Investigation of system properties related to MYCN oncogene expression in neuroblastoma

By
Nicholas Shipillis

A thesis submitted to the University of Manchester for the degree of Doctor of Science

in the school of
Chemical Engineering and Analytical Science
Systems Biology Doctoral Training Centre

2012
# TABLE OF CONTENTS

CONTENTS PAGES ........................................................................................................... 1-4

LIST OF FIGURES ............................................................................................................. 5-7

LIST OF TABLES ............................................................................................................... 8-9

ABREVIATIONS .................................................................................................................. 10-12

ABSTRACT .......................................................................................................................... 13

DECLARATION OF AUTHORSHIP ..................................................................................... 14

COPYRIGHT ....................................................................................................................... 15

ACKNOWLEDGEMENTS ..................................................................................................... 16

CHAPTER 1: INTRODUCTION ............................................................................................. 17

1.1. Classification of Neuroblastoma .................................................................................. 18

1.2. Biological features and markers of NB ....................................................................... 19
   1.2.1. NB markers present in serum or urine ................................................................. 19
   1.2.2. NB pathological markers ..................................................................................... 20
   1.2.3. NB cytogenetic markers ...................................................................................... 21
   1.2.4. Molecular markers associated with NB ............................................................... 23

1.3. The MYCN gene and NB ............................................................................................ 24
   1.3.1. MYCN gene amplification and NB ....................................................................... 24
   1.3.2. Association of MYCN amplification with NB-related chromosomal abnormalities and genes located at 2p24 ................................................................. 28
   1.3.3. Expression of MYCN and its significance for NB prognosis- studies based on NB cell-lines ........................................................................................................... 30
   1.3.4: Expression of MYCN and its significance for NB prognosis- studies based on NB clinical samples ........................................................................................................ 31
   1.3.5: Conclusions from studies of MYCN expression in NB ....................................... 33
   1.3.6: Difference of prognostic significance in NB between MYCN amplification and expression .............................................................................................................. 34

1.4. Structure, activity and control of MYCN. potential contributors towards heterogeneous MYCN transcript, protein and activity levels .......................................... 36
   1.4.1. Role of MYCN and the MYC family in normal cells ............................................. 36
   1.4.2. Structural features of the MYC family protein ..................................................... 38
   1.4.3. The myc-associated factor x (MAX) .................................................................... 40
1.4.4. Controlling MYCN and MYC family proteins at the genetic level in normal cells.................................................................41
1.4.5. Controlling MYCN and MYC family proteins at the protein levels in normal cells........................................................................43
1.4.6. The MYC/MAX/MAD network..............................................................................................................................................45
1.4.7. Regulation of MYC expression through feedback loops................47

1.5. Systems Biology approaches to biomedical investigations..................49
1.5.1. Top-down Systems Biology........................................................50
1.5.2. Bottom-up Systems Biology..........................................................51

1.6. qPCR data analysis at the DNA level..................................................55

1.7. Project aims and objectives.................................................................60

CHAPTER 2: MATERIALS AND METHODS.................................................64

2.1. Cell culture.......................................................................................64

2.2. Nucleic Acid Extraction......................................................................64
2.2.1. DNA.............................................................................................64
2.2.2. RNA..............................................................................................65
2.2.3. cDNA............................................................................................65

2.3. Quantitative Polymerase Chain Reaction (qPCR) ..................................66

2.4. Sample preparation for protein-content analysis..................................67

2.5. Enzyme Linked Immunosorbent Assay (ELISA) ..................................67
2.5.1. Normal ELISA...............................................................................69
2.5.2. Sandwich ELISA............................................................................70

2.6. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and western blots..........................................................71
2.6.1. Buffer/gel materials and recipes.......................................................71
2.6.2. SDS-PAGE gel protein separation..................................................71
2.6.3. Western blot....................................................................................72

2.7. Dry Laboratory Materials....................................................................73

CHAPTER 3 RESULTS: MODELLING MYCN EXPRESSION.........................74

3.1. Overview..........................................................................................74

3.2. Defining the model............................................................................75

3.3. Connection of the model to the expression of MYCN during the cell cycle....76

3.4. Building the model............................................................................78
3.4.1. Rate laws for the reactions..............................................................78
3.4.2. Model parameters ........................................................................................................... 81

3.5. Model validation .................................................................................................................. 84
  3.5.1. Single copy gene MYCN .............................................................................................. 84
  3.5.2. MYCN gene amplification ......................................................................................... 87
  3.5.3. Reaction fluxes over time ......................................................................................... 91

3.6. Metabolic Control Analysis (MCA) and Sensitivity Analysis ........................................... 94

3.7. Conclusions ...................................................................................................................... 100

CHAPTER 4: RESULTS: MEASURING MYCN DNA AND RNA ........................................... 103

4.1. Overview .......................................................................................................................... 103

4.2. Quantification of MYCN Gene Copy Number .................................................................. 104
  4.2.1. Cq values, test efficiencies and data transformation .................................................... 104
  4.2.2. Analysis and comparison of the HKGs ...................................................................... 106
  4.2.3. Calculating gene copy-numbers for MYCN ................................................................. 109

4.3. Quantification of MYCN transcript numbers ................................................................. 110
  4.3.1. Construction of standard curves for mRNA absolute quantification ......................... 110
  4.3.2. The role of the calibrator in the absolute quantification of cDNA particle numbers .... 111
  4.3.3. Absolute quantification of MYCN mRNA ................................................................. 112
  4.3.4. Relative quantification of MYCN mRNA ................................................................. 115

4.4. Conclusions ...................................................................................................................... 116

CHAPTER 5: RESULTS: MEASURING N-MYC AND MAX PROTEINS .................................... 118

5.1. Overview ......................................................................................................................... 118

5.2. Quantifying N-Myc protein levels in neuroblastoma cell-lines ...................................... 120
  5.2.1. Preliminary ELISA work ........................................................................................... 120
  5.2.2. ELISA tests comparing Biotin-conjugated and FITC-conjugated secondary antibodies ................................................... 122
  5.2.3. Comparison of N-Myc detection in total cell lysate and soluble-only cell-lysate .......... 125
  5.2.4. Investigation of N-Myc protein levels using 4 different anti-N-myc primary antibodies ................................................... 127
  5.2.5. Sandwich ELISA for investigation of N-Myc protein levels ................................... 129
  5.2.6. Detection of N-Myc protein the neuroblastoma cell-lines using SDS-PAGE and western blots ................................................... 134
  5.2.7. Quantification of N-Myc protein using fluorescence-based western blots ................ 140
  5.2.8. Cell lysis and processing of lysates ......................................................................... 141

5.3. Quantifying MAX protein levels in neuroblastoma cell lines ........................................ 150
LIST OF FIGURES

Figure 1.1. The Event-Free Survival (EFS) of patients in relation to their MYCN GCN. Page 25

Figure 1.2. Effects of MYC transcriptional activity Page 38

Figure 1.3. Basic structural features of MYC proteins and their binding partner MAX. Page 39

Figure 1.4. Competition for MAX within the MYC/MAX/MAD network. Page 46

Figure 1.5. Expression patterns of MYC/MAX/MAD network proteins during the cell-cycle in normal cells. Page 47

Figure 1.6. Negative feedback loop between MYC and SIRT1. Page 48

Figure 1.7. The generated map of N-Myc protein and associated molecules that are deregulated in a 4(+) neuroblastoma sample in comparison to a 1(-) neuroblastoma sample. Page 51

Figure 1.8. Some of the key features of a bottom-up approach as illustrated in a myc-related study Page 53

Figure 1.9. Flow chart with the proposed work-plan of the study. Page 63

Figure 3.1. Representation of the MYCN model. Page 76

Figure 3.2A-F. The concentration of the model species as they change during a time-course simulation. Page 86

Figure 3.3. The values of [NMyC_MAX_Promoter] at the end of a 24 hour time-course with MYCN DNA amplification from 2 to 2000 GCN per cell (6.64*10^-7µM to 6.64*10^-4µM concentration). Page 88

Figure 3.4. The values of [NMyC_MAX_Promoter] at the end of a 24 hour time-course with MYCN DNA amplification from 2 to 20 GCN per cell (6.64*10^-7µM to 6.64*10^-6µM). Page 89

Figure 3.5. The percentage of occupied promoters by N-myc/MAX dimer in relation to total available E-Box promoters after a 24-hour time-course at 20 GCN of MYCN per cell. Page 89

Figure 3.6A-F. The fluxes through the model reactions during a 24-hour time-course with the model defined under different sets of parameters. Page 93

Figure 3.7. The Flux Control Coefficients for the dimerisation reaction in relation to change of flux at steady state for transcription, translation, RNA degradation and protein degradation with different parameter sets. Page 97
Figure 3.8A-D. The concentration control coefficients of transcription, translation, RNA degradation and protein degradation on MAX, NMyc_MAX, NMyc_MAX_Promoter and Promoter.

Figure 3.9. Sensitivity Analysis and the response coefficients for NMyc_MAX_Promoter in relation to the model reactions.

Figure 4.1. A standard curve for MYCN cDNA using the MYCN template oligonucleotide.

Figure 5.1. N-myc protein detection for the 4 neuroblastoma cell-lines using FITC-conjugated secondary antibodies.

Figure 5.2A-B. N-myc protein detection with ELISA in the 4 neuroblastoma samples.

Figure 5.3. Standard curve of purified N-myc protein with ELISA.

Figure 5.4. N-myc detection in cell-lysates and soluble lysates (supernatant) of the neuroblastoma cell-lines.

Figure 5.5A-D. N-myc protein detection in the neuroblastoma cell-lines with 4 different primary antibodies.

Figure 5.6A-H. Sandwich ELISA results with all the possible combinations of the anti-N-myc antibodies.

Figure 5.7. N-myc protein detection with western blot.

Figure 5.8. Fluorescent western blot of the neuroblastoma samples and the N-myc purified protein.

Figure 5.9. Fluorescent western blot of the neuroblastoma samples and the N-myc purified protein.

Figure 5.10A-B. Standard curves for N-myc protein.

Figure 5.11. Fluorescent western blot of the neuroblastoma samples and the purified N-myc protein.

Figure 5.12. Fluorescent western blot of the neuroblastoma samples and the purified N-myc protein.

Figure 5.13. Standard curve based on the purified N-myc protein readings from Figure 5.12.

Figure 5.14. ELISA test for MAX protein quantification.

Figure 5.15. Standard curve of purified MAX protein.
Figure 5.16. Western blot image for recognition of the protein MAX.  Page 153

Figure 5.17A-B. Purified MAX-protein standard curve.  Page 153

Figure 5.18. Fluorescent western blot for MAX protein detection.  Page 155

Figure 6.1. Suggested future directions based on the work presented for this study  Page 183
LIST OF TABLES

Table 1.1. The different stages of NB with their definitions according to the International Neuroblastoma Staging System INSS. Page 18

Table 1.2. Risk groups for NB patients according to biological and clinical features Page 28

Table 1.3. Main characteristics of top-down and bottom-up systems biology approaches. Page 54

Table 2.1. The antibodies and positive controls used for the quantification of the N-Myc and MAX proteins with the ELISA method. Page 69

Table 3.1. The rate-laws that describe the reactions of the model. Page 79

Table 3.2. The ordinary differential equations (ODE) describing the rate of change of the model species. Page 80

Table 3.3. The ODEs describing the rate of change of each species at steady state (ss) conditions. Page 80

Table 3.4. Simplified equations for describing the rate of change of each species at steady state conditions. Page 80

Table 3.5. Parameter values of the model obtained from the literature. The values were adjusted to fit with the chosen units of the model. Page 82

Table 3.6. The parameter sets used for the different potential amounts of [N-myc] and [MAX] at steady state conditions. Page 84

Table 3.7. The equations for the calculated coefficients from MCA. Page 95

Table 4.1. The Cq values obtained from all the cell-lines and genes used in this study. Page 105

Table 4.2. Numerical analysis of the obtained qPCR DNA data. Page 106

Table 4.3. The results of 2-sample Kolmongorov-Smirnov comparisons between all the HKGs. Page 107

Table 4.4. The calculated absolute MYCN GCN in the four neuroblastoma cell-lines. Page 109

Table 4.5. The calculated MYCN mRNA molecules per cell for the neuroblastoma cell-lines. Page 113

Table 4.6. The efficiencies of the Standard Curves for the genes MYCN, SDHA and HPRT1. Page 115

Table 4.7. Relative amounts of MYCN mRNA.
Table 5.1. The key steps for normal and sandwich ELISA using a biotin-conjugated secondary antibody.

Table 5.2. The combinations of anti-N-Myc primary antibodies used for the sandwich ELISA tests.

Table 5.3. Quantification of the protein bands from Figure 5.7A using the ImageJ software.

Table 5.4. The relative amounts of N-myc protein per cell-line.

Table 5.5. The values obtained with the calculation for the N-Myc protein concentration.

Table 5.6. The calculated concentration of N-myc protein per cell.

Table 5.7. Maximum concentration for MAX protein per cell-line.

Table 5.8. The calculated values for the MAX protein present in the 4 NB cell lines according to the ELISA tests in section 5.3.1.

Table A.1. The associated assay components for the relevant DNA qPCR assays.

Table A.2. The associated assay components for the relevant cDNA qPCR assays.
### ABREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALK</td>
<td>Anaplastic Lymphoma Kinase</td>
</tr>
<tr>
<td>BCA</td>
<td>Bicinchoninic Acid</td>
</tr>
<tr>
<td>BDNF</td>
<td>Brain-Derived Neurotrophic Factor</td>
</tr>
<tr>
<td>bHLH</td>
<td>Helix-Loop-Helix</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumine</td>
</tr>
<tr>
<td>CBP</td>
<td>CREP Binding Protein</td>
</tr>
<tr>
<td>CDK1</td>
<td>Cyclin dependant kinase 1</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CDR BP</td>
<td>CRD Binding Protein</td>
</tr>
<tr>
<td>Cq</td>
<td>Quantification cycle</td>
</tr>
<tr>
<td>CV</td>
<td>coefficient of variance</td>
</tr>
<tr>
<td>DDX1</td>
<td>DEAD box polypeptide 1</td>
</tr>
<tr>
<td>DM</td>
<td>Double Minutes</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl Sulphoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Desoxiribonucleic Acid</td>
</tr>
<tr>
<td>ECL</td>
<td>Enhanced chemiluminescence</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EFS</td>
<td>Event-Free Survival</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-Linked Immunosorbent Assay</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence in-situ Hybridisation</td>
</tr>
<tr>
<td>FITC</td>
<td>fluorescein</td>
</tr>
<tr>
<td>GCN</td>
<td>Gene Copy-Numbers</td>
</tr>
<tr>
<td>GSEA</td>
<td>Gene Set Enrichment Analysis</td>
</tr>
<tr>
<td>GSK3</td>
<td>Glycogen synthase kinase 3</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-Transferase</td>
</tr>
<tr>
<td>HATs</td>
<td>Histone Acetytransferases</td>
</tr>
<tr>
<td>HKGs</td>
<td>Housekeeping Genes</td>
</tr>
<tr>
<td>HSRs</td>
<td>Homogeneously Staining Regions</td>
</tr>
<tr>
<td>HVA</td>
<td>Homovanillic acid</td>
</tr>
<tr>
<td>IGF2BP1</td>
<td>IGF-II mRNA binding protein 1</td>
</tr>
<tr>
<td>INPC</td>
<td>International Neuroblastoma Pathology Classification System</td>
</tr>
<tr>
<td>INRG</td>
<td>International Neuroblastoma Risk Group</td>
</tr>
</tbody>
</table>
INSS: International Neuroblastoma Staging System
kBp: Kilo Base pairs
kDa: Kilo Daltons
LOH: Loss of Heterozygosity
Lys: Lysine
LZ: Leucine Zipper
MAX: myc-associated factor X
MB: Myc Homology Boxes
MBp: Mega Base pairs
MCA: Metabolic Control Analysis
MKI: Mitosis-karyorrhexis index
mRNA: Messenger Ribonucleic Acid
MW: Molecular Weight
MYCN: v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)
NAG: Neuroblastoma Amplified Gene
NB: Neuroblastoma
NGF: Nerve Growth Factor
NLS: Nuclear Localisation Signal
NRQ: Normalised Relative Quantities
NT3: Neurotrophin-3
ODE: Ordinary Differential Equations
PBS: Phosphate Buffer Saline
qPCR: quantitative Polymerase Chain Reaction
Rb: Retinoblastoma protein
RNA: Ribonucleic Acid
RPL11: Ribosomal protein L11
RT-qPCR: Reverse transcription quantitative polymerase chain reaction
S-HRP: Streptavidin- Horse Radish Peroxidase
S: Survival
SKP2: S-phase kinase-associated protein 2
SDS-PAGE: Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Ser: Serine
SHRP: Streptavidin-Horse Radish Peroxidase
SIRT1: Sirnuin 1
ss: Steady state
TD: Transactivation Domain
Thr: Threonine
TMB: 3,3’,5,5’-Tetramethylbenzidine
TO: Template Oligonucleotide
Trk: Tyrosine Receptor Kinase
TRRAP: Transactivation\Transformation Associated Protein
VMA: Vanillylmandelic acid
ABSTRACT

Neuroblastoma (NB) is one of the most common cancers in infancy and childhood, and possession of amplified MYCN gene sequence (gene locus 2p24) is related to aggressive disease and poor prognosis, with a clear distinction established regarding the survival of patients based on the gene copy-number of MYCN. However, the expression of MYCN has been reported to vary between patients of even the same NB subgroups and more importantly its significance in relation to NB prognosis is still not clearly established with various reports presenting contradicting results.

In this study, a bottom-up Systems Biology approach is suggested for studying both the significance of MYCN expression, but also the distribution of control in the related pathways that regulate this expression. An initial model was constructed describing the basic steps in MYCN expression and it was parameterised with values obtained from the literature. The results from the model simulations and analysis generated the hypothesis that the expression of MYCN cannot be used exclusively as a predictive factor without taking into account the relative levels of its dimerisation partner, the protein MAX. Additionally, it was predicted that the amplification of MYCN had a more pronounced relative effect at lower rather than at higher MYCN gene copy numbers.

In order to create separate models for 4 NB cell-lines, it was necessary to perform absolute measurements for MYCN at the DNA, mRNA and protein level, as well as for the MAX protein. The MYCN gene copy numbers were measured using the qPCR method, while a new data analysis method was suggested for performing absolute quantification with the use of house-keeping genes, appropriate statistical methods and no reference samples. The relative amounts of MYCN mRNA were also measured using qPCR and the results obtained were in agreement with the suggested levels from the literature.

The absolute measurement of the N-Myc and MAX proteins was attempted using two complementary methods, western blots and ELISA. A series of optimisation experiments and data analysis steps were taken that resulted in the refinement of the experimental conditions to the point where they can be used for successful quantification of the absolute levels of the N-Myc protein. Alternatively, the procedure used for the MAX protein proved problematic and was not as successful.

Overall, this study was successful in becoming the first step for an expanded bottom-up systems biology study regarding the significance of MYCN expression in NB. The combination of both the modelling and experimental parts of this work illustrated some of the potential benefits of Systems Biology approaches in studying disease. In this case the resulting model, once fully parameterised with experimental data, can be expanded in a number of suggested ways that address questions like the role and control of MYCN expression in relation to the cell-cycle deregulation or multi-drug resistance in NB, giving in the process a better understanding regarding suitable treatment targets for individual NB cases.
Declaration of Authorship

I confirm that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright Statement

- The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright including for administrative purposes.

- Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

- The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

- Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy, in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations and in The University’s policy on presentation of Theses.
Acknowledgements

There are many people that I would like to acknowledge for their help and contribution in this study, perhaps too many to mention individually. The members of the individual Philip Day and Hans Westerhoff groups, including the PIs that were my supervisors, have contributed as much as anyone. I would also like to specially mention and express my gratitude for the help I received by David Brough and his group, as well as members of stuff from the campus like Sheona Drummond, Kieran Smallbone and Nadia Luheishi. Additionally, the whole of DTC Integrated Systems Biology, especially Lynne Davies, Mara Nardelli and Gerold Baier, as well as all the fellow co-phd students for supporting each other through this experience. Finally, I wish to acknowledge all my friends and family that I have completely ignored while trying to finish this phd. I wish only the best to everyone and I will thank you all in person as soon as i get the chance.
Chapter 1

INTRODUCTION

Neuroblastoma (NB) is a complex and multifactorial disease that accounts for the highest percentage of extra-cranial solid tumours in children and also 7-10% from all childhood cancers in general (Maris, Hogarty et al. 2007). More than 40% of the children diagnosed as having NB belong to the high-risk group that has 50% mortality in a five-year period (Cohn, Pearson et al. 2009; Matthy, Reynolds et al. 2009), while in total it accounts for 15% of childhood deaths in the U.S. (Ishola and Chung 2007).

NB usually develops during embryogenesis. Neuroblasts are derived from neural crest cells, the primordial cells of the sympathetic nervous system, with which they share similarities (Maris, Hogarty et al. 2007). The neural crest cells normally reach their mature form as defined by their location in the human body before birth, while in NB they form self-generating, undifferentiated cells. The most common location for these cells is the adrenal medulla due to their linage, but they can also be found on the sympathetic ganglia chain in the region between the neck and the pelvis (Hiyama, Yokoyama et al. 2000; Ishola and Chung 2007).
1.1: Classification of Neuroblastoma

NB has more than one stage with clinical and biological manifestations that vary between each stage. A number of staging systems are used around the world that employ different characteristics and biomarkers of the disease for stratification, like the International Society of Pediatric Oncology Europe Neuroblastoma Group (Cecchetto, Mosseri et al. 2005) and Children’s Oncology Group, with the most commonly used classification system being the International Neuroblastoma Staging System (INSS) with the different stages shown in Table 1.

Table 1:1: The different stages of NB with their definitions according to the International Neuroblastoma Staging System INSS. Adapted from (Ishola and Chung 2007)

<table>
<thead>
<tr>
<th>Stage</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Localised tumour, grossly complete resection, with/out microscopic residual disease, negative ipsilateral lymph nodes</td>
</tr>
<tr>
<td>2A</td>
<td>Localised tumour, grossly incomplete resection, positive ipsilateral non-adherent lymph nodes, negative contralateral lymph nodes</td>
</tr>
<tr>
<td>2B</td>
<td>Localised tumour with/out grossly complete resection, positive ipsilateral non-adherent lymph nodes, negative contralateral lymph nodes</td>
</tr>
<tr>
<td>3</td>
<td>Unresectable unilateral tumour infiltrating across the midline with/out regional lymph involvement OR midline tumour with contralateral regional lymph node involvement OR midline tumour with bilateral extension by infiltration (unresectable) or by lymph node involvement</td>
</tr>
<tr>
<td>4</td>
<td>Any primary tumour with dissemination to distant lymph nodes, bone, bone marrow, liver, skin or other organs</td>
</tr>
<tr>
<td>4S</td>
<td>Localised primary tumour (as defined for stages 1, 2A and 2B) with dissemination limited to skin, liver and bone marrow (limited to infants less than 1 year age)</td>
</tr>
</tbody>
</table>

More recently, Cohn and colleagues performed a large-scale study where they attempted to combine the markers and features used by various staging systems and analyse their statistical and clinical significance in a group of 8,800 patients from different regions in the world. Their results were used to produce the International
Neuroblastoma Risk Group (INRG) system that could provide a universally accepted and used platform in terms of pre-treatment risk stratification of NB. Based on the results of the study, the chosen markers were the age of the patient, the histologic category and grade of differentiation of the tumor, MYCN amplification status, aberration at chromosomal location 11q and the ploidy status of the tumour (Cohn et al., 2009).

1.2: Biological features and markers of NB

One reason for the existence of numerous NB staging systems is that NB has a number of biomarkers and associated biological and genetic abnormalities whose significance is not always universally established and accepted. These markers can be generally grouped into serum and urine markers, patho-biological manifestations and cytogenetics’ abnormalities, while several molecular markers can also help in providing a clearer image about the stage and prognosis of NB patients.

1.2.1: NB markers present in serum or urine

From the biomarkers present in serum and/or urine, the levels and the ratio of the catecholamine metabolites vanillylmandelic acid (VMA) with homovanillic acid (HVA) is potentially the most reliable. Their presence is closely related to the neural origins of NB cells, while their ratio (VMA/HVA) is indicative of the presence and severity of NB, especially if it is below the value of one. A mass-screening program in Japan tried to establish a closer association between the VMA/HVA ratio and the presence/severity of NB (Sawada, Hirayama et al. 1984). The results were deemed
originally to be a success, but further studies have revealed the tendency for oversensitivity of these screens that can lead to identification of false positives. Nonetheless, this is still widely used in Japan, while in general establishing this ratio is also considered a cautious mean for verification of the presence of NB as indicated by other signs and/or its severity (Tanaka, Iehara et al. 2005; Iehara, Hosoi et al. 2006). Additional markers that belong in this category include the enzyme lactate dehydrogenase (Otsu, Hirata et al. 1985), the iron-storing protein ferritin (Hann, Stahlhut et al. 1985) and the glycolytic enzyme neuron specific enolase (Odelstad, Pahlman et al. 1981).

1.2.2: NB pathological markers

In terms of pathological manifestations, one of the best established and characterised systems that examine the differentiation and alteration-patterns of NB cells was established originally by Shimada and colleagues in 1984 (Shimada, Chatten et al. 1984) and was later refined in its present form (Shimada, Ambros et al. 1999). The whole basis of characterisation is the altered morphology of the NB tumours as seen through histologic examination. The original approach incorporated the Schwannian stroma development and the “mitosis-karyorrhexis index” (MKI) into the age-linked classification of neuroplastic tumors (neuroblastoma, ganglioneuroblastoma, ganglioneuroma) and their prognosis. Further studies have suggested the incorporation of mitotic rate and tumour-associated calcification into this system in order to enhance its prognostic accuracy (Joshi, Cantor et al. 1992; Joshi, Cantor et al. 1992) that eventually resulted in the standards suggested by the International
Neuroblastoma Pathology Classification System (INPC) (Shimada, Ambros et al. 1999).

1.2.3: NB cytogenetic markers

The cytogenetic alterations of NB include numerous chromosomal abnormalities. Despite the universal agreement on the existence of most of these, there are varying opinions regarding their severity and the weight carried by some of them in terms of prognosis and in relation to other genetic changes of NB (Brodeur 2003).

A common abnormality in many NB cases is the loss of heterozygosity (LOH) in a number of chromosomes. More specifically, chromosome 11q showed allelic loss in 43% of the examined cases (Guo, White et al. 1999), which is the highest percentage for NB, and this event was linked to worse outcome (Attiyeh et al., 2005). Furthermore, it was also associated with 14q LOH (or more specifically 14q23-32) that was shown to be present in 23% of the examined cases (Thompson, Seifried et al. 2001).

Both of these events were found to be inversely linked to LOH of 1p (an event closely related to amplification of the oncogene MYCN) while the predictive significance of some of these events, both in terms of disease development and mortality has been limited. No specific association has been found in terms of prognosis for 14q LOH, even though it is detected in a number of NB cases of
different stage and as such implying that important gene-suppression genes are located at this region (Takayama et al., 1992; Thompson et al., 2001).

On the other hand, Chromosome 1 is a location of high activity in NB. Firstly, 1p36 is usually deleted in cases of sporadic NB, leading to the hypothesis that potentially tumour-suppressor genes are clustered in the specific location and this deletion could be necessary for predisposition to NB. The notion of predisposition was rejected, though, by the discovery that 1p36 is not linked to familial NB (Maris, Kyemba et al. 1996) while the exact location that is deleted has been mapped to 1p36.3 with an estimated frequency of 25% in NB patients and 87% in NB cell-lines (White et al., 2005). The 1p36 deletion has been linked to worse outcome for NB patients (Attiyeh et al., 2005) or at least reduced chance for Event Free Survival (EFS) but with no significance for Overall Survival (OS) (Maris et al., 2000). Furthermore a broader examination of the genetic-linkage to predisposition in NB on a genome-scale has identified 16p12-13 as a more probable candidate (Maris, Weiss et al. 2002) while LOH at the specific location has been linked with favourable prognosis (Furuta et al., 2000). Additionally, as already mentioned, deletion of the short arm of chromosome 1 (or arm 1p) is closely associated with amplification of oncogene MYCN. The percentage of occurrence of this event greatly varies according to different studies, however it is agreed that it is commonly found in advance cases of NB (Brodeur 2003).

However, LOH or deletions are not the only karyotypic abnormalities associated with NB. On 17q there is frequent observation of gain of genetic material that has
been reported to have a frequency among NB-patients ranging from 66.7% (Bown et al., 2001) to 86.8% (Brinkschmidt et al., 2001), while it was closely linked to adverse outcome for NB patients (Bown et al., 2001). It is considered the most widespread genetic abnormality in NB and even though it can occur independently, in many cases it is linked to unbalanced transfer of genetic material from 1p to 17q (Van Roy, Laureys et al. 1997). Finally, the general ploidy status of a patient is also considered to be a useful prognostic factor for children under the age of 1 year. Near diploid (or hyperdiploid) status is usually associated with poor prognosis while almost-triploid status with good prognosis (Look, Hayes et al. 1984; Look, Hayes et al. 1991).

1.2.4: Molecular markers associated with NB

A number of molecular markers are associated with NB and its different stages and are used as prognostic factors. Sometimes even the same gene can be the subject of a number of deregulations and/or mutations that are associated with different clinical outcomes. General models have been proposed trying to provide a comprehensive set of features (pathological, cytogenetic and oncogenic) mapping the disease through it’s different stages and the relevant prognosis, however new discoveries in particular in this section will probably provide more variables that could be used for forming more precise or different models.

The Tyrosine Receptor Kinases (Trk) A/B/C are transmembrane receptors with an established link with NB prognosis. Each one has a different ligand and subsequently
involved with different pathways functions: the nerve growth factor (NGF), the brain-derived neurotrophic factor (BDNF) and neurotrophin-3 (NT3) for TrkA, B and C respectively (Patapoutian and Reichardt 2001). Their expression has been linked to different stages of NB and also MYCN expression. For TrkA it was found that it is highly expressed in lower stages of NB with higher survival rate and no MYCN amplification (Kogner, Barbany et al. 1993; de Souza, Sanabani et al. 2011). Alternatively, expression of TrkB in its full form was found to correlate significantly with stages also exhibiting MYCN amplification (Nakagawara, Arima-Nakagawara et al. 1993). Finally, TrkC seems to be sometimes co-expressed with TrkA, but never with TrkB, without any significant differentiation found between cases with both TrkA and TrkC or only TrkA (Ryden, Sehgal et al. 1996).

The glycoprotein CD44 is found on the cell surface and it is involved in cell-cell interactions and adhesion. A high correlation has been demonstrated between lack of expression of CD44 and NB types with poor prognosis (Munchar, Sharifah et al. 2003), while it has been suggested that high expression is found in more differentiated tumours with positive outcome (Tang, Robinson et al. 2004, Combaret, Gross et al. 1996).

1.3: The MYCN gene and NB

1.3.1: MYCN gene amplification and NB

Perhaps the most important oncogene and general biomarker of NB is MYCN. The correlation of MYCN amplification with advance stages of NB was firstly reported
in 1984 (Brodeur, Seeger et al. 1984) while its significance as a prognostic factor was first examined a year later (Seeger, Brodeur et al. 1985). Based on the results of the study by Seeger and colleagues, it was suggested that patients having more than ten MYCN gene copy-numbers (GCN) per haplotype have poor prognosis while having less than three MYCN GCN per haplotype had good prognosis. Patients having between 4-9 MYCN GCN were mostly associated with good prognosis, but not to the levels of 1-3 MYCN GCN (Seeger, Brodeur et al. 1985). Similar results were also obtained in a study where patients with 3-9 copies were found to have a high rate of survival (89%) while for higher copy numbers the survival rate dropped significantly (Iehara, Hosoi et al. 2006).

Figure 1.1: The Event-Free Survival (EFS) of patients in relation to their MYCN GCN. In this study, patients with 3-9 MYCN GCN had an 89% EFS probability, while for patients with more than 10 MYCN GCN there was an almost 40% drop, indicated by the red arrow. Adapted from (Iehara, Hosoi et al. 2006)
The greatest strength of using MYCN amplification as a prognostic factor in NB lies with the fact that its importance transcends the patients’ age and stage, as well as the presence of most of the other biological features associated with NB. For example, in the study by Seeger and colleagues the examination consisted of samples from patients belonging to all stages of NB with some of the patients from Stages 2, 3 and 4 having MYCN amplification. From the MYCN-amplified subset, all the patients had significantly worse progression-free survival regardless of NB stage. Additionally, for stage 2 tumours there was a higher statistical correlation with metastasis in comparison to stage 2 non-MYCN amplified patients while for stage 4 patients the tumour progression was significantly faster in the presence of MYCN amplification (Seeger et al., 1985). Similar results were obtained from a study that included a greater number of samples while also examining the significance of MYCN amplification in relation to the age of patients. The results showed that other than stage 4S patients, the presence of MYCN amplification is a factor for unfavourable prognosis, even more so for patients with otherwise favourable NB-related characteristics like being less than one year old or generally having a favourable NB stage like 1 or 2 (Tonini et al., 1997). In that respect, Iehara and colleagues also provided additional evidence suggesting that the presence of MYCN amplification is a strong factor for poor prognosis in infants (Iehara et al., 2006).

Furthermore, the amplification of MYCN has been shown to be either mutually exclusive with favourable NB molecular abnormalities like expression of TrkA or its presence having a higher significance than the presence of other favourable characteristics like the expression of CD44. Performing a multivariate analysis in regard to these markers and the NB stage of the patients, the NB stage and MYCN
amplification were found to have the highest significance in terms of survival (Kramer et al., 1997). Similar results were demonstrated from multivariate analysis of additional NB markers like lactate dehydrogenase (LDH) or Neuron Specific Enolase (NSE) in relation to their prognostic significance, with MYCN amplification exhibiting the highest relevance in terms of poor prognosis (Rubie et al., 1997).

Finally, the importance of DNA ploidy status in relation to MYCN amplification in terms of prognosis has also been the subject of a number of studies, with the original suggestion being that there was a potential correlation between MYCN amplification and diploid NB cells in terms of disease aggression (Cohn et al., 1990). Since then, various studies have provided similar observations. For example, Bagatell and colleagues both verified the significance of MYCN amplification and also suggested that the ploidy status of NB tumour cells is of high importance in terms of prognosis since patients with diploid tumours showed significantly worse outcome than patients with hyperdiploid tumours (Bagatell et al., 2009). Similar results were also obtained by Schneiderman and colleagues, where both MYCN amplification and ploidy status exhibited a significant influence on NB prognosis, where patients with diploid and hypodiploid NB tumours had lower Event-Free Survival (EFS) and Overall Survival (OS) that was statistically significant (Schneiderman et al., 2008).
Table 1.2: Risk groups for NB patients according to biological and clinical features. In most cases, MYCN amplification is a strong enough feature to provide a high-risk prognosis of the patient, regardless of other biological and clinical features or even the stage of the disease. The stage of NB is based on the International Neuroblastoma Staging System (INSS), “amp” is for amplified, DI is for DNA index and ploidy status, FH for favourable histology and UH for unfavourable histology. Adapted from Brodeur 2003.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Low Risk</th>
<th>Intermediate Risk</th>
<th>High Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>2A, 2B</td>
<td>Age &lt;1 year or age 1-21 years and MYCN non-amp or age 1-21 years and MYCN amp + FH</td>
<td>None</td>
<td>Age 1-21 years and MYCN amp + UH</td>
</tr>
<tr>
<td>3</td>
<td>None</td>
<td>Age &lt;1 and MYCN non-amp or age 1-21 years and MYCN non-amp + FH</td>
<td>Age 0-21 years and MYCN amp or age 1-21 years and MYCN non-amp + UH</td>
</tr>
<tr>
<td>4</td>
<td>None</td>
<td>Age &lt;1 and MYCN non-amp</td>
<td>Age &lt;1 and MYCN amp or age 1-21 years</td>
</tr>
<tr>
<td>4S</td>
<td>MYCN non-amp, FH, DI&gt;1</td>
<td>MYCN non-amp, UH, DI=1</td>
<td>MYCN amp</td>
</tr>
</tbody>
</table>

1.3.2: Association of MYCN amplification with NB-related chromosomal abnormalities and genes located at 2p24

With MYCN amplification having a higher significance than most other factors associated with NB, several studies have tried to identify the precise connections between these biomarkers. To begin with, it has been demonstrated the strong association between gain of material at 17q with deletion at 1p (P<0.001) and amplification of MYCN (P<0.001) (Bown, Cotterill et al. 1999). This amplification also correlates with and maps to the presence of double minutes (DM) and homogeneously staining regions (HSRs) in NB (Solovei, Kienle et al. 2000), two changes that potentially enhance MYCN transcription and activity. The amplicon
that contains MYCN has been found to vary in size going up to 1MB, however a
conserved 130kb sequence that was found to be present in most cases has been
proposed as the potential location of the MYCN gene (Reiter and Brodeur 1996).

The chromosomal location of MYCN, 2p24 is potentially a hub of neuroblastoma-
related genes, as well. With the amplicon containing the MYCN gene varying in size
(up to 1MBp) with the MYCN gene typically taking around 10kBp (Bown 2001), a
number of genes have been found on a regular basis to be co-amplified with MYCN
in NB. One of these genes is the DEAD box polypeptide 1 (DDX1) of the DEAD
(Asp-Glu-Ala-Asp) box protein family that is involved in a number of processes.
While its co-amplification with MYCN in a number of patients is not disputed, there
is disagreement about its prognostic significance. Some studies have suggested that
when co-amplified with MYCN in advance-stage NBs it is a marker for favourable
prognosis (Weber et al., 2004) while others have either contradicted these findings
and suggested that DDX1 amplification is of no prognostic significance, with or
without MYCN co-amplification (De Preter et al., 2005). Furthermore the precise
role of its amplification and relevant significance in terms of prognosis for NB is
unclear (Kaneko, Ohira et al. 2007), as is the significance of its expression (de
Souza, Sanabani et al. 2011).

Another co-amplified gene is the Neuroblastoma Amplified Gene (NAG) that was
firstly identified by Wimmer and colleagues and was mapped to be directly after the
DDX1 gene, both in a telomeric position to the MYCN gene (Wimmer, Zhu et al.
1999). The exact role of the gene and its protein is not clearly established, while
there are also contradictory studies regarding its relevance in prognosis for NB and any exact link to the level of MYCN amplification (Scott, Board et al. 2003).

Finally, Anaplastic Lymphoma Kinase (ALK) is another gene linked to NB that is mapped to the 2p24 amplicon. ALK is a tyrosine receptor that it is found both as a full length or a fusion protein and it is primarily involved in neural development (Allouche, 2007). Originally it was suggested that its impact on NB is minimal, but a study by Mossé and colleagues has found a significant link between ALK alterations, both in terms of mutations and amplification, and familial NB predisposition (Mosse et al., 2008). Of potentially great interest is also the suggestion that ALK is amplified independently to MYCN rather than as a “domino effect” of MYCN deregulation (Caren et al., 2008).

1.3.3: Expression of MYCN and its significance for NB prognosis- studies based on NB cell-lines.

Even though the amplification of MYCN at the DNA level is considered to be one of the most reliable prognostic biomarkers for NB stratification, the expression of MYCN and the significance of its RNA and protein levels for NB are still not clear. Studies based on established NB cell-lines have showed that there is a correlation between over-expression of MYCN and enhanced in-vitro growth potential of cells (Gross, Miescher et al. 1994). This was true regardless of MYCN amplification at the genetic level, even though the cell-lines with amplification had both higher levels of N-Myc expression and higher growth rate of the cells (Wada, Seeger et al. 1993).
Furthermore, down-regulation of MYCN expression using MYCN antisense constructs caused decreased proliferation and growth of NB cell-lines (Schmidt, Salwen et al. 1994).

1.3.4: Expression of MYCN and its significance for NB prognosis- studies based on NB clinical samples.

Alternatively, studies based on clinical samples have not produced a universally-accepted conclusion on the subject. In one of the first studies on the subject, Hiyama and colleagues investigated the correlation between MYCN expression and patient survival using 41 primary samples. The correlation between poor prognosis and MYCN amplification and over-expression was statistically significant, while the detection of the N-Myc protein in NB patients was described by the authors as “one of the most unfavourable prognostic factors in neuroblastoma” (Hiyama, Hiyama et al. 1991).

Similar results were found in a study where the analysis included samples from NB patients classified at different stages of the disease (Chan, Gallie et al. 1997). The expression of the N-Myc protein was found to be independent of amplification at the genetic level, with the samples presenting heterogeneity of expression even within the same NB stage. Additionally, samples without MYCN amplification at the genetic level showed over-expression of the N-Myc protein, but without reaching the levels of expression the samples with MYCN amplification had, thus being in agreement with the observations on established cell-lines (Wada, Seeger et al. 1993).
Their results also showed that there was a statistically significant correlation between the N-Myc protein expression and the survival of the patients, suggesting that the N-Myc protein expression was not only a strong prognostic factor for the outcome of the patient, but also that it was independent of the other stratification factors like age, stage and MYCN gene status. Alaminos and colleagues, while investigating 99 clinical samples and 12 NB cell-lines, reached the same conclusion regarding the significance of MYCN expression where they reported identical patient survival with MYCN over-expression, regardless of the levels of MYCN gene copy-numbers (Alaminos, Gerald et al. 2005). These results were seen as supportive of the outcome of a previous study by the same group where the expression pattern of several genes showed better correlation among samples showing MYCN over-expression than ones that didn’t (Alaminos, Mora et al. 2003).

Other studies using clinical samples contradicted these conclusions, though. In a study by Slavc and colleagues, samples from 45 untreated NB patients were examined for MYCN amplification and 23 of them for MYCN expression (Slavc, Ellenbogen et al. 1990). Their results showed that in the absence of MYCN amplification, the expression of MYCN could not be established as a reliable prognostic marker. Additionally, in a study by Cohn and colleagues, it was theorised that the disparity between the conclusions that could be drawn regarding the prognostic significance of MYCN expression was at least partly related to the lack of homogeneity between the different studies regarding the characteristics and NB stage of the samples. For this reason, samples from 69 patients at stage 3 or 4 with no MYCN amplification were used for analysis of MYCN expression and correlation with the 5-year survival of the patients. The expression of MYCN was
heterogeneous even between samples from the same stage of NB, with a clear
distinction between patients that were less than 1 year old and patients that were
older than 1 year. Additionally, for probably the first time, over-expression of
MYCN in patients with no MYCN amplification was correlated to increased
survivability, even though this observation was not statistically significant (Cohn,
London et al. 2000). A study by de Souza and colleagues used samples from 64 NB
patients belonging to different risk groups for examining the mRNA expression of
different NB biomarkers, including MYCN. The results showed that no correlation
could be found between MYCN or DDX expression and survival outcome for this
heterogeneous group of patients (de Souza, Sanabani et al. 2011).

1.3.5: Conclusions from studies of MYCN expression in NB

So far the message given regarding the expression of the MYCN RNA and the
N-Myc protein in NB and its prognostic significance is not clear. The reliability of
using the expression of the MYCN gene at the RNA and protein level as a prognostic
factor varies from study to study, while the significance of the MYCN expression
as a positive or negative indicator to a patient’s prognosis is also a matter of
dispute. Even though the methodology used in terms of quantifying the MYCN RNA
levels has gradually changed from northern blots to the more accurate real-time
quantitative PCR (RT-qPCR), the predominant method used for the quantification of
the N-Myc protein levels was the SDS-PAGE gels and western blot. The protein
results were interpreted in a qualitative manner with no attempt for absolute
quantification, even without acknowledgement of any limitations related to
densitometry methods (Gassmann et al., 2009). Finally, in these studies biological
replicates were included in the form of samples from patients of the same NB stage or risk group, but no technical replicates were reported. Overall, this makes the reported N-Myc protein results potentially untrustworthy in terms of accurately establishing the role of MYCN protein expression in NB.

However, there are certain qualitative characteristics that appear in multiple studies and have not been contradicted so far. For example, the expression of MYCN is heterogeneous between NB cell-lines and patients even if they belong in the same risk group or stage. Additionally, the expression of MYCN in samples with MYCN amplification is higher than samples with MYCN over-expression but no amplification. From this information it becomes clear that either different regulatory processes are involved in the expression and regulation of MYCN between patients, or the same processes are not acting to the same extend between different NB samples, due to potentially additional oncogenic deregulations. In that respect any of the multiple pathways that have been implicated in regulating MYCN at different levels from DNA to protein activity could be at least partly responsible for the observed phenomena already described.

1.3.6: Difference of prognostic significance in NB between MYCN amplification and expression

Alternatively, a number of potential theories might be able to explain these discrepancies between the results of the studies trying to understand the significance of MYCN expression in NB. Firstly, it could be argued that the MYCN amplification
at the genetic level is simply a quantitative measure of upstream deregulations that are responsible for the relevant neuroblastoma severity. In such a case the effect of the expression of the N-Myc protein would be of partly irrelevant consequence and would account for the variations in its suggested significance. It is already known for example that there are two different paths for MYCN amplification resulting in either Double Minutes (DMs) or Homogeneously Staining Regions (HSRs) and events associated to these pathways could carry more significance than what is already recognised (Solovei et al., 2000)

Furthermore, most of the MYCN-expression studies have provided qualitative results in terms of the N-Myc protein levels and tried to associate this expression with the MYCN gene copy numbers, stage and/or survival of the NB patients. In that respect, the non-quantitative nature of results and the usually binary association of MYCN expression to any of these factors means that potentially important information that would provide a clearer image, in the way of network interactions or relevant system properties, might have been missed. It could even be argued that studying NB markers and their significance in a way that emphasises simple ratios or threshold points of these markers, rather than incorporating a holistic approach in studying the relevant systems, is one of the reasons that despite the wealth of information about NB in general, the understanding of the underlying mechanisms is comparatively lacking. Finally, this lack of clarity and agreement regarding the significance of MYCN expression in NB patients could also be due to a combination of any of the potential factors already mentioned.
1.4: Structure, activity and control of MYCN in normal cells: potential contributors towards heterogeneous MYCN transcript, protein and activity levels.

MYCN belongs to the MYC family of transcription factors and to exert its biological activity it needs to heterodimerise with Max and employ additional co-transcriptional factors, while its targets are involved in a number of important cellular functions. Despite the fact that MYCN amplification is one of the most important adverse prognostic factors in NB, the details of the mechanisms involved in the expression of MYCN are not fully understood, at least in the context of NB. However, the available information regarding many of the interactions and control mechanisms of the MYC gene-products for normal cells is substantial (Vita and Henriksson 2006).

1.4.1: Role of MYCN and the MYC family in normal cells

The MYC family members of transcription factors include c-myc and l-myc along with MYCN (Henriksson, Lüscher et al. 1996). They are responsible for the regulation of up to 10% of the genome (Dang, O'Donnell et al. 2006) and their transcriptional targets have a wide range of roles in the mammalian cell that include apoptosis, cell cycle proliferation to cell differentiation and adhesion (Figure 1.2). Additionally, the expression of the MYC family members is potentially regulated by mechanisms that prevent the co-expression of more than one member at significant levels. For example, it has reported that c-myc was present at higher levels when co-expressed with MYCN. However when MYCN was over-expressed, either in
MYCN amplified, MYCN non-amplified or MYCN transfected NB cell-lines, the expression of c-myc dropped significantly (Breit and Schwab 1989).

Primarily, the MYC proteins promote the transcription of their target genes, while on occasion they have been found to be involved in opposing processes or pathways (Vita and Henriksson 2006). Specifically, one of the main results of MYCN action in a normal cell is to help the cell move from the G1 to the S phase in the cell cycle, during which time its expression reaches its peak while during the rest of the cell cycle it returns to low levels (Kelly, Cochran et al. 1983). The role of assisting cell growth and differentiation has been tested in various mouse models. For example, in transgenic mice overexpressing c-myc in pancreatic beta cells in vivo, the observations included increased cell proliferation and hyperplasia (Laybutt et al., 2002). On the other hand, it has also been demonstrated that over-expression of MYC proteins in normal beta cells promotes apoptosis, a finding that is shared in other studies (Evan et al., 1992; Cano et al., 2008), while this effect can be countered by overexpression of anti-apoptotic factors like Bcl2 (Aubry and Charron, 2000) and Bcl-(x)L (Pelengaris et al., 2002). These results lead to suggestions that the potentially carcinogenic effect of MYC over-expression is combined with either additional deregulations that prevent the proper function of safety mechanisms against MYC over-expression or the presence of survival/anti-apoptotic factors (Hurlin and Dezfouli, 2004).
Figure 1.2: Effects of MYC transcriptional activity. With the MYC family members regulating up to 10% of the genome, their targets are involved in many important processes like cell-cycle progression, cell growth and division and metabolism. Alternatively, in cases with MYC deregulation there is also a broad range of effects that include genomic instability, immortalisation and independence from growth factors. Adapted from (Vita and Henriksson 2006).

1.4.2: Structural features of the MYC family proteins

The MYCN gene is located at chromosomal location 2p24 and the protein encoded has two variants, one 456 aa (58kDa) long and another at 464aa (64kDa). The MYCN protein shares the same general structure with all myc family members as shown in Figure 1.3 (Lu, Pearson et al. 2003). Their major features are the basic Helix-Loop-Helix (bHLH) motif, the Leucine Zipper (LZ) domain, the nuclear localisation signal (NLS), the DNA binding site and the MYC Homology Boxes (MB) 1-4 of which 1 and 2 are located in the Transactivation Domain (TAD) while 3 and 4 are located in the space between TAD and NLS (Wenzel and Schwab 1995; Cowling and Cole 2006; Liu and Levens 2006).
Figure 1.3: Basic structural features of MYC proteins and their binding partner MAX. Shown here are the Transactivation Domain (TAD), the Nuclear Localisation Signal (NLS), the basic Helix-Loop-Helix (B-HLH) motif and the Leucine Zipper (Zip). Adapted from (Lu, Pearson et al. 2003)

All these structural features play important roles in the way the MYC family proteins interact with other molecular entities and exert their transcriptional activity. The MYC Homology Boxes act primarily as platforms for the recruitment of various co-factors assisting the transcriptional activity of the MYC proteins by targeting specific gene promoters. Different co-factors are employed at the MBs depending on the cellular environment while each MB box is responsible for the transactivation of specific categories of functions within the cell. For example MB1 and MB2 are involved in inducing apoptosis and also blocking cell differentiation, while some of the notable co-factors associated with them are H-Ras (Freytag, Dang et al. 1990; Evan, Wyllie et al. 1992), the Transactivation\Transformation Associated Protein TRRAP (McMahon, Van Buskirk et al. 1998) and the set of p300/CBP (Vervoorts, Luscher-Firzlaff et al. 2003; Faiola, Liu et al. 2005). MBIII on the other hand recruits co-factors that target genes vital for cell transformation (Herbst, Hemann et al. 2005) while MB4 targets assist cells in progressing through the G2 phase of the cell-cycle (Cowling, Chandriani et al. 2006). Finally, the N-Myc protein (and MYC proteins in general) exert their transcriptional activity via the mechanisms of
chromatin acetylation and remodelling where the binding of RNA polymerase and subsequently the initiation of transcription is promoted (Adhikary and Eilers 2005; Cowling and Cole 2006).

1.4.3: The myc-associated factor x (MAX)

The bHLH and the LZ are the structural characteristics that allow the heterodimerisation of the MYC proteins with the myc-associated factor x (MAX) (Wenzel et al., 1991). MAX is a protein with 160aa length and it is required for the transcriptional activity of MYCN while a disruption of the regular structure of these motifs prevents the dimerization between the MYC proteins and MAX (Kato et al., 1992). The MAX gene is located at chromosomal location 14q23 while its RNA has a half-time of approximately three hours, making it more stable than the corresponding myc RNA (Wagner et al., 1992). Additional structural features of importance of the MAX protein include a nuclear localisation signal located at the c-terminus of the polypeptide (Kato et al., 1992) and two major phosphorylation sides at the N-terminus (Ser-2 and Ser-11) that are targeted by Casein Kinase II and affect the MAX protein activity (Bousset et al., 1993; Koskinen et al., 1994).

In the heterodimer form, the DNA-binding domain targets and binds to the E-Box that has the sequence 5’-CACGTG -3’ (Wenzel and Schwab 1995). Max can also form dimers with other proteins that belong to the MAD family of proteins as part of the MYC/MAX/MAD network or even form homodimers. These dimers have transcriptional repression activity, in contrast to the MYC-MAX dimers (Hurlin,
Queva et al. 1995; Hooker and Hurlin 2006). A study by Blackwood and colleagues has found that the expression of MAX protein in non-NB cell-lines is stable during the cell-cycle and at lower levels than the N-Myc protein (Blackwood et al., 1991). These results were reproduced in a qualitative manner in terms of MYCN expression in NB cell-lines, even though the provided data was not of the highest quality since they partly consisted of badly copied western blot images (Raschella et al., 1994). The NB cell-lines used in this study were either with or without MYCN amplification and the results were reported to be similar. Based on these results, the suggestion was made that the activity of the MYC/MAX dimers is dependent on MYC rather than MAX. An observation related to this suggestion is that without knowing the exact kinetic parameters of the relevant interactions between the MYC and MAX proteins, as well as any associated co-factors and DNA, it is difficult to ignore the possibility that at higher expression levels of the MYC proteins (and especially N-Myc in NB) the MAX protein could be depleted. In such a case the MAX protein levels present a threshold of maximum potential activity for the MYC/MAX dimers while it has to be emphasised that this assumption considers every other related factor as non-limiting which may not be the case either.

1.4.4: Controlling MYCN and MYC family proteins at the genetic level in normal cells

With the MYC proteins being involved in many important cell functions, their deregulation in any way can have a major impact on cellular behaviour and dynamics. As such, the activity of MYC proteins is tightly controlled at different levels in a number of ways that include regulation of transcription of the MYC gene,
mRNA stability, different post-translational modification mechanisms, MYC protein stability, competition for dimerisation with MAX and negative or positive feedback loops.

Starting at the genetic level, the MYC gene promoters are subject to activation from a number of pathways in the cell that include the E2F family of transcription factors (Thalmeier, Synovzik et al. 1989) and NF-κB protein complex (Duyao, Buckler et al. 1990) among others. With many of these of pathways being involved in different forms of cancer, it is not surprising that MYC expression is also deregulated under these conditions (Ponzielli, Katz et al. 2005; Liu and Levens 2006). Under normal conditions it is imperative that mitogenic stimulation occurs prior to transcriptional activation of the MYC gene. While the transcription levels of the MYC gene are substantially elevated during G1, they drop back to normal by the time the cell has entered the S phase (Kelly, Cochran et al. 1983; Liu and Levens 2006).

The mRNA of MYC family members was originally thought to be either rapidly translated or degraded. However, a coding region instability determinant (CRD) sequence was found to be important in stabilising c-myc mRNA during translation arrest. One way this occurs is via the interaction with the CRD Binding Protein (CRD BP) that caps CRD in the absence of ribosomes in order to prevent mRNA degradation by endonuclease targeting, while ribosomal binding at the mRNA is strong enough to displace CRD BP and continue with translation (Lemm and Ross 2002; Sparanese and Lee 2007). The importance of CRD was reinforced when (Weidensdorfer, Stohr et al. 2009) also found that a complex based on IGF-II mRNA
binding protein 1 (IGF2BP1) binds at the CRD and prevents degradation of c-myc mRNA.

1.4.5: Controlling MYCN and MYC family proteins at the protein level in normal cells

The numerous post-translational modifications of the MYC proteins allow the cell to regulate their stability and activity. To begin with, the phosphorylation of MYC proteins at different amino acids, as well as the sequence and timing of these events, has been shown to be of major importance to their function and stability. Two of the first phosphorylation sites recognised were one at the acidic domain and one near the basic domain, with both sites being targets of the protein kinase CK2 (Luscher, Kuenzel et al. 1989). While not all the details of this modification are clearly established, it is interesting to note that the CK2-targeted sites are involved in MYC stabilisation (Channavajhala and Seldin 2002) in addition to potentially co-regulating cell-cycle progression (Li, Dobrowolska et al. 1999).

Another cascade of phosphorylation events that has attracted a lot of interest involves the amino acids Ser-58 and Thr-62, both within the MB1. The primary interest is firstly due to the number of different kinases that are associated with the phosphorylation of these two amino acids, including CDK1, MAPK and JNK, which indicates that the cell is regulating phosphorylation at the specific site by different mechanisms that depend on the cell state. Additionally, the phosphorylation of these two sites is a sequential event where firstly the one at Thr-62 is required so that
GSK3 can phosphorylate Ser-58 (Sears 2004; Yeh, Cunningham et al. 2004). Subsequently, Thr-62 phosphorylation is required for MYC activation while the Ser-58 phosphorylation is believed to acts as a signal for protein degradation (Sjostrom, Finn et al. 2005). As such, under normal conditions this MYC proteins have a window to act before they are degraded. Alternatively Ras has been shown to be involved in the inhibition of Ser-58 phosphorylation, resulting in reduced N-Myc protein degradation (Yaari, Jacob-Hirsch et al. 2005) and could hence be considered a factor that prolongs MYCN activity. Finally, phosphorylation of these 2 amino acids is linked with some of the mechanisms of ubiquitylation of MYC like the SCF-SKP2 dependent degradation pathway (Welcker, Orian et al. 2004; Yada, Hatakeyama et al. 2004).

Ubiquitylation of the MYC proteins can also occur via additional mechanisms that do not depend on phosphorylation. For example the ubiquitin-ligase complex SCF-SKP2 interacts with MYC and it has a dual role: first to assist the transcriptional activity of the MYC protein, but also to make MYC a target for degradation at the proteosome (Kim, Herbst et al. 2003; von der Lehr, Johansson et al. 2003). Additionally, HECTH9 targets Lys-63 for polymerisation resulting in enhanced recruitment of CBP/p300 (Adhikary and Eilers 2005). In addition to CBP/p300, a number of Histone Acetytransferases (HATs) that are recruited by MYC not only serve in assisting with the transcriptional activity of MYC, but also in making the MYC protein itself a target for modifications at its lysine residues (Vervoorts, Luscher-Firzlaff et al. 2003). These lysine residues serve as targets for both ubiquitylation and acetylation, and these two types of modifications have been found to oppose each other, as for example with stimulated acetylation that increases
the stability of MYC proteins (Patel, Du et al. 2004) or partially inhibits recruitment of CBP/p300 (Faiola, Liu et al. 2005).

**1.4.6: The MYC/MAX/MAD network**

In normal cells the activity of MYCN is also regulated by competition for heterodimerisation with MAX. With MYC proteins being part of the so-called MYC/MAD/MAX network there is competition for binding to MAX between the MYC family members on one hand and Mad1, Mxi1, Mad3, Mad4 (these being members of the MAD family), Mnt and Mga on the other, as well as MAX homodimerising with itself (Hurlin, Queva et al. 1995). One of the main characteristics of these proteins is that they all contain a bHLH/LZ region that acts as the point of interaction with MAX, while in addition they also contain a recognition site for the E-Box domain. It is also established that the MAD, Mnt and Mga proteins act as transcription repressors, having at least a partial overlap with the transcriptional targets of the MYC family (Figure 1.4). In theory this would make the Mad/Mnt/Mga proteins act as tumor suppressors, however it has been indicated that this is not necessarily the case. On one hand the expression patterns of the members of the two families are synchronised in such a way that there is almost no co-expression of proteins from both families at any stage of the cell-cycle while their role in terms of transcriptional activation and repression has been established to be opposite from each other. As such it is more probable that these two families exist as part of a higher level network that regulates cell differentiation rather than as direct, co-expressed competitors (Hurlin, Queva et al. 1995).
In specific, the expression levels of N-Myc protein in differentiated cells are low while for the MAD family members, Mnt and Mga there are alternating patterns depending on the stage of the cell-cycle and also the type of cell. On the other hand the expression of MYC is upregulated either following mitotic stimulation in the cell-cycle (essentially $G_0/G_1 \rightarrow S$ in differentiating cells) or to a lesser extend when it induces apoptosis, as shown in Figure 1.5 (Kelly, Cochran et al. 1983; Rottmann and Luscher 2006). Additionally, a theoretical model was proposed regarding the relative expression of the MYC/MAX/MAD members, their interactions and dimer levels through the cell cycle. The proposal suggested that in tumours with MYC deregulation the levels of MYC/MAX are constantly high during the cell-cycle at the
expense of the MAD/Mnt/Mga proteins, even though the exact details for these levels have not been experimentally established for all the involved genes (Hooker and Hurlin 2006).

Figure 1.5: Expression patterns of MYC/MAX/MAD network proteins during the cell-cycle in normal cells. These are mostly based on experimental observations and suggest that competition for MAX-binding between the MYC and MAD family members is regulated partly in a temporal manner. The red line indicates the suggested MYCN expression pattern in NB, peak 1 the expression of Mad3 during the S phase and peak 2 the increased expression of Mad3 and myc prior to terminal differentiation. Adapted from (Rottmann and Luscher 2006)

1.4.7. Regulation of MYC expression through feedback loops

Finally, the last level of regulation for MYC proteins comes in the form of feedback loops. One of the recently reported proteins that forms a negative feedback loop with MYC is the ribosomal protein RPL11, a transcription target for MYC proteins (Dai,
Arnold et al. 2007). The regulation can potentially occur via a number of mechanisms, but the ones for which there is evidence so far involve competition for binding at the MBII box between RPL11 and various essential co-factors like TRRAP (Dai, Arnold et al. 2007) or disruption of MYC mRNA stability (Dai, Sears et al. 2007). Additionally, the protein deacetylase SIRT1 has also been identified as part of another feedback mechanism where the SIRT1 gene is targeted for transcription by c-MYC and the SIRT1 protein deacetylases and destabilises c-MYC (Yuan et al., 2009, Figure 1.6). However, in a recent publication it was suggested that for the N-Myc protein SirT1 is actually a stabilising factor while functioning via a positive feedback loop and potentially a cancer therapy target (Marshall et al., 2011). This falls in line with the ambiguous role of SIRT1 in cancer development, where it is suggested that spatio-temporal elements related to its activity might provide useful information regarding its contribution to specific cases (Fang and Nicholl 2011).

Figure 1.6: Negative feedback loop between MYC and SIRT1. SIRT1 is a transcriptional target of MYC proteins, but part of its role is the deacetylation of MYC proteins that leads to the degradation, thus forming a negative feedback loop. Adapted from (Yuan, Minter-Dykhouse et al. 2009)
1.5: Systems Biology approaches to biomedical investigations

Having numerous biomarkers and characteristics associated with NB makes gaining a complete understanding of certain aspects of the disease a difficult task. This is also true with many other conditions where the experimental results examined in isolation do not provide an adequate understanding of the biological situation at hand, or can even be contradictive between studies with no apparent explanation, like in this case the significance of MYCN expression in NB patients. The reason for this in many cases is the lack of knowledge and understanding of the relevant systems properties.

Systems biology is a discipline that studies the function of cells and organisms through the dynamic processes that link the individual molecules by the means of mathematical and computational models (Bruggeman and Westerhoff 2007). In principle, the fundamental characteristic of living organisms that Systems Biology is build upon is that every system is more than the sum of its parts. With the mathematical description and examination of supramolecular structures, it becomes possible to discover and analyse emergent network properties like robustness and fragility that can help better understand the physiological behaviour of organisms (Kitano 2004). Various principles have been discovered and methodologies were developed in order to help examine system properties. These include among others the metabolic control exerted by enzymes at different steps of a network (Westerhoff and Kell 1987) in addition to the hierarchical regulation control of the related genes, that together can account for the control exerted by gene expression for the involved processes (ter Kuile and Westerhoff 2001).
1.5.1: Top-down Systems Biology

Two main approaches are followed in terms of systems biology modelling and investigation. The most popular one is the top-down approach, where the starting point is physiological observations and organism-wide datasets from various levels like the genome, transcriptome and proteome. The main advantage of such an approach is the acquisition of a complete image from these datasets and their relation to physiology. Additionally, the ease of generating such datasets with emerging technologies and reduced costs has contributed to the popularity of the top-down approach in systems biology (Hecker, Lambeck et al. 2009). These models are usually phenomenological, since the mechanisms for many of the interactions between the molecular entities that are described are not known explicitly, but their description is based on virtual mechanisms build with the purpose of replicating the data at hand (Bruggeman and Westerhoff 2007).

An example of a top-down approach is found in the study regarding the global genomic and proteomic analysis of two NB samples: one from stage 1 without MYCN amplification and one from stage 4 with MYCN amplification (Chen et al., 2010). By comparing the datasets from the two samples and identifying differentially regulated genes at either the transcript or protein level, the authors were able to identify pathways that are potentially affected in stage 4 NBs with MYCN amplification. This analysis was performed using software like Gene Set Enrichment Analysis GSEA (Subramanian, Tamayo et al. 2005) and MetaCore (Nikolsky, Ekins et al. 2005) that utilise specific algorithms and curetted data in order to identify pathways that are involved with the deregulated genes from the data. A map was
generated at the end with all the deregulated molecules identified from these datasets and their potential associations, thus providing useful information about relevant pathways that might help understand the processes that differentiate the two NB stages (Figure 1.7).

Figure 1.7: The generated map of N-Myc protein and associated molecules that are deregulated in a 4(+) neuroblastoma sample in comparison to a 1(-) neuroblastoma sample. This was the product of a top-down approach where complete datasets from the two samples at the transcriptome and proteome level were compared for significantly-different levels of expression and then analysed with tools using algorithms based on annotated data (GSEA and MetaCore). Indicated are biomarkers already associated with NB prognosis like TrkA and N-Myc, or family members of other NB-associated biomarkers like DDX21. The lines indicate association between different molecules and their colour specifies the type of association: red for suppression, green for activation, grey for unspecified association and the arrow showing the direction of the association. Furthermore, red triangular nodes indicate transcription factors, while circular nodes indicate differentially expressed molecules as identified by the specific study (red for over-expressed, blue for under-expressed). Adapted from Chen, Song et al. 2010.

1.5.2: Bottom-up Systems Biology

Alternatively, a bottom-up approach can be followed for studying a system. This methodology requires at least a good understanding of the underlying mechanistic
relationships between the molecular species of the model that can then be described with the appropriate mathematical relations (rate laws). The requirement of quantitative data is imperative, since they are used for assigning the parameter values that define the model. The final aim of this type of methodology is to combine the separate reactions and sub-networks into the complete system of the organism in order to model its complete behaviour (Bruggeman and Westerhoff 2007). Overall, these two complementary methodologies start from opposite ends in terms of the level of examination (molecular vs physiological) and attempt to eventually reach the starting point of each other, while in doing so they discover and analyse the emerging network properties of the system that is under examination.

An example of a bottom-up study would be the investigation on the effect of Rb-E2F on the restriction point of the cell-cycle (Yao et al., 2008). A model was build describing the dynamic relation between the molecules involved in this part of the cell-cycle and the parameters were set based on values found in the literature or following parameter scans that were aiming to assign a specific behaviour to the separate reactions as reported in the literature. The theoretical analysis of the model suggested that the behaviour of Rb-E2F gives rise to a bistable switch at the R-point based on the levels of expression of E2F as affected by the duration of stimulation and the levels of growth factors. Even though this was a simplified cell-cycle model, the generated hypotheses were verified experimentally with single-cell experiments and thus provided useful information regarding the dynamic properties of the R-point and the molecules involved (Figure 1.8).
Figure 1.8: Some of the key features of a bottom-up approach as illustrated in a study from (Yao, Lee et al. 2008). The dynamic relations between the molecular species are illustrated on the diagram of the pathway, while some of the simulation results of the model indicate a bistability (2 steady states: activation and maintenance) in the system in relation to the presence of E2F that eventually drives the cell through to the next step of the cell-cycle. Adapted from (Yao, Lee et al. 2008).
Table 1.3: Main characteristics of top-down and bottom-up systems biology approaches. Each method has its own strengths and is primarily determined by the starting point-of-view and the available experimental data for validation of the proposed models and their generated hypotheses. Adapted from (Logan, Kelly et al. 2010).

<table>
<thead>
<tr>
<th>Main characteristics of top-down and bottom-up systems biology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top-down systems biology</strong></td>
</tr>
<tr>
<td>• From big picture down to fine detail</td>
</tr>
<tr>
<td>• Inductive approach</td>
</tr>
<tr>
<td>• Starts with global system data (‘omics)</td>
</tr>
<tr>
<td>• Uses computer modeling and statistics</td>
</tr>
<tr>
<td>• Proposes interactions/mechanisms to explain global data (hypotheses)</td>
</tr>
<tr>
<td>• Performs experiments to test hypotheses</td>
</tr>
<tr>
<td>• Iterative cycle leads to a better model each time</td>
</tr>
<tr>
<td><strong>Advantages:</strong></td>
</tr>
<tr>
<td>– Comprehensive (potentially complete) datasets</td>
</tr>
<tr>
<td>– Focuses on system-level behavior and phenotype</td>
</tr>
<tr>
<td><strong>Disadvantages:</strong></td>
</tr>
<tr>
<td>– Models not grounded in biological knowledge</td>
</tr>
<tr>
<td>– Statistics may not accurately reflect biology</td>
</tr>
</tbody>
</table>

| **Bottom-up systems biology**                                    |
| • From fine detail up to big picture                            |
| • Deductive approach                                            |
| • Starts with local mechanistic data (e.g., kinetics)            |
| • Uses formulation and integration                              |
| • Generates a mechanistic model of the subsystem                 |
| • Combines subsystem models into a global model (modular approach) |
| • Accumulation leads to bigger model each time                  |
| **Advantages:**                                                   |
| – Models grounded in biological knowledge                       |
| – Parameters are determined experimentally                       |
| **Disadvantages:**                                                |
| – Difficult to obtain accurate mechanistic data                  |
| – Combining models lead to cumulative errors                     |
In either of these situations, building a model is an iterative process that requires continuous validation of the experimentally-obtained data with the theoretical predictions and hypothesis generated by the models and *vice versa*. Each one has an effect on the other, where the model-generated hypothesis can define the nature of the performed experiments, while these in turn should produce data that is used for the validation and refinement of the model in order to provide a better representation of the observed biological behaviour and more accurate predictions (Logan, Kelly et al. 2010).

### 1.6: qPCR data analysis at the DNA level

In either of the two Systems Biology approaches, it is imperative that the experimental data used for building the models or validating the generated hypotheses are both accurate and properly analysed. As part of this project, one of the quantified parameters would be the gene copy-number (GCN) of MYCN. Traditionally the methods used for detecting the MYCN amplification in NB have been fluorescence in-situ hybridisation (FISH) and Southern blotting. In 2002 a number of oncogenes related to NB were used in a study showing that quantitative Polymerase Chain Reaction (qPCR) could be used for accurately establishing gene copy-numbers (GCNs) from clinical samples while having additional benefits in relation to the previously used methods (De Preter, Speleman et al. 2002).

In the same study the authors also suggested the absolute quantification of GCNs without the use of standard curves from foreign genetic material but rather solely
based on the use of the appropriate reference or housekeeping genes (HKGs) for each study and comparisons with a test sample. A proposal ensued to modify the classic $2^{-\Delta C_q}$ formula (Pfaffl 2001, Livak and Schmittgen 2001) into Equation 1.1:

$$\frac{(1+E)^{\Delta C_q_{\text{gene}}}}{(1+E)^{\Delta C_q_{\text{reference gene}}}}$$

(Equation 1.1)

The quantification is based on the $\Delta C_q$ ratios of the gene of interest and the HKGs between the sample and control material, while $E$ represents the efficiency of each individual test that can be set at a fixed value for all genes under consideration. This formula was modified further by producing Equation 1.2 that was part of the software qBase (now qbase\textsuperscript{plus}) where multiple HKGs were used for normalisation of the relative quantification of a target gene (NRQ represents normalised relative quantities, Hellemans, Mortier et al. 2007).

$$NRQ = \frac{\prod_{i=1}^{n} e^{\Delta C_q_{\text{gene}}}}{\sqrt{\prod_{i=1}^{n} e^{\Delta C_q_{\text{reference gene}}}}}$$

(Equation 1.2)

These formulae are based among others on two concepts that are commonly examined in the field of qPCR data analysis; the use of appropriate housekeeping genes and the appropriate calculation and utilisation of the qPCRs’ efficiencies. We believe these concepts can be expanded upon and help provide an alternative methodology for data analysis in order to move further away from GCN ranking estimations to achieve absolute quantification of the GCN of an oncogene using multiple HKGs with or without a control sample.
The use of HKGs for normalisation purposes has been paramount in the field of qPCR data analysis. This is reflected in the basic $2^{\Delta\Delta C_q}$ algorithm used for calculating relative GCNs and which forms the basis for a number of different calculation variants. It is evident from the formula that the outcome is partly dependent on the HKGs and as a result there have been reports for different methodologies of how to choose the most suitable HKGs for a study. Some of these studies produced new formulae and software to assist in achieving this goal, with some of the most popular being GeNorm (Vandesompele, De Preter et al. 2002), BestKeeper (Pfaffl, Tichopad et al. 2004) and NormFinder (Andersen, Jensen et al. 2004). These methodologies were concerned primarily with results at the mRNA level and sought to help establish the levels of relative rather than absolute stability among sets of HKGs. This is understandable due to the nature of mRNA expression profiles of any gene and the heterogeneity found from sample to sample or even from cell to cell (Vandesompele, De Preter et al. 2002).

On the other hand, when choosing a HKG for DNA level qPCR studies there is the additional prerequisite for the ideal genes that are single copy genes per haplotype. What this translates to is that at a certain level of measurement two HKGs should give values that are not significantly different after appropriate statistical analysis. For these values, classic parametric or non-parametric statistical methods should be applicable and can help establish sets of HKGs displaying non-significantly different GCNs within and across samples.
Additionally, it has been argued that the correct level for statistical comparisons of qPCR data is not at the Cq level, but rather at a "linearised" level where the data is transformed by the formula in Equation 1.3 (Livak and Schmittgen, 2001):

\[2^{-Cq}\]  

(Equation 1.3).

The data at this “linear” or “relative concentration” format as represented here depend exclusively on the Cq values obtain from each test. This is due to the base 2 being applied to each power-transformation which represents the ideal and maximum theoretical amplification efficiency per qPCR cycle where the target genetic material is produced geometrically.

The use of predefined efficiencies for qPCR data transformation, and the most appropriate method for establishing the actual efficiency of a test have also been subjects of investigation in the field. To begin with, the efficiency of a qPCR test was defined by Pfaffl as

\[E = 10^{\frac{1}{\text{slope}}}\]  

(Equation 1.4)

where the slope is taken from the linear part of a standard curve representing the Cq values vs \(\log_{10}\) concentration of the target genetic material (Pfaffl 2001). Other definitions and methods of manually calculating the efficiency of a test have also been suggested (Khan-Malek and Wang 2011) or even incorporated in the form of \(E=10^{\frac{1}{\text{slope}}}-1\) by software like the latest versions of REST (Pfaffl, Horgan et al. 2002).
Furthermore, each gene is tested by a different set of experimental conditions and as such their amplifications have different efficiencies. For reasons such as interference from sample impurities these efficiencies can go higher than the theoretical maximum, while for practical purposes the optimum range is considered to be +/- 10% of the maximum theoretical (Nolan, Hands et al. 2006). When concerned with the relative concentration values, even the smallest differences in the efficiencies of individual tests could have a high impact on the outcome of any calculations and comparisons between genes or even replicates of the same gene. Software like BestKeeper or REST take this into account, while Yuan and colleagues compared the outcomes of the original $2^{-\Delta\Delta Cq}$ formula versus one with adjusted efficiencies for each test and highlighted the differences in the outcome (Yuan, Wang et al. 2008). As such, the specific efficiency for each test is a parameter that has to be taken into account when performing analysis of qPCR data, whether it is to compare HKGs or to quantify GCNs of target genes, regardless of the method used to establish said efficiencies.

As part of this study, the MYCN GCN would be established on 4 NB cell-lines based on the selection of appropriate HKGs from a set of 4 candidates. The Cq values would be transformed into “relative concentration” values using the specific efficiency for each test as calculated by the method suggested by (Pfaffl 2001). The data would then used to test statistically significant differences between the HKG means across all 4 cell-lines and then quantification of the MYCN GCNs would be performed for each cell-line. Due to the nature of the data and the assumptions related to static copy number that can be made for the HKGs at the DNA level (unlike mRNA which is subject to regulation and where copy numbers can vary), the
quantification could be of absolute GCN values without the need for a reference sample. Assuming certain criteria are met, this methodology could be used for choosing appropriate HKGs at the DNA level by performing absolute quantification without the need for control samples.

1.7: Project aims and objectives

Based on the research reported in the literature regarding the expression of the MYCN oncogene in NB tumours and its potential contribution to prognosis, the aim of this project was to investigate the significance of this expression using a Systems Biology approach (Figure 1.9). The reported literature suggests that there is a clear discrepancy between the MYCN gene expression (RNA and protein) and MYCN gene amplification, in terms of their significance in relation to staging, prognosis and outcome of NB patients. While the effect of MYCN amplification is considered to be strongly related to increased N-Myc protein activity, several alternative possibilities have been suggested where, for example, the significance of MYCN gene amplification doesn’t lie in the subsequent effects of the N-Myc protein activity but rather some unknown, upstream events that have as one of their consequences the amplification of MYCN at the genetic level and the associated effects to NB (section 1.3.6).

Additionally, while the characterisation of MYCN gene and transcript numbers is performed exclusively in a quantitative manner, in the vast majority of the reported cases for the N-Myc protein the quantification is of qualitative nature. This limits the
suggestions and conclusions that can be associated with a potentially functional role of the N-Myc protein in NB. Finally, with the emergence of the Systems Biology field it has become evident that it is important to take into account the relevant systems associated with the biological entities under examination since potentially novel and emergent network properties might help explain behaviours that cannot be fully understood when individually examining said entities. For this happen, it is imperative that quantitative and properly analysed data is acquired that can then be used for parameterising and testing the relevant mathematical models.

Based on this information, and keeping in mind the financial and practical limits associated with this PhD program, we decided to examine the significance of MYCN amplification and expression from a Systems Biology point of view. The novelty of this method would be the first reported use of mathematical models describing the association of relevant molecules and biological entities, while all the information and data used for parameterising the model would be of quantitative nature and the mathematical models would be used in order to examine related systems properties. In specific:

1. A bottom-up model would be created linking the MYCN GNC to the MYCN transcript and N-Myc protein numbers, with the protein MAX also included in the model as the binding partner of N-Myc protein. The kinetic relations of the model would be described using the appropriate rate-laws that in this case were chosen to be mass-action. The parameters of the model would be found from the literature, with any remaining parameters calculated accordingly. The aim of the model would be to investigate the effect of MYCN
amplification under the specified conditions and help in the hypothesis generation regarding the potential significance of the expression and activity of MYCN. This model would eventually form the basis for separate models describing individual NB cell-lines that would be compared between them in order to provide information regarding changes in systems properties and the levels of control of the different reactions of the model, thus helping in deciding at which level MYCN might be more sensitive to treatment. As part of this work, one of the long term views aims was to incorporate this model into the cell-cycle model by Yao and colleagues and as such the chosen units for describing the model parameters and variables were taken from there (Yao et al., 2008).

2. Experimental data would be acquired from 4 NB cell-lines in order to provide a unique expression profile for different NB cases. According to the literature, these 4 cell-lines (SHEP1, Kelly, NB-EB, SH-SY5Y) have different profiles in terms of MYCN GCN and expression. Quantitative measurements for MYCN gene, transcript and protein numbers would be made in order to acquire values that would be used in validating and refining the model parameters. Ideally the obtained values would be of absolute nature in the form of number of molecules or concentration per cell. The available methodologies included qPCR and reverse transcription (RT) qPCR, enzyme-linked immunosorbent assay (ELISA) and western blots. Emphasis would be placed upon both obtaining accurate data with the available means and also appropriate data analysis.

3. The obtained experimental data would be used to recalculate the model parameters for each cell line and create a unique version of the model for
each one. The models would then be analysed in order to validate if they support any generated hypotheses from the original model, while also investigate whether they create new hypotheses themselves, both as individual models but also in comparison to each other. Potential differences in the systems properties and control coefficients of the reactions are of particular interest, while any generated hypotheses would help define any additional experimental work.

Figure 1.9: Flow-chart with the proposed work-plan of the study.
2.1: Cell culture

FourThe neuroblastoma cell-lines, SHEP-1, Kelly, NB-EB and SH-SY5Y, were used in this study. The cells were maintained in 25ml flasks in RPMI 1640 media with L-glutamine and NaHCO₃ (PAA Laboratories, Yeovil, U.K.) that was supplemented with 10% Foetal Calf Serum, 100 µg/mL streptomycin and 100U/mL penicillin (all supplementary materials were from Sigma-Aldrich, Gillingham, U.K.). This will be referred to in subsequent sections of this document as “Full Media”. For freezing purposes the cells were immersed in FCS supplemented with 10% Dimethyl Sulphoxide DMSO (BDH Laboratories, East Grinstead, U.K.) incubated for 24 hours storage at -80°C prior to storage in liquid nitrogen. Phosphate Buffered Saline (PBS) was used for washing the cells prior to any treatment including the use of Trypsin-EDTA for detaching the cells from the flask surface (both reagents from PAA Laboratories). The growth conditions were 37°C and 5% CO₂.

2.2: Nucleic Acid Extraction:

2.2.1: DNA:

The cells were washed with PBS and treated with Trypsin-EDTA. A volume of Full Media equal to 3 times the volume of Trypsin-EDTA used was added once the cells were detached from the flask. They were then collected by centrifugation at 1.5k500 rpm200g for 4 minutes and DNA was extracted using the procedure and reagents
from the ISOLATE Genomic DNA Kit (Bioline, London, U.K.). The extracted genomic DNA was re-suspended in 50µl dH₂O and stored at -20°C for future use.

2.2.2: RNA

The cells were collected in the same way, washed once with PBS and suspended in 1ml TRIzol (Invitrogen, Paisley, U.K.) and 0.2ml chloroform (BDH, acquired via VWR, East Grinstead, U.K.) before breaking them down by vortexing. After a 2-minute incubation at room temperature (RT) they were centrifuged for 15 minutes at 13k rpm/14000g and the supernatant was transferred into a fresh tube containing equal volume of 70% ethanol (BDH). The RNA from the cells was then extracted according to the protocol for the RNeasy Mini Kit (Qiagen), and finally suspended in 50µl dH₂O and stored at -20°C.

2.2.3: cDNA

The RNA samples were DNase treated for elimination of DNA molecules using the protocol and reagents from the DNA-free kit (Ambion, Warrington, U.K.). From the resulting DNA-free RNA, 2µg was used for reverse transcription into cDNA using SuperScript 3 III Reverse Transcriptase (Invitrogen) and the corresponding protocol, while the reaction was primed by using oligo(dT)₁₂₋₁₈. The concentration and purity of the collected genetic material (DNA, RNA, DNA-free RNA, cDNA) was checked at each step using the NanoDrop ND1000 Spectrophotometer (Thermo Scientific, Langenselbold, Germany). All genetic material was stored at -20°C and used for subsequent downstream applications within 3 months of extraction.
2.3: Quantitative Polymerase Chain Reaction (qPCR)

The specific qPCR assays for each gene of interest were designed using the ‘Universal Probe Library Assay Design Centre’ from Roche, while the chosen primers were checked for specificity using the Basic Local Alignment Search Tool (BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Each assay was run on 384 well plates in 10µl reaction volumes comprising: 5µl Light Cycler 480 Probe Master Mix (Roche Applied Sciences, West Sussex, UK), 200pM forward and reverse primers (Metabion, Martinsried, Germany), the corresponding probe from the Roche Universal Probe Set Library-Human (Roche Applied Sciences) and 4µl of material to be tested. The material was one of DNA, cDNA, template oligonucleotide (Metabion) or dH2O (negative control).

The qPCR reactions were run on the Light Cycler® 480 Real-Time PCR System (Roche Applied Sciences) and the program used consisted of a single step at 95°C for 5mins, followed by 50 cycles at 95°C for 10seconds and each and a single cycle at 60°C for 30seconds, with a final cooling temperature at 40°C. Calibration graphs were constructed for each gene to be tested by doing serial dilutions at known concentrations of either the template oligonucleotide or genetic material and then plotting the Cp values versus the log10 concentration of the sample. The list of genes and the relevant sequences of the primers and template oligonucleotides used are found in Appendix 1.
2.4: Sample preparation for protein-content analysis

The lysis buffer used at this stage consisted of 50mM Trizma Base, 50mM β-glycosphosphate, 1mM Na3VO4, 5mM NaF, 1%, TX-100 at pH7.5 (all products from Sigma-Aldrich). Protease Inhibitor Cocktail Set 1 (Merck, Nottingham, U.K.) at 1X final concentration was added to the lysis buffer prior to being used.

The cells were plated 24 hours in advance in a 6-well plate so that they would >80% confluent at the time of lysis and collection. They were washed twice with PBS and then appropriate amount of Lysis Buffer was added and incubated at 4°C for 20 minutes with occasional shaking. The cells were scraped from the well, collected in eppendorf tubes, centrifuged at 13,000rpm/14000g, and 4°C for 15mins and the supernatant was transferred in a separate tube. The total protein content for each sample was then measured using the Bicinchoninic Acid (BCA) Protein Assay (Fisher Scientific, Loughborough, UK). The protein content of the lysates was measured in technical triplicates at two different dilutions (1:5 and 1:10) with the concentration taken as the average between the readings.

2.5: Enzyme Linked Immunosorbent Assay (ELISA)

For the purpose of this protocol, all incubation steps were done at 4°C with shaking at 60rpm unless stated otherwise. Between each step the plate was washed 4 times with “Wash Buffer” consisting of PBS with 0.05% v/v Tween 20 (Sigma Aldrich) unless stated otherwise. “Block Buffer” as mentioned in the protocols consists of PBS with 1% w/v BSA (Sigma-Aldrich). All volumes used were 50µl per well
unless stated otherwise and all antibodies were diluted to the desired concentration using Block Buffer. At all stages the plate was covered on top with parafilm and completely with aluminium foil as such the wells were always devoid of light during the incubation steps. For each sample a minimum of 3 technical replicates were tested. The antibodies and protein standards that were used for these protocols are shown in Table 2.1:
Table 2.3: The antibodies and positive controls used for the quantification of the N-Myc and MAX proteins with the ELISA method.

<table>
<thead>
<tr>
<th>Name</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYCN Monoclonal Antibody (M02) Clone 4H4 (mouse monoclonal, 1mg/ml)</td>
<td>Abnova/ Caltag Medsystems Ltd (Buckingham, UK)</td>
</tr>
<tr>
<td>N-Myc (3C165) Antibody (mouse monoclonal, 200µg/ml)</td>
<td>Santa Cruz Biotechnologies (Heidelberg, Germany)</td>
</tr>
<tr>
<td>N-Myc (H-50) Antibody (rabbit polyclonal, 200µg/ml)</td>
<td>Santa Cruz Biotechnologies</td>
</tr>
<tr>
<td>N-Myc antibody (rabbit polyclonal, 200µg/ml)</td>
<td>New England Biolabs (Hitchin, UK)</td>
</tr>
<tr>
<td>Max (S20) antibody (rabbit polyclonal, 200µg/ml)</td>
<td>New England Biolabs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goat anti-mouse IgG-FITC sc2010 (400µg/ml)</td>
<td>Santa Cruz Biotechnologies</td>
</tr>
<tr>
<td>Mouse anti-rabbit IgG-FITC sc2359 (400µg/ml)</td>
<td>Santa Cruz Biotechnologies</td>
</tr>
<tr>
<td>Goat anti-mouse IgG-Biotin sc2039 (400µg/ml)</td>
<td>Santa Cruz Biotechnologies</td>
</tr>
<tr>
<td>Mouse anti-rabbit IgG-Biotin sc2491 (400µg/ml)</td>
<td>Santa Cruz Biotechnologies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Myc purified protein 110ng/µl (H00004613-P01)</td>
<td>Abnova, Caltag Medsystems Ltd</td>
</tr>
<tr>
<td>Max purified protein 0.5mg/ml (NBP1-44497)</td>
<td>Abnova/ Novus Biologicals (Cambridge, UK)</td>
</tr>
</tbody>
</table>

2.5.1: Normal ELISA:

The samples were incubated in a 96-well plate F-96 Certified Maxisorp Nunc Immunoplate (Thermo Scientific) O/N at the desired concentration. The plate was then washed and a minimum of 50µl Block B buffer was added for a 1-hour
incubation, followed by a 1-hour incubation with detecting antibody (anti-NMyc or anti-MAX) and then 20 minutes with the appropriate secondary antibody.

When using a FITC-conjugated secondary antibody the plate would be read straight after the secondary antibody incubation with the Synergy HT Multi-Mode Microplate Reader (NorthStar Scientific Ltd, Leeds, U.K.). The settings were 485nm (+/-20) for excitation and 528nm (+/-20) for absorbance. If the secondary antibody was Biotin-conjugated then these additional steps were followed by: firstly a 20 minute incubation with 0.5µg/ml Streptavidin-Horse Radish Peroxidase SHRP (RnD Systems, Abington, U.K.) followed by a 20 minute incubation with TMB Substrate (BD Biosciences, Oxford, U.K.) and finally a minimum 20 minute incubation with 1M H\textsubscript{2}SO\textsubscript{4} (Sigma-Aldrich) that was added to the plate without washing out the TMB Substrate. A colorimetric measurement was then performed at 450nm HT Multi-Mode Microplate Reader.

2.5.2: Sandwich ELISA:

The capture antibody was incubated O/N in a 96 well plate and was followed by a 1-hour incubation with Block buffer, a 2-hour incubation with the samples, a 1-hour incubation with the detection antibody and then a 20-minute incubation with the secondary antibody. The screening of the results that followed depended on the secondary antibody used and it was done as described for the Normal ELISA method.
2.6: Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and Western Blots

2.6.1: Buffer/gel materials and recipes:

5x Loading Buffer: 10% w/v SDS, 10nM β-mercaptoethanol, 50% v/v Glycerol, 0.4M Tris-HCl, pH6.8, 0.025% w/v Bromophenol Blue (all from Sigma-Aldrich)

1x Running Buffer: 25mM Trizma Base, 1.92M Glycine, 0.1% w/v SDS (all from Sigma-Aldrich)

1x Transfer Buffer: 24mM Trizma Base, 1.93M Glycine (all from Sigma-Aldrich)

Resolving Gel: dH2O, 1.5M Tris-HCl pH8.8, 10% w/v SDS, 30% w/v Acrylamide, 0.8% w/v Bis-Acrylamide Stock Solution, 37.5:1 (National Diagnostics Ltd, Hessle, U.K.), ammonium persulfate (for electrophoresis, ≥ 98%), TEMED (Biorad, Hemel Hemstead, U.K.)- all from Sigma-Aldrich unless otherwise stated. The final concentrations for each material were altered based on the required final percentage of SDS, as defined by the molecular weight of the protein to be detected.

Stacking Gel: The same components were used as for the resolving gel, with the exception of replacing the 1.5M Tris-HCl pH8.8 with 0.5M Tris-HCl pH6.8.

2.6.2: SDS-PAGE gel protein separation:

The resolving gel was prepared at either 10% SDS (for MYCN-myc protein detection) or 12% SDS (for MAX protein detection). The samples were mixed with PBS and 5x Loading Buffer for the desired concentration and incubated at 95°C for 5 minutes prior to being loaded to the gel. Each gel also included the Precision Plus
Protein Ladder Plus (Biorad). The gel was run at 70V for 40 minutes and then at 150V for 1 hour.

2.6.3: Western bBlot:

During the next step the samples were transferred for 45 minutes at 15V from the gel to a nitrocellulose membrane that was soaked in Transfer Buffer. The membrane was then incubated for 1 hour in 5% Milk (Marvel) and 0.1% Tween 20 (Sigma Aldrich) in PBS, washed 10 times with 0.1% Tween 20/PBS and then incubated O/N at 4°C with the corresponding primary antibody in either 5% Milk/0.1% Tween 20/PBS or 0.1% BSA/0.1% Tween 20/PBS. The primary antibodies were the same used for ELISA as listed in the previous section. The membrane was then washed for an hour with the same buffer before incubated with the appropriate secondary antibody (1:1000 dilution in 5% Milk/PBS/0.1% Tween or 0.1% BSA/0.1% Tween 20/PBS) for 1 hour under the same conditions. The primary antibodies were the same used for ELISA as listed in the previous section.

For chemiluminescence-based detection methods, the secondary antibodies used were either Rabbit Anti-Mouse HRP-conjugated or Goat Anti-Rabbit HRP-conjugated at 1:1000 dilution (Dako, Ely, U.K.). Following a multi-step 20-minute wash with 0.1% Tween 20/PBS the membranes were then incubated for 1a minute with 2ml ECL Plus Western Blotting Detection Reagent (Amersham, Little Chalfont, U.K.). Finally, the membranes were exposed to film (CL-XPosure TM Film, Thermo Scientific) for the appropriate time interval before being developed.
For fluorescence-based methods, the secondary antibodies used were the IRDye 800CW Goat anti-Mouse IgG (H+L), IRDye 800CW Goat anti-Rabbit IgG (H+L), IRDye 680LT Goat anti-Mouse IgG (H+L) and IRDye 680LT Goat anti-Rabbit IgG (H+L) antibodies (LI-COR Biotechnology, Cambridge, U.K.). All were at 1:5000 dilution in 0.1% BSA/0.1% Tween 20/PBS and the incubations were performed under no-light conditions, while the membranes were visualised with the Odyssey® Infrared Imaging System (LI-COR Biotechnologies). The antibodies used for loading controls were the Histone H3 Mouse monoclonal Ab (96C10) from NEB and the GAPDH rabbit polyclonal Ab (ab8485) from Abcam (Cambridge, U.K.).

2.7: Dry Laboratory Materials

The model was created, simulated and analysed using the Complex Pathway Simulator (COPASI) software (Hoops, Sahle et al. 2006). The graphs were created using Microsoft Excel 2007 and the figures were created or modified using Microsoft PowerPoint 2007. Statistical analysis of the data was performed using OriginPro8.5 (OriginLab, Northampton, U.S.A.). The quantification of the band intensities from the western blot images was performed using the software ImageJ (http://imagej.nih.gov/ij/).
Chapter 3

RESULTS: MODELLING MYCN EXPRESSION

3.1: OVERVIEW

The significance of MYCN expression in NB is still a matter of controversy as is shown from the results of various conflicting reports regarding the levels, predicted outcome and statistical significance of MYCN expression across various subtypes of NB. These reports were primarily concerned with the observed MYCN mRNA and/or protein levels in a mostly qualitative manner, especially for the N-Myc protein levels. As such, they were trying to match these qualitative observations with the physiological description of the patients and their Event-Free Survival (EFS) or Survival (S), without any real gain in understanding of the mechanistic significance at the molecular level of the MYCN expression.

Alternatively, this is a subject that would be best suited for a systems-based investigation, but following a bottom-up methodology as already described in Chapter 1. This approach was chosen since the starting-point would be the MYCN DNA level as one of the most reliable NB biomarkers, while establishing the kinetic relations and investigating the systems properties associated with the relevant reactions that lead to the activity of the N-Myc/MAX dimer. The resultant mathematical model would give us a tool that would be used for analysing the predicted outcome of the system under different conditions and perturbations, such
as MYCN amplification. Finally, analysis of the systems properties under the different sets of conditions could provide useful information regarding potential shifts of importance between reactions under changing conditions.

3.2: Defining the model: rate laws and parameters

The model would describe the path of MYCN going from the gene level to the activity level in conjunction with MAX as a heterodimer. The reactions of MYCN transcription, translation, dimerisation and promoter binding would be included in the model. Additionally, mRNA and N-Myc protein degradation reactions would also be included, however that was not the case with feedback loops since they are another area of conflicting information that would complicate the model unnecessarily at this stage. Finally, the transcription and translation reactions were not broken down to the different steps described in Chapter 1, but rather kept as a single, unidirectional reaction each. Based on the simulation and analysis results, if any of these steps would be deemed suitable then it can be expanded according to the interactions described in the literature. The model would first incorporate the following species and interactions as shown in Figure 3.1:
3.3: Connection of the model to the expression of MYCN during the cell-cycle

One of the aims of this project was to parameterise the model describing MYCN expression with data and kinetic values found in the literature. With this achieved, the model would then be validated with accurate, experimentally obtained data that would also be used in order to examine any suggestions or theories that might arise from performing simulations with the model. In terms of the model parameterisation, it would be necessary to take into consideration the expression pattern of MYCN throughout the cell cycle in normal cells. This, is illustrated in figure 1.5 showing a complex expression profile that changes according to the stage of the cell cycle. As such, ideally the stage of the cells used for obtaining the material for the experiments would be accurately established and then construct a model with changing behaviour according to the stage of the cell cycle.
However, there were certain practicalities that needed to be considered before proceeding with this task. The main one was that the experimental data would be acquired from cells in un-synchronous culture and as such the obtained results would correspond to a quasi steady-state rather than any specific part of the cell cycle. In other words, simply harvesting the cells and obtaining their DNA, RNA and protein status does not give data that can be directly referred to any part of Figure 1.5, but rather correspond to average values from all the cells at potentially all stages of the cell cycle. The major negative aspect of this data-collecting method is that the results could potentially be irrelevant or misleading in relation to the actual dynamic properties of MYCN expression during the cell cycle. That is something that would have to be validated in later stages of the project with the refinement of both the experimental techniques and the model.

The decision taken, though, was to proceed with a quasi steady-state methodology for mainly two reasons. Firstly, some of the methods for collecting the appropriate MYCN-related material (DNA, RNA, protein) would have to be set up and optimised for the first time in the particular lab where they would take place. It was deemed more important that this optimisation process occurs prior to any validation of cell-synchronisation or single cell-based protocols since not having accurate methods for performing these measurements would invalidate the most important part of the project. The second reason for following this methodology was that the vast majority of the published work regarding the expression of MYCN in neuroblastoma was following the same procedure without trying to synchronise the cells under examination or link the obtained results to specific parts of the cell cycle. This would mean for example, that kinetic parameters obtained from the literature in
relation to RNA and protein degradation rates would again correspond to a quasi-steady state rather than any specific part of the cell-cycle (Amy and Bartholomew, 1987; Cohn et al., 1990). Finally, while, the protein expression levels associated with both cell lines and clinical samples were also obtained from un-synchronised samples (Section 1.3).

In conclusion, for reasons associated with both lab-based practicalities and the nature of the available data and kinetic parameters related to this study, the decision was taken to create a model for MYCN expression at a quasi steady-state. On the other hand, having the associated limitations in mind, one of the long-term goals of this project would be to eventually switch from this quasi steady-state model to one that more accurately reflects the different expression patterns of MYCN during the cell-cycle.

3.4: Building the model:

3.4.1: Rate laws for the reactions

The kinetic reactions were described using mass-action kinetics. The rate laws describing each reaction are shown on Table 3.1:
Table 3.4: The rate-laws that describe the reactions of the model.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Rate law</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcription</td>
<td>$k_1 \cdot DNA$</td>
</tr>
<tr>
<td>Translation</td>
<td>$k_2 \cdot RNA$</td>
</tr>
<tr>
<td>Dimerisation</td>
<td>$k_{-3} \cdot NMyc_MAX - k_3 \cdot NMyc \cdot MAX$</td>
</tr>
<tr>
<td>Promoter binding</td>
<td>$k_{-4} \cdot NMyc_MAX_Promoter - k_4 \cdot NMyc_MAX \cdot Promoter$</td>
</tr>
<tr>
<td>RNA degradation</td>
<td>$k_5 \cdot RNA$</td>
</tr>
<tr>
<td>Protein degradation</td>
<td>$k_6 \cdot NMyc$</td>
</tr>
</tbody>
</table>

By merging the rate laws the following into balance equations describe the rate of change of each species becomes as stated in (Table 3.2):

Table 3.5: The ordinary differential equations (ODE) describing the rate of change of the model species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYCN RNA</td>
<td>$\frac{dRNA}{dt} = k_1 \cdot DNA - k_5 \cdot RNA$</td>
</tr>
<tr>
<td>N-Myc protein</td>
<td>$\frac{dN_Myc}{dt} = k_2 \cdot RNA + k_{-3} \cdot NMyc_MAX - k_3 \cdot NMyc \cdot MAX - k_6 \cdot$</td>
</tr>
<tr>
<td>N-Myc/MAX dimer</td>
<td>$\frac{dNMyc_MAX}{dt} = -k_{-2} \cdot NMyc_MAX + k_3 \cdot NMyc \cdot MAX + k_{-4} \cdot N. \cdot NMyc_MAX \cdot Promoter$</td>
</tr>
<tr>
<td>N-Myc/MAX dimer bound to E-Box</td>
<td>$\frac{dNMyc_MAX_Promoter}{dt} = k_4 \cdot NMyc_MAX \cdot Promoter - k_4 \cdot NM$</td>
</tr>
</tbody>
</table>

Assuming steady state conditions where there is no change of these species over time ($\frac{dx}{dt} = 0$, where x is any of molecular species of the model shown in Table 3.3) the equations can be presented in the following forms:
Table 3.6: The ODEs describing the rate of change of each species at steady state (ss) conditions.

<table>
<thead>
<tr>
<th>Species</th>
<th>Rate law at steady state (ss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA</td>
<td>$k_1 * DNA = k_3 * RNA_{ss}$</td>
</tr>
<tr>
<td>N-Myc protein</td>
<td>$k_2 * RNA_{ss} + k_{-2} * NMye_{ss} * MAX_{ss} = k_1 * NMye_{ss} * MAX_{ss} + k_6 * $</td>
</tr>
<tr>
<td>N-Myc/MAX dimer</td>
<td>$k_4 * NMye_{MAX_{ss}} * Promoter_{ss} + k_{-4} * NMye_{ss} * MAX_{ss} + k_4 * NMye_{ss} * Prom</td>
</tr>
<tr>
<td>N-Myc/MAX dimer bound to E-Box</td>
<td></td>
</tr>
</tbody>
</table>

Adjusting the equations for the N-Myc protein and the N-Myc/MAX dimer we get the final versions shown in Table 3.4. These simplified versions were obtained by merging some of the equations of Table 3.3, for example the equation for the N-Myc/MAX dimer bound to the promoter with the equation for the N-Myc/MAX dimer, and sequentially for the N-Myc protein.

Table 3.7: Simplified equations for describing the rate of change of each species at steady state conditions.

<table>
<thead>
<tr>
<th>Species</th>
<th>Simplified Rate law at steady state (ss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA</td>
<td>$RNA_{ss} = k_1 * DNA/k_3$</td>
</tr>
<tr>
<td>N-Myc protein</td>
<td>$NMye_{ss} = k_2 * RNA_{ss}/k_6$</td>
</tr>
<tr>
<td>N-Myc/MAX dimer</td>
<td>$NMye_{MAX_{ss}} = k_3 * NMye_{ss} * MAX_{ss}/k_{-2}$</td>
</tr>
<tr>
<td>N-Myc/MAX dimer bound to E-Box</td>
<td>$NMye_{ss} * Promoter_{ss} = k_4 * NMye_{ss} * MAX_{ss} * Promoter_{ss}/k_{-4}$</td>
</tr>
</tbody>
</table>

These equations were used for the steady state calculations described in this chapter.
3.4.2: Model parameters

The next step was to find from the literature reported parameter values for the rate laws of the model. “Parameters” are defined as the time-independent values that describe a model, while “variables” are the time-dependent values associated with the model. As such it was important to firstly choose the units with which all parameters would be normalised. In this case the chosen units were hours (h) for time, litres (l) for volume and micromolar (µM) for concentration for reasons of consistency and compatibility, since one of the potential uses of the model would be the incorporation into the cell-cycle model from Yao and colleagues that already includes MYC (Yao et al., 2008).

The parameter values shown in Table 3.5 were obtained from the literature and were adjusted in order to be compatible with the chosen units. As an example, Fernandez and colleagues provided the number of potential transcription initiation sites as that include an E-Box ast 2224 per cell (Fernandez et al., 2003). Using a cell volume of $5 \times 10^{-12}$ L this number becomes 0.00074µM.

These values and the rate laws of the model were used for calculating any remaining unknown parameters. To do this, the model was assumed to have reached steady state, describing initially a system without MYCN amplification. As such we set the MYCN gene copy-number as 2 per normal diploid cell or a concentration of $6.63 \times 10^{-7}$ µM. Additionally, the N-Myc protein degradation rate chosen was the one established for non-amplified neuroblastoma cell-lines (Cohn, Salwen et al. 1990),
rather than the one for a cell-line with MYCN amplification (Slamon, Boone et al. 1986).

<table>
<thead>
<tr>
<th>Parameter/Species</th>
<th>Value</th>
<th>Units</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{3eq}$</td>
<td>0.013</td>
<td>1/µM</td>
<td>(Park, Chung et al. 2004)</td>
</tr>
<tr>
<td>$k_3$</td>
<td>$6.4 \times 10^6$</td>
<td>1/(µM*h)</td>
<td>(Ecevit, Khan et al. 2010)</td>
</tr>
<tr>
<td>$k_4$</td>
<td>760</td>
<td>1/µM*h</td>
<td>(Park, Chung et al. 2004)</td>
</tr>
<tr>
<td>$k_{4eq}$</td>
<td>6.6</td>
<td>1/µM</td>
<td>(Park, Chung et al. 2004)</td>
</tr>
<tr>
<td>$k_5$</td>
<td>1.4</td>
<td>1/h</td>
<td>(Amy and Bartholomew 1987)</td>
</tr>
<tr>
<td>$k_6$</td>
<td>0.4</td>
<td>1/h</td>
<td>(Cohn, Salwen et al. 1990)</td>
</tr>
<tr>
<td>Promoter</td>
<td>0.00074</td>
<td>µM</td>
<td>(Fernandez et al., 2003)</td>
</tr>
<tr>
<td>Cell volume</td>
<td>$5 \times 10^{-12}$</td>
<td>L</td>
<td>(Cohen and Studzinski 1967)</td>
</tr>
</tbody>
</table>

Starting with the bound E-Boxes related to promoter sites (NMyc_MAX_Promoter in the model), in the same publication Fernandez and colleagues reported that 58% of them were found to be occupied in cell-lines with no MYC amplification (Fernandez et al., 2003). In our model it would translate into 0.00043µM at steady state, leaving the free promoter concentration at 0.00031µM for the same conditions. Based on these numbers the value of the [NMyc_MAX/ss] was then calculated to be 0.21µM.

The rest of the parameters of the model could not be defined in such a precise manner. Starting with the mRNA concentration, it is reported in the literature that there is no specific value that can be assigned to a specific subtype of neuroblastoma patient or even a precise correlation with the MYCN gene copy numbers, but rather
it is patient (or in our case cell-line) specific. This meant that the transcription rate $k_1$ could not be calculated based on pre-reported values either. As such, the number of 100 MYCN mRNA$_{ss}$ molecules per cell ($3.3 \times 10^{-5}$µM concentration) was initially chosen and $k_1$ was calculated as $70h^{-1}$. These values would be adjusted once the appropriate experimental values were obtained.

Finally, the potential [NMycN-myc$_{ss}$] and [MAX$_{ss}$] concentrations can exist within a broad spectrum of values. The calculated product of these two was found to be $16\mu$M$^2$, indicating that the minimum total concentration is $8\mu$M. This is both a potentially high value, but it also presents a degree of freedom that can potentially change the significance of the related reactions. It was decided that different sets of values for NMycN-myc$_{ss}$ and MAX$_{ss}$ were to be tested and these ranged from [MAX$_{ss}$]/[MYCN$_{ss}$]=64 to [MYCN$_{ss}$]/[MAX$_{ss}$]=64. Table 3.6 presents the relevant parameter values that were affected under these conditions. In order to have the system with one of these parameter sets we would need to change the $k_2$ and initial [MAX].
Table 3.9: The parameter sets used for the different potential amounts of $[\text{N-Myc}]$ and $[\text{MAX}]$ at steady state conditions. The first two columns show the steady state concentrations of free NMycN-myC and MAX and $k_2$ is the translation rate required to achieve this $[\text{N-Myc}]_{ss}$ assuming that MYCN RNA reaches a steady state value of $3.3e^5\mu\text{M}$. The columns indicating the total values of N-Myc and MAX take into account the bound species for the NMyc_MAX and the NMyc_MAX_Promoter species (the dimer bound to a transcription initiation site). The last column is the ratio of the free N-Myc and MAX proteins at steady state and it will be used as the identification number for describing each set of these parameters in subsequent simulations and analysis.

<table>
<thead>
<tr>
<th>$[\text{NMycN-myC}_{ss}]$</th>
<th>$[\text{MAX}_{ss}]$</th>
<th>$k_2$ (1/h)</th>
<th>Total $[\text{NMycN-myC}]$</th>
<th>Total $[\text{MAX}]$</th>
<th>$[\text{NMycN-myC}<em>{ss}]/[\text{MAX}</em>{ss}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>32</td>
<td>$6.00*10^3$</td>
<td>0.71</td>
<td>32.21</td>
<td>0.015625</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>$1.20*10^4$</td>
<td>1.21</td>
<td>16.21</td>
<td>0.0625</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>$2.40*10^4$</td>
<td>2.21</td>
<td>8.21</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>$4.80*10^4$</td>
<td>4.21</td>
<td>4.21</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>$9.60*10^4$</td>
<td>8.21</td>
<td>8.21</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>$1.92*10^5$</td>
<td>16.21</td>
<td>1.21</td>
<td>16</td>
</tr>
<tr>
<td>32</td>
<td>0.5</td>
<td>$3.94*10^5$</td>
<td>32.21</td>
<td>0.71</td>
<td>64</td>
</tr>
</tbody>
</table>

3.5: Model Validation

3.5.1: Single copy gene MYCN

Once the model was constructed and parameterised, it was tested for validation purposes at both MYCN non-amplified and MYCN amplified conditions. The method of choice was to examine the model behaviour within a 24 hour time-course simulation using the parameter sets from Table 3.6. The COPASI software (Hoops et al., 2006) was used in order to perform these simulations and comparisons were made to relevant results found in the literature. In this case, the study by Fernandez and colleagues investigated the E-Box/promoter occupancy in cell-lines with and without MYC amplification and the results from that study would be used to validate the model. Specifically, they reported that in a cell-line without MYC amplification, the percentage of occupied promoters with the E-Box sequence was at 58%, while
for a cell-line with MYC amplification the percentage rose to 99%, while also a number of seemingly random promoters were also bound by the myc-MAX protein dimer (Fernandez et al., 2003).

As illustrated in Figure 3.2A-F, in all cases the species' concentrations approach a steady state within 24 hours. For the MYCN mRNA levels the different parameter sets have no effect as expected since it is not affected by any of the changed parameters. The biggest shifts in values are seen with N-Myc and MAX proteins that are inversely related, as expected from the calculations. Finally, for the promoter-bound dimer and free promoter, the only difference is in the time needed to reach the same point. These differences were relatively small and all the progressions fell within the time-frame of this simulation. These results show that in terms of validation, when simulated with non-amplified MYCN the model exhibits a quantitative behaviour that corresponds to the reported results as found in the literature in terms of E-Box occupancy (Fernandez et al., 2003). However, it would also be necessary to validate the model’s behaviour in situations with MYCN amplification and compare with the results reported from the same study. This verifies that the model is functional while all the species concentrations approach the calculated steady state values within 24 hours.
Figure 3.6A-F: The concentration of the model species as they change during a time-course simulation. The model time is 0 to 24 hours, while the parameter sets are defined by the [N-myc]/[MAX] ratio as illustrated in the figure key. The MYCN RNA levels are constant for all of the simulations, while the final levels of N-myc and MAX proteins are inversely related. For the NMyc_MAX dimer, NMyc_MAX_Promoter and Promoter species the only change is the shift in the time-frame needed to reach the same point. Overall, all species are showing to be approaching the steady-state values within 24 hours.

A) [RNA]  
B) [NMyC]  
C) [MAX]  
D) [NMyc_MAX]  
E) [NMyc_MAX_Promoter]  
F) [Promoter]  

Key:  
- 64  
- 16  
- 4  
- 1  
- 0.25  
- 0.0625  
- 0.015625
3.5.2: MYCN gene copy-number amplification

These results showed that in terms of validation, when simulated with non-amplified MYCN the model exhibits a quantitative behaviour that corresponds to the reported results as found in the literature in terms of E-Box occupancy (Fernandez et al., 2003). However, it would also be necessary to validate the model’s behaviour in situations with MYCN amplification and compare with the results reported from the same study. For this purpose, we simulated for the 6 parameter sets presented in Table 3.6 (and illustrated by the colour code of the key of Figures 3.3, 3.4 and 3.5) the effect of increasing MYCN GCN from 2 to 2000 per cell (6.64*10^-7 to 6.64*10^-4 μM concentration) on the final values of NMyc_MAX_Promoter at 24 hours. The plot from Figure 3.3 illustrates the outcome of this investigation.

The results of this simulation indicate that amplification of MYCN at the genetic level under the different theoretical conditions could not result in itself in the same outcome in terms of N-myc activity. The relative amounts of MAX protein are as important in this model since they impose the limitation as to how much NMyc_MAX dimer can be formed and subsequently how much promoter binding can be achieved by the same amounts of N-myc protein. When the parameter values used predict high levels of MAX, then the outcome of even minor amplification becomes almost complete occupation of the E-Box promoters. In this case, varying MYCN GCN from 2 to 20 per cell the percentage of occupied promoters at 20 MYCN GCN and high MAX is 92% of total promoter concentration ([N-myc ss]/[MAX ss] = 0.015625), while the equivalent value for low MAX is 79% ([N-myc ss]/[MAX ss]=32) (Figures 3.4 and 3.5). This suggests that even relatively minor
MYCN amplification (up to 20 GCN) can have a much higher impact than MYCN amplification beyond that limit (>20 GCN), regardless of the parameter sets or conditions.

Figure 3.7: The values of [NMyc_MAX_Promoter] at the end of a 24 hour time-course with MYCN DNA amplification from 2 to 2000 GCN per cell (6.64*10^{-7} µM to 6.64*10^{-4} µM concentration). This simulation was performed with the parameter sets of Table 3.6 that defined different amounts of available MAX protein for the system and are colour coded as illustrated by the key (the ratio of [N-myc_{ss}]/[MAX_{ss}] at steady state without MYCN GCN amplification).
Figure 3.8: The values of [NMyc_MAX_Promoter] at the end of a 24 hour time-course with MYCN DNA amplification from 2 to 20 GCN per cell (6.64*10^-7 µM to 6.64*10^-6 µM). This simulation was performed with the parameter sets of Table 3.6 that defined different amounts of available MAX protein for the system and are colour coded as illustrated by the key (the ratio of [N-myc]s/[MAXs] at steady state without MYCN GCN amplification). At low levels of available MAX in the system (high [N-myc]s/[MAXs] ratio), the number of occupied promoters becomes lower even with the same amplification at the genetic level.

Figure 3.9: The percentage of occupied promoters by N-myc/MAX dimer in relation to total available E-Box promoters after a 24-hour time-course at 20 GCN of MYCN per cell. With excess MAX present, the production of higher amounts of N-Myc protein has a bigger impact on the binding of E-Box promoters, with a difference of 13% between the two extreme sets of parameters.
Furthermore, in terms of validation of the model, the qualitative behaviour exhibited in terms of occupancy of the E-Box promoters by the N-Myc/MAX protein dimers is different than what is reported by Fernandez and colleagues (Fernandez et al., 2003). In their study, they reported that for the cell-line used with MYC amplification and myc protein overexpression the level of E-Box occupancy was at 99% plus there was further occupancy of other, seemingly random promoters. In the case of the model constructed for this study, the observed result is that only with the parameter set that has MAX protein levels much higher that N-Myc protein levels the level of E-Box occupancy approaches, but does not reach, the one reported by Fernandez and colleagues. What makes this difference potentially significant is that firstly the published study reported not only complete occupancy of the E-Box promoters, but also of other, random promoters while in this case the dynamics of the model do not allow for a qualitatively similar behaviour. Furthermore, this model behaviour was observed while having much higher levels of MAX protein than N-Myc protein, something that comes in contrast with what is reported in the literature for neuroblastoma cell-lines (Raschella et al., 1994). It has to be noted though, that the results from the study by Raschella and colleagues have to be taken with caution due to the lack of clarity on their published figures, even though they are partly in agreement with the report by Blackwood and colleagues (Blackwood et al., 1991). As such, we can conclude that the model in its present form is qualitatively not in agreement with already published results found in the literature, at least in terms of cell-lines with MYCN amplification and over-expression, and it requires improvements.
A number of factors could be responsible for the observed discrepancy between the model and the published results. The three most important ones are perhaps the relative simplicity of the model, the model being wrong in terms of either the portrayed interactions or the assumed quasi steady-state, and finally that at least some of the parameters used, as found in other published studies, are incorrect. For the purposes of this project, and falling in line with the associated experimental aspects, the quantitative data obtained for MYCN at the DNA, RNA and protein level would be used to refine the model and its parameters and thus investigating one of the possible reasons for the model being in agreement with previously published data.

3.5.3: Reaction fluxes over time

The results of monitoring the fluxes through the model reactions during the same time-course are presented in Figure 3.6A-H. Starting with transcription and RNA degradation, it was verified that the specific parameter sets tested for these time-course simulations have no effect at this part of the model with the flux through these steps being unaffected by the parameters changed. Regarding N-myc protein translation and degradation, the fluxes increase with the parameters favouring higher production of MYCN protein and in both cases they approach a steady state within the time frame of the simulations. The flux through the promoter binding step peaks early (approximately within the 2-hour mark), before approaching zero by the end of the simulation.
The most unexpected news from these simulations was the results of the flux and its changes through time for the dimerisation reaction between the N-Myc and MAX proteins. The simulation shows continuous fluctuations during the period of this time-course, with a different pattern shown for each parameter set. The reason behind this behaviour is that the model is “stiff”, where the time-scale of related unidirectional reactions is significantly different. In this case these reactions are the translation reaction and the forward step of the dimerisation between N-Myc and MAX proteins. The fluctuations exhibited are due to the simulation either over or undershooting the reaction flux between the different time-points and then attempting to correct the error before the next time-frame of the simulation. This numeric instability ceases to exist once the model is simulated for more than 24h and it approaches steady state.
Figure 3.10A-F: The fluxes through the model reactions during a 24-hour time-course with the model defined under different sets of parameters. The transcription flux and RNA degradation fluxes are constant with all parameter sets, while the fluxes of the translation and protein degradation reactions increases as the system is defined as having more N-myc protein. The promoter binding flux reaches a peak at approximately 2 hours and then drops to zero, with the only difference being the height of the peak. Finally, the flux through the dimerisation reaction shows continuous perturbations that differ between models, suggesting that this is a potentially unstable reaction during at least part of this time-frame. The perturbations are due to the level of difference between the rate of dimerisation and the rates of transcription and translation, with the former being much faster.
3.6: Metabolic Control Analysis (MCA) and Sensitivity Analysis

Metabolic Control Analysis is a mathematical methodology for analysing how system variables are affected by its parameters. In order to do so it takes into account all the reactions of a model and calculates how the concentrations of the variables in the system and the fluxes have an impact on each other. As such, it is a useful tool for getting information on the global and local properties of a system.

The MCA provides 3 types of coefficients associated with various aspects of the model. The Concentration Control and Flux Control coefficients describe the network properties of the model while the Elasticity coefficient describes how reactions change locally according to perturbations of the system. The Flux Control Coefficient $C^J_p$ shows how much a parameter $p$ controls a flux $J$ at steady state, assuming all other parameters of the system are unchanged. In the same way, the Concentration Control Coefficient $C^X_p$ shows how much control a parameter has on the concentration of a variable, again assuming that all the rest of the system parameters stay constant. Since the change of any parameter potentially has an effect anywhere within the system and these calculations describe the effects, they are considered to be providing information about the system properties of the network. Finally, Elasticity Coefficients $e^v_p$ show how much the change of any metabolite $X$ or parameter $p$ affects the rate $v$ of a reaction it is involved in. Due to the fact that this considers only directly associated variables it is considered an indicator of local properties of the system. The equations for calculating all 3 coefficients are shown in Table 3.7, while it is also worth noting that these coefficients are linked with
summations theorems as presented in equations 3.1 and 3.2 (Westerhoff and Kell 1987).

<table>
<thead>
<tr>
<th>MCA function</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flux Control Coefficient</td>
<td>$C_P^f = \left( \frac{d \ln J}{d \ln P} \right)^{xz}$</td>
</tr>
<tr>
<td>Concentration Control Coefficient</td>
<td>$C_P^c = \left( \frac{d \ln X}{d \ln P} \right)^{xz}$</td>
</tr>
<tr>
<td>Elasticity Coefficient</td>
<td>$\varepsilon_X^{ij} = \frac{d \ln v_i}{d \ln X}$ or $\varepsilon_P^{ij} = \frac{d \ln v_i}{d \ln P}$</td>
</tr>
</tbody>
</table>

\[
\sum_{i=1}^{n} C_{ei}^{ij} \ast \varepsilon_{X_{ik}}^{ij} = 0 \quad \text{(Equation 3.1)}
\]

\[
\sum_{i=1}^{n} C_{ei}^{Ri} \ast \varepsilon_{X_{ik}}^{Ri} = -\delta_{ik}^{ij} \quad \text{if } j = k \text{ then } \delta_{ik}^{ij} = 1, \text{ in all other cases it equals to zero} \quad \text{(Equation 3.2)}
\]

Metabolic control analysis (MCA) and sensitivity analysis were performed using the in-built COPASI functions in order to examine the system properties of the model. From the results already presented, it is obvious that when reactions with 0 flux at steady state are involved, then the expected outcome will be either 0, $-\infty$ or $+\infty$. COPASI allows the projection of “unscaled” values for each function of the MCA, however this would differentiate the values and the underlying calculations from the classic description of the functions as well as potentially provide false impressions with values seemingly too high or low.
With this analysis we examined how the MCA coefficients change, if at all, with the different parameter sets shown in Table 3.6. Starting with the elasticity coefficients, the only changes observed were for the Dimerisation flux and the effect from N-myc protein, MAX protein and the NMyc_MAX dimer, with the values obtained having very high absolute values (in the region of $10^{15}$). This is due to the dimerisation reaction reaching equilibrium, where in such a case the elasticities have a value of infinity since the flux is zero. This is something that can also cause numerical instability in the model, as shown in section 3.4.3.

The Flux Control Coefficients didn’t show any significant changes for any reactions between different parameter sets other than for the flux of dimerisation. The effect of change of flux of transcription, translation, RNA degradation and protein degradation are shown to have varying effects on the steady state flux of dimerisation based on the set of parameters used. Perhaps the most interesting feature is that again the values can go from positive to negative and vice versa (Figure 3.7). Additionally, when MAX is present at the highest levels then the effect that the change of flux of these reactions can have is much higher due to the lower levels of N-myc protein. On the contrary, when N-myc is present in high concentration then a change of flux for the aforementioned reactions has minimum impact on the dimerisation flux.
Figure 3.7: The Flux Control Coefficients for the dimerisation reaction in relation to change of flux at steady state for transcription, translation, RNA degradation and protein degradation with different parameter sets. The x-axis is at a logarithmic scale in order to allow for better representation of the results.

With Concentration Control Coefficients, the biggest change is observed for MAX, where the higher the \([\text{N-my}c_{ss}] / [\text{MAX}_{ss}]\) ratio the bigger the effect is, with transcription and translation having a negative value and RNA degradation and protein degradation having a positive value. Regarding the N-my_c_MAX dimer, the NMyc_MAX_Promoter and the Promoter, the effects shown by these four reactions are much less pronounced. For most of the parameter sets, the values remain constant (albeit at different levels) while they start converging towards zero with the higher \([\text{N-my}c_{ss}] / [\text{MAX}_{ss}]\) ratios (Figure 3.8A-D).
Figure 3.8A-D: The concentration control coefficients of transcription, translation, RNA degradation and protein degradation on MAX, NMyc_MAX, NMyc_MAX_Promoter and Promoter. The biggest changes in these values are found in relation to the MAX protein. For the other, the coefficients remain at approximately constant levels until N-myc ss starts reaching higher levels than MAX ss, at which point the various coefficients start converging to zero. In all cases, the coefficients of transcription were the same as the ones of translation, while the same behaviour was observed between the coefficients of RNA degradation and the ones of Protein Degradation.
A sensitivity analysis for NMyc_MAX_Promoter was performed in relation to the reactions of the model in order to obtain the relevant response coefficients. The effect of the promoter binding reactions (forward and reverse) was found to be almost completely unchanged by the different sets of parameters defining the model. On the other hand, NMyc_MAX_Promoter was shown to be becoming less sensitive to the rest of the reactions with higher [N-mycss]/[MAXss] ratios. From those reactions transcription, translation and dimerisation (forward) had positive values at 0.43 or lower while RNA degradation, protein degradation and dimerisation (reverse) had negative values at -0.43 or higher (Figure 3.9).

Figure 3.9: Sensitivity Analysis and the response coefficients for NMyc_MAX_Promoter in relation to the model reactions. Other than the promoter binding reactions (forward and reverse) that stay practically at constant levels, NMyc_MAX_Promoter becomes less sensitive to the rest of the reactions with the higher amounts of N-mycss protein available to the system. The x-axis is at a logarithmic scale in order to allow for better representation of the results. In this case, the same behaviour was exhibited by the coefficients for translation, dimerization (forward) and transcription on one hand and protein degradation, dimerization (reverse) and RNA degradation on the other.
3.7: Conclusions

A model was built that describes the pathway of MYC from the genetic to the activity level in the form of N-Myc/MAX dimer bound to the E-Box-related promoter (section 3.2). The model was assigned to a quasi steady-state rather than any specific part of the cell-cycle as depicted in figure 1.5. The quasi steady-state was chosen for practical reasons that are primarily related to experimental considerations, both in terms of this project but also in terms of the relevant reports found in the literature. The shortcomings of this decision were acknowledged and trying to eliminate these negative aspects would be part of the long-term aims of this project (Section 3.3).

The rate laws were described using ordinary differential equations and mass-action kinetics (section 3.4.1). Most of the model parameters were taken from the literature, while the transcription and translation rates were calculated according to assigned steady-state values for MYCN mRNA and MAX protein. A degree of uncertainty existed with regards the potential values of N-Myc and MAX proteins at steady-state (and linking to that the rate of translation for the N-Myc protein) and a range of potential values was calculated for all involved parameters and variables (section 3.4.2).

A number of simulations were performed using the defined parameter sets from Table 3.6. In terms of validating the model, this was done by performing time course simulations and comparing the level of E-Box occupancy by the N-Myc/MAX
protein dimers as shown by the model with what is reported in the literature (Fernandez et al., 2003). Testing all the different parameter sets from Table 3.6, with single-copy MYCN gene the results from the simulations matched the ones reported in the study by Fernandez and colleagues. However, with increased MYCN gene copy-numbers, the level of E-Box occupancy by N-Myc/MAX protein dimer reaches a plateau at approximately 92% and this is qualitatively different from the reported values and general behaviour. The model in its current state predicted that the levels of MAX protein, at least in relation to the N-Myc protein, affect the degree of promoter occupancy by the N-Myc/MAX dimer, which is contrast with what is suggested in the literature (Raschella et al., 1994). Potential reasons for these discrepancies between the model behaviour and what is reported or suggested in the literature were identified, making the acquisition of in-house experimental results that would allow refining and re-evaluating of the model of paramount importance (section 3.5.2).

This behaviour was also exhibited in the range of MYCN amplification from 1 to 10 gene copy-numbers per haplotype. Additionally, the model simulations suggested that the effect of MYCN amplification at lower levels (MYCN GCN ≤ 20) is more pronounced than the effect of amplification beyond that point (MYCN GCN≥20), with the majority of E-Box promoters already occupied by the N-Myc/MAX dimer when MYCN GCN reaches 20 per cell (section 3.5.2).

From the reaction fluxes of the reactions, the N-Myc/MAX dimerisation step showed numerous fluctuations before reaching a stable value at approximately 0 for all the
tested parameter sets. This was due to the stiffness of the model where the rates of translation and forward dimerisation reactions are different by a large margin, resulting in numerical instability (section 3.5.3).

In terms of network properties, the system was analysed using the Metabolic Control Analysis frame of algorithms, as well as with sensitivity analysis for the levels of the bound promoter in relation to the model reactions. All the results indicated that the relative levels of MAX protein to N-Myc protein at steady state conditions have an effect in the behaviour and properties of the model. This, along with the validation results, strongly suggested that it would be necessary to establish these values experimentally before any conclusions can be drawn regarding the behaviour of the model and its system properties (section 3.6).
4.1: OVERVIEW

The focus of this project is the MYCN oncoprotein and the significance of its expression in NB, especially in relation to the MYCN gene copy-numbers (GCN). A bottom-up systems approach was chosen as the method of investigating the subject and a mathematical model was built and analysed as presented in Chapter 3. To provide accurate parameters for the model, experimentally obtained values for MYCN DNA, mRNA, and protein were generated that would refine the model and validate the parameters of some of the model reactions that were obtained from the literature.

In this chapter, the aim is to acquire absolute values for the MYCN DNA and mRNA molecules per cell in four neuroblastoma cell-lines. The method used for data generation was qPCR, while for the analysis of the DNA data the methodology described in Chapter 1. The novelty of this suggested methodology is the use of relevant and appropriate statistical methods for comparison of the acquired qPCR data from the potential house-keeping genes (HKGs) across all the cell-lines to be examined, in order to find statistically suitable HKGs that can then be used for absolute quantification of GOI without the use of reference samples. Finally, multiple HKGs were used for the quantification at both the DNA and mRNA level to
provide higher reliability to the results, as previously suggested (Vandesompele, De Preter et al. 2002).

4.2: QUANTIFICATION OF MYCN GENE COPY NUMBER

4.2.1: Cq values, test efficiencies and data transformation

The genes to be tested other than MYCN were CD200, PolR2D, GPR15 and TP53BP2. CD200 is a gene encoding a type-1 membrane glycoprotein and it is located at 3q12-q13, PolR2D encodes the polymerase (RNA) II (DNA directed) polypeptide D and the gene is located at 2q21, GPR15 encodes the G-protein coupled receptor 15 and the gene is located at 3q11.2-q13.1 while TP53BP2 encodes the tumor protein 52 binding protein 2 that is located at 1q41. The efficiencies of the qPCR tests for each gene were calculated after triplicate analyses and were found to be 1.74 for MYCN, 2.10 for CD200, 2.07 for GPR15, 1.86 for POLR2D and 1.91 for TP53BP2. The efficiency values for the HKGs fell within the accepted empirical limits, while the value for MYCN was slightly lower.

Following the suggestion by Livak and Schmittgen, but incorporating the specific efficiency from each gene, the $E^{Cq}$ formula was used to transform the data for each cell-line (Livak and Schmittgen 2001). Cq stands for quantification cycle, the point where fluorescence surpasses a specific threshold limit (or DNA equivalence point) and is optically measured by the qPCR equipment. It is a term suggested by Bustin and colleagues to provide a standard name to be used across the qPCR field rather than different commercial terms that were in use up to that point (Bustin et al.,
The Cq values obtained are shown in Table 4.1, while the results of the numerical analysis of the data for the HKGs across all cell lines, both at Cq and “relative concentration” level are presented in Table 4.2. CV stands for “coefficient of variance” and is a measure of the distribution of the data set around their calculated mean. As such, a lower CV value indicates a more accurate (or less dispersed) data set.

Table 4.1: The Cq values obtained from all the cell-liners and genes used in this study.

<table>
<thead>
<tr>
<th>Cell-line</th>
<th>Gene</th>
<th>MYCN</th>
<th>CD200</th>
<th>GPR15</th>
<th>POLR2D</th>
<th>TP53BP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHEP1</td>
<td></td>
<td>36.07</td>
<td>26.83</td>
<td>27.55</td>
<td>28.54</td>
<td>26.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>36.07</td>
<td>26.71</td>
<td>27.51</td>
<td>28.58</td>
<td>26.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>36.23</td>
<td>26.80</td>
<td>27.55</td>
<td>28.65</td>
<td>26.77</td>
</tr>
<tr>
<td>Kelly</td>
<td></td>
<td>26.2</td>
<td>26.98</td>
<td>27.53</td>
<td>29.87</td>
<td>26.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26.29</td>
<td>26.97</td>
<td>27.34</td>
<td>29.73</td>
<td>26.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26.25</td>
<td>26.94</td>
<td>27.48</td>
<td>29.70</td>
<td>26.51</td>
</tr>
<tr>
<td>NB-EB</td>
<td></td>
<td>34.88</td>
<td>26.37</td>
<td>26.99</td>
<td>28.48</td>
<td>25.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td>34.78</td>
<td>26.55</td>
<td>27.03</td>
<td>28.24</td>
<td>25.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>34.9</td>
<td>26.49</td>
<td>26.94</td>
<td>28.29</td>
<td>25.82</td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td></td>
<td>35.85</td>
<td>27.12</td>
<td>27.82</td>
<td>28.83</td>
<td>26.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35.66</td>
<td>27.25</td>
<td>27.59</td>
<td>29.29</td>
<td>26.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>36.23</td>
<td>27.36</td>
<td>27.60</td>
<td>28.79</td>
<td>26.67</td>
</tr>
</tbody>
</table>
Table 4.2: Numerical analysis of the obtained data. The values were both at Cq and “relative concentration” format that was obtained via the $E^{-Cq}$ transformation, using the specific efficiency associated with each HKG’s test. CV stands for “coefficient of variance” and is a measure of the distribution of the data set around their calculated mean. As such, lower CV indicates a more accurate (less dispersed) data set.

<table>
<thead>
<tr>
<th>HKG</th>
<th>CD200</th>
<th>GPR15</th>
<th>PolR2D</th>
<th>TP53BP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>26.86</td>
<td>27.41</td>
<td>28.91</td>
<td>26.44</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.32</td>
<td>0.30</td>
<td>0.62</td>
<td>0.40</td>
</tr>
<tr>
<td>CV</td>
<td>1.21</td>
<td>1.08</td>
<td>2.16</td>
<td>1.50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HKG</th>
<th>CD200</th>
<th>GPR15</th>
<th>PolR2D</th>
<th>TP53BP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.21E-09</td>
<td>2.18E-09</td>
<td>1.61E-08</td>
<td>3.71E-08</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>5.07E-10</td>
<td>4.77E-10</td>
<td>5.47E-09</td>
<td>9.79E-09</td>
</tr>
<tr>
<td>CV</td>
<td>22.98</td>
<td>21.86</td>
<td>33.95</td>
<td>26.37</td>
</tr>
</tbody>
</table>

4.2.2: Analysis and comparison of the HKGs

Firstly, the type of distribution for the complete dataset of each HKG needed to be established (n=12, 3 technical replicates per cell-line) to choose between parametric and non-parametric tests. This was achieved using the Saphiro-Wilk test that showed HKGs GPR15 and TP53BP2 did not have a normal distribution across the 4 cell-lines and as such the non-parametric tests were selected.

The means of the HKGs across all cell lines were tested for significant difference at p=0.05 using the Kruskal-Wallis test (n=12 per HKG) and the test showed that there was a significant difference between these means at this selected p value. This
indicated that at least one of our HKGs is potentially mutated at the DNA level and should potentially be excluded for calculations regarding the GCN of MYCN. However the Kruskal-Wallis test did not provide specific information regarding comparisons between separate pairs of HKGs, so additional testing was needed.

For this purpose comparisons were performed between all possible pairs of the HKGs using the two-sample Kolmogorov-Smirnov test and as such establish specifically which of the HKGs’ means were significantly different from each other. This test was chosen over the Mann-Whitney since it also accounts for different distributions among the compared datasets, which was the case with our data. The statistical comparison showed that the HKG pair of CD200 and GPR15 did not have significantly different means while all other HKGs combinations did (Table 4.3), suggesting that they could be used as the HKGs of this study.

<table>
<thead>
<tr>
<th></th>
<th>CD200</th>
<th>GPR15</th>
<th>PolR2D</th>
<th>TP53BP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD200</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>GPR15</td>
<td>0.87</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>PolR2D</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>TP53BP2</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>--</td>
</tr>
</tbody>
</table>

The result of this comparison provided evidence that the CD200 and GPR15 HKGs were both single-copy genes per haplotype. Strictly speaking, the tests have shown that the means from all the readings of CD200 and GPR15 across these
neuroblastoma cell-lines have no statistically-significant difference, even if their distribution is not the same. Theoretically, this points to the possibility of both genes having exactly the same levels of mutation in all of the cell-lines (gain or loss), in part due to their close proximity in terms of chromosomal location. This possibility is unlikely due to the low level of mutations reported in the chromosomal location 3q in neuroblastoma, that are also mapped outside of the location of the 2 HKGs (Satge, Moore et al. 2003). An alternative hypothesis would be that these two genes are not affected in the cell-lines under examination and can be considered to be truly single-copy genes. In such a case, where there can be confidence about both the stability of the HKGs across all the samples and also their precise copy-number per cell, this knowledge can be used to simplify the calculations for absolute quantification of the MYCN GCN without the need for either a standard curve or a reference sample. Karyotyping the cell-lines used in this study via FISH (fluorescent in-situ hybridisation) and comparing the results with more potential HKGs would help validating these findings and the stability of these two HKGs.

4.2.3: Calculating gene copy-numbers for MYCN

The MYCN gene copy-number for each cell-line was calculated in two ways. The first was by dividing the mean “relative concentration” of MYCN with the mean “relative concentration” of the CD200 and GPR15 HKGs. In addition to establish the range of potential values that could be acquired from this method, pair-wise comparisons were performed of the MYCN technical replicates (n=3) with the CD200 and GPR15 replicates (n=6). From these results the mean, standard deviation
(std) and CV were calculated for each cell-line (n=18). The results of these calculations are represented in Table 4.4.

Analysis of results from cell-lines SHEP-1, NB-EB and SH-SY5Y displayed no amplification of the MYCN gene, while the cell-line Kelly was shown to be amplified with more than 200 MYCN gene copies per haplotype. These results are in agreement with reports in the literature and thus further reinforcing the potential strength and use of this novel method.

Table 4.4: The calculated absolute MYCN GCN in the four neuroblastoma cell-lines. The HKGs CD200 and GPR15 were used for the calculations. For the pair-wise comparisons, the ratios of all the MYCN individual values against all the CD200 and GPR15 individual values per cell-line were calculated before establishing the mean, standard deviation and coefficient of variance. For the comparison of means, the average MYCN value was divided by the average value of all the CD200 and GPR15 readings. The obtained results show that the cell-lines SHEP-1, NB-EB and SH-SY5Y did not have MYCN amplification, while the cell-line Kelly is amplified for MYCN, and these observations were in agreement with various literature reports.

<table>
<thead>
<tr>
<th>Cell-line</th>
<th>MYCN GCN</th>
<th>Std</th>
<th>CV</th>
<th>MYCN GCN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHEP1</td>
<td>0.95</td>
<td>0.09</td>
<td>9.47</td>
<td>0.95</td>
</tr>
<tr>
<td>Kelly</td>
<td>233.16</td>
<td>11.67</td>
<td>5.00</td>
<td>232.93</td>
</tr>
<tr>
<td>NB-EB</td>
<td>1.40</td>
<td>0.08</td>
<td>5.71</td>
<td>1.39</td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>1.34</td>
<td>0.21</td>
<td>15.86</td>
<td>1.33</td>
</tr>
</tbody>
</table>
The values for the MYCN GCN calculated by pair-wise comparisons with the HKGs are practically identical to the ones obtained by simply calculating the ratio of the means. The CV is within biologically acceptable limits for all cell-lines, while the differences in the actual CV values stem partly from the differences in the distribution of the HKG data. Overall, this gives a better understanding of the spread of the values obtained by following this methodology and the precision of the final outcome.

4.3: QUANTIFICATION OF MYCN TRANSCRIPT NUMBERS

4.3.1: Construction of standard curves for mRNA absolute quantification

Standard curves were constructed using Template Oligonucleotides (TO) for each gene of interest to quantify the cell-lines’ cDNA. These genes were MYCN, SDHA (official full name: succinate dehydrogenase complex, subunit A, flavoprotein (Fp), gene found at chromosomal location 5p15) and HPRT1 (official name: hypoxanthine phosphoribosyltransferase 1, gene found at chromosomal location Xq26.1). As a first step for constructing these graphs, serial dilutions of the TO were performed and then assayed using a Light Cycler 480 thermal cycler as described in section 2.3. The amount of molecules loaded per reaction was calculated by transforming the concentration from moles/litre to particles/litre based on the Avogadro’s number, and then scaling-down to the number of molecules per reaction. The graphs were constructed by plotting the observed Cq values vs the $\log_{10}$(particle number) as shown in Figure 4.1.
**Figure 4.1:** A standard curve for MYCN cDNA using the MYCN template oligonucleotide. The corresponding equation of the standard curve and the coefficient of determination for the equation are shown.

4.3.2: The role of the calibrator in the absolute quantification of cDNA particle numbers

A mid-point dilution of TO (1pM or 240,000 molecules per well) for each standard curve was selected and used as a calibrator in subsequent runs. The calibrator was used to adjust the relevant equation for the conditions of the present test. To be more precise, each standard curve comes in the format of

\[ Y = AX + C \]  \hspace{1cm} (Equation 4.1)

where \( Y \) equals Cq value, \( X \) the \( \log_{10}(\text{particle number}) \), \( A \) is the slope of the graph and \( C \) is a constant that is defined by the experimental conditions.

Under ideal conditions there would be no inter-run variability, however, this is not always the case. The consistency of the tests was checked by performing replicates.
of the standard curve until the value of $A$ from Equation 4.1 was showing a standard deviation of no more than 5% (arbitrary cut-off value). For the MYCN standard curve the value of $A$ was taken as -3.86.

Once the slope $A$ was consistent within certain limits, the variability of $C$ from test to test could be accounted for with the use of a calibrator, and therefore equation 4.2 was developed:

$$C = Y - AX \quad \text{(Equation 4.2)}$$

where $Y$ and $X$ are taken from the calibrator values and the new $C$ is found that corresponds to the specific experimental conditions of the test.

### 4.3.3: Absolute quantification of MYCN mRNA

The Cq readings from each cell-line were incorporated into the Equation 4.1 and the number of MYCN cDNA particles was calculated. To find the number of MYCN mRNA particles per cell, it was necessary to take into account cell-lysis, RNA extraction and reverse transcription of mRNA into cDNA. Assuming 100% efficiency in all steps, the calculated values for MYCN mRNA molecules per cell for each cell-line are presented in Table 4.5:
Table 4.5: The calculated MYCN mRNA molecules per cell for the neuroblastoma cell-lines. These values were obtained via the use of MYCN standard curves that were constructed with the Cq values from serial dilutions of MYCN TO.

<table>
<thead>
<tr>
<th>Cell-line</th>
<th>MYCN mRNA molecules per cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHEP1</td>
<td>0</td>
</tr>
<tr>
<td>Kelly</td>
<td>$7.4 \times 10^{-5}$</td>
</tr>
<tr>
<td>NB-EB</td>
<td>$1.1 \times 10^{-5}$</td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>$3.5 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

These values were deemed to be erroneous for the reason that they are too low as absolute numbers. This could be due to any of a number of potential sources of error that have an effect on the efficiencies of cell-lysis, RNA extraction, reverse transcription and construction of the calibration graphs. To eliminate the possibility of impurities or problematic reagents for the qPCR reactions, all the reagents were re-ordered, fresh genetic material was extracted and the tests were repeated using the same protocols as described in Chapter 2. The results from the new tests were identical, with the calculated values for the MYCN RNA copies per cell being erroneously low. This suggested that potentially the problem lies with either the set-up of the tests (including the potential unsuitability of any of the reagents) or the calculations used for estimating the number of RNA particles.

Following from these findings, and after going through the calculations on several occasions to check their validity, the next potential source of error to investigate would be the concentration of the reagents used for the qPCR assay, namely the primers and the TOs. With regards to the TOs, it is imperative that the concentration of the material used for constructing the standard curve is accurately established and
in that respect we attempted to do the same with the MYCN TO to validate the concentrations suggested by the provider.

To measure the correct concentration, we needed to establish first if it falls within the detection limits of the available equipment, the NanoDrop1000 Spectrophotometer. The MYCN cDNA template oligonucleotide has 62 bases length (Appendix 1). Using the values provided by the Current Protocols webpage (http://www.currentprotocols.com/tools/dnarnaprotein-molecular-weight-calculator) this TO has a weight of 19096Da or $3.170948088 	imes 10^{-20}$ grams. The stock solution for the TOs is 10µM, which in this case translates to $3.17 	imes 10^{-4}$ ng/µL. This value was below the minimum detection limit of the available equipment and this was also validated in practise where none of the template oligonucleotide samples was giving any readings.

Due to our inability to accurately establish the concentration of TO with the available means, the decision was taken to try and eliminate the potential interference of the TO by performing relative quantification of the MYCN RNA copy numbers in relation to the ones from selected HKGs. This methodology calculates the relative amounts of the GOI RNA copy-numbers in comparison to the HKGs RNA copy numbers without the use of a standard curve from TOs.
4.3.4: Relative quantification of MYCN mRNA

To perform the relative quantification of MYCN mRNA in the neuroblastoma cell-lines, the mRNA levels of the HKGs SDHA and HPRT1 were also established using RT-qPCR. These two HKGs were selected based on previous work in the lab by Dr Jennifer Logan where a number of HKGs were tested across 9 NB cell-lines and the most stable genes were chosen with the GeNorm tool (Vandesompele, De Preter et al. 2002). Standard curves were constructed for the HKGs in order to establish the efficiency of each test using Equation 1.4 as described in Chapter 1, section 1.6. These efficiencies are shown in Table 4.6:

<table>
<thead>
<tr>
<th>Gene</th>
<th>Test efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYCN</td>
<td>1.918</td>
</tr>
<tr>
<td>SDHA</td>
<td>2.002</td>
</tr>
<tr>
<td>HPRT1</td>
<td>1.976</td>
</tr>
</tbody>
</table>

The efficiency for each gene was then used in order to transform the Cq values into “relative amounts” using the $E^{-Cq}$ formula. The mean of the two genes was established and the ratio of MYCN over the HKG mean was found for each cell-line. Finally, the MYCN:HKGs ratio obtained for each cell-line was compared with the one for SH-SY5Y and the results are shown in Table 4.7. The standard deviation was
calculated between the values from 2 biological replicates. These values are the relative mRNA values to be used with the model described in Chapter 3.

Table 4.7: Relative amounts of MYCN mRNA. The comparisons were in relation to the MYCN mRNA in the SH-SY5Y cell-line. The MYCN values were normalised against 2 HKGs (SDHA and HPRT1), while the results were obtained from 2 biological replicates.

<table>
<thead>
<tr>
<th>Cell-line</th>
<th>Relative MYCN</th>
<th>St. Dev.</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHEP1</td>
<td>&lt;0.01</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Kelly</td>
<td>237.97</td>
<td>2.15</td>
<td>0.9</td>
</tr>
<tr>
<td>NB-EB</td>
<td>10.43</td>
<td>0.32</td>
<td>3.07</td>
</tr>
</tbody>
</table>

4.4: Conclusions

The MYCN GCN was successfully measured in the 4 NB cell-lines used in this study. The qPCR data were transformed using the individual test efficiencies (section 4.2.1), while the statistical significance of the variability between the potential HKGs was tested with appropriate tests (section 4.2.2). The chosen HKGs that showed no significantly different means across all cell-lines were used for calculating the GCN of MYCN across all 4 NB cell-lines (section 4.2.3). Despite the proximal chromosomal location of the selected HKGs, the low inter cell-line variability in combination to the reported stability of the location 3q in NB would suggest that these two HKGs are truly single copy. Expanding these tests to a larger pool of potential HKGs and karyotyping via FISH the NB cell-lines used in this study would help clarify this issue and validate the results.
The novelty of the methodology for accurately establishing the gene copy numbers of MYCN in all the NB cell-lines was the use of appropriate statistical methods for correctly choosing potential HKGs that do not have a statistically significant difference in their levels across all samples to be tested. Once these HKGs were identified, they were used to calculate the GCN of MYCN across all cell-lines without the use of reference samples. The obtained results were accurate and in agreement with the cell-line profiles reported in the literature (Oude Luttikhuis et al., 2000; Edsjo et al., 2004; Chesler et al., 2006; Gheeya et al., 2009) and as such this study can be considered a successful first step in validating this novel methodology.

The absolute MYCN mRNA levels were not successfully established in the 4 NB cell-lines, due to problems that could potentially be associated with the TO used for the construction of the calibration graphs (section 4.3.3). Relative quantification was then successfully performed using 2 HKGs and the cell-line SH-SY5Y as the baseline, with the obtained relative amounts showing good agreement between biological replicates (section 4.3.4).
Chapter 5

RESULT: MEASURING N-MYC AND MAX PROTEINS

5.1: OVERVIEW

The simulation results of the model for MYCN expression from Chapter 3 showed that it would be necessary to accurately measure the levels of the N-Myc and MAX proteins, or at least their relative ratio. The degree of uncertainty of the model due to the unavailability of certain parameters as well as the absolute levels of the MAX protein, provided the possibility for a range of potential N-Myc and MAX protein values that could exist within the parameter space of the model, but would affect both the model behaviour and the system properties depending on the actual values.

The aim of this part of the project was to accurately measure the absolute levels of N-Myc and MAX proteins, ideally obtaining the corresponding values for the numbers of each protein per cell for all the neuroblastoma cell-lines used for this study (SHEP1, Kelly, NB-EB and SH-SY5Y). Two complementary methods were used, the Enzyme Linked Immunosorbent Assay (ELISA) and Sodium Dodecyl Sulphate Polyacrylamite Gel Electrophoresis (SDS-PAGE) with western bloting. These two methods were used because they offer different advantages for accurate quantification of protein expression.
ELISAs present the opportunity to include many technical replicates, allowing for statistical analysis of the acquired data. Additionally, as a technique it has high sensitivity and can detect low amounts of the protein of interest, while the results come in the form of numerical measurements. However, the specificity of the used antibodies cannot be determined exclusively via the ELISA tests, thus making the results unreliable unless the specificity of the antibody is verified by other means.

Alternatively, western blots allow for the visualisation of the detected molecular species via binding to primary antibodies used for each experiment, allowing for comparisons between samples and confirmation of the specificity of these antibodies. The main methods of visualisation of the identified, protein-containing bands are chemiluminescence and fluorescence, while various densitometry software packages have been developed for measurement of the bands’ intensity for more accurate quantification. One of the main disadvantages of western blots is the difficulty for including adequate numbers of technical replicates for performing statistical analysis of the obtained densitometry values, while the field of densitometry itself in relation to western blots is not clearly established.
5.2: Quantifying N-Myc protein levels in neuroblastoma cell-lines

5.2.1: Preliminary ELISA work

Initial ELISA experiments were performed to determine the appropriate reagents and experimental conditions that would generate reliable and reproducible results (the protocol can be found in section 2.5). All 4 anti-N-Myc antibodies listed were tested. These antibodies will be referred to from now on by the name of the company from which they were purchased- namely Abnova, New England Biolabs (NEB), Santa Cruz (SC) monoclonal or Santa Cruz (SC) polyclonal. In all experiments using these primary antibodies, the secondary antibodies used were FITC-conjugated.

Several experimental parameters were varied in an attempt to optimise the data generated using this panel of primary antibodies. However the obtained data were not reproducible, between both technical and biological replicates, with the coefficient of variance between technical replicates being at times more than 100% as illustrated in Figure 5.1. This was the case for the samples of the 4 neuroblastoma cell-lines as well as the purified N-Myc protein that was used as a positive control. The parameter conditions that were varied included 1) antibody concentrations (both primary and secondary) 2) different composition dilution buffers for the samples 3) incubation times for all the steps and 4) detection sensitivity of the plate-reader. The variability between replicates was reduced through using 200µl of Block Buffer per well instead of 50µl, diluting antibodies in Block Buffer instead of PBS (both primary and secondary antibodies) and an increase in the number of technical replicates (Block Buffer is PBS with 1% w/v BSA, section 2.5).
In order to minimise the high levels in variation observed when using FITC-conjugated secondary antibodies, a biotin-conjugated secondary antibody was used as an alternative reagent. The substitution of the FITC-conjugated secondary antibody for a biotin-conjugated secondary antibody has the advantage of including a further amplification step, namely the inclusion of incubation with 3,3’,5,5’-Tetramethylbenzidine (TMB) as a substrate of Streptavidin- Horse Raddish Peroxidase (S-HRP) that binds on the biotin-conjugated secondary antibody, which can increase the signal to noise ratio and thereby minimise the variability between data replicates.

Figure 5.11: N-Myc protein detection for the 4 neuroblastoma cell-lines using FITC-conjugated secondary antibodies. This level of variability was typical for all the experimental conditions conducted using FITC-conjugated secondary antibodies. Abnova anti-N-Myc antibody (2.5µg/ml) was used and the samples were present at a concentration of 500ng/µl total protein. A.f.u. stands for arbitrary fluorescent units, while the error bars represent the standard deviation between 4 technical replicates.
5.2.2: ELISA tests comparing Biotin-conjugated and FITC-conjugated secondary antibodies

In all experiments using the FITC-conjugated secondary antibody, the anti-N-mycMyc primary antibody from Abnova consistently gave the lowest variability between technical and biological replicates. Therefore, the Abnova primary antibody was used to compare the results obtained with either FITC or Biotin-conjugated secondary antibodies. The comparative analysis was performed using 500ng/µl of total cell lysates from each of the 4 neuroblastoma cell-lines, plus a negative control (PBS alone). The Abnova anti-N-mycMyc antibody was used at a final concentration of 2.5µg/ml (1:400 dilution) and the secondary antibodies at 1µg/ml (1:400 dilution). Each sample was loaded in 4 technical replicates (this was the case for all samples analysed by with ELISA in this chapter).

Figure 5.2 shows a marked difference between the results obtained with the FITC-conjugated and the Biotin-conjugated secondary antibodies. The coefficient of variance (CV) with the FITC-conjugated secondary antibody was above 100% while with the biotin-conjugated secondary antibody it was less than 25%, with a repeat of the same experiment confirming those results. A standard curve was also produced using purified N-mycMyc protein, the Abnova anti-N-Mmyc primary antibody and the biotin-conjugated secondary antibody. In contrast to the results obtained when using the FITC-conjugated secondary antibody, the serial dilutions of the sample gave clear strong measurements and the reproducibility of the technical replicates was high (CV less than 4%) (Figure 5.3). Based on these results, all subsequent ELISA experiments were performed with the Biotin-conjugated
secondary antibody. In all experiments using the FITC-conjugated secondary antibody, the anti-N-Myc primary antibody from Abnova consistently gave the lowest variability between technical and biological replicates. Therefore, the Abnova primary antibody was used to compare the results obtained with either FITC or Biotin-conjugated secondary antibodies. Comparative analysis was performed using 500ng/µl of total cell lysates from each of the 4 neuroblastoma cell-lines, plus a negative control (PBS alone). The Abnova anti-N-Myc antibody was used at a final concentration of 2.5µg/ml (1:400 dilution) and the secondary antibodies at 1µg/ml (1:400 dilution). Each sample was loaded in 4 technical replicates (this was the case for all samples analysed by ELISA in this chapter).

Figure 5.2 shows a marked difference between the results obtained with the FITC-conjugated and the Biotin-conjugated secondary antibodies. The coefficient of variance (CV) with the FITC-conjugated secondary antibody was above 100% while with the biotin-conjugated secondary antibody it was less than 25%, with a repeat of the same experiment confirming those results. A standard curve was produced using purified N-Myc protein, the Abnova anti-N-Myc primary antibody and the biotin-conjugated secondary antibody. In contrast to the results obtained when using the FITC-conjugated secondary antibody, the serial dilutions of the sample gave clear measurements and the reproducibility of the technical replicates was high (CV less than 4%) (Figure 5.3). Based on these results, all subsequent ELISA experiments were performed with the Biotin-conjugated secondary antibodies.
Figure 5.12: N-Myc protein detection with ELISA in the 4 neuroblastoma cell-lines N-Myc levels were detected using the Abnova anti-N-Myc primary antibody (2.5µg/ml) and either (A) the FITC-conjugated secondary antibody or (B) the Biotin-conjugated secondary antibody (both at 2µg/ml). The samples were at 500ng/µl concentration of total protein. In (A) the CV for most samples is >100% while in (B) it is always <25%. The values were normalised against the PBS-only negative control and the error bars represent the standard deviation between 4 technical replicates.

Figure 5.13: Standard curve of purified N-Myc protein with ELISA. The test was performed using the Abnova anti-N-Myc primary antibody (2.5µg/ml) and the biotin-conjugated secondary antibody (25ng/ml), while the readings are from serial dilutions of the sample. The error bars represent the standard deviation between 4 technical replicates and the CV of the samples was less than 2%.
5.2.3: Comparison of N-Myc detection in total cell lysate and soluble-only cell-lysate

Quantification of the N-Myc protein levels in the neuroblastoma cell-lines with the biotinylated secondary antibody consistently indicated the presence of significant amounts of N-Myc protein in the SHEP1 cell-line. This was not in agreement with the N-Myc protein levels for SHEP1 reported in the literature. Several studies have shown that this cell-line does not express detectable levels of N-Myc protein (Cui et al., 2005; Bray et al., 2009; Peirce and Findley, 2009; Porro et al., 2010). In all these studies the N-Myc protein levels were determined via western blotting although the antibodies used were different than the ones utilised in this study. Furthermore, the obtained results for the MYCN mRNA levels for the SHEP1 cell-line (section 4.3.4) were near zero and significantly lower for SHEP1 than the other cell-lines (>100 times less than the MYCN mRNA levels in SH-SY5Y). Taken together, these results suggested that further optimisation should take place in order to eliminate the possibility of non-specific binding from the Abnova anti-N-Myc primary antibody.

A reason for the potential non-specific binding of the Abnova anti-N-Myc primary antibody was the presence of non-solubilised material (or cellular debris) in the full cell-lysate that could interact with the primary antibodies, even if they are monoclonal. This would lead to the high readings by the SHEP1 cell-lysate, in addition to also increasing the readings from all the cell-lines and thus giving artificially high numbers for the N-Myc protein levels. To investigate the possibility that this non-solubilised material is responsible for the SHEP1 high readings, the decision was taken to compare the results from the full lysate from each cell-line
against the results from the soluble fraction of the lysate from each cell-line that does not contain any insoluble debris. Therefore, two sets of samples were prepared from each cell-line; one was the total lysate and the other one was only the soluble part of the cell-lysate that was collected as the supernatant following a centrifugation step at 14000g for 15 minutes at 4°C. The antibody concentrations for the ELISA test, both for the Abnova anti-N-Myc primary and the biotin-conjugated secondary, were the same as tests previously described in this chapter, while all the wells were loaded with 300ng/µl of either total or soluble cell lysate.

The results showed that the projected N-Myc protein levels drop significantly in the soluble part of the cell-lysates (Figure 5.4). However the results from the SHEP1 samples were suggesting the presence of N-Myc protein at similar levels as for the NB-EB and SH-SY5Y cell-lines, both in the full-lysate and the soluble part of the lysate, making these results inconclusive regarding the specificity of detection of the Abnova anti-N-Myc primary antibody and the levels of N-Myc protein in these cell lines.
Figure 5.14: N-Myc detection in cell-lysates and soluble lysates (supernatant) of the neuroblastoma cell-lines. The test was performed using the Abnova anti-N-Myc primary antibody (2.5µg/ml) and the biotin-conjugated secondary antibody (2µg/ml), while the samples were at 300ng/µl total protein concentration. The samples marked as “soluble lysate” were from the supernatant of the full lysates after a 14k rpm, 15-minute centrifugation at 4°C. The values were normalised against the PBS-only negative control and the error bars represent the standard deviation between 4 technical replicates.

5.2.4: Investigation of N-Myc protein levels using 4 different anti-N-Myc primary antibodies

With the cell-line SHEP1 giving consistently high readings with the Abnova anti-N-Myc primary antibody, the cell lysates were tested against all 4 anti-N-Myc primary antibodies already mentioned (Abnova, New England Biolabs (NEB) and Santa Cruz (SC) monoclonal or polyclonal). The aim was to investigate if any of these antibodies gave results for SHEP1 that were more in agreement with the cDNA results obtained in section 4.3.4 and what is reported in the literature. The primary antibodies were all used at 2µg/ml except for the Abnova primary antibody (2.5µg/ml) and the samples used were from full lysates at 100ng/µl concentration of total protein.
Figure 5.15A-D: N-Myc protein detection in the neuroblastoma cell-lines with 4 different primary antibodies. The Abnova primary antibody (A) was at 2.5µg/ml and the rest of the primary antibodies (B-D) at 2µg/ml. Full-lysates were used from each sample at 100ng/µl total protein concentration. The values were normalised against the PBS-only negative control and the error bars represent the standard deviation between 4 technical replicates.
The results obtained showed that despite the differences between the results from each primary antibody, all of them were potentially binding in a non-specific manner since SHEP1 gave consistently a strong signal (Figure 5.5A-D).

### 5.2.5: Sandwich ELISA for investigation of N-Myc protein levels

Normal ELISA tests showed potentially non-specific binding from all the primary anti-N-Myc antibodies used, as indicated by the N-Myc protein detection levels in the SHEP1 cell-line. Sandwich ELISAs with all the available combinations of these primary anti-N-Myc antibodies would be performed to try and increase the specificity of the data and increase the signal-to-noise ratio. In general, sandwich ELISA requires two primary antibodies recognising different parts of the same protein to work in tandem and perform in ideal for them conditions in order to recognise said target. In contrast, normal ELISA tests only one primary antibody is used for identifying the target molecule (Table 5.1). This results in sandwich ELISA being a more specific, but less sensitive method than normal ELISA and it could be suitable for the purposes of establishing the N-Myc protein levels in the four cell lines used in this study. Sandwich ELISAs were performed with all the available combinations of the primary anti-N-Myc antibodies. This assumed that there were no common non-specific products recognised by two or more antibodies, while ideally the capture and detection anti-N-Myc antibodies would not compete for the same or adjacent binding sites on the N-Myc protein and hinder each other’s binding.
Table 5.11: The key steps for normal and sandwich ELISA using a biotin-conjugated secondary antibody. In sandwich ELISA the detection of the protein of interest requires recognition by 2 different primary antibodies, the second of which (detection antibody) must be recognised in turn by the secondary antibody. Before each step other than the first incubation (step 1) and the addition of $\text{H}_2\text{SO}_4$ there is a wash step for removal of any unbound material.

<table>
<thead>
<tr>
<th>Normal ELISA</th>
<th>Sandwich ELISA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Coating wells with sample (O/N)</td>
<td>1) Coating wells with capture anti-N-mycMyc antibody (O/N)</td>
</tr>
<tr>
<td>2) Block</td>
<td>2) Block</td>
</tr>
<tr>
<td>3) Primary antibody</td>
<td>3) Sample</td>
</tr>
<tr>
<td>4) Secondary antibody</td>
<td>4) Detection anti-N-mycMyc antibody</td>
</tr>
<tr>
<td>5) S-HRP</td>
<td>5) Secondary antibody</td>
</tr>
<tr>
<td>6) TMB</td>
<td>6) S-HRP</td>
</tr>
<tr>
<td>7) $\text{H}_2\text{SO}_4$</td>
<td>7) TMB</td>
</tr>
<tr>
<td>8) $\text{H}_2\text{SO}_4$</td>
<td></td>
</tr>
</tbody>
</table>

Sandwich ELISAs have not been previously done with combinations of the primary antibodies available for this study. According to the technical details provided by their perspective suppliers, each one of these antibodies had been raised against recombinant N-Myc polypeptides that correspond to a different part of the N-Myc protein of human origin. For the Abnova antibody this was against the 1-100 amino acid (aa) sequence, the Santa Cruz monoclonal antibody was raised against the 327-339 aa sequence, the Santa Cruz polyclonal antibody was raised against the 136-185 aa sequence and finally the NEB antibody was detecting the sequence around the lysine at location 351 of the N-Myc protein. This information suggested that there was no overlap between any combination of these antibodies due to the sides they recognise. A properly functioning sandwich ELISA with appropriate combinations of these antibodies would be expected to reduce problems regarding non-specific binding in relation to the N-Myc protein.
The primary antibody combinations were chosen based exclusively on the source of origin for each one. In this case the primary antibodies raised in mouse would only be paired with the antibodies raised in rabbit. As such the only possible combinations were Abnova (mouse) with SC polyclonal (rabbit), Abnova (mouse) with NEB (rabbit), SC monoclonal (mouse) with SC polyclonal (rabbit) and SC monoclonal (mouse) with NEB (rabbit) as presented in Table 5.2. Both antibodies from each set were used as either capture or detection antibody, resulting in two sets of readings from each pair. The antibodies were at the concentrations already mentioned (2.5µg/ml for the Abnova, 2.0µg/ml for the rest), while the samples from the full cell-lysatess and soluble cell-lysatess were at 100ng/µl total protein concentration.

<table>
<thead>
<tr>
<th>Capture Antibody</th>
<th>Detection Antibody</th>
<th>Secondary Antibody</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnova (mouse)</td>
<td>SC polyclonal (rabbit)</td>
<td>Anti-rabbit</td>
</tr>
<tr>
<td></td>
<td>NEB (rabbit)</td>
<td>Anti-rabbit</td>
</tr>
<tr>
<td>SC monoclonal (mouse)</td>
<td>SC polyclonal (rabbit)</td>
<td>Anti-rabbit</td>
</tr>
<tr>
<td></td>
<td>NEB (rabbit)</td>
<td>Anti-rabbit</td>
</tr>
<tr>
<td>SC polyclonal (rabbit)</td>
<td>Abnova (mouse)</td>
<td>Anti-mouse</td>
</tr>
<tr>
<td></td>
<td>SC monoclonal (mouse)</td>
<td>Anti-mouse</td>
</tr>
<tr>
<td>NEB (rabbit)</td>
<td>Abnova (mouse)</td>
<td>Anti-mouse</td>
</tr>
<tr>
<td></td>
<td>SC monoclonal (mouse)</td>
<td>Anti-mouse</td>
</tr>
</tbody>
</table>

The results showed that all the tests with any combination of antibodies were still producing significant readings with the SHEP1 total cell-lysate and soluble cell-lysate (Figure 5.6A-H). Additionally, in most cases the full cell-lysatess had higher
readings than the soluble cell-lysates, verifying that the material recognised by the antibodies, potentially both the N-Myc protein and random products, is partly lost with centrifugation. This was most pronounced with the NB-EB cell-lines where there is consistently at least more than a 50% drop in the readings from the full lysate to the soluble cell-lysate. Overall, based on the results obtained by that point, the decision was taken to switch to SDS-PAGE gels and western blots in order to visualise the products recognised by the antibodies used in this series of experiments.
Figure 5.16A-H: Sandwich ELISA results with all the possible combinations of the anti-N-Myc antibodies. The x-axis shows the samples tested and the y-axis the levels of N-Myc protein (arbitrary fluorescence units). The values were normalised against the PBS-only negative control and the error bars represent the standard deviation between 4 technical replicates.

Key
- Full cell-lysate
- Soluble cell-lysate
5.2.6: Detection of N-Myc protein in the neuroblastoma cell-lines using SDS-PAGE and western blots

The basic steps of this methodology involve denaturing the proteins present in the test-samples within a polyacrylamide gel containing Sodium Dodecyl Sulphate (SDS) that are then free to move and separate according to their size in the direction of the applied electrical current. Once complex protein mixtures are resolved the proteins are transferred on a nitrocellulose membrane where the proteins of interest are detected using appropriate primary and secondary antibodies. A commonly used method for signal detection is chemiluminescence and the resulting image of the protein bands recognised by the primary antibody is developed on light-sensitive film.

For the first experiment two each sample was loaded in duplicate in a 10% SDS gel. These were loaded with full cell lysate containing either 5µg or 10µg total protein amount. For the detection of the N-Myc protein on the nitrocellulose membrane the Abnova anti-N-Myc antibody was used at 1µg/ml (1:1000 dilution) while after the application of the chemiluminescent cocktail ECL the film was developed for 1 hour.

The results confirmed that a number of non-specific products are recognised with this antibody (Figure 5.7A). The intensity of the N-Myc protein band at 62kDa from all the cell-lines is qualitatively in agreement with the cDNA measurements: SHEP1 has almost no N-Myc protein, Kelly has by far the most while NB-EB and SH-SY5Y have low but detectable levels. Despite having partial similarities, the recognised
non-specific bands were not all the same between cell-lines, thus making it unlikely that they are the same products.

Subsequent experiments were done with the same methodology while using the NEB anti-N-Myc primary antibody. The antibody concentration was at 1µg/ml (or 1:1000 dilution). However, even with 25µg total protein content loaded per sample, the only band detected was the potential N-Myc band with the Kelly cell-line without any non-specific bands visible (Figure 5.7B). This indicated that the NEB anti-N-Myc primary antibody was potentially more specific but less sensitive in comparison to the Abnova anti-N-Myc primary antibody. Additionally, the difference in the results between ELISA and western blot when using the NEB antibody could be attributed to the higher sensitivity of the former method over the latter. In other words, the NEB antibody might still be exhibiting non-specific recognition, but at levels that are below the threshold of detection for western blots and above the threshold of detection for ELISA.
Figure 5.17: N-Myc protein detection with Western Blot. The Abnova anti-N-Myc primary antibody (A) and NEB anti-N-Myc primary antibody (B) were used, both at 1:1000 dilution. The loaded material was from full cell lysates of each cell-line, while in gel B 10µg total protein was loaded per lane. The bands on the left indicate the molecular-weight marker.

The software ImageJ was used in order to acquire densitometry readings from the bands present in Figure 5.7A where the Abnova anti-N-Myc primary antibody was used. The values would be used to investigate whether there was a linear relationship between the amount of total protein per lane, the amount of N-Myc protein detected and the obtained densitometry readings. The answer to this would
help establish if there is a stable ratio between the total signal and the signal that can be attributed to the N-Myc protein per cell-line regardless of the loaded total material, by doing intracellular comparisons. This would then be used to infer the N-Myc related signal for each sample on the ELISA tests. Additionally, the intercellular ratio of the N-Myc protein for each cell-line towards the amount of N-Myc protein present in SH-SY5Y would also be established by comparing the intensity of the N-Myc bands. This relative quantification method would follow on from the data analysis method used when performing relative quantification of the cDNA values, where SH-SY5Y was the baseline sample.

The results in Tables 5.3 and 5.4 show that neither of these hypotheses were correct. The ratio of N-Myc protein towards the total detected signal from all the bands per cell-line changes according to the total protein amount loaded (5µg or 10µg). This suggested that there could be no extrapolation of this ratio towards the ELISA readings since there was no stable relation between the loaded amount or concentration per method and their output. This resulted in having two readings from each cell-line, making it impossible to establish an accurate relation between the signal given by N-Myc protein and the total signal. Additionally, Table 5.4 shows the results of performing relative quantification of the amount of N-Myc protein in these cell-lines by using the reading from SH-SY5Y as the baseline. This was done in order to adopt the procedure followed in the relative quantification of cDNA values. These calculations produced significantly different results depending on the chosen total amount of loaded protein (5µg or 10µg), with all cell-lines showing a 3- or 4-times drop in the relative amounts of N-Myc protein between the values from 5µg and 10µg total protein.
Table 5.13: Quantification of the protein bands from Figure 5.7A using the ImageJ software. The columns with the Total readings and N-Myc readings show the absolute values from the relevant bands after removal of the background readings.

<table>
<thead>
<tr>
<th>Cell-line</th>
<th>Total protein amount</th>
<th>Total reading (absolute)</th>
<th>N-Myc reading (absolute)</th>
<th>N-Myc (% of total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHEP 5ug</td>
<td>6841</td>
<td>52</td>
<td></td>
<td>0.76</td>
</tr>
<tr>
<td>SHEP 10ug</td>
<td>20111</td>
<td>204</td>
<td></td>
<td>1.01</td>
</tr>
<tr>
<td>Kelly 5ug</td>
<td>10902</td>
<td>3516</td>
<td></td>
<td>32.26</td>
</tr>
<tr>
<td>Kelly 10ug</td>
<td>36007</td>
<td>16552</td>
<td></td>
<td>45.97</td>
</tr>
<tr>
<td>NBEB 5ug</td>
<td>4444</td>
<td>635</td>
<td></td>
<td>14.29</td>
</tr>
<tr>
<td>NBEB 10ug</td>
<td>11199</td>
<td>2500</td>
<td></td>
<td>22.32</td>
</tr>
<tr>
<td>SHSY5Y 5ug</td>
<td>5693</td>
<td>165</td>
<td></td>
<td>2.91</td>
</tr>
<tr>
<td>SHSY5Y 10ug</td>
<td>18998</td>
<td>2455</td>
<td></td>
<td>12.92</td>
</tr>
</tbody>
</table>

Table 5.14: The relative amounts of N-Myc protein per cell-line. These values were obtained by comparing the absolute readings from each cell-line to the readings from the SH-SY5Y cell-line. The values used for the calculations are the N-Myc reading (absolute) column from Table 5.3.

<table>
<thead>
<tr>
<th>Cell-line</th>
<th>Relative amount of N-Myc</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHEP (5ug)</td>
<td>0.32</td>
</tr>
<tr>
<td>SHEP (10ug)</td>
<td>0.08</td>
</tr>
<tr>
<td>Kelly (5ug)</td>
<td>21.31</td>
</tr>
<tr>
<td>Kelly (10ug)</td>
<td>6.74</td>
</tr>
<tr>
<td>NBEB (5ug)</td>
<td>3.85</td>
</tr>
<tr>
<td>NBEB (10ug)</td>
<td>1.02</td>
</tr>
</tbody>
</table>

The data analysis suggested the need for implementing additional steps for a more precise and consistent quantification of the N-Myc protein levels. Firstly, the construction of a standard curve from densitometry readings using serial dilutions of
the purified N-Myc protein would help establish the relation between the amount of N-Myc protein present and the measured value with each primary antibody used. The detection and saturation limits for each primary antibody would be established and help provide more confidence in the obtained values from the samples.

Additionally, new methods for western blot membrane visualisation would need to be utilised. The chemiluminescence methods like ECL have a low saturation limit that results in a small linear range. Furthermore, the heterogeneity in terms of image-development conditions can lead to substantial variability in the results, with some of the contributing factors being the film development conditions and equipment, the film scanner and the background correction methods (Gassmann, Grenacher et al. 2009). This could be a problem in this study since the expected differences in N-Myc protein expression between the different cell-lines were significant. One alternative methodology was the near-infrared fluorescence system Odyssey® Infrared System (Li-Cor, Cambridge, U.K.). This method utilises secondary antibodies that are conjugated with near-infrared dyes that provide higher sensitivity and a bigger linear range of detection in comparison to the chemiluminescence method (Wang, Wade et al. 2007). Furthermore, the images are always developed and scanned under the same conditions since both steps are done with the Odyssey® System. The visualisation software allows manually setting the filter sensitivities and scanning parameters, allowing for consistent use of the desired settings. Finally, the secondary antibodies can have different-colour fluorochromes attached on them (red or green) allowing for simultaneous detection of both the protein of interest and the protein detected for loading normalisation purposes.
5.2.7: Quantification of N-Myc protein using fluorescence-based western blots

For the first fluorescence-based western blot the Abnova anti-N-Myc primary antibody was used with 25µg of total protein content from the soluble lysate of the samples. The main reason was that at higher concentrations of total protein content, the samples were turning significantly viscous with the addition of the loading buffer and that created difficulties in terms of loading the samples in equal amounts on the gels. Furthermore, a western blot image with total cell lysate had already been acquired with the Abnova antibody (section 5.2.6) and as such it was deemed advantageous for comparison purposes to have one with only the soluble fraction of the lysate.

Additionally, serial dilutions of the purified N-Myc protein were included that would be used for constructing a standard curve in order to validate the detection and saturation limits of this method. The purified protein had a fused GST-tag at the N-terminus and a total MW of 77.11kDa (Abnova) which is a higher molecular weight than the N-Myc protein from the cells. The purified protein was loaded on the SDS-PAGE gel at serial dilutions between 20ng and 1.25ng. The bands corresponding to detection by the Abnova anti-N-Myc primary antibody (1:1000 dilution) were coloured red, while the bands from the anti-GAPDH antibody (1:2500 dilution) that was used for testing the loading efficiency would be green (GAPDH protein MW 36kDa). The colour of the bands was due to the colour of the dye attached on the secondary antibodies, with the ones available for this study being a red anti-mouse and a green anti-rabbit secondary antibodies.
The western blot image (Figure 5.8) showed that the Abnova antibody does not recognise the N-Myc protein at 20ng or lower. Additionally, the expected bands for the N-Myc protein as visualised in Figure 5.7A had much lower intensity, something that was more prominent with the sample from the Kelly cell-line. This suggested that at least part of the N-Myc protein content was present in the non-soluble part of the cell-lysate. Finally, some of the non-specific material targeted by this antibody was also lost with the lysate separation into soluble and non-soluble parts.

5.2.8: Cell lysis and processing of lysates

The results from the previous western blot indicated that the whole cell-lysate should be used for the SDS-PAGE gels since a significant amount of N-Myc protein was
lost with the exclusion of the non-soluble part of the lysate. For this purpose, a
different lysis buffer was used (RIPA Buffer, Sigma-Aldrich) and 25µg total protein
per cell line per well were prepared for loading on the SDS-PAGE gel. On adding
the sample-buffer to the lysates some of the samples became very viscous, therefore
impairing the consistent loading of the gel. Purified N-Myc protein was also loaded
on the gel, with the range of protein being from 200ng to 25ng. The NEB anti-N-
Myc primary antibody was used for testing the western blot from this gel (1:1000
dilution, green bands) along with the anti-histone H3 primary antibody (NEB) for
checking the loading efficiency of the samples (1:2000 dilution, red bands). The
results are presented in Figure 5.9.

Figure 5.19: Fluorescent western blot of the neuroblastoma samples and the N-Myc purified protein. The
NEB anti-N-Myc primary antibody was used for detecting the N-Myc protein (1:1000 dilution, green
bands) and the anti-histone 3 primary antibody (1:2000 dilution, red bands) for establishing the loading
efficiency between samples. The cell samples were from the full lysates at 25µg total protein content, while
the loaded amount of purified N-Myc protein was from 25ng to 200ng. The purified N-Myc protein is
detected at a higher MW due to the attached GST tag at the N-terminus.
The disparities in the intensity of the control protein (red bands) suggest that the efficiency loading between the samples was uneven. Additionally, only the sample from the Kelly cell-line produced a clear and visible band at the expected MW for N-Myc. However, the N-Myc protein standard was visible at the specific range (25-200ng) and densitometry measurements were acquired from the bands in order to construct a standard curve for the N-Myc protein under these specific experimental conditions. The data were fit in a straight line passing through the origin of the axis and the relevant equation and coefficient of determination were calculated. Using all 5 points the fitness of such a line with the data was relatively poor ($R^2 = 0.8537$, Figure 5.10A) while using the values from N-Myc protein at 25ng to 150ng the fitness of the line with the data was significantly higher ($R^2 = 0.9762$, Figure 5.10B).
Figure 5.10A-B: Standard curves for N-Myc protein. The values were obtained from the measurements of the corresponding bands of the fluorescent western blot with the NEB anti-N-Myc antibody (1:1000) from Figure 5.9. A straight line was generated with the acquired data that passed through the origin of the axis, with the equation and the coefficient of determination presented. Figure A) includes all the N-Myc values while Figure B) the readings from N-Myc protein from 25ng to 150ng. The lines and equations were acquired through the relevant Microsoft Excel 2007 functions.

An additional step was added for the treatment of the cell lysates, where they were subjected to vigorous pipetting in order to break down the DNA from the cells that is responsible for the viscosity of the samples. This method proved to be more successful and the gel loading did not present any problems. The samples (25µg total protein per cell-line from full lysate) were then tested with both the NEB anti-N-Myc and the Abnova anti-N-Myc primary antibodies at the same conditions. Both gels also included serial dilutions of the purified N-Myc protein.
The western blot with the NEB antibody did not recognise any bands from the cell-lysates, in contrast with the full detection of the purified protein (Figure 5.11). Additionally, the signal from the purified protein was strong enough that it had already reached the saturation point at the highest amount loaded for the N-Myc protein (200ng).

The western blot with the Abnova anti-N-Myc antibody detected a band at the expected size for the N-Myc protein for both the cell samples and the purified protein (Figure 5.12). Of concern was the appearance of a band of similar intensity between the samples that was directly below the potential N-Myc band. Additionally
the sample from SHEP1 cell-line also showed a visible band at the projected N-Myc MW.

Figure 5.12: Fluorescent western blot of the neuroblastoma samples and the purified N-Myc protein. The Abnova anti-N-Myc primary antibody was used for detecting the N-Myc protein (1:1000 dilution, red bands) and the anti-GAPDH antibody for checking the loading efficiency of the samples (1:2500, green bands). The samples were from the full cell lysates that were subjected to vigorous pipetting and the loaded amount was 25µg total protein content per sample. The loaded amount of the purified N-Myc protein was from 25ng to 200ng.

A standard curve was constructed with the results from the purified N-Myc protein of Figure 5.12. Only 4 measurements were obtained (there was no recognised band at 25ng of N-Myc protein) and a straight line going through the origin of the 2 axis was plotted whose coefficient of determination was 0.9198 (Figure 5.13).
Figure 5.13: Standard curve constructed with the purified N-Myc protein readings from Figure 5.12. The N-Myc protein was detected using the Abnova anti-N-Myc primary antibody (1:1000 dilution). A straight line passing through the origin of the axis was drawn and its equation and coefficient of determination are displayed.

\[
y = 41.795x \\
R^2 = 0.9198
\]

Using the equation of the purified N-Myc standard curve and the acquired values for the N-Myc bands from the samples, the number of N-Myc particles per cell was calculated. The calculations were based on the following procedure: the standard curve equation was first used in order to calculate the value of “x” from the standard curve of the equation for each cell-line, which corresponds to the value of N-Myc protein (ng) present in 25µg of total protein content. Additionally, the number of cells from each cell-line used for lysis, the total lysate volume and the protein concentration for each lysate were also known. Based on this information and assuming 100% efficiency for cell-lysis step (all the cells counted were lysed without any problems), the number of cells per sample that would be needed to give 25µg of total protein was calculated and from there the amount of N-Myc protein (in nanograms) per cell for each cell-line. Finally, in order to convert this value into N-Myc protein molecules per cell, the Avogadro’s Number and the molecular weight of N-Myc protein (62kDa) were used, while also accounting for the correct unit.
transformation of nanograms to grams. This value was also converted in µM in accordance to the concentration units used for the N-Myc model described in Chapter 3. The intermediate values of the calculations, as well as the final values of N-Myc in terms of number of particles per cell and concentration are given in Tables 5.5 and 5.6.

Table 5.5: The values obtained with the calculation for the N-Myc protein concentration. The “Fluorescence” column represents the value obtained for the N-Myc band in each sample after subtraction of the background readings from Figure 5.11. The values for “N-Myc (ng)” are calculated from the equation of the standard curve from Figure 5.13 and represent the total mass of N-Myc protein present in 25µg of total protein per sample. This value is used to calculate the amount of N-Myc protein (ng) per cell.

<table>
<thead>
<tr>
<th>Cell-line</th>
<th>Fluorescence</th>
<th>N-Myc (ng)</th>
<th>N-Myc (ng/cell)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHEP1</td>
<td>1704</td>
<td>40.78</td>
<td>1.1x10^{-3}</td>
</tr>
<tr>
<td>Kelly</td>
<td>10318</td>
<td>246.87</td>
<td>8.0x10^{-3}</td>
</tr>
<tr>
<td>NB-EB</td>
<td>4750</td>
<td>113.66</td>
<td>4.3x10^{-3}</td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>3382</td>
<td>80.93</td>
<td>2.7x10^{-3}</td>
</tr>
</tbody>
</table>

Table 5.6: The calculated concentration of N-Myc protein per cell. These values were obtained based on the measurements from the western blot shown in Figure 5.11 while the numbers in each column were rounded to one decimal place. The units for the concentration were chosen in accordance to the concentration units used for the N-Myc model.

<table>
<thead>
<tr>
<th>Cell-line</th>
<th>N-Myc copy number (x10^6 per cell)</th>
<th>[N-Myc] (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHEP1</td>
<td>10.6</td>
<td>35.3</td>
</tr>
<tr>
<td>Kelly</td>
<td>77.4</td>
<td>258.2</td>
</tr>
<tr>
<td>NB-EB</td>
<td>41.7</td>
<td>138.9</td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>26.4</td>
<td>88.1</td>
</tr>
</tbody>
</table>

The results obtained for the concentration of N-Myc per cell and presented in Table 5.6 should be subject to further examination. Firstly, the cell-line SHEP1 gave a
band at the expected MW level for the N-Myc protein and this was not expected for 2 reasons: the lack of detection of MYCN cDNA from this cell-line that have been observed as part of this study, as well as the various reports in the literature that show no detectable N-Myc protein levels in this cell-line, making it a suitable negative control for MYCN expression studies (Cui et al., 2005; Bray et al., 2009; Peirce and Findley, 2009; Porro et al., 2010). Additionally, the standard curve that was constructed based on 4 points, while the coefficient of determination was relatively low. Finally, there were no technical replicates for either the samples or the purified N-Myc protein used for the construction of the standard curve, nor biological replicates for the samples. This resulted in the inability to perform any meaningful statistical analysis on the data. The reasons for the lack of replicates were of practical nature. These results came at the end of this project with no extra time and limited funds being available. Thus there was no opportunity for repeating the last experiment. Otherwise, reproducing these results would have been the next step in this series of optimisation experiments, both in order to qualitatively validate the results but also in order to acquire a reasonable amount of replicates that would allow for statistical analysis and evaluation of the numerical results.

Alternatively, this was one of the first times that absolute quantification of N-Myc protein in neuroblastoma cell-lines has been attempted using the two complementary techniques described in this section. As part of the conducted research, the methodology has been refined and with specific additions and alterations, initially quality controls and increase of technical and biological replicates, it could produce accurate and statistically acceptable values for N-Myc protein levels.
5.3: Quantifying MAX protein levels in neuroblastoma cell-lines

5.3.1: ELISA-based MAX protein quantification

ELISA tests were performed using the biotin-conjugated secondary antibodies in order to quantify the MAX protein levels in the 4 neuroblastoma cell-lines. The cell lysates were at 100ng/µl and the anti-MAX antibody was used at a concentration of 2µg/ml (Figure 5.14). Additionally, purified MAX protein (Novus) was used as a positive control in serial dilutions of 1ng/µl to 0.0625ng/µl in order to construct a standard curve (Figure 5.15). Finally, the anti-MAX primary antibody used was from NEB and was at 2µg/ml concentration for the experiment.

The results showed no-significant variability between the technical replicates for both the samples and the controls. The Kelly cell-line had the highest readings for the MAX protein while the other 3 cell-lines were at similar levels between them. Additionally, the results from the serial dilutions of the purified MAX protein produced a standard curve giving a straight-line fit for the range of values we tested. The coefficient of determination for this line was high ($R^2=0.9882$).
Figure 5.14: ELISA test for MAX protein quantification. The anti-MAX primary antibody (NEB) was at 2µg/ml and the cell lysates at 100ng/µl. The values were normalised against the results of the used negative control (PBS). The error bars represent the standard deviation between 4 technical replicates for each sample.

![Graph showing MAX protein quantification](image)

Figure 5.15: Standard curve of purified MAX protein. The ELISA test was performed using the anti-MAX primary antibody (NEB) at 2µg/ml and the biotin-conjugated secondary antibody (2µg/ml). The error bars represent the standard deviation between 4 technical replicates while the CV was less than 5% for all points.

![Graph showing standard curve](image)

The experience from the N-Myc protein experiments, where the specificity of the ELISA results was not verified by western blots, also in addition to the non-specific binding demonstrated by the anti N-Myc antibodies, were the catalysts in deciding to
check the specificity of the anti-MAX primary antibody before using the values obtained by ELISA for any calculations. For that purpose SDS-PAGE gels and western blots were used in order to get a visual representation of the recognised products and check the specificity of the anti-MAX primary antibody.

5.3.2: SDS-PAGE and western blot-based MAX protein quantification

Initially a western blot with the anti-MAX primary antibody at 1:1000 dilution was used. The 12% SDS-concentration gel was loaded with 10µg total protein from the lysate from each cell-line, as well as serial dilutions of the purified MAX protein from 200ng to 12.5ng. The western blot image showed that the antibody recognised the purified protein clearly, but didn’t recognise anything from the cell lysates (Figure 5.16). Using the software ImageJ to measure the intensity of the bands from the purified protein and plotting against the amount of MAX protein we get the graphs shown in Figure 5.17A-B. A straight line based on all 5 readings and passing through the origin of the axis gave a low coefficient of determination ($R^2=0.8205$, Figure 5.17A), however on removing the point at 200ng of MAX protein, the line fitted the data in a significantly more accurate manner ($R^2=0.9835$, Figure 5.17B). This suggested that the chemiluminescence saturation point for the amount of MAX protein is between 100ng and 200ng. The main concern regarding the results of this methodology, though, was the lack of recognition from the anti-MAX primary antibody of any protein from the actual cell samples, even after repeated tests.
Figure 5.16: Western blot image for recognition of the protein MAX. Cell lysates (10μg) and purified MAX protein were tested with an anti-MAX primary antibody from NEB (1:1000 dilution). The amount of purified MAX protein loaded is indicated above the corresponding lanes.

Figure 5.17 A-B: Purified MAX-protein standard curve. The graph was based on the values obtained via densitometry readings with ImageJ of the bands of the purified MAX protein shown in Figure 5.16. The top graph (A) includes the readings for all 5 serial dilutions of the MAX protein while graph (B) includes only the readings from the lower 4 amounts of MAX protein (12.5→100ng).
Switching to the Odyssey® Infrared Imaging System, the western blots were repeated using both the full and the soluble part of the cell lysates, all times at 25µg total protein content. Additionally, each gel included serial dilutions of the purified MAX protein with the loaded amounts ranging between 12.5ng to 200ng. Another experimental parameter that was evaluated was the anti-MAX primary antibody dilution factor that was also increased to 1:500. In all cases the anti-MAX antibody would not detect anything from the cell lysates (full or soluble), while the purified MAX protein always gave a clear signal (green bands, Figure 5.18). The possibility of problematic loading or sample degradation was always tested by the anti-histone 3 antibody (at 1:2000 dilution, red bands) with no potential problems detected (Figure 5.18). Based on these results, the conclusion was that if the MAX protein is present at these cell-lines, it is at less than 12.5ng MAX protein per 25µg total protein content. Using the same calculations described for the estimation of the N-Myc protein levels, the maximum potential levels of MAX in each cell-line are presented in Table 5.7.

As a verification step, using the results for each cell line from the ELISA tests, as well as the equation obtained from the serial dilutions of the MAX protein standard from the same test (Figures 5.14 and 5.15), the amount of MAX protein per sample was calculated as ng of MAX protein per 100 ng of total protein content. This value was then scaled up to 25mg of total protein content and the results are shown in Table 5.8. The final values calculated indicate that theoretically the levels of MAX protein for all cell-lines are high enough to be detectable by western blots, both by chemiluminescent and fluorescent, since they are at worst at the same level as the detected band with the lowest amount of purified MAX protein (12.5ng). However,
with this is not being the case the possibility that the ELISA test is also taking into account non-specific products, as was the case with the N-Myc protein, was reinforced. As such, the ELISA results cannot be used in isolation and further optimisation must take place for all the methods described in this section.

**Figure 5.18**: Fluorescent Western Blot for MAX protein detection. Both full cell lysates (25µg total protein) and purified MAX protein (12.5ng\(\rightarrow\)200ng) were loaded on the gel, while the NEB anti-MAX primary antibody (1:500 dilution) was used for MAX protein detection (green bands). The anti-histone 3 antibody was used for detection of loading efficiency (1:2000, red bands).

**Table 5.7**: Maximum concentration for MAX protein per cell-line. The upper limit presented was based on the lowest amount of purified MAX protein detected by the anti-MAX primary antibody (NEB) which was 12.5ng. This was taken to be the highest potential amount of MAX in 25µg total protein content for each cell-line.

<table>
<thead>
<tr>
<th>Cell-line</th>
<th>[MAX] (x10^6 molecules/cell)</th>
<th>[MAX] (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHEP1</td>
<td>&lt;10.2</td>
<td>&lt;34</td>
</tr>
<tr>
<td>Kelly</td>
<td>&lt;12.3</td>
<td>&lt;41</td>
</tr>
<tr>
<td>NB-EB</td>
<td>&lt;14.4</td>
<td>&lt;48</td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>&lt;12.9</td>
<td>&lt;43</td>
</tr>
</tbody>
</table>
Table 5.8: The calculated values for the MAX protein present in the 4 NB cell lines according to the ELISA tests in section 5.3.1. For all cell lines, the projected values are at or above the threshold of detection of the western blots as illustrated in figures 5.16 and 5.18

<table>
<thead>
<tr>
<th>Cell-line</th>
<th>MAX protein (ng/100ng total protein)</th>
<th>MAX protein (ng/25µg total protein)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHEP1</td>
<td>0.053</td>
<td>13.25</td>
</tr>
<tr>
<td>Kelly</td>
<td>0.088</td>
<td>22</td>
</tr>
<tr>
<td>NB-EB</td>
<td>0.048</td>
<td>12</td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>0.049</td>
<td>12.25</td>
</tr>
</tbody>
</table>

5.4: Conclusions

In this chapter the optimisation of the methodology for the quantification of the N-Myc and MAX proteins was presented. The results from the ELISA methodology suggested that the anti-N-Myc antibodies potentially show non-N-Myc specific detection in both normal and sandwich ELISA tests (sections 5.2.4 and 5.2.5). This could not be verified with the ELISA method and the decision was taken to test the specificity of the antibodies using western blots.

The results from chemiluminescence-based detection methods showed that the 2 anti-N-Myc antibodies tested (Abnova and NEB) either detected a number of bands per sample (Abnova) or only a single band (NEB) from the Kelly cell-line at the expected molecular weight for the N-Myc protein. Additionally, analysis of the band intensities from the image with the Abnova anti-N-Myc primary antibody suggested that there was no linear relationship between the band intensities and the detected
protein amount (section 5.2.6). The decision was taken to switch to a fluorescence-based detection method that was reported to offer a wide linear range of signal-to-protein ratio, while using the Odyssey® Infrared Imaging System for visualising the western blots.

Following the investigation of the effects of different treatment methods for the samples used for the SDS-PAGE and western blotting methods, the amount of N-Myc protein in the 4 NB cell-lines was quantified following detection with the Abnova anti-N-Myc primary antibody. However, the lack of biological and technical replicates, as well as the relatively low quality of the standard curve used for the interpolation of the amounts of N-Myc protein, make it necessary to obtain additional replicates, both biological and technical, prior to accepting these results (section 5.2.8).

The MAX protein was firstly detected using the ELISA method, which also produced a high-quality standard curve using a purified MAX protein as a positive control (section 5.3.1). However, subsequent tests with SDS-PAGE and western blotting, both chemiluminescence and fluorescence-based, did not manage to produce any signal from any of the 4 NB cell-lines under various experimental conditions. Alternatively, the purified MAX protein gave consistently a strong signal at different amounts. Based on these results, the only numbers that could be calculated were the highest possible levels of MAX protein that is present in these cell-lines (section 5.3.2).
6.1: General Discussion

The aims of this study were to investigate the significance of and the system properties related to the expression of the MYCN oncogene in NB tumours with the use of a bottom-up systems approach. The initial model was populated by parameters obtained from the literature. Experimental values from 4 NB cell-lines at the DNA, RNA and protein level for MYCN were used to define the model and form further models for each NB cell-line. This lead to the analysis and comparisons of the results between the individual cell-lines and assisted a better understanding of any potential differences in the significance of MYCN expression and the relevant control mechanisms.

6.1.1: Modelling MYCN expression and activity as a heterodimer with MAX

As a first target of these studies, a model was constructed for the expression of MYCN and the binding of E-Box promoters by the N-Myc/MAX dimer. There were mainly two reasons for keeping the starting model relatively simple. Firstly the desire to examine the significance of MYCN expression when related exclusively to its transcriptional activity resulting from the binding of E-Box-related promoters and secondly to parameterise most of the model with values found from the literature and
produce a general model as a starting point before making adjustments according to the experimental data obtained from each NB cell-line used in this study.

For the initial model, the binding of the N-Myc/MAX protein dimer was employed as the endpoint of the model. The literature search performed when trying to find data for parameterising the model returned values regarding the E-Box promoter occupancy by the MYC/MAX protein dimer (Fernandez et al., 2003), but no specific or universal values linked either the levels of N-Myc protein or E-Box promoter occupancy to any form of quantified transcriptional activity. Including an additional step in the model without any kind of values to parameterise it could raise questions regarding the model’s function and usability, at least in the initial stages. Furthermore, once the model was constructed, parameterised with our own experimentally-obtained data and improved according to our validation results, then the addition of a transcriptional activity step would be a relatively simple task, depending on the depicted complexity. At that stage, additional Luciferase activity experiments as described in the literature could be performed to produce N-Myc activity data that would help with the proper parameterisation of the transcription step in the model (Lu et al., 2003).

With bottom-up approaches in modelling it is paramount that the associations between molecules are clearly established so that they can be described in a precise manner by the model. In the case of MYC family members in general, despite suggestions for them either having a potential role without MAX (Steiger, Furrer et al. 2008) and that when over-expressed MYC binds promoters randomly in addition
to the ones in the vicinity of an E-Box (Fernandez, Frank et al. 2003), the only clearly established function for MYC is the binding and transcriptional activation of E-Box-containing promoters in association with MAX.

The MYC proteins in general have been shown to have a complex expression pattern that is associated with the cell cycle (Figure 1.5). Ideally, a model should consider this behaviour and correspond directly to a specific state within the cell cycle and thus the specific expression levels of myc. However, the available experimental techniques and the planned experiments for this project dictated that the calculations for the parameters of the model would assume quasi steady-state conditions. In specific, the collection of experimental data regarding the MYCN DNA, RNA and N-Myc protein status would be generated from NB cells collected from unsynchronised cell culture. As such, the data would represent an average of values across all stages of the cell cycle rather than any specific stage and the initial model and its parameterisation had to reflect this condition.

Starting with the DNA value, 2 GCN per cell was chosen for a non-MYCN amplified NB cell-line. For the RNA levels there was no specific value that could be associated with any cell-line or subtype, with reported variation in MYCN RNA levels even within the same subtypes of NB patients (Cohn, London et al. 2000). Thus the arbitrary number of 100 mRNA molecules per cell was chosen that would again be adjusted according to the experimental data from each cell-line (section 3.3.2).
The levels of both the N-myc and the MAX proteins presented a degree of freedom for the potential values that could fit in the model. Originally the protein MAX was included in the model at high values that would not have a negative effect on the activity of N-myc protein. The reason for doing so was that MAX has not been taken into consideration in the studies regarding the significance of MYCN amplification and expression in NB. In practise this would make the protein MAX an un-needed molecular species from a modelling perspective. Additionally, the levels of MAX protein have not been established in a quantitative manner in NB cell-lines. Instead, they were reported to be lower in some NB cell lines following qualitative investigations via western blots (Blackwood et al., 1991; Raschella et al., 1994). Furthermore, the kinetic parameters found in the literature for MYC binding on the E-Box sequence are for the MYC-MAX dimer (Park, Chung et al. 2004). As a result, the MAX protein was included in the model at concentrations calculated based on the relevant kinetic interactions and parameters (section 3.3.2).

The validation procedure of the model took the form of time-course simulations (section 3.5), with and without MYCN amplification while using all potential parameters and MAX protein values (Table 3.6), with the results verified against the experimental observations made by Fernandez and colleagues (Fernandez et al., 2003). In all cases, the amount of N-Myc/MAX protein dimers binding on the E-Box promoters was the quantity describing the model outcome. For the non-amplified MYCN, the concentration of the bound E-Box promoters was as reported in the study by Fernandez and colleagues (section 3.5.1). However, with MYCN
amplification the level of promoter occupancy exhibited by the model never reached the value reported in the same study, where the N-Myc/MAX protein dimer was detected not only binding 99% of the E-Box promoters, but it was also found at seemingly random promoter locations as well (section 3.5.2).

The partial lack of agreement between the model behaviour and the results reported in the literature is an expected outcome for a number of reasons. Primarily, this is the first model of this type of system that is also not fully parameterised from published and peer-reviewed data due to the lack of such data for all the steps and species found in the model. Furthermore, not all the published parameters were obtained under the same experimental conditions as each other, or even experimental conditions reflecting accurately the biological situation under examination. One example of the latter would be the establishment of the different rates describing the association of myc and MAX proteins and the E-Box promoters that was produced using partial constructs and without taking in consideration any co-factors that might influence these interactions (Park et al., 2004). These issues taken together made the model building and model validation procedures difficult. On the other hand, the model behaviour does match some of the data while the whole procedure can be considered the first step in a multi-step path to accurately tackle the questions at hand, taking into consideration both experimental and theoretical realities in a true Systems Biology approach.

An additional observation from these simulations was that the model behaviour changed according to the used parameter values and the concentration of MAX, as
presented from the simulations where all parameters were kept constant and the MYCN GCN was increased (section 3.5.2). The binding of the E-Box promoters was dependent on the relative amounts of MAX present. At high levels of MAX protein the available promoters are almost completely occupied by the N-Myc/MAX protein dimers, while on the other end of the parameter spectrum, with low amounts of MAX protein and high amounts of N-Myc protein, the binding of the E-Box promoters was significantly lower. However, in all cases the model predicted a relatively higher impact on E-Box binding by moderate MYCN GCN increase (≤20 GCN) rather than higher levels of MYCN amplification (>20GCN) since all of MAX is depleted while there are comparatively very few E-Box promoters available to bind (section 3.5.2).

These results formed the first hypothesis generated by the model, that in addition to examining the levels of expression of MYCN, it might be necessary to establish the levels of the MAX protein before any predictions can be made regarding the significance of MYCN expression, at least in relation to the binding of E-Box promoters. This is something that to our knowledge is not found in practise when examining the significance of MYCN expression in either cell lines or samples from patients. As already presented and discussed in detail in sections 1.3.3 through to 1.3.5, despite the agreement regarding the significance of MYCN amplification at the genetic level, there is no clear understanding regarding the significance of MYCN expression and/or over-expression in relation to prognosis and outcome in neuroblastoma. It is important to notice that for all the studies presented in sections 1.3.4 and 1.3.5, in no occasion the levels of MAX protein were taken into account and evaluated alongside the levels of MYCN RNA or N-Myc protein levels. As a
result, the experimental procedures of this study were altered accordingly with the aim of investigating the hypothesis that the levels of MAX protein could influence the effect of MYCN expression in relation to binding the E-Box promoters (Chapter 5).

The system properties of the model were examined using the MCA toolkit (section 3.6). While some of the observed behaviour can be attributed partly to the numerical instability of the dimerisation flux due to the scale-difference of the related reactions, it was also clear that the system properties would change based on the relative amounts of N-Myc and MAX proteins. The flux through the dimerization step becomes almost completely unaffected by changes to N-Myc protein levels when the parameters are defined for high MYCN expression and low levels of MAX protein, with the opposite behaviour exhibited at low expression of MYCN and high levels of MAX protein. Similarly the Concentration Control and Response coefficients for the bound N-myc/MAX dimer on the E-Box promoter were also diminishing in a similar manner. Taken together, these results reinforced the hypothesis generated regarding the behaviour of the system where both the expression of MYCN and the relative levels of the MAX protein that are present in the system are important in defining the significance of MYCN expression.

6.1.2: MYCN DNA measurements

The experimentally obtained values for the MYCN GCN of 4 NB cell-lines (SHEP1, SH-SY5Y and NB-EB with no MYCN amplification, and Kelly with more than 100
MYCN GCN per haplotype) are in agreement with what is reported in the literature in terms of the levels of MYCN amplification (Edsjo, Nilsson et al. 2004; Horvilleur, Bauer et al. 2008; Gheeya, Chen et al. 2009). The methodology presented in section 1.8 was used for acquiring the MYCN DNA GCN results for the four NB cell-lines. This method expands upon previously reported work in the field of qPCR data analysis, and especially the selection of HKGs and their utilisation along with each qPCR test’s efficiency. The argument presented here was that in addition to checking relative stability between HKGs at the mRNA level, the potential HKGs at the DNA level are expected to be single-copy genes per haplotype and as such at an appropriate numerical level their mean values should be found not to significantly differ by the appropriate parametric or non-parametric statistical tests. Following from suggestions found in the literature, this was accepted to happen at the Efficiency$^{Cq}$ “relative concentration” level. The novelty of the suggested methodology is that with the use of appropriate statistical tests for validating the potential HKGs, with two or more appropriate HKGs found then the need to include control samples for the purpose of absolute quantification of the gene copy numbers of target genes ceases to exist.

The obtained data for the gene of interest (MYCN) and the four potential HKGs (GPR15, CD200, PolR2D, TP53BP2) from 4 neuroblastoma cell-lines were transformed using the Efficiency$^{Cq}$ formula and were used in subsequent calculations (section 4.2). The transformed data were tested for their type of distribution and then compared with the appropriate non-parametric analysis of variance. Following this analysis, two HKGs were found with non-significantly different means (section 4.2.2). The selected HKGs (CD200 and GPR15) are located
at superimposed chromosomal locations (3q12-q13 for CD200 and 3q11.2-q13.1 for GPR15) and that could suggest the possibility of sharing similar numerical values due to similar local levels of amplification. The data provided evidence that both the inter and intra cell-line variability of these HKGs was within acceptable statistical levels (section 4.2.2) while it is reported that the specific chromosomal location is not mutated in neuroblastoma (Satge et al., 2003). Based on this, the specific HKGs can be accepted for the purposes of this study as a first step in validating this methodology. Further validation steps would be the testing of this methodology with a greater pool of potential HKGs as well as examining the complete karyotype of the cell lines under consideration via the FISH (fluorescence in-situ hybridisation) methodology or indeed by use of CNV analysis by SNP microarrays in order to verify that there is no amplification at chromosomal location 3q.

Having identified two potentially acceptable HKGs, the need for a standard curve or a control sample looses useful merit, allowing for absolute quantification of the gene-of-interest by simply comparison of the two single-copy HKGs. To generate an idea of the spread of potential results from this comparison, the analysis was repeated in a pair-wise fashion with all the technical replicates of both the gene of interest and the two HKGs, a practise that is already established in the field of qPCR data analysis. This gave a coefficient of variance of less than 16% for one cell-line and less than 10% for the rest, which provided further support for the accuracy of this methodological approach (section 4.2.3).
Overall, this simplified method for calculating GCN without the use of reference samples suggests a way for testing potential HKGs at the DNA level that is based on classic statistical methods. Once the criterion regarding the HKGs is met, then the calculation of the absolute GCNs of a gene of interest can be performed without the need for control samples or calibration graphs. This is a simple and clear methodology that could provide a more cost-effective route to perform accurate, quantitative assessment of gene copy-numbers based on qPCR data.

6.1.3: MYCN RNA measurements

The qPCR methodology used in this study for the absolute quantification of the MYCN mRNA molecules for the NB cell-lines produced values that were considered to be inaccurate due to the calculated values of MYCN mRNA molecules per cell being in the region of $10^{-5}$ (section 4.3.3). The basis of the methodology for absolute quantification was on standard curves constructed with the obtained Cq values from the serial dilutions of a MYCN TO, that were then used for interpolating the absolute values of MYCN mRNA for the 4 NB cell-lines.

A number of potential reasons could be responsible for obtaining such low MYCN RNA numbers in terms of absolute quantification. The first parameter checked was the correctness of the relevant calculations used to produce the final values. These were checked, verified and acknowledged independently as correct and appropriate for quantifying the levels of MYCN RNA per cell. This was also the case regarding
the design and validity of the set-up for the qPCR assay (primer design and conditions), which was also confirmed to be appropriate. Additionally, the quality of most of the reagents used (everything other than the relevant MYCN template oligonucleotide that was omitted due to the associated costs) was checked by ordering fresh reagents. These were used for repeating all the tests by using the same methodology and identical results were obtained. This brought the concentration and integrity of the TO in question as one of the last parameters left to validate. However, the means available for this project did not allow the verification of the TO concentration with an independent methodology and the decision was taken to establish the relative amounts of mRNA in these cell-lines, using the cell-line SH-SY5Y as the baseline (section 4.3.3).

For establishing the relative MYCN mRNA values, two HKGs (SDHA and HPRT1) were used and the efficiency of the test of each gene was used in the calculations (both MYCN and the HKGs), a practise that is suggested in the literature and is used in qPCR data analysis software like REST (Pfaffl, Horgan et al. 2002). The obtained values were within acceptable variation limits (CV >5%) in terms of biological replicates, while the relative MYCN expression levels of the cell-lines were in agreement with the reported levels in the literature, at least for the cell-lines SHEP1, Kelly and SH-SY5Y (Edsjo, Nilsson et al. 2004; Horvilleur, Bauer et al. 2008).

Overall, the successfully performed relative quantification of the MYCN RNA levels for these NB cell-lines provided data that can be used for refining the model presented in Chapter 3. In terms of dealing with the problems associated with the
absolute quantification of the MYCN RNA molecules, assuming no limitations in time and money, there are a number of possible steps to take. First of all, fresh TO would be ordered and used for repeating the experiments. Additionally, the cell lysis, RNA extraction and RT procedures would be repeated and their efficiency validated in terms of reproducibility and quality of the extracted material. Finally, specific mass spectrometry methodologies as suggested in the literature can be employed in order to accurately measure the concentration of the TO (Van Ness et al., 2003).

6.1.4: N-Myc and MAX protein measurements

The aim of this part of the project was to accurately measure the levels of N-Myc and MAX proteins in the four NB cell-lines used in this study (Chapter 5). This was required for both the correct parameterisation of the model and also the validation of the model-generated hypothesis that the relative levels of MAX protein to the N-Myc protein are also important in terms of predicting the significance of MYCN expression in relation to binding to E-Box promoters. Two complementary methods, western blots and ELISA tests, were used and optimised for the appropriate experimental conditions, data acquisition and data analysis methods, aiming to achieve in this way the desirable results (Chapter 5).

A multi-step optimisation process was originally followed with the ELISA method. The ELISA tests were performed using 4 primary anti-N-Myc antibodies, 2 monoclonal (from Abnova and Santa Cruz Biotechnologies) and 2 polyclonal (from Santa Cruz Biotechnologies and New England Biolabs). In the normal ELISA tests
with any of the primary anti-N-Myc antibodies, the SHEP1 cell-line consistently showed high readings indicating the presence of N-Myc protein in amounts comparable to the SH-SY5Y and NB-EB cell-lines (section 5.2.2). That was qualitatively not in agreement with the obtained MYCN RNA results for these cell-lines (section 4.3.4) where SHEP1 had less than a hundred-fold lower amount of MYCN RNA than SH-SY5Y (practically no detectable amount of MYCN RNA). Furthermore, it has also been demonstrated in numerous studies that the SHEP1 cell-line does not have any detectable levels of N-Myc protein (Cui et al., 2005; Bray et al., 2009; Peirce and Findley, 2009; Porro et al., 2010).

To investigate the possibility that any material from the insoluble fraction of the cell lysates was responsible for the readings in SHEP1 the tests were repeated with only the soluble fraction. The results were similar, having lower overall readings, but still showing a significant reading from SHEP1 (section 5.2.3). The next step was to try sandwich ELISAs with all the possible combinations of the primary antibodies. According to the information provided by the supplier of each primary antibody, these antibodies were raised against different parts of partial or full recombinant N-Myc proteins. This indicated a lower probability of interference between the antibodies when used in tandem. Furthermore, care was taken when selecting pairs between the primary antibodies so that no two antibodies raised in the same species were used together (Table 5.2). Qualitatively, the sandwich ELISA results for all combinations of the primary antibodies were no different than normal ELISA tests in terms of showing that the SHEP1 cell-line had easily detectable levels of N-Myc protein (section 5.2.4). With these results in mind, and taking into consideration that the majority of the published studies dealing with N-Myc protein levels were using
western blots as the method of choice (even with different antibodies than the ones used here), the decision was taken to switch to western blots.

Western blots visualised via chemiluminescence (ECL) verified that a number of non-N-Myc bands were detected for all cell-lines (Figure 5.7). The densitometry-based measurements of the intensities of the N-Myc bands and the analysis of the data suggested that this method of visualisation was not adequate for absolute quantification of the N-Myc protein using western blots (section 5.2.6 Tables 5.3 and 5.4). This was due to the inconsistencies presented between the intensities of the recognised bands from the same cell-line at different concentrations, indicating that there is no linear relationship between the amount of the recognised protein and the corresponding intensity of the band. The low saturation levels of the densitometry measurements of the visible bands resulted in a non-linear relationship between the amount of protein present and the corresponding signal. An additional source of error between different experiments can also be the variation in exposure conditions between different blots that can have both human and equipment-related origins.

Fluorescence-based methods for western blots can help reduce these two sources of variability. Using fluorescence-based methods for western blot visualisation the upper limit of N-myc detection from the samples did not exceed the saturation point of this methodology, as observed from the measurements of the purified N-myc protein control (section 5.2.7 and 5.2.8). A problem was encountered with higher amounts of total protein content (25mg per lane) from the full lysate samples that were becoming viscous with the addition of the loading buffer. This created
difficulties both in terms of consistent loading and running of the gels. The first measure taken was to use only the soluble part of the lysate for SDS-PAGE (separated by centrifugation as described in section 5.2.3), however the results suggested that a significant loss of N-Myc protein occurred with this method (section 5.2.7). A lysis buffer with a different composition was also tried (RIPA buffer from Sigma) that however did not help with the viscosity problem. This problem was probably caused by the release of DNA molecules after cell-lysis, and a physical method was preferred for breaking them down rather than a chemical one (such as adding DNAse to the lysis buffer). The samples were subjected to vigorous pipetting and this method produced good results in terms of dealing with the sample problems allowing for the use of total lysates at high amounts of protein content (section 5.2.8).

Due to lack of time and money the last western blots were not repeated while the relevant experimental and data analysis details caused some concerns regarding the validity of the data. Primarily, the standard curve that was used for interpolating the N-Myc protein values in the cell-lines included only 4 data points that did not present a good linear fit, especially in comparison to what is reported in the literature (Wang, Wade et al. 2007), even though it was an improvement over the chemiluminescence methods. Additionally, a band at the correct molecular weight for N-Myc was visualised for the SHEP1 cell-line, creating questions for the accuracy of these results since no N-Myc protein was expected for SHEP1. Finally, as a single data set this is an unacceptably low sample number that does not allow for any meaningful statistical analysis and validation of the obtained results.
of quantitative, analysed and statistically validated data in our model, the final data presented in section 5.2.8 can only be considered as initial data that require additional biological and technical replicates (Tables 5.5 and 5.6).

Initially, the results obtained for the MAX protein via the ELISA method were not associated with any problems. The calibration graph was of excellent quality with a very high coefficient of determination and very low variability between replicates. Furthermore, the replicates for the cell lines also had low variability (section 5.3.1) and the obtained values were qualitatively mostly in agreement with what exists in the literature, other than the Kelly cell-line showing higher readings than the other cell-lines (Blackwood et al., 1991; Raschella et al., 1994). In light of the problems encountered in the pursuit for quantifying the N-Myc protein levels via the ELISA method, additional western blots were run in an attempt to visualise the recognised products of the anti-MAX antibody used in this study. This was done using both chemiluminescent and fluorescent detection methods (section 5.3.2). In both cases, despite getting a clear signal from the purified MAX protein used as a positive control, there was not detection signal from any of the cell-lines’ lysate. The highest possible concentration of MAX protein per cell for each cell-line was calculated based on the lowest detectable amount of purified MAX protein that was used as positive control (table 5.7). In comparison the values calculated based on the western blots were below the values obtained from the ELISA method (Table 5.8), suggesting that the readings observed via the ELISA method were not exclusively due to the presence of the MAX protein, but rather due to non-specific binding of the antibody. In comparison, the reported studies that investigated the MAX protein levels also used the western blot methodology for qualitative measurements, while
also employing different anti-MAX primary antibodies than the one used for this study (Blackwood et al., 1991; Raschella et al., 1994).

Overall, the quality of the acquired data at the protein level did not allow for the refinement of the model parameters. However, the preliminary results obtained for the protein levels of both N-Myc and MAX proteins suggest that the level of MAX is significantly lower than N-Myc, especially for the Kelly cell-line. This raises various possibilities regarding the biological behaviour of the NB cells in relation to the MYCN expression.

It is understandable that the kinetic parameters used in this model are not entirely accurate for reasons mentioned in relevant publications, where for example only partial sequences from MAX and MYC were used rather than the complete proteins with all the associated modifications and relevant co-factors (Park, Chung et al. 2004). This means that the kinetic parameters can potentially be faster and as such accommodate higher amounts of N-myc protein that would function via this pathway.

However, the observed potential difference in the protein levels would support suggestions regarding alternative modes of action for N-Myc, especially at higher concentrations. Two reports that have already been mentioned show that over-expressed myc binds randomly to promoters (Fernandez, Frank et al. 2003) and that MYC in general has a pathway independent of MAX (Steiger, Furrer et al. 2008).
Additionally, the simulations of the suggested MYCN gene expression model have already predicted that there is higher relative impact at low levels of MYCN amplification and increase in expression, (section 3.4.2) meaning that the potentially excessive amounts of N-Myc protein present are not fully employed in this pathway. It follows that if the levels of N-Myc protein in NB cell-lines (or patients) like Kelly are fully relevant rather than simply exhausting their related network, an alternative mode of action probably takes place than the one involving dimerisation with MAX for binding on E-Box promoters. This would be an interesting question to answer regarding the role and significance of MYCN amplification and expression in NB.

**6.2: FUTURE DIRECTIONS**

This study can form the basis for incorporating various MYCN and NB related aspects in the model and expand the scope of the investigation accordingly as it will be discussed in the following section (Figure 6.1). These aspects include among others the temporal elements relevant to MYCN expression, the relation of MYCN to the cell-cycle in general and the MYCN-related multi-drug resistance in NB.

It is imperative that great consideration must be given to a number of key features and the main questions to be answered before any decisions regarding the direction of the research are taken. Perhaps one of the most important aspects is the availability of specific experimental techniques for acquiring data that can be used for both refining the model and testing the generated hypotheses, as shown by some
of the limitations of this study. For example, additional techniques or quality controls should be incorporated to achieve absolute quantification of the MYCN transcript and N-Myc protein levels.

Finally, it is also important to expand the means of simulation and analysis of the model. COPASI is a very useful tool for simulation and analysis of models since it incorporates a number of complex algorithms for performing tasks like MCA, optimisation and parameters scans and stochastic simulations. However it lacks certain features that other tools like Mathematical Laboratory MATLAB® (The MathWorks, Cambridge, U.K.) provide that include tools for creating plots of high quality and tools for bifurcation analysis, while it is also difficult to introduce certain features in the model like periodic elements that would describe for example the expression of MYCN in normal cells in a normal cell cycle. These are features that will potentially be needed for future work on the subject, where the impact of various modifications and aberrations in relation to MYCN will examined and compared to normal-cell behaviour. Utilising both tools, COPASI and MATLAB, would be the preferable way of conducting future modelling work.

6.2.1: Refinement and additional analysis of the existing model

The repetition and refinement of the methodology used for accurate measurement of the N-Myc and MAX proteins would be desirable, as well as accurately measuring MYCN mRNA levels. The acquired data will be used for fitting the parameters of the original model into separate parameter sets that correspond to the individual NB
cell lines. This can also be done with primary samples in the future, depending on the availability of samples and if it would be deemed to be useful. To reduce the non-specific binding that drastically increases the noise in ELISA, the samples could be subjected to a process known as “preclearing”. This is simply the washing of the samples with a random antibody that would bind to any proteins that are susceptible to interacting randomly with immunoglobulins. The result of this method is increase of the detection specificity and general improvement of the signal-to-noise ratio. Furthermore, alternative methods for protein measurements would include mass spectrometry and nuclear magnetic resonance. Both methods are more expensive and time-consuming that the ones used in this study, but can offer more information about the proteins under investigation that can be useful depending on the questions asked, like the phosphorylation status at specific sides of a protein or the interactions between different molecules of interest.

6.2.2: Controlling MYCN expression

The analysis and comparison of the individual models can provide evidence for potential shifts in the significance and the control coefficients of the different constituent reactions and especially the ones related to the regulation of MYCN expression. These steps can be expanded in any of the ways presented in Chapter 1, section 1.4.2 that assist in the processing and stability of the transcription and translation products of the MYCN gene. For example if the model predicts that there will be higher impact on the MYCN-related activity by interfering with the N-Myc protein stability, the phosphorylation of the N-Myc protein at positions Ser-58 and Thr-62 could be a subject for investigation. It has already been reported that Ras
interferes with the N-Myc protein stability and could provide a potential target (Yaari et al., 2005). Additionally, a theoretical study has suggested a model that describes the effect of Ras on the accumulation of N-Myc molecules via the action of Erk and PI3K (Lee et al., 2008). This model could provide a first version of an expanded translation step for the N-Myc protein where the effect on network properties will be examined in relation to any interference with these N-Myc phosphorylation steps, while any generated hypotheses would be validated experimentally.

6.2.3: Temporal aspects of MYCN presence in the cell

One of the assumptions made for constructing the model and the way the experimental data was meant to be used is that the system reaches quasi-steady state. This is not necessarily true in terms of the expression pattern of MYC proteins during the cell-cycle of non-NB cells (Kelly, Cochran et al. 1983). Additionally, the units of the model were chosen in order to be compatible with the cell-cycle model of a different study with the provision that once the models from this project were properly parameterised with experimental data, they would be merged with this specific cell-cycle model (Yao et al., 2008). These are some of the temporal aspects that could be further explored in the future, based on the results of this study.

6.2.4: Expression of MYCN

The expression of MYCN and c-myc is a signal-dependent event, peaking at G1 in normal cells when its role is to push the cell through to S phase. As such it has been
suggested that changes in the temporal aspects of this expression profile, rather than changes in the absolute levels, could be more important for pushing the cell towards a tumour state by preventing exit from the cell-cycle into a G0 or apoptosis stage (Kelly, Cochran et al. 1983). To investigate this possibility, the specific expression profiles for the different cell-lines can be established experimentally using cell-synchronisation protocols that include among others serum starvation methods and double thimidine blocks. Constructing the expression profiles through the cell cycle for each cell-line will provide valuable information regarding the potential significance of the presence of N-Myc protein and its activity. Furthermore it would also provide information about the potential blocking of the MAX-related activity of the MAD family members that are responsible for the termination of transcription of many common targets that are shared with the myc family members (Rottmann and Luscher 2006).

6.2.5: MYCN and cell-cycle progression

The cell-cycle model from Yao and colleagues is one of the few MYC-related cell-cycle models to be published (Yao et al., 2008). The interactions defining MYC in that model are a one-step MYC production and a number of reactions where MYCN is responsible for inducing the production of other cell-cycle molecules. For the purposes of that study this description was sufficient, however the incorporation of the model built for this project could provide certain advantages. With the additional steps provided in terms of the MYCN expression, it will be feasible to investigate the effect and importance of these steps in a bigger system rather than in isolation, giving to the researchers suggestions as to which steps are truly important for cell
cycle progression in relation to the MYCN expression and activity steps. It will be important to note how different network properties will also change following the merging of the two models, especially for a model that would be describing the expression of the Kelly cell-line due to its high levels of MYCN amplification and expression.

These observations and any hypotheses generated for the individual cell lines will be investigated experimentally. The methodology described in the specific publication can be followed in its entirety if there are no practical restrictions (Yao et al., 2008). A starting point would be the growth of cells under different FCS concentrations in the media in order to provide different growth-factor concentrations and then measure the expression of MYCN at both the transcript and the protein level. Additionally, it will also be useful to do so for MAX and if possible for the additional factors present in the model like E2F and Rb to provide helpful information leading towards a complete understanding and validation of the system under investigation.

6.2.6: Pharmacological studies in NB

A study by Paffhausen and colleagues investigated the phenomenon of multi-drug resistance of NB patients with MYCN amplification by performing different dose-response experiments with anti-tumour drugs like paclitaxel and cisplatin that affect different parts of the cell-cycle (Paffhausen et al., 2007). The study was laboratory-based and generated important data that defined some of the differences in behaviour
between a cell-line expressing MYCN and one that does not. This work could form the basis for a Systems Biology study on the same subject that would include a higher number of experimental samples as well as the support of relevant mathematical modelling and systems-based investigation.

Starting from the experimental aspects, the specific study used the SHEP1 cell-line in both the wild-type form and also as a stable clone transfected with a MYCN vector to over-express MYCN. The result was the comparison of the drug effects between a non MYCN-amplified and an artificially MYCN-amplified NB cell-line, while it would be interesting to investigate the outcome of these experiments on the other 3 NB cell-lines used in this study, especially to Kelly that has the highest levels of MYCN amplification and expression. Additionally, stable clones of the other cell-lines can be created that either over-express MYCN, using the procedure followed for creating the SHEP$^{\text{MYCN}}$ cell line used in the paper, or under-express MYCN as MYCN-gene knock-outs using a siRNA methodology. Woo and colleagues have presented as part of their study the methodology they followed for successfully creating cell-lines with MYCN gene knock-outs (Woo et al., 2008). Additionally, a commercially available set of siRNAs could be acquired in the form of an siRNA SmartPool for MYCN from Dharmacon (Fisher Scientific, Loughborough, U.K.), a methodology that is used successfully in an affiliated laboratory.

In terms of the relevant modelling, to incorporate and examine the effects of all the drugs used in the study it would be necessary to either add the rest of the cell-cycle reactions to the model or use a different and complete cell cycle model to which
MYCN and its related interactions would be incorporated. One such model is from Conradie and colleagues who also suggested a new methodology for analysis of dynamic properties and emergent behaviours in networks (Conradie et al., 2010). It is vital to appreciate that the incorporation of the model of this study into this cell cycle model would be a more complicated process than doing the same with the model from Yao and colleagues. This is due to the substantial difference in the size of the models, while MYCN (or even MYC) is not part of the Conradie model and it needs to be incorporated in a manner that does not disturb the reported function of the model.

Overall, various experimental and modelling avenues can be pursued based on the ideas and outcome of the presented work. Each subject could provide additional information on a number of themes regarding the expression, regulation and function of the N-Myc protein in different subtypes of NB. The interaction between lab-based and mathematical-based experimentation is a vital aspect in improving our understanding of the emergent properties of all the relevant networks and mechanisms responsible for the MYCN-related aspects of NB. This is a prerequisite for reaching the stage where each NB patient will be receiving the appropriate personalised treatment.
6.3: Conclusions and study summary

Overall the study has been successful in meeting some of the targets it had set. Even though the extent to which the benefits of using both experimental and computational approached in scientific research has not reached the desired level, some evidence has been provided none-the-less. In specific:

1. A general model for MYCN expression was created following the originally-set design parameters and using the COPASI software. These included creating a model to relate the MYCN gene, transcript and protein numbers and the binding to the E-Box promoters in association with the protein MAX,
while taking the model parameters from the literature. The simulations and analysis investigated the degree of freedom of the model in relation to the relative levels of MYCN and MAX proteins and generated among others a hypothesis regarding the potential significance of MYCN expression that could be tested experimentally (Chapter 3).

2. The absolute quantification of MYCN GCN was successful and a new methodology was suggested for analysing qPCR data at the DNA level, while incorporating relevant statistical tests without the use of extra samples as standards (Chapter 4).

3. Despite the problems associated with absolute quantification of the MYCN mRNA levels, relative quantification was achieved with good agreement between biological replicates, as well as with the reported levels in the literature (Chapter 4).

4. The N-myc protein levels were established, but not to a satisfactory level of accuracy and confidence. However, a detailed methodology involving two complementary techniques was optimised and presented while additional biological and technical replicates are required for validation of the results obtained (Chapter 5).
REFERENCES


Solovei, I, D Kienle, G Little, R Eils, L Savelyeva, M Schwab, W Jager, C Cremer, and T Cremer. “Topology of Double Minutes (dmins) and Homogeneously Staining


# APPENDIX

## Table A1: The associated assay components for the relevant DNA qPCR assays

<table>
<thead>
<tr>
<th>Gene name (official)</th>
<th>UPL Probe#</th>
<th>Chromosomal location</th>
<th>Assay Component</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MYCN</strong> (v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian))</td>
<td>45</td>
<td>2p24</td>
<td>Forward Primer</td>
<td>ctaaagttgttgacgttg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reverse Primer</td>
<td>tggaaacttggaaacacaa</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amplicon</td>
<td>ctaaagttgttgatggagtggcctcgctgttgaagtcacctgtgtgttctcaagtctcctcctaagtttccaa</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Template oligonucleotide</td>
<td>ctaaagttgtagccttgagcctctgctgttgaagtcacctgtgtgttctcaagtcttccaa</td>
</tr>
<tr>
<td><strong>GPR 15</strong> (G protein-coupled receptor 15)</td>
<td>51</td>
<td>3q11.2-q13.1</td>
<td>Forward Primer</td>
<td>ctgtgccagcatctgttt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reverse Primer</td>
<td>tccttgtgagagaagcttg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amplicon</td>
<td>ctgtgccagcatctgtttatctcctactctgtgtgttccaagtttccaa</td>
</tr>
<tr>
<td><strong>TP53BP2</strong> (Tumor protein p53 binding protein 2)</td>
<td>25</td>
<td>1q42.1</td>
<td>Forward Primer</td>
<td>gtcacagtgtcgcgtgtg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reverse Primer</td>
<td>gttggtggtggtggtg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amplicon</td>
<td>gtcacagtgtcgcgtgtgtaactgtggtggtggtggtg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Template oligonucleotide</td>
<td>gtcacagtgtcgcgtgtgtaactgtggtggtggtggtg</td>
</tr>
<tr>
<td><strong>CD200</strong> (CD200 molecule)</td>
<td>88</td>
<td>3q12-q13</td>
<td>Forward Primer</td>
<td>atgggcacagctgttac</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reverse Primer</td>
<td>cttcccaacgacatgtag</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amplicon</td>
<td>atgggcacagctgttacgactctctctctctctctcataagacaccctgagtccagggag</td>
</tr>
<tr>
<td><strong>POLR2D</strong> (Polymerase (RNA) II (DNA directed) polypeptide D)</td>
<td>25</td>
<td>2q21</td>
<td>Forward Primer</td>
<td>gaggagttctaatctgttcgg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reverse Primer</td>
<td>tcagagcgccgctgtacg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amplicon</td>
<td>gaggagttctaatctgttcggaggggtcagtcatcatctgcctgca</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Template oligonucleotide</td>
<td>gaggagttctaatctgttcggaggggtcagtcatcatctgcctgca</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>gaggagttctaatctgttcggaggggtcagtcatcatctgcctgca</td>
</tr>
</tbody>
</table>

APPENDIX
<table>
<thead>
<tr>
<th>Gene name (official)</th>
<th>UPL Probe#</th>
<th>Chromosomal location</th>
<th>Assay Component</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYCN (v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian))</td>
<td>55</td>
<td>2p24</td>
<td>Forward Primer</td>
<td>tgaacctttgattttgacccagtacccgagccggtctatattccaga</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reverse Primer</td>
<td>cgagacggggtctatattccaga</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amplicon</td>
<td>tgaacctttgattttgacccagtacccgagccggtctatattccagagacaacccctgtcctttgtggacgttggcacugggaaggtcacatgctac</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Template oligonucleotide</td>
<td>tgaacctttgattttgacccagtacccgagccggtctatattccagagacaacccctgtcctttgtggacgttggcacugggaaggtcacatgctac</td>
</tr>
<tr>
<td>HPRT1 (hypoxanthine phosphoribosyltransferase 1)</td>
<td>73</td>
<td>Xq26.1</td>
<td>Forward Primer</td>
<td>tgaacctttgattttgacccagtacccgagccggtctatattccaga</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reverse Primer</td>
<td>cgagacggggtctatattccaga</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amplicon</td>
<td>tgaacctttgattttgacccagtacccgagccggtctatattccagagacaacccctgtcctttgtggacgttggcacugggaaggtcacatgctac</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Template oligonucleotide</td>
<td>tgaacctttgattttgacccagtacccgagccggtctatattccagagacaacccctgtcctttgtggacgttggcacugggaaggtcacatgctac</td>
</tr>
<tr>
<td>SDHA (Succinate dehydrogenase complex, subunit A, flavoprotein variant)</td>
<td>69</td>
<td>5p15</td>
<td>Forward Primer</td>
<td>aagaagccctttgaggagca</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reverse Primer</td>
<td>cgagacggggtctatattccaga</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amplicon</td>
<td>aagaagccctttgaggagacacacactggaggacaggacacaccctgtcctttgtggacgttggcacugggaaggtcacatgctac</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Template oligonucleotide</td>
<td>aagaagccctttgaggagacacacacugctcctttgtggacgttggcacugggaaggtcacatgctac</td>
</tr>
</tbody>
</table>