

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>

**A therapeutic approach to chronic myeloid leukaemia using short hairpin RNA molecules**

Al-Mazedi, Maryam

*Awarding institution:*  
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

**END USER LICENCE AGREEMENT**



**Unless another licence is stated on the immediately following page** this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

**Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

This electronic theses or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>

**Title:** A therapeutic approach to chronic myeloid leukaemia using short hairpin RNA molecules

**Author:** Maryam Al-Mazedi

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

#### END USER LICENSE AGREEMENT



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. <http://creativecommons.org/licenses/by-nc-nd/3.0/>

You are free to:

- Share: to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

#### Take down policy

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# **A Therapeutic Approach to Chronic Myeloid Leukaemia Using Short Hairpin RNA Molecules**

By

Maryam S. Al-Mazedi

A thesis submitted in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy at King's College London

The Institute of Pharmaceutical Science,  
School of Biomedical and Health Sciences,  
King's College London,  
Hodgkin's Building,  
Guy's Campus,  
London  
SE1 1UL

September 2011

## Abstract

Chronic myeloid leukaemia (CML) was one of the first cancers to be linked to a chromosomal abnormality, the Philadelphia chromosome. This chromosome results in a translocation between chromosomes 9 and 22, where the ABL gene on chromosome 9, a tyrosine kinase, is translocated to the BCR gene region on chromosome 22 giving rise to an abnormal BCR/ABL fusion gene. The resultant fusion gene has an abnormally upregulated tyrosine kinase activity that results in an increase in the proliferation of immature white blood cells, thus leading to the development of CML. There are several breakpoints that can occur in the BCR gene two of which give rise to 95% of CML cases. These fusion points are called the  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$  depending on where the chromosomal breakages occur in the BCR gene. The aim of the project was to establish a new method of treatment for CML through the use of RNAi to abolish the increase in tyrosine kinase activity of the abnormal fusion gene product.

The K562 and KCL22 cell lines incorporating the  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$  fusion points respectively were used in this project. The fusion points were cloned and sequenced. The human U6 and H1 promoters, were selected for the production of the antisense molecules and were obtained by PCR of K562 cDNA. Short hairpin RNA molecules were designed to the sequences of the  $\beta 3\alpha 2$  and the  $\beta 2\alpha 2$  fusion points. These designed shRNA molecules were synthesized as oligonucleotides and were incorporated into a reverse PCR primer. Cassettes containing shRNA molecules and a respective promoter were produced by means of PCR and the products cloned into pB12mcs-eGFP vector, which was used as a GFP reporter system. Constructs were then transfected into the appropriate cell lines, and expression studies including qPCR and Western blot analysis were conducted, to examine the effects of the designed shRNA constructs to their target sites on both mRNA and protein levels. In addition, these experiments also indicate the efficiency of the construct and also their specificity to their targets.

qPCR and Western blot analysis, show that both the shRNA molecules designed against the  $\beta 3\alpha 2$  and the  $\beta 2\alpha 2$  fusion points, efficiently induced RNAi based gene silencing to their target sites. In addition, the designed constructs show high specificity to only its target sites and not to other unrelated or related genes. These results point towards the use of molecular modulation of gene expression as a promising strategy for potential CML therapy.

## **Acknowledgements**

I would like to thank my supervisors Peter Eagles and Mark Sanderson for their guidance, help, and advice through out the current project. Special thanks to Amer Qureshi for all his help, advice, and bearing with my continuous questions, and great thanks to Derek knight for all his help.

Special thanks to my friends: Kasia (always getting me the best chocolates and selection of teas ever), Vibah, Aseem, Oishik, Dima, Ashvin, Mads, Martin, Nicola, and Alina for helping me, being there for me, giving me support whether I need it or not, and putting a smile on my face all the time. I would also like to thank Ay Lin and Martina for their help in my Western Blots, and Anna for helping me with qPCR data.

Finally, I would like to thank my parents for their support throughout my years of study.

# Table of Contents

<b>ABSTRACT</b> .....	<b>2</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>4</b>
<b>TABLE OF CONTENTS</b> .....	<b>5</b>
<b>TABLE OF FIGURES</b> .....	<b>9</b>
<b>TABLE OF TABLES</b> .....	<b>12</b>
<b>LIST OF ABBREVIATIONS</b> .....	<b>13</b>
<b>CHAPTER 1: INTRODUCTION</b> .....	<b>17</b>
1.1 HAEMATOPOIESIS AND LEUKAEMIA .....	17
1.2 CHRONIC MYELOID LEUKAEMIA (CML).....	18
1.2.1 DEFINITION OF CML .....	19
1.2.2 CLINICAL STAGING OF CML .....	19
1.3 THE MOLECULAR BIOLOGY OF CML .....	21
1.3.1 THE ABL PROTEIN .....	22
1.3.2 THE BCR PROTEIN .....	24
1.3.3 THE BCR-ABL FUSION GENE.....	26
1.3.4 THE p210 <sup>BCR-ABL</sup> FUSION PROTEIN .....	30
1.3.5 BCR-ABL ACTIVITY IN CML.....	32
1.3.5.1 The RAS Pathway.....	33
1.3.5.2 The Phosphatidylinositol 3' Kinase Pathway.....	36
1.3.5.3 The JAK/STAT Pathway.....	39
1.3.5.4 Other Proteins Associated with BCR-ABL Activity.....	41
1.3.5.5 BCR-ABL and Protein Pathways: A Conclusion.....	41
1.4 CML AND THERAPY.....	42
1.4.1.1 Current Drug Treatment for CML.....	42
1.4.1.2 Mechanisms of Imatinib Drug Resistance in CML.....	44
1.4.1.2.1 BCR-ABL Independent Mechanisms of Drug Resistance.....	45
1.4.1.2.2 BCR-ABL Dependent Mechanisms of Drug Resistance.....	46
1.4.2 SECOND AND THIRD GENERATION TKI .....	46
1.5 MODULATION OF GENE EXPRESSION .....	47

1.5.1 DNA TRIPLE HELICES .....	48
1.5.2 ANTISENSE OLIGONUCLEOTIDES .....	49
1.5.3 RIBOZYMES.....	50
1.5.4 RNA INTERFERENCE .....	51
1.5.4.1 The Mechanism of RNAi.....	51
1.5.4.2 MicroRNAs.....	54
1.5.4.3 siRNAs.....	54
1.5.4.4 shRNAs.....	55
1.5.5 THE USE OF RNAi AS A THERAPEUTIC TOOL.....	58
1.5.6 RNAi AND CML .....	60
1.6 CELL LINES AS A MODEL FOR CML.....	62
1.7 AIM OF THE PROJECT .....	63
<b>CHAPTER 2: MATERIALS AND METHODS.....</b>	<b>65</b>
2.1 MOLECULAR BIOLOGY.....	65
2.1.1 RNA Extraction and Purification .....	65
2.1.2 DNA Extraction and Precipitation.....	65
2.1.3 Quantification of Nucleic Acid Concentration.....	66
2.1.4 cDNA Synthesis.....	66
2.1.5 Polymerase Chain Reaction .....	67
2.1.6 Agarose Gel Electrophoresis.....	70
2.1.7 Restriction Enzyme Digest .....	71
2.1.8 DNA Ligation.....	72
2.1.9 Transformations.....	73
2.1.10 Plasmid DNA Isolation and Purification .....	74
2.1.11 Glycerol Stocks.....	74
2.1.12 DNA sequencing .....	75
2.1.13 Generation of Blunt Ends.....	75
2.2 AT CLONING .....	76
2.3 GENERATION OF PLASMIDS.....	77
2.3.1 Construction of the shRNA Cassettes.....	81
2.3.2 Cloning of Promoter Constructs .....	82
2.4 TISSUE CULTURE.....	87
2.4.1 Cell line Maintenance.....	87
2.4.2 Optimisation of Cell Line Transfection .....	90
2.4.2.1 Chemical based transfections.....	89
2.4.2.2 Non-chemical based transfections.....	91
2.5 FLUORESCENCE-ACTIVATED CELL SORTER (FACS) ANALYSIS.....	94
2.6 REAL TIME PCR.....	97
2.7 WESTERN BLOT ANALYSIS.....	103
2.7.1 Cell lysis and protein extraction.....	104



2.7.2 SDS-Polyacrylamide Gel Electrophoresis.....	104
2.7.3 Western Blotting .....	105

**CHAPTER 3: GENERATION OF A REPORTER GENE SYSTEM FOR EXPRESSION OF SHRNA CASSETTES TARGETING THE BCR-ABL FUSION POINTS.....109**

3.1 INTRODUCTION.....	109
3.2 RESULTS .....	110
3.2.1 Sequencing of the $\beta 3\alpha 2$ and the $\beta 2\alpha 2$ Fusion Points .....	110
3.2.1.1 PCR Amplification of the $\beta 3\alpha 2$ and the $\beta 2\alpha 2$ Fusion points.....	108
3.2.1.2 AT Cloning of the $\beta 3\alpha 2$ and the $\beta 2\alpha 2$ Products in the pGEM@T Easy Vector System.....	111
3.2.1.3 Analysis of the $\beta 3\alpha 2$ and the $\beta 2\alpha 2$ Fusion Point sequences.....	112
3.2.2 Preparation of the pEGFP-N1 Vector for Cloning of the shRNA Cassettes.....	115
3.2.2.1 Digestion and Removal of the CMV Promoter from the pEGFP-N1 vector .....	113
3.2.2.2 Self Ligation and Cloning of the Modified pEGFP-N1 vector.....	115
3.2.3 Preparation of the pB12mcs-eGFP Vector .....	119
3.2.3.1 Sequencing of the pB12mcs-eGFP vector.....	117
3.2.3.2 Digestion and Purification of the pB12mcs-eGFP Vector.....	117
3.2.4 Amplification of the U6 and H1 Promoters.....	121
3.2.4.1 Cloning of the U6 and H1 Promoters in the pB12mcs-eGFP Vector.....	120
3.2.4.2 Sequence Analysis of the U6 and H1 Promoters.....	122
3.2.5 Construction of the shRNA Molecules .....	127
3.2.5.1 Designing of the shRNA Molecules.....	125
3.2.5.2 Synthesis of the U6 and H1 shRNA Cassettes.....	126
3.2.5.3 Cloning of the shRNA Cassettes in the pB12mcs-eGFP Vector.....	128
3.2.5.4 Conformation by PCR of the shRNA Cassettes and the Sequencing of the Cassettes..	129
3.3 DISCUSSION.....	137

**CHAPTER 4: EXPRESSION OF THE SHRNA CONSTRUCTS IN K562 AND KCL-22 CELL LINES.....139**

4.1 INTRODUCTION.....	139
4.2 RESULTS .....	139
4.2.1 Transfection of the shRNA Constructs in K562 and KCL-22 Cell lines .....	139
4.2.2 Transfection by Chemical Methods and FACS Analysis.....	140
4.2.3 Transfection Through Nucleofection .....	145
4.2.4 Real Time PCR Analysis of the Transfected Cell Lines.....	151
4.2.4.1 The Generation of Standard Curves.....	153
4.2.4.2 Effects of shRNA $\beta 3\alpha 2$ and shRNA $\beta 2\alpha 2$ in K562 and KCL-22 cells.....	161
4.2.4.3 Effects of shRNA1 $\beta 2\alpha 2$ in KCL-22 and K562 cell lines.....	164
4.2.4.4 Effects of the shRNA constructs on normal BCR and ABL mRNA levels.....	167
4.2.5 Protein analysis of the K562 transfected Cell Line by Western Blotting .....	171
4.2.6 Protein analysis of the KCL-22 transfected Cell Line by Western Blotting.....	174
4.3 DISCUSSION.....	178

<b>CHAPTER 5: DISCUSSION .....</b>	<b>180</b>
5.1 DISCUSSION.....	180
5.1.1 DEVELOPMENT AND EXPRESSION OF shRNA CONSTRUCTS TARGETING THE BCR-ABL VARIANTS.....	181
5.1.2 EFFICIENCY OF shRNA CONSTRUCTS TARGETING THE $\beta 3\alpha 2$ FUSION POINT.....	185
5.1.2.1 Delivery and production of the shRNA $\beta 3\alpha 2$ Constructs.....	186
5.1.2.2 Effects of the shRNA $\beta 3\alpha 2$ Constructs on $\beta 3\alpha 2$ mRNA and protein levels.....	187
5.1.2.3 Conclusion.....	188
5.1.3 EFFICIENCY OF shRNA CONSTRUCTS TARGETING THE $\beta 2\alpha 2$ FUSION POINT.....	189
5.1.3.1 Delivery and production of the shRNA $\beta 2\alpha 2$ Construct.....	189
5.1.3.2 Effects of the shRNA $\beta 2\alpha 2$ Constructs on $\beta 3\alpha 2$ mRNA and protein levels.....	190
5.1.3.3 Conclusion.....	191
5.1.4 siRNA AGAINST THE BCR-ABL VARIANTS.....	191
5.1.5 FUTURE DEVELOPMENTS .....	193
<b>APPENDIX 1: THE MOLECULAR WEIGHT MARKERS USED IN THE CURRENT PROJECT .....</b>	<b>196</b>
<b>APPENDIX 2: THE SECONDARY SHRNA STRUCTURES .....</b>	<b>200</b>
<b>APPENDIX 3: THE PROTEIN MOLECULAR WEIGHT MARKER USED IN THE CURRENT PROJECT .....</b>	<b>205</b>
<b>APPENDIX 4: THE 3A2 AND B2A2 SEQUENCES.....</b>	<b>206</b>
<b>APPENDIX 5: THE RAW SEQUENCING DATA OF THE B3A2 AND B2A2 FUSION POINT SEQUENCES IN THE PGEM®-T EASY VECTOR.....</b>	<b>208</b>
<b>APPENDIX 6: THE HUMAN U6 AND H1 PROMOTER SEQUENCES.....</b>	<b>210</b>
<b>APPENDIX 7: ELECTROPHORETIC ANALYSIS OF REPRESENTATIVE COLONY SCREENS PCR PRODUCTS OF THE U6 SHRNA AND H1 SHRNA CONSTRUCT ..</b>	<b>212</b>
<b>APPENDIX 8: THE RAW SEQUENCING DATA OF THE U6 AND H1 PROMOTERS INSERTED INTO THE PB12MCS-EGFP PLASMID.....</b>	<b>218</b>
<b>APPENDIX 9: RAW DATA OF SEQUENCING OF THE SHRNA CONSTRUCTS .....</b>	<b>222</b>
<b>APPENDIX 10: FACS GENERATED DATA OF TRANSFECTION OF K562 AND KCL-22 CELL LINES.....</b>	<b>256</b>
<b>BIBLIOGRAPHY .....</b>	<b>263</b>

## Table of Figures

FIGURE 1.1 CLASSIFICATION OF THE WHITE BLOOD CELLS (HOFFBRAND ET AL., 2006).....	18
FIGURE 1.2 THE RECIPROCAL TRANSLOCATION BETWEEN CHROMOSOMES 9 AND 22 PRODUCING THE PHILADELPHIA CHROMOSOME. ....	22
FIGURE 1.3 THE STRUCTURAL ORGANIZATION OF THE C-ABL PROTEIN. ....	23
FIGURE 1.4 THE STRUCTURAL ORGANIZATION OF THE BCR PROTEIN. ....	24
FIGURE 1.5 THE BREAKPOINTS WITHIN THE <i>BCR</i> GENE FOUND ON CHROMOSOME 22.....	27
FIGURE 1.6 THE <i>BCR</i> AND <i>ABL</i> GENES, BREAKPOINTS, AND TRANSCRIPTS.....	28
FIGURE 1.7 SCHEMATIC REPRESENTATION OF THE STRUCTURAL MOTIFS OF THE NORMAL BCR PROTEIN AND THEIR ASSOCIATION WITH THE BREAKPOINT CLUSTER REGIONS.....	30
FIGURE 1.8 SCHEMATIC REPRESENTATION OF THE STRUCTURAL MOTIFS OF THE P210 BCR-ABL PROTEIN.....	32
FIGURE 1.9 A SCHEMATIC DIAGRAM OF THE RAS PATHWAY AND THE EFFECTS OF THE BCR-ABL ON IT.....	34
FIGURE 1.10 SCHEMATIC DIAGRAM OF THE PI3K PATHWAY.....	38
FIGURE 1.11 A SCHEMATIC DIAGRAM OF THE JAK/STAT PATHWAY.....	40
FIGURE 1.12 THE MECHANISM OF BCR-ABL INHIBITION BY IMATINIB.....	44
FIGURE 1.13 SCHEMATIC DIAGRAM OF RNAi MECHANISM.....	53
FIGURE 1.14 THE MECHANISMS OF ENDOGENOUS AND EXOGENOUS EXPRESSION OF miRNA, shRNA, AND siRNA.....	57
FIGURE 2.1 MAP OF THE PGEM®T EASY VECTOR SYSTEM. THE DIAGRAM WAS OBTAINED FROM THE PROMEGA WEBSITE.....	77
FIGURE 2.2 PEGFP-N1 VECTOR MAP. THE DIAGRAM WAS OBTAINED FROM THE CLONTECH WEBSITE.....	79
FIGURE 2.3 THE PB12MCS-EGFP VECTOR MAP. THE PLASMID WAS CONSTRUCTED AND GIVEN BY DR. ADAM RODAWAY.....	80
FIGURE 2.4 SCHEMATIC DIAGRAM OF THE SYNTHESIS OF THE shRNA CASSETTES.....	84
FIGURE 2.5 A FLOW DIAGRAM OF THE OVERALL PROCEDURES UNDERTAKEN TO GENERATE THE PLASMIDS CONTAINING THE shRNA CASSETTES.....	86
FIGURE 2.6 THE HAEMACYTOMETER. (THE DIAGRAM WAS TAKEN FROM THE KLINIK ANDROLOGI WEBSITE).....	89
FIGURE 2.7 THE BASIC PRINCIPLE OF THE FACS SCAN. (SOURCE <a href="http://WWW.ABCAM.COM">WWW.ABCAM.COM</a> ).....	95
FIGURE 2.8 FACS HISTOGRAM GRAPH.....	97
FIGURE 2.9 REAL TIME PCR USING SYBR GREEN I METHODOLOGY.....	99
FIGURE 2.10 REAL TIME PCR IN THE PRESENCE OF TAQMAN PROBES.....	100
FIGURE 3.1 ELECTROPHORETIC ANALYSIS OF PCR REACTION USING $\beta$ -ACTIN PRIMERS ON K562 cDNA.....	111

FIGURE 3.2 ELECTROPHORETIC ANALYSIS OF THE b2A2 PCR PRODUCT FROM KCL-22 cDNA.	112
FIGURE 3.3 ELECTROPHORETIC ANALYSIS OF THE b3A2 PCR PRODUCT FROM K562 cDNA. ...	113
FIGURE 3.4 THE K562 AND KCL-22 FUSION POINT SEQUENCING RESULTS.....	114
FIGURE 3.5 ELECTROPHORETIC ANALYSIS OF pEGFP-N1 DIGESTIONS .....	116
FIGURE 3.6 ELECTROPHORETIC ANALYSIS OF PCR PRODUCTS FROM THE SELF-LIGATED pEGFP-N1 VECTOR.....	118
FIGURE 3.7 ELECTROPHORETIC ANALYSIS OF THE DIGESTED pB12MCS-EGFP VECTOR .....	120
FIGURE 3.8 ELECTROPHORETIC ANALYSIS OF THE PCR PRODUCTS OF THE U6 AND H1 HUMAN PROMOTERS.....	122
FIGURE 3.9 ELECTROPHORETIC ANALYSIS OF THE PCR PRODUCTS OF THE pB12MCS-U6-EGFP PLASMID COLONY SCREEN .....	123
FIGURE 3.10 ELECTROPHORETIC ANALYSIS OF THE PCR PRODUCTS OF THE COLONY SCREEN OF THE pB12MCS-H1-EGFP VECTOR.....	124
FIGURE 3.11 ELECTROPHORETIC ANALYSIS OF THE PCR PRODUCTS OF THE pB12MCS-U6-EGFP PLASMID (COLONY 11) .....	125
FIGURE 3.12 ELECTROPHORETIC ANALYSIS OF THE PCR PRODUCTS OF THE EXTRACTED pB12MCS- H1-EGFP PLASMIDS (COLONIES 3,4, 8, AND 9).....	126
FIGURE 3.13 THE SECONDARY STRUCTURE OF shRNA1 ANTISENSE OLIGONUCLEOTIDE TARGETING THE b3A2 FUSION POINT .....	128
FIGURE 3.14 ELECTROPHORETIC ANALYSIS OF PCR PRODUCTS FROM shRNA CASSETTE SYNTHESIS.....	129
FIGURE 3.15 ELECTROPHORETIC ANALYSIS OF THE COLONY SCREEN PCR PRODUCTS OF THE U6 shRNA1 CONSTRUCT WITH T7 FORWARD AND T3 REVERSE PRIMERS.....	130
FIGURE 3.16 ELECTROPHORETIC ANALYSIS OF THE pB12MCS-EGFP EXTRACTED PLASMID WITH THE U6 shRNA1 CONSTRUCT TARGETING THE b2A2 FUSION GENE.....	132
FIGURE 4.1 HISTOGRAM GENERATED BY FACS DEMONSTRATING THE TRANSFECTION EFFICIENCY OF THE NIMT FeOfECTION/YELLOW TRANSFECTION REAGENT IN K562 CELL LINES. ....	141
FIGURE 4.2 HISTOGRAM GENERATED BY FACS DEMONSTRATING THE TRANSFECTION EFFICIENCY OF THE NIMT FeOfECTION/YELLOW TRANSFECTION REAGENT IN KCL-22 CELL LINE.....	142
FIGURE 4.3 HISTOGRAM GENERATED BY FACS DEMONSTRATING THE TRANSFECTION EFFICIENCY OF THE NANOFECTIN TRANSFECTION REAGENT IN K562 CELL LINE.....	143
FIGURE 4.4 HISTOGRAM GENERATED BY FACS DEMONSTRATING THE TRANSFECTION EFFICIENCY OF THE NANOFECTIN TRANSFECTION REAGENT IN KCL-22 CELL LINE. ....	144
FIGURE 4.5 NUCLEOFECTION OF THE K562 CELL LINES WITH THE pB12MCS PLASMIDS.....	147
FIGURE 4.6 NUCLEOFECTION OF THE KCL-22 CELL LINES WITH THE pB12MCS PLASMIDS. ....	148
FIGURE 4.7 HISTOGRAM GENERATED BY FACS DEMONSTRATING THE TRANSFECTION EFFICIENCY OF NUCLEOFECTION IN THE K562 CELL LINE.....	149
FIGURE 4.8 HISTOGRAM GENERATED BY FACS DEMONSTRATING THE TRANSFECTION EFFICIENCY OF NUCLEOFECTION IN THE KCL-22 CELL LINE.....	150

FIGURE 4.9 ELECTROPHORETIC ANALYSIS OF PCR PRODUCTS OF GAPDH PRIMERS USING K562 CDNA .....	152
FIGURE 4.10 ELECTROPHORETIC ANALYSIS OF PCR REACTIONS USING THE ABL AND BCR PRIMERS ON K562 CDNA.....	153
FIGURE 4.11 ELECTROPHORETIC ANALYSIS OF PCR REACTION USING b3A2 AND b2A2 PRIMERS WITH K562 AND KCL-22 CDNA, RESPECTIVELY.....	154
FIGURE 4.12 THE QPCR OF GAPDH STANDARDS AND THE GENERATED STANDARD CURVE. ....	156
FIGURE 4.13 THE QPCR OF BCR STANDARDS AND THE GENERATED STANDARD CURVE.....	157
FIGURE 4.14 THE QPCR OF ABL STANDARDS AND THE GENERATED STANDARD CURVE.....	158
FIGURE 4.15 THE QPCR OF b3A2 STANDARDS AND THE GENERATED STANDARD CURVE. ....	159
FIGURE 4.16 THE QPCR OF b2A2 STANDARDS AND THE GENERATED STANDARD CURVE. ....	160
FIGURE 4.17 AN EXAMPLE OF AN ELECTROPHORETIC ANALYSIS OF THE PCR REACTION USING RPL13 PRIMERS ON TRANSFECTED K562 CELLS CDNA.....	162
FIGURE 4.18 THE PERCENTAGE OF EXPRESSION OF b3A2 MRNA LEVELS IN K562 CELL LINES TRANSFECTED WITH SHRNA MOLECULES TARGETING THE b3A2 FUSION POINT. ....	165
FIGURE 4.19 PERCENTAGE OF EXPRESSION OF b2A2 MRNA LEVELS IN KCL-22 CELLS AFTER TREATMENT WITH b3A2 SHRNA CONSTRUCTS.....	166
FIGURE 4.20 THE PERCENTAGE OF EXPRESSION OF b2A2 MRNA LEVELS IN KCL-22 CELL LINES TRANSFECTED WITH SHRNA1 <sup>b2A2</sup> CONSTRUCT. ....	168
FIGURE 4.21 PERCENTAGE OF b3A2 MRNA LEVELS IN K562 CELLS AFTER TREATMENT WITH b2A2 SHRNA CONSTRUCTS.....	169
FIGURE 4.22 THE PERCENTAGE OF EXPRESSION OF THE NORMAL <i>BCR</i> AND <i>ABL</i> GENES IN K562 CELL LINE TRANSFECTED WITH SHRNA CONSTRUCTS TARGETING THE b3A2 FUSION GENE. ....	170
FIGURE 4.23 THE PERCENTAGE OF EXPRESSION OF THE NORMAL <i>BCR</i> AND <i>ABL</i> GENES IN KCL-22 CELL LINE TRANSFECTED WITH THE SHRNA1 <sup>b2A2</sup> CONSTRUCT. ....	171
FIGURE 4.24 WESTERN BLOT ANALYSIS OF THE b3A2 FUSION PROTEIN FROM K562 TRANSFECTED CELL LINE. ....	172
FIGURE 4.25 WESTERN BLOT ANALYSIS OF THE b3A2 FUSION PROTEIN FROM K562 TRANSFECTED CELL LINE. ....	174
FIGURE 4.26 WESTERN BLOT ANALYSIS OF THE b2A2 FUSION PROTEIN FROM THE KCL-22 TRANSFECTED CELL LINE.....	175
FIGURE 4.27 WESTERN BLOT ANALYSIS OF THE b2A2 FUSION PROTEIN FROM KCL-22 TRANSFECTED CELLS. ....	177

## Table of Tables

TABLE 1.1 THE CHARACTERISTICS OF CML CLINICAL PHASES.....	20
TABLE 2.1 COMPONENTS OF THE PCR MASTER MIX. ....	67
TABLE 2.2 COMPONENTS OF THE FAST START HIGH FIDELITY PCR SYSTEM.....	68
TABLE 2.3 PRIMER SETS USED WITHIN THE PROJECT.....	69
TABLE 2.4 RESTRICTION ENZYMES USED IN THIS PROJECT AND THEIR INCUBATION BUFFERS. ....	72
TABLE 2.5 THE DESIGNED SHRNA TRANSCRIPTS. ....	82
TABLE 2.6 PRIMER SEQUENCES USED TO CONSTRUCT THE SHRNA CASSETTES.....	85
TABLE 2.7 TAQMAN PRIMERS AND PROBE SETS USED IN THE PROJECT. ....	101
TABLE 2.8 PARAMETERS USED FOR REAL TIME PCR CONDUCTED IN THIS PROJECT.. ....	102
TABLE 2.9 COMPONENTS OF THE QPCR MASTER MIX 20 $\mu$ L REACTION MIXTURE. ....	103
TABLE 2.10 COMPONENTS OF THE QPCR MASTER MIX 20 $\mu$ L.....	103
TABLE 2.11 COMPONENTS OF THE LYSIS BUFFER AND THE SAMPLE LOADING BUFFER. ....	105
TABLE 2.12 RECIPES OF THE SOLUTIONS USED IN THE PROCESS OF SDS-PAGE ELECTROPHORESIS AND WESTERN BLOTTING. ....	107
TABLE 2.13 A LIST OF THE PRIMARY AND SECONDARY ANTIBODIES USED IN THE WESTERN BLOTS IN THIS PROJECT.....	108
TABLE 3.1 A SUMMARY OF THE SYNTHESIS AND SEQUENCING OF THE SHRNA CONSTRUCTS.....	133
TABLE 3.2 THE SHRNA SEQUENCING RESULTS OF THE PB12MCS-U6-SHRNA-EGFP CONSTRUCTS. ....	134
TABLE 3.3 THE U6 SEQUENCING RESULTS OF THE PB12MCS-U6-SHRNA-EGFP CONSTRUCTS. ....	135
TABLE 3.4 THE SHRNA SEQUENCING RESULTS OF THE PB12MCS-H1-SHRNA-EGFP CONSTRUCTS. ....	136
TABLE 3.5 THE H1 SEQUENCING RESULTS OF THE PB12MCS-H1-SHRNA-EGFP CONSTRUCTS.....	136
TABLE 4.1 A SUMMARY OF DIFFERENT TRANSFECTION METHODS ON THE TRANSFECTION EFFICIENCY OF K562 AND KCL-22 CELL LINES. ....	145
TABLE 4.2 A SUMMARY OF THE TRANSFECTION EFFICIENCY OBTAINED FROM FACS ANALYSIS AND CELL VIABILITY OF K562 AND KCL-22 CELL LINES BY NUCLEOFECTION WITH THE PB12MCS-U6-EGFP VECTOR.....	146
TABLE 4.3 A SUMMARY OF THE THREE NUCLEOFECTIONS CONDUCTED ON EACH OF THE HUMAN LEUKAEMIC CELL LINES, K562 AND KCL-22, AND THE CELL VIABILITY IN EACH OF THE TRANSFECTED CELL GROUPS.	151
TABLE 4.4 THE QPCR DATA OF THE b3A2 mRNA LEVELS OF THE TRANSFECTED K562 CELLS. ....	164
TABLE 4.5 THE QPCR DATA OF THE b2A2 mRNA LEVELS OF THE TRANSFECTED KCL-22 CELLS. ....	167
TABLE 4.6 STATISTICALLY ANALYSED DATA FROM THE WESTERN BLOT CONDUCTED ON THE b3A2 FUSION PROTEIN EXTRACTED FROM K562 CELLS THAT WERE TRANSFECTED WITH SHRNA1 <sup>B3A2</sup> , SHRNA3 <sup>B3A2</sup> , SHRNA1 <sup>B2A2</sup> , AND THE BLOT PROBED WITH C-ABL AB, AND BAND INTENSITIES MEASURED.....	173
TABLE 4.7 STATISTICALLY ANALYSED DATA FROM WESTERN BLOT CONDUCTED ON THE b2A2 FUSION PROTEIN FROM KCL-22 CELLS TRANSFECTED WITH WITH SHRNA1 <sup>B2A2</sup> , SHRNA1 <sup>B3A2</sup> , SHRNA3 <sup>B3A2</sup> , AND THE BLOT PROBED WITH b2A2 AB, AND BAND INTENSITIES MEASURED.....	176

## List of Abbreviations

<b>A</b>	<b>Adenine</b>
<b>AAV</b>	<b>Adeno-associated viral</b>
<b>A<sub>260nm</sub></b>	<b>Absorbance at 260nm</b>
<b>Ab</b>	<b>Antibody/s</b>
<b>ALL</b>	<b>Acute lymphoid leukaemia</b>
<b>AML</b>	<b>Acute myeloid leukaemia</b>
<b>ATP</b>	<b>Adenosine triphosphate</b>
<b>bcr</b>	<b>Breakpoint cluster region</b>
<b>BM</b>	<b>Bone marrow</b>
<b>BSA</b>	<b>Bovine serum albumin</b>
<b>C</b>	<b>Cytosine</b>
<b>cDNA</b>	<b>Complimentary DNA</b>
<b>CLL</b>	<b>Chronic lymphoid leukaemia</b>
<b>CML</b>	<b>Chronic myeloid leukaemia</b>
<b>CrkL</b>	<b>Crk Like</b>
<b>DEPC</b>	<b>Diethylpyrocarbonate</b>
<b>cfu</b>	<b>Colony forming Unit</b>
<b>CML</b>	<b>Chronic Myeloid Leukaemia</b>
<b>CMV</b>	<b>Cytomegalovirus</b>
<b>DNA</b>	<b>Deoxyribonucleic acid</b>
<b>dNTP</b>	<b>Deoxyribonucleotide triphosphate</b>
<b>DMSO</b>	<b>Dimethylsulfoxide</b>
<b>dsRNA</b>	<b>Double stranded RNA</b>
<b>DTT</b>	<b>Dithiothreitol</b>
<b>E.Coli</b>	<b><i>Escherichia coli</i></b>
<b>ECL</b>	<b>Enhanced chemiluminescence</b>
<b>EDTA</b>	<b>Ethylenediaminetetraacetic acid</b>
<b>eGFP</b>	<b>Enhanced Green Fluorescent protein</b>

<b>FACS</b>	<b>Fluorescence-Activated Cell Sorter</b>
<b>FBS</b>	<b>Foetal bovine serum</b>
<b>G</b>	<b>Guanine</b>
<b>GAP</b>	<b>GTPase activating protein</b>
<b>GAPDH</b>	<b>Glyceraldehyde-3-phosphate dehydrogenase</b>
<b>GDP</b>	<b>Guanosine diphosphate</b>
<b>GEFs</b>	<b>Guanine nucleotide exchange factors</b>
<b>GFP</b>	<b>Green fluorescent protein</b>
<b>GMCSFR</b>	<b>Granulocyte monocyte colony stimulating factor receptor</b>
<b>Grb2</b>	<b>Growth factor receptor-bound protein 2</b>
<b>GTP</b>	<b>Guanosine triphosphate</b>
<b>H<sub>2</sub>O</b>	<b>Water</b>
<b>HEPES</b>	<b>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid</b>
<b>hr</b>	<b>hour/s</b>
<b>INF</b>	<b>Interferon</b>
<b>IPTG</b>	<b>Isopropyl-β-D-thio-galactoside</b>
<b>IRS1</b>	<b>Insulin receptor substrate 1</b>
<b>JAK</b>	<b>Janus kinase</b>
<b>KDa</b>	<b>Kilo-Daltons</b>
<b>Kb</b>	<b>Kilobase</b>
<b>LB</b>	<b>Lysogeny broth</b>
<b>Max</b>	<b>Maximum</b>
<b>M-bcr</b>	<b>Major breakpoint cluster region</b>
<b>m-bcr</b>	<b>minor breakpoint cluster region</b>
<b>μ-bcr</b>	<b>micro breakpoint cluster region</b>
<b>MCR</b>	<b>Multiple Cloning Region</b>
<b>μg</b>	<b>Micro-gram</b>
<b>miRNA</b>	<b>MicroRNA</b>
<b>miRNP</b>	<b>MicroRNA protein complex</b>
<b>Min</b>	<b>Minutes</b>
<b>μl</b>	<b>Micro-Litre</b>



<b>M-MLV</b>	<b>Moloney Murine Leukaemia Virus</b>
<b>ng</b>	<b>Nanograms</b>
<b>nt</b>	<b>Nucleotides</b>
<b>OD</b>	<b>Optical density</b>
<b>PBS</b>	<b>Phosphate buffered saline</b>
<b>PCR</b>	<b>Polymerase chain reaction</b>
<b>PDK</b>	<b>Phosphoinositide-dependent kinase</b>
<b>PFA</b>	<b>Paraformaldehyde</b>
<b>pg</b>	<b>Picogram</b>
<b>Ph</b>	<b>Philadelphia chromosome</b>
<b>Ph<sup>+</sup></b>	<b>Philadelphia chromosome positive</b>
<b>PIP2</b>	<b>Phosphoinositide-4, 5-triphosphate</b>
<b>PIP3</b>	<b>Phosphoinositide-3, 4, 5-triphosphate</b>
<b>PI3K</b>	<b>Phosphatidylinositol 3-kinases</b>
<b>Pol III</b>	<b>Polymerase III</b>
<b>Pri-miRNA</b>	<b>primary miRNA</b>
<b>PRL13</b>	<b>Ribosomal protein large 13</b>
<b>PVDF</b>	<b>Polyvinylidene fluoride</b>
<b>qPCR</b>	<b>quantitative PCR</b>
<b>RISC</b>	<b>RNA-induced silencing complex</b>
<b>RNA</b>	<b>Ribonucleic acid</b>
<b>RNAi</b>	<b>RNA interference</b>
<b>Rpm</b>	<b>Revolutions per minute</b>
<b>RT-PCR</b>	<b>Reverse transcription polymerase chain reaction</b>
<b>SCID</b>	<b>Severe Combined Immunodeficiency</b>
<b>SD</b>	<b>Standard deviation</b>
<b>SDS</b>	<b>Sodium dodecyl sulfate</b>
<b>SDS-PAGE</b>	<b>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</b>
<b>SH3</b>	<b>Src-homology 3</b>
<b>SHIP1</b>	<b>SH2 domain-containing inositol 5-phosphatase</b>
<b>shRNA</b>	<b>Short hairpin RNA</b>

<b>siRNA</b>	<b>Short interfering RNA</b>
<b>SOS</b>	<b>Son of sevenless protein</b>
<b>STAT</b>	<b>Signal transducer and activator of transcription</b>
<b>T</b>	<b>Thymine</b>
<b>TAE</b>	<b>Tris-acetic acid EDTA</b>
<b>TBS</b>	<b>Tris buffer saline</b>
<b>TE</b>	<b>Tris-Ethylenediaminetetraacetic acid</b>
<b>TFO</b>	<b>Triplex forming oligonucleotides</b>
<b>TKI</b>	<b>Tyrosine kinase inhibitor</b>
<b>UPL</b>	<b>Universal probe library</b>
<b>UV</b>	<b>Ultra violet</b>
<b>WBC</b>	<b>White blood cell</b>
<b>w/v</b>	<b>Weight per volume</b>
<b>X-gal</b>	<b>5-bromo-4-chloro-3-indolyl-b-D-galactoside</b>

# Chapter 1 : Introduction

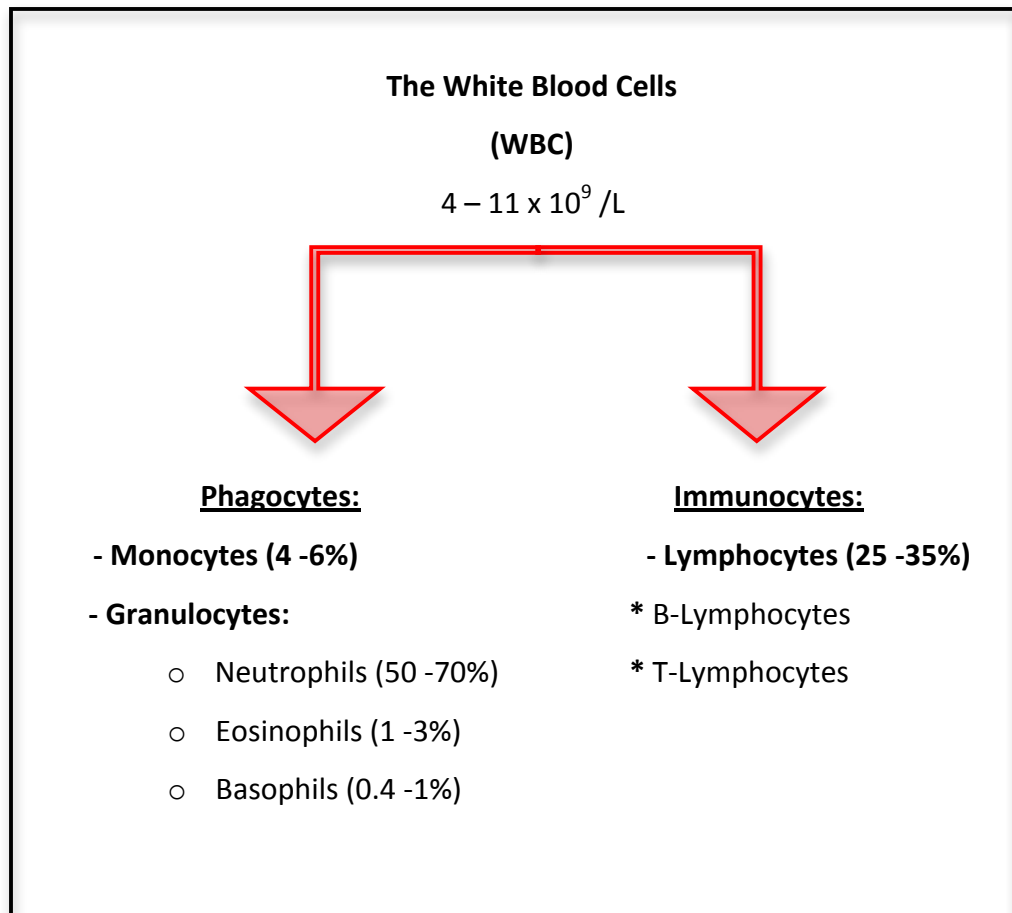
## 1.1 Haematopoiesis and Leukaemia

Haematopoiesis is defined as the process of white blood cell (WBC) production. The process starts with the haematopoietic stem cell, then goes through several steps of differentiation and proliferation, involving several haemopoietic growth factors and cytokines. Mature WBC are then produced.

There are two main types of WBC, the myeloid or the phagocytic lineage and the lymphoid or the immunocytic lineage. The myeloid lineage gives rise to the monocytes and granulocytes, the latter includes the neutrophils, basophils and eosinophils, whereas the lymphoid lineage gives rise to the B- and T-lymphocytes (Hoffbrand *et al.*, 2006). Figure 1.1

Leukaemias or blood cancers are a group of malignancies that involve the haemopoietic system (Sobecks and Theil, 2003) in particular the WBC. In all leukaemias the WBC proliferation dramatically increases; however, differentiation decreases to a point where the bone marrow (BM) becomes hypercellular and immature malignant cells are seen in the peripheral blood.

Four types of leukaemia exist depending on the cell lineages affected, the types of malignant cells seen and the disease progression. These are the acute myeloid leukaemia (AML), the chronic myeloid leukaemia (CML), the acute lymphoid leukaemia (ALL) and the chronic lymphoid leukaemia (CLL) (Hoffbrand *et al.*, 2006). This project considers CML, therefore only this type of leukaemia will be discussed in detail.



**Figure 1.1 Classification of the White Blood Cells (Hoffbrand et al., 2006).**

A normal adult has 4 -11x10<sup>9</sup>/L WBC in their peripheral blood. WBC can be mainly divided into phagocytes and immunocytes. The phagocytes are further divided into the monocytes and the granulocytes. Monocytes are agranulocytic cells making up 4-6% of the total WBC count, while the granulocytes consist of neutrophils, eosinophils and basophils that make up 50-70%, 1-3% and 0.4-1% of the total WBC count respectively. The immunocytes, on the other hand, consist of the T- and B- lymphocytes that make up 25-35% of the total WBC count.

## **1.2 Chronic Myeloid Leukaemia (CML)**

In 1845 Chronic Myeloid leukaemia (CML) was independently described and recognized by three German pathologists Virchow, Craigie and Bennett in Edinburgh, (Chopra *et al.*, 1999). However, it was not until the 1920s that the epidemiology of the disease was recognized (Chopra *et al.*, 1999). The incidence of CML is approximately 1 in 100,000 (Jemal A. *et al.*, 2002) with the

disease affecting both sexes. However, it is slightly higher in males (Sobecks and Theil, 2003) and increases with age (Michor *et al.*, 2006). The median age of presentation of CML is between 50 and 60 years, nevertheless, 12-30% of cases are diagnosed above 60 (Kalidas *et al.*, 2001).

### **1.2.1 Definition of CML**

CML is a myeloproliferative disorder that arises due to a clonal process involving the haematopoietic stem cells. The disease primarily affects the cells of the myeloid lineage, however, as the disease progresses it also affects the erythroid, megakaryocytic and lymphoid lineages (Faderl *et al.*, 1999).

The disease is characterised by having a severe increase in the WBC count, a mild to moderate anaemia and a hypercellular BM with a marked proliferation of all the granulocytic elements. The latter, leads to BM failure and cytopenias that will in turn lead to haemorrhage, recurrent infections and organ failure as the disease progresses to its advanced stages (Hoffbrand *et al.*, 2006).

### **1.2.2 Clinical Staging of CML**

CML is known to be a triphasic disorder, and the chronic phase is the first phase of the disease. 50% of Patients are usually asymptomatic at this stage of the disorder. However, if symptoms persist, they can include fatigue, weight loss, abdominal fullness, bleeding, sweats, splenomegaly, hepatomegaly (Advani and Pendergast, 2002), anaemia, leukocytosis, and thrombocytosis (Faderl *et al.*, 1999). It is then followed by the accelerated phase and subsequently the blastic phase (Sawyers, 1999). The last two phases are sometimes joined together and considered to be advanced-phase CML. The duration of the three phases together, if untreated, has a median span of 2

to 5 years. Usually the chronic phase lasts from 2-5 years; the accelerated phase can last up to one year, whereas the blastic phase may lasts for 3-6 months (Sessions, 2007). Patients may be diagnosed at any of the three stages. Table 1.1 summarizes the main characteristics of each phase. 85% of Patients are usually diagnosed in the chronic phase and 50% of them are diagnosed from routine tests where patients are asymptomatic (Faderl *et al.*, 1999): they visit their physicians for other unrelated issues, and end up being diagnosed through an abnormally elevated WBC (Sessions, 2007).

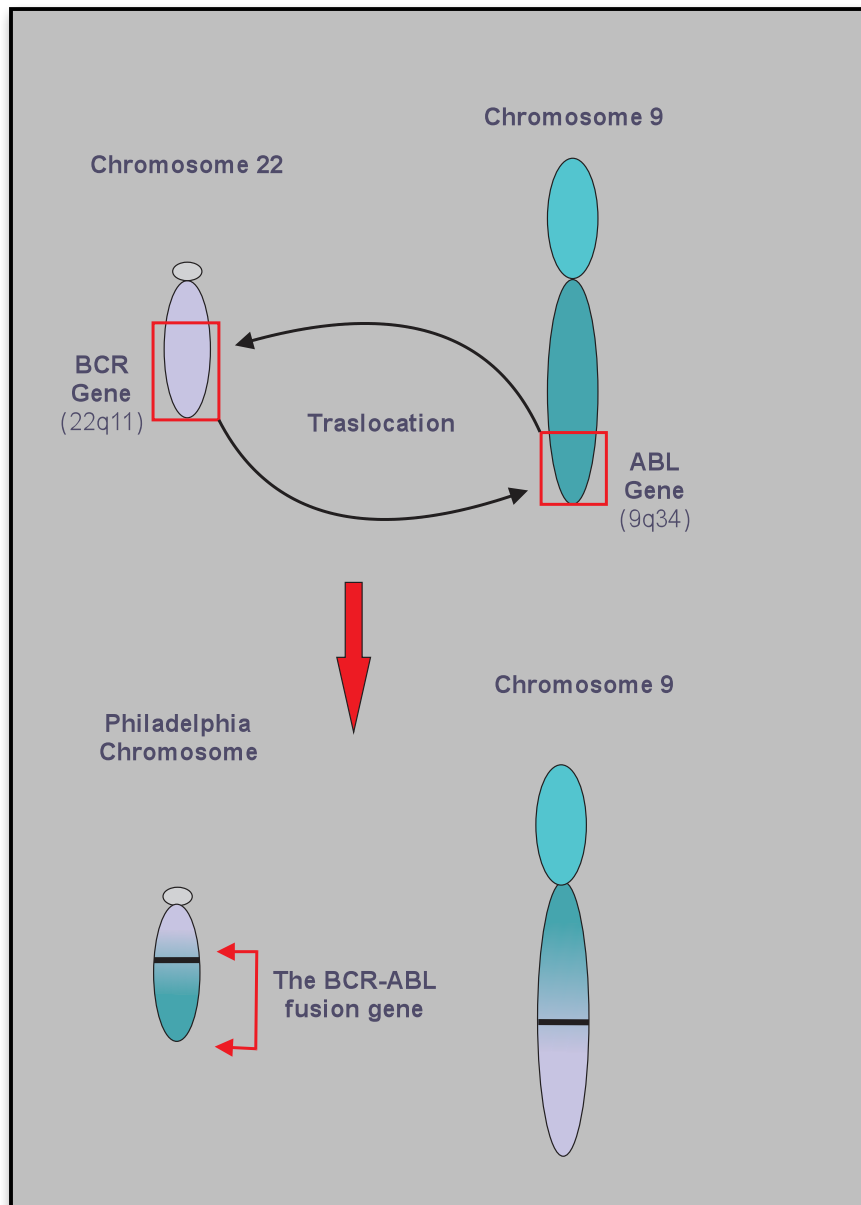
Name of CML clinical phase	Main characteristics
Chronic phase	<ul style="list-style-type: none"> <li>• Usually an asymptomatic phase. If symptoms persist, they may include splenomegaly, anaemia and hyperviscosity.</li> <li>• This phase is easily controlled by conventional therapy.</li> <li>• Lasts for an average of 3 years.</li> </ul>
Accelerated phase	<ul style="list-style-type: none"> <li>• WBC doubling every five days, hard to control by conventional therapy.</li> <li>• 20% of blasts and promyelocytes are found in the blood and BM.</li> <li>• 30% basophils and eosinophils are seen in the blood.</li> <li>• Splenomegaly in consistently increasing.</li> <li>• Anaemia and thrombocytopenia are non-respondent.</li> </ul>
Blastic phase	<ul style="list-style-type: none"> <li>• 30% of blasts and promyelocytes are found in the blood and BM.</li> <li>• Patients experience splenic and bone pain in addition to lymphadenopathy.</li> <li>• Median survival rate 3-6 months.</li> </ul>

**Table 1.1 The characteristics of CML clinical phases. (Hoffbran et al., 2006).**

### 1.3 The Molecular Biology of CML

CML was one of the first cancers to be associated with a chromosomal abnormality (Sawyers, 1999). In 1960 Nowell and Hungerford noticed an abnormally small chromosome in patients with CML (Nowell, 2007). Further studies by Rowley J.D in 1973, using quinacrine fluorescent and Giemsa staining band technique, revealed that the abnormal chromosome in these patients was produced due to a balanced translocation between chromosomes 9 and 22. This new chromosome came to be known as the Philadelphia chromosome (Ph): named after the city where it was first discovered. The chromosome plays a significant role in the development of the disease and it is found to be present in 95% of all CML cases (Evans and Sillibourne, 1996 and Wertheim *et al.*, 2002). Patients with t(9;22) translocation are termed Ph positive (Ph<sup>+</sup>).

In 1982 the first aspects of the molecular biology of CML were subsequently discovered. In that year a research group in the USA showed that the gene on chromosome 9, which was translocated to chromosome 22, was the human *ABL* (Heisterkamp *et al.*, 1982 and Heisterkamp and Groffen, 2002) proto-oncogene (Advani and Pendergast, 2002). It was then found that this gene was translocated into a limited region, around 5.8 kilo bases (KB), on chromosome 22. This region was termed the breakpoint cluster region (bcr) (Groffen *et al.*, 1984). Hence, the name of the gene on chromosome 22 is now known as the *BCR* gene. Figure 1.2 shows the translocation between chromosomes 9 and 22.



**Figure 1.2 The Reciprocal translocation between Chromosomes 9 and 22 producing the Philadelphia Chromosome.**

The *ABL* gene located in chromosome 9 is translocated on chromosome 22, where the *BCR* gene is located via a reciprocal translocation. This results in the Philadelphia chromosome that bears the abnormal *ABL-BCR* fusion gene.

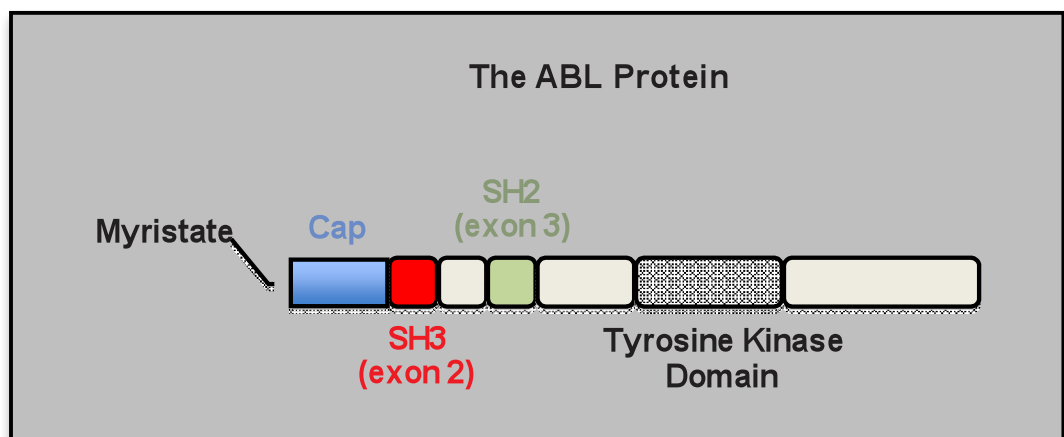
### 1.3.1 The ABL Protein

The normal *ABL* gene is composed of 11 exons (Laurent *et al.*, 2001), and transcribes the c-ABL protein. This 145 Kilo-Daltons (KDa) protein is a non-receptor tyrosine kinase. c-ABL has roles in assembling signalling proteins,



growth factor signalling, cytoskeletal organization, lipid metabolism, and transcriptional regulation (Chopra *et al.*, 1999). Two alternative first exons are present (1a and 1b) and depending on the alternative splicing of the transcribed RNA of the human *c-ABL* gene, two splice variants are present (Shtivelman *et al.*, 1986). Furthermore, the *c-ABL* type 1a is the spliced form of the protein and does not include a myristolation signal. In both variants, at the amino-terminus, a Cap region is present, then an Src-homology 3 (SH3) domain, followed by an SH2 domain, a tyrosine kinase domain, and finally a long carboxyl-terminal extension, which is also called the last exon region (Hantschel, 2004). The *c-ABL* protein also contains both nuclear localization and F-actin binding motifs. This explains the observations that the *c-ABL* protein can be seen in both the nucleus and the cytoplasm (Laurent *et al.*, 2001).

In addition, Pluk *et al.*, 2002, showed that the *c-ABL* protein is autoregulated, and the Cap, SH3, SH2 and the SH3-SH2 linker catalytic domain, play a role in autoinhibition of the normal *c-ABL* protein. The structural organization of this protein can be seen in Figure 1.3.

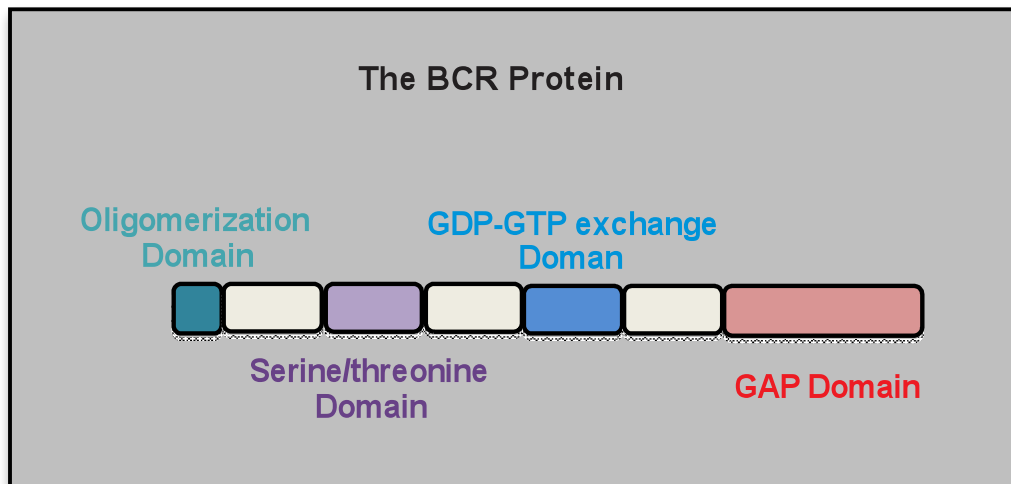


**Figure 1.3 The Structural Organization of the *c-ABL* protein.**

The *c-ABL* protein is 145 KDa, and consists of several structural motifs. These motifs include the Cap, the SH3, the SH2, and the tyrosine kinase domain. The protein is autoregulated and the kinase domain is autoinhibited in the normal protein.

### 1.3.2 The BCR Protein

The *BCR* gene spans 130kb and is composed of 23 exons. The gene is found to have two alternative exons; an alternative exon 1 and an alternative exon 2, which are usually spliced out (Laurent *et al.*, 2001). The gene encodes a 160 KDa protein. The protein has several structural domains that are illustrated in Figure 1.4. From the N-terminus of the BCR protein the following domains are present; an oligomerization domain (Zhao *et al.*, 2002), a serine/threonine kinase domain (Maru, 1991), SH2-binding domains (Laurent *et al.*, 2001), and a GDP-GTP exchange domain (Chopra *et al.*, 1999). At the C-terminus, a domain showing homology to the GTPase activating protein, GAP domain, is present (Diekmann *et al.*, 1991). The first two domains are encoded from the first exon of the gene and are preserved in all BCR-ABL variants (Laurent *et al.*, 2001), as all breakages in the *BCR* gene occur downstream to exon 1.



**Figure 1.4 The Structural Organization of the BCR Protein.**

The BCR protein is composed of several structural motifs. These include the oligomerization domain, the serine/threonine domain and the GDP-GTP exchange domain on the N-terminus of the protein. At the C-terminus, the GAP domain is present.

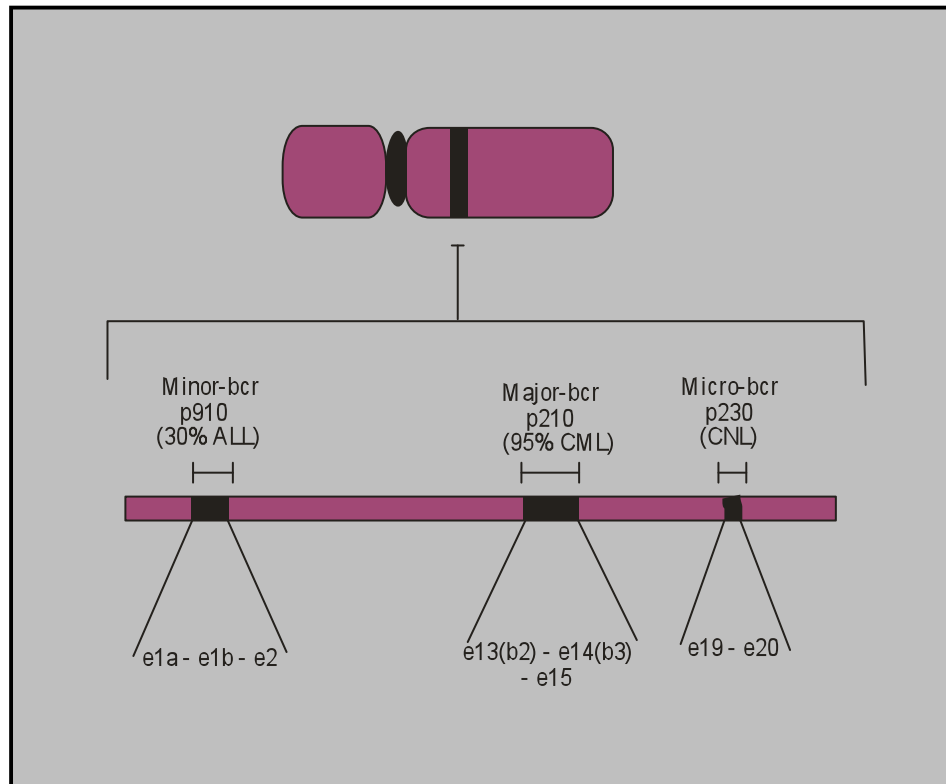
The oligomerization domain (a coiled-coil motif) facilitates hetero-oligomerization with the BCR protein. In the abnormal CML fusion protein this

motif allows homo-dimerization of the fusion protein that is necessary for the activity (Faderl *et al.*, 1999) and localization of the abnormal BCR-ABL protein. Deletions of this portion of the protein were found to decrease its binding to F-actin, because a F-actin binding motif is found within the ABL portion of the protein. It is found that the oligomerization domain is responsible for the cytoplasmic localization of the fusion protein (Laurent *et al.*, 2001) to one of the SH2-binding domains of the c-ABL protein. This interaction is thought to be essential for the oncogenic activation of the BCR-ABL protein (Laurent *et al.*, 2001). The GDP-GTP exchange domain serves as an important control element for the RAS family, and is a GDP-GTP exchange factor for other GTP binding proteins, whereas the GAP domain is associated with the RAS-related protein family, RAC and RHO. The GAP domain controls the rate of GTP hydrolysis of the active RAS proteins to their inactive form, bound to GDP (Faderl *et al.*, 1999). The structure of the BCR protein, therefore, makes it a multifunctional signal transducing molecule (Chopra *et al.*, 1999 and Laurent *et al.*, 2001). And because of the presence of the GDP-GTP exchange factor and GDP domains, see Figure 1.4, the protein seems to be especially concerned with phosphorylation, GTP binding of various proteins in several pathways that are involved in cytoskeletal organization, cell growth, and normal development (Laurent *et al.*, 2001). These pathways are further discussed in sections 1.3.5.1-1.3.5.4.

In terms of BCR expression, the protein is constantly expressed in various tissues and cells, indicating its role in general cell metabolism rather than being specific to a particular tissue (Maru, 1991). The protein's mRNA levels were found to be highest in brain and hematopoietic cells. Interestingly, the BCR protein is expressed in the first stages of myeloid differentiation, and the levels of the protein decreases significantly with leukocyte maturation (Laurent *et al.*, 2001). Finally, the protein is usually cytoplasmic. However, it has also been found to be associated with condensed DNA.

### 1.3.3 The *BCR-ABL* Fusion Gene

When the *c-ABL* gene is translocated from chromosome 9 to chromosome 22, the gene will be under the regulation of the *BCR* promoter (Shah *et al.*, 1991 and Zhu *et al.*, 1990). This translocation gives rise to an abnormal upregulated tyrosine kinase activity of the *BCR/ABL* fusion gene. The breakpoint within the *c-ABL* gene occurs within the first intron of this gene, leading to the fact that the resultant fusion transcripts always include the portion of the *ABL* gene from exon 2 onwards (Shtivelman *et al.*, 1985). The *BCR* part of the fusion gene has three different points at which a chromosomal breakage can occur. Three breakpoints exist, and are called the Major breakpoint cluster region (M-bcr), the minor breakpoint cluster region (m-bcr) and the micro breakpoint cluster region ( $\mu$ -bcr) (Chopra *et al.*, 1999 and Melo, 1997); this is how the *BCR* gene got its name. Figure 1.5 demonstrates these breakpoints. The M-bcr spans a 5.8kb region within the *BCR* gene (Groffen *et al.*, 1984). The breakpoints here span 5 exons historically named b1-b5, now known as exons 12 to 16 (Melo, 1996), and all breakpoints within this region are found to be intronic (Heisterkamp *et al.*, 1985).

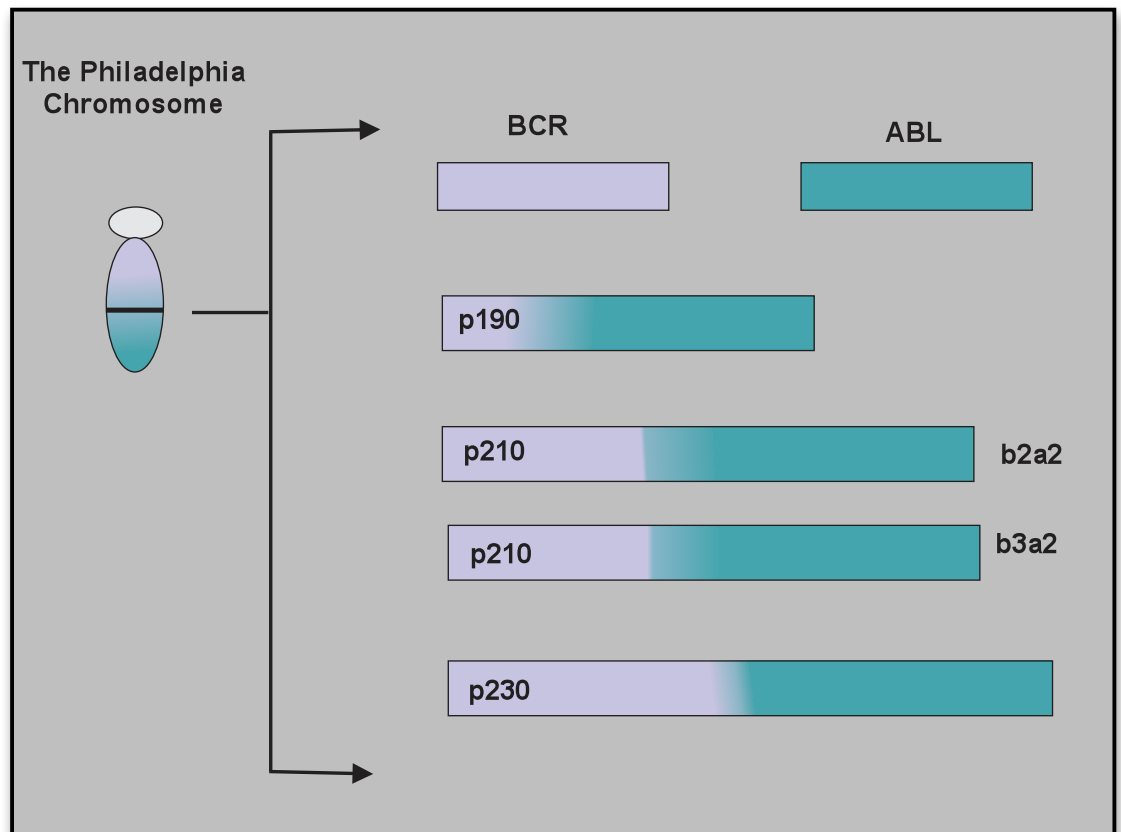


**Figure 1.5 The Breakpoints within the *BCR* gene found on chromosome 22.**

The position of the *BCR* gene is indicated by the dark transverse band shown on the long arm of chromosome 22. The gene is then further amplified to show the three points at which breakages may occur in the chromosome. These are the minor-bcr, the major-bcr and the micro-bcr breakpoints. The minor-bcr is located between exons 1 and 2, fusions to the *ABL* gene at this point give rise to the p190 fusion protein that is responsible for 30% of ALL cases. The major-bcr is between exons 13 and 15 and when fused to the *ABL* gene the p210 fusion protein results. This protein is responsible for 95% of CML cases. Finally, the micro-bcr is between exons 19 and 20, and fusions to the *ABL* gene at this point give rise to the p230 fusion protein that is associated with chronic neutrophilic leukaemia. Note that exons 13 and 14 are also known as  $\beta 2$  and  $\beta 3$  respectively.

The M-bcr occurs between exons 13 and 14 or 14 and 15 giving rise to two variants of the p210 protein that are associated with 95% of CML cases. The m-bcr occurs between the alternative exons 1 (1b and 1a) and exon 2. Breaks at this point give rise to the p190 protein that is associated with 15-30% of adult ALL and around 3-5% of childhood ALL. Finally, the  $\mu$ -bcr occurs between exons 19 and 20 of the *BCR* gene. The resultant protein is the p230 (Melo, 1997 and

Van Der Burg *et al.*, 2004) and it is found to be associated with chronic neutrophilic leukaemia (Advani and Pendergast, 2002 and Quackenbush *et al.*, 2000). Figure 1.6 demonstrates the different BCR-ABL fusion proteins. Note that the numbering of each protein corresponds to their size in kDa (Pane *et al.*, 2002).



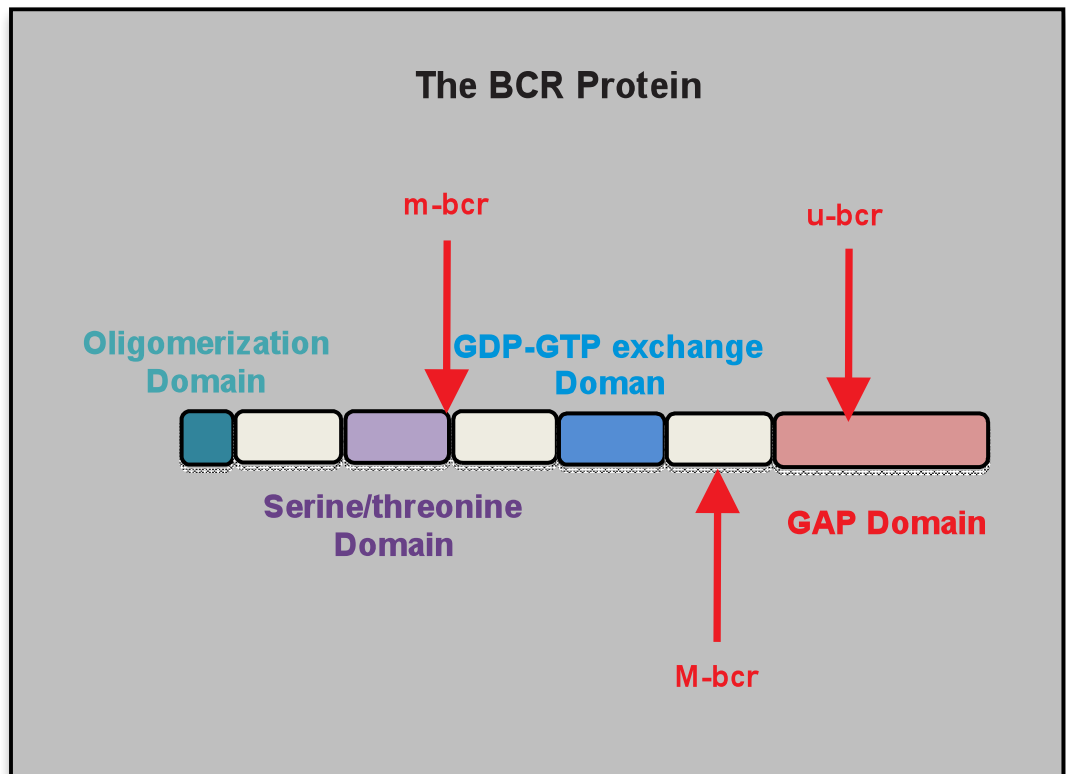
**Figure 1.6** The *BCR* and *ABL* genes, breakpoints, and transcripts.

Depending on where the chromosomal breakage occurs in the *BCR* gene, four transcriptional variants can result. Breakages in the m-bcr region give rise to the e1a2 variant that is transcribed into a 190 kDa protein, p190. Breaks in the M-bcr region may give rise to the b3a2 or the b2a2 variants both of which transcribe a 210 kDa protein, p210. Finally, breakages in the  $\mu$ -bcr region give rise to the e19a2 variant that transcribes a 230 kDa protein, p230.

Each of the different BCR-ABL proteins is associated with a different type of leukaemia, as demonstrated above. This strongly suggests that the type of leukaemia, phenotype outcome, depends highly on the domains of the *BCR*

gene included. The p190 is the smallest of the BCR-ABL proteins and is associated with cases of ALL whereas the p230, the largest of the BCR-ABL proteins, is associated with chronic neutrophilic leukaemia, which is considered to be a less aggressive form of leukaemia in comparison with ALL and CML (Melo, 1996 and Advani and Pendergast, 2002). Hence, the less the amount of the *BCR* gene included, the more aggressive the leukaemia.

The amount of the *BCR* gene being encoded in the BCR-ABL fusion protein correlates with the BCR domains present in the latter. Thus, in p190 protein only the oligomerization domain and the serine/threonine domains are present, whereas in the p210 protein, the previous domains in addition to the GDP-GTP exchange domain are present. Finally, the p230 protein includes all the domains present in the p210 protein and part of the GAP domain. This suggests the presence of a correlation between the structure of the BCR-ABL fusion protein and the disease phenotype (Melo, 1996). Figure 1.7 illustrates the relation between breakpoint cluster regions and the BCR domains. This project will only take into account the p210 BCR/ABL fusion protein due to its significant association with CML. Therefore only the M-bcr region will be further discussed.



**Figure 1.7 Schematic representation of the structural motifs of the normal BCR protein and their association with the breakpoint cluster regions.**

The red arrows show the positions of the three breakpoints. The first arrow represents the m-bcr. Breakpoints at this position results in a p190 bcr-abl fusion protein that retains the oligomerization domain and the serine/threonine domain. The second arrow indicates the M-bcr. Breakpoints here give rise to the p210 BCR-ABL fusion proteins. These proteins include the domains present in the p190 protein and the GDP-GTP exchange domain. The p210 fusion protein is associated with CML and has two subsequent variants ( $\beta 2\alpha 2$  and  $\beta 3\alpha 2$ ). Finally, the third arrow labelled  $\mu$ -bcr points to the last breakpoint. Breakpoints at this position results in the p230 BCR-ABL protein. This protein has all the domains with the exception of the GAP domain, where only a part of it is present.

### 1.3.4 The p210<sup>BCR-ABL</sup> Fusion Protein

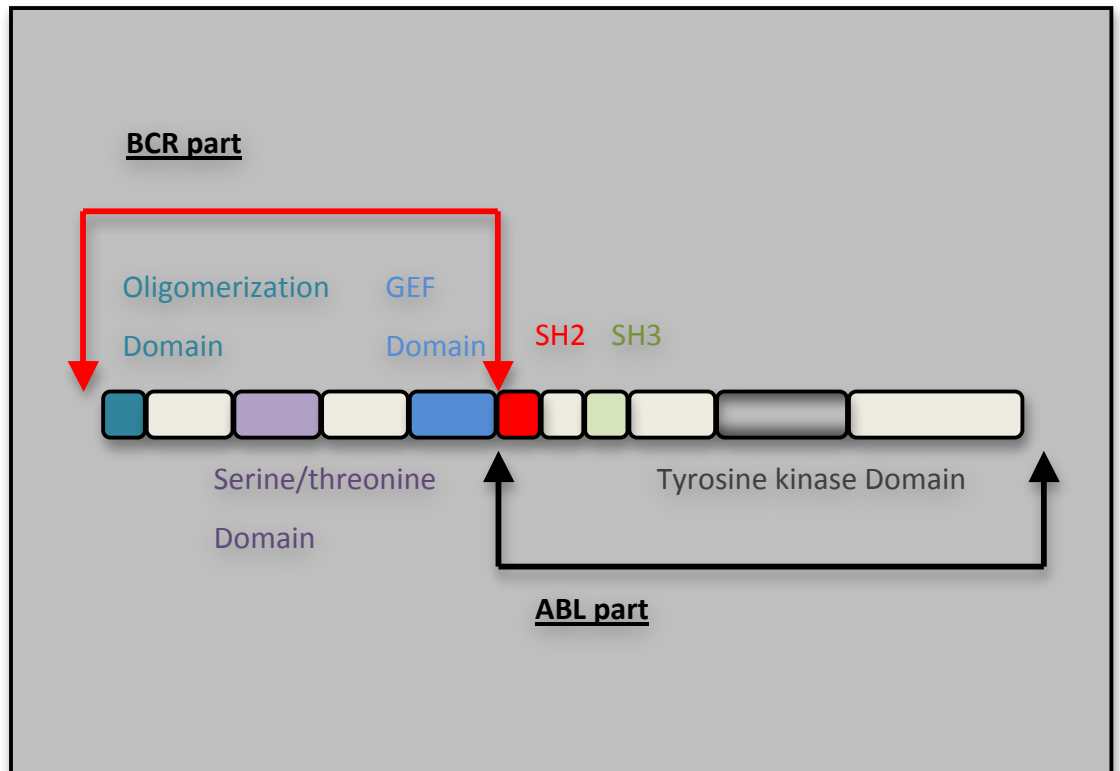
The M-bcr is around 5.8 kb in size (Pane *et al.*, 2002) and as figure 1.5 demonstrated the breakage occurs between exons 13/14 or 14/15, thus



retaining the coding region of the *BCR* gene from exon 1 to exon 13 or 14; the last two are also known as exons b2 and b3 respectively (Melo, 1996 and Chopra *et al.*, 1999). Moreover, breakages on the *ABL* gene occur between the alternative exons 1b and 1a, which are then spliced out.

The net result of the t(9;22) translocation is that the second exon of the *ABL* gene will fuse (at the gene's 3' end) to either exon 13 or 14 of the *BCR* gene (at its 5' end) (Pane *et al.*, 2002 and Laurent *et al.*, 2001). Hence, giving rise to the two BCR-ABL alternative transcripts, the  $\beta 2\alpha 2$  and the  $\beta 3\alpha 2$  variants, with the second variant having a longer transcript. (Advani and Pendergast, 2002) Both variants encode the p210 BCR-ABL protein in which the ABL tyrosine kinase is constitutively active. This constitutive activation is the result of oligomerization domain of the *BCR* gene that promotes tetramerization of the fusion protein (Faderl *et al.*, 1999), thus allowing the Serine/threonine domain of the BCR protein to bind to the SH2 domain of the ABL protein leading to activation of its tyrosine kinase domain (Laurent *et al.*, 2001). Figure 1.8 demonstrates the BCR and ABL domains included in the p210 BCR-ABL protein.

On the basis of the more *BCR* sequence included in the protein, the less aggressive the leukaemia, mentioned earlier, scientists were interested to examine if this was true for the p210 variants (Mills *et al.*, 1988, Leibowitz *et al.*, 1991, and Shepherd *et al.*, 1995). The  $\beta 3\alpha 2$  being initiated from a breakage in the 3' end of the M-bcr is the longer transcript, whereas the  $\beta 2\alpha 2$  was produced due to breakages in the 5' end of the M-bcr yielding shorter transcripts (Melo, 1997). Studies that examined the presence of a correlation between the p210 BCR-ABL variants and CML disease prognosis had different results. Some studies found that a correlation does exist; those studies implied that the  $\beta 3\alpha 2$  variant had a better prognosis and a longer chronic phase than  $\beta 2\alpha 2$  (Mills *et al.*, 1988 and Leibowitz *et al.*, 1991), others found no evidence (Shepherd *et al.*, 1995). From this point onward the BCR-ABL protein will only refer to p210 variant.



**Figure 1.8 Schematic representation of the structural motifs of the p210 BCR-ABL protein.**

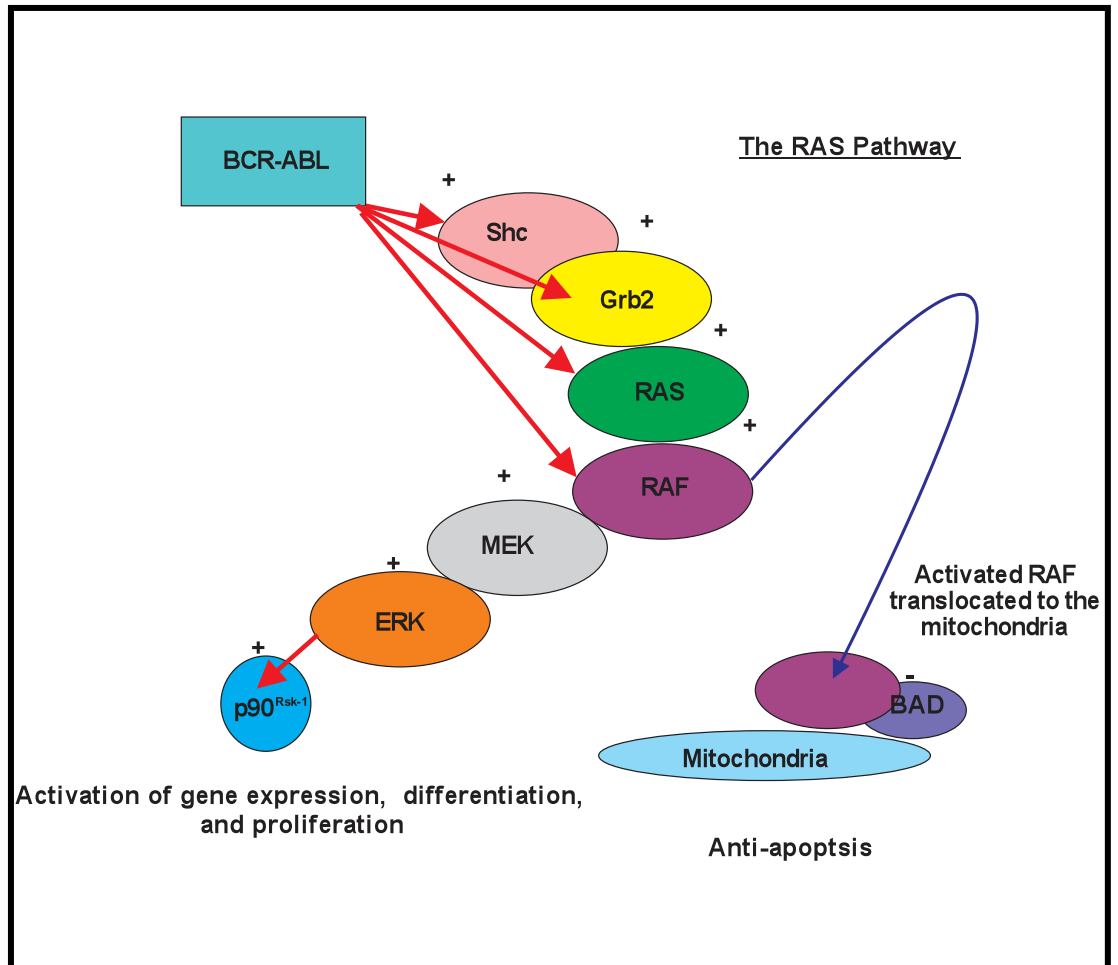
The BCR part of the p210 fusion protein is indicated by the red arrows and include the oligomerization, the serine/threonine, and the GEF domains. The ABL part of the fusion protein is indicated by the black arrows and include the SH2, the SH3, and the tyrosine kinase domains.

### 1.3.5 BCR-ABL Activity in CML

How the BCR-ABL protein's activity leads to leukaemogenesis is of great interest. There is no single protein, or one signalling pathway that has been identified to cause leukaemogenesis. In fact a multi-network of signalling pathways and different proteins are involved in this process (Heisterkamp and Groffen, 2002). Therefore, the main pathways involved in leukaemogenesis of CML due to the deregulation of the *BCR-ABL* gene will be described below.

### 1.3.5.1 The RAS Pathway

This pathway is a central signal transduction pathway that transmits signals from several cell surface receptors to several transcription factors in the nucleus. Ras protein is a small protein that is a common molecule for several other signalling pathways. The protein is in its active form when bound to GTP, and is inactive when bound to GDP. Thus, the regulation of the RAS protein can be achieved by GTPase activating proteins and GEFs (guanine nucleotide exchange factors). Mutations that result in the continuous expression of the active RAS protein are observed in around 30% of cancers (Steelman *et al.*, 2004). The Ras pathway is demonstrated in Figure 1.9.



**Figure 1.9 A Schematic diagram of the RAS pathway and the effects of the BCR-ABL on it.**

Illustrated is the RAS pathway. Under normal conditions the RAS pathway will be activated through the activation of Shc via a stimulated cell surface receptor (not shown here). The Shc+ will recruit Grb2, activate it, and in turn Grb2+ activates the downstream GDP bound RAS to the GTP bound RAS. RAS+ will activate RAF that in turns activates MEK that will activate ERK. Finally ERK+ activates p90<sup>Rsk-1</sup> that will make its way to the nucleus and cause activation of various transcriptional factors. It should also be noted that RAF can translocate to the mitochondrial membrane and through the interactions with Bcl2 family, especially BAD, cause anti-apoptosis effects by inactivation of the pro-apoptotic protein BAD. Finally, the red arrows from the fusion protein itself demonstrate the points at which the BCR-ABL protein activates the RAS pathway.

For activation of RAS, a cell surface receptor that can be a growth factor, mitogen, or cytokine receptor, should be first simulated. This will lead to the Shc adaptor protein to be associated with the simulated receptor. The Shc adaptor protein will recruit the Grb2 (growth factor receptor-bound protein 2) and SOS (son of sevenless protein). This interaction leads to loading the membrane bound RAS with a GTP, hence its activation (Steelman *et al.*, 2011). The activated RAS protein undergoes conformational change that leads to the translocation of the RAF protein to the cell's membrane where it binds to the active RAS and will be activated by phosphorylation. The RAF protein is involved in proliferation, differentiation, apoptosis induction, and further to cytokine stimulation. In addition, over expression of this protein abolishes cytokine dependence in the haematopoietic cells.

Furthermore, the activated RAF will in turn activate MEK1 that in turn activates ERK1 and ERK2 through phosphorylation (Steelman *et al.*, 2011 and Steelman *et al.*, 2004). This results in phosphorylation of transcription factors, proliferation, and apoptosis inhibition (Steelman *et al.*, 2004), which is achieved directly through the activated ERK where it translocate to the nucleus phosphorylation several transcription factors that in turn bind to promoters of growth factors and cytokines. The activated ERK can further activate downstream p90<sup>Rsk-1</sup> that will influence gene expression (Steelman *et al.*, 2011).

Activated RAF protein can be translocated to the mitochondria (Hindley and Kolch, 2002) leading to the alteration of the mitochondrial distribution of another downstream protein called BAD (Steelman *et al.*, 2004), a pro-apoptotic protein that will be inactivated (Henson and Gibson, 2006). This alteration has been seen to prevent apoptosis in hematopoietic cells (Steelman *et al.*, 2004). Thus, the activation of this pathway plays a role in neoplasia and to be more precise in leukaemogenesis. In CML the BCR-ABL protein has direct effects on many of the proteins within the RAS pathway. The effects of the BCR-ABL protein is through direct activation of Grb2 (Puil *et al.*, 1994, Chopra *et al.*, 1999 and Steelman *et al.*, 2011), SOS (Laurent *et al.*, 2001 and Steelman *et al.*, 2011), RAS and RAF (Steelman *et al.*, 2004 and Steelman *et al.*, 2011).

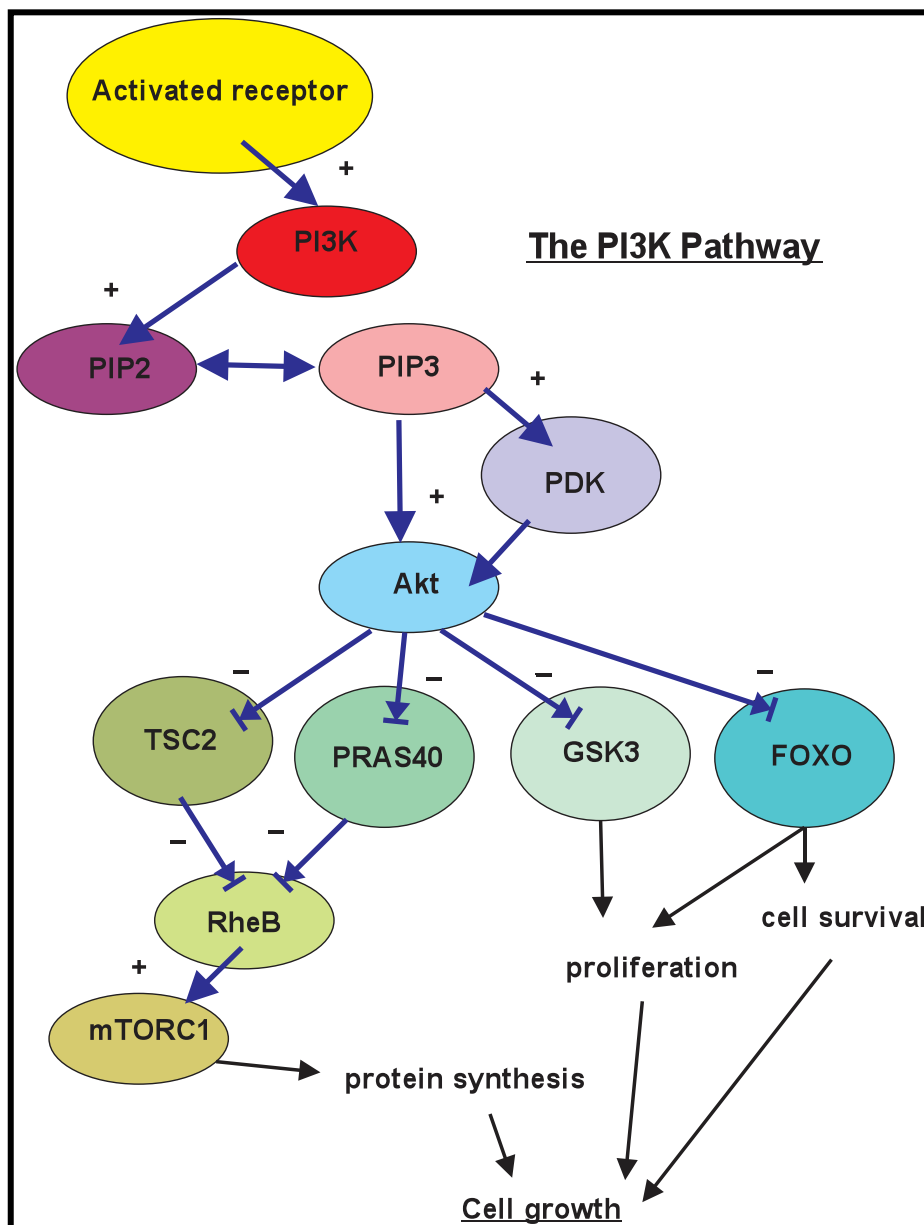
See Figure 1.9. In summary, it can be concluded that the BCR-ABL fusion protein can play a role in CML development through the activation of the RAS pathway.

### **1.3.5.2 The Phosphatidylinositol 3' Kinase Pathway**

In CML, BCR-ABL protein is seen to activate the PI3K pathway (Zhu *et al.*, 2011, Naughton *et al.*, 2009, Steelman *et al.*, 2004, Chopra *et al.*, 1999, and Faderl *et al.*, 1999). Phosphatidylinositol 3' kinase (PI3K) is a protein made of 2 subunits a catalytic subunit (p110) and a regulatory subunit (p85) (Chopra *et al.*, 1999). The protein's activity is involved with cell division, cytoskeletal organization, and inhibition of apoptosis (Steeleman *et al.*, 2004). The catalytic subunit of the protein is the part that interacts with receptor and non-receptor protein tyrosine kinases leading to the activation of PI3K (Chopra *et al.*, 1999). In addition PI3K can also be activated through RAS (Steeleman *et al.*, 2004). Once activated, the active PI3K will act on its substrate PIP2 (phosphoinositide-4, 5-triphosphate) generating PIP3 (phosphoinositide-3, 4,5-triphosphate) that will lead to the activation of PDK (Phosphoinositide-dependent kinase) and Akt. In addition, Akt is also a substrate for the active PDK, hence it will be activated by that protein (Georgescu, 2010).

From this point the activated Akt protein will lead to the phosphorylation of several target sites, these will include the TSC2 and PRAS40, GSK3, and FOXO. TSC2 and PRAS40 are two upstream inhibitors of mTORC1. mTORC1 is a protein that activates mTORC1, which is responsible for protein synthesis in cells. When Akt phosphorylates TSC2 and PRAS40, it will lead to their inactivation, hence there will be no negative regulation of mTORC1 and an increase in the activation of mTORC1 is established. The net result here is an increase in protein synthesis (Georgescu, 2010). Akt also negatively regulates GSK3, a protein that usually signals apoptosis. When Akt phosphorylates GSK3, the latter is inhibited leading to cell proliferation (Henson and Gibson, 2006 and Georgescu, 2010). Finally, FOXO proteins induce cell cycle arrest and apoptosis. They achieve this

through regulating the expression of various cell cycle regulators, and through up-regulation of pro-apoptotic members of the Bcl2 family. The active Akt protein again has a negative inhibitory effect on FOXO, thus leading to cell survival, proliferation, and apoptosis resistance (Georgescu, 2010). Figure 1.10 summarises the PI3K pathway.



**Figure 1.10 Schematic Diagram of the PI3K pathway.**

The PI3K pathway as demonstrated is rather a complex pathway. It is initiated by the activation of PI3K that in turn leads to a series of lipid-protein and protein-protein interactions. These interactions lead to the activation of the protein Akt. From here the protein negatively regulates further downstream substrates that usually inhibit cell growth, proliferation and protein synthesis. Akt negatively regulates downstream proteins leading to the net result of cell growth and anti-apoptosis. This complex pathway is activated in Philadelphia positive cells, hence, the proliferation and apoptotic resistance of these cells.



### 1.3.5.3 The JAK/STAT Pathway

CML also affects the JAK (janus kinase) family of kinases which are activated by signal transduction from cytokine receptors such as IL-3 and GM-CSFR (granulocyte monocyte colony stimulating factor receptor). Activated JAK protein will lead to activation of the STAT (signal transducer and activator of transcription) protein through phosphorylation. In turn the activated STAT proteins dimerize and move into the nucleus where they direct transcription and regulate gene expression (Constantinescu *et al.*, 2008) through binding to specific DNA elements (Henson and Gibson, 2006 and Chopra *et al.*, 1999). The down-regulation of the JAK/STAT pathway is achieved through the CIS/SOCS proteins. Where the CIS proteins inhibit STATs and the SOCS proteins bind and inhibit JAKs (Steelman *et al.*, 2004). The activity of the JAK/STAT pathway is highly increased in CML (Laurent *et al.*, 2001), and the BCR-ABL protein can directly activate STAT (Steelman *et al.*, 2004, Laurent *et al.*, 2001 and Noor *et al.*, 2011). This results in the growth factor independence observed in CML (Laurent *et al.*, 2001). Figure 1.11 gives a schematic diagram of this pathway.

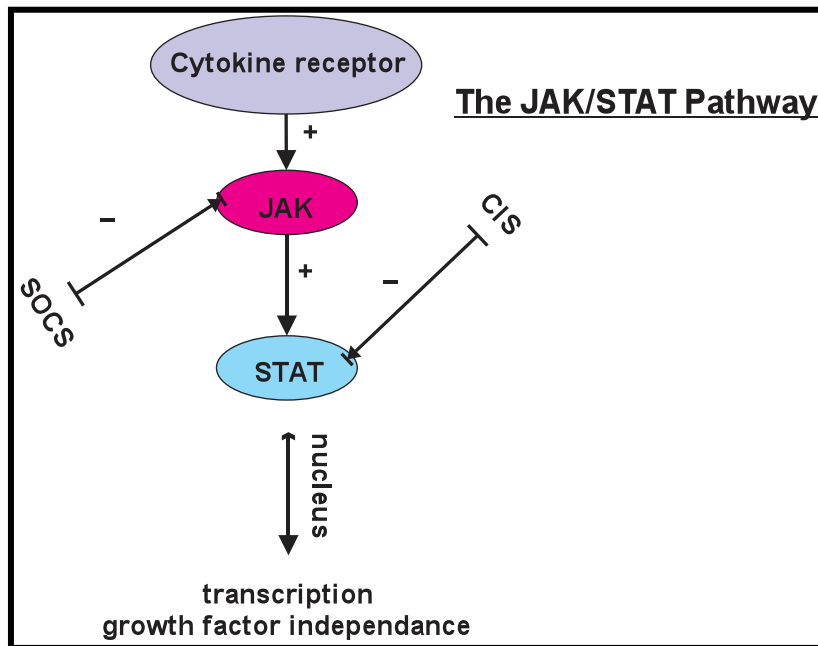


Figure 1.11 A Schematic diagram of the JAK/STAT pathway.

The JAK/STAT pathway is activated in CML, this is due to the direct activation effects of the BCR-ABL fusion protein on STAT. giving a net results of increased transcription and growth factor independence observed in CML cells. CIS/SOCS proteins negatively regulate the pathway.

The JAK/STAT pathway has also been found to be activated in BCR-ABL negative myeloproliferative leukaemias, thus emphasising the important role of this pathway in leukaemogenesis. In these leukaemias, the activation of the JAK/STAT pathway is initiated through a mutated JAK, where the protein is constitutively activated and bypasses the need for cytokine receptor activation (Constantinescu *et al.*, 2008). Consequently, scientists have taken an interest in developing Anti-cancer drugs targeting this mutated protein (Knoops and Constantinescu, 2011 and Noor *et al.*, 2011). In summary, the JAK/STAT pathway is involved in leukaemogenesis, and due to the effects of the *BCR-ABL* gene on STAT, the pathway is continuously activate in CML.

#### **1.3.5.4 Other Proteins Associated with BCR-ABL Activity**

Other proteins that are associated with the activity of BCR-ABL can include Crk Like (CrkL) and SH2 domain-containing inositol 5-phosphatase (SHIP1) proteins. CrkL protein has been noted to be associated with the fusion BCR-ABL protein, however, how this protein plays a role in haematopoietic diseases and CML is still not understood (Sriram and Birge, 2010). In spite of this, the interaction between CrkL and the BCR-ABL protein is confirmed and the phosphorylation levels of CrkL are considered as a prognostic parameter for CML patients (La Rosee *et al.*, 2008 and Xu *et al.*, 2011). A study done by Sriram *et al.* in 2011 has revealed that a tyrosine phosphorylation site on the CrkL is constitutively phosphorylated by BCR-ABL, this further confirms that the CrkL plays a role in CML; still more light is yet to be shed in this area.

SHIP1 protein has been shown to regulate haematopoiesis in mice and when disrupted leads to the development of myeloproliferative disorders (Laurent *et al.*, 2001). It is a negative regulator of the PI3K pathway by hydrolysing 5' phosphate of the PI3K substrate, PIP3, to its inactive form PIP2 (Ruschmann *et al.*, 2010). It has also been documented that the BCR-ABL protein inhibits SHIP1 (Steelman *et al.*, 2004) by tyrosine phosphorylation. This phosphorylation signals the proteasomal degradation of SHIP1 (Ruschmann *et al.*, 2010), thus abolishing its negative control on the PI3K pathway.

#### **1.3.5.5 BCR-ABL and Protein Pathways: A Conclusion**

In conclusion, the BCR-ABL has an upregulated tyrosine kinase activity. The BCR-ABL protein interacts with many other proteins that are involved in different pathways (Brehme *et al.*, 2009) leading to increased gene expression, protein synthesis, cell proliferation, differentiation, growth factor independence, and resistance to apoptosis. These effects all lead to the developments of leukaemic transformation and the development of CML.

The BCR-ABL activated pathways involve the RAS pathway, the PI3K pathway, the JAK/STAT pathway and other proteins such as CrkL and SHIP. These pathways are non-linear and are very complex as discussed above. The BCR-ABL protein can cause activation of these pathways at different levels and can also deactivate some negative regulatory events. This makes targeting the BCR-ABL fusion protein a highly appealing approach for CML therapy.

## **1.4 CML and Therapy**

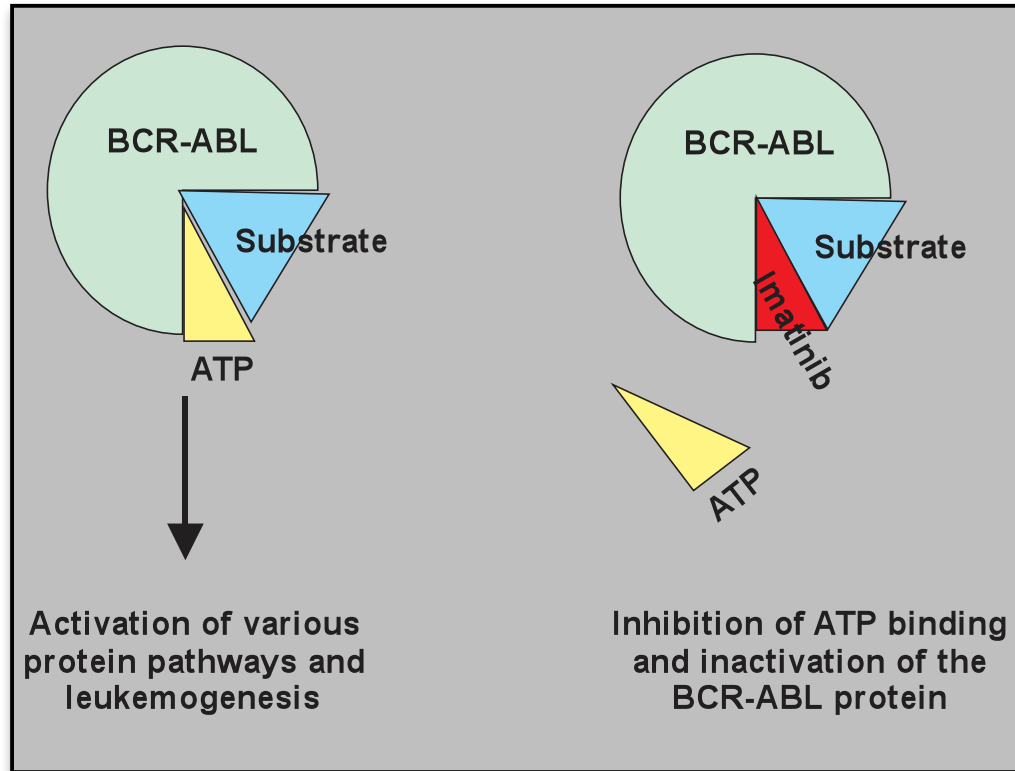
Like all cancers, the treatment approach in CML is multiple sessions of chemotherapy via the usage of drug cocktails (Hoffbrand *et al.*, 2006). Drugs for the disease have been developed over the years, differing in their actions and target sites. Historically, drugs such as busulfan and hydroxyurea, which fall under the name of cytoreductive chemotherapy, were used to control the WBC of patients with CML. However, this kind of treatment has not stopped the disease progressing to the accelerated phase and the blastic phase (Fausel, 2007). Then interferon (IFN) was developed and used. This drug showed a complete haematologic response in 50% of patients. However, IFN's adverse effects, toxicity, cost effectiveness, and inconvenience limited its benefits. Therefore, new strategies of treatment were required (Fausel, 2007).

### **1.4.1.1 Current Drug Treatment for CML**

With further understanding of the molecular pathology of CML, new drugs have been designed and developed. These drugs aimed at inhibiting the ABL tyrosine kinase activity (Milojkovic and Apperley, 2008, Puttini *et al.*, 2008) therefore, are named tyrosine kinase inhibitors (TKI). Three generations of TKI exist: first generation TKI include Imatinib, which to date is the most popular drug for CML. Second generation TKI include Dasatinib and Nilotinib. Finally,

the third generation TKI include Ponatinib and Bosutinib. Ponatinib is in clinical phase II (Vaidya *et al.*, 2011), while Bosutinib has just finished phase III clinical trials (Shieh *et al.*, 2011).

Imatinib, also known as STI571, is a drug made by Novartis and has a commercial name of Gleevec or Glivec (Druker, 2002). The drug works by competing with adenosine triphosphate (ATP) for its binding site to the *BCR/ABL* gene product. This inhibits the tyrosine kinase activity of the mutated gene, hence inhibiting its autophosphorylation, and subsequent activation of all its protein substrates that are involved in the various pathways discussed earlier. This will finally lead to the decrease in WBC proliferation and cell survival (Fausel, 2007 and Stentoft *et al.*, 2001). See Figure 1.12. Chronic phase patients on imatinib showed a 90% complete haematologic response. The drug initially showed promising outcomes, however, cases of drug resistance quickly emerged (Steinberg, 2007 and Talpaz *et al.*, 2006).



**Figure 1.12 The mechanism of BCR-ABL inhibition by Imatinib.**

In CML, BCR-ABL is able to phosphorylate downstream proteins when bound to ATP, and is therefore termed “active” in this state. The BCR-ABL protein switches from being active and inactive. When BCR-ABL is in its inactive state, Imatinib binds to the protein at its ATP binding site, locking the protein in its inactivated form, consequently, inactivating the protein pathways responsible for leukaemogenesis.

#### **1.4.1.2 Mechanisms of Imatinib Drug Resistance in CML**

Drug resistance to Imatinib can be classified into primary and secondary resistance. Patients with primary resistance illustrate a lack of effectiveness to the drug from the start of treatment. Those with secondary/acquired resistance show a loss in haematologic, cytogenetic and molecular response. The mechanisms by which resistance to Imatinib develop can be either BCR-ABL dependent or independent (Vaidya *et al.*, 2011).

#### 1.4.1.2.1 BCR-ABL Independent Mechanisms of Drug Resistance

BCR-ABL independent mechanisms include the persistence of insensitive haematopoietic stem cells; these cells seem to be persistent due to their over expression of the BCR-ABL protein (Vaidya *et al.*, 2011). Plasma enzymes inactivate Imatinib by binding and metabolizing it. This also demonstrates that the drug's plasma levels do not reflect the actual concentrations of the drug within the CML cells (Gambacorti-Passerini *et al.*, 2003). Another BCR-ABL independent mechanism includes drug efflux. Drug efflux alters the intracellular availability of the drug. It has been shown that some CML patients resistant to Imatinib over express proteins, such as P-glycoprotein, responsible for imatinib efflux. When those patients were treated with P-glycoprotein inhibitor their sensitivity to Imatinib was increased (Jabbour *et al.*, 2011). Finally, Imatinib resistance independent from BCR-ABL can also be due to the activation of the pathways implicated in CML due to mutations in the different proteins involved in these pathways. Many studies have demonstrated that mutations in proteins involved in any of the three main pathways involved in leukaemogenesis can independently lead to CML or a myeloproliferative disorder. Nambu *et al.*, 2010, showed that one mechanism of independent resistance was the phosphorylation of ERK1 and 2. Scientists here developed a K562R cell line that is highly resistant to imatinib and had phosphorylated ERK1/2, STAT 5 and Akt proteins. When treated with Imatinib, the drug reduced the phosphorylation of all proteins with the exception of ERK1 and ERK2. A study done by Esposito *et al.*, 2011, found that the levels of SHIP1 were decreased in cell lines resistant to imatinib, suggesting it as another mechanism for independent resistance. Another study showed that five cell lines, all resistance to Imatinib, had an increase in activity of the PI3K pathway due to a mutation in PI3K making the protein continuously active (Quentmeier *et al.*, 2011).

#### **1.4.1.2.2 BCR-ABL Dependent Mechanisms of Drug Resistance**

The BCR-ABL dependent resistance to Imatinib can be classified into three categories. Resistance due to mutations within the tyrosine kinase domain, outside the tyrosine kinase domain, and resistance due to over expression of BCR-ABL. Mutations within the tyrosine kinase domain all result in the inability of the drug to bind to its target site on the BCR-ABL protein (Vaidya *et al.*, 2011). An example is the T315I mutation that is the first BCR-ABL mutation to be discovered leading to Imatinib resistance (Gorre *et al.*, 2001). Mutations outside the kinase domain include mutations in the Cap, SH3, SH2, and SH3-SH2 linker domains. These domains are usually responsible for autoinhibition of the c-ABL protein; therefore, when mutated they will lead to the activation of the tyrosine kinase (Sherbenou *et al.*, 2010 and Vaidya *et al.*, 2011). Resistance due to over-expression of the BCR-ABL protein develops because the BCR-ABL protein is over expressed to a point where it can no longer be inhibited by the therapeutic dosage (Vaidya *et al.*, 2011).

As demonstrated a wide range of resistant mechanisms exist against the first generation TKI Imatinib leading to drug resistance in CML. Mutations leading to drug resistance can occur at any stage of the disease and is not limited to the accelerated or blast phase of the disorder (Roche-Lestienne *et al.*, 2002). Therefore, other TKI were developed. These include second and third generation TKI (Vaidya *et al.*, 2011).

#### **1.4.2 Second and third generation TKI**

New second generation TKIs for relapsed patients, resistant to Imatinib, have been developed. These include a drug called Dasatinib (Steinberg, 2007 and Talpaz *et al.*, 2006), developed by Bristol Pharmaceuticals in Switzerland, and FDA approved in October 2010 (Shieh *et al.*, 2011), and another called Nilotinib. The latter is developed by Novartis in the USA (Vaidya *et al.*, 2011),



and was FDA approved in June 2010 (Eskazan *et al.*, 2011 and Faber *et al.*, 2010 and Shieh *et al.*, 2011). These drugs were developed due to the cases of CML patient relapse. Nilotinib shows 10-25 fold more potency than the drug Imatinib, whereas Dasatinib shows between 100-300 fold greater activity. CML drug resistance to Nilotinib has also developed, and in some cases can be overcome by treatment with Dasatinib (Okabe *et al.*, 2011). Nevertheless, the first BCR-ABL mutation to give rise to imatinib drug resistance remains resistant to all three drugs discussed above (Skaggs *et al.*, 2006).

Third generation drugs, Ponatinib and Bosutinib, are in clinical phases with Bosutinib recently just completed phase III clinical trials (Shieh *et al.*, 2011). Data from the third generation TKI are yet to be collected and evaluated. Meanwhile, TKI available for CML treatment can be used for CML patients in any phase of the disease, with always a better prognosis when used in the chronic phase.

However, these drugs are only treatments not cures for CML. Patients tend to relapse, and the issue of drug side effects still remain. These will include toxicity, fluid retention, cardiac toxicity, platelet dysfunction and bleeding, immunosuppression, proinflammatory effects, and gastrointestinal side effects (Shieh *et al.*, 2011). The only cure available for this disease to date is allogeneic BM transplants (Fausel, 2007). Nevertheless, BM transplants are limited to younger patients (patients with CML are usually diagnosed between ages 50 – 60) and the availability of an appropriate HLA matching donor (Advani and Pendergast, 2002). Thus, making modulation of gene expression a promising strategy for potential CML treatment.

## **1.5 Modulation of gene expression**

Various strategies exist for modulating gene expression, and expression can be increased, decreased or even abolished (Lavrovsky *et al.*, 1997). For the

purpose of this study, only those that decrease or abolish gene expression will be considered.

Strategies that abolish gene expression can act at different molecular levels. These will include intervention at pre-transcription (DNA), post transcription (RNA), or post translational levels (protein). In past years, studies in down-regulating gene expression have focused on specific areas. These are, the formation of DNA triple helices, antisense, ribozyme, and RNA interference (RNAi) technologies. Each approach has its own advantages and disadvantages, depending on the nature of the pathological condition (Karagiannis and El-Osta, 2005 and Lavrovsky *et al.*, 1997).

### **1.5.1 DNA triple helices**

DNA triple helices, also known as triplex forming oligonucleotides (TFO), are short oligonucleotides designed to be complementary to a regulatory region in a gene (Duca *et al.*, 2008). Hence the binding of the TFO to their specific target sites in the major groove of the double-stranded DNA, results in the formation of a triple helix (Arya, 2011). The resultant triple helix will block the binding of various DNA binding proteins, such as endonucleases, transcription factors, DNA methylating enzymes and polymerases, resulting in down regulation of gene expression (Lavrovsky *et al.*, 1997). One example of the use of TFO technology to target the fusion gene in CML is a study done by Rapozzi *et al.* 2002. In this study a 13-mer TFO targeting the BCR-ABL promoter region was designed. The TFO<sup>BCR-ABL</sup> showed down regulation of the BCR-ABL mRNA expression in K562 cell lines.

Although TFO are attractive tools for modulating gene expression, limitations do exist. These include the dependency of the presence of an oligopyrimidine/oligopurine in the DNA target site for binding, stability of the triple helix (Duca *et al.*, 2008), inefficient cellular uptake, insufficient nuclear internalization and oligonucleotide aggregation (Rapozzi *et al.*, 2002).

### 1.5.2 Antisense oligonucleotides

This technology involves molecules that modulate gene expression by targeting mRNA sequences, and uses antisense oligonucleotides of single-stranded DNA or RNA molecules. These sequences are 13-25 nt long, complementary to a specific site on the mRNA molecule of the gene of interest, and the binding results in the formation of sequence specific double helices. The use of synthesized phosphodiester oligonucleotides have proven to be of limited use. This is because these oligonucleotides are easily degraded by intracellular endo- and exo-nucleases. Therefore, attempts to chemically modify these synthesized oligonucleotides have been made with the overall purpose to gain more stable, nuclease resistant, high mRNA target hybridization affinity oligonucleotides. Dias and Stein reviewed these chemical modifications, in 2002 (Dias and Stein, 2002).

The double-helix formation may then down-regulate gene expression in two ways: one is by blocking the translation of the bonded mRNA molecule, and the other, is through the activation of cellular ribonuclease called RNase-H. This ribonuclease specifically cleaves RNA that is in a RNA-DNA hybrid. Theoretically, only the desired gene will be affected, making this technology an effective and powerful tool in gene expression modulation of various diseases and disorders (Dias and Stein, 2002 and Lavrovsky *et al.*, 1997).

The use of antisense oligonucleotides has been applied to target the mRNA of the *BCR-ABL* gene in CML. Though it was reported to be successful in reducing the mRNA expression in SCID mouse tissue, previously injected with Ph+ CML blast crisis cell line, and increased the duration of survival in treated mice (Agarwal and Gewirtz, 1999). Its clinical applications have been held back due to lack of stability, and the requirements of high dosages to achieve an anti-leukaemic effect. These dosages can also lead to unspecific binding and off-target toxicity (Quintas-Cardama, 2008).

### 1.5.3 Ribozymes

Ribozymes are RNA structures with a specific catalytic activity. They are able to bind to other RNA molecules; thus breaking and/or forming covalent bonds with extraordinary specificity, thereby accelerating the rate of these reactions (Castanotto *et al.*, 2002).

Naturally occurring ribozymes act in a *cis* manner. These include hammerhead and hairpin ribozymes, that are derived from plants, RNase P (Kijima *et al.*, 1995), and group I and group II ribozymes that are found as introns in organelles and bacteria (James and Gibson, 1998). However, ribozymes can be also engineered to act in a *trans*-acting manner (Castanotto *et al.*, 2002 and Kijima *et al.*, 1995).

In the hammerheads, hairpins and group I introns, the specificity of a ribozyme to a particular target site depends on base-pairing between the ribozyme and its RNA target (Kijima *et al.*, 1995). In the RNase P ribozymes, it depends on the pairing of a guide RNA with the RNA target (Castanotto *et al.*, 2002), while in cases of group II introns it is contingent on the pairing of the ribozyme to its target DNA (Lewin and Hauswirth, 2001).

Once the ribozyme is attached to its target site, it will form a specific complex that will catalyse the hydrolysis of the phosphodiester bonds, hence cleaving the target RNA (Lavrovsky *et al.*, 1997). These properties make ribozymes useful tools in gene therapy (Kijima *et al.*, 1995) especially against HIV and cancer (Mulhbacher *et al.*, 2010). Ribozymes have been also previously used against the mRNA of the *BCR-ABL* gene. A reduction in the mRNA levels was observed in several studies (Lange *et al.*, 1994, Mendoza-Maldonado *et al.*, 2002), however, other studies didn't replicate these findings but showed nonspecific cleavage of the normal c-ABL mRNA (James *et al.*, 1996). Limitations of ribozymes exist, these include ribozyme stability, efficient target binding, co-localization of target and ribozyme, and specificity of delivery (Castanotto *et al.*, 2002 and James and Gibson, 1998).

## 1.5.4 RNA interference

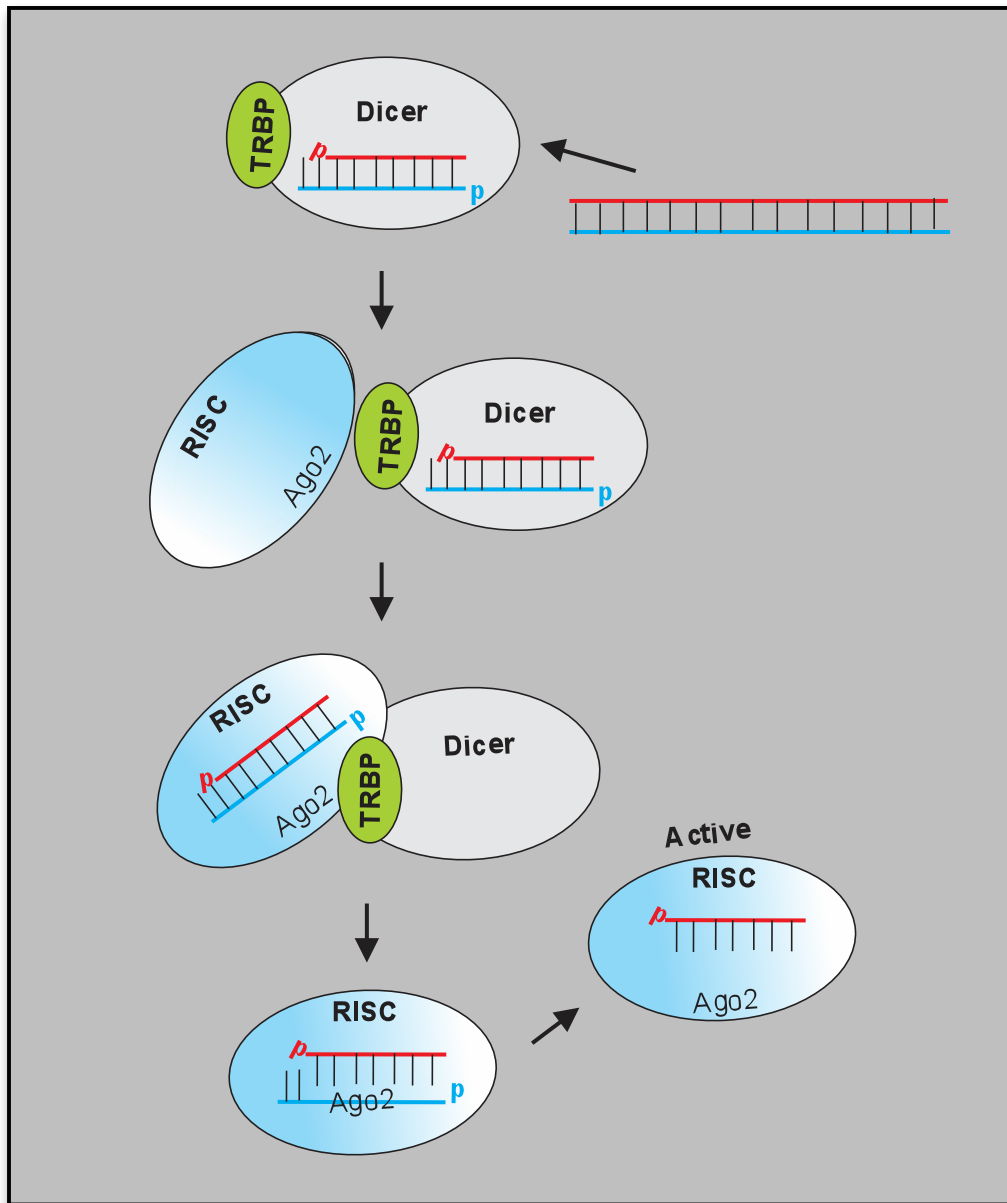
In 1998, Fire *et al.* observed that double stranded RNA (dsRNA) in *C.elegans* caused a substantially more effective RNA interference than using either an individual sense or anti-sense strand to modify gene expression. From that point on the process of RNA interference (RNAi) was discovered. RNAi quickly become evident as an ancient process. It is now known to be the oldest and most ubiquitous antiviral system in plants and animals, where RNAi is seen to be linked with post-transcriptional gene silencing in the above organisms (Sharp, 2001 and Elbashir *et al.*, 2001).

### 1.5.4.1 The Mechanism of RNAi

Basically, RNAi is the process whereby a double-stranded RNA (dsRNA) induces sequence specific post-transcriptional gene silencing. The dsRNA (>30 bp) does this by generating specific 21- and 22- nucleotide (nt) RNA fragments (Elbashir *et al.*, 2001). These fragments are generated by the cleaving action of a RNase-III family nuclease called "dicer", which is believed to be the first initiating step in RNAi (Bernstein *et al.*, 2001) and the fragments are called short interfering RNAs (siRNAs) (Cullen, 2002, and Elbashir *et al.*, 2001).

Dicer is associated with another protein called TRBP. The last recruits the RNA endonuclease Ago2 (Chendrimada *et al.*, 2005). Ago2 is also the catalytic component of the RNA-induced silencing complex (RISC) (Matranga *et al.*, 2005). The double stranded siRNA gets loaded onto Ago2 nuclease, and the passenger strand (sense strand) of the siRNA molecule is cleaved by the action of this protein. This cleavage occurs 9-10 nt counting from the 5' end of this strand (Rand *et al.*, 2005). This leads to the activation of the RISC complex that is guided by the siRNA guide strand (anti-sense strand) to its complementary target mRNA causing its degradation by the endonuclease activities of Ago2 (Matranga *et al.*, 2005).

In summary, the Dicer generated siRNA fragments are the mediators for RNAi and they do not exist independently. They are found associated with a number of proteins, which are still being studied in order to shed more light on their structure and how they contribute in the RNAi pathway (Wang *et al.*, 2009). Together these molecules form complexes called silencing complexes that recognize their single-stranded mRNA targets in a highly sequence specific manner. The silencing complexes will then cause direct cleavage of their target sites, thus destroying any mRNAs that perfectly match the guiding siRNA strands. This will prevent the mRNA molecule(s) from being translated into protein. It should also be noted that even with a single nucleotide mismatch between the siRNA and its target, the effect is greatly diminished and may be even eliminated entirely (Dykxhoorn *et al.*, 2003), making RNAi a highly precise and very specific tool. Figure 1.13 illustrates this process. Since the discovery of RNAi, this process has been used for various loss of function studies and other therapeutic approaches, which will be discussed later on. To date the application of RNAi has been mediated through synthetic siRNA, microRNA, and short hairpin RNA (shRNA) molecules.



**Figure 1.13 Schematic diagram of RNAi mechanism.**

dsRNA is recognised by the protein Dicer to which it binds. Once bound, the TRBP protein associated with Dicer recruits the RISC complex and the dsRNA gets loaded onto the Ago2 protein within the complex. The endonuclease action of Ago2 cleaves the passenger strand, shown in blue, and the RISC is then activated. The guide strand, shown in red, directs the activated complex to its complementary mRNA. The antisense strand binds to its target mRNA and through the nuclease action of Ago2, the target mRNA will be cleaved 10-12 nt from the 5' end of the guide strand, thus RNAi is achieved.

#### **1.5.4.2 MicroRNAs**

MicroRNAs (miRNAs) are a group of endogenously expressed small RNA molecules that are encoded by genes present in the host's genome. miRNAs do not lead to protein production, instead they act as regulators for mRNA expression. miRNAs are single stranded 21-23 nt long molecules that are produced by the cleavage of 70-90 nt shRNA precursor by the enzyme Dicer. This shRNA precursor is also known as pre-miRNA (Terasawa *et al.*, 2011). The miRNA molecules go on to associate with RISC that in turn induces translational repression by blocking the ribosomal translation (Karagiannis and El-Osta, 2005), and leading to inhibition of the translation of the target mRNA. Translational inhibition occurs when the miRNAs are partially complementary to the 3' untranslated region of their target sites. However, if the miRNAs are completely complementary to their targets, they will act in a way similar to siRNA molecules. Hence, by causing cleavage and degradation of their targets, see figure 1.14.

miRNA molecules have been identified in many multi-cellular organisms such as fruit flies, plants and humans (Pushparaj and Melendez, 2006 and Karagiannis and El-Osta, 2005). A study by Terasawa *et al.*, 2011, demonstrated that synthetic pre-miRNA based-shRNA had more potent RNAi activity than their corresponding siRNA molecules. In addition, they also observed that the guide strand of these pre-miRNA molecules was more efficiently incorporated into RISC. Thus, giving scientists further options in designing RNA molecules with more potent RNAi.

#### **1.5.4.3 siRNAs**

As stated earlier, siRNA molecules are the first discovered naturally occurring mediators of an RNAi response. They are 21- 22 nt long and are the products of dsRNA precursors. Synthetic siRNA molecules have been



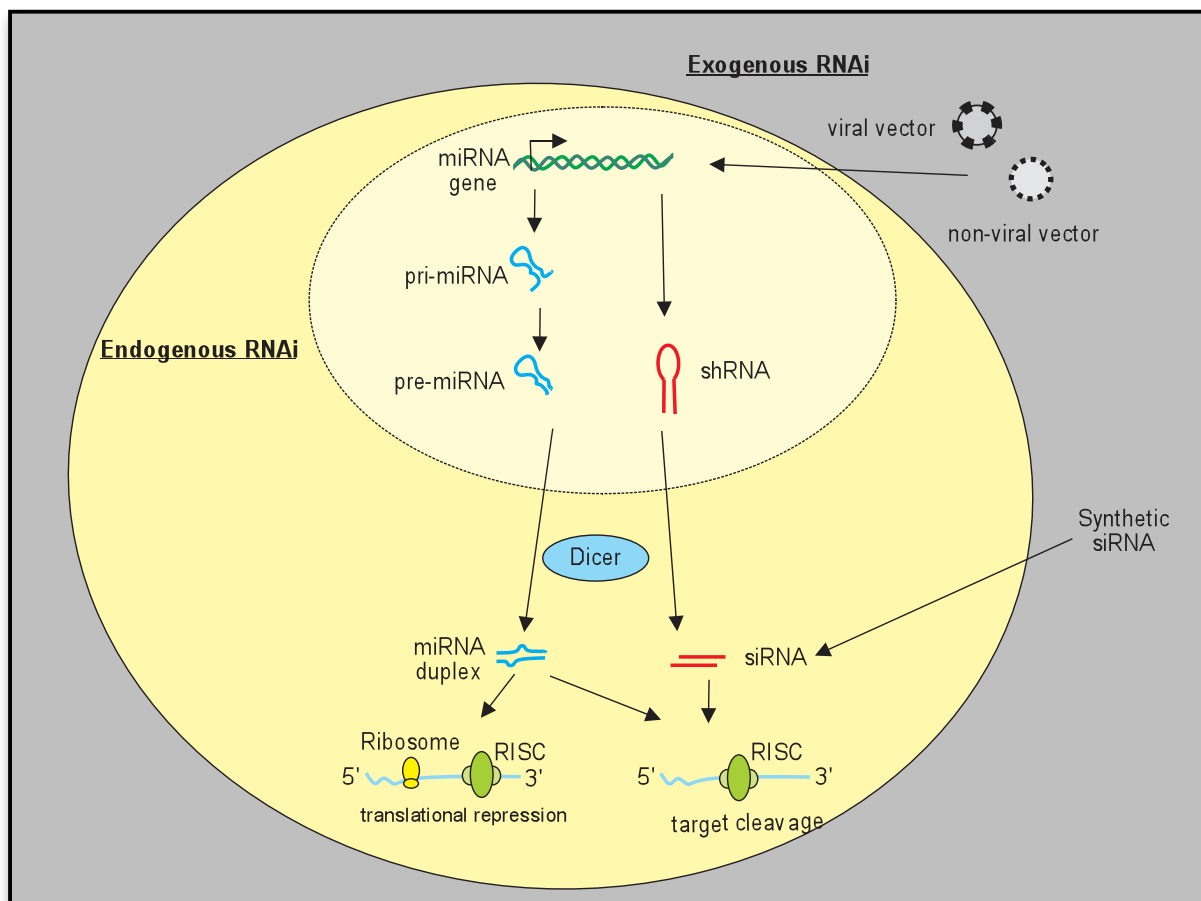
developed and used for loss of function studies or as therapeutic tools to modulate gene expression, of either an abnormal or an over expressed wild type gene (Dykxhoorn *et al.*, 2003). These synthetic siRNA can be introduced exogenously or by the expression of viral or non-viral vectors (Pushparaj and Melendez, 2006), see figure 1.14. The classical siRNA structures proposed by Elbashir *et al.* in 2001 are siRNA molecules that are 19 nt in length with 3' overhang of 2 nt. A fixed length, symmetry in structure and the 3' overhang, were structural design rules to achieve optimum RNAi. However, these rules were observed in *Drosophila* embryonic extract. However, a recent review by Chang *et al.*, 2011, shed light that siRNA structures used in mammalian cell lines can undergo more flexibility in design and chemical modifications based on the different siRNA structures that have been used with successful induction of RNAi to their target genes. Hence, the previously proposed fixed rules can be altered in mammalian cell lines, giving the advantage of developing siRNA molecules with more potent RNAi, reduced off target responses, and enhanced cellular delivery.

#### **1.5.4.4 shRNAs**

These RNA molecules are characterized by base-paired stems and a loop region, hence their name hairpin. shRNA is also processed by Dicer into siRNA that are 21 - 23 nt duplexes with a two base overhang at the 3' ends. The processing of such molecules yields mature products similar to the naturally occurring miRNA. Thus, competition between synthetic shRNA and naturally occurring miRNA for the endogenous RNAi machinery may result. However, to bypass this, the best approach is to select potent shRNA molecules capable of down-regulation of target gene when expressed in low levels (Rossi, 2008).

The use of synthetic siRNA is appropriate when transient inhibition of gene expression is satisfactory. However, in cases of prolonged gene silencing, several or continuous siRNA administrations will be required in order to

maintain a long-term inhibition of gene expression. Expression vector systems have been developed where the DNA is transcribed and processed into short RNA within the cells, see Figure 1.14. In these systems the expression of shRNA molecules are favourable (Karagiannis and El-Osta, 2005). These systems usually use RNA polymerase III promoters (Pol III), especially the U6 small nuclear RNA and the H1 RNase P RNA promoters. These promoters are known for their simplicity, and strength of expression (Rossi, 2008). They do not add any extra bases that might affect the functional activity of the transcript, and their expression is well known to be constitutive and ubiquitous (Ohkawa and Taira, 2000). Hence, the combination of an expression vector system that uses a Pol III promoter to induce a high RNAi through the expression of shRNA molecules is a recommended system when long term gene silencing is required.



**Figure 1.14 The Mechanisms of Endogenous and Exogenous expression of miRNA, shRNA, and siRNA.**

The process of RNAi can be achieved from either endogenous or exogenous triggers. The endogenous triggers of RNAi are the miRNA molecules. These are expressed from non-protein genes. miRNA are initially expressed as a long primary miRNA (pri-miRNA) that is further processed by an RNase-III type endonuclease enzyme called Drosha and its cofactor DGCR8 (Han *et al.*, 2004) into the miRNA precursors known as pre-miRNA. The pre-miRNA then exits the nucleus and gets processed by Dicer to imperfect 21- 23 nt long duplexes. These duplexes are taken up by RISC that cleaves the passenger strand and the mature miRNA directs the activated complex to its target mRNA where translational repression occurs. It should be noted that if the miRNA perfectly binds to its target mRNA it will lead to nuclease cleavage of its target. Exogenous RNAi is induced through the expression of shRNA constructs or the exogenous administration of synthetic siRNA. shRNA can also be administered into the cells via the use of viral vectors. Dicer processes the shRNA molecules into siRNA duplexes that are loaded into RISC. Within RISC the passenger strand is cleaved and the active complex is directed to the target mRNA. Complementary binding between the guide strand and target leads to the cleavage of the latter and the achievement of RNAi.

### 1.5.5 The use of RNAi as a therapeutic tool

With the developments of siRNA and shRNA based technologies, several research groups have directed their work and interests in using the RNAi technology for therapeutic applications. These applications included viral infections, neurodegenerative disorders, cancer (Karagiannis and El-Osta, 2005), dominant gene disorders (Seyhan, 2011) and targeting downstream modulators of disease (Pushparaj and Melendez, 2006). Nevertheless, several limitations do exist for the use of RNAi as a therapeutic agent. These limitations include the stability of the siRNA molecules, and the major drawback of delivery into tissues in addition to the specificity of delivery.

To overcome these limitations, various chemical modifications to the backbone of the siRNA molecules can be introduced. These modifications have been proven to increase the time life of the short RNA molecules. For the issues of delivery and specificity of siRNA/shRNA molecules, several delivery systems have been developed these can be mainly divided into viral and non-viral systems. Non-viral systems include polymer and lipid based systems. Here the short RNA molecules complexes with the transfecting reagent to form lipoplexes, dendriplexes, and polyplexes depending on the vector used, cationic lipids, dendrimer, or cationic polymers respectively (Shegokar *et al.*, 2011).

Viral vectors, on the other hand, have also been developed. These vectors provide sustained expression of their transcripts and have higher transfection efficiencies. Many vectors exist and the choice of the most suitable vector to be used highly depends on the therapeutic aim and tissue to be targeted. The most frequently used viral vectors will include adenoviral vectors, adeno-associated viral (AAV) vectors, retroviral vectors, and lentiviral vectors (Davidson and McCary, 2011). These vectors have the advantage of transducing both dividing and non-dividing cells. The adenoviral vectors are usually used for the delivery of shRNA or miRNA *in vitro*. However, they do not usually integrate into their target cell's genome. Thus, are used for short-term

and single gene expressions. Three generations of adenoviral vectors exist. AAV vectors offer stable and high-level gene expression. They don't usually integrate into their host genome but exist as episomal forms. There are several serotypes of AAV, thus giving a choice for selection of the most appropriate type depending on the cells of interest. Retroviral vectors, on the other hand, yield long-term expression of the transcript of interest. This is because the viral RNA is reverse transcribed into double-stranded DNA upon entry into the host cells. The transcribed DNA will then integrate into the host's chromosome. Nevertheless, with time, persistent gene expression may not be sustained; this may be due to the occurrence of silencing transcriptional units. Another drawback in retroviral vectors is the risk of T-cell leukaemia development that has been previously observed in patients. Finally, lentiviral vectors, a subgroup of retroviruses, can produce persistent transgene expression, which can also be silenced over time. These vectors permit the integration in their host's genome within intronic regions of active transcriptional units. (Liu *et al.*, 2011).

With all the recent developments in RNAi, the technology has been rapidly improving since its discovery in 1998. Over the years many studies have been now approved and are in pre-clinical and clinical trials. In addition some RNAi therapeutic approaches have already completed phase II. These studies cover a wide range of diseases and disorders. Their clinical setting for RNAi therapy include 7 drugs for ocular and retinal disorders, 8 drugs for different types of cancers, 3 drugs for kidney disorders, 2 drugs for LDL lowering, and 5 drugs for antiviral infections (reviewed by Davidson and McCarty, 2011). These are promising developments for RNAi as a constantly evolving technology to become the therapeutic tool of the future.

### 1.5.6 RNAi and CML

CML was characterized as the first cancer to be linked with a chromosomal abnormality that is constant in 95% of CML cases. The disease results mainly from overexpression of the c-ABL tyrosine kinase activity. For many years the BCR-ABL abnormal protein have been an attractive site for modulators of gene expression. These include antisense oligonucleotides, ribozymes and RNAi, including both siRNA and shRNA molecules.

In this project, gene modulation of the expression of the abnormal *BCR-ABL* gene through RNAi is of interest. Several studies have attempted to target the BCR-ABL chimeric protein through RNAi by the use of either synthetic siRNA or shRNA molecules. These include studies done by Mahmoodabady *et al.*, 2010, Myssina *et al.*, 2009, Li *et al.*, 2003, and Arthanari *et al.*, 2010 for shRNA molecules, and Rapozzi and Xodo, 2004, Wohlbold *et al.*, 2003, and Bártová *et al.* 2005 for siRNA molecules. The studies using shRNA molecules targeted both fusion points,  $\beta 3\alpha 2$  (Mahmoodabady *et al.*, 2010, Myssina *et al.*, 2009, Li *et al.*, 2003, and Arthanari *et al.*, 2010) and the  $\beta 2\alpha 2$  (Li *et al.*, 2003). The studies used chemical and physical methods of transfection, and different shRNA sequences that were regulated and expressed from different promoters and vector systems. However, all publications reported a decrease in both mRNA and protein levels after shRNA treatment of only the  $\beta 3\alpha 2$  variant. The mRNA and protein levels of the  $\beta 2\alpha 2$  variant, however, were not affected (Li *et al.*, 2003). The findings of these publications will further be discussed in detail in chapter 5.

Other studies have also considered RNAi technology either as a therapeutic tool for CML, through targeting proteins down stream BCR-ABL, or to understand the many pathways involved in leukaemogenesis by loss of function analysis. Examples include a study done by Machado-Neto *et al.* 2011, that used an shRNA-lentiviral delivery system to study the knockdown effects of a protein called IRS1 (insulin receptor substrate 1) on the PI3K-Akt/mTOR and MAPK pathways in K562 cell lines. This study was done on the basis that

IRS1 is continuously phosphorylated and associated with BCR-ABL in K562 cell line (Traina *et al.*, 2003). Silencing of IRS1 did down-regulate the associated pathways, causing a decrease in cell proliferation and colony formation, but apoptosis was not affected.

Another study conducted by Albers *et al.* 2011, demonstrated the use of RNAi based systems for loss of function analysis. The group found that Raf1 protein is a crucial mediator of BCR-ABL dependent CML, through activation of the MAPK/ERK cascade. In this study the shRNA was expressed from a retroviral vector in a coupled expression system. Another study done by Scherr *et al.* 2006, have also demonstrated that SHP2, STAT5 and Grb2 proteins can be therapeutic targets in cases of drug resistance. They have identified these target proteins through the use of a lentiviral mediated RNAi. The study also demonstrated that co-expression of shRNA against the BCR-ABL and SHP2 proteins from a single lentiviral vector induces stronger colony formation inhibition when compared to using shRNA alone. Elmaagacli *et al.* 2005, found that the use of two siRNA molecules had additive effects in the induction of apoptosis in CML cells, in comparison to when each siRNA was used individually. In this study two synthetic siRNAs were co-transfected. One siRNA was against the *BCR-ABL*, while the other targeted the *Wilms' tumor* genes, a gene found to be abnormally over-expressed in leukemic cells.

Other studies have found that the use of RNAi technology can sensitize the drug resistant leukaemic cells to TKI, such as Imatinib and Nilotinib. These studies include those done by Wohlbold *et al.*, 2003, Rumpold *et al.*, 2005, and Koldehoff *et al.* 2010, respectively. In these studies the RNAi mediated knockdown was against the BCR-ABL transcripts, giving a basis to combining both RNAi technology and TKI, thus targeting the *BCR-ABL* fusion gene molecularly at both mRNA and protein levels. A study conducted by Mendonca *et al.* 2010, showed that siRNA against the *BCR-ABL* gene could be co-encapsulated with Imatinib giving a new therapeutic approach for CML and at the same time bypassing the effects of drug resistance.

In conclusion, since the discovery of RNAi by Fire *et al.* in 1998, the technology has been advancing at a high speed. A lot has been unravelled

about RNAi mediated gene silencing, the proteins involved in the process, and the different mechanisms of the different triggers of RNAi. The technology is rapidly becoming a standard tool for loss of function studies and an appealing therapeutic approach for a wide range of diseases and disorders including cancer in general and targeting leukemic fusion proteins in particular (Thomas *et al.*, 2006).

## 1.6 Cell Lines as a Model for CML

Since the establishment of the first CML cell line in 1975 by Lozzio and Lozzio, many other CML cell lines have been established. There are around 30 CML cell lines as listed in a review by Drexler in 1994, some of which express the  $\beta 3\alpha 2$  BCR-ABL variant, whereas others express the  $\beta 2\alpha 2$  fusion protein. This wide range of CML cell lines differ morphologically and phenotypically, giving scientists the opportunity to study the pathobiology of CML and the role of the BCR-ABL protein in the development of leukaemogenesis (Drexler *et al.*, 1999).

In 1975 the first CML cell line, the K562 cell line, was established (Lozzio and Lozzio, 1975). These cells express the  $\beta 3\alpha 2$  BCR-ABL variant that is a free-floating suspension cell line (Lozzio and Lozzio, 1975). Cytologically, K562 cells are composed of undifferentiated myeloid cells (Lozzio and Lozzio, 1979) and molecularly, are a triploid cell line (Drexler *et al.*, 1999). K562 cells were used over the years since its establishment for various studies including those that shed light on the nature of *BCR-ABL* fusion gene (Heisterkamp *et al.*, 1983), or those that used the cell line for gene modulation (Wilda *et al.*, 2002, Rapoozi *et al.*, 2002, Mahmoodabady *et al.*, 2010, and Arthanari *et al.*, 2010). Because the cell line is easy to grow and maintain, in addition, molecularly they express the *BCR-ABL* gene in an oncogenic amplification manner (20 – 30 fold), this cell line provides a suitable CML model for gene modulation studies; therefore, the K562 cell line was selected for the purposes of the current study.



The KCL-22 cell line, on the other hand, expresses the  $\beta 2\alpha 2$  BCR-ABL variant. Kubonishi and Miyoshi, 1983, first established this cell line in 1983. KCL-22 cells are also free-floating in suspension, cytologically the cells are composed of immature undifferentiated haematopoietic cells; and molecularly it is a diploid cell line with double Ph chromosome, in addition to a normal chromosome 9 and 22 (Kubonishi and Miyoshi, 1983). The double Ph chromosomes are not surprising as this cell line (and K562) was established from a CML patient in blast phase. In the CML blast phase/crisis other chromosomal abnormalities occur including trisomy #8, isochromosome of the long arm of chromosome #17, and double Ph chromosomes are observed (Kubonishi and Miyoshi, 1983, Drexler, 1994, and Drexler *et al.*, 1999), hence, actually giving evidence of the leukaemic origin of the KCL-22 cells. Rapozzi and Xodo, in 2004 used the KCL-22 cell line as a leukaemic model in attempt to down-regulate the expression of the  $\beta 2\alpha 2$  BCR-ABL by siRNA molecules, and Ohmine *et al.*, 2003, used these cells to establish a new Imatinib-resistant cell line, KCL-22/SR in 2003. A publication by Esposito *et al.*, in 2011 used a KCL-22/R (imatinib resistant) cell line to study the mechanisms of imatinib resistant in CML. The cells are also easy to grow and maintain, and have a rate of doubling every 24hr (Kubonishi and Miyoshi, 1983), thus, for the purpose of this study, this cell line was chosen as it serves as an appropriate model for CML expressing the  $\beta 2\alpha 2$  BCR-ABL fusion protein.

## 1.7 Aim of the project

As previously mentioned, RNAi is becoming a tool for therapeutic approaches against several diseases and disorders, including cancer. CML was the first cancer to be characterized by the presence of an abnormal Philadelphia chromosome that transcribes an abnormal protein through as the *BCR-ABL* oncogene. This abnormal fusion gene has two variants; the  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$ , both playing an important role in leukaemogenesis of CML. TKI are

currently the first line of treatment for CML. However, first and second generations of TKI have failed to overcome the development of leukaemic drug resistance in relapsed patients, making modulation of the expression of the *BCR-ABL* oncogene a favourable approach. Using RNAi mediated gene silencing of the *BCR-ABL* oncogene through stable expression of shRNA<sup>BCR-ABL</sup> would be an excellent therapeutic approach for CML. The combination of two or more shRNA, targeted against one fusion variant or both together (in some cases of CML both fusion variants are expressed simultaneously (Adler *et al.*, 2009)), would be another therapeutic approach. This is based on the studies mentioned earlier where siRNA molecules had additive effects on CML. Finally, shRNA molecules against CML can also be used to sensitize leukemic cells to TKI, or the shRNA constructs can be administrated with TKI to establish complete hematologic response for patients with CML.

This project aims to decrease or abolish the expression of the two BCR/ABL variants. It aims to accomplish this by inducing RNAi through designing shRNA molecules against the two BCR-ABL transcripts. The designed molecules were then expressed and delivered via a non-viral system in two CML cell lines; one for each variant. The shRNA molecules to be expressed were designed to be as symmetrical as possible, with an 8 nt long loop, and were under the regulation of the human Pol III promoters, U6 or H1. The efficacy and specificity of the shRNA constructs were established as a prelude to these molecules being incorporated in a lentiviral system that would be tested on cell lines, and on primary CML cells from patient samples *in vitro*.

## **Chapter 2 : Materials and Methods**

### **2.1 Molecular Biology**

#### **2.1.1 RNA Extraction and Purification**

Cell lines were grown and  $1 \times 10^6$  cells were used for RNA extraction. RNA extraction was done using the RNeasy Mini Kit (Qiagen, UK). The extraction was done according to the manufacturer's instructions. The extracted RNA was further treated with DNase in order to remove any traces of contaminating DNA. This was done using the TURBO DNA-free™ Kit (Ambion Applied Biosystems, UK), following the manufacturer's protocol. The amount of the extracted RNA and the purity of the sample obtained are described in section 2.1.3.

#### **2.1.2 DNA Extraction and Precipitation**

To the DNA containing solution, an equal volume of phenol/chloroform mixture was added. The mixture was vortexed for a couple of seconds, kept at room temperature for 15 minutes, then centrifuged for 5 minutes at 14,000g. The layer containing the DNA, the top layer, was transferred to a fresh eppendorf. To that 1/10 the volume of 10M ammonium acetate and 2.5 the volume of 100% ethanol was added. The solution was vortexed for a couple of seconds, incubated on ice for 30 minutes, and then centrifuged for 5 minutes at 14,000g. The supernatant was removed, and the pellet was washed with 70% ethanol. The DNA pellet was air dried for 20 minutes, and then resuspended in an appropriate volume of TE buffer or DEPC H<sub>2</sub>O. The amount of the extracted and purified DNA, and the purity of the sample, is described in section 2.1.3.

### 2.1.3 Quantification of Nucleic Acid Concentration

The concentration of nucleic acids was quantified by measuring the diluted sample's absorbance at 260nm. This was done by the use of a GeneQuant spectrophotometric measurement (Pharmacia Biotech, Sweden), and depending on what the nucleic acid pellet was resuspended in, TE buffer or DEPC H<sub>2</sub>O was used as a blank. The following formula was used to estimate the nucleic acid concentration:

$$1.0 \text{ OD}_{260} = 44\text{ng}/\mu\text{l for RNA}$$

$$1.0 \text{ OD}_{260} = 50\text{ng}/\mu\text{l for DNA}$$

The purity of the samples was estimated through the OD<sub>260</sub>/OD<sub>280</sub> ratio. Ratios of 1.8 for DNA and 2.0 for RNA were considered to be highly pure.

### 2.1.4 cDNA Synthesis

cDNA synthesis is the production of a DNA complementary strand from an existing RNA strand. This process can be obtained by a reverse transcription (RT) reaction. The reaction is accomplished by the use of a reverse transcriptase enzyme that extends an RNA hybridised primer. Since, eukaryotic mRNA is 3' polyadenylated, a polydeoxythymidine oligomer (16-18 bases) can be used as a primer for this reaction, consequently giving rise to RNA-DNA hybrids (Gerard *et al.*, 1997).

For the RT reaction the following mix was prepared: a volume equivalent to 1-2µg of RNA, 1µl of oligo dT<sub>16-18</sub> (2µg/µl, Synthesized by the Molecular Biology Unit at Hodgkin Building, King's College London), and RNase free water (Qiagen, UK) to a total volume of 45µl was added. The mixture was incubated at 70°C for 10 minutes (Thermal Reactor- HYBAID Omnigene, UK). This was done to denature the RNA secondary structure.

The 45µl mixture was left on ice for 5 minutes and the following were added: 5µl of 10mM deoxynucleotide triphosphate (dNTP) (Promega, UK), 20µl of first strand 5x Buffer (250mM Tris-HCl: pH 8.3, 375mM KCl, 15mM MgCl<sub>2</sub>:

Invitrogen, UK), 10µl of 1M DTT (Invitrogen, UK), 5µl of RNasin Ribonuclease Inhibitor, 20-40u/µl (Promega, UK), and 5µl of M-MLV reverse transcriptase, 200u/µl (Promega, UK). The final volume of the RT mixture was 100µl which was incubated at 37°C for one hour (Techne DB, DRI-Block®, U.S.A).

### 2.1.5 Polymerase Chain Reaction

PCR involves the amplification of a DNA segment at high temperature by the use of sequence specific primers. The DNA template is initially denatured, primers then anneal to the template and a heat stable DNA polymerase extends them. This temperature cycle is repeated for a number of times, usually 30-40 cycles. The product of the first amplification becomes the template for the next one, and the amount of product increases exponentially, theoretically, two fold for each cycle.

All PCR reactions were carried out either with PCR master mix (Promega, UK) or Fast Start High Fidelity PCR system (Roche, UK). Tables 2.1 and 2.2 list all components that were used with the above PCR polymerases respectively.

Component	Volume per 25µl reaction
20ng of DNA	xµl
10µM Forward Primer	1µl
10µM Reverse Primer	1µl
PCR Master Mix	12.5µl
H <sub>2</sub> O	Add to 25µl

Table 2.1 Components of the PCR Master Mix (25µl reaction mixture).

Component	Volume per 25µl reaction
20ng of DNA	xµl
10µM ForwardPrimer	1µl
10µM Reverse primer	1µl
10mM dNTP	1.25µl
10x fast start high fidelity buffer	2.5µl
Fast start high fidelity <i>Taq</i> polymerase enzyme	1.25 U
H <sub>2</sub> O	Add to 25µl

**Table 2.2 Components of the Fast Start High Fidelity PCR system (25µl reaction mixture).**

The preparation of the reaction master mix was performed on ice, and the high fidelity *Taq* polymerase was added just before the initiation of the reaction.

The parameters set for the PCR systems are shown below and all sets of primers used for this project can be seen in table 2.3.

The following PCR conditions were applied for the reactions done by the use of the PCR Master Mix and the Fast Start high Fidelity PCR system:

Initial denaturation at 96°C for 5 minutes	
30 cycles of →	Denaturation at 95°C for 1 minute
	Annealing (temperature depending on the primer set, see table 2.3) for 50 seconds
	Elongation at 72°C for 15 seconds x length of the amplicon/Kb
Final extension at 72°C for 5 minutes	

The PCR products were analysed by agarose gel electrophoresis.

PCR Primer Set	Primer Set Sequence	Annealing temperature used
$\beta$ 3 $\alpha$ 2 primers*	Forward 5' d-AGCTTCTCCCTGACATCCGTGGA'3 Reverse 5' d-CCCATTGTGATTATAGCCTAAGA'3	60°C
$\beta$ 2 $\alpha$ 2 primers*	Forward 5' d-GCTACGGAGAGGCTGAAGAA'3 Reverse 5' d-CGTGATGTAGTTGCTTGGGA'3	60°C
U6 primers	Forward 5' d-CCC <b>AAGCTT</b> CAAGGTCGGGCAGGAAGAGGGCCTA'3 Reverse 5' d-AT <b>GGATCC</b> TAGTATATGTGCTGCCGAAGCGAGCAC'3	64°C
H1 primers	Forward 5' d-CCC <b>AAGCTT</b> CGAACGCTGACGTCATCAAC'3 Reverse 5' d-AT <b>GGATCC</b> GAAAGAGTGGTCTCATAACAAGTT'3	64°C
CMV-340	Forward 5' d-TTCCAAGTCTCCACCCATTGACG T'3	60°C
CMV-207	Forward 5' d-CCAATTGGCAGTACATCAAG'3	60°C
EGFPN1	Reverse 5' d-GTCCAGCTCGACCAGGATG'3	60°C
Sp6	5' d-ATTTAGGTGACACTATAG'3	50°C
pEGFP-N1	Forward 5' d-AACCCAATAGGGACTTTCC'3 Reverse 5' d-AAGTCGTGCTGCTTCATGTG'3	57°C
B-actin primer N	5' d-CTGAGTCTCCTTTGGAAGTCTGCAG '3	60°C
B-actin primer K	5' d-GAGCACAGAGCCTCGCCTTTC'3	60°C
B-actin primer L	5' d-GGATCTTCATGAGGTAGTCAGTCAGG'3	60°C
PRL13 Primers	Forward 5' d-TGAAGGAGTACCGCTCCAAA'3 Reverse 5' d-GGGTGGCCAGTTTCAGTTC '3	60°C
T3 reverse	5' d-GCAATTAACCCTCACTAAAGG'3	50°C
T7 forward	5' d-TAATACGACTCACTATAGGG'3	50°C

**Table 2.3 Primer sets used within the project.** Primers were aimed to be 21-23bp with a GC content of 48-60%. SGD web primer design software was used to design the primers. Red coloured sequences in U6 and H1 primer sets, indicate added restriction sites. These are HindIII and BamHI restriction sites located in the forward and reverse primers respectively. Primers indicated by (\*) were taken from the following publications, (Campanini F *et al.*, 2001) for the  $\beta$ 3 $\alpha$ 2 primer set and (Goh H.G *et al.*, 2006) for the  $\beta$ 2 $\alpha$ 2 primer set.

### 2.1.6 Agarose Gel Electrophoresis

DNA molecules carry a negative charge as a consequence of their phosphate backbone. Therefore, if DNA molecules are placed in an electrical field they will migrate towards the positive pole. An agarose gel is used for support, and comprises a network of pores, through which DNA molecules can travel. The pore size of a gel is determined by the agarose concentration, the greater the concentration of the gel, the smaller its pores.

To make a 1% gel, 0.5g of agarose powder (Fisher Bioreagents, UK) was weighed (Mettler R1210) and transferred into a clean conical flask that was then filled with 50ml of 1x TAE buffer (40mM Tris-Acetate-EDTA, 1mM EDTA pH 8.5, Sigma, Germany). The mixture was then heated on maximum power for 2 minutes in a microwave oven (Hotline, 930W Bejam, Korea). 10 $\mu$ l of ethidium bromide (1mg/ml, MBL International, USA) was then added and the solution was mixed.

An electrophoretic boat was placed horizontally into an electrophoretic tank (Max fill, Fisher brand, UK). A comb with the desired number of wells, was placed into the boat. The agarose was then poured into the boat and the gel was allowed to set. Once set, the comb was removed and the electrophoretic tank was then filled with 1x TAE buffer to a point where all the wells were covered. Samples were then mixed with 6x loading buffer (0.25% bromophenol blue, and 30% glycerol in 1x TAE buffer, BioLabs inc, England) and loaded into the wells. 300-450ng of the appropriate molecular weight marker was also loaded onto the gel. See Appendix 1 for all molecular weight markers used in this project.

The tank was then connected to an electrophoresis 3500 power supply (Pharmacia, UK) and a current between 90-100 volts was applied. Once the samples had migrated three-quarters of the way through the gel, the current was stopped and the gel was viewed under UV light (Gene flash, Syngene Bio Imaging, USA), and a picture was taken (Video graphic printer, UP-850, Sony, Japan). The DNA fragments of the size of interest were cut out of the agarose gel



and were purified by the use of the Wizard® SV Gel and PCR Clean-Up System according to the manufacturer's instructions (Promega, UK). The concentration of the purified DNA fragments was then estimated by means of OD<sub>260nm</sub> measurement.

### **2.1.7 Restriction Enzyme Digest**

DNA digestions were carried out on PCR products or plasmids. According to the manufacturer's protocol, 1 unit of the enzyme is required to digest 1µg of DNA in one hour under the optimum temperature and buffer conditions (Promega and New England Biolabs, UK). In this project, DNA fragments of interest were digested for a duration of 3 hours at 37°C with the appropriate restriction enzyme at its optimum 10x buffer. All restriction enzymes and their buffers used in this study are listed in table 2.4.

Restriction Enzyme	Restriction Buffer (100% suitable)	Manufacturer
BamHI	Buffer B: 6mM Tris-Cl pH 7.5, 6mM MgCl <sub>2</sub> , 50mM NaCl, 1mM DTT Or Multicore buffer: 25mM Tris-acetate pH 7.8, 100mM K(Ac), 10mM Mg(Ac) <sub>2</sub> , 1mM DTT	Promega
EcoRI	Buffer H: 90mM Tris-Cl pH 7.9, 6mM MgCl <sub>2</sub> , 50mM NaCl Or Multicore buffer	Promega
HindIII	Buffer E: 6mM Tris-Cl pH 7.5, 6mM MgCl <sub>2</sub> , 100mM NaCl, 1mM DTT	Promega
SacI	Buffer J: 10mM Tris-Cl pH 7.5, 7mM MgCl <sub>2</sub> , 50mM KCl, 1mM DTT	Promega
SnaBI	Buffer 4: 20mM Tris-acetate pH 7.9, 50mM K(Ac), 10mM Mg(Ac) <sub>2</sub> , 1mM DTT	New England Biolabs
XhoI	Buffer B: 6mM Tris-Cl pH 7.5, 6mM MgCl <sub>2</sub> , 150mM NaCl, 1mM DTT	Promega

**Table 2.4 Restriction Enzymes used in this project and their incubation buffers.**

### 2.1.8 DNA Ligation

Purified and digested PCR products were ligated into the vector pB12mcs-eGFP, that was kindly given by Dr Adam Rodaway (King's College London, Cardiovascular Division). The plasmid was also digested with the same restriction enzymes as that used for digesting the PCR products to be inserted. Following gel purification of DNA and estimation of its concentration by means of OD<sub>260nm</sub>, a molar ratio 10:1 insert to vector was used to establish ligation. To do so, the following equation was used:

$$\text{ng of insert} = (\text{ng of vector} \times \text{kb of insert size} / \text{kb of vector size}) \times (\text{molar ratio of insert : vector})$$

The ligation reactions were done in a total volume of 10 $\mu$ l, where 200ng of the vector, xng of the insert, 1 $\mu$ l of T4 DNA ligase enzyme (3U/ $\mu$ l), 1 $\mu$ l of 10x ligase buffer (Promega, UK) and a volume of x $\mu$ l of DEPC treated water were added. The reaction was left for overnight ligation at 4°C. The ligation reaction was further assessed by means of PCR, where the ligation mixture was diluted by 1000 and used as a DNA template. Primer sets used were either both located within the plasmid or at least one was within the plasmid while the other was within the insert. In addition, a negative control was established by the use of the same set of primers to PCR the wild type plasmid.

### **2.1.9 Transformations**

50 $\mu$ l of JM109 competent cells (>10<sup>7</sup> cfu/ $\mu$ g, Promega, UK) were thawed on ice for 5 minutes. 5 $\mu$ l of the ligation mixture was added and the cells were kept in a 1.5ml Eppendorf tube on ice for another 20 minutes. To heat shock the cells, the tube was then transferred to a heat block at 42°C (Grant BT1 Block Thermostat, UK) for 45-55 seconds then back on ice for another 2 minutes. 350 $\mu$ l of SOC medium (BioLine Ltd, UK) was then added to the cells, and the cells were incubated at 37°C for 45-90 minutes.

200 $\mu$ l of transformed competent cells were then spread on prepared LB agar plates. These plates were prepared from 16g of LB agar (Sigma, UK) that was dissolved in 500ml of Milli Q H<sub>2</sub>O (ELGA LA620, UK), autoclaved, then further boiled, left to cool down, and 60 $\mu$ g/ml of ampicillin was added. LB agar plates were stored at 4°C for a month from the date of preparation. The plates were placed overnight in a 37°C incubator (Luckham R300), and were examined for colony growth the next day.

By the use of the appropriate set of primers, the grown colonies were screened by means of PCR. A negative control, wild type plasmid, was also included. Positive colonies that contain the inserted DNA fragment were picked up from the agar plate and further grown in 20ml of LB broth medium (Fisher Scientific, UK). The LB medium was prepared by dissolving 12.5g of LB broth in

500ml of Milli Q H<sub>2</sub>O, autoclaved, and 60µg/ml of ampicillin was added just prior to bacterial colony addition. The 50ml tubes containing the LB medium were left overnight at 37°C in a shaker incubator (Gallenkamp Orbital incubator).

### **2.1.10 Plasmid DNA Isolation and Purification**

Solutions used include:

- Cell suspension solution (P1): 25mM Tris-HCL, pH 8, 10mM EDTA with 50mM glucose and DNase-free RNase (100µg/µl).
- Cell lysis solution (P2): 0.2N NaOH, 1% SDS.
- Potassium acetate solution (P3): 20 Volumes of 5M potassium acetate, 3 volume of Glacial acetic acid and 2 volumes of H<sub>2</sub>O.

Mini and Maxipreps were conducted on transformed bacterial cells, which were incubated in 20ml or 500ml of LB media respectively containing the appropriate antibiotic. The cells were grown overnight in a shaking incubator at 37°C and were harvested by centrifugation at 5000rpm at 4°C for 10 min (Mini) or 20 min (Maxi). The supernatant was discharged and the DNA isolation was performed by the use of the Qiagen Plasmid Mini or Endofree Plasmid Maxi kit that was conducted according to the manufacturer's protocol. The purity and the amount of the isolated DNA were determined as described earlier in section 2.1.3.

### **2.1.11 Glycerol Stocks**

Upon confirmation of clones of interest by means of PCR colony screening, each clone was picked up from a positive single colony and grown overnight in LB media containing a suitable antibiotic. 900µl of the overnight grown bacterial colony was pipetted into a 1.5ml Eppendorf tube containing 100µl of

glycerol (Sigma, UK). The tube was inverted several times and the glycerol stock was stored at  $-80^{\circ}\text{C}$ .

### **2.1.12 DNA sequencing**

5 $\mu\text{l}$  of plasmid DNA, at a concentration of 100ng/ $\mu\text{l}$ , was enclosed in a 0.5ml Eppendorf tube and sent to Geneservice Ltd at UCL London for sequencing. Geneservice Ltd provided common primers for sequencing such as those that bind to T7 or Sp6 regions. However, for sequencing with custom made primers, 5 $\mu\text{l}$  at 3.2pmol/ $\mu\text{l}$  of each primer per reaction was sent separately with the DNA samples to be sequenced. Sequencing results were analysed using the Sequence Scanner software v1.0 (Applied Biosystems, UK).

### **2.1.13 Generation of Blunt Ends**

Some restriction enzymes generate blunt ends directly, such as *Stu*I and *Sna*BI. However, other enzymes generate sticky ends with some base overhangs. To achieve blunt end cloning or even vector self-ligation through blunt ends, these overhangs must be converted to blunt ends. This can be achieved by either filling up the overhangs with complimentary bases or by removal of the excess bases. In this study, the Klenow fragment (3'  $\rightarrow$  5' exo-), kindly given by Dr. Cathy Joyce, the Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT., USA, was used to create blunt ends by removal of the 3' overhangs.

To do so, 1 $\mu\text{g}$  of digested vector, was incubated with 1 $\mu\text{l}$  of Klenow fragment in 3-4 $\mu\text{l}$  of 10x buffer 2 (1x buffer 2: 10mM Tris-HCL pH 7.9, 50mM NaCl, 10mM  $\text{MgCl}_2$ , 1mM DTT), 33 $\mu\text{M}$  of each dNTP and  $x\mu\text{l}$  of DEPC  $\text{H}_2\text{O}$  to a total of 30-40 $\mu\text{l}$ . The reaction was left at room temperature for 15 minutes then stopped by heating at  $75^{\circ}\text{C}$  for 25 minutes. The Klenow treated vector, was then run on an agarose gel, where the band representing the vector was

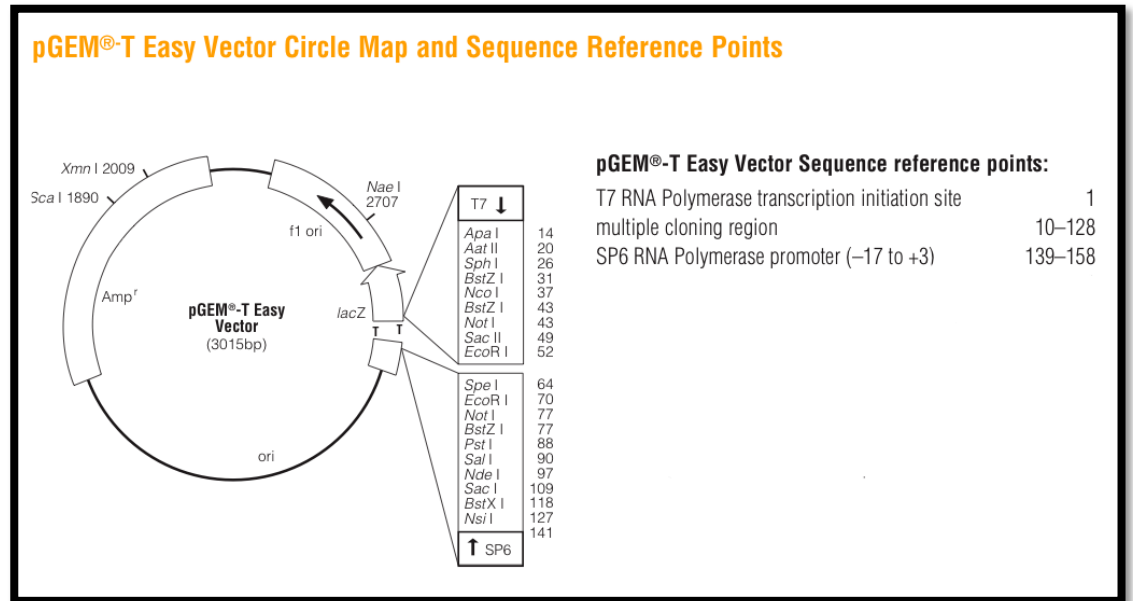
purified by the use of the Wizard® SV Gel and PCR Clean-Up System according to the manufacturer's protocol. The concentration of the purified DNA fragments was then estimated by means of OD<sub>260nm</sub> measurement.

## 2.2 AT Cloning

PCR products generated by the use of *Taq* polymerase enzyme have an "A" residue overhang on the 3' ends. Therefore, unmodified PCR products can be cloned into a linearised vector having a "T" overhang of the 3' end. Based on these facts, unmodified PCR products were cloned into the pGEM-T vector (Promega, UK). The AT cloning was done by the use of the pGEM-T easy vector kit (Promega, UK) and was conducted according to the manufacturer's protocol. The pGEM®T easy vector system is shown in Figure 2.1. This vector has a *LacZ* gene within its multiple cloning region (MCR). The *LacZ* gene expresses β-galactosidase enzyme upon induction with IPTG. This means that bacterial colonies containing the pGEM®T vector when grown on agar plates containing both IPTG and X-gal will produce β-galactosidase, and these bacterial colonies will be blue in colour. However, when a DNA fragment is inserted into the MCR, the *LacZ* gene is disrupted and bacterial colonies containing the vector and insert will be visible as white colonies, hence, allowing blue-white colony screening.

Pre-prepared LB agar plates with 100µg/ml ampicillin, were further plated with a 100µl of 100mM IPTG (promega, UK) and 20µl 50µg/ml X-Gal (Promega, UK) 30 minutes prior to plating the transformed bacterial cells. Colonies were left to grow overnight at 37°C, and the white colonies were screened by means of PCR.

Positive PCR screened colonies were further grown overnight in LB broth containing ampicillin and the plasmids were then extracted. The purified extracted plasmids were quantified and were then sent to Geneservice Ltd at UCL London for sequencing. The primers T7 forward and Sp6 reverse were used to conduct this sequencing.



**Figure 2.1** Map of the pGEM<sup>®</sup>-T Easy Vector system. The diagram was obtained from the Promega website.

The pGEM<sup>®</sup>-T easy vector system was used in this study for the purposes of AT cloning. As the Map indicates, the vector has a “T” overhang within the MCR in which the DNA fragment of interest, having an “A” overhang, can be inserted.

### 2.3 Generation of Plasmids

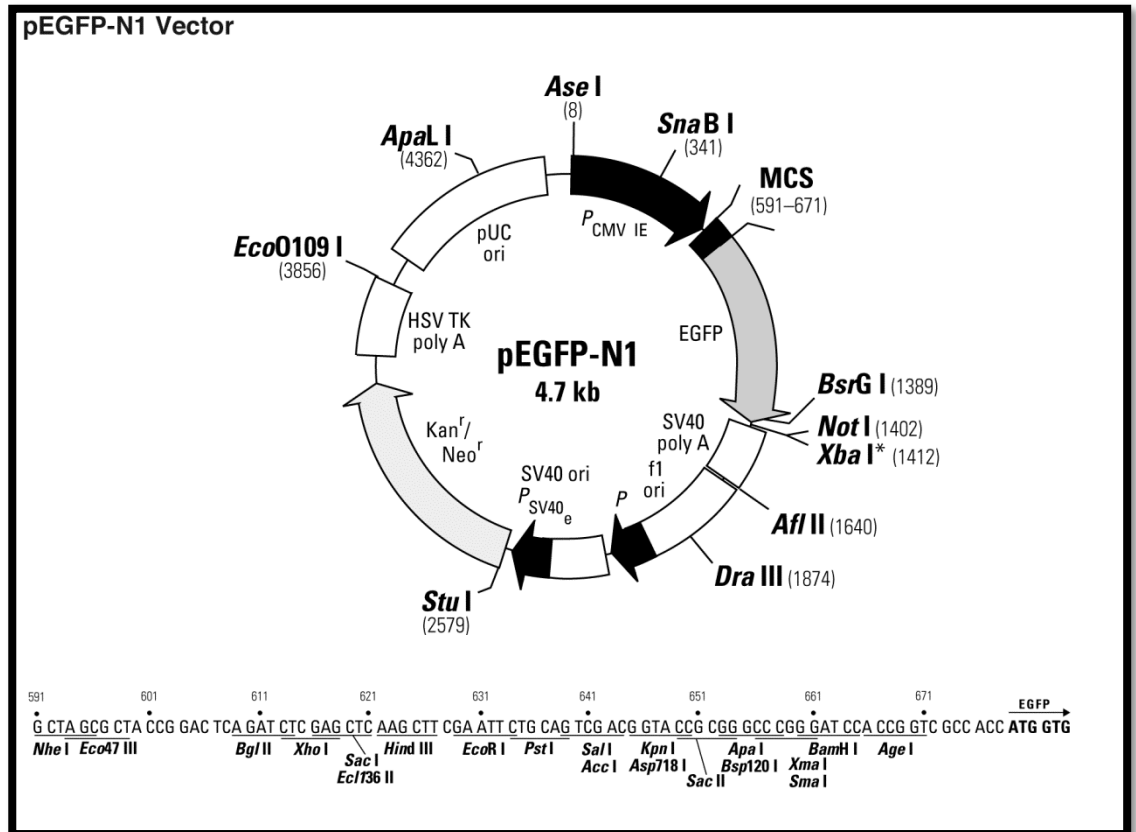
Two plasmids were used for the purpose of incorporating shRNA molecules, pEGFP-N1 plasmid and pB12mcs-eGFP plasmid. The pEGFP-N1 vector (Clontech, UK) has an enhanced GFP, and includes a human cytomegalovirus (CMV) promoter. For the purposes of this project the CMV promoter was not used. Instead either the human U6 or the human H1 promoter was used. The pEGFP-N1 plasmid with the CMV promoter removed was used as a negative control during expression experiments. The pEGFP-N1 vector, where either the U6 or H1 promoter had replaced the CMV promoter, would be used as the positive control. The pEGFP-N1 vector is shown in figure 2.2.

JM109 bacterial cells were transformed with the pEGFP-N1 vector. The cells were then grown overnight in 500ml of LB with the appropriate antibiotic.

1ml of the grown bacteria was stored at  $-80^{\circ}\text{C}$  as glycerol stock, and a maxi prep was then conducted on the transformed bacterial cells. The extracted pEGFP-N1 plasmid was resuspended in an appropriate volume of Milli Q water, quantified and stored at  $-20^{\circ}\text{C}$ .

The pB12mcs-eGFP vector was kindly given by Dr. Adam Rodaway. This vector also contains an enhanced GFP, however, it has no promoter. This is advantageous, as the wild type vector can be used as a negative control. Upstream of the eGFP is the MCR, in which the U6, the H1 human promoters, and the shRNA cassettes can be individually cloned. The vector with only the U6 or the H1 promoter can then serve as positive control. A map of the pB12mcs-eGFP plasmid is shown in Figure 2.3. Before deciding which restriction enzymes were to be used, the vector was sequenced to confirm its identity.



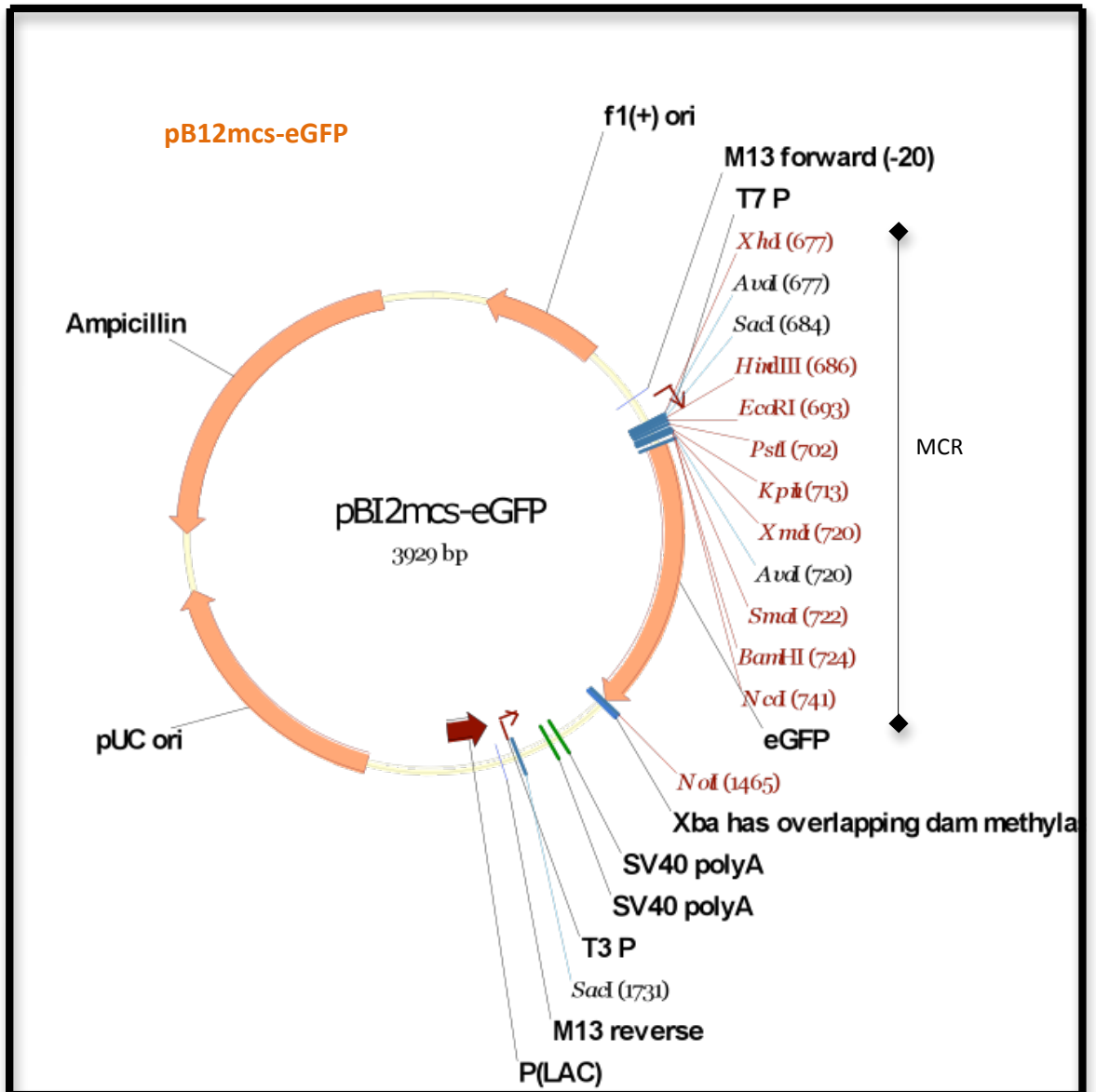


**Figure 2.2 pEGFP-N1 Vector Map.** The diagram was obtained from the Clontech website.

Top: The map of the pEGFP-N1 vector.

Bottom: The sequence and the restriction sites of the MCR of the pEGFP-N1 vector.

The pEGFP-N1 vector was selected to clone the shRNA cassettes. The vector has an eGFP gene that gives its high expression and bright fluorescence in mammalian cells.



**Figure 2.3 The pBI2mcs-eGFP Vector Map. The plasmid was constructed and given by Dr. Adam Rodaway**

The pBI2mcs-eGFP vector is 3029bp in size, has an ampicillin resisted gene, and an eGFP gene that gives a higher and brighter expression of this protein in mammalian cells. The wild type vector can be used as a negative control. Positive controls can be constructed through the insertion of a U6 or an H1 human promoter into the MCR.

### 2.3.1 Construction of the shRNA Cassettes

Short hairpin RNA (shRNA) molecules were designed according to the Cold Spring Harbor Lab guidelines. shRNA molecules were designed to be of 50-52 nt long and were to be cloned with, and downstream of, the U6 or H1 human promoters. The U6-shRNA primer transcripts were initialised with a “G” nucleotide, of the U6/H1-small nuclear RNA Pol III transcript, and termination was mediated by a stretch of 6 “T” nt at the end of the transcript.

Each of the designed shRNA molecules can be divided into four regions. These are the anti-sense region, the loop, and sense and termination regions. The sense region corresponds to 21-22 nt and it’s sequence is the same as the target DNA, this region was designed to be complementary to the fusion points of the *BCR-ABL* gene. Thus, aiming to target the fusion gene only and not the normal *BCR* or *ABL* genes. The anti-sense region is complimentary to the sense part. Both the sense and anti-sense parts of the shRNA transcript are separated by an 8 nt long loop. The final region is made up of a stretch of “Ts”, because the U6 and H1 promoters require an oligo T termination signal. The sequences of the designed shRNA transcripts can be found in table 2.5.

Brummelkamp and colleagues in 2002 studies the effects of the loop size on the RNAi potency of shRNA molecules that had the same sense and anti-sense sequences, but either had five, seven, or nine long nt loop sequences. Their study showed that the differences in loop size and sequence had different RNAi effects, where shRNA molecules with a five nt loop were inactive, those with seven were moderately active, and shRNAs with 9 nt long gave around 90% knockdown of the gene of interest (Brummelkamp *et al.*, 2002). Furthermore, in 2010, Schopman and colleagues demonstrated that moderate or even weak shRNA inhibitors could be transformed into strong potent inhibitors by simply changing the loop size and structure. In this study scientists found that shRNA molecules with an open loop structure (no sequences of the loop are complementary to each other at the sense and anti-sense part of the shRNA) exhibit stronger RNAi gene silencing (Schopman *et*

al., 2010). On this basis, an 8 nt long open loop structure was used for the shRNA molecules designed in the current work.

Once the designing of the shRNA molecules was complete, the secondary structures of these molecules were predicted by the use of the program 'mfold'. These secondary structures can be seen in Appendix 2, except for  $\beta 3\alpha 2$  shRNA1 structure that is demonstrated in Figure 3.13.

Fusion point	shRNA sequences 5'→3'	
$\beta 3\alpha 2$	shRNA1 <sub>(21nt)</sub>	gcagaguuca <u>aa</u> agcccuuca <u>ga</u> agcuug <u>uga</u> agggcu <u>uu</u> ugaacucugcttttt
	shRNA2 <sub>(21nt)</sub>	cagaguuca <u>aa</u> agcccuuca <u>ga</u> agcuug <u>cuga</u> agggcu <u>uu</u> ugaacucugcttttt
	shRNA3 <sub>(22nt)</sub>	gcagaguuca <u>aa</u> agcccuuca <u>ga</u> agcuug <u>cuga</u> agggcu <u>uu</u> ugaacucugcttttt
$\beta 2\alpha 2$	shRNA1 <sub>(21nt)</sub>	caauaaggaa <u>ga</u> agcccuuca <u>ga</u> agcuug <u>uga</u> agggcu <u>uc</u> uccuuauugcttttt
	shRNA2 <sub>(21nt)</sub>	aaauaaggaa <u>ga</u> agcccuuca <u>ga</u> agcuug <u>uga</u> agggcu <u>uc</u> uccuuauuttttt
	shRNA3 <sub>(22nt)</sub>	caauaaggaa <u>ga</u> agcccuuca <u>ga</u> agcuug <u>uga</u> agggcu <u>uc</u> uccuuauugcttttt

**Table 2.5 The designed shRNA transcripts.**

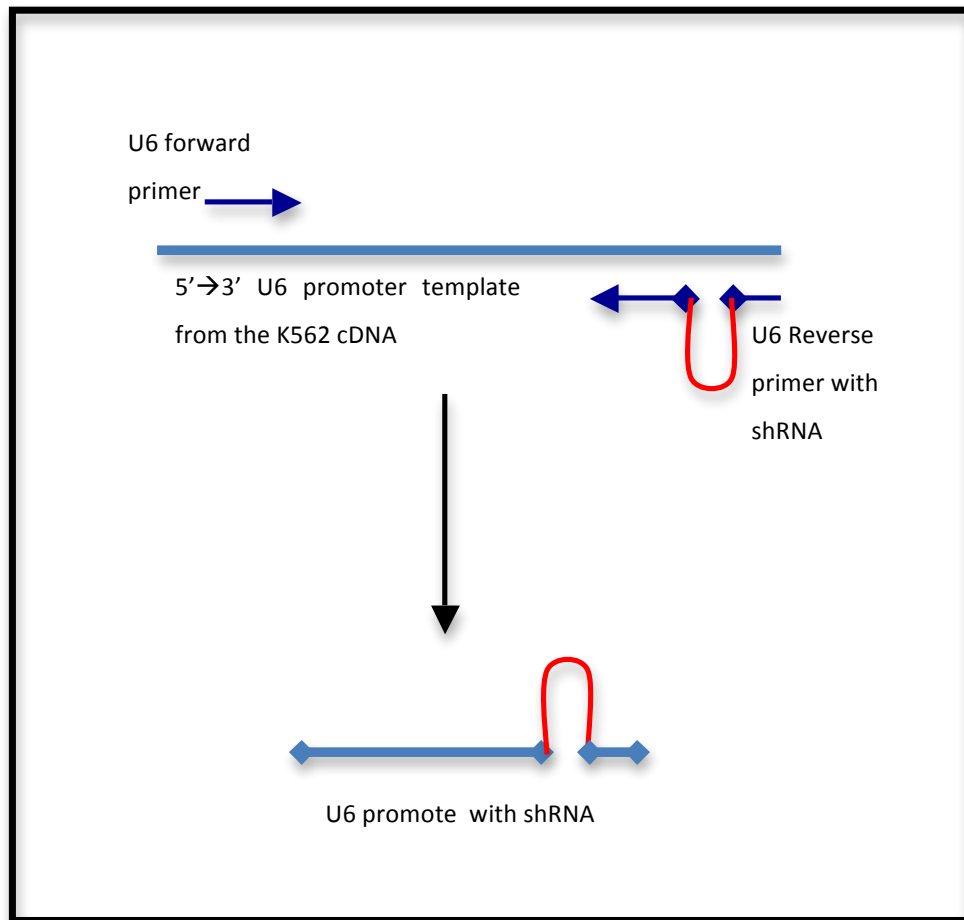
The above table demonstrates that three different shRNA molecules were designed for each fusion point. shRNA1 and 2 are of 21 nt long, whereas shRNA3 is 22 nt long. shRNA1 has 11 nt upstream the fusion point and 10 nt downstream, shRNA2 had 10 nt upstream the fusion point and 11 nt downstream, whereas shRNA3 has 11 nt on either sides of the fusion point. The sequences shown in blue represent the anti-sense part of the molecule and the fusion points within this part are underlined and in bold. Any sequence differences between the three shRNA molecules are shown in green. The sequences in red represent the sense part. The sequences in black and underlined correspond to the hairpin loops, and the oligo T termination signal is in black.

### 2.3.2 Cloning of Promoter Constructs

PCR was used to produce the U6/shRNA and H1/shRNA cassettes. These were established by incorporating the shRNA sequence within a reverse primer

that had 28 nt of its sequence complimentary to the U6 or H1 promoters. HindIII and BamHI restriction sites were also incorporated into the forward and reverse primers respectively. Finally, PCR, conducted on K562 cDNA, was used to produce the whole cassettes. The sequences for these primers are shown in table 2.6.

The U6/shRNA and the H1/shRNA cassettes were cloned into the pB12mcs-eGFP plasmid. The plasmids were then sent for DNA sequencing to confirm the correct sequence. A schematic diagram for the steps involved in the synthesis of the shRNA cassette is shown in Figure 2.4. A flow diagram showing the overall procedures undertaken to generate the pB12mcs-eGFP containing the shRNA cassettes can be seen in Figure 2.5.

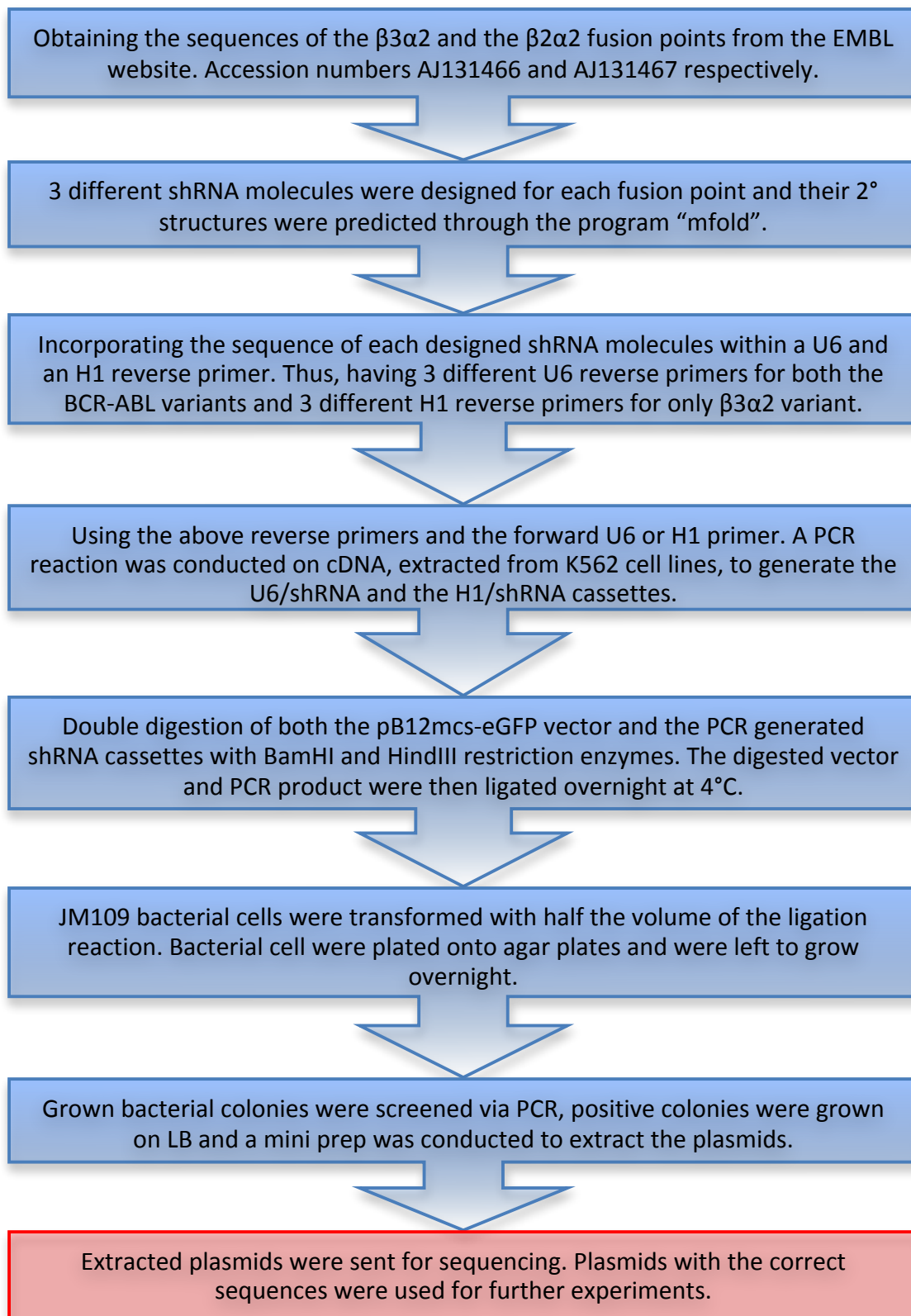


**Figure 2.4 Schematic Diagram of the Synthesis of the shRNA cassettes.**

The diagram illustrates how a shRNA cassette under the regulation of the U6 promoter was synthesised. cDNA extracted from the K562 cell line was used as a DNA template to amplify the human U6 promoter. To synthesise the shRNA cassette (top part of the diagram) the shRNA molecule was synthesised as an oligonucleotide that had 27bp complimentary to the U6 promoter. This oligo was part of a U6 reverse primer, which could be used together with the U6 forward primer in a PCR reaction. The resultant PCR product contained the U6 promoter linked to the shRNA (bottom part of the diagram). The reaction was conducted for every designed shRNA molecule, and the same technique was used to produce shRNA cassettes regulated under the human H1 promoter.

Primer Name	Primer sequence	Melting temperature
U6 Forward	5' d-CCCAAGCTTCAAGGTCGGGCAGGAAGAGGGCCTA'3	64°C
$\beta 3\alpha 2$ shRNA1 U6 reverse	5' d-ATGGATCCAAAAAGCAGAGTTCAAAGCCCTTCACAAGCTT CTGAAGGGCTTTTGAAGCTCTGCGTAGTATATGTGCTGCCGAAGCGAGCAC'3	68.1°C
$\beta 3\alpha 2$ shRNA2 U6 reverse	5' d-ATGGATCCAAAAACAGAGTTCAAAGCCCTTCAGCAAGCT TCCTGAAGGGCTTTTGAAGCTCTGGTAGTATATGTGCTGCCGAAGCGAGCAC'3	68.3°C
$\beta 3\alpha 2$ shRNA3 U6 reverse	5' d-ATGGATCCAAAAAGCAGAGTTCAAAGCCCTTCAGCAAGC TTCCTGAAGGGCTTTTGAAGCTCTGCGTAGTATATGTGCTGCCGAAGCGAGCA C'3	68.9°C
$\beta 2\alpha 2$ shRNA1 U6 reverse	5' d-ATGGATCCAAAAACAATAAGGAAGAAGCCCTTCACAAGCT TCTGAAGGGCTTCTCCTTATTGGTAGTATATGTGCTGCCGAAGCGAGCAC'3	67°C
$\beta 2\alpha 2$ shRNA2 U6 reverse	5' d-ATGGATCCAAAAAATAAGGAAGAAGCCCTTCACCAAGCTT CGTGAAGGGCTTCTCCTTATTGTAGTATATGTGCTGCCGAAGCGAGCAC'3	67.1°C
$\beta 2\alpha 2$ shRNA3 U6 reverse	5' d-ATGGATCCAAAAACAATAAGGAAGAAGCCCTTCACCAAGCT TCGTGAAGGGCTTCTCCTTATTGGTAGTATATGTGCTGCCGAAGCGAGCAC' 3	68.2°C
H1 forward	5' d-CCCAAGCTTCGAACGCTGACGTCATCAAC'3	64°C
$\beta 2\alpha 2$ shRNA1 H1 reverse	5' d-ATGGATCCAAAAACAATAAGGAAGAAGCCCTTCACAAGCTT CTGAAGGGCTTCTCCTTATTGAAAGAGTGGTCTCATACAGAACTT'3	66.8°C
$\beta 2\alpha 2$ shRNA2 H1 reverse	5' d-ATGGATCCAAAAAATAAGGAAGAAGCCCTTCACCAAGC TTCGTG AAGGGCTTCTCCTTATTGAAAGAGTGGTCTCATACA GAACTT'3	66.9°C
$\beta 2\alpha 2$ shRNA3 H1 reverse	5' d-ATGGATCCAAAAACAATAAGGAAGAAGCCCTTCACCAAGCT TCGTGAAGGGCTTCTCCTTATTGAAAGAGTGGTCTCATACAGA ACTT'3	68.2°C

**Table 2.6 Primer sequences used to construct the shRNA cassettes.** This table represents all the primers used to produce the shRNA cassettes. The sequences within the primers in red show the incorporated restriction enzyme sites, HindIII and BamHI, located in the forward and reverse primers respectively.



**Figure 2.5 A Flow Diagram of the Overall Procedures Undertaken to Generate the Plasmids Containing the shRNA Cassettes.**

This diagram summarises the steps and experiments conducted to construct the shRNA cassettes, directed against the *BCR-ABL* fusion genes, and to incorporate these cassettes into a suitable vector.



## **2.4 Tissue Culture**

Two types of cell lines were used for the purposes of this study. These are the K562 cell line (kindly given by Dr. David Grimwade, Guy's tower, King's College London) and the KCL-22 cell line (DSMZ, Germany). These are CML cell lines, and as explained in chapter 1, the K562 cells are CML cells with the  $\beta 3\alpha 2$  fusion point, whereas the KCL-22 cells have the  $\beta 2\alpha 2$  fusion point.

### **2.4.1 Cell line Maintenance**

#### **2.4.1.1 Thawing and recovery of cell lines**

A vial of CML cell lines was removed from the liquid nitrogen freezer (RS series, Jencons pls) and placed immediately in a 37°C water bath (Grant Instrument Ltd, England). The cells were washed with 2ml warm RPMI 1640 medium with L-glutamine (PAA, Austria), 10% Foetal Bovin Serum (FBS) (Sigma, UK) and 550µl of penicillin/streptomycin antibiotic (10,000iu PEN/ml, 10,000ug STREP/ml, MP Biomedicals, France), and centrifuged for 5 minutes at 1000 rpm (Beckman CS-R6 centrifuge, UK) at room temperature. The supernatant was discharged and the pellet resuspended in 1ml of medium. Cells were then transferred to a 75cm<sup>2</sup> culture flask that contained 30ml of medium and incubated in a humidified 37°C, 5% CO<sub>2</sub> incubator (Binder, USA).

#### **2.4.1.2 Feeding and Passageing of cell lines**

For the cells to be passaged they must be confluent, that is their count should be about  $2.5 \times 10^6$  cells/ml in a 250ml flask; cells were fed every 2-3 days till confluency was reached.

#### - **Feeding of Cell lines**

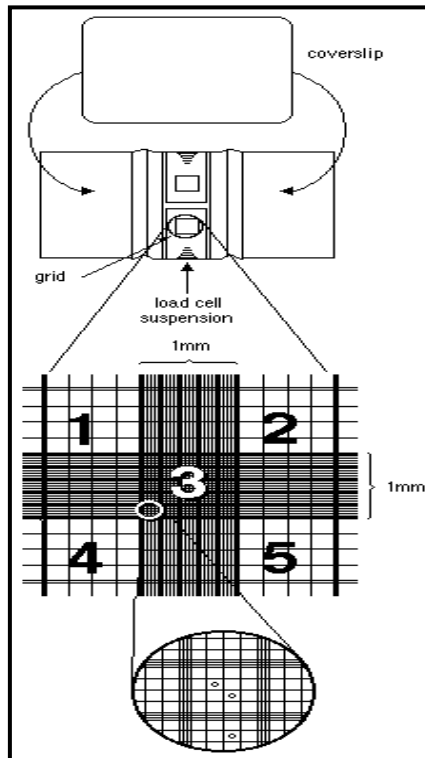
Cells were counted and if non-confluent, one third of the medium in the flask was aseptically removed and replaced by an equal volume of fresh warm media. The flask was gently swirled and returned to the incubator. On the days cells were not fed, they were checked by gently swirling the flask, to resuspend them, and by observing any media colour changes which reflected their metabolic growth.

#### - **Passaging of Cell lines**

When cells were confluent, half the cells in suspension were aseptically transferred to a new culture flask. To each flask an equal amount of warm medium was added and flasks were returned to the incubator.

### **2.4.1.3 Counting of cell lines**

The process of cell counting is an important procedure through which optimisation of culture conditions and the performance of accurate quantification experiments can be conducted. To count the cells a haemocytometer (Bright-Line, Hausser Scientific, USA) was used. The device is a thick glass slide with a central counting chamber. The chamber is divided into two portions each made up of a 3 x 3 mm etched grid that is further divided into nine secondary squares, 1 x 1 mm each. The four corner squares and the central one are used for counting; they are divided into 16 and 25 tertiary squares respectively. A cover slip is also included with the device. See figure 2.6.



**Figure 2.6 The Haemocytometer.** (The diagram was taken from the Klinik Andrologi website)

(Top) The central portion of the haemocytometer, the counting chamber, is divided into two portions. Each portion has its counting grid. (Bottom) One of the grids is amplified showing five counting boxes. The outer four are further divided into 16 tertiary boxes, while the central one comprises 25 boxes. A central tertiary box is further enlarged showing two cells within it. The bent arrows on the top show where the cover slip is placed prior to loading the cell suspension. The straight arrow in the middle shows the groove in which the cell suspension is loaded.

The haemocytometer and the cover slip were wiped with 70% ethanol and the cover slip placed on the counting chambers. A 1:1 dilution for the cells with trypan blue (0.4%, Sigma, UK) was prepared and 10 $\mu$ l of the dilution was loaded into each counting chamber by a pipette. The suspension was drawn under the slip by capillary action, and cells, in the centre and in two of the corner squares, were counted under a light microscope (100x magnification, Nikon Eclipse TS1000, Japan) and with the aid of a hand-held counter. Whole blue-stained cells were excluded in the count, since only those staining yellow are viable. The following formula was used to calculate the cell number:

**“Cells/ml = average count of squares x dilution factor x 10<sup>4</sup>”**

In some cases the Invitrogen automated cell counter, Countess<sup>®</sup>, was also used to count the cells. Here cells were again prepared at a 1:1 dilution factor with trypan blue (0.4%, Invitrogen, UK) and were loaded into both sides of the Countess<sup>™</sup> cell counting chamber slide. The manufacturer’s protocol was followed in order to count the cells and predict the percentage of cell viability.

#### **2.4.1.4 Freezing of cell lines**

Cells for freezing were first transferred into 50ml test tubes, centrifuged for 5 minutes at 1000 rpm (GS-6R Centrifuge, Beckman, USA) at room temperature. The supernatant was discharged and the pellet was resuspended in 10ml of warm medium. Cells were counted, spun as stated previously and were resuspended in a volume of freezing medium, 90% FBS with 10% DMSO, (Sigma, UK) to obtain a concentration of  $2 \times 10^6$  cells/ml. Cells were pipetted in 1ml aliquots into 2ml cryovials. The vials were placed in a 1°C cryo freezing chamber (Nalgene<sup>™</sup>, USA) to allow a gradual decrease in temperature thus reducing the loss of cells. The chamber containing the cells was left overnight in a -80°C freezer (Revco Scientific Inc, USA). The following day, cells were placed in a -150°C liquid nitrogen Dewar (HC series, Taylor-Wharton, UK).

#### **2.4.2 Optimisation of Cell Line Transfection**

The term transfection is usually used to indicate a process in which nucleic acids are intentionally introduced into cells. To date various methods for transfection are present and can be categorised as chemical and non-chemical based transfections. In this project, the K562 and KCL22 cell lines were transfected with the U6/shRNA or H1/shRNA pB12mcs-eGFP plasmids in order to observe and study the effects of the constructed shRNA cassettes. However,

the K562 and KCL22 cell lines are cells in suspension and therefore are considered to be hard to transfect (Ovcharenko *et al.*, 2005, Oliveira and Goodell, 2003 and Merkerova *et al.*, 2007). Hence several transfection methods were tested to establish an acceptable transfection protocol giving a good efficiency.

#### **2.4.2.1 Chemical based transfections**

These transfections depend on the interactions of chemical materials with the DNA of interest in order to internalise the latter into the targeted cell lines. Several transfection reagents were tested, these included the followings:

- SuperFect® Transfection Reagent (QIAGEN, UK): This transfecting reagent has a specifically designed activated dendrimer that possesses a defined spherical core with radiating branches that are positively charged at the ends. The reagent assembles the DNA particles into compact positively charged structures that in turn bind to the negatively charged receptors found on the surfaces of eukaryotic cells, hence internalising the target DNA molecules into the cells.
  
- FuGENE® HD Transfection Reagent (Roche, UK): This transfecting reagent consists of a proprietary blend of lipids and other components that are supplied in 80% ethanol, sterile-filtered, glass packaged vials. The reagent complexes with the target DNA molecules, thus transporting the DNA molecules into the cells during transfection.
  
- Lipofectamine™ 2000 Transfection Reagent (Invitrogen, UK): This is a cationic lipid based reagent. The reagent is a mix of cationic and neutral lipids that form unilamellar positively charged liposome vesicles. When mixed with DNA, the cationic liposomes will absorb the DNA molecules then facilitate their

internalisation into the target cells by means of fusion to the cell's plasma membrane, thus forming an endocytic vesicle.

- TransIT®-2020 Transfection Reagent (Mirus Bio, USA): According to the manufacturer, this transfecting reagent is lipid based, however it is not liposomal. Therefore, the reagent complexes with the DNA and causes it to be internalized into the cells during transfection.
- HilyMax Transfection Reagent (Dojindo Molecular Technologies, USA): is a highly purified lipid of a single molecule. The reagent forms cationic liposomes that complexes with the negatively charged DNA molecule. The DNA-HilyMax complexes electrostatically bind to the cell's membrane and the DNA molecules enter the cells by endocytosis.
- NIMT®FeOfection Yellow Transfection Reagent (GENOVIS, SWEDEN): This transfection reagent is composed of superparamagnetic nanoparticles. The nanoparticles are composed of an iron oxide core and a positive surface. This composition facilitates the formation of non-covalent bonds between the nanoparticles and the DNA, and the resultant particle/DNA complexes are then internalized into the target cell lines.
- CalPhos Mammalian Transfection kit (Clontech): The action of this transfection reagent depends on the formation of calcium phosphate precipitate. In this method, a saline buffer 2x HEPES containing the phosphate ions, negatively charged, is combined with the DNA plasmid calcium solution, positively charged. This finally results in the production of a fine precipitate that is taken up by the target cell lines when added.
- Nanofectin Kit (PAA): This transfection reagent is formed of two components, a positively charged polymer that has the capacity to bind to DNA and a porous nanoparticles. The unique size of the DNA-nanoparticle complex favours the

uptake of the complex by the target cells. In addition, the complex aids to protect the DNA from nuclease degradation.

To determine the amount of reagent needed for DNA transfection for each of the above transfection reagents, the pEGFPN-1 vector was initially used followed by the pB12mcs-U6-eGFP vector. The optimization reactions were conducted according to each manufacturer's protocol.

#### **2.4.2.2 Non-chemical based transfections**

Several non-chemical based transfection methods exist, however, only nucleofection, a subset of electroporation, was considered. The latter is a transfection method that depends on the discharge of electrical impulses that leads to the creation of transient micro-pores in the membrane of the target cell line. Nucleofection is a transfection method first developed by Amaxa to target cell lines that are difficult to transfect, such as primary cell lines, cells in suspension and non-dividing cells.

In nucleofection, the DNA molecules are directly delivered into the nucleus of the target cell. This requires both a nucleofector device and a cell-type specific nucleofector reagent. The Amaxa® Cell Line Nucleofector® Kit V and the Amaxa nucleofector device, Germany, were used. The control pmaxGFP® vector, supplied with the kit, was used to determine the appropriate nucleofection program to be used on the K562 and KCL22 cell lines. Furthermore, the pEGFPN-1 was initially used to determine the amount of plasmid DNA needed to optimize the nucleofection reaction. Once that was achieved, the constructed pB12mcs-U6-eGFP vector was used to optimize the amount of plasmid DNA needed to transfect the target cell lines. All nucleofection reactions were conducted according to the manufacturer's protocol.

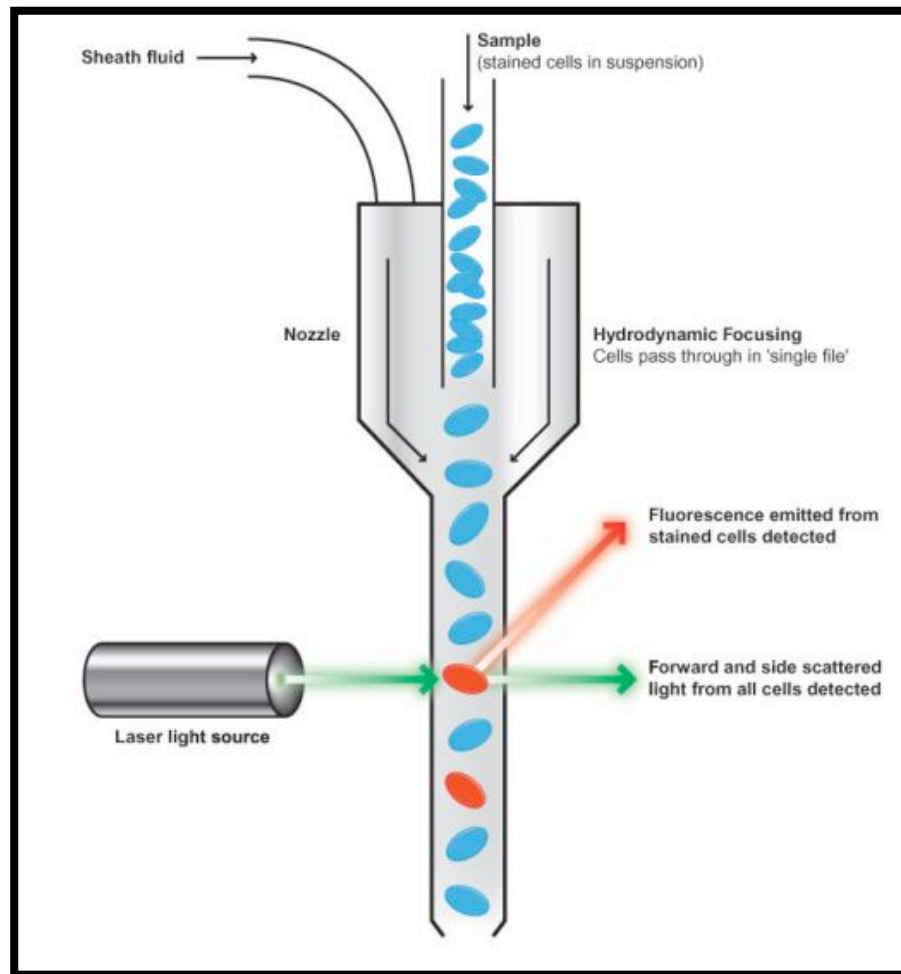
To assess the efficiency of nucleofection a 6-well plate with the transfected cell lines was observed under a fluorescence microscope (Nikon, UK). Once,

fluorescence was detected, the cells were counted and  $3 \times 10^5$  cells were fixed as described below for FACS analysis.  $1 \times 10^6$  cells were used for RNA extraction and  $1 \times 10^6$  cells were used for western blotting.

## **2.5 Fluorescence-Activated Cell Sorter (FACS) Analysis**

FACS scan is a type of flow cytometry that involves the analysis and sorting of single cells or particles, in this case the K562 and KCL22 cell lines, depending of their physical and chemical properties. The cells of interest are suspended in a narrow precisely defined liquid stream also known as the sheath buffer. When the cell sample is passed through the sheath buffer, the cells will undergo hydrodynamic focusing. This type of pressure causes the cells of interest to be individually separated as single droplets that flow through the machine's cuvette. Furthermore, a laser beam of a single wavelength is directed on each cell in the sample. The laser beam passing through the flow stream causes the illuminating light to be scattered and fluorescence will be emitted. Two types of light scattering exist, forward scatter and side scatter. The first is linked with the cell's volume, whereas the second is more dependent on the internal properties of the cell. All light and fluorescent emissions are collected and directed to several special detectors. The detectors will proportionally convert these light pulses into electrical pulses, which can then be amplified by a linear or logarithmic amplifier. The electrical pulses can also be digitized for further computer processing and storage (Robinson, 2004). In the current work the cytometer used was a FACScan from Becton Dickinson, UK. Figure 2.7 illustrates the basic principles of the FACS scan.

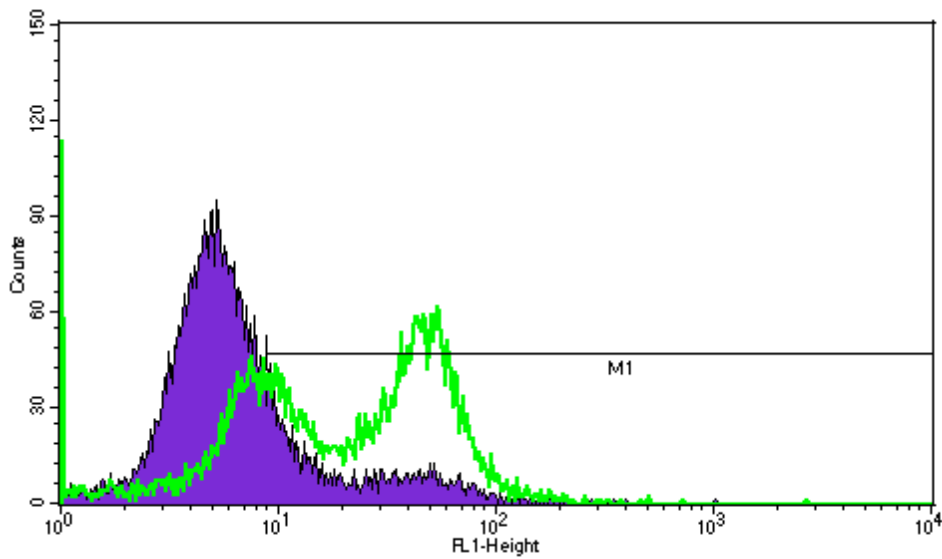




**Figure 2.7** The basic principle of the FACS scan. (Source [www.abcam.com](http://www.abcam.com))

This figure illustrates the principles of the FACS scan. The cells of interest can be stained with a specific dye or express a fluorescent protein, for example eGFP. The sample cells are passed through a stream of liquid called the sheath fluid/buffer and through the process of hydrodynamic focusing, the cells will be separated into individual droplets that pass through a thin narrow glass cuvette. Once in the cuvette, a laser beam is directed in each individual cell, this causes the beam to separate into scattered light, in forward and side scattered, and emitted fluorescence. The emission is collected through a detector that will proportionally convert the collected light pluses into digitalized electrical pulses that can be further analysed by computer software.

As discussed previously, the K562 and KCL22 cell lines were transfected with the pB12mcs-U6-eGFP vector. The plasmid expresses eGFP that can be detected by FACS. The cell lines were centrifuged at 1,000 rpm (Beckman TJ-6 centrifuge, Ireland), the supernatant discharged and the pellet was washed with phosphate buffer saline (PBS) (Invitrogen, UK). The pellet was then resuspended in 500µl of PBS and the cells fixed by the addition of 500µl of 4% paraformaldehyde (4g PFA, (Sigma, UK) dissolved in 100ml of PBS, heated up to 60°C with moderate stirring, and the pH was adjusted between 7.2-8. Once the solution was clear, the heat was turned off, the solution was cooled down, filtered, and aliquoted into 15ml tubes. The solution was then stored at -20°C). The parameters were set for the type of cells used in this project, the K562 and KCL-22, and the cell suspensions were analysed using the FACS machine. The obtained data was analysed by the use of the computer software "CellQuest Pro" (Becton Dickinson, UK), and the collected data was plotted in a single dimension to produce a histogram. A two-dimension plot in the form of a dot plot was also used, this was done when acquiring the data to ensure that the initial control sample fell in one group of cells. Figure 2.8 shows a graph generated from the CellQuest Pro software in the form of a histogram.



**Figure 2.8 FACS Histogram Graph.**

This figure shows a Histogram graph that was generated through the use of the “CellQuest Pro” software. The Histogram shows two samples, a control sample, in purple (this sample represents cells that have not been treated by any means), and a test sample overlaid in green. Both samples contain the same number of cells (10,000). The M1 Line corresponds to the gated cells. The histogram above shows that above 52% of cells have been transfected successfully.

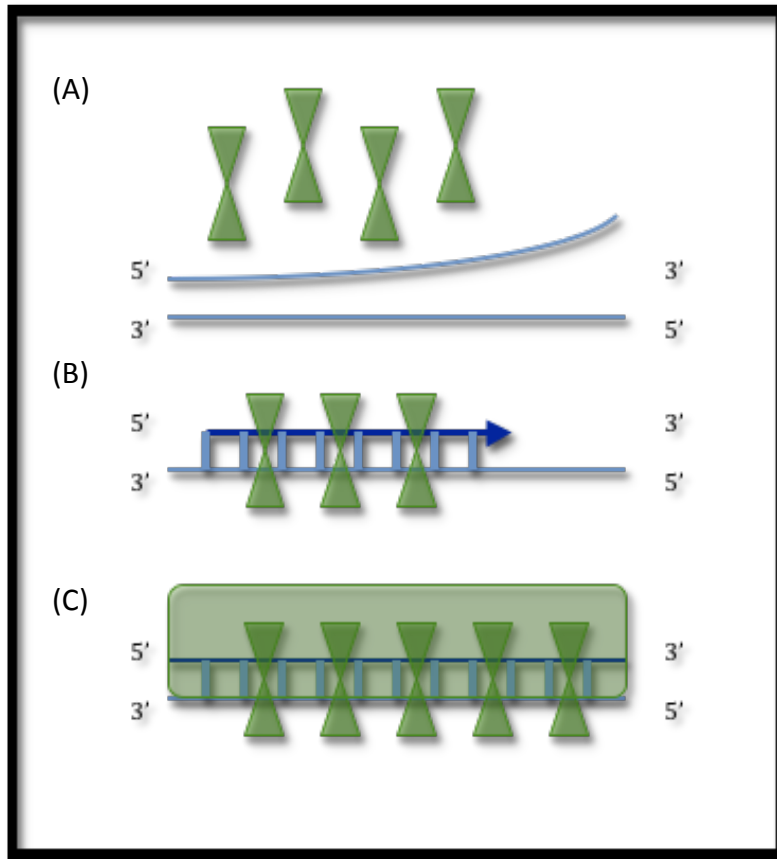
## 2.6 Real Time PCR

Real time PCR or quantitative PCR (qPCR) is a highly sensitive technique that allows simultaneous amplification and accurate quantification of a specific nucleic acid sequence from DNA or cDNA targets. PCR products in this technique are detected through using either fluorescent dyes, which bind to double stranded DNA, or by using fluorescently labelled probes that are sequence specific. Consequently, quantification is achieved by a determination of the cycle number at which the sequence of interest is first detectable.

In this study, real time PCR was used for gene expression analysis, the Roche Light Cycler 480 (Roche, UK) was used and PCR product quantification was done through the use of SYBR Green I and Taqman probes. SYBR Green I is

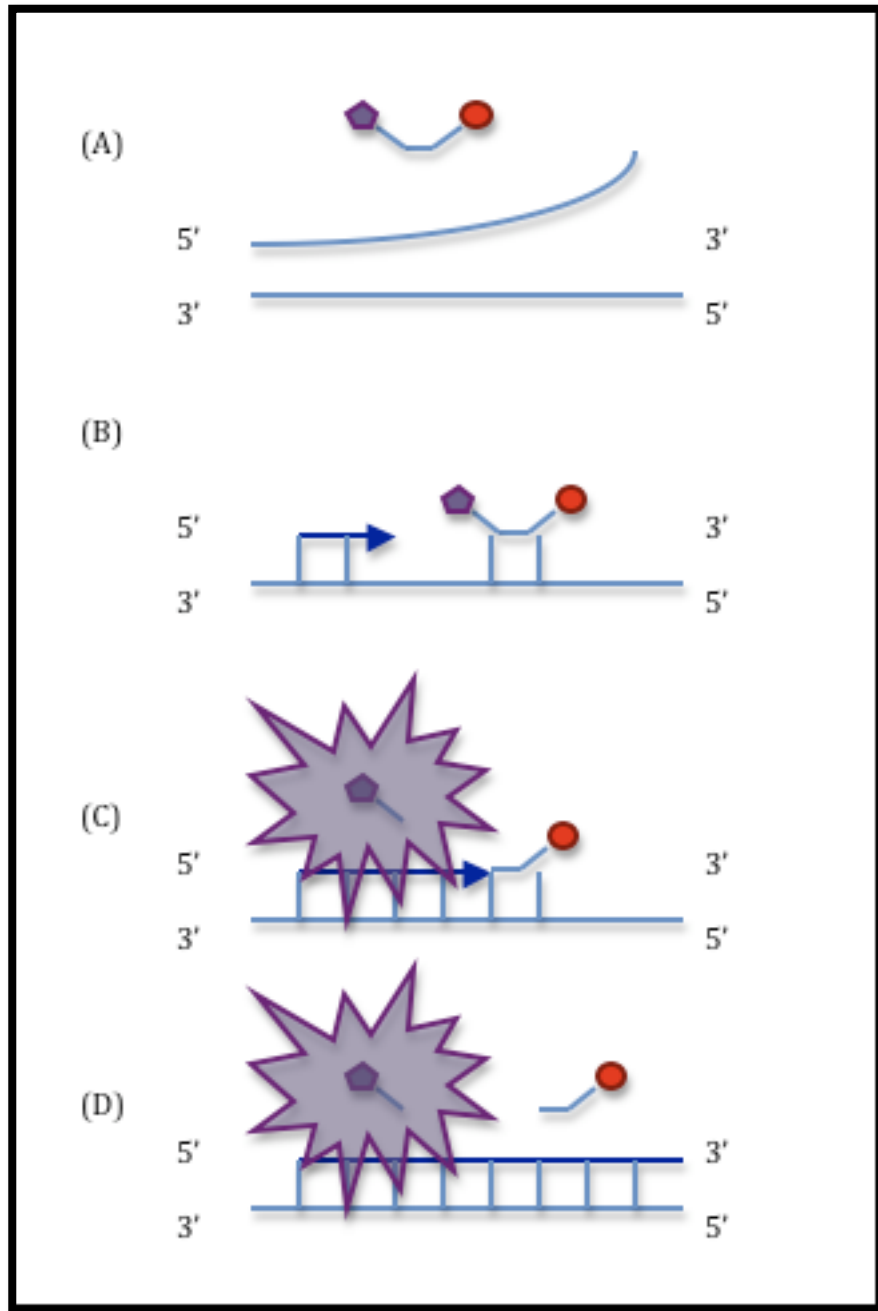
a dye that binds to any double stranded DNA molecule, and emits a fluorescent signal of a particular wavelength. The detection of its fluorescence is achieved at the extension phase of the qPCR. Then with the accumulation of the double stranded PCR products, an increase of the fluorescent signal is also detected at the end of each cycle. This is due to the fact that the amount of fluorescence/signal is proportional to the amount of the accumulating PCR product. Figure 2.9 illustrates the basic principles of real time PCR when conducted using SYBR Green I.

Taqman probes also called hydrolysis probes are sequence specific oligonucleotides that carry a fluorophore moiety at the 5' end and a quencher moiety at the 3' end. Once the probe is cleaved by the 5' → 3' exonuclease activity of *Taq* polymerase (Holland P.M *et al.*, 1991), the fluorophore and the quencher moiety will be separated and a fluorescent signal will be detected. Again the signal is proportional to the amount of PCR product being accumulated. This process occurs during the annealing/extension phase of the qPCR. Figure 2.10 illustrates the basic principles of real time PCR using the Taqman probes. Taqman primers and probes were designed by the use of Roche Universal probe library (UPL) software and Primer3 software. Table 2.7 shows all Taqman primers and probes used in this project. The same primers were used without the probes when conducting real time PCR using SYBR Green I.



**Figure 2.9 Real time PCR using SYBR Green I methodology.**

SYBR Green dye binds to any double stranded DNA regardless its sequence. The dye will only fluoresce once bound to DNA and is excited via blue light. In (A) the DNA is denatured and the synthesis of double stranded DNA starts in (B). Once the double stranded DNA is being synthesised the SYBR Green dye starts to bind to the newly synthesised DNA molecules. The fluorescent signal increases to its maximum when the DNA synthesis is complete and the maximum amount of dye is bound (C). Hence, the fluorescent signals in this type of qPCR are measured at the end of each elongation phase.



**Figure 2.10 Real time PCR in the presence of Taqman probes.**

Taqman probes or hydrolysis probes are sequence specific that bind to DNA at a specific sequence, the probe has both a dye at its 5' end and a fluorescence quencher on its 3' end. In (A) the DNA is being denatured and the probe binds to its target site in (B) while DNA amplification begins. As the primer is being extended and the double stranded DNA is being synthesised to a point where it reaches the Taqman probe, the 5' nuclease activity of the polymerase enzyme will cleave the probe, this releases the dye and a fluorescent signal can be detected. This is demonstrated in part (C). Finally, in (D) the DNA synthesis is complete. The fluorescence in this type of qPCR is measured at the end of the elongation phase.

Taqman primer sets		Primer and probe sequences
<i>GAPDH</i> (226bp)	Forward	5' d-GAAGGTGAAGGTCGGAGTC'3 (NM_002046.3)
	Reverse	5' d-GAAGATGGTGATGGGATTC'3 (NM_002046.3)
	Probe	5' d-[6FAM]CAAGCTCCCGTTCTCAGCC[BHQ1]'3
<i>ABL gene*</i> (77bp)	Forward	5' d-GGGCTGCAAATCCAAGAAG'3 (NM_005157.3)
	Reverse	5' d-ATGCTACTGGCCGCTGAA'3 (NM_005157.3)
	Probe	UPL probe number 65
<i>BCR gene*</i> (91bp)	Forward	5' d-TCAGCCACTGGATTTAAGCAG'3 (NM_021574.2)
	Reverse	5' d-CGCGCTCTTTGCTTTAT'3 (NM_021574.2)
	Probe	UPL probe number 10
$\beta$ 3 $\alpha$ 2 fusion point (190bp)	Forward	5' d-GTCCACTCAGCCACTGGATT'3 (AJ131466)
	Reverse	5' d-TGTTATCTCCACTGGCCACA'3 (AJ131466)
	Probe	5' d-[TAM]AACTCCAAGGAAAACCTTCTCGCTGGACCC[BHQ2]'3
$\beta$ 2 $\alpha$ 2 fusion point (183bp)	Forward	5' d-GATGCTGACCAACTCGTGTG'3 (AJ131467)
	Reverse	5' d-GTTGGGGTCATTTTCACTGG'3 (AJ131467)
	Probe	5' d-[TxRd]TCGTTGGAAGCAAGGAAAACCTTCTCGCT[BHQ2]'3

**Table 2.7 Taqman primers and probe sets used in the project.** This table summarises all Taqman primer sets and their appropriate probes that have been used in this study. The sequence of each set is shown. Primer sets marked with (\*) were designed by the use of the UPL software (<http://www.roche-applied-science.com/sis/rtpcr/upl/ezhome.html>), while unmarked sets were designed using primer3 output software. The PCR product band sizes are shown in column 1 and the accession numbers of the transcripts are shown between brackets in column 3.

For qPCR reactions the Roche LightCycler® 480 SYBR Green I Master and the Roche LightCycler® 480 Probe Master mixes were used. Reactions were conducted on a 96 well plate (Roche, UK), with each qPCR reaction repeated in duplicate. The 96 well plate was covered with an adhesive seal and the reactions were conducted under the selection of the SYBR Green I or the multi-hydrolysis probe program with the SYBR Green or the Taqman template respectively. The parameters for the real time PCR are shown below in table 2.8 and the contents of each qPCR reaction can be found in table 2.9 and 2.10. Genes analysed by qPCR post transfection included the *GAPDH* gene, a house keeping gene, the normal *ABL* and *BCR* genes, to see whether or not the shRNA

constructs have an effect on the normal genes, and the abnormal fusion genes. To achieve this, the first step was to generate standard curves for each gene. These curves are then used to determine the cross point value of the cDNA levels of the genes of interest in the transfected cell lines. The concentrations of each gene in ng were obtained from the Roche Lightcycler 480 software its self. The resultant concentrations where normalized with GAPDH. Statistical analysis was then conducted on the normalized data. The student t-test was used to calculate significance. Finally, the percentage of mRNA expression and depletion was calculated. “Microsoft Excel” was used to conduct all calculations.

(A) Parameters for the Multi-hydrolysis probe program:

One cycle of initial denaturation at 95°C for 5 minutes	
40 cycles of amplification → (Quantification)	Denaturation at 95°C for 45 seconds
	Annealing temperature at 60°C for 1 minute
	Extension at 72°C for 1 second
One cycle of cooling at 40°C	

(B) Parameters for the SYBR Green I program:

One cycle of initial denaturation at 95°C for 5 minutes	
40 cycles of amplification → (Quantification)	Denaturation at 95°C for 15 seconds
	Annealing temperature at 60°C for 10 seconds
	Extension at 72°C for 30 seconds
	A single acquisition mode at 80°C for 1 second
One cycle of Melting that goes up to 95°C for 1 second then down to 60°C for another 1 second	
One cycle of cooling at 40°C	

**Table 2.8 Parameters used for real time PCR conducted in this project.** (A) Shows the parameters used for real time PCR using the multi-hydrolysis probe program, while (B) shows the real time PCR parameters used when the SYBR Green I program was conducted.



Component	Volume per 20 $\mu$ l reaction
cDNA	2-4 $\mu$ l
10 $\mu$ M of Forward primer	1 $\mu$ l
10 $\mu$ M of Reverse primer	1 $\mu$ l
0.2 $\mu$ M of probe	1 $\mu$ l
Roche LightCycler <sup>®</sup> 480 Probe Master mix	10 $\mu$ l
Nuclease free H <sub>2</sub> O	x $\mu$ l

**Table 2.9 Components of the qPCR Master Mix (20 $\mu$ l reaction mixture).** The preparation of master mix solution was done on ice.

Component	Volume per 20 $\mu$ l reaction
cDNA	2-4 $\mu$ l
10 $\mu$ M of Forward primer	1 $\mu$ l
10 $\mu$ M of Reverse primer	1 $\mu$ l
Roche LightCycler <sup>®</sup> 480 SYBR Green I Master mix	10 $\mu$ l
Nuclease free H <sub>2</sub> O	x $\mu$ l

**Table 2.10 Components of the qPCR Master Mix (20 $\mu$ l reaction mixture).** The preparation of master mix solution was done on ice.

## 2.7 Western Blot Analysis

Western blots are an immunoanalytical method used for the purposes of protein analysis. The technique depends on antigen-antibody interactions in order to detect certain proteins of interest. The above technique allows a quantitative as well as a qualitative analysis of proteins.

### **2.7.1 Cell lysis and protein extraction**

1 x 10<sup>6</sup> cells (K562 or KCL-22) were pipetted into a 1.5ml Eppendorf tube, washed once with PBS, then a pellet was obtained by centrifugation of the cells for 5 minutes at 14,000 g at 4°C. The supernatant was removed and 100µl of cell lysis buffer (Table 2.11) was added and the cells lysed by pipetting vigorously several times. The lysed samples were then kept on ice for 30 minutes, centrifuged for 5 minutes at 14,000 g at 4°C and the supernatant, containing the extracted proteins, was pipetted to a new Eppendorf. The lysate was then mixed with 1x sample Loading buffer (Table 2.11) in a 1:1 ratio and the samples were stored at -20°C.

### **2.7.2 SDS-Polyacrylamide Gel Electrophoresis**

A ready prepared 12% polyacrylamide gel (Serva Electrophoresis, Germany) was placed in a tray and filled with 1 x Running buffer, Table 2.11 demonstrates the components of all buffers and solutions that are used during the running of the polyacrylamide gel and the Western blotting protocol. To denature the extracted proteins, previously described in section 2.7.1, the protein samples were placed in a heating block at 100°C and were allowed to boil for 3-4 minutes. The samples were left to cool down, and then 20-30µl was loaded onto the gel. 10µl of high range Rainbow marker (GE Healthcare, UK), see Appendix 3, was also loaded. The gel was then run at 100V for 90 minutes or when samples reached to the bottom of the gel.

Solution Name	Components	Storage Conditions
Lysis Buffer	50mM Tris-HCL (Sigma, UK) 1% Triton 150 NaCL (Sigma, UK) 1mM EDTA (Sigma, UK) - For every 10ml of buffer one tablet of complete, mini, EDTA-free protease inhibitor cocktail tablets (Roche, UK) was added.	Once the protease inhibitor tablet was added the solution was stored at -20°C for 12 weeks and at 4°C for 2 weeks.
Sample Loading buffer	62.5mM Tris-HCL pH 6.8 25% Glycerol 2% SDS 0.01% Bromophenol Blue(Bio-RAD) 5% β-mercaptoethanol	4°C

**Table 2.11 Components of the Lysis buffer and the Sample loading buffer.** The Sample loading buffer is also known as the Laemmli loading buffer.

### 2.7.3 Western Blotting

The cassette containing the polyacrylamide gel was cracked open, the wells removed and the gel washed once with Milli Q water for 5 minutes. The gel was then covered with 1x Transfer buffer and placed on a rotary shaker for 20 minutes. The PVDF transfer membrane was treated according to the manufacturer's instructions (Thermo Scientific, UK), and the transfer cassettes, sponge pads, 6 pieces of Whatman paper were placed in a large container and equilibrated in 500ml of 1x transfer buffer for a few minutes.

The transfer sandwich was assembled, placed in the transfer tank with a block of ice and a magnetic stirring pin. The tank was placed on a magnetic stirrer, the whole apparatus was placed in a cold room at 4°C, and the transfer was conducted overnight at 20-30V.

Once the transfer was completed, the membrane was washed once with Milli Q water. The membrane was then dipped for a couple of seconds in 50ml of 1% (w/v) Ponceau S solution in 5% acetic acid (Sigma,UK). Once the proteins were observed, the membrane was scanned and the image was kept to assist the loading of the samples. The membrane was washed with TBS-Tween, washing buffer, for 10 minutes. Then the membrane was blocked with blocking buffer (see Table 2.12 for the composition of all solutions used in the processes of SDS-PAGE electrophoresis and Western Blotting) for 1 hour at room temperature.

The blocking buffer was then discharged and the membrane was incubated with the primary antibody, diluted in 15ml of blocking buffer, overnight at 4°C on a rotary shaker (Table 2.13 illustrates all antibodies used for the Western blots conducted in this study). The following day, the primary antibody was discharged and the transfer membrane was washed 3 times with washing buffer, 5 minutes each time at room temperature. The membrane was then incubated with the secondary antibody, again diluted in 15ml of blocking buffer, for 1 hour at room temperature. The secondary antibody was then removed and the membrane was washed for another 3 times in washing buffer, 5 minutes each time at room temperature. Finally, the blots were developed using the Pierce® ECL Western Blotting Substrate Kit (Thermo Scientific, UK). This was done according to the manufacturer's protocol. The membrane was then placed in a film cassette, in such a way that the treated protein side of the PVDF membrane can be exposed to the X-ray film (Fujifilm, UK). In the dark room and under a red safe light, the X-ray film was cut, to the appropriate size of the membrane, placed on the treated side of the membrane and the cassette was closed. The time of exposure varied from 1-60 minutes, depending on the intensity of the signal. Band intensities were analysed through the program "ImageJ", and all calculations were done by the use of "Microsoft Excel" software.

<b>Solution name</b>	<b>Components</b>	<b>Additional Information</b>
10x Running Buffer (Laemmli Buffer)	25mM Tris-HCL 200mM Glycine 0.1% (wt/v) SDS - dH <sub>2</sub> O as required	The buffer was diluted to 1x before use
1x Transfer Buffer (1L)	100ml of: 1.92M Glycine 0.25M Tris-HCL 200ml of: MetOH - 700ml of dH <sub>2</sub> O	
TBS-Tween (Wash Buffer)	1M Tris-HCL pH6.8 5M NaCL 1% Tween dH <sub>2</sub> O as required	
Blocking Buffer	5% (wt/v) Non Fat Milk (Marvel) in 1x TBS-Tween 5% (wt/v) BSA (Sigma) in 1x TBS-Tween	

**Table 2.12 Recipes of the solutions used in the process of SDS-PAGE electrophoresis and Western Blotting.** The above table shows the composition of all the buffers used for of SDS-PAGE electrophoresis and Western Blotting.

Primary Antibody	Dilution used	Secondary Antibody	Dilution used	Blocking Buffer used
c-ABL (Cell Signalling)	1:1000	Anti-Rabbit Secondary Antibody	1:1000	5% (wt/v) BSA in 1x TBS-Tween
$\beta$ 2 $\alpha$ 2 junction specific (Cell Signalling)	1:1000	Anti-Mouse Secondary Antibody	1:1000	5%(wt/v) Non Fat Milk in 1x TBS-Tween
GAPDH (Cell Signalling)	1:1000	Anti-Rabbit Secondary Antibody	1:1000	5% (wt/v) BSA in 1x TBS-Tween

**Table 2.13** A list of the primary and secondary antibodies used in the Western Blots in this project. Antibodies used were all purchased from Cell signalling.

## **Chapter 3 : Generation of a Reporter gene system for expression of shRNAi cassettes targeting the BCR-ABL fusion points**

### **3.1 Introduction**

The *BCR-ABL* fusion genes result from a reciprocal translocation between chromosomes 9 and 22. The two most common resultant fusion points associated with CML are the  $\beta 3\alpha 2$  and the  $\beta 2\alpha 2$  variants. These variants are characterised by an abnormally up-regulated Tyrosine kinase activity that causes the development of CML (Advani and Pendergast, 2002). Consequently, targeting of these fusion points as a therapeutic approach for CML could be rewarding.

In this project, short hairpin RNA molecules have been designed to target the two BCR-ABL variants. To achieve this, a reporter gene system was needed to monitor the synthesis and expression of the shRNA constructs. The shRNA oligonucleotides were designed and constructed in such a manner that they were under the regulation of either the human U6 or H1 promoters. The constructed cassettes were further incorporated into an expression vector that had the gene for the enhanced green fluorescent protein (eGFP). eGFP is a mutated form of GFP that has been optimised to have a brighter fluorescence and a higher expression in mammalian cells than GFP (Siemering K. R *et al.*, 1996), and for the purposes of this project eGFP acted as a reporter gene for the shRNA constructs. Two plasmids were considered in order to generate the reporter gene system. The first was the pEGFP-N1 plasmid; however by using this plasmid it proved difficult to construct the system. The second plasmid was the pB12mcs-eGFP plasmid and constructing the reporter gene system in this plasmid was successful. This chapter shows the results obtained from attempts to construct such a system.

## **3.2 Results**

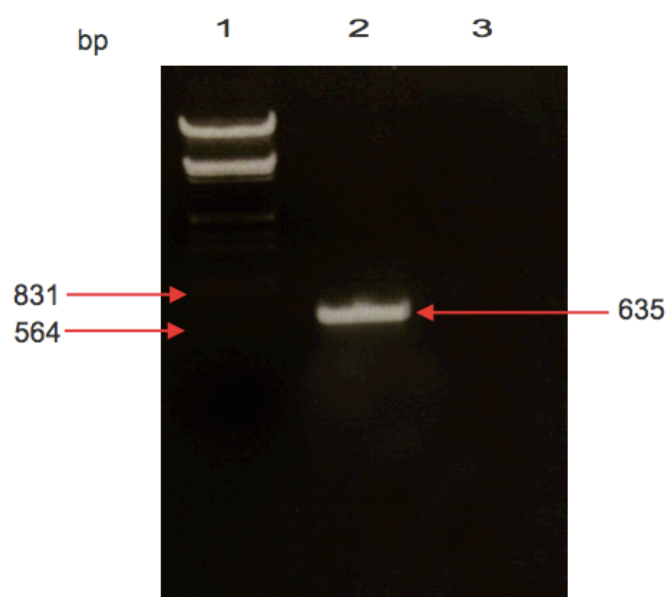
### **3.2.1 Sequencing of the $\beta 3\alpha 2$ and the $\beta 2\alpha 2$ Fusion Points**

The  $\beta 3\alpha 2$  and the  $\beta 2\alpha 2$  fusion genes present in the K562 and KCL-22 cell lines used in this study were sequenced. This step was done to verify that the sequences of the fusion points matched the sequences found in the EMBL website, accession numbers AJ131466 and AJ131467 respectively (Appendix 4).

#### **3.2.1.1 PCR Amplification of the $\beta 3\alpha 2$ and the $\beta 2\alpha 2$ Fusion Points**

Total RNA was extracted from K562 and KCL-22 cell lines. The RNA was quantified and a reverse transcriptase reaction was conducted. To determine whether the reaction worked, a PCR reaction using  $\beta$ -actin primers K, L and N, sequences found in Table 2.2, was conducted on the synthesised cDNA of both cell lines. When using the  $\beta$ -actin primers K and L as a pair, the  $\beta$ -actin gene PCR product is 635bp, whereas pairing primers N and L gives rise to a PCR product of 512bp. The last set of primers only works if genomic DNA is present in the sample; this is due to the fact that primer N is intronic. Figure 3.1 shows the results of the PCR reaction conducted on the synthesised cDNA from the K562 cell lines and shows that cDNA synthesis is successful with no genomic DNA contamination. Similar results were obtained with the KCL-22 cell line.





**Figure 3.1 Electrophoretic Analysis of PCR reaction using  $\beta$ -actin primers on K562 cDNA.**

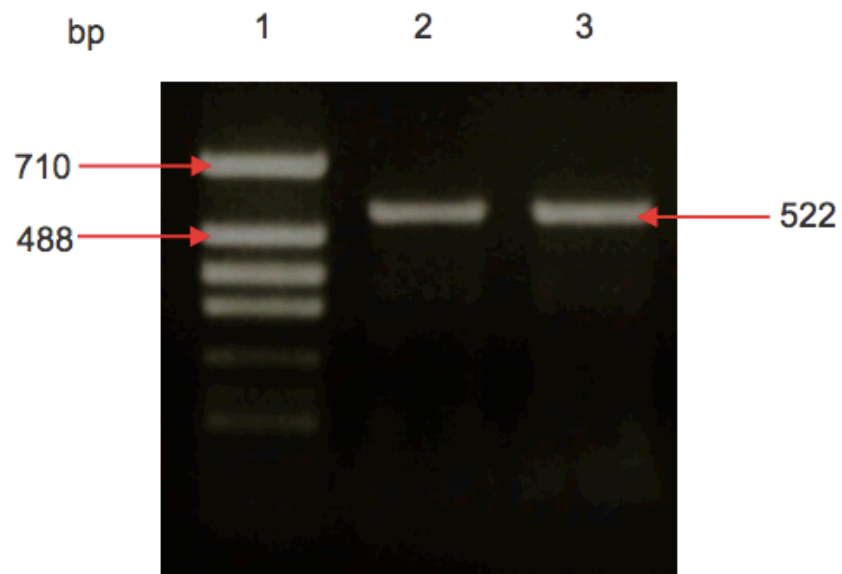
Lane 1:  $\lambda$ /EcoRI/HindIII marker

Lane 2: PCR products of the  $\beta$ -actin primers K and L

Lane 3: PCR products of the  $\beta$ -actin primers N and L (negative control)

$\beta$ -actin PCR products were analysed via electrophoresis using a 1% agarose gel. The 635bp band in lane 2 indicates that the RNA purification and K562 cDNA synthesis was successful. Lane 3 is the negative control and indicates the absence of any genomic DNA contamination.

Once the synthesis of cDNA was confirmed for the K562 and KCL-22 cell lines, a PCR reaction using primer set  $\beta3\alpha2$  and  $\beta2\alpha2$ , sequences found in Table 2.2, was conducted. This PCR was done to amplify the region around the  $\beta3\alpha2$  and  $\beta2\alpha2$  fusion points, and PCR products of 381bp and 522bp in size are expected to be produced, respectively. The PCR products are shown in Figures 3.2 and 3.3. The  $\beta3\alpha2$  and  $\beta2\alpha2$  PCR products were then purified, quantified, and 500ng was used for AT cloning. The remaining  $\beta3\alpha2$  and  $\beta2\alpha2$  PCR products were stored at  $-20^{\circ}\text{C}$ .



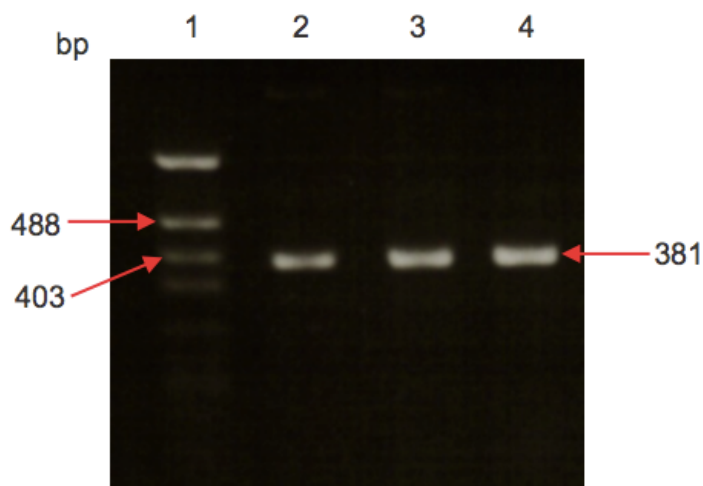
**Figure 3.2 Electrophoretic Analysis of the  $\beta 2\alpha 2$  PCR product from KCL-22 cDNA.**

Lane 1: pBluescript Hpa II marker

Lane 2: PCR products

Lane 3: repeat of lane 2

$\beta 2\alpha 2$  PCR products were analysed via electrophoresis on a 2% agarose gel. The bands present in lanes 2 and 3 fall in the size range of interest, that is 522bp. This indicates that the KCL-22 cell line contains the  $\beta 2\alpha 2$  fusion point.



**Figure 3.3 Electrophoretic Analysis of the  $\beta 3\alpha 2$  PCR product from K562 cDNA.**

Lane 1: pBluescript Hpa II marker

Lane 2: PCR products

Lanes 3 and 4: repeat reactions

$\beta 3\alpha 2$  PCR products were analysed via electrophoresis on a 2% agarose gel. The bands present in lanes 2-4 fall in the size range of interest, that is 381bp. This indicates that the K562 cell line contains the  $\beta 3\alpha 2$  fusion point.

### **3.2.1.2 AT Cloning of the $\beta 3\alpha 2$ and the $\beta 2\alpha 2$ PCR Products in the pGEM<sup>®</sup>T Easy Vector System**

The amplified and purified  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$  PCR products were cloned in the pGEM<sup>®</sup>T easy vector system. This was done through the process of AT cloning. The PCR products have an 'A' overhang, and the pGEM<sup>®</sup>T easy vector has a complimentary 'T' overhang in its multiple cloning region (MCR) for easy insertion of the PCR products.

### 3.2.1.3 Analysis of the $\beta 3\alpha 2$ and the $\beta 2\alpha 2$ Fusion Point Sequences

The sequences of the cloned  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$  fusion points were analysed and compared to those sequences found in the EMBL website, accession numbers AJ131466 and AJ131467 respectively, Figure 3.4. The results confirmed that the sequences of the fusion points found in the K562 and KCL-22 cell lines used in this study matched their sequences found in the EMBL website. The full sequences of  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$  fusion points are shown in Appendix 4, and the raw sequencing data are shown in Appendix 5.

```
(a)
aagcagagtt caaaagccct tcagcggcca
aagcagagtt caaaagccct tcagcggcca

(b)
tgaccatcaa taaggaagaa gcccttcagc ggccagtagc
tgaccatcaa taaggaagaa gcccttcagc ggccagtagc
```

**Figure 3.4 The K562 and KCL-22 Fusion Point Sequencing Results**

(a) Top: Boxed and underlined is the  $\beta 3\alpha 2$  fusion point sequence found in the EMBL website (EMBL/AJ131466)

Bottom: Boxed and underlined is the sequencing result of the  $\beta 3\alpha 2$  fusion point sequence found in the K562 cell line used in this study.

(b) Top: Boxed and underlined is the  $\beta 2\alpha 2$  fusion point sequence found in the EMBL website (EMBL/AJ131467)

Bottom: Boxed and underlined is the sequencing result of the  $\beta 2\alpha 2$  fusion point sequence found in the KCL-22 cell line used in this study.

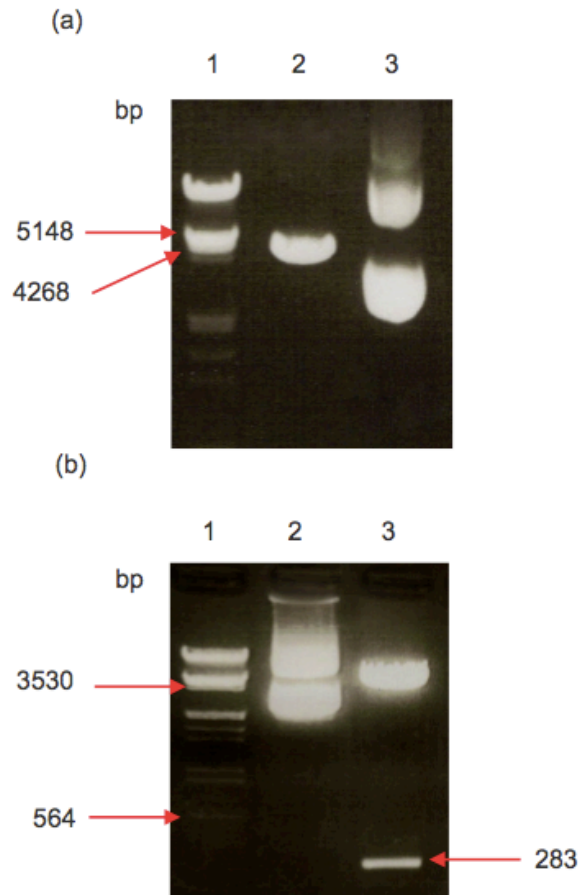
### **3.2.2 Preparation of the pEGFP-N1 Vector for Cloning of the shRNA Cassettes**

The first step was to remove the CMV promoter. Once the promoter was removed the prepared vector was self-ligated, creating a negative control. The vector was subsequently used for cloning the human U6 promoter (positive control), the human H1 promoter (positive control), and the shRNA cassettes.

#### **3.2.2.1 Digestion and Removal of the CMV Promoter from the pEGFP-N1 Vector**

For total removal of the CMV promoter, the restriction enzyme *AseI* could be used with any other restriction site found in the MCR. However the restriction site of the *AseI* enzyme is also found within the shRNA cassettes. So, instead the enzyme *SnaBI*, which has a site within the CMV promoter, and does not cut within the shRNA cassettes, was used.

2 $\mu$ g of the pEGFP-N1 vector was diluted and then digested with *SnaBI* enzyme for 3 hours at 37°C with the appropriate buffer, Figure 3.5 gel (a) lane 2. The vector was then precipitated, resuspended in Milli Q water and further digested with *HindIII* for another 3 hours at 37°C with the appropriate buffer. The digested vector was then precipitated, resuspended in Milli Q water and was run on a 1% agarose gel, Figure 3.5 gel (b) lane 3. The double digestion of the *SnaBI* and *HindIII* sites should give products of 283bp and ~4.4Kb. The presence of the 283bp band is clearly seen confirming that the double digestion has worked. The digested vector, band size of ~4.417Kb, was cut out of the gel, and was purified by the Wizard® SV kit. The vector was then quantified.



**Figure 3.5 Electrophoretic Analysis of pEGFP-N1 Digestions**

Gel (a)

Lane 1:  $\lambda$ /EcoRI/HindIII marker

Lane 2: pEGFP-N1 vector digested with SnaBI

Lane 3: Wild type pEGFP-N1 vector

Gel (b)

Lane 1:  $\lambda$ /EcoRI/HindIII marker

Lane 2: Wild type pEGFP-N1 vector

Lane 3: pEGFP-N1 vector digested with SnaBI and HindIII

To determine the success of the pEGFP-N1 digestion by the restriction enzyme SnaBI, 1 $\mu$ l of the digested vector was run on a 1% agarose gel. Gel (a) lane 2. A single band, of a size approximately 4.7Kb is seen compared with the undigested wild type vector in lane 3. Gel (b), lane 3, two bands of approximately 283bp and 4.4KB are present, thus confirming that both digestions have worked. Lane 2 is the wild type vector that was run as a negative control.

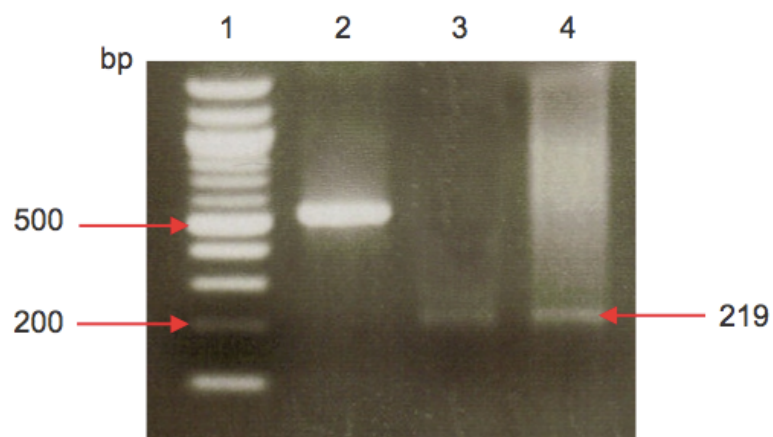
### 3.2.2.2 Self Ligation and Cloning of the Modified pEGFP-N1

#### Vector

The ends generated from the SnaBI restriction enzyme are blunt ends, whereas the ends generated from the HindIII restriction enzyme are sticky with a couple of bases overhang, therefore the ends are incompatible for joining together. To resolve this issue the sticky ends were converted to blunt ones. A number of strategies were tried in order to achieve this self ligation.

The sticky ends of a vector can be either filled in with complementary bases or the overhang bases can be digested. To fill in the overhang created was considered including strategies such as filling the gap with *Taq* polymerase or treating the ends with the large Klenow fragment, and the Klenow fragment ( $3' \rightarrow 5'$  exo-). *Taq* polymerase was first used to try to fill in the sticky ends. Several attempts were conducted but the vector failed to self ligate, indicating that the reaction was not successful. The large Klenow fragment retains its  $5' \rightarrow 3'$  polymerase activity, and its  $3' \rightarrow 5'$  exonuclease activity, but lacks its  $5' \rightarrow 3'$  exonuclease activity. The pEGFP-N1 plasmid was treated with this enzyme several times under different time and temperature conditions. Again the treated vector failed to self ligate. Finally, the Klenow fragment ( $3' \rightarrow 5'$  exo-) that retains its  $5' \rightarrow 3'$  polymerase activity, and lacks both its  $3' \rightarrow 5'$  and its  $5' \rightarrow 3'$  exonuclease activities was used and proved successful.

After treatment with the Klenow fragment ( $3' \rightarrow 5'$  exo-), the treated pEGFP-N1 plasmid was purified and 500ng were used for self ligation. The self ligation reaction was done for an hour at room temperature, and overnight at  $16^{\circ}\text{C}$  and  $4^{\circ}\text{C}$ . The reaction was analysed by PCR using CMV-207 forward and EGFPN1 reverse primers. PCR of the wild type vector was included as a control. When using this set of primers a band of 530bp is expected in the wild type vector, whereas a band of 219bp is expected in the double digested, self-ligated vector. The results of this PCR can be seen below in Figure 3.6.



**Figure 3.6 Electrophoretic Analysis of PCR products from the Self-ligated pEGFP-N1 vector**

Lane 1: 100bp DNA Ladder from NEB

Lane 2: PCR of wild type pEGFP-N1 vector using CMV-207 forward and EGFPN-1 reverse

Lane 3: PCR of self-ligated pEGFP-N1. The reaction was done at 16°C

Lane 4: PCR of self-ligated pEGFP-N1. The reaction was done at 4°C

A 2% agarose gel was used to analyse the PCR products of the wild type and self-ligated pEGFP-N1 vectors. Lane 2 has shown the PCR product of the wild type vector, with a single band of ~530bp. Whereas, lane 4, shows a band ~219bp that corresponds to the double digested, self-ligated pEGFP-N1 vector.

Half the amount of the ligation mix was used to transform 50µl of JM109 bacterial cells. The cells were then plated on agar plates with the appropriate antibiotic and were kept overnight in an incubator at 37°C. The colonies on the plates were screened by means of PCR. The PCR was done using the set of primers CMV-207 forward and EGFPN1 reverse. However, all screened colonies were of wild type vector. Therefore, a higher concentration of ligation mix was used to transform bacterial cells. This also resulted in the growth of wild type colonies only, so the ligation reaction was digested with the enzyme XhoI. This was done to decrease the chances that a wild type vector would be taken into the bacterial cells, as an XhoI site is only present in the wild type vector and not in the self-ligated one. This experiment was repeated three times, and in all three no bacterial colonies grew on the agar plates. Consequently, work on



this vector was stopped and another vector was selected to generate a reporter system

### **3.2.3 Preparation of the pB12mcs-eGFP Vector**

This vector was selected to construct the reporter gene system for the shRNA cassettes after work on the pEGFP-N1 vector has been stopped. The vector has an eGFP gene and no promoter. Therefore, the wild type vector was used as the negative control.

#### **3.2.3.1 Sequencing of the pB12mcs-eGFP Vector**

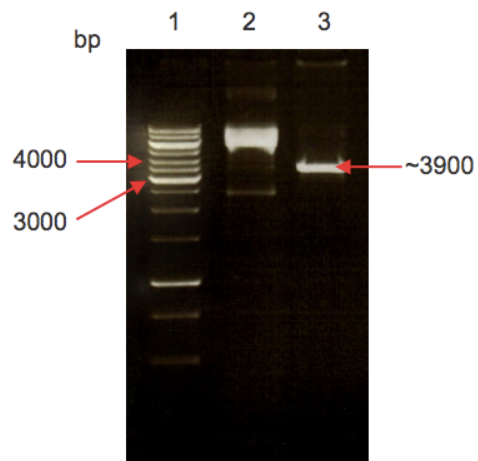
The plasmid was sent for sequencing using the T7 forward primer and the T3 reverse primer and a new primer called GFPneo. The sequencing results confirmed that the plasmid matched its map shown in materials and methods chapter. However, some variation existed in the MCR; there was no EcoRI restriction site but an EcoRV instead. Based on the sequence analysis, the restriction sites for the enzymes HindIII and BamHI were selected, and the next step was to start preparing the pB12mcs-eGFP plasmid for insert intake.

#### **3.2.3.2 Digestion and Purification of the pB12mcs-eGFP Vector**

50µl of JM109 bacterial cells were transformed with 50ng pB12mcs-eGFP. The transformed bacterial cells were grown overnight in 500ml of LB with the appropriate antibiotic. 1ml of the grown bacteria was stored at -80°C as glycerol stock and then a maxi prep took place and the extracted plasmid was quantified. 2µg of the pB12mcs-eGFP vector was diluted and digested with BamHI enzyme. To confirm digestion, the digested plasmid was then analysed

through gel electrophoresis, Figure 3.7. The digested vector was then purified, further digested with the enzyme HindIII, purified and quantified.

The double digestion of the BamHI and HindIII sites should give products of 62bp and 3.887Kb. The presence of the 62bp band cannot be seen when the digestion reaction is analysed on the 1% agarose gel. Only the 3.887Kb band can be seen. Therefore, to ensure that the restriction enzyme HindIII digests the pB12mcs-eGFP plasmid efficiently, 2µg of the wild type pB12mcs-eGFP vector was freshly diluted, and was then digested separately with the HindIII enzyme. 5µl of this digestion was then analysed through electrophoresis. The digested pB12mcs-eGFP plasmid appeared as a single band, in comparison to the supercoiled wild type plasmid, hence confirming that the digestion conditions were correct (results are not shown).



**Figure 3.7 Electrophoretic Analysis of the digested pB12mcs-eGFP vector**

Lane 1: GeneRuler™ 1Kb DNA Ladder

Lane 2: Wild type pB12mcs-eGFP vector

Lane 3: BamHI digested pB12mcs-eGFP vector

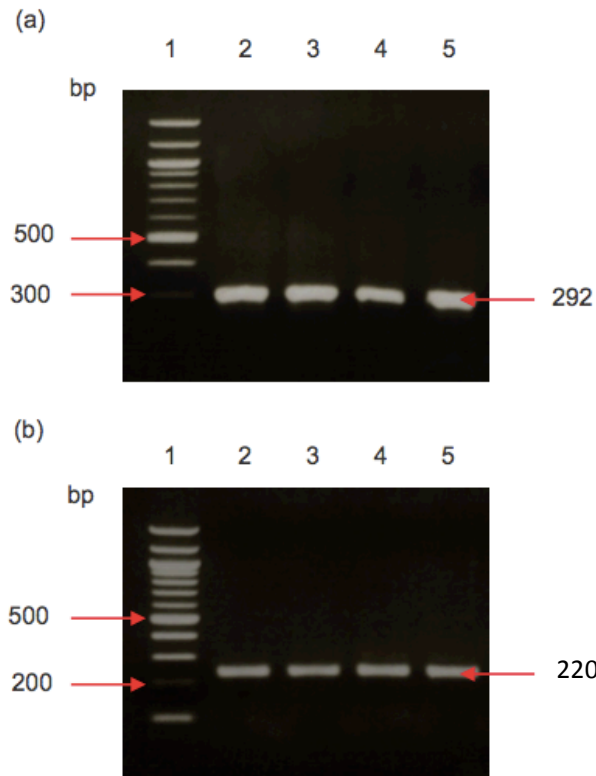
A clear single band of ~3.9Kb in size can be seen in lane 3, upon comparison with the multiband wild type vector observed in lane 2. The multiband appearance is the result of the wild type vector being supercoiled. Overall, the gel indicates digestion of the pB12mcs-eGFP vector with the enzyme BamHI.

### 3.2.4 Amplification of the U6 and H1 Promoters

Once the pB12mcs-eGFP plasmid was digested and purified, the next step was to construct the positive controls by incorporating the U6 human promoter or the H1 human promoter in the MCR of the digested pB12mcs-eGFP plasmid. The sequences of both promoters were found through the EMBL website and the sequences had accession number M14486 for the U6 promoter, and accession number X16612 for the H1 promoter. The complete sequences of these promoters can be found in Appendix 6.

From the sequences found in the website, a U6 primer set and an H1 primer set were designed. The primer sequences were shown in Table 2.2. These primers were designed in such a way that a restriction site for the enzyme HindIII was incorporated in the forward primers, and a restriction site for the BamHI enzyme was incorporated in the reverse primers. PCR using these sets of primers was conducted on cDNA from K562 cell lines as a template. The size of the PCR products for these reactions was expected to be 292bp for the U6 promoter, and 219bp for the H1 promoter. Figure 3.8 demonstrates the electrophoretic analysis of the PCR products of the U6 and H1 human promoters respectively.

The PCR products of the U6 and H1 human promoters were purified, digested with the enzyme BamHI for 3 hours at 37°C with the appropriate buffer. The digested promoters were then precipitated and digested again for another 3 hours at 37°C with the HindIII enzyme and its appropriate buffer. Once, this digestion was achieved, the promoters were precipitated, resuspended in 10µl of Milli Q water, and were then quantified.



**Figure 3.8 Electrophoretic Analysis of the PCR products of the U6 and H1 human promoters**

(a) Top:

Lane 1: 100bp DNA Ladder from NEB

Lanes 2-5: U6 promoter PCR products from four reactions

(b) Bottom:

Lane 1: 100bp DNA Ladder from NEB

Lanes 2-5: H1 promoter PCR products from four reactions

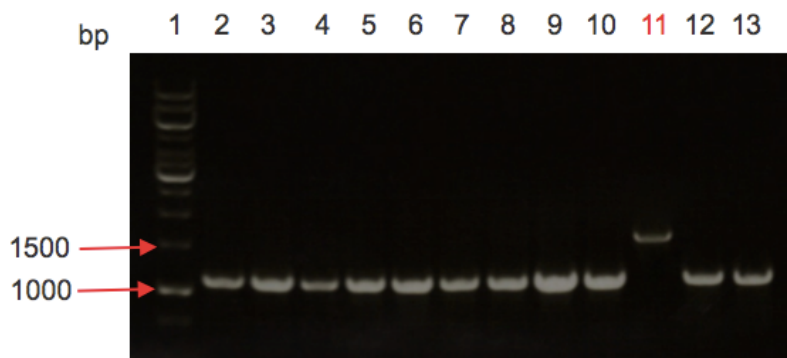
The U6 promoter is 292bp in size, whereas the H1 promoter is 220bp. In lanes 2-5 in gel (a), bands of ~300bp are present, whereas in lanes 2-5 in gel (b), bands just above 200bp are present. These findings confirm that the PCR products are of the size expected.

### **3.2.4.1 Cloning of the U6 and H1 Promoters in the pB12mcs-eGFP Vector**

A 1:10 vector to insert ratio was used to clone the prepared promoters into the double digested pB12mcs-eGFP vector. The ligation reaction was done

overnight at 4°C in a total volume of 10µl. Half the volume of the ligation reaction was used to transform 35µl of JM109 bacterial cells. The cells were heat shocked and were left in a shaking incubator at 37°C for 2 hours before being plated on agar plates with the appropriate antibiotic.

The plate was examined for the growth of bacterial colonies and colonies were screened by means of PCR T7 forward and T3 reverse primers. The wild type pB12mcs-eGFP vector was also included as a negative control. PCR products were analysed by electrophoresis, and are illustrated in Figures 3.9 and 3.10. Positive colonies were grown overnight in 20ml of LB with the appropriate antibiotic in a shaker incubator.



**Figure 3.9 Electrophoretic Analysis of the PCR products of the pB12mcs-U6-eGFP plasmid Colony Screen**

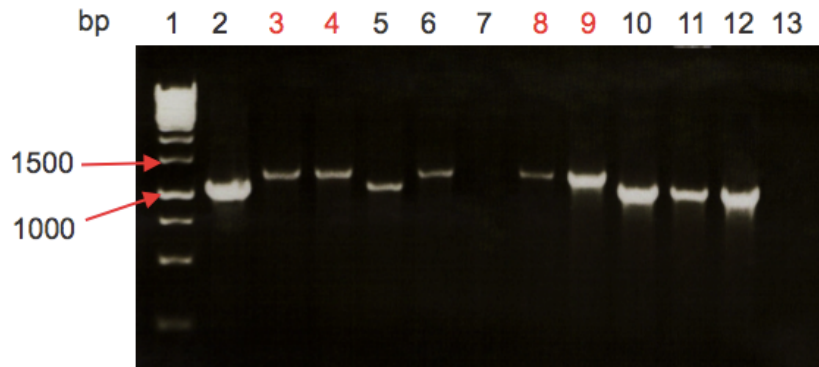
Lane 1: GeneRuler™ 1Kb DNA Ladder

Lane 2: Wild type pB12mcs-eGFP vector PCR product

Lanes 3-13: Colony screen PCR products

Lane 11: Positive Colony

The primer set T7 forward and T3 reverse was used for the above colony screen and the expected PCR product for the wild type vector is ~1050bp, lane 2. The wild type vector was included as a negative control. Lanes 3-10 and 12-13 all have PCR products matching the size of the wild type vector, and were considered to be negative screens. However, in lane 11, highlighted in red, a band of size ~1310bp (~1050bp for the wild type and 292bp for the U6 promoter) is present. This colony was considered to be a positive colony, and was picked up from the agar plate and grown in 20ml of LB, with the appropriate antibiotic.



**Figure 3.10 Electrophoretic Analysis of the PCR products of the Colony Screen of the pB12mcs-H1-eGFP vector**

Lane 1: GeneRuler™ 1Kb DNA Ladder

Lane 2: Wild type pB12mcs-eGFP vector PCR product

Lanes 3,4,6,8, and 9: PCR products of a positive colony screen

Lanes 5 and 10-12: PCR products of a negative colony screen

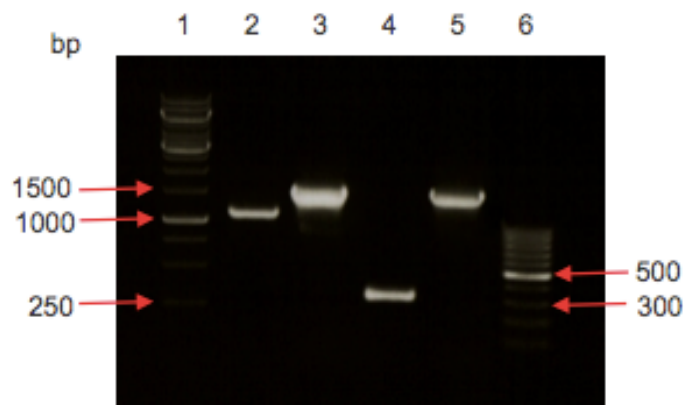
Lanes 7 and 13: PCR did not work

Lane 2 illustrates the PCR product of the wild type pB12mcs-eGFP vector. The band size of this PCR product is of around 1050bp when using the primer set T7 forward with T3 reverse. The band size for the positive colonies were around 1270bp, which is the standard size of the wild type PCR product (~1050) plus the size of the H1 human promoter (220bp). When comparing the PCR products of the colonies screened, the colonies in lanes 5 and 10-12 were negative, whereas colonies in all other lanes were positive (lanes 7 and 13 are excluded as the PCR reaction did not work here). The lanes highlighted in red are the colonies that were selected and grown in 20ml of LB with the appropriate antibiotic.

### 3.2.4.2 Sequence Analysis of the U6 and H1 Promoters

Plasmid extraction took place by means of mini preps, and plasmids were analysed by PCR first. Three sets of primers were used; for U6 analysis T7 forward with T3 reverse, T7 forward with U6 reverse, and U6 forward with T3 reverse. While the set of primers that were used for the plasmids positive for the H1 promoter were T7 forward with T3 reverse, T7 forward with H1 reverse, and H1 forward with T3 reverse. The PCR products of the reactions were

analysed by gel electrophoresis, and are shown in Figures 3.11 and 3.12. Once, positive colonies were confirmed by PCR the plasmids were sent to Geneservice LTD for sequencing.



**Figure 3.11 Electrophoretic Analysis of the PCR products of the pB12mcs-U6-eGFP plasmid (colony 11)**

Lane 1: GeneRuler™ 1Kb DNA Ladder

Lane 2: PCR product of the wild type pB12mcs-eGFP vector using primer set T7 forward and T3 reverse (negative control)

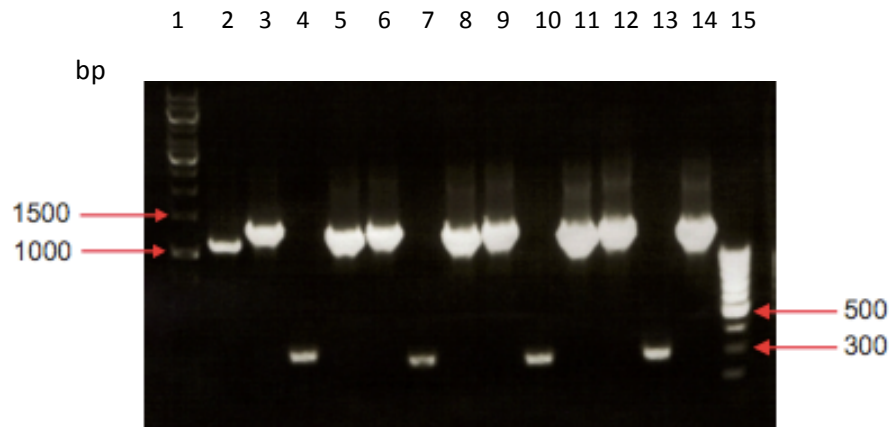
Lane 3: PCR product of the pB12mcs-U6-eGFP vector using primer set T7 forward with T3 reverse

Lane 4: PCR product of the pB12mcs-U6-eGFP vector using primer set T7 forward with U6 reverse

Lane 5: PCR product of the pB12mcs-U6-eGFP vector using primer set U6 forward with T3 reverse

Lane 6: GeneRuler™ 100bp DNA Ladder

To confirm that the plasmid had the right insert in the right orientation, three PCR reactions were conducted on that plasmid. The three different sets of primers that were used were T7 forward with T3 reverse, T7 forward with U6 reverse and U6 forward and T3 reverse. If a U6 promoter has been inserted the sizes of the PCR products expected for the primer sets are 1320bp, 310bp, and 1350bp respectively. The negative control lane 2, was a PCR product of the wild type vector with the first set of primers, giving rise to a ~1050bp band. The electrophoretic analysis illustrates that the resultant PCR products are of the sizes expected, thus confirming that the plasmid in colony 11 had the correct insert.



**Figure 3.12 Electrophoretic Analysis of the PCR products of the extracted pB12mcs-H1-eGFP plasmids (colonies 3,4, 8, and 9)**

Lane 1: GeneRuler™ 1Kb DNA Ladder

Lane 2: PCR product of the wild type pB12mcs-eGFP vector, ~1050bp in size, primer set T7 forward with T3 reverse (negative control)

Lanes 3, 6, 9, and 12: PCR product of the pB12mcs-H1-eGFP vectors, primer set T7 forward with T3 reverse. Colonies 3, 4, 8 and 9 respectively

Lanes 4, 7, 10 and 13: PCR product of the pB12mcs-H1-eGFP vectors, primer set T7 forward with H1 reverse. Colonies 3, 4, 8 and 9 respectively

Lanes 5, 8, 11, and 14: PCR product of the pB12mcs-H1-eGFP vectors, primer set H1 forward with T3 reverse. Colonies 3, 4, 8 and 9 respectively

Lane 15: GeneRuler™ 100bp DNA Ladder

To confirm that the plasmids had the right insert in the right orientation, three PCR reactions were conducted on each plasmid. The three different sets of primers that were used in the reactions were T7 forward with T3 reverse, T7 forward with H1 reverse and H1 forward and T3 reverse. The sizes of the PCR products expected for the primer sets were 1270bp, 260bp, and 1300bp respectively. The analysis illustrates that the resultant PCR products are of the sizes expected, thus confirming that the extracted plasmids had the correct insert.

The sequencing results were analysed and the results for the pB12mcs-U6-eGFP and pB12mcs-H1-eGFP used in this study are found in Appendix 8. The sequences of the U6 and H1 human promoters were compared to their sequences found in EMBL. Those plasmids with a 100% sequence match were



grown in 500ml of LB medium, and a maxi prep was conducted. The extracted plasmids were purified, quantified and stored at -20°C.

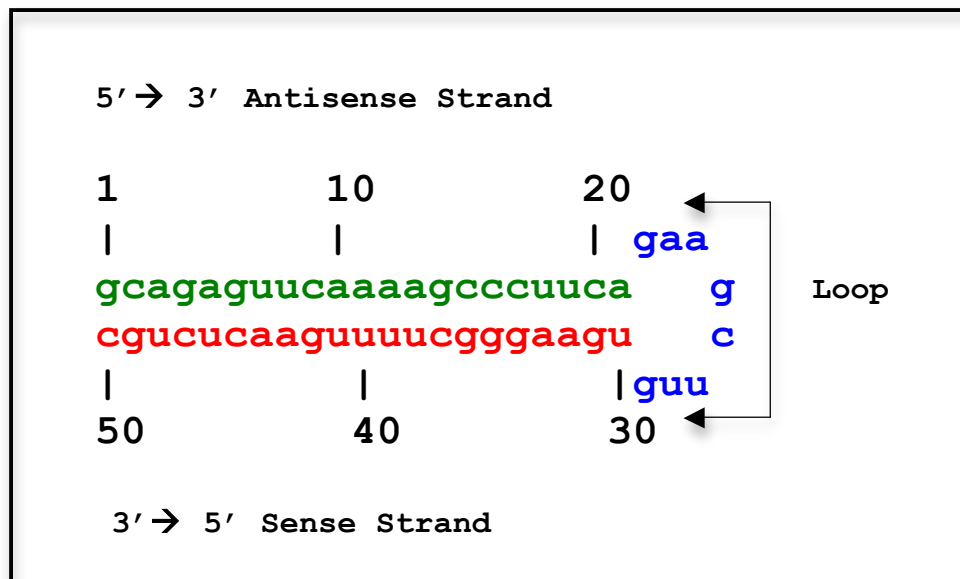
### **3.2.5 Construction of the shRNA Molecules**

Following preparation of the pB12mcs-eGFP vector for insert uptake and the construction of the positive control vectors. The next step was to design the shRNA molecules targeting the two different fusion genes, followed by construction of the shRNA cassettes.

#### **3.2.5.1 Designing of the shRNA Molecules**

The sequences of the  $\beta 3\alpha 2$  and the  $\beta 2\alpha 2$  fusion points found in EMBL website, accession numbers AJ131466 and AJ131467, were used as a basis for designing the shRNA molecules. The full sequences are shown in Appendix 4. For each fusion point three different shRNA molecules were designed, two molecules were 21nt long, whereas the third molecule was 22nt long.

The 21nt long shRNA molecule, shRNA1, had 11nt upstream of the fusion point and 10nt downstream, whereas shRNA2 had 10nt upstream of the fusion point and 11nt downstream. In the case of the 22nt long molecule, shRNA3, 11nt were on either side of the fusion points. Figure 3.13 demonstrates the predicted secondary structure of the shRNA1 molecule targeting the  $\beta 3\alpha 2$  fusion point. The predicted secondary structures for all the remaining shRNA molecules are given in Appendix 2.



**Figure 3.13 The Secondary Structure of shRNA1 Antisense Oligonucleotide targeting the  $\beta 3\alpha 2$  fusion point**

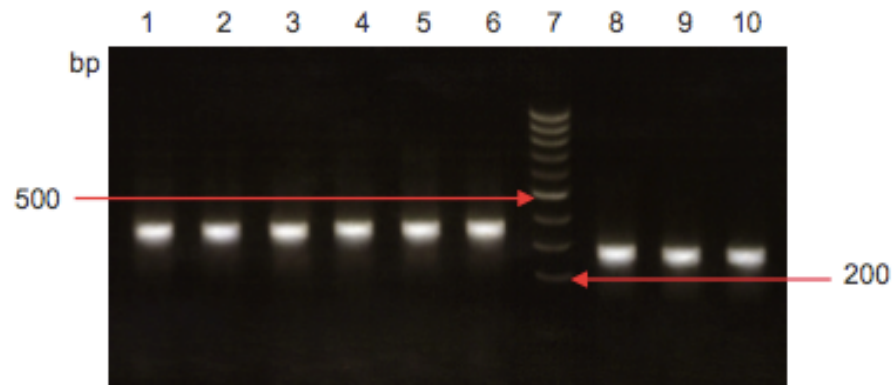
The figure shows the predicted secondary hairpin structure of the designed shRNA1 molecule targeting the  $\beta 3\alpha 2$  fusion point. The nucleotide sequence from 1-21, in green, represents the antisense strand, whereas the sense strand corresponds to the nucleotide sequence from 30-50, in red. The 8nt loop is in blue.

### 3.2.5.2 Synthesis of the U6 and H1 shRNA Cassettes

Once the secondary structures of the shRNA molecules were checked, U6 and H1 human promoter reverse primers, which contained the designed shRNA molecule sequences, were constructed and produced. The sequences of these primers are given in chapter 2 in Table 2.6.

To synthesise the shRNA cassettes, the U6 forward or the H1 forward primer was used together with the U6 or the H1 shRNA reverse primer to conduct a PCR reaction. K562 cDNA was used as a template, and the resultant PCR products were analysed through agarose gel electrophoresis. Figure 3.14 shows the analysis of these PCR products. The bands of the correct size were cut out of the gel, and the shRNA cassettes were purified. The shRNA cassettes were then quantified and digested with the restriction enzymes, BamHI and

HindIII respectively. The shRNA cassettes were purified and quantified after each digestion. The final shRNA cassettes were then stored at -20°C.



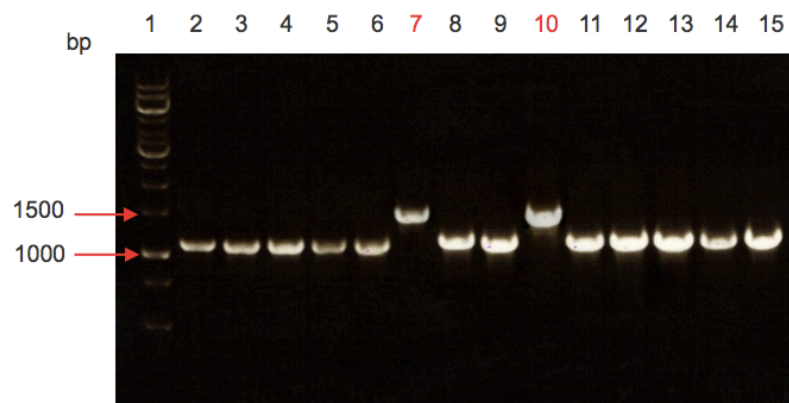
**Figure 3.14 Electrophoretic analysis of PCR products from shRNA cassette synthesis**

- Lane 1: PCR product of the U6  $\beta 3\alpha 2$  shRNA1 cassette
- Lane 2: PCR product of the U6  $\beta 3\alpha 2$  shRNA2 cassette
- Lane 3: PCR product of the U6  $\beta 3\alpha 2$  shRNA3 cassette
- Lane 4: PCR product of the U6  $\beta 2\alpha 2$  shRNA1 cassette
- Lane 5: PCR product of the U6  $\beta 2\alpha 2$  shRNA2 cassette
- Lane 6: PCR product of the U6  $\beta 2\alpha 2$  shRNA3 cassette
- Lane 7: GeneRuler™ 100bp DNA Ladder
- Lane 8: PCR product of the H1  $\beta 2\alpha 2$  shRNA1 cassette
- Lane 9: PCR product of the H1  $\beta 2\alpha 2$  shRNA2 cassette
- Lane 10: PCR product of the H1  $\beta 2\alpha 2$  shRNA3 cassette

The PCR products of the shRNA cassettes were analysed by electrophoresis on a 2% agarose gel. Lanes 1-3 show the PCR products of the shRNA cassettes designed to target the  $\beta 3\alpha 2$  fusion point under the regulation of the U6 promoter. Lanes 4-6 show the PCR products of the shRNA cassettes designed to target the  $\beta 2\alpha 2$  fusion point under the regulation of the U6 promoter. The PCR products should be around 340bp in size, and when analysed above, the visualized bands are of the correct size. Lanes 8-10 show the PCR products of the shRNA cassettes designed to target the  $\beta 3\alpha 2$  fusion point that are under the regulation of the H1 promoter. The expected size of these PCR products is around 290bp.

### 3.2.5.3 Cloning of the shRNA Cassettes in the pB12mcs-eGFP Vector

The digested, purified shRNA cassettes were ligated into the pB12mcs-eGFP vector. The reaction was conducted overnight at 4°C with a 1:10 vector to insert ratio. Colony screening by means of PCR, using the T7 forward and the T3 reverse primer set, was conducted and the wild type pB12mcs-eGFP vector was included as a negative control. An example of the colony screen conducted on one of the U6 shRNA constructs is shown in Figure 3.15. Representative electrophoretic analysis for all the remaining shRNA cassettes can be found in Appendix 7.



**Figure 3.15 Electrophoretic Analysis of the colony screen PCR products of the U6 shRNA1 construct with T7 forward and T3 reverse primers**

Lane 1: GeneRuler™ 1Kb DNA Ladder

Lane 2: PCR product of the wild type pB12mcs-eGFP vector

Lanes 3-15: PCR products of the colony screen

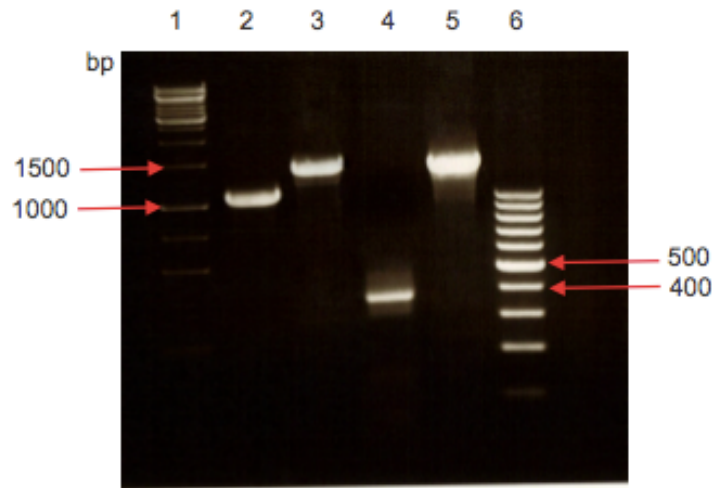
Lanes 7 and 10: Positive colonies

This electrophoretic analysis was conducted on a 2% agarose gel. The above analysis is of the PCR products of the colony screen conducted on the U6 shRNA2 cassette that targets the  $\beta 2\alpha 2$  fusion point. The wild type vector gives a PCR product of ~1050bp. Any colony with the same band size indicates that the insert was not taken up by the vector. However, colonies 7 and 10, which are highlighted in red, show bands of ~1390bp, indicating that the vector has an insert of 340bp. These colonies were therefore picked up further analysis.

### **3.2.5.4 Confirmation by PCR of the shRNA Cassettes and the Sequencing of the Cassettes**

The positive pB12mcs-eGFP-shRNA colonies were grown and purified plasmids were screened by three sets of primers prior to sequencing them. The sets of primers were the T7 forward with the T3 reverse primer, the T7 forward with the appropriate U6 or H1 shRNA reverse primer, and the U6 or H1 forward with the T3 reverse primer. PCR products were analysed by electrophoresis, and colonies positive in all three screens, with the correct band sizes, were sent for sequencing. Figure 3.16 shows an example of the analysis for one of the positive colonies. Representative analyses of the remaining colonies can be seen in Appendix 7.

The sequencing results were analysed and the sequences of both the U6 or H1 promoter and the shRNA molecules were checked. The raw sequencing data of all shRNA cassettes in Appendix 9 and Table 3.1 gives a brief summary and overview about all the shRNA cassettes that have been sequenced. A summary of the sequencing results of the shRNA constructs that were tested, and used for further expression studies can be found in Tables 3.2 – 3.5. Tables 3.2 and 3.4 show the sequencing results of the shRNA part of the constructs. Table 3.3 shows the sequencing results of the U6 part of the constructs and 3.5 the H1 part. The shRNA cassettes that were successfully synthesised included the U6 shRNA1; the U6 shRNA3 that targeted the  $\beta3\alpha2$  fusion point; the U6 shRNA1 that targeted the  $\beta2\alpha2$  fusion point, the H1 shRNA2; and the H1 shRNA3 that targeted the  $\beta2\alpha2$  fusion point. Due to time limitations, the synthesis of further shRNA cassettes was stopped at this point.



**Figure 3.16 Electrophoretic Analysis of the pB12mcs-eGFP Extracted Plasmid with the U6 shRNA1 Construct Targeting the  $\beta 2\alpha 2$  Fusion Gene**

Lane 1: GeneRuler™ 1Kb DNA Ladder

Lane 2: PCR product of the wild type pB12mcs-eGFP vector with the primer set T7 forward and T3 reverse (negative control)

Lane 3: PCR product of the primer set T7 forward with the T3 reverse

Lane 4: PCR product of the primer set T7 forward with the U6 shRNA1 <sup>$\beta 2\alpha 2$</sup>  reverse

Lane 5: PCR product of the primer set U6 forward with the T3 reverse

Lane 6: GeneRuler™ 100bp DNA Ladder

The electrophoretic analysis on a 2% agarose gel shows three PCR products from three different sets of primers. A negative control Lane 2 is included, which is the PCR product of the wild type plasmid with the primer set T7 forward and T3 reverse and a band size of ~1050bp is seen. In the case of a positive colony a band size of ~1390bp is expected, when using the same primer set. This is illustrated in lane 3. Lane 4 is the PCR product of the primer set T7 forward with the U6 shRNA1 <sup>$\beta 2\alpha 2$</sup>  reverse. Here a band size of 380bp is expected and seen. For the last set of primers, the U6 forward and T3 reverse, a band size of ~1420bp is expected (lane 5).

shRNA Construct	Promoter	Sequenced	Sequence outcome *	Used for expression studies	Number of colonies sequenced
$\beta 3\alpha 2$ shRNA1	U6	Yes	2 mismatches	Yes	4
$\beta 3\alpha 2$ shRNA2	U6	Yes	Inconclusive	No	3
$\beta 3\alpha 2$ shRNA3	U6	Yes	3 mismatches	Yes	3
$\beta 2\alpha 2$ shRNA1	U6	Yes	1 mismatch	Yes	4
$\beta 2\alpha 2$ shRNA2	U6	No	-	-	-
$\beta 2\alpha 2$ shRNA3	U6	No	-	-	-
$\beta 2\alpha 2$ shRNA1	H1	Yes	ShRNA sequence was missing	No	13
$\beta 2\alpha 2$ shRNA2	H1	Yes	2 mismatches	No due to time limitations	3
$\beta 2\alpha 2$ shRNA3	H1	Yes	3 mismatches	No due to time limitations	5

**Table 3.1 A summary of the synthesis and sequencing of the shRNA constructs.**

Nine constructs were designed in total, six regulated by the U6 promoter, and three by the H1 promoter. The U6-shRNA constructs targeted both fusion points, however, the H1-shRNA constructs only targeted the  $\beta 2\alpha 2$  fusion point. Several colonies for each construct were sequenced and only those with the correct shRNA sequence, and least number of promoter mismatches, were selected for further studies. \* All mismatches indicated are within the promoter sequences and not the shRNA molecules.

	DNA Fragment	Sequence
1	<b>β3α2 shRNA1<sub>21nt</sub></b>	AAAAAAGCAG AGTTCAAAAG CCCTTCACAA GCTTCTGAAG GGCTTTTGAA CTCTGC
2	<b>β3α2 shRNA3<sub>22nt</sub></b>	AAAAAAGCAG AGTTCAAAAG CCCTTCAGCA AGCTTCCTGA AGGGCTTTTG AACTCTGC
3	<b>β2α2 shRNA1<sub>21nt</sub></b>	AAAAACAAT AAGGAAGAAG CCCTTCACAA GCTTCTGAAG GGCTTCTTCC TTATTG
<b>Sequencing results</b>		
4	<b>U6 β3α2 shRNA1</b>	AAAAAAGCAG AGTTCAAAAG CCCTTCACAA GCTTCTGAAG GGCTTTTGAA CTCTGC
5	<b>U6 β3α2 shRNA3</b>	AAAAAAGCAG AGTTCAAAAG CCCTTCAGCA AGCTTCCTGA AGGGCTTTTG AACTCTGC
6	<b>U6 β2α2 shRNA1</b>	AAAAACAAT AAGGAAGAAG CCCTTCACAA GCTTCTGAAG GGCTTCTTCC TTATTG

**Table 3.2 The shRNA sequencing results of the pB12mcs-U6-shRNA-eGFP constructs.**

The sequences of the originally designed shRNA molecules are shown in rows 1-3. Rows 4-6 show the consensus results of the shRNA molecules that were inserted into the pB12mcs-eGFP vector. The sequencing results confirm that the shRNA molecules found in the synthesised constructs completely match their original designed sequence. The sequences are given as the reverse sequence 3' → 5'.



DNA Fragment	Sequence				
<b>U6 promoter</b> (Accession number M14486)	CAAGGTCGGG ATATACGATA AACACAAAGA TGGGTAGTTT TTACCGTAAC AAAGGACGAA	CAGGAAGAGG CAAGGCTGTT TATTAGTACA GCAGTTTTAA TTGAAAGTAT ACACCGTGCT	GCCTATTTCC AGAGAGATAA AAATACGTGA AATTATGTTT TTCGATTTCT CGCTTCGGCA	CATGATTCCT TTAGAATTAA CGTAGAAAAGT TAAAATGGAC TGGCTTTATA GCACATATAC	TCATATTTGC TTTGACTGTA AATAATTTCT TATCATATGC TATCTTGTGG TA
<b><u>Sequencing Results</u></b>					
<b>U6 β3α2 shRNA1</b>	CAAGGTCGGG ATATACGATA AACACAAAGA TGGGTAGTTT TTACCGTAAC AAAGGACGAA	CAGGAAGAGG CAAGGCTGTT TATTAGTACA GCAGTTTTAA TTGAAAGTAT GCACCGTGCT	GCCTATTTCC AGAGAGCAA AAATACGTGA AATTATGTTT TTCGATTTCT CGCTTCGGCA	CATGATTCCT TTAGAATTAA CGTAGAAAAGT TAAAATGGAC TGGCTTTATA GCACATATAC	TCATATTTGC TTTGACTGTA AATAATTTCT TATCATATGC TATCTTGTGG TA
<b>U6 β3α2 shRNA3</b>	CAAGGTCGGG ATATACGATA AGCACAAAGA TGGGTAGTTT TTACCGTAAC AAAGGACGAA	CAGGAAGAGG CAAGGCTGTT TATTAGTACA GCAGTTTTAA TTGAAAGTAT GCACCGTGCT	GCCTATTTCC AGAGAGATAA AAATACGTGA AATTATGTTT TTCGATTTCT CGTTTCGGCA	CATGATTCCT TTAGAATTAA CGTAGAAAAGT TAAAATGGAC TGGCTTTATA GCACATATAC	TCATATTTGC TTTGACTGTA AATAATTTCT TATCATATGC TATCTTGTGG TA
<b>U6 β2α2 shRNA1</b>	CAAGGTCGGG ATATACGATA AACACAAAGA TGGGTAGTTT TTACCGTAAC AAAGGACGAA	CAGGAAGAGG CAAGGCTGTT TATTAGTACA GCAGTTTTAA TTGAAAGTAT GCACCGTGCT	GCCTATTTCC AGAGAGATAA AAATACGTGA AATTATGTTT TTCGATTTCT CGCTTCGGCA	CATGATTCCT TTAGAATTAA CGTAGAAAAGT TAAAATGGAC TGGCTTTATA GCACATATAC	TCATATTTGC TTTGACTGTA AATAATTTCT TATCATATGC TATCTTGTGG TA

**Table 3.3 The U6 sequencing results of the pB12mcs-U6-shRNA-eGFP Constructs.**

This table summarises the sequencing results from the pB12mcs-U6-eGFP constructs. The sequences shown above are of only the constructs that have been further used in expression studies described in the following chapter. The top of the table illustrates the human U6 promoter sequence found in the EMBL website. Below are the consensus results of the U6 promoter from each construct. Any mismatches in the sequenced data are highlighted in red. The sequences are given as the forward sequence 5' → 3'.

	DNA Fragment	Sequence
1	<b><math>\beta</math>2<math>\alpha</math>2 shRNA2<sub>21nt</sub></b>	AAAAAAAAATA AGGAAGAAGC CCTTCACCAA GCTTCGTGAA GGGCTTCTTC CTTATTG
2	<b><math>\beta</math>2<math>\alpha</math>2 shRNA3<sub>22nt</sub></b>	AAAAACAAT AAGGAAGAAG CCCTTCACCA AGCTTCGTGA AGGGCTTCTT CTTATTGG
<b>Sequencing results</b>		
3	<b>H1 <math>\beta</math>2<math>\alpha</math>2 shRNA2</b>	AAAAAAAAATA AGGAAGAAGC CCTTCACCAA GCTTCGTGAA GGGCTTCTTC CTTATTG
4	<b>H1 <math>\beta</math>2<math>\alpha</math>2 shRNA3</b>	AAAAACAAT AAGGAAGAAG CCCTTCACCA AGCTTCGTGA AGGGCTTCTT CTTATTGG

**Table 3.4 The shRNA sequencing results of the pB12mcs-H1-shRNA-eGFP constructs.**

The sequences of the originally designed shRNA molecules are shown in rows 1 and 2. Rows 3 and 4 show the consensus results of the shRNA molecules that were inserted into the pB12mcs-eGFP vector. The sequencing confirms that the shRNA molecules found in the synthesised constructs completely match their original designed sequence. The sequences are given as the reverse sequence 3' → 5'.

DNA Fragment	Sequence
<b>H1 promoter</b> (Accession number X16612)	GAACGCTGAC GTCATCAACC CGCTCCAAGG AATCGCGGGC CCAGTGTCAC TAGGCGGGAA CACCCAGCGC GCGTGC GCC TGGCAGGAAG ATGGCTGTGA GGGACAGGGG AGTGGCGCC TGCAATATTT GCATGTCGCT ATGTGTTCTG GGAAATCACC ATAAACGTGA AATGTCTTTG GATTTGGGAA TCTTATAAGT TCTGTATGAG ACCACTCTTT
<b>Sequencing Results</b>	
<b>H1 <math>\beta</math>2<math>\alpha</math>2 shRNA2</b>	GAACGCTGAC GTCATCAACC CGCTCCAAGG AATCGCGGGC CCAGTGTCAC TAGGCGGGAA CACCCAGCGC GCGTGC GCC TGGCAGGAAG ATGGCTGTGA GGGACAGGGG AGTGGCGCC TGCAATATCT GCATGTCGCT ATGTGTTCTG GGAAATCACC <b>G</b> TAAACGTGA AATGTCTTTG GATTTGGGAA TCTTATAAGT TCTGTATGAG ACCACTCTTT
<b>H1 <math>\beta</math>2<math>\alpha</math>2 shRNA3</b>	GAACGCTGAC GTCATCAACC CGCTCCAAGG AATCGCGGGC CCAGTGTCAC TAGGCGGGAA CACCCAGCGC GCGTGC GCC TGGCAGGAAG ATGGCTGTGA GGGACAGGGG AGTGGCGCC TGCAATATCT GCATGTCGCT ATGTGTTCTG GGAAATCACC <b>G</b> TAAACGTGA AATGTCTTTG GATTTGGG <b>A</b> TCTTATAAGT TCTGTATGAG ACCACTCTTT

**Table 3.5 The H1 sequencing results of the pB12mcs-H1-shRNA-eGFP Constructs.**

This table summarises the sequencing results obtained from the pB12mcs-H1-eGFP constructs. The sequences shown above are of only the constructs that have been further used in expression studies described in the following chapter. The top of the table illustrates the human H1 promoter sequence found in the EMBL website. Below shows the consensus results of the H1 promoter from each construct. Any mismatches in the sequenced data are highlighted in red. The sequences are given as the forward sequence 5' → 3'.

### 3.3 Discussion

This chapter describes how a reporter gene system was generated. The pB12mcs-eGFP vector was used for this purpose. The vector had an enhanced GFP that is considered to be advantageous when using the system in mammalian cell lines. The plasmid was used to clone several synthesised shRNA cassettes. These shRNA molecules were designed to target the fusion points,  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$ , of the Philadelphia chromosome.

Cell lines chosen for use in this project were the human K562 cell line that bears the  $\beta 3\alpha 2$  fusion point, and the KCL-22 cell line that has the  $\beta 2\alpha 2$  fusion point. It was confirmed that both these cell lines contained the fusion point of interest by means of PCR and it was further confirmed that their sequences matched the  $\beta 3\alpha 2$  and the  $\beta 2\alpha 2$  sequences, that are found in EMBL website. Upon this confirmation, three short hairpin RNA molecules for each fusion point were designed. The secondary structures of the designed antisense oligonucleotides were predicted and upon confirmation that the designed shRNA molecules do have a hairpin like secondary structure, the molecules were synthesised. Synthesis was achieved by PCR, where the sequence of the designed antisense molecule was linked to a reverse primer for the U6 or the H1 human promoters. The shRNA cassettes were therefore, synthesised in a manner where the molecules were downstream to the human U6 or H1 promoters.

The U6 and H1 promoters were selected due to the advantage of their natural presence in human cell lines and the ease of signalling their termination, where a stretch of 4 to 6 'T' nucleotides causes the promoter to terminate. The synthesised shRNA cassettes were then cloned into the double digested pB12mcs-eGFP vector. Positive colonies including both the vector and inserted shRNA cassettes were sequenced and those that have the U6 promoter and the correct shRNA sequence were selected for further expression studies.

The next step was to examine and test how efficient and specific the designed shRNA molecules were to their target sites. The following chapter deals with this issue, where expression studies were conducted on the generated pB12mcs-U6-shRNA-eGFP constructs.

## **Chapter 4 : Expression of the shRNA Constructs in K562 and KCL-22 Cell Lines**

### **4.1 Introduction**

To examine the inhibitory effects that the synthesised shRNA constructs might have on their target sites, these constructs were expressed in two human cell lines K562 and KCL-22. The K562 cell line expresses the  $\beta 3\alpha 2$  fusion protein, whereas the KCL-22 cell line expresses the  $\beta 2\alpha 2$  fusion protein (Drexler *et al.*, 1999). The expression of the shRNA constructs was monitored through the reporter gene system, which has been described previously.

The effects of the synthesised constructs were then examined at both the RNA and protein levels, and the specificity of the shRNA molecules were also examined. In this part of the study, various chemical methods to transfect the synthesised shRNA constructs were undertaken; in addition, one type of physical transfection was also examined: nucleofection, which is a form of electroporation. In summary, this chapter deals with examining the expression of the generated shRNA constructs and their inhibitory effects on their target *BCR-ABL* fusion gene.

### **4.2 Results**

#### **4.2.1 Transfection of the shRNA Constructs in K562 and KCL-22 Cell lines.**

The K562 and KCL-22 cell lines are CML cell lines bearing the  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$  BCR-ABL variants respectively. Leukaemic cell lines are characterised by having

a growth advantage over normal cells, due to their rapid range of proliferation. These cells are unable to mature to functional non-dividing cells, hence leading to their accumulation (Koeffler and Golde, 1980). Lozzio and Lozzio, 1975, first reported the K562 cell line that was established from a pleural effusion of a terminal patient with CML in blast phase. The KCL-22 cell lines were established in the same manner as K562. Kubonishi and Miyoshi, 1983, first reported KCL-22 cell line in 1983.

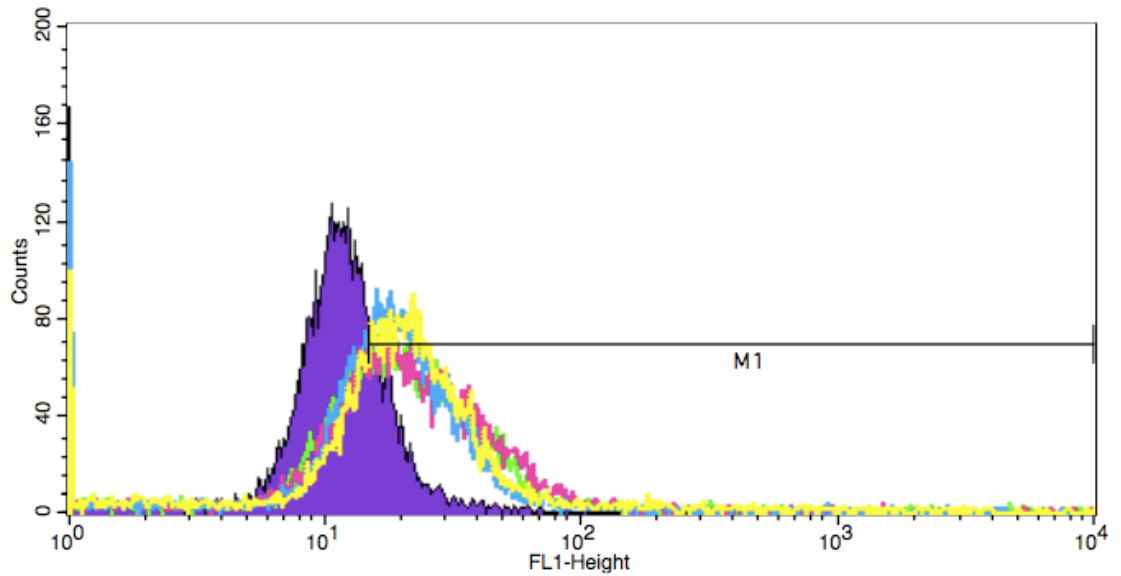
These cell lines grow as single-cell suspension cultures (Lozzio and Lozzio, 1975 and Kubonishi and Miyoshi, 1983) and primary cells and cells in suspensions are hard to transfect (Ovcharenko *et al.*, 2005 and Oliveira and Goodell, 2003). Therefore, in order to select the best method of transfection, with the highest transfection efficiency, both chemical and physical methods were considered.

In this project, two of the three shRNA constructs designed to target the mRNA sequence of the  $\beta 3\alpha 2$  fusion variant were studied. While, one construct that targets the  $\beta 2\alpha 2$  fusion variant was examined.

#### **4.2.2 Transfection by Chemical Methods and FACS Analysis**

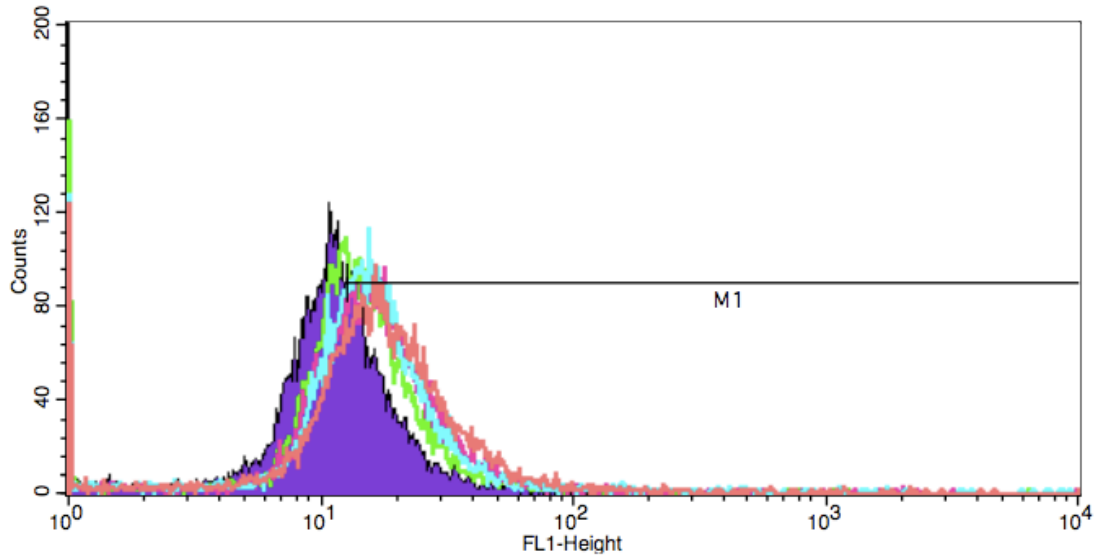
A range of chemical reagents was used to assess transfection efficiency into cell lines, and FACS analysis was performed after transfection. For FACS,  $3 \times 10^5$  cells were counted and the “Cell Quest” software used for analysis. The parameters were adjusted according to the cell line type when the FACS analysis was conducted. Examples of the generated data for the different transfection reagents tested are shown in Figures 4.1 - 4.4. Figures 4.1 and 4.2 demonstrate the highest transfection efficiencies achieved when using a chemical transfection reagent, while Figures 4.3 and 4.4 demonstrate transfection reactions that did not work. Table 4.1 summarises the transfection data obtained when using the different types of chemical transfection reagents. The highest levels of transfection in both K562 and KCL-22 cells were

seen when using feOfection/Yellow as a transfection reagent (K562 48% and KCL-22 38%).



**Figure 4.1 Histogram generated by FACS demonstrating the transfection efficiency of the NIMT FeOfection/Yellow transfection reagent in K562 cell lines.**

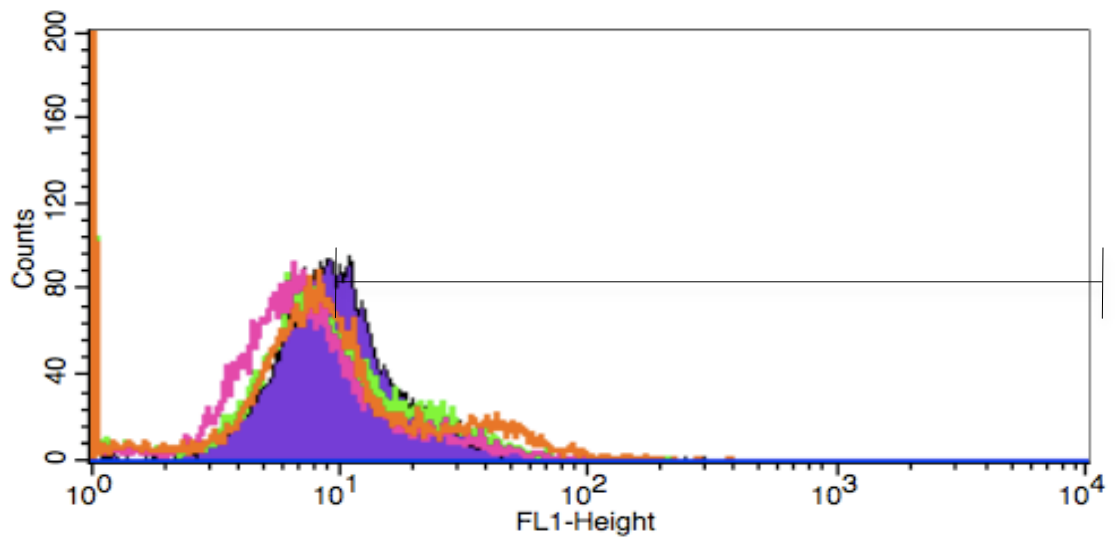
A histogram generated by the program CellQuest by the use of FACS for the K562 cell line transfected with the pEGFP-N1 vector by NIMT FeOfection/Yellow transfection reagent. The control, non-transfected cells, is shown in purple followed by the readings of the different dilutions of the transfection reagent, green, pink, blue, and yellow. These have the following dilutions of the transfection reagent 100, 120, 140, and 160 $\mu$ l, respectively, mixed with 5 $\mu$ l of DNA. The estimated transfection efficiencies are 43.4, 46.56, 43.5, and 48.7% respectively.



**Figure 4.2 Histogram generated by FACS demonstrating the transfection efficiency of the NIMT FeOfection/Yellow transfection reagent in KCL-22 cell line.**

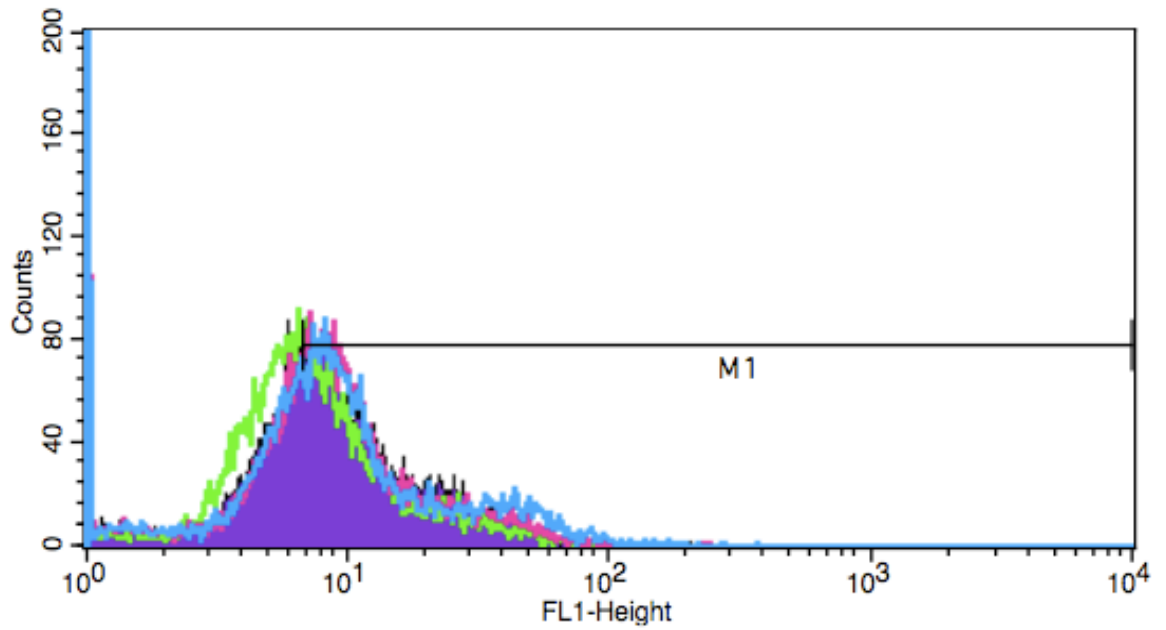
A histogram generated by the program CellQuest by the use of FACS for the K562 cell line transfected with the pEGFP-N1 vector by NIMT FeOfection/Yellow transfection reagent. The control, non-transfected cells, is shown in purple followed by the readings of the different dilutions of the transfection reagent, green, pink, blue, and red. These have the following dilutions of the transfection reagent 100, 120, 140, and 160 $\mu$ l, respectively, mixed with 5 $\mu$ l of DNA. The estimated transfection efficiencies are 19.53, 29.01, 29.33, and 34.68% respectively.





**Figure 4.3 Histogram generated by FACS demonstrating the transfection efficiency of the Nanofectin transfection reagent in K562 cell line.**

A histogram generated by the program CellQuest by the use of FACS for the K562 cell line transfected with the pEGFP-N1 vector by Nanofectin transfection reagent. The control, non-transfected cells, is shown in purple followed by the readings of the different nanofectin ratios 2 (8μl), 3 (12.8μl), and 4 (16μl) that were mixed with 4μl of DNA. These are shown in green, pink, and red respectively.



**Figure 4.4 Histogram generated by FACS demonstrating the transfection efficiency of the Nanofectin transfection reagent in KCL-22 cell line.**

A histogram generated by the program CellQuest by the use of FACS for the KCL-22 cell line transfected with the pEGFP-N1 vector by Nanofectin transfection reagent. The control, non-transfected cells, is shown in purple followed by the readings of the different nanofectin ratios 2 (8 $\mu$ l), 3 (12.8 $\mu$ l), and 4 (16 $\mu$ l) that were mixed with 4 $\mu$ l of DNA. These are shown in green, pink, and blue respectively.

Name of Kit	Supplier	Transfection efficiency	
		K562	KCL-22
Superfect	Qiagen	21-31%	Didn't work
Lipofectamin 2000	Invitrogen	17.7-27%	14.8-18.5%
Hilymax	Dojindo	10-15%	Didn't work
TransIT-2020	Mirus	21.8-40.5%	25.5-26.5%
NIMT FeOfection/Yellow*	Genovis	43-48.7%	29-38%
FuGENE HD	Roche	16-32%	25-30%
Nanofectin	PAA	Didn't work	Didn't work
Lentiphos HT	Clontech	Didn't work	Didn't work

**Table 4.1 A summary of different transfection methods on the transfection efficiency of K562 and KCL-22 cell lines.**

Several transfection kits were tested in order to transfect a control vector, the pEGFP-N1 vector, into the K562 and KCL-22 cell lines. The above table shows a list with all the transfection reagents that have been tested and the range of resultant transfection efficiency obtained from each kit. Some kits didn't work at all, where no fluorescent cells were visible under the fluorescent microscope nor any difference between the control and transfected sample using FACS. \*The highest transfection efficiency was by using the NIMT FeOfection/Yellow kit that gave an efficiency of up to 48.7% in the K562 cell lines.

### 4.2.3 Transfection Through Nucleofection

Chemical methods of transfecting K562 and KCL-22 cells did not give high efficiencies so nucleofection was tried. Nucleofection is a form of electroporation, where specific electrical impulses are used to weaken the target cell's membrane. Pores in the cell's membrane are created, through which DNA fragments can be internalized. In nucleofection, the DNA of interest is mixed with a special reagent (its composition is unknown as the manufacturer does not supply this information), which is then added to the cell line, and a wave of electrical impulses is introduced. The Amaxa

nucleofector system was used in this study. It should be noted that the company supplies different transfecting reagent kits and nucleofector programs depending on the cell line type.

For the K562 cell line a commercially available reagent and two nucleofection programs were available. Both programs were used with the K562 cells and control vector supplied with the nucleofection Kit to determine which program was best for this cell line. However, in the case of the KCL-22 cells, neither reagent nor programs were available. Therefore, the same reagent and programs, previously tested for the K562 cells, were used for the KCL-22 cell line. The program T-016 was found to give best transfection efficiencies and cell viability for both cell lines. The transfection efficiency and cell viability results obtained from both nucleofection programs are summarized in Table 4.2 below.

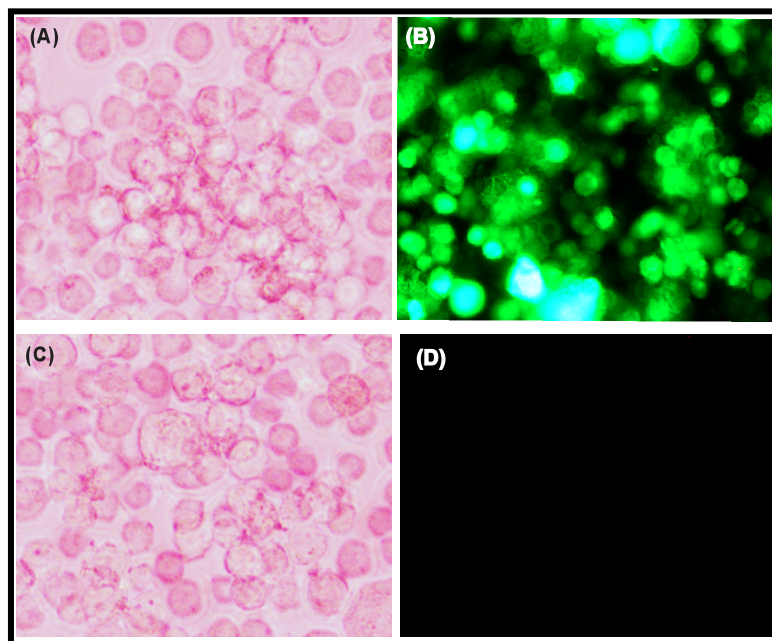
	<b>K562 cell line</b>	<b>T-003</b>	<b>T-016</b>
1	Transfection efficiency	40.4%	54.7%
2	Viability	76%	90%
	<b>KCL-22 cell line</b>	<b>T-003</b>	<b>T-016</b>
3	Transfection efficiency	35%	40%
4	Viability	78%	90%

**Table 4.2 A Summary of the transfection efficiency obtained from FACS analysis and cell viability of K562 and KCL-22 cell lines by Nucleofection with the pB12mcs-U6-eGFP vector.**

K562 and KCL-22 cell lines were transfected using the Amaxa cell line nucleofector® kit V and two nucleofection programs, T-003 and T-016. Each program was for nucleofecting a different K562 clone. The two programs were tested because the clone of the given K562 cell line was unknown. Transfection and viability results for both cell lines showed that the T-016 program had higher transfection efficiency and viability than the T-003. This was observed in both cell lines, consequently, the T-016 program was used for all following nucleofections.

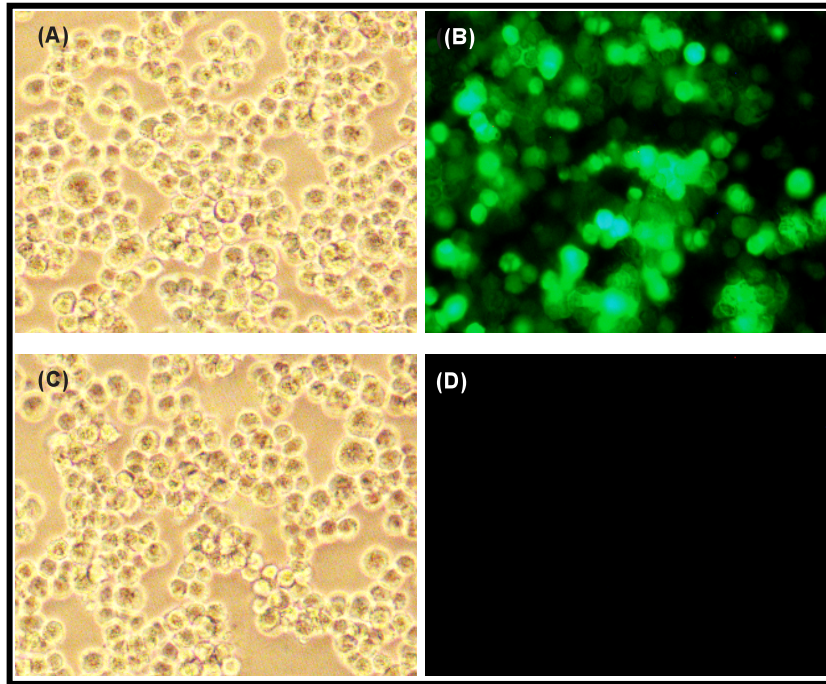
The T-016 program was used in all subsequent studies. To investigate the transfection efficiency with pB12mcs-eGFP vectors, groups of  $2.5 \times 10^6$  cells were transfected with the following: negative control (pB12mcs-eGFP

wildtype), positive control (pB12mcs-U6-eGFP), shRNA1 <sup>$\beta 3\alpha 2$</sup> , and shRNA3 <sup>$\beta 3\alpha 2$</sup>  constructs for the K562 cell line, and negative control (pB12mcs-eGFP wildtype), positive control (pB12mcs-U6-eGFP), and shRNA1 <sup>$\beta 2\alpha 2$</sup>  construct, for the KCL-22 cell line. Successful transfection was confirmed by detection of GFP fluorescence of the target cell lines under the fluorescent microscope, see Figures 4.5 and 4.6. Once the fluorescence of the eGFP was detected, the percentage of cell viability was calculated, through the use of an automated cell counter.



**Figure 4.5 Nucleofection of the K562 cell line with the pB12mcs plasmids.**

(A) Shows the K562 cell line under the fluorescent microscope 24 hours after nucleofection with the pB12mcs-U6-eGFP plasmid (positive control, U6 promoter), in bright field. (B) Shows the cells when excited with blue light, fluorescing in green. (C) Shows the K562 cells transfected with wild type pB12mcs-eGFP vector (negative control, no promoter) in bright field. The same cells are shown in (D) after excitation with blue light.



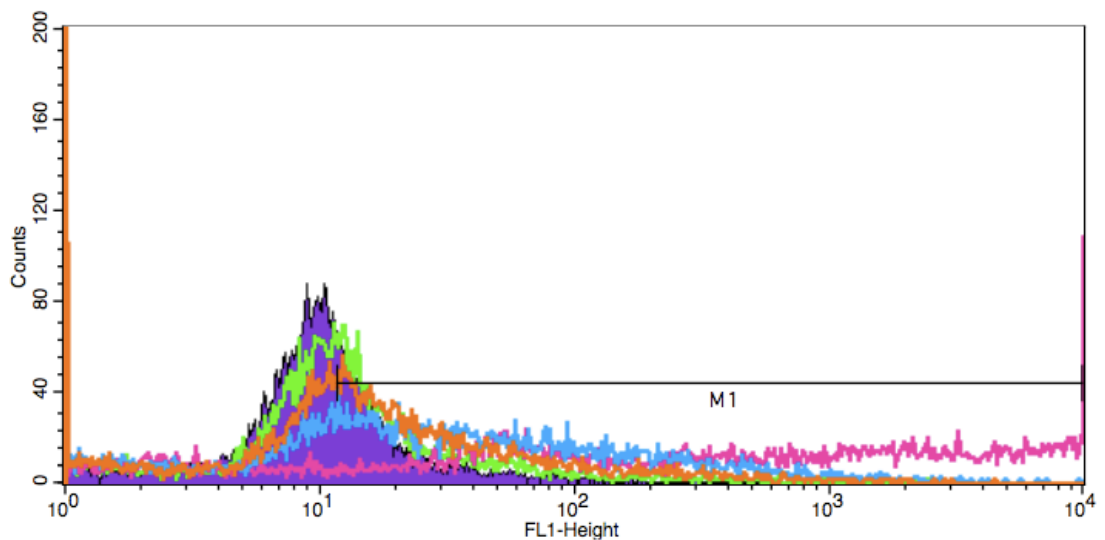
**Figure 4.6 Nucleofection of the KCL-22 cell line with the pB12mcs plasmids.**

(A) Shows the KCL-22 cell line under the fluorescent microscope 24 hours after nucleofection with the pB12mcs-U6-eGFP plasmid (positive control, U6 promoter), in bright field. (B) Shows the cells when excited with blue light, fluorescing in green. (C) Shows the KCL-22 cells transfected with wild type pB12mcs-eGFP vector (negative control, no promoter) in bright field. The same cells are shown in (D) after excitation with blue light.

Once successful transformation was achieved, cell viability was determined and  $3 \times 10^5$  cells from an original number of  $2.5 \times 10^6$  transfected cells were counted and fixed and the transfection efficiency was determined by FACS analysis. Five groups of K562 cell were tested: non-transfected cells, cells transfected with the negative control vector (wild type pB12mcs-eGFP no promoter), cells transfected with the positive control vector (pB12mcs-U6-eGFP), cells transfected with pB12mcs-U6-shRNA1 <sup>$\beta 3\alpha 2$</sup> -eGFP, and cells transfected with pB12mcs-U6-shRNA3 <sup>$\beta 3\alpha 2$</sup> -eGFP. Four groups of KCL-22 cell lines were tested: non-transfected cells, cells transfected with the negative control vector (wild type pB12mcs-eGFP), cells transfected with the positive control vector (pB12mcs-U6-eGFP), and cells transfected with pB12mcs-U6-shRNA1 <sup>$\beta 2\alpha 2$</sup> -eGFP. Figures 4.7 and 4.8 are examples of the FACS results

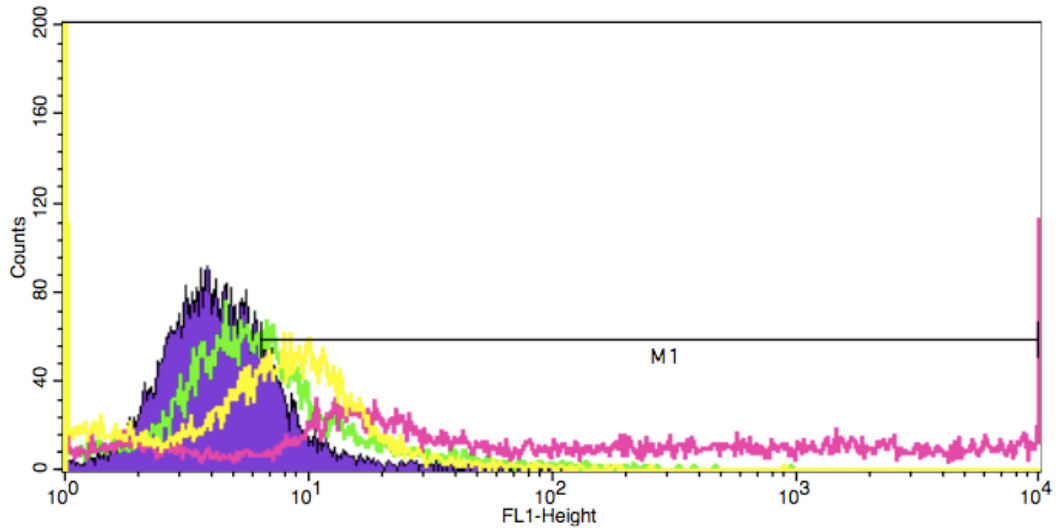
generated from the nucleofections of K562 and KCL-22 groups, and Table 4.3 summarises the transfection efficiency results of three successful nucleofections conducted on each of the cell lines. All the other data is given in Appendix 10.

It was interesting to observe that the shRNA constructs also generated fluorescent signals indicating that the stop signal did not cause termination of the U6 promoter. Advantage of the observed GFP fluorescence generated was taken in measuring efficiency of transfection.



**Figure 4.7 Histogram generated by FACS demonstrating the transfection efficiency of nucleofection in the K562 cell line.**

The histogram is a representation of one of three nucleofections conducted on the K562 cell lines, transfection number 2 in Table 4.3. In purple are the non-transfected K562 cells. In pink are the K562 cells transfected with the positive control vector (pB12mcs-U6-eGFP). In green are the K562 cells transfected with the negative control, pB12mcs-eGFP wild type vector. In blue are the cells transfected with the shRNA1<sup>β3α2</sup> construct, and in orange are the cells transfected with the shRNA3<sup>β3α2</sup> construct. The transfection efficiencies for each with regards to the non-transfected cells are 57%, 11.63%, 35%, and 23% respectively. The transfection efficiencies with regards to the cells transfected with the negative control are 45.08% for the positive control and 23.29% for shRNA1<sup>β3α2</sup> construct and 11.39% for shRNA3<sup>β3α2</sup> construct.



**Figure 4.8 Histogram generated by FACS demonstrating the transfection efficiency of nucleofection in the KCL-22 cell line.**

The histogram is a representation of one of three nucleofections conducted on the KCL-22 cells, transfection number 2 in Table 4.3. In purple are the non-transfected KCL-22 cells; in pink are the cells transfected with the positive control vector (pB12mcs-U6-eGFP); in green are cells transfected with the negative control (pB12mcs-eGFP wild type vector), and in yellow are the cells transfected with the shRNA1 <sup>$\beta 2\alpha 2$</sup>  construct. The transfection efficiencies for each with regards to the non-transfected cells are 64%, 19.08%, and 29.1% respectively. The transfection efficiencies with regards to the cells transfected with the negative control are 44.9% for the positive control and 10.02% for shRNA1 <sup>$\beta 2\alpha 2$</sup>  construct.

Clearly the negative control in this transfection exhibiting GFP fluorescence and will be discussed latter.



K562 Cell Line	(A) Transfection efficiency of the positive control plasmid (pB12mcs-U6-eGFP)	(B) Cell viability of cells transfected with:				
		Non transfected cells	Negative control	Positive control	shRNA1 $\beta3\alpha2$	shRNA3 $\beta3\alpha2$
Transfection 1	28%	99%	89%	90%	86%	80%
Transfection 2	57%	98%	90%	92%	84%	82%
Transfection 3	72%	99%	90%	90%	87%	84%
KCL-22 cell line	(C) Transfection efficiency of the positive control plasmid (pB12mcs-U6-eGFP)	(D) Cell viability of cells transfected with:				
		Non transfected cells	Negative control	Positive control	shRNA1 $\beta2\alpha2$	
Transfection 1	53%	98%	90%	89%	87%	
Transfection 2	64%	98%	88%	91%	86%	
Transfection 3	67%	99%	90%	88%	87%	

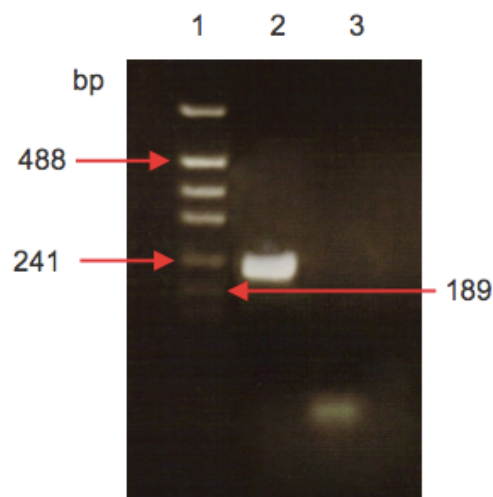
**Table 4.3 A Summary of the three nucleofections conducted on each of the human leukaemic cell lines, K562 and KCL-22, and the cell viability in each of the transfected cell groups.**

Each K562 cell transfection consisted of 5 groups; non-transfected cells, cells transfected with the wild type pB12mcs-eGFP vector (negative control), cells transfected with the pB12mcs-U6-eGFP vector (positive control), cells transfected with the shRNA1 <sup>$\beta3\alpha2$</sup>  construct, and cells transfected with the shRNA3 <sup>$\beta3\alpha2$</sup>  construct. The transfection efficiency (positive control plasmid) obtained from FACS analysis for each transfection is shown in part (A) of the table. The cell viability for each group of transfected cells is indicated in part (B). The bottom parts of the table (C) and (D) illustrate a similar table but for the KCL-22 cell line with cells transfected with the wild type pB12mcs-eGFP vector (negative control), cells transfected with the pB12mcs-U6-eGFP vector (positive control), cells transfected with the shRNA1 <sup>$\beta2\alpha2$</sup>  construct.

#### 4.2.4 Real Time PCR Analysis of the Transfected Cell Lines

qPCR was used to study the effects of the generated shRNA constructs at the mRNA levels of the fusion gene of interest. To do so, several qPCR primers were designed. Primers that spanned the fusion points,  $\beta3\alpha2$  and the  $\beta2\alpha2$ , and their appropriate probes were designed by the program 'primer3'. Primers that amplified the normal *BCR* and *ABL* genes and the *GAPDH* reference gene were generated through the Roche UPL library software, which has a set of 250

probes. All primers gave rise to PCR products that were no more than 230bp in size. Primers were tested for their specificity to their target sequences. This was achieved through conducting a PCR reaction on cDNA extracted from K562 and KCL-22 cell lines. The electrophoretic analysis of the products of these PCR reactions can be seen in figures 4.9 - 4.11. Figure 4.9 shows an electrophoretic analysis of PCR products with GAPDH primers using K562 cDNA as a template. Figure 4.10 shows the electrophoretic analysis of PCR products with ABL and BCR primers. The electrophoretic analysis with  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$  primers are shown in Figure 4.11.



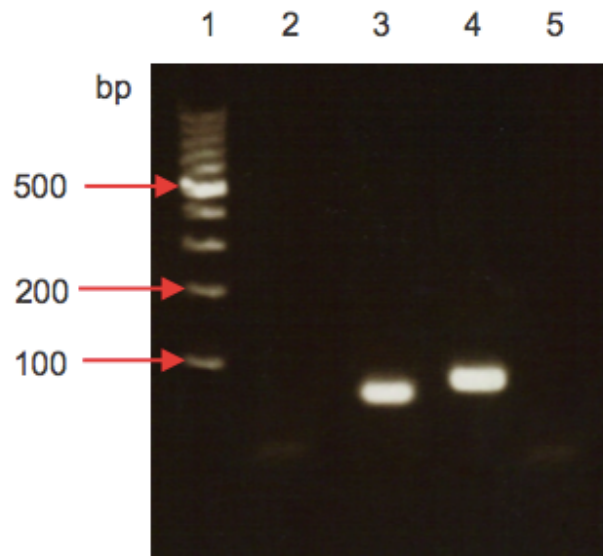
**Figure 4.9 Electrophoretic analysis of PCR products of GAPDH primers using K562 cDNA**

Lane 1: pBluescript Hpa II marker

Lane 2: PCR product of GAPDH primers using K562 cDNA

Lane 3: PCR product of GAPDH primers using MilliQ H<sub>2</sub>O

GAPDH PCR products were analysed by the use of a 2% agarose gel. Lane 2 shows a 226bp band, which is the size of the GAPDH PCR product. PCR in lane 3 is a negative control.



**Figure 4.10 Electrophoretic analysis of PCR reactions using the ABL and BCR primers on K562 cDNA**

Lane 1: GeneRuler™ 100bp DNA ladder

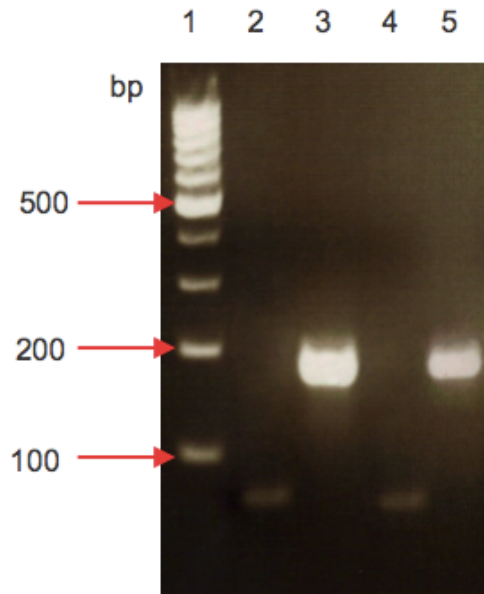
Lane 2: PCR product of MilliQ H<sub>2</sub>O using the ABL primers (negative control)

Lane 3: PCR product of K562 cDNA using the ABL primers

Lane 4: PCR product of K562 cDNA using the BCR primers

Lane 5: PCR product of MilliQ H<sub>2</sub>O using the BCR primers (negative control)

The PCR products of ABL and BCR primers were analysed through electrophoresis using a 2% agarose gel. In Lane 3, the ABL PCR product of 77bp is present, in lane 4, a BCR PCR product of 91bp is shown. PCR in lanes 2 and 5 didn't give a product.



**Figure 4.11 Electrophoretic analysis of PCR reaction using  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$  primers with K562 and KCL-22 cDNA, respectively.**

Lane 1: GeneRuler™ 100bp DNA ladder

Lane 2: PCR product of MilliQ H<sub>2</sub>O using the  $\beta 3\alpha 2$  primers (negative control)

Lane 3: PCR product of K562 cDNA using the  $\beta 3\alpha 2$  primers

Lane 4: PCR product of MilliQ H<sub>2</sub>O using the  $\beta 2\alpha 2$  primers (negative control)

Lane 5: PCR product of KCL-22 cDNA using the  $\beta 2\alpha 2$  primers

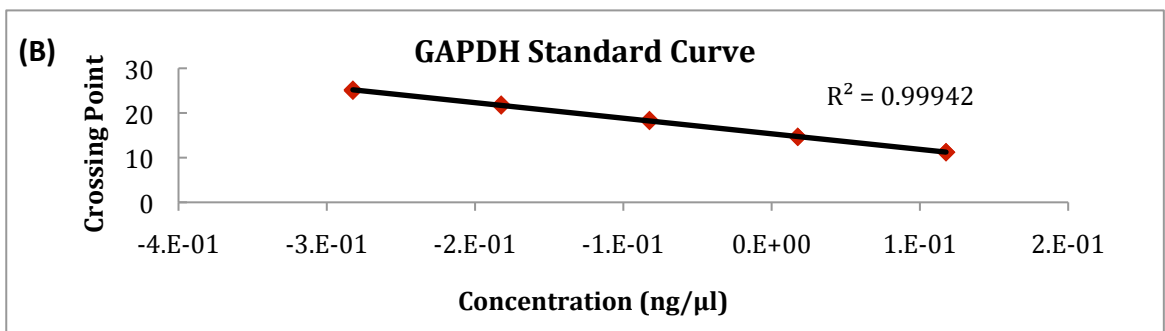
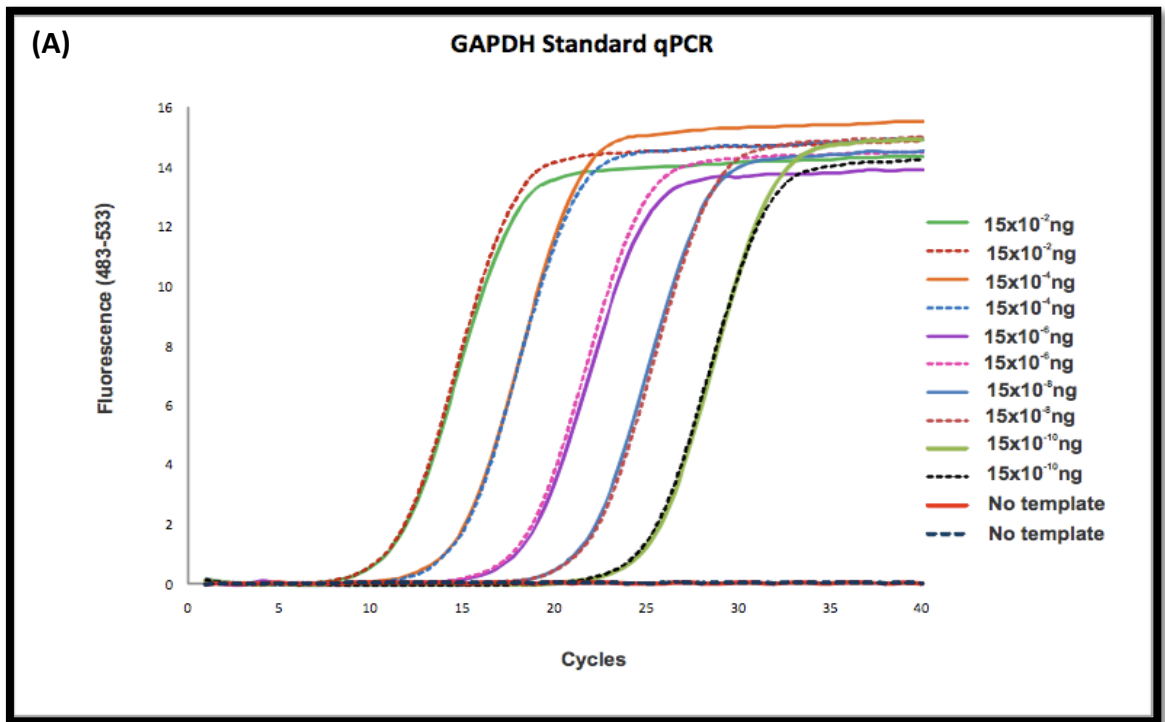
The PCR products with  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$  primers were analysed through electrophoresis using a 2% agarose gel. In Lane 3, a  $\beta 3\alpha 2$  PCR product of 190bp is present, in lane 5, a  $\beta 2\alpha 2$  PCR product of 183bp is shown. PCR in lanes 2 and 4 didn't give a product.

Electrophoretic analysis of all hydrolysis primers confirmed the specificity of the primers, as only one band of the right size is present in each PCR reaction. In addition, no bands were observed in the negative controls when MilliQ H<sub>2</sub>O was used as a template.

A reference gene, GAPDH, is included in this analysis as a control. The normal BCR and ABL levels were also examined; this was done to investigate whether the shRNA constructs had an inhibitory effect on mRNA production of the normal genes as well as the fusion genes.

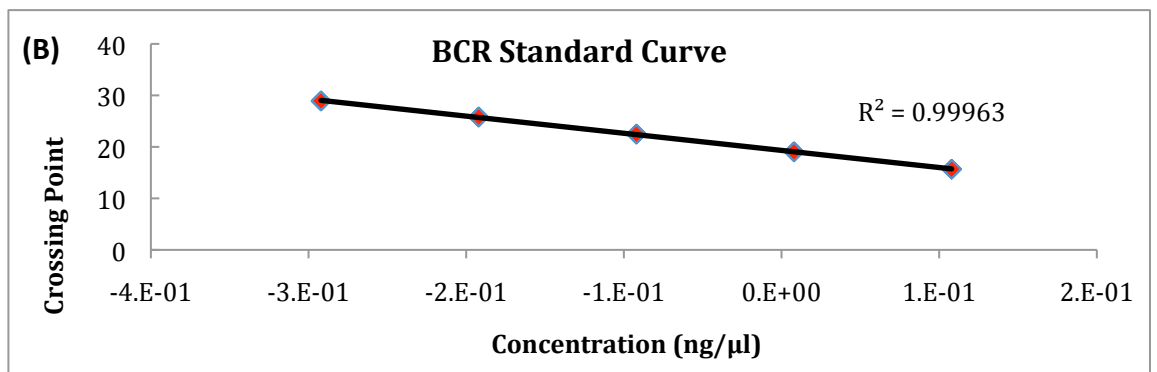
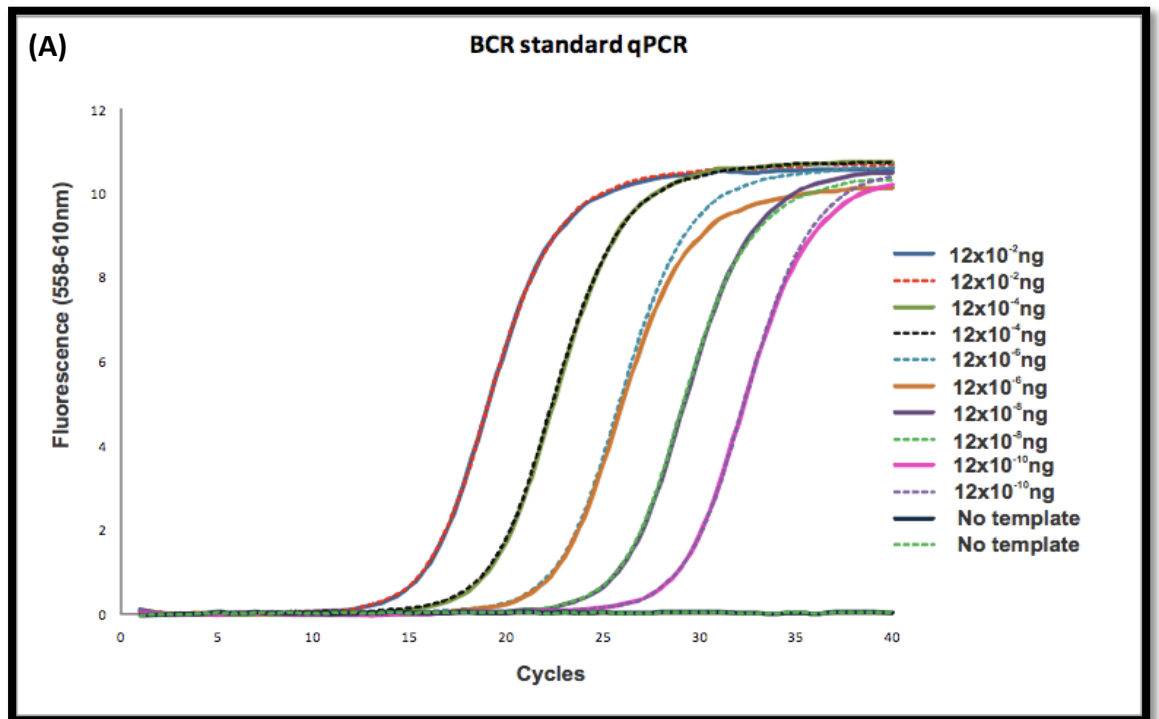
#### 4.2.4.1 The Generation of Standard Curves

Preliminary experiments used SYBR green I methodology to validate the primers. Once validated standard curves for each gene product were generated by the use of Taqman and UPL probes and primers. This was achieved by conducting a qPCR on several dilutions of plasmids in the case of the  $\beta 3\alpha 2$  and the  $\beta 2\alpha 2$  fusion points or on several dilutions of a PCR product, in case of the *GAPDH*, normal *BCR*, and normal *ABL* genes. The starting point for the concentration of each gene ranged from 10-30ng (per 100 $\mu$ l), then each sample was diluted 100x to give rise to a total of 5 samples. The qPCR results were analysed through the Roche LightCycler 480 software. Figures 4.12-4.16 show the generated standard curves.



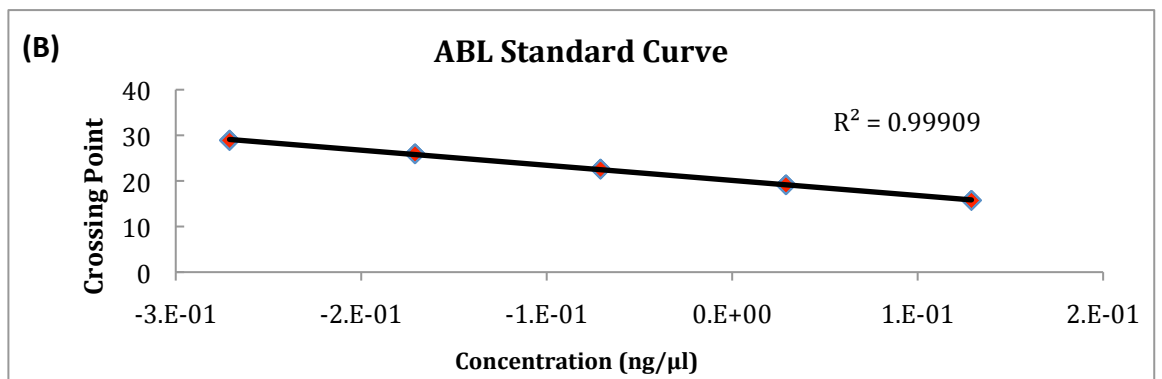
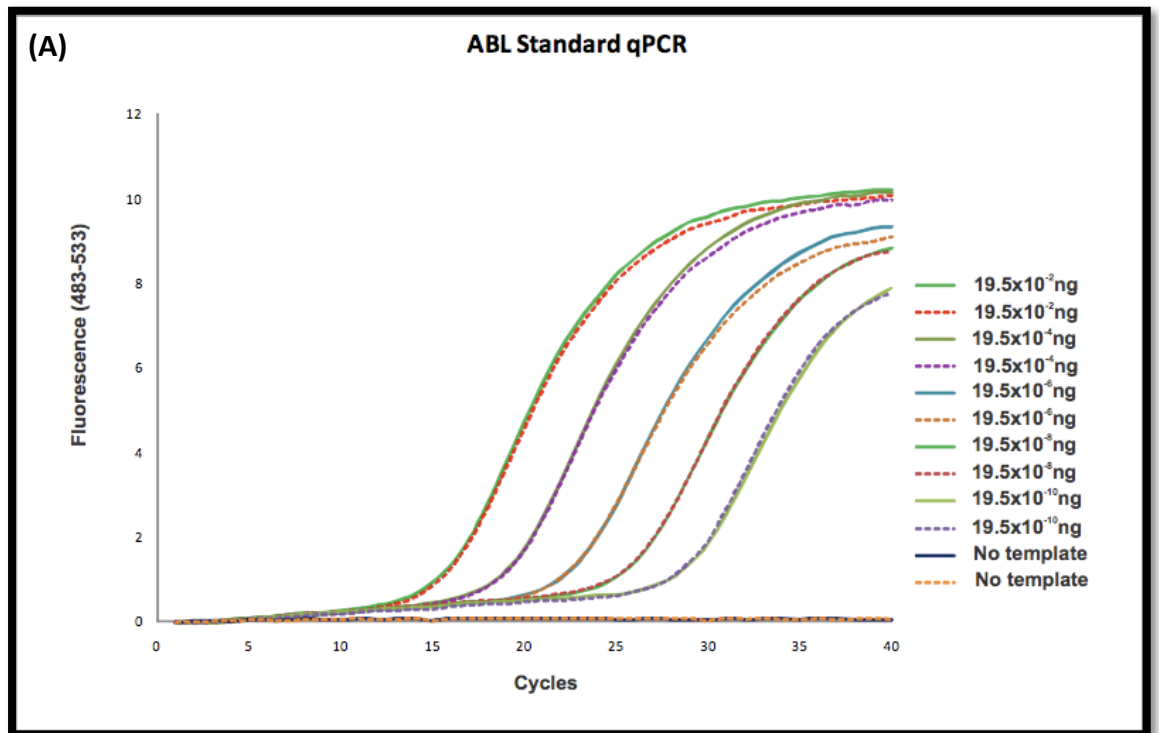
**Figure 4.12 The qPCR of GAPDH standards and the generated Standard Curve.**

The qPCR results of the GAPDH standards are shown in (A). The qPCR was conducted on five dilutions of a GAPDH PCR product. The starting concentration, of 0.15ng/μl of a purified GAPDH PCR product, and the dilutions used are also shown here. The qPCR reactions were performed in duplicate, and the results analysed by the Light Cycler 480 software to determine the values of the GAPDH standard curve in (B). The standard curve is a least square fit line drawn through all dilutions, which is the mean as the reactions were done in duplicate. The calculated amplification efficiency was 91.3%. This curve was further used to calculate the concentration of the GAPDH gene product in the cell lines of interest.



**Figure 4.13 The qPCR of BCR standards and the generated Standard Curve.**

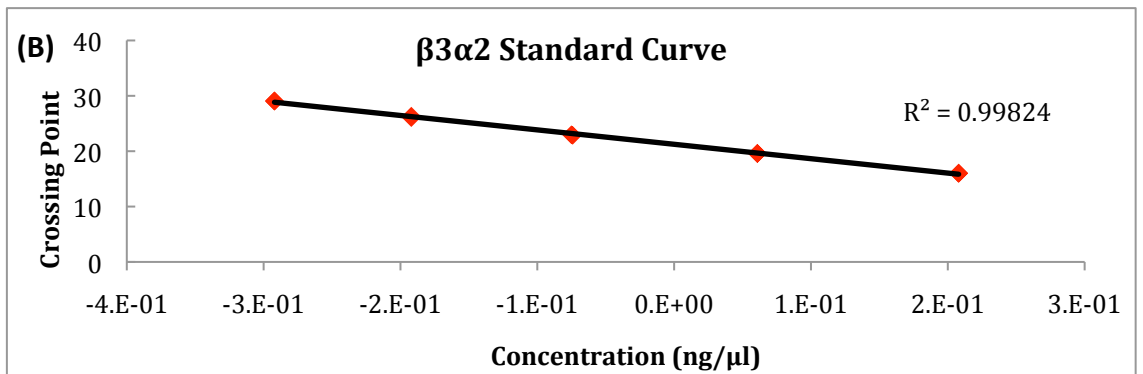
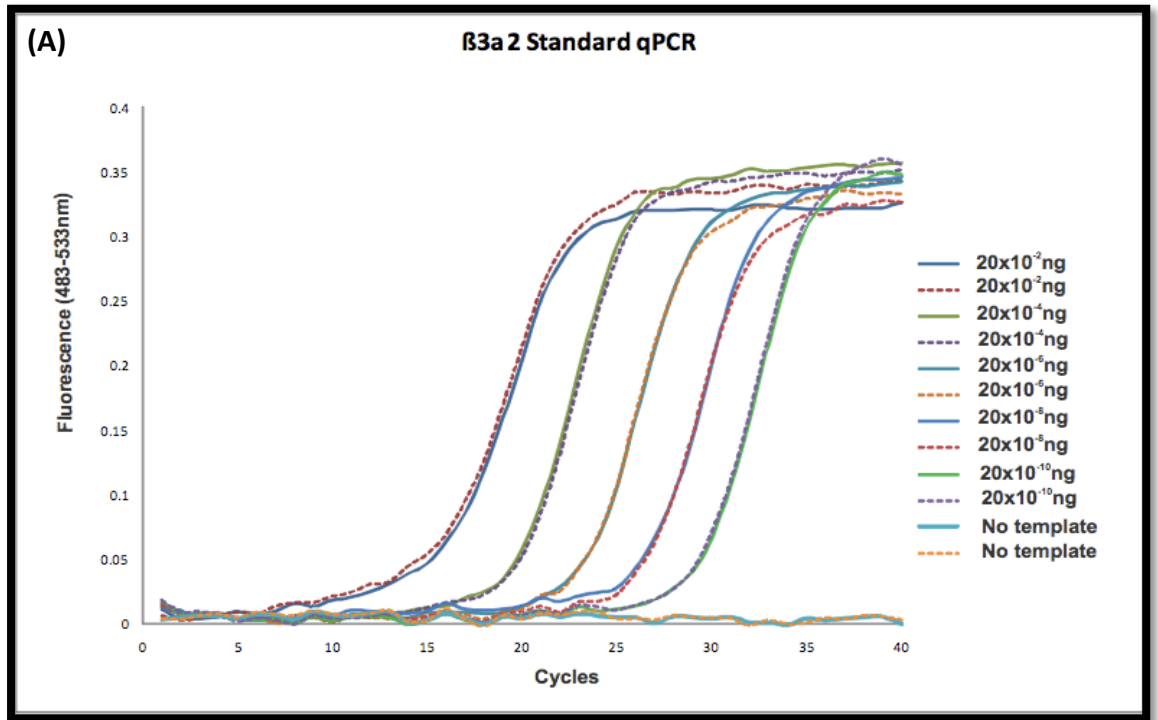
The qPCR results of the BCR standards are shown in (A). The qPCR was conducted on five dilutions of a BCR PCR product. The starting concentration, of 0.12ng/μl of a purified BCR PCR product, and the dilutions used are also shown here. The qPCR reactions were performed in duplicate, and the results analysed by the Light Cycler 480 software to determine the values of the BCR standard curve in (B). The standard curve is a least square fit line drawn through all dilutions, which is the mean as the reactions were done in duplicate. The calculated amplification efficiency was 96.4%. This curve was further used to calculate the concentration of the *BCR* gene product in the cell lines of interest.



**Figure 4.14 The qPCR of ABL standards and the generated Standard Curve.**

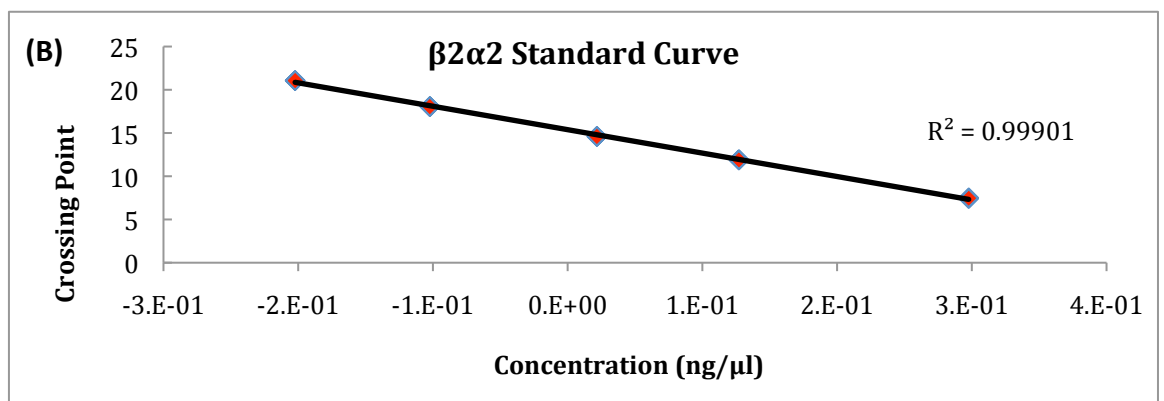
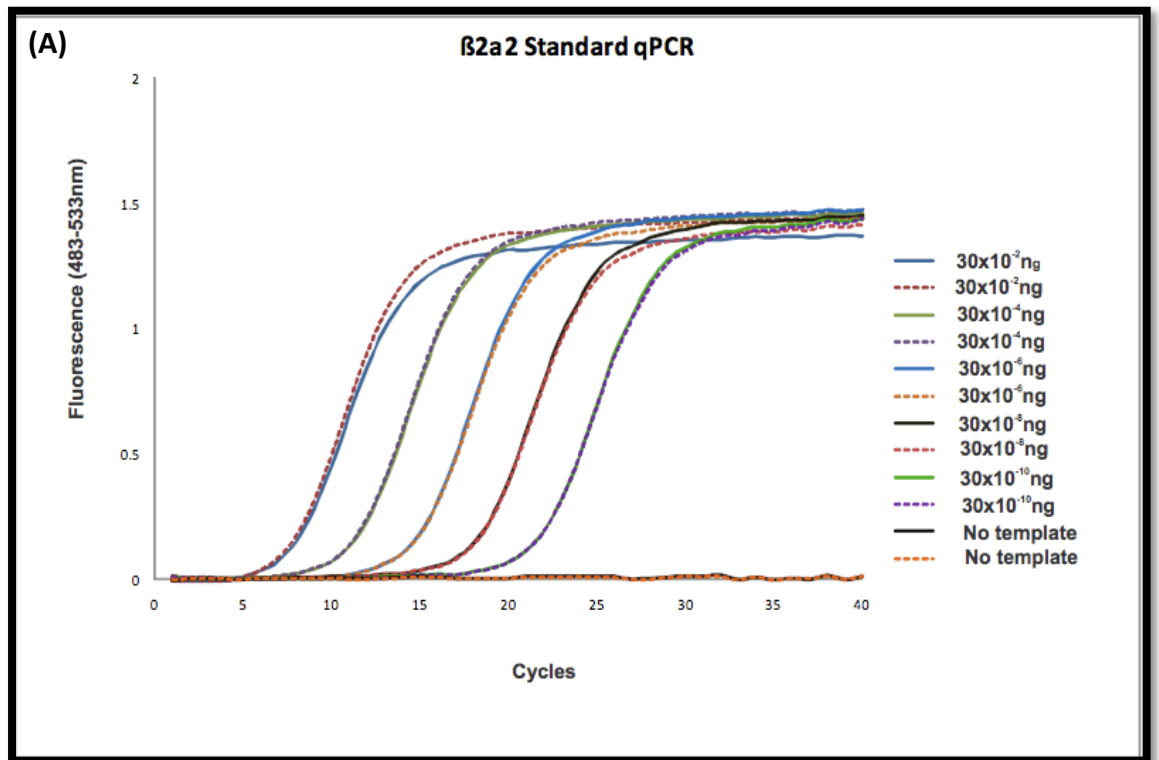
The qPCR results of the ABL standards are shown in (A). The qPCR was conducted on five dilutions of a ABL PCR product. The starting concentration, of 0.195ng/μl of a purified ABL PCR product, and the dilutions used are also shown here. The qPCR reactions were performed in duplicate, and the results analysed by the Light Cycler 480 software to determine the values of the ABL standard curve in (B). The standard curve is a least square fit line drawn through all dilutions, which is the mean as the reaction was done in duplicate. The calculated amplification efficiency was 96.6%. This curve was further used to calculate the concentration of the ABL gene product in the cell lines of interest.





**Figure 4.15 The qPCR of  $\beta 3\alpha 2$  standards and the generated Standard Curve.**

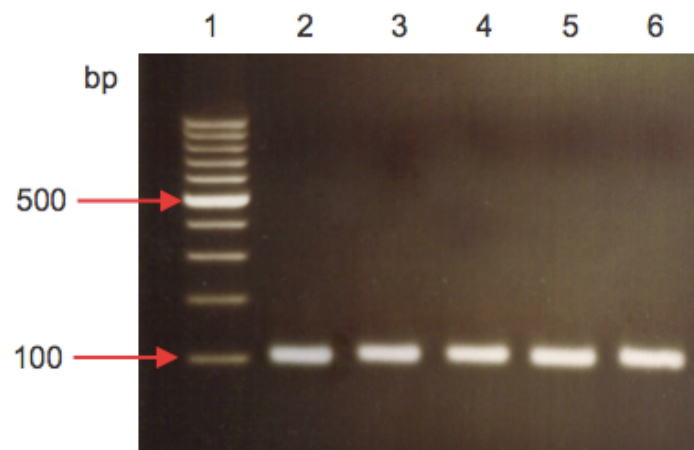
The qPCR results of the  $\beta 3\alpha 2$  standards are shown in (A). The qPCR was conducted on five dilutions of a  $\beta 3\alpha 2$  containing plasmids. The starting concentration, of  $0.2\text{ng}/\mu\text{l}$  of a  $\beta 3\alpha 2$  construct, and the dilutions used are also shown here. The qPCR reactions were performed in duplicate, and the results analysed by the Light Cycler 480 software to determine the values of the  $\beta 3\alpha 2$  standard curve in (B). The standard curve is a least square fit line drawn through all dilutions, which is the mean as the reaction was done in duplicate. The calculated amplification efficiency was 95%. This curve was further used to calculate the concentration of the  $\beta 3\alpha 2$  fusion gene product in the cell lines of interest.



**Figure 4.16 The qPCR of  $\beta 2\alpha 2$  standards and the generated Standard Curve.**

The qPCR results of the  $\beta 2\alpha 2$  standards are shown in (A). The qPCR was conducted on five dilutions of a  $\beta 2\alpha 2$  containing plasmid. The starting concentration, of  $0.3 \text{ ng}/\mu\text{l}$  of a purified  $\beta 2\alpha 2$  construct, and the dilutions used are also shown here. The qPCR reactions were performed in duplicate, and the results analysed by the Light Cycler 480 software to determine the values of the  $\beta 2\alpha 2$  standard curve in (B). The standard curve is a least square fit line drawn through all dilutions, which is the mean as the reaction was done in duplicate. The calculated amplification efficiency was 89.7%. This curve was further used to calculate the concentration of the  $\beta 2\alpha 2$  fusion gene product in the cell lines of interest.

Once the standard curves were established,  $2.5 \times 10^6$  cells, K562 or KCL-22, were nucleofected, and  $3 \times 10^5$  cells were used for FACS analysis, and  $1 \times 10^6$  cells were used for total RNA extraction, followed by cDNA synthesis. Another sample of  $1 \times 10^6$  of transfected cells was used for protein extraction. A PCR reaction on the newly synthesised cDNA, to confirm that the RT-PCR had worked, was routinely done. Primers for the RPL13 gene (ribosomal protein L13 gene), a housekeeping gene, were also used in this reaction, and a band size of 96bp was expected; Figure 4.17 shows an example of the electrophoretic analysis. Normal ABL and BCR levels were also examined; this was done to investigate whether the shRNA constructs reduced the mRNA levels of the normal genes. Five qPCR reactions (for *GAPDH*, *ABL*, *BCR*,  $\beta 3\alpha 2$ , and  $\beta 2\alpha 2$ ) were done in duplicate for each sample. The standard curves were used for each transfection to determine the concentrations of genes of interest, the results were analysed by the Roche LightCycler 480 software. The software calculates the mRNA concentration levels. The levels were analysed and the percentages of depletion were calculated.



**Figure 4.17 An example of an electrophoretic Analysis of the PCR reaction using RPL13 primers on transfected K562 cells cDNA.**

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: PCR product of the non-transfected K562 cDNA

Lane 3: PCR product of the K562 transfected cDNA wild type pB12mcs-eGFP

Lane 4: PCR product of the K562 transfected cDNA wild type pB12mcs-U6-eGFP

Lane 5: PCR product of the K562 transfected cDNA wild type pB12mcs-U6-shRNA1<sup>β3α2</sup>-eGFP

Lane 6: PCR product of the K562 transfected cDNA wild type pB12mcs-U6-shRNA3<sup>β3α2</sup>-eGFP

RPL13 PCR products were analysed via electrophoresis using a 2% agarose gel .The 96bp band in lanes 2-6 indicates that the RNA purification and K562 cDNA synthesis were successful in all the transfections including the non-transfected K562 control.

GAPDH was used as a reference gene, and the qPCR data generated in this reaction were used to normalize the data obtained from the other genes.

#### 4.2.4.2 Effects of shRNA1<sup>β3α2</sup> and shRNA3<sup>β3α2</sup> in K562 and KCL-22 cells

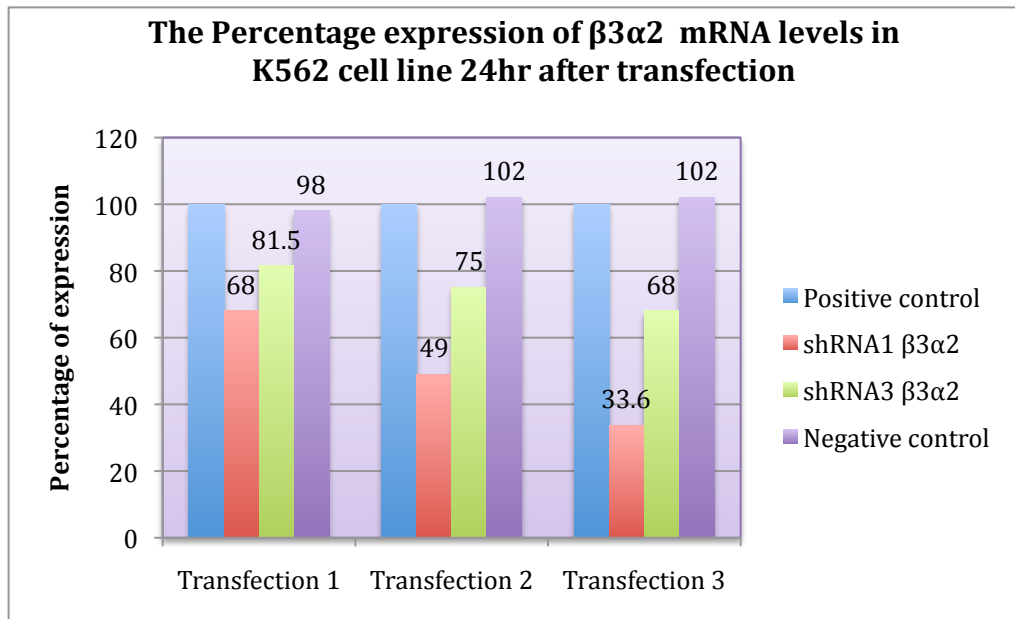
To study the effects of the shRNA<sup>β3α2</sup> constructs, the RNAi molecules were transfected into K562 cells that express the target site, the β3α2 BCR-ABL variant. To examine the specificity of the constructs KCL-22 cells expressing the β2α2 variant that has the same *ABL* sequence as β3α2 were used. To measure the β3α2 mRNA levels, qPCR was conducted using primers specific only to the β3α2 fusion gene. The forward primer binds to exon 14 of the *BCR* part of the gene, while the reverse primer binds to the *ABL* part. Similarly, the β2α2 variant was measured using primers specific only for this fusion gene. The forward primer binds to exon 13 of the *BCR* part, while the reverse primer binds to the *ABL* part of the fusion gene.

Table 4.4 demonstrates the normalized qPCR data generated from the K562 cell lines transfected with the shRNA1<sup>β3α2</sup> and shRNA3<sup>β3α2</sup> constructs. The data were then converted to percentage of expression with regards to the positive control and a bar chart is shown in Figure 4.18. Note, the 'negative' control are the cells transfected with the wildtype vector pB12mcs-eGFP, whereas the 'positive' control are cells transfected with the pB12mcs-U6-eGFP vector.

Exp	Efficiency	`Negative` Control (pg*)	shRNA1 <sup>β3α2</sup> (pg*)	shRNA3 <sup>β3α2</sup> (pg*)	`Positive` Control (pg*)
1	28%	0.0043	0.0030	0.0036	0.0045
2	57%	5.1176	2.4459	3.746	5.0037
3	72%	7.7561	2.5397	5.1694	7.5672

**Table 4.4 The qPCR data of the β3α2 mRNA levels of the transfected K562 cells.**

The RNA extraction was conducted 24hr after each transfection. The generated data from the qPCR reactions of the β3α2 gene is presented as concentrations in pg. The results from three transfections (Exp. 1-3) with their corresponding transfection efficiencies are shown. The data are normalized with regards to the reference gene, *GAPDH*. The `negative` controls are cells transfected with pB12-eGFP and the `positive` controls are cells transfected with pB12mcs-U6-eGFP. Similar values are seen with the `negative` control and `positive` control because qPCR was measuring only the β3α2 fusion point. \* Per pg of *GAPDH*.



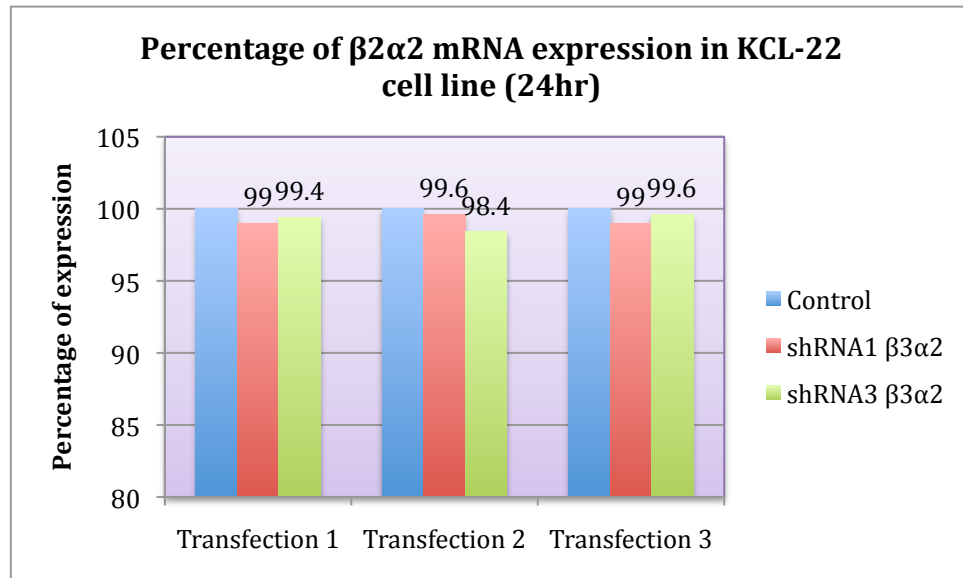
**Figure 4.18** The percentage of expression of  $\beta3\alpha2$  mRNA levels in K562 cell line transfected with shRNA molecules targeting the  $\beta3\alpha2$  fusion point.

The figure shows the results for three transfection experiments. The 'positive' controls are K562 cells transfected with the pB12mcs-U6-eGFP vector and are arbitrarily chosen to be 100%. The percentages of expression of all other samples were calculated with regards to the positive control. The 'negative' controls are the K562 cells treated with the wild type vector (pB12mcs-eGFP). K562 cells transfected with shRNA1 <sup>$\beta3\alpha2$</sup>  and shRNA3 <sup>$\beta3\alpha2$</sup>  showed  $\beta3\alpha2$  mRNA expression levels of 33.6-68% ( $p < 0.034$ ) and 68-81.5% ( $p < 0.023$ ), respectively.

The data show a reduction in the  $\beta3\alpha2$  mRNA levels when using both constructs. However, shRNA1 <sup>$\beta3\alpha2$</sup>  constantly showed a higher level of depletion in comparison with shRNA3 <sup>$\beta3\alpha2$</sup> . The third transfection reaction had the highest transfection efficiency of 72%, and shRNA1 <sup>$\beta3\alpha2$</sup>  decreased the  $\beta3\alpha2$  mRNA levels to 33.6% in comparison with shRNA3 <sup>$\beta3\alpha2$</sup>  that had 68% expression.

To test the specificity of the  $\beta3\alpha2$  shRNA molecules, the KCL-22 cell line was treated with the constructs and the normalized data were converted into percentage of expression with regards to the positive control, the KCL-22 cells transfected with the pB12mcs-U6-eGFP. This is shown in Figure 4.19. The levels

of the  $\beta 2\alpha 2$  mRNA expression were on average 99.2% of the controls, which indicates that the shRNA1 <sup>$\beta 3\alpha 2$</sup>  and shRNA3 <sup>$\beta 3\alpha 2$</sup>  do not target the  $\beta 2\alpha 2$  gene.



**Figure 4.19** Percentage of expression of  $\beta 2\alpha 2$  mRNA levels in KCL-22 cells after treatment with  $\beta 3\alpha 2$  shRNA constructs.

The figure shows the results of three transfection experiments. The controls are the KCL-22 cells treated with the pB12mcs-U6-eGFP vector and are arbitrarily chosen to be 100%. The data obtained from the cells treated with shRNA1 <sup>$\beta 3\alpha 2$</sup>  and shRNA3 <sup>$\beta 3\alpha 2$</sup>  were converted to percentages in regards to the 'positive' control. The expression levels of the  $\beta 2\alpha 2$  mRNA was of 99-99.6% ( $\pm 2$  SD) for shRNA1 <sup>$\beta 3\alpha 2$</sup>  and of 98.4-99.6 ( $\pm 0.5$  SD) for shRNA3 <sup>$\beta 3\alpha 2$</sup> .

#### 4.2.4.3 Effects of shRNA1 <sup>$\beta 2\alpha 2$</sup> in KCL-22 and K562 cell lines

shRNA1 <sup>$\beta 2\alpha 2$</sup>  constructs were transfected into KCL-22 cells in order to study their effects on their target site; KCL-22 cells express the  $\beta 2\alpha 2$  gene. To study the specificity of the shRNA molecule, K562 cells, which express the other BCR-ABL variant ( $\beta 3\alpha 2$ ), were used for transfection. The  $\beta 2\alpha 2$  mRNA levels were measured by qPCR with the use of primers specific to only the  $\beta 2\alpha 2$  gene. Table 4.5 demonstrates the normalized qPCR data of the effects of shRNA1 <sup>$\beta 2\alpha 2$</sup>

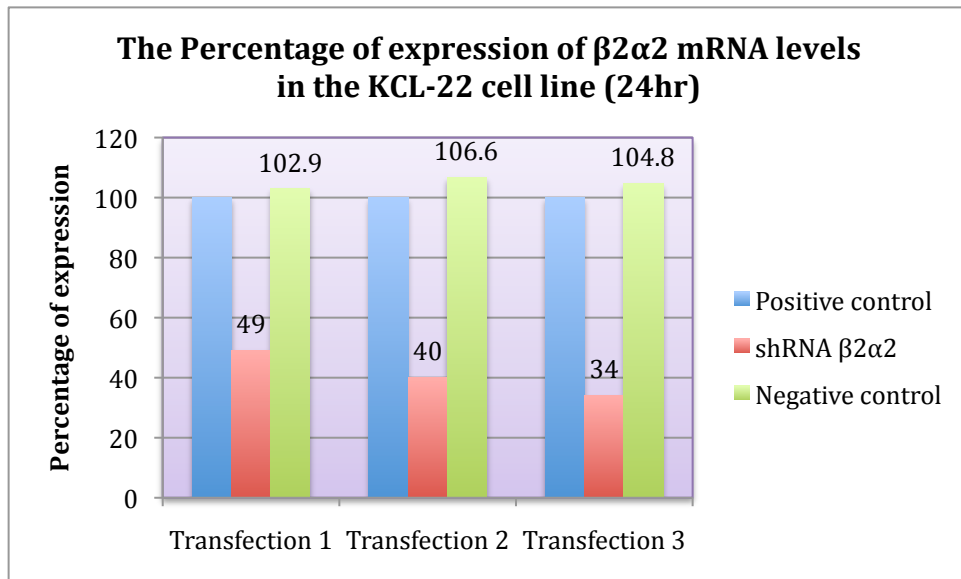


on  $\beta 2\alpha 2$  fusion point in the KCL-22 cell line. These data were calculated as percentages of expression with regards to the 'positive' control. The 'positive' control was transfected with the pB12mcs-U6-eGFP vector and the 'negative' control was transfected with the pB12mcs-eGFP vector. These data were used to generate the bar chart shown in Figure 4.20. The data showed a reduction of the  $\beta 2\alpha 2$  mRNA expression. The expression levels of the  $\beta 2\alpha 2$  mRNA treated with shRNA1 <sup>$\beta 2\alpha 2$</sup>  reduced to a maximum of 66% when the transfection efficiency of nucleofection was 67%.

Exp	Effeciency	'Negative' control (pg*)	shRNA1 <sup><math>\beta 2\alpha 2</math></sup> (pg*)	'Positive' control (pg*)
1	53%	5.072	2.3719	4.8399
2	64%	17.7292	6.6635	16.6261
3	67%	14.6786	4.816	14.2657

**Table 4.5 The qPCR data of the  $\beta 2\alpha 2$  mRNA levels of the transfected KCL-22 cells.**

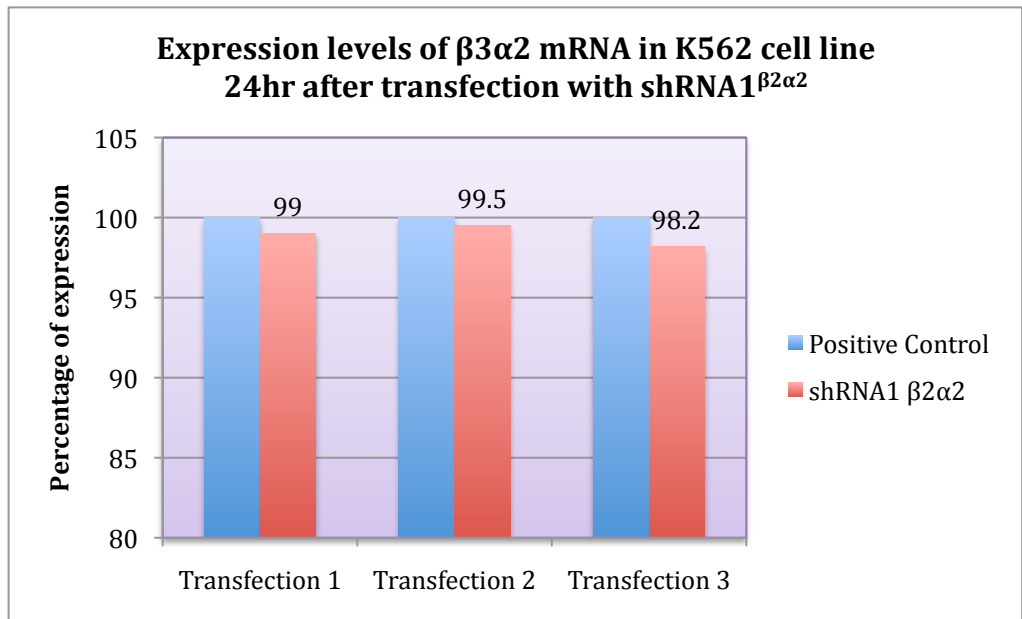
The RNA extraction was conducted 24hr after each transfection. The generated data from the qPCR reactions of the  $\beta 2\alpha 2$  gene is presented as concentrations in pg. The results from three transfections (Exp. 1-3) with their corresponding transfection efficiencies are shown. The data are normalized in regards to the reference gene, *GAPDH*. The 'negative' controls are cells transfected with pB12-eGFP and the 'positive' controls are transfected with pB12mcs-U6-eGFP. Similar values are seen with the 'negative' control and 'positive' control because qPCR was measuring only the  $\beta 2\alpha 2$  fusion point. \* Per pg of *GAPDH*.



**Figure 4.20** The percentage of expression of β2α2 mRNA levels in KCL-22 cell line transfected with shRNA1<sup>β2α2</sup> construct.

The figure shows the results of three transfection experiments on KCL-22 cell line. The 'positive' and 'negative' controls were transfected with pB12mcs-U6-eGFP and pB12mcs-eGFP vectors respectively. The positive control was chosen arbitrarily to be 100%. The normalized data were then calculated into percentages of expression in regards to the positive control. KCL-22 cells treated with shRNA<sup>β2α2</sup> showed expression levels of 34-49%. (p<0.005)

The specificity of the shRNA1<sup>β2α2</sup> construct was also tested in the K562 cell line, Figure 4.21, where the β3α2 mRNA levels of expression were on average 98.9% of the controls. The results demonstrate that the shRNA1<sup>β2α2</sup> is specific for the β2α2 target site.



**Figure 4.21 Percentage of  $\beta3\alpha2$  mRNA levels in K562 cells after treatment with  $\beta2\alpha2$  shRNA constructs.**

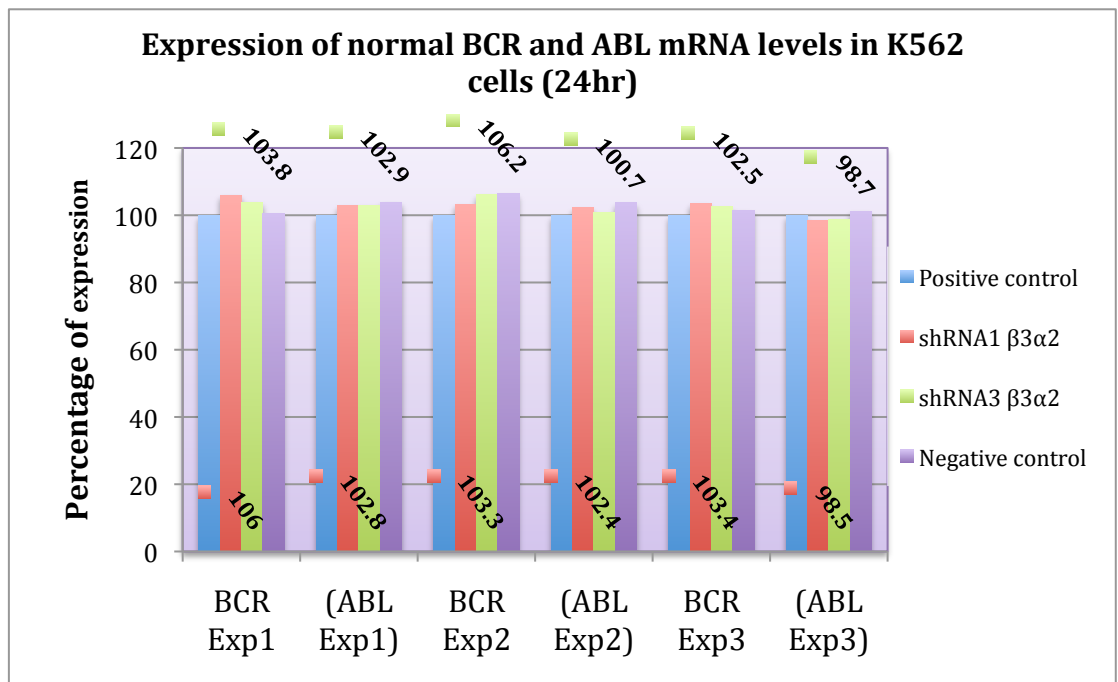
The figure shows the results of three transfection experiments. The control is the K562 cells treated with the pB12mcs-U6-eGFP vector and are arbitrarily chosen to be 100%. The data obtained from the cells treated with shRNA1 $\beta2\alpha2$  were converted to percentages in regards to the 'positive' control. The expression levels of the  $\beta3\alpha2$  mRNA were of 98.2-99.5% ( $\pm 0.66$  SD).

#### **4.2.4.4 Effects of the shRNA constructs on normal BCR and ABL mRNA levels**

To further confirm the specificity of the shRNA constructs to their target sites, the levels of the normal *ABL* and *BCR* genes were examined in K562 cells when transfected with shRNA1 $\beta3\alpha2$  and shRNA3 $\beta3\alpha2$ , and in KCL-22 cells when transfected with shRNA1 $\beta2\alpha2$ . This was done by qPCR with primers that are specific only for the normal *BCR* and *ABL* genes. In each primer set, at least one primer was binding to an exon not present in the *BCR-ABL* fusion genes.

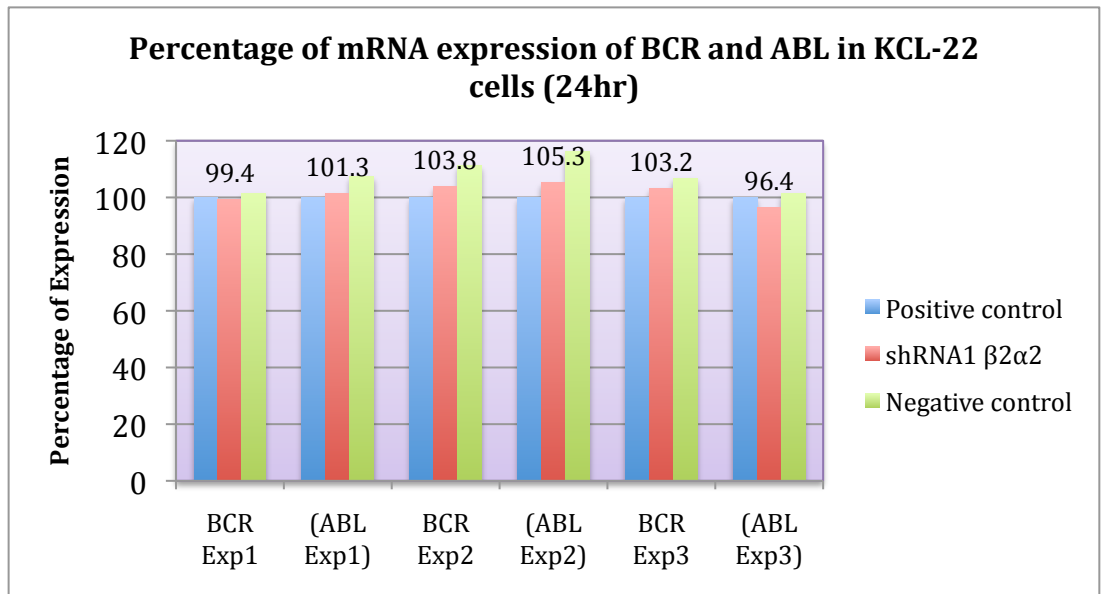
*GAPDH* was used as a reference gene, and the data generated were normalized. The mRNA levels of the normal *ABL* and *BCR* genes were not

affected by any of the shRNA constructs. This is shown in Figures 4.22 and 4.23. Figure 4.22 shows the levels of the *BCR* and *ABL* genes in relation to the positive control in transfected K562 cells. Figure 4.23 shows results for KCL-22 cells. The 'negative' controls are the cells transfected with the wildtype vector pB12mcs-eGFP, whereas the 'positive' controls are cells transfected with the pB12mcs-U6-eGFP vector.



**Figure 4.22** The percentage of expression of the normal *BCR* and *ABL* genes in K562 cell line transfected with shRNA constructs targeting the  $\beta3\alpha2$  fusion gene.

This figure shows the results of three transfection reaction experiments (Exp 1-3). The 'positive' and 'negative' controls are K562 cells transfected with pB12mcs-U6-eGFP and pB12mcs-eGFP vectors, respectively. The 'positive' control was selected arbitrarily to be 100%; the remaining data generated for all other samples were converted to percentages with regards to the positive control. *BCR* and *ABL* mRNA expression levels for shRNA1 <sup>$\beta3\alpha2$</sup>  ranged from 103.3-106% ( $\pm 1.5$  SD), and 98.5-102.8% ( $\pm 1.88$  SD), respectively. For shRNA3 <sup>$\beta3\alpha2$</sup>  *BCR* and *ABL* mRNA expressions ranged from 102.5-106.2% ( $\pm 2.4$  SD), and for 98.7-102.9% ( $\pm 2.1$  SD), respectively. The percentage of expression is shown in the bar chart, red for shRNA1 <sup>$\beta3\alpha2$</sup>  and green for shRNA3 <sup>$\beta3\alpha2$</sup> .

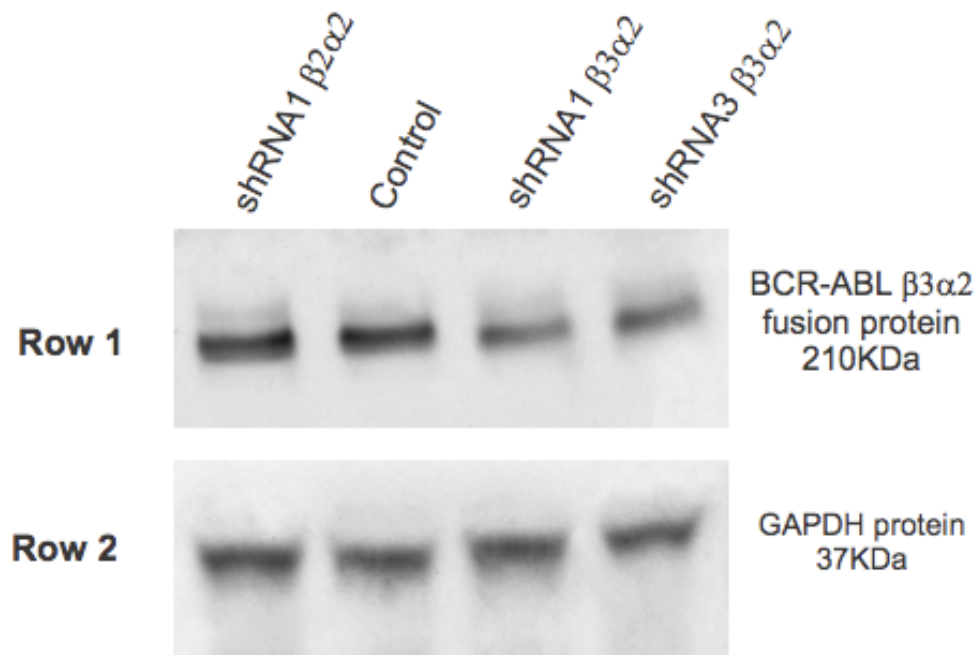


**Figure 4.23** The percentage of expression of the normal *BCR* and *ABL* genes in KCL-22 cell line transfected with the shRNA1 <sup>$\beta 2\alpha 2$</sup>  construct.

This figure shows the results of three transfection reaction experiments (Exp 1-3). The 'positive' and 'negative' controls are KCL-22 cells transfected with pB12mcs-U6-eGFP and pB12mcs-eGFP vectors, respectively. The 'positive' control was selected arbitrarily to be 100%; the remaining data generated for all other samples were converted to percentages with regards to the 'positive' control. *BCR* and *ABL* mRNA expression levels for shRNA1 <sup>$\beta 2\alpha 2$</sup>  ranged from 99.4-103.8% ( $\pm 2.37$  SD), and 96.4-105.3% ( $\pm 4.46$  SD), respectively.

#### 4.2.5 Protein analysis of the K562 transfected Cell Line by Western Blotting

To observe the effects of the constructed shRNA molecules on their target genes at the protein level, Western blots were conducted. Proteins extracted from the K562 cell line transfected with pB12mcs-U6-eGFP and all pB12mcs-U6-shRNA-eGFP vectors, targeting both *BCR-ABL* genes, were examined. Figure 4.24 shows the Western blot analysis conducted on the K562 cell lines.



**Figure 4.24 Western blot analysis of the  $\beta 3\alpha 2$  fusion protein from K562 transfected cell line.**

The figure shows a representative Western blot analysis of the  $\beta 3\alpha 2$  protein from the transfected K562 cell line. The first band in Row 1 is the  $\beta 3\alpha 2$  protein from K562 cells transfected with pB12mcs-U6-shRNA1 <sup>$\beta 2\alpha 2$</sup> -eGFP plasmid. The control band row 1 refers to the  $\beta 3\alpha 2$  protein from cells that were transfected with the 'positive' control plasmid, pB12mcs-U6-eGFP. The last two bands in Row 1 show the  $\beta 3\alpha 2$  proteins from the K562 cell line treated with shRNA1 <sup>$\beta 3\alpha 2$</sup>  and shRNA3 <sup>$\beta 3\alpha 2$</sup>  constructs. Row 2 shows the GAPDH housekeeping gene that was included as a control for loading and was further used to normalise the results. The blot was probed with an anti *c-ABL* antibody (row 1) and anti *GAPDH* antibody (row2).

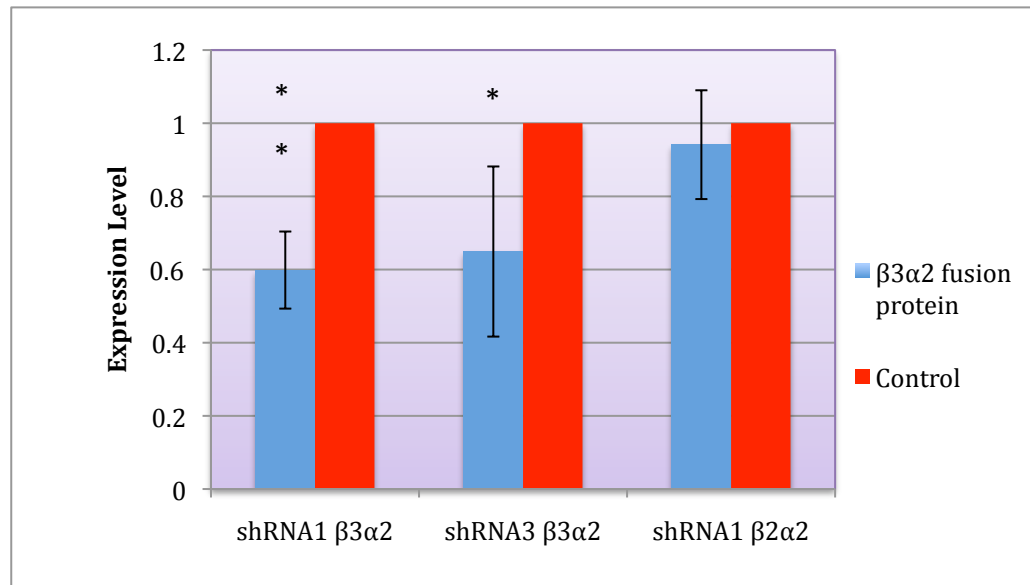
The band intensities for all samples were measured and normalized with regards to the *GAPDH* housekeeping gene. The data generated were converted to percentage of expression with regards to the 'positive' control, which was chosen to be 100%. Upon analysis of the band intensities a reduction in the  $\beta 3\alpha 2$  fusion protein was observed. Each western blot was repeated three times. Table 4.6 shows the statistical analysis done on the results obtained from this Western blot.

Transfected shRNA molecule	Measurement of band intensities		Percentage of reduction	SD ( $\pm$ )	Student t-test P value
	$\beta$ 3 $\alpha$ 2 fusion protein	Control			
shRNA1 $\beta$ 3 $\alpha$ 2	1.90	3.16	40%	0.105	<0.002
shRNA3 $\beta$ 3 $\alpha$ 2	2.10	3.16	33.5%	0.232	<0.014
shRNA1 $\beta$ 2 $\alpha$ 2	3.00	3.16	5.1%	0.148	<0.163

**Table 4.6 Statistically Analysed data from the Western blot conducted on the  $\beta$ 3 $\alpha$ 2 fusion protein extracted from K562 cells that were transfected with shRNA1 <sup>$\beta$ 3 $\alpha$ 2</sup>, shRNA3 <sup>$\beta$ 3 $\alpha$ 2</sup>, shRNA1 <sup>$\beta$ 2 $\alpha$ 2</sup>, and the blot probed with *c-ABL* Ab, and band intensities measured.**

The arithmetic means for the results obtained from three Western blots were first calculated. A reduction in the  $\beta$ 3 $\alpha$ 2 protein treated with shRNA molecules specific for the gene showed a reduction in comparison with the control. The percentage of reduction was then calculated, column 4. The results showed 40% reduction in protein levels when treated with shRNA1 <sup>$\beta$ 3 $\alpha$ 2</sup> and 33.5% with shRNA3 <sup>$\beta$ 3 $\alpha$ 2</sup>. No significant reduction is observed when treated with shRNA1 <sup>$\beta$ 2 $\alpha$ 2</sup>, column 4 row 3. The SD and the student t-test are shown in columns 5 and 6.

Both the shRNA <sup>$\beta$ 3 $\alpha$ 2</sup> molecules reduced the levels of  $\beta$ 3 $\alpha$ 2 protein. It was found that a higher reduction in the protein level was observed with shRNA1 <sup>$\beta$ 3 $\alpha$ 2</sup> construct in comparison with shRNA3 <sup>$\beta$ 3 $\alpha$ 2</sup>. In addition, the shRNA1 <sup>$\beta$ 2 $\alpha$ 2</sup> molecule targeting the  $\beta$ 2 $\alpha$ 2 fusion gene showed no effects on the levels of the  $\beta$ 3 $\alpha$ 2 fusion protein. These results support the previous qPCR findings (previous section) and demonstrate that shRNA1 <sup>$\beta$ 3 $\alpha$ 2</sup> and shRNA3 <sup>$\beta$ 3 $\alpha$ 2</sup> are specific for their target sites. Figure 4.25 demonstrates the generated Western blot data in the form of a bar chart.



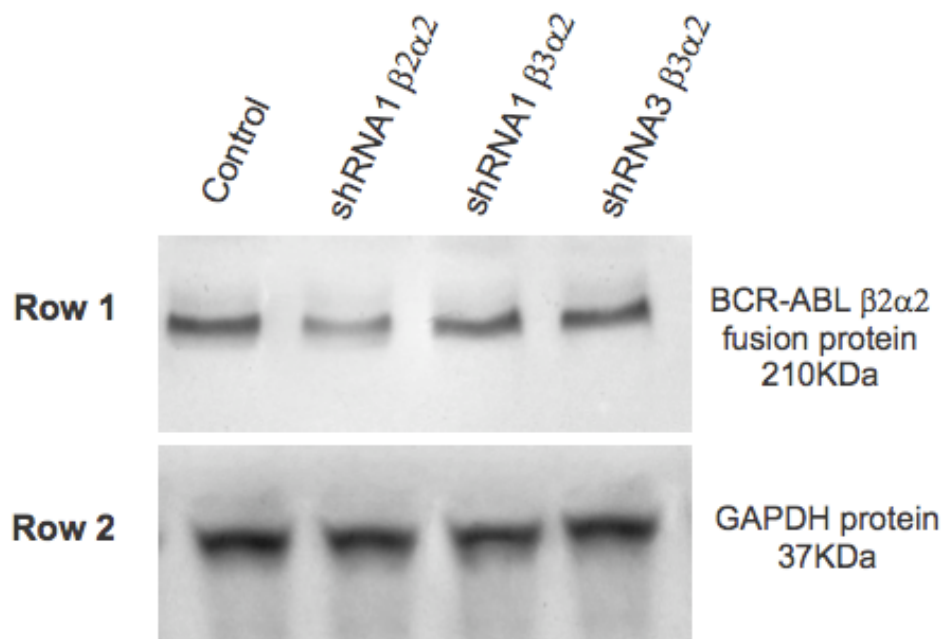
**Figure 4.25 Western blot analysis of the  $\beta3\alpha2$  fusion protein from K562 transfected cell line.**

The figure shows Western blot results from three different transfection experiments. Each bar is generated from the average of three Western blot repeats. The 'positive' control, cells transfected with the pB12mcs-U6-eGFP vector, was selected to be 1 and the data from other transfected cells were calculated with regards to the 'positive' control. The chart demonstrates that shRNA1 <sup>$\beta3\alpha2$</sup>  and shRNA3 <sup>$\beta3\alpha2$</sup>  constructs reduced  $\beta3\alpha2$  protein levels by 40 and by 33.5% respectively. The data also show that the shRNA1 <sup>$\beta2\alpha2$</sup>  molecule has little effect on the  $\beta3\alpha2$  fusion gene product. \*\*p < 0.002 and \*p < 0.01

#### **4.2.6 Protein analysis of the KCL-22 transfected Cell Line by Western Blotting**

To observe the effects of shRNA1 <sup>$\beta2\alpha2$</sup>  on its target gene at the protein level, three Western blots were conducted. Proteins extracted from the KCL-22 cell line transfected with pB12mcs-U6-eGFP and pB12mcs-U6-shRNA1 <sup>$\beta2\alpha2$</sup> -eGFP vectors, were examined. Figure 4.26 is a representation of a Western blot analysis done on the KCL-22 cell lines.





**Figure 4.26 Western blot analysis of the  $\beta 2\alpha 2$  fusion protein from the KCL-22 transfected cell line.**

The figure shows a representative Western blot analysis of the  $\beta 2\alpha 2$  protein from the transfected KCL-22 cell line. The first band in Row 1 refers to the control band that is the  $\beta 2\alpha 2$  protein from cells transfected with the 'positive' control plasmid, pB12mcs-U6-eGFP. The remaining bands in row 1 show the fusion protein from KCL-22 cells transfected with pB12mcs-U6-shRNA-eGFP plasmids. Row 2 shows the GAPDH housekeeping gene that was included as a control for loading and was further used to normalise the results. The blot was probed with anti  $\beta 2\alpha 2$  junction specific antibody (row 1) and anti GAPDH antibody (row 2).

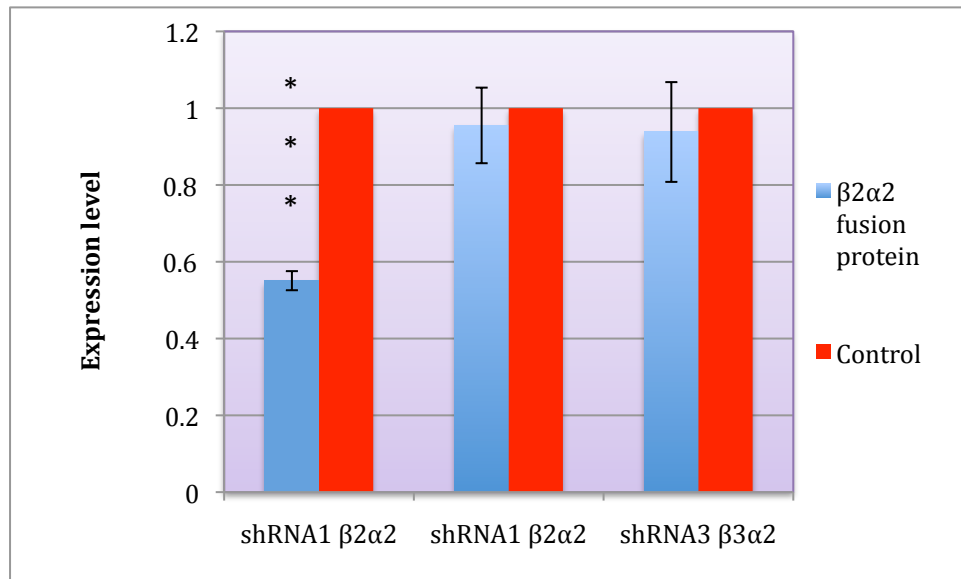
Band intensities for all samples were measured and normalized with regards to the *GAPDH* housekeeping gene. The data generated were converted to percentages of expression with regards to the 'positive' control, which was chosen to be 100%. Table 4.7 shows the statistical analysis done on the results obtained from this Western blot.

Transfected shRNA molecules	Measurement of band intensities		Percentage of Reduction	SD ( $\pm$ )	Student t-test p value
	$\beta$ 2 $\alpha$ 2 fusion protein	Control			
shRNA1 $\beta$ 2 $\alpha$ 2	1.7	3.1	45%	0.0248	0.0001
shRNA1 $\beta$ 3 $\alpha$ 2	3.0	3.1	4.5%	0.0984	0.1335
shRNA3 $\beta$ 3 $\alpha$ 2	2.9	3.1	6.2%	0.1299	0.1240

**Table 4.7 Statistically analysed data from Western blot conducted on the  $\beta$ 2 $\alpha$ 2 fusion protein from KCL-22 cells transfected with shRNA1 <sup>$\beta$ 2 $\alpha$ 2</sup>, shRNA1 <sup>$\beta$ 3 $\alpha$ 2</sup>, shRNA3 <sup>$\beta$ 3 $\alpha$ 2</sup>, and the blot probed with  $\beta$ 2 $\alpha$ 2 Ab, and band intensities measured.**

The arithmetic means for the results obtained from the Western blots were first calculated. The percentage of reduction was then calculated, column 4. The results showed a 45% reduction in protein levels when treated with shRNA1 <sup>$\beta$ 2 $\alpha$ 2</sup>. No significant reduction is observed with cells treated with shRNA1 <sup>$\beta$ 3 $\alpha$ 2</sup> or shRNA3 <sup>$\beta$ 3 $\alpha$ 2</sup>, column 4 rows 3 and 4. The SD and the student t-test are shown in columns 5 and 6.

The data were used to generate the bar chart shown in figure 4.27.



**Figure 4.27** Western blot analysis of the  $\beta 2\alpha 2$  fusion protein from KCL-22 transfected cells.

The bar chart shows Western blot results from three different transfection experiments. Each bar is generated from the average of three Western blot repeats. The 'positive' control, cells transfected with the pB12mcs-U6-eGFP vector, was selected to be 1 and the data from other transfected cells were calculated with regards to the positive control. The chart demonstrates that the  $\text{shRNA1}^{\beta 2\alpha 2}$  construct is specific for its fusion point and has reduced its target's protein levels to 55%. In the case of the  $\text{shRNA1}^{\beta 3\alpha 2}$  and  $\text{shRNA3}^{\beta 3\alpha 2}$  constructs, little effect on the  $\beta 2\alpha 2$  protein levels was seen. \*\*\* $p < 0.0001$ .

The data generated from measuring band intensities of the Western blots, show a reduction in  $\beta 2\alpha 2$  BCR-ABL fusion protein only when cells are treated with  $\text{shRNA1}^{\beta 2\alpha 2}$ , and not with  $\text{shRNA1}^{\beta 3\alpha 2}$  and  $\text{shRNA3}^{\beta 3\alpha 2}$ . Hence, confirming the specificity of the constructs to their target gene.

### 4.3 Discussion

In this chapter the effects of some of the constructed shRNA molecules were tested and studied. This was conducted through *in vitro* studies. Three of the constructed shRNA molecules were transfected by nucleofection into K562 and KCL-22 cell lines. Following confirmation of successful transfection by fluorescence microscopy, FACS analysis was performed to quantify transfection and the transfection efficiencies were calculated. mRNA and protein extractions were conducted and the effects of the shRNA constructs were examined at the mRNA levels using qPCR and protein levels using Western blotting.

FACS results have indicated that the negative control, pB12mcs-eGFP with no promoter, gave GFP fluorescent signals. This background fluorescence maybe due to cell damage by the physical process of nucleofection, or due to intrinsic fluorescence that are present in all cells due to mitochondria, lysosomes, aromatic amino acids, and other endogenous fluorophores such as NADPH (Srivastava *et al.*, 2011 and Mosiman *et al.*, 1997).

The qPCR results showed that the mRNA levels of the targeted genes were reduced. This was true for both cell lines. In the K562 cell line the shRNA1 <sup>$\beta 3\alpha 2$</sup>  construct was more efficient in reducing its target's mRNA levels than the shRNA3 <sup>$\beta 3\alpha 2$</sup>  construct. The first exhibited a 66.4% depletion in the concentration of the  $\beta 3\alpha 2$  mRNA levels, whereas the second only exhibited a 32% depletion. The specificity of both constructs was also tested, and it was found that both construct were specific to their target gene. This was concluded because the shRNA <sup>$\beta 3\alpha 2$</sup>  constructs did not reduce the mRNA levels of the  $\beta 2\alpha 2$  variant when transfected into KCL-22 cell line. In addition, the shRNA <sup>$\beta 3\alpha 2$</sup>  construct did not affect the normal levels of the *BCR* and *ABL* genes.

In the case of the second BCR-ABL variant  $\beta 2\alpha 2$ , the KCL-22 cell line showed a reduction in the mRNA levels of the  $\beta 2\alpha 2$  fusion point, when transfected with the shRNA1 <sup>$\beta 2\alpha 2$</sup>  construct. The construct caused a maximum

of 66% depletion in the mRNA levels. The specificity of this construct was also tested, where the construct was used to transfect the K562 cell line. qPCR results of this transfection demonstrated that there was no significant reduction in the mRNA levels of the  $\beta 3\alpha 2$  fusion point. In addition, the  $\text{shRNA1}^{\beta 2\alpha 2}$  construct did not affect the normal expression levels of the *BCR* and *ABL* genes when transfected into the KCL-22 cells, only reductions in the  $\beta 2\alpha 2$  gene were observed. Hence, confirming the specificity of the synthesised  $\text{shRNA1}^{\beta 2\alpha 2}$  molecule to its target site.

Studies conducted at the protein level were performed by Western blotting. The results obtained from the Western blots with  $\text{shRNA}^{\beta 3\alpha 2}$  and  $\text{shRNA}^{\beta 2\alpha 2}$  constructs confirmed those from the qPCR, where reductions of both BCR-ABL fusion proteins were observed. In the case of the  $\beta 3\alpha 2$  fusion protein,  $\text{shRNA1}^{\beta 3\alpha 2}$  showed a higher reduction in the protein levels in comparison with  $\text{shRNA3}^{\beta 3\alpha 2}$ . Statistical analysis showed that the reduction levels were significant. The Western blots also confirmed the qPCR findings that the synthesised shRNA molecules were specific to their target sequence.

In conclusion, the results shown in this chapter give evidence that the constructed shRNA molecules do indeed specifically target their mRNA sequences causing a reduction in both mRNA and protein levels.

## Chapter 5 : Discussion

### 5.1 Discussion

Chronic myeloid leukaemia (CML) was one of the first cancers to be linked to a chromosomal abnormality (Sawyers, 1999), the Philadelphia chromosome. This chromosome results from a translocation between chromosomes 9 and 22, where the *ABL* gene on chromosome 9, a tyrosine kinase, is translocated to the *BCR* gene region on chromosome 22 giving rise to an abnormal *BCR-ABL* fusion gene (Rowley, 1973). The resultant fusion gene has an abnormally upregulated tyrosine kinase activity (Weisberg and Griffin, 2000) that results in an increase in the proliferation of immature white blood cells, thus leading to the development of CML. There are several breakpoints that can occur in the *BCR* gene two of which give rise to 95% of CML cases (Evans and Sillibourne, 1996). These fusion points are called the  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$  depending on where the chromosomal breakages occur in the *BCR* gene (Van Der Burg *et al.*, 2004).

Like all cancers, the treatment approach in CML is multiple sessions of chemotherapy via the usage of drug cocktails (Hoffbrand *et al.*, 2006). Drugs for the disease have been developed over the years, differing in their actions and target sites. With further understanding of the molecular pathology of CML, new drugs have been designed and developed. These drugs have aimed at inhibiting the *ABL* tyrosine Kinase activity (Milojkovic and Apperley, 2008, Puttini *et al.*, 2008) and, are named tyrosine kinase inhibitors (TKI). However, these drugs are only treatments, not cures, for CML; patients tend to relapse, and the issue of drug toxicity still remains. The only cure available for this disease is allogeneic bone marrow (BM) transplants (Fausel, 2007),

Various drugs exist for CML and the TKI Imatinib is considered to be the first line of treatment for patients with the disorder. Imatinib (Gleevec<sup>®</sup>), a TKI that was developed by Novartis, USA, has shown promising outcomes with

patients establishing a 90% complete hematologic response upon treatment with the drug. However, cases of drug resistance have occurred leading to patient relapse, which led to the development of other tyrosine kinase inhibitors such as Nilotinib (Tasigna®), produced by Novartis, USA (Shieh *et al.*, 2011), and Dasatinib (Sprycel®) (Steinberg, 2007, Talpaz *et al.*, 2006, and Shieh *et al.*, 2011), produced by Bristol Pharmaceuticals, Switzerland. Unfortunately, both drugs display toxic side effects and the issue of drug resistance still remains, favouring the need for development of further drugs and other therapeutic approaches, and making the modulation of gene expression a promising strategy for potential CML treatment.

This project aims to develop an alternative therapeutic approach for CML by the use of shRNA that target and down regulate the expression of the two BCR-ABL variants accounting for 95% of CML cases.

### **5.1.1 Development and Expression of shRNA constructs targeting the BCR-ABL variants**

K562 and KCL-22 cell lines were selected for the purpose of this study. The K562 cell line is a CML cell line with the  $\beta 3\alpha 2$  variant (Lozzio and Lozzio, 1975), whereas the  $\beta 2\alpha 2$  variant is present in the KCL-22 cell line (Kubinishi and Miyoshi, 1983). shRNA molecules were designed based on the sequences of these fusion points, and delivered endogenously along with a reporter gene system. The first attempt to construct a reporter gene system used the pEGFP-N1 vector. This vector has an eGFP and a CMV promoter. The eGFP is a mutated form of GFP that is found to be thermostable at 37°C in mammalian cultured cells. Hence, it has greater and improved levels of fluorescence when compared the wild type protein (Siemering *et al.*, 1996), making its use advantageous in the current study. Also, the CMV promoter was to be removed from the vector and the shRNA molecules placed under the regulation of the human U6 and H1 promoters. This was because these

promoters are members of the RNA Polymerase III (Pol III) promoter family, and transcribe successfully any sequence no more than 400 bases (Tuschl, 2002) without addition of extra bases flanking sequences of interest. Previous studies have shown that chemically and T7-synthesized shRNAs exhibit 3' or 5' overhang extensions, which tend to reduce the efficacy of the synthesized shRNA (Paddison *et al.*, 2002). In addition, transcription with U6 and H1 promoters is easily terminated by a stretch of 6 'T' residues, making them ideal for the expression of siRNAs and shRNAs (Tuschl, 2002). Furthermore, these promoters can transcribe small RNA transcripts that lack the 5' cap and 3' polyadenosine (Wu *et al.*, 2005).

Attempts to remove part of the CMV promoter and re-circularize the double digested pEGFP-N1 vector were undertaken and were successful (previously shown in chapter 3, sections 3.2.2.1 and 3.2.2.2). However, it proved difficult to transform E.coli with the altered pEGFP-N1 plasmid. In every attempt, colony screening showed that E.coli were transformed with only the wild type vector. This may be due to the fact that not all the CMV promoter was removed and the part still intact within the vector was toxic for the growth of E.coli bacterial cells. This explanation was considered due to the fact that the wild type vector always got cloned, rather than the altered one. In cases where attempts to eliminate the presence of the wild type vector in the sample was undertaken, none of the vectors were cloned. In future studies, it would be interesting to see which part of the CMV promoter is toxic to the cells. A previous study done by Trudel *et al.* in 1996 found that a part of the GATA-1 eukaryotic transcription factor was toxic to bacterial cells. This toxic peptide was then used to develop a positive selection vector called pGATA (Clonesure<sup>TM</sup>), where inserts caused a disturbance to the gene encoding the peptide. Something similar may be done after identifying the toxic part of the CMV promoter.

Because of the above difficulties, another plasmid was considered in order to construct the reporter gene system. This was the pB12mcs-eGFP plasmid, which has an eGFP as a reporter gene, and lacks a promoter. Therefore, it was



a suitable choice because the human U6 or the H1 promoter could be inserted into it.

The shRNA expression cassettes, containing both the promoter and shRNA molecules, were synthesised by PCR, and inserted into the pB12mcs-eGFP plasmid. The shRNA molecules designed in this study were designed to be as symmetrical as possible to aid the binding to their target site and cleavage of the correct gene. All constructs were sequenced before testing in expression studies. To conduct expression studies, the pB12mcs-U6-shRNA-eGFP constructs were delivered into K562 and KCL-22, suspension cell lines. Because primary cells and cells in suspension are hard to transfect (Ovcharenko *et al.*, 2005, Oliveira and Goodell, 2003 and Merkerova *et al.*, 2007), several chemical transfection agents were tested, none was found to give good transfection efficiency (45-50% for both cell lines). Consequently, other methods of transfection were used. It was found that nucleofection, a form of electroporation, where the DNA molecules of interest are directly delivered into the target cell's nucleus gave acceptable transfection efficiencies, a maximum of 72% for K562 and a maximum of 67% for KCL-22. In addition nucleofection gave a good percentage of cell viability, an average of 89.3% for K562 and 90.8% for KCL-22. The 'negative' control (pB12mcs-eGFP with no promoter) showed auto-fluorescence; therefore, the transfection efficiencies with regards to the 'negative' control, as well as the non-transfected cells were calculated. However, the transfection efficiencies shown in the current study are those with regards to the non-transfected cell lines. The auto-fluorescence observed on the 'negative' control maybe due to cell damage or due to intrinsic levels of fluorescence due to mitochondria, lysosomes, aromatic amino acids, and other endogenous fluorophores such as NADPH (Srivastava *et al.*, 2011 and Mosiman *et al.*, 1997). Cell damage might have occurred due to the physical nature of nucleofection. Thus, suggesting that other methods of transfection avoiding cell damage and at the same time giving high transfection efficiencies are required for CML cell lines. In future developments, attempts to quench auto-fluorescence through staining with

trypan blue before FACS analysis can be applied (Srivastava *et al.*, 2011 and Mosiman *et al.*, 1997).

Merkerova *et al.*, 2007, examined two physical transfection methods including nucleofection on CML primary cells and three chemical methods to transfect leukemic cell lines and primary CML cells. Their study reported that two chemical methods gave 35-40% efficiency while the third one gave no satisfactory transfection in leukaemic cell lines; the efficiency for the last was not reported nor the type of the cell lines used in the study. The efficiency of nucleofection using the program T-16 on primary CML cells was 25-29%. Another study done by Mahmoodabady *et al.*, 2010, also compared nucleofection to other chemical methods in K562 cells. The study found that nucleofection gave higher transfection efficiencies, 50%, when compared to lipofectamine and CaCl<sub>2</sub> methods, <50% and <20% respectively; cell viability was not mentioned. Nevertheless, both studies give evidence supporting the results in the current study on the difficulty of transfecting leukaemic cell lines and that nucleofection is a more efficient method of transfection in regards to CML cell lines.

The reason behind the K562 and KCL-22 cell lines being difficult to transfect maybe due to the fact that these cells are morphologically spherical cells (Bártová *et al.*, 2005) that float within the media, in contrast to adherent cells that become immobilized and adhere to the bottom of the culture flask. Adherent cells become flatted and spread, gaining more surface area in comparison to the spherical K562 and KCL-22 cell lines. This would result in easier, more abundant take up of the chemically manipulated DNA complexes used in transfection. No program, or a transfection kit, was available for the KCL-22 cell line, therefore the same program and Kit used for the K562 cell line were tested and were found to give satisfactory results, average of 61.3% for transfection efficiency and 90.8% for viability.

### 5.1.2 Efficiency of shRNA constructs targeting the $\beta 3\alpha 2$ fusion point

Two shRNA constructs, shRNA1 <sup>$\beta 3\alpha 2$</sup>  and shRNA3 <sup>$\beta 3\alpha 2$</sup> , under the regulation of the U6 promoter were tested against the  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$  variants. The  $\beta 3\alpha 2$  and  $\beta 2\alpha 2$  BCR-ABL mRNA expression levels were examined 24 hr after transfection, and were quantified by the use of real time PCR. The results of this reaction demonstrated a decrease in the levels of only  $\beta 3\alpha 2$ . The results show that the mRNA expression levels of  $\beta 3\alpha 2$  fusion product were depleted by 66.4% with shRNA1 <sup>$\beta 3\alpha 2$</sup>  and 38% with shRNA3 <sup>$\beta 3\alpha 2$</sup> . Further studies showed that the constructs were specific for the  $\beta 3\alpha 2$  fusion point and that they did not target the  $\beta 2\alpha 2$  fusion point. The different efficiencies of shRNA1 <sup>$\beta 3\alpha 2$</sup>  and shRNA3 <sup>$\beta 3\alpha 2$</sup>  could be due to the fact that the U6 promoter expressing the shRNA3 <sup>$\beta 3\alpha 2$</sup>  molecule had 3 mismatches compared to 2 mismatches in the U6 promoter regulating the expression of the shRNA1 <sup>$\beta 3\alpha 2$</sup>  molecule. Mismatches occurred in the U6 promoter even though a high fidelity *Taq* polymerase was used (Fast Start High Fidelity PCR system, Roche, UK). For future studies, it would be interesting to examine what effect the mismatches in U6 may have on the effectiveness of the promoter. Another reason could be due to the difference in sequences of the shRNA molecules, or how the molecules bind to their targets in the cells in regards to the target's secondary structure; the shRNA1 <sup>$\beta 3\alpha 2$</sup>  sequence may give the molecule a better efficiency to induce RNAi based silencing than shRNA3 <sup>$\beta 3\alpha 2$</sup> .

Western blot results supported the real time PCR results and demonstrated a reduction in the protein levels and confirmed the specificity of the shRNAs. This reduction was 40% and 33.5% for shRNA1 <sup>$\beta 3\alpha 2$</sup>  and shRNA3 <sup>$\beta 3\alpha 2$</sup> , respectively, in relation to the positive control. These results demonstrate that the shRNA molecules tested had an effect on both mRNA and protein levels of the *BCR-ABL* fusion gene.

### 5.1.2.1 Delivery and production of the shRNA <sup>$\beta 3\alpha 2$</sup> Constructs

Mahmodabady *et al.*, 2010, used shRNA targeting the  $\beta 3\alpha 2$  variant in K562 cell lines. Their publication reported that they have established 50% transfection efficiency with the use of electroporation, however, Lipofectamine<sup>TM</sup> 2000 that gave lower transfection efficacy was used to carry out their transfection experiments (the actual percentage of transfection efficiency was not reported but only a bar chart represented the results). Arthanari *et al.*, 2010, also transfected K562 cells with shRNA constructs; transfection was done through chemical methods that involved lipofectamine showing transfection efficiencies between 20-50%. In the current study, K562 cells were transfected by nucleofection and a maximum transfection efficiency of 72% was established when using non-viral vectors for shRNA delivery. Other studies have achieved higher transfection efficiencies for shRNA delivery including those using a lentiviral system. Examples include studies done by Myssina *et al.*, 2008 and Li *et al.*, 2003. In general lentiviral vectors, have higher transfection efficiencies than other non-viral systems, and the advantage of long stable expression. However, these vectors are highly toxic (Wang *et al.*, 2008).

Different promoters can be used to produce shRNA molecules, however, as discussed earlier, the small nuclear RNA Pol III promoters have been used to develop transcripts to induce RNAi, especially the U6 and H1 promoters (Wu *et al.*, 2005). Mahmodabady *et al.*, 2010, used the H1 promoter to express shRNA molecules used in their study. Myssina *et al.*, 2009, used a CMV promoter to produce their shRNA molecules. The CMV promoter is a Pol II promoter; these promoters have also been shown to produce shRNA molecules (Rossi, 2008). A study done by Li and Mahato, 2009, to see the effects of H1, U6, and CMV promoters on shRNA mediated gene silencing, found that when the same shRNA molecule was produced from the three different promoters, there was no significant difference in gene silencing between the U6 and CMV promoters. However, H1 had less effect than CMV and U6. Other studies have

also reported that the U6 promoter is more potent than H1 (An *et al.*, 2006 and Snyder *et al.*, 2009). In the present study, the U6 promoter was used to produce shRNA molecules.

#### **5.1.2.2 Effects of the shRNA <sup>$\beta 3\alpha 2$</sup> Constructs on $\beta 3\alpha 2$ mRNA and protein levels**

Mahmodabady *et al.*, 2010, showed significant reduction in the BCR-ABL mRNA levels, up to 42%, 24hr after shRNA transfection. However, to measure the mRNA levels, RT-PCR was performed and PCR was used to amplify the RT-PCR products. A more accurate way of measuring  $\beta 3\alpha 2$  mRNA levels is through qPCR as attempted in this study. Myssina *et al.*, 2009, showed that their designed shRNA against the  $\beta 3\alpha 2$  variant had an effect on mRNA and protein levels, where a 75% and 95% reduction, respectively were reported. However, this was measured 48hr after transfection on a 100% transfected cell population (FACS was used to select GFP+ cells) that was transfected with high concentrations of shRNA ( $5\mu\text{g}/10^6$  cells), which was expressed from a lentiviral system. These can be the reasons for achieving the high levels of reduction. This was also the case in the study conducted by Li *et al.*, in 2003 that used a lentiviral vector to express shRNAs, under the regulation of the U6 promoter. The latter, as discussed previously, also yields high levels of gene expression, hence the observation of >90% depletion in both mRNA and protein levels. Nevertheless, in the current study a non-viral system has demonstrated successful depletion of both mRNA (66.4%) and protein levels (40%) when under regulation of the U6 promoter, and a shRNA concentration of  $2\mu\text{g}/2.5 \times 10^6$  cells. The effects of the shRNA constructs used in the current work might have greater effects if a lentiviral system were used.

Arthanari *et al.*, 2010, have shown that an shRNA expressed in the K562 cell line leads to a decrease in  $\beta 3\alpha 2$  mRNA levels of 70% and protein of 75%. mRNA levels were measured after 48hr and protein levels after 96hr, when cells were

treated with 10µg of shRNA. A further decrease of mRNA to 85% and protein to 90% was observed at 30µg of shRNA. The depletion levels were measured after 48 and 96hr in comparison to only 24hr in the present work. These high levels of depletion might be due to this reason. Another explanation for achieving such high depletion levels could be due to the high shRNA vector concentrations used by Arthanari *et al.*, 2010. As they have used 10-30µg of shRNA vector per  $1.5 \times 10^6$  cells, which is 5-15x more the concentration of shRNA, used in the present study (2µg/ $2.5 \times 10^6$ ). Greater levels of depletion might have been accomplished in the current project, if higher shRNA vector concentrations had been used.

In the current work, the shRNA<sup>β3α2</sup> constructs demonstrated high specificity to their target site because these molecules induced RNAi mediated gene silencing only to their target sites and not to other genes such as *BCR-ABL* fusion point (β2α2), the normal *ABL*, and *BCR* genes, in which 48-50% similarity in sequence exist. Other studies testing shRNA molecules against the β3α2 either considered their molecules specific on the basis that the levels of *GAPDH* housekeeping were not affected, or by having a scramble sequence of the same shRNA used that did not target any human gene (Mahmodabady *et al.*, 2010). Myssina *et al.*, 2009 and Li *et al.*, 2003, demonstrated specificity of their shRNA molecules in regards to the β2α2 variant and did not consider the normal *ABL* and *BCR* genes.

### 5.1.2.3 Conclusion

In conclusion, the shRNA molecules targeting the β3α2 fusion gene showed specific reduction of their target site mRNA (66.4%) and protein levels (40%), which were achieved 24hr after transfection with the shRNA constructs and with a relatively low concentration of shRNA (2µg/ $2.5 \times 10^6$  cells). The *BCR-ABL* protein has been reported to have a long half-life (Myssina *et al.*, 2008 and Wilda *et al.*, 2002) that is estimated to be >48hr (Spiller *et al.*, 1998).

Therefore, higher levels of reduction at protein levels can be expected for the shRNAs used in the current study over a longer period of time.

### **5.1.3 Efficiency of shRNA constructs targeting the $\beta 2\alpha 2$ fusion point**

The  $\beta 2\alpha 2$  BCR-ABL mRNA expression levels were examined 24 hours after transfection, and were quantified by the use of real time PCR. The results demonstrated a decrease in the level of the  $\beta 2\alpha 2$  BCR-ABL variant. One shRNA constructs, shRNA1 <sup>$\beta 2\alpha 2$</sup> , under the regulation of the U6 promoter was tested. The results show that the mRNA expression levels of the  $\beta 2\alpha 2$  fusion product, and not  $\beta 3\alpha 2$  were depleted by 66% with shRNA1 <sup>$\beta 3\alpha 2$</sup> . Western blot results demonstrated a reduction in the protein levels of 45%, in relation to the 'positive' control. These results show that the shRNA molecules tested had an effect on both mRNA and protein levels of the *BCR-ABL* fusion gene.

#### **5.1.3.1 Delivery and production of the shRNA <sup>$\beta 2\alpha 2$</sup> Construct**

The current study demonstrated that KCL-22 cell lines can be transfected by means of nucleofection, and the shRNA1 <sup>$\beta 2\alpha 2$</sup>  construct achieved a relatively high transfection efficiency of 67% by this method. shRNA1 <sup>$\beta 2\alpha 2$</sup>  was cloned in a non-viral vector, pB12mcs-eGFP plasmid, and was under the regulation of the snRNA Pol III U6 promoter. Li *et al.*, 2003, constructed a  $\beta 2\alpha 2$  shRNA molecule under the regulation of the U6 promoter, however, the construct was inserted into a lentiviral system, and the transduction efficiency reached ~100% in Meg01 leukaemic cell line. Meg01 is a CML cell line that expresses the  $\beta 2\alpha 2$  variant, and was developed by Ogura *et al.*, in 1985. This is a differentiated cell line where mostly large immature megakaryocytes predominate, unlike the

KCL-22 cell line, used in the current study, that is an undifferentiated haematopoietic cell line (Kamamoto *et al*, 1986). In summary, Li *et al.*, 2003, used a different cell line and a viral system to express their shRNA constructs that can both explain the high transduction efficiency achieved in their study, in comparison to the non-viral vectors used in the current study.

### **5.1.3.2 Effects of the shRNA <sup>$\beta 2\alpha 2$</sup> Constructs on $\beta 2\alpha 2$ mRNA and protein levels**

In the present study a reduction in the  $\beta 2\alpha 2$  mRNA levels of 66% and protein levels of 45% was observed 24hr after transfection with shRNA1 <sup>$\beta 2\alpha 2$</sup> . In contrast Li *et al.*, 2003, showed no significant reduction in the  $\beta 2\alpha 2$  variant. This maybe the result of differences in the shRNA structure and sequence used in their study. The current study used an shRNA molecule of 21 nt long and an open stem loop of 8 nt. Li *et al.*, used the same shRNA sequence but with a 9 nt loop and a G overhang at the 5' end on the antisense strand. It has been reported that the loop size, structure, and its sequence, affects the knockdown efficiencies of shRNA molecules (Brummelkamp *et al.*, 2002 and Schopman *et al.*, 2010). Schopman and colleges, 2010, have found that improvements on the 9 nt long stem loop used in Brummelkamp *et al.*, 2002, studies are achievable by using different structures of the stem loop. Schopman *et al.*, 2010, demonstrated that even weak shRNA inhibitors could be turned into strong gene silencers by simply differing the loop length, sequence, or structure. Consequently, these could be the reasons why the shRNA molecules used in Li *et al.*, 2003, study demonstrated lack of  $\beta 2\alpha 2$  gene silencing in comparison to current work.

In the current study, qPCR results also demonstrate that the shRNA1 <sup>$\beta 2\alpha 2$</sup>  construct was specific to its target sequence and no significant reductions in the mRNA expression levels were observed in related and non-related genes, including normal *ABL*, normal *BCR*, *GAPDH* (when transfected into KCL-22 cells)



and the  $\beta 3\alpha 2$  fusion gene. The protein analysis confirmed that the shRNA1 <sup>$\beta 2\alpha 2$</sup>  construct was specific to its target site, supporting the qPCR results.

### 5.1.3.3 Conclusion

In conclusion, the current study demonstrates successful and specific targeting of the  $\beta 2\alpha 2$  variant by a non-viral system expressing shRNA molecules under the regulation of the U6 promoter. The effects of the shRNA used in the present work might have greater effects if a lentiviral system was used. Lentiviral systems have higher transfection efficiencies than non-viral systems; this is mainly because lentiviruses have the ability to integrate into the host's genome, establishing long-term stable expression of the integrated DNA sequence. However, drawbacks exist for these types of system, which include toxicity and oncogenesis (Hacein-Bey-Abina *et al.*, 2008 and Davidson and McCary, 2011).

### 5.1.4 siRNA against the BCR-ABL Variants

Other studies have targeted the *BCR-ABL* fusion genes through siRNA technology have for example Scherr *et al.*, 2003, Koldehoff *et al.*, 2010, Wohlbold *et al.*, 2003, Bártová *et al.*, 2005, and Withey *et al.*, 2005 using siRNA targeting the  $\beta 3\alpha 2$  variant, and Rapozzi and Xodo, 2004 for targeting the  $\beta 2\alpha 2$  variant. Findings from Scherr *et al.*, 2003, demonstrated that the siRNA tested had reduced the mRNA levels up to 87% after 24hr. However, this efficiency was obtained from a co-transfection of both a BCR-ABL expressing vector, which was fluorescently labelled, and chemically synthesised siRNA molecules in HeLa cells. It was further reported, in the same publication, that when the same siRNA was transfected into K562 cell lines, the mRNA levels of the *BCR-ABL* fusion gene were reduced up to 24.8% after 24hr. Withey *et al.*, 2005, reported a 84% reduction of mRNA levels in K562 cell line. However, these

results were obtained from PCR amplification of the cDNA of the transfected cell lines. A more precise method to measure the mRNA levels would be to use qPCR, which was used in the current study. The shRNA molecules used in the current study have higher potency, sustainable effects using low copy number, and less off-target effects in comparison with siRNAs (Rao *et al.*, 2009).

Studies have also been conducted where antisense technology and drug treatment have been combined to try and reduce CML effect. For example Koldehoff *et al.*, 2010 and Wohlbold *et al.*, 2003 showed that siRNA targeting the BCR-ABL fusion point had additive effects with Nilotinib and Imatinib. In addition, the study done by Wohlbold *et al.*, 2003, suggested that targeting the mRNA by the process of RNAi sensitizes the drug resistant leukaemic cells to TKI. The findings of these studies suggest future strategies combining both molecular targeting of the *BCR-ABL* fusion gene at the mRNA level by the use of RNAi and using TKI on the protein level.

Bártová *et al.*, 2005, demonstrated the use of siRNA against the *BCR-ABL* fusion gene. However, the siRNA designed in this study targeted the *BCR-ABL* downstream of the fusion point, thus also targeting the normal *c-ABL*. In addition, the transfection levels of the siRNA only ranged between 25-30% which is considered to be quite low.

A study done by Rapozzi and Xodo, in 2004, showed that significant reductions in mRNA and protein levels of the  $\beta 2\alpha 2$  BCR-ABL variant could be achieved with single stranded (40% for mRNA and 28% for protein) and double stranded siRNA (20-40% for mRNA and 83% for protein) within 24hr. Analysis was carried out 24 and 48hr after two electroporation treatments. Thus, high doses of siRNA were required to achieve these results. In addition, the effects of the siRNA molecules only lasted for 72hr after treatment, indicating the transient effect of siRNA molecules.

In conclusion, RNAi technology is a powerful tool for down regulation of genes of interest. In the current project, this technology was used to down regulate the expression of the abnormal *BCR-ABL* gene, more precisely the up-regulated tyrosine kinase activity of the *ABL* part of the fusion gene. The constructed shRNA molecules, targeting both BCR-ABL variants, show

specificity to their target sites and caused depletion in both mRNA and protein levels. The U6 promoter showed ease of production and construction of shRNA expression cassettes. The promoter also displayed efficient synthesis of shRNA molecules. In this context, shRNA constructs demonstrated in the current work can be promising tools for CML therapy when used alone or coupled with TKI's.

### **5.1.5 Future developments**

Future developments to be considered in this project include the testing of the other constructs that have been synthesised and sequenced; transfection for longer periods of time, such as 48 and 72 hours; and a comparison the effects of different concentration of the shRNA molecules on their target sites. Once an appropriate time period and concentration has been achieved, it would also be appealing to compare the effects of the same shRNA molecules but under the regulation of the two different human promoters, the U6 and H1, as only the U6 promoter was tested in the current project. The effects of the shRNA on growth inhibition, cell proliferation, and differentiation could be also investigated.

It would also be intriguing to examine the tested shRNA constructs in an animal model. To do so, first, a suitable delivery system needs to be chosen. These systems can be either viral or non-viral systems, with each option having its own pro and cons. Non-viral vectors avoid issues such as immunogenicity and insertional mutagenesis; however, these vectors suffer from low transduction efficiency, and non-permanent expression. Viral vectors, on the other hand, overcome the disadvantages of non-viral vectors, in terms of efficiency and provide stable expression, but, give rise to issues relating to toxicity, safety, immunogenicity, and specificity of delivery. Other vectors to consider are the new hybrid vectors, engineered to have qualities of more than one vector (Wang *et al.*, 2008). Rumpold *et al.*, 2005, constructed a non-viral transposon-based vector that demonstrated stable expression of siRNAs, that

caused knockdown to their target genes. However, the system was described to have a laborious production, systematic instability and had a pathogenicity issue. Therefore, great care should be taken when deciding what type of vector system is chosen to express the shRNA molecules *in vivo*.

Secondly, once the appropriate vector system has been chosen, another concern is the method and specificity of delivery to the target cells. The target cells in CML will not only be the cancerous cells but also the abnormal hematopoietic stem cells that are hard to identify. Recently, researchers at Lund University have shown that CML stem cells might be isolated through a cell surface marker called IL-1 receptor accessory protein (Luo, 2010). This can be advantageous, as it can help the identification of stem cells with the Philadelphia chromosome. These stem cells can be therefore isolated and used for the conduction of *ex vivo* studies, providing a way to achieve specific targeting of the abnormal cells of interest. Another aspect to consider is cell-penetrating peptides. This technology has been successfully considered as an application for *in vivo* delivery of various molecules including shRNAs (Wang *et al.*, 2008). A study done by Arthanari *et al.*, 2010, showed that by combining the cell penetrating peptide HIV-Tat with the membrane lytic peptide LK-15, higher transfections efficiencies could be achieved for delivery of shRNA targeting the *BCR-ABL* fusion gene in K562 cell lines. These findings make the technology of cell-penetrating peptides an appealing area to explore when considering *in vivo* studies.

The next stage would be to choose the appropriate animal model. This is of great importance. It would be highly interesting to see how the shRNA molecules react *in vivo* and how the animal reacts to such treatment. The Severe Combined Immunodeficiency (SCID) mouse or murine CML-like models could be considered. The SCID mouse results from a mutation that impairs the cell development in the lymphoid lineage making the animal severely immunodeficient. This leads to the animal accepting human exogenic engraftments without initiating an immune response. A study done by Wang *et al.*, in 1998 showed the establishment of high levels of engraftments by SCID mice with

haematopoietic cells from patients with CML. Thus, making this model highly considerable for future studies.

Other murine CML-like models have also been developed to study CML through *in vivo* studies. These include mice where both transgenic and retroviral transduction approaches have been employed, the latter seems to be more promising in regards to establishing a realistic murine CML-like model (Van Etten, 2002). A study done by Wertheim *et al.*, in 2002 has shown a successful attempt to generate a CML-like murine model. Scientists in this study developed a protocol to transfuse high titre retroviral vectors into murine bone marrow causing the haematopoietic stem cells to have an increased expression of the BCR-ABL protein. Furthermore, lethally irradiated mice were transplanted with the transfused bone marrow cells leading to the development of fatal myeloproliferative disorders in 100% of the mice. Therefore, this model can also be considered. However, more research into this area should be conducted in order to select the best CML model before clinical trials are attempted.

## Appendix 1: The Molecular Weight Markers Used in The Current Project

Five types of molecular weight markers were used to determine band sizes. The  $\lambda$ /EcoRI/HindIII marker, and the GeneRuler™ 1Kb marker were used for determination of large band sizes and on a 1% agarose gel. The pBluescript Hpa II marker, the 100bp DNA Ladder, and the GeneRuler™ 100bp marker were used for smaller size bands ranging from 1000bp to as small as 100bp in size. These markers were usually used on 2% agarose gels.

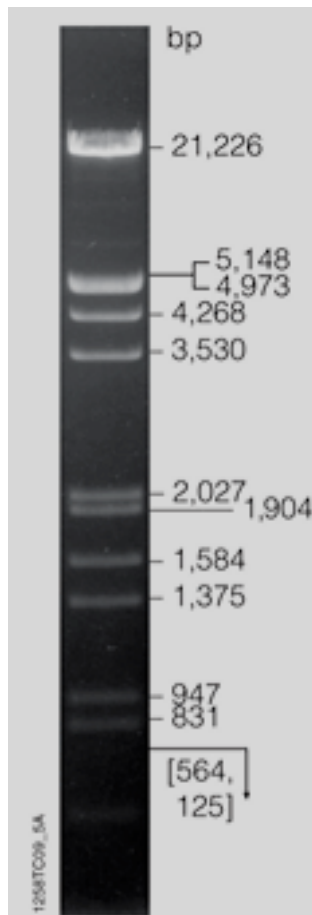


Figure 67 The  $\lambda$ /EcoRI/HindIII Marker. Taken from the Promega website

## GeneRuler™ 1 kb DNA Ladder

O'GeneRuler™ 1 kb DNA Ladder,  
ready-to-use

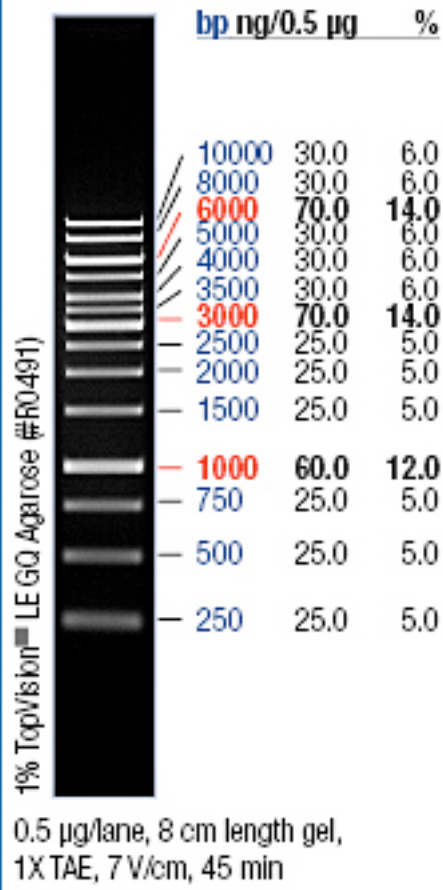


Figure 68 The GeneRuler™ 1Kb marker. Taken from the Thermo Scientific website

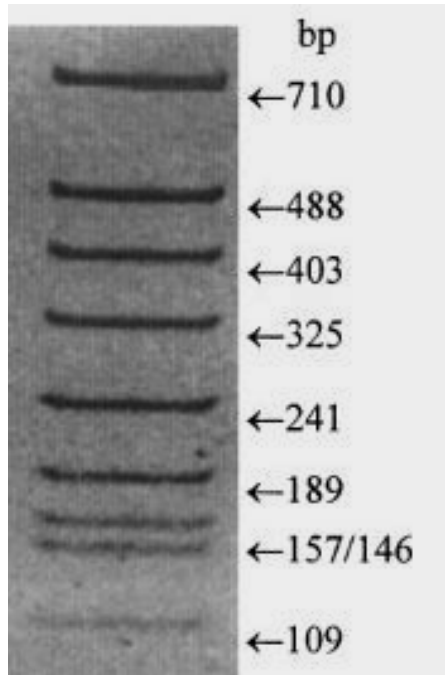


Figure 69 The pBluescript HpaII Marker.

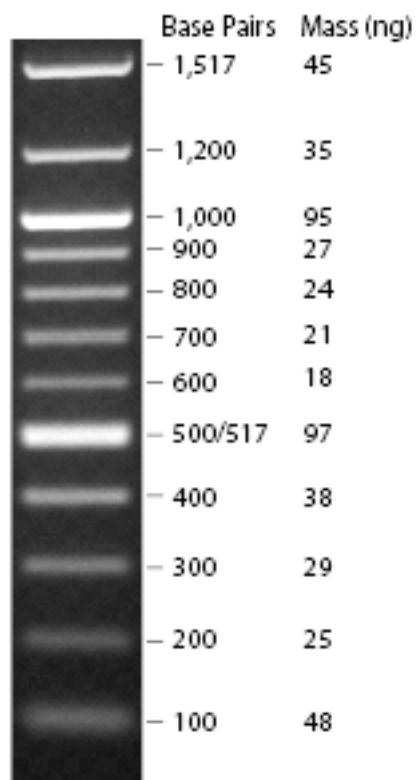


Figure 70 The 100bp DNA Ladder. Taken from the NEB website



## GeneRuler™ 100 bp DNA Ladder

O'GeneRuler™ 100 bp DNA Ladder,  
ready-to-use

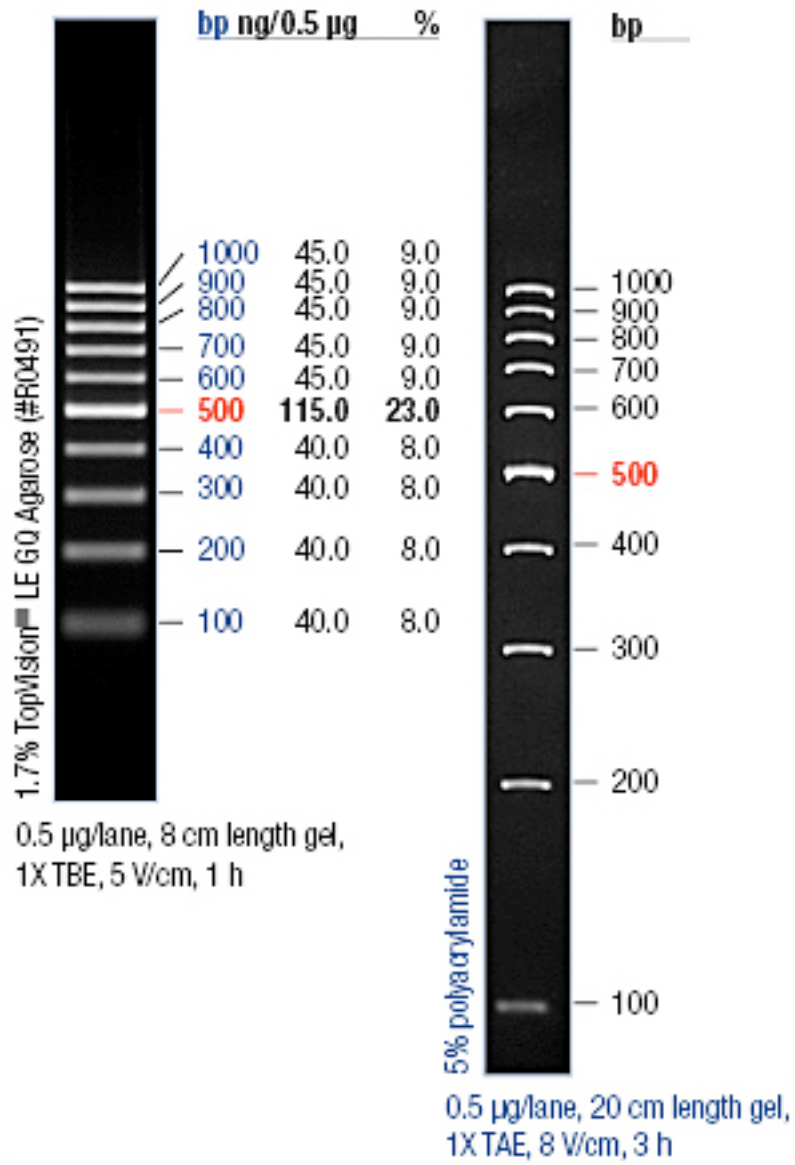


Figure 71 The GeneRuler™ 100bp Marker. Taken from the Thermo Scientific website

## Appendix 2: The Secondary shRNA Structures

The following figures are the secondary structures predictions of the designed shRNA molecules. The program “mfold” was used to predict these secondary structures. The program is found in the following website:

(<http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form>)

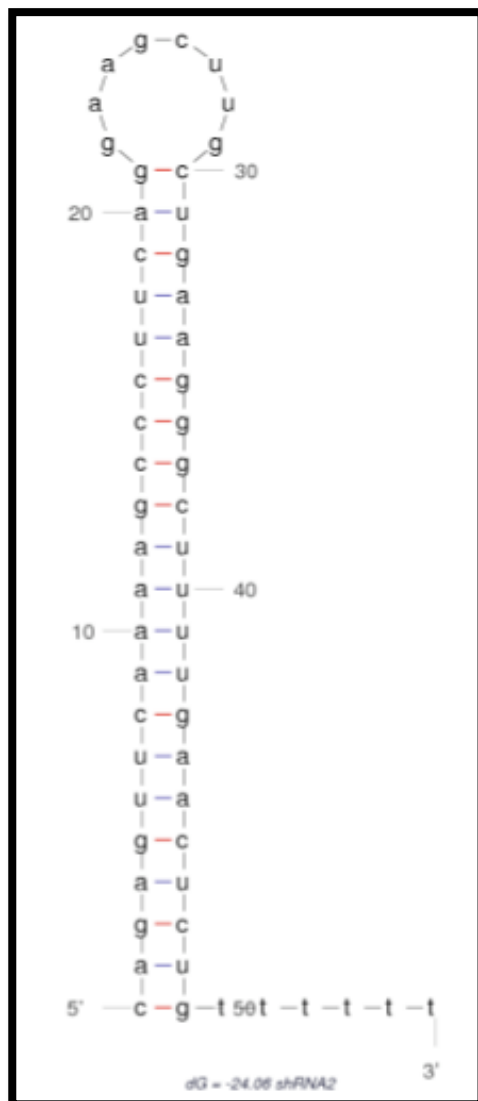


Figure 72 The Secondary structure prediction of shRNA2<sup>β3α2</sup> by “mfold”

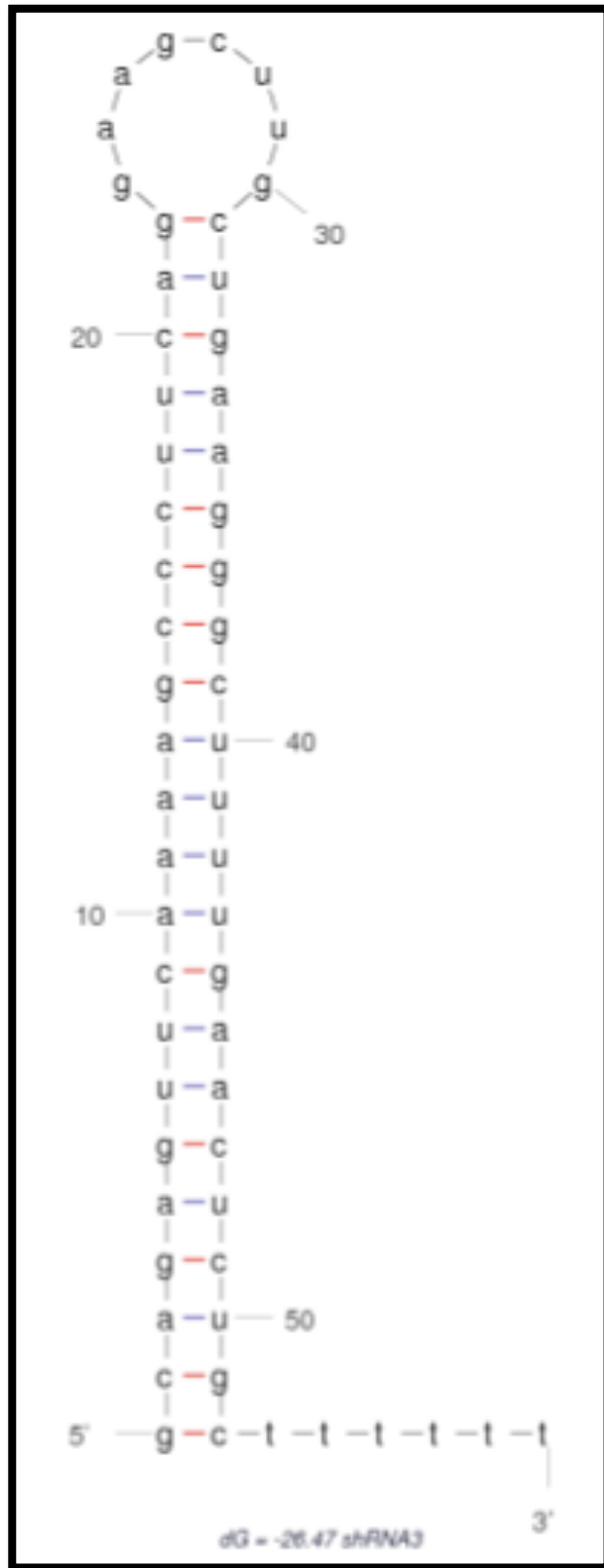


Figure 73 The Secondary structure prediction of shRNA3<sup>β3α2</sup> by “mfold”

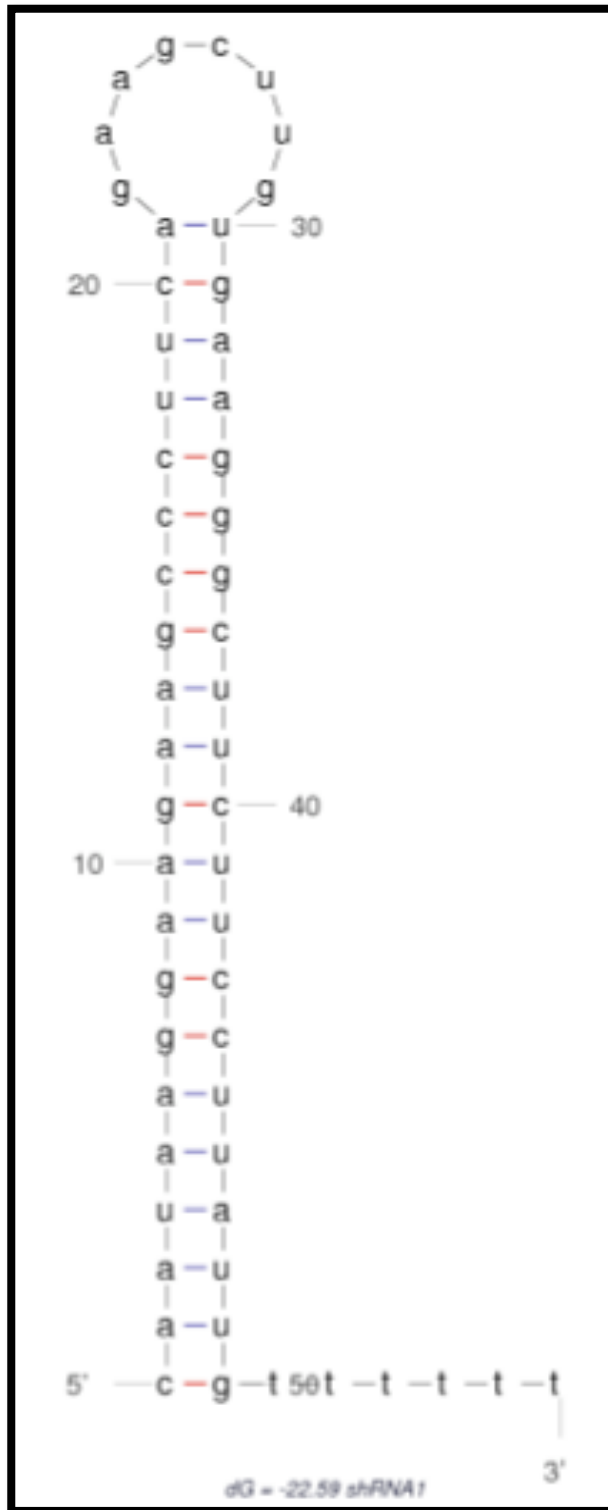


Figure 74 The Secondary structure prediction of shRNA1<sup>β2α2</sup> by “mfold”

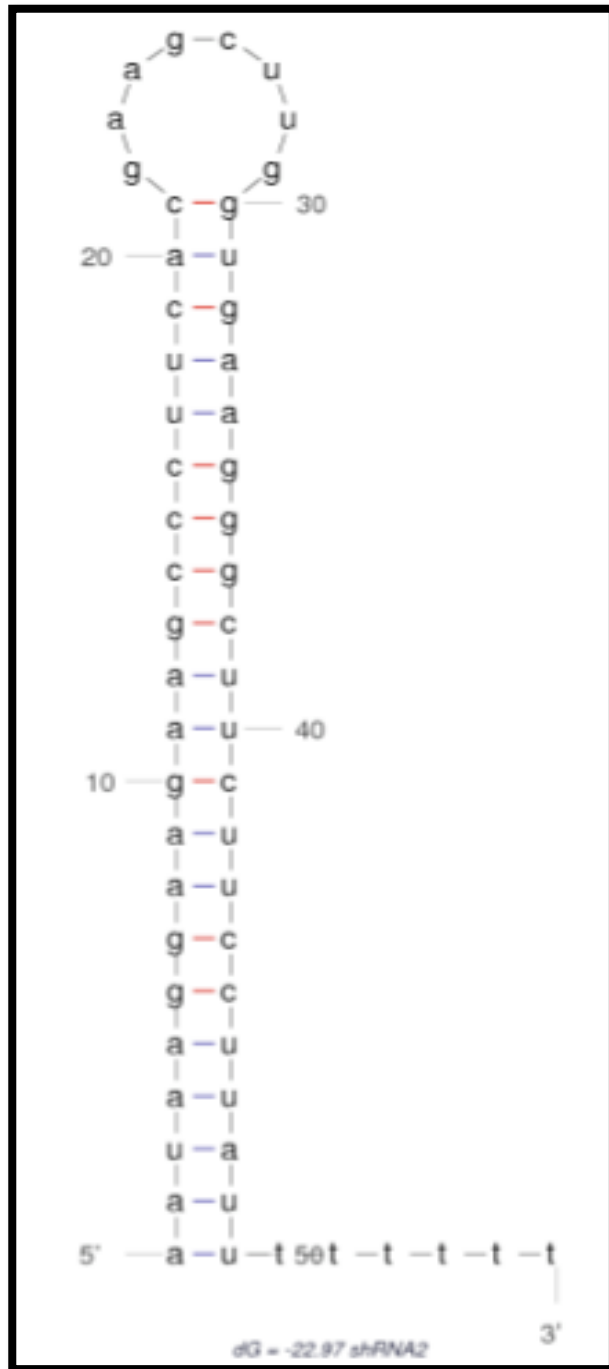


Figure 75 The Secondary structure prediction of shRNA2<sup>β2α2</sup> by "mfold"

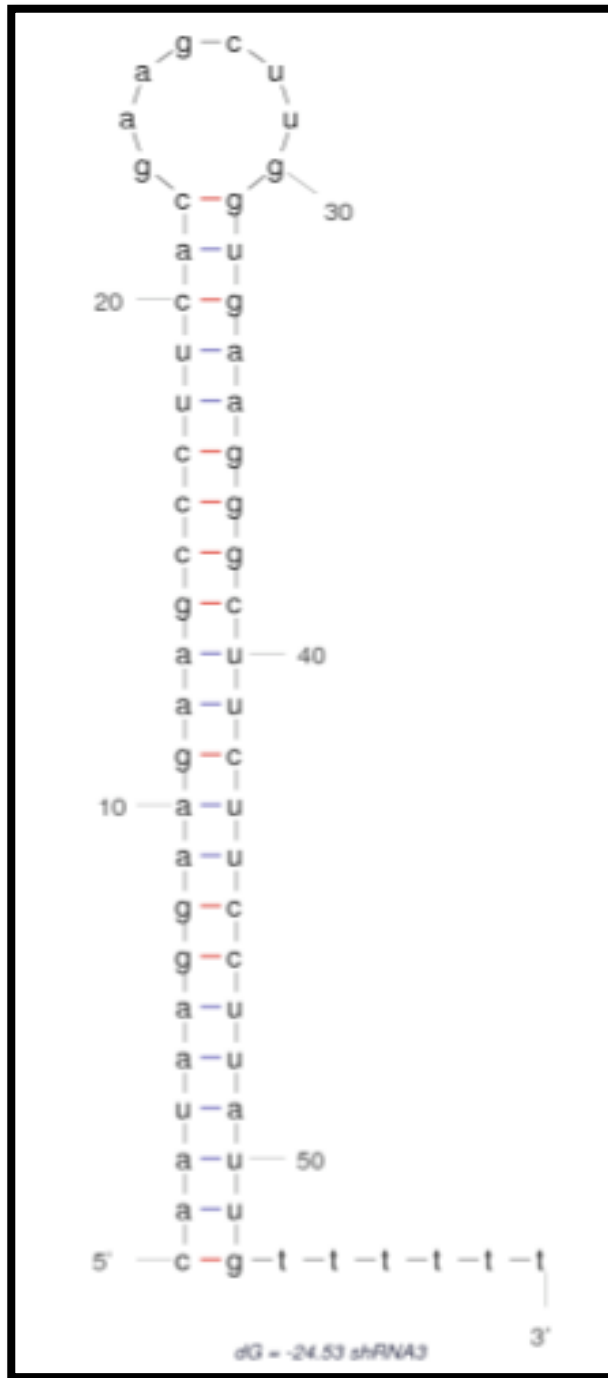


Figure 76 The Secondary structure prediction of shRNA3<sup>β2α2</sup> by “mfold”

### Appendix 3: The Protein Molecular Weight Marker used in the Current Project

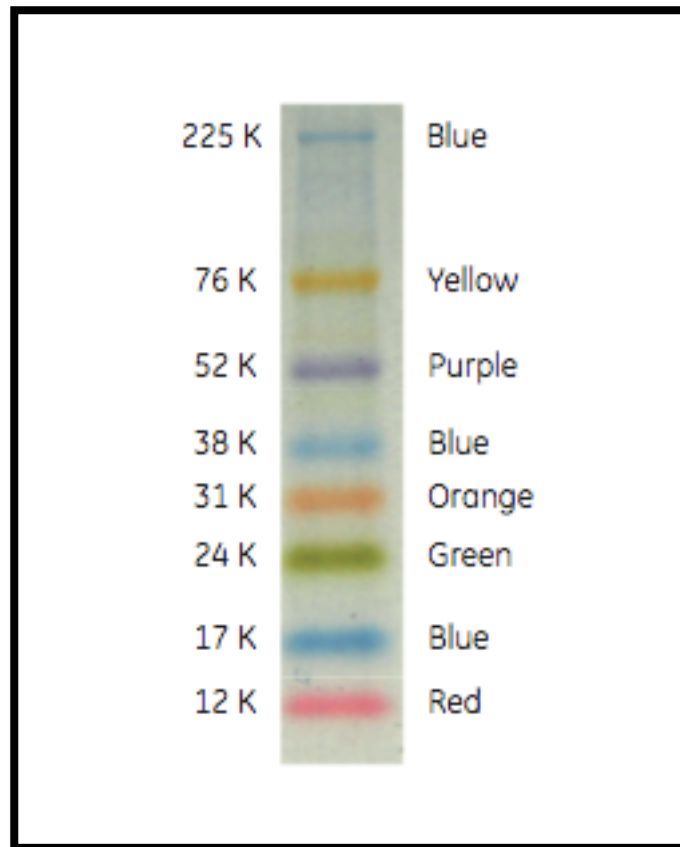


Figure 77 The High-Range Rainbow protein molecular weight marker.

## Appendix 4: The $\beta 3\alpha 2$ and $\beta 2\alpha 2$ Sequences

- The  $\beta 3\alpha 2$  fusion point sequence from the EMBL website (Accession number AJ131466).

```
1   gcgaacaagg gcagcaaagc tacggagagg ctgaagaaga agctgtcggg gcaggagtca
61  ctgctgctgc ttatgtctcc cagcatggcc ttcaggggtg acagccgcaa cggcaagagt
121 tacacgttcc tgatctcctc tgactatgag cgtgcagagt ggagggagaa catccgggag
181 cagcagaaga agtgtttcag aagcttctcc ctgacatccg tgggagctgca gatgctgacc
241 aactcgtgtg tgaaactcca gactgtccac agcattccgc tgaccatcaa taaggaagat
301 gatgagtctc cggggctcta tgggtttctg aatgtcatcg tccactcagc cactggattt
361 aagcagagtt caaaagcctc tcagcggcca gtagcatctg actttgagcc tcagggctctg
421 agtgaagccg ctcgttggaa ctccaaggaa aaccttctcg ctggaccagc tgaaaatgac
481 cccaaccttt tcgttgact gtatgatttt gtggccagtg gagataaac tctaagcata
541 actaaagggtg aaaagctccg ggtcttaggc tataatcaca atggggaatg gtgtgaagcc
601 caaaccaaaa atggccaagg ctgggtccca agcaactaca tcacgccagt caacagtctg
661 gagaaacact cctggtacca tgggcctgtg tcccgcaatg ccgctgagta tctgctgagc
721 agcgggatca atggcagctt cttggtgcgt gagagtgaga gcagtcctgg ccagaggtcc
781 atctcgctga gatacgaagg gaggggtgtac cattacagga tcaaacctgc ttctgatggc
841 aagctctacg tctcctccga gagccgcttc aacaccctgg ccgagttggt tcatcatcat
901 tcaacgggtg cgcacgggct catcaccacg ctccattatc cagcccaaaa gcgcaacaag
961 cccactgtct atgggtgtgc ccctaactac gacaagt
```

The fusion point is underlined and highlighted in yellow. The forward (in green) and reverse (in blue) primers to amplify the fusion point are highlighted.



- The  $\beta 2\alpha 2$  fusion point sequence from the EMBL website (Accession number AJ131467).

```

1   gcgaacaagg gcagcaaggc tacggagagg ctgaagaa ga agctgtcggg gcaggagtca
61  ctgctgctgc ttatgtctcc cagcatggcc ttcaggggtgc acagccgcaa cggcaagagt
121 tacacgttcc tgatctcctc tgactatgag cgtgcagagt ggagggagaa catccgggag
181 cagcagaaga agtgtttcag aagcttctcc ctgacatccg tggagctgca gatgctgacc
241 aactcgtgtg tgaaactcca gactgtccac agcattccgc tgaccatcaa taaggaaga
301 gcccttcagc ggccagtagc atctgacttt gagcctcagg gtctgagtga agccgctcgt
361 tggaaactcca aggaaaacct tctcgtgga cccagtgaaa atgaccccaa ccttttcggt
421 gactgtgatg attttggtgc cagtggagat aacactctaa gcataactaa aggtgaaaag
481 ctccgggtct taggtataa tcacaatggg gaatggtgtg aagcccaaac caaaaatggc
541 caaggctggg tccaagcaa ctacatcacg ccagtcaaca gtctggagaa acactcctgg
601 taccatgggc ctgtgtcccg caatgccgct gagtatctgc tgagcagcgg gatcaatggc
661 agcttcttgg tgcgtgagag tgagagcagt cctggccaga ggtccatctc gctgagatac
721 gaagggaggg tgtaccatta caggatcaac actgcttctg atggcaagct ctacgtctcc
781 tccgagagcc gttcaacac cctggccgag ttggttcac atcattcaac ggtggccgac
841 gggctcatca ccacgtcca ttatccagcc ccaaagcgca acaagccac tgtctatggt
901 gtgtcccca actacgacaa gt

```

The fusion point is underlined and highlighted in yellow. The forward and reverse primers to amplify the fusion point are highlighted in green and blue respectively.

## Appendix 5: The Raw Sequencing Data of the $\beta 3\alpha 2$ and $\beta 2\alpha 2$ Fusion Point Sequences in the pGEM<sup>®</sup>-T Easy Vector

- 3 colonies were sequenced, and all showed similar results. The  $\beta 3\alpha 2$  fusion point sequence for a representative colony is shown below, colony A5. The fusion point is marked by “|”:

```
NNNNNNNNNNNNNNNNATGNNNNNNGCCGCCATGGCGGCCGCGGAATTCGATTAGCTTCTCCCTGGCAT
CCGTGGNAGCTTCTCCCTGGCATCCGTGTCTTAGGCTATAATCACAATGGGGCTCCCTGGCATCCGTGGAG
GATCTTAGGCTATAATCACAATGGGACTAGCTTCTCCCTGGCATCCGTGGCTCGTAGTGTCTGGCGAGCA
GTTTGTGTTGATTTAACTGTTGAGGCTTAGGCTATAATCACAATGGGACTAGCTTCTCCCTGGCATCCGTGG
AGCCGCAGATGCTGACCAACTCGTGTGTGAAACTCCAGACTGTCCACAGCATTCCGCTGACCATCAATAAG
GAAGATGATGAGTCTCCGGGGCTCTATGGGTTTCTGAATGTCATCGTCCACTCAGCCACTGGATTTAAGCA
GAGTTCAA | AGCCCTTCAGCGGCCAGTAGCATCTGACTTTGAGCCTCAGGGTCTGAGTGAAGCCGCTCGT
TGGAAGTCCAAGGAAAACCTTCTCGCTGGACCCAGTGAAAATGACCCCAACCTTTTCGTTGCACTGTATGA
TTTTGTGGCCAGTGGAGATAACACTCTAAGCATAACTAAAGGTGAAAAGCTCCGGGTCTTAGGCTATAATC
ACAATGGGAATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGG
ATGCATAGCTTGAGTATTCTATAGTGCACCTAAATAGCTTGCGGTAATCATGGTCATAGCTGTTTCCCTGT
GTGAAATTGTTATCCGCTCACAATTCACACACACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTG
NCTAATGAGTGAGCTAACTCACATTAATTNGCGTTGCGCTCACTGCCCCGCTTTCAGTCGGGAAACCTGTC
GTGCCAGCTGCATTAATGAATCNGGCCNACGCGCGGGGAGAGCGGNTTNGCGNNTGGNCGCTCTTCCGCT
TCCTCGCTCACTGACTCGCTGCGCTCGGNCGTTTCGGCTGCGGCGAGCGNNNNCAGCTCACTCNAAGCGGN
ANNNNGTNTCCNCAGAATCAGNNNNNACGCAGGAAANAACATGNTGANCNANAAGNCNAGCNAANNCNNN
AANCNNAAAANGCNGNNNNNNNNNNNTTTNNNNNNNNCGCCNCCNNGNNNNNNNNCNAANNNGNNNNTC
ANNNNNANNNNNANNNNNNNNNNNANNNNCNGNNNNCCNNNNNNNNNTCCNNNNGNNGNNNNCNNNCNNN
NNNNN
```

- 3 colonies were sequenced, and all showed similar results. The  $\beta 2\alpha 2$  fusion point sequence for a representative colony is shown below, colony KCL-22 `1'. The fusion point is marked by “|”:

NNNNNNNNNNCNNNNNNNNNTNCNNCNGCCGCCATGGCGGCCGCGGAATTCGATTGCTACGGAGAGG  
 CTGAAGAGGAAGCTGTCGGAGCAGGAGTCACTGCTGCTGCTTATGTCTCCAGCATGGCCTTCAGGGTGCA  
 CAGCCGCAACGGCAAGAGTTACACGTTCTGATCTCCTCTGACTATGAGCGTGCAGAGTGGAGGGAGAACA  
 TCCGGGAGCAGCAGAAGAAGTGTTTCAGAAGCTTCTCCCTGACATCCGTGGAGCTGCAGATGCTGACCAAC  
 TCGTGTGTGAAACTCCAGACTGTCCACAGCATTCCGCTGACCATCAATAAGGAAG | AAGCCCTTCAGCGGC  
 CAGTAGCATCTGACTTTGAGCCTCAGGGTCTGAGTGAAGCCGCTCGTTGGAACCCAAGGAAAACCTTCTC  
 GCTGGACCCAGTGAAAAAGACCCCAACCTTTTCGTTGCACTGTATGATTTTGTGGCCAGTGGAGATAACAC  
 TCTAAGCATAACTAAAGGTGAAAAGCTCCGGGTCTTAGGCTATAATCACAATGGGGAATGGTGTGAAGCCC  
 AAACCAAAAATGGCCAANGCTGGGCCCAAGCAACTACATCACGAATCACTAGTGAATTCGCGGCCGCTG  
 CAGGTGACCATANGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCATATAGTGCACCTAAA  
 TAGCTTGCGTAATCATGGNCATAGCTGTTTCCTGTGTGAAATGTTATCCGCTCACANTTNCNCACAACA  
 TACGAGCCGGAAGCATAAAGTGNAAGCCTGGGGTGCCATAATGAGTGAGCTAACTCACATTAATNGCGTTG  
 CGCTCACTGCCCCTTTCCAGTCGGGAAANCNGNCNNGNCNGCTGCATTANNNAATCNGCCANCGCNCNG  
 NNAAGGCGGTTGCGNATTGGGNGCTCTTTCCGCTTCTCCGNNCACTNNACTCGNNGCGCTCGGNCGNNCN  
 GNNNN

## Appendix 6: The Human U6 and H1 Promoter Sequences

- The Human snRNA U6 Pol III promoter sequence from the EMBL Website (Accession number M14486).

```
1  cccagtgga  aagacgcgca  ggcaaaacgc  accacgtgac  ggagcgtgac  cgcgcgccga
61  gcccaaggtc  gggcaggaag  agggcctaatt  tcccatgatt  cttcatatt  tgcataatcg
121  atacaaggct  gttagagaga  taattagaat  taatttgact  gtaaacacaa  agatattagt
181  acaaaatcgc  tgacgtagaa  agtaataatt  tcttggttag  tttgcagttt  taaaattatg
241  ttttaaatg  gactatcata  tgcttaccgt  aacttgaaag  tatttcgatt  tcttggtttt
301  atatatcttg  tggaaaggac  gaaacaccgt  gctcgcttcg  gcagcacata  tacta  aaatt
361  ggaacgatac  agagaagatt  agcatggccc  ctgcaagg  atgacacgca  aattcgtgaa
421  gcgttcata  ttttacatc  aggttgtttt  tctgttttta  catcaggttg  tttttctggt
481  tggttttttt  tttacaccac  gtttatacgc  cggcgcacgg  tttacca
```

The forward and reverse primers to amplify the U6 promoter are highlighted in yellow and green respectively.

- **The Human RNA H1 Pol III promoter sequence from the EMBL Website (Accession number X16612).**

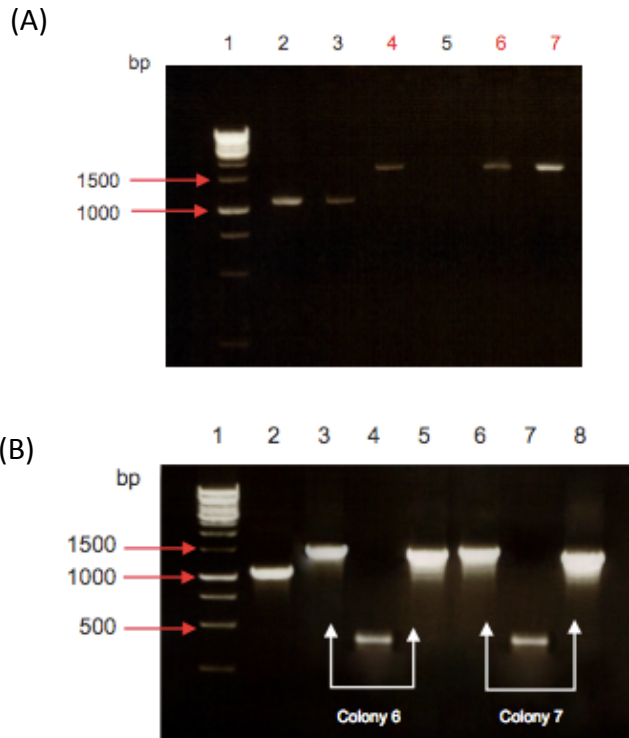
```

1   ttataggag ctgaaggaa gggggtcaca gtaggtggca tcgttccttt ctgactgcc
61  gcccccgca tgccgtccc cgatattgag ctccgaacct ctgcacctgc cgccgccggt
121 gtcctcgtcg cgccgcgccc ccatggaatt cgaacgctga cgtcatcaac ccgctccaag
181 gaatcgcggg cccagtgtca ctaggcggga acaccagcg cgcggtgcgcc ctggcaggaa
241 gatggctgtg agggacaggg gagtggcgcc ctgcaatatt tgcattgtcg tatgtgttct
301 gggaaatcac cataaacgtg aaatgtcttt ggatttggga atcttaag ttctgtatga
361 gaccactctt tccataggg cggagggaa ctcattcagt gggccacgag ctgagtgcgt
421 cctgtcactc cactcccatg tcccttggga aggtctgaga ctagggccag aggcggccct
481 aacagggctc tccctgagct tcggggaggt gatttcccag agaacggggc tccgcgagag
541 gtcagactgg gcaggagatg ccgtggacct cgcccttcgg ggaggggcc ccgagatgcc
601 tcctttgccg gagcttgaaa cagactcacg gccagcgaag tgagttcaat ggctgaggtg
661 aggtaccccg caggggacct cataaccxaa ttcagactac tctcctccgc ccatttttgg
721 aaaaaaaaaa aaaaaaaaaa aacaaaacga aaccgggccg ggcgcggtgg ttcacgccta
781 taatcccagc actttgggag gccgagggcg gcggtacaca aggtcaggag gtcgagacca
841 tccaggctaa cacggtgaaa cccccccca tctctactaa aaaaaaaaaa tacaaaaaat
901 tagccattag ccgggcgtgg tggcggcgcc ctataatccc agctacttgg gaggctgaag
961 cagaatggcg tgaaccggg aggcggacgt tgcagtgagc cgagatcgcg ccgactgcat
1021 tccagcctgg gcgacagagc gagtctcaaa aaaaaaa

```

The forward (in yellow) and reverse (in green) primers to amplify the H1 promoter are highlighted.

## Appendix 7: Electrophoretic Analysis of Representative Colony Screens PCR Products of the U6 shRNA and H1 shRNA Construct



**Figure 78** Electrophoretic analysis of colony screen PCR products of the U6 shRNA1 <sup>$\beta$ 3 $\alpha$ 2</sup> construct

Gel (A): PCR products of the U6 shRNA1 <sup>$\beta$ 3 $\alpha$ 2</sup> construct using the T7 forward and T3 reverse primers

Lane 1: Gene Ruler™ 1KB DNA Ladder

Lane 2: PCR product of wild type pB12mcs-eGFP vector (~1050bp)

Lanes 3-7: PCR products of the colony screen

Lanes 3, 6, and 7: positive colonies (~1390bp)

Gel (B): PCR products of pB12mcs-U6-shRNA1 <sup>$\beta$ 3 $\alpha$ 2</sup>-eGFP extracted plasmids

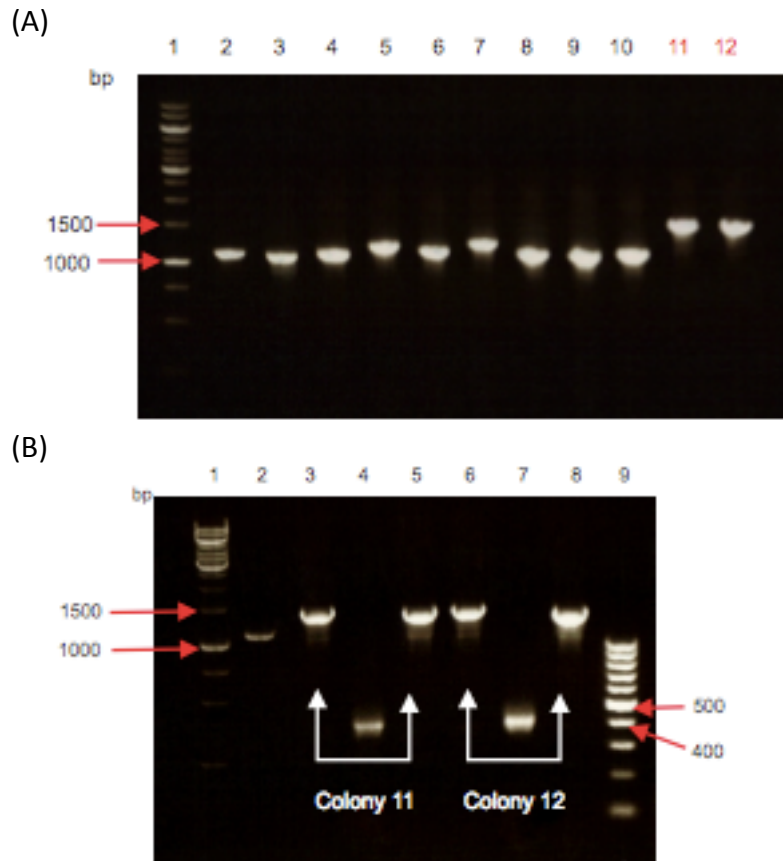
Lane 1: Gene Ruler™ 1KB DNA Ladder

Lane 2: PCR product of wild type pB12mcs-eGFP vector (~1050bp)

Lanes 3 and 6: PCR product of the primer set T7 forward and T3 Reverse (~1390bp)

Lanes 4 and 7: PCR products of the primer set T7 forward with U6 shRNA1 <sup>$\beta$ 3 $\alpha$ 2</sup> reverse (~380bp)

Lanes 5 and 8: PCR products of the primer set U6 forward and T3 reverse (~1420bp)



**Figure 79** Electrophoretic analysis of colony screen PCR products of the U6 shRNA2 <sup>$\beta$ 3 $\alpha$ 2</sup> construct

Gel (A): PCR products of the U6 shRNA2 <sup>$\beta$ 3 $\alpha$ 2</sup> construct using the T7 forward and T3 reverse primers

Lane 1: Gene Ruler™ 1KB DNA Ladder

Lane 2: PCR product of wild type pB12mcs-eGFP vector (~1050bp)

Lanes 3-12: PCR products of the colony screen

Lanes 11 and 12: positive colonies (~1390bp)

Gel (B): PCR products of pB12mcs-U6-shRNA2 <sup>$\beta$ 3 $\alpha$ 2</sup>-eGFP extracted plasmids

Lane 1: Gene Ruler™ 1KB DNA Ladder

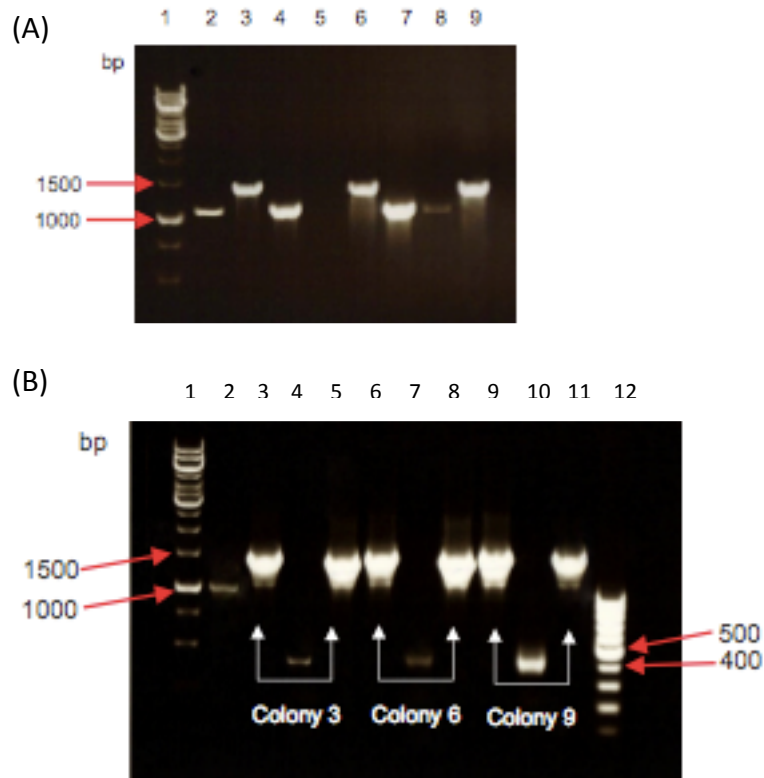
Lane 2: PCR product of wild type pB12mcs-eGFP vector (~1050bp)

Lanes 3 and 6: PCR product of the primer set T7 forward and T3 Reverse (~1390bp)

Lanes 4 and 7: PCR products of the primer set T7 forward with U6 shRNA2 <sup>$\beta$ 3 $\alpha$ 2</sup> reverse (~380bp)

Lanes 5 and 8: PCR products of the primer set U6 forward and T3 reverse (~1420bp)

Lane 9: Gene Ruler™ 100bp DNA Ladder



**Figure 80** Electrophoretic analysis of colony screen PCR products of the U6 shRNA3<sup>β3α2</sup> construct

Gel (A): PCR products of the U6 shRNA3<sup>β3α2</sup> construct using the T7 forward and T3 reverse primers

Lane 1: Gene Ruler™ 1KB DNA Ladder

Lane 2: PCR product of wild type pB12mcs-eGFP vector (~1050bp)

Lanes 3-9: PCR products of the colony screen

Lanes 3, 6, and 9: positive colonies (~1390bp)

Gel (B): PCR products of pB12mcs-U6-shRNA3<sup>β3α2</sup>-eGFP extracted plasmids

Lane 1: Gene Ruler™ 1KB DNA Ladder

Lane 2: PCR product of wild type pB12mcs-eGFP vector (~1050bp)

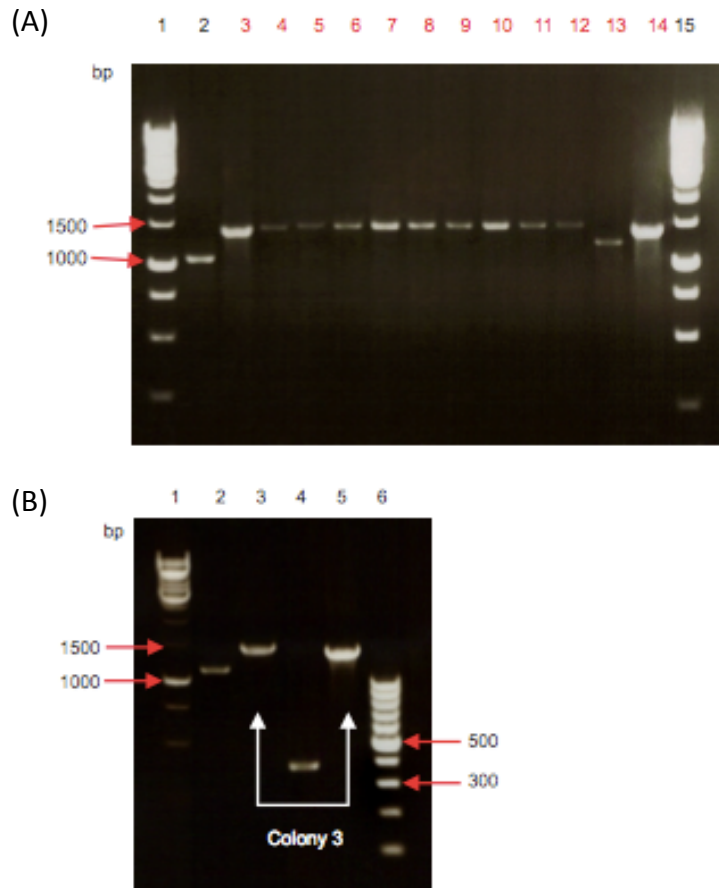
Lanes 3, 6, and 9: PCR product of the primer set T7 forward and T3 Reverse (~1390bp)

Lanes 4, 7, and 10: PCR products of the primer set T7 forward with U6 shRNA3<sup>β3α2</sup> reverse (~380bp)

Lanes 5, 8, and 11: PCR products of the primer set U6 forward and T3 reverse (~1420bp)

Lane 12: Gene Ruler™ 100bp DNA Ladder





**Figure 81** Electrophoretic analysis of colony screen PCR products of the H1 shRNA1<sup>β2α2</sup> construct

Gel (A): PCR products of the H1 shRNA1<sup>β2α2</sup> construct using the T7 forward and T3 reverse primers

Lanes 1 and 15: Gene Ruler™ 1KB DNA Ladder

Lane 2: PCR product of wild type pB12mcs-eGFP vector (~1050bp)

Lanes 3-14: PCR products of the colony screen

Lanes 3-14: positive colonies (~1340bp)

Gel (B): PCR products of pB12mcs-H1-shRNA1<sup>β2α2</sup>-eGFP extracted plasmid

Lane 1: Gene Ruler™ 1KB DNA Ladder

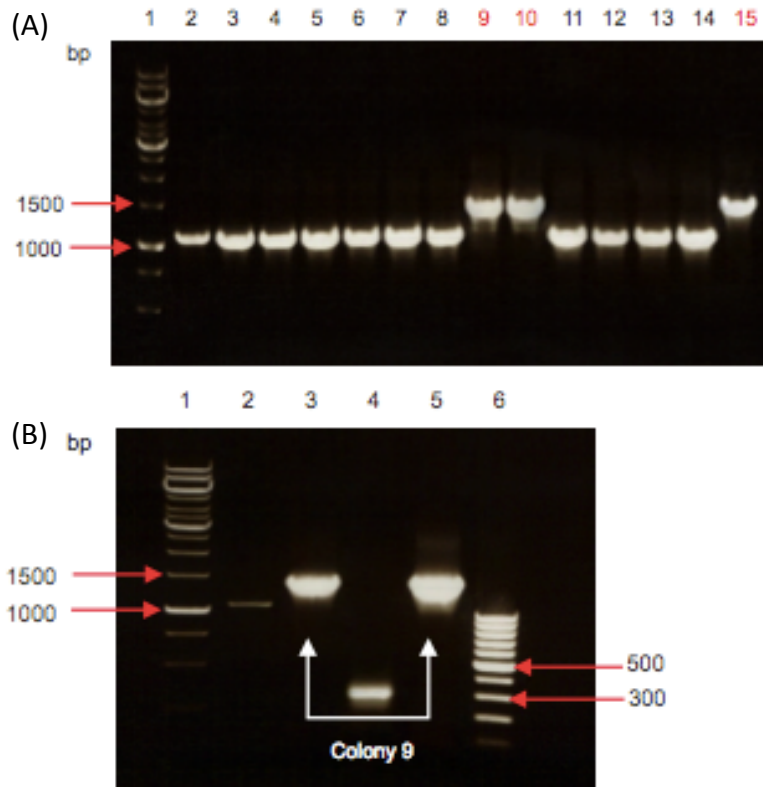
Lane 2: PCR product of wild type pB12mcs-eGFP vector (~1050bp)

Lane 3: PCR product of the primer set T7 forward and T3 Reverse (~1340bp)

Lane 4: PCR products of the primer set T7 forward with U6 shRNA1<sup>β3α2</sup> reverse (~330bp)

Lane 5: PCR products of the primer set H1 forward and T3 reverse (~1370bp)

Lane 6: Gene Ruler™ 100bp DNA Ladder



**Figure 82** Electrophoretic analysis of colony screen PCR products of the H1 shRNA2<sup>β2α2</sup> construct

Gel (A): PCR products of the H1 shRNA2<sup>β2α2</sup> construct using the T7 forward and T3 reverse primers

Lane 1: Gene Ruler™ 1KB DNA Ladder

Lane 2: PCR product of wild type pB12mcs-eGFP vector (~1050bp)

Lanes 3-15: PCR products of the colony screen

Lanes 9-10, and 15: positive colonies (~1340bp)

Gel (B): PCR products of pB12mcs-H1-shRNA2<sup>β2α2</sup>-eGFP extracted plasmid

Lane 1: Gene Ruler™ 1KB DNA Ladder

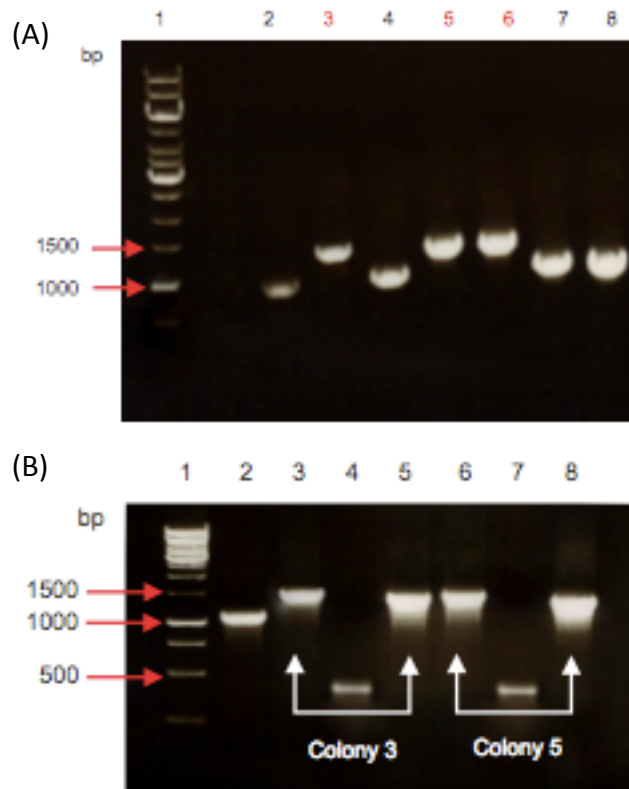
Lane 2: PCR product of wild type pB12mcs-eGFP vector (~1050bp)

Lane 3: PCR product of the primer set T7 forward and T3 Reverse (~1340bp)

Lane 4: PCR products of the primer set T7 forward with H1 shRNA2<sup>β3α2</sup> reverse (~330bp)

Lane 5: PCR products of the primer set H1 forward and T3 reverse (~1370bp)

Lane 6: Gene Ruler™ 100bp DNA Ladder



**Figure 83 Electrophoretic analysis of colony screen PCR products of the H1 shRNA3<sup>β2α2</sup> construct**

Gel (A): PCR products of the H1 shRNA3<sup>β2α2</sup> construct using the T7 forward and T3 reverse primers

Lane 1: Gene Ruler™ 1KB DNA Ladder

Lane 2: PCR product of wild type pB12mcs-eGFP vector (~1050bp)

Lanes 3-8: PCR products of the colony screen

Lanes 3, 5, and 6: positive colonies (~1340bp)

Gel (B): PCR products of pB12mcs-H1-shRNA3<sup>β2α2</sup>-eGFP extracted plasmid

Lane 1: Gene Ruler™ 1KB DNA Ladder

Lane 2: PCR product of wild type pB12mcs-eGFP vector (~1050bp)

Lanes 3 and 6: PCR product of the primer set T7 forward and T3 Reverse (~1340bp)

Lanes 4 and 7: PCR products of the primer set T7 forward with H1 shRNA2<sup>β3α2</sup> reverse (~330bp)

Lanes 5 and 8: PCR products of the primer set H1 forward and T3 reverse (~1370bp)

## Appendix 8: The Raw Sequencing Data of the U6 and H1 Promoters Inserted into the pB12mcs-eGFP Plasmid

- Sequencing of the U6 promoter inserted into the pB12mcs-eGFP
  1. Colony number 5. The following U6 sequence is given as the forward sequence 5' → 3' and is highlighted in yellow. The U6 sequence shows no mismatches to the U6 sequence found in the EMBL website (accession number M14486).

```
NNNNNNNNNNNNNGNNNATCTCGAGAGCTTCAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTT
CATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATA
TTAGTACAAAATACGTGACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTAAAAATTATGTTTTAA
AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATCTTGTGGAAAG
GACGAAACACCGTGCTCGCTTCGGCAGCACATATACTAGGATCCACAGCCACCATGAGTAAAGGAGAAGAA
CTTTTCACTGGAGTTGTCCCAATCTTGTGTTGAATTAGATGGTGATGTTAATGGGTACAAATTTTCTGTCAG
TGGAGAGGGTGAAGGTGATGCAACATACGGAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACCTAC
CTGTTCCATGGCCAACTTGTCACTACTCTCACTTATGGTGTTCATGCTTTTCAAGATATCCAGATCAT
ATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAGAGGACCATCTTCTTCAA
GGACGACGGGAAC TACAAGACACGTGCTGAAGTCAAGTTTGAGGGAGACACCCCTCGTCAACAGGATCGAGC
TTAAGGGAATCGATTTCAAGGAGGACGGAAACATCCTCGGCCACAAGTTGGAATACAAC TACNACTCCCAC
AACGTATACATCATGGCCGACAAGCAAAGAACGGCATCAAAGCCAAC TCAAGACCCGCCACAACATCGA
AGACGGCGGCGTGCAACTCGCTGATCATTATCAACAAAATAC TCCAATTGGCGATGACCCTGTCTTTTAC
CAGACAACCATTACCTGTCCACACAATCTGCCCTTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGNC
CTTCTTGAGTTGTACGGCTGCTGGGNNTACN NATGNNATGNNANTATACAAANNANTCGANCNTANNA
CTNNANNNGTCTGTANTACGTANATCNNANNTGATAANNANNNAT TGATNANTTNGNNNNNNNNANTANNN
GCNNNAAAANGNNTTNANTTNGNGNAANTNGGANGNNNTGNTNNNNNANCNNTNNNNNNNCNNNANNAAN
NTNNNNNNNNC NNNNNNTNNNNNTNNGNNNNNNNNNNNNNGNNNNNNANTTTTTNTNNNN
```

The following is the same U6 sequence, however, it is given as the reverse sequence 5' → 3'.

```
NNNNNNNNNNNNNNANNGGNNNNCTTAAGTCGCGCCGCGAATTA AAAAACCTCCCACACCTCCCCCTGA
ACCTGNNNNNTAAAATGAATGCAATTGTTGTTGTTAACTTGT TTTATTTGCAGCTTATAATGGTTACAAATAA
AGCAATAGCATCACAATTTACAAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAACT
CATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTATTT
GTATAGTTCAATCCATGCCATGTGTAATCCCAGCAGCCGTTACAACTCAAGAAGGACCATGTGGTCTCTCT
TTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGTCA
TCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCCGCTCTTCGATGTTGTGGCGGGT
CTTGAAGTTGGCTTTGATGCCGTTCTTTTGCTTGTTCGGCCATGATGTATACGTTGTGGGAGTTGTAGTTGT
```

ATTCCAACCTTGTGGCCGAGGATGTTTCCGTCCCTTGAATCGATTCCCTTAAGCTCGATCCTGTTGACG  
 AGGGTGTCTCCCTCAAACCTTGACTTCAGCACGTGTCTTGTAGTTCCTCGTCCTTGAAGAAGATGGTCTT  
 CTCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATCTTG  
 AAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTAGTG  
 CAAATAAATTTAAGGGTAAGTTTTCCGTATGTTGCATCACCTTCACCTCTCCACTGACAGAAAATTTGTA  
 CCCATTAACATCACCATCTAATTCACAAGAATTGGGACAACCTCCNGTGAAAAGTCTTCTCCTTTACTCA  
 TGGTGGCTGTGGATCCTAGTATATGTGCTGCCGAAGCGAGNCGGNGCTTCGNCCTTCCACANANNTATN  
 AANNCANNAATCGAAATACCTTCAANTNNGNAAGCANATGANNTCCATTTTNAACNTAANTTTAAACT  
 GCNAACTNCCNNGAANNNNNNNNNTTCNACNNCNCNNNNNNNNNAANNNNNTNNNNNNNNNAANNANNA  
 NNNNANNNCNNNNNNNNNNNTNCNNANTGCNAANNNGNNNNNNNCNNGNNANNGNCNNNNNTNNNNNNNCN  
 NNNNNNN

2. Colony number 11. The following U6 sequence is given as the forward sequence 5' → 3' and is highlighted in yellow. The U6 sequence shows no mismatches to the U6 sequence found in the EMBL website (accession number M14486).

NNNNNNNNNNNNNNNNNNCTCGAGAAGCTTCAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCCT  
 TCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAGAT  
 ATTAGTACAAAATACGTGACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTAAAATTTATGTTTTA  
 AAATGGACTATCATATGCTTACCGTAACTTGAAAAGTATTTTCGATTTCTTGGCTTTATATATCTTGTGGAAA  
 GGACGAAACACCGTGCTCGCTTCGGCAGCACATATACTAGGATCCACAGCCACCATGAGTAAAGGAGAAGA  
 ACTTTTCACTGGAGTTGTCCCAATCTTGTGTAATTAGATGGTGATGTTAATGGGTACAAATTTCTGTCA  
 GTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACCTTACCCTTAAATTTATTTGCACTACTGGAAAAC  
 CCTGTTCCATGGCCAACACTTGTCACTACTCTCACTTATGGTGTCAATGCTTTTCAAGATATCCAGATCA  
 TATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAGAGGACCATCTTCTTCA  
 AGGACGACGGGAAC TACAAGACACGTGCTGAAGTCAAGTTTGGAGGAGACACCCTCGTCAACAGGATCGAG  
 CTTAAGGGAATCGATTTCAAGGAGACGGAACATCCTCGGCCACAAGTTGGAATACAAC TACAAC TCCCA  
 CAACGTATACATCATGGCCGACAAGCAAAAAGAACGGCATCAAAGCCAACCTTCAAGACCCGCCACAACATCG  
 AAGACGGCGGCGTGCAACTCGCTGATCATTATCAACAAAATAC TCCAATTGGCGATGACCCTGTCTTTTA  
 CCAGACAACCATTACCTGTCCACACAATCTGCCCTTCGAAAAGATCCCAACGAAAAGAGAGACCACATGGN  
 CCTTCTTGAGTTTGTAAACGGCTGCTGGGATTACNCATGGNATNNGANTATACNAATAANTCNANNNTAN  
 ANTATAGTGAGTCGTANTACGTANATCNANNGANNAGNNNNNNNATNNNTTTNGNNNNNNCNNNNNANAN  
 GCNNNAAAAANGNTTTNNTNGNNNAANTGNNGNNTNGNNTNNNTNNNNCNNNNNNCNCANNNNANNTNA  
 NNNNNNNNNNGNNTNNNNNTNNNNNNNNNNNNNGNNNNNNNNNNNNNNNNNN

The following is the same U6 sequence, however, it is given as the reverse sequence 5' → 3'.

NNNNNNNNNNNNNNANNGGNNAACTTAAGTCGGGCCGGAATTAAAAACCTCCACACCTCCCCCTG  
 AACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATA  
 AAGCAATAGCATCACAATTTCAAAAATAAAGCATTTTTTCTCACTGCATTCTAGTTGTGGTTTTGTCCAAAC

TCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTATT  
 TGATAGTTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCTCTC  
 TTTTCGTTGGGATCTTTTCGAAAGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGTC  
 ATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCCGCTTTCGATGTTGTGGCGGG  
 TCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTCGGCCATGATGTATACGTTGTGGGAGTTGTAGTTG  
 TATTCACACTTGTGGCCGAGGATGTTTCCGTCCTCCTTGAAATCGATTCCCTTAAGCTCGATCCTGTTGAC  
 GAGGGTGTCTCCCTCAAACCTTGACTTCAGCACGTGTCTTGTAGTTCCTCGTCTCCTTGAAGAAGATGGTCC  
 TCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCTGTCGCCCTTCATATGATCTGGATATCTT  
 GAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTAGT  
 GCAAATAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAAATTTGT  
 ACCCATTAACATCACCATCTAATTCACAAGAATTTGGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTACTC  
 ATGGNGGCTGTGGATCCTAGTATATGTGCTGCCGANGCAGCACNGTTTTCGTCTTTCNCANANNNNNAA  
 NNNNAAANCGAANTANTTTTCAGTANGNAGCNTATGANNNTCCATTTNAAACNNNANTTNAACTGCAACTN  
 NCCNNNAANTNNNNNTNNNACNTNNNNNNNNNTNNNNANNNNNNNNNNNGNNANNNNNNNNNNNNCNNANN  
 CNNNNCNNNNNGCANNNNANNNNNNNNNNNNNNNNNNTTTNNNNNNNNNAANNNNNNNN

Another 6 colonies were also sequenced and were found to have mismatches (raw sequences are not shown).

- Sequencing of the H1 promoter inserted into the pB12mcs-eGFP

1. Colony number 8. The following U6 sequence is given as the forward sequence 5' → 3' and is highlighted in yellow. The U6 sequence shows no mismatches to the U6 sequence found in the EMBL website (accession number X16612).

NNNNNNNNNNNNNGNNATCTCGAGAAGCTTGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCC  
 CANTGTCACTAGGCGGGAACACCAGCGCGCTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAG  
 TGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATGCTTTGGAT  
 TTGGGAATCTTATAAGTTCTGTATGAGACCCTCTTTGGATCCACAGCCACCATGAGTAAAGGAGAAGAA  
 CTTTTCACTGGAGTTGTCCCAATCTTGTGTAATAGATGGTGATGTTAATGGGTACAAATTTCTGTACAG  
 TGGAGAGGGTGAAGGTGATGCAACATACGGAAAACCTACCCTTAAATTTATTTGCACTACTGGAAAACCTAC  
 CTGTTCCATGGCCAACACTTGTCACTACTCTCACTTATGGTGTTCATGCTTTTCAAGATATCCAGATCAT  
 ATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAGAGGACCATCTTCTTCAA  
 GGACGACGGGAAC TACAAGACACGTGCTGAAGTCAAGTTTGGAGGAGACACCCCTCGTCAACAGGATCGAGC  
 TTAAGGGAATCGATTTCAAGGAGGACGGAAACATCCTCGGCCACAAGTTGGAATACAAC TACAACCTCCAC  
 AACGTATACATCATGGCCGACAAGCAAAAAGACGGCATCAAAGCCAACCTTCAAGACCCGCCACAACATCGA  
 AGACGGCGGCGTCAACTCGTGATCATTATCAACAAAATACCTCAATTTGGCGATGACCCTGTCTTTTAC  
 CAGACAACCATTACCTGTCCACACAATCTGCCCTTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTC  
 CTCTTTGAGTTTGTAAACGGCTGCTGGGATTACACATGGCATGGNTGAACTATACAAATAACTCGAGCCTCT  
 AGAACTATAGTGAGTCGTANTACGTANATCCNGACATGATAGATACATTGATGAGTTTGGNNAANNACACT  
 AGAATGCAGTGAAAAATGNTTNATTTGNGAAATTTNGNGATGCTNTGCTTNNNNGNNNNNTNNNAGCTGCAN  
 TNAANNANTNNCAACNNNNNTGCATTTNNNTTNNNNNNNNNGNNNNNNNNNNNNNNNNNTNNNAANNNTTTNNNNN  
 CNGNNCNNNNNTNNNANNNNNCNNNNNNNNNNNNNNNNNNNNNNCNTNGNTNCNNNNNNNGNNGGGN

The following is the same H1 sequence, however, it is given as the reverse sequence 5' → 3'.

```
NNNNNNNNNNNNNGNNANNGGTAATCTTAAGTCGCGGCCGGAATTAAAAACCTCCCACACCTCCCC
TGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAA
TAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAA
ACTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTA
TTTGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCTC
TCTTTTCGTTGGGATCTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGG
TCATCGCCAATTGGACTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCCGCTCTCGATGTTGTGGCG
GGTCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTGCGCCATGATGTATACGTTGTGGGAGTTGTAGT
GTATTCCAACCTGTGGCCGAGGATGTTTCCGTCCTCCTTGAATCGATTCCCTTAAGCTCGATCCTGTTG
ACGAGGGTGTCTCCCTCAAACCTTGACTTCAGCACGTGTCTTGTAGTTCCCGTCGTCCTTGAAGAAGATGGT
CCTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATC
TTGAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTA
GTGCAAATAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAATTT
GTACCCATTAACATCACCATCTAATTCAACAAGAATTGGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTAC
TCATGGTGGNTGTGNNCCGAAAGAGTGNNNCATACNGAACTTATANATCCCAAATCCAANACNTTTCACG
TTTATGGTGATTTNCCANANNCNTANGNCNTGCANNNNTGCNNNCNNCACNCCNNTCCNNNNNNAN
CNTCTTNCNNCGGNNNANGNNNCNNGNNNTCCNNCNANNGNNNCTNNNNNCNATNNNNNNNNNNNNN
ANNNNNNCNCCNCCANCTTNNNNNNNCCNNNNANCCNNNNNNNCNNNNNNNCCNNNNNNNN
```

Another 4 colonies were also sequenced and were found to have mismatches (raw sequences are not shown).





TGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCTCTC  
TTTTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGT  
ATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCCGCTTCGATGTTGTGGCGGG  
TCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGCGGCCATGATGTATACGTTGTGGGAGTTGTAGTTG  
TATTTCCAACCTGTGGCCGAGGATGTTTCCGTCCTCTTGAAATCGATTCCCTTAAGCTCGATCCTGTTGAC  
GAGGGTGTCTCCCTCAAACCTTGACTTCAGCACGTGCTTGTAGTTCCCGTCGTCCTTGAAAGAAGATGGTCC  
TCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAAGAAGTCGTGCCGCTTCATATGATCTGGATATCTT  
GAAAAGCATTTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTCCAGTAGT  
GCAAATAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAATTTGT  
ACCCATTAACATCACCATCTAATCAANAAGAAATGGNACAACCCAGTGAAAANTNCTTCTCCTTTACTC  
ATGGTGGCTGTGGATCCAAAAAGCNNNNNTCAAAGCCCTTNNCAGCTTCTGAANGGNTTTTGAACCTCN  
CTAGTNNATGNGCTGNCGAAGCGANNNNGGNGNNTTCGTNNNTTCNNCAANATANNATAAGNCAAGAAATC  
GAAANACNTTNNNGNACNNNNNNNNGANAGNCCNTTTTTAAANNNNNNTNANNNGNANNNNNNAA  
NTNNNNNTNNNNNNCNNNNNTNNNNNNNANNNNTNNNTNNNNNNAANNNNNNAANNNNNN

- Colony 4, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNGNTATCTCGAGAAGCTTCAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCCT  
TCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAGAT  
ATTAGTACAAAATACGTGACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTAAAATTATGTTTA  
AAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATCTTGTGGAAA  
GAACGAAACACTGTGCTCGCTTCGCAAGAAATTAACAAGAAAAGTTCAAAGCCCTTCAAACCTGGGAA  
AGGGTTTTTAAACTCGGTTTTTTGGATCCCCACCCCCAGGAGTAAGGAAAAAAATTTTTTCATTGAATT  
TGCCCCATTTCTGGTTAAATTAAGGGGGAGGTTAAGGGGAACAAATTTTCGGCCANGGAAAAGGGTGAAG  
GNGATGCACCTTACGAAAAATTNACCCTTAAATTTATTTGCACTATGGAAAACTACCGGTCNTTGNCCA  
ACNNTGGTCATTACTTCNATTTAGNGGTCNATTTGGTTTTCAAATTTCCGAATANNATGAANCGNCCAC  
CTTCTTNNNAANNCCCTTGCNNAAGGAANNNGTANGAANAGGACNTCCTCCTCANGACNAANGGAANTT  
NCNNGANNCTNCCGGAGTCCAGGTTGAAGGNAGANNCCCTCCNNNNNGAATCAANCTAANGGAANTCAAT  
TTCAAGGAGNANNNAATCNCNCTCGNCNNNANNNGGNAANANNNTACAACCCCCACCGANACTTCNGNN  
CNNANAGGCAAAGGAANGNNTCNAGNCAACTTNNGNCCNNCCCNACATCCAAAACGTCNGNNTGCNNCN  
NNTNNTCTTNNCANCAATGACTCAANTNNNNATGNCNNNCNTTTTACNNNNACNNTNNNTNCCNCCN  
NNCNGCCNNTNNANNNNCCNNNNANANNNNNNCNNNNNGNTCNNNTNNNTTNNNTNNGNCNCTGNNNT  
NNNNNTGNNNGNNTNANTANNCANNNANNNNNNNNNNNNNANNNNGNANNNNTNNNNNNCNCNANNN  
NNNANCNTNNTGNTTNNNNNNNNNCNNNNNNNNNNNNNTNNNNNNNNNNNTGCNTATNNNCNTNNTN  
NCNNGNNNNNNNNNNNNNN

- Colony 4, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNGNANNGGNNNNCTTAAGTCGCGGCCGGAATTA AAAAACCTCCCACACCTCCCCCT  
GAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAAT  
AAAGCAATAGCATCAAAATTTCAAAATAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAA  
CTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTAT  
TTGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCTCT  
CTTTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGT

CATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCGTCTTCGATGTTGTGGCGG  
GTCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTGCTGCGCCATGATGTATACGTTGTGGGAGTTGTAGTT  
GTATTTCCAACTTGTGGCCGAGGATGTTTCCGTCCTCCTTGAATCGATTCCCTTAAGCTCGATCCTGTTGA  
CGAGGGTGTCTCCCTCAAACCTGACTTCAGCACGTGTCTTGTAGTTCCCGTCGTCCTTGAAGAAGATGGTC  
CTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATCT  
TGAAAAGCATTTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTAG  
TGCAAATAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAATTTG  
TACCATTAAACATCACCATCTAATTCACAAGAATTTGGGACAACCCAGTGAAAAGTTCTTCTCCTTTACT  
CATGNNNGTGTNNNCCAAAAAGCAGAGTTCAAAGCCCTTCACNAGCTTCTGANGGCTTTTGAACCTGCG  
TAGTANATGTGCTGCNNANNNNAGNNNNAGNNGTNTNNTNNTTCCNNAANAANNNNNTNAANNNNNAANN  
NNAANANTTTNNAGTACGNANGNNNNNNNNNCNNTTNAANNNNNNNNNNNNANNNNCCNNANNNNN  
NNNNNNNNNNNTTNNNNNNNTNNNNNNNNNNNAANNANNNNNANNNN

- Colony 6, the forward sequence is given from 5' → 3'

NNNNNNNNNNNCGCNTGCNCCGCGCCATGGCGGCCGCGGGAATTCGATTTGGAATTCAAGGTCGGGC  
AGGAAGAGGGCCTATTTCCCATGATTCCTTCATATCTGCATATACGATACAAGGCTGTAGAGAGATAATT  
AGAAATTAATTTGACTGTAAACACAAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGG  
GTAGTTTGCAGTTTAAATTAATGTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCCG  
ATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAGCACCGTGCCTTCGGCAGCACATATACTAGCA  
ATAAGGAAGAAGCCCTTCAGAAGCTTGTGAAGGGCTTCTTCCCTTATGTTTTTGGATCCATAATCACTAG  
TGAATTCGCGGCCGCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTAT  
TCTATAGTGCNCCTAAANNATAATTTTAAACTGCAAACTACCCNGAAATTATTACTTTCTACGTTACGT  
ATTTTGTACTAATATCTTTGTGTTTACNGTCAAATTAATTCTAATTATCTCTCTAACAGCCNTGTATCGTA  
TAT

- Colony 6, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNCGGTTGGGAGCTCTCCATATGGTCGACCTGCAGGCGGCCGGAATTCACTAGTGATT  
ATGGATCCAAAAACAATAAGGAAGAAGCCCTTCAACAAGCTTCTGAAGGGCTTCTTCCCTTATTGTAGTATA  
TGTGCTGCCGAAGCGAGCACGGTGCTTCGTCTTTCCACAAGATATATAAAGCCAAGAAATCGAAACTT  
TCAAAGTTACGGTAAGCATATGATAGTCCATTTTAAAAATAATTTTAAACTGCAAACTACCCAAGAAAT  
ATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCT  
CTAACAGCCTTGATCGTATATGCAGATATGAAGGAATCATGGGAAATAGGCCCTTCTCCTGCCGACCTT  
GAATTCAAATCGAATTCGCCGCGCCCATGGCGGCCGGGAGCATGCGACGTGGGCCCAATTCGCCCTA  
TAGTGAGTCGTATNAA

- Colony 7, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNNNNNTGNNNAGNNNNCCATATGGTCGACCTGCAGGCGGCCGNAATTCACTAG  
TGATTTGGAATTCAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGAT  
ACAAAGCTGTTANAGAGATAATTAGAATTAATTTGACTGTAAACACAAAAGATATTAGTACAAAATACGTGA

CGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTAAAAATTATGTTTTAAAAATGGACTATCATATGCC  
TACCGTAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGTGCTCG  
CTTCGNCGACACANATACTAGCAGAGTTCAAAAGCCCTTCAGAAGCTTGTGAAGGGCTTTTGAACCTCTGCT  
TTTTTGGATCCATAATCGAATTCGCCGCGCCGCGCATGGCGNCCGGNAGCATGCGACGCTCGGGCCCAATTCG  
CCCTATAGTGAGTCGNATTAANN

- Colony 7, the reverse sequence is given from 3' → 5'

NNNNNNNNNNC  
GAGTTCAAAGCCCTCNAAGCTTCTGAAGGGCTTTGAACTCTGCTAGTATATGTGCTGCCGAAGCGAGC  
ACGGNGTTTCGTCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCAT  
ATGAGAGTCCATTTTAAACATAATTTTAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACG  
TATTTTGTACTAATATCTTTGTGTTTACAGGCAAATTAATTCCTAATTATCTCTCTAACAGCTTTGTATCGT  
ATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCCTGCCGACCTTGAATTCCAAATCACTAGT  
GAATTCGCGGCCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCNNAGCTTGAGTATT  
CTATAGNGTCACCTAAAANN

1. The following are sequences from colonies with the pB12mcs-U6-shRNA2<sup>β3α2</sup>-eGFP plasmid. The sequencing results were inconclusive.

- Colony 11, the forward sequence is given from 5' → 3'

NNNNNNNNNNG  
TGTTCCAGGGCTTTAAATAGCTGGTGGTGAACCAATATTCGTGCTCGCTTCGGCAGCACATATACTAG  
CAGAGTTCAAAGCCCTCAGGAAGCTTGTGAAGGGCTTTTGAACCTCTGTTTTTTGGATCCACAGCCACC  
ATGAGTAAAGGAGAAGAAGCTTTTCACTGGAGTTGTCCCAATTCCTGTTGAATTAGATGGTGATGTTAATGG  
GTACAAATTTTCTGTCTAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACCTTACCCTTAAATTTATTT  
GCACTACTGGAAAACCTACCTGTTCATGGCCAACACTTGTCACTACTCTCACTTATGGTGTTCAATGCTTT  
TCAAGATATCCAGATCATATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCCCTGAGGGATACGTGCAGGA  
GAGGACCATCTTCTTCAAGGACGACGGGAACCTACAAGACACGTGCTGAAGTCAAGTTTGGAGGAGACACC  
TCGTCAACAGGATCGAGCTTAAGGAATCGATTTCAAGGAGGACGAAAACATCCTCGGCCACAAGTTGGAA  
TACAACATAACTCCCAACGTATACATCATGGCCGACAAGCAAAAGAACGGCATCAAAGCCAACCTCAA  
GACCCGCCACAACATCGAAGACGGCGGCGTGAACCTCGCTGATCATTATCAACAAAATACTCCAATTTGGCG  
ATGACCCTGTCTTTTACCAGACAACATTACCTGTCCACACAATCTGCCCTTTCNAAAGATCCCAACGAA  
AAGAGAGACCACATGGTCTTCTTGAGTTTGTAAAGGCTGCTGGGATTACACATGGCATGNNTGANCTATA  
CAAATAACTCGAGCTCTAGAACTATAGTGAGTCTGATTACGTANATCCAGACNTGATAAGATACATTTGAT  
GANNTTGGNCAANNANANTANAATGCAGTGAANNAATGCTTTANTTNGANNNTTNNGATGCTANTGNT  
TTANTNGNNNNNTANNAGCTGCAATANANCNAGTNNNNNNNNATGCATTCNNTTATGNTTCANNNCAGN  
NNGGNNNNNNNTNTTNNNNNNNNNNC  
NANGNNANTNNNNNTTNNNNNANNNNNN

- Colony 11, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNNNGNAANNGGNNATCTTNNTCGCGGCCGCGAATTA AAAAACCTCCACACCTCCCCCT  
GAANNNNNNNNNNNAAATGAATGCAATTGTTGTTGTTAACTTGTATTATTGNAGCTTATAATGGTTACAAAT  
AAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTTGTCCAAA  
CTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTAT  
TTGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAACTCAAGAAGGACCATGTGGTCTCT  
CTTTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGT  
CATCGCCNATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCCGCTCTTCGATGTTGTGGCGG  
GTCTTGAAGTTGGCTTTGATGCCGTTCTTTTGCTTGTCGGCCATGATGTATACGTTGTGGGAGTTGTAGTT  
GTATTC AACTTGTGGCCGAGGATGTTTCCGTCCTCCTTGAAATCGATTCCCTTAAGCTCGATCCTGTTGA  
CGAGGGTGTCTCCCTCAAACCTGACTTCAGCACGTGCTTTGTAGTTCCCGTCGTCCTTGAAGAAGATGGTC  
CTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTGAAGAAGTCGTGCCGCTTCATATGATCTGNATATCT  
TGAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGNNACNGGTAGTTTTCCAGTAG  
TGCAATAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCTCTCCACTGACAGAAATTTGT  
ACCCATTAACATCACCATCTAATTCAACAAGAANTNNNNANTNNNNGAAAAGTNNNTTCNNCTTTACTCATG  
GTGGNTGTGNCAAAAAACAGANTTNNANNCNNTNNNANCTNCNANGGNTTTGAANTCTGTANNNNAT  
GNGCTGNCNAAGNNNNNCANNNNGGNNNCNCANNNANTNANNCCNGANNANGNCTGTANNNNNNNCN  
NNNTNNNNNCNANNTNANNNNNNNNNNTNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNCNGNNCNNNNNNNNNN

- Colony 12, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNNNGNTATCTCGAGNGCTTCAAGGTCGGGCAGGAAGAGGGCCTATGTGTACAGAC  
ATTTNNNTNNNNGGGCTTTAAATAGCTGGTGGTGGAAACCAATATTCGTGCTCGCTTCGGCAGCACATATAC  
TAGCAGAGTTCAAAGCCCTTCAGGAAGCTTGCTGAAGGGCTTTTGAACCTCTGTTTTTTGGATCCACAGCC  
ACCATGAGTAAAGGAGAAGAACTTTCACTGGAGTTGTCCCAATCTTGTTGAATTAGATGGTGTATGTTAA  
TGGGTACAAATTTCTGTCACTGGAGAGGGTGAAGGTGATGCAACATACGAAAACCTTACCCTTAAATTTA  
TTTGCCTACTGGAAAACCTACCTGTTCCATGGCCAACACTTGTCACTACTCTCACTTATGGTGTCAATGC  
TTTTCAAGATATCCAGATCATATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCA  
GGAGAGGACCATCTTCTTCAAGGACGACGGGAACATAAGACACGTGCTGAAGTCAAGTTTGGGGAGACA  
CCCTCGTCAACAGGATCGAGCTTAAGGGAATCGATTTCAAGGAGGACGAAAACATCCTCGGCCACAAGTTG  
GAATACAAC TACAACCTCCACAACGTATACATCATGGCCGACAAGCAAAAGAAGGCATCAAAGCCAACCTT  
CAAGACCCGCCACAACATCGAAGACGGCGGCGTCAACTCGCTGATCATTATCAACAAAATACTCCAATTG  
GCGATGACCTGTCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAAC  
GAAAAGAGAGACCACATGGTCTTCTTGAGTTTGTAAACGGCTGCTGGGATTACACATGGCATGGATGAACT  
ATACAAATAACTCNAGCCTCTAGAACTATANTGAGTCGTATTACGTANATCCAGACATGATAAGATACNTT  
GATGANTTTGGACAAACNCACTNNNANGCAGTGNAANAAATGCCTTNNTTTTGNNAANNNTNGNNGATGCT  
NTNNNTTANNNGNANCATNNAAGCNCANTNAAANNANNTTNNNANNANNNNGNANTNCNNTTNNNNNT  
TCANGNNCAGNNNNNGNNNNNNNGNTTTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NTNNNNNNNTTANNGNNGNNAANNNNNNNNNNNNNNNNNNNNNNN

- Colony 12, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNNNNNNNNNGNNNANNGGNTANTCTTAAGTCGCGGCCGCGAATNAAAAACCTCCCACACCTCCC  
CCTGANNNNNNNNNNNNNATGAATGCAATTGTTGTTGTTAACTTGT'TTNNNNCAGCTNATAATGGTTACAAA  
TAAAGCAATAGCATCACAAATTCACAAATAAAGCATT'TTTTTCACTGCATTCTAGTTGTGGTTTTGTCCAA  
ACTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTA  
TTTGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCTC  
TCTTTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGG  
TCATCGCNAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCCGCTCTCGATGTTGTGGCG  
GGCTTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTGCGCCATGATGTATACGTTGTGGGAGTTGTAGT  
TGTATTCCAAC'TGTGGCCGAGGATGTTTCCGTCCTCCTTGAATCGATTCCCTTAAGCTCGATCCTGTTG  
ACGAGGGTGTCTCCCTCAAAC'TTGACTTCAGCACGTGTCTTGTAGTTCCCGTCGTCCTTGAAGAAGATGGT  
CCTCTCCTGCACGTATCCCTCAGGNTGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATC  
TTGAAAAGCATTTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGNCATGGAACAGGTAGTTTCCAGTA  
GTGCAAATAAATTTAAGGGTAAGTTTCCGTATGTTGCATCNACCTTCACCCCTCCTCACTGACAGAAAAT  
TTGTACCCATTAACATCACCATCTAATTCACAAGAAATTTGGGACAACTCCNGTGAAAAGTTCTTCTCNCNT  
TNACTCATGNNNNCTGTGNATCNAAAAACAGAGNTCAAANGCCCTTCNNNNAAGNNTTCTGNNNGGN  
TTTTNNAACTCTNGNNNNNNNANNNTNCTGNCNANNNNNNACNAANNNTNNNNNANNCNNCANNNNNA  
TTTNNANNCCNNNGNAANNANNNNNTNGNNNCNNNTNNNNCCNNNTTNNNNNNNNNNNTTNNNNNNNN  
NNANNNNCCNNNGNNNNNCNCNNNNANTNNCNAATTTNNCCNNNNANNNNNNNANNNNNNATANN

- Colony 15, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNNNNNNATCTCGAGNGCTTCAAGGTCGCGCAGGAAGAGGGCCTATGTGTACAGACATTTGT  
TCCAGGGGCTTTAAATAGCTGGTGGTGGAACTCAATATTCGTGCTCGCTTCGGTAGCACATATACTACAGA  
GTTCAAAGCCCTTCAGGAAGCTTGTCTGAAGGGCTTTTGAAC'TGT'TTTTTGTATCCACAGCCACCATGA  
TTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATCTTGTGTAATTAGATGGTGATGTTAATGGGTAC  
AAATTTTCTGTCTAGTGGAGAGGNTGAANNTGATGCAACATACGGAACCTTACCCTTAAATTTATTTGCAC  
TACTGGAAAAC'TACTGTTCCATGGCCAACACTTGTCACTACTCTCACTTATGGAGTTCAATGCTGTGCAA  
GATATCCNGATCATATGAANCNGCNGNCTTCTTCANNAGCGCCATGCCTGAGGGATACGTGCGNGAAAGGA  
CCATCTTCTTCAAGGACGACNGGAAC'TACAAGACACGTGCTGNGNCAAGTTTGGGGANACTCCCTCGTCA  
ACCGGAACGAGCTTAAGGNAATCNATTTCAATGAAGACCGAANCCTCCTCGACCAGANTTGAATACNNC  
TACNNCTCCCACANCNNANACNTCNNGNNNNACNNNNNAGNAACGGTCTCACNNCCNACTTCNAGACCCN  
CNNNNNCATCCAANACNGCNNNNNTGAAANTNGCTGATCATTGTCCACAAAATACTCGAATTGGNNNTGACC  
CTGANNTTTTACCANANCACCATAACCGGACCACAGANCTGCNCTTTCNAANNANCCAGACGAAAAGNN  
ANNNNAGNTGGTCNTTTCNNTGAGTNGTNACCNTGCTGNCATNNCACNTGNCNGGGANGAACTATANAAG  
TANCTCNAGNCNTANAACNATNCTGAGTGGTAATNCGNANATNCNNNACATGCANNAGATACATTGATGA  
GNNTNGGNAANNCCNNCTNNAANNNNNNNNNNNNNCAANTGCTNNANTTTGGGCAGNNNTNNNNNAN  
NNNANTAGNNTTNNNNNNNNNNCCNNNNNNANNNNTNNCCATNNANGNCCNNNNANNNNCNGTNNNGNN  
NNNNNACNGNNCGANNNNNNNNNNNNNNNNNNNGNNTTNNTNANTNNNNNNNNNNNNNNNNNNNTN  
NNNNNNNNNNNNNNNNNTCNNNNCNNNNNNNNNNNNANNTTNNNNNGNN

- Colony 15, the reverse sequence is given from 3' → 5'

```
NNNNNNNNNNNNNNNNNNNNGGGATAATCTTAAGTCGCGNNNGCGAATTNNNNCCCTCCNANACCTCCCC
TGANNNNNNNNNAGAAAATGAATGCAAGTGGTGGTGTAACTTGT'TNNTTGCAGCTTATAATGGTTACAAAT
AAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAA
CTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTAT
TTGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAACTCAAGAAGGACCATGTGGTCTCT
CTTTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGT
CATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCGTCTTCGATGTTGTGGCGG
GTCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTGCGCCATGATGTATACGTTGTGGGAGTTGTAGTT
GTATTCCAACTTGTGGCCGAGGATGTTTCCGTCCTCCTTGAATCGATTCCCTTAAGCTCGATCCTGTTGA
CGAGGGTGTCTCCCTCAAACCTGACTTCAGCACGTGCTTGTAGTTCCCGTCGTCCTTGAAGAAGATGGTC
CTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATCT
TGAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTAG
TGCAATAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCTCTCCACTGACAGAAAATTTG
TACCATTAAACATCACCATCTAATTCACAANAATTTGGGACAACCCNGTGAAAAGTTCTTCTCCTTTACT
CATGGNGGCTGTGNTCCAAAAAACANAGTTCAAAAGCCCTTCAGCAAGCTTCNTGAANGNNTTTTGAA
CTNNNNAGTATATGTGCTACCGAAANCNNNNNNNNNANNNNNNNTNCCACCNCANNCTATTNNAGCCCC
NNGNNNNNNANGNCNGTNNNNCNNNNNNNNCNNNNNNNNCANNNNANNTNNNNCNNNAANNNGNCNAA
NNGNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

2. The following are sequences from colonies with the pB12mcs-U6-shRNA3<sup>β3α2</sup>-eGFP plasmid.

- Colony 3, the forward sequence is given from 5' → 3'

(shRNA sequence was inconclusive)

```
NNNNNNNNNNNNNNNNNNNGNNATCTCGAGNGCTTCAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATT
CNNNNNNATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTNGAATTAATTTGACTGTAAGCACAAA
GATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTAAAAATTATGTT
TTAAAAATGGACTATCATATGCTTACCGTAACTTGAAGTATTTTCGATTTCTTGGCTTTATATATCTTGTGG
AAAGGACGAAGCACCGTGCTCGTTTCGGCAGCACATACTACGCAGAGTTCAAAGCCCGTGNNAAACCT
TGCCCCNTNNCTTTGCACTANGGANCATGGGGTCCCTCNTTTCGTTGGGATCTTTCGAAAAGGTTGACTG
TGTTGNCNNCCACGGTTGTCTGGTNNNNGGACAGGGTCATCTGGGATTGCATNATTTTGCAGATAATATG
CTGCGAGTTACGCGCCGCCCTCTCAATGATNTTAAAANTTCTTGAATTTGGTGTGATGCCNNNGNTTTG
NTTGCCGGCCNTGATGTANNCCCTGCTGGAGNTGTAGTTTAAATCCCACTAGCGGCNGANNATGTNACCGA
CCNCCTTGNCGTCGANACCCCTTATGCTCGAGCCAGGTGACNAGGGAGNCTCCATCANACTAGACTTCACAC
GAGTCTTGTGANTCCCGTCGAANTTGATGAAAAAGGCCNCCTCTGCTCATATCCTCANGCATGGCGCTTT
CTGNNAATCNGCCGCTACNNTGATCCGGACAANNTGAAAANCATNGNCCACNAGNANNNGNANNANTGN
TGAGNGNTNGNCCATGGNACAACGGNNNNNNCANNAGNATTTGNATATNNATTTANNGGAAANGNNTNCNN
NNAANTTNCNACTCNCCTTCTCCCCATCNCCTGAANGNNANANTGNANGCANNACCAGCCNNNTTGNCGNT
TNNACNNNAANTGGGNNCAACTCCGNNGGNNAN'TNNNTCANCC'TANNNNANGNNGGCNNGGNAANC
```

GACNNANGACNAATTTNCANANTCCCTTTCAGGANNNTCANNANANNCCATNANNACCNANTNNAANAA  
TATGNGCNCNTNNGNNNANGNNNNNGGNNNNNTNNTCCGNTNTNCCCNNNNANTNNTNCANANGCNN  
NCNNANCCNNCNATTTGNNNNNNNNNNNNNNNCNNNNNNNNCCNNNNNNNANNNNNNNNNNN  
TNTNACCNNNNNNANNNNNNNNTNANNNNNNNNNNNNNNNNNNNN

- Colony 3, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNNNAGGNAATCTTAAGTCGCGGCCGGAATAAAAACCTCCACACCTCCCCCT  
GAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTATGCAGCTTATAATGGTTACAAAT  
AAAGCAATAGCATCACAAATTCACAAATAAGCATTTTTTTCCTGACTGCATTCTAGTTGTGGTTTGTCCAAA  
CTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTAT  
TTGTATAGTTTCATCCATGCCATGTGTAATCCCAGCAGCCGTACAACTCAAGAAGGACCATGTGGTCTCT  
CTTTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGT  
CATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCCGCTCTTCGATGTTGTGGCGG  
GTCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTCGGCCATGATGTATACGTTGTGGGAGTTGTAGTT  
GTATTCCAAC TTGTGGCCGAGGATGTTTCCGTCCTCCTTGAATCGATTCCCTTAAGCTCGATCCTGTTGA  
CGAGGGTGTCTCCCTCAAAC TTGACTTCAGCACGTGTCTTGTAGTTCCCGTCGTCCTTGAAGAAGATGGTC  
CTCTCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATCT  
TGAAAAGCATTGAAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTAG  
TGCAATAAAATTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAATTTG  
TACCATTAAACATCACCATCTAATTCACAAGAATTTGGGACAACCCAGTGAAAAGTTCTTCTCCTTTACT  
CATGGNGGCTGTGGATCNAAAAAGCANNAGTTCAAAGCCCNTTCAGCANGCTTNN'TNNGGC'TTTGAAN  
TCTGCTANNANATGTGCTGNCGAANCNANCNNGNGCTCNNCNTTTCNNNAGANNNNNAANNANAAANN  
ANNACNTTNNAGNCCGNNNCNNA TGNNNNNCNNTNNNNNNNNNNNNANTGNNNNNNNNNNNNNNTTNN  
CNNNNNNNNNGNNANNAANNNTTNNNNNNNNNN

- Colony 6, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNNNTATCTCGAGANGCTTCAAGGTCGGGCAGGAAGAGGGCTATTTCCCATGATT  
CCTTCNTATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAGCACAAA  
GATATTAGTACAAAATACGTGACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTAAAAATATGTT  
TTAAAATGGACTATCATATGCTTACCGTAACTTGAAGTATTTTCGATTTCTTGGCTTTATATATCTTGTGG  
AAAGGACGAAGCACCGTGCTCGTTTCGGCAGCACATATACTAGCAGAGTTCAAAAAGCCCTTCAAGAACCTT  
GCTGAAGGGCTTTTGAACCTGGATTTNTGGATCCCCCTTTTCNTGAGATCTTTCGAAAAGCTTTACTGT  
GTGNACGGCCCATGTCTGGCTGGNNNAGGACAGGGTCATCGCNAATTTGGATNTNTGTTGATAATGATCAG  
CGAGTTATGCGCCGCCNCTTCGNNACCCTTAAATTTCTTGCATTTGGTGTGATGCCCTGTTTTGTTTGC  
CGGCCATGATGTANNTCTTGCCTGAGGTGTAGTTGGATTCCNACTAGTGGCCGATCATGTTACCGGCTCC  
NTGACTTCGATACNCTCANGCTCGATCCAGTNGANNAGGNGNCACCNTCNNCTTACNTCNACACGNGN  
NNTGNAATTTCCGCTGANNTTGANNANGATGNANNNCNCGNTCGTATCNCNTCAAGCNTGGNGCTNNNGA  
NNAANTCNGACGCNNNNNATNANCTNNTATCTTGAAGAAGCATTGAACACTCNAANTGAGANTAGTGANA  
NNNGNNNNNCATGNNAACGGNATNTNTCCNGANTTGAATANATNNNNAGNGTAAANTATCNNNATGTGC  
NANNCNCTNNCCCTCTCCACCTGACAGACNANTTGNCNNNTANCAGTCCNCTTTANNATTCANCNNA  
TTGGNACNAATTCNNTGACNATTTCTNCNANCTTNACTCAAGGNGGANNNGNNTCCANNANNTNANTN  
NNAANCCNTNNNAAGTCTCNTGNANGNATNANNNTTCNGCTNANNNNNTCNTNNNGAAANNN

ACTNGNNNNNTTCNNNNNTTNCNNNNNAAATNNTTNNNNNCNNNNNANNNNNNNNNNNNTNCCC  
GNTTANNNNNNNNNANNNNNNANNNNNNNNTTTNANNNNNNANNANTNNNNNNNNNGNNNNNNN  
NNNNNNNNNNNNNNNNAN

- Colony 6, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNGANNNGNNNTCTTAAGTCGCGGCCGGAATAAAAACCTCCACACCTCCCCCT  
GANNNNNNNNANAAAATGAATGCAATTGTTGTTGTTAACTTGTATTATTGNAGCTTATAATGGTTACAAATA  
AAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTCACATGCATTCTAGTTGTGGTTTGTCCAAAC  
TCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTATT  
TGTATAGTTTCATCCATGCCATGTGTAATCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCTCTC  
TTTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGTC  
ATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCCGCTCTTCGATGTTGTGGCGGG  
TCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTGCGCCATGATGTATACGTTGTGGGAGTTGTAGTTG  
TATTCCAACCTGTGGCCGAGGATGTTTCCGTCCTCTTGAATCGATTCCCTTAAGCTCGATCCTGTTGAC  
GAGGGTGTCTCCCTCAAACCTTGACTTCAGCACGTGCTTGTAGTTCCCGTCGTCCTTGAAGAAGATGGTCC  
TCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATCTT  
GAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTAGT  
GCAAATAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTCACCTCTCCACTGACAGAAAATTTGT  
ACCCATTAACATCACCATCTAATTCAACAAGAATTGGGANNACTCCNGTGAAAAGTTCTTCTCCTTACTC  
ATGGNGGCTGTGGATCCAAAAAGCAGAGTTCAAANNCNTTCANCAAGCTTNGAAGGGCTTTTNNACTC  
TGCTANNATATGTGCNNCCNAAACGNNNNNNNGGNNNTNCNTCTTNNNCNAGAATATNNNAAGNCNAANA  
NNCNANNNANTTNCNANNTTANNNGAANCNNNNNGNNNCNTTNNNANNNNNNNNNNNNNNNNNNNNA  
NTATNNNNNCNCCNNGNANTTNNNNNAAAAANNNNNN

- Colony 9, the forward sequence is given from 5' → 3'

(shRNA sequence was inconclusive)

NNNNNNNNNNNNNNNATCTCGAGAAGCTTCAAGGTCGGGCAGGAAGAGGGCTATTTCCCATGATTCTT  
TCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAGCACAAAGAT  
ATTAGTACAAAATACGTGACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTAAAAATTATGTTTA  
AAATGGACTATCATATGCTTACCGTAACTTGAAGTATTTTCGATTTCTTGGCTTTATATATCTTGTGGAAA  
GGACGAAGCACCGTGCTCGTTTCGGCAGCACATACTAGCAGAGTTCAAAAGCCCGTGAAGACCCCTTGCT  
GAGTGACTTTTGAAC TAAGGATTTTGGGGTCCCTCTTTTCCCTTGGGATCTTTCGAAAGGGTTTATTGTGTG  
GACGGGCCATGGTTGNCTGGTTAAAGGACAGGGTCATCGCCAATTGGAGTATTTTGTGATAATGATCAGC  
GAGTTGCACGCCCGCTCTTCGANGTTGTGNCGGGTCTTGAATTTGGCTTTGATGCCGTTCTTTTGTCTGC  
CGGCCATGATGTATACNTTGTGTGAGTTGTAGTTGTATTCCNACTTGTGGCCGAGGATGTTACCGTCTCC  
TTGACNTCGATACCCCTTANGCTCGATCCAGTTGACGAGGGNGTCTCCCTCANACTTGACTTCANACGNGN  
TNTGTNANTCCCGTCGACCTTGAANAGANGANNCTCCTGCNCGTATCCCTCAAGCATGGCGCTNTTGAN  
NAANNCGNGNCNTNCNNATGANCGNNANNTCTTGAAGCATTTGAACACTATAAGTGAGANNAGTGACA  
GNGTTGGCNCATGNACACGGTANNTTNNAGNATTTGCAAATANNNTAAAGNTAAGTNNNTCCNNANGTTN  
GCATCCCTTCTCCCTCTCCCTGACAGAANATTTGGACCCNNNNCGTCNCTTCTNATTCANCAATAAN  
TGGGACAACCTCCNGTGNNNAGTTNNNNNCNCTTTACTCAANNNGNNTGTGGNATNCNNNNATCGNAGNNNT



TNNAACGNCCTNNCAGGCNANTCNNNNNGAANNNTNNGANC'TTNGNNAANTNNNNNCNNNNAA  
NNNNANCANNNGNNNNNTNNNCNTNTNNNNNNNANNANNNTNNANNCNNNNANNCCNNNNNCNTNNNN  
NNNNNNNGNAANGNNANNNNNANNANNNTNNAANNNNNNNNNNNNANNNNNCNAANNNNNN  
NNNNNCNNNNNNNNNNAN

- Colony 9, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNGNNANNGNTANTCTTAAGTCGCGCCGCGAATTAACCACTCCACACCTCCCCCT  
GAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTGTTTATTGCAGCTTATAATGGTTACAAAT  
AAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAA  
CTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTAT  
TTGTATAGTTCAATCCATGCCATGTGTAATCCCAGCAGCCGTACAACTCAAGAAGGACCATGTGGTCTCT  
CTTTTCGTTGGGATCTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGT  
CATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCCGCTTTCGATGTTGTGGCGG  
GTCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGCGCCATGATGTATACGTTGTGGGAGTTGTAGTT  
GTATTTCAACTTTGTGGCCGAGGATGTTTCCGTCCTCCTTGAAATCGATTCCCTTAAGCTCGATCCTGTGGA  
CGAGGGTGTCTCCCTCAAACCTGACTTCAGCACGTGTCTTGTAGTTCCCGTCGTCCTGAAGAAGATGGTC  
CTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATCT  
TGAAAAGCATTTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTAG  
TGCAATAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAATTTG  
TACCCATTAACATCACCATCTAATTCACAAGAATTGGGACAACCTCCNGTGAAAAGTTCTTCTCCTTTACT  
CATGGTGGCTGTGGATCCAAAAAGCANNANNNTCAAAGCCCTTCAGCAAGCTTNCNGAANGGCTTTTGA  
ACTCTGCTAGNANATGNGCTNCNAAACNNNNACGGNCNTCNCNTTNNNANNNNNNNANNNNNNAANN  
NAANTANTTTCANNTACGNAAGCNNNNNNNNNCNTNNNANNNNNNNTAAACNNNNNNNNNNNTNNNT  
NNNNNNNNNNNNNNNNNNNNNCNNAANNNTNNNTNGGNNN

- The followings are the raw sequence data of all colonies sequenced for the shRNA constructs targeting the  $\beta 2\alpha 2$  fusion point and under the regulation of the U6 promoter.

1. The following are sequences from colonies with the pB12mcs-U6-shRNA1 <sup>$\beta 2\alpha 2$</sup> -eGFP plasmid

- Colony 7, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNATGCNCNGGCCGCCATGGCGCCGCGGAATTCGATTTGGAATTCAAGGTCGGGCA  
GGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTA  
GAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGG

TAGTTTGCAGTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTCGA  
TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAGCACCGTGCCTCGCTTCGGCAGCACATATACTAGCAA  
TAAGGAAGAAGCCCTTCAGAAGCTTGTGAAGGGCTTCTTCCTTATGTTTTTTGGATCCATAATCACTAGT  
GAATTCGCGGCCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATT  
CTATANNGNNCCTAANN

- Colony 7, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNCCANGCGTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGGAATTCAGT  
GATTATGGATCCAAAAACAATAAGGAAGAAGCCCTTCACAAGCTTCTGAAGGGCTTCTTCCTTATGTAG  
TATATGTGCTGCCGAAGCGAGCACGGTGCTTCGTCTTTCCACAAGATATATAAAGCCAAGAAATCGAAAT  
ACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTAAAACATAAATTTAAAACGCAAACTACCCAAGA  
AATTATTACTTTCTACGTACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCATATAT  
CTCTCTAACAGCCTTGTATCGTATATGCAGATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCGA  
CCTTGAATTCCAAATCGAATTCGCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGC  
CCTATAGTGAGNCCGTATNAA

- Colony 10, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNATGNNCGCCGCNATGGCGCGCGGAATTCGATTTGGAATTCAAGGTCGGCAG  
GAAGANNNNCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAG  
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTGGGT  
AGTTTGCAGTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTCGAT  
TTCTTGGCTTTATATATCTTGTGGAAAGAACGAAACACTGTGCTCGCTTCGGCAGCACATATACTAGCAAT  
AAGGAAGAAGCCCTTCAGAAGCTTGTGAAGGGCTTCTTCCTTATGTTTTTTGGATCCATAATCACTAGT  
AATTCGCGGCCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATTC  
TATANNGNNCCTANN

- Colony 10, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNANGCGTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGGAATTCAGT  
ATGGATCCAAAAACAATAAGGAAGAAGCCCTTCACAAGCTTCTGAAGGGCTTCTTCCTTATTGTAGTATA  
TGTGCTGCCGAAGCGAGCACAGTGTTCGTCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATCTT  
TCAAGTTACGGTAAGCATATGATAGTCCATTTAAAACATAAATTTAAAACGCAAACTACCCAAGAAAT  
ATTACTTTCTACGTACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCATATATCTCT  
CTAACAGCCTTGTATCGTATATGAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCGACCTT  
GAATTCAAATCGAATTCGCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTA  
TAGTGAGNCGTATNAN



- Colony 3, the reverse sequence is given from 3' → 5'

```
NNNNNNNNNNNNNNNNNNAGNNAATCTTAAGTCGCGGCCGGAATTA AAAAACCTCCACACCTCCCC
TGAANCNGAAACANAAAATGAATGCAATTGTTGTTGTTAACTTGTATTATGCAGCTTATAATGGTTACAAA
TAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTCACCTGCATTCTAGTTGTGGTTGTCCAA
ACTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTA
TTGTATAGTTTCCATGCCATGTGTAATCCCAGCAGCCGTTACAACTCAAGAAGGACCATGTGGTCTC
TCTTTTCGTTGGGATCTTTTCGAAAGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGG
TCATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCCGCTCTTCGATGTTGTGGCG
GGTCTTGAAGTTGGCTTTGATGCCGTTCTTTTGCCTGTGCGGCCATGATGTATACGTTGTGGGAGTTGAGT
TGATTCCAACTTGTGGCCGAGGATGTTTCCGCTCCTTGAATCGATTCCCTTAAGCTCGATCCGTGTG
ACGAGGGTGTCTCCCTCAAACCTTGACTTCAGCACGTGTCTTGTAGTTCCCGTCGCTTGAAGAAGATGGT
CCTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATC
TTGAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTTAGTTTCCAGT
ANTGCAAATAAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCNCCCTCTCCACTGACAGAAAATT
TGTANCCATTAACATCACCATCANNATTCAACAAGAATTGGGACAACCTCCAGTGAAAAGTTCTTCTNNTTAC
TCANGGNGGNGGGATCCAAAAAANANNNAANN
```

- The followings are the raw sequence data of all colonies sequenced for the shRNA constructs targeting the  $\beta 2\alpha 2$  fusion point and under the regulation of the H1 promoter.

1. The following are sequences from colonies with the pB12mcs-H1-shRNA1 <sup>$\beta 2\alpha 2$</sup> -eGFP plasmid. All sequences were missing the shRNA sequence.

- Colony 3, the forward sequence is given from 5' → 3'

```
NNNNNNNNNNNNNNNNNNATCTCGANNGCTTGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGN
NNNANNGNNNCTAGGCGGGAACACCCAGCGCGCTGCGCCCTGNNAGGAAGATGGCTGTGAGGGACAGGGG
AGTGGCGCCCTGCAATATTTGCATGTGCTATGTGTTCTGGGAAATCACCGTAAACGTGAAATGCTTTGG
ATTTGGGAATCTTATAAGTTCTGTATGAGACCACTCTTTGCGATCCACAGCCACCATGAGTAAAGGAGAAG
AACTTTTCACTGGAGTTGTCCCAATCTTGTGTAATTAGATGGTGATGTTAATGGGTACAAATTTTCTGTG
AGTGGAGAGGGTGAAGGTGATGCAACATACGGAACCTTACCTTAAATTTATTTGCACTACTGGAACCT
ACCTGTTCCATGGCCAACACTTGTCACTACTCTCACTTATGGTGTCAATGCTTTTCAAGATATCCAGATC
ATATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCCGAGGGATACGTGCAGGAGAGGACCATCTTCTTC
AAGGACGACGGGAAC TACAAGACACGTGCTGAAGTCAAGTTTGAGGGAGACACCCCTCGTCAACAGGATCGA
GCTTAAGGGAATCGATTTCAAGGAGGACGGAACATCCTCGGCCACAAGTTGGAATACAAC TACAACCTCC
ACAACGTATACATCATGGCCGACAAGCAAAAAGAACGGCATCAAAGCCAACCTTCAAGACCCGCCACAACATC
GAAGACGGCGCGTGCAACTCGCTGATCATTATCAACAAAATACTCCAATTGGCGATGACCCTGTCTTTT
ACCAGACAACCATACCTGTCCACAACTGTGCCCTTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGG
TCCTTCTTGTAGTTGTAACGGCTGCTGGGATTACACATGGNATGGNTGAACTATACAATAACTCGAGCCT
```

CTAGAACTATAGTGAGTCGTATTACGTAGATCCAGACATGANANNANNCATTGATGANNTTNGNNNNNNNA  
CNACTAGAATGCNNNGAAAAATGNTTTNNNTTGNNAANNTGNGATGCNNNGCNTTNNNTNNNAANCATN  
NANCNGCAAATNAANCANTNNNNNANNNNNNTNNNTTNNNNNTTNGNNNGGNNNNNNNNNNNNNNNTTT  
NNNNNCNNGGNNNCNNNNNTTNNN

- Colony 3, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNNNNNAGGGTAATCTTAAGTCGCGGCCGGAATTA AAAAACCTCCCACACCTCCCCCT  
GAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAAT  
AAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTTACTGCATTCTAGTTGTGGTTTTGTCCAAA  
CTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTAT  
TTGTATAGTTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAACTCAAGAAGGACCATGTGGTCTCT  
CTTTTCGTTGGGATCTTTGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGT  
CATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCCGCTCTTCGATGTTGTGGCGG  
GTCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTGCGCCATGATGTATACGTTGTGGGAGTTGTAGTT  
GTATTTCAAACCTGTGGCCGAGGATGTTTCCGTCCTCCTTGAATCGATTCCCTTAAGCTCGATCCTGTTGA  
CGAGGGTGTCTCCCTCAAACCTGACTTCAGCACGTGTCTTGTAGTTCCCGTCGTCCTGAAGAAGATGGTC  
CTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATCT  
TGAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTAG  
TGCAATAAAATTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAATTTG  
TACCATTAAACATCACCATCTAATTCACAAGAATTTGGGACAACCCAGTGAAAAGTTCTTCTCCTTTACT  
CATGGNGGCTGTGGATCCGAAAGANTGGNNTCATA CAGAACTTANANANNCCAAATNCNAAAGACNTTNN  
ACGTTACGGNGATTTNCCAGAACNNTANCGACATGCAAATNTGNNGGNNCCNNNTNCCNNNNCCNNNNAN  
NNNNNNNNNNNNCGNNNNNNNNCNGNNNNNNNGC  
NNNNNNNNNNNNNNNNNNNNNANGNNNNAN  
NACNNCNNNTNCANNNNNNNNN

- Colony 4, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNNNGGNNATCTCGA **GAAGCTTGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGC**  
**GGGCCAGTGTCACTAGGCGGGAACACCCAGCGCGCTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAG**  
**GGGAGTGGCGCCCTGCAATATTTGCATGTGCTATGTGTTCTGGGAAATCACCGTAAACGTGAAATGTCTT**  
**TGGATTTGGGAATCTTATAAGTTCTGTATGAGACCACTCTTT**CGGATCCACAGCCACCATGAGTAAAGGAG  
AAGAACTTTTCACTGGAGTTGTCCCAATTTCTTGTGAATTAGATGGTGATGTTAATGGGTACAAATTTTCT  
GTCAGTGGAGAGGGTGAAGGTGATGCAACATACGAAAACCTACCCTTAAATTTATTTGCACTACTGGAAA  
ACTACCTGTTCCATGGCCAACACTTGTCACTACTCTCACTTATGGTGTCAATGCTTTTCAAGATATCCAG  
ATCATATGAAGCGGCACGACTTCTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAGAGGACCATCTTC  
TTCAAGGACGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAGGGAGACACCCTCGTCAACAGGAT  
CGAGCTTAAGGAATCGATTTCAAGGAGGACGAAACATCCTCGGCCACAAGTTGGAATACAAC TACAAC  
CCCACAACGTATACATCATGGCCGACAAGCAAAAGAACGGCATCAAAGCCAAC TCAAGACCCGCCACAAC  
ATCGAAGACGGCGCGTCAACTCGCTGATCATTATCAACAAAATACTCCAATTGGCGATGACCCTGTCTCT  
TTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAAGATCCCAACGAAAAGAGAGACCACA  
TGGNCCTTCTTGAGTTTGTAAACGGCTGCTGGGATTACACATGGCATGGNTGAAC TATACAATAACTCGAG  
CCTCTAGAACTATANNGAGTCGTANTACGTANATCCNGACATGANNNNATACATGTAGTGTGTTNGNNNNN  
CCNCNNNANNGC  
NNNGAAANAAATGCTTTNNNNNNNA  
ANNNNNNTGC  
NNNNNTT  
NNNNNNNAC  
NTNNAN

CTGCANAANNNTACNNNNCNNNNNTNNTTTTNNNNNNNNNNNNNNNNNNNGNNNNNTTNNNNNNN  
NNNNNNNTANNNNN

- Colony 4, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNGNNNAGGNTAATCTTAAGTCGCGGCCGGAATAAAAACCTCCCACACCTCCCCCTG  
AACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTATGCAGCTTATAATGGTTACAAATA  
AAGCAATAGCATCACAATTTTCAAAATAAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTTGTCCAAAC  
TCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTATT  
TGATAGTTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAACTCAAGAAGGACCATGTGGTCTCTC  
TTTTCGTTGGGATCTTTTCAAAGGCGAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGTC  
ATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCCGCTCTTCGATGTTGTGGCGGG  
TCTTGAAGTTGGCTTTGATGCCGTCTTTTGTCTGTGGCCATGATGTATACGTTGTGGGAGTTGTAGTTG  
TATTCCAACCTGTGGCCGAGGATGTTCCGTCCTCTTCAAATCGATTCCCTTAAGCTCGATCCTGTTGAC  
GAGGGTGTCTCCCTCAAACCTTGACTTCAGCAGGTGTCTTGTAGTTCCCGTCGCTTGAAGAAGATGGTCC  
TCTCCTGCAGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTGTGCCGCTTCATATGATCTGGATATCTT  
GAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTAGT  
GCAAATAAATTAAGGGTAAGTTTCCGATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAATTTGT  
ACCCATTAACATCACCATCTAATCAACAAGAATTGGGACAACCTCAGTGAAGTTCTTCTCTTTACTC  
ATGGNNGCTGTGGNTCCGAAAGANNGNCTCATAACAAGTATATANATCCCAAATCCNAAGACATTTNNC  
GTNNGNNGATTNCCANAACNNNTANCGACATGCAATNTGCNGNGCCNNNNCCNNNNCCNNNNANNNNTN  
NN  
NNNNNNNNAANN

- Colony 5, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNGGNNATCTCGAGANGCTTGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGG  
CCCAGTGTCACTAGGCGGGAACACCCAGCGCGGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGG  
AGTGGCGCCCTGCAATATTTGCATGTGCTATGTTCTGGGAAATCACCGTAAACGTGAAATGCTTTGG  
ATTTGGGAATCTTATAAGTTCTGTATGAGACCCTCTTTCGGATCCACAGCCACCATGAGTAAAGGAGAAG  
AACTTTTCACTGGAGTTGTCCAAATCTTGTGTAATTAGATGGTGATGTTAATGGGTACAAATTTTCTGTG  
AGTGGAGAGGGTGAAGGTGATGCAACATACGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACT  
ACCTGTTCCATGGCCAACACTTGTCACTACTCTCACTTATGGTGTCAATGCTTTTCAAGATATCCAGATC  
ATATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCTGAGGGATACGTGCAGGAGAGGACCATCTTCTTC  
AAGGACGACGGGAACACAAGACAGTGTGTAAGTCAAGTTTGAAGGAGACACCCCTCGTCAACAGGATCGA  
GCTTAAGGGAATCGATTTCAAGGAGGACGGAACATCCTCGGCCACAAGTTGGAATACAACATAACTCCC  
ACAACGTATACATCATGGCCGACAAGCAAAGAACGGCATCAAAGCCAACCTTCAAGACCCGCCACAACATC  
GAAGACGGCGGCTGCAACTCGCTGATCATTATCAACAAAATCTCCAATTGGCGATGACCCTGTCTTTT  
ACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTTCAAAGATCCCAACGAAAAGAGAGANCACATGG  
TCCTTCTTGAGTTTGTAAAGGCTGTGGGATTACACATGGCATGNTGAACTATACANTAACTCGAGCCTCT  
AGAACTATAGTGAGTCGTATTACGTAGATCCAGANATGANANNTACATTTGATGANTNNGNNNNNNACANT  
ANNNGCANNNAAAAAATGNNNNAATNNAANNNNNNTGCTNTGNTTNNNTNACNTNNNANCTGCANNA  
NCANNNNNCNCNN  
NNNTANNATNN

- Colony 5, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNNNACNGGNNNNCTTAAGTCGCGGCCGCGAATTAAAAACCTCCCACACCTCCCCCTG  
AACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTATGTCAGCTTATAATGGTTACAAATA  
AAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTTGTCCAAAC  
TCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTATT  
TGTATAGTTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAACTCAAGAAGGACCATGTGGTCTCTC  
TTTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGTC  
ATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCGTCTTCGATGTTGTGGCGGG  
TCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTCGCCATGATGTATACGTTGTGGGAGTTGTAGTTG  
TATTCCAACCTGTGGCCGAGGATGTTTCCGTCCTCCTTCAAATCGATTCCCTTAAGCTCGATCCTGTTGAC  
GAGGGTGTCTCCCTCAAACCTTGACTTCAGCACGTGCTTGTAGTTCCCGTCGTCCTTGAAGAAGATGGTCC  
TCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATCTT  
GAAAAGCATTTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTAGT  
GCAAATAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCTCTCCACTGACAGAAAATTTGT  
ACCCATTAACATCACCATCTAATTTCAACAAGAAATGGGACAACCTCCAGTGAAGGTTCTTCTCTTTACTC  
ATGGNGGCTGNNNNCCGAAAGNNNGNCTCATAAGAACTTATANANTCCNAANNCAAGACATTTAC  
GTTACGNGATTTNCCANANNNATANGACATGCAANTATGCNNNNCCNNNTCCNNNTCCNNNNNNNTN  
TNNNNNGNNNNNNNNNNNGGNNNNCGNCNNNNNNNNNNNNNNNTNNNNNNNGNNNNATNANNNCAG  
NNNNCANNNTTNNNNNN

- Colony 6, the forward sequence is given from 5' → 3'

NNNNNNNNNGNNNNNGNNATCTCGANNCTTGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGG  
GCCAGTGTCACTAGGCGGGAACACCCAGCGCGCTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGG  
GAGTGGCGCCCTGCAATATTTGTCATGTGCTATGTGTTCTGGGAAATCACCGTAAACGTGAAATGTCTTTG  
GATTTGGGAATCTTATAAGTTCTGTATGAGACCCTCTTTCCGATCCACAGCCACCATGAGTAAAGGAGAA  
GAACTTTTCACTGGAGTTGTCCCAATTCTTGTGTAATTAGATGGTGATGTTAATGGGTACAAATTTTCTGT  
CAGTGGAGAGGGTGAAGGTGATGCAACATACGAAAACCTTACCCTTAAATTTATTTGCACTACTGAAAAC  
TACCTGTTCCATGGCCAACACTTGTCACTACTCTCACTTATGGTGTCAATGCTTTTCAAGATATCCAGAT  
CATATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAGAGGACCATCTTCTT  
CAAGGACGACGGGAACATAAGACACGTGCTGAAGTCAAGTTTGGAGGAGACACCCCTCGTCAACAGGATCG  
AGCTTAAGGGAATCGATTTCAAGGAGGACGAAAACATCCTCGGCCACAAGTTGGAATACAACATAACTCC  
CACAACGTATACATCATGGCCGACAAGCAAAAGAACGGCATCAAAGCCAACTTCAAGACCCGCCACAACAT  
CGAAGACGGCGCGTGAACCTCGCTGATCATTATCAACAAAATACTCCAATTGGCGATGACCCTGTCTTTT  
TACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTTCGAAAGATCCCAACGAAAAGAGAGACCACATG  
GTCTTTCTTGTAGTTTGTAAACGGCTGCTGGGATTACACATGGNATGGNNGAACTATACAAATAACTCGAGCC  
TCTAGAACTATANNAGTCGTANTACGTAGATCCNGACNTGATANNNNCNTNGATGANTTTNGNNNNNNNN  
NNAGANGCNNNAAAAAATGNTTNANTNNNAAATNNNNNTGCNNTGNTTTNNNTNGNANNTNNNNCTGCAN  
TAANNANNTNNANNNCNNNNNNNTTNNNTNNNNNNNNNNNNNGNNNNNTTNNNNNCNNNNNNNNNT  
TANNNNNNNN

- Colony 6, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNNNANNGGNNCTTAAGTCGGCGCCGCGAATTA AAAAACCTCCCACACCTCCCCTG  
 AACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTATGCAGCTTATAATGGTTACAAATA  
 AAGCAATAGCATCACA AATTCACAAATAAAGCATTTTTTTCAC TGCATTCTAGTTGTGGTTTGTCCAAAC  
 TCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTATT  
 TGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAACTCAAGAAGGACCATGTGGTCTCTC  
 TTTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGTC  
 ATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCCGCTTTCGATGTTGTGGCGGG  
 TCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTCGGCCATGATGTATACGTTGTGGGAGTTGTAGTTG  
 TATTCCAAC TTTGTGGCCGAGGATGTTTCCGTCCTCCTTGAAATCGATTCCCTTAAGCTCGATCCTGTTGAC  
 GAGGGTGTCTCCCTCAAAC TTAGCTTCAGCAGGTGTCTTGTAGTTCCCGTCGTCCTTGAAAGAAGATGGTCC  
 TCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCTGCCGCTTCATATGATCTGGATATCTT  
 GAAAAGCATTTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTCCAGTAGT  
 GCAAATAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCTCTCCACTGACAGAAAATTTGT  
 ACCCATTAACATCACCATCTAATTCACAAGAATTTGGGACAAC TCCAGTGAAAAGTTCTTCTCCTTTACTC  
 ATGGNGNTGTGGATCCGAAAGAGTNNNTCATA CAGA ACTTATANANTCCCAAANN CNAANNANATTTNAC  
 GTTACGNGATTTCCAGAACACNTANCGACATGCAANNNTGNNNNNNCCNNNNCCNNNNCCTCNNANNT  
 NNNNNNNGNNC NNNNNNCGNNNNNCNNNC NNNNNNNNNNNNNNNNTNNNNNNNNNGNNNNNANNCA  
 GNNNNNTCAANANNNNNN

- Colony 7, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNGGNNATCTCGANNGCTT **GANGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCC**  
**CAGTGTCACTAGGCGGGAACCCAGCGCGCTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAG**  
**TGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCGTAAACGTGAAATGTCTTTGGAT**  
**TTGGGAATCTTATAAGTTCTGTATGAGACCCTCTTT**CGGATCCACAGCCACCATGAGTAAAGGAGAAGAA  
 CTTTTCACTGGAGTTGTCCCAATCTTGTGTAATTAGATGGTGATGTTAATGGGTACAAATTTTCTGTGAC  
 TGGAGAGGGTGAAGGTGATGCAACATACGAAA ACTTACCCTTAAATTTATTTGCACTACTGAAA ACTAC  
 CTGTTCCATGGCCAACACTTGTCACTACTCTCACTTATGGTGTCAATGCTTTTCAAGATATCCAGATCAT  
 ATGAAGCGGCACGACTTCTTCAAGAGCGCATGCTGAGGGATACGTGCAGGAGAGGACCATCTTCTTCAA  
 GGACGACGGGA ACTACAAGACACGTGCTGAAGTCAAGTTTGAGGGAGACACCCCTCGTCAACAGGATCGAGC  
 TTAAGGGAATCGATTTCAAGGAGGACGGAACATCCTCGGCCACAAGTTGGAATACA ACTACA ACTCCCAC  
 AACGTATACATCATGGCCGACAAGCAAAGAACGGCATCAAAGCCA ACTTCAAGACCCGCCACAACATCGA  
 AGACGGCGGGT GCAACTCGCTGATCATTATCAACAAAATACTCCAATTGGCGATGACCCTGTCTTTTAC  
 CAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCAACGAAAAGAGAGACCACATGGNC  
 CTCTTGAGTTGTAAACGGCTGCTGGGNTTACACATGGNATGGNTGAACTATACNAATAACTCGAGCCTCT  
 AGAACTATNNGAGTCGTANTACGTANATCCAGACATGATANNNNNATTGATGANTNNNNNNNNNNANNAN  
 NNNCANNNNAAAANANGNNTTNATTNNGAAATNGNGTGCNNTGCTTNATNNNNNNNTNNNANCTGCAATA  
 AANCNNGNNANNNNNNCNNNNNTCNTNNNNNNNNNNNNNNNNNNNNNGNNNNNNNTTNANNNTCNGNN  
 NNNNNNNNNNN



- Colony 7, the reverse sequence is given from 3' → 5'

NNNNNNNNNANNNNGGNNACNGGNTATCTTAAGTCGCGGCCGGAATTA AAAAACCTCCACACCTCCCC  
TGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTATGCAGCTTATAATGGTTACAAA  
TAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTTGTCCAA  
ACTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTA  
TTGTATAGTTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCTC  
TCTTTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGG  
TCATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCCGCTTTCGATGTTGTGGCG  
GGTCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTGCGCCATGATGTATACGTTGTGGGAGTTGTAGT  
TGTATTTCCAACCTGTGGCCGAGGATGTTTCCGTCCTCCTTGAAATCGATTCCCTTAAGCTCGATCCTGTTG  
ACGAGGGTGTCTCCCTCAAACCTGACTTCAGCACGTGTCTTGTAGTTCCCGTCGTCCTTGAAGAAGATGGT  
CCTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATC  
TTGAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTA  
GTGCAAATAAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAATTT  
GTACCCATTAACATCACCATCTAATTCAACAAGAATTGGGANANTCCAGTGAAAAGTTCTTCTCCTTACT  
CATGGNNGCTGTGGNTCCGAAAGNANTGGNCTCATACAGAACTTANANANTCCNNAATCCNNGACATTTCA  
CGNNNCGGNGATTNCCANANNNNNNANNGACNTGCAANNNTGCNNNNGNCNNNTCCNNNNCCNNNNNNNTN  
NTNNNNNNNNNNNNNNNNNNNGNNNNNCNNNNNNNNNNNNNNNNNNNCNNNNANGGNNANNNNNNCAGCN  
NNTNANNNNNNNNNNN

- Colony 8, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTCGAGNAGCTTGAACGCTGACGTCNTCNACACCATAAGTTCTGTA  
TGAGACCCTCTTTCCAATAAGGAAGAAGCCCTTCAGAAGCTTGAACGCTGACGTCATCAACCCGCTCCAA  
GGAATCGCGGGCCAGTGTCACTAGGCGGGAACCCAGCGCGCTGCGCCCTGGCAGGAAGATGGCTGTG  
AGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTGCTATGTGTTCTGGGAAATCACCGTAAACGTGA  
AATGTCCTTTGGATTTGGGAATCTTATAAGTTCTGTATGAGACCACTCTTTTCGGATCCACAGCCACCATGAG  
TAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATCTTGTGTAATTAGATGGTGATGTTAATGGGTACA  
AATTTTCTGTGAGTGGAGAGGGTGAAGGTGATGCAACATACGGA AAAACTTACCTTAAATTTATTTGCACT  
ACTGGA AAACTACCTGTTCCATGGCCAACACTTGTCACTACTCTCACTTATGGTGTTC AATGCTTTTCAAG  
ATATCCAGATCATATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAGAGGA  
CCATCTTTTCAAGGACGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAGGGAGACACCCCTGTC  
AACAGGATCGAGCTTAAGGGAATCGATTTCAAGGAGGACGGAACATCCTCGGCCACAAGTTGGAATACAA  
CTACNACTCCCAACGTATACATCATGGCCGACAAGCAAAAAGAACGGCATCAAAGCCAACTTCAAGACCC  
GCCACAACATCGAAGACGGCGGCTGCAACTCGCTGATCATTATCAACAAAATACTCCAATTGGCGATGAC  
CCTGTCTTTTACCAGACAACCATACCTGTCCACACAATCTGCCCTTTTCGAAAGATCCCNACGAAAAGAG  
AGACCACATGGNNCTTNNAGTTTGNAAACGGCNGCTGNNANNCNATGNNNTGNNANNATAACNAANNANTC  
GANCNTCTANANTNNANNNAGNCGNNTNCNTANANNCNNANNNNANANANNCNNNTNATNANNNNNNNNN  
NNTANNGCNANANNANGNNTNNTNNNNNNNNNNGNNTGNNNATNNNNCNTNANCNNNNNNNAANNANN

- Colony 8, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNNNNNANNGGGTANTCTTAAGTCGCGGCCGCGAATTA AAAAACCTCCCACACCTCCCC  
 TGAACCTGAAACATAAAATGAATGCAATTGTTGTGTGTTAACTTGTATTATTGCAGCTTATAATGGTTACAAA  
 TAAAGCAATAGCATCACAAATTCACAAATAAAGCATT TTTTTTCACTGCATTCTAGTTGTGGTTTTGTCCAA  
 ACTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTA  
 TTTGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCTC  
 TCTTTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGG  
 TCATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCCGCTCTCGATGTTGTGGCG  
 GGCTTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTGCGCCATGATGTATACGTTGTGGGAGTTGTAGT  
 TGTATTCCAAC TTGTGGCCGAGGATGTTTCCGTCCTCCTTGAAATCGATTCCCTTAAGCTCGATCCTGTTG  
 ACGAGGGTGTCTCCCTCAAAC TTGACTTCAGCAGCTGTCTTGTAGTTCCCGTCGTCCTTGAAAGAAGATGGT  
 CCTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATC  
 TTGAAAAGCATTTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTA  
 GTGCAAATAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCTCTCCACTGACAGAAAATTT  
 GTACCCATTAACATCACCATCTAATTTCAACAAGAAATGGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTAC  
 TCATGGNGGCTGTGGATCCGAAAGANNNGNCTCATACAGAACTNNNAGANTCCAAAATCCNAAGACATTTN  
 NCGTTTANGNNATTTCCANANNNNNAGCNACNTGCAANTNTGCNGGNNCNCCTCCNNTCTCNNANNTN  
 NTCNNNNNNNNNNNGNNNNNNNGNNNNNCCNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
 NNNNCAGNNNNNCANNNNN

The following sequences were inconclusive:

- Colony 9, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNTAGGCGGCCATGAGCTCGGGGCTCGAGTTATTCCAGTGCCANAATTC TTTCTTCA  
 TTCNTGGNTTTATATCTTTGCCTTTCAGGGCATCATATTAACATAACTTCAAATTTTCTTCTTTGTCT  
 TTCTTCTGCACTACTTTTTCGGTTTTTTTCTTTGCGCAGAAAATAGGATCCGGGGGTACCTCTTAATTA  
 CTGGCCTCATGGGCTTCCGCTCACTGCCCGCTTTCAGTCGGGAAACCTGTCTGTCAGCTGCATTAACA  
 TGGTCATAGCTGTTTCTTGCCTATTGGGCGCTCTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCG  
 TTCGGGTAAAGCCTGGGGTGCCTAATGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAGGCCGCGT  
 TGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGG  
 CGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCC  
 GACCCTGCCGCTTACCGGATACCTGTCCGCTTCTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCAC  
 GCTGTAGGTATCTCAGTTCCGTTAGGTCGTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTTCAG  
 CCCGACCGCTGCGCTTATCCGGTAACATCGTCTTGAGTCCAACCCGGAAGACACGACTTATCGCCACT  
 GGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGT  
 GGCTAACTACNGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA  
 AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAACACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCA  
 GCANATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTCTACGGGGTCTGACGCTCAGNG  
 GANCGAACTCACGTTNNGGATTTTGGNCATGAGATTATCAAANGATCTTCACCTANATCCTTTTAAAT  
 TAAAATGANNTNATCNNTCTANNNTATATANGANTAANNNNNNTGACNNNNNNNGCTNNTCNNNNGCAC  
 NNNNNCAGCNATNNNNATTCGNTCATCNTANNNNCNGNNNNCCNCCNCCNNGNNNANACTACGANNGGNNN



CTANATCCTTTTAAATTAAAATGANNTTNATCNNCTANNNATATATANGANTAANNNNNNTGACNNTNNNN  
NGCTNNTCNNNGCACNNNNNCAGCNATNNNNNATTCGNTCATCNTANNNNCNGNNNNCCCNNNCNGNNA  
NACTACGANNGGNNNGNNNANNNCNGNCCNNGNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNTNCCNNNNNN  
TNTNNNGCANNNANCCANNNNCNNNNNNNNNNNN

- Colony 10, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNNNNNTCTANATNCNANCTGCNGGCATGCGNNTTGGCNTAATCNGGCCATAGCTGT  
TTCCTGTGNNAGATTGTTATCCGCTCANATTCCACACAACATACGAGCCGGAATCNTAAAGTGGAAAGCCN  
GGAGTGCCTAATGAATGACTAACTCACATTAATTCGCTTGCCTCACTGCCCGCTTCCAGTCNGCAAACC  
TGTCNGCCANCTGCATTAATGAATCNGCCAAGGCGGGGAGAGCGGTTTGTAGTATTGNGCNCATCC  
GCTTCTCNCNACTGACTCNCCTGCGCTCGNGCGTTGNGCTGCGNAGAGCGGTATCGGGTCACTCANAGGC  
AGNAATACGGTTATCNAGAGAATCAGGGGATAACGCNGGAAANAACATGTGAACAAAAGGCCGNCNAAAGG  
CCANGACCCGTAAAAAGGCCGCGTTGCTGGCAGTTTGCCATAGGCTCCNNCCCTGTACNAGCATCANAGA  
AATCGACGTTNAGTCNAGGTGGTAAAGGGATTGGACTATAANNATACTAGGGTGTNCCTGGAAGCAC  
TCTCGCGCGCTCACCTGTTCCCACTCT

- Colony 11, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNGNTATCTCGANNGCTTGNNGCTGACGTCATCAACCCGNTCCAAGGAATCGCGGGCCC  
AGTGTCACTAGGCGGGAACACCAGCGCGCTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGT  
GGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCGTAAACGTGAAATGTCTTTGGATT  
TGGGAATCTTATAAGTTCTGTATGAGACCACCTTTTCGGATCCACAGCCGATGAGTAAAGGAGAAGAAC  
TTTTCACTGGAGTTGTCCCAATCTTGTGTAATTAGATGGTGTGTTAATGGGTACAAAATTTCTGTCACT  
GGAGAGGGTGAAGGTGATGCAACATACGAAAACCTTACCCTTAAATTTATTTGCACTACTGGAAAACCTACC  
TGTTCCATGGCCAACTTGTCACTACTCTCACTTATGGTGTCAATGCTTTTCAAGATATCCAAATCATA  
TGAAGCGGCACGACTTCTTCAAGAGCGCCNTGCCTGANGGATACGTGCAGGAGAGGACCATCTTCTTCAAG  
GACGACGGGAAC TACAAGACACGTGCTGAAGTCAAGTTTGAGGGAGACACCCTCGTCAACAGGATCGAGCT  
TAANGGAATCGATTTCAAGGAGGACGGANACATCCTCGGCCACAAGTTGGAATACAAC TACAAC TCCACA  
ACGTATACNTCATGGCCGACAAGCAAAGAACGGCATCAAAGCCAACTTCAAGACCCGCCACAACATCNAA  
GACGGCGCGGTGCAACTCGCTGATCATTATCNACANNATACTCCAATTGGCGATGACCCTGTCTTTTACC  
AGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCC  
TTCTTGAGTTTGTAAACNNNTGCTGGGATTACACATGNNTGNNTGAACTATACANATAACTCGAGCCTCTA  
GAACTATAGTGAGTCGNATTACGTANATCCAGANNGANANANACATTGATGAGTTTGGACNNNCNAACT  
AGAATGCANNGAANAATAATGCTTTATTTGNGAAATTTGTGANGCTANTGCTCTATTTGTANNNNTANANC  
TGCAANNANNCANNTTAANANNNNANTGCNTTCNTTTNANNNTTNNAGGTNNGGNNNNNGGNNNNNTTT  
TNNNTNNGNCCNACTANATACCNNNNANCNNNNNNNTCNNNNNTGNNCCNTNNNNNNNNNNNNNNNNNN  
NNCNNNNNTANCATNN  
NNNNNNNNNNNCNNGNNNGNNCTANNGN

- Colony 11, the reverse sequence is given from 3' → 5'

NNNNNNNNGNNANNGGNTATCTTAAGTCGCGGCCGGAATTA AAAAACCTCCCACACCTCCCCGTAACCT  
GNNNCNNNAATGAATGCAATTGTTGTTGTTAACTTGTGTTTATTGCANCTTATAATGGTTACAAATAAAGCA  
ATAGCATCACAAATTTACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTGTCCAAATCATC  
AATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTATTTGTAT  
AGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCTCTCTTTTC  
GTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGTCATCGC  
CAATTGGAGTATTTGTTGATAATGATCAGCGAGTTGCACGCCGCGTCTTCGATGTTGTGGCGGGTCTTG  
AAGTTGGCTTTGATGCCGTTCTTTTGCTTGTGCGCCATGATGTATACGTTGTGGGAGTTGTAGTTGTATTC  
CAACTTGTGGCCGAGGATGTTTCCGTCCTCCTTGAATCGATTCCCTTAAGCTCGATCCTGTTGACGAGGG  
TGCTCCTCAAACCTTGACTTCAGCACGTGTCTTGTAGTTCCCGTCGTCCTTGAAGAAGATGGTCCCTCC  
TGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATCTTAAAA  
GCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAAACAGGTAGTTTTCAGTAGTGAAA  
TAAATTTAAGGGTAAGTTTTCCGTATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAATTTGTACCCA  
TTAACATCACCATCTAATTTCAACAAGAATTGGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTACTCATGGT  
GGCTGTGGATCCGAAAGAGTGGTCTCATAACAGAACTTATAAGATTCCCAAATCCAAAGACATTTACAGTTT  
ACGGTGATTTCCAGAACACATAGCGACATGCAATATTGCNNGNNCCACTCCCCGTCCNTCACAGCCATC  
TTNCTGNCAGGGNGCACGCNNGCTGGGNGTTCCCGCCTANNGNNNCTGNNNGNGATTCCTNNNNCGNNN  
GATGACGTCAGCNTTCAGCTNNTCGANATTNNCCNNNNATCCCTANNACCCNNNNCCNNNAGNGANNCNN  
NANTACNNNNCGNCCNNNNNNNNNNNNNNNTTNNNNGNCNNGNANNGNNNNACCNGNNNNNTCCNANTNANCN  
NNNNNNNGNNNNNTCCNNNNNNNNNNNNNNNGNANNNNNCNNAANNNCNNNNCCNNNNAATNNNCCNTT  
CCNANN

- Colony 12, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNCTAGGCGGCCNTGAGCTCGGGGCTCGAGTTATTCCAGTGCCAAAATTTCTTCTTCA  
TCCNNGGNTNTATATCTTTGCCTTTTCAGGGCATCATATTAACATAACTTCAAAATTTTCTTCTTTGTC  
TTCTTCTGACCACTTTTTTCGGTTTTTTTCTTTGCGCAGAAAATAGGATCCGGGGGTACCTCTTAATTA  
CTGGCCTCATGGCCTTCCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAACA  
TGGTCATAGCTGTTTCTTGGCTATTGGGCGCTCTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTGCG  
TTCGGGTAAAGCTGGGGTGCCTAATGAGCAAAAGCCAGCAAAAGCCAGGAACCGTAAAAGGCCCGCT  
TGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGG  
CGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCC  
GACCCTGCCGCTTACCGGATACCTGTCCGCTTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCAC  
GCTGTAGGTATCTCAGTTCCGTTAGGTCGTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTTCAG  
CCCGACCGCTGCGCTTATCCGGTAACATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACT  
GGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGT  
GGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA  
AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCA  
GCANATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCANN  
GGAACGAAAACCTACGTTANGGATTTTGGNCATGAGATTATCAAAAAGGATCTTCNCNANATCCTTTTAA  
ATTAANNANNNTTAAATCAATCTAAAGTANATATGAGTAACTNGNCTGACNNTTACCNANGNTNNNCAN  
NGANGCACCTANNNANNNATNTGNCTANTTCCNNTCATCANNANNNCNNTNANNCCNNTNNNNNANANN

CNNACNANCCGGNNNNNNNNNNCTNNNNNNNNNNANNNNCNNNNNNNCNCNNNNNNNNNNNNNNNTN  
NNANNNNNNNNNNNNNNNNNANNGNNNNNNNN

- Colony 12, the reverse sequence is given from 3' → 5'

NNNNNCNNNNNNNNNNNNNNANNNNCAAGNCNGCGNNNNNNNAAGGCNACNNGNGACNGCTCGTCATGA  
GGGTGAGTNGANNGATTGTNNTGTTGNTGNAACCGANNNTGGNNNTGAAGNGGNAAAGAAAAGAATT  
AGAAGCACAGATTTNAGATGTAGCCCTTNNTTGTNNNGNGTTCNTNATTGGGNTTGNAAACCTTTTNGTGT  
NNCNTATCATGNCNCCAGCTNTCATGTACGACCTNNGCCTTGATGCCTTCNGNAGNTTCTTGGTATANTGA  
TTCATGTTACGTCGNGTNNCNTGNCCNCGTCTTACACCGTAATGACTGGGATGACCCTGTTTTCACTGGG  
ATTATTTCGANTTGCCGCACTTCGCGCTCTGGTCATGGTGCTNNGGNNTGAAGANNCCGCATCGGAACTCG  
CATTACANNAGNGGGACNGCCCGAAGGGTGAAGGNAGCGCNNNTNNNATGTTNTGGCGGGTCTATCACGGN  
CTATTCTNGCCGTTCTTATGCATGTCGCCGATGATGGNACAATGNTGGGANAAGGAGCTGANNNTCAATNG  
TTTACNACGANGTTTCATAACNTATTGANGTTTATTCCTTAGGCGCACTCTTGC TGACAATGTGCGCGGA  
ACCCCAATCTACTTCAGTNCGGGANTGTGTCCAACCAGNCNTCCNNNCANGAGACGNTACTCTCTGATAC  
NCGTCTCNNNANNNTGNNNTCTTAAANNNTANTNTNNNTTNTNNGATGTNNNNCTTTATNCCNCTATTT  
GNACAGNTAGTGTNCGTGNATGCACNNTCGCNACGGGNTGGNAGTTNNCANGCNGNNAACANNNGNG  
GNNNNATANNNGAANCNNCTGNATNTCANTATCNGNCTANCTTAAAATCGTCNCCNNAACAACATCCCAT  
CNNNANNATTGTGAGTCTCGCTAGAAGATCGNNTCATTAGTTCATGAGGNNNNNAANANNANAAGGNCTCA  
GANANC

- Colony 13, the forward sequence is given from 5' → 3'

NNNNNNNNNNNGNNTAGGCGGCCATGAGCTCGGGGCTCGAGTTATTCAGTGCCAAAATTTCTTCTTCAT  
CCATGGCTTTATATCTTTGCCTTTCAGGGCATCATCATTAAACATAAATTCAAATTTTCTTCTTTGTCT  
TCTTCTGCACCACTTTTTCGGTTTTTCTTTGCGCAGAAAATAGGATCCGGGGGTACCTCTTAATTAAT  
GGCCTCATGGGCCTTCCGCTCACTGCCCGCTTTCAGTCCGGAAACCTGTCTGCCAGCTGCATTAACATG  
GTCATAGCTGTTTCTTGCATATTGGGCGCTCCTCCGCTTCTCGTCACTGACTCGCTGCGCTCGGTGCTT  
CGGGTAAAGCCTGGGGTGCCTAATGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCCGCTG  
CTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCG  
AAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCTGTCCGA  
CCCTGCCGCTTACCGGATACCTGTCCGCTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCACGC  
TGANGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCGCTTACGC  
CGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCGGTAAGACACGACTTATCGCCACTGG  
CAGCAGCCACTGGTAAACAGGATTAGCANAGCGAGGTATGTANGCGGTGCTACAGAGTTCTTGAAGTGGTG  
CCTAACTACNGCTACNCTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCNGAAA  
AAGAGTTGGTAGCTCTTGATCCGGCAAACAAANNACCGCTGGTAGCGNNNNNTTTTGTGTTGCAAGNN  
CANANTANNCGNNAANAAAANNNTNNNANAANANCNTTTGANCTTTTCNANGGNTCNGANGNNNNNN  
GNNNAAACTCACGTTNNGGNNNNNNNNNNNNNNNTNNNNANNNNNCNNNNNCNTTTANNNNNNAAN  
NANTTNNNCNNNNNNNGNATNNNANNANNGNNNNNGTNCNNNNNNNNCNGNGNNGNNNNNNNNNNNTNN  
NNNTNNNNNNNNCNNNNNCNNNNNNNNCNNNNNNNGNNNNNNNNNGGNNNNNNNNCNNNCCNNNNNNNNAN  
CNNNNNNNNNNGNNNTNCNNNTN

- Colony 13, the reverse sequence is given from 3' → 5'

NNNNNTNNNGNNNNNNNNNNNNNNNNNAAGNCGCGGCCGANCANNNGNNNNNTNNNNCACCTCGGCA  
TGAAGCTGTNNNAAAAGNNNGAANGNANTGCTGTGTTANTTGTATTATGTCNGNNTGTNNAGGTTACAAATG  
AAGAAATAGAATCTNNAATTCAGAAAAAGTGCCTTTTTTTCNNGTANNCTNNTTNNGGNNTGTCNNNC  
TTNTCAATGGGNTTAANTGNCTGGAGCTACGTAGTACNACTCACGATNGGCCANAGGCTNCAGATATTTG  
TAATATTCATCCATGNTATGGGGCTCCCNACGCCGTTACAACTCCAGAANGACCGTGANNACTCTCTT  
GTCNTTGGGATCAATCGACATGGCACATTGTGCGGACACGTCAGGGTGTGTCAGGNAAGGAGAGGTTTAC  
CGCCNATTGGAGNATTTTGATGACAATGCCCGNCGGGTGGCAGGCCGCCAGGAATGTTGTGTCAACCC  
NNGAAGTTGGCTNTGATGNNNTCTTAAGCTNNTCGCCNATGATGCNNACNNGGAAAGAAGGACCTGCAT  
TCCNAAATGTGNNGAGGATGTTTCANNCTCTTGANACTAATTCCTTAAGCTCACTTCTGTGGACAAGGG  
GGGCCAATCCCACTTGACTTCNNNTGGANGTNGTCCCTCCNGCCGTCCATGTAGAANATGNTCCCCCT  
CGCACTATCCCTCANGNNTGNCGAATCTNGAANAANTAAGTGCCNCNTCNTNTGATCTGGACCTCTGNAA  
ANNTTTGNACNNGCTANCGNNANNGNNNNCENNAGTNNNNANGNNANAGNANNCTACAGNACGNNTAT  
AANTGAAGGACNAGTNTCNGATATGNNAAATGATCTTCACCC'TNTGAACTNTCCGGAANANNNGNACCC  
CTGACCNTCTCCAT'TNNATTGCANNNGGANTNGNGACNACNCNNGGANNNTGATCTCNTNANNNNNGN  
ANNCCGANNANANNNNANNNNGCNTCTCATNNANANNTTNCAGAANTGCCCTNNNTCENNACNTNNNACGN  
TNTCNNNNANNNCCTNNAGCCTGNNNNNNNNNNGNATCAT'TNNNGGTNGCCNCCNNNNNGTCCNTCNA  
NTNNNATCCATCTNTCANANGNNNNNAACGNACACNCTANGCT

- Colony 14, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNTAGGCGGCCNTGAGCTCGGGGCTCGAGTTATTCCAGTGCCAAAATTTCTTCTTC  
ATCCNNNNNNNTATATCTTTGCTTTTCAGGGCATCATCATTAAACATAACTTCAAATTTTCTTCTTGT  
CTTCTTCTGCACCACTTTTTTCGGTTTTTTTCTTTGCGCAGAAAATAGGATCCGGGGGTACCTCTTAATTA  
ACTGGCCTCATGGGCCTTCCGCTCACTGCCCGCTTTCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAAC  
ATGGTCATAGCTGTTTCTTGCCTATTGGGCGCTCTCCGCTTCCCTCGCTCACTGACTCGCTGCGCTCGGTC  
GTTCCGGTAAAGCCTGGGGTGCCTAATGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGG  
TTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTG  
GCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTC  
CGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCA  
CGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTCA  
GCCCCAGCGCTGCGCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACAGACTTATCGCCAC  
TGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGG  
TGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGG  
AAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGTGGTAGCGGTGGTTTTTTTGTGTTGCAAGC  
AGCANATTACGCGCAGAAAAAAAANGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAN  
TGGAACGAACTCACGTTAANGNATTTTGGNCATGAGATTATCAAAAAGGATCTTACCTAGATCNNTTT  
NAATTAANANGANTTTTNANTCANTCNNAAGTATATNNGANTAAANTNNCTGACNNTTACCNATGCNTAN  
NNNNAGCACNATNNNAGNATNNNNNNATNNNNATNNNNNNNGNCTGANCCCCGNNNNNANNNANNTA  
CNNNNCNGNNNNNNANANCNNNNCNNNNNNANNNNCNNNNCNNNNNNNACNNNNNNNNNNNNCA  
NNANNANCCNNNNNNNNANN

- Colony 14, the reverse sequence is given from 3' → 5'

NNNNNNNNNACNNNGNTCCNNNNNNNAAAGNNNGNGCGGAGCNNNANGNNNTGAGNCCNGGTAANNTGA  
GGGANNNNNGNNNNNTNCCNGGGNNAGNNGNNGNANGNNGNNNTNNAGAAGACAAAGAAGAAGATTGTG  
AAGTTNAGGTTAATGAGGATGCCCTGGAAGGCGAAGATAATAAANNCATGNANGACAANAATTTTGGGCAC  
TGGAATAAECTNAGCCCCNAGCTCATGGCGCGCTAGGCCTTGACGNNCTTCNGCENNATTCGCCCTATNNT  
GAGACGGATTACGTCGCGCTCACTGGCCGTCGTTNTACANCGCCGTGACNGGNAACCCCTGGCGTTACCN  
AACTTANCGACTTGAGCENATCCCCCTTCGCCANCTGGCNTAGAGCGAAGAGGCCCGCCGAAACNNNN  
NTCNCAACANNGCGCNGCCTGATGGGAATGGGAGCGCCCTGTAGCGGCCACTCGCCNGTCTCGGNTACN  
TTCGATTTNTAANGGATTTTGGCGATTTTCGGCATATTGNNACANAATGATCTGATTTANCNAAAANATAA  
CGCGAATTTTAAACGNANNNTANCGCTTACATTTANGTGGNACTCCACGGGAAAATGTGTGCGNGANCCCTA  
TCTGTCTATGATGTGCANTCCCCCTCCAAAATGTATACGCTCATGAAAACACTAANCCTGATAACTGTTTC  
GGATGATATTGAAAAGGAAGAGTAGGACGCACACAACATATCNGTGTNCCTGTANNCTTTCNTCCAGC  
ANGTTTGCCTTCCTGNTTTNGCNNNNNTAAANGNNNNNGNAAGTAGTANNGGTGANNATNANNNGTGC  
ACGAGANTGGNNATGNNANCATNCTANCCGNNNCNNCANNNNNANNNTNCCCNANNACNNNNNTGN  
NNNTGCACNTTTTGATTGNTTTGNCTTTTGNNNNNNNNTATNNCGTATTAGCNNNNNNNN

- Colony 15, the forward sequence is given from 5' → 3'

NNNNNNNNNANGNNNATCTCGAGAGCTTCTGAGGGCTTCTTCCCTTATTGGAAGAGTGGTCTCATAACAGAA  
CTTCAACCTCGTTGATGACGTCAGCGTTCACGCTTGAACGCTGACGTTCTCAATTTCTTCTTGGAACTCT  
GCCTGCAGCGTCTTTACGCGGTACCCCCAGCGCGAGAGCACCCCTGGGGGGAAGATGGCTGGGAGGGATGG  
GGAGTGGCGCCCTGCAATATTTGCCTGTCTCTATGTGGTCTGGGAAGTCGTGCCAAACGTGAAATGCTTTT  
GGATTTGGGAATCTTATAAGTCTGTATGAGACTCTTTCCGATCCACAGCCACCATGACTAAAGGAGAA  
GAACTTTTCCCGGGANGTGTCCGATTCTTGTGAATTAGATGGTGTGTTAATGNNACNAATTTTCTGT  
CGNGTNNNNNNANGNNNATGCCNCNTNCNGAAAACCTTACCCTTAAATTTATTTGCACTACTGGAAAA  
TACCTGTTCCATGGTCAACACTTGTGCTACTCTCACTGATGGCGTTCAATGCTTTTCTNGAGATCCNAAN  
CATATGAAGCGNNNCGACNTCTTCNANANCGCCATGCCTGANGNATACGTGCATGANAGGACCNTCTTCTT  
NNAGGACTACNNTAACTACAAGACACGTGCTGAAGTCNAGNNTGAGGGAGACACCCCTCGTCAACAGGATCN  
AGCTTACGGGAATCGATTTCAAGGAAGACNGAAACNTCCTCCGCCACAAGTTGGAATACCNCTACTGNTCC  
CACNCGTANACATCATGNCCGACAAACAAAAGAACGGCCTCNGAGCCACCTACNTGACCCGCCGCAACAT  
CNAATACGGCGGCNCGCAANTCGTGTATCNTTATCACNNAATNCTCCAGATGNCGATGACCCTGTCCNNTT  
ACCTGACAACCTNTACCTGNCCACANAATCTGCCCTTNNAAAGATCCCATCTAATAGATAGACACATGGNC  
NNNNNGAGNTNGAAACGGATCGCNGGGATACNCNTGGNNTGNNTNACCTATACAGANNCTCGANNCACTAN  
AACTNNANNGNAGNCGNANTACGNNCNTCCNGANNTGATAANANTNCCTTGANNNNNNNNNNNAANCNNNA  
ACNNNNANGCNNNNANNAACGCTTTATTTNTNAANTNNNNNTNNNNCNNNNNNNANTNGTNNNNNTNNN  
NNGCTGCNNNANNAGNTCANNNCATCATTGCNNTCNNNNNNNNNNCNGNNNCNNNNNNNNNGNNNNNNNA  
NNNNNNNNNNNNNNNNNNNNNNNGNNNNCNNNNNNNNCCNNNANNNNNNNNNTNNNNNTNNNNNNNGTNN  
NNNTANNNTNGCNNGANNANGNNNNNNNNNNCNNNNNNNNNN



- Colony 15, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNNNGGNATCTTAAGTCGCGGCCGGAATTAACCACTCCACACCTCCCCCTGAACC  
 TGANACNTAAATGAATGCAATTGTTGTTGTTAACTTGTATTATGCAGCTTATAATGGTTACAAATAAAGC  
 AATAGCATCACAAATTTACAAATAAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAACTCAT  
 CAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTATTTGTA  
 TAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAACTCAAGAAGGACCATGTGGTCTCTCTTTT  
 CGTTGGGATCTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGTCATCG  
 CCAATTGGAGTATTTGTTGATAATGATCAGCGAGTTGCACGCCCGCTTTCGATGTTGTGGCGGGTCTT  
 GAAGTTGGCTTTGATGCCGTTCTTTGCTTGTTCGCCATGATGTATACGTTGTGGGAGTTGTAGTTGTATT  
 CCAACTTGTGGCCGAGGATGTTCCGTCCTCCTTCAAATCGATTCCCTTAAGCTCGATCCTGTTGACGAGG  
 GTGTCTCCCTCAAACCTTGACTTCAGCACGTGCTTGTAGTTCCCGTCGTCCTTGAAGAAGATGGTCTCTC  
 CTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATCTTGAAA  
 AGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTCCAGTAGTGCAA  
 ATAAATTTAAGGTAAGTTTTCCGTATGTTGCATCACCTTCACCTCTCCACTGACAGAAAATTTGTACCC  
 ATTAACATCACCATCTAATTCAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTACTCATGG  
 TGGCTGTGGATCCGAAAGANTGTCTCATAACAGAACTTATAAGATTCCCAAATCCAAAGACATTTACGTTT  
 ACGGTGATTTCCAGAACNCNTAGCGACATGCAATATTGCAGGGCGCANNCTCCCTGTCCNTCACANCCA  
 TCTTNTGNCAGGNNNACGCNCGCTGGGNGNTCCGNNANNGANNCTGGNNNGCATCNNAGNNNNNTNGA  
 TGACGNCANNNCANNTGAANGCTGANNNCNNNNNNNNNNNNNNNNNNNNANACCNNNNNTTCATAGNN  
 NANCCNNNNNNNCNNCANNNTNNNCNNNNNNNNNNCCNNNNNNNANNNNNANNNNNNNNNNNNNNN  
 NCNATNNNCNNNNNNNNNGNNNNNNNNNTNNAANNNNGNNNNNNNNNNANNCNNNNN

2. The following are sequences from colonies with the pB12mcs-H1-shRNA<sup>β2α2</sup>-eGFP plasmid. All sequences were missing the shRNA sequence.

- Colony 9, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNGGNATCTCGAGAGCTTGANCGCTGACGTCATCAACCCGCTCCAAGGAATCGCGG  
 GCCAGTGTCACTAGGCGGGAACACCAGCGCGCTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGG  
 GAGTGGCGCCCTGCAATATCTGCATGTGCTATGTGTTCTGGGAAATCACCGTAAACGTGAAATGTCTTTG  
 GATTTGGGAATCTTATAAGTCTGTATGAGACCACTCTTCAATAAGGAAGAAGCCCTTCACGAAGCTTGG  
 TGAAGGGCTTCTTCCCTATTTTTTTTGGATCCACAGCCGATGAGTAAAGGAGAAAGAACTTTTCNCTGG  
 AGTTGTCCNATCTTGTGTAATTAGATGGTATGTTAATGGGGTACNAAATTTCTGTGAGGGGANAAGG  
 GNGAAGGAGAGGCAACAGTACGGGAAAACTTTACNNNAAATTTATTTTGCCTACTGAAAACCTACC  
 GTTCCATGGCCNANACTTGTCACTACTCTCACTNTATGGNTGTTCAATGCTTTTCAAGATATCCAGATCA  
 TATGAAGCGGCNCGACTTCTTCAAGAGCGCCNTGCCTGAANGGATACGTNCANGAGGACCATCTTCTTC  
 NNNGGACGACGNNNCTACAAGACACGNNNTGNNGNCNAGCTTTGAGNNNACANNCCGGTCAACNGGNT  
 CGANNTTAANNGGAATCGATTTNNNNGNANGACGNNNCNNNNNNNNNNANNNNNNNNNNNCTACANN  
 NNNNNNNNNCNCATGNNNGANNNGNNNAGNANNNGNNNTNNNNCNNNTNNNNNNCNNNTNNNNN  
 ANNNANNNNNNNNNNNNCNNNNNNNNNATCNNNNNANNNNNNNNNNNNGNNGATNANCCNNNNN  
 NTNNNNNNNNNNNNNNNNNNANNNNNNNCNNNNANNTCNGNCCNNNTNNNNN

- Colony 9, the reverse sequence is given from 3' → 5'

NNNNNNNNNTNNNNGNNNNAGGGTAATCTTAAGTCGCGGCCGGAATAAAAAACCCTCCACACCTCCCC  
CTGAACCTGAAACATAAAATGAATGCAATTGTTGTGTGTTAACTTGTATTATGCAGCTTATAATGGTTACAA  
ATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTCACCTGCATTCTAGTTGTGGTTTGTCCA  
AACTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTT  
ATTTGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCT  
CTCTTTTCGTTGGGATCTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGG  
GTCATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCCGCTTTCGATGTTGTGGC  
GGGTCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTGCGCCATGATGTATACGTTGTGGGAGTTGTAG  
TTGTATTCCAACCTGTGGCCGAGGATGTTTCCGTCCTCCTTGAATCGATTCCCTTAAGCTCGATCCTGTT  
GACGAGGGTGTCTCCCTCAAACCTTGACTTCAGCACGTGTCTTGTAGTTCCCGTCGTCCTTGAAGAAGATGG  
TCCTCTCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATAT  
CTTGAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGT  
AGTGCAAATAAATTAAGGGTAAGTTTTCCGTATGTTGCATCACCTTCACCTCTCCACTGACAGAAAATT  
TGTAACCATTAACATNNNNNTCAATTCACAAGAATTGGGACAACCTCAGTGAAAAGTTCTTCTCCTTTA  
CTCATGNNGCTGTGGATCCAAAAANNAAGGNAAGAAGCCNTCACCNAGCTTCGTGAANGGCTNCTNCT  
TANTGAAANANNGNNTCATAACNGAAGTTNTANANNNNNANTCNNGANNNTNACNNNNNNNATTCNNAN  
NNNANCNNNTGCANAANNTGNNGNNNCNNNNCNNNCNTCNNNNNCTTNNNNNNNCANNNGNCNNNT  
NNNNNCNNCANNNGNANNNGNNNNNNANNNCNNNNNNNGNNNN

- Colony 10, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNNNTATCTCGA GAAGCTTCGTGAAGGGCTTCTTCTTATTAAAAAGAGTGGTCTCA  
TACAGAACTTATAAGATTCCCAAATCCAAAGACATTCACGTTTACGGTGATTTCCAGAACACATAGCGA  
CATGCAGATATGCAGGGCGCCACTCCCCTGTCCCTCACAGCCATCTTCTGCCAGGGCGCACGCGCGCTG  
GGTGTTCGCGCTAGTGACACTGGGCCCGCATCCTTGGAGCGGGTTGATGACGTACGCGTTCAAGCTTG  
GTGAAGGGCTTCTTCTTATTTTTTTTGGATCCACAGCCACCATGAGTAAAGGAGAAGAACTTTTCACTGG  
AGTTGTCCCAATCTTGTGTTGAATTAGATGGTGATGTTAATGGGTACAAATTTTCTGTGAGTGGAGAGGGTG  
AAGGTGATGCAACATACGAAAACTTACCCTTAAATTTATTTGCACTACTGAAAACTACCTGTCCATGG  
CCAACACTTGTCACTACTCTCACTTATGGTGTTCATGCTTTTCAAGATATCCAGATCATATGAAGCGGCA  
CGACTTCTTCAAGAGCGCCATGCTGAGGGATACGTGCAGGAGAGGACCATCTTCTTCAAGGACGACGGGA  
ACTACAAGACACGTGCTGAAGTCAAGTTGAGGGAGACACCCCTCGTCAACAGGATCGAGCTTAAGGGAATC  
GATTTCAAGGAGGACGGAAACATCCTCGGCCACAAGTTGGAATACAAC TACAACCTCCACAACGTATACAT  
CATGGCCGACAAGCAAAGAACGGCATCAAAGCCAACCTTCAAGACCCGCCACAACATCGAANACGGCGCGG  
TGCAACTCGCTGATCATTATCACAAAATACTCCAATTGGCGATGACCTGTNCNTTTACCANANANCATTA  
CCTGTCCNCACATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATNNNNNNNAGTTNNACGGC  
TGCTGGNANNNNNATGNNNNNGANTATNCAANANTCNANCNNNANACTNNNTGNNTCNANTNCGTNNAT  
CNNANNGATANANNCNTNTNANNNNNNNNNNNACTNNNANGNNNNAAAANGNTNANTNNGANNNNNNNTG  
NNTNGTNTNTNNNTNANNNNNANNNNGNNANNNNNNNNNGNNNNNNNTNNGTNNGNCNGNNNNGGNNNN  
NTTNNNTNNGNCNACTNNNTNCCNNNNNN

- Colony 10, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNNNGNANNAGGGTAATCTTAAGTCGCGGCCGGAATTA AAAAACCTCCCACACCTCCCC  
CTGAACCTGAAACATAAAATGAATGCAATTGTTGTGTGTTAACTTGTATTATGCAGCTTATAATGGTTACAA  
ATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTCCTGCTTCTAGTTGTGGTTTTGTCCA  
AACTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTT  
ATTTGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCT  
CTCTTTTCGTTGGGATCTTTCGAAAGGGCAGATGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGG  
GTCATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCCGCTTTCGATGTTGTGGC  
GGGTCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTGCGCCATGATGTATACGTTGTGGGAGTTGTAG  
TTGTATTCCAACCTGTGGCCGAGGATGTTTCCGTCCTCCTTGAATCGATTCCCTTAAGCTCGATCCTGTT  
GACGAGGGTGTCTCCCTCAAACCTTGACTTCAGCACGTGTCTTGTAGTTCCCGTCGTCCTTGAAGAAGATGG  
TCCTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATAT  
CTTGAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGT  
AGTGCAAATAAATTTAAGGGTAAGTTTTCCGTATGTTGCATCACCTTCACCTCTCCACTGACAGAAAATT  
TGTAACCATTAACATCACCATNNNNNTCAACAAGAATTGGNACACTCCAGTGAAGAAGTNNCTNCTTTACT  
CATGGNGGNTGTNTCCAAAAAAAATANNNANAAGCCNTTCNCNAGCTTGAACGCTGACGTCATCANNCN  
GNTNAGNNCNCGNCCANNNTNCTNGNNNNNNNCCNNNNNGNNNNNNNCCNNNNGANANGNNTGNNNNNNN  
GGNANNNNNNNNCCNNCNGCNNNNCNNNNNNNNTNNGNAANNCCNNAANNNNNNNNNNNNNNNTNNNNANN  
TNNANNNNNNNNNNNNNNNCCNNNTTTNNNNNN

The following sequence was inconclusive:

- Colony 15, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNNNGGNATCTCGAGNGCTTCGTGAAGGGCTTCTTCTTATTA AAAAGAGTGGTCTC  
ATACAGAACTTATAAGATTCCCAAATCCAAAGACATTTACGTTTACGGTGATTTCCAGAACACATAGCC  
ACATGCAGATATTCAGGGCGCCACTCCCTGTCCCTCACAGCCATCTTCTGCCAGGGCGCACGCCGCT  
GGGTGTTCCCGCTAGTGACACTGGGCCCGGATTCCTTGGAGCGGGTTGATGACGTCAGCGTTCAAGCTT  
GGTGAAGGGCTTCTTCTTATTTTTTTTGGATCCACAGCCACCATGAGTAAAGGAGAAGAACTTTTCACTG  
GAGTTGTCCCAATCTTGTGTAATTAGATGGTGATGTTAATGGGTACAAATTTTCTGTGTCAGTGGAGAGGGT  
GAAGGTGATGCAACATACGGAAAACCTTACCCTTAAATTTATTTGCACTACTGGAAAACCTGTTCCATG  
GCCAACACTTGTCACTACTCTCACTTATGGTGTCAATGCTTTTCAAGATATCCAGATCATATGAAGCGGC  
ACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAGAGGACCATCTTCTTCAAGGACGACGGG  
AACTACAAGACACGTGCTGAAGTCAAGTTTGGAGGAGACACCCTCGTCAACAGGATCGAGCTTAAGGGAAT  
CGATTTCAAGGAGGACGGAACATCCTCGGCCACAAGTTGGAATACAACCTACAACCTCCACAACGTATACA  
TCATGGCCGACAAGCAAAAGAACGGCATCAAAGCCAACCTTCAAGACCCGCCACAACATCGAAGACGGCGGC  
GTGCAACTCGCTGATCATTATCAACAAAATACTCCAATTGGCGATGACCCTGTCTTTTACCAGACAACCA  
TTANNNGNACACAATCTGCCCTTTCGAAAGATCCCAACGAAAGAGAGACCACATGNNCCTTCTTGAGTT  
TGNAACGGCTGCTGGGATTACNCATGGCATGNTGNANTATNCNAANNNNNGAGNCTANANTATNNNGAG  
TCGTATNCGTANATCNGNNNTGATAGANNNTGNANNTNNNNNNNNNANCTNNNANGNNNNNAAAAANG  
NNNNNTNNNNNNNNNGNNNTNGNNNNNNGNACNNTNAGCNCANNACANNNNNNNNNNNNCCNNCCTNNNTNNN  
TNNNNNNNNNGGNANNNNNNNNNNNNTTTNNNN

- Colony 15, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNANGNANNAGGGTAATCTTAAGTCGCGGCCGCGAATTAATAAACCTCCCACACCTCCCC  
 TGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTATTGCAGCTTATAATGGTTACAAA  
 TAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTTGTCCAA  
 ACTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTA  
 TTTGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCTC  
 TCTTTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGG  
 TCATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCCGCTCTCGATGTTGTGGCG  
 GGCTTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTGCGCCATGATGTATACGTTGTGGGAGTTGTAGT  
 TGTATTCCAACCTGTGGCCGAGGATGTTTCCGTCCTCCTTGAATCGATTCCCTTAAGCTCGATCCTGTTG  
 ACGAGGGTGTCTCCCTCAAACCTTGACTTCAGCAGCTGTCTTGTAGTTCCCGTCGTCCTTGAAGAAGATGGT  
 CCTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATC  
 TTGAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTA  
 GTGCAAATAAATTTAAGGGTAAGTTTCCGTATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAATTT  
 GTACCCATTAACNNCACCATCTAATTCAACAAGAATTGGGACAACCTCCAGTGAAGAAGTNCCTTNTCCTTTAC  
 TCNTGGTGGCTGNGGNTCCAAAAAATAANNAAGAAGCCNNTTACCANCTTGAACGCTGACGTATCA  
 NNCGCTCNNTTGNNGGNCNNNNNCNNTAGNNGGNNCCNAGCGNNGNNGCCNNNNGNNNANGNTN  
 NNNNGNNNGNNGNNGNNNNNCANNNCTGCANNTCNCNNNNNTTNNNNNAANNNNNNNNNNNNNNNNN  
 NANNNNNNNNNANNNNNNNNTNANANCNNCNTNTNNNNNN

3. The following are sequences from colonies with the pB12mcs-H1-shRNA3<sup>β2α2</sup>-eGFP plasmid. All sequences were missing the shRNA sequence.

- Colony 3, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNNGNTATCTCGAGAAGCTTGAACGCTGACGTCATCAACCGCTCCAAGGAATCGC  
 GGGCCAGTGTCACTAGGCGGGAACACCAGCGCGCTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAG  
 GGGAGTGGCGCCCTGCAATATCTGCATGTGCTATGTGTTCTGGGAAATCACCGTAAACGTGAAATGTCTT  
 TGGATTTGGGGATCTTATAAGTTCTGTATGAGACCACTCTTTCCAATAAGGAAGAAGCCCTTACGGAAGCT  
 TGGTGAAGGGCTTCTTCTTATTGTTTTTGGATCCACAGCCACCATGAGTAAAGGAGAAGAACTTTTAC  
 TGGAGTTGTCCCAATCTTGTGTAATTAGATGGTATGTTAATGGGTACAAATTTCTGTGAGTGGAAAGG  
 GTGAAGGTGATGCAGCATAACGAAAACCTTACCCTTAAATTTATTTGCACTACTGGAAAACCTACCTGTTCCA  
 TGGCCAACACTTGTCACTACTCTCACTTATGGTGTCAATGCTTTTCAAGATATCCAGATCATATGAAGCG  
 GCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAGAGGACCATCTTCTTCAAGGACGACG  
 GGAACTACAAGACACGTGCTGAAGTCAAGTTTGGAGGAGACACCCTCGTCAACAGGATCGAGCTTAAAGGGA  
 ATCGATTTNAGGAGGACGGANACATCCTCGGNCACNAGTTGGAATACNACTACNACTCCCACAACGTATA  
 CATCATGGNCGACAAGCAAAAGAACGGCATCANAGCCAACCTTCAAGACCCGCCACAACATCGAAGACGGCG  
 GCGTGCAACTCGCTGATCATTATCAACAAAATACTCCAATTGGCGATGACCCTGTCTTNTTACCANNNNNN  
 ATTACCTGTCCNCACAATCTGCNCTTTGAAAGANCNNNACGANANANAGACNNNNNTGNNCNTTCTNTGAGT  
 GTACGGCTGCTGNNATTANNATGCNNGNATGANNTNNACNNNNNCTCGNNGCTCNAGACNNTAGNNAGN  
 NNNNNNACNNCNANCNNANNNNANNGANNNTTNTNTNNNNNNANNNNNNNNNANNNNNNNNAANNAN  
 NNNTNATNNNANNGGNNNGNNGNNTNNANNTNNANCTGCANNAANCAGNNNNNNNN

- Colony 3, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNNNNNNGGNTATCTCGAGACCCAAGCTTGAACGCTGACGTCATCAACCCGCTCCAAGG  
AATCGCGGGCCAGTGTCACTAGGCGGGAACACCCAGCGCGCTGCGCCCTGGCAGGAAGATGGCTGTGAG  
GGACAGGGGAGTGGCGCCCTGCAATATCTGCATGTCGCTATGTGTTCTGGGAAATCACCGTAAACGTGAAA  
TGCTTTTGGATTTGGGAATCTTATAAGTTCTGTATGAGACACTCTTTCCAATAAGGAAGAAGCCTTTCACG  
AAGCTTGGTGAAGGGCTTCTTCTTATTGTTTTTTGGATCCACAGCCACCATGAGTAAAGGAGAAGAACTT  
TTCAGTGGAGTTGTCCCAATCTTGTGAATTAGATGGTGTATGTTAATGGGTACAAAATTTCTGTGTCAGTGG  
AGAGGGTGAAGGTGATGCAACATACGGAAAACCTACCCCTTAAATTTATTTGCACTACTGGAAAACCTACCTG  
TTCCATGGCCAACACTTGTCACTACTCTCACTTATGGTGTTCATGCTTTTCAAGATATCCAGATCATATG  
AAGCGGCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAGAGGACCATCTTCTTCAAGGA  
CGACGGGAAC TACAAGACACGTGCTGAAGTCAAGTTTGGAGGAGACACCCCTCGTCAACAGGATCGAGCTTA  
ANGGAATCGATTTCAAGGAGGACGGAAACATCCTCGGCCACAAGTTGGAATACAAC TACTCCCACAACG  
TATACNTCATGNCGACAGCAAAGANGGCATCAAAGCCAACCTCAGACCCGCACACATCGANANGCGCGTGCA  
NTCGCTGANCATATCACAATACTCCATTGCNATGACCNNGNNNTNACAGANACNTNACTGNCCACACATCT  
GNCCTTTCNAANATCCNANNAANANAGACCNNATGNNCNCNTNNNNNNNACGNCTGCTGGNATNNNNNTG  
GCATGNNGACNANNNANNANNNTNNNCNNNTANANTNNNNNNNNCNCNTNACNNAATCNCNNNGNNANATNNNT  
NNGANNNNNNNNNCNCNANGNNNNNAANNNNNNNNNNNNNNNNNCNCNTNNNNNNNNNNNNNNNNNNNN  
NNNTNNNNNNNNNNNGNNNNNNNNNNNNNNNN

- Colony 5, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNANNNNNNNGGNNNNNCTCGAGACCCAAGCTTGAACGCTGACGTCATCAACCCGCTCCA  
AGGAATCGCGGGCCAGTGTCACTAGGCGGGAACACCCAGCGCGCTGCGCCCTGGCAGGAAGATGGCTGT  
GAGGGACAGGGGAGTGGCGCCCTGCAATATCTGCATGTCGCTATGTGTTCTGGGAAATCACCGTAAACGTG  
AAATGTCTTTGGATTTGGGAATCTTATAAGTTCTGTATGAGACACTCTTTCCAATAAGGAAGAAGCCTTTC  
ACGAAGCTTGGTGAAGGGCTTCTTCTTATTGTTTTTTGGATCCACAGCCACCATGAGTAAAGGAGAAGAA  
CTTTTCACTGGAGTTGTCCCAATCTTGTGAATTAGATGGTGTATGTTAATGGGTACAAAATTTCTGTGTCAG  
TGNAGAGGGTGAAGGTGATGCAACATACGGAAAACCTACCCCTTAAATTTATTTGCACTACTGGAAAAC TAC  
CTGTTCCATGGCCAACACTTGTCACTACTCTCACTTATGGTGTTCATGCTTTTCAAGATATCCAGATCAT  
ATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAGAGGACCATCTTCTTCAA  
GGACGACGGGAAC TACAAGACACGTGCTGAAGTCAAGTTTGGAGGAGACACCCCTCGTCAACAGGATCGAGC  
TTAAGGGAATCGATTTCAAGGAGGACGGAAACATCCTCGGCCACNAGTTGGAATACAAC TACTCCCACA  
ACGTATACATCATGGCCGACAGCAAAGANGCATCAAAGCCAACCTCNCANACCCGCACACATCGANACGNG  
CGTGNACTCGCTNANCATATCACAATACTCAATTGCGATGACCTGTCTTTTACNNACACCATAACNGTCC  
NNCATCTGNCNTNNNNNNCCACNAAAANAACNNNNGGNCNTNCTGANTTACGCTGCTGGNANNNNNNGC  
ATGNGACNNNCNATACTCNAGCCTCNCNANNNANNNNCNCNTNNNCNNNTCCGANTGNNNNNNNTGTGNN  
NNGNNNNNNNNANATGNNNNANAANNNNTATTNATNNNNNTTGNTNNNNNNANNNNNNNAAANNNAN  
NCNCNNCNCNNNTNTNTNNGTNNGGNNGNNGNNNNNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTNN  
NNNNNNNNNNNGNNNNNNNNNTNNNNNNNNNNNNNNANNGNANNANNNNNGNTNNNNNNNGCNCN  
NTNNNNNNANNNNNNNNNNNNNCNCNNNNAN

- Colony 5, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNNNNGGNANNAGGGTAATCTTAAGTCGCGGCCGCGAATTA AAAAACCTCCCACACCTCCCC  
 TGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTATTGCAGCTTATAATGGTTACAAA  
 TAAAGCAATAGCATCACAAATTCACAAATAAAGCATT TTTTTTCACTGCATTCTAGTTGTGGTTTTGTCCAA  
 ACTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTTA  
 TTTGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCTC  
 TCTTTTCGTTGGGATCTTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGG  
 TCATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCCGCTCTCGATGTTGTGGCG  
 GGCTTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTGCGCCATGATGTATACGTTGTGGGAGTTGTAGT  
 TGTATTCCAAC TTGTGGCCGAGGATGTTTCCGTCCTCCTTGAAATCGATTCCCTTAAGCTCGATCCTGTTG  
 ACGAGGGTGTCTCCCTCAAAC TTGACTTCAGCAGCTGTCTTGTAGTTCCCGTCGTCCTTGAAAGAAGATGGT  
 CCTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATATC  
 TTGAAAAGCATTTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGCCATGGAACAGGTAGTTTTCCAGTA  
 GTGCAAATAAATTTAAGGGTAAGTTTTCGATGTTGCATCACCTTCACCTCTCCACTGACAGAAAATNN  
 GTACCCATTAACATCACCATCTAATTCACAAGAAT TNNNNNCTCCAGTGAAGTTCTTCTCCTTTACTC  
 ATGGNGGCTGTGGATCCAAAAANANAAGNAAGAGCCNCTCNCNNTT CNTGAAAGNTTCTTCNNTNNNNN  
 NNGNCTCATAACGACTNNANATNCNNTANNNTNNNNNT TNNCGTTTACGNGATTNCCNANNNNNANNNNC  
 GNNNANTGNNNNNNNNNNNNNNNNNNNNNNNNNTNNNNNNNNNCTGNNNCNCNNNNNNNGNNNNNNNTCN  
 NNNNNNNN

- Colony 6, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNNNNGGNTATCTCGAGACCCAAGCTTGAACGCTGACGTCATCAACCCGCTCCAAGG  
 AATCGCGGGCCAGTGTCACTAGGCGGGAACACCCAGCGCGCTGCGCCCTGGCAGGAAGATGGCTGTGAG  
 GGACAGGGGAGTGGCGCCCTGCAATATCTGCATGTCGCTATGTGTTCTGGGAAATCACCGTAAACGTGAAA  
 TGTCTTTGGATTTGGGAATCTTATAAGTTCTGTATGAGACACTCTTTCCAATAAGGAAGAAGCCTTTCAG  
 AAGCTTGGTGAAGGGCTTCTTCTTATTGTTTTTTGGATCCACAGCCACCATGAGTAAAGGAGAAGAACTT  
 TTCCTGGAGTTGTCCCAATCTTGTGTAATTAGATGGTGATGTTAATGGGTACAAAATTTCTGTCTAGTGG  
 AGAGGGTGAAGGTGATGCAACATACGGAAAAC TTACCCTTAAATTTATTTGCACTACTGGAAAAC TACCTG  
 TTCCATGGCCAACTTGTCACTACTCTCAC TTATGGTGTTCAATGCTTTTCAAGATATCCAGATCATATG  
 AAGCGGCACGACTTCTTCAAGAGCGCCATGCC TGAGGGATACGTGCAGGAGAGGACCATCTTCTTCAAGGA  
 CGACGGGAAC TACAAGACACGTGCTGAAGTCAAGTTT GAGGGAGACACCTCGTCAACAGGATCGAGCTTA  
 ANGGAATCGATTTCAAGGAGGACGGAAACATCCTCGGCCACAAGTTGGAATACAAC TACACTCCCACAACG  
 TATACNTCATGNCGACAGCAAAGANGGCATCAAAGCCA ACTTCAGACCGCACACATCCGANANGCGCTGCA  
 NTCGCTGANCATATCACAACTCCATTGCNATGACCNGNNNTNACAGANACNTNACTGNCCACACATCT  
 GNCCTTTCNAANATCCNANNAAAAANAGACCNNATGNNC NNTNNNNNNNACGNC TGTGGNATNNNNNTG  
 GCATGNNGACNANNNANNANNTNNNCNNTANANTNNNNNNNNC NNTNCNNAATC NNNNGNNAATNNNT  
 NNNGANNNNNNNNNC NNNANGNNNNNAANNNNNNNNNNNNNNNNNNC NNTNNNNNNNNNNNNNNNNNN  
 NNNNTNNNNNNNNNNNGNNNNNNNNNNNNNNNNN

- Colony 6, the reverse sequence is given from 3' → 5'

NNNNNNNNNNNTNTANGNNNNAGGGTAATCTTAAGTCGCGGCCGCGAATTAAAAACCTCCCACACCTCCC  
CCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACA  
AATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTTCACTGCATTCCTAGTTGTGGTTTGTCC  
AAACTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGT  
TATTTGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTC  
TCTCTTTTCGTTGGGATCTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAG  
GGTCATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCGCCGCTCTTCGATGTGTGG  
CGGGTCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGTCCGCCATGATGTATACGTTGTGGGAGTTGTA  
GTTGTATTCCAACCTGTGGCCGAGGATGTTTCCGTCCTCCTTGAAATCGATTCCCTTAAGCTCGATCCTGT  
TGACGAGGGTGTCTCCCTCAAACCTTGACTTCAGCACGTGTCTTGTAGTTCCCGTCGTCCTTGAAGAAGATG  
GTCCTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATA  
TCTTGAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGTGTGGNCATGGAACNGTAGTTTTCCAGT  
ANTGCAAATAAATTTNNGNNAAGTTTTCNGTATGNTGCATCACTTCACCTCTCNCTGACAGNNNNNTNGT  
ACCCATTAANATCANCATCTAATNAACAAGANTNNNCNACTCCAGTGAAAAGTNTCTCNTTACTCATGNNG  
CTGNNGNCNAAAANAANNANNAGAAGCCNTNNCANNTNNNNAANNCTCNNTNAANNNNNNNNNNNANGAC  
TNNANATNNNTCANNNNNNNNTNNGGATNNNNCNNNNNNNNGNANNNNCNNNNNNNNNNNNANCNNTT  
CNGCNGNNNNNNNNNNNNNNNNNNNNNATCTNANNNNNANNANNNNN

The following sequences was inconclusive:

- Colony 4B, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNGNNANTCTCGAGACCCAAGCTTGAACGCTGACGTCATCAACGAGGTTGAAGTTCT  
GTATGAGACCCTCTTTCCAATAAGGAAGAAGCCCTTACGAAGCTTGGTGAAGGCTTCTTTCTTATTGT  
TTTTTGGATCCACAGCCACCATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTGAA  
TTAGATGGTGATGTTAATGGGTACAAATTTTCTGTCAAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAA  
ACTTACCCTTAAATTTATTTGCACTACTGGAAAACCTACCTGTTCCATGGCCAACACTTGTCACTACTCTCA  
CTTATGGTGTTCATGCTTTTCAAGATATCCAGATCATATGAAGCGGCACGACTTCTTCAAGAGCGCCATG  
CCTGAGGGATACGTCAGGAGAGGACCATCTTCTCAAGGACGACGGGAACCTACAAGACACGTGCTGAAGT  
CAAGTTTGAAGGAGACACCCCTCGTCAACAGGATCGAGCTTAAGGGAATCGATTTCAAGGAGGACGGAAA  
TCCTCGGCCACAAGTTGGAATACAACCTACAACCTCCACAACGTATACATCATGGCCGACAAGCAAAGAAC  
GGCATCAAAGCCAACCTTCAAGACCCGCCACAACATCGAAGACGGCGGCTGCAACTCGCTGATCATTATCA  
ACAAAATACCTCAATTTGGCGATGACCCTGTCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCC  
TTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAAACGGCTGCTGGGATTACA  
CATGGNATGGNTGAACTATACAAATAACTCGAGCCTCTAGAACTATAGTGAGTCGTATTANNNNNTCCAGA  
CATGATAAGATACATTTGATGAGTTTGGACNAACCACANTAGANGNAGTGAATAATGCTTANTTNGGAAAN  
TTNNNTGCNNNTNNTNANTNGTANCATNNNGCTGCANNAANNNNNNACACANNNNNNNTNCATNNNNNTNNT  
NNNNNNNGNNGGNNNGGAGNNTNNNNNTNNNNGNNNNNNNTNNNTTACCNGNTANNCNNNNNNCNGNTT  
NNNNNNNNNNNGNNGNNNTGNNNANCANNNNNNNNNNNNNNNNNNNNCNCNNNNNNNNNNANNNNNNNN  
NNNNNNNNANAN

- Colony 4B, the reverse sequence is given from 3' → 5'

NNNNNNNNNTNTNNGNAANNNGNTANNCTTAAGTCGCGGCCGGAATTAATAAAACCTCCCACACCTCCCC  
CTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTATGCAGCTTATAATGGTTACAA  
ATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTCCTGACTGCATTCTAGTTGTGGTTTGTCCA  
AACTCATCAATGTATCTTATCATGTCTGGATCTACGTAATACGACTCACTATAGTTCTAGAGGCTCGAGTT  
ATTTGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCCGTTACAAACTCAAGAAGGACCATGTGGTCT  
CTCTTTTCGTTGGGATCTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGG  
GTCATCGCCAATTGGAGTATTTTGTGATAATGATCAGCGAGTTGCACGCCCGCTTTCGATGTTGTGGC  
GGGTCTTGAAGTTGGCTTTGATGCCGTTCTTTTGTCTGCGCCATGATGTATACGTTGTGGGAGTTGTAG  
TTGTATTCCAACCTGTGGCCGAGGATGTTTCCGTCCTCCTTGAAATCGATTCCCTTAAGCTCGATCCTGTT  
GACGAGGGTGTCTCCCTCAAACCTGACTTCAGCACGTGTCTTGTAGTTCCCGTCGTCCTTGAAGAAATGG  
TCCTCTCCTGCACGTATCCCTCAGGCATGGCGCTCTTGAAGAAGTCGTGCCGCTTCATATGATCTGGATAT  
CTTGAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAGTGTGGCCATGGAACAGGTAGTTTTCCAGTA  
NTGCAAATAAATTTANGGTAAGTTTCCGTATGNTGCATCACTTCACCTCNNCACTGACNGAAAATTTGTA  
CCCATTANATCACCATCTANTCAANANAATTGGGACANTCNGTGAAAAGTCTTTCNTNCNGGNGNTNGGNN  
ANNNNNNNNNANNNANNNANNNANNCCTTCACNNNGNTNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNN  
NACNNANTNCANNCNTGANGANGTCNNCNTCAGTCGNNNNNNNTNNCNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNCNNNTNNNNNNNTCCNGNNNNNNNNNNNNNNNNNNNNCNGAANCNNNNNNNNNNNNNNNNNN  
NNNCNCNNTTNCNNNNNNNNNNNNANANNNCNNNNCNNNN

- Colony 5B, the forward sequence is given from 5' → 3'

NNNNNNNNNNNNNNNNNNNGNNANCTCGANNGCTTGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGC  
GGGCCAGTGTCACTAGGCGGGAACACCCAGCGCGCTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAG  
GGGAGTGGCGCCCTGCAATATCTGCATGTGCTATGTGTTCTGGGAAATCACCGTAAACGTGAAATGTCTT  
TGGATTTGGGAATCTTATAAGTCTGTATGAGACCACTCTTCAATAAGGAAGAAGCCCTTCACGAAGCTT  
GGTGAAGGGCTTCTTCTTATTTTTTTTGGATCCACAGCCACCATGAGTAAAGGAGAANAACTTTTACTG  
GAGTTGTCCCAATCTTGTGTAATAGATGNNGATGTTAATGGGTACAAATTTCTGTGAGTGGAGANGGT  
GAAGGTGATGCAACGTACGGAAAACCTTACCCTTAAATTTATTTGCACTACTGGAAAACCTGTTCCATG  
NCCAACACTTGTCACTACTCTCACTTATGGTGTCAATGCTTTTCAAGATATCCAGATCATATGAAGCGNC  
ACGACTTCTTCAAGAGCGCCATGCCTGANGGATACGTGCAGGAGAGGACCATCTTCTTCANGGACNACGGN  
AACTACNAGACACGTGCTGAANTCAAGTTTGGGGAGACACCCTCGTCACAGGATCGAGNTAANGGAATCG  
ATTTCAGGAGGACGGAACATCCTCNCNCNAGTTGANACACTACACTCCCACACGTATACNTCNTGNCNAN  
AGCAAAGACGNATCAAGCAACTCANACCCGCNCAACATCGANACGNGCNNNNNTCGNTGATCATTATCAN  
NNAATACNNCANGGNNATNACCNTGTCNTTTACAGANNCTTACNNNCACNCATCNGCCCTTNCGAGANCC  
NNNAAANANNACNANTNGNCCNTNTNNTTGNCCGNNNNNNNNNCANGNNNNNNNNNNCNCNNNNNNNGGCNN  
NNNNNNNNNNNTNNNNNNNNANNAANNNNNNNTTANNNNNNNCCCNNNNANNNNGNANNNNNNNNNN  
NNNNGTAGNNNNNNNNNNANNNNNNNNNANNNNNNNNNNNNNNNNNNNNTNNTNNTCNCNNNNNGNANNNGN  
NGNNNNNNNNNNANANNATCNCNNNNNNNNNNNTNNTGNCNNNNNNNCNNNNNNNGN





## Appendix 10: FACS Generated Data of Transfection of K562 and KCL-22 cell lines

- FACS data generated from nucleofection of K562 cells

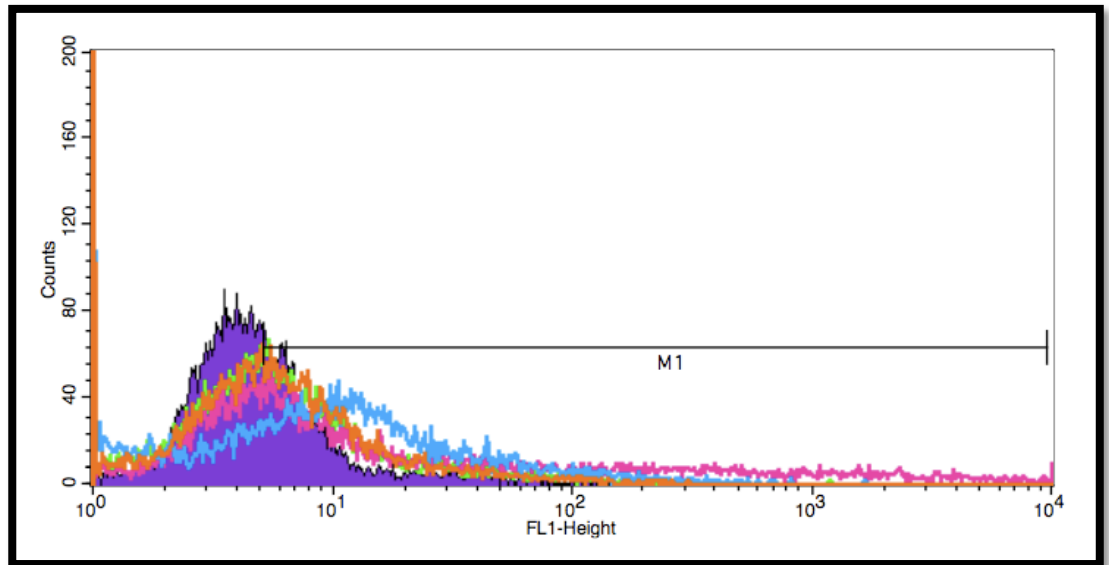


Figure 84 FACS histogram showing the transfection efficiency in K562 cells, transfection number 1. The results as shown below:

Sample	Transfection efficiency with regards to non-transfected cells		Transfection efficiency with regards to negative control cells	
	(35.95%)	-Purple	(47.40%)	-Green
Positive control -Blue	28.06%		28%	
shRNA1 <sup><math>\beta^3\alpha^2</math></sup> - Pink	26.04%		26.07%	
shRNA3 <sup><math>\beta^3\alpha^2</math></sup> -Orange	14.01%		2.53%	

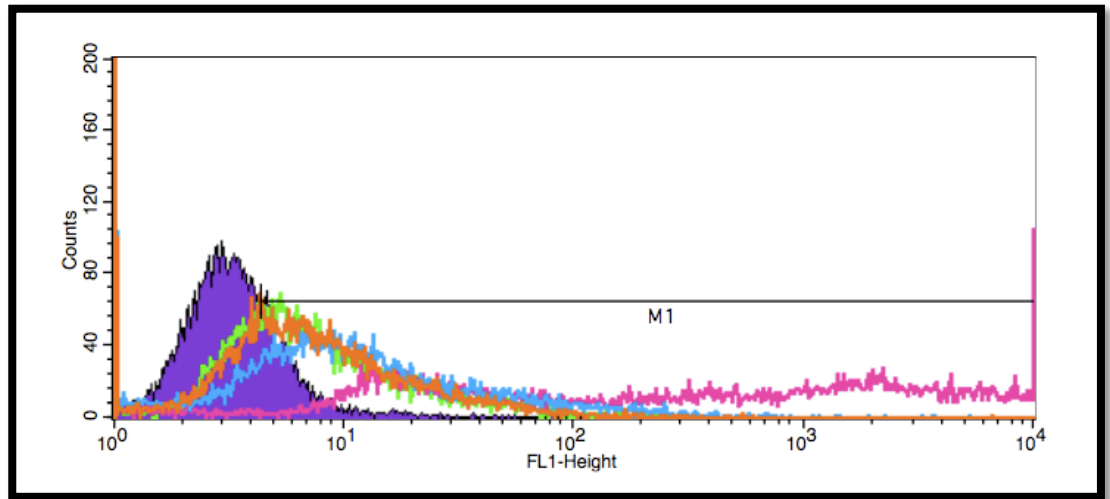


Figure 85 FACS histogram showing the transfection efficiency in K562 cells, transfection number 2. The results as shown below:

Sample	Transfection efficiency regards to non-transfected cells	Transfection efficiency regards to negative control cells
	(23.05%) -Purple	(68.33%) -Green
Positive control -Pink	72.15%	27.17%
shRNA1 <sup><math>\beta^3\alpha^2</math></sup> - Blue	54.18%	8.9%
shRNA3 <sup><math>\beta^3\alpha^2</math></sup> -Orange	47.72%	2.44%

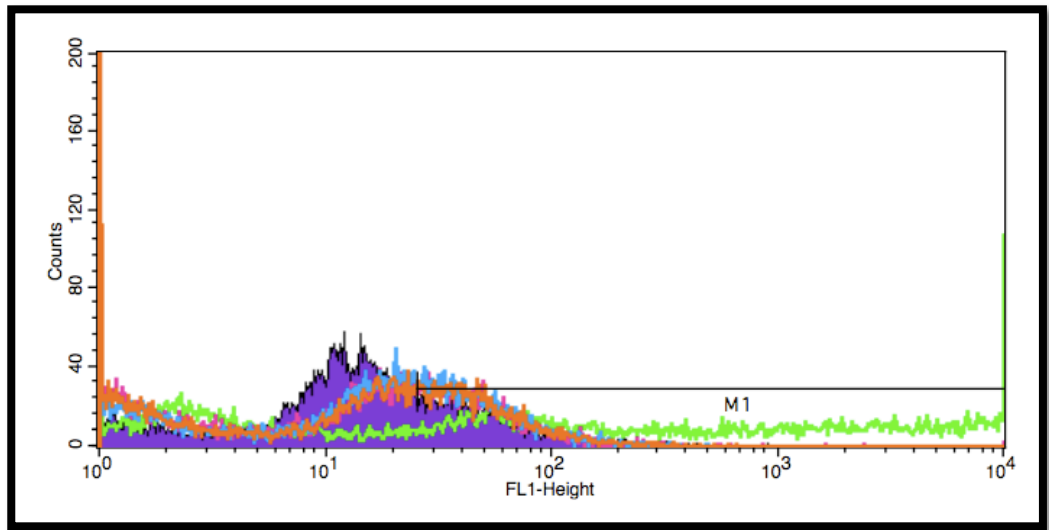


Figure 86 FACS histogram showing the transfection efficiency in K562 cells, using shRNA1 <sup>$\beta$ 2 $\alpha$ 2</sup>. The results as shown below:

Sample	Transfection efficiency regards to non-transfected cells (21.70%) –Purple
Positive control -Green	43.18%
shRNA1 <sup><math>\beta</math>2<math>\alpha</math>2</sup> - Pink	5.77%
shRNA1 <sup><math>\beta</math>2<math>\alpha</math>2</sup> - Blue	9.86%
shRNA1 <sup><math>\beta</math>2<math>\alpha</math>2</sup> - Orange	6.77%

- FACS data generated from nucleofection of KCL-22 cells

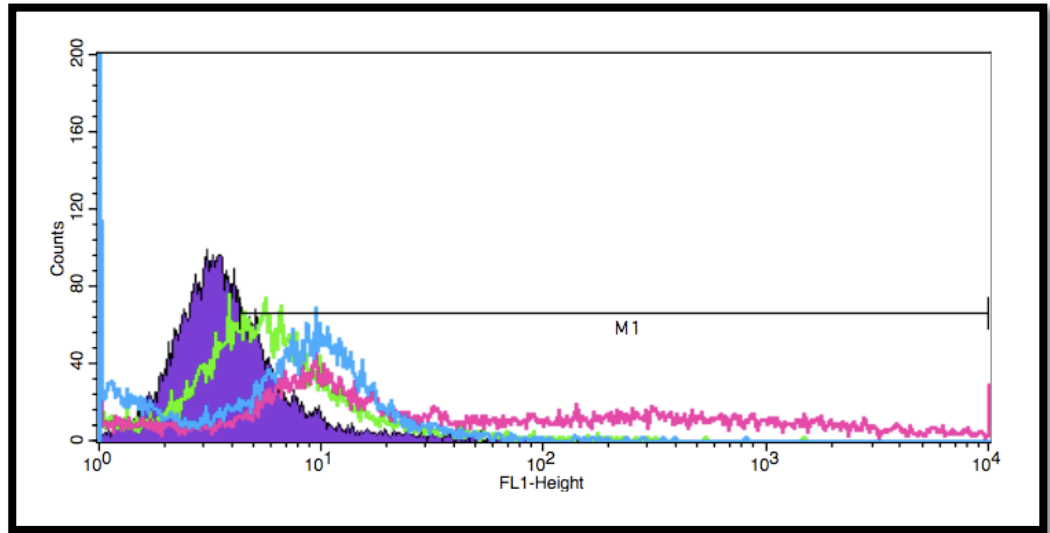


Figure 87 FACS histogram showing the transfection efficiency in KCL-22 cells, transfection number 1. The results as shown below:

Sample	Transfection efficiency regards to non-transfected cells	Transfection efficiency regards to negative control cells
	(32.29%) -Purple	(58.69%) -Green
Positive control -Pink	53.14%	26.74%
shRNA1 <sup>β2α2</sup> - Bue	28.33%	1.93%

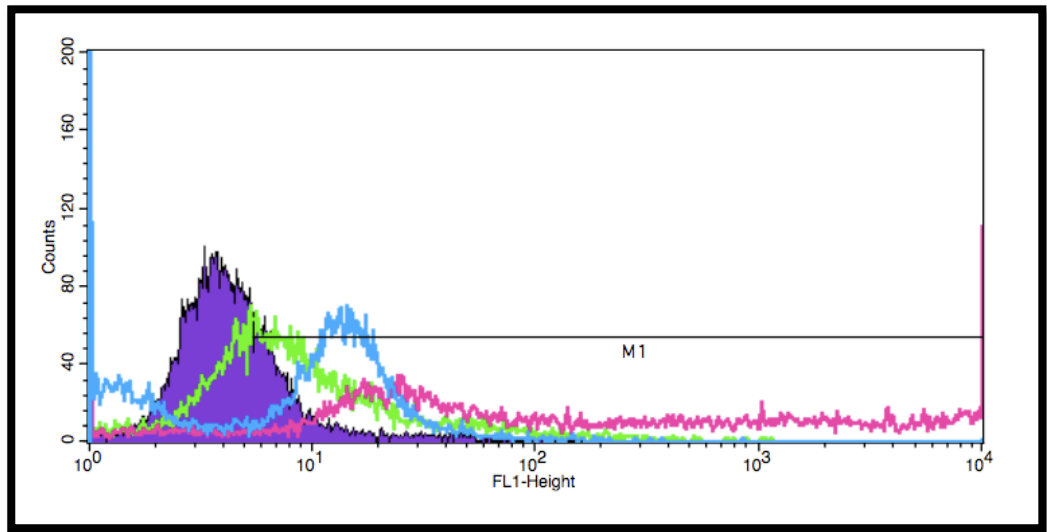


Figure 88 FACS histogram showing the transfection efficiency in KCL-22 cells, transfection number 3. The results as shown below:

Sample	Transfection efficiency regards to non-transfected cells (26.06%) -Purple	Transfection efficiency regards to negative control cells (61.66%) -Green
Positive control -Pink	67%	31.35%
shRNA1 <sup>β3α2</sup> - Blue	36.94%	1.34%

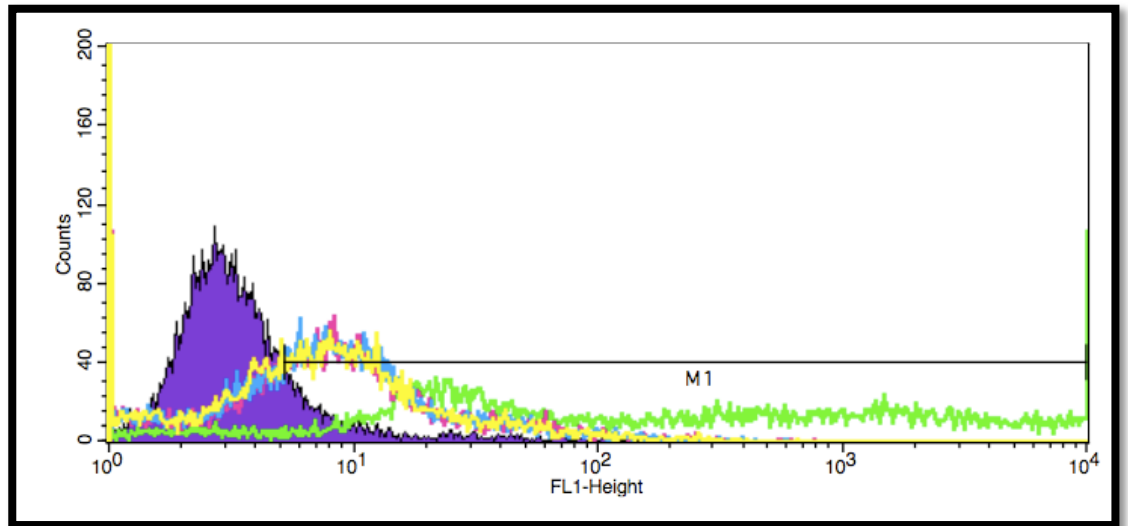


Figure 89 FACS histogram showing the transfection efficiency in KCL-22 cells, using shRNA1 <sup>$\beta$ 3 $\alpha$ 2</sup>. The results as shown below:

Sample	Transfection efficiency regards to non-transfected cells (18.87%) – Purple
Positive control -Green	79.27%
shRNA1 <sup><math>\beta</math>3<math>\alpha</math>2</sup> - Pink	50.36%
shRNA1 <sup><math>\beta</math>3<math>\alpha</math>2</sup> - Blue	51.35%
shRNA1 <sup><math>\beta</math>3<math>\alpha</math>2</sup> - Yellow	44.29%

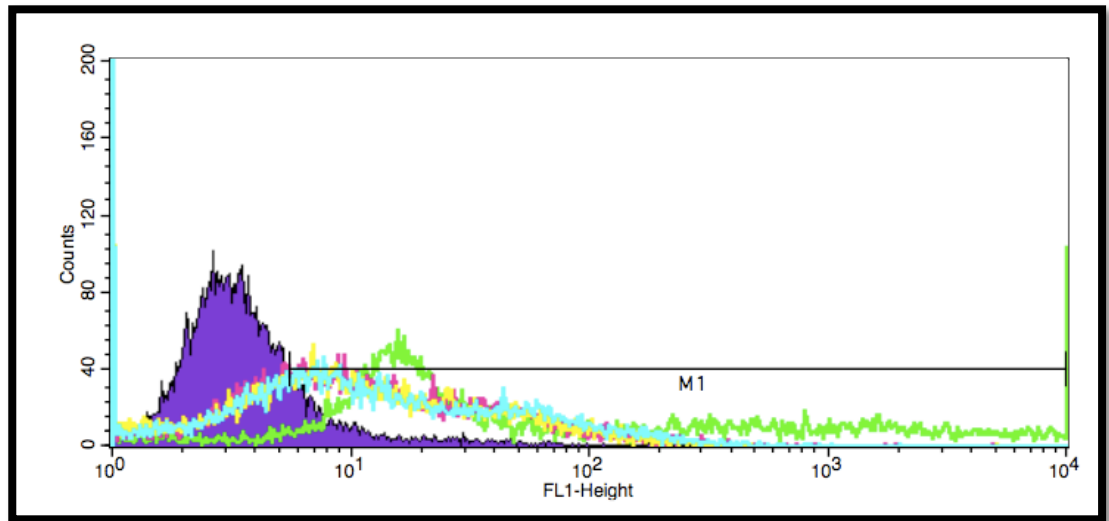


Figure 90 FACS histogram showing the transfection efficiency in KCL-22 cells, using shRNA3<sup>β3α2</sup>. The results as shown below:

Sample	Transfection efficiency with regards to non-transfected cells (15.44%) – Purple
Positive control -Green	77.49%
shRNA3 <sup>β3α2</sup> - Pink	53.67%
shRNA3 <sup>β3α2</sup> - Yellow	51.81%
shRNA3 <sup>β3α2</sup> - Blue	54.88%



## Bibliography

Advani, A. S. and A. M. Pendergast (2002). "Bcr-Abl variants: biological and clinical aspects." Leuk Res 26(8): 713-720.

Adler, R., S. Viehmann, et al. (2009). "Correlation of BCR/ABL transcript variants with patients' characteristics in childhood chronic myeloid leukaemia." Eur J Haematol 82(2): 112-118.

Agarwal, N. and A. M. Gewirtz (1999). "Oligonucleotide therapeutics for hematologic disorders." Biochim Biophys Acta 1489(1): 85-96.

Albers, C., A. L. Illert, et al. (2011). "An RNAi based system for loss of function analysis identifies Raf1 as a crucial mediator of BCR-ABL driven leukemogenesis." Blood.

An, D. S., F. X. Qin, et al. (2006). "Optimization and functional effects of stable short hairpin RNA expression in primary human lymphocytes via lentiviral vectors." Mol Ther 14(4): 494-504.

Arthanari, Y., A. Pluen, et al. (2010). "Delivery of therapeutic shRNA and siRNA by Tat fusion peptide targeting BCR-ABL fusion gene in Chronic Myeloid Leukemia cells." J Control Release 145(3): 272-280.

Arya, D. P. (2011). "New approaches toward recognition of nucleic acid triple helices." Acc Chem Res 44(2): 134-146.

Bartova, E., A. Harnicarova, et al. (2005). "Nuclear topography and expression of the BCR/ABL fusion gene and its protein level influenced by cell differentiation and RNA interference." Leuk Res **29**(8): 901-913.

Bernstein, E., A. A. Caudy, et al. (2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference." Nature **409**(6818): 363-366.

Brehme, M., O. Hantschel, et al. (2009). "Charting the molecular network of the drug target Bcr-Abl." Proc Natl Acad Sci U S A **106**(18): 7414-7419.

Brummelkamp, T. R., R. Bernards, et al. (2002). "A system for stable expression of short interfering RNAs in mammalian cells." Science **296**(5567): 550-553.

Campanini, F., M. A. Santucci, et al. (2001). "Competitive polymerase chain reaction as a method to detect the amplification of bcr-abl gene of chronic myeloid leukemia." Haematologica **86**(2): 167-173.

Castanotto, D., J. R. Li, et al. (2002). "Intracellular ribozyme applications." Biochem Soc Trans **30**(Pt 6): 1140-1145.

Chang, C. I., H. A. Kim, et al. (2011). "Structural Diversity Repertoire of Gene Silencing Small Interfering RNAs." Oligonucleotides.

Chendrimada, T. P., R. I. Gregory, et al. (2005). "TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing." Nature **436**(7051): 740-744.

Chopra, R., Q. Q. Pu, et al. (1999). "Biology of BCR-ABL." Blood Rev **13**(4): 211-229.

Constantinescu, S. N., M. Girardot, et al. (2008). "Mining for JAK-STAT mutations in cancer." Trends Biochem Sci **33**(3): 122-131.

Cullen, B. R. (2002). "RNA interference: antiviral defense and genetic tool." Nat Immunol **3**(7): 597-599.

Davidson, B. L. and P. B. McCray, Jr. (2011). "Current prospects for RNA interference-based therapies." Nat Rev Genet **12**(5): 329-340.

Dias, N. and C. A. Stein (2002). "Antisense oligonucleotides: basic concepts and mechanisms." Mol Cancer Ther **1**(5): 347-355.

Diekmann, D., S. Brill, et al. (1991). "Bcr encodes a GTPase-activating protein for p21rac." Nature **351**(6325): 400-402.

Drexler, H. G. (1994). "Leukemia cell lines: in vitro models for the study of chronic myeloid leukemia." Leuk Res **18**(12): 919-927.

Drexler, H. G., R. A. MacLeod, et al. (1999). "Leukemia cell lines: in vitro models for the study of Philadelphia chromosome-positive leukemia." Leuk Res **23**(3): 207-215.

Druker, B. J. (2002). "STI571 (Gleevec) as a paradigm for cancer therapy." Trends Mol Med **8**(4 Suppl): S14-18.

Duca, M., P. Vekhoff, et al. (2008). "The triple helix: 50 years later, the outcome." Nucleic Acids Res **36**(16): 5123-5138.

Dykxhoorn, D. M., C. D. Novina, et al. (2003). "Killing the messenger: short RNAs that silence gene expression." Nat Rev Mol Cell Biol **4**(6): 457-467.

Elbashir, S. M., W. Lendeckel, et al. (2001). "RNA interference is mediated by 21- and 22-nucleotide RNAs." Genes Dev **15**(2): 188-200.

Elmaagacli, A. H., M. Koldehoff, et al. (2005). "WT1 and BCR-ABL specific small interfering RNA have additive effects in the induction of apoptosis in leukemic cells." Haematologica **90**(3): 326-334.

Eskazan, A. E., T. Soysal, et al. (2011). "Chronic myeloid leukemia patients with F317L BCR-ABL kinase domain mutation are resistant to dasatinib: Is that true for all the patients?" Leuk Res **35**(9): e145-146.

Esposito, N., I. Colavita, et al. (2011). "SHP1 expression accounts for resistance to imatinib treatment in Philadelphia chromosome-positive cells derived from patients with chronic myeloid leukemia." Blood.

Evans H. and Sillibourne J. (1996). "Detection of the BCR-ABL Gene in CML/ALL patients." Promega Notes Magazine **57** 21.

FACS figure. (<http://www.abcam.com/ps/CMS/Images/Flow-Cytometry-Diagram2.jpg>). [Accessed August 2011].

Faber, E., R. Mojzikova, et al. (2010). "Major molecular response achieved with dasatinib in a CML patient with F317L BCR-ABL kinase domain mutation." Leuk Res **34**(4): e91-93.

Faderl, S., M. Talpaz, et al. (1999). "The biology of chronic myeloid leukemia." N Engl J Med **341**(3): 164-172.

Fausel C. (2007). "Targeted Chronic Myeloid Leukemia Therapy: Seeking a Cure." Supplement to Journal of Managed Care Pharmacy **13** S8-S12.

Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*." Nature **391**(6669): 806-811.

Gambacorti-Passerini, C. B., R. H. Gunby, et al. (2003). "Molecular mechanisms of resistance to imatinib in Philadelphia-chromosome-positive leukaemias." Lancet Oncol **4**(2): 75-85.

Georgescu, M. M. (2010). "PTEN Tumor Suppressor Network in PI3K-Akt Pathway Control." Genes Cancer **1**(12): 1170-1177.

Gerard, G. F., D. K. Fox, et al. (1997). "Reverse transcriptase. The use of cloned Moloney murine leukemia virus reverse transcriptase to synthesize DNA from RNA." Mol Biotechnol **8**(1): 61-77.

Gorre, M. E., M. Mohammed, et al. (2001). "Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification." Science **293**(5531): 876-880.

Goh, H. G., J. Y. Hwang, et al. (2006). "Comprehensive analysis of BCR-ABL transcript types in Korean CML patients using a newly developed multiplex RT-PCR." Transl Res **148**(5): 249-256.

Groffen, J., J. R. Stephenson, et al. (1984). "Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22." Cell **36**(1): 93-99.

Hacein-Bey-Abina, S., A. Garrigue, et al. (2008). "Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1." J Clin Invest **118**(9): 3132-3142.

Han, J., Y. Lee, et al. (2004). "The Drosha-DGCR8 complex in primary microRNA processing." Genes Dev **18**(24): 3016-3027.

Hantschel, O. and G. Superti-Furga (2004). "Regulation of the c-Abl and Bcr-Abl tyrosine kinases." Nat Rev Mol Cell Biol **5**(1): 33-44.

Heisterkamp, N. and J. Groffen (2002). "Philadelphia-positive leukemia: a personal perspective." Oncogene **21**(56): 8536-8540.

Heisterkamp, N., J. Groffen, et al. (1982). "Chromosomal localization of human cellular homologues of two viral oncogenes." Nature **299**(5885): 747-749.

Heisterkamp, N., K. Stam, et al. (1985). "Structural organization of the bcr gene and its role in the Ph' translocation." Nature **315**(6022): 758-761.

Heisterkamp, N., J. R. Stephenson, et al. (1983). "Localization of the c-ab1 oncogene adjacent to a translocation break point in chronic myelocytic leukaemia." Nature **306**(5940): 239-242.

Henson, E. S. and S. B. Gibson (2006). "Surviving cell death through epidermal growth factor (EGF) signal transduction pathways: implications for cancer therapy." Cell Signal **18**(12): 2089-2097.

Hindley, A. and W. Kolch (2002). "Extracellular signal regulated kinase (ERK)/mitogen activated protein kinase (MAPK)-independent functions of Raf kinases." J Cell Sci **115**(Pt 8): 1575-1581.

Hoffbrand A.V., Moss P.A.H. and Pettit J.E. (2006). *Essential haematology* (Fifth Edition) Blackwell Publishing. **Chapters 7, 8 and 13.**

Holland, P. M., R. D. Abramson, et al. (1991). "Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of *Thermus aquaticus* DNA polymerase." Proc Natl Acad Sci U S A **88**(16): 7276-7280.

Jabbour, E., S. Branford, et al. (2011). "Practical advice for determining the role of BCR-ABL mutations in guiding tyrosine kinase inhibitor therapy in patients with chronic myeloid leukemia." Cancer **117**(9): 1800-1811.

Jemal, A., B. I. Graubard, et al. (2002). "The association of blood lead level and cancer mortality among whites in the United States." Environ Health Perspect **110**(4): 325-329.

James, H. A. and I. Gibson (1998). "The therapeutic potential of ribozymes." Blood **91**(2): 371-382.

James, H. A., C. M. Twomey, et al. (1996). "Specificity of ribozymes against the bcr-abl mRNAs in vitro." Biochem Soc Trans **24**(3): 409S.

Kalidas, M., H. Kantarjian, et al. (2001). "Chronic myelogenous leukemia." JAMA **286**(8): 895-898.

Kamamoto, T., T. Ohkubo, et al. (1986). "Establishment of two Ph1 chromosome-positive cell lines, KPB-M8 and KPB-M15." Jpn J Clin Oncol **16**(2): 107-115.

Karagiannis, T. C. and A. El-Osta (2005). "RNA interference and potential therapeutic applications of short interfering RNAs." Cancer Gene Ther **12**(10): 787-795.

Kijima, H., H. Ishida, et al. (1995). "Therapeutic applications of ribozymes." Pharmacol Ther **68**(2): 247-267.

Klinik andrologi website.

(<http://klinikandrologi.blogspot.com/2011/01/oligozoospermia.html>)

[accessed: 13<sup>th</sup> of Dec. 2007].

Knoops L., and S. N. Constantinescu. (2011). "JAK inhibitors in myeloproliferative neoplasms." Belg J Hemato **2**: 27:35.

Koeffler, H. P. and D. W. Golde (1980). "Human myeloid leukemia cell lines: a review." Blood **56**(3): 344-350.

Koldehoff, M., L. Kordelas, et al. (2010). "Small interfering RNA against BCR-ABL transcripts sensitize mutated T315I cells to nilotinib." Haematologica **95**(3): 388-397.

Kubonishi, I. and I. Miyoshi (1983). "Establishment of a Ph1 chromosome-positive cell line from chronic myelogenous leukemia in blast crisis." Int J Cell Cloning **1**(2): 105-117.

Lange, W., M. Daskalakis, et al. (1994). "Comparison of different ribozymes for efficient and specific cleavage of BCR/ABL related mRNAs." FEBS Lett **338**(2): 175-178.

La Rosee, P., S. Holm-Eriksen, et al. (2008). "Phospho-CRKL monitoring for the assessment of BCR-ABL activity in imatinib-resistant chronic myeloid leukemia or Ph+ acute lymphoblastic leukemia patients treated with nilotinib." Haematologica **93**(5): 765-769.



Laurent, E., M. Talpaz, et al. (2001). "The BCR gene and philadelphia chromosome-positive leukemogenesis." Cancer Res **61**(6): 2343-2355.

Lavrovsky, Y., S. Chen, et al. (1997). "Therapeutic potential and mechanism of action of oligonucleotides and ribozymes." Biochem Mol Med **62**(1): 11-22.

Leibowitz, D., K. Young, et al. (1991). "Sequence-specific DNA-binding proteins within the MbcR on the Ph1 chromosome." Genes Chromosomes Cancer **3**(4): 308-312.

Lewin, A. S. and W. W. Hauswirth (2001). "Ribozyme gene therapy: applications for molecular medicine." Trends Mol Med **7**(5): 221-228.

Li, M. J., R. McMahon, et al. (2003). "Specific killing of Ph+ chronic myeloid leukemia cells by a lentiviral vector-delivered anti-bcr/abl small hairpin RNA." Oligonucleotides **13**(5): 401-409.

Li, F. and R. I. Mahato (2009). "Bipartite vectors for co-expression of a growth factor cDNA and short hairpin RNA against an apoptotic gene." J Gene Med **11**(9): 764-771.

Liu, Y. P. and B. Berkhout (2011). "miRNA cassettes in viral vectors: Problems and solutions." Biochim Biophys Acta.

Lozzio, C. B. and B. B. Lozzio (1975). "Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome." Blood **45**(3): 321-334.

Lozzio, B. B. and C. B. Lozzio (1979). "Properties and usefulness of the original K-562 human myelogenous leukemia cell line." Leuk Res **3**(6): 363-370.

Luo K.J. (2010). "Homing on in CML stem cells." Science-Business X-change Cover story: Targets and Mechanisms. 1-2.

Machado-Neto, J. A., P. Favaro, et al. (2011). "Knockdown of insulin receptor substrate 1 reduces proliferation and downregulates Akt/mTOR and MAPK pathways in K562 cells." Biochim Biophys Acta **1813**(8): 1404-1411.

Maru, Y. and O. N. Witte (1991). "The BCR gene encodes a novel serine/threonine kinase activity within a single exon." Cell **67**(3): 459-468.

Matranga, C., Y. Tomari, et al. (2005). "Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes." Cell **123**(4): 607-620.

Melo, J. V. (1997). "BCR-ABL gene variants." Baillieres Clin Haematol **10**(2): 203-222.

Melo, J. V. (1996). "The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype." Blood **88**(7): 2375-2384.

Mendonca, L. S., J. N. Moreira, et al. (2010). "Co-encapsulation of anti-BCR-ABL siRNA and imatinib mesylate in transferrin receptor-targeted sterically stabilized liposomes for chronic myeloid leukemia treatment." Biotechnol Bioeng **107**(5): 884-893.

Mendoza-Maldonado, R., L. Zentilin, et al. (2002). "Purging of chronic myelogenous leukemia cells by retrovirally expressed anti-bcr-abl ribozymes with specific cellular compartmentalization." Cancer Gene Ther **9**(1): 71-86.

Merkerova, M., H. Klamova, et al. (2007). "Targeting of gene expression by siRNA in CML primary cells." Mol Biol Rep **34**(1): 27-33.

Michor, F., Y. Iwasa, et al. (2006). "The age incidence of chronic myeloid leukemia can be explained by a one-mutation model." Proc Natl Acad Sci U S A **103**(40): 14931-14934.

Mills, K. I., E. D. MacKenzie, et al. (1988). "The site of the breakpoint within the bcr is a prognostic factor in Philadelphia-positive CML patients." Blood **72**(4): 1237-1241.

Milojkovic D. and Apperley J. (2008). "State-of-the-Art in the Treatment of Chronic Myeloid Leukaemia." Current Opinion in Oncology **20** 112-121.

Mosiman, V. L., B. K. Patterson, et al. (1997). "Reducing cellular autofluorescence in flow cytometry: an in situ method." Cytometry **30**(3): 151-156.

Mulhbacher, J., P. St-Pierre, et al. (2010). "Therapeutic applications of ribozymes and riboswitches." Curr Opin Pharmacol **10**(5): 551-556.

Myssina, S., G. V. Helgason, et al. (2009). "Combined BCR-ABL inhibition with lentiviral-delivered shRNA and dasatinib augments induction of apoptosis in Philadelphia-positive cells." Exp Hematol **37**(2): 206-214.

Nambu, T., N. Araki, et al. (2010). "Contribution of BCR-ABL-independent activation of ERK1/2 to acquired imatinib resistance in K562 chronic myeloid leukemia cells." Cancer Sci **101**(1): 137-142.

Naughton, R., C. Quiney, et al. (2009). "Bcr-Abl-mediated redox regulation of the PI3K/AKT pathway." Leukemia **23**(8): 1432-1440.

Nowell, P. C. (2007). "Discovery of the Philadelphia chromosome: a personal perspective." J Clin Invest **117**(8): 2033-2035.

Noor, S. M., R. Bell, et al. (2011). "Shooting the messenger: targeting signal transduction pathways in leukemia and related disorders." Crit Rev Oncol Hematol **78**(1): 33-44.

Ogura, M., Y. Morishima, et al. (1985). "Establishment of a novel human megakaryoblastic leukemia cell line, MEG-01, with positive Philadelphia chromosome." Blood **66**(6): 1384-1392.

Ohkawa, J. and K. Taira (2000). "Control of the functional activity of an antisense RNA by a tetracycline-responsive derivative of the human U6 snRNA promoter." Hum Gene Ther **11**(4): 577-585.

Ohmine, K., T. Nagai, et al. (2003). "Analysis of gene expression profiles in an imatinib-resistant cell line, KCL22/SR." Stem Cells **21**(3): 315-321.

Okabe, S., T. Tauchi, et al. (2011). "Dasatinib preferentially induces apoptosis by inhibiting Lyn kinase in nilotinib-resistant chronic myeloid leukemia cell line." J Hematol Oncol **4**(1): 32.

Oliveira, D. M. and M. A. Goodell (2003). "Transient RNA interference in hematopoietic progenitors with functional consequences." Genesis **36**(4): 203-208.

Ovcharenko, D., R. Jarvis, et al. (2005). "High-throughput RNAi screening in vitro: from cell lines to primary cells." RNA **11**(6): 985-993.

Paddison, P. J., A. A. Caudy, et al. (2002). "Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells." Genes Dev **16**(8): 948-958.

Pane, F., M. Intrieri, et al. (2002). "BCR/ABL genes and leukemic phenotype: from molecular mechanisms to clinical correlations." Oncogene **21**(56): 8652-8667.

pEGFP-N1 vector diagram. ([www.pkclab.org/PKC/vector/pEGFPN1.pdf](http://www.pkclab.org/PKC/vector/pEGFPN1.pdf)). [Accessed on May 2008].

pGEM<sup>®</sup>T easy vector system diagram. (<http://www.promega.com/resources/product-guides-and-selectors/protocols-and-applications-guide/cloning/>). [Accessed on April 2008].

Pluk, H., K. Dorey, et al. (2002). "Autoinhibition of c-Abl." Cell **108**(2): 247-259.

Puil, L., J. Liu, et al. (1994). "Bcr-Abl oncoproteins bind directly to activators of the Ras signalling pathway." EMBO J **13**(4): 764-773.

Pushparaj, P. N. and A. J. Melendez (2006). "Short interfering RNA (siRNA) as a novel therapeutic." Clin Exp Pharmacol Physiol **33**(5-6): 504-510.

Puttini, M., S. Redaelli, et al. (2008). "Characterization of compound 584, an Abl kinase inhibitor with lasting effects." Haematologica **93**(5): 653-661.

Quackenbush, R. C., G. W. Reuther, et al. (2000). "Analysis of the biologic properties of p230 Bcr-Abl reveals unique and overlapping properties with the oncogenic p185 and p210 Bcr-Abl tyrosine kinases." Blood **95**(9): 2913-2921.

Quentmeier, H., S. Eberth, et al. (2011). "BCR-ABL1-independent PI3Kinase activation causing imatinib-resistance." J Hematol Oncol **4**: 6.

Quintas-Cardama, A. (2008). "Experimental non-ATP-competitive therapies for chronic myelogenous leukemia." Leukemia **22**(5): 932-940.

Rand, T. A., S. Petersen, et al. (2005). "Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation." Cell **123**(4): 621-629.

Rao, D. D., J. S. Vorhies, et al. (2009). "siRNA vs. shRNA: similarities and differences." Adv Drug Deliv Rev **61**(9): 746-759.

Rapozzi, V., S. Cogoi, et al. (2002). "Antigene effect in K562 cells of a PEG-conjugated triplex-forming oligonucleotide targeted to the bcr/abl oncogene." Biochemistry **41**(2): 502-510.

Rapozzi, V. and L. E. Xodo (2004). "Efficient silencing of bcr/abl oncogene by single- and double-stranded siRNAs targeted against b2a2 transcripts." Biochemistry **43**(51): 16134-16141.

Robinson J.P. (2004). "Flow cytometry" Encyclopedia of Biomaterials and Biomedical Engineering 630-640. ([www.abdserotec.com/uploads/Flow-Cytometry.pdf](http://www.abdserotec.com/uploads/Flow-Cytometry.pdf)). [Accessed on September 2008].

Roche-Lestienne, C., V. Soenen-Cornu, et al. (2002). "Several types of mutations of the Abl gene can be found in chronic myeloid leukemia patients resistant to STI571, and they can pre-exist to the onset of treatment." Blood **100**(3): 1014-1018.

Rossi, J. J. (2008). "Expression strategies for short hairpin RNA interference triggers." Hum Gene Ther **19**(4): 313-317.

Rowley, J. D. (1973). "Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining." Nature **243**(5405): 290-293.

Rumpold, H., A. M. Wolf, et al. (2005). "RNAi-mediated knockdown of P-glycoprotein using a transposon-based vector system durably restores imatinib sensitivity in imatinib-resistant CML cell lines." Exp Hematol **33**(7): 767-775.

Ruschmann, J., V. Ho, et al. (2010). "Tyrosine phosphorylation of SHIP promotes its proteasomal degradation." Exp Hematol **38**(5): 392-402, 402 e391.

Sawyers, C. L. (1999). "Chronic myeloid leukemia." N Engl J Med **340**(17): 1330-1340.

Scherr, M., A. Chaturvedi, et al. (2006). "Enhanced sensitivity to inhibition of SHP2, STAT5, and Gab2 expression in chronic myeloid leukemia (CML)." Blood **107**(8): 3279-3287.

Scherr, M., K. Battmer, et al. (2003). "Specific inhibition of bcr-abl gene expression by small interfering RNA." Blood **101**(4): 1566-1569.

Schopman, N. C., Y. P. Liu, et al. (2010). "Optimization of shRNA inhibitors by variation of the terminal loop sequence." Antiviral Res **86**(2): 204-211.

Sessions, J. (2007). "Chronic myeloid leukemia in 2007." J Manag Care Pharm **13**(8 Suppl A): 4-7.

Seyhan, A. A. (2011). "RNAi: a potential new class of therapeutic for human genetic disease." Hum Genet.

Shah, N. P., O. N. Witte, et al. (1991). "Characterization of the BCR promoter in Philadelphia chromosome-positive and -negative cell lines." Mol Cell Biol **11**(4): 1854-1860.

Sharp, P. A. (2001). "RNA interference--2001." Genes Dev **15**(5): 485-490.

Shegokar, R., L. Al Shaal, et al. (2011). "SiRNA delivery: challenges and role of carrier systems." Pharmazie **66**(5): 313-318.

Shepherd, P., R. Suffolk, et al. (1995). "Analysis of molecular breakpoint and m-RNA transcripts in a prospective randomized trial of interferon in chronic myeloid leukaemia: no correlation with clinical features, cytogenetic response, duration of chronic phase, or survival." Br J Haematol **89**(3): 546-554.

Sherbenou, D. W., O. Hantschel, et al. (2010). "BCR-ABL SH3-SH2 domain mutations in chronic myeloid leukemia patients on imatinib." Blood **116**(17): 3278-3285.

Shieh, M. P., M. Mitsuhashi, et al. (2011). "Moving on up: Second-Line Agents as Initial Treatment for Newly-Diagnosed Patients with Chronic Phase CML." Clin Med Insights Oncol **5**: 185-199.

Shtivelman, E., B. Lifshitz, et al. (1985). "Fused transcript of abl and bcr genes in chronic myelogenous leukaemia." Nature **315**(6020): 550-554.



Shtivelman, E., B. Lifshitz, et al. (1986). "Alternative splicing of RNAs transcribed from the human *abl* gene and from the *bcr-abl* fused gene." Cell **47**(2): 277-284.

Siemering, K. R., R. Golbik, et al. (1996). "Mutations that suppress the thermosensitivity of green fluorescent protein." Curr Biol **6**(12): 1653-1663.

Skaggs, B. J., M. E. Gorre, et al. (2006). "Phosphorylation of the ATP-binding loop directs oncogenicity of drug-resistant BCR-ABL mutants." Proc Natl Acad Sci U S A **103**(51): 19466-19471.

Snyder, L. L., I. Ahmed, et al. (2009). "RNA polymerase III can drive polycistronic expression of functional interfering RNAs designed to resemble microRNAs." Nucleic Acids Res **37**(19): e127.

Sobecks R.M. and Theil K. (2003). Leukaemias. The Cleveland Clinic Foundation. [Internet Site]. Accessed on the 25<sup>th</sup> of April. 2008.  
([www.clevelandclinicmeded.com/medicalpubs/diseasemanagement/hematology/chleuk/chleuk.htm#ref1](http://www.clevelandclinicmeded.com/medicalpubs/diseasemanagement/hematology/chleuk/chleuk.htm#ref1).)

Spiller, D. G., R. V. Giles, et al. (1998). "The influence of target protein half-life on the effectiveness of antisense oligonucleotide analog-mediated biologic responses." Antisense Nucleic Acid Drug Dev **8**(4): 281-293.

Sriram, G. and R. B. Birge (2010). "Emerging roles for *crk* in human cancer." Genes Cancer **1**(11): 1132-1139.

Sriram, G., C. Reichman, et al. (2011). "Phosphorylation of Crk on tyrosine 251 in the RT loop of the SH3C domain promotes Abl kinase transactivation." Oncogene.

Srivastava, G. K., R. Reinoso, et al. (2011). "Trypan Blue staining method for quenching the autofluorescence of RPE cells for improving protein expression analysis." Exp Eye Res.

Steelman, L. S., R. A. Franklin, et al. (2011). "Roles of the Ras/Raf/MEK/ERK pathway in leukemia therapy." Leukemia **25**(7): 1080-1094.

Steelman, L. S., S. C. Pohnert, et al. (2004). "JAK/STAT, Raf/MEK/ERK, PI3K/Akt and BCR-ABL in cell cycle progression and leukemogenesis." Leukemia **18**(2): 189-218.

Steinberg, M. (2007). "Dasatinib: a tyrosine kinase inhibitor for the treatment of chronic myelogenous leukemia and philadelphia chromosome-positive acute lymphoblastic leukemia." Clin Ther **29**(11): 2289-2308.

Stentoft J., N. Pallisgaard, et al. (2001). "Kinetics of BCR-ABL Fusion Transcript Levels in Chronic Myeloid Leukemia patients treated with STI571 Measured by Quantitative Real-Time Polymerase Chain Reaction." European Journal of Haematology **67** 302-308.

Talpaz M., N.P. Shah, et al. (2006). "Dasatinib in Imatinib-Resistant Philadelphia Chromosome-Positive Leukemias." The New England Journal of Medicine **354** 2531-2541.

Terasawa, K., K. Shimizu, et al. (2011). "Synthetic Pre-miRNA-Based shRNA as Potent RNAi Triggers." J Nucleic Acids **2011**: 131579.

Thomas, M., J. Greil, et al. (2006). "Targeting leukemic fusion proteins with small interfering RNAs: recent advances and therapeutic potentials." Acta Pharmacol Sin **27**(3): 273-281.

Traina, F., J. B. Carneiro, et al. (2003). "BCR-ABL binds to IRS-1 and IRS-1 phosphorylation is inhibited by imatinib in K562 cells." FEBS Lett **535**(1-3): 17-22.

Trudel, P., S. Provost, et al. (1996). "pGATA: a positive selection vector based on the toxicity of the transcription factor GATA-1 to bacteria." Biotechniques **20**(4): 684-693.

Tuschl, T. and A. Borkhardt (2002). "Small interfering RNAs: a revolutionary tool for the analysis of gene function and gene therapy." Mol Interv **2**(3): 158-167.

Vaidya, S., K. Ghosh, et al. (2011). "Recent developments in drug resistance mechanism in chronic myeloid leukemia: A review." Eur J Haematol.

Van Der Burg M., T.S. Poulsen, et al. (2004). "Split-Signal FISH for Detection of Chromosome Aberrations in Acute Lymphoblastic Leukemia." Leukemia **18** 895-908.

Van Etten, R. A. (2002). "Studying the pathogenesis of BCR-ABL+ leukemia in mice." Oncogene **21**(56): 8643-8651.

Wang, H. W., C. Noland, et al. (2009). "Structural insights into RNA processing by the human RISC-loading complex." Nat Struct Mol Biol **16**(11): 1148-1153.

Wang, J. C., T. Lapidot, et al. (1998). "High level engraftment of NOD/SCID mice by primitive normal and leukemic hematopoietic cells from patients with chronic myeloid leukemia in chronic phase." Blood **91**(7): 2406-2414.

Wang, Q. Z., Y. H. Lv, et al. (2008). "The design of vectors for RNAi delivery system." Curr Pharm Des **14**(13): 1327-1340.

Weisberg, E. and J. D. Griffin (2000). "Mechanism of resistance to the ABL tyrosine kinase inhibitor STI571 in BCR/ABL-transformed hematopoietic cell lines." Blood **95**(11): 3498-3505.

Wertheim, J. A., J. P. Miller, et al. (2002). "The biology of chronic myelogenous leukemia: mouse models and cell adhesion." Oncogene **21**(56): 8612-8628.

Wilda, M., U. Fuchs, et al. (2002). "Killing of leukemic cells with a BCR/ABL fusion gene by RNA interference (RNAi)." Oncogene **21**(37): 5716-5724.

Withey, J. M., S. B. Marley, et al. (2005). "Targeting primary human leukaemia cells with RNA interference: Bcr-Abl targeting inhibits myeloid progenitor self-renewal in chronic myeloid leukaemia cells." Br J Haematol **129**(3): 377-380.

Wohlbold, L., H. van der Kuip, et al. (2003). "Inhibition of bcr-abl gene expression by small interfering RNA sensitizes for imatinib mesylate (STI571)." Blood **102**(6): 2236-2239.

Wu, M. T., R. H. Wu, et al. (2005). "Simple and efficient DNA vector-based RNAi systems in mammalian cells." Biochem Biophys Res Commun **330**(1): 53-59.

Zaree Mahmoodabady, A., H. R. Javadi, et al. (2010). "Bcr-abl silencing by specific small-interference RNA expression vector as a potential treatment for chronic myeloid leukemia." Iran Biomed J **14**(1-2): 1-8.

Zhao, X., S. Ghaffari, et al. (2002). "Structure of the Bcr-Abl oncoprotein oligomerization domain." Nat Struct Biol **9**(2): 117-120.

Zhu, Q. S., N. Heisterkamp, et al. (1990). "Unique organization of the human BCR gene promoter." Nucleic Acids Res **18**(23): 7119-7125.

Zhu, X., L. Wang, et al. (2011). "TGF-beta1-induced PI3K/Akt/NF-kappaB/MMP9 signalling pathway is activated in Philadelphia chromosome-positive chronic myeloid leukaemia hemangioblasts." J Biochem **149**(4): 405-414.

Xu, N., Z. Ouyang, et al. (2011). "[Clinical significance of CRKL protein phosphorylation level in the treatment of chronic myeloid leukemia with imatinib]." Zhonghua Xue Ye Xue Za Zhi **32**(1): 25-28.