Linköping University | Department of Physics, Chemistry and Biology
Master thesis, 30 hp | Educational Program: Physics, Chemistry and Biology
Spring term 2021 | LITH-IFM-A-EX—21/3981--SE

Investigating Minor States of the Oncoprotein N-MYC, with Focus on Proline Cis/Trans Isomerisation using NMR Spectroscopy

Frida Haugskott

Examinator, Maria Sunnerhagen
Tutor, Alexandra Ahlner

Avdelning, institution
Division, Department

Department of Physics, Chemistry and Biology
Linköping University

Datum
Date
2021-06-18

Språk
Language

☐ Svenska/Swedish
☒ Engelska/English

☐ _____

Rapporttyp
Report category

☐ Licentiatavhandling
☒ Examensarbete
☐ C-uppsats
☐ D-uppsats
☐ Övrig rapport

☐ _____

ISBN

ISRN: LITH-IFM-A-EX--21/3981--SE

Serietitel och serienummer **ISSN**
Title of series, numbering _____

URL för elektronisk version

Titel
Title

Investigating Minor States of the Oncoprotein N-MYC, with Focus on Proline Cis/Trans Isomerisation using NMR Spectroscopy

Författare
Author

Frida Haugskott

Sammanfattning
Abstract

MYC is a family of three regulator genes that codes for transcription factors. Expression of Myc proteins from MYC genes is found to be deregulated in 70 % of all cancer forms. The three human homologs C-Myc, N-Myc and L-Myc are mainly associated with cancer in the lymphatic system, nerve tissues and lung cancer, respectively.

Even though N-Myc is associated with Neuroblastoma, the cancer variant that is most common among children, the field is focused towards C-Myc. The activation of C-Myc begins with phosphorylation of Serine 62, followed by trans-to-cis isomerisation of Proline 63. Then Threonine 58 becomes phosphorylated leading to that Serine 62 is dephosphorylated and subsequent cis-to-trans isomerisation of Proline 63, and C-Myc is marked for degradation.

Cis-trans isomerisation is necessary for regulation of gene expression, and is therefore important to understand. Since N-Myc and C-Myc have identical sequences between residues 47 to residue 69, the hypothesis is that N-Myc is activated in the same manner, but this has not been confirmed.

In this project the first 69 amino acids of N-Myc were analysed with NMR spectroscopy. This resulted in a near complete assignment of the major conformation, and of the alternative minor conformations as well. The traditional assignment experiments HNCACB, HN(CO)CACB, HNCO, HN(CA)CO in combination with CCH-TOCSY and HN(CCO)C revealed that the majority of the minor configurations can be explained by cis/trans isomerisation of prolines. In addition, the protein was analysed with direct carbon detected NMR spectroscopy to be able to detect the prolines.

Nyckelord
Keyword
Myc, N-Myc, Intrinsically disordered protein (IDP), Proline, cis-trans isomerization, minor conformation, NMR, Cancer

Announcement

Due to the Global Covid-19 pandemic, some parts of the laboratory work for this thesis has been adapted to meet the rules and recommendations both from Linköping University and the Public Health Agency of Sweden. For example, the protein purification was made after the NMR measurements, to ensure that all steps of the investigation were learned. The protein measured by NMR was made by research preparatory student Johanna Hultman before the start of this thesis, to ensure that the critical NMR experiments could be pursued within the time range of this thesis. The quality of the protein preparation made by Frida Haugskott later during the thesis was excellent.

Abstract

MYC is a family of three regulator genes that codes for transcription factors. Expression of Myc proteins from MYC genes is found to be deregulated in 70 % of all cancer forms. The three human homologs C-Myc, N-Myc and L-Myc are mainly associated with cancer in the lymphatic system, nerve tissues and lung cancer, respectively.

Even though N-Myc is associated with Neuroblastoma, the cancer variant that is most common among children, the field is focused towards C-Myc. The activation of C-Myc begins with phosphorylation of Serine 62, followed by trans-to-cis isomerisation of Proline 63. Then Threonine 58 becomes phosphorylated leading to that Serine 62 is dephosphorylated and subsequent cis-to-trans isomerisation of Proline 63, and C-Myc is marked for degradation.

Cis-trans isomerisation is necessary for regulation of gene expression, and is therefore important to understand. Since N-Myc and C-Myc have identical sequences between residues 47 to residue 69, the hypothesis is that N-Myc is activated in the same manner, but this has not been confirmed.

In this project the first 69 amino acids of N-Myc were analysed with NMR spectroscopy. This resulted in a near complete assignment of the major conformation, and of the alternative minor conformations as well. The traditional assignment experiments HNCACB, HN(CO)CACB, HNCO, HN(CA)CO in combination with CCH-TOCSY and HN(CCO)C revealed that the majority of the minor configurations can be explained by cis/trans isomerisation of prolines. In addition, the protein was analysed with direct carbon detected NMR spectroscopy to be able to detect the prolines.

Acronyms and Abbreviations

A – Alanine
AE – Anion exchange chromatography
bHLH – basic helix loop helix
C – Cysteine
CSP – Chemical shift perturbation
D – Aspartic acid
E – Glutamic acid
F – Phenylalanine
G – Glycine
H – Histidine
HSQC – Heteronuclear single quantum coherence Myc-
I – Isoleucine
IDP – intrinsically disordered region
L – Lysine
L – Leucine
M – Methionine
Max – Myc associated factor X
N – Asparagine
NMR – Nuclear magnetic resonance
ON – Over night
P – Proline
P44L – N-Myc 1-69 P44L mutant
Q – Glutamine
R – Arginine
S – Serine
SSP – Secondary structure propensity
T – Threonine
V – Valine
W – Tryptophan
WT – Wild type
WT – N-MYC 1-69 wild type
Y – Tyrosine

Acronyms and Abbreviations

1 Introduction

1.1 Why do we need to study proteins?

The human Genome consists of roughly 3 billion base pairs and approximately 25000 protein coding genes. In a healthy and functioning body, the regulation of which genes that are expressed and the levels of proteins in the body is strictly regulated, and the amount of a particular protein in a cell is balanced between the rate of synthesis and the biochemical degradation pathways. (Contents of Essentials of Cell Biology, 2014).

However, many diseases origins from misfunction of transcription factors resulting in deregulation of genes and/or misfunction of proteins. In Sweden alone 65000 persons was diagnosed with Cancer in 2017, and Cancer was the main cause of death for people under 80 years of age. Cancer is not one disease, Cancer is an collective name for around 200 different diseases, that origins from uncontrolled cell growth and cell division. If untreated the cell growth can lead in to tumors and potentially death. The cancer treatments that are common today is surgery, immunotherapy, radiation therapy and chemotherapy, whom all have different side effects and affects non-carcinogenic cells badly. (Cancerfonden, 2021; World health organisation, 2021)

The Myc family consists of three proteins coded from different MYC genes, which are involved in around 70 % of all cancer forms (Duffy et al. 2021). Myc is a transcription factor that promotes healthy cell growth, differentiation and proliferation when correctly regulated. When deregulated Myc instead drive tumorigenesis (Hanahan and Weinberg, 2011). One homolog from the Myc family, C-Myc, is activated by phosphorylation of Serine 62, followed by trans to cis isomerisation of Proline 63. Then Threonine 58 becomes phosphorylated, leading to that Serine 62 is dephosphorylated, and cis to trans isomerisation of Proline 63, and then the protein is degraded by targeting to the proteasome (Cohn *et al*., 2020).

1.2 Aim of this project

The aim of this thesis is to investigate the minor states of N-Myc 1-69^{WT}, and couple the minor states to the major states. The family of Myc proteins is associated with many types of cancer. Better understanding on how the different states of the protein are populated, what leads to the formation of minor states, and what function these states may have, will hopefully in the future lead to the finding of a cure to cancer. NMR is uniquely able to identify and characterise minor states in solution. In this thesis, I have analysed the first 69 N-terminal amino acids of N-Myc, comprising the phosphorylated Myc degron, with nuclear magnetic resonance spectroscopy.

2 Theory

2.1 Intrinsically Disordered proteins

Historically the general understanding was that a fixed three-dimensional structure was necessary for biological functions of a protein. This was questioned and abandoned in the 70's due to increasing knowledge about protein dynamics (Bu and Callaway, 2011). It is today accepted that proteins exists in ensembles of structures with either intrinsically disordered regions (IDR) or that the whole protein lacks a constrained structure, an intrinsically disordered protein (IDP). These highly dynamic regions can be linked to allosteric regulation and enzyme catalysis (Bu and Callaway, 2011).

Intrinsically disordered proteins are characterised by their lack of a stable structure, instead they tend to exist in conformational ensembles. Hydrophilic and charged amino acids, like *P, E, S, Q, K* are promoting disorder in a structure, while hydrophilic and uncharged amino acids, like *C, W, L, I, Y, F, M, H, N*, are known to instead promote structure, and *A, G, R, D, T* are indifferent to structure or disorder. (Oldfield & Dunker, 2014). The amino acids can be ordered, from order promoting to disorder promoting, according to W, F, Y, I, M, L, V, N, C, T, A, G, R, D, H, Q, K, S, E, P (Campen et al. 2008).

IDRs and IDPs are common in disease related proteins, causing for example Cancer, diabetes, cardiovascular disease, neurodegenerative diseases and amyloidosis. Two well-known IDPs are alpha-synuclein and tau, both involved in Alzheimer's disease. (Uversky et al. 2008)

IDPs are very heterogenous and are therefore challenging to characterize, they are protease hypersensitive and their abnormal migration in SDS-Page can be used as an indicator of disorder (Schramm, et al., 2019). NMR spectroscopy has played an important role in understanding the intrinsically disordered native state and also in understanding how intrinsically regions contribute to biological functions. (Gibbs *et al*., 2017)

2.2 The Myc Family

The Myc family comprises three human homologs C-Myc, N-Myc and L-Myc, which are global regulator genes that codes for transcription factors. C-Myc was first discovered and plays an important role in Burkitt Lymphoma, a variant of cancer in the lymphatic system. Overexpression of N-Myc is mainly associated with neuroblastoma, meaning cancer in the nerve tissues, especially common among children. Excess L-Myc is often found in patients with lung cancer. And this is not all, Myc is deregulated in approximately 70% of all cancer forms (Duffy et al. 2021). Myc is classified as an intrinsically disordered protein, and the sequence among with the composition of each amino acid can be found in Figure 1.

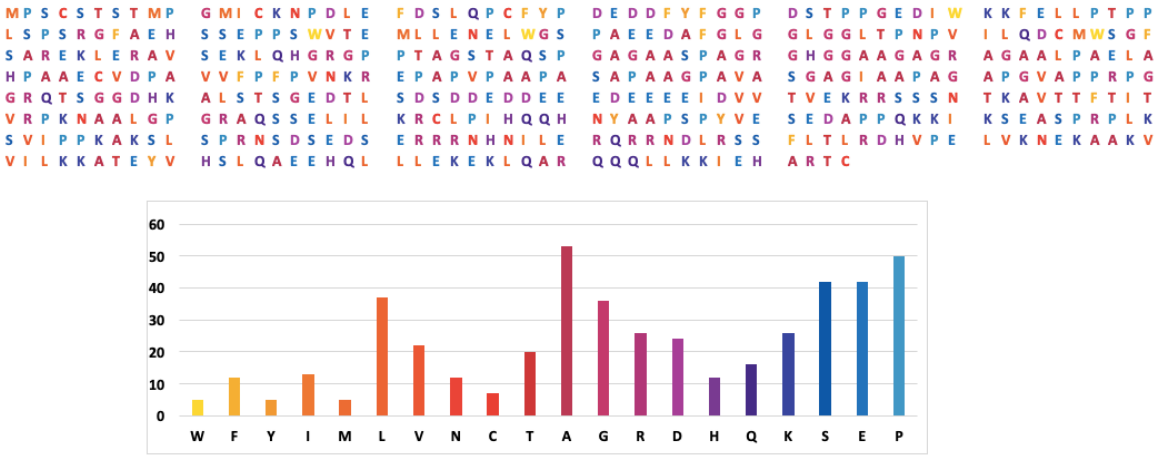


Figure 1 - *The sequence of human N-Myc (1-464) P04198, coloured from order promoting in yellow to disorder promoting in blue, according to the scale described by Campen et al. 2008. Notice that the majority of amino acids is in either disorder promoting (Q to P) or neutral (A to D). The x axis is the different amino acid, and the y-axis is the number of times each amino acid occurs in the sequence.*

The most well studied variant of the Myc family is C-Myc, however the structure of N-Myc and C-Myc is very similar and have identical parts, se figure 2.

Score	Expect	Method	Identities	Positives	Gaps
59.7 bits(143)	4e-19	Compositional matrix adjust.	31/56(55%)	40/56(71%)	5/56(8%)

```
N-Myc:  15  KNPDLEFDSLQPCFYPDEDDFYFGGPDST-----PPGEDIWKKFELLPTPPLSPSR  65
            +N DL++DS+QP FY DE++ ++      +       P EDIWKKFELLPTPPLSPSR
C-Myc:  10  RNYDLDYDSVQPYFYCDEEENFYQQQQQSELQPPAPSEDIWKKFELLPTPPLSPSR  65
```

Figure 2 – Sequence alignment of the N-terminal regions of N-Myc (P04198) and C-Myc (P01106), aligned using pBlast.

The proteins expressed from Myc family genes belong to the superfamily basic helix-loop-helix leucine zipper (bHLH-LZ). The Leucine zipper (LZ) motif enables dimerization with Max, and the basic helix-loop-helix (bHLH) motif allows Myc proteins to bind to DNA. After Myc dimerizes with Myc associated factor X (MAX), it binds to enhancer boxes (E-boxes). This interaction alters the gene expression, resulting in either up- or downregulation. That the Myc-Max heterodimer attach to the

E-Box, appears to be the primary pathway of how Myc regulates the gene expression, however Myc has also been reported to attach to non E-box DNA regions, and that Myc binds to the promotor of ribosomes, as well as that Myc binds independently of Max,. (Duffy et al. 2021)

Myc regulates around 15% of all genes in the human genome, i.e 2000-4000 genes. In healthy tissues the expression of Myc is very controlled, however, as mentioned before the expression of Myc is altered in around 70 % of all malignancies. Myc was previously categorized as "undruggable" due to its intrinsically disordered properties, resulting in lack of hydrophobic pockets and lack of catalytic activity. Myc is also located in the cell nucleus, thus cannot monoclonal antibodies be used as a therapeutical agent. Myc is one of the most studied IDPs, and the focus of enable suppression of Myc includes; Blocking Myc-Max interaction, prevent Myc-Max from binding to DNA, prevent expression and/or translation of Myc, Stabilising Max homodimers and finding synthetic lethal partners for Myc. (Duffy et al. 2021)

Myc consists of 3 major parts; the N terminus region containing the transactivation domain, the central region responsible for stability control and nuclear localisation and the C terminus region associated with MAX interaction and DNA Binding.

All three Myc variants have highly conserved regions, so called Myc-Boxes (MB). C-Myc and N-Myc has six MB: MB0, MBI, MBII, MBIIIa, MBIIIb and MBIV, and L-Myc lacks MBIIIa resulting in total of five MB. Each MB have a role for the function of Myc and binds to a specific protein, MB0 binds to the transcription elongation complex. MBI is involved in proteasome degradation and ubiquitination of Myc, MBII is responsible for tumour initiation and important for Myc mediated gene transcription. MBIIIa and MBIIIb plays a role in regulation of apoptosis and interaction with the protein WDR5, respectively. MBIV binds to chromatin, and interacts with proteins regulating the cell cycle. (Duffy et al. 2021)

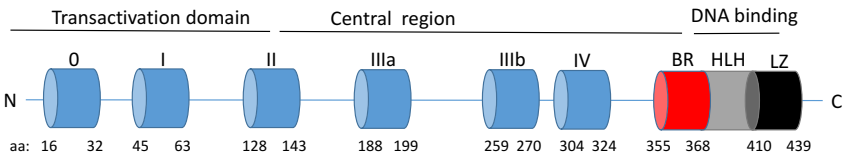


Figure 3 - *Domain structure of C-Myc. Blue shows Myc boxes, red the basic region, grey the helix-loop-helix and black the leucine zipper. Figure inspired by Duffy et al, 2021.*

2.3 Prolines and cis/ trans isomerisation

The majority,~99.7%, of peptide bonds is in trans conformation (Weiss *et al*. 1998), this due to less steric repulsions between atoms and favourable electrostatic interactions (Wedemeyer *et al*. 2002), se figure 4. Prolines on the other hand have a higher preference to forms cis conformations, this because of that they forms a cyclic structure, that the backbone nitrogen is linked to the C^{δ}, resulting in a pyrrolidine ring, se figure 5. The steric similarity between carbonyl carbon and C^{α} respective C^{δ} in the two conformations leads to increased cis population for prolines, of approximately 5-10 %, compared to the other amino acids that occurs in cis conformation in less than 0.5 %. (Mateos *et al*., 2020)

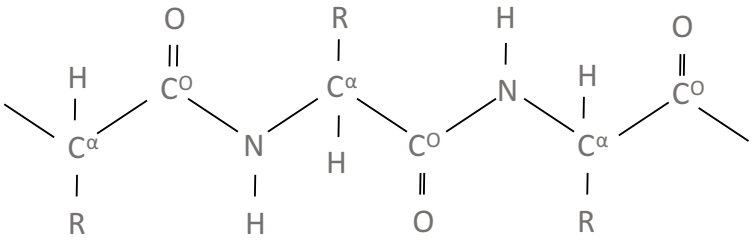


Figure 4 – Overview of the primary structure of a protein. R is the functional group that differentiate each amino acid. All the C^{α} and the functional R groups are facing the opposite direction = trans conformations.

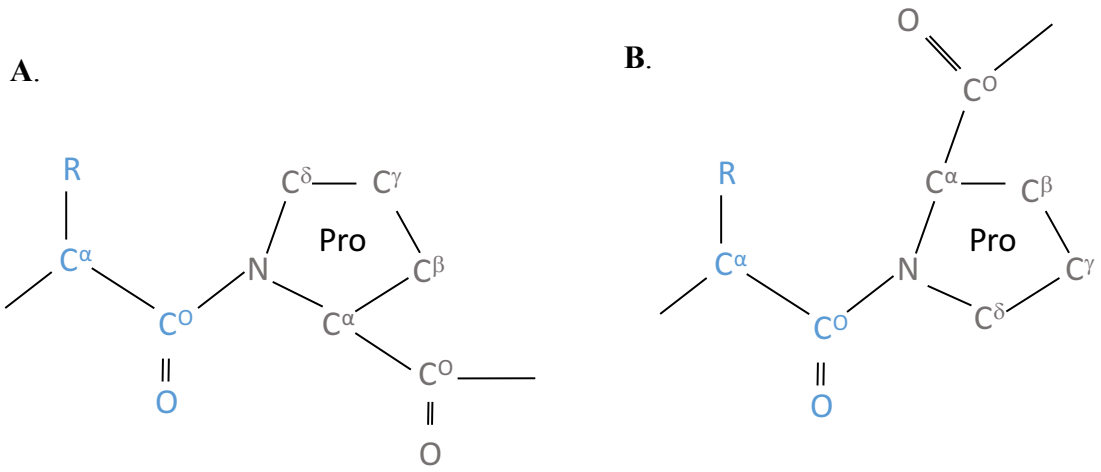


Figure 5 – Structure of proline, A: trans configuration and B: cis configuration.

Prolines is highly abundant in IDPs, however their role is very elusive. Dependent on cis peptide bond, prolines can alter the chemical shift of residues located ± 3 away from the proline. These altered chemical shifts give rise to alternative conformations, so called minor states. The chemical shift alterations of Prolines, dependent on if cis or trans configuration can be found in Table I. Proline isomerization is an active and important field of research, due to its role in protein folding and function (Shen *et. al*, 2010).

Table I- Average secondary chemical shift and the standard deviations for $^{13}C^{\beta}$, $^{13}C^{\alpha}$ and backbone of prolines proceeded by a cis/ trans peptide bond. Table from (Shen *et. al*, 2010).

	$<\Delta\delta_{cis}$	$<\Delta\sigma_{cis}$	$<\Delta\delta_{trans}$	$<\Delta\sigma_{cis}$
$^{1}H^{\alpha}$	0.17	0.53	-0.03	0.39
$^{13}C^{O}$	-1.78	1.45	-0.26	1.58
$^{13}C^{\alpha}$	-0.19	1.07	0.55	1.47
$^{13}C^{\beta}$	2.42	1.27	0.21	0.97
$^{13}C^{\gamma}$	24.41	0.74	27.45	0.86

Detection of $^{13}C^{O}$ has proven to be very useful when studying IDPs, the 2D CON experiment correlates $^{15}N(i)$ with $^{13}C^{O}(i-1)$ is particularly useful since the proline residues is separated from other amino acids (Bermel *et al.,* 2012: Murrali *et al.*, 2018). Dependent on which amino acid that are preceding the proline the shift is altered, meaning that all prolines after an serine is located in the same region, resulting in a fingerprint pattern. (Murrali *et al.*, 2018)

Prolines within a protein can exist in two different stable conformations cis or trans, and have the ability to isomerize, i.e. alter between these conformations, this is a very slow process due to the large energy barrier that isolates the different states, and it is necessary with a catalyst for switching between the states, at a speed relevant to signalling within a cell. The most well-known catalyst is peptidyl-prolyl cis/trans isomerases (PPIases), and they can, by isomerisation, regulate N-Myc stability and transcriptional activity. (Cohn *et al*., 2020)

Proline isomerisation is a key function for post-translational control of cellular proteins. A predominant hypothesis in the field is that c-Myc functional regulation is influenced by two proline phosphorylation events, phosphorylation of S62 and phosphorylation of T58, together with cis-trans conversions in the adjacent P63. Firstly, C-Myc is activated downstream of growth stimuli by a kinase that phosphorylates S62 when P63 is be in trans. An isomerase (PPIase) binds to the S62 and P63-trans MYC and catalyses a conformational change from trans to cis, which sterically prevents phosphatase activity. C-Myc is now transcriptionally active and has increased DNA binding to E-box promoters. GSK3 kinase phosphorylates T58, which promotes dephosphorylation of S62, this step requires cis to trans isomerisation of P63. E3 ubiquitin ligase targets phosphorylated T58 and C-Myc is directed towards degradation. (Cohn *et al*., 2020)

2.4 Nuclear magnetic resonance spectroscopy

Two of the most common methods when studying the three-dimensional structure of proteins, are Nuclear magnetic resonance spectroscopy (NMR) and X-ray crystallography. Both methods has pros and cons, and X-ray crystallography is especially useful if the protein is large, but requires the protein to be packed in crystals. The other method, NMR, analyses the protein in solution, which is more accurate to the natural state of proteins. The drawback of NMR is that the protein must be isotopically labelled to yield a detectable signal (Sekhar and Kay, 2019) .

2.4.1 NMR Theory

The atom nucleus has a magnetic moment, or spin as it is also called, which depend on the number of paired and unpaired protons and neutrons, and gives rise to the spin quantum number *I*. For example, ^{13}C has 6 protons and 7 neutrons, i.e. even number of protons and odd number of neutrons, making it an NMR active half integer with spin ½, more examples can be seen in table II.

Table II – Different Atoms and their composition of protons and neutrons, together with the nuclear spin number (*I*).

Example	**Neutrons**	**Protons**	**Spin (*I*)**
^{16}O	8 = Even	8 = Even	0
^{2}H	1 = Odd	1= Odd	1 = Integer
^{13}C	7 = Odd	6 = Even	½ = Half integer
^{15}N	8 = Even	7 = Odd	½ = Half integer

The magnetic moment $\boldsymbol{\mu}$ is proportional to the spin angular momentum vector $\boldsymbol{I}$, according to equation below, where γ is the gyromagnetic ratio, a constant for each kind of atom, se table III. The nuclear spin results in a magnetic field around the nucleus, and when for example a nuclei with spin 1/2 is placed in an external magnetic field, the nuclei will align itself with the external field and precess around the axis , and can populate two different energy states. A radio-frequent pulse is necessary, so that the magnetic moment flips from the z-axis to the x,y -plane of the external field, and thereby can be detected.

$$\boldsymbol{\mu} = \gamma \cdot \mathrm{I}$$

Table III - *Overview of different nuclei's and their gyromagnetic ratio, and the natural abundance. Data from Bernstein et al, 2004.*

Nuclei	γ **$10^7 T^{-1}s^{-1}$**	**Natural abundance**
^{1}H	26.75	99.985 %
^{2}H	4.11	0.015 %
^{12}C	-	98.93 %
^{13}C	6.73	1.108 %
^{14}N	1.93	99.636 %
^{15}N	-2.71	0.37 %

High gyromagnetic ratio leads to better detection in an NMR experiment, and therefore is the most favoured element ^{1}H. The magnetisation follows the equation below, where γ is the gyromagnetic ratio, h is Planck's constant, Ns is the number of spins, B_o is magnetic field strength, k is the Boltzmann constant, and T is temperature.

$$Mo \approx \gamma^2 h^2 Ns\, B_o / 4kT$$

Since the gyromagnetic ratio is squared, a small difference in ratio results in a much larger difference in magnetisation. Therefore is direct carbon detected experiments to prefer in cases were the resolution or signal to noise is to low when looking at nitrogen detected experiments.

2.4.2 Chemical shift

The chemical shift of a nucleus varies by the chemical environment where the nucleus is located. If the nucleus is close to an electronegative atom, the electronegative atom drags the electrons toward itself resulting in a de-shielded nucleus with a higher chemical shift. The resonance frequency of a nucleus depends on where in the molecule the nucleus is positioned, or in other words how the local electron distribution is in the molecule. The equation of chemical shift can be seen below, there v is the measured frequency, and $vref$ is the frequency of the standard compound.

$$\delta = \frac{10^6 \cdot (v - v_{ref})}{v_{\text{ref}}}$$

2.4.3 Isotope labelling

The most common isotopes in proteins are ^{1}H, ^{12}C, ^{14}N and ^{16}O. Among them is only ^{1}H a suitable atom to use, ^{12}C and ^{16}O has even number of protons and neutrons i.e. they have spin I=0 and is therefore not detectable, ^{14}N is NMR active but has I=1, resulting in more than two energy levels which complicates the spectra.
The traditional way of solving this is to control the sources of nitrogen, carbon and oxygen during expression of the protein. If $^{15}NH_4Cl$ and ^{13}C-glucose is the only

sources of nitrogen and carbon, ^{15}N and ^{13}C will be incorporated in the cells and expressed in proteins.

2.4.4 2D NMR experiments

In a one-dimensional proton spectrum many peaks will overlap, resulting in broad peaks comprising many proton signals which cannot be interpreted individually. Since proteins have very many protons, this is often the case. For example, N-Myc 1-69 has approximately 500 hydrogen atoms, see appendix 1.

The magnetisation can be transferred between two different NMR active nucleus, and labelled with the chemical shift from both, resulting in a 2D spectra with less overlaps. The most common 2D experiment for proteins is the Heteronuclear Single Quantum Coherence (HSQC), where the ^{1}H is exited, and then the magnetisation is transferred along the spin-spin coupling to the neighbouring ^{15}N. After processing the data, each peak in the spectrum corresponds to one single amino acid. The ^{15}N-HSQC spectra (hereafter referred as HSQC) can be used as a fingerprint for the protein, and some of the information that can be extracted from a HSQC spectrum, is if the protein is folded or not, if the sample is pure, and how an added ligand interacts with the protein. One drawback of an HSQC experiment is that prolines can't be detected due to the fact that they lack an amide proton, because of their ring structure, se figure 6.

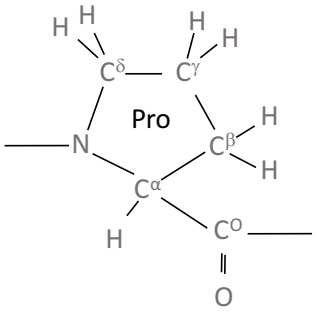


Figure 6 – Overview of proline structure. The backbone nitrogen is covalently attached to both C^{α} and C^{δ}.

^{13}C has higher gyromagnetic ratio than ^{15}N, which results in better detection in an NMR experiment. Therefore, was the direct carbon detection experiments C_CON and N_(H)CAN used, that not only results in better dispersion, they also display the prolines. The principle of an C_CON and an N_(H)CAN experiment can be seen in figure 7.

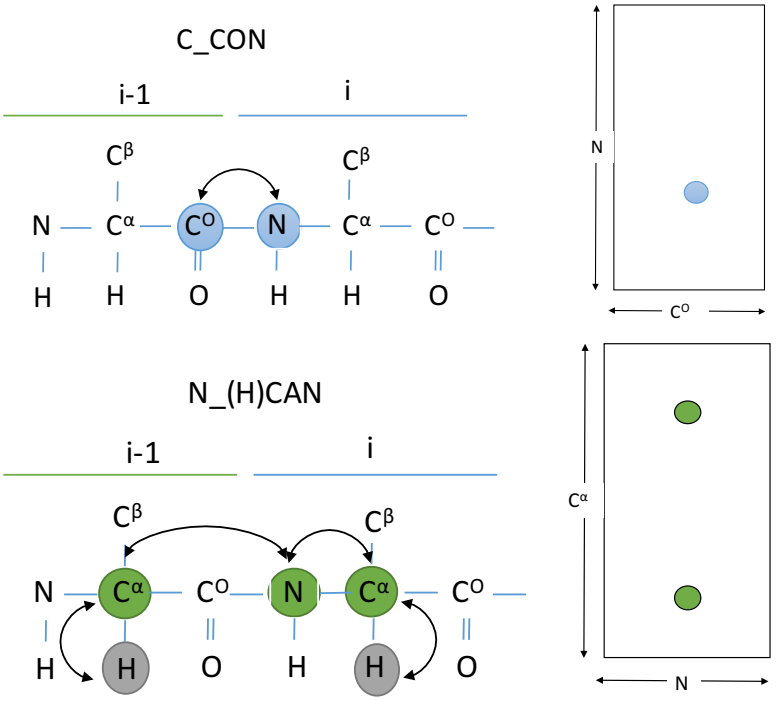


Figure 7 – To the upper left is the setup of an C_CON experiment, where the magnetisation is transferred from the carbonyl carbon from residue i-1 to the amide of residue i. To the lower left is an N_(H)CAN experiment, where the magnetisation starts on the $^{1}H^{\alpha}$ hydrogen, but do not detect on it, and then is labelled with the chemical shift of the $^{13}C^{\alpha}$ and then to the nitrogen. To the right is an illustration over the resulted spectrum for C_CON and N_(H)CAN, respectively.

2.4.5 3D NMR experiments

To be able to know which peak that corresponds to each amino acid, the peaks must be assigned. This can be done in different ways, manually, semi-manually or automatically. In this thesis this was done manually in a software called NMRFAM Sparky. For the sequential assignment method, which is the most commonly used, several different spectra are required, to be able to determine which amino acid who is responsible for each signal. The experiments used in this thesis are listed below. The name of each experiment indicates what are being measured, for example the HNCA experiment correlates ^{1}H, ^{15}N and $^{13}C^{\alpha}$ for residue i and residue i-1. A common and useful experiment is the HNCACB, that correlates ^{1}H, ^{15}N and both $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ for residue i and residue i-1. If the name contains a parenthesis, it means that the magnetisation is transferred through this atom, but the atom is not detected. One example of this is the HN(CA)CO, whom correlates ^{1}H, ^{15}N and ^{13}CO from residue i and residue i-1, more examples can be found in table IV.

Table IV - Overview of all experiments performed in this thesis.

Experiment	Correlates nuclei
HSQC	N(i) and NH(i)
HN(CA)CO	N(i), NH(i), C^{o}(i) and C^{o}(i-1)
HNCACB	N(i), NH(i), C^{α}(i), C^{α}(i-1), C^{β}(i) and C^{β}(i-1)
HN(CO)CACB	N(i), NH(i), C^{α}(i-1) and C^{β}(i-1)
HNCO	N(i), NH(i), C^{o}(i-1)
N_CON	N(i) and C^{o}(i-1)
N_CAN	N(i), C^{α}(i), C^{α}(i-1)
C_CON	N(i) and C^{o}(i-1)
N_(H)CAN	N(i), C^{α}(i), C^{α}(i-1)
2D 15N Plane iHNCA XP	N(i) and NH(i)
2D 15N Plane HNCO PX	N(i) and NH(i)
HN(CCO)C	N(i), NH(i) and all C(i-1)
HN(CCO)H	N(i), NH(i) and all CH(i-1)
CCH-TOCSY	All C(i) and all CH(i)

The principle of assigning is rather straight forward in theory; one finds a starting amino acid with deviant shifts, and then find the correlating amino acid before or after that one in the sequence, and build up fragments of shorter sequences, then compare these fragments with the amino acid sequence and when all amino acids are assigned the process is completed. By stepping through each amino acid, one can assign a large protein, se figure 8.

The detected chemical shifts is compared to chemical shifts in literature, e.g BMRB. The fact that some amino acids have similar shifts makes the process much more difficult, so to make the procedure easier; one can lay spectra on top of each other, the combination of HNCO and HN(CA)CO is often used for assigning the Carbonyls and HNCACB and HN(CO)CACB for assigning C^{α} and C^{β}, these spectra can be found in figure 9 and figure 10, respectively.

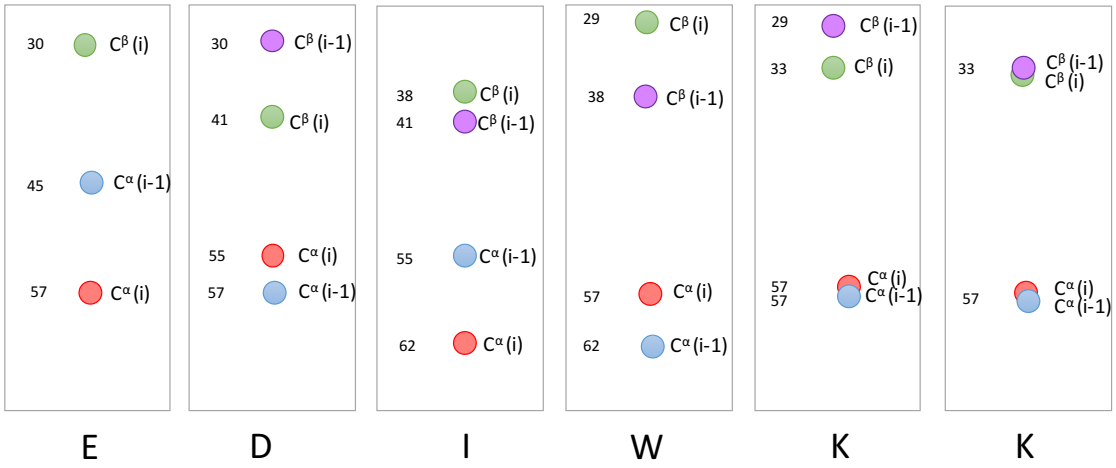


Figure 8 – Example of the assigning process, The chemical shift of $C^{\alpha}(i)$ and $C^{\beta}(i)$ cohere with the chemical shift of $C^{\alpha}(i-1)$ and $C^{\beta}(i-1)$ of the next amino acid in the sequence.

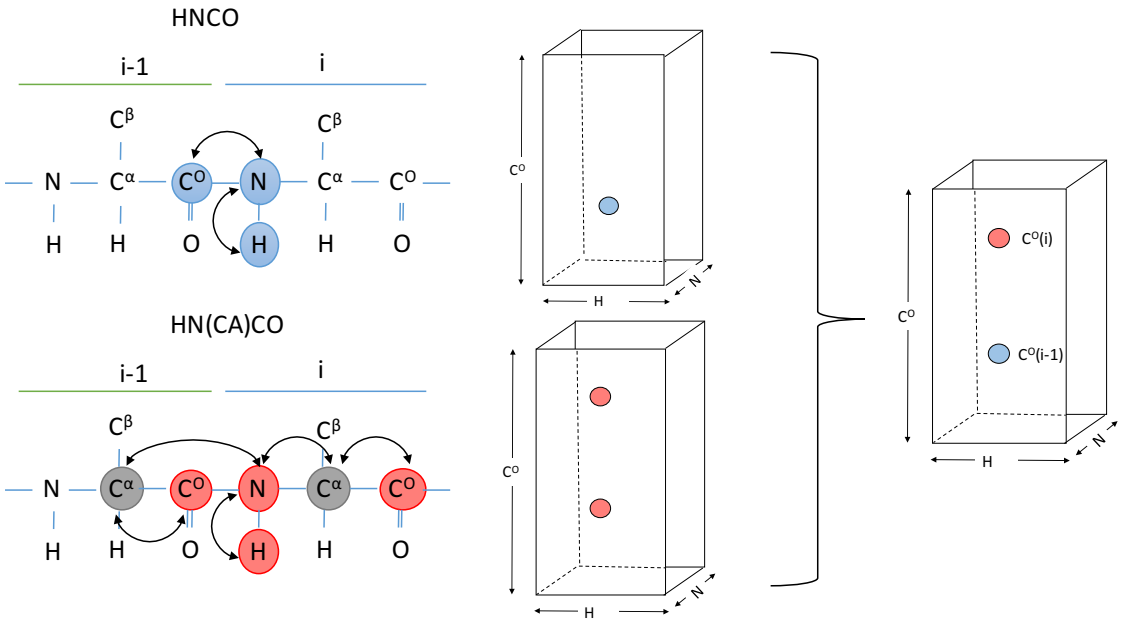


Figure 9 – Principle of determining CO shift. Upper left shows HNCO and bottom left shows HN(CA)CO, in the middle is an illustration of how the result looks in three dimensions and to the right is the overlay of the two spectra, where the two carbonyls have different colours and therefore easy to distinguish.

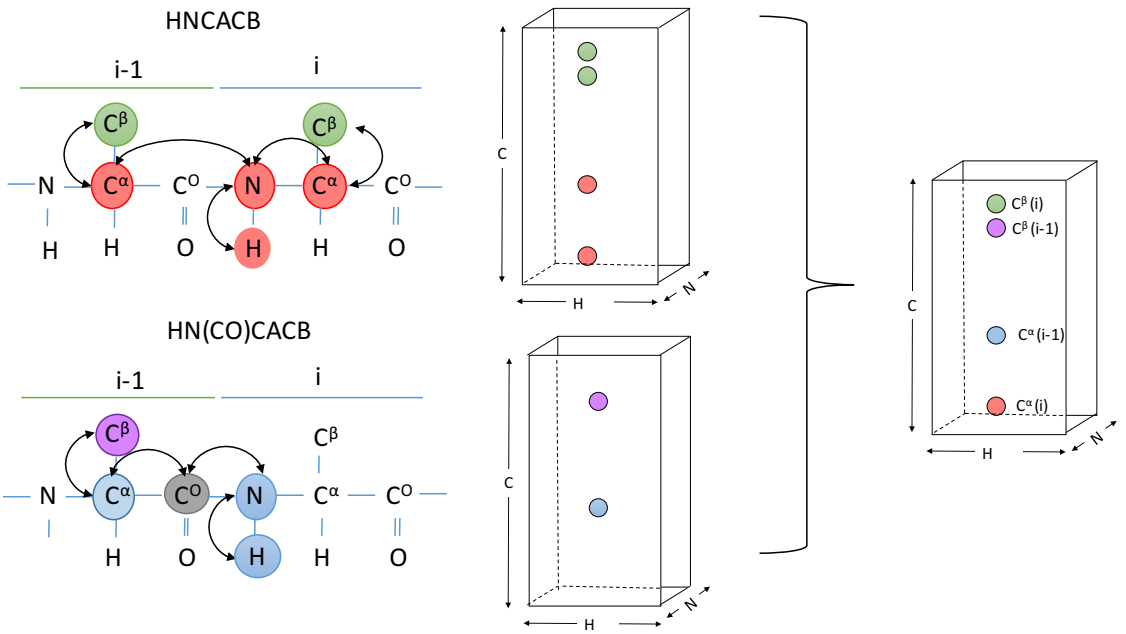


Figure 10 – Principle of determining C^{α} and C^{β} shift. Upper left shows HNCACB and bottom left shows HN(CO)CACB, in the middle is an illustration of how the result looks in three dimensions and to the right is the overlay of the two spectra where the four carbons have different colours and therefore easy to distinguish.

2.5 Methodological theory

2.5.1 Gel electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) is a method for analysing the size and purity of proteins and DNA. It is based on the principle that different sized charged molecules will migrate differently when placed in a electric field. SDS works as a denaturant and unfolds the proteins, so that the shape of the protein does not affect the migration length. The protein is dyed with example Coomassie blue staining and compared to a mixture of proteins with already known molecular weights.

2.5.2 Protein separation by chromatography

Not only the protein of interest is produced, the cellular host produces many proteins and other impurities that need to be removed, to get accurate measurements of the protein of interest. The principle of chromatography, is that a mobile phase containing the protein, carries through a stationary phase at different speeds, dependent pf the nature of the mobile and the stationary phase, which causes a separation. (Thermofisher – Ask a Scientist, 2019).

In this thesis three different types of chromatography were used; Immobilized metal ion chromatography (IMAC), Ion exchange chromatography and reverse IMAC. IMAC, as the name suggests, builds on the principle that metal ion is immobilized in the stationary phase, and that the proteins in the mobile phase is separated based on affinity. In this particular case, a His-tag (6 histidine residues in a row) was incorporated in to the expression vector, meaning that the expressed N-Myc(1-69) will have additionally histidine's at the N-terminus. Histidine have a strong affinity for metal ions such as Zn^{2+}, Ni^{2+}, Co^{2+}, and Cu^{2+}in aqueous solution. And the stationary phase consists of Ni^{2+} beads, resulting in that N-Myc(1-69) among with some other proteins with affinity for Ni^{2+}will bind, and the rest will only flow through the column. An elution buffer with higher salt concentration, was used to weaken the electrostatic interactions, to make the proteins loosen from the Ni^{2+} beads. Ion exchange chromatography, is based on the overall charge of the protein, the stationary phase has one charge whom attracts the opposite charged proteins from the mobile phase.

Reverse IMAC, is exactly the same principle as an regular IMAC, the difference is that this step is done after the his tag is cleaved of from the protein of interest, so this works as a purifications step to remove the His-tag since it binds to the stationary phase and the protein stays in the mobile phase.

2.5.3 Protein characterisation

To be able to determine the amount of protein after purification, Nanodrop was used, the device measures the absorption of UV light. Proteins that contain the amino acids; tryptophan, tyrosine and cysteine will absorb light at the wavelength 280 nm, the amount of absorbed light is exponentially dependent on how much of these amino

acids whom is present in the sample. The protein concentration can then be calculated using Lambert-Beer's Law, where A is absorbance at 280 nm, ε is the extinction coefficient, c is the protein concentration and l is the path length of light. The extinction coefficient for WT and P44L was estimated using ProtParam, se appendix 1 for data.

$$A = \varepsilon \cdot l \cdot c$$

2.6 Computer software

2.6.1 NMRFAM – SPARKY

NMRFAM – SPARKY (Sparky) are a computer software provided by UCSF, Sparky is a popular NMR analysis tool, including functions such as; Automated peak picking, Semi-automated Assignment, Graphical assignment assistant and Structure prediction to mention a few (*NMRFAM-Sparky Distribution)*. In this thesis, Sparky have mainly been used to assign peaks to the corresponding amino acids.

2.6.2 PINT

PINT is a software created by Patrik Lundström and Co-workers (Ahlner *et al*, 2013; Niklasson *et al*, 2017) . The main use of PINT in this thesis, was for integrating each peak and thereby calculate the volume, with the purpose to determine how much is in each state.

3 Experimental procedures

In the following section, a description of the performed experiments can be found. The protein production was, as stated in "Announcement", pre prepared by Johanna, due to limited student access to the lab because of the Covid-19 pandemic. However, during the course of my thesis I also made a preparation of MYC according to the same protocol.

3.1 Protein Production

The N-Myc 1-69 WT or P44L construct was cloned into pNH-Trxt vector using ligation independent cloning as a fusion protein together with the protein TRX, the vector have resistance for Kanamycin. The plasmids was transformed into DH5α cells, and grown in medium containing 2g ^{13}C glucose, 1.5g $^{15}NH_4Cl$, 1mM $CaCl_2$, 1mM $MgSO_4$, 10 ml Gibco Vitamin Solution, 50µg/ml kanamycin and 34 µg/ml chloramphenicol per liter growth medium, until OD_{600} reached 0.6 and the cells was induced with 0.5mM IPTG, followed by harvest by centrifugation and frozen in liquid nitrogen and stored in Eppendorf tubes at -80°C.

3.2 Protein Purification

N-Myc(1-69)^{WT} and N-Myc(1-69)^{P44L} was purified in the exact same way. In this section only N-Myc(1-69)^{WT} will be described.

The five buffers listed in table V was prepared.

Table V – Buffers for protein purification**.**

Buffers for protein purification					
Key components: 50 mM **$NaHPO_4$ (**pH 8), 10 mM **Tris-HCl** (pH 8), 5% **glycerol** and 10 mM **B-ME**					
	Lysis Buffer	**Wash buffer**	**Elution buffer**	**Dialysis 1 buffer**	**Dialysis 2 buffer**
NaCl 5 M	100 mM	100 mM	100 mM	45 mM	100 mM
Imidazole pH 7.5, 2 M	10 mM	10 mM	10 mM		
DNase1	5 U/ml				

Cell lysis

Harvested cells from the N-Myc(1-69)^{WT} expression batch, was thawed on ice, lysed in Lysis buffer by sonication (2.5 seconds 30% amp and 7.5 second off, for total four minutes). Centrifuged at 18000g at 4°C for 30 minutes. The pellet was saved in the refrigerator, and the supernatant was incubated with 6 ml Ni^{2+} beads and equilibrated with Wash buffer, for one hour at rotation.

IMAC
The protein-bead mixture was applied to a colomn, and the flow through was collected. After the beads was washed with 10 column volumes (CV) Wash buffer, Elution buffer was added, firstly in 6x5 ml fractions (called e1.1 – e1.6) followed by 2x10 ml fractions (called e2.1and e2.2). All fractions were analysed with SDS-PAGE, and the fractions containing protein, e1.1-e1.6, was pooled together in one dialysis tube, and dialysed in Dialysis 1 buffer ON, cut-off 3.5 KDa.

Anion exchange chromatography
Ion exchange chromatography was performed in the ÄKTA PURE 25M system. Fractions containing protein was chosen from the chromatogram and analysed with SDS-Page and the fractions containing protein was pooled together, with 1 um TEV, and dialysed ON in Dialysis 2 Buffer, cut-off 2 kDa.

Reverse IMAC
The protein-sample was re-mixed with the Ni^{2+} beads used for IMAC, and incubated for an hour at 4°C, on rotation. The sample was applied to the same columns as in the IMAC step, washed with 10 CV Wash buffer and eluted with Elution buffer in steps of 3*5 ml (e1.1-e1.3) and 2*10 ml (e2.1 and e2.2). All fractions were analysed using SDS-PAGE. The flow through was concentrated using concentration cells to a volume of 2-3 ml each.

Gel filtration
Gel filtration was done in the ÄKTA system and was collected in 98x1 ml fractions. From the chromatogram, relevant fractions for N-Myc 1-69 was chosen and analysed with SDS page, and the concentration was measured using NanoPhotometer® NP80. The fractions containing more than 10 uM was quick-frozen in liquid Nitrogen and stored in -80°C.

3.3 Nuclear Magnetic Resonance spectroscopy experiments

3.3.1 Preparation of NMR samples

Firstly the buffer of the purified protein was changed from the $NaHPO_4$ containing dialysis 2 buffer to a MES based buffer (20 mM MES monohydrate (pH 6.5), 100 mM NaCl, 5% glycerol and 5 mM DTT). The dialysis was done in 3x500 ml step for 3h, 6h and ON, respectively. The sample was concentrated using a concentration cell and a 1 kDa cutoff filter.

The concentrated protein was mixed with 50 µl D_2O, 1 µl TCEP (0.5 M pH 7.5), 0.2 µl NaN_3 and 450 µl sterile filtered MES buffer, to a final volume of 3 ml and a final protein concentration of 630 mM. The sample was packed on ice and shipped to Swedish NMR centre in Gothenburg.

3.3.2 NMR spectroscopy experiments

The following spectra of N-Myc(1-69)^{WT} was recorded on Brucker 800 At Swedish NMR centre: HSQC, HNCACB, HN(CO)CACB, HNCO, HN(CA)CO, N_CON, N_(H)CAN, N_CAN, C_CON, CCH-TOCSY, HN(CCO)C, HN(CCO)H and the two 2D ^{15}N Plane: iHNCA XP and HNCO PX.

3.4 Data Analysis

3.4.1 Calculated chemical shift perturbations (CSP)

The CSP for the minor states, $\Delta\delta_{MINOR}$, was calculated using the following equation, where δN and δH is the difference in amid-nitrogen and amid-proton for the major and minor state, respectively and Rscale is equal to 6.5 (Andrèsen,. *et al.* 2011)

$$\Delta\delta_{MINOR} = \sqrt{\left(\frac{\delta N}{R_{Scale}}\right)^2 + \delta_{NH}^2}$$

3.4.2 Secondary structure propensity (SSP)

SSP was calculated using the Perl script written by Joseph Marsh, Department of Biochemistry, University of Toronto (Marsh, J. A. et al. 2006) Where the chemical shifts for $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{3}C^{O}$, $^{1}H^{N}$ and ^{15}N was used as input.

3.4.3 Estimating population distribution

To determine the relative populations between major and minor states the equation below was used. E.g. volume S3 /(Volume S3 + Volume S103 + volume S203).

$$\frac{Peak\ volume}{the\ sum\ of\ all\ volumes\ from\ the\ same\ peak}$$

4 Results and Discussion

4.1 Expression and Purification of N-Myc 1-69

The purification was made for both N-Myc(1-69)^{WT} and N-Myc(1-69)^{P44L}, however in this section only the results N-Myc(1-69)^{WT} will be presented, because only this protein variant was analysed further. The corresponding figures and calculations for the P44L mutated protein are found in appendix 2.

All SDS-Page gels can be found in figure 11, 11A comes from the IMAC purification step, and resulted in bright bands at approximately 25 kDa, which is higher than the theoretical molecular weight of 19.5 kDa. This behaviour of excessive migration is known for IDPs because of their low hydrophobicity and high net charge whom often results in large electrophoretic migration. There are proteins both in the pellet and in the supernatant, and the numerous band throughout the lane, indicates that there are other proteins still in the sample.

The second purification step, the Anion exchange, figure 11B, has removed the majority of the impurities, however some impurities can still be seen in forms of weak bands in almost every well. The densest band corresponds to N-Myc and has an approximate molecular weight of 25 kDa.

The reverse IMAC, figure 11C, clearly shows that TEV have successfully cleaved of the tag, i.e. a difference between cleaved and un-cleaved Myc. The un-cleaved protein still has a M_W of 25 kDa, after the His-Tag is cleaved of by the TEV protease, the well with the cleaved protein shows two bands, one belonging to the TEV protease and one band below 14.4 kDa at approximately 10 kDa belonging to N-Myc. In the flow-through and the two washing steps no protein was found, which indicates that the previous steps have removed a lot of impurities and that this protocol works well with N-Myc.

The SDS-PAGE gel from the gel filtration, figure 11D, shows that the loading fraction for the gel filtration contains at least two different proteins, indicated by the two bands in the Load well, however the gel filtration have successfully removed the impurity because only one band is seen in the fractions.

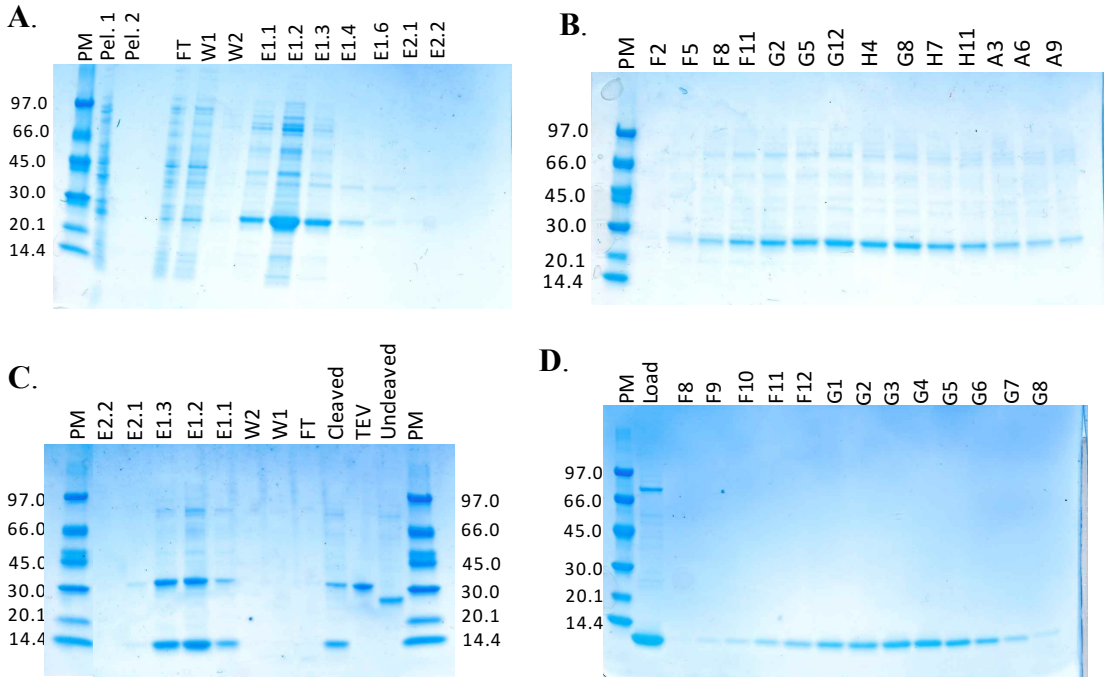


Figure 11 - SDS-PAGE gels of N-Myc 1-69^{WT} **A**: forward IMAC, **B**: Anion exchange, **C**: Reverse IMAC, **D**: Gel filtration. PM is the protein marker used for identification of the migration length. FT is the flow through, W means washing fractions and E is elution fractions.

The protein concentration of the fractions from gel filtration were calculated using Lambert beers law and can be seen in table VI. Fractions F11- G7 was quick-frozen in liquid nitrogen and stored in -80°C.

Table VI – Protein concentrations of N-Myc 1-69^{WT} calculated using Lambert beers law.

Fraction	**F8**	**F9**	**F10**	**F11**	**F12**	**G1**	**G2**	**G3**	**G4**	**G5**	**G6**	**G7**	**G8**
A260/280	1.845	1.845	1.845	1.605	1.207	1.137	0.953	0.893	0.992	1.179	1.513	1.640	2.212
A280	0.057	0.057	0.057	0.088	0.126	0.212	0.279	0.325	0.268	0.18	0.123	0.085	0.054
c (µM)	6.62	6.62	6.62	10.23	14.64	24.64	32.45	37.77	31.14	20.92	14.29	9.88	6.28
Saved in	-	-	-	1 ml	1 ml	1 ml	1 ml	1 ml	1 ml	1 ml	1 ml	1 ml	-

4.2 NMR Results

The nomenclature used in this section and further on is that the main peak is labelled with the one letter code for the amino acid and the position, i.e. G39. The corresponding small peak with the highest intensity is named in the same way, except that 100 is added to the position, i.e G139, and in the few cases were two small peaks was observed the second small peak with the lower intensity is named with additional 100, i.e. G239.

4.2.1 Assignments of major peaks

Firstly, all major peaks in the ^{15}N - HSQC spectrum was assigned, according to the principle described in figure 8. This was done using the spectra HNCO ontop of HN(CA)CO and HN(CO)CACB ontop of HNCACB, as illustrated in figure 9 and 10, respectively. The assigned HSQC for the major states can be found in figure 12. All residues except M1, and the 12 prolines were assigned, the reason that the methionine was missing can be due to dynamics at the N-terminus and the reason that all prolines was missing is because they lack amide-protein, and therefor is not detectable in a ^{15}N -HSQC.

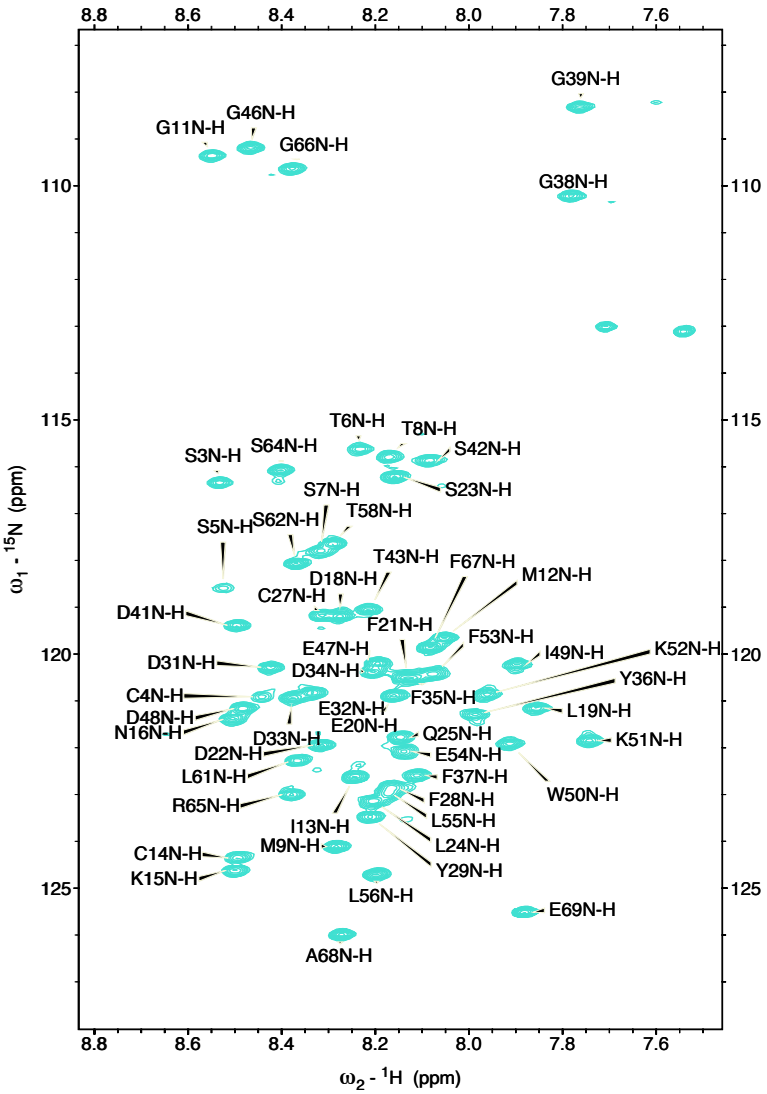


Figure 12 – HSQC of major states of N-Myc 1-69 WT showing major state, all residues are assigned except Methionine 1 and all the proline residues. The two peaks around ^{1}H~7.6 and ^{15}N~113, belongs to the sidechain of Asn or Gln, but is not of interest for this thesis.

4.2.2 Assignments of minor peaks

The next step was to assign the minor peaks. Before assigning the whole HSQC that contained approximately 30 minor un-known peaks, two other spectra with less information was assigned; the two 2D ^{15}N Plane: iHNCA XP and HNCO PX. The advantage of assigning these spectra first is that they only displays the amino acid before or the amino acid after a proline, resulting in fewer un-known peaks. This was done using the spectra HN(CO)CACB ontop of HNCACB and HNCO ontop of HN(CA)CO, exactly like the major spectra was assigned. One example of an assignment of an amino acid located at the position before a proline can be found in figure 13, The major peak Q25, and the minor peak Q125, have a difference of ~0.2 ppm in ^{13}CA shift , ~0.8 in ^{13}CB shift and ~0.4 ppm for ^{13}CO shift, and have a difference of approximately 0.25 ppm in ^{1}H shift and approximately 2.3 ppm in ^{15}N shift.

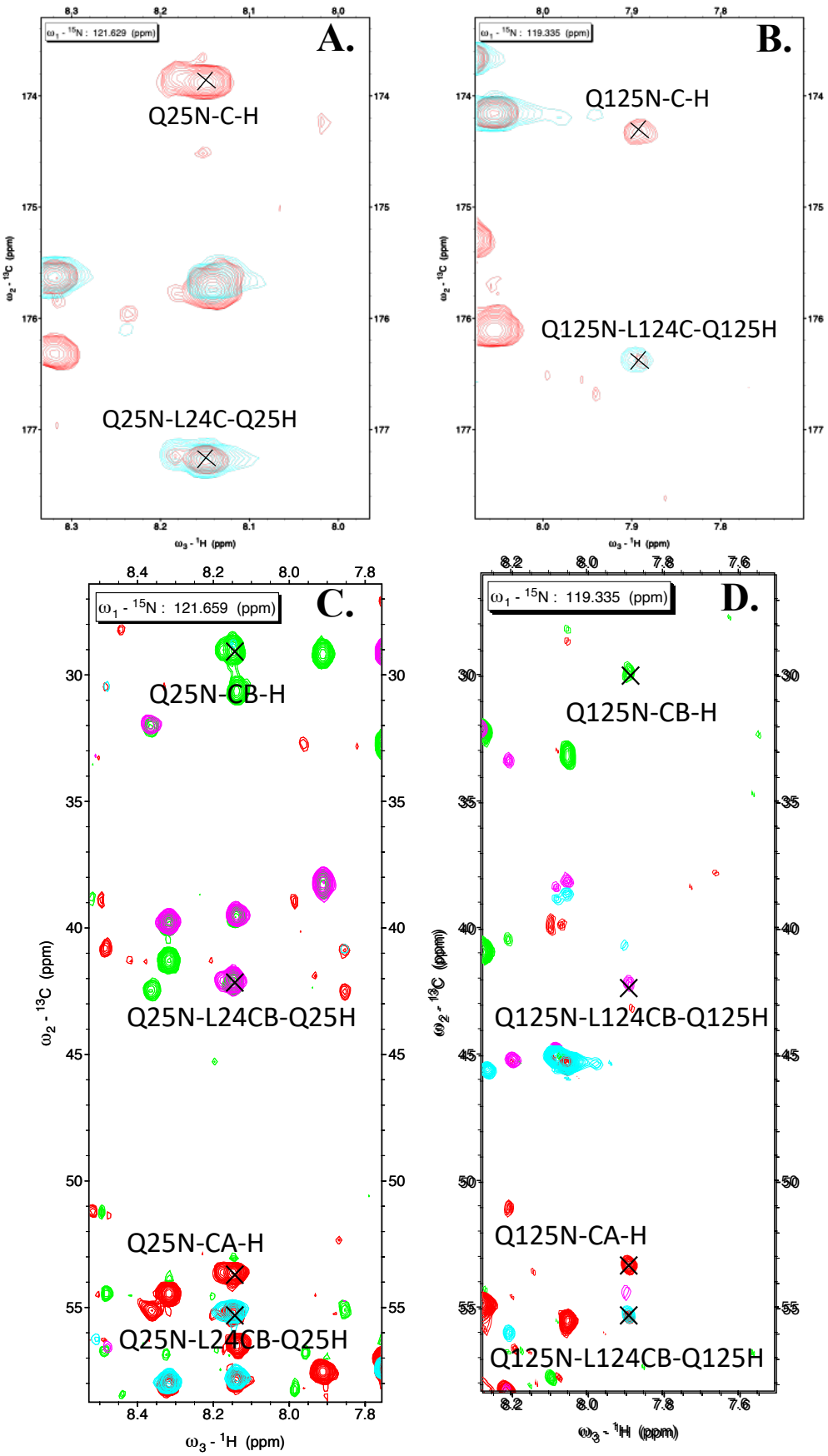
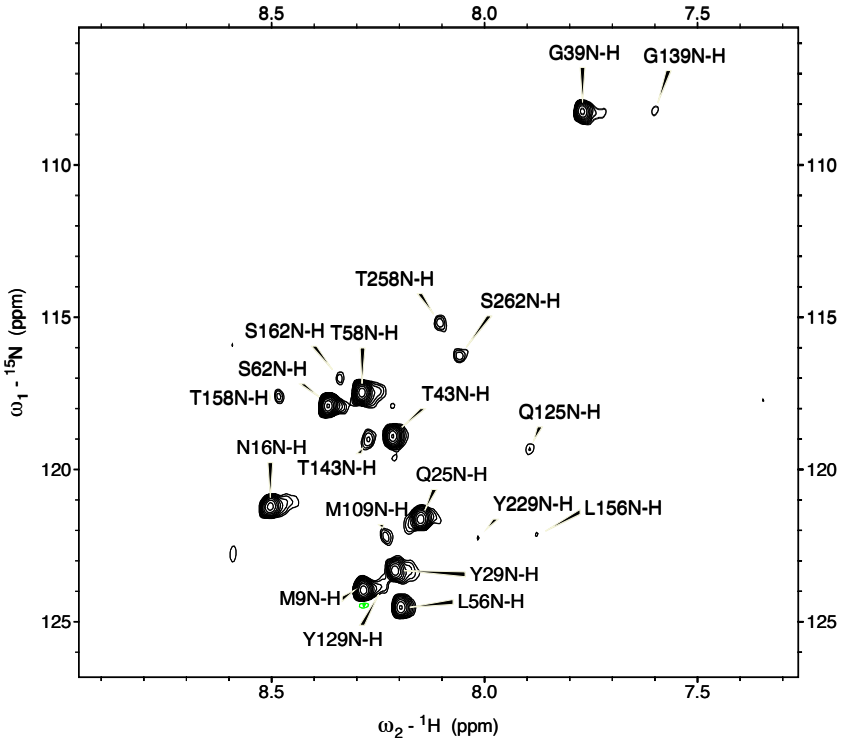


Figure 13 – A and **B**: HNCO and HN(CA)CO overlaid spectra for Q25 and Q125, respectively. **C** and **D**: HNCACB and HN(CO)CACB overlaid spectra for Q25 and Q125, respectively.

The two complete assigned 2D ^{15}N Planes: iHNCA XP and HNCO PX spectra can be found in figure 14 and figure 15. The chemical shifts for each minor peak were compared to the chemical shifts of already assigned major states, with the purpose to find unique matches. Some peaks were easier to assign, for example T58 is located in between two prolines (P57 and P59), and therefore is the minor states belonging to T58(T158 and T258) visible in both the two 2D ^{15}N Plane: iHNCA XP and HNCO PX.



M P S C S T S T M P G M I C K N P D L E F D S L Q P C F Y P
D E D D F Y F G G P D S T P P G E D I W K K F E L L P T P P
L S P S R G F A E

Figure 14 –Above: 2D ^{15}N Plane iHNCA XP spectrum, all amino acids before a proline is detected. Bottom figure shows the amino acid sequence, where prolines is coloured in red and found minor states, M109, Q125, Y129 Y229, G139, G239, T143, L156, T158, T258, S162 and S262 is coloured in blue.

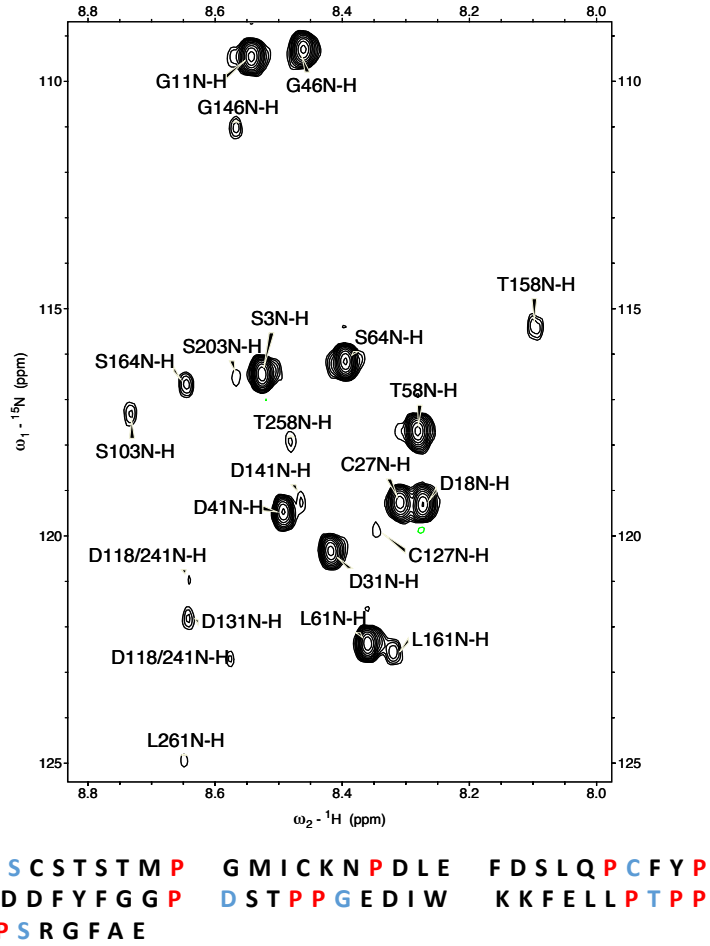


Figure 15- 2D ^{15}N Plane HNCO PX, shows all the amino acids after a proline. Bottom figure shows the amino acid sequence, where prolines is coloured in red and found minor states, S103, C127, D131, D141, G146, G246, T158, T258, L161, L261 and S164 is coloured in blue. Notice the two peaks in bottom right corner labelled D118/241, meaning that that the D118 and D241 was undistinguishable and therefore is the two peaks named D118/D241.

During the assignment of the two 2D ^{15}N Planes: iHNCA XP and HNCO PX, it became obvious that some major states had two minor states. The intensities of proline CO shift from HNCO data was plotted against residue number, and can be found in figure 16. The minor peak with the highest intensity was labelled as +100, and the minor state with the lower intensity was labelled with +200, for example G139 and G239.

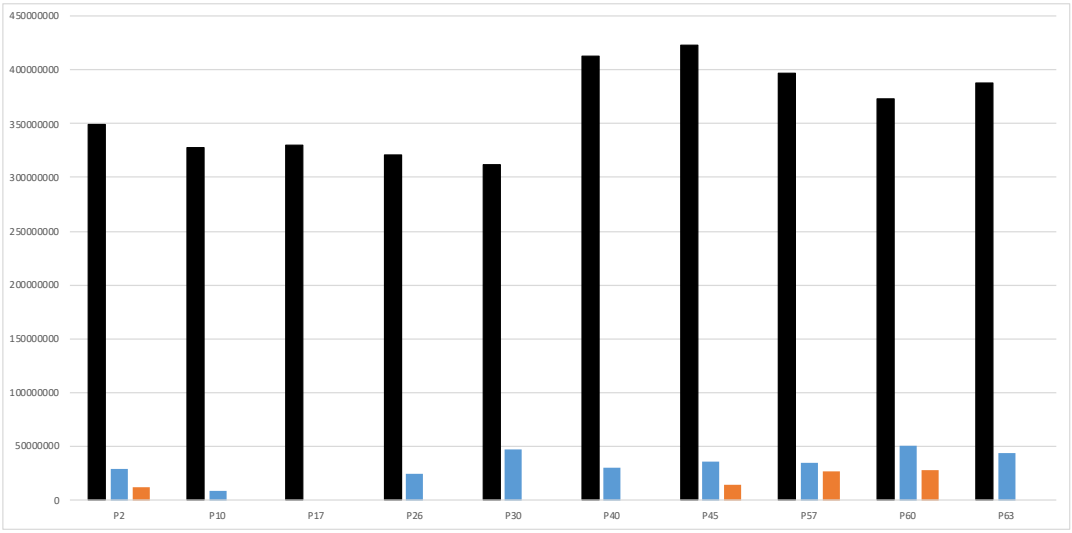


Figure 16 – Intensity for proline peaks obtained from HNCO spectrum. Major state is represented in black, and the highest intensity of minor states (+100) in blue, and the lowest intensity of minor state (+200) in orange.

After the two 2D ^{15}N Plane: iHNCA XP and HNCO PX were assigned the peaks was transferred to the HSQC and 22 out of the minor states was then assigned, the remaining minor states was assigned by connecting their C^{α}, C^{β}, and C^{O} chemical shifts for either residue i or residue i-1 to already assigned minor or major states, like the principle illustrated in figure 8. The complete assigned HSQC including minor peaks can be found in figure 17.

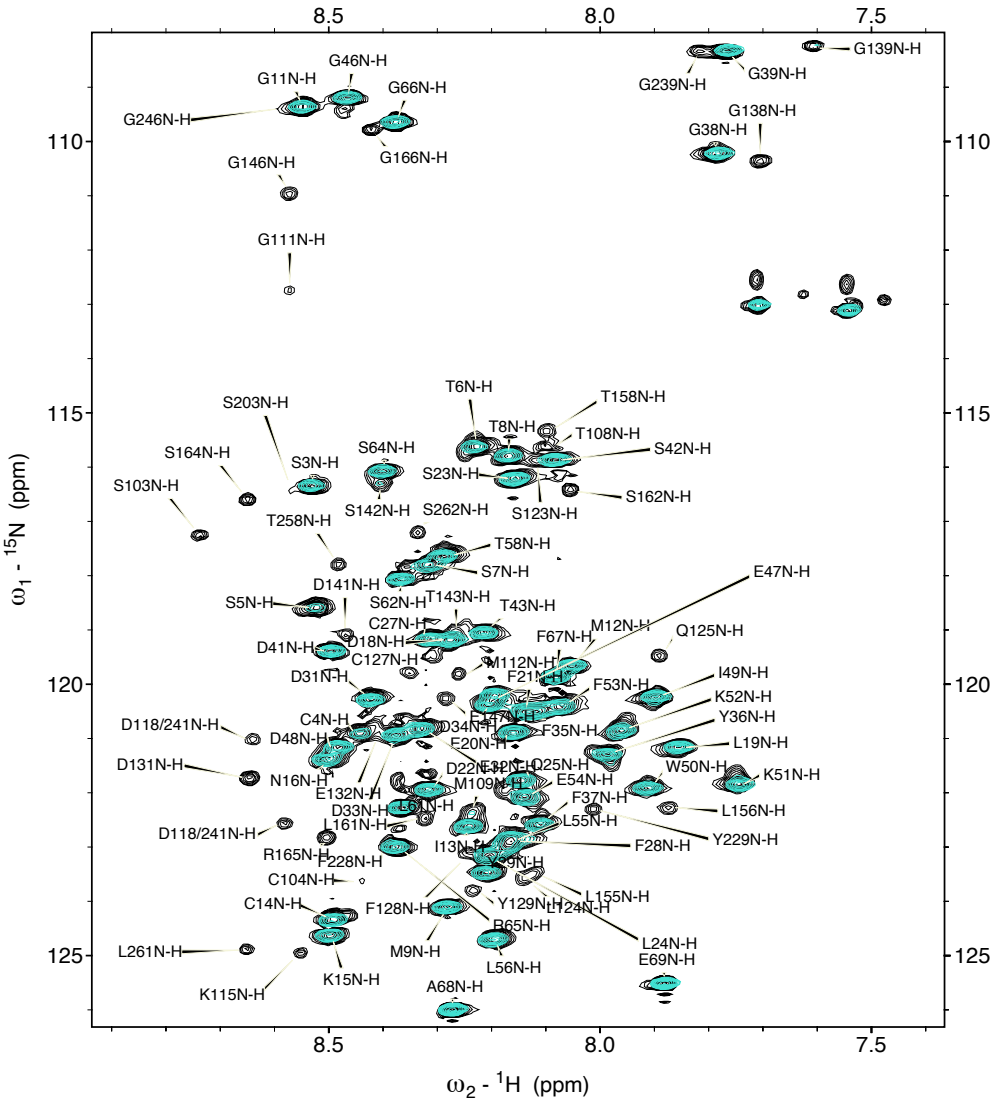


Figure 17 – Assigned HSQC spectrum of N-Myc 1-69 of major and minor states. The HSQC for major peaks (figure 12) is placed on top, coloured in blue.

Another example of an assignment, for residue 47, can be found in figure 18. The major peak E47, and the corresponding minor E147, has very similar $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C^{O}$ and ^{15}N shift, but have a difference of approximately 0.2 ppm in ^{1}H shift. The major and minor peak for residue 47 is located around ^{15}N ~ 121 ppm and ^{1}H ~ 8.1 ppm in the HSQC. The i-1 minor state from E147, G146, is located around ^{15}N ~111 ppm and ^{1}H~8.6 ppm in the HSQC.

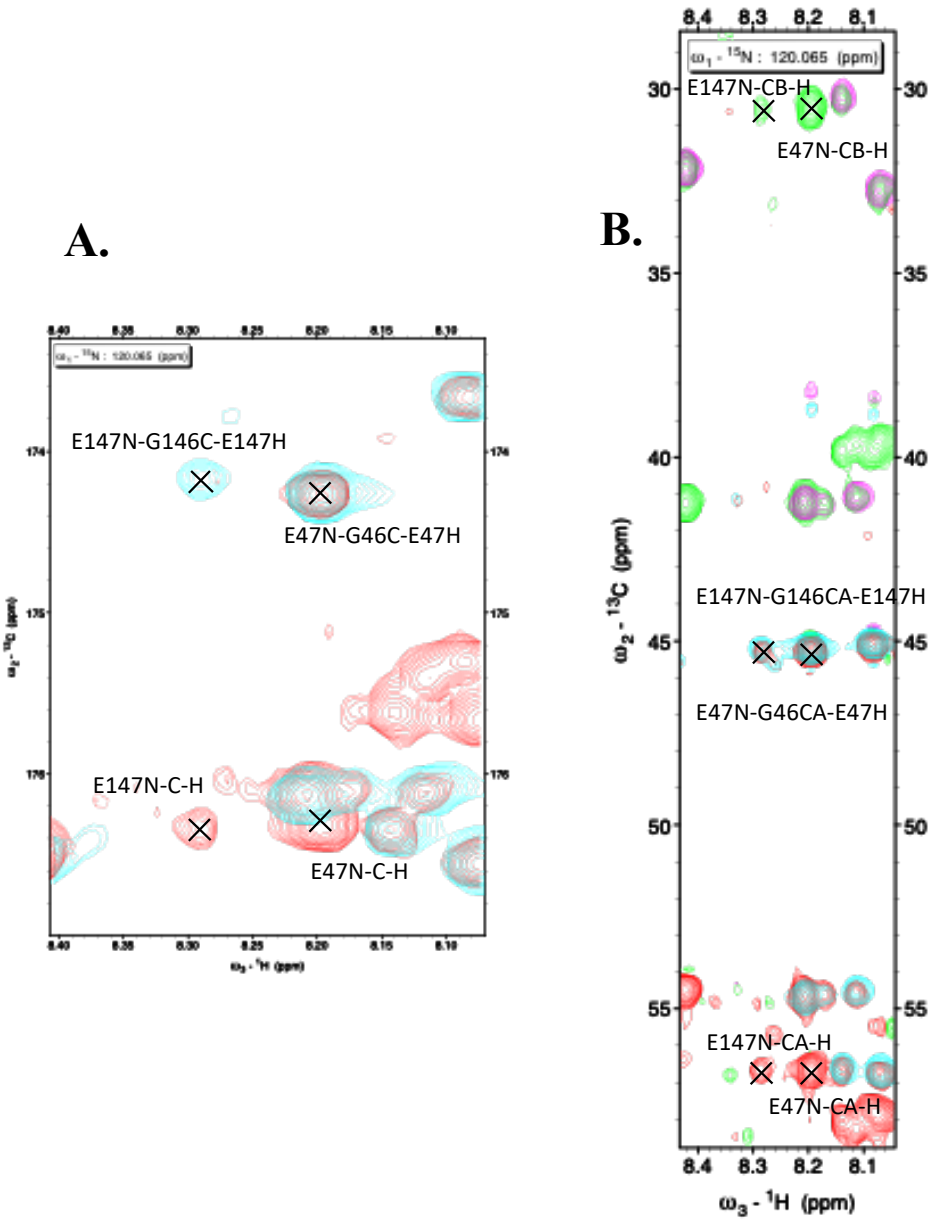


Figure 18 – A: HNCO and HN(CA)CO overlaid spectra **B**: HNCACB and HN(CO)CACB overlaid spectra for major state E47 and minor state E147.

Of practical reasons, not every minor assignment can be represented with complete strips, instead they will be presented graphically in figure 19, that shows the difference in $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift between the major and the minor state. The minor $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift is subtracted from the major $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift, respectively, and plotted against residue number. This to get an overview of how the chemical shifts differs between major and minor state in the sequence.

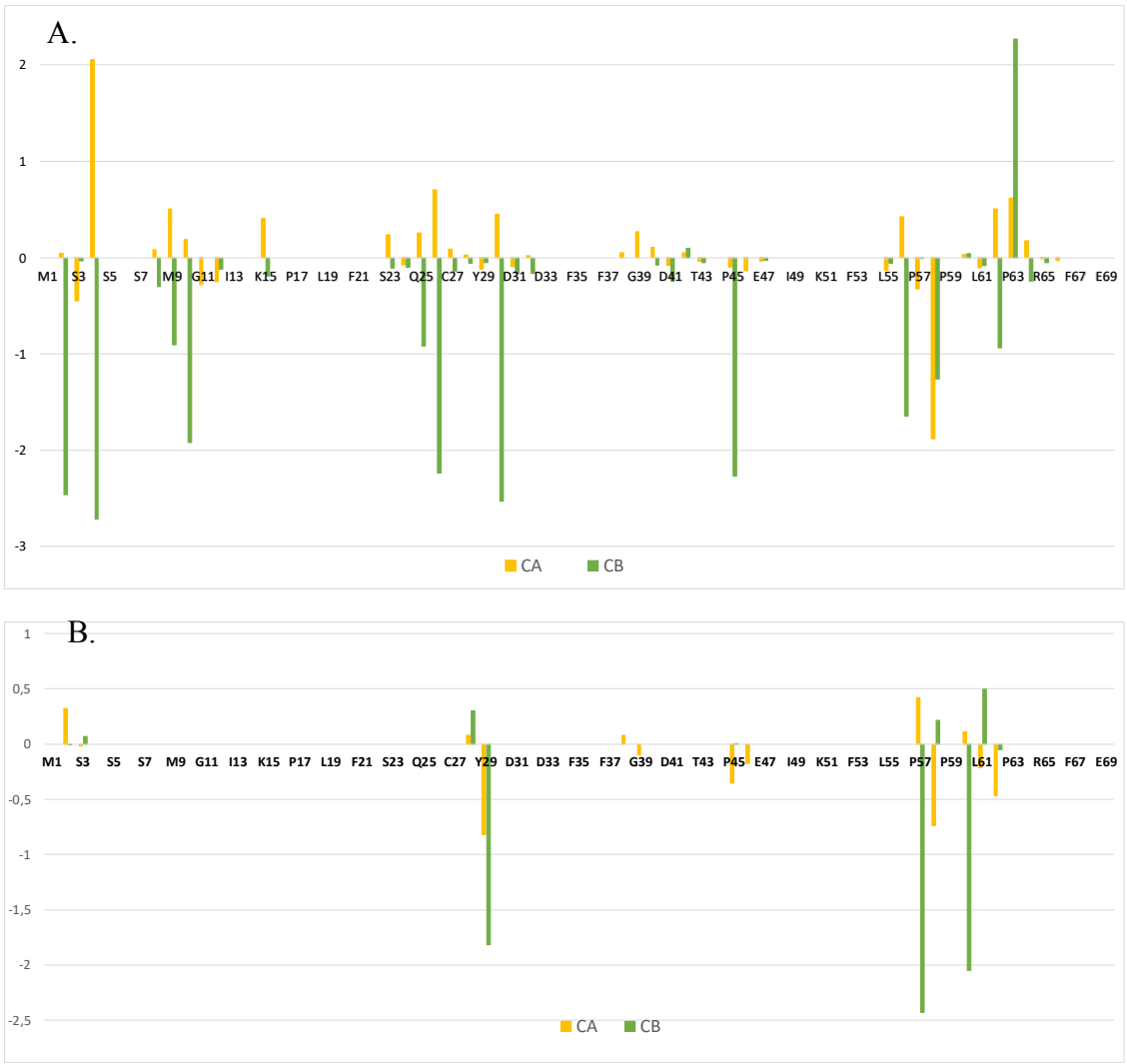


Figure 19 – Overlay of the difference in in $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shifts between Major states and A: Minor states with higher intensity and B: minor states with lower intensity. The shift from the minor peak is subtracted from the major peak, $^{13}C^{\alpha}$ shift difference in yellow and $^{13}C^{\beta}$ shift difference in green.

The difference between minor and major state can be presented as calculated CSP, from the equation described in section 3.4.1 plotted against their assigned residue number (figure 20), seven values is exceeding one standard deviations away from the mean.

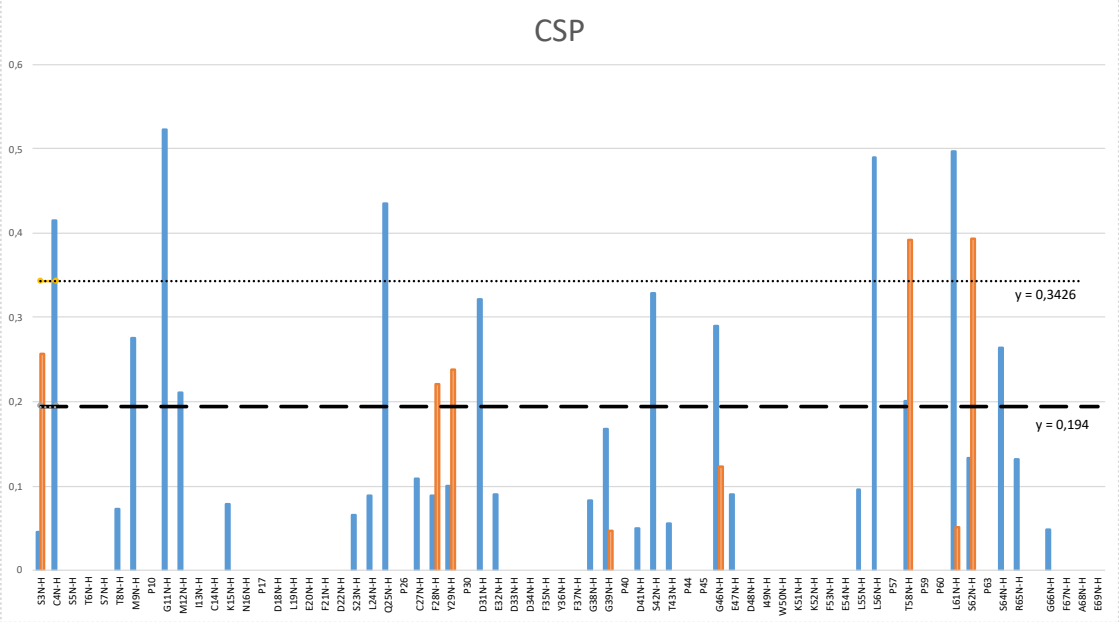


Figure 20 -Calculated CSP for N-Myc 1-69, Blue shows one minor state, and orange shows the other minor state. The dashed line is the mean and the dotted line is one standard deviation away from the mean.

The CSP is between 0 and 0.57 for all residues, and higher in the N- terminus and C-terminus and lower in the middle. The CSP is calculated from the difference in ^{15}N and $^{1}H^{N}$ chemical shift between the minor and the major state, and is therefore can be seen as a representation of how far away the minor peaks are from the major peaks in the HSQC spectrum.

The next step was to look at the C_CON, N_CON, N_CAN and N_(H)CAN spectrum to investigate the prolines, since they are not detectable in HSQC or other proton detected experiments, due to lacking a amide-proton. The C_CON and N_CON are similar experiment, and shows the N(i) and C^{o}(i-1) correlation, the difference is which nuclei that are detected, for C_CON the nucleus is ^{13}C and for N_CON the nucleus is ^{15}N. The result was almost the same, except that C_CON had higher signal to noise and is therefore the one presented in this thesis. The difference between N_CAN and N_(H)CAN is that the magnetisation starts at the $^{1}H^{\alpha}$ nuclei, but do not detect on it, for the N_(H)CAN experiment. The result was that N_(H)CAN had higher signal to noise ratio and is the variant presented in this thesis. All amino acids except M1 could be assigned in the C_CON spectrum, se figure 21. All prolines are displayed in the bottom right corner of figure 21a, and that area is magnified in figure 21b. All the prolines corresponds well to the sectioning behaviour described in (Murrali *et al.*, 2018), i.e prolines preceded by a proline have the lowest ^{15}N shifts but high ^{13}C shifts, and prolines proceeded by a threonine have the highest ^{15}N shifts and the lowest ^{13}C shifts.

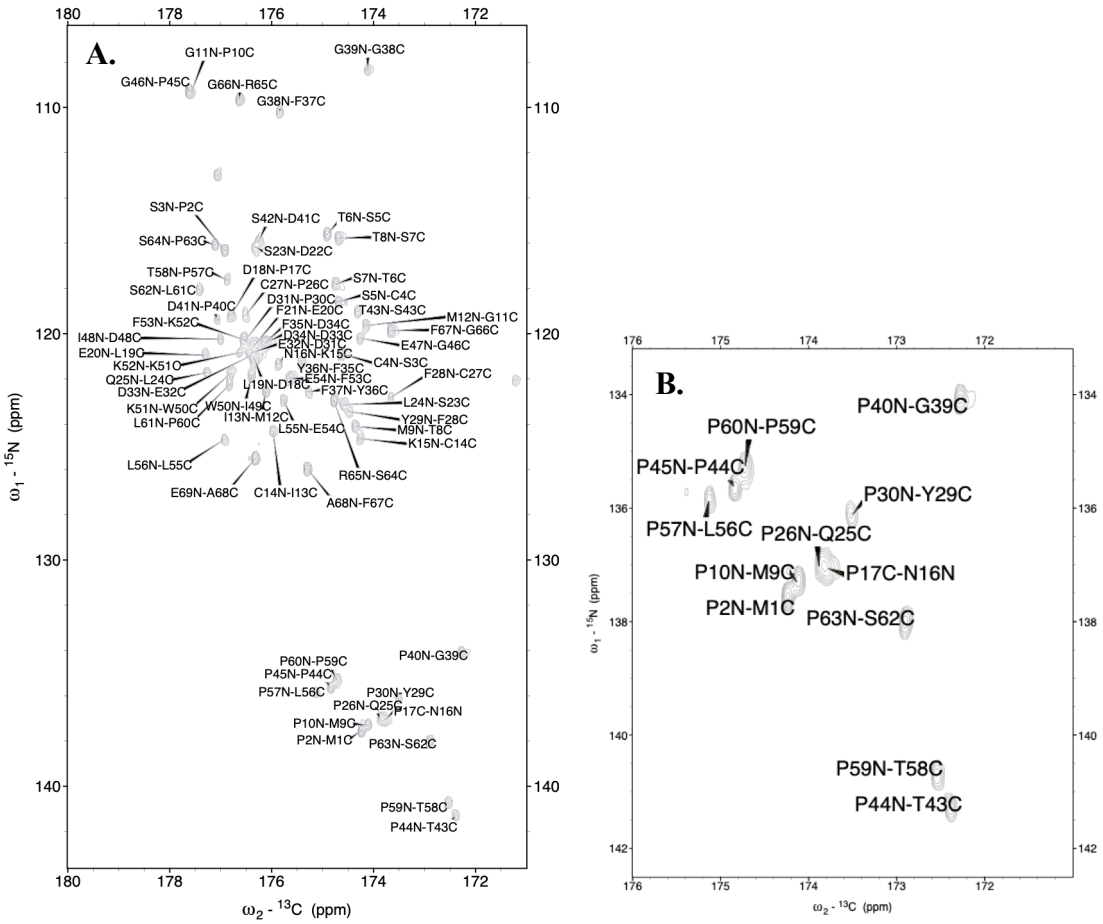


Figure 21 - C_CON spectrum of N-Myc 1-69, showing the $^{13}C^{O}$(i-1)-^{15}N(i) correlation, **A**: whole spectrum and **B**: The proline region, due to the peculiar shifts of prolines all prolines results in a well-separated region.

Next was to assign the N_(H)CAN spectra, and especially the proline region, this was successfully done and all C^{α} and C^{δ} shifts for prolines was detected, se figure 22.

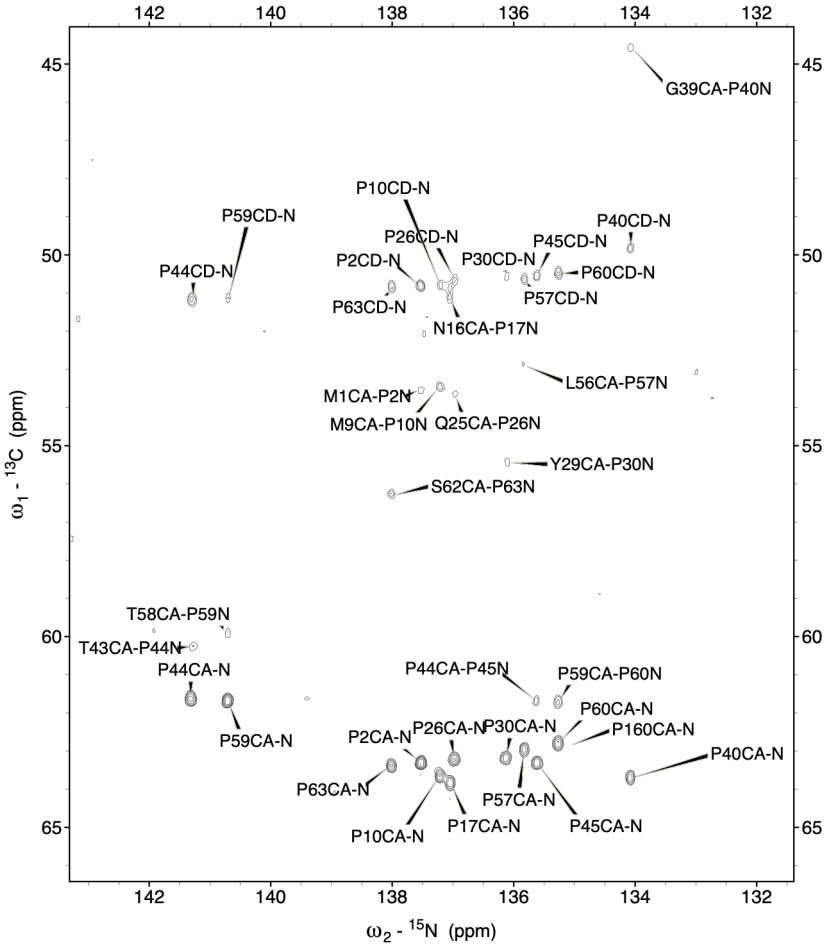


Figure 22 – N_(H)CAN spectrum of N-Myc 1-69 zoomed in to proline region, a correlation experiment between N(i), C^{α}(i) and C^{α}(i-1).

The direct carbon detected spectra C_CON and N_(H)CAN, figure 21 and figure 22, did not provide better sensitivity than the protein detected spectra and the minor states unfortunately was undetectable. However, the major states was detected and all the prolines was displayed with good sensitivity, so the carbon detected experiment definitely has potential when investigating prolines, but optimisation is necessary to be able to see the minor states.

To be able to establish if the prolines are in cis or trans configuration the C^{β} and C^{γ} chemical shift must be detected, this was done using the HN(CCO)C spectrum and the CCH-TOCSY spectrum, for examples se figure 23 and figure 24 respectively. The HN(CCO)C spectrum was used to assign the shifts for all minor states and all the major states except P44 and P57, because they precedes a proline and is therefore not visible in a HN(CCO)C spectrum. The C^{α}, C^{β}, C^{γ} and C^{δ} chemical shift for the two residues P44 and P57 are from the assigned CCH-TOCSY spectrum.

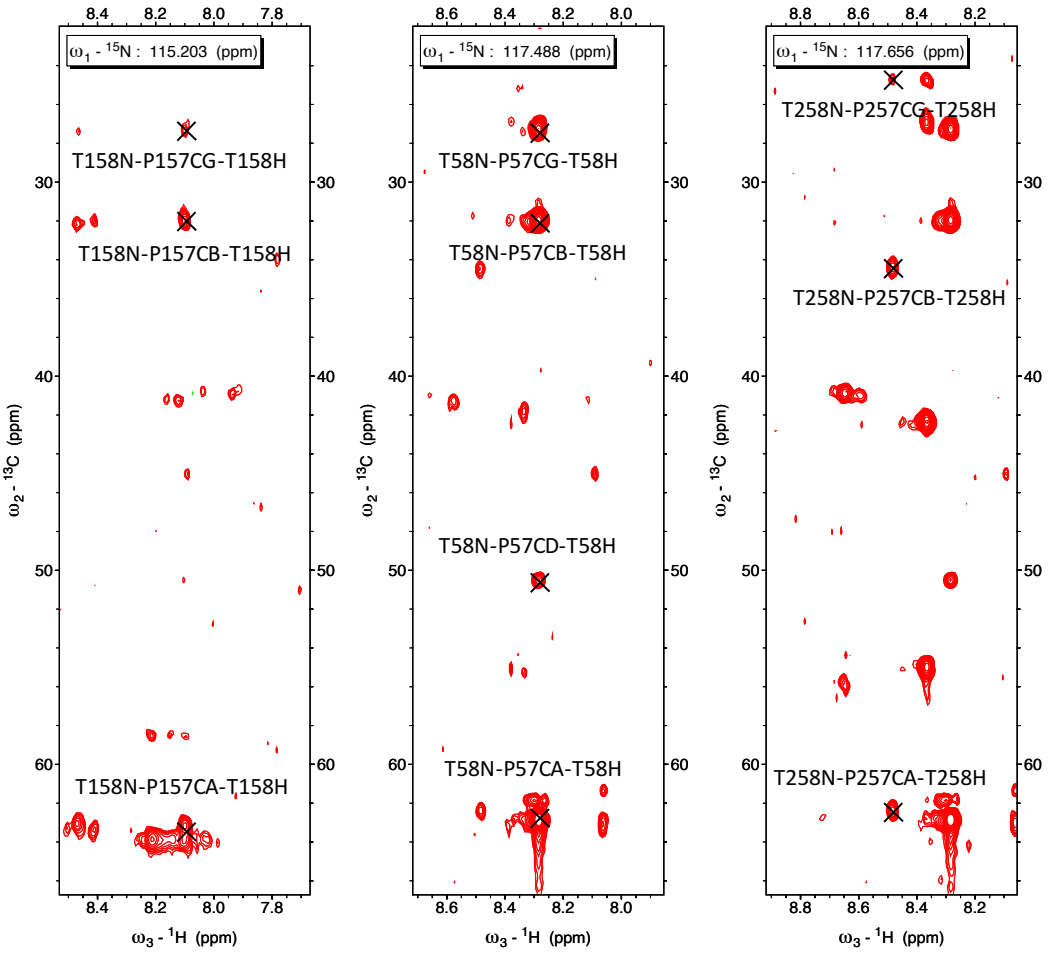


Figure 23 – Strips from HN(CCO)C spectra for P57, P157 and P257. Notice the difference in C^{γ} (CG) chemical shift for P257 ≈ 24 ppm, compared to P57 and P157 ≈ 27 ppm.

The HN(CCO)C spectra was of mixed quality, some regions have very good resolution and signal to noise ratio, and some have not. The assigned C^{γ} chemical shifts $^{\gamma}$, can be found in table VII. The literature values of prolines is $^{13}C^{\beta}$(cis)= 33.8±1.2 ppm, $^{13}C^{\beta}$(trans)= 31.8±1 ppm, $^{13}C^{\gamma}$ (cis)= 24.4± 0.7 ppm, $^{13}C^{\gamma}$ (trans)= 27.4± 0.9 ppm (Shen et al, 2010). The C^{γ} is the best indicator of cis vs trans configuration, due to the largest difference between the states. The C^{γ} of minor proline states P102 and P126 could not be detected and therefore makes it hard to say if they are in cis or trans configuration even though the C^{β} indicates trans.

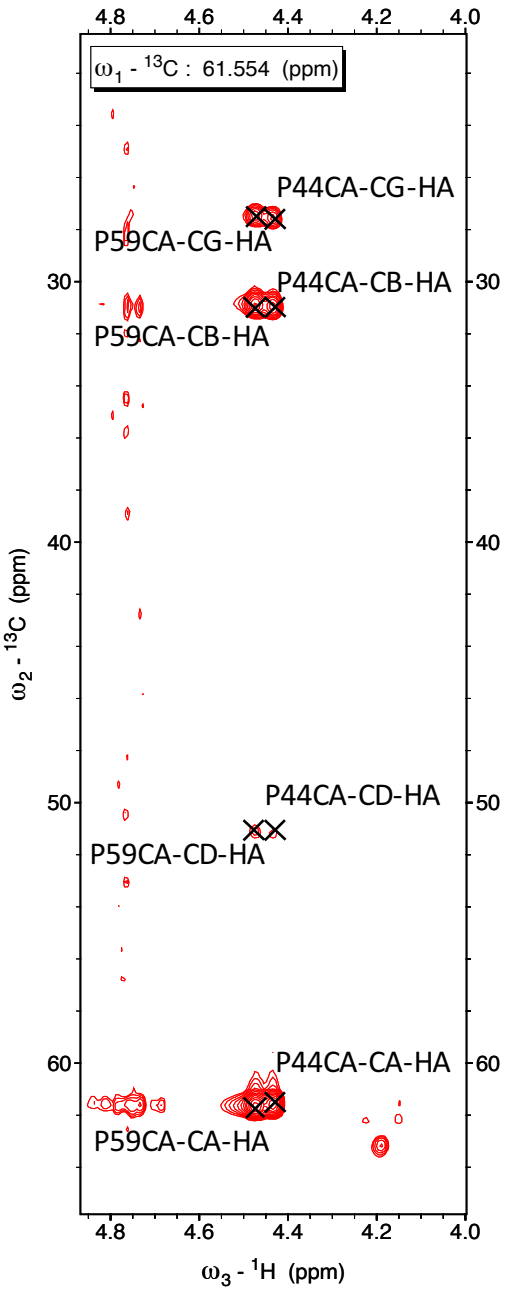


Figure 24 – CCH-TOCSY for P44 and P59.

All the detected chemical shifts for $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ for prolines can be found in table VII. The $^{13}C^{\gamma}$ is missing for P102, P202, P110, P117, P126, P245 and the two indistinguishable peaks called P117/240. When comparing the measured chemical shifts of $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ and comparing them to literature values, all chemical shifts for the major states (P2, P10, P17, P26, P30, P40, P44, P45, P57, P59 P60 and P63) suggest that they are in trans configuration to the previous amino acid. Out of the minor states, P102, P110, P130, P145, P257, P260 and P263 corresponds to cis configuration and P202, P126, P140, P245, P157 and P160 to trans configuration compared to literature values.

Table VII – Measured chemical shifts for assigned $^{13}C^{\beta}$ and $^{13}C^{\gamma}$, from HN(CCO)C and CCH-TOCSY spectrum. The compared literature values of prolines are $^{13}C^{\beta}$(cis)= 33.8±1.2, $^{13}C^{\beta}$(trans)= 31.8±1, $^{13}C^{\gamma}$ (cis)= 24.4± 0.7, $^{13}C^{\gamma}$ (trans)= 27.4± 0.9, according to the Shen et al, 2010. Cis configurations highlighted with grey background.

	$^{13}C^{\beta}$	**Literature**	$^{13}C^{\gamma}$	**Literature**
P2	32.291	Trans	27.444	Trans
P102	34.639	Cis	-	
P202	32.165	Trans	-	
P10	32.16	Trans	27.461	Trans
P110	34.024	Cis	-	
P17	32.212	Trans	-	
P26	32.194	Trans	27.312	Trans
P126	32.396	Trans	-	
P30	32.184	Trans	27.227	Trans
P130	34.755	Cis	24.692	Cis
P40	32.107	Trans	27.111	Trans
P140	32.223	Trans	27.223	Trans
P44	30.930	Trans	27.531	Trans
P45	32.080	Trans	27.439	Trans
P145	34.189	Cis	25.734	
P245	32.041	Trans	-	
P57	32.080	Trans	27.378	Trans
P157	32.112	Trans	27.342	Trans
P257	34.506	Cis	24.781	Cis
P59	30.876	Trans	27.478	Trans
P60	32.057	Trans	27.364	Trans
P160	32.008	Trans	27.248	Trans
P260	34.165	Cis	24.662	Cis
P63	32.155	Trans	27.337	Trans
P163	34.405	Cis	24.924	Cis
P117/240	34.599	Cis	-	
P117/240	33.967	Cis	-	

All assigned minor states, and how they are connected are illustrated in figure 25. The result from table VII is added and represented in different colours, and the two 2D ^{15}N Plane: iHNCA XP and HNCO PX, figure 14 and figure 15, is represented in underlined residues.

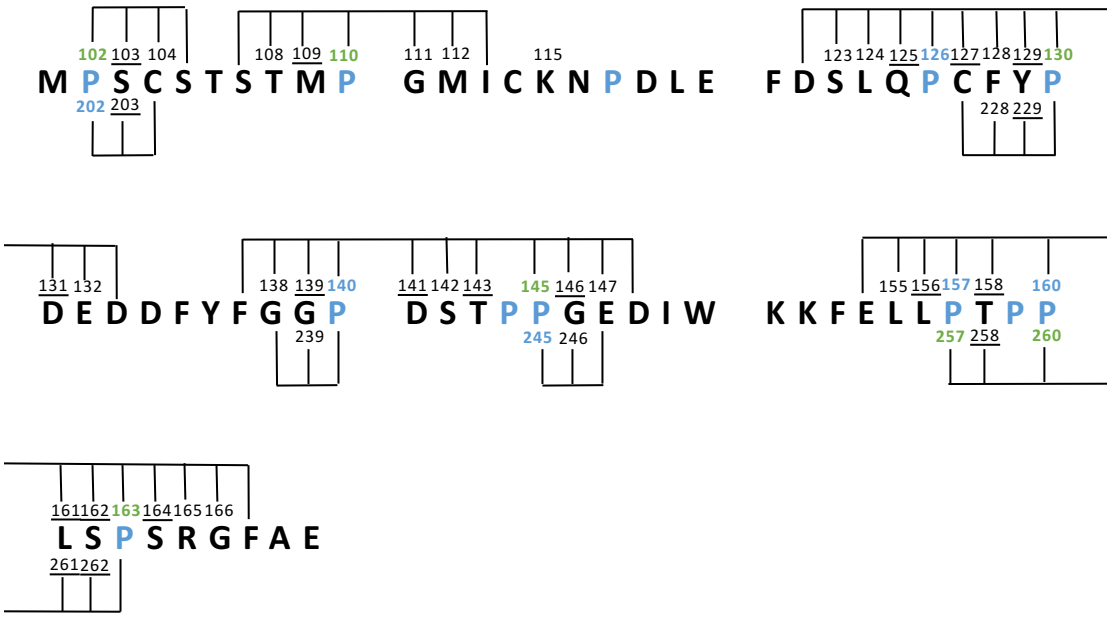


Figure 25 – Overview of all minor states of N-Myc. In the middle is the one-letter code for the protein sequence, which represents the major state. Above the sequence is the minor state with the highest intensity, named 102, 103 and so on, and the bottom row is the minor states with the lowest intensity. Prolines are highlighted in green and blue dependent on if they are found in cis or trans configuration, respectively, se table VII. The two peaks D118/D241 could not be distinguished in between, and are removed from this figure. The lines indicates connected peaks, for example; P2 is found in two conformations, one indicating cis (102) and one indicating trans (202), these two minor states is connected to other minor states of the following S2 (S103 and S203, respectively), and this continues until no minor states is no longer visible and have joined the major state. The underlined minor states is found in the two 2D ^{15}N Plane: iHNCA XP and HNCO PX.

P2 have two alternative conformations, P102 and P202, where P102 has the higher intensity (figure 16), and therefore can be interpreted as most populated. P102 is in cis configuration and P202 is in trans configuration, when compared to literature (Table VII), so the following peaks S103 and C104 can be explained by cis configuration of P102, S203 on the other hand cannot be explained by a proceeding cis proline. P10 has a minor state in cis configuration, thus the minor states T108, M109, G111 and M112 can be derived from cis/trans isomerisation. Even though no minor state of P17, could be detected, the two possible states marked in yellow in table VII, both indicate cis configuration. So the peaks K115 can be explained by a close cis proline, even if the peaks D118/D241 could not be distinguished from each other.

The C^{β} chemical shift for P126 indicates trans configuration, however no C^{γ} shift is detected, and the C^{β} shift is only 0.2 ppm away from being interpreted as a cis configuration, therefor making the uncertainty of cis or trans configuration, very high. The minor states S123, L124, and Q125 have $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shift that differs the most for Q125 and decreases more and more for L124 and S123, implicating that P126 might after all be in cis configuration. The same $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ pattern is true for C127,

F128 and Y129, whom decreases away from P126. P130 is cis configuration, and since the chemical shifts for both C^{β} and C^{γ} indicates cis makes it more certain. Y229 have a larger chemical shift difference than F228. The pattern of decreasing chemical shift away from a proline is also indicating that these two minor states comes from cis configuration of proline 30.

The chemical shifts for P140 statistically are in trans configuration, and the shift difference of the previous residues G139 and G138 have very small shift differences. However the two indistinguishable P117/240 peaks marked in yellow in table VII, implicates that there are a minor state of P40 that is in cis configuration and can be the explanation of at least the presence of one minor state of G39. Minor states D141, S142 and T143 can be due to either cis P240, or that P44 exists in a minor cis state, this could not be measured in the set-up of experiments, so no conclusion over minor states of P44 can be drawn.

P45 exists in two different minor states, whereof one is in cis (P145) and one in trans (P245). The minor cis conformation have been confirmed with both the C^{β} and C^{γ} shift, and the minor trans conformation determination have only been based on the C^{β} shift. Both minor states G146 and E147 can be explained by cis configuration of P45 and their difference in $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift is decreasing away from P145, which is also implicating cis/trans isomerisation. G246 have also a small $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ difference, but is not proceeded by a cis proline, one plausible explanation is that it exists a minor state of P44 in cis configuration, as mentioned before, but this has not been proven in any way.

P57 also exists in two different minor states, P157 and P257. P157 is in trans according to both C^{β} and C^{γ}, and P257 is in cis confirmed both by C^{β} and C^{γ}. The minor states for L55 and L56, named L155 and L156, can be explained by the cis conformation of P257. The $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ difference in chemical shift values of L155 and L156 is increasing towards residue 57, implicating that they are in fact a result from cis configuration of P257. Regarding T158 and T258 it is hard to draw any conclusions regarding which proline they might be affected by, since P257 is in cis, as well as P260, and no data is detected for eventual minor states of P59.

P163 is in cis configuration and together with the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shift difference pattern for L161, S162, S164, R165, and G166 which is increasing until residue 63, and then decreasing, highly suggests that these minor peaks is because of cis configuration of P163.

4.2.3 Estimating population distribution

To determine how populated each state was, the assigned HSQC was loaded in to the software PINT, and the volume of each peak was calculated by integration. Each volume was then divided with the sum of all the volumes from the same peak, as described in section 3.4.3, to get the relative populations in percent. This was done for all peaks and plotted against residue number in figure 26.

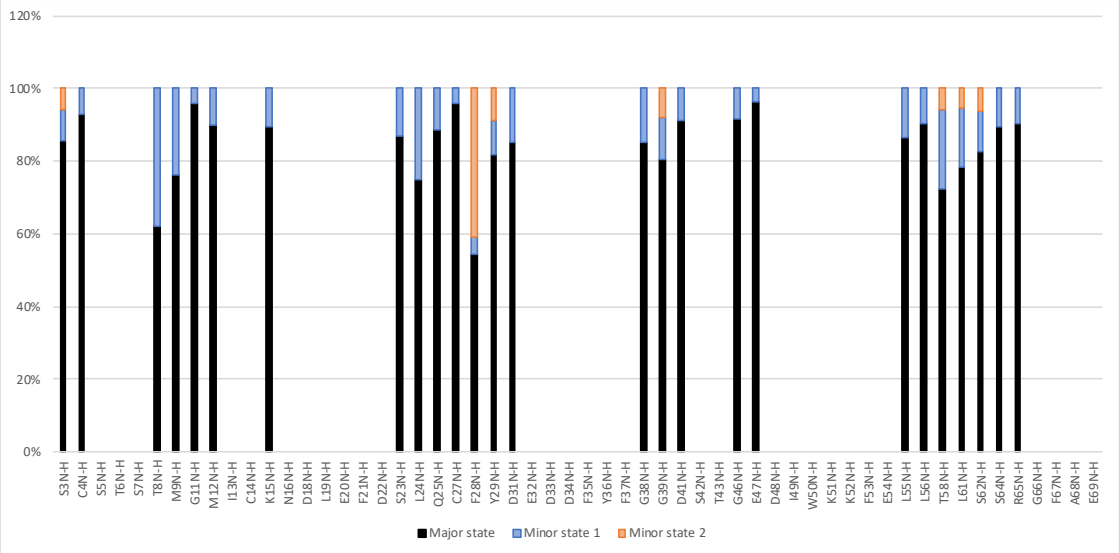


Figure 26 – Representation over the ratios of peak volumes plotted against residue number. Black shows the major state of N-Myc and blue and orange shows the two minor states, respectively. The volume of each peak is calculated using PINT. Notice that the minor states S142 and T143 has been removed due to not sensible enough to be visible in PINT, the peaks included in this plot had a score fitting of 3 out of 4 or higher.

All volumes for the major state are approximately 80% except for F28, that only is 57% due to very populated minor state 2, i.e. F228.

The $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C^{O}$, ^{15}N and H^{N} values for all assigned spectra was added into the pearl script for calculation of the SSP, to see if the minor and minor states affects the secondary structure differently. The SSP for minor and major states is plotted against residue in figure 27. A correct folded α- helix have a SSP value of +1, compared to a folded β-sheet of -1, values around 0, indicate random coil. The score for the minor and major state ranges in between +0.15 and -0.6, i.e. more against a β-sheet folding.

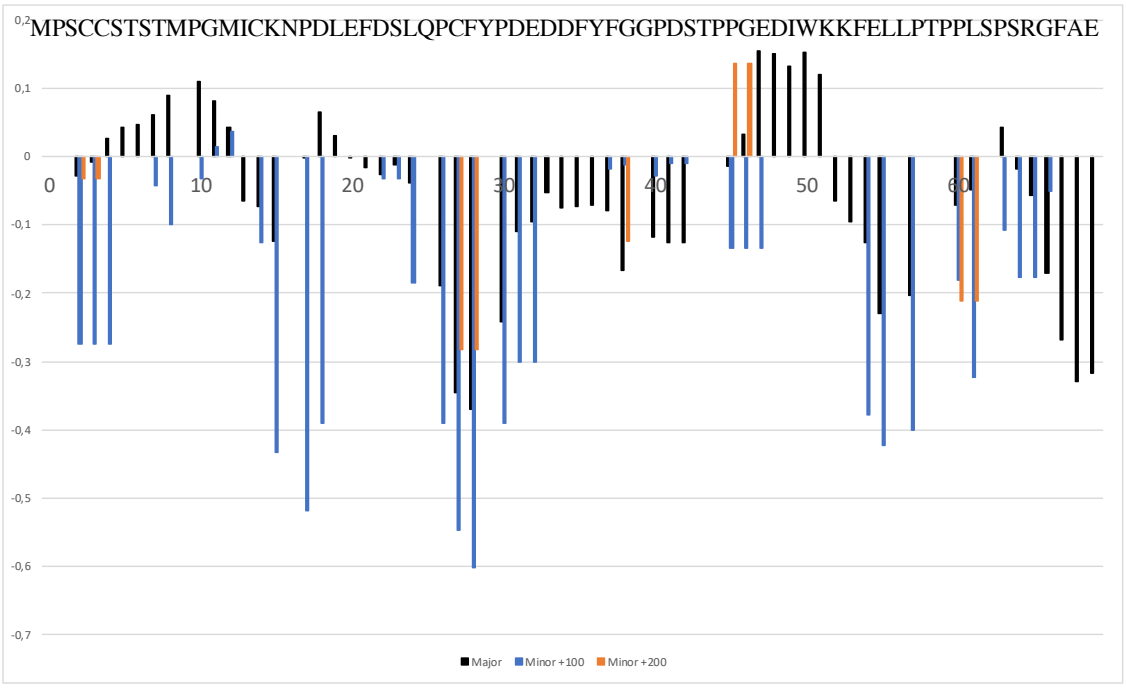


Figure 27- SSP score of Major states in black, minor state and high intensity in blue and minor state and low intensity in orange. The SSP for P257 was removed from this plot, because the SSP was below -1.

To easier see the difference in SSP between the Major and minor state, a plot over ΔSSP was made, se figure 28, where the SSP for the major state is subtracted from the SSP for the minor state.

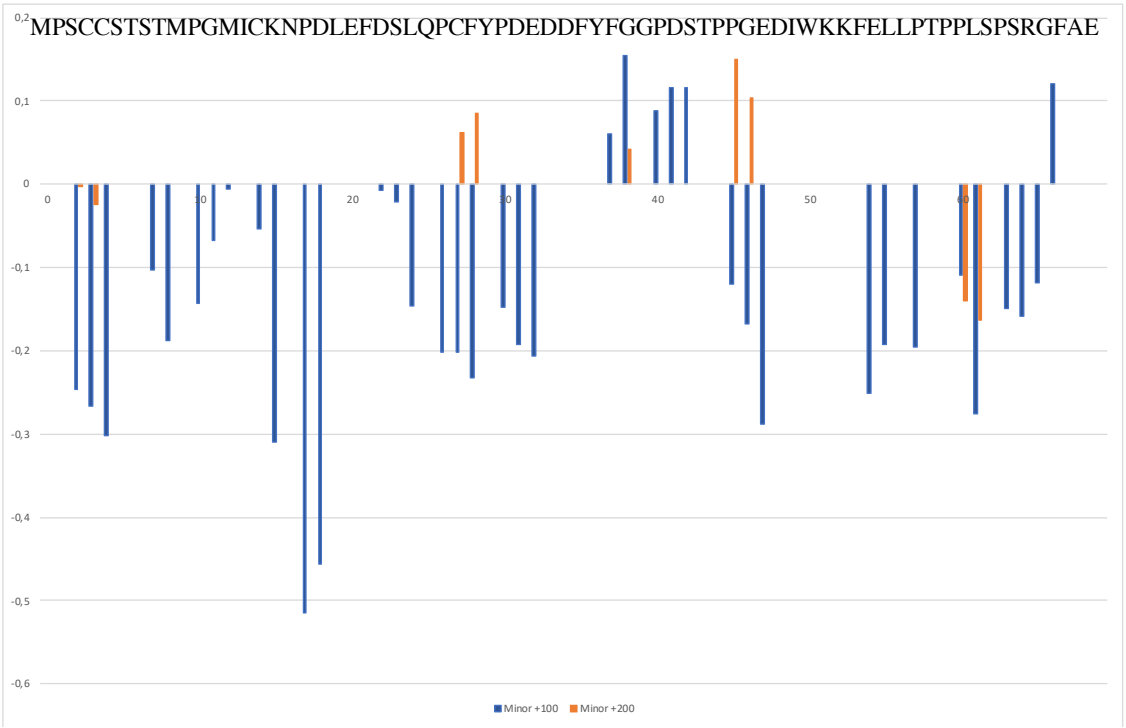


Figure 28 – Difference in SSP for minor stares compared to major states. Minor state labelled +100 in blue, minor state labelled +200 in orange.

The ΔSSP score for +100 minor state overall indicates more β-sheet folding, except for residue 37-42. The ΔSSP for +200 minor state indicate more β-sheet folding in the N terminus and C-terminus of the protein, but a an increased α-helical propensity in the middle.

5 Conclusions and Future prospects

The majority of the minor peaks, could be explained by a close located cis proline, this is new information that was not known before. To confirm that the minor states is correctly assigned, it would be good to do some experiments that connects over the prolines, example (H)N(CA)NH that connects three backbone Nitrogen's; N(i), N(i-1) and N(i+1). Some of the C^{γ} values for prolines in the HN(CCO)C experiment was undetectable, therefore it would be interesting to repeat this experiment , and run it for a longer time at more optimized conditions to be able to see all the carbons. The carbon detected experiments was interesting, but need to be optimized for N-Myc, to enable detection of the minor peaks.

When comparing the SSP for N-Myc with published data for C-Myc (Andresen et al, 2012), the SSP for N-Myc have higher propensity for β-Sheet structure in general. The largest difference in SSP, are located in the regions where the prolines in the sequences does not agree. For example, residue 30, which is a proline in N-Myc shows β-Sheet propensity, and residue 30 in C-Myc is an Asparagine which shows α-helical propensity. This trend with the largest difference in SSP where the prolines in the sequences differentiates, is consistent in the whole protein. This indicates that prolines plays an important role in the secondary structure formation.

The most interesting experiment to do is to phosphorylate T58 and S62 in separate experiments and see if/how the cis /trans behaviour of prolines changes, especially considering P63. This to get a better understanding of how phosphorylation's affects the prolines and in the future understand the mechanism of activation and deactivation of N-Myc regulation.

N-Myc has proven to not only be easier to produce, but also gives a better sensitivity and dispersion of peaks in the NMR spectra compared to C-Myc. And since N-Myc and C-Myc have similar sequences, especially in the region around T58, that is together with S62 the two known phosphorylation sites that regulates the degradation, many more experiments and researcher should be focused on the N-Myc variant.

Acknowledgements

I firstly wants to thank Alexandra Ahlner for always taking the time to explain, and to answer questions, to provide support and encouragement and drive the project forward. And I would like to thank Johanna Hultman for helping me with everything that involves protein production and purification, and for her endless patience when it comes to questions.

And I also want to thank Ulrika Brath and Cecilia Persson at Swedish NMR center in Gothenburg, for recording the NMR experiments and provide good discussions and help with interpreting the Data.

I am grateful for the opportunity to attend the interesting presentations and discussions at the weekly meetings in Eleonore von Castlemur and Maria Sunnerhagen research groups.

And lastly, many thanks to my examiner Maria Sunnerhagen, for financially support this project and for offering knowledge and enthusiasm not only about this protein, but also for research in general.

References

Andresen, C. et al. (2012) 'Transient structure and dynamics in the disordered c-Myc transactivation domain affect Bin1 binding', Nucleic acids research, 40(13). doi: 10.1093/nar/gks263.

Andrèsen, C. *et al.* (2011) 'Transient structure and intrinsic disorder in the c-Myc transactivation domain and its effects on ligand binding'. Nucleic acid res. 40(13):6353-66.

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

Ashok Sekhar and Lewis E. Kay (2019) 'An NMR View of Protein Dynamics in Health and Disease'. Annu. Rev. Biophys. 48:297–319. doi: 10.1146/annurev-biophys-052118-115647

Ask A Scientist Staff (2019) *What is Chromatography and How Does It Work?* Available at: https://www.thermofisher.com/blog/ask-a-scientist/what-is-chromatography/ (Accessed: 26 May 2021).

Becker, W. Hardin, J. Kleinsmith, L. Bertoni, G. (2009) The World of the Cell. 7th edition. Pearson Education.

Bernstein, M.A.; King, K.F.; Zhou, X.J. (2004). *Handbook of MRI Pulse Sequences*. San Diego, CA: Elsevier Academic Press. p. 960. ISBN 0-12-092861-2 –2

Bermel, W. *et al.* (2012) 'Exclusively heteronuclear (13) C-detected amino-acid-selective NMR experiments for the study of intrinsically disordered proteins (IDPs)', *Chembiochem: a European journal of chemical biology*, 13(16). doi: 10.1002/cbic.201200447.

BMRB - Biological Magnetic Resonance Bank (no date). Available at: https://bmrb.io/ref_info/stats.php?restype=aa&set=full (Accessed: 24 May 2021).

Campen, A. et al. (2008) '*TOP-IDP-Scale: A New Amino Acid Scale Measuring Propensity for Intrinsic Disorder'. Protein & Peptide Letters 15:* 956 – 963. doi 10.2174/092986608785849164.

Cancerfonden (2021). Om cancer. https://www.cancerfonden.se/om-cancer. (Accessed: 17 May 2021).

Cohn, G. M. *et al.* (2020) 'PIN1 Provides Dynamic Control of MYC in Response to Extrinsic Signals', *Frontiers in Cell and Developmental Biology*, 8. doi: 10.3389/fcell.2020.00224.

Contents of Essentials of Cell Biology (2014). Available at: https://www.nature.com/scitable/ebooks/essentials-of-cell-biology-14749010/122996928/ (Accessed: 17 May 2021).

Gibbs E. et.al (2017). '*Application of NMR to studies of intrinsically disordered proteins*'. Archives of Biochemistry and Biophysics 628 57e70. doi: 10.1016/j.abb.2017.05.008

Hanahan D, Weinberg RA (2011). 'Hallmarks of cancer: The next generation'. Cell. 144(5):646–74. doi: 10.1016/j.cell.2011.02.013

Mateos, B. et al. (2020) '*The Ambivalent Role of Proline Residues in an Intrinsically Disordered Protein: From Disorder Promoters to Compaction Facilitators*', Journal of molecular biology, 432(9). doi: 10.1016/j.jmb.2019.11.015.

Marsh, J. A. et al. (2006) '*Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation'*, Protein science: a publication of the Protein Society, 15(12), p. 2795.

Michael J. Duffy, Shane O'Grady, Minhong Tang, John Crown. 2021. '*MYC as a target for cancer treatment'*. Elsevier Ltd. doi 10.1016/j.ctrv.2021.102154

Murrali, M. G. *et al.* (2018) 'Proline Fingerprint in Intrinsically Disordered Proteins', *Chembiochem: a European journal of chemical biology*, 19(15). doi: 10.1002/cbic.201800172.

Niklasson, M., Otten, R., Ahlner, A., Andrésen, C., Schlagnitweit, J., Petzold, K. & Lundström, P. (2017). 'Comprehensive analysis of NMR data using advanced line shape fitting'. Journal of Biomolecular NMR, 69(2), 93-99

NMRFAM-Sparky Distribution (no date). Available at: https://nmrfam.wisc.edu/nmrfam-sparky-distribution/ (Accessed: 28 May 2021).

Oldfield Christopher J. and Dunker A. Keith. (2014). *'Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions'. Annual Review of Biochemistry* 83:553–84. doi 10.1146/annurev-biochem-072711-164947.

Schramm A. et al (2019). 'An arsenal of methods for the experimental characterization of intrinsically disordered proteins – How to choose and combine them?' Arch. Biochem. Biophys. 676(June):108055

Shen, Y and Bax, A. (2010) 'Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts', *Journal of biomolecular NMR*, 46(3), p. 199.

Uversky VN, Oldfield CJ, Dunker AK .2008.*'Intrinsically disordered proteins in human diseases: introducing the D2 concept*'. Annual Review of Biophysics. 37: 215–46. doi:10.1146/annurev.biophys.37.032807.125924

Wedemeyer, W. J., Welker, E. and Scheraga, H. A. (2002) 'Proline cis-trans isomerization and protein folding', *Biochemistry*, 41(50). doi: 10.1021/bi020574b.

Weiss, M. S., Jabs, A. and Hilgenfeld, R. (1998) 'Peptide bonds revisited', *Nature structural biology*, 5(8). doi: 10.1038/1368.

World Health Organization. Cancer. https://www.who.int/health-topics/cancer#tab=tab_1. (Accessed: 25 May 2021).

Zimei Bu, David J.E. Callaway, 2011. Chapter 5 - Proteins MOVE! Protein dynamics and long-range allostery in cell signaling, Advances in Protein Chemistry and Structural Biology, Academic Press, Volume 83, Pages 163-221.

Appendix

A1. ProtParam for N-Myc(1-69)^{WT} & N-Myc(1-69)^{P44L}

Molecular weight: 7678.60 N-Myc(1-69)^{WT} , 7694.64 N-Myc(1-69)^{P44L}
Theoretical pI: 3.95 for both N-Myc(1-69)^{WT} & N-Myc(1-69)^{P44L}
Extinction coefficients: 8605 $M^{-1}cm^{-1}$ for both N-Myc(1-69)^{WT} & N-Myc(1-69)^{P44L} (at 280 nm measured in water)

Amino acid composition:

	N-Myc(1-69)^{WT}		N-Myc(1-69)^{P44L}	
Ala (A)	1	1.4%	1	1.4%
Arg (R)	1	1.4%	1	1.4%
Asn (N)	1	1.4%	1	1.4%
Asp (D)	7	10.1%	7	10.1%
Cys (C)	3	4.3%	3	4.3%
Gln (Q)	1	1.4%	1	1.4%
Glu (E)	5	7.2%	5	7.2%
Gly (G)	5	7.2%	5	7.2%
His (H)	0	0.0%	0	0.0%
Ile (I)	2	2.9%	2	2.9%
Leu (L)	5	7.2%	6	8.7%
Lys (K)	3	4.3%	3	4.3%
Met (M)	3	4.3%	3	4.3%
Phe (F)	6	8.7%	6	8.7%
Pro (P)	12	17.4%	11	15.9%
Ser (S)	7	10.1%	7	10.1%
Thr (T)	4	5.8%	4	5.8%
Trp (W)	1	1.4%	1	1.4%
Tyr (Y)	2	2.9%	2	2.9%
Val (V)	0	0.0%	0	0.0%

Atomic composition:

		N-Myc(1-69)^{WT}	N-Myc(1-69)^{P44L}
Carbon	C	345	346
Hydrogen	H	502	506
Nitrogen	N	78	78
Oxygen	O	109	109
Sulfur	S	6	6

A2. Protein purification for N-Myc(1-69)^{P44L}

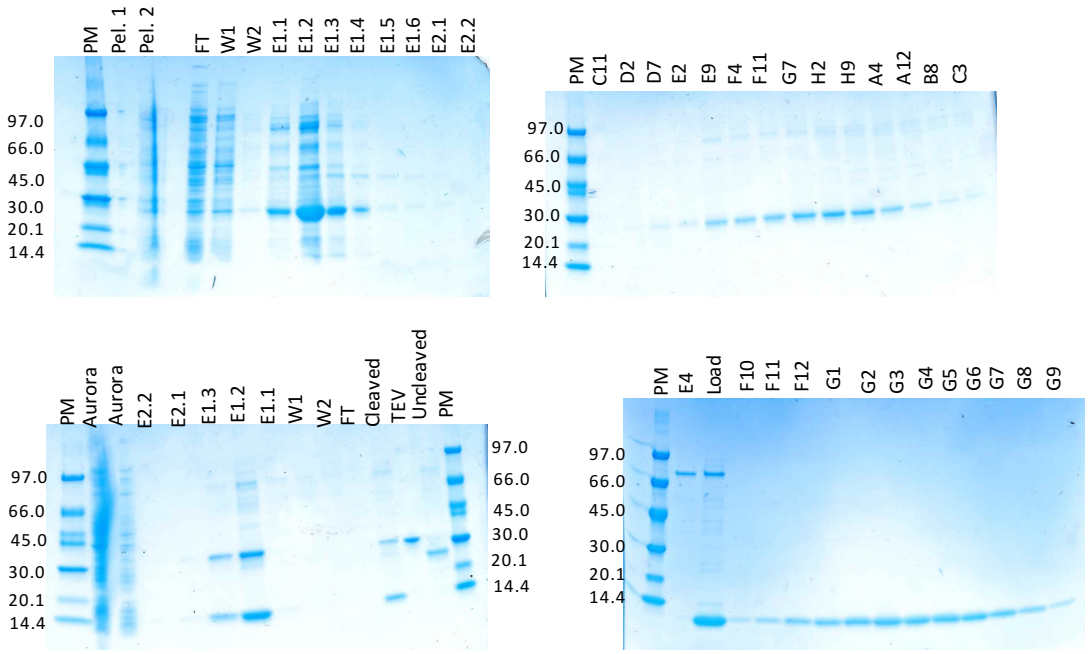


Figure 29: SDS-PAGE gels of N-Myc 1-69P44L A: IMAC, B: Anion exchange, C: Reverse IMAC, D: Gel filtration.

Fraction	F10	F11	F12	G1	G2	G3	G4	G5	G6	G7	G8	G9
A260/280	0.667	0.758	0.717	0.824	0.744	0.763	0.769	0.767	0.772	0.865	0.858	0.745
A280	0.031	0.063	0.161	0.247	0.425	0.539	0.556	0.457	0.334	0.201	0.109	0.057
c (um)	3.60	7.32	18.71	28.70	49.39	62.64	64.61	38.81	38.81	23.36	12.67	6.62
Saved in	-	1 ml	1 ml	1 ml	1 ml	2 x 0.5 ml	2 x 0.5 ml	1 ml	1 ml	1 ml	1 ml	-