity of this region becomes apparent. This suggests a propensity for 'folding – upon – binding' that is typical of IDPs and is observed in N-myc TAD in the context of Aurora A interaction or in the presence of the helicity-promoting region. This agrees with $^{13}C\alpha$ secondary chemical shifts that indicate helical propensity, and the NMR relaxation data (^{1}H-^{15}N-nOe, R_1 and R_2), which indicates a reduced degree of freedom in motions of S^{76}–W^{88} residues. ^{1}H-^{15}N-nOe values fall between 0.2 and 0.4, which is typical for a protein with conformational plasticity (Konrat, 2014). The relaxation data for N-myc's closest paralogue – c-myc$_{1\text{-}88}$ and c-myc$_{1\text{-}88}$ phosphorylated on serine 62 (pS^{62}-c-myc$_{1\text{-}88}$) are consistent with ^{1}H-^{15}N-nOe values shown for the N-myc TAD, confirming that these two proteins are both IDPs (Andresen et al., 2012; Helander et al., 2015). For comparison, a protein of similar size to the N-myc TAD (17 kDa) but globular in nature, such as VirB7 which forms part of the bacterial secretion system, has ^{1}H-^{15}N-nOe values that average well above 0.5 (Souza et al., 2011).

The presence of secondary or tertiary structure restricts anisotropic environment of individual residues and thus enhances cross-relaxation which leads to faster relaxation rates (Konrat, 2014). R_2 values for the N-myc TAD are in the range of 4.0–6.0 s^{-1} whereas for c-myc$_{1\text{-}88}$ and pS^{62}-c-myc$_{1\text{-}88}$ they are in the range of 15–20 s^{-1}, with the difference in the rates being attributed to the difference in the molecular weight of these two proteins (Helander et al., 2015). A variation in R_2 values for different residues in the sequence points to regions of restricted local mobility, such as regions of transient secondary structure. W^{50}–F^{53}, L^{87} and G^{89} and I^{111}– E^{124} all have higher R_2 rates in comparison to their neighbouring residues. In *apo* c-myc$_{1-88}$, regions 1–28 and 50–62 show high values for R_2 in comparison to neighbouring regions. Whereas we did observe some differences within W^{50}–P^{57}, there was no increase in R_2 rates for the very N-terminal part of the N-myc TAD (Andresen et al., 2012). This is particularly striking as this region encompasses part of MB0 (R^{10}–E^{28} in c-myc) which is conserved between N- and c-myc.

The majority of residues within the S^{76}–W^{88} region with helical propensity are missing from relaxation data analysis. W^{77}–N^{85} peaks diminish in intensity for longer delays in the relaxation experiments, thus they cannot be reliably used to obtain R_2 value. In IDPs this is typical of transient secondary structure regions, which undergo chemical exchange on intermediate timescale between folded and unfolded state, making these residues difficult to detect (Helander et al., 2015; Konrat, 2014). As this region is known to have helical propensity in N-myc TAD and is predicted to be an α-helix, it is consistent with an intermediate chemical exchange of two conformers.

We also demonstrated the presence of a second, previously unreported, region with helical propensity at the C-terminus of the N-myc TAD, encompassing residues A^{122}–E^{132} and falling within the region of MBII, with the predictions of TALOS-N, AlphaFold, CSI and dynamics for this region are more confident than the predictions for S^{76}–W^{88}.

4.3.2 Secondary structures in the context of partner binding

The regions with IDPs with secondary structural propensities are often involved in interaction with the binding partners (Wright and Dyson, 2009). The N-terminal α-helix, featured in the crystal structure when bound to Aurora A, is transient in nature, as documented by our far-UV CD data, and is predicted with low confidence for this region in the modelling data. The sequence identity is not conserved in myc paralogues, however a proximal region of c-myc is observed to be an α-helix in the crystal structure of c-myc$_{96-125}$ in complex with TAF1-TBP (Richards et al., 2016; Wei et al., 2019). The second site, spanning residues A^{122}–E^{132} has not been featured in any crystal structures, however, together with MB0, is has been implicated in the interaction with TRRAP which recruits multiple histone acetyltransferases through such as GNC5, responsible for histone acetylation (McMahon et al., 1998; Kalkat et al., 2018).

The mechanism by which IDPs bind to their targets through transient secondary structures has received a lot of attention recently, and these regions have been named molecular recognition elements (MoRFs; Radivojac et al., 2007; Garner et al., 1999). The regions S^{76}–W^{88} and A^{122}–E^{132} both fit the criteria of a MoRF, with the presence α-helix promoting amino acids (alanine, glutamate, arginine, leucine; Pace and Scholtz, 1998). MoRFs often exist as random coils in solution and become folded upon interacting with a binding partner, most commonly assuming an α-helical conformation (Radivojac et al., 2007; Garner et al., 1999; Wright and Dyson, 2009). The transition from unstructured to structured is advantageous, as it allows conformational flexibility and binding affinities that are low with fast *on* and *off* rates. This grants IDPs binding promiscuity and allows rapid modulation of the signal which is critical in the processes that IDPs regulate, such as mitogenic responses, gene expression and cell cycle (Wright and Dyson, 2009).

The mechanisms by which IDPs bind to their interacting partners through MoRFs is poorly understood, however two alternative theories have been proposed. In ‘induced folding’ theory, an IDP binds to its target protein in disordered form, initially forming non-specific hydrophobic interactions (Figure 4.15, top panel). Gradual interaction with the binding partner ‘molds’ secondary or tertiary structure on an IDP. In the ‘conformational selection’ model the IDP exists in a state of equilibrium between its folded and unfolded state. The target protein selects for and interacts with the folded population of the IDP (Figure 4.15, bottom panel; Wright and Dyson, 2009). The initial binding event seems to be mainly driven by hydrophobic interactions; however it has been reported that initial matching can also be mediated by electrostatic interactions and then hydrophobic interactions drive folding (Wright and Dyson, 2009; Arai, 2018).

Whether either of the MoRF sites on the N-myc TAD are examples of ‘induced fit’ or ‘conformational selection’ models remains to be answered. The helical propensity present within these two regions could drive both processes. Increased helical propensity lowers the entropic barrier to forming an α-helix upon binding but could also indicate a minor population of N-myc with a helical conformation, working in favour of both ‘induced fit’ or ‘conformational selection’ models. It has also been demonstrated that the two mechanisms are not mutually exclusive (Arai, 2018). Either models can be correct and be utilised within the N-myc TAD, depending on which site is driving the interaction or the cellular and physiological context.

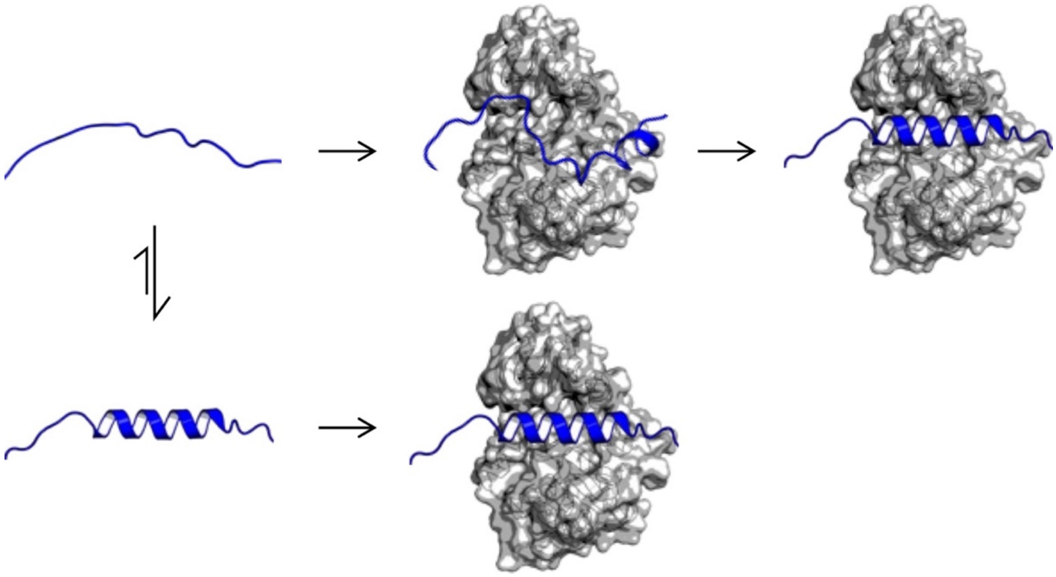


Figure 4.15. 'Induced folding' vs. 'conformational selection' model for an IDP binding to its binding partner.
In the 'induced fit' model (top row) an IDP interacts with a binding partner first in unfolded/random coil form (middle panel). The initial interaction has low affinity, mainly stabilised by hydrophobic interactions. As binding occurs an IDP assumes a stable conformation, often with secondary structure (right panel). In the 'conformational selection' model (bottom row) the IDP exists as an equilibrium between folded and unfolded states. A binding partner selects for and binds to a minor conformation of folded IDPs. Adapted from (Wright and Dyson, 2009).

4.3.3 The accuracy of model predictions

As CS-Rosetta utilises TALOS-N dihedral angle predictions based on the CSI, it is expected that the predictions generated by both algorithms would predict similar secondary structure (Shen et al., 2008; Shen et al., 2009b; Shen et al., 2010). The two models generated indicated that CS-Rosetta is more conservative in its estimation of secondary structures as it not only utilises dihedral angles to guide the model building process, but also scans N-myc TAD-identical polypeptide strings in the depositories. For example, the CS-Rosetta model predicts an α-helix in the position E^{81}–L^{86}, whereas TALOS-N places an α-helix from W^{77}–E^{84}. For the C-terminal α-helix the prediction from CS-Rosetta spans A^{122}–A^{128}, whereas for TALOS-N the prediction is between residues A^{122}–E^{132}.

Models generated by AlphaFold and CS-Rosetta are characterised by low confidence, which is mostly the reflection of the lack of N-myc or similar structures deposited in the database. There are many other structure prediction tools, all of them depend on utilising previously deposited structures to learn and to predict *de novo* structures. This introduces a significant bias against IDPs, firstly restricting the machine learning of model algorithms and secondly, reducing the confidence in predictions for these proteins. The C-terminal α-helix structure is not featured in any deposited structures, and it has high confidence scores, mainly due to its primary amino acid sequence, often seen in PBD as α-helix. An α-helix for the sub-set of residues S^{75}–W^{88} is predicted by both algorithms because it is featured in the depository, however its confidence score is lower, reflecting the small number of structures that feature this sequence.

4.3.4 Distinct chemical environment of S^{76} – W^{88}

This work interrogated extensively the presence of two distinct chemical environments for residues participating in the α-helix S^{76}–$W^{88,}$ which was apparent in ^{1}H-^{15}N HSQC of N-myc$_{64\text{-}137}$ construct. No definite conclusion could be drawn regarding the nature of these two distinct chemical environments. The effect of proline isomerisation, oligomer-formation and slow exchange between random coil and α-helical structure within this region could all contribute to this phenomenon, in non-mutually exclusive manner.

Proline is characterised by the incorporation of the side chain imino ring into its main backbone through a covalent bond. This rigidity causes a small energy difference between *cis* and *trans* peptidyl-prolyl bond conformations, making proline far more likely than other amino acids to exchange spontaneously between these two isomers on the seconds to minutes timescale (Eyles, 2001; Lu and Zhou, 2007). Native proteins usually favour the *trans* isomer, which forms 95% of all peptidyl-proline bonds seen in crystal structures (Weiss et al., 1998). The isomeric heterogeneity of peptidyl-prolyl bond cause neighbouring residues to experience two distinct chemical environments, resulting in two observable peaks per residue with a minor population representing a local *cis*- isomer and the major population representing a local *trans*-isomer of peptidyl-prolyl bond (Weiss et al., 1998; Lu and Zhou, 2007). With prolines capping S^{76}–S^{90}, an increased proportion of the *cis*-isomer in one or more of these Pro residues could give rise to two distinct chemical environments through the entire length of the α-helix region. In addition, examination of other residues which neighbour prolines (*e.g.* A^{92}–G^{98}) reveal a similar pattern of major and minor population peaks, confirming that proline isomerisation is a possible mechanism behind this effect.

Another possible mechanism that could cause two distinct chemical environments is localised dimerisationof N-myc TAD or formation of higher order oligomer. The sequence of the S^{76}–S^{90} region contains a number of hydrophobic amino acids: Leu, Met, Trp and Val which could drive dimerisation through water expulsion mechanism (Radivojac et al., 2007). The visual inspection of N-myc$_{64\text{-}137}$ following the collection of NMR data of 300 μM sample did not reveal precipitation which could indicate aggregation. There is no evidence in the literature that N-myc TAD can dimerise or oligomerise, apart from forming dimer of heterodimers with its partner Max, through its C-terminal DNA binding domain (Nair and Burley, 2003). Recording ^{1}H–^{15}N HSQC spectra for N-myc$_{64\text{-}137}$ at two different concentrations (300 μM vs 150 μM) and the comparison of major-to-minor H-N correlated peak heights indicate the mechanism driving these two distinct chemical environments is not concentration-dependent. Z-Z exchange experiment showed no chemical exchange between the two species, indicating any oligomerisation taking place would be a very slow process exceeding seconds timescale. Slow oligomerisation processes can occur and are not uncommon and is seen for example in the formation of α-synuclein fibres (Ghosh et al., 2018).

Chemical exchange between random coil and α-helix, was a hypothesised mechanism behind the presence of two distinct chemical environments, since it is commonly seen in IDPs and the region affected has been shown to have helical propensity (Dyson and Wright, 2005). The data collected however does not support this hypothesis either. Z-Z exchange mechanism shows no signs of chemical exchange. The analysis of R_2 data of minor peaks reveal a species that relaxes more slowly, implicating that this species is smaller in size. This could be explained either by the presence of monomer – dimer, or by the presence of smaller N-myc construct in NMR reaction. Z-Z exchange experiment indicates no exchange at time scales where oligomerisation usually takes place. In addition, the intensity of the peaks was independent of the concentration. Mass spectrometry of N-myc$_{64\text{-}137}$ indicating that there is only a single species present.

The phenomenon observed behind the presence of these two species requires further investigation. Careful analysis of ^{1}H-^{15}N HSQC of N-myc TAD features minor population peaks for some of the residues, *e.g.* W^{77}, V^{78}, T^{79}, however for the majority of the peaks participating in the helix, the second minor population are not observed. The shorter construct has more favourable NMR dynamics with longer T_2 and this translates into better signal-to-noise ratio, allowing to observe the minor population peaks. The N-myc TAD, due to its size, has less favourable signal-to-noise ratio and thus the resonant frequencies of the minor population peaks are practically undetectable. If this is the case, this would likely indicate that these peaks experience the dual chemical environment, which could be associated with the helical propensity and not the result of proteolytic degradation.

Whichever mechanism gives a rise to these two distinct chemical environments, it is possible that it is occurring on longer time scales, than the ones probed through NMR. To exclude the possibility of α-helix formation, NOESY experiment could be conducted. α-helix would give a raise to cross-peaks of non-neighbouring residues, arising from the presence of this secondary structure. To observe the behaviour of prolines, ^{13}C- and ^{15}N-detected spectra could be used to directly establish the presence of *cis* and *trans* peptidyl prolyl bonds in this smaller construct. It is likely that each of the processes interrogated (proline isomerisation, oligomerisation and chemical exchange between α-helix and a random coil) are not mutually exclusive and can all contribute to the distinct chemical environment experienced by these residues.

5 N-myc TAD as a kinase substrate

5.1 Introduction

5.1.1 Methods to study phosphorylation

PTMs are very important in the context of IDPs. PTMs can alter interactome profile, induce profound conformational changes within the IDP, signal for different functional outcomes, modulate signalling, amongst other mechanisms (Bah and Forman-Kay, 2016). Myc proteins are subjected to extensive PTMs that lead to protein degradation, however the hierarchy and the importance of these PTMs is still an open debate in the field. Phosphorylation of serines/threonines is one of the most common and most studies PTM, and it features heavily in regulating myc protein stability (Ramazi and Zahiri, 2021). This chapter interrogates known and putative phosphorylation mechanisms that regulate the stability of N-myc protein.

There are several methods to study phosphorylations. Universal kinase assays measure the amount of phosphorylation product, the amount of ATP used, or the amount of ADP formed, and detection is usually coupled with signal amplification, for example through fluorescence. While this approach quantifies kinase activity regardless of the type of the enzyme and substrate used, it is limited in determining modification sites and the number of modifications. Detection of specific phospho-sites through immunoblotting requires highly specific antibodies, which are costly and may be impacted by cross-reactivity, especially if phospho-sites are proximal to each other. Immunoblotting techniques also become labour-intensive if multiple modifications are investigated. Mass spectrometry can provide in-depth analysis of phosphorylation of respective sites. Intact mass spectrometry can inform of the number of modifications, and protease digest allows peptide mapping of these phospho-sites. The drawback of the technique is the identification relies on the presence of the protease sites, and if multiple samples are analysed, the cost of this experiment can become prohibitive. NMR spectroscopy allows for real-time observation of phosphorylation reaction. In ^{1}H-^{15}N HSQC, each residue has a single peak of specific resonant frequencies of correlated amide and amide protons (apart from prolines, asparagines, glutamines and tryptophans). Phosphorylation of serines, threonines and tyrosines brings about very characteristic changes in the resonant frequency of these peaks, with ^{1}H δ of ~9.2 and ^{15}N δ of 118 – 119, referred to in this work as phospho-species region (Bienkiewicz and Lumb, 1999). The complete backbone assignment of the N-myc TAD allows us to analyse enzymatic reactions on per-residue basis. In addition, in contrast to other spectroscopic methods, NMR delivers a very small energy to induce transition between the energy states, and thus a sample remains intact, allowing for multiple reactions to be interrogated in a sequential fashion.

5.1.2 Myc proteins and PTMs

As described in detail in the introduction, the main PTM that govern N-myc stability is hierarchical phosphorylation of S^{62} and T^{58} by Cdk1:cyclin B and GSK3, respectively (Sears et al., 2000; Yeh et al., 2004), however other sites have also been suggested as potential phospho-sites. One other residue that is listed in PhosphoSitePlus® database, documenting phospho-sites and is coming from phosphoproteomic data, is S^{64}, putatively phosphorylated by Aurora A kinase (Zhou et al., 2013). S^{64} lies within the conserved MBI region, close to well-characterised S^{62} site. Another less interrogated putative phospho-site is T^{43}, phosphorylated by Plk4 (Byrne et al., 2020). This site is positioned immediately next to P^{44} site, which is mutated in a sub-set of N-myc-driven cancers (P^{44L}; Williamson et al., 2005; Tate et al., 2019). As Plk4 is a proline-directed kinase, it is possible that this mutation abrogates Plk4-specific phosphorylation which might be tumour-suppressive in nature.

5.1.3 Signalling network involved in myc stability

Upon undergoing a series of phosphorylations and other PTMs, c-myc is subjected to rapid degradation by SCF^{Fbw7} complex, to prevent uncontrolled myc-driven signalling. To bypass this regulatory mechanism cancer cells commonly harbour mutations in the myc degradation pathway, that either directly or indirectly prevents myc protein degradation.

Two kinases, Aurora A and Plk1, are both known to be important in preventing the degradation of myc protein. Aurora A becomes activated in late G_2 with its activity peaking in mitosis, performing pleiotropic functions including centrosome separation, mitotic entry, bipolar-spindle assembly, chromosome alignment, amongst others (Marumoto et al., 2005). The full activation of Aurora A is complex and depends on subcellular context and cell cycle time point. However, likewise to other kinases, the phosphorylation of T^{288} within the activation segment in *cis* enhances catalytic activity of Aurora A by 15 fold (Tavernier et al., 2021). Active Aurora A then phosphorylates Plk1 on position T^{210}, contributing to its activation (Combes et al., 2017). Plk1 has an N-terminal bilobal kinase domain and a C-terminal PBD, composed of two highly homologues polo box motifs (Seong et al., 2002). The PBD is important for phospho-peptide recognition of Plk1 substrates and directing Plk1 subcellular localisation (Elia et al., 2003). Humans have four paralogues of Plk: Plk1, Plk2, Plk3 and Plk4. Whereas the three former ones share a similar architecture of PBD, Plk4 is characterised by pseudo-PBD with only one polo box present (Park et al., 2010).

Plk1 is controlled both spatially and temporarily and its expression peaks at M stages of cell cycle. Similarly to Aurora A, Plk1 has pleiotropic functions within M stage of cell cycle, including centrosome maturation, bipolar spindle formation, mitotic entry and cytokinesis, amongst others (Park et al., 2010; Archambault and Glover, 2009).

5.1.4 Hierarchy of N-myc degradation prevention

Interactions between N-myc, Aurora A, Plk1 and Fbw7, reveals a complex network of N-myc regulation. From the literature, Aurora A antagonises the binding of Fbw7 to N-myc, thus reducing the probability of SCF^{Fbw7}- mediated degradation, by binding to the region just C-terminal of the phosphodegron in phospho-independent manner (Figure 5.1; Richards et al., 2016). There is also another layer of signalling in which Aurora A, acting together with Plk1, contributes in an indirect fashion to the stability of N-myc. Plk1 is activated by phosphorylation of T^{210} residue by Aurora A:Bora complex (Macůrek et al., 2008). In this mechanism Bora brings together Aurora A and Plk1 within the proximity of each other. Plk1 then phosphorylates Fbw7 and promotes its autoubiquitination and proteasomal degradation, however the authors did not assess the ability of Plk1 to bind to N-myc (Xiao et al., 2016). Plk1 phosphorylates [D/E] – [X] – [p-S/T] – [Φ] – [X] – [D/E] motifs and this sequence is absent from N-myc TAD (Alexander et al., 2011; Dou et al., 2011; Hegemann et al., 2011; Nakajima et al., 2003; Heinrich et al., 2013; Kettenbach et al., 2011) and therefore this region is unlikely to be a Plk1 substrate. This simplified network illustrates intricate dependencies that regulate myc protein stability and gives an idea how cancer cells could potentially synergistically enhance myc protein stability indirectly, through disruption of any of these components (Figure 5.1). It highlights the necessity of studying these interactions both independently and in tandem, to pharmacologically target myc protein stability.

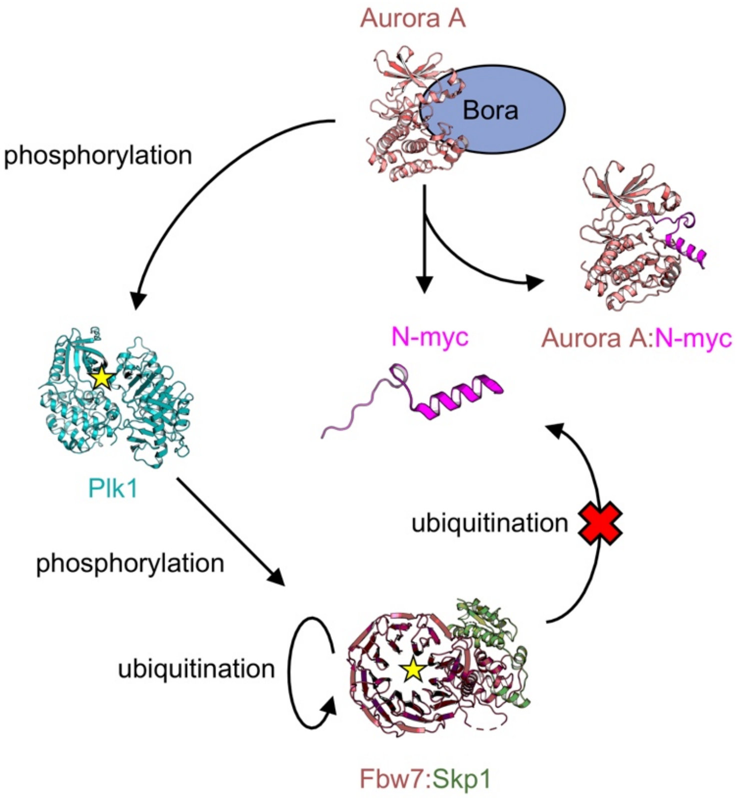


Figure 5.1. Proposed model of mechanism enhancing myc protein stability. Aurora A and Polo box domain kinase 1 (Plk1) work in concert to enhance stability of myc protein. Aurora A competes with Fbw7 for the binding to the region close to the phosphodegron region of N-myc. In addition, Aurora A, together with scaffolding protein Bora, also phosphorylates Plk1 on its activation loop which activates Plk1. Plk1 then phosphorylates Fbw7 causing its autoubiquitination and degradation. Yellow stars indicate phosphorylation (Richards et al., 2016; Xiao et al., 2016; Seki et al., 2008; Macůrek et al., 2008).

5.2 Results

5.2.1 Myc degradation pathway

5.2.1.1 N-myc as a substrate of Cdk's

Cdk1:cyclin B, which becomes activated at G_2/M transition, has been shown to phosphorylate S^{62} *in vivo* in N-myc (Sjostrom et al., 2005). To test this reaction in a simpler system, an *in vitro* phosphorylation assay was utilised. The assay was performed at 35 °C for 1 hour in 96-well plate with the reactions terminated by the addition of SDS-PAGE loading dye. The modification was monitored by observing the changes in the SDS-PAGE mobility between treated *vs.* untreated peptides. The addition of phosphate group causes retarded electrophoretic mobility and produces distinctive higher-molecular weight bands.

From the previous work performed in our group, 28-89 fragment of N-myc TAD was a sufficient region to form a stable crystallizable complex with Aurora A, therefore a panel of N-myc$_{28\text{-}89}$ peptides was used in the kinase assay. The peptides were phosphorylated at different positions: N-myc$_{28\text{-}89}$, N-myc$_{28\text{-}89}$ p-T^{58}, N-myc$_{28\text{-}89}$ p-S^{62}, N-myc$_{28\text{-}89}$ p-S^{64}, N-myc$_{28\text{-}89}$ p-T^{58} p-S^{62} (referred to as WT, p-T^{58}, p-S^{62}, p-S^{64} and p-T^{58} p-S^{62}, respectively; Richards et al., 2016) to monitor any phospho-site depended phosphorylations by Cdks. For example, Cdk1 is a proline-directed kinase involved in cell cycle progression and regulation and its phosphorylation of S^{62} acts as a priming event for phosphorylation of T^{58} by GSK3 (Sears et al., 2000; Yeh et al., 2004). The objective of this experiment was to assess the specificity and thus viability of *in vitro* studies and test other Cdk's ability to phosphorylate other sites surrounding a phospho-degron or other sites. In total three cell cycle-related kinases (Cdk1:cyclin B, Cdk2: cyclin A, Cdk4:cyclin D$_3$) and a single transcription-related kinase (Cdk9:cyclin T$_1$) were tested. The analysis of potential Cdk sites within N-myc sequence reveals in total of three [S/T] – [P] sites (T^{43}, T^{58} and S^{62}) and four additional phosphoryable residues that are not followed by proline in the sequence (S^{64}, S^{71}, S^{72} and S^{75}).

The three Cdks involved in the cell cycle phosphorylated N-myc to varying degrees. Cdk1:cyclin B readily phosphorylated N-myc peptides *in vitro*, regardless of pre-existing phosphorylations (left panel, Figure 5.2 A, + depicts kinase added and – depicts no kinase added). The phosphorylation was not limited to the S^{62} site, as evidenced by extra bands seen in p-S^{62} peptide. Cdk1:cyclin B targets the S^{62} site, however p-S^{62} peptide was also phosphorylated, highlighting the potential issue with the specificity of these kinases under the experimental conditions. Among all peptides treated with Cdk1:cyclin B, p-S^{64} peptide was the species that showed the most phosphorylation-related bands and with highest levels of motility retardation (two; Figure 5,2 A left panel).

Treatment of the peptides with Cdk2:cyclin A produces less phosphorylation than with Cdk1:cyclin B, albeit all of the peptides tested were still modified to some extent, even the p-T^{58} p-S^{62} peptide (Figure 5.2 A middle panel). The least efficient phosphorylation of the peptides is seen with Cdk4:cyclin D$_3$ with very little difference in band appearance between treated (+) and untreated (-) peptides (Figure 5.2 A, right panel). Without the presence of positive control in place it is difficult whether this is a kinase-specific effect or whether it is due to its low intrinsic catalytic activity.

An interaction between c-myc and Cdk9:cyclinT (pTEFb) has been documented in the literature, and it is thought to be mediated through cyclin T (Eberhardy and Farnham, 2003). It has not been determined whether this interaction exists in the context of myc acting as binding partner or as a substrate for this complex. Similar to Cdk1:cyclin B, the most substantial phosphorylation is seen on is p-S^{64}, indicating the addition of bulky negatively charged phosphates, makes other sites more accessible for kinases or it acts a priming event for both of the kinases, however literature documents no such case and the function of this site is obscure. S^{64}, which belongs to a conserved MBI, has been listed as a non-frequent missense mutation in blood cancers (Baluapuri et al., 2020; Bahram et al., 2000). The remaining peptides are phosphorylated to the similar extent, implicating multiple sites that can be targeted by this kinase.

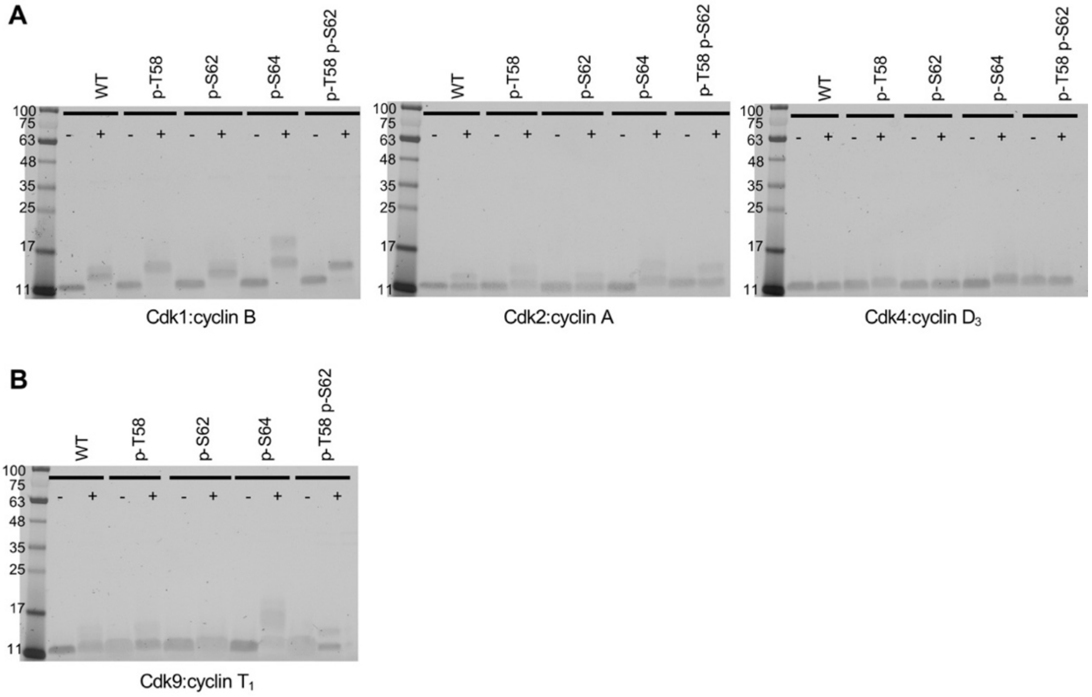


Figure 5.2. N-myc TAD phosphorylation by Cdk:cyclin complexes.
A. SDS-PAGE of N-myc$_{28\text{-}89}$ peptides treated with Cdk1:cyclin B, Cdk2:cyclin A and Cdk4:cyclin D3. N-myc$_{28\text{-}89}$ peptides with different phospho-sites were treated *in vitro* with Cdks. Phosphorylation alters the motility of peptides, retarding their migration through polyacrylamide gel. **B**. SDS-PAGE of N-myc$_{28\text{-}89}$ peptides treated with Cdk9:cyclin T1 (pTEFb). - and + indicate untreated and treated peptides, respectively.

The lack of discrimination of phosphorylation events in this assay raises the questions of specificity of this assay *in vitro*. According to the literature, Cdk1:cyclin B should be phosphorylating all peptides apart from p-S^{62} and p-T^{58} p-S^{62}, as per *in vivo* data, however it appears that all peptides are phosphorylated by Cdk1 to some extent. Another drawback of this approach is lack of correlation between number of bands observed and number of phosphorylated sites, which prohibits any quantitative analysis. Overall, the assay indicated the lack of kinase specificity in *in vitro* studies, where kinases can phosphorylate sites that are sub-optimal.

5.2.1.2 Phosphorylation of S^{62} by ERK1

The above study has revealed the need to use a phosphorylation assay that would inform which sites specifically are targeted. and the relative efficiency of phosphorylation at each site. NMR is the best-suited technique for this, and the N-myc TAD backbone assignment allows to monitor the modifications in real time. Due to de-shielding caused by electronegative oxygens in phosphate groups, phosphorylated residues have characteristic resonant frequencies on ^{1}H-^{15}N HSQC spectrum, usually at ~9 ppm in ^{1}H dimension and ~120 ppm at ^{15}N dimension (Selenko et al., 2008; Bienkiewicz and Lumb, 1999). In addition, phosphorylation alters the chemical environment of neighbouring residues which is realised in the changes in peak intensities or in chemical shift perturbations (CSPs) of neighbouring residues (Selenko et al., 2008). CSP is calculated using the formula

$$CSP = \sqrt{(1H\delta_{REF} - 1\mathrm{H}\delta_{OBS})^2 + ((15N\delta_{REF} \; - \; 15N\delta_{OBS})/6.5)^{2^{\frac{1}{2}}}},$$

where δ_{REF} and δ_{OBS} is chemical shift of the reference and altered frequencies. The reference spectra were collected at 10 °C using a standard 5 mm NMR tube, prior to the addition of 0.3 μM of desired enzymes, after which the reactions were incubated for varying amount of time at 25 °C and ^{1}H-^{15}N HSQC spectra was recorded at 10 °C to achieve optimal signal-to-noise ratio. NMR buffer was supplemented with 4 mM $MgCl_2$ and 1 mM ATP added freshly into the reaction. All enzymes were either made in-house or obtained commercially from MRC PPU Reagents and Services.

The goal of this segment of work, was to establish the degradation pathway *in vitro* and to monitor these reactions in real time. The second objective was to obtain di-phosphorylated N-myc (p-T^{58} and p-S^{62}) to be used in the interaction studies with Fbw7. Because of the previous data indicating the lack of specificity of Cdk1:cyclin B towards N-myc *in vitro*, ERK1 was used instead. ERK1 is a proline-directed kinase which becomes stimulated by mitogenic signals and it phosphorylates specifically S^{62} in c-myc, both *in vivo* and *in vitro* (Sears et al., 2000; Yeh et al., 2004). Figure 5.3 A illustrates ^{1}H-^{15}N HSQC of S^{62} and the residues surrounding this site, prior to the addition of the kinase (pink) and post-incubation with ERK1 (brown). Only the residues that undergo a significant CSP (above 0.05) were shown in the Figure 5.3

A. S^{62} becomes fully phosphorylated within 45 mins (time needed for ERK1 incubation and the collection of the spectrum). S^{62} undergoes CSP and its resonant frequencies change to the frequencies that are characteristic of phosphorylated residues (downfield shift in the ^{1}H dimension of ^{1}H-^{15}N HSQC spectrum). Residues surrounding this site: L^{61}, S^{64}, R^{65} and G^{66} experience significant CSPs, indicating that their chemical environment has changed, due to the addition of the phosphorylation on the nearby residue (Figure 5.3 B). There were some other observable CSPs that were not associated with ERK1 activity directly and were positioned closely to histidine residues (H^{70} and H^{136}, Figure 5.3 B). These changes are most likely due to changes in protonation of histidines as pH in the reaction decreases due to ATP hydrolysis.

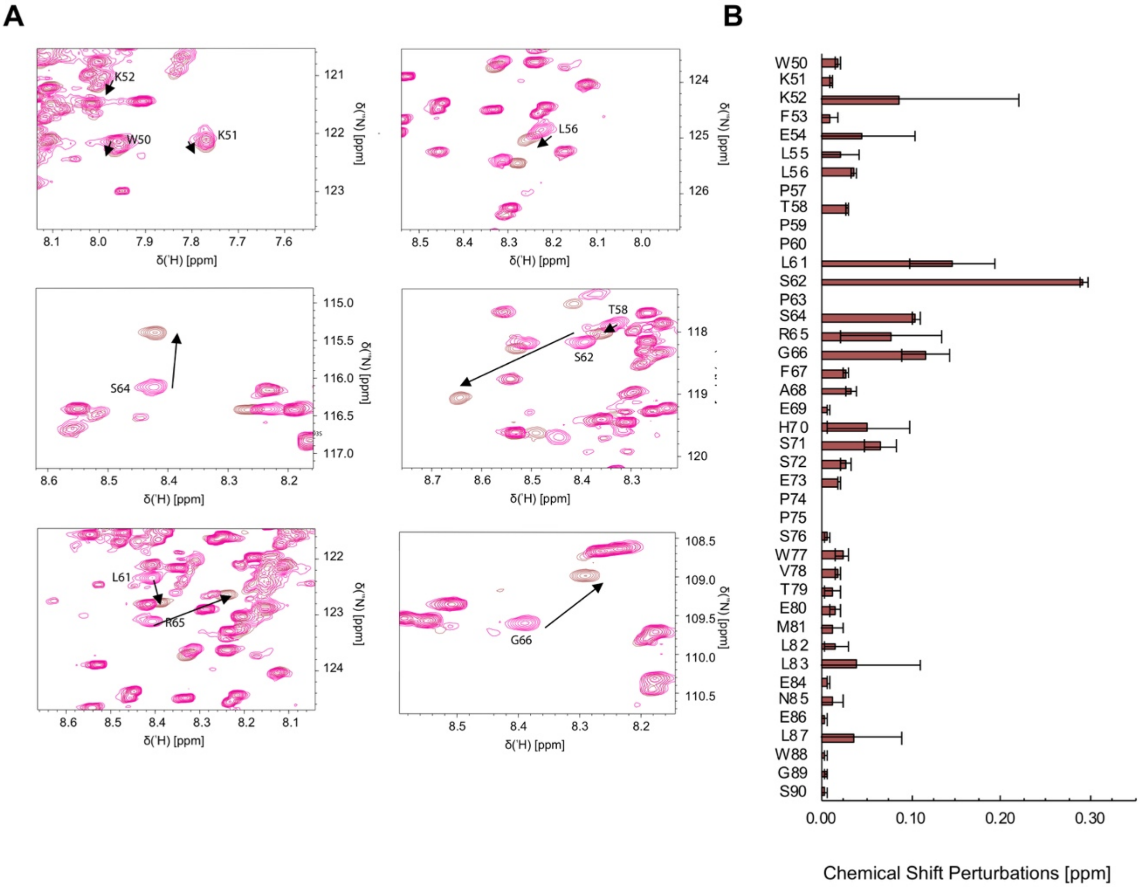


Figure 5.3. Phosphorylation of N-myc TAD with ERK1.
A. Overlay of N-myc TAD ^{1}H-^{15}N HSQC of reference spectrum (pink) and N-myc TAD spectrum following 45 min incubation with ERK1 (brown). **B.** Average CSPs of residues spanning W^{50} – S^{90}, calculated from 4 experiments. The remaining residues within N-myc TAD sequence did not experience significant CSPs. Error bars indicate standard deviation, which was calculated from four biological replicates.

There are five sites within N-myc TAD that either meet the criteria for the ERK1 minimal consensus site (p[S/T] – [P]) or ERK1 optimal consensus site ([P] – [L] – p[S/T] - [P]), where p stands for the residue targeted for phosphorylation by ERK1 (black boxes; Figure 5.4 A; Gonzalez et al., 1991). We therefore examined if these sites were targeted. CSPs and changes to peak intensities suggest that sites T^{43}, T^{58} and T^{103} were also partially phosphorylated, albeit at slower time scales. Unlike the p-S^{62} site, which was characterised by CSPs of neighbouring residues, T^{43} underwent diminishing of the peak intensities over the course of 3 hours 30 mins with concomitant resurgence of a new peak localised in the phospho-species region of ^{1}H-^{15}N HSQC (Figure 5.4 B and C). The neighbouring residues underwent modest CSPs and peak intensity diminishing (Figure 5.4).

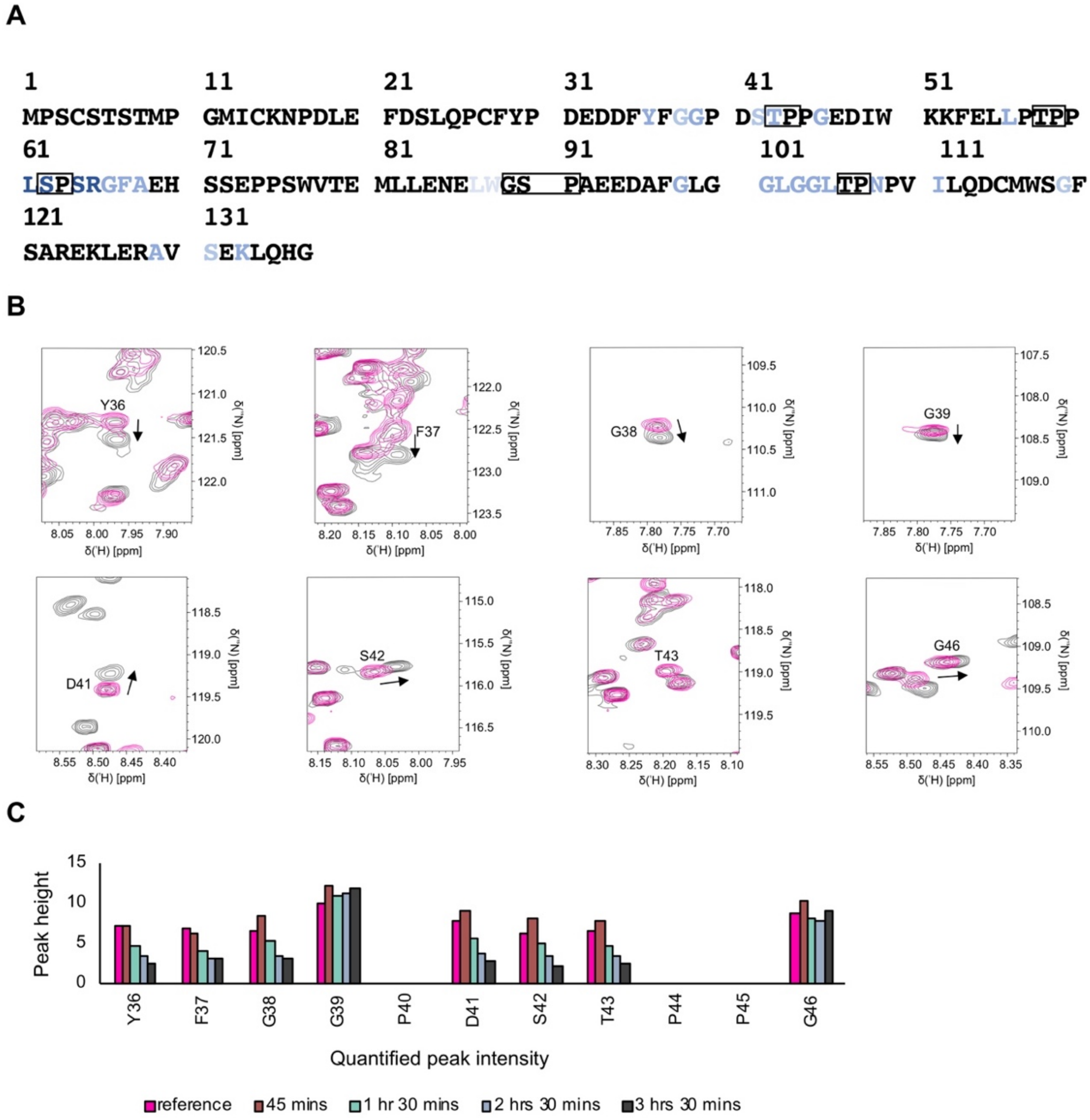


Figure 5.4. Phosphorylation of T^{43} site of N-myc TAD.
A. Amino acid sequence of N-myc TAD with the highlighted consensus sites for ERK1 (black rectangles). Residues in blue indicate sites with observable CSPs or diminishing peak intensities. **B.** Overlay of the reference (pink) and endpoint (grey) ^{1}H-^{15}N HSQC of ERK phosphorylating Y^{36} – G^{46} of N-myc TAD. The endpoint was collected after 3 hrs 30 mins of the incubation with ERK1. **C.** Quantified peak intensity of each time-points (reference – pink, brown– 45 mins, green – 1 hrs 30 mins, blue – 2 hrs 30 mins, grey – 3 hrs 30 mins).

From visual inspection of ^{1}H-^{15}N HSQC it was evident that T^{106} was also phosphorylated. The quantification of CSP or changing peak intensity for this site was not possible due to peak overlap with other residue (data not shown), however the quantification of these two parameters of residues G^{98} – I^{111} indicates changes in their chemical environment (Figure 5.5 A). These peaks are not characterised by uniform diminishing of the peak intensities as T^{43} site, but they undergo moderate CSP over the time course investigated (Figure 5.5 B).

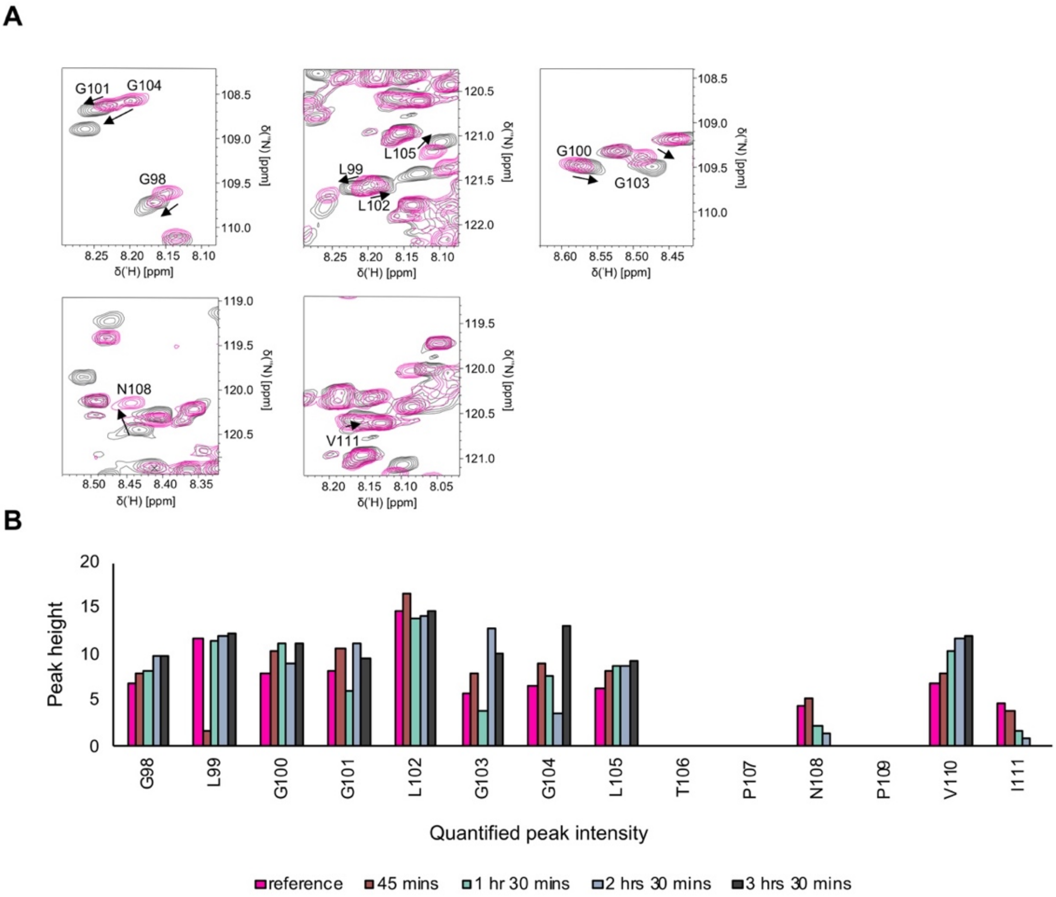


Figure 5.5. Phosphorylation of T^{106} site.
A. Overlay of the reference (pink) and end-point (grey, following 3 hrs 30 mins incubation) ^{1}H-^{15}N HSQC of ERK1 phosphorylating G^{98} – I^{111} of N-myc TAD. **B.** Peak height quantification at each time-point investigated (reference – pink, brown – 45 mins, green – 1 hrs 30 mins, blue – 2 hrs 30 mins, gray – 3 hrs 30 mins).

5.2.1.3 Phosphorylation of N-myc TAD by GSK3

The next step in the myc degradation pathway is phosphorylation of T^{58} by GSK3 family of enzymes. As discussed previously, GSK3 is unique amongst kinases, as it requires a priming phosphorylation in +4 residue to achieve its full activation (Fiol et al., 1987). In the context of myc proteins, GSK3 phosphorylates T^{58}, following S^{62} phosphorylation by mitogen-stimulated kinases (Sears et al., 2000; Yeh et al., 2004).

Initial reports of GSK3 targeting this site came out in the late 80's and early 90's when recombinant protein techniques where not commonly used, and enzymes were extracted from tissues, often cross-contaminated with other isoforms (Sears et al., 2000; Cohen and Frame, 2001; Pulverer et al., 1994; Lutterbach and Hann, 1994; Gregory et al., 2003b). To address the question which specific isoform of GSK3 targets this site *in vitro*, we used recombinant GSK3α and GSK3β separately. *In vitro* phosphorylation of T^{58} was done by addition of 0.3 μM of GSK3β to a sample of N-myc TAD that was first phosphorylated on S^{62} by ERK1, and the subsequent reaction was monitored using NMR. Figure 5.6 shows the overlay of ^{1}H-^{15}N HSQCs, with p-S^{62} spectrum used as the reference spectrum (brown). It was impossible to analyse the behaviour of T^{58} peak due to its overlap with S^{7} residue, however the analysis of the neighbouring residues (W^{50}, K^{51}, K^{52}, F^{53}, E^{54}, L^{56}, L^{61}) revealed modest CSP, apart from L^{56} residue, which experienced moderate CSP (Figure 5.6 A and B). As ERK1 continues its catalytic activity, it was difficult to discern whether the changes observed were the result of ERK1 or GSK3 activity.

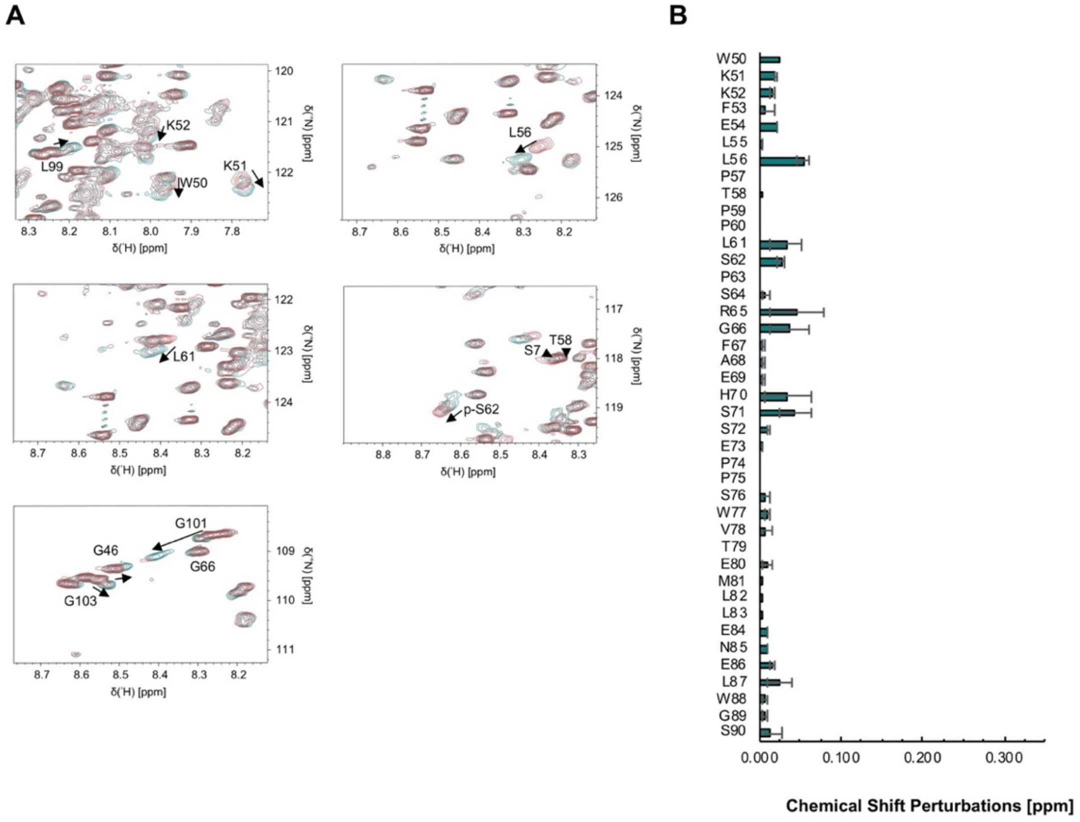


Figure 5.6. Chemical shift perturbations (CSPs) as the result of GSK3β activity. A. Overlay of reference ^{1}H-^{15}N HSQC spectrum (brown, following N-myc TAD phosphorylation by ERK1) and the spectrum following GSK3β phosphorylation (teal). **B.** Average of CSPs calculated from three biological repeats of residues W^{50} – S^{90}. Error bards indicate standard deviation.

It was apparent that even if this site was being phosphorylated by GSK3β, the level of phosphorylation is less in comparison to ERK1 and no other sites were targeted by GSK3β. Even with longer incubations times (5 hrs) there was no further changes to the spectrum, apart from phosphorylation by ERK1, as evidenced by CSPs observed which very similar to CSPs observed with ERK1 only. Due to the technical limitations of this assay, ERK1-mediated phosphorylation could not be quenched during the incubation with GSK3β, thus it was likely that ERK1 continued its catalytic activity on N-myc TAD.

To address the peak overlap of S^7 and T^{58}, a new N-myc TAD construct was designed and purified that incorporated an S^{7A} mutation. ^1H-^{15}N HSQC of N-myc TAD^{S7A} overlapped with N-myc TAD, except for the mutation site and its surrounding residues (Figure 5.7 A). N-myc TAD^{S7A} was subjected to the same experimental procedures, however GSK3α was used instead as means to phosphorylate T^{58}. T^{58} underwent a significant CSP of T^{58} with the change of its resonant frequencies into the region of phospho-species (Figure 5.7 B). The new peak is broad and has a low signal-to-noise ratio, however the site seems to be completely phosphorylated as T^{58} peak completely disappears from the spectrum. Unlike in the case for the phosphorylation of S^{62} by ERK1, neighbouring residues (W^{50}, K^{51}, K^{52}, F^{53}, E^{54}, L^{56}, L^{61}) displayed very little change in their resonant frequencies or peak intensities (Figure 5.7 B and C) and were very similar in magnitude to calculated CSPs when GSK3β isoform was used. Considering similarity of CSPs in neighbouring residues, it is conceivable that GSK3β was also phosphorylating T^{58}, however due to the peak overlap with S^{7A} this was not clearly observed. It is also possible that GSK3β was less catalytically active than Gsk3α, or that the modest changes in the resonant frequencies could be caused by the abundance of proline residues surrounding T^{58}, making it sterically more distant than the residues surrounding S^{62} site.

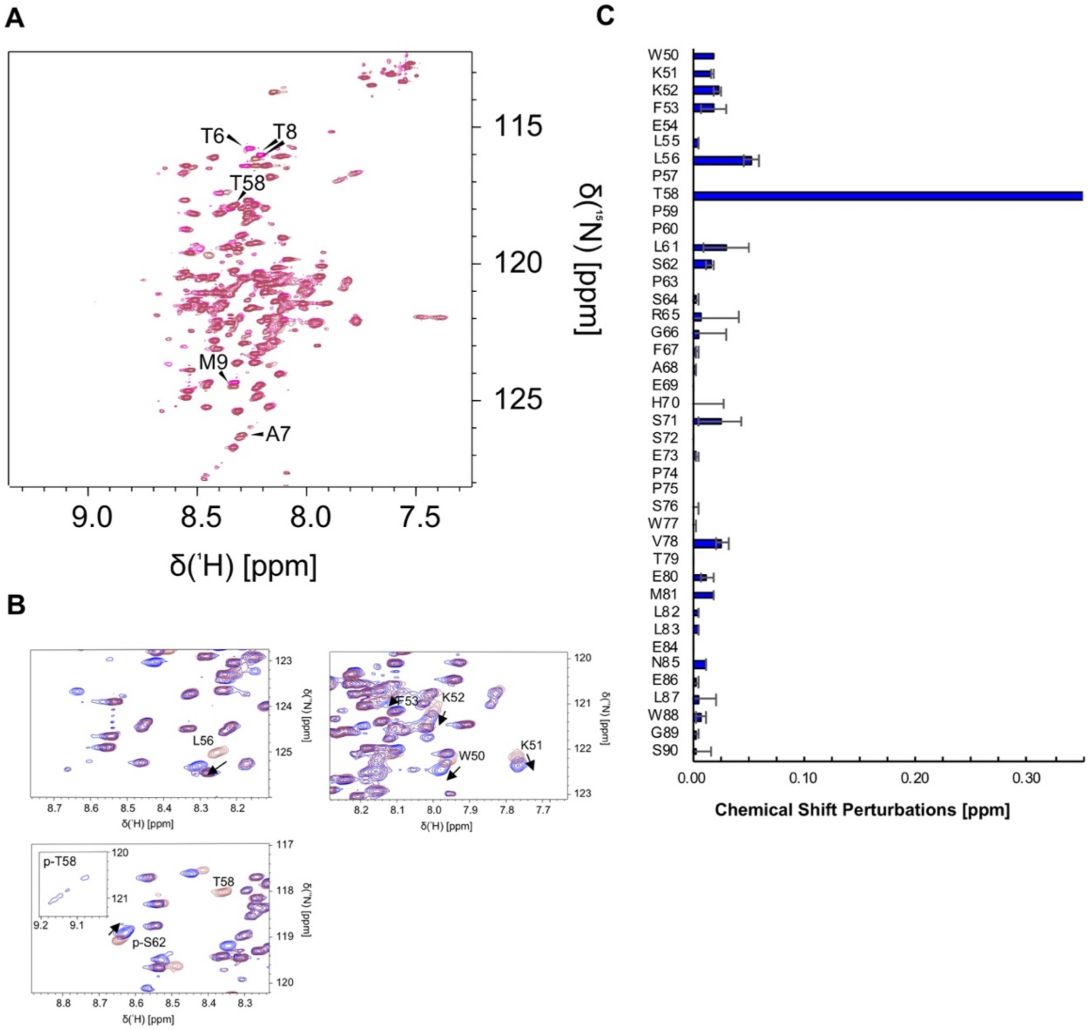


Figure 5.7. Phosphorylation of N-myc TAD by GSK3α
A. Overlay of ^{1}H-^{15}N HSQC spectrum of N-myc TAD (brown) and N-myc TAD^{S7A} (pink). The resonant frequencies are similar between two protein constructs with the exception of the residues surrounding the mutated site. **B.** Overlay of ^{1}H-^{15}N HSQC reference spectrum, following ERK1 phosphorylation of S^{62} (brown) and spectrum following GSK3α phosphorylation (blue). **C.** Quantification of CSPs following GSK3α phosphorylation. Error bars are standard deviation calculated from three biological repeats. T^{58} CSP is 0.45 but was clipped for the purposes of clarity.

To confirm that GSK3β requires priming S^{62} phosphorylation, phosphorylation reactions were set up in 96-well plates using unphosphorylated N-myc TAD as a substrate and the data was analysed using intact mass spectrometry. The data confirms that GSK3β is reliant on phosphorylation of S^{62} site for the efficient phosphorylation of T^{58} (Figure 5.8). There is low level of phosphorylation present after 3 hrs 30 mins, with a species of molecular weight 80 Da higher than unphosphorylated N-myc TAD. After 7 hrs there is no increase in the abundance of this phosphorylated species, indicating that the reaction has reached its limit, most likely to do exhaustion of ATP, as there is still some unphosphorylated species left. The low level of phosphorylation is consistent with overexpressed GSK3β being able to phosphorylate T^{58} *in vivo*, when S^{62A} mutation is present (Welcker et al., 2022).

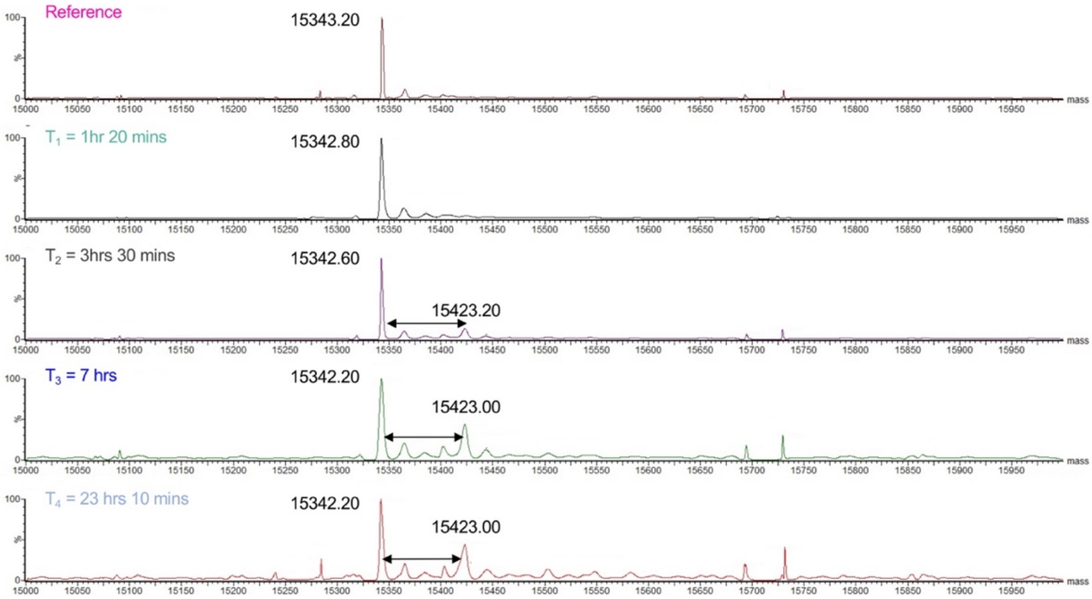


Figure 5.8. Intact mass spectrometry analysis of N-myc TAD phosphorylated with GSK3β.
Unphosphorylated N-myc TAD was incubated with GSK3β without prior phosphorylation of S^{62} and samples were incubated at 25 °C for a period of time. Specific time points which matched time points observed in NMR experiments (t=0, after 1hr 20 mins, 2 hrs 30 mins, 7 hrs and following overnight incubation) were interrogated by intact mass spectrometry. Courtesy of Rachel George.

5.2.1.4 Further steps in the N-myc degradation pathway

The subsequent PTMs in the myc degradation pathway are poorly understood. According to Yeh et al. (2004), following GSK3 phosphorylation of T^{58}, p-S^{62}-P^{63} bond becomes isomerised by Pin1, a prerequisite event to the dephosphorylation of S^{62} by *trans*-favouring PP2A (Sears et al., 2000; Yeh et al., 2004). To assess these PTMs *in vitro*, we generated di-phosphorylated N-myc TAD^{S7A} using the protocol used to monitor p-S^{62} and p-T^{58} phosphorylation using NMR. All enzymes were added to the final concentration of 0.3 μM, incubated at 25 °C and the spectra were collected at 10 °C. In-house made Pin1 (courtesy of Eoin Leen) was added first with ^{1}H-^{15}N HSQC recorded pre-and post Pin1 addition (purple and green, respectively; Figure 5.9 A). Although prolines are undetectable in ^{1}H-^{15}N HSQC spectrum, the conformational change caused by the bond isomerisarion would cause observable CSPs in neighbouring residues, in particular p-S^{62}. There was very little change between the two spectra (Figure 5.9 A). Pin1 might be catalytically inactive, all of the bonds present are in *trans* isoform already, or that this site is not the substrate for this enzyme.

Without appropriate controls it was impossible to determine which of these hypotheses were correct.

We then sought to dephosphorylate p-S^{62} site. The literature indicates PP2A as the enzyme targeting this site, however as PP2A was not commercially available, we resorted to using its closest paralogue – PP1. Both enzymes belong to the same family of phosphatases, with three subunits: structural, regulatory and catalytic, named A, B and C, respectively (Moorhead et al., 2007). Each of these subunits have multiple splice variants and substrate specificity is conferred by the combination of those, therefore establishing a consensus site for these enzymes is very challenging (Virshup and Shenolikar, 2009) .

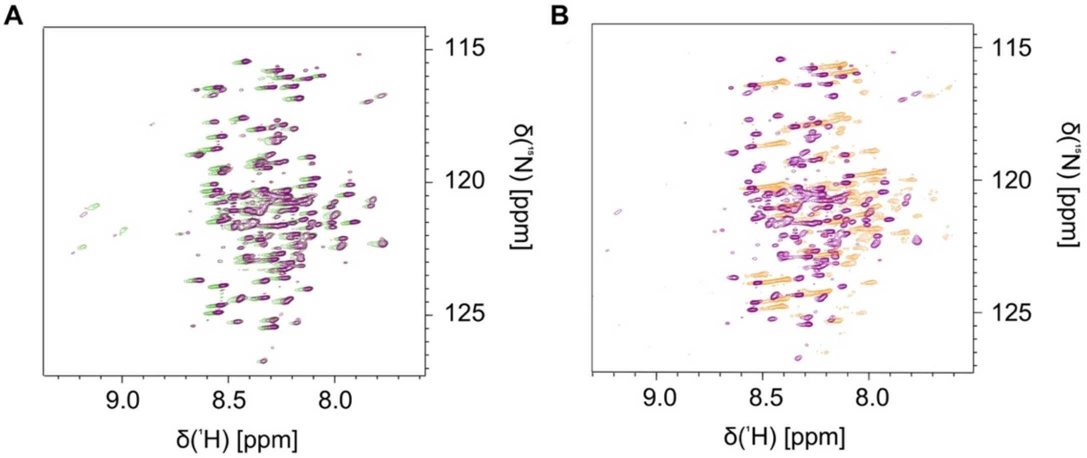


Figure 5.9. Further modifications of N-myc TAD.
A. Overlay of ^{1}H-^{15}N HSQC of ERK1 - and GSK3α-phosphorylated N-myc TAD^{S7A} (purple) and spectrum collected following the addition of Pin1 (green)**. B.** Overlay of ^{1}H-^{15}N HSQC of ERK1 - and GSK3α-phosphorylated N-myc TAD (purple) and spectrum following the incubation of PP2A (orange). Following Pin1 or PP2A addition, the protein was incubated at 25 °C for 30 mins and the spectra were acquitted at 10 °C for 42 mins using 600 MHz Oxford Instruments, equipped with QCI-P-cryoprobe (5mm).

The addition and incubation with PP1 rendered an atypical spectrum, with some peak intensities lost, distorted and broadened with a uniform upfield shift in both of the dimensions (Figure 5.9 B orange). This loss of sensitivity and signal usually indicates problems with a sample, albeit no precipitation within the NMR tube was observed in this instance. Detailed quantitative analysis of this spectrum was not possible, and it was difficult to assess the CSPs or changes in peak intensities. Overall, this approach was difficult as it required a series of PTM reactions with all the enzymes catalytically active at the end of the experiment. For example, enzyme cross-reactivity could be an issue. Our data indicates that ERK1 is a promiscuous enzyme in *in vitro* context, that could not only phosphorylate N-myc TAD, but could also target Pin1 and PP2A, rendering them inactive. Ideally, following each PTM, the specific enzymatic reaction should be quenched either by using a specific inhibitor or by removing the enzymes using SEC.

5.2.1.5 The impact of phosphorylation on the interaction with Fbw7-Skp1

Fbw7 has been demonstrated to be responsible for recognition of c-myc and N-myc phospho-degrons and mediate their ubiquitination and targeting them for proteosomal degradation (Yada et al., 2004; Sjostrom et al., 2005). The utilisation of NMR to monitor N-myc TAD PTMs was a reliable and reproducible system to obtain a di-phosphorylated N-myc TAD, which was required to test the interaction with Fbw7-Skp1. To establish the importance of individual PTMs, unphosphorylated, N-myc TAD p-S^{62} and N-myc TAD p-S^{62} p-T^{58} were generated. The reactions were stopped by snap freezing in liquid nitrogen. Analytical SEC was employed to test the ability of N-myc species to form a complex with Fbw7-Skp1. Analytical SEC was a method of choice because it does not involve the presence of any tags and the size of Fbw7-Skp1 is beyond the technical capabilities of NMR. Prior to analytical SEC, N-myc TAD phosphorylated at different residues was mixed in equivalent molar ratio with recombinant Fbw7-Skp1 (courtesy of Mohd Syed Ahanger) to the total volume of 100 μL. The proteins were incubated, by slow rotation in at 4 °C for two hours prior to SEC to allow for the equilibrium to be reached. Fractions were collected every 0.5 mL and analysed using SDS-PAGE.

Unphosphorylated N-myc TAD, N-myc $TAD^{p\text{-}S62}$ and N-myc $TAD^{p\text{-}T58\ p\text{-}S62}$, as well as Fbw7-Skp1, were analysed through analytical SEC without the addition of the second part of the complex to establish their normal elution profiles (Figure 5.10). Due to limited availability of the N-myc TAD samples, N-myc TAD^{S7A} was used in one of the runs instead of N-myc TAD. Fbw7-Skp1 is a large complex, prone to dimerisation, that elutes at 10.5 mL (Figure 5.10, blue trace). N-myc TAD and N-myc TAD^{S7A} have very similar elution profiles (Figure 5.10 pink *vs.* green trace) indicating that the amino acid substitution does not impact the volume of elution. The di-phosphorylated N-myc TAD (gray) elutes at lower elution volume and could be the reflection of change in a size of the protein (Figure 5.10). There was also an extra elution peak at ~17.3 mL elution volume, which was undetectable on SDS-PAGE gel. As the samples were snap frozen following the phosphorylation reactions, this is likely to be ATP that also absorbs strongly at 280 nm.

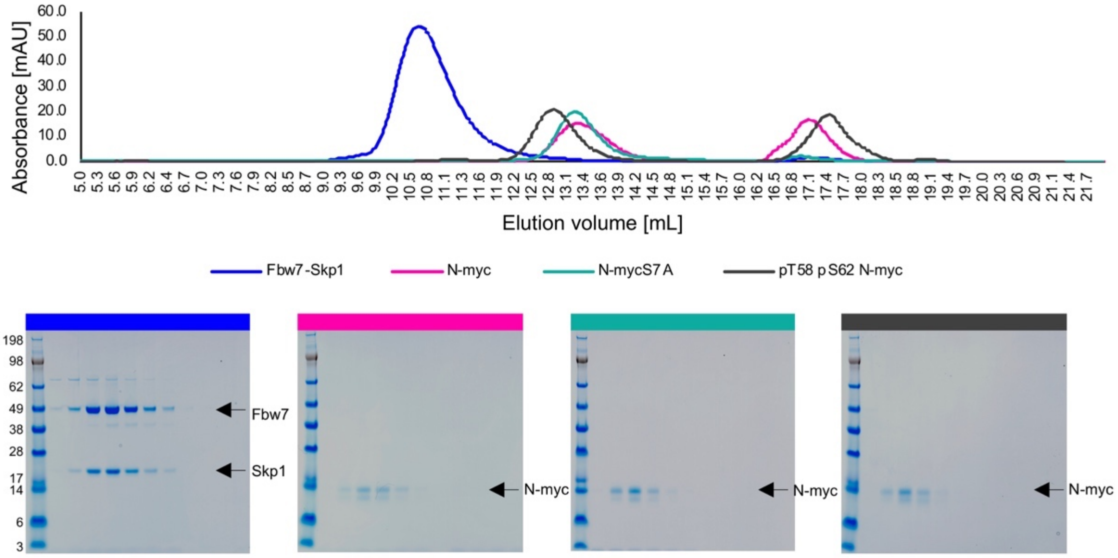


Figure 5.10. Control runs of analytical size exclusion chromatography (analytical SEC).
The colour of the traces correspond to the colours of SDS-PAGE images. Proteins were run individually on Superose 12 10/300 GL column. Peak fractions were collected at 0.5 mL and resolved on SDS PAGE and stained with Coomassie blue.

We then incubated N-myc TAD, N-myc TAD^{S7A}, N-myc TAD$^{p\text{-}T58\ p\text{-}S62}$ with Fbw7-Skp1 and tested their elution profiles (Figure 5.11). Unphosphorylated N-myc TAD does not form a stable complex with Fbw7-Skp1 (Figure 5.11 A, light green trace). Despite long incubation time, these proteins elute separately in analytical SEC, at exactly same elution volumes as individual proteins, which is also confirmed by their respective SDS-PAGE.

We then incubated N-myc TAD$^{p\text{-}S62}$ and Fbw7-Skp1 together. The chromatogram indicated these two species do not form a stable complex (Figure 5.11 B). There is a moderate shift in N-myc TAD$^{p\text{-}S62}$ elution profile (light blue) to lower elution volumes in comparison to N-myc TAD (pink), however Fbw7-Skp1 does not shift towards higher elution volumes (blue), indicating that there is no complex formation between the two species. This was confirmed by SDS-PAGE analysis where there was little overlap in fractions of elutions between N-myc TAD$^{p\text{-}S62}$ and Fbw7-Skp1. The interaction between the two species becomes apparent when N-myc TAD is di-phosphorylated on T^{58} and S^{62} (Figure 5.11 C) with N-myc peak almost completely disappearing (Figure 5.11 C, grey) and co-eluting with Fbw7-Skp1 (Figure 5.11 C, red). Interestingly, the complex between Fbw7-Skp1 and N-myc TAD$^{p\text{-}T58\ p\text{-}S62}$ elutes later than the Fbw7-Skp1 complex alone, suggesting that the interaction with N-myc TAD makes the complex more compact despite an increase in its molecular weight.

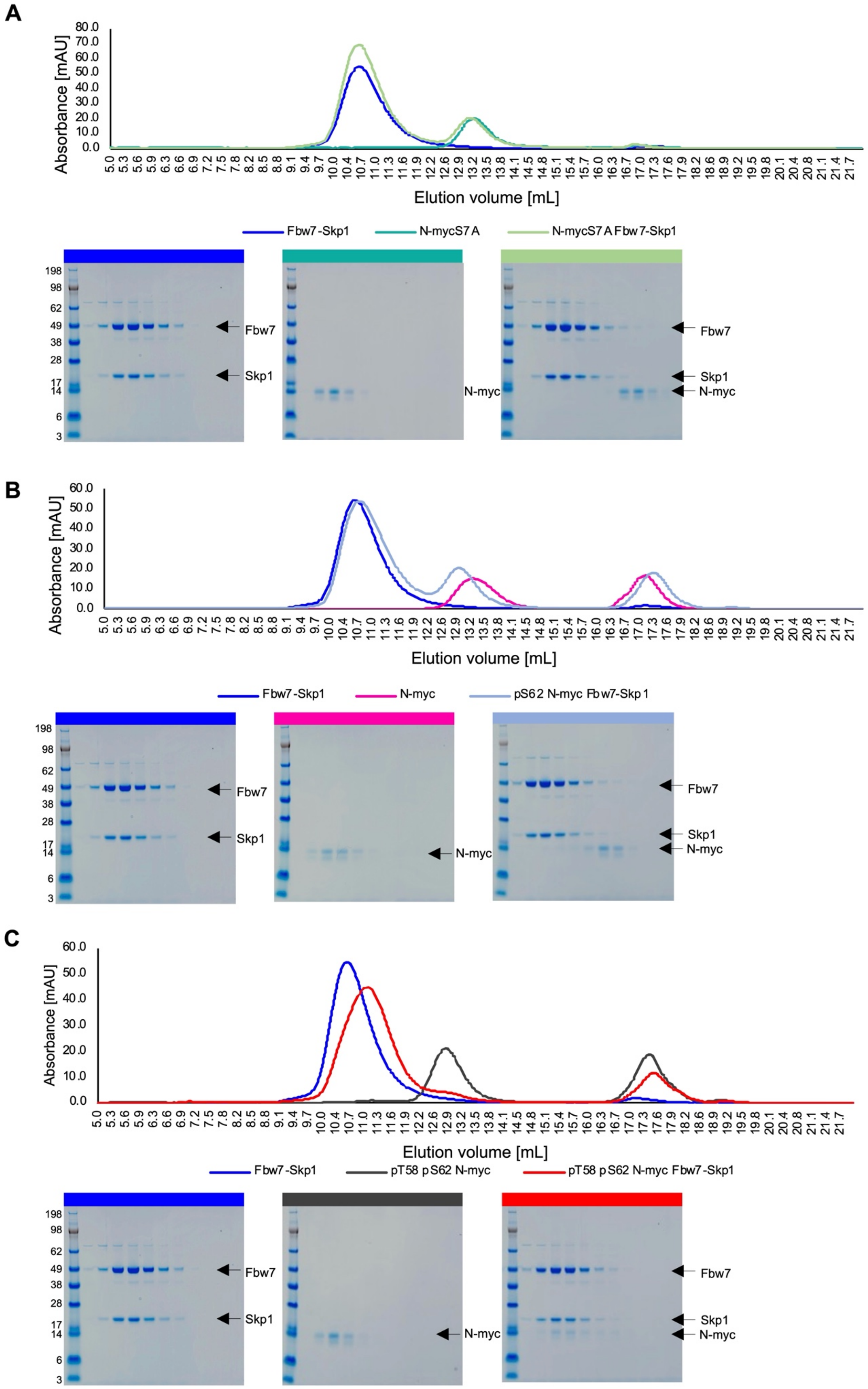

A
Absorbance [mAU]
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
5.0 5.3 5.6 5.9 6.3 6.6 6.9 7.2 7.5 7.8 8.2 8.5 8.8 9.1 9.4 9.7 10.0 10.4 10.7 11.0 11.3 11.6 11.9 12.2 12.6 12.9 13.2 13.5 13.8 14.1 14.5 14.8 15.1 15.4 15.7 16.0 16.3 16.7 17.0 17.3 17.6 17.9 18.2 18.6 18.9 19.2 19.5 19.8 20.1 20.4 20.8 21.1 21.4 21.7
Elution volume [mL]
Fbw7-Skp1
N-mycS7A
N-mycS7A Fbw7-Skp1
198
98
62
49
38
28
17
14
6
3
Fbw7
Skp1
N-myc
Fbw7
Skp1
N-myc
B
Absorbance [mAU]
60.0
50.0
40.0
30.0
20.0
10.0
0.0
5.0 5.3 5.6 5.9 6.3 6.6 6.9 7.2 7.5 7.8 8.2 8.5 8.8 9.1 9.4 9.7 10.0 10.4 10.7 11.0 11.3 11.6 11.9 12.2 12.6 12.9 13.2 13.5 13.8 14.1 14.5 14.8 15.1 15.4 15.7 16.0 16.3 16.7 17.0 17.3 17.6 17.9 18.2 18.6 18.9 19.2 19.5 19.8 20.1 20.4 20.8 21.1 21.4 21.7
Elution volume [mL]
Fbw7-Skp1
N-myc
pS62 N-myc Fbw7-Skp1
198
98
62
49
38
28
17
14
6
3
Fbw7
Skp1
N-myc
Fbw7
Skp1
N-myc
C
Absorbance [mAU]
60.0
50.0
40.0
30.0
20.0
10.0
0.0
5.0 5.3 5.6 5.9 6.3 6.6 6.9 7.2 7.5 7.8 8.2 8.5 8.8 9.1 9.4 9.7 10.0 10.4 10.7 11.0 11.3 11.6 11.9 12.2 12.6 12.9 13.2 13.5 13.8 14.1 14.5 14.8 15.1 15.4 15.7 16.0 16.3 16.7 17.0 17.3 17.6 17.9 18.2 18.6 18.9 19.2 19.5 19.8 20.1 20.4 20.8 21.1 21.4 21.7
Elution volume [mL]
Fbw7-Skp1
pT58 pS62 N-myc
pT58 pS62 N-myc Fbw7-Skp1
198
98
62
49
38
28
17
14
6
3
Fbw7
Skp1
N-myc
Fbw7
Skp1
N-myc

Figure 5.11. The impact of N-myc TAD phosphorylation on its interaction with Fbw7-Skp1.
A. Overlay of chromatograms of Fbw7-Skp (blue), N-myc TAD^{S7A} (teal) and N-myc TAD^{S7A} in complex with Fbw7-Skp1 (light green). Peak fractions were analysed by SDS-PAGE. **B.** Overlay of chromatograms of Fbw7-Skp1 (blue), N-myc TAD (pink) and N-myc TAD $^{p\text{-}S62}$ in complex with Fbw7-Skp1 (light blue) with the corresponding SDS-PAGE of peak fractions, seen below. **C.** Overlay of chromatograms of Fbw7-Skp1 (blue), N-myc TAD$^{p\text{-}T58\ p\text{-}S62}$ (grey) and N-myc TAD$^{p\text{-}T58\ p\text{-}S62}$ in complex with Fbw7-Skp1 (red). All proteins were incubated together in equimolar amounts for two hours prior to analytical SEC.

To determine the molecular weight of the complexes formed, analytical SEC standards (BioRAD) were run on the column using the same buffer and experimental set up. The elution profile of the standard is shown in Figure 5.12 A. The elution volumes of the standards, as well as protein complexes, were converted to average distribution constant (K_{AV}) according to the formula:

$$K_{AV} = \frac{V_e - V_0}{V_c - V_0}$$

where V_e is the volume of elution specific to each protein, V_0 is the void volume, specific to the column used and V_c is the total bed volume (Size Exclusion Chromatography, Principles and Methods, Cytvia, 2020). K_{AV} was plotted against the molecular weight in kDa and the regression line was plotted using Excel: Trendline operation (Figure 5.12 B, green dotted line). K_{AV} of investigated species was also plotted on the graph. Their molecular weight was then established using interpolation and the regression equation (Figure 5.12 B).

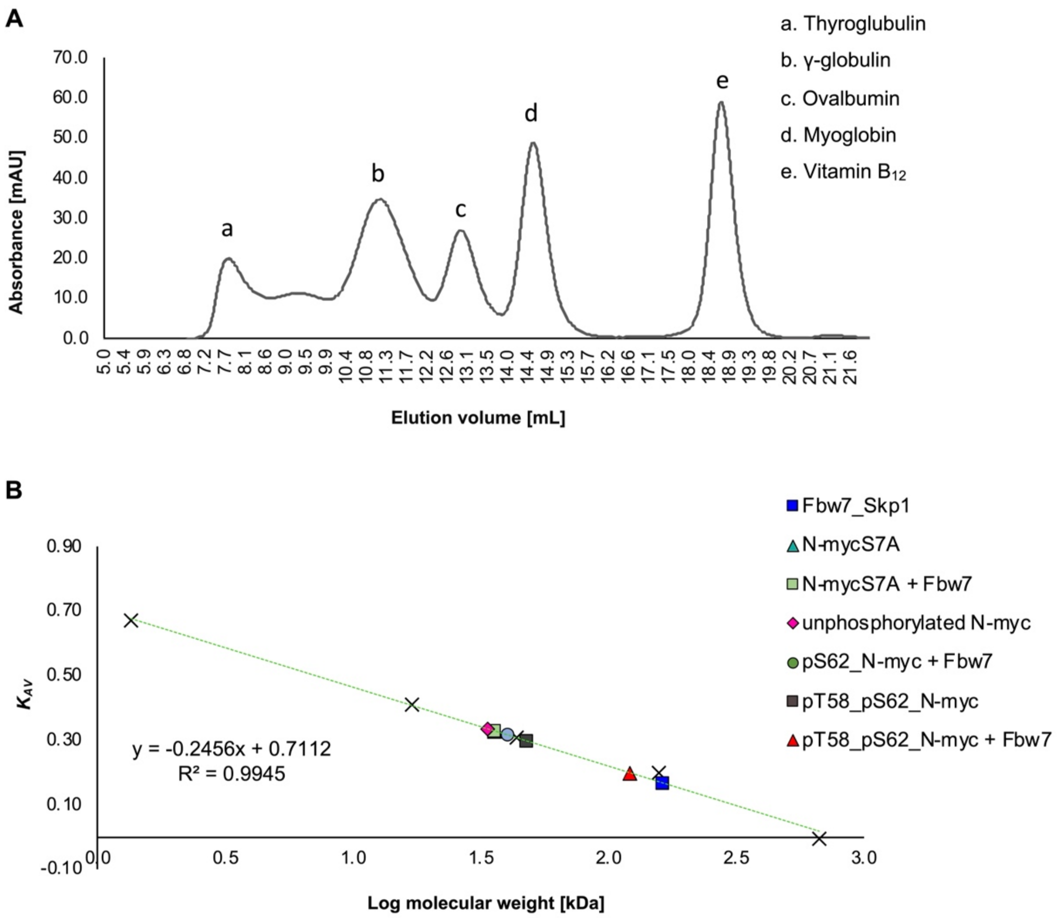


Figure 5.12. Molecular weights of Fbw7-Skp1:N-myc TAD complexes.
A. Chromatogram of analytical SEC of protein standards (BioRAD), with elution volume plotted against A_{280} absorbance. **B.** Calibration curve of protein standard plotted as log molecular weight against K_{AV} parameter (green) with the regression line equation. K_{AV} of the investigated species are also plotted and their molecular weights were calculated using the equation.

All of the N-myc species have similar molecular weight, estimated to be around 35 kDa, which is significantly higher than the expected weight of 17 kDa (Figure 5.12 B). This is typical for IDPs, as due to their lack of structure they are able to move faster through the stationary phase and elute at lower elution volumes (Graether, 2019). The molecular weight calculated from the regression line equation for the Fbw7-Skp1:N-myc TAD$^{pT58\ pS62}$ complex was estimated to be 122 kDa, whereas the calculated molecular weight of Fbw7-Skp1 is 163 kDa (Figure 5.12 B, red triangle and blue square). The actual molecular weight of the species should be 69 kDa for Fbw7-Skp1 alone and 84 kDa for Fbw7-Skp1:N-myc TAD, assuming a 1:1:1 stoichiometry.

Molecular weight standards are globular proteins; thus they are less geared for estimating weights of IDPs which have different hydrodynamic properties, and they escapes a typical SEC elution profile. In addition, the elution volumes of analytical SEC are the resultants of not only molecular weights but a hydrodynamic radius – thought of as a radius of the protein, caused by its shape, molecular weight and its hydration shell (La Verde et al., 2017). N-myc forms extended conformations with impact the elution profiles of protein that form a complex with it, which explains the discrepancy between the estimated molecular weight of Fbw7-Skp1:N-myc TAD complex and its weight following the estimation from the experimental data.

5.2.2 The regulation of N-myc degradation through phosphorylation

5.2.2.1 N-myc TAD as a substrate for Aurora A

Both Plk1 and Aurora A has been suggested as kinases that stabilise N-myc against proteasomal degradation. The crystal structure of Aurora A with N-myc indicates that Aurora A competes with Fbw7 for the binding to N-myc TAD in phosphorylation-independent manner and prolongs N-myc half-life (Richards et al., 2016). The consensus site for Aurora-A phosphorylation is [R/K/N] – [R] – [X] – [p-S/T] - Φ, where X is any amino acid and Φ is any hydrophobic amino acid apart from prolines (Ferrari et al., 2005) and this consensus sequence is not present in the N-myc TAD. The curated database of phosphorylation sites (PhosphoSitePlus®) indicates that S^{64} is phosphorylated by Aurora A, as evidenced by large scale proteomics data from cancer cells (Zhou et al., 2013; Wu et al., 2010). The co-crystal structure of Aurora-A and N-myc lacks electron density for this region, we therefore tested this putative phosphorylation using Aurora $A_{122\text{-}403}$ (courtesy of Selena Burgess and Matt Batchelor) using the same protocol as for N-myc TAD degradation pathway interrogation.

^{1}H-^{15}N HSQC data indicates that Aurora A does not phosphorylate this site, or any other site on N-myc TAD (Figure 5.13). Visual analysis of ^{1}H-^{15}N HSQC revealed no significant CSPs between the reference (black) and the spectrum recorded post-Aurora A addition (teal; Figure 5.13) and prolonged incubation for 3 hrs did not result in any changes in peak intensities or CSPs. The exact same batch of kinase was shown to be catalytically active by phosphorylation of TACC3 using a similar, NMR-monitored, kinase assay (unpublished data, Matt Batchelor).

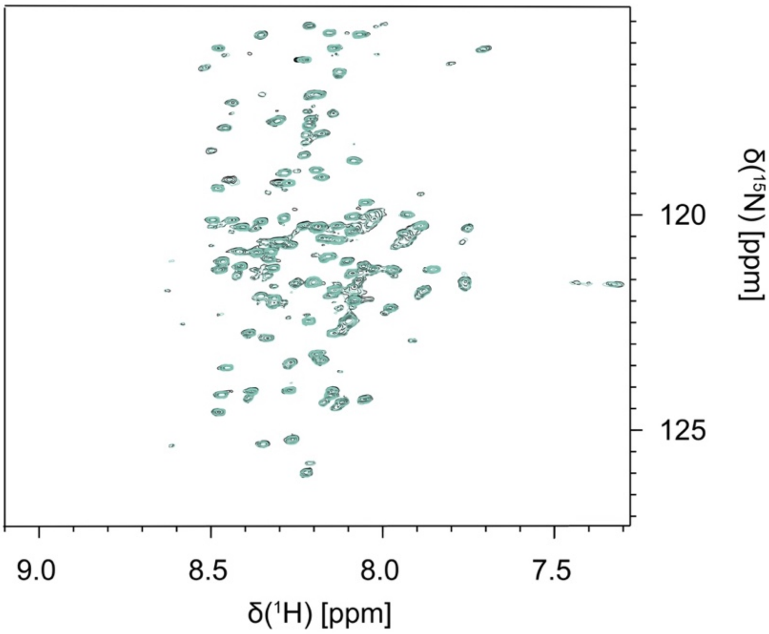


Figure 5.13. Investigating N-myc TAD as a substrate of Aurora A.
Overlay of ^{1}H-^{15}N HSQC reference spectrum (black) and the spectrum collected after 1 hr incubation with Aurora A (teal). Th spectrum was collected on 600 MHz Oxford Instruments, equipped with QCI-P-cryoprobe (5mm) with 120 μM of N-myc TAD and 0.3 μM of Aurora A at 10 °C.

Aurora A is a basophilic kinase which requires R/K residue in -4 position, the criteria not met by S^{64} site. This indicates that N-myc is not a substrate for Aurora A which is in agreement with the studies by (Richards et al., 2016).

5.2.2.2 N-myc TAD as a substrate for Plk1

Plk1 is another kinase implicated in the regulation of N-myc and myc protein stability, thus we tested whether Plk1 has an ability to phosphorylate N-myc TAD, despite a lack of its consensus sites within TAD sequence ([D/N/E/Y] –[X] – [S/T] – [F/ Φ; no P] – [Φ /X]; Alexander et al., 2011; Dou et al., 2011; Hegemann et al., 2011; Kettenbach et al., 2011; Nakajima et al., 2003). Plk1 has also been proposed to phosphorylate S^{62} site in c-myc, demonstrated by *in vivo* assays (Tan et al., 2013). We utilised in-house Plk1$_{13\text{-}345}$ T210V construct of Plk1 (courtesy of Eoin Leen). Despite the presence of non-phosphoryable valine in its activatory site, this construct remains catalytically active in comparison to WT levels (Kothe et al., 2007). NMR was utilised to monitor this reaction with ^{1}H-^{15}N HSQC recorded pre- and post-kinase addition. There was no apparent phosphorylation of any of the residues, as observed in ^{1}H-^{15}N HSQC experiments and the reference spectrum (black) and the spectrum collected following 1 hour incubation (red) were similar (Figure 5.14 A). Incubation for longer periods of time did not yield positive results and N-myc TAD remained unphosphorylated (data not shown). A few observable differences affected residues that are histidines or are in their proximity (data not shown), which is likely to be due to the change in histidine protonation following a change in buffer pH due to ATP hydrolysis.

To confirm the results using orthogonal techniques, we utilised ADPSensor^{TM} Universal Kinase Activity Assay Kit (BioVision Inc.) which detects fluorescence when its probes interact with ADP, indicating catalytic reaction. We tested the same batch of in-house Plk1 together with a well-characterised substrate of Plk1, casein and against N-myc TAD^{S7A} (Figure 5.14 B; Park et al., 2010). Plk1$_{13\text{-}345}$ T210V phosphorylates casein (blue circles; Figure 5.14), and the fluorescence recording after cycle 31 was above the detection range of the equipment (Figure 5.14 B, blue trace). We also observed fluorescence emission for Plk1$_{13\text{-}345}$ T210V with N-myc TAD^{S7A}, however the trace was flat and did not increase with the time of the reaction (red squares, Figure 5.14). The analysis of the raw data revealed high basal signal of the background control (no kinase present), which likely was due to the hydrolysis of ATP to ADP which occurs with time even without the presence of enzymatic reactions. Taken together with the negative results of NMR data and the lack of Plk1 consensus sites in N-myc TAD sequence makes N-myc TAD an unlikely candidate as a substrate for Plk1. The inclusion of the negative control (a protein that is not a target of Plk1) in the kinase assay would eliminate the possibility that an increase in emission in Figure 5.14 B is caused by the promiscuity of the kinase. Overall, the data suggests that if there is a phosphorylation taking place, it is at a low level in comparison to the positive control.

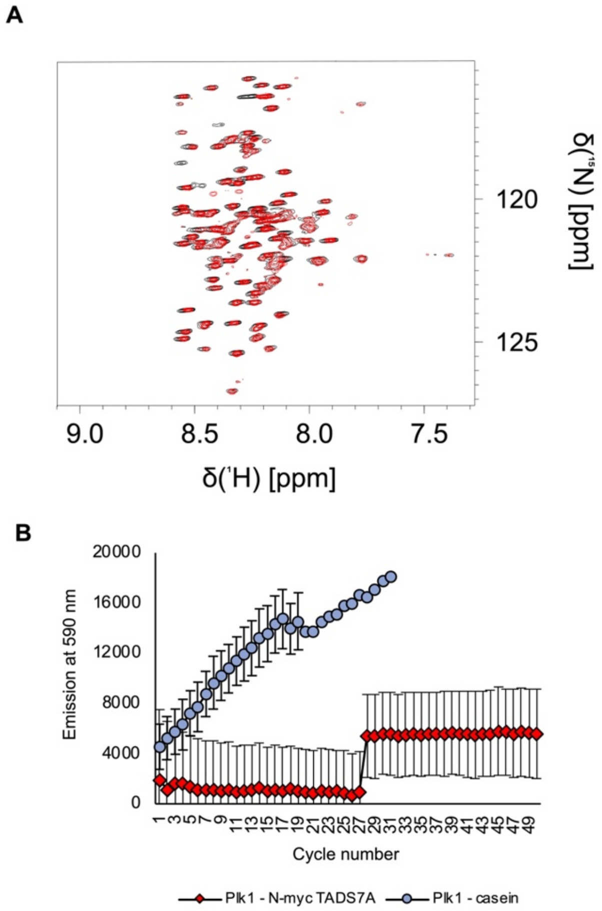


Figure 5.14. Investigating N-myc TAD^{S7A} as a substrate for Plk1.
A. Overlay of the reference ^{1}H-^{15}N HSQC (black) and the spectrum following the addition of Plk1 and the incubation for 1 hr (red). **B.** Kinase assay of Plk1 with casein (blue) and Plk1 with N-myc TAD^{S7A} (red). Error bars indicate standard deviation calculated from three biological repeats.

5.3 Discussion

5.3.1 N-myc TAD degradation signalling

Our data indicates that ERK1 phosphorylates S^{62} site of N-myc. The N-myc TAD$^{p\text{-}S62}$ residue resonates at an unusual frequency, rarely seen in other phosphorylated species, however the analysis of c-myc$_{1\text{-}88}$ p-S^{62} reveals a highly similar positioning of the resonance peak in ^{1}H-^{15}N HSQC spectrum (Helander et al., 2015). S^{62} is known as stabilizing phosphorylation and at the same time is a pre-requisite to the phosphorylation of T^{58}. The juxtaposition of proline residue following S^{62} makes this site a likely substrate for proline-directed kinases. Indeed, multiple kinases have been proposed including: ERK1, ERK2, c-Jun N-terminal Kinase (Jnk), Cdk2:cyclin E, Cdk2:cyclin 2$_A$ and Cdk1:cyclin B (Sears et al., 2000; Hydbring and Larsson, 2010; Noguchi et al., 1999; Helander et al., 2015; Sjostrom et al., 2005). We decided to use

ERK1 to phosphorylate this site, partially due to the difficulties in obtaining Cdk1:cyclin B and partially because it is a well-established kinase that targets this site in c-myc (Sears et al., 2000).

The number of putative kinases listed as potential enzymes targeting this site raises the question of specificity. In physiological conditions, the activity of kinase to substrates is controlled in various ways, through spatial-temporal manner, differential expression patterns, distal docking motifs on substrates and through different binding partners, scaffolding proteins, to name a few. *In vitro* studies are limited in their ability to assess all of these, as they work on a very simplified system that does not exist in the physiological context. For example, the ERK family of kinases recognise their substrates through a DEF motif, which is localised away from the targeted phospho-sites. The general consensus sequence of a DEF motif is R/K_{2-3} - X_{1-6} – Φ – X – Φ and in c- and N-myc this consensus site can be met by K^{55} – K – F – E – L – L – P^{57} and R^{65} – R – S – G – L – C^{70} however the latter is not conserved in N-myc (Jacobs et al., 1999). It is not known how ERK1 recognises either of these proteins *in vivo*.

The data presented here begins to answer a long-standing question of which isoform of GSK3 phosphorylates T^{58} site. GSK3α and β have both been suggested (Lutterbach and Hann, 1994; Pulverer et al., 1994). Studies done in tumour models and cancer cell lines do not distinguish between the two isoforms and knock-out studies using pharmacological agents rarely discriminate between the two isoforms (Kazi et al., 2018). Our data indicates that both GSK3α and GSK3β are able to phosphorylate T^{58} position of N-myc TAD. Our analysis also indicates that p-S^{62} is indeed a pre-requisite for the phosphorylation by GSK3 as previously reported (Lutterbach and Hann, 1994). The abrogation of T^{58} phosphorylation prolongs c- and N-myc half-life substantially from ~13 mins to 63 mins, indicating the role of this modification in myc protein degradation (Sears et al., 2000). However, an S^{62A} mutation also prolongs c-myc half-life in proliferating cells, which indicates that the main function of S^{62} phosphorylation is propagation of myc along its degradation pathway. This would also fit the data which shows that phosphomimic mutation of S^{62} do not occur frequently in cancer, as potentially they could be detrimental to myc protein stability.

Cis-trans p-S^{62}-P^{63} peptidyl prolyl isomerisation bond has been proposed as the next step in c-myc degradation pathway *in vivo* (Yeh et al., 2004). Due to proline amino acid structure, isomerisation induces global conformational change in proteins (Lu and Zhou, 2007). Prolines are undetectable in a ^{1}H-^{15}N HSQC, however any isomerisation would be apparent from CSPs of neighbouring residues. We could not detect such changes in our experiment, in residues surrounding p-S^{62}- P^{63} site or any other sites. The structural basis of Pin1 interaction with myc proteins remains controversial. It was demonstrated that Pin1 interacts with higher affinity with unphosphorylated c-myc$_{1-88}$ than with c-myc$_{1-88}$ p-S^{62} (Helander et al., 2015). This is in contrast to *in vivo* studies where p-S^{62} phosphorylation is a pre-requisite for the binding of Pin1 (Yeh et al., 2004). Pin1 can interact both with N-myc T^{58A} and S^{62A} mutants however these mutations have no impact on N-myc degradation (Sjostrom et al., 2005).

The dephosphorylation of S^{62} has been implicated as the final step, necessary for the ubiquitination of myc protein by SCF^{Fbw7} complex. Our data shows that N-myc TAD can be phosphorylated by ERK1 and then by either of GSK3 isoforms and the doubly phosphorylated species on T^{58} and S^{62} form a complex with Fbw7-Skp1. Whereas N-myc TAD phosphorylated on S^{62} shows weak affinity to Fbw7-Skp1, the addition of phosphorylation on T^{58} allows these two components to form a stable species. This is in agreement with (Yada et al., 2004), where c-myc$^{T58A\ S62A}$ was resistant to binding to Fbw7, resistant to ubiquitination and resistant to proteasomal degradation. This is also in agreement with data obtained from the crystal structure of Fbw7 with a phospho-degron sequence of cyclin E and with structural and biochemical data on the c-myc:Fbw7 interaction, published as the data in his work was just collected (Hao et al., 2007; Welcker et al., 2022). Both cyclin E and c-myc have two phosphodegrons within their sequences, with each phosphodegron possessing two residues spaced 4 amino acids apart, that when phosphorylated, allow a high affinity interaction with Fbw7 WD40 domain. The overlay of these phosphodegrons indicate a very similar mode of binding. The N-terminal phospho-residue (+1) of each phospho-degron interacts with R^{465}, R^{479}, R^{505}, and Y^{519} within WD40 domain and the second phospho-residue (+4 position) nestles within S^{462}, T^{463}, R^{441} and R^{478} Figure 5.15 A (Hao et al., 2007; Welcker et al., 2022).

The N-terminal phosphodegron of c-myc is conserved in N-myc (Figure 5.15 B) which makes it likely that is interacts with WD40 domain in a similar fashion. The C-terminal phosphodegron of c-myc is only partially homologous (Figure 5.15 B), however the important residues are conserved either in their identity or chemistry. In the c-myc structure, the residue immediately preceding first phospho-site is negatively charged

E^{243}, which mediates an important interaction with R^{689} on Fwb7. In N-myc the equivalent residue is D^{257} (Figure 5.15). The subsequent residue, which is a phospho-site, is conserved between c- and N-myc. The second phospho-site in N-myc is S^{362}, in contrast to threonine in c-myc (T^{248}). The site-directed mutagenesis revealed that c-myc T^{248S} is still able to bind to Fbw7 (Welcker et al., 2022). Based on this analysis, and the data from analytical SEC, it is likely that both phophodegron sites of N-myc interact with Fbw7 in a similar fashion to c-myc. All of the data gathered in this work and in this field raises questions if Pin1 isomerisation and S^{62} dephosphorylation is required in the degradation pathway as has been orthodoxy in the field for decades.

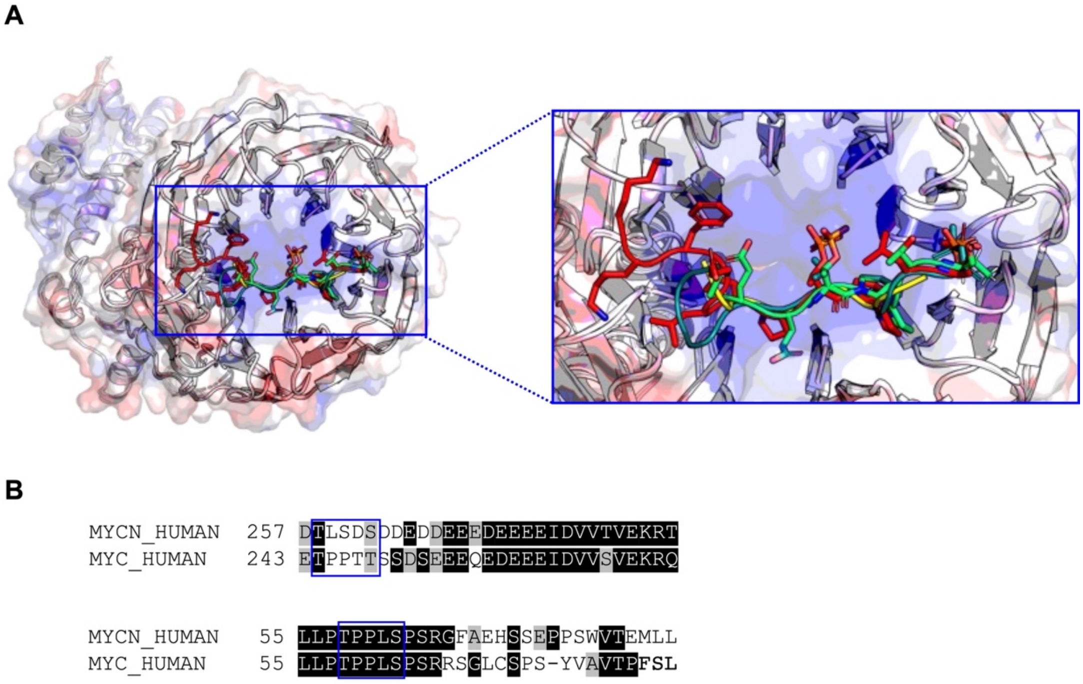


Figure 5.15. Interaction of myc protein phosphodegrons with Fbw7-Skp1 complex.
A. Fbw7-Skp1 (surface, with red indicating +ve charged residues and blue indicating -ve charged residues) with phosphodegrons: cyclin E (S^{58} – K^{65}; lime green), cyclin E (L^{374} – G^{385}; teal), c-myc (K^{51} – S^{62}; red) and c-myc (H^{241} – T^{248}; yellow). **B.** Sequence alignment of c-myc and N-myc N- and C- phosphodegrons. Conserved residues are highlighted in black and residues with similar chemistry are highlighted in grey. Residues which are not conserved are highlighted in white. Blue rectangles indicate the positioning of both of the phosphodegrons withing c- and N-myc protein sequences.

6 N-myc interactions

6.1 Introduction

6.1.1 N-myc known and putative interactions

Multiple binding proteins have been shown to interact with myc proteins, however, very little data is available on the structural basis of these interactions (Kalkat et al., 2018). The difficulty in studying interactions by structural biology lies in the intrinsically disordered nature of myc and the dynamic nature of the interaction of myc with its non-constitutive binding partners. This dynamism is expressed in two ways. Firstly, affinities tend to be relatively low (>1 μM), and secondly, at least in one case (Pin1), myc has been shown to form a highly dynamic fuzzy complex with its binding partner (Helander et al., 2015).

This chapter presents a preliminary interaction data that has been collected for pTEFb, Plk1 and TAF1-TBP proteins, using recombinant proteins. It is often the case that myc "interactions" are determined in cells, using myc immuno-precipitation coupled to identification by immunoblotting or mass-spectrometry approaches. This approach can be prone to false positives as many "interactions" are indirect. Two of the interactors chosen for the work (Plk1 and pTEFB) outlined below have been shown to interact with either myc or N-myc by other groups using such an approach (Büchel et al., 2017; Kalkat et al., 2018; Gargano et al., 2007). The final interactor, the TAF1-TBP complex, had been shown to interact *in vitro* with c-myc, and during the course of this PhD a crystal structure of c-myc with TBP:TAF1 was determined. Three main assays used in this work were pull-down assays, ITC and NMR titrations, selected to test the feasibility of the study, assess the quantitative nature of the interaction with minimal number of tags present and lastly to obtain per-residue resolution of the interaction, respectively.

6.2 Results

6.2.1 Interaction of N-myc TAD with TAF1-TBP

The crystal structure published under the accession code: 6E16 (Wei et al., 2019) indicated that TAF1-TBP interacts with c-myc$_{96\text{-}125}$ in the manner that excludes DNA binding (Figure 6.1 A; Wei et al., 2019), raising a possibility that this interaction might have a regulatory role. The region of c-myc seen in the crystal structure falls outside the conserved MBs, however more detailed analysis of the aligned fragments of N- and c-myc revealed a low level of identity between the two sequences (Figure 6.1 B).

Both fragments are known to form an α-helix when bound to Aurora A and TAF1-TBP, for N- and c-myc, respectively (Richards et al., 2016; Wei et al., 2019). In addition, Piskacek and colleagues documented a 9 amino acid motif within this region which is reasonably well conserved across the four Yamanaka transcription factors (Oct3/4, Sox2, Klf4, and c-myc) which are required for induced pluripotency, possibly suggesting shared mechanism of transactivation between these factors (Piskacek et al., 2020). If it is that the myc-TAF1-TBP is a critical interaction conserved across TF families, we reasoned that it should be conserved within the myc family of TFs.

To test whether N-myc has an ability to interact with TAF1-TBP, we utilised a strategy that was successful in obtaining the crystal structure of c-myc and TAF1-TBP (Wei et al., 2019). We used a recombinant TAF1-TBP construct, in which TAF1$_{6-71}$ was fused to TBP$_{6-240}$ through a triple GSSS linker (Table 2.7). The construct used in obtaining the crystal structure was TAF1$_{9-40}$-c-myc$_{96-125}$-TBP$_{61-240}$, however we decided to use TAF1$_{6-71}$-3x(GSSS)-TBP$_{6-240}$ which was used in NMR titrations and biolayer interferometry with an affinity of 4.7 μM (Wei et al., 2019). In the cloning of our construct, two TEV-NIa protease sites were mistakenly introduced into the fusion proteins and following the TEV-NIa protease cleavage, three different species were generated (uncleaved or cleaved at either of the sites).

To interrogate the interaction of N-myc with TAF1-TBP, a biotin-based pull-down assay was used (Chapter 2.4.4). Biotinylated-N-myc 25-meric TAD peptides (Table 2.5) were used as a bait on Strep-Tactin®, in final concentration of 10 μM. Peptides were incubated for 2 hrs with 15 μM of TAF1-TBP (prey), prior to the elution with 6x SDS-PAGE loading dye. Aurora A was also incubated with N-myc$_{68-89}$ peptide as positive control and His-GFP was used as negative control. We also incubated TAF1-TBP with no bait present to control for non-specific binding of the prey to the resin. Qualitative analysis of de-stained SDS-PAGE gel did not reveal any N-myc TAD region that would strongly interact with TAF1-TBP, beyond non-specific interaction observed between TAF1-TBP and Strep-Tactin® resin (Figure 6.2 C).

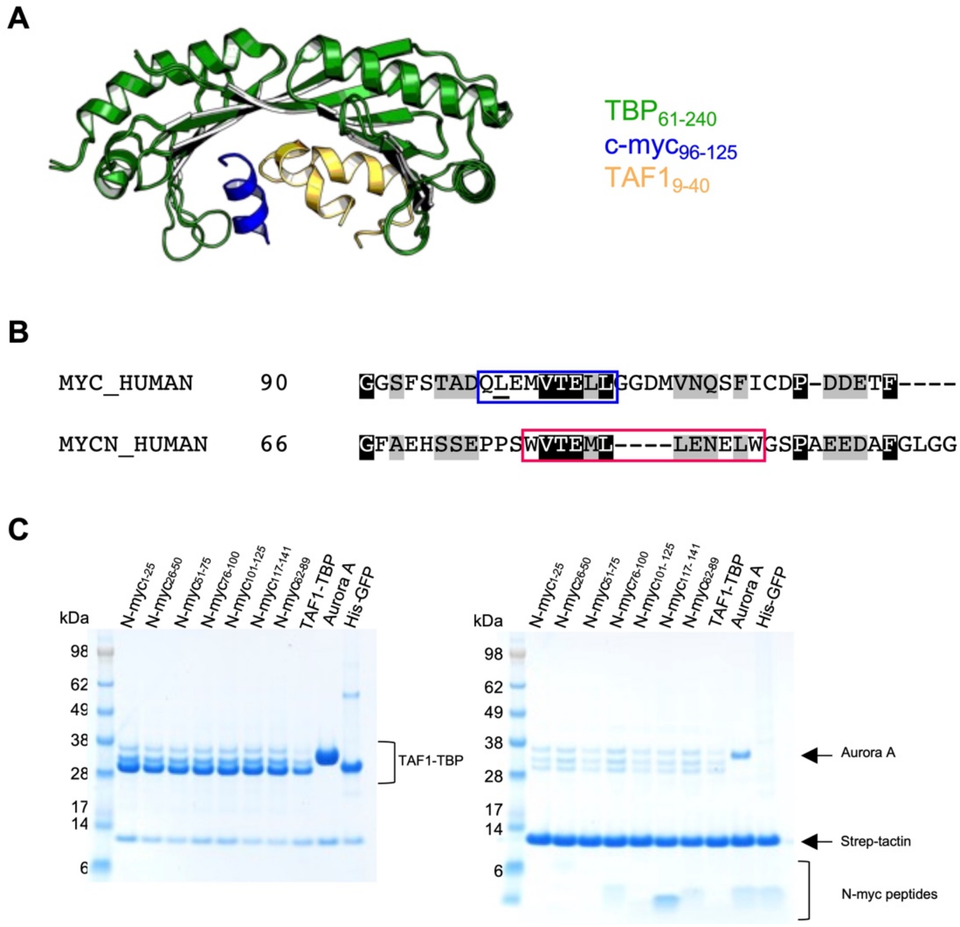


Figure 6.1. N-myc TAD interaction with TAF1-TBP.
A. Crystal structure of a fusion protein of TAF1$_{1\text{-}41}$ (green), c-myc$_{96\text{-}125}$ (blue) and TBP$_{61\text{-}240}$ (yellow) linked by 3x(GSSS) linkers (not shown). **B.** Sequence alignment of fragments of c- and N-myc forming α-helix as seen in crystal structures (Wei et al., 2019; Richards et al., 2016). Green and red box indicate the extent of α-helix as observed in the crystal structures of c- and N-myc, respectively. **C.** Biotin-Strep-Tactin® pull-down assay of N-myc TAD 25-meric biotinylated peptides and TAF1-TBP used as a bait with input (left) and pull-down (right) shown.

To eliminate the issue of three different sizes of TBP-TAF1 construct in the assays, TAF1-TBP was expressed and purified again, omitting TEV-Nia cleavage of 6x His-tag from the purification protocol. We assumed that the presence of this small flexible tag should not impact the affinity studies. We employed ITC, to qualitatively assess weaker interactions, using N-myc TAD as a titrand and TAF1-TBP as a titrant. To ensure that saturation was achieved in the case of low binding affinity, the concentration of TAF1-TBP was over 10-fold more concentrated (200 μM) than N-myc TAD (15 μM). To minimise artefacts due to the heat-of-dilution, both proteins were dialysed overnight into the same buffer (1x TBS, 1 mM TCEP). Raw data was integrated and thermograms were plotted for heat-of-dilution reaction (with the buffer used as a titrand) and the three repeats. There was no signal observed that would indicate any interaction between the two proteins, and the magnitude of the signal of each of the repeats was similar to heat transferred during heat-of-dilution (Figure 6.2).

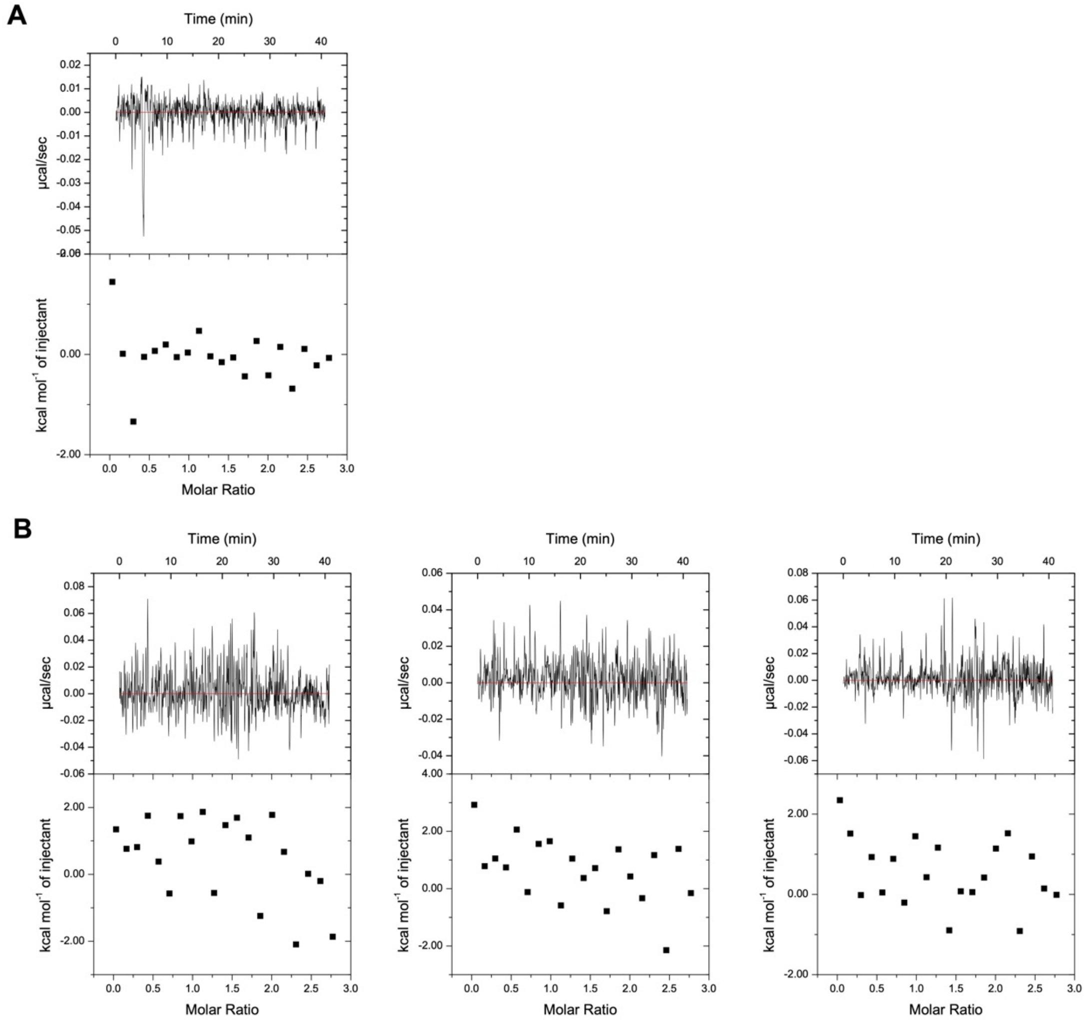


Figure 6.2. ITC data of N-myc TAD (titrand) and TAF1-TBP (titrant).
A. Heat-of-dilution control. TAF1-TBP was injected into the cell containing the buffer to control for the heats associated with the dilution of higher-concentrated species. **B.** Three replicates of ITC titration between N-myc TAD (titrand) and TAF1-TBP (titrant) with the concentration as 15 μM and 200 μM, respectively.

Strong interactions can give ‘weak’ ITC signal, when endo- and exo-thermic components of the interaction are similar in magnitude. To ensure that this was not the case with N-myc TAD and TAF1-TBP, NMR titrations were used as on orthogonal technique to monitor any potential interaction on per-residue basis. NMR titrations rely on collecting consecutive ^{1}H-^{15}N HSQC spectra, with an increasing molar ratio of unlabeled binding partner-to-^{15}N-labelled N-myc TAD. Interaction is reflected in changes to the spectrum as the concentration of a titrant is increased, usually realised in CSPs or diminishing peak intensities of affected residues. To achieve full saturation of the signal, ^{1}H-^{15}N HSQC spectra were recorded at 0, 0.5, 1.0, 1.5 and 2.5 molar ratio of TAF1-TBP-to-N-myc TAD. The spectra were recorded at 25 °C, which was selected based on temperature used for TAF1-TBP titrations published in Wei et al. (2019) but was associated with the loss of observable N-myc TAD peaks (Figure 6.3 A).

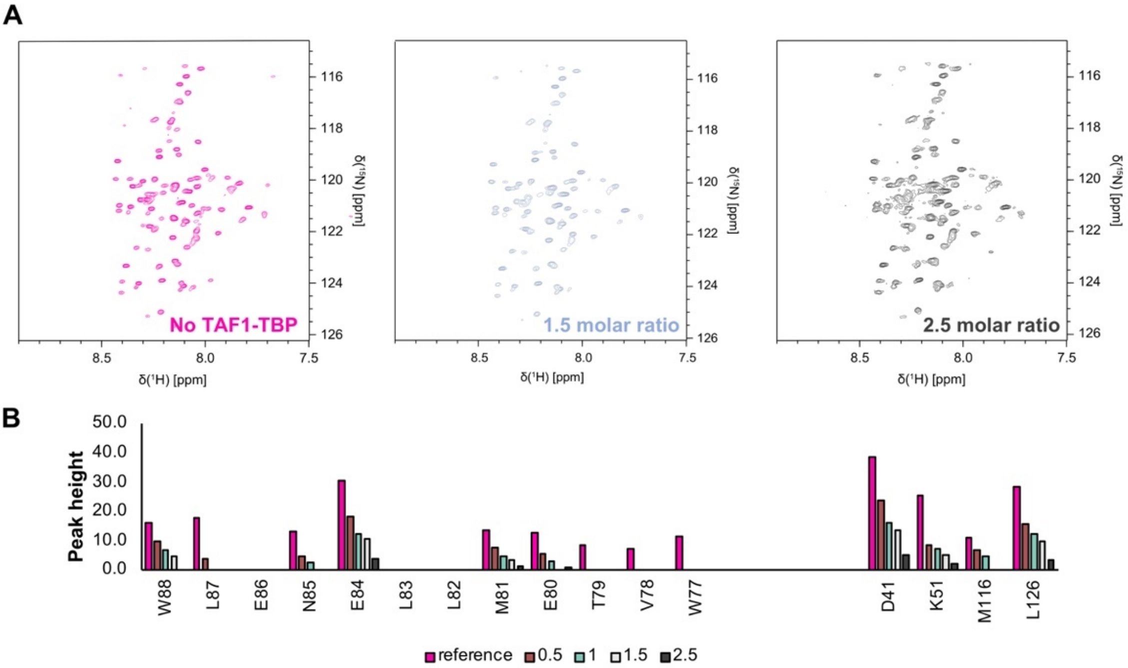


Figure 6.3. NMR titration of ^{15}N N-myc TAD and TAF1-TBP.
A. ^{1}H-^{15}N HSQC reference spectra (pink) and of various points of N-myc TAD and TAF1-TBP titration points, blue (middle) 1.5 molar ratio of TAF1-TBP to N-myc and black (right) 2.5 molar ratio of TAF1-TBP to N-myc **B.** Peak height quantification of N-myc TAD residues equivalent to the residues in c-myc$_{96\text{-}125}$ and randomly selected N-myc TAD peaks (D^{41}, K^{51}, M^{116} and L^{126}), not associated with TAF1-TBP interaction. Numbers indicate molar ratio of TAF1-TBP-to-N-myc TAD.

The analysis of spectra obtained following each titration point indicated no presence of CSPs, however there was an overall dampening of the signal in the mid-titration in comparison to the end-point (blue *vs.* black, Figure 6.3 A). We quantified the heights of the peaks of W$_{77}$-W$_{88}$ residues, which were thought to potentially interact with TAF1-TBP. We also did the same analysis on the peaks selected from different regions of the N-myc TAD sequence that were thought not to participate in the interaction (D$_{41}$, K$_{51}$, M$_{116}$ and L$_{126}$). The quantitative assessment of all of the peaks showed an overall decrease in the signal intensity with the increasing molar ratio of TAF1-TBP-to-N-myc TAD, possibly due to peak broadening (Figure 6.3 B). This effect was much more pronounced for W$_{77}$-W$_{88}$ residues in comparison to the random residues of N-myc TAD (Figure 6.3 B). The reduction of peak intensities as the concentration of TAF1-TBP increased could be explained by few different factors. The titration of increasing concentrations of TAF1-TBP causes dilution of ^{15}N N-myc TAD, leading to a decrease in the signal intensities. The interaction between two partners that falls within the intermediate chemical exchange could also result in the peak intensity loss, however, in this case the signal should return upon reaching the saturation at higher titration points (Helander et al., 2015). This could be further complicated by N-myc forming latent secondary structures upon binding to TAF1-TBP, which causes the W^{77}-W^{88} region to exist in two different conformations (α-helical and random coil), thus splitting the signal amongst fractional populations.

We employed three different orthogonal techniques to interrogate this interaction and none of the data generated was unambiguously positive evidence for an interaction between N-myc and TAF1-TBP. It is possible that the TAF1-TBP construct used is incompatible with N-myc TAD binding. The crystal structure only features the N-terminal region of TAF1 (residues 9 - 40; Wei et al., 2019), whereas in our construct, TAF1 is extended to the residue 71. The lack of appropriate positive controls with c-$\text{myc}_{96\text{-}125}$ limits the scope of analysis that can be performed on the data. However, given that the none of the data were indicating the interaction N-myc TAD and TAF1-TBP we decided that this interaction was not a priority for further work.

6.2.2 Interaction of N-myc TAD with Plk1

Plk1 has been repeatedly shown to interact with both c-myc and N-myc by immuno-precipitation approaches (Kalkat et al., 2018; Büchel et al., 2017; Popov et al., 2010; Koch et al., 2007). Plk1 has also been shown to prevent N-myc degradation through promoting autoubiquitination of Fbw7 (Xiao et al., 2016). Although the previous data presented in this work has indicated that N-myc is not a substrate of Plk1, the question whether Plk1 and N-myc interact as direct binding partners remains unanswered. This sub-chapter demonstrates the challenging nature of studying IDP interactions. To study the interaction between N-myc TAD and Plk1, a 'reductive approach' was undertaken, in which the individual proteins were broken down into smaller subcomponents to identify specific sites driving the interaction.

To first test the feasibility of this study using recombinant proteins, a biotin-Strep-Tactin® pull-down assay was set up using biotinylated N-myc 25-meric peptides as a bait (Table 2.5) and untagged full-length $\text{Plk1}_{3\text{-}603}$ as prey (courtesy of Eoin Leen). N-myc TAD fragments: 25-50, 62-89 and 75-100 all pulled down Plk1 (Figure 6.4). To control for the unspecific binding of Plk1 to the resin, Plk1 was incubated with the resin and no bait (background; Figure 6.4 A and B). ImageJ was utilised to quantify band intensities and the ratio of intensities of Plk1-to-Strep-Tactin® was calculated, with three biological replicates performed. This experiment confirmed that Plk1 interacts significantly with N-myc$_{25-50}$ and N-myc$_{62-89}$ (Figure 6.4 A and B).

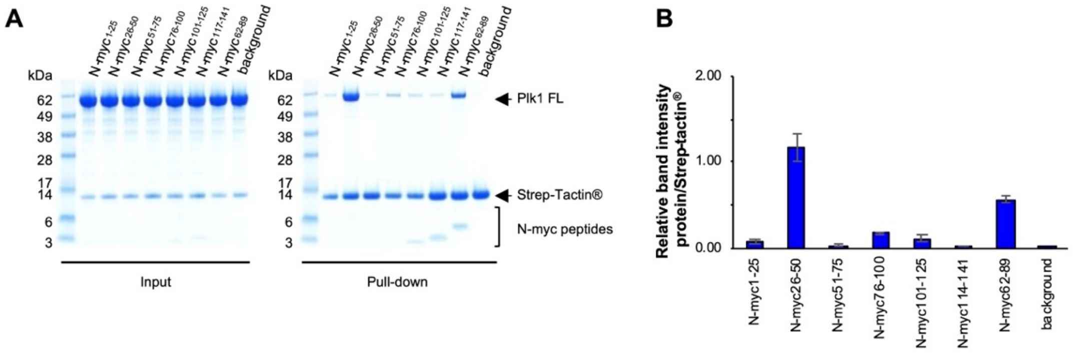


Figure 6.4. N-myc interacts with Plk1 FL.
A. Pull-down assay of biotinylated N-myc 25-meric peptides (bait) and untagged full-length Plk1 (Plk1 FL; prey). N-myc peptides were immobilised on Strep-Tactin® resin and incubated with Plk1 FL. Background indicate Plk1 FL incubated with the resin only. **B** Quantification of biotin-Strep-Tactin® pull-down assay of biotinylated N-myc 25-meric peptides and Plk1 FL. ImageJ was used to quantify the intensity of the prey and Strep-Tactin$^{®}$ bands in pull-down assay and the ratio of prey-to-Strep-Tactin$^{®}$ was quantified. Errors indicate standard deviation calculated from three biological repeats.

To further test which Plk1 domain is contributing to this interaction, biotin-based pull-down assays were repeated, this time using recombinant Plk1$_{3-330}$ (Plk1 Kinase domain - KD) and Plk1$_{326-603}$ (Plk1 Polo box domain – PBD). Both proteins were generated in-house, courtesy of Dr Eoin Leen. To minimise protein demands, only 25-meric N-myc peptides that showed to interact with Plk1 FL were used as prey (N-myc$_{26-50}$, N-myc$_{62-89}$ and N-myc$_{76-100}$) and N-myc$_{51-75}$ was used as a negative control. We also included the background controls which were Plk1 domains incubated with the resin only. As a positive control, N-myc$_{26-50}$ was incubated with Plk1 FL. Visual inspection of SDS-PAGE gels indicated the level of signal, comparable to a non-specific binding of the negative control (Figure 6.5 A). The pull-down quantification using ImageJ and the band intensities did not yield consistent results, characterised by a high standard deviation (Figure 6.5 B). Notably, the PBD showed equal binding to the negative control peptide (N-myc$_{51-75}$), consistent with non-specific binding to all peptides.

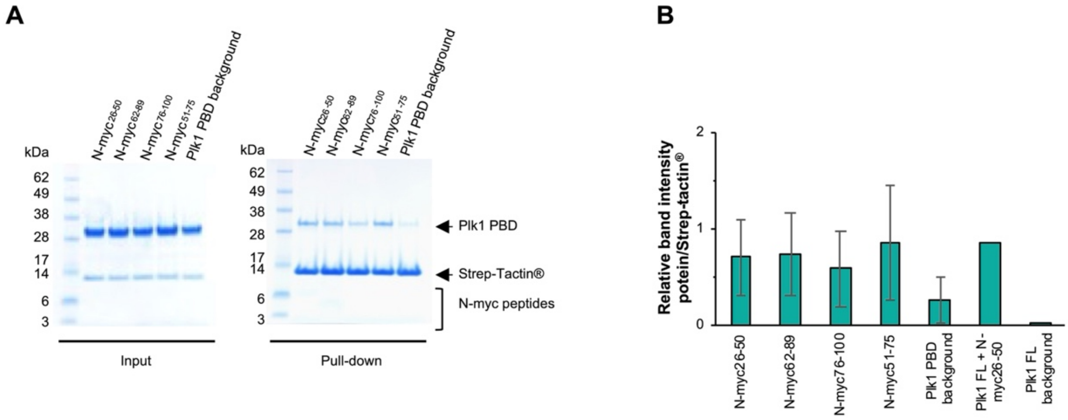


Figure 6.5. N-myc pull-down with Plk1 PBD
A. Coomassie staining of pull-down between biotinylated N-myc 25-meric peptides (bait) and Plk1 PBD (prey). The bait was immobilised on Strep-Tactin® resin, incubated extensively with Plk1 PBD and washed thoroughly**.** Plk1 PBD background indicates the reaction where the protein was incubated with the resin only. **B.** ImageJ was used to quantify the intensity of Plk1 PBD bands and Strep-Tactin® bands and the ratio of Plk1 PBD bands-to- Strep-Tactin® band was calculated. Error bars are standard deviation calculated from four biological repeats.

Pull-downs indicated that Plk1 PBD was not the main domain mediating the interaction between N-myc TAD and Plk1, thus Plk1 KD was investigated next using same pull-down set-up as above. Unlike Plk1 PBD, Plk1 KD showed a clear, reproducible enrichment with N-myc$_{28-89}$ and N-myc$_{25-50}$, and much weaker interaction with the negative control peptide (N-myc$_{51-75}$), consistent with the finding reported for Plk1 FL (Figure 6.6 A). Band intensity quantification of the pull-down fraction showed high Plk1 KD-to-Strep-Tactin® ratio, comparable to the interaction between Plk1 FL and N-myc$_{26-50}$ (Figure 6.6 B).

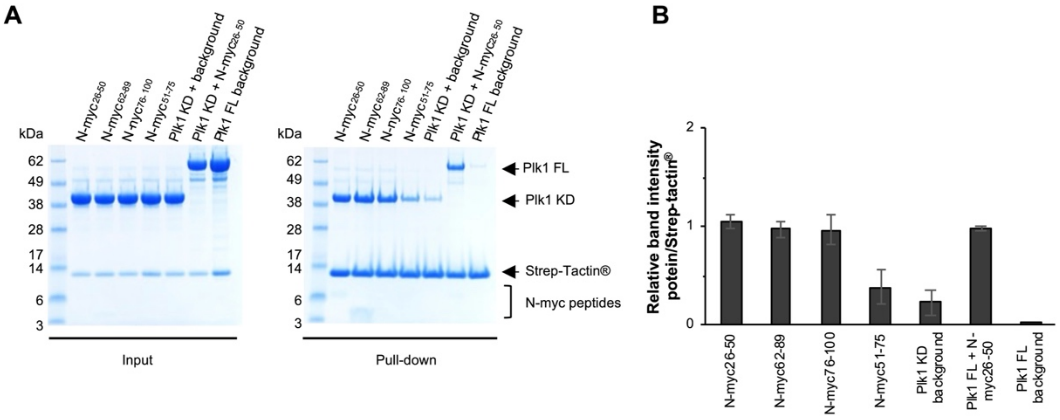


Figure 6.6. N-myc pull-down with Plk1 KD.
A. Pull-down of biotinylated N-myc 25-meric peptides and Plk1 KD. The peptides were immobilised on Strep-Tactin$^{®}$ resin and incubated with Plk1 KD. **B.** Quantification of biotin-Strep-Tactin$^{®}$ pull-down fractions between biotinylated N-myc

25-meric peptides and Plk1 KD. Error bards are standard deviation calculated from three biological repeats.

To quantitatively assess this interaction, we utilised ITC with N-myc TAD as titrand and Plk1$_{13\text{-}345}$ T210V as a titrant (courtesy of Eoin Leen). As Plk1$_{13\text{-}345}$ T210V expressed at much higher yields than Plk1$_{3\text{-}330}$, it was used in ITC, which requires high amounts of titrant to reach binding curve saturation, especially in the context of IDP interactions, which are characterised by lower affinities. IDP interactions are also characterised by high avidity, i.e., possessing multiple binding sites, where many residues mediate weak interactions that cumulatively give a raise to biologically-relevant binding (Dyson and Wright, 2005; Uversky, 2013). To assess if this is the case for N-myc and Plk1 KD, recombinant N-myc TAD was utilised, instead of N-myc peptides. To ensure the saturation of the signal in a case of weak interaction, 15 μM of N-myc TAD and 180 μM of Plk1$_{13\text{-}345}$ T210V was used. A day prior to the experiment, the proteins were dialysed overnight into the same buffer (20 mM HEPES pH 7.2, 150 mM NaCl, 1 mM TCEP, 2 mM $MgCl_2$, 10% v/v glycerol). We did not manage to observe a binding curve between the two binding partners, with heat signatures similar to the heat-of-dilution control (Figure 6.7). Due to the signal corresponding to the noise, it was not possible to fit the binding curve to the thermogram.

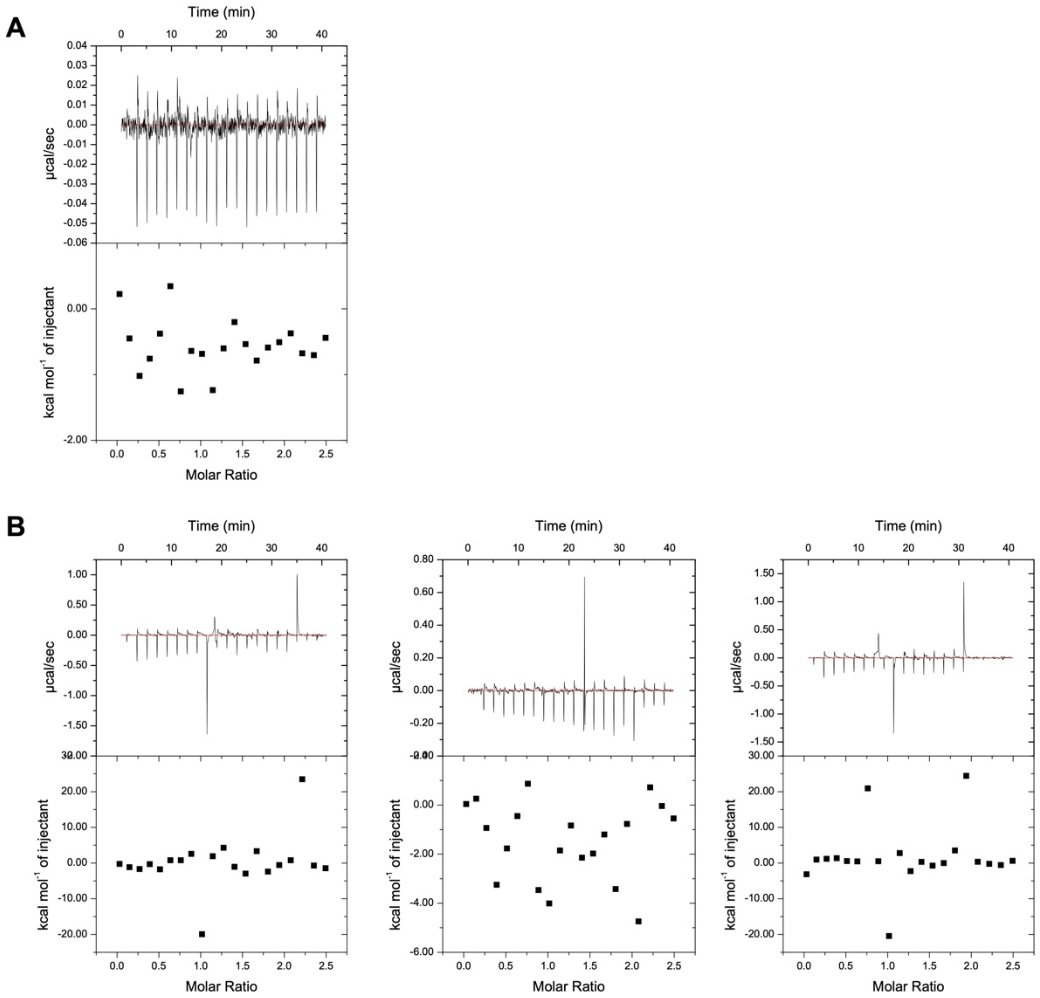


Figure 6.7. ITC of N-myc TAD (titrand) and Plk1$_{45\text{-}345}$ T210V (titrant).
A. Heat-of-dilution control of the buffer (cell) and Plk1$_{45\text{-}345}$ T210V (syringe). **B.** Three biological repeats of ITC titrations between N-myc TAD, used as titrand in the final concentration of 15 μM and Plk1$_{45\text{-}345}$ T210V (titrant) in the final concentration of 180 μM.

To probe this interaction further, ITC experiment was set-up using N-myc TAD or N-myc$_{64\text{-}137}$ as a titrand in concentrations of 13 and 15 μM, respectively and His-SUMO Plk1$_{367\text{-}603}$ H538A K540A (referred to as His-SUMO Plk1 PBD), used as a titrant in final concentration of 160 μM (Figure 6.8). H538A K540A are important residues mediating phospho-peptide motif recognition with PBD, and the crystal structure of Plk1 PBD has been obtained by mutating these residues to alanines (Elia et al., 2003). The proteins concentrations used in this assay were low, due to a protein loss that occurred during the dialysis into ITC buffer (1x TBS, 5 % v/v glycerol, 1 mM TCEP, 5% v/v DMSO) that occurred a day prior to the ITC experiment. We observed large heat signatures and number of erratic points in the thermograms when performing heat-of-dilution control, which was likely caused by a mismatch of DMSO between the cell and the syringe. It appears there is some saturation being reached in the thermogram of full-length N-myc TAD (Figure 6.8 C), however the isotherm following the data integration is ambiguous.

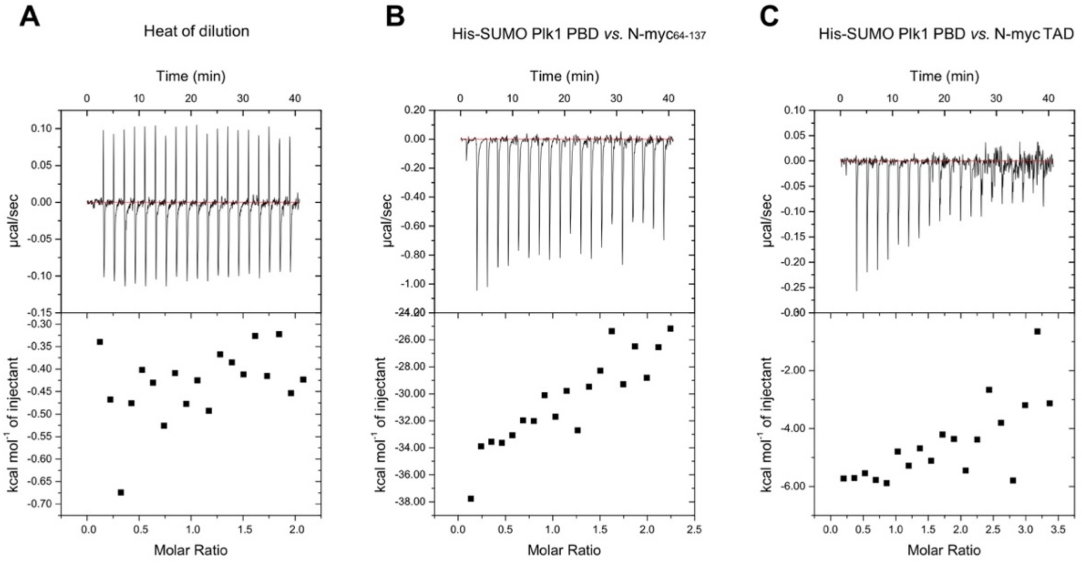


Figure 6.8. ITC titrations of N-myc TAD and N-myc$_{64-137}$ and N-myc TAD with His-SUMO Plk1$_{367-603}$ H538A K540A (His-SUMO Plk1 PBD).
A. Heat-of-dilution control with His-SUMO Plk1 PBD titrated into the matching buffer (1x TBS, 5% v/v glycerol, 1 mM TCEP, 5% v/v DMSO). Large heat signatures are likely a result of the presence of 5% v/v DMSO in the buffer. **B.** Titration of His-SUMO Plk1 PBD (titrand) into N-myc$_{64-137}$ (titrant). **C.** Titration of His-SUMO Plk1 PBD (titrand) into N-myc TAD (titrant). All experiments were performed at 25 °C, using 2 μL injection volumes. Each titration was performed with only one biological repeat, due to the protein loss during overnight dialysis.

To further probe the interaction between Plk1 PBD and N-myc TAD, NMR titrations were utilised as means to follow the interaction on per-residue basis with 0.73, 1.0 and 2.0 molar ratio of Plk1 PBD-to-^{15}N N-myc TAD. To achieve the best signal-to-noise ratio, all the experiments were recorded at 10 °C. Despite that, ^{1}H-^{15}N HSQC spectra were characterised by low signal-to-noise ratio, with signal dissipating even at low molar ratios (Figure 6.9 A). The visual inspection of the sample following data collection revealed no precipitation, indicating that this effect was not due to sample instability. There was some evidence that Plk1 PBD interacts with N-myc TAD at 2:1 molar ratio. This suggests that even at the end point of the NMR titration (2.0 molar ratio) with the signal-to-noise ratio significantly diminished, only half of the sites on N-myc TAD were saturated. Plk1 PBD is also known to dimerise, thus potentially increasing the size of the complex to ~70 kDa, which exceeds the technical capabilities of routine NMR (Raab et al., 2022; Cordeiro et al., 2020; Elia et al., 2003).

High molecular weight complexes are characterised by slow molecular tumbling times which decreases R_2 and prolongs R_1. This causes FID to dissipate quickly and limits the number of subsequent pulsing, which makes it challenging to collect a data with high signal-to-noise ratio and high resolution.

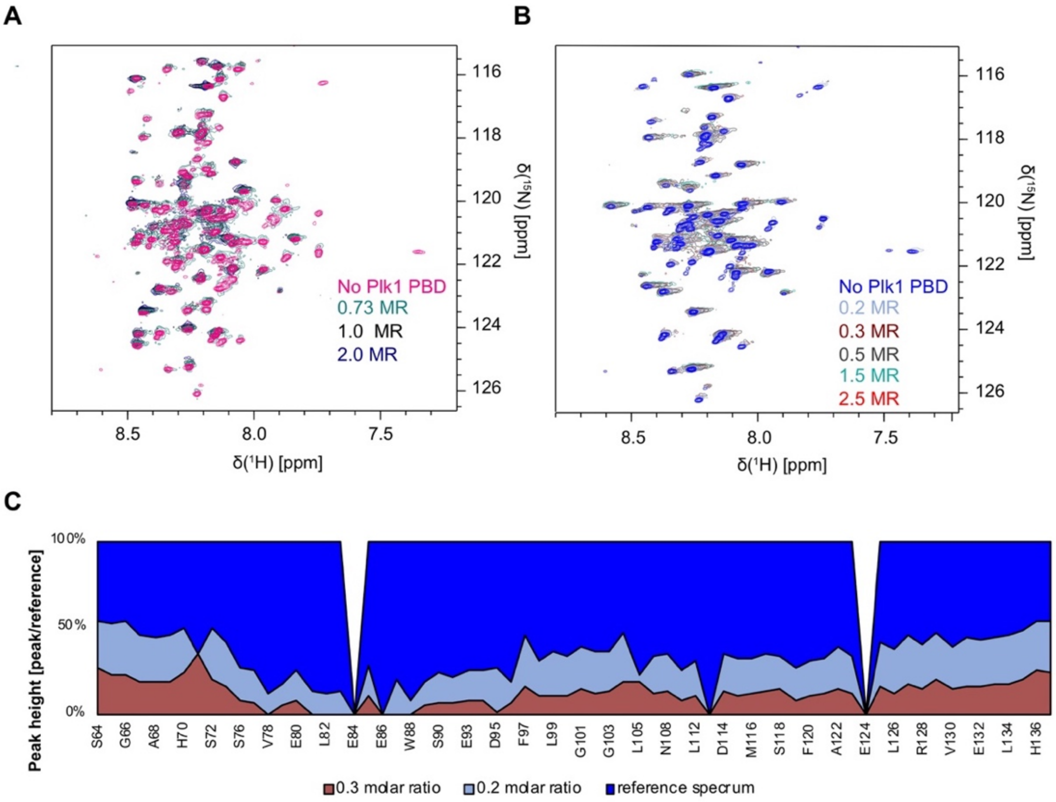


Figure 6.9. N-myc TAD NMR titrations with Plk1 PBD.
A. Overlay of ^{1}H-^{15}H HSQC spectra of ^{15}N N-myc TAD with no Plk1 PBD present (pink) and Plk1 PBD at various Plk1 PBD-to-N-myc TAD ratios (teal, black and blue for 0.73, 1 and 2 molar ratio, respectively). **B.** Overlay of ^{1}H-^{15}H HSQC spectra of ^{15}N N-myc$_{64-137}$ with no Plk1 PBD present and present at 0.2, 0.3, 0.5, 1.5 and 2.5 molar ratio of Plk1 PBD-to-N-myc$_{64-137}$ (blue, light blue, brown, gray, teal and red, respectively). **C.** Quantification of peak height of N-myc$_{64-137}$ for three titration points (no Plk1 PBD, 0.2 and 0.3 molar ratios). Peak heigh of the reference spectrum (blue) was normalised to 100%. Reference peak height at 0% indicate that peak intensity was too weak for quantification.

Based on the data gathered from pull-downs between N-myc peptides and Plk1 FL, we speculated that the interaction resembles the interaction with Aurora A, with two binding sites, one localised to MB0 and the second one localised to W$_{77}$-W$_{88}$ (Richards et al., 2016). To remove one of the binding sites on N-myc TAD, we resorted to using a shorter N-myc$_{64-137}$ construct, to reduce the size of the putative complex formed. Titration points were selected at smaller molar ratio increments to observe and record the initial stages of binding reaction, with 0.2, 0.3, 0.5, 1 and 1.5 molar ratio of Plk1 PBD to N-myc$_{64-137}$ selected (Figure 6.9 B). Similarly to the titration with full-length N-myc TAD, there was an overall signal dissipation starting at higher molar ratio of 0.5, thus peak heights were quantified for no Plk1 PBD, 0.2 and 0.3 molar ratio for all residues. Peak heights were normalised with reference spectrum peak heights set to 100% (Figure 6.9 C). Following a careful analysis of individual peaks, we did not observe any CSPs in response to Plk1 PBD binding. There was an overall decrease in peak heights across all the residues, with peak heights of W^{77}-W^{88} particularly afflicted by the drop of intensity, with M^{81}, L^{82}, L^{83}, E^{86}, L^{87} and W^{88} disappearing completely, either at 0.2 or 0.3 molar ratio. It is possible that it is caused by the interaction with Plk1 PBD, however this region is also characterised by low peak intensities prior to the addition of any binding partners. This pattern of signal dissipation was very similar to the NMR titrations performed with TAF1-TBP, where even low molar ratio titration causes loss of signal-to-noise and is commonly observed in NMR of IDPs with the combination of factors contributing to this phenomenon. IDPs often form secondary structures within their MoRFs regions, which in itself can lead to a decrease in peak intensities (Sharma et al., 2019).These regions usually also are the source of interactions, which also leads to decrease in peak intensities, especially if the interaction fall into intermediate chemical exchange.

The work presented in this chapter highlights the challenges in studying IDP interactions. To characterise the interaction in detail, one might have to use a simplified system of recombinant proteins, which strips the interaction of many other auxiliary proteins that might support the interaction. Using more simple approach with small peptides can also result in lack of interaction observed, due to removal of other important sites within protein. Our pull-down assay results with the peptides indicated that Plk1 KD is the major source of interaction between N-myc TAD and Plk1. However, we could not confirm this binding event using ITC. There might be few reasons behind this. Firstly, the version of pull-down assay used in this work would not only show the interacting protein, but also the protein that precipitated on the Strep-Tactin® beads. Peptides N-myc$_{25-50}$ and N-myc$_{62-89}$ are hydrophobic in nature and particularly prone to precipitating on the beads. In addition, it is likely that N-myc interacts with both domain and potentially at their interface and reducing the interaction to individual domains reduced the avidity of the entire interaction and thus can give false negative results.

6.2.3 Interaction of N-myc TAD with pTEFb

The interaction between myc proteins and pTEFb has pursued for decades however so far, no study has measured the binding affinity or characterised the structural basis of the interaction. The previous chapter explored the possibility that N-myc TAD is a substrate for pTEFb and this chapter explores the putative complex formation between the two species.

A HTRF® assay was utilised to assess the binding affinity of this putative interaction. HTRF® assays requires two different tags. The proximity of these two species causes the donor and the acceptor to come in close contact and allows FRET-based transfer of energy. Biotinylated N-myc$_{62\text{-}89}$ peptides phosphorylated at different positions (Table 2.5) were utilised, as well as either FLAG-tagged Cdk9:cyclin T_1 (hereafter referred to pTEFb) or FLAG-tagged cyclin T_1. Streptavidin-conjugated dye (Streptavidin-XL665) and anti-FLAG antibody-conjugated terbium were added as the acceptor and donor, respectively (Cisbio). Upon the recognition of the species, anti-FLAG antibody-conjugated terbium emits a fluorescence at 620 nm which is transferred onto the donor (Streptavidin-XL665), which in turn emits fluorescence at 665 nm. The ratio of 665-to-620 is then calculated and high HTRF ratio is indicative of high 665 nm emission, due to donor and acceptor being brought into proximity of each other. To eliminate the possibility of unbound acceptor and donor interacting, their concentration was kept at equimolar ratio to their respective interactors.

Both pTEFb and cyclin T_1 were tested for the interaction with biotinylated N-myc$_{28\text{-}89}$ peptides, however they did not indicate any measurable interaction (Figure 6.10 A and B, respectively). AFF_4, which is known to interact with pTEFb, was used as a positive control (Schulze-Gahmen et al., 2013). Whereas AFF_4 produced a high HTRF ratio, none of the N-myc peptides produced significant HTRF ratio, apart from N-myc$_{28\text{-}89}$ p-T^{58} interacting with cyclin T_1, which has shown high standard deviation (teal circle, Figure 6.10 A).

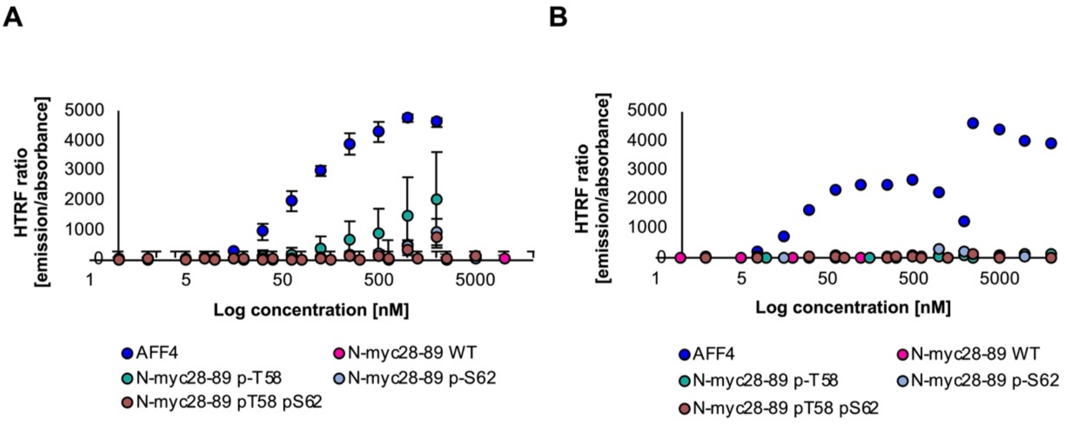


Figure 6.10. HTRF assay of biotinylated N-myc$_{28-89}$ peptides.
A. HTRF assay of biotinylated N-myc$_{28-89}$ peptides and cyclin T$_1$. AFF$_4$ was used as a positive control. **B.** HTRF assay between biotinylated N-myc$_{28-89}$ peptides and pTEFb. The concentration of pTEFb and cyclin T$_1$ was varied to achieve a saturation in case of weak binding. Error bars are standard deviation calculated from three biological repeats.

As a HTRF assay requires two different tags, it could be that either tag obscures the interaction site or that the interaction itself was preventing binding of donor and acceptor antibodies to bind to their respective tags. To remedy this issue, pull-down assays using either FLAG or glutathione resins were employed. We utilised recombinant 3x FLAG N-myc TAD and Gst-tagged pTEFb.

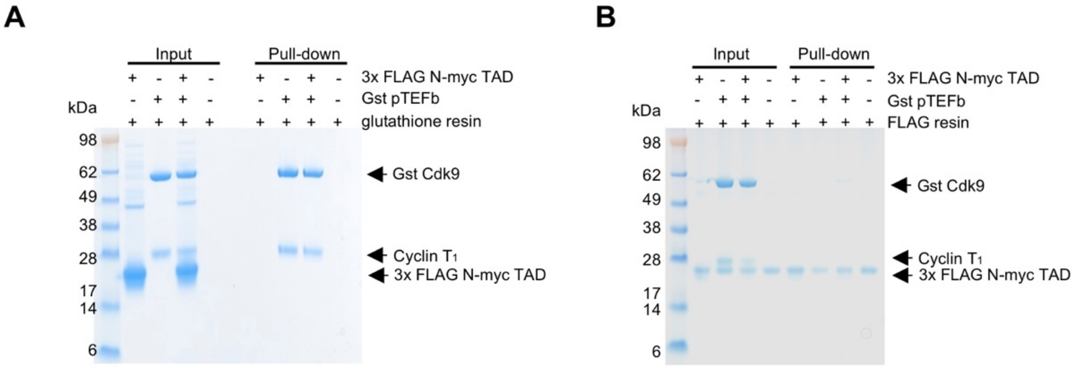


Figure 6.11. Pull-down assays of 3x FLAG N-myc TAD and Gst pTEFb.
A. Glutathione resin pull-down assay with Gst pTEFb used as bait and 3xFLAG N-myc TAD used as prey. **B.** FLAG resin pull-down assay with and 3x FLAG N-myc TAD immobilised on the resin as a bait and Gst pTEFb used as a prey. In both pull-down prey and bait were incubated for two hours prior to the extensive washing procedure. The elutions were performed by the addition of 4x SDS loading dye.

None of the pull-downs indicated an interaction between N-myc TAD and pTEFb (Figure 6.11 A and B). There was no enrichment of prey in any of these pull-downs as observed through extensive Coomassie staining. To complete the investigation, immunoblotting was performed against FLAG- and Gst tags, respectively, however no signal was observed for the presence of the baits in the respective pull-downs (data not shown).

To confirm the negative findings regarding N-myc TAD and pTEFb interaction, a final set of pull-down assays was conducted, this time utilizing FLAG-tagged recombinant N-myc proteins spanning different lengths of TAD (Figure 6.12 A). Following the incubation of 10 μg of Gst pTEFb (bait) with 20 μg of Gst FLAG N-myc TAD proteins (prey) and an extensive washing, there was no enrichment of the beads with FLAG-tagged N-myc recombinant proteins. The utilisation of anti-FLAG immunoblotting did detect the enrichment of N-myc$_{22\text{-}137}$, N-myc$_{64\text{-}137}$ and N-myc TAD on the beads (Figure 6.12 B). N-myc$_{48\text{-}89}$ and N-myc$_{1\text{-}64}$ were not detected in the pull-down (Figure 6.12 B), indicating that the C-terminal region of TAD is critical for the interaction. The binding to N-myc$_{64\text{-}137}$ was very weak, indicating that the N-terminal region, potentially harbours a second site.

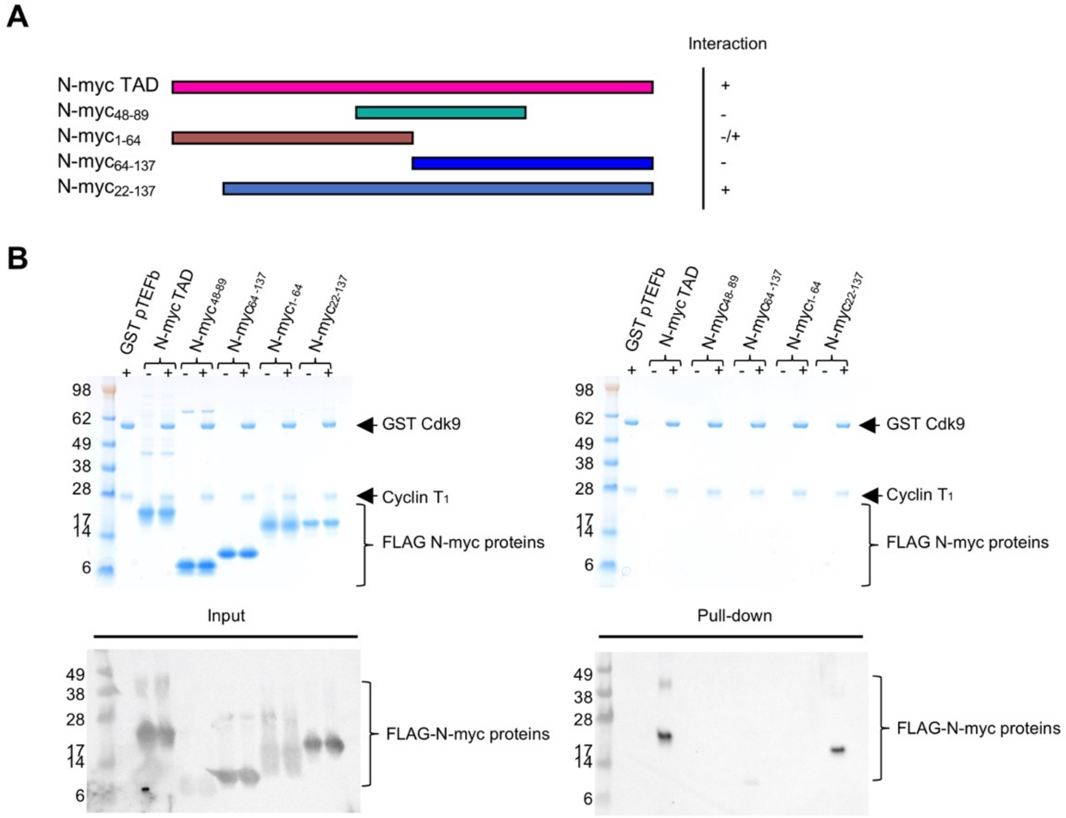


Figure 6.12. Pull-down assays of Gst pTEFb and FLAG-tagged N-myc proteins spanning the length of TAD.
A. Schematic representation of FLAG-N-myc TAD (pink) recombinant proteins and FLAG-tagged TAD truncations used in the assay. **B.** Coomassie staining of pull-down assay (top) with input (left) and pull-down (right) fractions. Gst pTEFb was immobilised on glutathione resin as a bait, and FLAG-tagged N-myc recombinant proteins were used as prey. Anti-FLAG Western blotting (bottom) of the stained fractions.

6.3 Discussion

6.3.1 Study of N-myc TAD interactions

This chapter has interrogated a range of putative interactions which have been implicated either for N-myc, or have been characterised in the context of c-myc. We utilised a broad range of biophysical techniques to investigate the binding both qualitatively and quantitatively. NMR was employed to study N-myc TAD interactions. Apart from its difficulty in data processing and analysis, we encountered typical problems that are associated with IDP interaction plasticity. In this work NMR titration suffer from poor signal-to-noise ratio even at low titration points, which is a common feature of both TAF1-TBP and PLK1 PBD titrations (Helander et al., 2015; Andresen et al., 2012). IDPs are characterised by increased interaction surface and high avidity of interactions, with multiple binding sites collectively contributing to stable binding (Uversky, 2013; Wright and Dyson, 2015). In addition, when studying the interactions through NMR, IDPs are usually isotopically labelled and thus observed. In the case of interaction with larger binding partner, the favourable properties such as fast tumbling rates and fast R_2 rates are foregone, as an IDP becomes incorporated into a larger complex through the interaction (Uversky, 2013). Ultimately, the study of IDPs in the context of a single binding partner, although necessary for drug design, is even less physiologically relevant than the study of globular folded proteins. IDPs tend to bind multiple partners at the same time on short time scales, which raises the possibility that a lot of these interactions are occurring due to the presence of auxiliary proteins (Tompa and Fuxreiter, 2008; Uversky, 2013).

Unsurprisingly, that there are only a handful of available myc structures, with only a single structure of N-myc, obtained using crystallography. These interactions were obtained following extensive research of observed myc properties in cells, giving these phenomena a structural basis. For example, the positive correlation between destabilisation of N-myc and Aurora A inhibition was long recognised, albeit only explained in molecular detail through the crystal structure of the complex of these two proteins (Otto et al., 2009; Richards et al., 2016). This example serves to underline that complex cellular system studies must act in concert with basic structural and biophysical research to fully comprehend myc biochemistry.

6.3.2 TAF1-TBP interaction

The region of c-myc interaction with TBP and TAF1 seen in the crystal structure lies outside the conserved region of MBs, however the alignment of a short region of N- and c-myc indicated some level of homology (Wei et al., 2019). c-myc residues that were directly implicated in the interaction were L^{99}, M^{101}, V^{102}, T^{103}, L^{105}, L^{106}. Especially, the four C-terminal c-myc residues were important for the interaction and these are retained in N-myc sequence (equivalent residues: V^{78}, T^{79}, L^{82} and L^{83}; Wei et al., 2019). Three orthogonal techniques used in this study did not confirm this interaction between N-myc TAD and TAF1-TBP.

As we have used a marginally different construct to the one used in the crystallography studies, with TAF1 elongated from 9-40 to 46-71, it was possible that the extra C-terminal residues of this protein could obstruct the binding between N-myc and TAF1-TBP. However, the analysis of the overlay of the crystal structures of $TAF1_{9-40}$-c-myc_{96-125}-TBP_{61-240} and $TAF1_{9-71}$- TBP_{61-240} indicated that residues 46-71 of TAD point away from c-myc surface and thus should not impact the binding between N-myc TAD and TAF1-TBP (not shown).

NMR titrations showed overall decrease in peak intensities, for all the residues, with this effect pronounced for W^{77} - W^{88} residues. Helander and colleagues (2015) and Andresen and colleagues (2012) observed the interaction between c-myc_{1-88} and Pin1 and c- myc_{1-88} and Bin, respectively through a combination of diminishing peak intensities and modest CSPs (Helander et al., 2015). Their conclusions were the result of c-myc assuming the dynamics of c-myc_{1-88}-Pin1 complex and continuously exchange between free and bound state (Helander et al., 2015; Uversky, 2013). The lack of signal recovery seen is their work, as well as ours, is consistent with the formation of the fuzzy complex, however without a further investigation this remains a hypothesis. This region is subjected to chemical exchange, either due to interchange between monomeric and dimeric state or due to exchange between folded and unfolded states. Without any orthogonal techniques confirming the interaction between N-myc TAD and TAF1-TBP, we concluded that withing the scope of reagents interrogated, there is no interaction between these two partners. The data could be improved by the incorporation of a positive control in form of c-myc_{96-125} peptide in ITC and pull-down assay to assess that TAF1-TBP construct used is an appropriate choice.

The analysis the sequences of known transcription factors indicated most of them feature a stretch of 9 amino acids which are sufficient to trigger transcription (Piskacek et al., 2020). This sequence is also present in Yamanaka factors (Oct4, Klf1 and Sox2; Takahashi and Yamanaka, 2006). In c-myc (E^{100}MVTELLGG^{109}) mediates the interaction with TAF1-TBP and in N-myc the equivalent region (S^{76}WVTEMLLE^{83}), was found not to interact with TAF1-TBP. It is possible that N-myc utilises different mechanism of transcriptional regulation to c-myc, to trigger similar physiological outcomes. For example, ‘TBP-free’ transcription can occur, with TAF11 proteins associating in different combinations and forming different TFIID fractions without the presence of TBP (Wieczorek et al., 1998). The substitution of c-myc gene sequence for N-myc gene sequence allows mice to grow into adulthood and reproduce normally, with no apparent developmental abnormalities, indicating that N-myc can functionally replace c-myc in mice (Malynn et al., 2000; Zimmerman et al., 1986). On the other hand, a knock-out of either c- or N-myc leads to the embryonic lethality at the early murine embryonic development stage, indicating that functional redundancy of these two TFs *in vivo* is not possible, at least at developmental stage (Davis et al., 1993).

6.3.3 Interaction with Plk1

This work focused on interrogating the interaction between N-myc TAD and Plk1 further by including the interaction analysis of the full length Plk1, as well as interrogation of the individual domains in binding to N-myc TAD. Pull-down assays using individual domains showed at least two binding sites for Plk1 on N-myc TAD and indicated that Plk1 KD is the main source of interaction with N-myc TAD, however these results were not reproduced through other orthogonal techniques, with the conflicting data on the main domain driving this interaction. It is possible that the Plk1 KD is a site of a single interaction site, which is mainly mediated by MB0 and thus it gives positive results when appropriate length N-myc peptides are used. Regarding the Plk1 PBD, it is possible that the entire interaction is mainly avidity driven, with each individual residue contributing little affinity and collectively giving rise to overall interaction. This would explain why the interrogation of Plk1 PBD with the peptides does not indicate a strong level of interaction, but the full-length N-myc TAD with Plk1 PBD does indicate interaction.

Further characterisation of the interactions should rely on establishing the binding parameters between N-myc TAD and the full-length Plk1. This could be achieved through ITC, or to mitigate this protein-demanding technique, through biolayer interferometry. The pull-down assay indicated the sites of the interaction to be MB0 and the fragment of N-myc spanning residues 62-89, and this strongly resembles the interaction between N-myc TAD and Aurora A (Richards et al., 2016). Competition studies would help to establish qualitatively that these sites are indeed involved in binding both kinases.

NMR, which is a technique most suited to studying the behaviour of IDPs, was not able to provide a definitive answer as to the interaction between N-myc TAD and Plk1 PBD. Due to the nature of an IDP, the observed signal dissipation could be a result of different dynamic processes occurring at different time scales and contributing to the diminishing of the signal in ^{1}H-^{15}N HSQC spectra. The decrease in the peak intensities could be minimised by a careful design of the experimental parameters. We already employed low temperature at data collection to minimise proton exchange with bulk water and minimised the concentration of glycerol in the sample buffer, to enhance the overall tumbling rate of the putative complex. Further steps could include sample deuteration to minimise spin diffusion and reduce the relaxation rates of ^{15}N (Sattler and Fesik, 1996), To mitigate poor signal-to-noise ratio due to a chemical exchange, TROSY-based pulse sequences can be implemented (Pervushin et al., 1997). The signal recorded in NMR is the average of T_2 of different components contributing to the relaxation (Pervushin et al., 1997). In a normal experiment, R_2 of these components is averaged out through decoupling, however in large molecules, this can lead to attenuation of the signal or complete signal disappearance if one of these components has fast R_2 and the other one has slow R_2 (Fernández and Wüthrich, 2003). In TROSY, instead of averaging the R_2 relaxation components, only the signal with slow R_2 is considered. Although these halves the signal collected, enhanced signal-to-noise compensates for the loss (Fernández and Wüthrich, 2003; Pervushin et al., 1997). Whereas this approach was not necessary for N-myc TAD backbone assignments, due to its already slow relaxation rates, TROSY could help to overcome the poor signal-to-noise ratio when N-myc is bound to a larger protein.

Lastly, the most challenging aspect to study would be obtaining the crystal structure of the complex. This would require establishing whether the interaction between N-myc TAD and Plk1 is ‘fuzzy’ in nature, and which regions specifically form such interactions and which regions are the minimal-interacting regions, as a common approach to crystallising these proteins is to remove more dynamic regions that are not seen in electron density maps. Another difficulty is crystallising full-length Plk1 and thus far, only one crystal structure is available which features KD and PBD expressed separately in *Danio rerio* and *Drosophila melanogaster*, respectively (Xu et al., 2013).

6.3.4 Interaction with pTEFb

The literature documents the interaction between c-myc and cyclin T_1, which was demonstrated through pull-down assay between Gst-tagged c-myc TAD fragments as a bait and cyclin T_1 as a prey, immunoprecipitated from HeLa lysates (Eberhardy and Farnham, 2002). Further validation of interactions was also performed using co-immunoprecipitation method and lysates from various cell lines (Kanazawa et al., 2003; Gargano et al., 2007). Although no further biochemical and structural data has been presented thus far, the correlation between c-myc and pTEFb function has been well-established. c-myc impacts the proximal pause release, through the recruitment of pTEFb, as evidenced by co-immunoprecipitation experiments in murine embryonic stem cells. The pharmacological abrogation of c-myc-Max heterodimer with reduces c-myc occupancy at the active genes, concomitant with a reduction of serine 2 phosphorylation of CTD (Rahl et al., 2010). c-myc turnover has also been shown to be important for the recruitment of pTEFb to active promoter sites (Jaenicke et al., 2016).

The main limitation of previous interaction studies of myc protein, was the utilisation of immunoblotting as means of detection, Protein-specific antibodies (anti-cyclin T_1 in this case) only capture their epitope and any other potential factors that may facilitate or even mediate the interaction. The remaining literature evidence linking c-myc to transcriptional elongation through the regulation of pTEFb, does not probe directly the nature of this interaction (Rahl et al., 2010) (Kanazawa et al., 2003; Gregory et al., 2003a).

In contrast to these data, our work focused on using recombinant proteins to directly probe the interaction between N-myc and pTEFb. The data collected indicate no or very weak interaction between N-myc TAD and pTEFb or cyclin T_1. FRET-based HTRF assay and an array of pull-downs did not indicate any interaction between the two proteins, apart from immunoblotting detection. pTEFb was a challenging complex to express and purify, with high levels of precipitation. The complex required glycerol in its buffer, which due to its increased viscosity, negatively impacts R_1 and R_2 and diminishes signal-to-noise ratio in NMR studies. Due to negative preliminary data and challenges in obtaining recombinant proteins, further investigation was not undertaken.

7 Final remarks

7.1 Summary of the thesis

N-myc is a potent oncogene in multiple cancers, but thus far there is no pharmacological intervention that would target its oncogenic properties. This work was focused on addressing some of the gaps in our understanding of N-myc TAD function and structure. We utilised NMR, which we deemed to be the most appropriate technique to study this IDP, to probe the dynamics and interactions of N-myc TAD. A substantial amount of work was put towards the backbone assignment of N-myc TAD, which is a foundation for all NMR-related experiments.

The backbone assignment confirmed that the N-myc TAD is an IDP, with narrow ^{1}H dispersion, peak overlap due to low sequence complexity and averaging out of chemical environment of due to fast dynamics. Using the backbone assignment and CSI, we confirmed the presence of α-helix spanning residues W^{77} – W^{88}. We also characterised previously unreported α-helix present at the C-terminus of TAD (A^{122} – E^{132}). The former α-helix participates in binding of N-myc TAD to Aurora A (Richards et al., 2016). The significance of the latter α-helix is not known, however it lies within MBII, the region which is implicated in the interaction with TRRAP (McMahon et al., 1998). The relaxation experiments performed on the N-myc TAD, revealed a degree of flexibility, however there were regions of restricted local motions, which corresponded to the presence of the α-helices. The assignment of a shorter TAD fragment, N-myc$_{64\text{-}137}$ revealed that the residues participating in $W_{77\text{-}W88}$ helix experience two distinct chemical environments. Following the extensive interrogation, we could not definitely confirm the mechanism behind these chemical environments. It is likely that multiple processes which undergo chemical exchange contribute to this phenomenon.

The second part of the work focused on investigating N-myc TAD in the context of its interactions, both as a substrate and as a binding partner. We reconstituted part of the N-myc degradation pathway including two sequential phosphorylation events *in vitro*, as demonstrated through changes in the chemical shifts. N-myc TAD was phosphorylated on S^{62} by ERK1 and then on T^{58} by GSK3 kinases. Further steps, which included isomerisation of the p-S^{62}-P^{63} bond by Pin1 and the dephosphorylation of p-S^{62} by PP1, were not observed. In agreement with the recently published data, Fbw7-Skp1 complex can bind p-S^{62} N-myc TAD, however the addition of phospho-group on T^{58} greatly enhances the interaction and leads to the formation of a stable complex, as per analytical SEC results.

The studies of the interactions of N-myc TAD yielded interesting results. The data gathered demonstrated the lack of interaction between N-myc TAD and both TAF1-TBP and pTEFb, in the context of reagents used for the work. N-myc TAD interacts with the full-length Plk1, as well as both of the domains – PBD and KD. Initially, the data from the pull-down assays indicated Plk1 KD as a major site of interaction with Plk1, however further data revealed that Plk1 PBD is as likely to participate in this interaction. It underlines the importance of utilising multiple orthogonal techniques to study interactions. In particular, IDPs are characterised by a high avidity of binding, where each low-affinity site contributes to biologically significant interaction in additive manner. To study interactions mediated by high avidity, full proteins domains should be used to obtain strong enough binding for biochemical and structural studies.

Throughout the duration of this PhD, a few crystal structures of c-myc were published, c-myc$_{96-125}$ with TAF1-TBP, c-myc$_{58-64}$ with Fbw7, and c-myc$_{260-267}$ with Fbw7 (Wei et al., 2019; Welcker et al., 2022). No NMR structure of c-myc or N-myc has been published thus far and the only available data on BRMB, is the assignment of c-myc$_{1-88}$. This work addresses the bottleneck in the study of N-myc through NMR by providing a complete backbone assignment and allow further studies to commence. For example, experiments probing slower dynamic processes (μs – ms) could be performed. Structural studies, for example residual dipolar couplings can also be collected which would provide information of special orientation of individual vectors and may lead to the elucidation of structure/ensemble of N-myc TAD structures in solution.

7.2 Outlook of N-myc – based cancer therapies

Two major obstacles in drug design are the lack of fundamental understanding of biochemistry, and the technical difficulty in designing a molecule that will interact stably with an IDP. The bulk of our understanding into N-myc biochemistry and interactions is derived from the studies in cells. Although helpful in guiding target selection, they lack the resolution that is required to design pharmacological agents. The only known pharmacological intervention targeted against myc protein is 10058-F4 and its derivatives, which disrupts the c-myc:Max heterodimer and prevents c-myc-related transcription (Yin et al., 2003). However, this inhibitor suffers from poor tumour penetration, lack of specificity and rapid *in vivo* metabolism (Wang et al., 2013). Due to its lack of 3D structure, the most likely mechanism to inhibit myc function is to target its upstream or downstream targets, of which some promising ones are kinases that regulate myc stability. It has been demonstrated that Aurora A inhibitors which distort its active site, reduce N-myc half-life (Otto et al., 2009; Richards et al; 2016). ERK1 inhibitors are at clinical or pre-clinical trial stages however thus far none of them have been FDA approved (Kidger et al., 2018). It is not known whether such as inhibitor would shorten N-myc half-life, however as often with kinase inhibitors a multiagent treatment would be advantageous as tumours often develop further mutation to mitigate the pharmacological effects of the inhibitors. The Plk1 kinase inhibitor that has made it the closest to the clinic is Volasertib, which showed a lot of promise in the I/II phase of clinical trials but did not manage to show the same success in the III phase (Gutteridge et al., 2016). Ultimately, patient stratification will be important in administering of these drugs to maximise their effectiveness and provide rational combinational therapies.

Research into fundamental functions of N-myc will propel the discovery of successful therapies. This work addresses part of these foundations in the form of complete backbone assignment, studies of the dynamics and preliminary data on several potential N-myc interaction partners.

8 References

Ahlner A, Carlsson M, Jonsson B-H, et al. (2013) PINT: a software for integration of peak volumes and extraction of relaxation rates. *Journal of Biomolecular NMR* 56(3): 191-202.

Alexander J, Lim D, Joughin BA, et al. (2011) Spatial exclusivity combined with positive and negative selection of phosphorylation motifs is the basis for context-dependent mitotic signaling. *Sci Signal* 4(179): ra42.

Allevato M, Bolotin E, Grossman M, et al. (2017) Sequence-specific DNA binding by MYC/MAX to low-affinity non-E-box motifs. *PloS one* 12(7): e0180147.

Andresen C, Helander S, Lemak A, et al. (2012) Transient structure and dynamics in the disordered c-Myc transactivation domain affect Bin1 binding. *Nucleic Acids Res* 40(13): 6353-6366.

Arai M (2018) Unified understanding of folding and binding mechanisms of globular and intrinsically disordered proteins. *Biophysical Reviews* 10(2): 163-181.

Archambault V and Glover DM (2009) Polo-like kinases: conservation and divergence in their functions and regulation. *Nature Reviews Molecular Cell Biology* 10(4): 265-275.

Ayer DE, Kretzner L and Eisenman RN (1993) Mad: A heterodimeric partner for Max that antagonizes Myc transcriptional activity. *Cell* 72(2): 211-222.

Ayer DE, Lawrence QA and Eisenman RN (1995) Mad-max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. *Cell* 80(5): 767-776.

Babu MM (2016) The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochemical Society Transactions* 44(5): 1185-1200.

Bah A and Forman-Kay JD (2016) Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *J Biol Chem* 291(13): 6696-6705.

Bahram F, von der Lehr N, Cetinkaya C, et al. (2000) c-Myc hot spot mutations in lymphomas result in inefficient ubiquitination and decreased proteasome-mediated turnover. *Blood* 95(6): 2104-2110.

Baluapuri A, Hofstetter J, Dudvarski Stankovic N, et al. (2019) MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. *Mol Cell* 74(4): 674-687.e611.

Barna JCJ, Laue ED, Mayger MR, et al. (1987) Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. *Journal of Magnetic Resonance (1969)* 73(1): 69-77.

Batchelor M, Wolny M, Baker EG, et al. (2019) Dynamic ion pair behavior stabilizes single α-helices in proteins. *Journal of Biological Chemistry* 294(9): 3219-3234.

Bax A and Ikura M (1991) An efficient 3D NMR technique for correlating the proton and15N backbone amide resonances with the α-carbon of the preceding residue in uniformly15N/13C enriched proteins. *Journal of Biomolecular NMR* 1(1): 99-104.

Bax B, Carter PS, Lewis C, et al. (2001) The structure of phosphorylated GSK-3beta complexed with a peptide, FRATtide, that inhibits beta-catenin phosphorylation. *Structure* 9(12): 1143-1152.

Beltran H, Rickman DS, Park K, et al. (2011) Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer Discov* 1(6): 487-495.

Beltran H, Tomlins S, Aparicio A, et al. (2014) Aggressive variants of castration-resistant prostate cancer. *Clin Cancer Res* 20(11): 2846-2850.

Berger I, Tölzer C and Gupta K (2019) The MultiBac system: a perspective. *Emerging Topics in Life Sciences* 3(5): 477-482.

Berjanskii MV and Wishart DS (2007) The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. *Nucleic Acids Res* 35(Web Server issue): W531-537.

Bermel W, Bertini I, Felli IC, et al. (2012) Speeding up sequence specific assignment of IDPs. *J Biomol NMR* 53(4): 293-301.

Bertini I, Duma L, Felli IC, et al. (2004) A Heteronuclear Direct-Detection NMR Spectroscopy Experiment for Protein-Backbone Assignment. *Angewandte Chemie International Edition* 43(17): 2257-2259.

Bienkiewicz EA and Lumb KJ (1999) Random-coil chemical shifts of phosphorylated amino acids. *J Biomol NMR* 15(3): 203-206.

Bister K (1986) Structure and Function of myc and mil Oncogenes. In: Tanner W and Gallwitz D (eds) *Cell Cycle and Oncogenes.* Berlin, Heidelberg: Springer Berlin Heidelberg, 75-81.

Bjerke L, Mackay A, Nandhabalan M, et al. (2013) Histone H3.3. mutations drive pediatric glioblastoma through upregulation of MYCN. *Cancer Discovery* 3(5): 512-519.

Blackwood E and Eisenman R (1991) Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* 251(4998): 1211-1217.

Blackwood EM, Kretzner L and Eisenman RN (1992a) Myc and Max function as a nucleoprotein complex. *Curr Opin Genet Dev* 2(2): 227-235.

Blackwood EM, Lüscher B and Eisenman RN (1992b) Myc and Max associate in vivo. *Genes Dev* 6(1): 71-80.

Blancato J, Singh B, Liu A, et al. (2004) Correlation of amplification and overexpression of the c-myc oncogene in high-grade breast cancer: FISH, in situ hybridisation and immunohistochemical analyses. *Br J Cancer* 90(8): 1612-1619.

Bouchard C, Dittrich O, Kiermaier A, et al. (2001) Regulation of cyclin D2 gene expression by the Myc/Max/Mad network: Myc-dependent TRRAP recruitment and histone acetylation at the cyclin D2 promoter. *Genes & Development* 15(16): 2042-2047.

Brodeur GM, Seeger RC, Schwab M, et al. (1984) Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* 224(4653): 1121-1124.

Brown RS, Sander C and Argos P (1985) The primary structure of transcription factor TFIIIA has 12 consecutive repeats. *FEBS Letters* 186(2): 271-274.

Brubaker K, Cowley SM, Huang K, et al. (2000) Solution structure of the interacting domains of the Mad-Sin3 complex: implications for recruitment of a chromatin-modifying complex. *Cell* 103(4): 655-665.

Büchel G, Carstensen A, Mak KY, et al. (2017) Association with Aurora-A Controls N-MYC-Dependent Promoter Escape and Pause Release of RNA Polymerase II during the Cell Cycle. *Cell Rep* 21(12): 3483-3497.

Buck M (1998) Trifluoroethanol and colleagues: Cosolvents come of age. Recent studies with peptides and proteins. *Quarterly reviews of biophysics* 31: 297-355.

Burrell RA, McGranahan N, Bartek J, et al. (2013) The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 501(7467): 338-345.

Byrne DP, Clarke CJ, Brownridge PJ, et al. (2020) Use of the Polo-like kinase 4 (PLK4) inhibitor centrinone to investigate intracellular signalling networks using SILAC-based phosphoproteomics. *Biochem J* 477(13): 2451-2475.

Cammers-Goodwin A, Allen TJ, Oslick SL, et al. (1996) Mechanism of Stabilization of Helical Conformations of Polypeptides by Water Containing Trifluoroethanol. *Journal of the American Chemical Society* 118(13): 3082-3090.

Campen A, Williams RM, Brown CJ, et al. (2008) TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept Lett* 15(9): 956-963.

Chakrabartty A, Kortemme T and Baldwin RL (1994) Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Science* 3(5): 843-852.

Chandriani S, Frengen E, Cowling VH, et al. (2009) A Core MYC Gene Expression Signature Is Prominent in Basal-Like Breast Cancer but Only Partially Overlaps the Core Serum Response. *PloS one* 4(8): e6693.

Chang AT, Liu Y, Ayyanathan K, et al. (2015) An evolutionarily conserved DNA architecture determines target specificity of the TWIST family bHLH transcription factors. *Genes Dev* 29(6): 603-616.

Chebaro Y, Ballard AJ, Chakraborty D, et al. (2015) Intrinsically Disordered Energy Landscapes. *Scientific Reports* 5(1): 10386.

Chhabra S, Fischer P, Takeuchi K, et al. (2018) <sup>15</sup>N detection harnesses the slow relaxation property of nitrogen: Delivering enhanced resolution for intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* 115(8): E1710.

Chong S-H and Ham S (2013) Conformational Entropy of Intrinsically Disordered Protein. *The Journal of Physical Chemistry B* 117(18): 5503-5509.

Chrzan P, Skokowski J, Karmolinski A, et al. (2001) Amplification of c-myc gene and overexpression of c-Myc protein in breast cancer and adjacent non-neoplastic tissue. *Clinical Biochemistry* 34(7): 557-562.

Clubb RT, Thanabal V and Wagner G (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C′ chemical shifts in 15N□13C-labelled proteins. *Journal of Magnetic Resonance (1969)* 97(1): 213-217.

Cohen P and Frame S (2001) The renaissance of GSK3. *Nature Reviews Molecular Cell Biology* 2(10): 769-776.

Combes G, Alharbi I, Braga LG, et al. (2017) Playing polo during mitosis: PLK1 takes the lead. *Oncogene* 36(34): 4819-4827.

Conacci-Sorrell M, McFerrin L and Eisenman RN (2014) An Overview of MYC and Its Interactome. *Cold Spring Harbor perspectives in medicine* 4(1).

Cordeiro MH, Smith RJ and Saurin AT (2020) Kinetochore phosphatases suppress autonomous Polo-like kinase 1 activity to control the mitotic checkpoint. *Journal of Cell Biology* 219(12): e202002020.

Cowling VH and Cole MD (2006) Mechanism of transcriptional activation by the Myc oncoproteins. *Seminars in Cancer Biology* 16(4): 242-252.

Cowling VH, D'Cruz CM, Chodosh LA, et al. (2007) c-Myc Transforms Human Mammary Epithelial Cells through Repression of the Wnt Inhibitors DKK1 and SFRP1. *Molecular and Cellular Biology* 27(14): 5135-5146.

Cramer P (2019) Organization and regulation of gene transcription. *Nature* 573(7772): 45-54.

Dajani R, Fraser E, Roe SM, et al. (2001) Crystal structure of glycogen synthase kinase 3 beta: structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell* 105(6): 721-732.

Dang C (2012) MYC on the Path to Cancer. *Cell* 149(1): 22-35.

Dani C, Blanchard JM, Piechaczyk M, et al. (1984) Extreme instability of myc mRNA in normal and transformed human cells. *Proceedings of the National Academy of Sciences* 81(22): 7046-7050.

Das SK, Kuzin V, Cameron DP, et al. (2022) MYC assembles and stimulates topoisomerases 1 and 2 in a “topoisome”. *Molecular Cell* 82(1): 140-158.e112.

Davis AC, Wims M, Spotts GD, et al. (1993) A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. *Genes & Development* 7(4): 671-682.

Deiana A, Forcelloni S, Porrello A, et al. (2019) Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PloS one* 14(8): e0217889.

Delaglio F, Grzesiek S, Vuister GW, et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3): 277-293.

Delaglio F, Walker GS, Farley KA, et al. (2017) Non-Uniform Sampling for All: More NMR Spectral Quality, Less Measurement Time. *American pharmaceutical review* 20(4): 339681.

Department of Biology (n.d.) *Random Coil Chemical Shift calculator for IDPs*. Available at: https://www1.bio.ku.dk/english/research/bms/sbinlab/randomchemicalshifts1/ (accessed 26.10.2022).

Desbarats L, Gaubatz S and Eilers M (1996) Discrimination between different E-box-binding proteins at an endogenous target gene of c-myc. *Genes Dev* 10(4): 447-460.

Diercks T, Coles M and Kessler H (1999) An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments. *Journal of Biomolecular NMR* 15(2): 177-180.

Doble BW and Woodgett JR (2003) GSK-3: tricks of the trade for a multi-tasking kinase. *J Cell Sci* 116(Pt 7): 1175-1186.

Dou Z, von Schubert C, Körner R, et al. (2011) Quantitative mass spectrometry analysis reveals similar substrate consensus motif for human Mps1 kinase and Plk1. *PloS one* 6(4): e18793.

Drozdetskiy A, Cole C, Procter J, et al. (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Research* 43(W1): W389-W394.

Dunker AK, Lawson JD, Brown CJ, et al. (2001) Intrinsically disordered protein. *J Mol Graph Model* 19(1): 26-59.

Dyson HJ and Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6(3): 197-208.

Eberhardy SR and Farnham PJ (2002) Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. *J Biol Chem* 277(42): 40156-40162.

Ehara H, Yokoyama T, Shigematsu H, et al. (2017) Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* 357(6354): 921-924.

Elia AE, Rellos P, Haire LF, et al. (2003) The molecular basis for phosphodependent substrate targeting and regulation of Plks by the Polo-box domain. *Cell* 115(1): 83-95.

Elster A (2021) *MRIquestions.com* Available at: https://www.mriquestions.com.

Estrada DF, Conner M, Jeor SC, et al. (2011) The Structure of the Hantavirus Zinc Finger Domain is Conserved and Represents the Only Natively Folded Region of the Gn Cytoplasmic Tail. *Front Microbiol* 2: 251.

Eyles S (2001) Proline not the only culprit? *Nature structural biology* 8: 380-381.

Facey G (2016) Non-uniform Sampling (NUS). In: Available at: http://u-of-o-nmr-facility.blogspot.com (accessed 2022).

Farmer BT, 2nd, Venters RA, Spicer LD, et al. (1992) A refocused and optimized HNCA: increased sensitivity and resolution in large macromolecules. *J Biomol NMR* 2(2): 195-202.

Farrow NA, Zhang O, Szabo A, et al. (1995) Spectral density function mapping using 15N relaxation data exclusively. *Journal of Biomolecular NMR* 6(2): 153-162.

Felli IC and Pierattelli R (2022) 13C Direct Detected NMR for Challenging Systems. *Chemical Reviews*. DOI: 10.1021/acs.chemrev.1c00871.

Fernández C and Wüthrich K (2003) NMR solution structure determination of membrane proteins reconstituted in detergent micelles. *FEBS Letters* 555(1): 144-150.

Ferrari S, Marin O, Pagano MA, et al. (2005) Aurora-A site specificity: a study with synthetic peptide substrates. *Biochem J* 390(Pt 1): 293-302.

Fiol CJ, Mahrenholz AM, Wang Y, et al. (1987) Formation of protein kinase recognition sites by covalent modification of the substrate. Molecular mechanism for the synergistic action of casein kinase II and glycogen synthase kinase 3. *Journal of Biological Chemistry* 262(29): 14042-14048.

Foster MP, McElroy CA and Amero CD (2007) Solution NMR of large molecules and assemblies. *Biochemistry* 46(2): 331-340.

Frank SR, Schroeder M, Fernandez P, et al. (2001) Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. *Genes & Development* 15(16): 2069-2082.

Frenkiel T, Bauer C, Carr MD, et al. (1990) HMQC-NOESY-HMQC, a three-dimensional NMR experiment which allows detection of nuclear overhauser effects between protons with overlapping signals. *Journal of Magnetic Resonance (1969)* 90(2): 420-425.

Galea CA, Wang Y, Sivakolundu SG, et al. (2008) Regulation of Cell Division by Intrinsically Unstructured Proteins: Intrinsic Flexibility, Modularity, and Signaling Conduits. *Biochemistry* 47(29): 7598-7609.

Gargano B, Amente S, Majello B, et al. (2007) P-TEFb is a crucial co-factor for Myc transactivation. *Cell Cycle* 6(16): 2031-2037.

Garner E, Romero P, Dunker AK, et al. (1999) Predicting binding regions within disordered proteins. *Genome informatics* 10: 41-50.

Garson JA, McIntyre PG and Kemshead JT (1985) N-myc amplification in malignant astrocytoma. *Lancet* 2(8457): 718-719.

Gerlach Jennifer M, Furrer M, Gallant M, et al. (2017) PAF1 complex component Leo1 helps recruit Drosophila Myc to promoters. *Proceedings of the National Academy of Sciences* 114(44): E9224-E9232.

Ghosh S, Kundu A and Chattopadhyay K (2018) Small Molecules Attenuate the Interplay between Conformational Fluctuations, Early Oligomerization and Amyloidosis of Alpha Synuclein. *Scientific Reports* 8(1): 5481.

Gibbs EB, Cook EC and Showalter SA (2017) Application of NMR to studies of intrinsically disordered proteins. *Archives of Biochemistry and Biophysics* 628: 57-70.

Gibbs EB and Kriwacki RW (2018) Direct detection of carbon and nitrogen nuclei for high-resolution analysis of intrinsically disordered proteins using NMR spectroscopy. *Methods (San Diego, Calif.)* 138-139: 39-46.

Gilmour DS and Lis JT (1986) RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. *Molecular and Cellular Biology* 6(11): 3984-3989.

Gonzalez FA, Raden DL and Davis RJ (1991) Identification of substrate recognition determinants for human ERK1 and ERK2 protein kinases. *J Biol Chem* 266(33): 22159-22163.

Graether SP (2019) Troubleshooting Guide to Expressing Intrinsically Disordered Proteins for Use in NMR Experiments. *Frontiers in molecular biosciences* 5: 118-118.

Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* 1(6): 2876-2890.

Gregory MA, Qi Y and Hann SR (2003a) Phosphorylation by glycogen synthase kinase-3 controls c-myc proteolysis and subnuclear localization. *J Biol Chem* 278(51): 51606-51612.

Gregory MA, Qi Y and Hann SR (2003b) Phosphorylation by Glycogen Synthase Kinase-3 Controls c-Myc Proteolysis and Subnuclear Localization *. *Journal of Biological Chemistry* 278(51): 51606-51612.

Grzesiek S and Bax A (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society* 114(16): 6291-6293.

Guito J, Gavina A, Palmeri D, et al. (2014) The cellular peptidyl-prolyl cis/trans isomerase Pin1 regulates reactivation of Kaposi's sarcoma-associated herpesvirus from latency. *Journal of virology* 88(1): 547-558.

Hafsa NE, Arndt D and Wishart DS (2015) CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. *Nucleic Acids Research* 43(W1): W370-W377.

Hanahan D (1983) Studies on transformation of Escherichia coli with plasmids. *J Mol Biol* 166(4): 557-580.

Hann SR and Eisenman RN (1984) Proteins encoded by the human c-myc oncogene: differential expression in neoplastic cells. *Molecular and Cellular Biology* 4(11): 2486-2497.

Hao B, Oehlmann S, Sowa ME, et al. (2007) Structure of a Fbw7-Skp1-Cyclin E Complex: Multisite-Phosphorylated Substrate Recognition by SCF Ubiquitin Ligases. *Molecular Cell* 26(1): 131-143.

Hegemann B, Hutchins JRA, Hudecz O, et al. (2011) Systematic Phosphorylation Analysis of Human Mitotic Protein Complexes. *Science Signaling* 4(198): rs12-rs12.

Heinrich S, Geissen E-M, Kamenz J, et al. (2013) Determinants of robustness in spindle assembly checkpoint signalling. *Nature Cell Biology* 15(11): 1328-1339.

Helander S, Montecchio M, Pilstål R, et al. (2015) Pre-Anchoring of Pin1 to Unphosphorylated c-Myc in a Fuzzy Complex Regulates c-Myc Activity. *Structure* 23(12): 2267-2279.

Hellman M, Piirainen H, Jaakola V-P, et al. (2014) Bridge over troubled proline: assignment of intrinsically disordered proteins using (HCA)CON(CAN)H and (HCA)N(CA)CO(N)H experiments concomitantly with HNCO and i(HCA)CO(CA)NH. *Journal of Biomolecular NMR* 58(1): 49-60.

Henriksson M, Bakardjiev A, Klein G, et al. (1993) Phosphorylation sites mapping in the N-terminal domain of c-myc modulate its transforming potential. *Oncogene* 8(12): 3199-3209.

Herold S, Wanzel M, Beuger V, et al. (2002) Negative regulation of the mammalian UV response by Myc through association with Miz-1. *Mol Cell* 10(3): 509-521.

Hershko A and Ciechanover A (1998) THE UBIQUITIN SYSTEM. *Annual Review of Biochemistry* 67(1): 425-479.

Higman V (2012) *Protein NMR*. Available at: https://www.protein-nmr.org.uk/.

Hovmöller S, Zhou T and Ohlson T (2002) Conformations of amino acids in proteins. *Acta Crystallogr D Biol Crystallogr* 58(Pt 5): 768-776.

Hurlin PJ, Quéva C and Eisenman RN (1997) Mnt, a novel Max-interacting protein is coexpressed with Myc in proliferating cells and mediates repression at Myc binding sites. *Genes Dev* 11(1): 44-58.

Hydbring P and Larsson LG (2010) Tipping the balance: Cdk2 enables Myc to suppress senescence. *Cancer research* 70(17): 6687-6691.

Iakoucheva LM, Brown CJ, Lawson JD, et al. (2002) Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *Journal of Molecular Biology* 323(3): 573-584.

Imai K and Mitaku S (2005) Mechanisms of secondary structure breakers in soluble proteins. *Biophysics (Nagoya-shi)* 1: 55-65.

Iritani BM and Eisenman RN (1999) c-Myc enhances protein synthesis and cell size during B lymphocyte development. *Proceedings of the National Academy of Sciences of the United States of America* 96(23): 13180-13185.

Irwin MS and Park JR (2015) Neuroblastoma: paradigm for precision medicine. *Pediatr Clin North Am* 62(1): 225-256.

Jacobs JJ, Scheijen B, Voncken JW, et al. (1999) Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. *Genes Dev* 13(20): 2678-2690.

Jaenicke LA, von Eyss B, Carstensen A, et al. (2016) Ubiquitin-Dependent Turnover of MYC Antagonizes MYC/PAF1C Complex Accumulation to Drive Transcriptional Elongation. *Mol Cell* 61(1): 54-67.

Jensen MR, Ruigrok RWH and Blackledge M (2013) Describing intrinsically disordered proteins at atomic resolution by NMR. *Current Opinion in Structural Biology* 23(3): 426-435.

Jia L, Bickel JS, Wu J, et al. (2011) RBX1 (RING box protein 1) E3 ubiquitin ligase is required for genomic integrity by modulating DNA replication licensing proteins. *J Biol Chem* 286(5): 3379-3386.

Johnston LA, Prober DA, Edgar BA, et al. (1999) Drosophila myc regulates cellular growth during development. *Cell* 98(6): 779-790.

Jones D (2022) *PSIPRED v 3.2* Available at: http://bioinf.cs.ucl.ac.uk/psipred/.

Jumper J, Evans R, Pritzel A, et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873): 583-589.

Jung DY, Lee H and Suk K (2005) Pro-apoptotic activity of N-myc in activation-induced cell death of microglia. *J Neurochem* 94(1): 249-256.

Kalkat M, Resetca D, Lourenco C, et al. (2018) MYC Protein Interactome Profiling Reveals Functionally Distinct Regions that Cooperate to Drive Tumorigenesis. *Mol Cell* 72(5): 836-848.e837.

Kanazawa S, Soucek L, Evan G, et al. (2003) c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. *Oncogene* 22(36): 5707-5711.

Kato GJ, Barrett J, Villa-Garcia M, et al. (1990) An amino-terminal c-myc domain required for neoplastic transformation activates transcription. *Mol Cell Biol* 10(11): 5914-5920.

Kay LE, Ikura M, Tschudin R, et al. (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* 89(3): 496-514.

Kay LE, Torchia DA and Bax A (1989) Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28(23): 8972-8979.

Kazi A, Xiang S, Yang H, et al. (2018) GSK3 suppression upregulates β-catenin and c-Myc to abrogate KRas-dependent tumors. *Nat Commun* 9(1): 5154.

Keeler J (2004) Understanding NMR Spectroscopy. University of Barcelona.

Kettenbach AN, Schweppe DK, Faherty BK, et al. (2011) Quantitative phosphoproteomics identifies substrates and functional modules of Aurora and Polo-like kinase activities in mitotic cells. *Sci Signal* 4(179): rs5.

Kim JW, Tchernyshyov I, Semenza GL, et al. (2006) HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. *Cell Metab* 3(3): 177-185.

King N, Westbrook MJ, Young SL, et al. (2008) The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* 451(7180): 783-788.

Kjaergaard M and Poulsen FM (2011) Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *Journal of Biomolecular NMR* 50(2): 157-165.

Knoepfler PS, Cheng PF and Eisenman RN (2002) N-myc is essential during neurogenesis for the rapid expansion of progenitor cell populations and the inhibition of neuronal differentiation. *Genes & Development* 16(20): 2699-2712.

Knoepfler PS, Zhang X-y, Cheng PF, et al. (2006) Myc influences global chromatin structure. *The EMBO Journal* 25(12): 2723-2734.

Koch HB, Zhang R, Verdoodt B, et al. (2007) Large-scale identification of c-MYC-associated proteins using a combined TAP/MudPIT approach. *Cell Cycle* 6(2): 205-217.

Kohl NE, Kanda N, Schreck RR, et al. (1983) Transposition and amplification of oncogene-related sequences in human neuroblastomas. *Cell* 35(2, Part 1): 359-367.

Konrat R (2014) NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J Magn Reson* 241(100): 74-85.

Kornev AP, Haste NM, Taylor SS, et al. (2006) Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proceedings of the National Academy of Sciences* 103(47): 17783.

Kosol S, Contreras-Martos S, Cedeño C, et al. (2013) Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* 18(9): 10802-10828.

Kothe M, Kohls D, Low S, et al. (2007) Research Article: Selectivity-determining Residues in Plk1. *Chemical Biology & Drug Design* 70(6): 540-546.

Kovermann M, Rogne P and Wolf-Watz M (2016) Protein dynamics and function from solution state NMR spectroscopy. *Q Rev Biophys* 49: e6.

Kress TR, Sabò A and Amati B (2015) MYC: connecting selective transcriptional control to global RNA production. *Nature Reviews Cancer* 15(10): 593-607.

Krzywda S, Murshudov GN, Brzozowski AM, et al. (1998) Stabilizing bound O2 in myoglobin by valine68 (E11) to asparagine substitution. *Biochemistry* 37(45): 15896-15907.

Kukic P, Farrell D, McIntosh LP, et al. (2013) Protein Dielectric Constants Determined from NMR Chemical Shift Perturbations. *Journal of the American Chemical Society* 135(45): 16968-16976.

Küppers R (2005) Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev Cancer* 5(4): 251-262.

Küppers R and Dalla-Favera R (2001) Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene* 20(40): 5580-5594.

La Verde V, Dominici P and Astegno A (2017) Determination of Hydrodynamic Radius of Proteins by Size Exclusion Chromatography. *Bio Protoc* 7(8): e2230.

Landschulz W, Johnson P and McKnight S (1988a) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* 240(4860): 1759-1764.

Landschulz WH, Johnson PF and McKnight SL (1988b) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* 240(4860): 1759-1764.

Lange Oliver F, Rossi P, Sgourakis Nikolaos G, et al. (2012) Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proceedings of the National Academy of Sciences* 109(27): 10873-10878.

Lee W-H, Murphree AL and Benedict WF (1984) Expression and amplification of the N-myc gene in primary retinoblastoma. *Nature* 309(5967): 458-460.

Leone G, Sears R, Huang E, et al. (2001) Myc requires distinct E2F activities to induce S phase and apoptosis. *Mol Cell* 8(1): 105-113.

Levitt M (2013) *Spin dynamics.* Wiley.

Li S-C, Goto NK, Williams KA, et al. (1996a) α -Helical, but not β -Sheet, Propensity of Proline is Determined by Peptide Environment. *Proceedings of the National Academy of Sciences of the United States of America* 93(13): 6676-6681.

Li SC, Goto NK, Williams KA, et al. (1996b) Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. *Proceedings of the National Academy of Sciences* 93(13): 6676.

Li Y and Palmer AG (2010) Narrowing of Protein NMR Spectral Lines Broadened by Chemical Exchange. *Journal of the American Chemical Society* 132(26): 8856-8857.

Lins L and Brasseur R (1995) The hydrophobic effect in protein folding. *The FASEB Journal* 9(7): 535-540.

Lu H, Xue Y, Yu GK, et al. (2015) Compensatory induction of MYC expression by sustained CDK9 inhibition via a BRD4-dependent mechanism. *eLife* 4: e06535.

Lu K, Tao H, Si X, et al. (2018) The Histone H3 Lysine 4 Presenter WDR5 as an Oncogenic Protein and Novel Epigenetic Target in Cancer. *Frontiers in oncology* 8: 502-502.

Lu KP and Zhou XZ (2007) The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and disease. *Nat Rev Mol Cell Biol* 8(11): 904-916.

Lutterbach B and Hann SR (1994) Hierarchical phosphorylation at N-terminal transformation-sensitive sites in c-Myc protein is regulated by mitogens and in mitosis. *Mol Cell Biol* 14(8): 5510-5522.

Maciejewski MW, Schuyler AD, Gryk MR, et al. (2017) NMRbox: A Resource for Biomolecular NMR Computation. *Biophysical Journal* 112(8): 1529-1534.

Macůrek L, Lindqvist A, Lim D, et al. (2008) Polo-like kinase-1 is activated by aurora A to promote checkpoint recovery. *Nature* 455(7209): 119-123.

Mahani A, Henriksson J and Wright AP (2013) Origins of Myc proteins--using intrinsic protein disorder to trace distant relatives. *PloS one* 8(9): e75057.

Malynn BA, de Alboran IM, O'Hagan RC, et al. (2000) N-myc can functionally replace c-myc in murine development, cellular growth, and differentiation. *Genes & Development* 14(11): 1390-1399.

Mariani V, Biasini M, Barbato A, et al. (2013) lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics (Oxford, England)* 29(21): 2722-2728.

Marion D, Kay LE, Sparks SW, et al. (1989) Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *Journal of the American Chemical Society* 111(4): 1515-1517.

Martinez E, Palhan VB, Tjernberg A, et al. (2001) Human STAGA Complex Is a Chromatin-Acetylating Transcription Coactivator That Interacts with Pre-mRNA Splicing and DNA Damage-Binding Factors In Vivo. *Molecular and Cellular Biology* 21(20): 6782-6795.

Marumoto T, Zhang D and Saya H (2005) Aurora-A - a guardian of poles. *Nat Rev Cancer* 5(1): 42-50.

McMahon SB, Van Buskirk HA, Dugan KA, et al. (1998) The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. *Cell* 94(3): 363-374.

Meyer N and Penn LZ (2008) Reflecting on 25 years with MYC. *Nature Reviews Cancer* 8(12): 976-990.

Mittermaier A and Meneses E (2013) Analyzing protein-ligand interactions by dynamic NMR spectroscopy. *Methods Mol Biol* 1008: 243-266.

Modi V and Dunbrack RL (2019) Defining a new nomenclature for the structures of active and inactive kinases. *Proceedings of the National Academy of Sciences* 116(14): 6818.

Moorhead GBG, Trinkle-Mulcahy L and Ulke-Lemée A (2007) Emerging roles of nuclear protein phosphatases. *Nature Reviews Molecular Cell Biology* 8(3): 234-244.

Muhandiram DR and Kay LE (1994) Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *Journal of Magnetic Resonance, Series B* 103(3): 203-216.

Nair SK and Burley SK (2003) X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. *Cell* 112(2): 193-205.

Nakajima H, Ikeda M, Tsuchida N, et al. (1989) Inactivation of the N-myc gene product by single amino acid substitution of leucine residues located in the leucine-zipper region. *Oncogene* 4(8): 999-1002.

Nakajima H, Toyoshima-Morimoto F, Taniguchi E, et al. (2003) Identification of a consensus motif for Plk (Polo-like kinase) phosphorylation reveals Myt1 as a Plk1 substrate. *J Biol Chem* 278(28): 25277-25280.

Nau MM, Brooks BJ, Battey J, et al. (1985) L-myc, a new myc-related gene amplified and expressed in human small cell lung cancer. *Nature* 318(6041): 69-73.

Neri F, Zippo A, Krepelova A, et al. (2012) Myc regulates the transcription of the PRC2 gene to control the expression of developmental genes in embryonic stem cells. *Mol Cell Biol* 32(4): 840-851.

Nogales E, Louder RK and He Y (2017) Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annu Rev Biophys* 46: 59-83.

Noguchi K, Kitanaka C, Yamana H, et al. (1999) Regulation of c-Myc through phosphorylation at Ser-62 and Ser-71 by c-Jun N-terminal kinase. *J Biol Chem* 274(46): 32580-32587.

Northcott PA, Jones DTW, Kool M, et al. (2012) Medulloblastomics: the end of the beginning. *Nature Reviews Cancer* 12(12): 818-834.

Okubo T, Knoepfler PS, Eisenman RN, et al. (2005) Nmyc plays an essential role during lung development as a dosage-sensitive regulator of progenitor cell proliferation and differentiation. *Development* 132(6): 1363-1374.

Okuda M and Nishimura Y (2014) Extended String Binding Mode of the Phosphorylated Transactivation Domain of Tumor Suppressor p53. *Journal of the American Chemical Society* 136(40): 14143-14152.

Otto T, Horn S, Brockmann M, et al. (2009) Stabilization of N-Myc is a critical function of Aurora A in human neuroblastoma. *Cancer Cell* 15(1): 67-78.

Overhauser AW (1953) Polarization of Nuclei in Metals. *Physical Review* 92(2): 411-415.

Pace CN and Scholtz JM (1998) A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical Journal* 75(1): 422-427.

Palmieri L and Rastelli G (2013) αC helix displacement as a general approach for allosteric modulation of protein kinases. *Drug Discov Today* 18(7-8): 407-414.

Papoian GA (2008) Proteins with weakly funneled energy landscapes challenge the classical structure–function paradigm. *Proceedings of the National Academy of Sciences* 105(38): 14237.

Park J-E, Soung N-K, Johmura Y, et al. (2010) Polo-box domain: a versatile mediator of polo-like kinase function. *Cellular and Molecular Life Sciences* 67(12): 1957-1970.

Patange S, Ball DA, Wan Y, et al. (2022) MYC amplifies gene expression through global changes in transcription factor dynamics. *Cell Reports* 38(4): 110292.

Peifer M, Hertwig F, Roels F, et al. (2015) Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* 526(7575): 700-704.

Pervushin K, Riek R, Wider G, et al. (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America* 94(23): 12366-12371.

Petroni M, Veschi V, Prodosmo A, et al. (2011) MYCN Sensitizes Human Neuroblastoma to Apoptosis by HIPK2 Activation through a DNA Damage Response. *Molecular Cancer Research* 9(1): 67-77.

Petroski MD and Deshaies RJ (2005) Function and regulation of cullin–RING ubiquitin ligases. *Nature Reviews Molecular Cell Biology* 6(1): 9-20.

Peukert K, Staller P, Schneider A, et al. (1997) An alternative pathway for gene regulation by Myc. *The EMBO Journal* 16(18): 5672-5686.

Piskacek M, Havelka M, Jendruchova K, et al. (2020) The evolution of the 9aaTAD domain in Sp2 proteins: inactivation with valines and intron reservoirs. *Cell Mol Life Sci* 77(9): 1793-1810.

Popov N, Schülein C, Jaenicke LA, et al. (2010) Ubiquitylation of the amino terminus of Myc by SCFβ-TrCP antagonizes SCFFbw7-mediated turnover. *Nature Cell Biology* 12(10): 973-981.

Prendergast GC, Lawe D and Ziff EB (1991) Association of Myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. *Cell* 65(3): 395-407.

Pugh TJ, Morozova O, Attiyeh EF, et al. (2013) The genetic landscape of high-risk neuroblastoma. *Nature Genetics* 45(3): 279-284.

Pulverer BJ, Fisher C, Vousden K, et al. (1994) Site-specific modulation of c-Myc cotransformation by residues phosphorylated in vivo. *Oncogene* 9(1): 59-70.

Raab M, Matthess Y, Raab CA, et al. (2022) A dimerization-dependent mechanism regulates enzymatic activation and nuclear entry of PLK1. *Oncogene* 41(3): 372-386.

Radivojac P, Iakoucheva LM, Oldfield CJ, et al. (2007) Intrinsic Disorder and Functional Proteomics. *Biophysical Journal* 92(5): 1439-1456.

Rahl PB, Lin CY, Seila AC, et al. (2010) c-Myc regulates transcriptional pause release. *Cell* 141(3): 432-445.

Rahl PB and Young RA (2014) MYC and transcription elongation. *Cold Spring Harbor perspectives in medicine* 4(1): a020990.

Ramazi S and Zahiri J (2021) Post-translational modifications in proteins: resources, tools and prediction methods. *Database* 2021: baab012.

Richards MW, Burgess SG, Poon E, et al. (2016) Structural basis of N-Myc binding by Aurora-A and its destabilization by kinase inhibitors. *Proceedings of the National Academy of Sciences* 113(48): 13726-13731.

Rickman DS, Schulte JH and Eilers M (2018) The Expanding World of N-MYC-Driven Tumors. *Cancer Discov* 8(2): 150-163.

Rogers S, Wells R and Rechsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* 234(4774): 364-368.

Rosano GL and Ceccarelli EA (2014) Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology* 5: 172.

Roskoski Jr R (2016) Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. *Pharmacological research* 103: 26-48.

Roussel M, Saule S, Lagrou C, et al. (1979) Three new types of viral oncogene of cellular origin specific for haematopoietic cell transformation. *Nature* 281(5731): 452-455.

Sahu D, Bastidas M and Showalter SA (2014) Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods. *Analytical biochemistry* 449: 17-25.

Sainsbury S, Bernecky C and Cramer P (2015) Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology* 16(3): 129-143.

Saleh A, Schieltz D, Ting N, et al. (1998) Tra1p is a component of the yeast Ada.Spt transcriptional regulatory complexes. *J Biol Chem* 273(41): 26559-26565.

Sattler M and Fesik SW (1996) Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure* 4(11): 1245-1249.

Sawai S, Shimono A, Wakamatsu Y, et al. (1993) Defects of embryonic organogenesis resulting from targeted disruption of the N-myc gene in the mouse. *Development* 117(4): 1445-1455.

Schafer JM, Lehmann BD, Gonzalez-Ericsson PI, et al. (2020) Targeting MYCN-expressing triple-negative breast cancer with BET and MEK inhibitors. *Science Translational Medicine* 12(534): eaaw8275.

Schanda P, Forge V and Brutscher B (2006) HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. *Magn Reson Chem* 44 Spec No: S177-184.

Schlessinger A, Schaefer C, Vicedo E, et al. (2011) Protein disorder—A breakthrough invention of evolution? *Current Opinion in Structural Biology* 21: 412-418.

Schmidt EV (1999) The role of c-myc in cellular growth control. *Oncogene* 18(19): 2988-2996.

Schuhmacher M, Staege MS, Pajic A, et al. (1999) Control of cell growth by c-Myc in the absence of cell division. *Curr Biol* 9(21): 1255-1258.

Schulman BA, Carrano AC, Jeffrey PD, et al. (2000) Insights into SCF ubiquitin ligases from the structure of the Skp1–Skp2 complex. *Nature* 408(6810): 381-386.

Schulze-Gahmen U, Upton H, Birnberg A, et al. (2013) The AFF4 scaffold binds human P-TEFb adjacent to HIV Tat. *eLife* 2: e00327.

Schwab M, Alitalo K, Klempnauer K-H, et al. (1983) Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature* 305(5931): 245-248.

Schwarzinger S, Kroon GJA, Foss TR, et al. (2001) Sequence-Dependent Correction of Random Coil NMR Chemical Shifts. *Journal of the American Chemical Society* 123(13): 2970-2978.

Sears R, Nuckolls F, Haura E, et al. (2000) Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. *Genes & Development* 14(19): 2501-2514.

Seeger RC, Brodeur GM, Sather H, et al. (1985) Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas. *N Engl J Med* 313(18): 1111-1116.

Seki A, Coppinger JA, Jang CY, et al. (2008) Bora and the kinase Aurora a cooperatively activate the kinase Plk1 and control mitotic entry. *Science* 320(5883): 1655-1658.

Selenko P, Frueh DP, Elsaesser SJ, et al. (2008) In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. *Nat Struct Mol Biol* 15(3): 321-329.

Seo HR, Kim J, Bae S, et al. (2008) Cdk5-mediated phosphorylation of c-Myc on Ser-62 is essential in transcriptional activation of cyclin B1 by cyclin G1. *J Biol Chem* 283(23): 15601-15610.

Seong Y-S, Kamijo K, Lee J-S, et al. (2002) A Spindle Checkpoint Arrest and a Cytokinesis Failure by the Dominant-negative Polo-box Domain of Plk1 in U-2 OS Cells*210. *Journal of Biological Chemistry* 277(35): 32282-32293.

Sharma R, Sharma A, Patil A, et al. (2019) Discovering MoRFs by trisecting intrinsically disordered protein sequence into terminals and middle regions. *BMC Bioinformatics* 19(13): 378.

Sheiness D and Bishop JM (1979) DNA and RNA from Uninfected Vertebrate Cells Contain Nucleotide Sequences Related to the Putative Transforming Gene of Avian Myelocytomatosis Virus. *Journal of virology* 31(2): 514-521.

Shen Y and Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* 56(3): 227-241.

Shen Y, Bryan PN, He Y, et al. (2010) De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. *Protein science : a publication of the Protein Society* 19(2): 349-356.

Shen Y, Delaglio F, Cornilescu G, et al. (2009a) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR* 44(4): 213-223.

Shen Y, Lange O, Delaglio F, et al. (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proceedings of the National Academy of Sciences* 105(12): 4685.

Shen Y, Vernon R, Baker D, et al. (2009b) De novo protein structure generation from incomplete chemical shift assignments. *Journal of Biomolecular NMR* 43(2): 63-78.

Shetty A, Kallgren SP, Demel C, et al. (2017) Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. *Molecular Cell* 66(1): 77-88.e75.

Siegel R, Naishadham D and Jemal A (2012) Cancer statistics, 2012. *CA: A Cancer Journal for Clinicians* 62(1): 10-29.

Sjostrom SK, Finn G, Hahn WC, et al. (2005) The Cdk1 Complex Plays a Prime Role in Regulating N-Myc Phosphorylation and Turnover in Neural Precursors. *Developmental Cell* 9(3): 327-338.

Smith GE, Summers MD and Fraser MJ (1983) Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol Cell Biol* 3(12): 2156-2165.

Solyom Ba (2017) Fast NMR Data Acquisition: Beyonnd the Fourier Transform.

Souza DP, Andrade MO, Alvarez-Martinez CE, et al. (2011) A Component of the Xanthomonadaceae Type IV Secretion System Combines a VirB7 Motif with a N0 Domain Found in Outer Membrane Transport Proteins. *PLOS Pathogens* 7(5): e1002031.

Staller P, Peukert K, Kiermaier A, et al. (2001) Repression of p15INK4b expression by Myc through association with Miz-1. *Nature Cell Biology* 3(4): 392-399.

Stanton BR, Perkins AS, Tessarollo L, et al. (1992) Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop. *Genes Dev* 6(12a): 2235-2247.

Sutherland C (2011) What Are the bona fide GSK3 Substrates? *Int J Alzheimers Dis* 2011: 505607.

Takahashi K and Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126(4): 663-676.

Takeuchi K, Heffron G, Sun Z-YJ, et al. (2010) Nitrogen-detected CAN and CON experiments as alternative experiments for main chain NMR resonance assignments. *Journal of Biomolecular NMR* 47(4): 271-282.

Tan J, Li Z, Lee PL, et al. (2013) PDK1 Signaling Toward PLK1–MYC Activation Confers Oncogenic Transformation, Tumor-Initiating Cell Activation, and Resistance to mTOR-Targeted Therapy. *Cancer Discovery* 3(10): 1156.

Tate JG, Bamford S, Jubb HC, et al. (2019) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* 47(D1): D941-d947.

Tavernier N, Thomas Y, Vigneron S, et al. (2021) Bora phosphorylation substitutes in trans for T-loop phosphorylation in Aurora A to promote mitotic entry. *Nature Communications* 12(1): 1899.

Thomas LR, Wang Q, Grieb BC, et al. (2015) Interaction with WDR5 promotes target gene recognition and tumorigenesis by MYC. *Mol Cell* 58(3): 440-452.

Thompson RF, Walker M, Siebert CA, et al. (2016) An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* 100: 3-15.

Tompa P and Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends in Biochemical Sciences* 33(1): 2-8.

Tompa P, Schad E, Tantos A, et al. (2015) Intrinsically disordered proteins: emerging interaction specialists. *Current Opinion in Structural Biology* 35: 49-59.

Tzankov A, Xu-Monette ZY, Gerhard M, et al. (2014) Rearrangements of MYC gene facilitate risk stratification in diffuse large B-cell lymphoma patients treated with rituximab-CHOP. *Mod Pathol* 27(7): 958-971.

Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein science : a publication of the Protein Society* 11(4): 739-756.

Uversky VN (2009) Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J* 28(7-8): 305-325.

Uversky VN (2013) Unusual biophysics of intrinsically disordered proteins. *Biochimica et biophysica acta* 1834(5): 932-951.

Uversky VN (2015) The multifaceted roles of intrinsic disorder in protein complexes. *FEBS Letters* 589(19PartA): 2498-2506.

Uversky VN, Gillespie JR and Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41(3): 415-427.

Vallurupalli P (2009) Chemical Exchange.

Varadi M, Anyango S, Deshpande M, et al. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* 50(D1): D439-D444.

Venkatesh S and Workman JL (2015) Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews Molecular Cell Biology* 16(3): 178-189.

Vennstrom B, Sheiness D, Zabielski J, et al. (1982) Isolation and characterization of c-myc, a cellular homolog of the oncogene (v-myc) of avian myelocytomatosis virus strain 29. *Journal of virology* 42(3): 773-779.

Virshup DM and Shenolikar S (2009) From promiscuity to precision: protein phosphatases get a makeover. *Mol Cell* 33(5): 537-545.

Vogt PK, Bos TJ and Doolittle RF (1987) Homology between the DNA-binding domain of the GCN4 regulatory protein of yeast and the carboxyl-terminal region of a protein coded for by the oncogene jun. *Proceedings of the National Academy of Sciences of the United States of America* 84(10): 3316-3319.

Vranken WF, Boucher W, Stevens TJ, et al. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59(4): 687-696.

Wei Y, Resetca D, Li Z, et al. (2019) Multiple direct interactions of TBP with the MYC oncoprotein. *Nature Structural & Molecular Biology* 26(11): 1035-1043.

Weiss MS, Jabs A and Hilgenfeld R (1998) Peptide bonds revisited. *Nature structural biology* 5(8): 676-676.

Weissman AM (2001) Themes and variations on ubiquitylation. *Nat Rev Mol Cell Biol* 2(3): 169-178.

Welcker M, Orian A, Jin J, et al. (2004) The Fbw7 tumor suppressor regulates glycogen synthase kinase 3 phosphorylation-dependent c-Myc protein degradation. *Proceedings of the National Academy of Sciences of the United States of America* 101(24): 9085-9090.

Welcker M, Wang B, Rusnac DV, et al. (2022) Two diphosphorylated degrons control c-Myc degradation by the Fbw7 tumor suppressor. *Sci Adv* 8(4): eabl7872.

Wernig M, Meissner A, Cassady JP, et al. (2008) c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cell* 2(1): 10-12.

Wieczorek E, Brand M, Jacq X, et al. (1998) Function of TAFII-containing complex without TBP in transcription by RNA polymerase II. *Nature* 393(6681): 187-191.

Williamson D, Lu YJ, Gordon T, et al. (2005) Relationship between MYCN copy number and expression in rhabdomyosarcomas and correlation with adverse prognosis in the alveolar subtype. *J Clin Oncol* 23(4): 880-888.

Williamson M (2009) Applications of the NOE in Molecular Biology. *Annual Reports on NMR Spectroscopy* 65: 77-109.

Wishart DS, Sykes BD and Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* 31(6): 1647-1651.

Woody RW (1995) Circular dichroism. *Methods in Enzymology*. Academic Press, pp.34-71.

Wright PE and Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19(1): 31-38.

Wright PE and Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology* 16(1): 18-29.

Wu F, Wang P, Zhang J, et al. (2010) Studies of phosphoproteomic changes induced by nucleophosmin-anaplastic lymphoma kinase (ALK) highlight deregulation of tumor necrosis factor (TNF)/Fas/TNF-related apoptosis-induced ligand signaling pathway in ALK-positive anaplastic large cell lymphoma. *Mol Cell Proteomics* 9(7): 1616-1632.

Wu S, Cetinkaya C, Munoz-Alonso MJ, et al. (2003) Myc represses differentiation-induced p21CIP1 expression via Miz-1-dependent interaction with the p21 core promoter. *Oncogene* 22(3): 351-360.

Xia Y and Zhang X (2020) The Spectrum of MYC Alterations in Diffuse Large B-Cell Lymphoma. *Acta Haematologica* 143(6): 520-528.

Xiao D, Yue M, Su H, et al. (2016) Polo-like Kinase-1 Regulates Myc Stabilization and Activates a Feedforward Circuit Promoting Tumor Cell Survival. *Mol Cell* 64(3): 493-506.

Xiao Q, Claassen G, Shi J, et al. (1998) Transactivation-defective c-MycS retains the ability to regulate proliferation and apoptosis. *Genes Dev* 12(24): 3803-3808.

Yada M, Hatakeyama S, Kamura T, et al. (2004) Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. *Embo j* 23(10): 2116-2125.

Yao J, Chung J, Eliezer D, et al. (2001) NMR Structural and Dynamic Characterization of the Acid-Unfolded State of Apomyoglobin Provides Insights into the Early Events in Protein Folding. *Biochemistry* 40(12): 3561-3571.

Yeh E, Cunningham M, Arnold H, et al. (2004) A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells. *Nat Cell Biol* 6(4): 308-318.

Yoon SH, Kim SK and Kim JF (2010) Secretory production of recombinant proteins in Escherichia coli. *Recent Pat Biotechnol* 4(1): 23-29.

Young SL, Diolaiti D, Conacci-Sorrell M, et al. (2011) Premetazoan ancestry of the Myc-Max network. *Mol Biol Evol* 28(10): 2961-2971.

Zeller KI, Zhao X, Lee CW, et al. (2006) Global mapping of c-Myc binding sites and target gene networks in human B cells. *Proceedings of the National Academy of Sciences of the United States of America* 103(47): 17834-17839.

Zerbe O and Bader R (no date) *Peptide NMR*. Available at: https://www.chem.uzh.ch/zerbe/PeptidNMR.pdf (accessed 27.10.2022).

Zhang Q, Throolin R, Pitt SW, et al. (2003) Probing Motions between Equivalent RNA Domains Using Magnetic Field Induced Residual Dipolar Couplings: Accounting for Correlations between Motions and Alignment. *Journal of the American Chemical Society* 125(35): 10530-10531.

Zheng N, Schulman BA, Song L, et al. (2002) Structure of the Cul1–Rbx1–Skp1–F boxSkp2 SCF ubiquitin ligase complex. *Nature* 416(6882): 703-709.

Zhou H, Di Palma S, Preisinger C, et al. (2013) Toward a Comprehensive Characterization of a Human Cancer Cell Phosphoproteome. *Journal of Proteome Research* 12(1): 260-271.

Zhou XZ, Kops O, Werner A, et al. (2000) Pin1-Dependent Prolyl Isomerization Regulates Dephosphorylation of Cdc25C and Tau Proteins. *Molecular Cell* 6(4): 873-883.

Zhou Z, He C and Wang J (2015) Regulation mechanism of Fbxw7-related signaling pathways (Review). *Oncol Rep* 34(5): 2215-2224.

Zhu B, Mandal SS, Pham AD, et al. (2005) The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes Dev* 19(14): 1668-1673.

Zhu C, Gao Y, Li H, et al. (2016) Characterizing hydrophobicity of amino acid side chains in a protein environment via measuring contact angle of a water nanodroplet on planar peptide network. *Proceedings of the National Academy of Sciences of the United States of America* 113(46): 12946-12951.

Zimmerman KA, Yancopoulos GD, Collum RG, et al. (1986) Differential expression of myc family genes during murine development. *Nature* 319(6056): 780-783.

Zuiderweg ER and Fesik SW (1989) Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry* 28(6): 2387-2391.