The Role of Mitochondrial Folate Enzymes in Cancer

Aoife MacCooey

MSc 2015

The Role of Mitochondrial Folate Enzymes in Cancer

Aoife MacCooey B.Sc.

Supervisor: Dr Anne Parle-McDermott

School of Biotechnology, Dublin City University

A thesis submitted for the degree of Master of Science

September 2015

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of MSc is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ (Candidate) ID No.: _____ Date: _____

Table of Contents

Abbreviations

- **ALDH51A** Aldehyde Dehydrogenase51
- **APS** Ammonium Persulfate
- **ATIC** 5-Aminoimidazole-4-Carboxamide Ribonucleotide Formyltransferase
- **bp** Base Pair
- **cDNA** Complementary Deoxyribonucleic Acid
- **CVD** Cardiovascular Disease
- **DHF** Dihydrofolate
- **DHFR** Dihydrofolate Reductase
- **DHFRL1** Dihydrofolate Reductase like-1
- **DIP** Deletion Insertion Polymorphism
- **DMEM** Dulbecco's Modified Eagle Medium
- **DNA** Deoxyribonucleic Acid
- **DNTPs** Deoxy Nucleotide Phosphate
- **FBS** Fetal Bovine Serum
- **GAPDH** Glyceraldehyde 3-phosphate dehydrogenase
- **GC-MS** Gas Chromatography–Mass Spectrometry
- **GUS** Beta-Glucuronidase
- **HEK** Human Embryonic Kidney
- **LB** Luria Bertani
- **miRNA** MicroRNA
- **mRNA** messenger RNA
- **mtDNA** Mitochondrial DNA
- **MTHFD1** Methylenetetrahydrofolate dehydrogenase ($NADP^+$ dependent) 1
- **MTHFD2** Methylenetetrahydrofolate dehydrogenase ($NADP^+$ dependent) 2
- **MTHFD1L** Methylenetetrahydrofolate dehydrogenase ($NADP^+$ dependent) 1-like
- **MTHFD2L** Methylenetetrahydrofolate Dehydrogenase ($NADP^+$ Dependent) 2-Like

- **NTD** Neural Tube Defect
- **ORF** Open Reading Frame
- **PCR** Polymerase Chain Reaction
- **PVDF** Polyvinylidene Difluoride
- **RNA** Ribonucleic Acid
- **RT-qPCR** Reverse Transcriptase - Quantitative Polymerase Chain
- **RPS** Ribosomal Protein S13
- **SAM** S-Adenosylmethionine
- **SDS PAGE** Sodium dodecyl Sulfate Polyacrylamide Gel Electrophoresis
- **SHMT1** Serine Hydroxymethyltransferase (cytoplasmic isoform) 1
- **SHMT2** Serine Hydroxymethyltransferase (mitochondrial isoform)2
- **SNP** Singular Nucleotide Polymorphism
- **TBT** TATA Binding Protein
- **TEMED** Tetramethylethylenediamine
- **THF** Tetrahydrofolate
- **TYMS** Thymidylate Synthase
- **UPL** Universal Probe Library

Acknowledgements

I would like to sincerely thank my supervisor Dr Anne Parle-McDermott for her continued support and guidance throughout these past years, her door was always open. It really was very much appreciated :-)!! I would like to thank my post Doc Dr Linda Hughes for all her help and guidance also and for her great stories which always gave me a laugh! I would like to thank my former colleagues from the Nutritional Genomics group, Dr Mari Ozaki-Proof good things really do come in small packages, Dr. Stefano Minguzzi "Mr Fix" he can fix any machine none too big or small and Dr Alan Harrison the "funny man" famed for his great jokes in all the land ;-).

I would also really like to thank my family for their continued love and support throughout this whole process. I would especially like to thank my mum who has always been there for me with buckets of support. I would also really like to thank my love and partner in crime Gary for his love support and for always listening to me talk about biology even though he has no clue what I am saying, though he may have learned something these past few years!!

The Role of Mitochondrial Folate Enzymes in Cancer

By Aoife MacCooey

Abstract

Folic acid is an essential B vitamin, the metabolism of folic acid via one carbon metabolism results in the production of important components for the cell, such as DNA bases and methyl donor groups. The importance of mitochondrial one carbon metabolism has recently been highlighted with the discovery of the novel enzyme Dihydrofolate Reductase like 1 (DHFRL1) and the association of other mitochondrial enzymes, in particular Methylenetetrahydrofolate dehydrogenase ($NADP^+$ dependent) 1-like (MTHFD1L) with rapid proliferation and mortality in cancer. The novel DHFRL1 enzyme has been shown to be a much less active enzyme with a reduced affinity for dihydrofolate relative to DHFR. The MTHFD1L enzyme is responsible for the last step in the production of formate for cytoplasmic one carbon metabolism. Due to MTHFD1L's associations with rapid proliferation rate in cancer and mortality, a biomarker for its expression is desirable. Formate analysis by Gas chromatography–mass spectrometry (GC-MS) in Human Embryonic Kidney (HEK) 293 cells with modulated MTHFD1L expression demonstrated that knocking down the gene resulted in reduced formate levels and reduced cell growth. Similarly overexpressing the MTHFD1L gene in HEK 293 cells resulted in an increased formate level and growth rate relative to the controls. Investigation was undertaken into the amino acid differences between DHFR and DHFRL1 to begin to understand their functional relevance. It was identified that the arginine at amino acid position 24 may result in DHFRL1 having an altered structure, which may account in part for DHFRL1's reduced affinity for dihydrofolate. In addition, DHFRL1 and other mitochondrial folate enzymes, MTHFD1L, Methylenetetrahydrofolate dehydrogenase ($NADP^+$ dependent) 2 (MTHFD2), Serine Hydroxymethyltransferase (mitochondrial isoform) 2 (SHMT2) were found to be up-regulated in a metastatic cancer cell line but their cytosolic paralogues showed no such up-regulation. The results presented provide further evidence of the mitochondrial driven role in cancer progression and are supportive for the use of formate as a biomarker for mitochondrial gene up-regulation.

Chapter 1

Introduction

1.0 Overview

Cancer may be defined as the uncontrolled growth of cells and is a leading cause of death in Ireland and worldwide (WHO 2015a). The molecular mechanisms which underpin cancer are constantly being uncovered through cancer research. One such area of research which has recently garnered interest within the scientific community is cancer cell metabolism, specifically the expression and functionality of mitochondrial folate metabolic enzymes in relation to cancer (Jain *et al.* 2012, McEntee *et al.* 2011, Selcuklu *et al.* 2012, Nilsson *et al.* 2014)

One carbon metabolism is an essential biochemical pathway required for folate metabolism. One carbon metabolism is compartmentalised within the cell between the cytoplasm, the mitochondria and the nucleus and is responsible for the production of DNA bases, thymidine, adenine and guanine and the re-methylation of homocysteine to methionine for cellular methylation reactions (Fox and Stover 2008). Folate is an essential nutrient as it cannot be synthesised de novo and so it must be consumed through the diet from foods naturally rich in folate, folic acid fortified foods and supplements. The enzyme Dihydrofolate Reductase (DHFR) is responsible for converting folic acid and dihydrofolate to its biologicaly active form Tetrahydrofolate (THF) within the cell (Stover 2009). The conversion of folate to its biologically active form THF sets off a cascade of biochemical reactions resulting in the production of purines, thymidylate and methionine (Fox and Stover 2008). The essentiality of the DHFR enzyme in initiating this enzyme cascade has led to it being targeted with anti-folate chemotherapeutics in the treatment of cancer (Visentin *et al.* 2012). Up until recently it was thought there was only one active and expressed DHFR enzyme and a number of pseudogenes; however, it has come to light that there is a second expressed DHFR enzyme, DHFRL1 (McEntee *et al.* 2011, Anderson *et al.* 2011). DHFRL1 has been found to localise to the mitochondria where it is thought to partake in de novo thymidylate synthesis (Anderson *et al.* 2011). DHFRL1's activity has been shown to be dramatically lower than that of DHFR (McEntee *et al.* 2011). The presence of a second less active dihydrofolate enzyme which localises to the mitochondria may have implications for health and disease.

Although there is only an 8% difference in the protein sequences of DHFR and DHFRL1, it is thought that these differences contribute to the apparent different activities of the enzymes (McEntee *et al.* 2011). Further investigation is needed into the specific amino acid differences between DHFR and DHFRL1 to begin to understand their functional relevance.

One carbon metabolism is reliant on the provision of one carbon donor molecules which largely come in the form of formate originating in the mitochondria. It is estimated that up to 75% of the one carbon donor molecules required for cytoplasmic one carbon metabolism are derived from the mitochondria (Pike *et al.* 2010). Methylenetetrahydrofolate dehydrogenase ($NADP^+$ dependent) 1-like (MTHFD1L) is the only enzyme capable of catalysing the last step in the production of formate in the mitochondria (Tibbetts and Appling 2010). MTHFD1L and other mitochondrial folate related enzymes Serine hydroxymethyltransferase 2 (SHMT2) and methylenetetrahydrofolate dehydrogenase ($NADP^+$ dependent) 2 (MTHFD2) have recently been associated with proliferation and mortality within cancer. (Jain *et al.* 2012, Selcuklu *et al.* 2012, Nilsson *et al.* 2014). There is evidence to suggest that the mitochondrial folate pathway is a driving force in cancer cell proliferation (Jain *et al.* 2012). The enzyme MTHFD1L is of particular interest in relation to this with it being responsible for the provision of formate for cytoplasmic one carbon metabolism (Pike *et al.* 2010). The aim of this of this research thesis is to investigate the role which mitochondrial folate related enzymes play within the cell and the implications of this in relation to health and disease, with a particular emphasis on cancer.

1.1 Folic Acid and Metabolism

Folic acid is an essential water soluble vitamin also known as vitamin B_9. Folic acid is the synthetic form of the naturally occurring folate. The recommended daily allowance (RDA) for folic acid is 400µg per day and 600µg per day for pregnant women (Bailey and Gregory 1999). Folate is found naturally in dark green leafy vegetables, pulses and grains. Folic acid the synthetic variety of folate, is found in vitamin supplements and fortified foods such as breads and cereals. Both synthetic folic acid and naturally occurring folate may be referred to with the generic term

folate. Due to the essential nature of folate, both its consumption and the genes that metabolise it have been implicated in many diseases such as Neural tube defects (NTDs), Cancer, Cardiovascular Disease (CVD) and Alzheimer's (Parle-McDermott *et al.* 2009, Kirke *et al.* 2004, Sugiura *et al.* 2004, Jain *et a.l* 2012). In order to fully understand the mechanisms by which folate and its related genes are involved in the aetiology of such diseases a comprehensive view of the way in which folate is metabolised is needed.

1.1.1 Overview of One Carbon Metabolism

Folate is metabolised via the metabolic pathway known as one carbon metabolism. One carbon metabolism is a complex system of interdependent metabolic pathways which results in the synthesis of purines, thymidylate and the remethylation of homocysteine to methionine for cellular methylation reactions (Fox and Stover 2008). One carbon metabolism is compartmentalised within the cell between the mitochondria, the nucleus and the cytoplasm, Figure 1.1

Once folate is recruited into the cell it undergoes polyglutamation by the enzyme Folyl polyglutamatesynthase (FPGS) which involves the addition of up to eight glutamate molecules to prevent efflux from the cell. Folic acid is first metabolised by the enzyme Dihydrofolate Reductase (DHFR) to dihydrofolate (DHF) with NADPH as a co-factor. DHF is then reduced to tetrahydrofolate (THF), the biologically active form of folate. THF in the mitochondria is converted into 5,10 Methylene-THF by condensation of formaldehyde produced from the catabolism of glycine, serine, sarcosine and dimethylglycine, Figure 1.1 (Stover 2009). The 5,10 methylene-THF is then oxidised by the enzyme MTHFD2/L to form 10-formyl-THF which can then either be used to formylate MET-tRNA for protein synthesis within the mitochondria or it can be hydrolysed by the enzyme MTHFD1L to THF and formate (Stover 2009). Formate originating from the mitochondria is estimated to provide up to 75% of the one carbon units for cytoplasmic one carbon metabolism (Pike *et al.* 2010). Purines are produced by the incorporation of 10-formyl THF (produced from formate and THF) into the C2 and C8 carbons of the purine ring in the cytoplasm. 10-formyl THF can be reduced to 5,10 methylene-THF which is utilised by Thymidylate Synthase (TYMS) for the conversion of deoxy uridine

monophosphate (dUMP) to deoxy thymidine monophosphate (dTMP) in the cytoplasm, nucleus and mitochondria for thymidylate synthesis (Stover 2009). 5,10 methylene-THF can also be produced by the enzymatic conversion of serine and THF; the cleavage of serine in this reaction also serves to produce the amino acid glycine. 5, 10 methylene THF is further reduced to 5- methyl-THF which serves in the re-methylation of homocysteine to methionine for cellular methylation of DNA and proteins. In the nucleus 5,10 methylene-THF transfers its one carbon unit to dUMP generating dTMP which also results in the regeneration of DHF, Figure 1.1 (Stover 2009, Fox and Stover 2008).

As can be seen from Figure 1.1 the metabolism of folic acid is a complex process involving many enzymatic reactions resulting in the production of very important components for the cell, such as DNA bases and methyl donor groups (Fox and Stover 2008). Folic acid is an essential nutrient so its consumption and metabolism are of the utmost importance to the cell. Both low folate status and alterations in folate metabolising enzymes have been associated with disease development, as explored below in section 1.2.

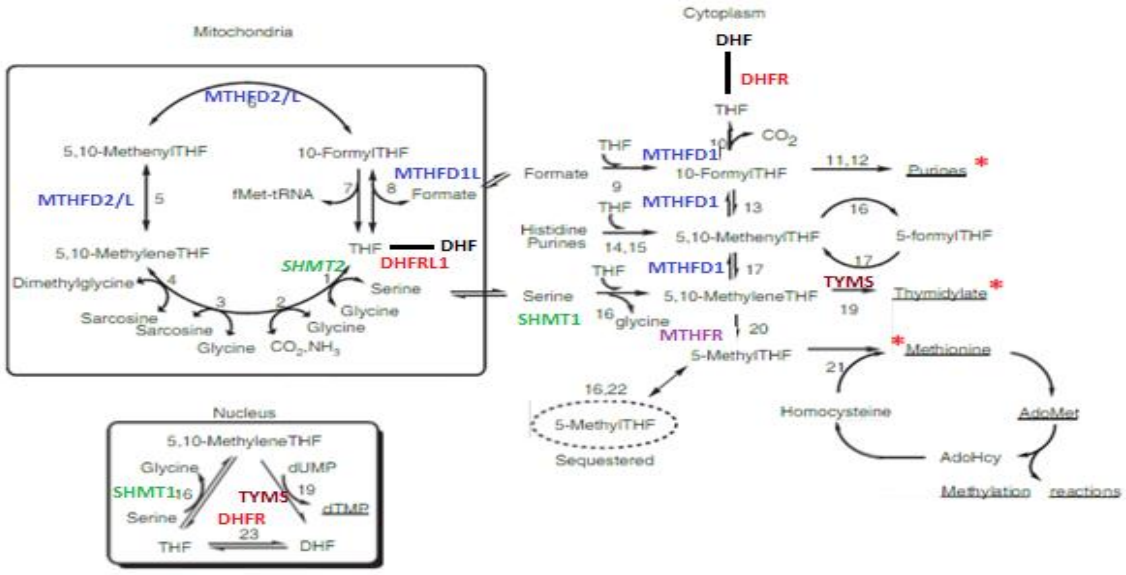


Figure 1.1 Compartmentalisation of folate metabolism. Folate metabolism is compartmentalised between the nucleus, cytoplasm and the nucleus. The active form of folate, THF, along with activated one carbon units undergo a series of enzymatic conversions which results in the production of purines, thymidylate and the re-methylation of homocysteine to methione. Paralogous enzymes are represented in the same colour; for example the MTHFD gene family consisting of mitochondrial MTHFD1L, MTHFD2, MTHFD2L and cytoplasmic MTHFD1 are all denoted in blue. * Denotes the main One Carbon Metabolism products. Image edited from Fox and Stover 2008.

Folate plays an integral role within the cell and without its consumption the health and wellbeing of the cell/tissue/person is compromised and disease states can arise. Folate metabolism results in the production of crucial products for the cell without which the cell cannot survive. Any disruption or alteration in folate intake or metabolism can have a deleterious effect; it is for this reason that folate has been implicated in playing a role in many diseases, as outlined below.

1.2.1 Neural Tube Defects

One of the most well-known conditions associated with low folate status are neural tube defects (NTDs). Folate is needed in times of rapid growth such as in the case of embryogenesis; perturbations in either folate consumption or metabolism can result in the development of congenital malformations such as NTDS (Fekete *et al.* 2010). NTDs occur when the neural tube in a developing embryo fails to close properly, resulting in a range of conditions which vary in severity, including anencephaly, spina-bifida, encephalocele, cranioschisis and iniencephaly (Botto *et al* .1999). Periconceptional folic acid intake has been shown to reduce the occurrence of NTDs by up to 72% (MRC Vitamin Research group 1991). The identification of the link between folate consumption and a decrease in NTDs resulted in the widespread fortification of many foods such as grains, cereals and milk both mandatorily and voluntarily throughout the world. The Centre for Disease Control and Prevention estimates that one third of flour produced throughout the world is fortified with folic acid (Centre for Disease Control and Prevention 2008). The Food Safety Authority of Ireland (FSAI) recommended mandatory fortification of our flour in 2006 due to the high incidence of NTD's (Food Safety Authority of Ireland 2006). However, in 2009 the FSAI decided to not proceed with mandatory fortification due to the decrease in NTD's which went from 1-1.5 to 0.93 per 1000 births (Food Safety Authority of Ireland 2009). This decrease in NTD occurrence was attributed to the fact that folate consumption had increased by 30% between 2006 and 2009 due to liberal voluntary fortification of a range of food products with folic acid in Ireland during the preceding years (Flynn *et al.* 2008). However, a recent publication by McDonnell *et al* (2015) found that between 2009 and 2011 NTD incidence in Ireland has increased to 1.04 per 1000

births. This increase in NTD rate is not statistically significant but does highlight the need for continued monitoring.

Aside from dietary factors there are also genetic elements at play in NTD development. It has been shown that people of different genetic origins appear to have different NTD rates; for instance Irish people experience a higher incidence than many other ethnic groups (Frey and Houser 2003). There have been over 80 genes implicated to varying degrees in NTD development (Green *et al*. 2009). Polymorphisms in folate related genes have been shown to impact the risk of NTD development; genes of particular interest are outlined below.

The folate metabolising enzyme MTHFR has a common polymorphism MTHFR 677C<T (Frosst *et al.* 1995). The homozygous TT variant of the gene results in the production of a thermo labile unstable enzyme. Both the TT and CT variants are risk factors for NTD development (Kirke *et al.* 2004). A 19 bp Deletion Insertion Polymorphism (DIP) in a noncoding region of the gene of the DHFR gene has also been associated with NTD risk. There have been conflicting reports as to the effect or functionality of this polymorphism with Johnson *et al* (2004) associating it as a risk factor for NTD development and Parle-McDermott *et al* (2007) associating it as protective against NTD development. In addition, a DIP in the MTHFD1L gene (rs3832406) and a single nucleotide polymorphism (SNP) in the MTHFD1 gene (R653Q) have also been associated as risk factors for NTD development (Parle-McDermott *et al.* 2009, Brody *et al.* 2002).

The exact mechanisms underlying the role of folate and its metabolising enzymes in NTD development are not fully understood. However, it appears that NTD risk and folate are closely linked (Beaudin and Stover 2009). NTD development appears to be multi-factorial with both genetics and environmental factors such as nutrition and folate status all contributing to the overall risk (either increasing or decreasing) (Beaudin and Stover 2009, Copp and Greene 2010).

1.2.2 Cardiovascular Disease

Folate status has also been implicated in playing a role in Cardiovascular Disease (CVD). CVD is the number one single cause of death worldwide (WHO 2015b). One of the leading types of CVD is coronary heart disease (CHD).There has been much debate over the role, if any, that folate plays in Coronary heart disease (CHD) development. Elevated homocysteine levels are associated with CHD risk (Marcus *et al.* 2007). Whether elevated homocysteine is causal or consequential of CHD is still up for debate. Early observational studies into the effect of homocysteine-lowering vitamins such as folate (recall how folate is required for the re-methylation of homocysteine to methionine) were positive (Clarke *et al.* 1991). A recent meta-analysis of 8 trials comprising of a total of 37,485 individuals involving B vitamin supplementation on the risk of CHD found there was no effect of B vitamin supplementation (inclusive of folate) on either CHD or stroke risk (Clarke *et al.* 2011). However, that the Meta analysis carried out by Clarke *et al* (2011) measured the occurrence of secondary events so the role of folic acid in primary prevention remains to be further elucidated.

In addition to the role of dietary folate, polymorphisms in folate related genes have also been associated with CHD risk. Work by Samani *et al* (2007) identified a polymorphism (rs6922269) within the MTHFD1L gene as a risk factor for CHD development, with as much as 23% increased risk associated per copy of the risk allele. The MTHFR 677C<T polymorphism is known to cause elevated homocysteine in individuals who are homozygous TT and has been implicated in CHD risk. A meta-analysis by Klerk *et al* (2002) found that individuals who were homozygous TT had a significantly increased risk of CHD compared to CC wild type homozygotes, particularly in a low folate setting. A more recent meta-analysis by Husemoen *et al* (2014) into the effect of the MTHFR 677C<T polymorphism on CVD found no causal relationship between homocysteine and CVD, consistent with the findings of Clarke *et al* (2011). However Husemoen *et al* (2014) couldn't exclude a causal relationship between the MTHFR polymorphism and ischemic heart disease.

1.2.3 Neurodegenerative Diseases

Folate status has also been linked to neurodegenerative diseases such as Alzheimer's and Parkinson's. Elevated homocysteine levels have been shown to correlate with poor cognitive function, Alzheimer's and dementia (Seshadri *et al.* 2002). Bryan *et al* (2002) found that folic acid supplementation resulted in improved memory processing. A randomised double blind control trial by Durga *et al* (2007) found that folic acid supplementation resulted in increased cognitive function. However, a Cochrane review into folic acid supplementation with or without Vitamin B 12 found that there is no consistent evidence to support or refute the role which folic acid and vitamin B 12 may play in dementia and Alzheimer's disease (Malouf *et al.* 2008).

As with many multi-factorial diseases, there is no one source or definitive cause of neurodegenerative disease; aside from nutrition genes are also thought to play a role in disease development. It is thought that the polymorphisms in both the MTHFD1L and MTHFR genes impact on homocysteine levels, with elevated homocysteine levels being associated with neurodegenerative diseases. A genome wide association study (GWAS) identified a single nucleotide polymorphism (SNP) rs11754661 in the MTHFD1L gene as being significantly associated with late onset Alzheimer's disease in a Caucasian population (Naj *et al.* 2010). Work by Ren *et al* (2011) found a similar association of the MTHFD1L SNP in a Han Chinese population. However, work by Ramirez-Lorca *et al* (2011) failed to find an association in a Spanish population, indicating that the effect may be population specific. The MTHFR 677C<T polymorphism has also been implicated as playing a role in neurodegenerative diseases. A recent meta-analysis by Zhu *et al* (2012) identified the 677C<T polymorphism of the MTHFR gene as being a risk factor for the development of Parkinson's disease. The findings outlined above again highlight the difficulty in identifying and confirming one specific contributor in the development and progression of multi-factorial diseases such as neuro-degenerative diseases and heart disease.

1.3 Folate and Cancer

Cancer is a leading cause of mortality worldwide. An average of 36,000 new cases of cancer were registered with the Irish National Cancer Registry between 2010-2012. The incidence of cancer in Ireland between 2010-2012 was 790 male and 747 female cases per 100,000 people per year. Although the incidence rate of cancer has increased between 1994-2012 the mortality rates have decreased by 1.4% and 1.1% for men and women respectively (National Cancer Registry 2014). The increased incidence and decreased mortality in Ireland over the last number of years demonstrates the positive impact cancer research has had on the advancement of techniques for cancer detection and treatment.

The role which folate and folate metabolising genes play within cancer development and progression is complex and has a history spanning over 60 years (Miller 2006). Cancer is another example of a multi-factorial disease, with genetics and environmental factors at play with exposure to chemicals and toxins, lifestyle factors and nutrition all affecting the risk or chance of cancer development. From a dietary point of view, much controversy exists over the role which folates play in cancer development, with evidence to support it both as a cancer preventative and a cancer causing agent (Terry *et al.* 2002, Larsson *et al.* 2004 Lajous *et al.* 2006, Cole *et al.* 2007). From a genetic point of view the primary folate metabolising enzyme DHFR has been targeted with anti-folate chemotherapy drugs in the treatment of cancer for many years (Visentin *et al*. 2012). However, our knowledge regarding folate genes and cancer is currently evolving with much renewed interest in the area, particularly towards mitochondrial folate genes which is the main focus of this research thesis. In order to get a comprehensive view of the relationship between folate and cancer, information regarding dietary folate intake and cancer risk and development is discussed below.

1.3.1 Dietary Folate and Cancer

Folate appears to exert a double mechanistic action on cancer growth and development. Due to the fact that folate metabolism leads to the production of DNA bases which are required in high amounts by rapidly proliferating cancer cells it can be easily understood how folate consumption can aid in cancer progression. Although folates promote the growth of pre-existing cancer there is much evidence to support that folate intake lowers cancer development risk i.e., the appearance of the primary tumour (Farber 1947, Giovanucci *et al.* 1998). Studies such as the nurse's study who recorded the dietary folate intake by semi quantitative questionnaire of 121,700 women over a 15 year period found that folate intake was correlated with a decreased cancer risk, with 400µg intake per day having a relative risk value of 0.69 (Giovannucci *et al.* 1998). Similar studies into the role of folate and primary cancer incidence by Terry *et al* (2002), Larsson *et al* (2004) and Lajous *et al* (2006) all correlated dietary folate intake with a decrease in cancer risk.

A clinical trial set up by Cole *et al* (2007) in 1994 set out to further prove the protective effect of folate in cancer, specifically in colon cancer. The trial was set up to assess the preventative effect of folic acid on the re-occurrence of adenomas (a pre-cancerous lesion in the colon) in individuals with a recent history of them. The results of the trial were quite un-expected, as it was found that folic acid consumption was associated with an increase in cancer risk (although this was not primary prevention). Cole *et al* (2007) found that folic acid supplementation was associated with increased adenoma incidence and an increased incidence of advanced lesions (Cole *et al.* 2007). The findings of the adenoma trial led to great concern about the implications of folate fortification of our food and this finding was highlighted by the FSAI as a one of the reasons for not undertaking mandatory folate fortification in 2009 (FSAI 2009). Figueiredo *et al* (2009) analysed the adenoma trial data in regard to the development of prostate cancer and found that men who received folic acid had an increased risk of developing prostate cancer relative to those who received a placebo. Analysis of clinical trial data from Norway by Ebbing *et al* (2009) found that those who received folic acid had an increased incidence of cancer and increased mortality relative to those who received a

placebo. In order to clarify the role of folates in cancer development a meta-analysis was undertaken by Vollset *et al* (2013) into the site specific cancer incidence in 50,000 individuals. The meta-analysis found that collectively there was no significant effect of folic acid supplementation either increasing or decreasing cancer risk over a 5 year period. Although, the authors did note that there was a slight increase in cancer risk in those supplemented with folic acid but that the increase could just as easily be contributed to random chance.

There are plausible molecular mechanisms to support folates as increasing and decreasing cancer risk. It appears that a delicate balance of folate is required. Folate deficiency disrupts the conversion of dUMP to dTMP leading to a shortage of thymidylate and an accumulation of uracil. Thymidylate is not strictly required for DNA synthesis as uracil can take its place (Blount *et al*. 1997, MacFarlane *et al.* 2011). The incorporation of uracil into DNA results in a number of excision and repair steps which if performed repeatedly can lead to chromosomal aberrations and strand breaks which may lead to carcinogenesis (Duthie 1999). In addition to this, a deficiency in folate may also impact cellular methylation as the folate derivative 5-methyl-THF is required for the conversion of homocysteine to methionine which in turn is converted to S-adenosylmethionine (SAM, referred to as Adomet in Figure 1.1) for cellular methylation reactions (Fox and Stover 2008). DNA methylation is associated with repression of gene expression. Hypo-methylation of important genes such as proto-oncogenes is a known feature of many cancers (Liang *et al.* 2015, Soes *et al.* 2014, Choi and Lee 2013). Hyper-methylation of tumour suppressor genes has also been associated with cancer: increased folate intake leads to the availability of methyl donor groups in theory allowing for hyper-methylation and repression of important genes (Berdasco and Esteller 2010). However, folate status and DNA methylation is not simply a cause and effect relationship. Evidence suggests that the effect nutrient availability has on DNA methylation is site specific and may depend on many factors such as gene, tissue, age etc. (Parle-McDermott and Ozaki 2011).

The relationship between folate and cancer is complex; folate deficiency can result in genetic damage and increases primary cancer risk, whereas folate consumption

when cancer has already been established results in accelerated growth and progression of the cancer (Kim 2007). Apart from folate consumption/status, more recent research has identified that the expression level of a number of mitochondrial folate enzymes is associated with increased proliferation and mortality rate in cancer. This is described further in Section 1.4 and is explored further in this research thesis.

1.4 Folate Metabolising Enzymes in Cancer and Health and Disease

The metabolism of folate by one carbon metabolism is reliant on the faithful activity of a plethora of enzymes, with one enzymes product acting as the next enzyme's substrate. One thing which is quite obvious from Figure 1.1 is that the compartmentalisation of folate metabolism within the cell has led to the evolution of a number of enzymes which share commonality but operate in different cellular locations. Genes or proteins that share function and sequence identity are known as paralogues and were created via gene duplication events (Koonin 2005). A number of paralogues operate in one carbon metabolism; for example, the enzyme serine hydroxymethyltransferase (SHMT) is responsible for the cleavage of serine to glycine. This reaction however, is carried out by different enzymes depending on cellular location, with SHMT2 carrying out reactions in the mitochondria and SHMT1 and SHMT2 α carrying out reactions in the cytoplasm and the nucleus (Anderson *et al.* 2012). The products of folate metabolism, purines, thymidylate and methionine are of paramount importance to the cell and are essential for the cell to grow and divide (Fox and Stover 2008).

The mitochondrion encompasses a number of key enzymes involved in one carbon metabolism. Figure 1.2 illustrates mitochondrial enzymes and demonstrates the roles carried out by these enzymes. The importance of mitochondrial folate enzymes is currently being uncovered within the scientific community. This is highlighted by the fact that the presence and functionality of two novel mitochondrial enzymes dihydrofolate reductase like 1 (DHFRL1) and Methylenetetrahydrofolate Dehydrogenase ($NADP^+$ Dependent) 2-Like (MTHFD2L)

was only discovered in 2011 (McEntee *et al.* 2011, Bolusani *et al.* 2011). Anderson *et al.* (2011) discovered the presence of a mitochondrial de novo thymidylate synthesis pathway with the newly discovered DHFRL1 enzyme thought to play a significant role in this pathway. McEntee *et al* (2011) demonstrated that the novel DHFRL1 enzyme is capable of reducing DHF to its biologically active form THF, albeit with a much reduced efficiency relative to DHFR. In addition to this, in 2012 mitochondrial folate enzymes were shown to correlate with cancer cell proliferation and mortality but cytoplasmic folate enzymes showed no such correlation (Jain *et al.* 2012). However, this was not the first time that mitochondrial folate enzymes were associated with cancer. The MTHFD1L enzyme was associated with cancer as early as 2004 by Sugiura *et al* (2004). As outlined above in section 1.2 MTHFD1L has been shown to play a role in many diseases such as NTDs, CHD and neurodegenerative diseases. The focus of this research thesis is on mitochondrial folate enzymes with a particular focus on MTHFD1L (due to its previous associations with cancer and human disease) and on DHFRL1, a newly discovered enzyme by the Nutritional Genomics group in 2011.

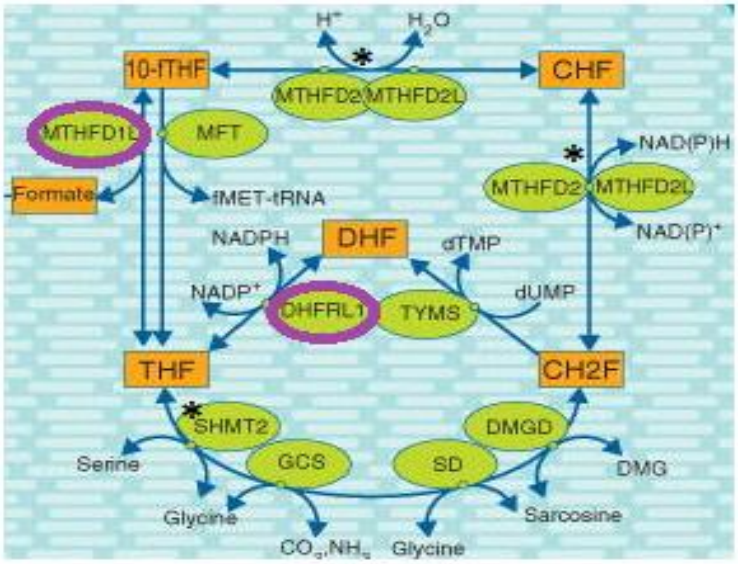


Figure 1.2 Mitochondrial One Carbon Metabolism. MTHFD1L is responsible for the conversion of 10-formyl-THF to formate which is then used as source of one carbon donor molecules in the cytoplasm. DHFRL1 is responsible for the reduction of DHF to THF for de novo thymidylate synthesis. SHMT2 cleaves serine to form glycine which is then further metabolised to form CH2F (also referred to as 5,10-methylene THF) which is used for thymidylate synthesis and as a substrate for MTHFD2/MTHFD2L. Genes of particular interest MTHFD1L and DHFRL1 are highlighted in purple. Genes of interest are denoted by*. Image edited from Scotti et al (2013).

Dihydrofolate Reductase (DHFR) is a folate metabolising enzyme responsible for initiating the one carbon metabolism enzyme cascade by reducing Dihydrofolate (DHF) to its biologically active form Tetrahydrofolate (THF) (Stover 2009). Up until 2011 it was widely accepted that the DHFR gene family consisted of one functional gene (DHFR) and four non-functional pseudogenes. The functional DHFR gene consists of six exons interspersed by five introns. In contrast, the four other members of the DHFR gene family are intronless and are thought to have originated from the re-integration of processed mRNA (Anagnou *et al.* 1984). Publications by McEntee *et al* (2011) and Anderson *et al* (2011) have changed the landscape of the DHFR gene family as we know it. It was proven that the former pseudogene DHFRP4 is an expressed enzyme known as dihydrofolate reductase like 1 or DHFRL1. Thus, proving that there is in fact nothing "pseudo" about DHFRL1 (DHFRP4) (McEntee 2011, Anderson 2011). The DHFR gene family is dispersed among the genome with DHFR, DHFRP1, DHFRP2, DHFRP3, and DHFRL1 (formerly DHFRP4) located on chromosomes 5, 18, 6, 2 and 3, respectively; see Figure 1.4 (Anagnou *et al.* 1988).

The role of DHFRL1 within the cell is thought to be in the provision of THF for de novo thymidylate synthesis within the mitochondria. The presence of de novo thymidylate synthesis within the mitochondria was confirmed by Anderson *et al* (2011). Both McEntee *et al* (2011) and Anderson *et al* (2011) reported that DHFRL1 was found to translocate to mitochondria despite the lack of a mitochondrial targeting sequence. This is in contrast to the DHFR enzyme which is thought to reside in the cytoplasm and the nucleus (Stover 2009). The enzymes DHFRL1, SHMT2 and TYMS are all involved in mitochondrial de novo thymidylate synthesis. DHFRL1 converts DHF to THF, one carbon activated molecules derived from serine and THF by SHMT2 are converted to thymidylate by (TYMS). Anderson *et al* (2011) demonstrated that de novo thymidylate synthesis is important to maintain mitochondrial DNA integrity (mtDNA). glyA Chinese hamster ovary (CHO) cells lacking the SHMT2 enzyme had 40% increased level of uracil in their mtDNA than the wild type (WT) cells. It was also found that transfection of the DHFRL1 gene into glyC CHO cells, an uncharacterised cell line auxotrophic for glycine, rescued the

glycine auxotrophy (Anderson *et al.* 2011). These results confirm the functional effects of DHFRL1 on mitochondrial one carbon metabolism.

On analysis of the amino acid residues of the DHFR and DHFRL1 enzymes, McEntee *et al* (2011) found that they share 92% homology, see Figure 1.3. Despite the homology between the two enzymes McEntee *et al* (2011) have demonstrated that the activity of DHFRL1 is much reduced in comparison to DHFR. DHFR has been shown to have a Km of 20.1μM for DHF, whereas DHFRL1 has been shown to have a much reduced affinity for DHF with a Km of 209.3μM. McEntee *et al* (2011) found that DHFRL1 was capable of compensating for DHFR knockout in both a mammalian and a bacterial system. DHFRL1 may also have a role in the control and expression of DHFR. Control and expression of the DHFR enzyme is tightly regulated, as disruption of DHFR supply can have serious effects on one carbon metabolism (Martianov 2007). At a post transcriptional level DHFR has been found to auto regulate its own expression by binding its own mRNA, thus preventing translation. One of the most significant findings in relation to DHFRL1 is the fact it has been shown to not only auto regulate its own expression by binding of its own mRNA but it is also capable of binding the mRNA of DHFR. Likewise DHFR has also found to be capable of binding the mRNA of DHFRL1 (McEntee *et al.* 2011). DHFR has been targeted with anti-folate chemotherpeutics in the treatment of cancer for over sixty years (Miller 2006). Treatment of cancer with the anti-folate drug methotrexate initially results in an up regulation of the DHFR enzyme due to the fact that the methotrexate binds the DHFR enzyme causing disassociation of the mRNA, leading to translation of the released mRNA (Ercikan-Abali *et al.* 1997). The different affinities of DHFR and DHFRL1 for folate suggest that they may have different affinities to anti-folate inhibitors. If methotrexate is not an effective inhibitor of DHFRL1 it could potentially hold onto the mRNA of DHFR effecting its expression. Investigation into the expression of DHFRL1 in cancer would help in its evaluation as a viable chemotherapy target. It would also provide further insight into the expression of mitochondrial folate enzymes in cancer, which shall be discussed further in section 1.4.2. Differential expression such as in the case of over expression of enzymes proteins etc., is a desirable feature of chemotherapeutic

targets. Overexpression of genes and proteins generally gives an advantage to a cancer cell. Blocking this overexpression can have a deleterious effect on the cancer cell leading to cell death e.g. HER2 overexpression and its targeted treatment with Trastuzumab (Herceptin) in the treatment of breast cancer (Vogel *et al.* 2002).

There are 15 amino acid differences between the human DHFR and DHFRL1 enzymes. Single amino acid changes in the DHFR enzyme have previously been shown to impact the sensitivity of the enzyme to anti-folates. Indeed, site directed mutagenesis of amino acid residue 31 of the DHFR enzyme from a phenylalanine to a serine has been shown to result in an 11 fold reduction in catalytic efficiency and a 100 fold reduced affinity for the anti-folate methotrexate (Schweitzer *et al.* 1989). Similarly, it has been shown that mutating the leucine at position 22 of the DHFR enzyme to a phenylalanine results in an 88 fold reduced affinity for methotrexate (Ercikan-Abali *et al.* 1996). Both positions 22 and 31 of the DHFRL1 enzyme are the same as DHFR, so it is unlikely that these residues would be a source of diversity between the enzymes. It is not fully known what the impact of the amino acid differences in DHFRL1 and DHFR has on the enzymes. It is known that DHFRL1 has a reduced affinity for folate in comparison to DHFR. Due to the number of amino acid differences in DHFRL1 (relative to DHFR) it is likely that the differences impact the enzyme in many ways both structurally and functionally. McEntee *et al* (2011) noted that the arginine at position 24 of the DHFLR1 enzyme may be a source of the significant differences observed in this enzyme. Gao *et al* (2013) performed predictive protein modelling into the conformation of the DHFRL1 enzyme which revealed that its structure may be closer to that of *E.coli* DHFR than to human DHFR. Molecular dynamic simulations into substrate enzyme binding affinities indicated that mutating the arginine at position 24 of the DHFRL1 enzyme to a tryptophan (as found in DHFR) would restore DHFRL1's binding affinity for folate (Gao *et al.* 2013) Further characterisation and investigation into the impact of specific amino acid differences between DHFR and DHFRL1 is needed in order to determine the contributions of the amino acids to the overall structure and function of the enzymes. In addition, analysis of DHFRL1's expression in a cancer cell line

model will allow for the further characterisation of the emerging mitochondrial driven role within cancer, which shall be explored further in section 1.4.3.

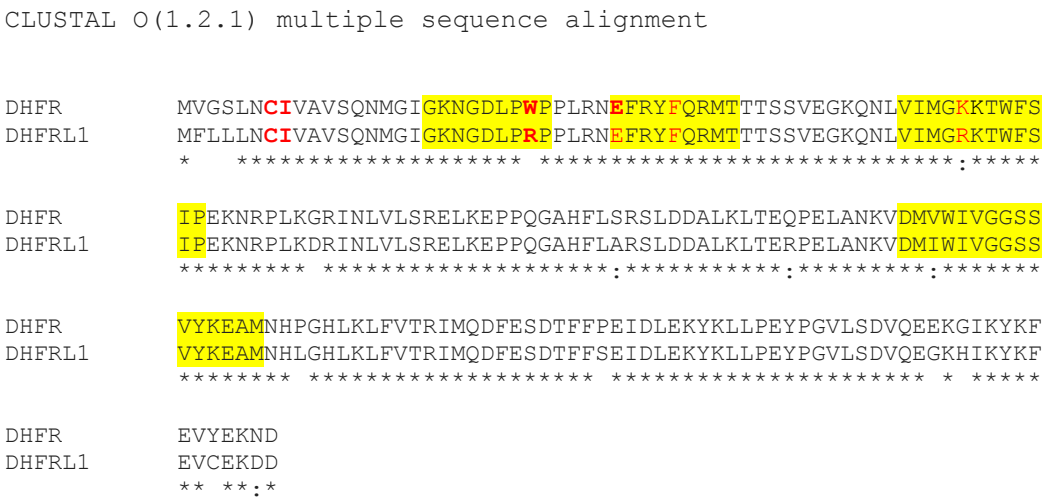


Figure 1.3 Clustal Omega alignment of DHFR and DHFRL1. There are 15 amino acid differences between DHFR and DHFRL1. The areas in yellow are the sequences required for catalytic activity of DHFR. Residues thought to be important for folate, NADPH and methotrexate binding are denoted in red (McEntee *et al* 2011).

DHFR Gene Family

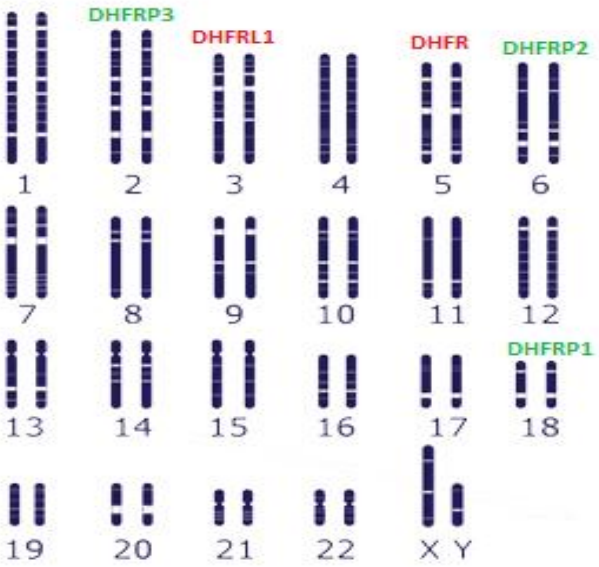


Figure 1.4 Chromosomal Location of the Human DHFR Gene Family. The two expressed enzymes DHFR and DHFRL1 are denoted in red and are located on chromosomes 3 and 5 respectively. The DHFR pseudo-genes, DHFRP1, DHFRP2 and DHFRP3 are denoted in green and are located on chromosomes 18, 6 and 2, respectively.

1.4.2 The Human MTHFD Gene family

Methylenetetrahydrofolate dehydrogenase ($NADP^+$ dependent) 1-like (MTHFD1L) is a mono-functional mitochondrial folate metabolising enzyme. MTHFD1L forms part of the MTHFD gene family consisting of the tri-functional cytoplasmic enzyme MTHFD1, and the bi-functional mitochondrial enzymes MTHFD2 and MTHFD2L, see Figure 1.5 (Tibbetts and Appling 2010). The MTHFD gene family is integral to the metabolism of folate via one carbon metabolism. Within the cytoplasm, the tri functional enzyme MTHFD1 utilises the one carbon units released from the mitochondria in the form of formate in three different ways. Formate is incorporated into the activated one carbon units cycle by its conversion to 10-formyl-THF by MTHFD1, thus providing the required substrate for purine synthesis. MTHFD1 can also reduce formate to 5,10-methylene-THF, where it can utilised in either thymidylate bio-synthesis or incorporated into the methyl cycle for the re-methylation of homocysteine to methionine (Tibbetts and Appling 2010). Within the mitochondria the enzymes MTHFD1L and MTHFD2/2L perform the 3 enzymatic reactions which the cytoplasmic MTHFD1 performs. MTHFD2 as stated previously is a bi-functional enzyme with 5,10-methenyl-THF (CH+-THF) cyclohydrolase and 5,10-methylene-THF (CH2-THF) dehydrogenase activity. MTHFD2 expression is absent in normal adult mammalian cells with the expression of the gene found in malignant transformed cells and in developing embryonic tissue where it is found to be essential (Tibbetts and Appling 2010). In 2011 Bolusani *et al* (2011) discovered the existence of the MTHFD2L gene responsible for the 5,10-methenyl-THF (CH+-THF) cyclohydrolase and 5,10-methylene-THF (CH2-THF) dehydrogenase activity within the mitochondria of normal adult mammalian cells .

It is estimated that up to 75% of cytoplasmic one carbon units originate in the mitochondria (Pike *et al.* 2010). The MTHFD1L gene performs the crucial last step in the supply of one carbon units for cytoplasmic one carbon metabolism by the conversion of 10-formyl-THF to formate (Tibbetts and Appling 2010). Since its discovery, MTHFD1L has been shown to play a role in many diseases such as NTDs, CHD and neurodegenerative diseases, as outlined in section 1.2. Most recently MTHFD1L has been associated with cancer, in particular the proliferation of rapidly

growing cancer cells. It appears that MTHFD1L, MTHFD2 and SHMT2 may work in conjunction with one another to allow for rapid proliferation and growth of cancer cells. Outlined below in section 1.4.3 is information regarding MTHFD1L and the appearance of a mitochondrial driven role within cancer.

1.4.3 Emergence of a Mitochondrial Driven Role in Cancer

MTHFD1L was first associated with cancer by Sugiura *et al* (2004) who found the gene to be up regulated in colon adenocarcinoma. Gene expression was analysed in 117 normal colon samples and 77 adenocarcinoma colon samples. It was found that MTHFD1L was 1.45 fold increased in benign adenoma tissue compared to normal colon tissue. However, the increase in MTHFD1L expression was more pronounced in the adenocarcinoma tissues with a 2.38 fold increase observed. The up-regulation of MTHFD1L in colon cancer was confirmed by examining its expression in normal colon cell lines (CCD841CoN) and cancerous colon cell lines (SW620 and HCT116). Once it was established that the over expression of MTHFD1L was linked to colon cancer, the authors set out to investigate what effect (if any) overexpressing MTHFD1L gene would have on the normal HEK 293 cell line. Interestingly it was found that over expression of MTHFD1L in HEK 293 led to increased "colony formation" or increased cell growth. Selcuklu *et al* (2012) found MTHFD1L to be up-regulated in breast cancer, where it is thought to be a target of miR 9, a micro RNA thought to possess tumour suppressor activity. Selcuklu *et al* (2012) also found MTHFD2 to be up regulated in breast cancer and found that the overexpression of miR 9 resulted in impaired cell growth and increased apoptotic activity. Most recently Jain *et al* (2012) associated MTHFD1L with cancer. The authors analysed the consumption and release (CORE) of 219 metabolites in the NCI-60 panel of cancer cell lines. Glycine consumption was found to be markedly increased in highly proliferative cancer cells whereas slowly proliferating cells were found to release glycine; see Figure 1.6. Glycine is a non-essential amino acid as it can be synthesised within the cell. Expression of the mitochondrial enzymes MTHFD1L, MTHFD2 and SHMT2 involved in the synthesis of glycine were associated with rapidly proliferating cancer cells whereas their cytoplasmic counterparts showed no such correlation, see Figure 1.6. It was established that glycine

consumption was a unique feature of rapidly proliferating cancer cells as normal cells with equivalent doubling times were found to release glycine (Jain *et al.* 2012).

Furthermore, Inhibition of SHMT2, a key enzyme in glycine biosynthesis, was found to greatly affect rapidly proliferating cells in the absence of glycine, whereas slowly proliferating cells remained largely unaffected. These findings indicate that glycine and its associated synthesising enzymes play a crucial role in driving cell growth and rapid proliferation. Moreover, Expression levels of these mitochondrial enzymes MTHFD1L, SHMT2 and MTHFD2 were analysed in a cohort of 1300 breast cancer patients. It was found that patients with above median level expression of mitochondrial enzymes (MTHFD1L, SHMT2 and MTHFD2) experienced greater mortality. These findings suggest that MTHFD1L and mitochondrial enzymes may play a key role in cancer growth and progression (Jain *et al.* 2012).

More evidence of the mitochondrial driven role in cancer cell growth and proliferation comes from Lehtinen *et al* (2013), who found that MTHFD2 was associated with increased mortality and poor outcome in breast cancer patients. Interestingly, it was shown that MTHFD2 may play a role in cancer invasion, migration and metastasis (Lehtinen *et al.* 2013). Most recently, A meta-analysis of available gene micro array sets by Nilsson *et al* (2014) encompassing over 20,000 genes and 1,981 tumour types in 19 cancer types identified MTHFD2 as the most highly scored protein associated with cancer. It was also found that knockdown of the MTHFD2 gene resulted in cell death in human cancer cells. Moreover, high expression of MTHFD2 was associated with increased mortality in breast cancer patients, consistent with previous associations of mitochondrial folate enzymes and increased mortality (Nilsson *et al.* 2014).

The uncovering of the up regulation of the mitochondrial folate pathway and its association with cancer presents us with the opportunity to further decipher the molecular mechanisms which occur during malignant transformation and cancer cell proliferation. Vasquez *et al* (2013) recently associated the expression of mitochondrial folate enzymes with increased methotrexate sensitivity in leukaemia patients, independent of DHFR expression. A biomarker for this mitochondrial

molecular gene signature would enable the stratification of patients based on their molecular profile into targeted drug therapies, hopefully leading to improved patients outcomes.

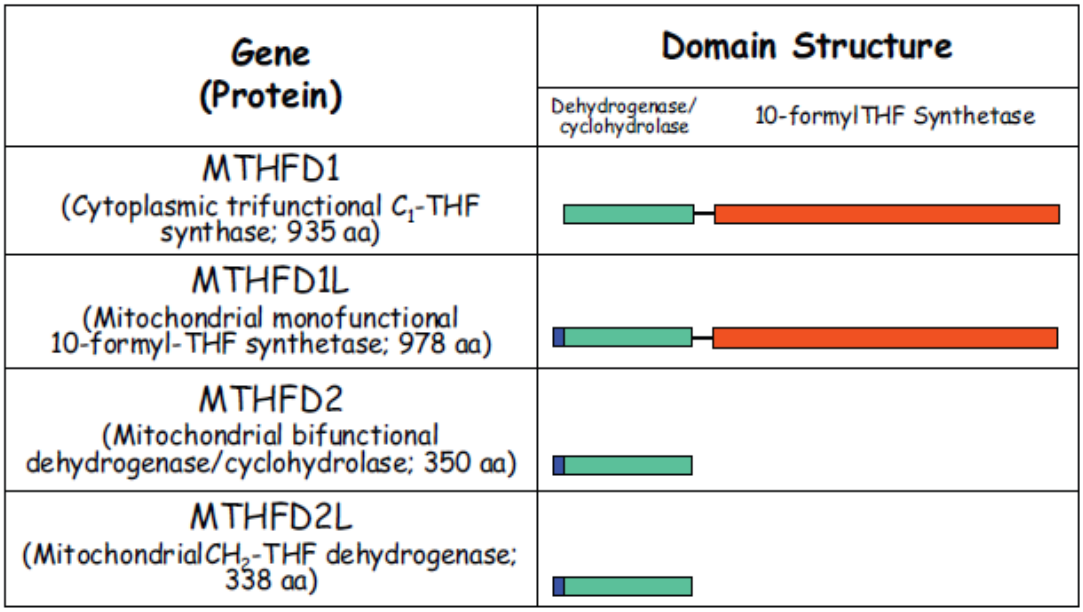


Figure 1.5 MTHFD Gene family domain structure. MTHFD1 is trifunctional with dehydrogenase, cyclohydrolase and 10-formyl THF synthetase activity. The dehydrogenase, cyclohydrolase domain of MTHFD1L is inactive meaning that it is monofunctional with 10-formyl THF synthetase activity. MTHFD2 and MTHFD2L are both bifunctional enzymes with dehydrogenase and cyclohydrolase activity. The blue domain represents the mitochondrial targeting sequence of MTHFD1L, MTHFD2 and MTHFD2L. Image taken from Tibbetts and Appling (2010).

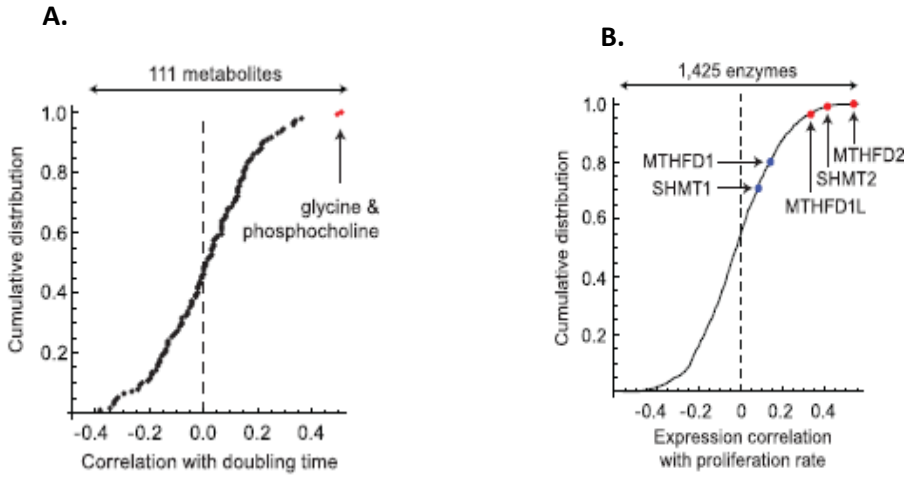


Figure 1.6. A. Correlation of metabolites with cell doubling times in the nci-60 panel. Glycine is highlighted as significant $p < 0.05$. **B.** Correlation of the nci-60 panel proliferation rate and expression of metabolic enzymes with MTHFD1L, MTHFD2 and SHMT2 are significantly correlated with proliferation rate but their cytoplasmic counterparts MTHFD1 and SHMT1 are non-significant. Image taken from Jain *et al* (2012).

1.5 Aims and Objectives

The overall aim of this thesis was to begin to elucidate the molecular mechanisms of mitochondrial folate enzymes in human disease. This aim was addressed through the following questions:

- **Aim 1:** Does manipulation of MTHFD1L expression have an impact on formate levels and cell growth? (Chapter 3).

  Objectives: Formate analysis by gas chromatography mass spectroscopy (GC-MS) in HEK 293 cells lines with modulated MTHFD1L expression.

- **Aim 2:** What is the functional relevance of amino acid 24 in DHFRL1? (Chapter 4)

  Objectives: Expression of recombinant DHFR, DHFRL1 and the mutant DHFR W24R and the DHFRL1 R24W proteins in an *E.coli* expression system. Analysis of the effect of these mutations on protein solubility and structure.

- **Aim 3:** Are folate enzymes differentially expressed in cancer metastasis? (Chapter 5).

  Objectives: RT-qPCR analysis of the folate related genes, DHFRL1, DHFR, MTHFD1, MTHFD2, MTHFD1L, SHMT1, SHMT2, TYMS, ATIC in primary and metastatic cancer cell lines.

Chapter 2

Methods and Materials

2.0 Materials

2.0.1 Protein Production and Analysis

Sigma Aldrich: Bis-acrylamide 40% w/v (Cat No. A9926), Ammonium Persulfate (Cat No. A3678-100G), Sodium Dodecyl Sulphate (Cat No. 71725), Dimethyl Sulfoxide (DMSO) (Cat No. D8418), EDTA (Cat No. EDS), Imidazole (Cat No. I5513), LB Broth (Cat No. L3033), Kanamycin (Cat No. K4000) Ampicillin (Cat No. A0166), Agar (Cat No. A5306), Yeast Extract (Cat No. Y1333), Tryptone (Cat No. T7293), NaCl (Cat No. S7653), L- Arabinose (Cat No. A3256),Gluthathione Agarose Beads (Cat No. G4510), Gluthathione (Cat No. G6013), Acetic Acid (Cat No. A6283), 2- Meracaptoethanol, (Cat No. M6250), Bradford Reagent (Cat No. B6916), Bovine Serum Albumin (Cat No. A3294),Glycine (Cat No. G8898), Tween (Cat No. P1379), Triton X (Cat No. X100), ColorBurst protein ladder (Cat No. C1992-1VL), His Select Nickel Affinity Gel (Cat No. P6611), Anti-Rabbit HRP secondary antibody (Cat No. A0545).

Clontech: HisTALON™ Gravity Column Purification Kit (Cat No. 635654)

Invitrogen: Bl21AI One Shot Chemically Competent Cells (Cat No. C6070-03),pDEST 15 Plasmid Vector (Cat No. 11802-014), pDEST 17 plamsid vector (Cat No11803-012). BP Clonase Enzyme (Cat No. 11789-020), LR Clonase Enzyme (Cat No. 11791-020), pDONR 221 (Cat No. 12536-017).

Bioline: Alpha Gold Efficiency Competent Cells (Cat No. BIO– 85027).

Thermo Scientific: 4- 20% w/v Tris Glycine Gels (Cat No. 25269), Super Signal West Femto Reagent (Cat no. 34096), Western Blot Stripping Buffer (Cat No. 46430), Western Blotting Filter Paper (Cat No. 84783), Carbenicillin (Disodium Salt) (Cat No. BP2648-5), Coomassie Brilliant Blue G250 Dye (Cat No 20279), PageRuler Plus Prestained Protein Ladder (Cat. No 26619), Pierce 1-Step Transfer Buffer (Cat No. 84731), PVDF Transfer Membrane (Cat No. 88518).

2.0.2 Molecular Biology

Agilent: QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Cat No. 210518).

Sigma Aldrich: Agarose (Cat No. A9539), Orange G Loading Dye (Cat No. O3756), RNA Loading Dye (Cat No. R1386), Taq Polymerase 5U/µl (D4545), DNAse treatment kit (Cat No: AMPD1-KIT).

Promega: dNTP's (Cat No. U1515)

New England Biolabs: 100bp ladder (Cat No. N3231L), 1kB DNA Ladder (Cat No. N0468S),.

Bioline: Plasmid Isolation Kit (Cat No. BIO-52026), RNA Extraction Kit (Cat No. BIO-52072), Random Hexamers (Cat No. BIO- 38028), Bioscript (Cat No. BIO-27036), Oligo dT (Cat No. BIO- 38029), Ribosafe RNAse inhibitor (Cat No. Bio-65027), Sensi Mix One Step Kit (Cat No. 76001).

Applied Biosystems: Taq man RT-qPCR assays (Cat no. 4331182) MTHFD1 (Hs 01068268_g1), MTHFD2 (Hs 00759197_s1), SHMT1 (Hs 00244618_m1).

Roche: Universal Probe Library 1-90 (Cat No. 04683633001).

2.0.3 Cell Culture

Sigma Aldrich: Trypsin EDTA (Cat No. T4049), Sodium Pyruvate (Cat No. S8636), Dulbecco's Modified Eagle Medium (DMEM) (Cat No. D5796), Cell Freezing Medium (Cat No. C6039), Propidium iodide (Cat No. P4864), Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Cat No. 15502), Fetal Bovine Serum (FBS) (Cat No. F9665).

Lennox: Puromycin Dihydrochloride (Cat No. CA 2856 0010)

Gibco: Trypan Blue (Cat No. 15250-061), Phosphate Buffer Saline 10X (Cat No. 14200).

Greiner Bio One: Tissue Culture Flasks 75cm^2 (Cat No. 658175CI), Tissue Culture Flasks 25 cm2 (Cat No. 690175 CI), 6 Well Plates (Cat No. 657160), Cell Scrapers (Cat No. 541070G).

Promega: Cell Titre 96 Non-Radioactive Cell Proliferation Assay (MTT) (Cat No. G4100).

2.0.4 Equipment

- Eppendorf pipettes (P10, P100, P1000) and Gilson Pipettes (P20) were used for micropipetting.
- Nano drop 100 Thermo Scientific
- GeneGnome Syngene Bio-imager
- DNR Mini-Bis Pro Bio-Imaging System
- Roche Lightcycler® 480
- Tecan i600 spectohotometer
- Theremo Scientific Pierce G2 Fast Blotter
- Thermo Scientific Forma Steri-Cycle Co_2 Incubator
- Innova®43 Incubator Shaker Series New Brunswick Scientific
- Applied Biosystems GeneAmp® PCR System 9700
- Thermo Scientific Forma Steri-Cycle Co2 Incubator
- Bioair Safeflow 1.2 Laminar Flow Biosafety Cabinet
- Branson Digital Sonifier

2.0.5 Stock Solutions

- **Coomassie Blue Stain:** 2.5g Coomassie, 100ml Acetic Acid, 300ml Methanol, 600ml H_2O.
- **Coomassie Blue De-stain –** 100ml Acetic Acid, 300ml Methanol, 600ml H_2O.
- **LB Broth –** 5 grams of tryptone, 5 grams of NaCl, 2.5grams of yeast extract, 500ml of H_2O.
- **LB Agar –** 15g of agar per litre of LB Broth
- **Tris Glycine Running Buffer (10 X):** 15.14 grams of Tris HCL, 5 grams of SDS, 71.3 grams of glycine (pH 8.3).
- **Transfer Buffer** (1 X) – 100 mL 10 X Transfer Buffer, 200 mL Methanol, 700 mL dH_2O.
- **Sample buffer** (4X)– 2.4 mL 1M Tris / HCl (pH 6.8), 4 mL Glycerol, 1 g SDS, 4 mg Bromophenol blue, 500 µL betamercaptoethanol, 3.1 mL dH_2O (pH 6.8).
- **TBS (10 X)**: 24.23 grams of Tris HCL, 80.06 grams of NaCl, 1 L of H_2O (pH 7.6).

- **TBS-T:** 50 mLs of 10 X TBS, 450 mL of H_2O, 250 µl Tween 20.
- **5% w/v Milk Blocking Buffer:** 2.5 grams of Non Fat Milk Powder, 50ml of 1 X TBST.
- **Bacterial Lysis Buffer:** 50 mM Potassium Phosphate (pH 7.8), 400 mM NaCl, 100 mM KCl, 10 % w/v Glycerol, 0.5 % w/v Triton X-100, 10 mM Imidazole (pH 7.8).
- **Orange G (10 X):** 0.1gram Orange G, 20 grams of sucrose, 50 ml of H_20
- **TBE (10 X):** 48.44 grams of Tris HCL, 12.37 grams of Boric Acid, 1.5 grams of EDTA, 500 mls of H_2O(pH 8.2).

2.1 Methods

2.1.1 Plasmid Preparations

A single colony of bacteria containing the desired plasmid of interest was picked and used to inoculate 5ml LB Broth containing a selective agent (100µg/ml ampicillin). The broth was incubated at 37°C shaking at 220 RPM overnight. The overnight culture was centrifuged at 1500 x g for 10 minutes to pellet the bacterial cells. The Plasmid DNA was then extracted from the bacteria using the Bioline Plasmid isolation kit (Cat No. BIO-52026) as per manufacturer's instructions.

2.1.2 Quantification of DNA and RNA

The Thermo Scientific Nano Drop 1000 apparatus was used to quantitate DNA and RNA. A 1 µl volume of DNA or RNA was loaded onto the pedestal of the apparatus and measured at 260nm. Nucleic acids free from contamination by either salt of organic compounds should have a 260:280 ratio of between 1.8-2.0.

2.1.3 Agarose Gel Electrophoresis

Agarose was added to 1 X Tris-Borate-EDTA (w/v) and boiled in a microwave oven until completely dissolved. The mixture was then cooled under running water until it was "hand hot". A 5µl volume of Ethidium bromide (10µg/ml) was added to a 150ml agarose gel (0.33 µg/ml). The agarose gel was set in a mould and wells were created using a standard gel comb. The DNA samples were mixed with Orange G loading dye and loaded into the wells. The appropriate DNA ladder was also loaded (either 100bp or 1kb). RNA was mixed with RNA loading buffer (Sigma Aldrich Cat No. R1386) and heated to 70 °C for 10 minutes; the tube containing the RNA was then put on ice prior to loading onto the agarose gel. Agarose gels were set at 90 volts for 40-50 minutes until resolved. Agarose gels were imaged and photographed under UV light with DNR Mini-Bis Pro Bio-Imaging System

2.1.4 Bacterial Cell Transformation

The pDEST 17 Plasmids (DHFRL1, DHFR or mutants DHFRL1 R24W and DHFR W24R) were transformed into BL21 AI One shot bacterial cells using the heat shock method. For each plasmid transformation one vial of the BL21 AI Bacterial cells was incubated with approximately 10ng of plasmid DNA in a final volume of 5µl for 30 minutes on ice. The cells were then placed in a water bath at 42 °C for 30 seconds to heat shock. Cells were immediately returned to ice and 250 µl of pre-warmed S.O.C (Invitrogen Cat No. C6070-03) medium was added. Cells were then placed in an incubator at 37° C shaking at 220 RPM for 1 hour. Cells were then plated on LB agar with a selective agent (100µg/ml ampicillin). Plates were incubated at 37 °C overnight.

2.1.6 Site Directed Mutagenesis

Site directed mutagenesis was performed to mutate the tryptophan at position 24 of the DHFR enzyme to an arginine (W24R) and to mutate the arginine at position 24 to a tryptophan in the DHFRL1 enzyme (R24W). The Agilent QuikChange Lightning Site-Directed Mutagenesis Kit was used to introduce the mutations (Product no. 210518). Primers were designed as per the instruction manual, see Table 1 page 33. The following reaction was set up to introduce the mutations into the DHFR and DHFRL1 genes:

- 5 µl of 10× reaction buffer
- 100 ng of pDEST 17 template (either DHFR or DHFRL1)
- 125 ng of oligonucleotide primer #1 X µl (DHFR/DHFRL1)
- 125 ng of oligonucleotide primer #2 (DHFR/DHFRL1)
- 1 µl of dNTP mix
- 1.5 µl of QuikSolution reagent
- H_2O was added to a final volume of 50 µl
- Lastly 1 µl of QuikChange Lightning Enzyme

The site directed mutagenesis reactions were then incubated as follows:

- 95°C for 2 minutes
- 95°C 20 seconds
- 60°C 10 seconds
- 68°C 30 seconds

X 18 cycles

- 68°C 5 minutes

Dpn I Digestion was carried out to digest the parental non mutant plasmid DNA. This was performed by the addition of 2 μl of Dpn I restriction enzyme to both the DHFR W24R and DHFRL1 R24W reactions. The digestion reaction was gently mixed and then incubated at 37°C for 5 minutes to digest the parental plasmid DNA. The mutated plasmids were then transformed into XL10 –Gold ultra-competent cells as follows. XL10 –Gold ultra-competent cells were thawed on ice and 2μl of β-ME mix from the site directed mutagenesis kit was added, the cells were swirled gently and incubated on ice for 2 minutes. 2μl's of the Dpn I treated plasmid DNA was added to the cells and incubated on ice for 30 minutes. The cells were then heat shocked for 30 seconds at 42°C and then returned to ice for 2 minutes. 500μl of preheated NZY+ broth was then added to each tube. The tubes were then incubated at 37°C for 1 hour with shaking at 225 RPM. 50μl's and 150μl's of each transformation were plated onto Lb ampicillin plates (100μg/ml). The transformation plates were incubated overnight a 37°C overnight. A plasmid preparation was performed as per section 2.1.1 and the isolated plasmids were sent for Sanger sequencing to confirm that the mutations had been inserted.

Table 1.0 DHFRL1 and DHFR Site Directed Mutagenesis Primers

	Sequence
DHFR W24R	**For**:GAACGGGGACCTGCCC**AGG**CCACCGCTCAGGAA **Rev:**TTCCTGAGCGGTGGCC**TGG**GCAGGTCCCCGTTC
DHFRL1 R24W	**For**:GAACGGGGACCTGCCC**TGG**CCGCCGCTCAGGAA **Rev**:TTCCTGAGCGGCGGCC**AGG**GCAGGTCCCCGTTC

2.1.7 Induction of Recombinant Protein Expression

Using the tip of a p10 pipette a single colony containing the plasmid pDEST 17 (DHFR, DHFRL1 or mutants W24R, R24W) was selected and cultured overnight in LB broth (ampicillin 100µg/ml) shaking at 220rpm at 37^{o}C. The next day a 1/20 dilution of the overnight culture into fresh LB broth was performed. The dilution allows the cells to quickly return to logarithmic growth. After approximately 1 hour and 40 minutes 1ml of culture was removed and placed in a spectrophotometer and measured at 600nm. The culture must have an optical density of between 0.4-0.5 (mid log phase) in order to proceed to the next step. The cultures were then split in two with 0.2% w/v of L-arabinose added to one of the cultures in order induce expression of the target protein, the un-induced culture acted as a control. In order to identify the time period at which the cells produce the greatest amount of protein, 1ml samples of induced and un-induced samples were taken at 0 hours, 2 hours, 4 hours, 6 hours and 24 hours after L-arabinose addition. Whole bacterial cell lysate was analysed for each time point by mixing with 4 X SDS PAGE buffer and heating to 99°C for 5 minutes and analysed by SDS PAGE as per section 2.1.9. After the desired time period of induction the cultures were centrifuged at 2,200 X G for 30 minutes at 4^{o}C. The LB broth was poured off and the pelleted cells were stored at -20 ^{o}C and used for purification as per 2.1.11 or were used for solubility analysis. For solubility analysis the cell pellets were weighed and B-PER Bacterial Protein Extraction Reagent (Cat no. 78243) was added in a ratio of 4mls/gram weight of bacteria. The bacterial samples were then lysed by sonication on ice in the Branson Digital sonifier apparatus at 30% amplitude for 3 X 10 seconds with a 30 second cooling period in between each sonication round. The bacterial/B-PER samples were then incubated at room temperature for 20 minutes and then centrifuged for 30 minutes at 10,000 x g at 4°C. This separated the samples into the soluble and insoluble fractions. All fractions were analysed by SDS and Western blot as per 2.1.9 and 2.1.10.

2.1.8 Purification of His Tagged DHFRL1 by His-Talon Affinity Purification

Samples Fractionation: Pelleted induced BL21 AI bacterial samples containing pDEST 17 were re-suspended in HisTALON xTractor Buffer in a ratio of 2ml of buffer per 100mg of cell pellet. Samples were incubated for 15 minutes on ice. The bacterial samples were then lysed by sonication on ice in the Branson Digital sonifier apparatus at 30% amplitude for 3 X 10 seconds with a 30 second cooling period in between each sonication round. Samples were then centrifuged for 20 minutes at 10,000 x g at 4°C. The supernatant containing the soluble protein fraction was removed an aliquot was saved for SDS PAGE and Western blot analysis the remaining soluble fraction was applied to the His-Talon affinity column.

Affinity Purification: All buffers were de-gassed prior to use by using 0.2µm membrane filter. The affinity resin was equilibrated with 6 bed volumes (6ml) of equilibration buffer, the buffer flowed off by gravity and was discarded. The clarified protein fraction was applied to the column in 5ml fractions; the flow through was collected and placed in a tube on ice. The first wash step was performed using 8 column volumes (8ml) of wash buffer, the wash step fraction was collected and stored on ice. The second was step was performed with 7 column volumes (7ml) of wash buffer this fraction was collected and stored on ice. The target recombinant His tagged protein was eluted using 6 ml of elution buffer collected in 1 ml fractions. Each 1ml fraction of eluted target protein was stored in 500 µl of glycerol to prevent damage of the protein by freeze thaw from storage at -20°C. All protein fractions were analysed by SDS PAGE and or Western Blot.

2.1.9 SDS Page Analysis Both pre-cast (**Thermo Scientific** Cat No. 25269) **and "homemade" SDS PAGE gels were used for protein analysis.**

- **SDS PAGE gel recipe**: To make a 10% w/v polyacrylamide gel the following was added. **Resolving Gel**: 4ml of **H_20**, 3.3ml of 30% w/v acrylamide, 2.5ml of 1.5M trizma base, 100ul of 10% w/v SDS, 100ul of 10% w/v APS and 4ul of TEMED. Once the gel is poured into the mould a layer of 70% w/v ethanol was added to ensure that the gel sets evenly. When the resolving get was set the ethanol was poured off and the stacking gel was added. **Stacking Gel:** 3.4ml of H_2O, 830ul of 30% w/v acrylamide, 630ul of 1m trizma base at pH 6.8, 50ul of 10% w/v SDS, 50ul of 10% w/v APS and 5ul of temed. A 12 well comb is added to the stacking gel in order to provide well's for the samples to be analysed.
- **Protein sample preparation:** Protein samples to be analysed were mixed with 4X sample buffer and heated at 95 °C for 5 minutes. The denatured protein samples were loaded onto an SDS PAGE gel (Thermo Scientific 4-20%) or a "homemade" 10% w/v SDS PAGE gel.
- **SDS PAGE analysis**: The SDS PAGE gels were run at 100 volts until samples had passed through the stacking layer of the gel. The voltage was then increased to 120 volts. The gel was run for the appropriate time for the ladder to resolve sufficiently which was approximately ~ 1 hour for the pre cast gels and 2 hours for the "homemade" gels. The SDS PAGE gel was then either used for western blot analysis (see below) or stained in coomassie blue for two hours while rocking gently. The SDS PAGE gels were then destained in destaining solution and then visualised and photographed under white light using the Mini Bis Pro Gel imager.

2.1.10 Western Blot Analysis

Two transfer methods were used for western blot analysis, the Biorad Mini Transblot Cell and the Thermo Scientific G2 Fast blotter.

- **Thermo Scientific G2 Fast blotter**

  The PVDF membrane was placed in the methanol for 30 seconds to charge. The PVDF membrane, the SDS PAGE gel and filter paper were then equilibrated in the Pierce 1 step transfer buffer for 10 minutes. The cassette for Pierce G2 fast blotter was assembled with the aforementioned pre-soaked materials, see figure 2.0 below. A roller was used in between the addition of each layer to eliminate any air bubble which could interfere with the transfer. The transfer was then set at 25 volts for 7 minutes.

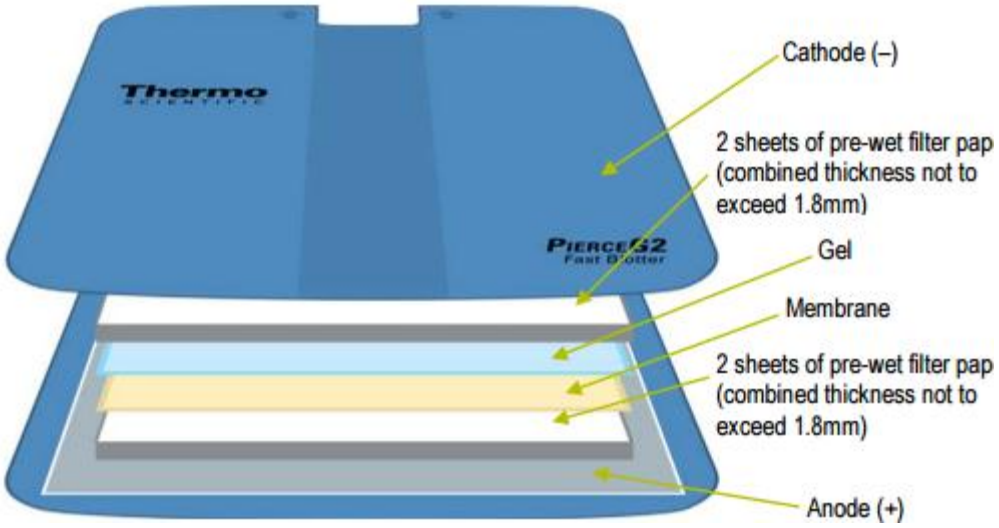


Figure 2.6 Schematic of Thermo Scientific G2 Fast Blotter. Transfers were set up with two layers of pre-soaked filter paper with PVDF membrane pre charged with methanol and soaked in transfer buffer with the SDS PAGE gel on top with two additional layers of filter paper over the gel. (Image taken from the fast blotter user manual May 2015 http://www.medical-supply.ie/products/attachments/62288%20Thermo%20Pierce%20G2%20Fast%20Blotter.pdf)

- **Biorad Mini Transblot Cell**

  The PVDF membrane was placed in the methanol for 15 seconds to “pre wet”. The PVDF membrane, the SDS PAGE gel, fibrous pads from the mini trans blot system and filter paper were then equilibrated in transfer buffer for 15 minutes at 4 °C. The cassette for the mini trans blot transfer system was assembled with the aforementioned pre-soaked materials. The ice block and stirring bar were added to the transfer tank. The transfer was set at 100 volts for 60 minutes in a 4 °C cold room.

- **Primary Antibody addition:** To confirm the transfer the membrane was then stained with ponceau S. The ponceau S stain was removed by washing with TBST. The membrane was then blocked for 1 hour gently rocking at room temperature in a 5% w/v milk TBST solution. The blocking buffer was then removed and the primary DHFR or his antibody was then added at a 1/10,000 dilution in a 5% w/v milk TBST solution and left at 4 °C gently rocking overnight.
- **Secondary Antibody addition:** The primary antibody was then removed and three consecutive 15 minute washes with TBST were performed to ensure removal of excess unbound primary antibody. The anti-rabbit secondary antibody was then added at a 1/50,000 dilution in TBST solution at room temperature and left to rock gently for one hour. The membrane was then further washed with TBST (3 x 15 minutes).
- **Western Blot Imaging:** The working solution of the Signal West Femto reagent was then prepared and 1ml was then applied to the membrane. The membrane was then imaged with the Gene Gnome Instrument (Syn Gene).

2.1.11 Bradford Assay

The Bradford assay allows for the quantification of protein in an unknown sample. The assay was carried out in a 96 micro well plate and the Tecan i600 spectrophotometer.tohotometer was used to measure the absorbance of the protein samples and standards.

- Protein standards were made up at 25µg, 125µg, 250µg, 500µg, 750µg, 1000µg, and 1500µg of bovine serum albumin (BSA).
- 5µl of each standard and protein of unknown concentration was added to a 96 well micro-plate in duplicate.
- 250µl of Bradford reagent was added to the samples and standards. The plate was wrapped in tinfoil, rocked gently for 30 seconds to mix and then incubated for 15 minutes at room temperature.
- The absorbance of the samples and standards was then read at 595nm.
- The concentration and absorbance of the standards was then plotted in Microsoft Excel. The equation of the line was then used to determine the concentration of the unknown protein samples.

2.2 Cell Culture

2.2.1 HEK 293 MTHFD1L Over expressed and Knockdown Cell Lines

Previously generated HEK 293 cells with modulated levels of MTHFD1L were utilised in this research thesis. The MTHFD1L knockdowns 1 and 2 were generated by knocking down the MTHFD1L gene using shRNA down regulation which could be essentially "turned on" by the addition of the inducer molecule IPTG. The MTHFD1L overexpressed HEK 293 contained a plasmid vector pcDNA 3.2 which constitutively over expressed the MTHFD1L gene. Human Embryonic Kidney 293 (HEK 293) cells were cultured in Dulbecco's modified Eagle medium DMEM with 10% fetal bovine calf serum, 200mM L-glutamine in a 5% CO_2 incubator at 37°C. The HEK293 shRNA MTHFD1L knockdown lines (1 and 2) and the associated non targeting shRNA control were grown in normal HEK 293 medium supplemented with puromycin (5 µg/ml) as a selective agent and with IPTG (200 µM) to induce the shRNA down regulation. The MTHFD1L overexpressed cell line and the empty vector control were grown in normal HEK293 medium supplemented with 500 µg/ml of the selective agent G418. Cells were cultured in 10ml of medium and seeded in 75 cm^2 tissue culture flasks. Cells were passaged by removing culture media, rinsing with 2mls of Trypsin-EDTA then incubating the cells with 3ml of Trypsin-EDTA at 37 °C for 5- 10 minutes. The trypsin is neutralised by the addition of 3ml of complete DMEM. Cells

are then centrifuged at 500 X G for 5 minutes to pellet the cells. Pelleted cells were then re-suspended in cell culture medium and seeded in 75 cm^2 tissue culture flasks.

2.2.2 Intracellular and Cell Medium Formate Quantification

In order to test the effect that altering the level of the MTHFD1L gene has on its metabolite production formate the HEK 293 cell lines with modulated levels of the MTHDFD1L gene were utilised. Prior to collection of samples all microfuge tubes were washed in ultra-pure H_2o to reduce the risk of formate contamination from the plastic. Optimisation was performed with the cell lines in order to identify the ideal seeding density for growth over five days. It was identified that a density of 2 X 10^6 was optimum. For formate quantification HEK 293 cell lines with altered MTHFD1L expression and associated controls (**1**.MTHFD1L overexpressed, **2**. MTHFD1L over expressed empty vector control, **3**. MTHFD1L shRNA knockdown 1, **4**. MTHFD1L shRNA knockdown 2,**5**. Knockdown control of non-targeting shRNA,**6**. normal HEK293 cells) were set up set up at a cell density of 2 X 10^6 in triplicate and seeded in 75 cm^2 tissue culture flasks. 300 µl's of cell culture medium was removed from the each cell line in duplicate every 12 hours for 5 days and stored at -80 °C. After 5 days had elapsed flasks set up in duplicate for each cell line were trypsinized and cells were counted (as per section 2.2.4). The cell pellet of each cell line was stored at -80 °C. The third flask set up of each cell line was used for RNA extraction to confirm MTHFD1L expression by RT-qPCR (as per section 2.3). All medium samples and cell pellets collected for the 5 day experiment were sent on dry ice to Dr Sean Brosnan in the University of Newfoundland, St. John's Canada for formate analysis by gas chromatography mass spectroscopy (GS-MS). Two approaches were taken in the normalisation of intracellular formate levels within the MTHFD1L over expressed and knockdown HEK 293 cell lines. Formate concentration for each replicate was divided by its corresponding protein concentration (Obtained via Bradford assay) or number of cells present on final day of experiment. Replicate values were averaged and the standard deviation values were obtained. The normalised formate values were plotted in Microsoft Excel for the MTHFD1L over expressed and knockdown cell lines. Statistical significance was determined by

Student's t-test using Microsoft Excel. The formate cellular medium concentrations were normalised to the volume of medium in each flask at the time of collection.

2.2.3 SW480 and SW620 Cell lines

SW480 and SW620 cells were cultured in Dulbecco's modified Eagle medium DMEM with 10% fetal bovine calf serum, 200mM L-glutamine in a 5% CO_2 incubator at 37 °C. . Cells were passaged by removing culture medium, rinsing with 2mls of Trypsin-EDTA then incubating the cells with 3ml of Trypsin-EDTA at 37 °C for 5- 10 minutes. The trypsin is neutralised by the addition of 3ml of complete DMEM. Cells are then centrifuged at 500 X G for 5 minutes to pellet the cells. Pelleted cells were then re-suspended in cell culture medium and seeded in 75 cm^2 tissue culture flasks.

2.2.4 Cell Enumeration

Cell viability counts were performed by the addition of the trypan blue at a 1:4 concentration; the cells were then left at room temperature for 5 minutes. The cells were then added to a haemocytometer and visualised at 100x magnification under a microscope. The number of live cells per square of the haemocytometer was then counted. Live cells appear white and the dead cells appear blue. The number of live cells was then calculated per ml of cells by multiplying by the dilution factor and then by 10^4.

2.3 RT-qPCR Analysis

2.3.1 RNA Extraction and cDNA Synthesis

Cell pellets were washed with 1ml of PBS and RNA was extracted using the Bioline RNA extraction kit as per manufacturer's instructions. The extracted RNA was quantified using the Nano drop 1000 (Thermo Scientific). The quality of the RNA was assessed by agarose gel electrophoresis. For DNAse treatment of RNA the following was performed, 1µl of Sigma 10x reaction buffer (R6273-1ML) and 1µl of Sigma DNAse 1(D5307 1000u/ml) to 2 µg's of RNA in a final volume of 10 µl. The DNAse reaction was incubated for 10 minutes at room temperature. The reaction

was inactivated by the addition of 1 µl of Sigma stop solution (S4809-1ML) and the RNA was heated to 70 °C for 10 minutes. Reverse transcription was performed by the addition of 2 µl (50 ng/ µl) of Bioline random hexamers primers (RHP-111C) and 4 µl (50µm) of oligo (dT) 18 primer mix (ODT-110L) to the DNAse treated RNA and incubated at 70 °C for 5 minutes, tubes were the placed on ice for 1 minute. Then remaining constituents of the reaction were added as follows, 1 x Bioline reaction buffer (MB-1091), 1µl of ribonuclease inhibitor (40,000u/ml R2520-2.5KU), 1µl of Bioline Bioscript (200u/µl, BIO-27036). The cDNA reaction was then incubated on the thermocycler as follows:

- 10 minutes at 25 °C
- 60 minutes at 42 °C
- 15 minutes at 70 °C
- hold at 4 ° C.

The cDNA was then either stored at -20 °C or used for the genomic DNA contamination assay and subsequent real time PCR analysis.

2.3.2 Genomic DNA Contamination Assay

Prior to RT-qPCR analysis of cDNA the MTHFD1 R653W intron flanking PCR assay was performed in order to ensure that the RNA from which the cDNA was synthesised was free of genomic DNA. The intron spanning assay allows for the detection of genomic DNA. The expected size of the PCR products of cDNA free from genomic DNA contamination is 232bp. The expected size of the PCR products containing genomic DNA is 330 bp. Contaminated cDNA would be represented by two PCR bands at 232 bp and 330bp. Although all RT-qPCR assays were intron spanning, genomic DNA may still interfere non-specifically and this assay acts as an extra control. A PCR master mix was made as per table 1.2. See Table 1.1 below for primer sequences.

Table 1.1 MTHFD1 R653Q Primer Sequences

MTHFD1 R653Q	Sequence
Forward Primer	5′- CACTCCAGTGTTTGTCCATG -3′
Reverse Primer	5′- GCATCTTGAGAGCCCTGAC-3′

Table 1.2PCR Master mix

Reagent	1X reaction
10 X PCR Buffer	5µl
$MgCl_2$	3µl
2.5 mM DNTP	4µl
Forward Primer	0.3µl
Reverse Primer	0.4µl
cDNA	1 µl
Taq Polymerase	0.2µl
H_2O	36.1µl

The PCR reaction was placed in the Applied Biosystems GeneAmp® PCR System 9700 set at the following:

- 95°C for 3 min
- 94°C for 30 s
- 58°C for 1 min
- 72°C for 1 min

35 Cycles

- 72°C for 10 min.

PCR products were analysed on a 1% w/v agarose gel by electrophoresis, as per section 2.1.3

2.3.3 Reverse Transcriptase Polymerase Chain Reaction (RT-qPCR)

RT-qPCR assays were carried out in white 96 well plates in the Roche Lightcycler® 480 machine. RT-qPCR assays were carried out using the Roche Universal Probe Library (UPL) and Applied Biosystem's Taq Man primer probe mix. A PCR master mix was made up for each gene to 15 µl and 1 µl of cDNA was added per reaction, as per Table 1.3. See Table 1.4 for primer sequences. The comparative E-method was used to determine the fold change between target genes relative to a reference gene (Tellman 2006).

Table 1.3 PCR master mix

Reagent	1 X Reaction
SensiMixII	8 µl
Probe	0.16 µl
Forward primer (0.2 µM/µl)	0.16 µl
Reverse primer (0.2 µM/µl)	0.16 µl
H_2O	5.52 µl

Each gene was analysed in triplicate and a negative control was also set up with 1 µl of H_2O in place of cDNA. The PCR reaction was up as follows:

- 95°C for 5 min
- 95°C for 30 sec
- 60°C for 30 sec
- 72°C for 1 sec

45 Cycles

Applied Biosystems Taq man assays were used in the analysis of MTHFD1 (product no. Hs01068268_g1), MTHFD2 (product no. Hs00759197_s1) and SHMT1 (HS00244618).

Table 1.4 Primer Sequences and UPL probes for RT-qPCR Assays

DHFR Primers, UPL Probe # 24	
Forward Primer	5'GGGGGAAAGCTGGAGTATTG3'
Reverse Primer	5'ACTATGTTCCGCCCACACAC 3'
DHFRL1, UPL Probe #89	
Forward Primer	5'CGGACCTTAGAAAGTCACACATC3'
Reverse Primer	5'GCGAAATTCCCTTCTTCAAA 3'
MTHFD1L, UPL Probe #42	
Forward Primer	5'GAGCTCTGAAGATGCATGGAG 3'
Reverse Primer	5'TGCTTCTGGAGGTTACAGCA 3'
SHMT2, UPL Probe #83	
Forward Primer	5' TTGCTGCCCTAGACCAGAG 3'
Reverse Primer	5'GACCAGCTGCCCACATCT 3'
TYMS, UPL Probe #43	
Forward Primer	5' CCCAGTTTATGGCTTCCAGT 3'
Reverse Primer	5'GCAGTTGGTCAACTCCTGT 3'
GUS, UPL Probe #57	
Forward Primer	5' ACGCCAGCTTCAAAGCAA 3'
Reverse Primer	5' TCACAGGAGAATCACTTCAACC 3'
ALDH51, UPL Probe #64	
Forward Primer	5' TCG AGA ACA GGG AAT AAC ATT G 3'
Reverse Primer	5' TGA TTT GGA TCC ATT GTT GC 3'
ATIC, UPL Probe #89	
Forward Primer	5' GAGGGACTGCAAAAGCTCTC 3'
Reverse Primer	5' GGA AATCCCGTCAACTCAGA 3'

Chapter 3

The effect of MTHFD1L Expression on Formate Production and Cell Growth in HEK 293 cells

3.0 Introduction

The main aim of this chapter is to characterise the effects altered Methylenetetrahydrofolate dehydrogenase ($NADP^+$ dependent) 1-like (MTHFD1L) expression has on HEK 293 cells. MTHFD1L is a monofunctional enzyme involved in folate metabolism. MTHFD1L is located in the mitochondria and is responsible for the crucial last step in the provision of one carbon units for the cytoplasm by the conversion of 10-formyl-THF to formate (Tibbetts and Appling 2010). Formate is utilised in one carbon metabolism in the production of purines, thymidylate and the re-methylation of homocysteine to methione, a donor group for cellular methylation reactions (Fox and Stover 2008). It is estimated that up to 75% of formate for one carbon metabolism originates in the mitochondria (Pike *et al.* 2010).

As discussed in section 1.2 folate is associated with many multifactorial diseases, aside from folate consumption genetic polymorphisms and gene expression of folate metabolising enzymes also contribute to the risk of disease such as neural tube defects (NTD's), cancer, Cardiovascular disease (CVD) and neurological diseases (Kirke *et al.* 2004, Samani *et al.* 2007, Cole *et al.* 2007). A delicate balance between formate supply and demand must be maintained within the cell for effective maintenance of one carbon metabolism. This is demonstrated by the work of Parle-McDermott *et al* (2009) who found that a deletion insertion polymorphism (DIP) (rs3832406) which effects alternative splicing of the MTHFD1L gene resulting in increased levels of the MTHFD1L enzyme is associated with Neural Tube Defect (NTD) risk. Impairment of folate metabolising enzymes is generally associated with increased NTD risk, the fact that a polymorphism in the MTHFD1L gene which results in greater enzyme levels is associated with NTD risk highlights the importance of metabolic balance (Kirke *et al.* 2004). It is thought the increased levels of MTHFD1L results in an increase in formate which causes a disruption to the flow and balance between cytoplasmic and mitochondrial one carbon metabolism, hindering cellular proliferation. The authors also hypothesised that excess formate may build up to toxic levels within developing embryos disrupting cellular proliferation (Parle-McDermott *et al.* 2009). Such is the importance of the MTHFD1L

gene Momb *et al* (2012) found that its absence was embryonic lethal in mice. The authors also found that knocking out the gene resulted in NTD development with 100% penetrance. Further proof of the role of the MTHFD1L gene in NTD's comes from Minguzzi *et al* (2014) who found that a single nucleotide polymorphism (SNP) rs7646 within the gene affects microRNA binding resulting in increased NTD risk. The SNP rs11754661 within the MTHFD1L gene has also been associated with Alzheimer's disease risk in American and Han Chinese populations, but studies by Ramirez-Lorca *et al* (2011) failed to find an association in a Spanish population (Naj *et al.* 2010, Ren *et al.* 2011, Ramirez-Lorca 2011). The role of the MTHFD1L gene in Alzheimer's disease is thought to be due to the effect its expression has on the re-methylation of homocysteine to methionine. Elevated levels of homocysteine have been associated with Alzheimer's disease, an impairment of the MTHFD1L gene may cause a reduction in formate and tetrahydrofolate (THF) from the mitochondria therefore disrupting the remethylation of homocysteine to methionine (Seshadri *et al.* 2002, Naj *et al.* 2010). In addition to its role in NTD's and Alzheimer's disease, MTHFD1L has also been associated with increased risk of Cardiovascular disease (CVD), again with its role thought to be due to the provision of one carbon molecules for re-methylation of homocysteine in the cytoplasm (Nilesh *et al.* 2007, Samani *et al.* 2007).

As discussed in section 1.4 MTHFD1L also has strong links to cancer growth and development. Sugiura *et al* (2004) found that MTHFD1L expression was elevated in colon cancer, they also found that over expressing the gene in HEK 293 cells lead to increased "colony formation", indicating the gene had an effect on the growth of the cells. Selcuklu *et al* (2012) also found that enzyme to be up-regulated in breast cancer. Most recently Jain *et al* (2012) associated MTHFD1L, along with other mitochondrial folate enzymes Methylenetetrahydrofolate dehydrogenase ($NADP^+$ dependent) 1-like (MTHFD2) and Serine hydroxy methyltranferase (SHMT2) with increased growth, proliferation and mortality in cancer. Moreover, findings by Fashfindar *et al* (2012) and Wang *et al* (2013) have found that the metabolite of MTHFD1L, formate may be a useful biomarker for staging and grading of tumours and in the detection of cancer metastasis.

HEK293 Cell lines with modulated MTHFD1L expression (increased and decreased) were previously generated in the Nutritional Genomics laboratory (Minguzzi 2013). Although HEK 293 cells are a non-cancerous "normal" cell line they are a good model for observing and characterising changes and can be utilised as a "test tube with a membrane". The MTHFD1L gene was overexpressed by transfection of HEK 293 cells with the Invitrogen plasmid pcDNA 3.2 which constitutively overexpresses the MTHFD1L gene. The MTHFD1L gene was knocked down by transfection of HEK 293 cells with shRNA (short hairpin RNA), two knockdown clones were created. The MTHFD1L knockdown cell lines have an IPTG inducible promoter system allowing for down regulation of the gene when desired. The inducible promoter system in the MTHFD1L knockdown cell lines limited the deleterious effects down regulation of the gene may have on the HEK 293 cell lines. Previous proteomic analysis of the HEK 293 cell lines with modulated MTHFD1L expression identified that a number of important DNA synthesis and repair enzymes were affected by MTHFD1L expression (Minguzzi 2013). An example of which is thymidylate synthase (TYMS), Minguzzi (2013) found that when MTHFD1L was over expressed the TYMS enzyme was down regulated relative to the control cell line. TYMS is known to translocate to the nucleus and the mitochondria for de novo thymidylate synthesis (Anderson *et al*. 2011 and 2012). Interestingly Minguzzi (2013) found proteins thought to be involved in complexing with TYMS at the nuclear lamina for thymidylate synthesis; Proliferating cell nuclear antigen (PCNA), Lamin A/C (LMNA) and Lamin B1 (LMNB1) were also down regulated (Anderson *et al.* 2012, Minguzzi 2013). Intriguingly, these proteins showed opposite trends were they were up-regulated when MTHFD1L was knocked down and down regulated when MTHFD1L was over expressed (Minguzzi 2013). It appeared that thymidylate synthesis was affected by MTHFD1L expression. It was hypothesised that the effects observed in the proteomic data were mediated by the metabolite of MTHFD1L, formate (Minguzzi 2013). Therefore one of the main aims of this chapter was to assess if altered MTHFD1L expression results in a detectable altered formate level which could have implications for formate as a biomarker, with secondary aims being detection of alterations in cell cycle and cell growth. Due to the fact that the expression of the MTHFD1L gene has been significantly associated with cancer cell growth and mortality, the detection of an

altered formate level within the HEK 293 cells with modulated MTHFD1L expression may have implications for the characterisation of formate as a biomarker. The detection and study of changes within the metabolism of a cell, tissue or organism is referred to as metabolomics and is fast emerging field within cancer research.

3.0.1 Metabolomics and Cancer

Genomics, transcriptomics, proteomics and now metabolomics are all utilised in the study and research of cancer in order to establish the changes that occur within the cell which enable it to become malignant. Aside from helping us attain a deeper understanding of cancer at a molecular level, these tools allow us to identify targets or biomarkers for the identification and treatment of cancer. Metabolic re-programming is a hallmark of cancer (Ward and Thompson 2012). Detecting changes in metabolites within cancer versus normal tissue can allow for an un-invasive means of cancer diagnosis and in some cases the development of targeted therapies (Patel and Ahmed 2015).

Metabolite profiling is often said to be the most reliable and coherent snap shot of the cell. Detection of changes at a DNA, transcription or protein level are indeed very valuable, however it thought that metabolic profiling is the closest representation to the phenotype of the cell/tissue, see Figure 3.1 (Patel and Armed 2015, Fien 2002) . Changes at a DNA or at a transcription level within a cell are not always represented at a protein or metabolic level due to regulatory processes within the cell. One of the earliest known changes detected in the metabolism of cancer cells is their propensity towards sugar as an energy source and aerobic glycolysis as an energy means (Warburg *et al.* 1927). Indeed, it is well documented that cancer cells readily consume sugar and this change in cancer cell metabolism can be used in conjunction with positron emission tomography (PET) scanning in the detection of tumours throughout the body particularly in the lungs (Vansteenkiste *et al*. 2002).

The main techniques employed for metabolite analysis are nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS). Mass spectrometry involves ionization of a compound and analysis based on the mass to charge ratio.

Different separation methods can be used in conjunction with mass spectrometry such as liquid chromatography (LC-MS), capillary electrophoresis (CE-MS) and gas chromatography (GC-MS) (Patel and Ahmed 2015).The method used in this chapter for formate analysis was GC-MS, see Figure 3.2 for GC-MS schematic.

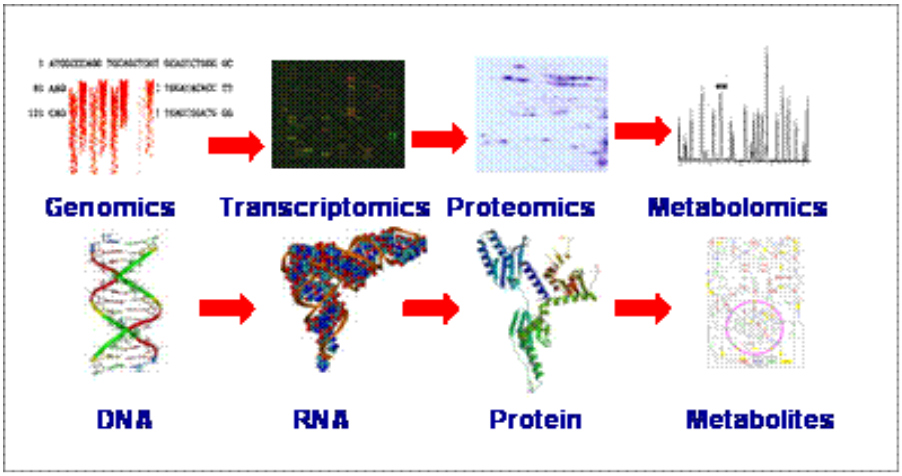


Figure 3.1 "Omics". There are four different molecular levels in which cells/tissue can be analysed and studied, DNA, RNA, protein and metabolic. The metabolite level is said to be the closest to the phenotype of the cell. (Image taken from http://www.metabolomics.bbsrc.ac.uk/background.htm)

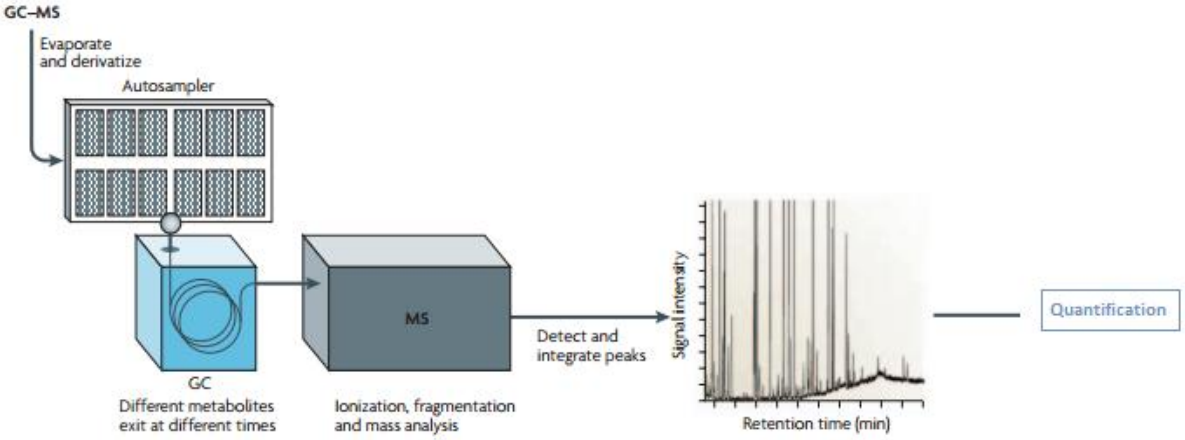


Figure 3.2 GC-MS Schematic for metabolite analysis. Samples are separated, ionized and molecules are detected by their mass to charge ratio. Accurate quantification of a metabolite can be achieved by the inclusion of an isotype dilution of the target metabolite (Lamarre *et al* 2014). Image edited from (Last *et al* 2007).

3.0.2 Aims

The main aim of this chapter was to characterise the effects altered MTHFD1L expression has on HEK 293 cells, with a view to assessing:

- If altered MTHFD1L expression results in a detectable altered formate level in HEK 293 cells. The premise of this was two-fold; 1. This would allow for the assessment of formate as a biomarker for MTHFD1L expression in regard to cancer. 2. This finding would allow for confirmation of the previous work by Dr Minguzzi who suggested that formate supply within the cell mediated the changes observed in DNA repair and synthesising enzymes in HEK 293 cells with altered MTHFD1L expression (Minguzzi 2013).
- The effect altered MTHFD1L expression has on the cell growth and the cell cycle in HEK 293 cells.

3.0.3 Objectives

In order to characterise the effect altered MTHFD1L expression has on HEK 293 cells the following procedures were undertaken.

- Cell cycle analysis of the MTHFD1L knockdown HEK 293 cell lines by flow cytometry using propidium iodide staining.
- Analysis of Formate in the MTHFD1L over expressed, knockdown and associated controls in both the cell medium of each cell line and intracellularly by GC-MS.
- Normalisation of formate levels to cell number and protein concentration for intracellular analysis and to medium volume for extracellular cell medium analysis.
- Cell proliferation analysis by cell counting.
- Confirmation of MTHFD1L in modulated HEK 293 cells by rt-qPCR.

3.1 Results

3.1.1 Cell Cycle Analysis of MTHFD1L Knockdown Cell Lines

Flow cytometric analysis was under taken in order to ascertain if knocking down the MTHFD1L gene had an effect on the cell cycle of the HEK 293 cells. Cells were permeabilised and fixed using ethanol and then stained using propidium iodide. Propidium Iodide is a DNA intercalating agent which also acts as a DNA fluorophore when excited in the range of 488 nm (Pozarowski and Darzynkiewicz 2004). Propidium iodide is impermeable to the cell membrane hence the need for ethanol permeabilastion. Cell cycle analysis by flow cytometry allows for the determination of what percentage of cells are in a given phase of the cell cycle. It can be seen from Figure 3.2 below, cell cycle analysis results in two gaussian peaks representing the G1 and G2/M phases respectively. The S phase of the cell cycle shares an overlap with both the G1 and G2/M phases. When the percentage of each cell in a given cell cycle phase are compared for the MTHFD1L knockdowns (1 and 2) and the associated non targeting shRNA control, very little difference was observed, see Figures 3.2 and 3.3.

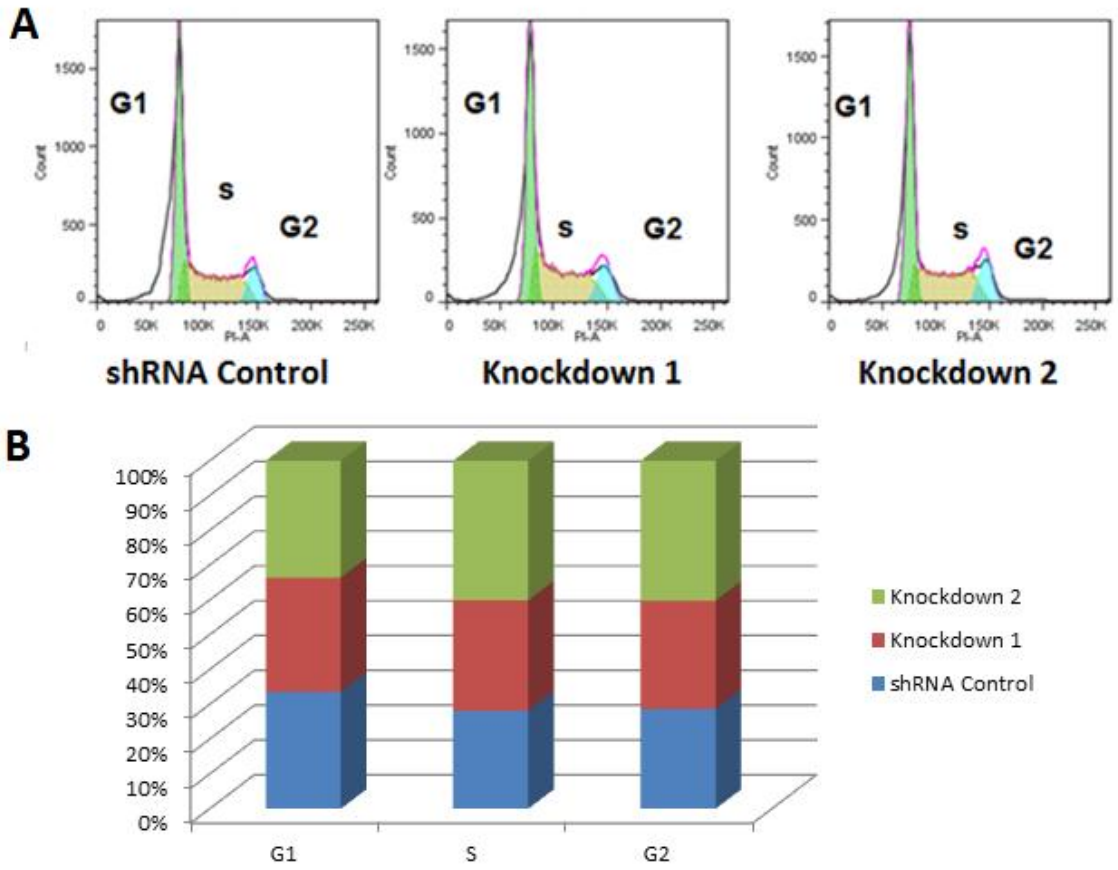


Figure 3.3 Cell Cycle Analysis by Flow Cytometry of MTHFD1L Knockdowns. A. The number of cells is represented on the y axis of the graph while the relative florescence of the cells is represented on the X axis. Cells which are in the G1 phase of the cell cycle have 1 copy of DNA, as the cell progresses through to the S phase of DNA synthesis the florescence increases accordingly. Cells in the G2 phase of the cell cycle are said to have twice as much DNA content and therefore twice the amount of florescence when compared to G1 cells. Very little difference can be observed between the shRNA control and knockdowns 1 and 2.**B.**The percentage of cells in each phase of the cell cycle can be seen. There appears to be very little difference between the two MTHFD1L knockdowns 1 and 2 the control cell line shRNA control in regards to percentage of cells in each phase of the cell cycle.

3.2. Intracellular Formate Analysis in HEK 293 Cells with Modulated MTHFD1L

3.2.1 Intracellular Formate Analysis in MTHFD1L Over expressed Cells

As per section 2.2.2 of Chapter 2 the HEK 293 cell lines with modulated MTHFD1L expression were set up at 2 x 10^6 in triplicate for 5 days. The pcDNA 3.2 empty vector HEK293 cell line acts a control for the MTHFD1L overexpressing cells. Unfortunately during the first day of the formate analysis experiment the empty vector cell line cells did not attach as well to the cell culture tissue flask in comparison to the MTHFD1L overexpressed cells. This resulted in fewer cell numbers in the empty vector cell line compared to the overexpressed. In order to act as an extra control normal HEK 293 cells were included in the analysis of the overexpressed cells. When intracellular formate levels are normalised to protein concentration it appears that the MTHFD1L overexpressing cell line has a 2.89 fold higher formate concentration when compared to the empty vector control. When the MTHFD1L over expressing cell line is compared to the normal HEK 293 cells it has 1.23 fold higher formate level, see Table 3.1 for fold differences and see Figure 3.4 for relative differences. Consistent with normalisation to protein concentration the MTHFD1L over expressing cells appear to have an increased formate concentration with it being 1.85 fold higher when compared to the empty vector control when normalised to cell number. When compared to the normal HEK 293 cells the MTHFD1L over expressing cell line has a 1.74 fold increased formate concentration, see Table 3.1 and Figure 3.5. When the intracellular formate data was normalised to protein concentration and cell number statistical analysis indicated that the increased formate level observed in the MTHFD1L overexpressing cells was statistically significant relative to the empty vector cells, see Table 3.1. However, relative to the normal HEK 293 cells statistical analysis indicated that the increased formate level in the MTHFD1L overexpressing cells was non-significant. The experiment would need to be repeated in order to conclusively identify if the MTHFD1L overexpressing cells have a statistically significant higher formate level than the controls. Despite the discordance with the statistical analysis results clear trends can be observed and it appears that overexpressing the MTHFD1L gene has had an effect on the metabolism of the HEK 293 cells

Table 3.1 Fold difference in Formate concentration normalised to protein and cell number

	Fold Difference		Fold Difference	
Cell Line	**Protein¥**	**p Value***	**Cell Numberɨ**	**p Value***
Overexpressed	2.89	**0.005**	1.85	**0.007**
Empty Vector Control	1		1	
	Fold Difference		**Fold Difference**	
Overexpressed	1.23	0.068	1.74	0.190
Normal HEK 293	**1**		**1**	

*As determined by T test (Significance ≤ 0.05)

¥ See Figure 3.4 for graphical representation of formate concentration (µM) per µg of protein.

ɨ See Figure 3.5 for formate concentration (µM) per cell

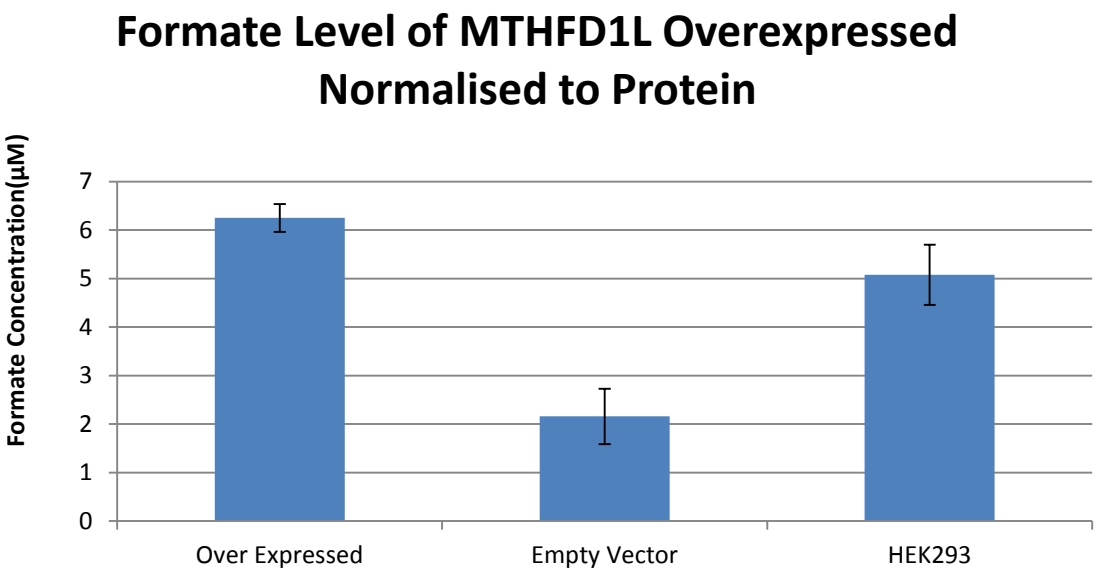


Figure 3.4 Formate Levels of MTHFD1L Overexpressed Cell line Normalised to Protein Concentration. It can be seen that Intracellular formate values obtained for the MTHFD1L overexpressed cell line appears to be much higher when compared to the formate values obtained for the empty vector with a 2.89 fold over expression (p=0.005) calculated as per Table 3.1. The formate level differences between the MTHFD1L over expressed and the HEK 293 cells indicate that the MTHFD1L had an increased formate level relative to the normal HEK 293 cell line, with a 1.23 fold increased concentration observed (p=0.068). However, the HEK 293 cells are grown in the absence of the selective agent G418 so are not the true control for the overexpressed cells but due to issues with the empty vector cell attachment, as explained previously in section 3.2.1 they included as an extra control.

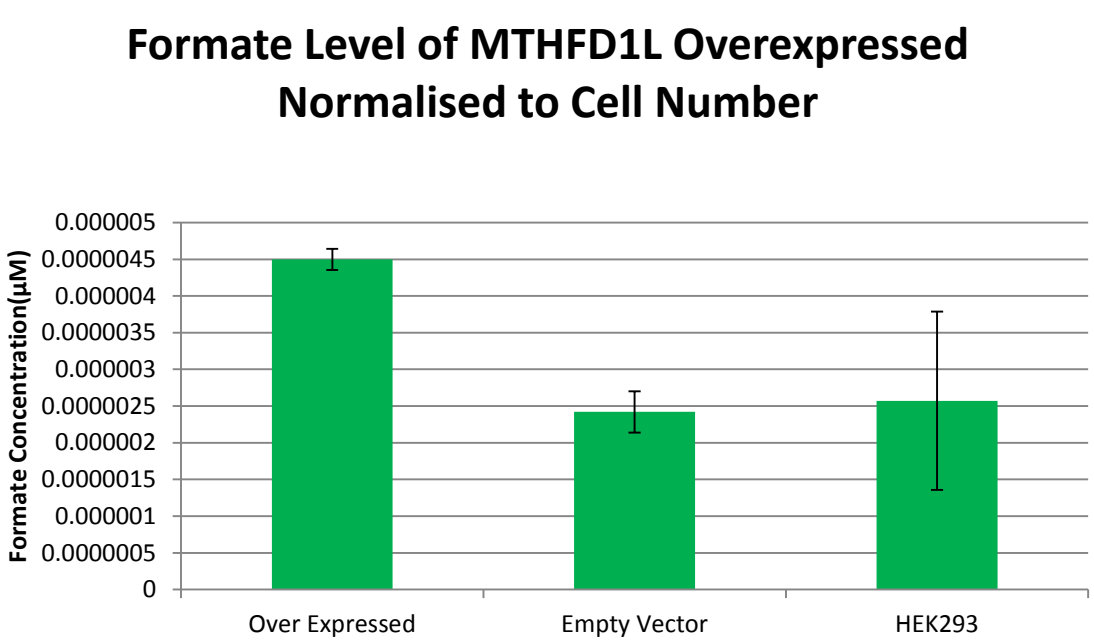


Figure 3.5 Formate Levels of MTHFD1L Overexpressed Cell line Normalised to Cell Number. It can be seen that when the formate levels are normalised to cell number that the formate level for MTHFD1L over expressing cells appears to be higher when compared to both the empty vector control with a 1.85 fold increased level calculated as per Table 3.1 (p=0.007) and the normal HEK 293 cells with a 1.74 fold increased level as per table 3.1 (p=0.190).

3.2.2 Intracellular Formate Analysis in MTHFD1L Knockdown Cells

The MTHFD1L knockdown cell lines were set up as per section 2.2.2 of Chapter 2. When intracellular formate levels are normalised to protein concentration it appears that the MTHFD1L shRNA knockdowns 1 and 2 have a -1.58 and -1.69 reduced formate concentration when compared to the non-targeting shRNA control, see Table 3.2 for fold differences and and Figure 3.6 for relative differences. These results indicate that knocking down the MTHFD1L gene has had a direct effect on intracellular formate concentration. When intracellular formate levels are normalised to cell number it appears that the MTHFD1L shRNA knockdowns 1 and 2 have a -1.02 and a -1.31 fold reduced formate concentration when compared to the non-targeting shRNA control, see Table 3.2 and Figure 3.7. When the intracellular formate data was normalised to protein concentration statistical analysis indicated that the changes observed were significant, Table 3.2. However, upon normalisation of the data to cell number statistical analysis indicated the changes observed in the MTHFD1L knockdown cell lines were non-significant. The experiment would need to be repeated in order to conclusively identify if the formate level of the MTHFD1L knockdown cell lines is statistically significantly lower than the shRNA control. Again, due to time constraints this could not be carried out. Despite the inconsistencies with the statistical analysis results and data normalisation clear trends can be observed and it appears that knocking down the MTHFD1L gene had an effect on formate production particularly in knockdown 2.

Table3.2. Fold difference in Formate concentration normalised to protein and cell number

	Fold Difference		Fold Difference	
Cell Line	**Proteinł**	**P Value***	**Cell NumberÏ**	**p Value***
Knockdown 1	-1.58	**0.007**	-1.02	0.409
Knockdown 2	-1.69	**0.003**	-1.31	0.106
shRNA Control	1		1	

*As determined by T test (Significance ≤ 0.05)

ł See Figure 3.6 for graphical representation of formate concentration (µM) per µg of protein.

Ï See Figure 3.7 for formate concentration (µM) per cell.

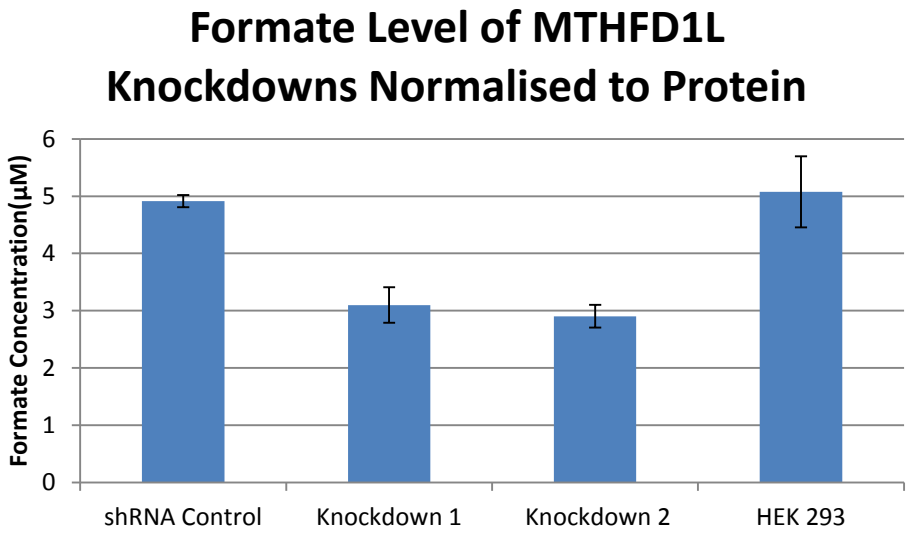


Figure 3.6 Formate Levels of MTHFD1L Knockdown cell line Normalised to protein. It can be seen that when the formate levels are normalised to protein concentration, that the formate level for MTHFD1L knockdown cell lines 1 and 2 appears to be lower when compared to the shRNA control with a -1.58 (p=0.007) and -1.69 (p=0.003) fold reduced formate concentration calculated (as per Table 3.2), respectively. The knockdowns 1 and 2 also appear to have a lower formate level relative to the normal HEK 293 cells. Although the normal HEK 293 cells are not the true control for the knockdown cell lines; the lower intracellular formate level observed in the knockdowns relative to the normal HEK 293 cells provides further evidence that knocking down the MTHFD1L gene has had an effect on the HEK 293 cells.

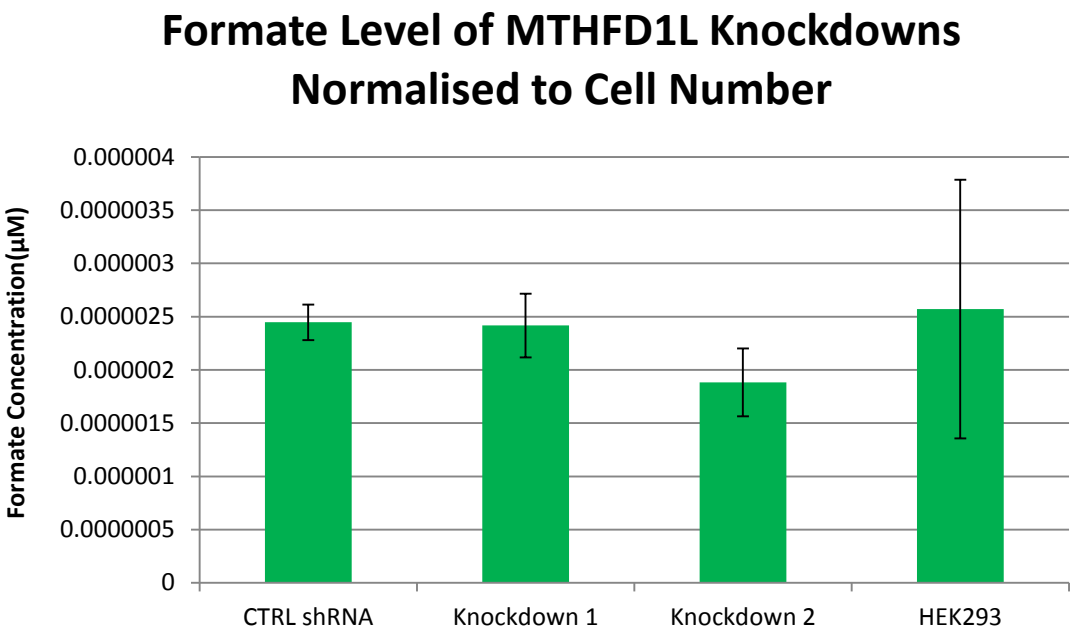


Figure 3.7 Formate Levels of MTHFD1L Knockdown cell line Normalised to cell number. When formate levels are normalised to cell number the MTHFD1L knockdown 2 has a -1.31 fold lower lower formate level p=.0.106 (as per Table 3.2) compared to the control shRNA control. The MTHFD1L knockdown 1 has a marginally reduced formate concentration (-1.02 fold as per Table 3.2) relative to the shRNA control p=0.409.

3.3 Formate Cellular Medium Analysis in HEK 293 Cells with Modulated MTHFD1L

3.3.1 Cell Medium Formate Analysis in MTHFD1L Overexpressing cells

The formate concentration in the cell medium of each cell line was monitored by collecting aliquots of medium every 12 hours for 5 days of growth, as per section 2.2.2 of Chapter 2. This allowed for the determination of formate concentration throughout the experiment as opposed to a final formate concentration. It can be seen from Figures 3.8 and 3.9 that both the MTHFD1L overexpressing cell line and the normal HEK 293 cells have similar formate level distribution patterns throughout the 5 day experiment. However, the formate level for the MTHFD1L over expressing cells peaks between days 2.5 and 3, and then begins to decline. Whereas the formate level for the normal HEK 293 cells peak at days 3 and 3.5 and like the MTHFD1L overexpressing cells then begins to decline. The cell medium formate level for the empty vector control cells, unlike the MTHFD1L over expressing and normal HEK 293 cells can be seen to increase steadily throughout the five day experiment. The different pattern in formate levels observed for the empty vector control may be due to the issues with growth and attachment of the cells in the first day of the experiment as explained above.

It can be seen from Figure 3.9 that the MTHFD1L over expressing cell line has the highest overall cellular medium formate concentration. This result is in concordance with the intracellular formate analysis with MTHFD1L over expressing cells also having the highest formate level, relative to the empty vector control and the normal HEK 293 cells.

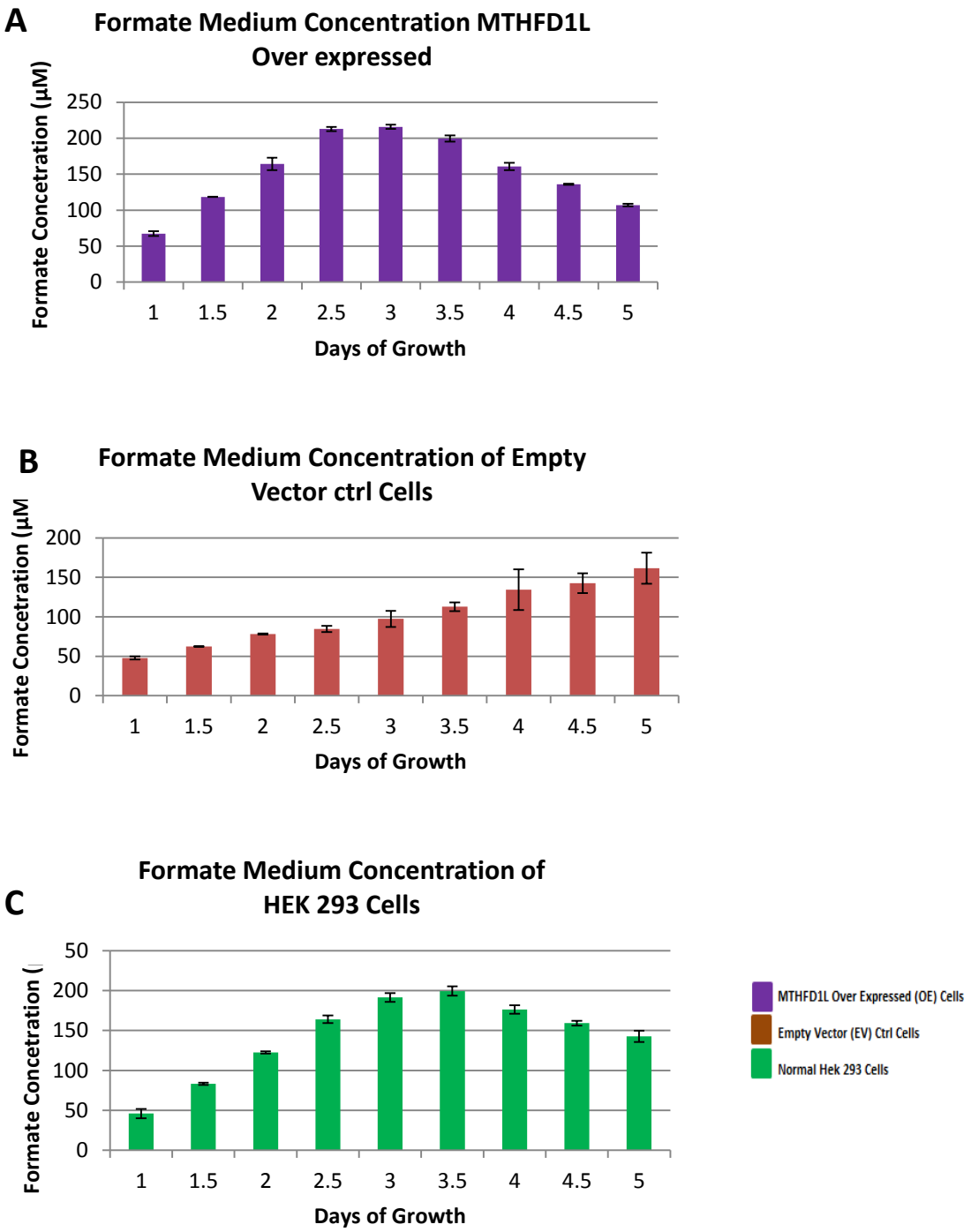


Figure 3.8 Formate Cellular Medium Concentrations. Sections A, B and C show the cell medium formate level for the MTHFD1L overexpressed, empty vector control and the normal HEK 293 cells throughout the 5 day experiment. It can be seen in section A and C that the formate distribution patterns are similar for both the MTHFD1L overexpressed and HEK 293 cells. The MTHFD1L overexpressed cell's medium concentration peaks at 216 µM of formate on day 3. The normal HEK 293 cell line has a peak formate concentration on day 3.5 with a concentration of 199 216 µM of formate The Empty vector control formate medium levels can be seen to increase steadily throughout the 5 days of growth, with the highest formate concentation observed on day 5 at 161µM of formate.

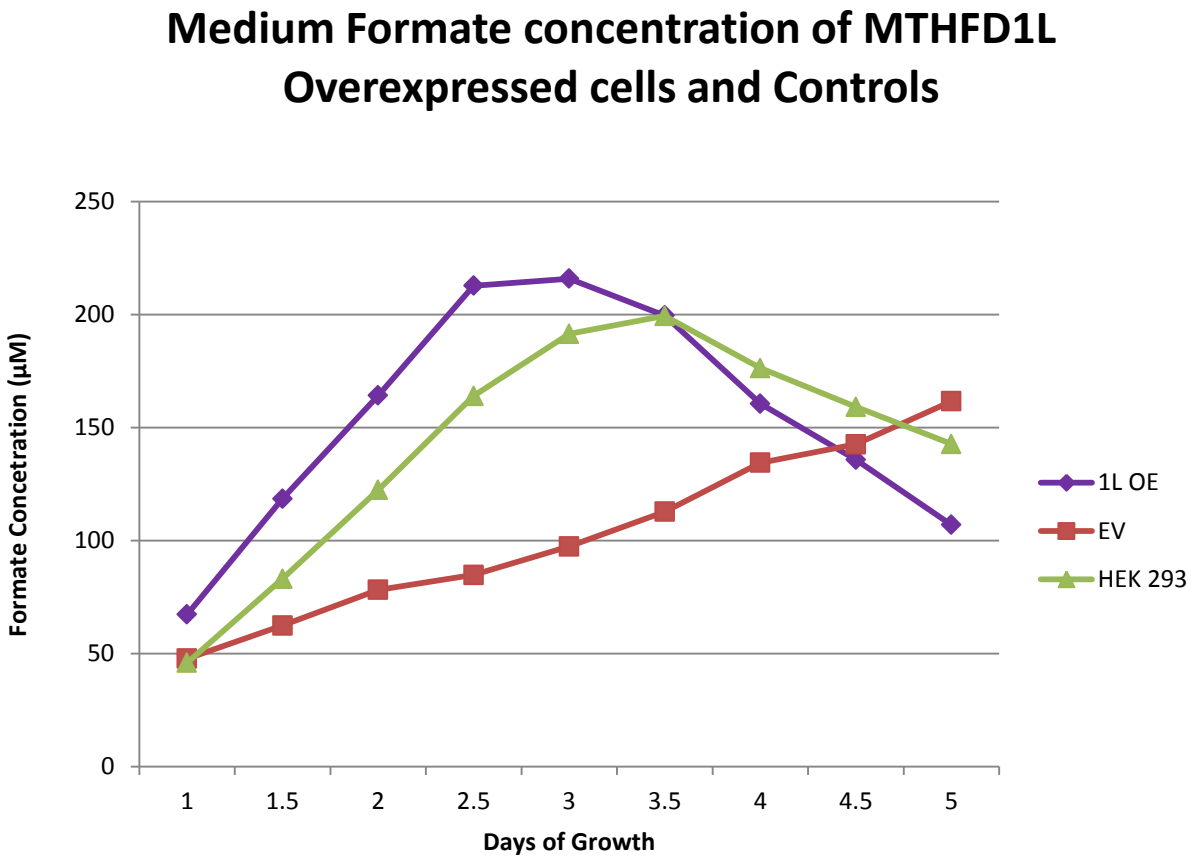


Figure 3.9 Cellular Medium Formate Concentrations for MTHFD1L Overexpressed Cell Line. The MTHFD1L over expressed and normal HEK 293 cell lines appear to have a similar formate distribution pattern. However it can be seen that the MTHFD1L overexpressed cell line has a higher cellular medium formate concentration in comparison to the HEK 293 cells from days 1 to 3.5 where it then begins to decline. The MTHFD1L overexpressed cell line also reaches the highest maximal value over all for the cellular medium formate concentration. The empty vector cells denoted by the red line appear to have consistently increasing formate concentration which may be attributed to the recovery and growth of the cells as mentioned previously.

3.3.2 Cell Medium Formate Analysis in MTHFD1L Knockdown Cells

The cellular medium formate analysis for the MTHFD1L knockdown cells and associated control indicated that knocking down the MTHFD1L had an effect on formate concentration within the cell medium. It can be seen from Figures 3.10 and 3.11 that both the MTHFD1L knockdown cells lines 1 and 2 exhibited a similar formate distribution pattern throughout the 5 day experiment. The non-targeting shRNA control appears to have an increasing formate concentration throughout the 5 day experiment where it then appears to plateau.

Figure 3.11 shows that the MTHFD1L knock down cell lines appear to have a much lower cellular medium formate concentration when compared to the non-targeting shRNA control. The shRNA control formate level starts off at a similar level to the MTHFD1L knockdowns, however by day 1.5 the formate level appears to increase well above that of the MTHFD1L knockdowns.

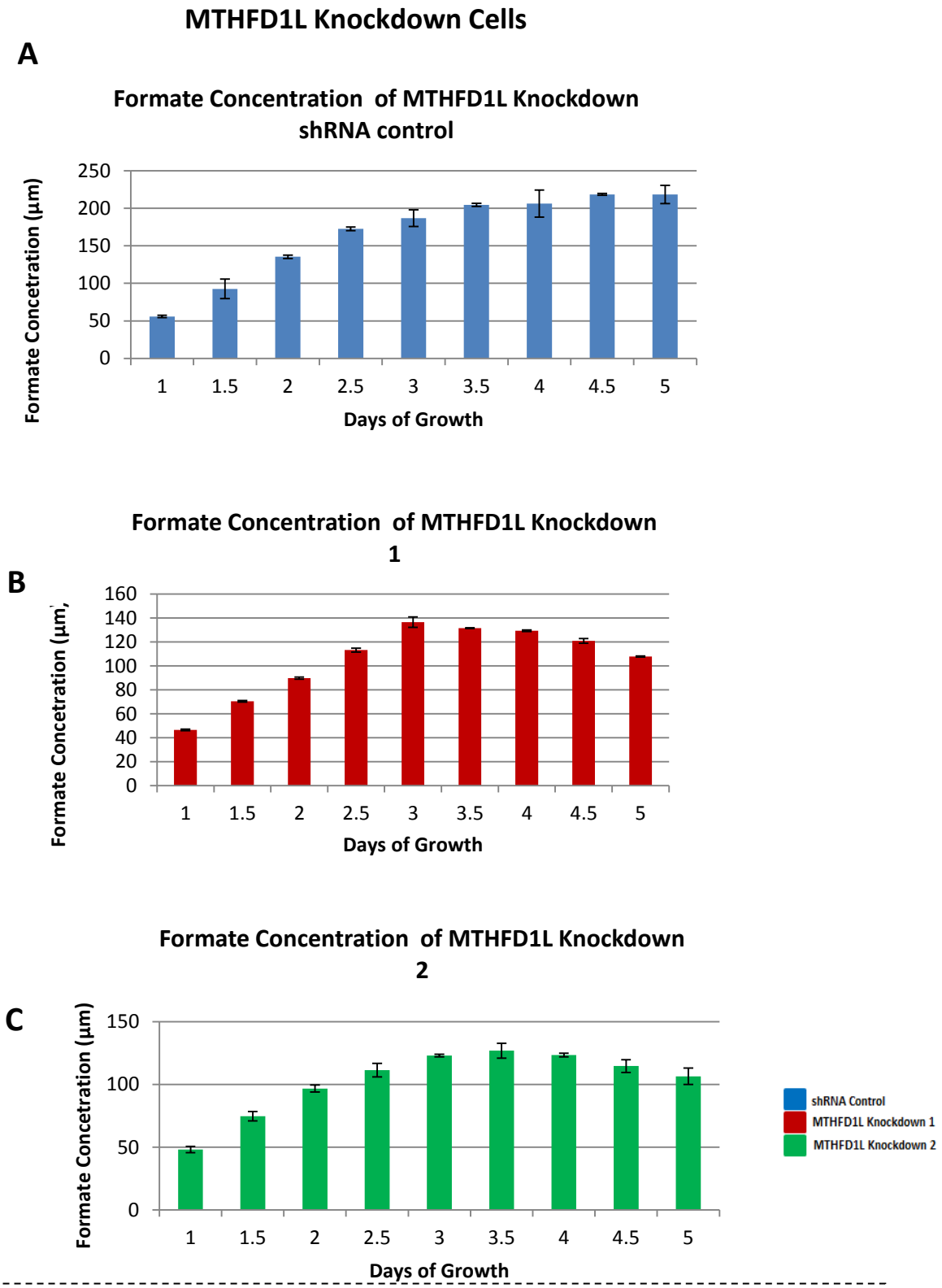


Figure 3.10 Formate Cellular Medium Concentrations. Sections A, B and C show the cell medium formate level for the shRNA control and the MTHFD1L knockdowns 1 and 2 throughout the 5 day experiment. It can be seen in section B and C that both the MTHFD1L knockdown cell lines 1 and 2 have a similar formate distribution patterns. The formate level of both the MTHFD1L knockdown cell lines increases steadily, plateaus and the declines. The formate level of the shRNA control cell line also increases steadily but does not exhibit a decline unlike the two knockdown cell lines.

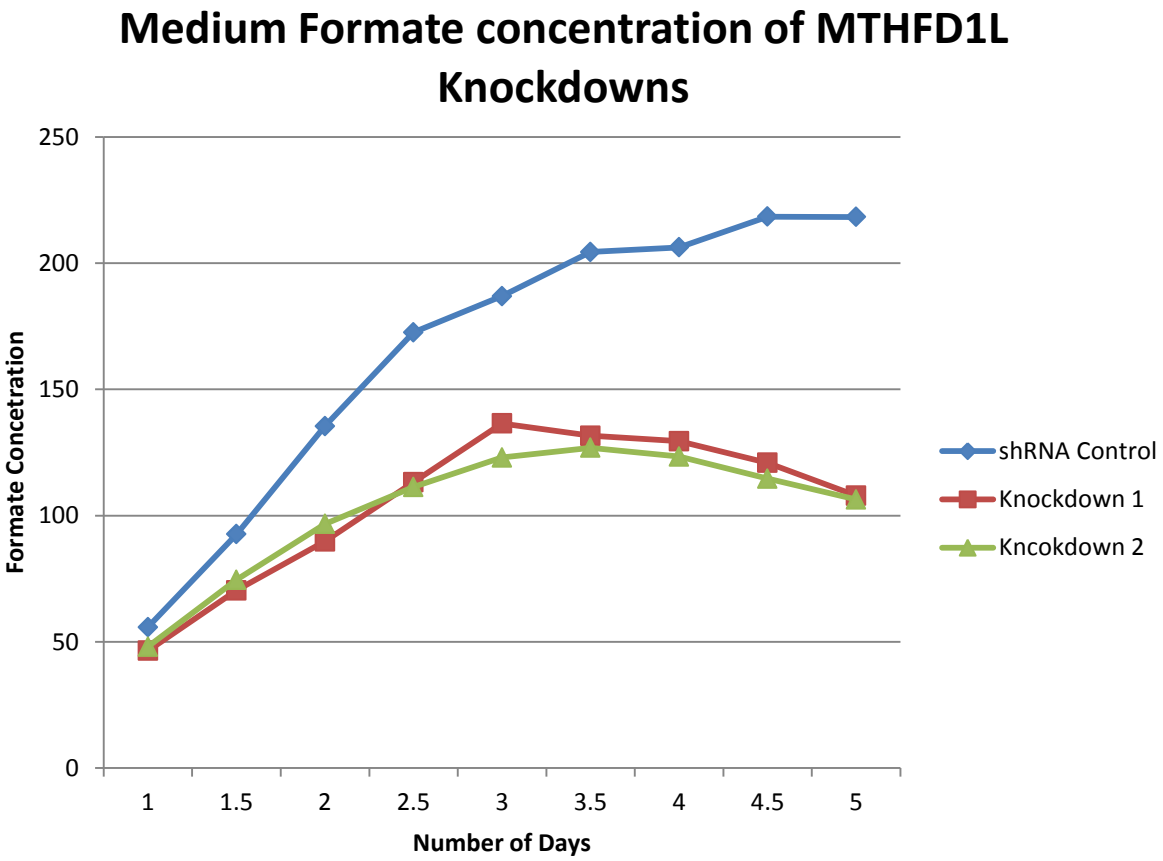


Figure 3.11 Formate Cellular Medium Concentrations. The shRNA control cell line, as denoted by the blues line, reaches the highest formate concentration over the 5 days of growth. The MTHFD1L knockdowns 1 and 2 have a similar formate distribution to each other throughout the 5 days of growth. The similarity between the two MTHFD1L knock down cell lines indicates that knocking down the MTHFD1L gene has affected the cells formate production.

3.4 Growth Analysis in HEK 293 Cells with Modulated Expression of MTHFD1L

Growth analysis of the HEK 293 cells with modulated MTHFD1L expression was achieved by counting the cell number for each cell line after a 5 day period of growth. Differential expression of the MTHFD1L gene was found to impact cell growth in HEK 293 cells. From Figure 3.12 it can be observed that the MTHFD1L over expressing cells had increased growth when compared to the normal HEK 293 cells and the empty vector control, see Table 3.3 for fold differences. It can be seen form Figure 3.13 knocking down MTHFD1L within the HEK 293 cells resulted in decreased cell growth in comparison to the control, see Table 3.4 for fold differences.

Table 3.3. Fold difference in growth rate of the MTHFD1L over expressed cell line relative to the empty vector control and the normal HEK 293 cells.

Cell Line	Cell Growth
Overexpressed	5.6
Empty Vector Control	1
	Fold Difference
Overexpressed	1.14
Normal HEK 293 ctrl	1

Table 3.4. Fold difference in growth rate of the MTHFD1L Knockdown cell lines relative to the shRNA control

	Fold Difference
Cell Line	**Cell Growth**
Knockdown 1	-1.81
Knockdown 2	-1.3
shRNA Control	1

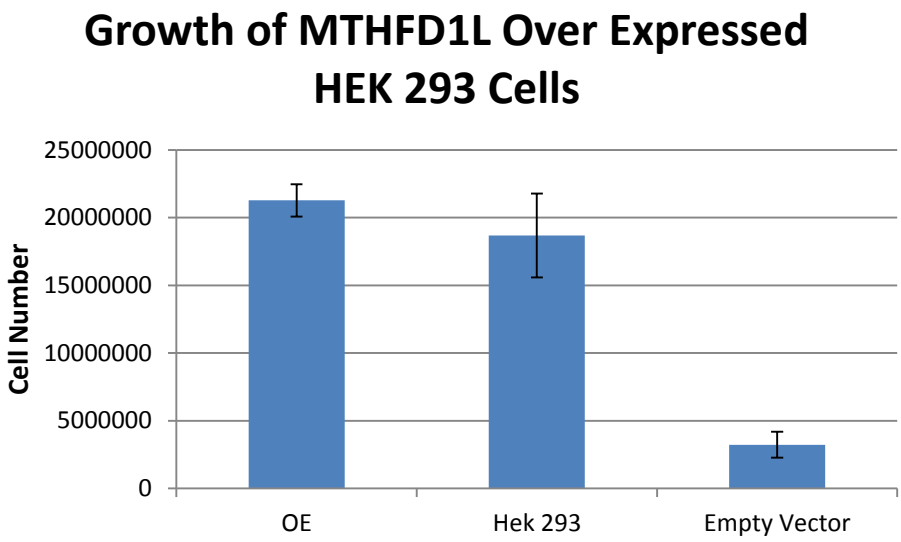


Figure 3.12 Growth of MTHFD1L Over expressed cells and associated control. It can be seen that MTHFD1L over expressed cells grew faster than both the HEK 293 cells and the empty vector control. The empty vector growth rate appears to be lower than the overexpressed and the normal HEK 293 cells. The empty vector control may have been affected by the initial attachment issues that occurred, or it may be that the selective agent G418 in which the cells are grown in has an effect on the growth rate of the cells and the overexpression of the MTHFD1L allows the cells to overcome this effect.

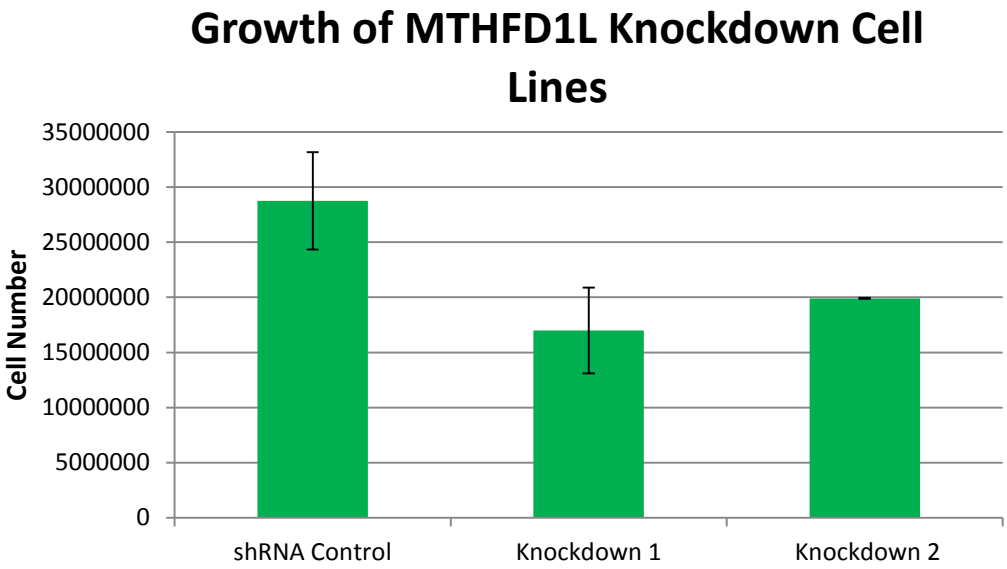


Figure 3.13 Growth of MTHFD1L knockdown cells and associated control. It can be observed that the MTHFD1L knockdown cell lines 1 and 2 grew slower than the non-targeting shRNA control. Knock down 1 had -1.81 fold decreased growth and knockdown 2 had -1.3 fold decreased growth relative to the shRNA control.

3.5 Confirmation of MTHFD1L Expression in the Knockdown and Overexpressed Cell lines

RT-qPCR expression analysis of the MTHFD1L gene was carried in all of the HEK 293 cells with modulated expression of the MTHFD1L gene and associated controls as per section 2.2.2 and 2.3 of Chapter 2. This analysis confirmed that the MTHFD1L 1 knockdown cell lines 1 and 2 had decreased MTHFD1L expression, having only 25% and 35% MTHFD1L expression relative to shRNA control cell line, respectively, see Figure 3.14 below. The MTHFD1L over expressed cell line contained pcDNA 3.2 vector which constitutively over expressed an "optimised" version of the MTHFD1L gene. The MTHFD1L overexpressed clone was optimised for codon expression in humans, meaning that the DNA sequence was different from the MTHFD1L gene but the amino acid and resultant protein had the same sequence as the MTHFD1L gene (Minguzzi 2013). RT-qPCR was carried out to detect the expression of the MTHFD1L overexpressed gene and endogenous MTHFD1L expression in both the MTHFD1L over expressed cell line and the empty vector see Figures 3.14, 3.15 and 3.16. The optimised MTHFD1L gene was confirmed as over expressed.

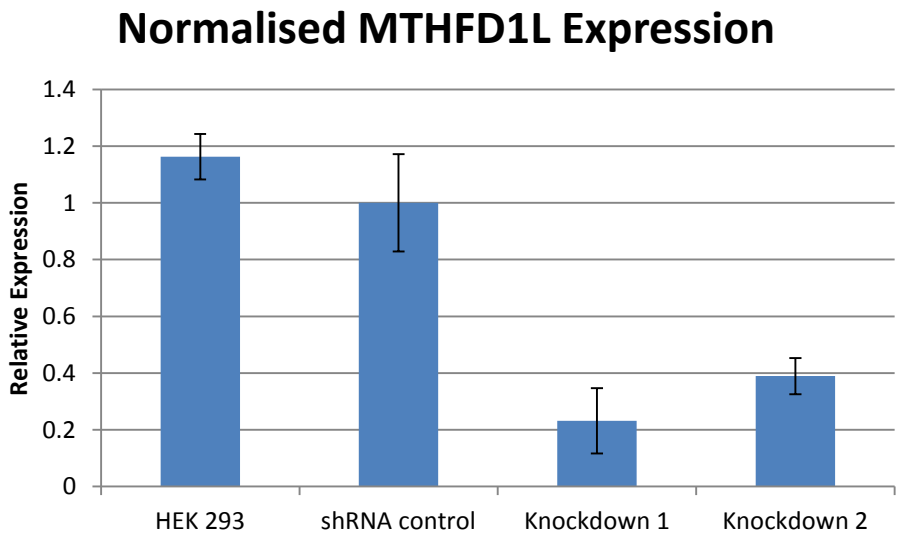
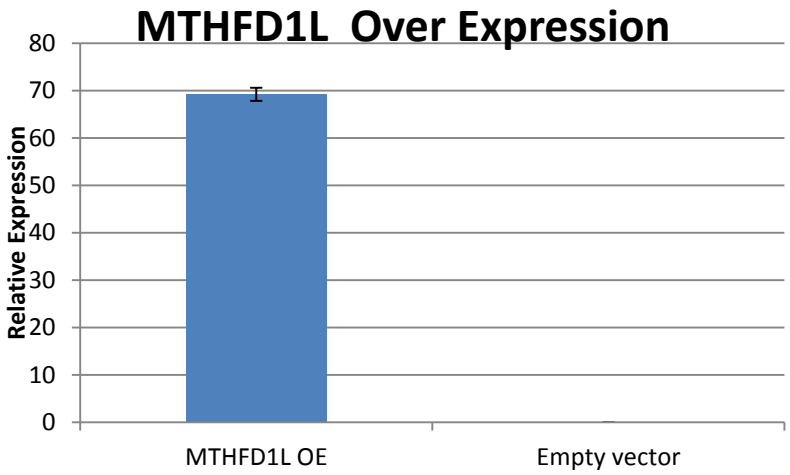


Figure 3.14 MTHFD1L expression in knockdowns and controls. Expression level of the MTHFD1L gene in knockdowns 1 and 2 is approximately 25% and 35% that of the shRNA control. MTHFD1L gene expression for each cell line was normalised to GUS gene expression.



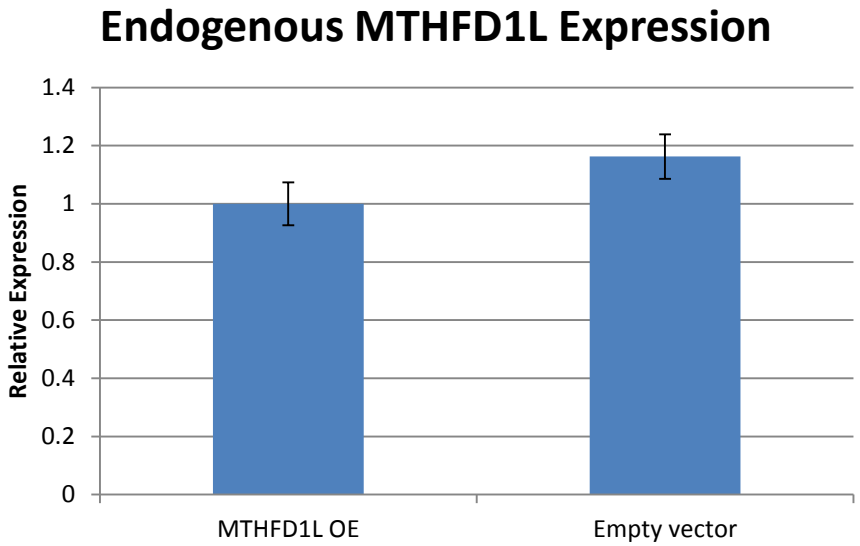


Figure 3.16 Relative Expression of endogenous MTHFD1L expression in the MTHFD1L overexpressed cell line and the empty vector. It can be seen that endogenous expression of the MTHFD1L gene is much the same between the empty vector control and the MTHFD1L overexpressing clone. Endogenous MTHFD1L expression is slightly lower in the MTHFD1L Opt clone but not significantly so.

3.6 Discussion

MTHFD1L is an essential enzyme responsible for the provision of one carbon donor molecules in the form of formate for cytoplasmic one carbon metabolism (Momb *et al*. 2012). MTHFD1L has been implicated in playing a role in many diseases such as NTD'S, CVD, Alzheimer's and cancer (Parle-McDermott *et al.* 2009, Minguzzi *et al.* 2014, Naj *et al.* 2010, Ren *et al.* 2011, Nilesh *et al.* 2007, Samani *et al.* 2007). In addition to this MTHFD1L has recently been implicated as playing a significant role in cancer proliferation and mortality (Sugiura *et al.* 2004, Jain *et al.* 2012). There is evidence to suggest that MTHFD1L is working in conjunction with other mitochondrial folate enzymes, SHMT2 and MTHFD2 in highly proliferative cancers (Jain et al. 2012, Nilsson *et al.* 2014). The up-regulation of the mitochondrial folate pathway is highly advantageous to a growing cancer cell with the mitochondrial enzymes providing the necessary components in the form of formate and glycine for the cell to grow divide and proliferate. The Warburg effect is a phenomenon known to occur in cancer cells, where energy production is switched from the energy efficient method of oxidative phosphorylation to the less energy efficient aerobic glycolysis (Warburg *et al.* 1927). The possible mechanisms behind the up-regulation of the mitochondrial folate pathway shall be discussed in detail in Chapter 5.

MTHFD1L's recent association with cancer growth, proliferation and mortality means that a biomarker for its expression would be extremely desirable (Jain *et al.* 2012, Nilsson *et al.* 2014, Selcuklu *et al.* 2012). The HEK 293 cells with modulated MTHFD1L expression act a good model for gauging formate levels both intracellulary and extracellulary. When the intracellular formate concentration was measured for the MTHFD1L over expressing cell line it was found that with normalisation to both cell number and protein concentration its formate level was 2.89 and 1.85 fold higher, respectively, relative to the empty vector control, Figure 3.4, 3.5. The normal HEK 293 cells were included as an extra control for the MTHFD1L over expressing cell line due to issues which occurred with the empty vector control cell line. The differences observed in normalisation between the normal HEK 293 cell line and the empty vector control cell line may be due to two

factors. One is that the empty vector is the true control and mimics the effect of the pcDNA 3.2 vector and the associated effects that would be exerted on the cells from having been transfected and grown in the G418 selective agent. It may be that the differences in formate level between the MTHFD1L overespressing cells and the empty vector are a true representation of the expression system and the effect that up regulating the MTHFD1L gene has on HEK 293 formate levels. Another factor which may be responsible for the discrepancies between the formate levels of the normal HEK 293 cells and the empty vector control is the fact that the empty vector cells did not attach as well as the MTHFD1L overexpressing and normal HEK 293 cells. However, normalisation to protein level in addition to cell number should have controlled for the difference in cell number. The experiment would have to be repeated to conclusively ascertain the effect of overexpressing the MTHFD1L gene relative to the empty vector control in HEK 293 cells. Unfortunately due to time constraints repetition of the experiment was not possible.

The MTHFD1L knockdown cell lines 1 and 2 exhibited lower intracellular formate concentration when normalised to protein and cell number. When the formate level of knockdown 1 was normalised to protein and cell number it had a -1.58 and a -1.02 fold decreased expression, respectively, when compared to the shRNA control. When the formate level of knockdown 2 was normalised to protein and cell number it had a -1.69 and a -1.31 fold decreased expression, respectively, when compared to the shRNA control. Although there are differences observed with the two normalisation methods, as evidenced by the statistical significance results in Table 3.2, an overall detectable decrease in formate concentration was observed, specifically for knockdown 2. The detection of altered formate levels supports Minguzzi's (2013) hypothesis that formate (either increased or decreased) mediated changes in expression of a number key enzymes within the MTHFD1L modulated HEK 293 cells (Minguzzi 2013). In relation to formate as a biomarker for MTHFD1L expression, both the knockdown and overexpressed cell lines provide further evidence that altered MTHFD1L gene expression effects formate production.

It was not known whether the intracellular formate levels would necessarily correlate with the cellular formate medium concentrations. However, consistent

with the intracellular formate data the MTHFD1L over expressed cells reached the highest maximal formate value when compared to the empty vector control and the normal HEK 293 cells. In Figure 3.8, it can be seen that the distribution pattern of formate was similar in the HEK 293 and the MTHFD1L over expressed cell line. Although the pattern is similar the over expressed MTHFD1L cell line has a much higher formate level on days 1 through 3 were the normal HEK 293 cells appears to peak on day 3.5. The fact that the cellular medium formate level in the MTHFD1L over expressing cells is higher than the normal HEK 293 cells from day one indicates that the overexpression of the MTHFD1L gene is having an effect on the metabolism of the cell. As can be seen in Figure 3.10 cell medium formate concentration for the MTHFD1L knockdowns 1 and 2 appears to be much lower than the shRNA control. The shRNA control's formate level appears to increase steadily and is seen to plateau on days 4.5 and 5. The MTHFD1L knockdown cell lines 1 and 2 cell medium formate level peaks between days 3 and 3.5 where it then begins to decline. These results demonstrate that altering MTHFD1L gene expression has resulted in a detectable change in the extra-cellular formate concentration. The detection of differences in formate level in the extracellular environment of the cell is very exciting and may have implications for the use of formate as a biomarker in the detections of altered MTHFD1L expression in diseases such as cancer.

In agreement with this finding a paper by Wang *et al* (2013) found formate to be significantly elevated in the mucosae of patients with oesophageal cancer versus healthy individuals when analysed by Nuclear Magnetic Resonance (NMR), (*p* 0.0005). Most interesting of all is that they found that the level of formate could be correlated to the stage and progression of the cancer at stages II,III,IV with significance of *p* 0.001, *p* 0.001 and *p* 0.0001 respectively. Wang *et al* (2013) also found that other metabolites involved in one carbon metabolism, uracil, methionine and glycine were also significantly linked to the stage and progression of oesophageal cancer, $p < 0.0001$, $p < 0.0001$ and $p < 0.0001$, respectively. Farshidfar *et al* (2012) also found that metabolomic analysis of serum samples from patients with adenocarcinoma identified that formate levels were significantly increased in patients with liver metastasis (*p*0.0005) (Farshidfar 2012).

Further confirmation that up regulating and knocking down the MTHFD1L had a significant impact on the HEK 293 cells comes from the fact that their growth rate was altered. The MTHFD1L knockdowns 1 and 2 had a growth rate of -1.81 and -1.3 fold that of the shRNA control, or in percentage terms knockdowns 1 and 2 had a growth rate that was reduced by 45% and 23% respectively. The MTHFD1L overexpressed cell lines had a 1.14 fold higher growth rate when compared to the normal HEK 293 cell line. The increased growth rate observed in the MTHFD1L over expressed HEK 293 cell line is in concordance with findings in the literature whereby Sugiura *et al* (2004) found that up regulating MTHFD1L resulted in "increased colony formation" in HEK 293 cells. Jain *et al (*2012) also showed a correlation with increased cell growth rate and expression of the MTHFD1L gene in various cancers. The fact that knocking down the MTHFD1L gene resulted in decreased cell growth in the HEK 293 cell lines is indeed very promising and provides evidence that the development of targeted therapies against MTHFD1L could be very effective in cancers which exhibit MTHFD1L over expression.

Previous work on the HEK 293 MTHFD1L cell lines found that the expression of the enzyme TYMS (thymidylate synthase) was decreased upon MTHFD1L over expression (Minguzzi 2013). TYMS is responsible for the production of the DNA base thymine. Minguzzi (2013) hypothesised that formate may act as a sensor for nucleotide supply. HEK 293 cells are a non-cancerous "normal" cell line. The decrease in TYMS expression in the HEK 293 MTHFD1L over expressed cell line was thought to be protective to prevent increased DNA synthesis and abnormal cell proliferation (Minguzzi 2013). When thymine is limited uracil is incorporated into DNA in its place. Consequently the incorporation of uracil in the place of thymidylate may cause strand breaks and subsequent abnormalities with DNA replication, increasing the risk of cancer development (Blount *et al.* 2002). The work by Wang and colleagues (2013) may demonstrate a similar scenario as in the MTHFD1L over expressed HEK293 cell line, as they observed an increased formate level indicating that MTHFD1L is up regulated. The fact that they observed an increased uracil level may be hypothetically due to the fact that MTHFD1L up regulation caused a decrease in TYMS gene expression levels and thymidine

concentration, in turn up regulating uracil concentration to take the place of thymidine; see Figure 3.17.

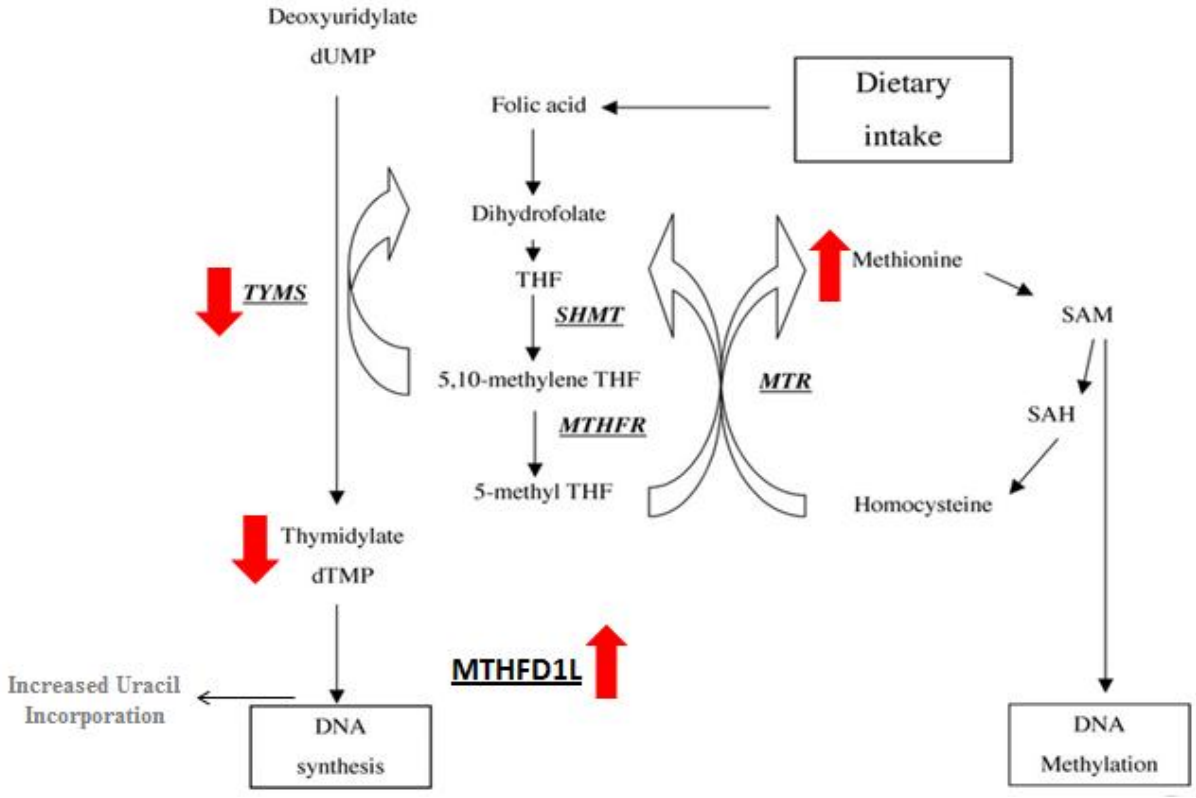


Figure 3.17 Scenario of the knock on effects MTHFD1L up regulation. Previous work in the HEK 293 cells found that Methylenetetrahydrofolate dehydrogenase ($NADP^+$ dependent) 1-like (MTHFD1L) over expression resulted in decreased thymidylate synthase (TYMS) enzyme levels. A decrease in TYMS enzyme levels may result in an increase in uracil concentration and may explain the observed increase in uracil concentration in oesophageal cancer by Wang *et al* (2013) *(*image edited from niclot *et al* 2006).

MTHFD1L is not solely responsible for formate production; however, it does produce up to an estimated 75% of formate required for the cell (Pike *et al.* 2010). The fact that increasing and decreasing MTHFD1L expression at genetic level resulted in changes in both intracellular and extra-cellular formate indicates that formate could be used as a biomarker for MTHFD1L expression in regard to cancer. MTHFD1L, along with other mitochondrial folate related enzymes, has been significantly associated with high growth and proliferation rates in cancer. MTHFD1L expression has also been associated with increased mortality in cancer patients (Jain *et al.* 2012, Sugiura *et al.* 2004). Work by Lamarre *et al* (2012) has demonstrated urinary and plasma formate as an accurate biomarker for vitamin B12 and folate deficiency in rats. As previously discussed, work by Wang *et al* (2013) has shown that formate level can be significantly correlated with stage and progression of oesophageal cancer (Wang *et al.* 2013). Results from this chapter have demonstrated that changes in MTHFD1L expression results in changes in formate concentration in both the intracellular and extracellular environment of the cell. The extra cellular environment is the ideal place for biomarker identification as the biomarker itself may then be excreted into bodily fluids such a blood, urine, mucus etc. Although urinary and plasma formate is an indication of folate and vitamin B 12 status, it is not known whether an increase in MTHFD1L expression, such as in the case of cancer would result in detectable urinary or plasma formate changes. However, what is known is the work by Farshidfar *et al* (2012) who found that formate was detected in the serum of patients and was correlated with liver metastasis. The results in this chapter provide further support for formate as a biomarker for MTHFD1L expression. As mentioned previously MTHFD1L is thought to be up-regulated in conjunction with other mitochondrial folate enzymes MTHFD2 and SHMT2 and all three enzymes have been associated with highly proliferative cancers (Jain *et al.* 2012). Vasquez *et al* (2013) recently associated the expression of SHMT2 and MTHFD2 with increased methotrexate sensitivity in leukaemia patients, independent of DHFR expression. Vasquez *et al* (2013) also identified that 25% of cancers have an up-regulation of folate related enzymes. This information taken collectively means that formate may be a biomarker for not just MTHFD1L expression but for expression of other folate related enzymes MTHFD2, and SHMT2.

It is notable that MTHFD1L is the rate limiting step in the mitochondria to convert 10-formylTHF to formate (Tibbets and Appling 2010). Therefore, in a cancer setting the detection of elevated formate levels may be indicative of the up-regulation of not only MTHFD1L, but of SHMT2 and MTHFD2, given recent findings from Jain *et al* (2012). If formate is an accurate biomarker for MTHFD2 and SHMT2 expression the findings by Vasquez et al (2013) on the expression of these enzymes with the increased methotrexate sensitivity in cancer patients is of particular relevance. Therefore, one might hypothesise that up-regulation of formate may also act as a biomarker to identify individuals who would benefit from methotrexate treatment.

To conclude, the HEK 293 cells with modulated MTHFD1L expression allowed for the analysis of changes at the metabolite level which stemmed from controlled alterations at the expression level of the MTHFD1L gene. Measuring both the intracellular formate levels and cellular medium formate concentration levels allowed for a comprehensive view of the formate status of each cell line with altered MTHFD1L expression. Using two methods for formate data normalisation, protein concentration and cell number also allowed for a robust assessment of the formate status of each cell line. In addition to this, analysis of extra cellular formate in the cell medium of each cell line allowed for a time lapse view of the formate level in a given cell line. These results demonstrate that overexpressing the MTHFD1L gene in HEK 293 cell resulted in an increased intracellular formate level between 1.23-2.89 fold, increased extra-cellular formate level and an increased cell growth rate by 1.14 fold. Knocking down the MTHFD1L gene in HEK 293 cells resulted in an intracellular formate level of between -1.02-1.69 fold that of the control and decreased extra-cellular formate level. Knocking down the MTHFD1L gene also resulted in a 23% and 45% reduction in growth rate in HEK 293 cells. These results provide support for the hypothesis that altered formate levels in the MTHFD1L HEK 293 cell lines mediated changes in DNA repair and synthesising enzymes (Minguzzi 2013). They also provide further evidence of the impact expression of the MTHFD1L gene has on the growth rate of cells. Most importantly this work provides further evidence for the use of formate as a biomarker for MTHFD1L expression in cancer.

Chapter 4

Investigation into the Effect of Amino Acid Differences between DHFR and DHFRL1

4.0 Introduction

In 2011 it was discovered that there is a second dihydrofolate reductase (DHFR) enzyme, dihydrofolate reductase like 1 (DHFRL1). DHFRL1 shares 92% homology with the DHFR enzyme (McEntee *et al.* 2011, Anderson *et al.* 2011). DHFR is involved in folate metabolism via the one carbon metabolism biochemical pathway. DHFR reduces folic acid and dihydrofolate (DHF) to its biologically active form tetrahydrofolate (THF) using NADPH as a co-factor for the reaction. One carbon metabolism results in the production of purines, thymidylate and methyl donor groups for cellular methylation reactions (Fox and Stover 2009). Although there is only an 8% amino acid difference between the two enzymes, it has been shown that DHFRL1 activity is 10 times less than that of DHFR (McEntee *et al.* 2011). McEntee *et al* (2011) have demonstrated that DHFRL1 is capable of compensating for DHFR knockout in both a bacterial and a mammalian system. DHFRL1 may also play a role in the control of expression of both itself and of DHFR. DHFR is known to bind its own mRNA to prevent translation; it has also been shown that DHFRL1 is capable of binding its own mRNA and the mRNA of DHFR (McEntee *et al.* 2011). In the treatment of cancer DHFR has been the target of anti-folate chemotherapy for over 60 years (Miller 2006). The discovery of a second DHFR enzyme, DHFRL1, means that the novel enzyme could also potentially be a chemotherapeutic target in the treatment of cancer in a similar manner as DHFR. Treatment of cancer with the anti-folate drug methotrexate initially results in an up regulation of the DHFR enzyme due to the fact that the methotrexate binds the DHFR enzyme causing disassociation of the mRNA leading to translation of the released mRNA (Ercikan-Abali *et al.* 1997). The different affinities of DHFR and DHFRL1 for folate suggest that they may have different affinities to anti-folate drugs which could potentially impact the control of expression of both DHFR and DHFRL1.

DHFRL1 has been shown to translocate to the mitochondria where it partakes in mitochondrial one carbon metabolism (McEntee *et al.* 2011, Anderson *et al.* 2011). DHFRL1 is involved in the de novo thymidylate synthesis pathway within the mitochondria. DHFRL1 is responsible for the reduction of DHF to THF in the mitochondria. THF and one carbon activated molecules derived from serine by the

enzyme serine hydroxymethyltransferase 2 (SHMT2) are converted to thymidylate by the enzyme thymidylate synthase (TYMS). This reaction also serves to regenerate DHF for further cycles of thymidylate synthesis (Anderson *et al.* 2011). The presence of a de novo thymidylate pathway within the mitochondria was only confirmed in 2011 (Anderson *et al.* 2011). Thymidylate is the only non-essential DNA base with uracil taking its place when thymidylate supply is limited (MacFarlane *et al.* 2011). Mitochondrial derived thymidylate is thought to be necessary to maintain the integrity of mitochondrial DNA (mtDNA) by preventing the incorporation of uracil (Anderson *et al.* 2011). mtDNA damage and TYMS inhibition has previously been associated with apoptosis. Due to these previous associations Anderson *et al* (2011) hypothesised that inhibition of thymidylate synthesis may result in increased mtDNA damage and cell death. As stated previously, DHFRL1 could potentially be a chemotherapeutic target. If Anderson *et al's* (2011) hypothesis is true DHFRL1 would be an ideal target in the treatment of cancer. Further investigation would be needed to assess the possibility of DHFRL1 as a viable target in the treatment of cancer.

4.1 Amino Acid differences between DHFRL1 and DHFR

Of the 15 amino acid differences between DHFRL1 and DHFR only one amino acid at position 24 is thought to be important for the binding of DHF and NADPH (McEntee *et al.* 2011). The DHFR enzyme has a tryptophan (W) at position 24 whereas DHFRL1 has an arginine (R). McEntee *et al* (2011) noted the W24R amino acid change as significant and hypothesised that it may be a source of the DHFRL1's reduced catalytic activity. The tryptophan at position 24 is highly conserved in the DHFR enzyme across many species. Indeed work by Beard *et al* (1991) found that site directed mutagenesis of the tryptophan at position 24 of human DHFR to a phenylalanine caused a 25 fold reduced affinity of the enzyme for DHF and a 21 fold reduced affinity for NADPH. The phenylalanine substitution also resulted in a 50% decrease in DHFR enzyme stability and a dramatic reduction in enzyme efficiency (Beard *et al.* 1991). Similarly work on mouse DHFR has also shown the importance of the tryptophan at position 24 of the enzyme. Thillet *et al* (1988) performed site directed mutagenesis on mouse DHFR, producing a number of

mutants one of which resulted in a tryptophan (W) to arginine (R) substitution at position 24. The W24R mutation in mouse DHFR resulted in over a 100 fold reduced affinity for folate, with the wild type enzyme having a Km of 0.09µM for folate whilst mutant had a Km of 10 µM. The authors also found that introducing a W24R mutation in mouse DHFR affected its ability to bind methotrexate, a known anti-folate drug used frequently in chemotherapeutic regimens. It is thought that the tryptophan at position 24 forms a hydrogen bond with methotrexate and that switching this amino acid to an arginine disrupts this binding. Thillet *et al* (1988) found that the W24R mutation in mouse DHFR resulted in it having a Ki of 300nM for methotrexate whereas the wild type enzyme has a Ki of 0.004nM; this results in the DHFR mutant (W24R) having a 75,000 fold reduced affinity for methotrexate (Thillet *et al.* 1988). Although this work was carried out in mouse DHFR, this could have some bearing on the human DHFRL1 enzyme where an arginine is found at position 24 in the conserved catalytic activity site in place of the highly conserved tryptophan residue.

Gao *et al* (2013) performed a molecular dynamics simulation as to the effect of mutating the arginine at position 24 in DHFRL1 to a tryptophan. Free energy calculations into enzyme substrate affinity found that the calculated values were in concordance with experimental values for DHFR and DHFRL1. Most interestingly the calculated values for DHFRL1 R24W indicated that the mutation recovers the binding affinity for DHF to that of DHFR. Molecular modelling also indicated that wild type DHFRL1 has a conformation which is more similar to *E.coli* DHFR that to human DHFR.

Despite shared genetic homology, there are distinct differences between the paralogous enzymes DHFR and DHFRL1. DHFRL1 has been shown to translocate to the mitochondria to partake in the de novo synthesis of thymidylate (Anderson *et al.* 2011). Whereas, DHFR is thought to primarily reside in the cytoplasm partaking in cytoplasmic one carbon metabolism. Both enzymes are capable of the reduction of DHF to its biologically active form THF; however, their abilities to perform this function are somewhat different. DHFRL1's activity has been shown to be only one tenth that of DHFR (McEntee *et al.* 2011). In addition, molecular dynamic simulation

into the structure and conformation of the DHFRL1 enzyme indicates that its structure may be different from DHFR's (Gao *et al.* 2013). Investigation into the effects of the amino acid differences between the enzymes is needed in order to identify specific residues which may contribute to the diversity between the enzymes.

4.2 Aims/Objectives

The main aim of this chapter was to investigate the effects of specific amino acid differences between DHFR and DHFRL1in order to begin to understand their functional relevance.

Objectives:

- Predictive protein modelling of DHFR, DHFRL1 and examination of amino acid substitutions in DHFRL1 R24W and DHFR W24R to ascertain the effect if any on predicted binding or folding.
- Expression of recombinant DHFR and DHFRL1 with purification of the recombinant enzymes.
- Analysis of the recombinant proteins using SDS PAGE and Western blot.
- Site directed mutagenesis to introduce the DHFRL1 R24W and the DHFR W24R mutations into the DHFRL1 and DHFR genes.
- Optimisation of expression of the DHFRL1 R24W and the DHFR W24R mutant proteins. Analysis of the effect of these mutations on protein expression and solubility.

4.3 Results

4.3.1 Identification of Important Amino Acid Differences between DHFR and DHFRL1

Since there are 15 amino acid differences between DHFR and DHFRL1, and there is a possibility of two amino acids at each of the 15 sites. Mathematically there are up to 2^{15} (33,000) possible amino acid combinations that could theoretically be responsible for the differences in enzyme activity between the two proteins. Therefore a rational approach had to be taken to identify amino acids which may be of particular importance. A Clustal Omega alignment of the amino acid sequences of DHFRL1 and DHFR from human, cow, rat, mouse, chicken and chimpanzee was undertaken in order to identify conserved amino acids across species which are likely to be important. One amino acid which was found to be highly conserved in the DHFR enzyme was a tryptophan (W) at position 24. Figure 4.1 shows that DHFRL1 has an arginine (R) at position 24, whereas DHFR has a tryptophan at position 24 in all 6 species examined. In addition to this, McEntee *et al* (2011) had previously noted this residue may be of particular importance in regard DHFRL1'S reduced activity. Work by Gao *et al* (2013) also provided strong evidence in the form of predictive modelling as to the relevance of the tryptophan at position 24. In order to further assess the importance of the amino acids at position 24 of predictive modelling was undertaken using Swiss-Model (Arnold *et al.* 2006, Bordoli *et al.* 2009, Biasini *et al.* 2014). Interestingly the protein modelling predicted that the DHFRL1 wild type protein did not have a conserved NADPH binding site, Figure 4.2. As mentioned previously NADPH is a co-factor for the DHFR enzyme in its reduction of folate. Mutating the amino acid at position 24 in DHFRL1 from an arginine (R) to tryptophan (W) failed to restore the NADPH binding site. Interestingly changing the tryptophan at position 24 of the DHFR protein to an arginine resulted in the loss of the NADPH binding site, as determined by the predictive modelling, see Figure 4.3. Predictive modelling alone is not conclusive for the loss or conservation of the NADPH binding site in the DHFR and DHFRL1 proteins. However, these results provided further evidence as to the amino acid at position 24 being a likely candidate for causing the differences observed between

the DHFR and DHFRL1 proteins and for these reasons it was selected for experimental analysis.

```
CLUSTAL O(1.2.1) multiple sequence alignment


CHICKEN         -VRSLNSIVAVCQNMGIGKDGNLPWPPLRNEYKYFQRMTSTSHVEGKQNAVIMGKKTWFS
BOS TAURUS      MVRPLNCIVAVSQNMGIGKNGNLPWPPLRNEFQYFQRMTTVSSVEGKQNLVIMGRKTWFS
MOUSE           MVRPLNCIVAVSQNMGIGKNGDLPWPPLRNEFKYFQRMTTTSSVEGKQNLVIMGRKTWFS
RATTUS          MVRPLNCIVAVSQNMGIGKNGDLPWPLLRNEFKYFQRMTTTSSVEGKQNLVIMGRKTWFS
DHFRL1          MFLLLNCIVAVSQNMGIGKNGDLPRPPLRNEFRYFQRMTTTSSVEGKQNLVIMGRKTWFS
DHFR            MVGSLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFS
CHIMPANZEE      MVGSLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFS
                 .  **.****.*******:*:** * ****::******:.* ****** ****:*****

CHICKEN         IPEKNRPLKDRINIVLSRELKEAPKGAHYLSKSLDDALALLDSPELKSKVDMVWIVGGTA
BOS             IPEKNRPLKDRINIVLSRELKEPPKGAHFLAKSLDDALELIEDPELTNKVDVVWIVGGSS
MOUSE           IPEKNRPLKDRINIVLSRELKEPPRGAHFLAKSLDDALRLIEQPELASKVDMVWIVGGSS
RATTUS          IPEKNRPLKDRINIVLSRELKEPPQGAHFLAKSLDDALKLIEQPELASKVDMVWVVGGSS
DHFRL1          IPEKNRPLKDRINLVLSRELKEPPQGAHFLARSLDDALKLTERPELANKVDMIWIVGGSS
DHFR            IPEKNRPLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSS
CHIMPANZEE      IPEKNRPLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELASKVDMVWIVGGSS
                ********* ***:******** *:***:*::****** * : *** .***::*:***::

CHICKEN         VYKAAMEKPINHRLFVTRILHEFESDTFFPEIDYKDFKLLTEYPGVPADIQEEDGIQYKF
BOS             VYKEAMNKPGHVRLFVTRIMQEFESDAFFPEIDFEKYKLLPEYPGVPLDVQEEKGIKYKF
MOUSE           VYQEAMNQPGHLRLFVTRIMQEFESDTFFPEIDLGKYKLLPEYPGVLSEVQEEKGIKYKF
RATTUS          VYQEAMNQPGHLRLFVTRIMQEFESDTFFPEIDLEKYKLLPEYPGVLSEIQEEKGIKYKF
DHFRL1          VYKEAMNHLGHLKLFVTRIMQDFESDTFFSEIDLEKYKLLPEYPGVLSDVQEGKHIKYKF
DHFR            VYKEAMNHPGHLKLFVTRIMQDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKF
CHIMPANZEE      VYKIPRCSL----------------------------------------------------
                **:

CHICKEN         EVYQKSVLAQ
BOS TAURUS      EVYEKNN---
MOUSE           EVYEKKD---
RATTUS          EVYEKKD---
DHFRL1 HUMAN    EVCEKDD---
DHFR HUMAN      EVYEKND---
CHIMPANZEE      ----------
```

Figure 4.1 Clustal Omega Alignment of human DHFRL1 and DHFR amino acid protein sequences. The tryptophan at amino acid position 24 appears to be highly conserved across, chicken, cow, mouse, rat, chimpanzee and human DHFR. However, the tryptophan is not conserved in the DHFRL1 enzyme as an alanine is found at position 24 of the protein.

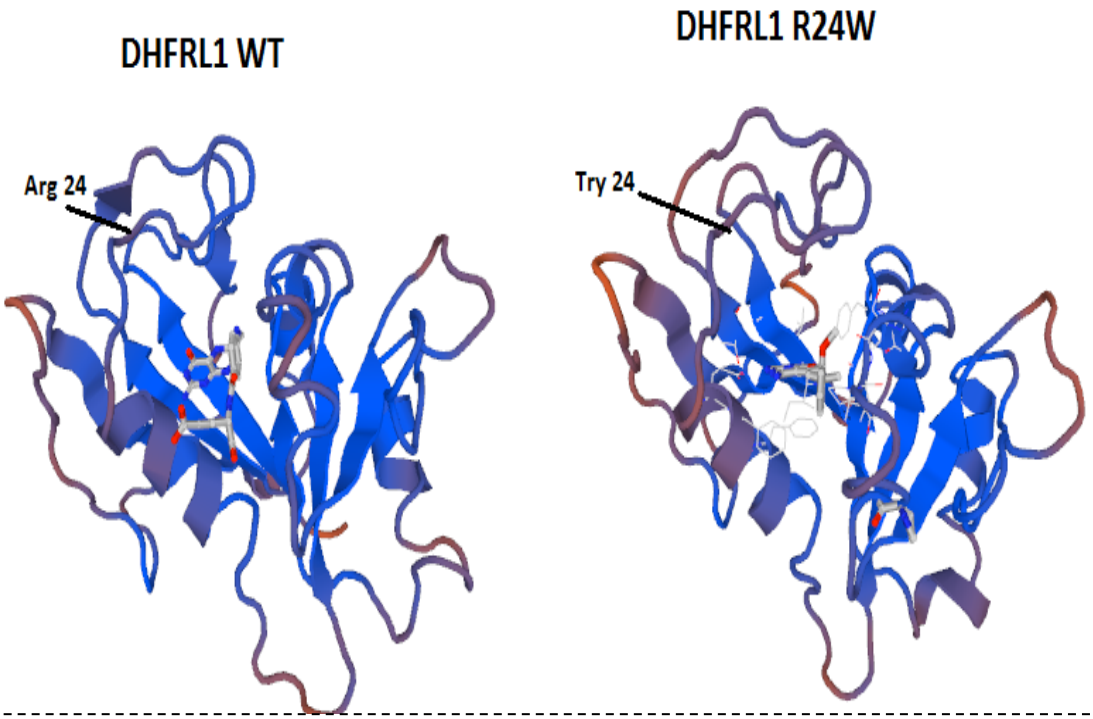


Figure 4.2 DHFRL1 WT and DHFRL1 R24W predicted protein structures. Predictive modelling indicates that DHFRL1 is capable of binding folic acid but the binding site for NADPH, a co-factor required for the reduction of folate is not conserved. Changing the amino acid of the DHFRL1 protein to that of the DHFR protein at position 24 from an arginine to a tryptophan fails to restore the NADPH binding site in the DHFRL1 protein.

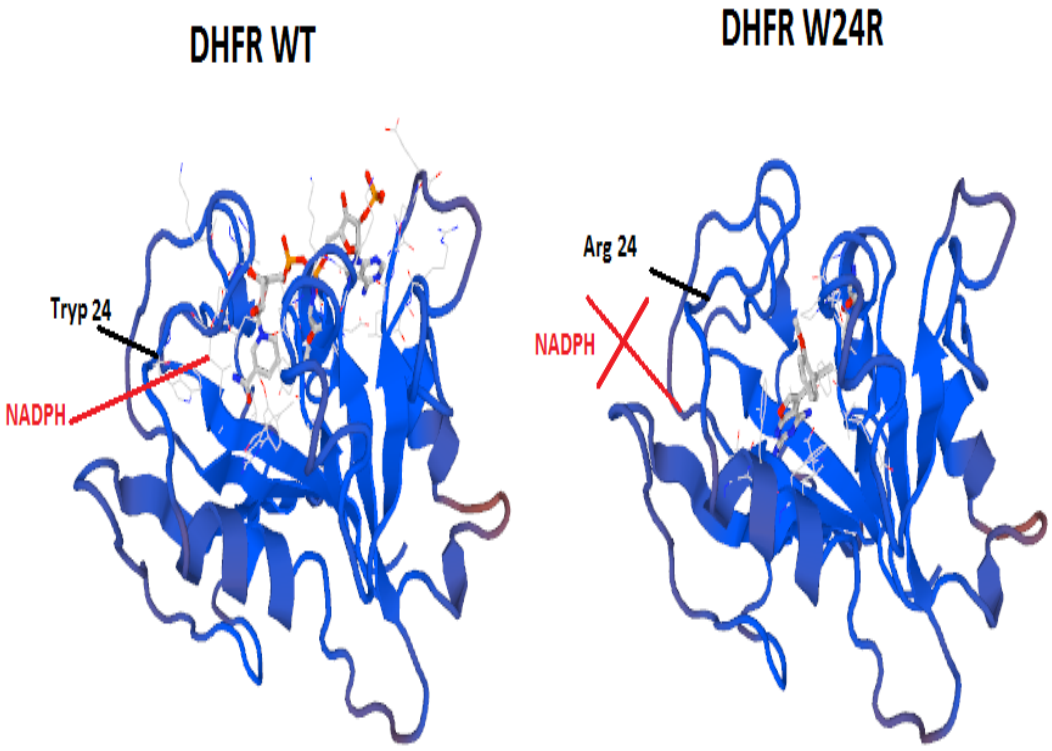


Figure 4.3 DHFR WT and DHFR W24R predicted protein structures. The single amino acid change at position 24 from a tryptophan (W) to an arginine (R) results in the loss of the NADPH binding site of the DHFR enzyme. Indicating the amino acid change affects the folding and conformation of the DHFR enzyme.

4.3.2 Expression and Purification of Recombinant DHFR and DHFRL1

DHFR and DHFRL1 ultimate orf clones in pDONR^{TM}221 were recombined with pDEST^{TM} 17 the resultant plasmid DNA was sequenced then transformed into the BL21 AI *E.coli* cells. The pDEST 17 vector encodes a His tag comprising of 6 histidine residues; this enables purification of the target proteins by affinity chromatography with either a nickel or cobalt column. The pDEST 17 vector encodes an inducible promoter; addition of L-arabinose to this system induces expression of the target protein. In the lab Dr Linda Hughes had previously shown the optimum temperature for protein induction was 37°C and the optimum concentration of the inducer molecule L- arabinose was 0.2% w/v. To examine the optimum time to collect the induced protein, 1ml was collected from induced and uninduced bacterial cultures at 2, 4 and 6 hours post induction, as per section 2.1.7 of Chapter 2. The samples were lysed in 4X SDS PAGE loading dye and analysed by SDS PAGE. It was found that the quantity of the target DHFRL1 and DHFR fusion proteins increased over time with its expression being highest 6 hours post induction, see Figures 4.4 and 4.5.

To purify the target DHFRL1 and DHFR his tagged proteins a His-TALON column containing immobilised cobalt was used. Prior to purification the bacterial cell pellets were weighed and the HisTALON xTractor Buffer was added in a ratio of 2mls of buffer per 100mg of bacteria and cells were lysed, as per section 2.1.8 of Chapter 2. Purification of the DHFR protein was successful, with the largest amount of protein being eluted in fraction 2. In contrast purification of the DHFRL1 protein yielded a small quantity of impure protein. Figure 4.6 shows that there is a number of contaminating proteins bands in the eluted fractions of the DHFRL1 protein. What is also of note is that, despite the fact the induction of the DHFRL1 protein appeared to be quite strong as analysed by whole cell lysate, there appears to be very little target DHFRL1 in the soluble protein fraction. Western blot analysis with an antiDHFR antibody confirmed that the purified protein fractions were DHFR and DHFRL1, see Figure 4.7 and 4.8. The presence of contaminating bands in the elution fractions of the DHFRL1 protein is likely due to the fact that there was very little DHFRL1 protein in the soluble fraction; this may have allowed native *E.coli* proteins to non-specifically bind to the column. Repeating the experiment showed that the

DHFR protein could be successfully purified by affinity chromatography whereas purification of the DHFRL1 protein even under increasing salt concentrations could not yield a clean purification. It became evident that the insolubility of the DHFRL1 recombinant protein was resulting in too little soluble protein being available for purification. Evidence pointing towards solubility issues came from the fact that good protein induction was observed in the whole cell lysate but very little protein was observed in the soluble protein fractions. DHFR and DHFRL1 only differ from each other by 15 amino acids. It was hypothesised that one or a number of the different amino acids contributed to the poor solubility of the DHFRL1 protein.

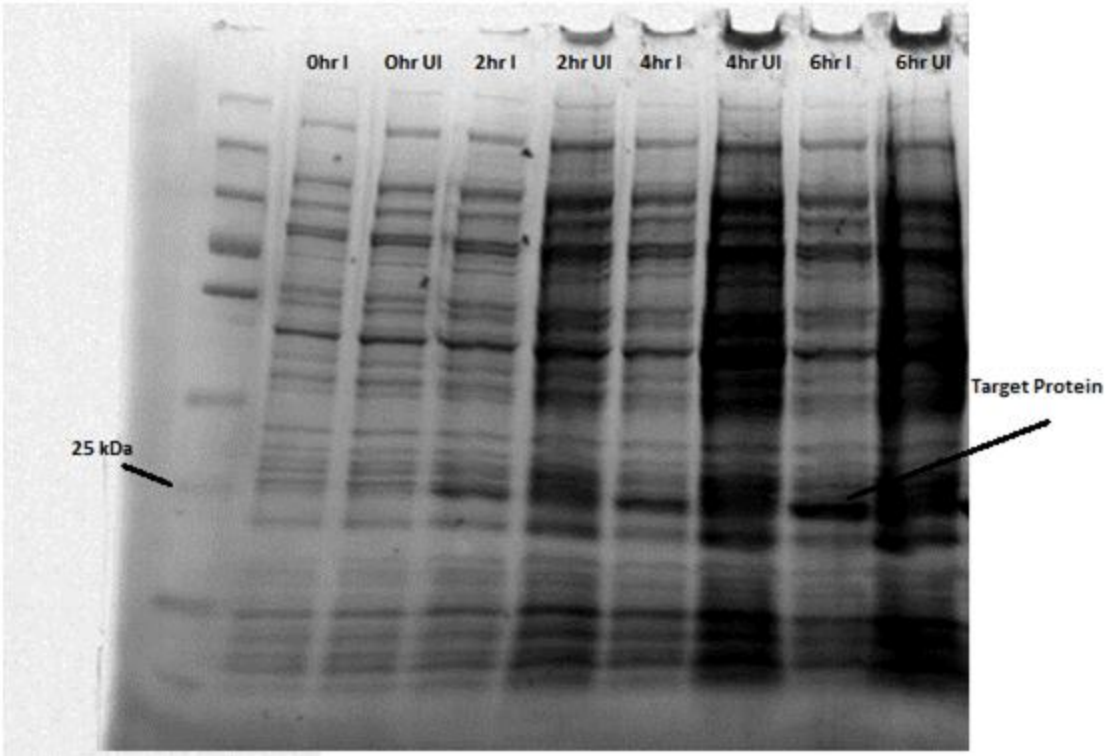


Figure 4.4 Time points of Induction of DHFRL1 HIS protein analysed by SDS PAGE and Stained in coomassie blue. Whole bacterial cell lysate was collected at 0, 2, 4 and 6 hours post induction in order to ascertain the optimum protein collection time point. Lane 1 contains the protein ladder. Lanes 2 and 3 contain DHFRL1 protein at 0 hours induced and un-induced. Lanes 4 and 5 contain DHFRL1 protein induced and un-induced at 2 hours. Lanes 6 and 7 contain DHFRL1 protein induced and un-induced at 4 hours. Lanes 8 and 9 contain DHFRL1 protein induced and un-induced at 6 hours. It can be seen that the largest amount of the target DHFRL1 protein is present at the 6 hour time point.

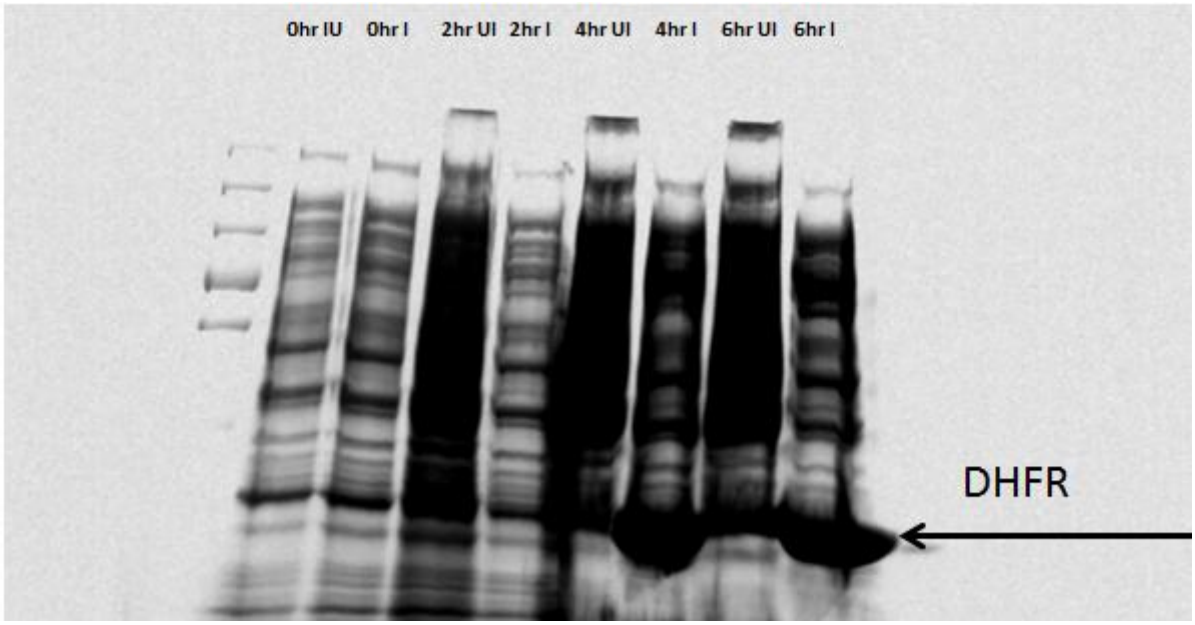


Figure 4.5 Time points of Induction of DHFR protein analysed by SDS PAGE. Whole bacterial cell lysate was collected at 0, 2, 4 and 6 hours post induction in order to ascertain the optimum protein collection time point. Lane 1 contains the protein ladder. Lanes 2 and 3 contain DHFR protein at 0 hours un-induced and induced. Lanes 4 and 5 contain DHFR protein un-induced and induced at 2 hours. Lanes 6 and 7 contain DHFR protein un-induced and induced at 4 hours. Lanes 8 and 9 contain DHFR protein un-induced and induced at 6 hours. It can be seen that the largest amount of the target DHFR protein is present at the 6 hour time point.

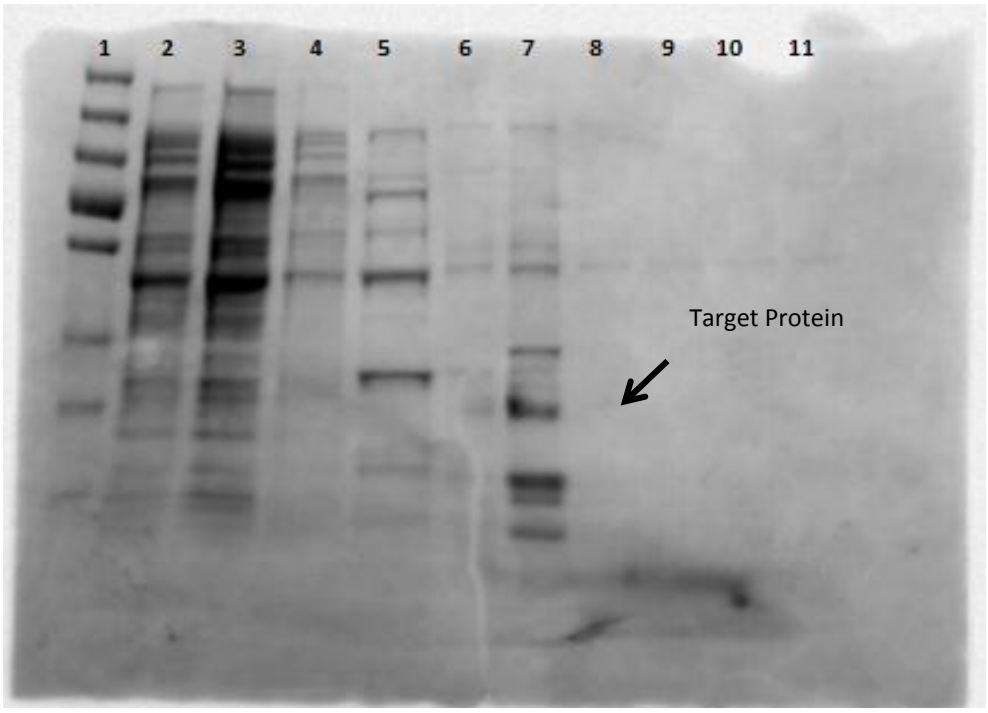


Figure 4.6 His-Talon Purification of DHFRL1 analysed by SDS PAGE. The SDS PAGE gel was stained with coomassie blue and visualised under white light. Lane 1 contains a protein ladder. Lane 2 contains the un-purified soluble protein fraction; little to no induced protein band can be seen in this fraction. Lane 3 contains the flow through fraction. Lanes 4 and 5 contain wash 1 and 2 fractions. Lanes 6 to 11 contain elution fractions 1-6 respectively. It appears that the purification process resulted in the enrichment of the target DHFRL1 His fusion protein as seen in lane 7. Although the target protein appears to have been enriched, it still remains quite impure.

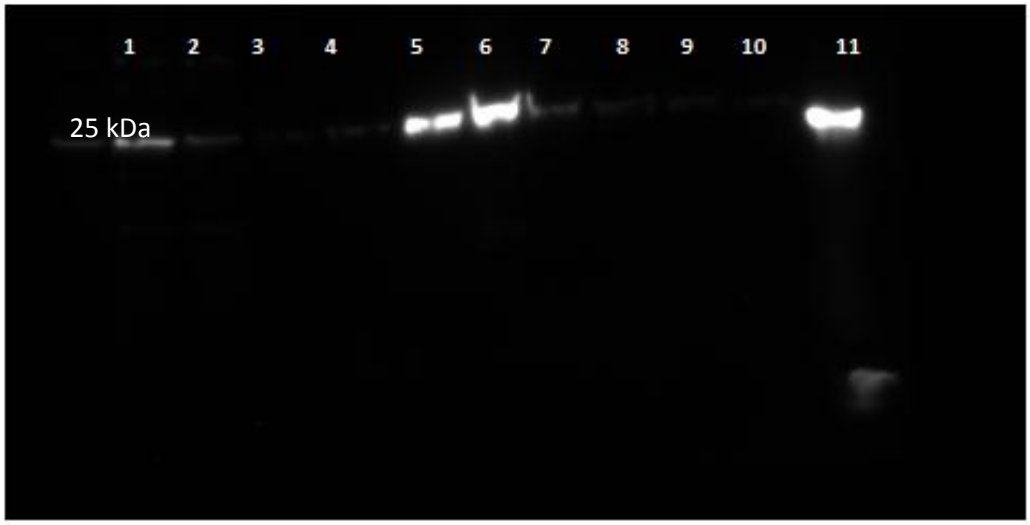


Figure 4.7 Western Blot analysis of Purified DHFRL1 with an antiDHFR antibody. Lane 1 contains the soluble induced protein fraction; a faint band can be observed, indicating there was very little target protein in the soluble fraction. Lanes 2, 3 and 4 contain the wash step fractions, lanes 5-10 contain elution fractions 1-6. Lane 11 contains the Sigma DHFR kit enzyme as a positive control. Consistent with SDS PAGE analysis, the majority of the target protein is present in elution fraction 2 where it appears to have been enriched as seen in lane 6.

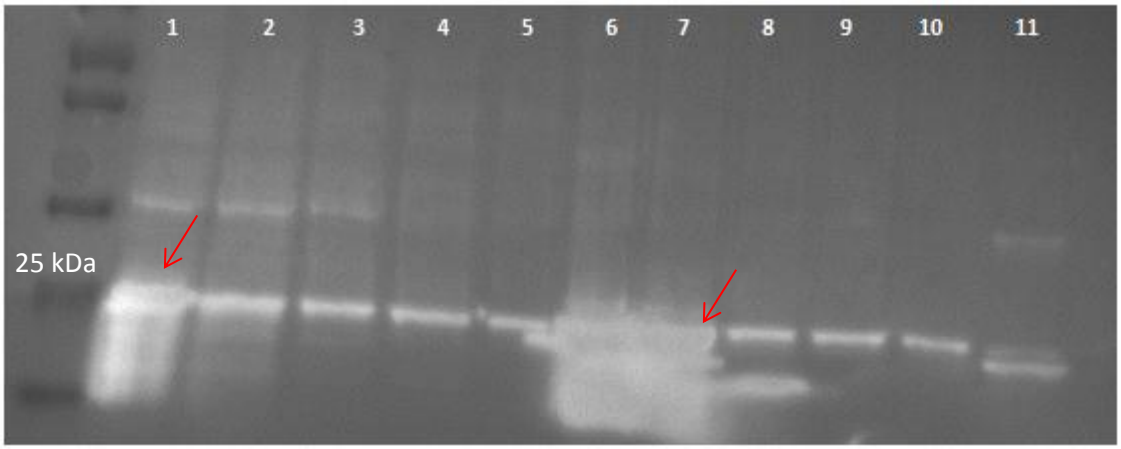


Figure 4.8 Western Blot analysis of Purified DHFR with an antiDHFR antibody. Lane 1 contains the soluble induced protein fraction; it can be seen that there is a large abundance of the DHFR protein in the soluble fraction. Lanes 2, 3 and 4 contain the wash step fractions, lanes 5-10 contain elution fractions 1-6. It can be seen that the majority of purified protein is in lanes 6 and 7, which contain elution fractions 2 and 3. Lane 11 contains the Sigma DHFR kit enzyme as a positive control..

4.3.3 Site Directed Mutagenesis of DHFR and DHFRL1

The R24W and the W24R mutations were inserted into the DHFRL1 and the DHFR genes, respectively by site directed mutagenesis as per section 2.1.6 of Chapter 2. The pDEST17 plasmids encoding the newly mutated DHFR and DHFRL1 genes were sent for Sanger sequencing to confirm that the mutations were correctly inserted. A Clustal Omega alignment was performed on the sequencing results against the DHFRL1 and the DHFR sequences to confirm the mutations were inserted in the correct positions; see appendices C1 and D1. The DNA sequencing results for the DHFRL1 R24W and the DHFR W24R were converted to the amino acid sequences. A Clustal Omega was performed which confirmed that the mutations in the DNA sequences resulted in the desired R24W and W24R changes in the amino acid sequence; see appendices C2 and D2.

4.3.4 Optimisation of DHFRL1 R24W and DHFR W24R Expression

The DHFRL1 R24W and the DHFR W24R pDEST 17 vectors were transformed into the BL21 AI cells for protein expression as per section 2.1.4 of Chapter 2. Expression was induced at 37°C with 0.2% w/v L-arabinose. A 1ml volume of culture was taken from the induced and un-induced cultures at 2, 4 and 6 hours to confirm that the optimum time point for protein collection was 6 hours, as with the wild type DHFRL1 and DHFR. The bacterial cell pellet was lysed in 4X SDS PAGE loading dye and analysed by SDS PAGE. It can be seen in Figure 4.9 that the amount of target protein increased over time with the largest amount of protein seen at 6 hours post induction, as was seen for the wild type DHFRL1 and DHFR proteins.

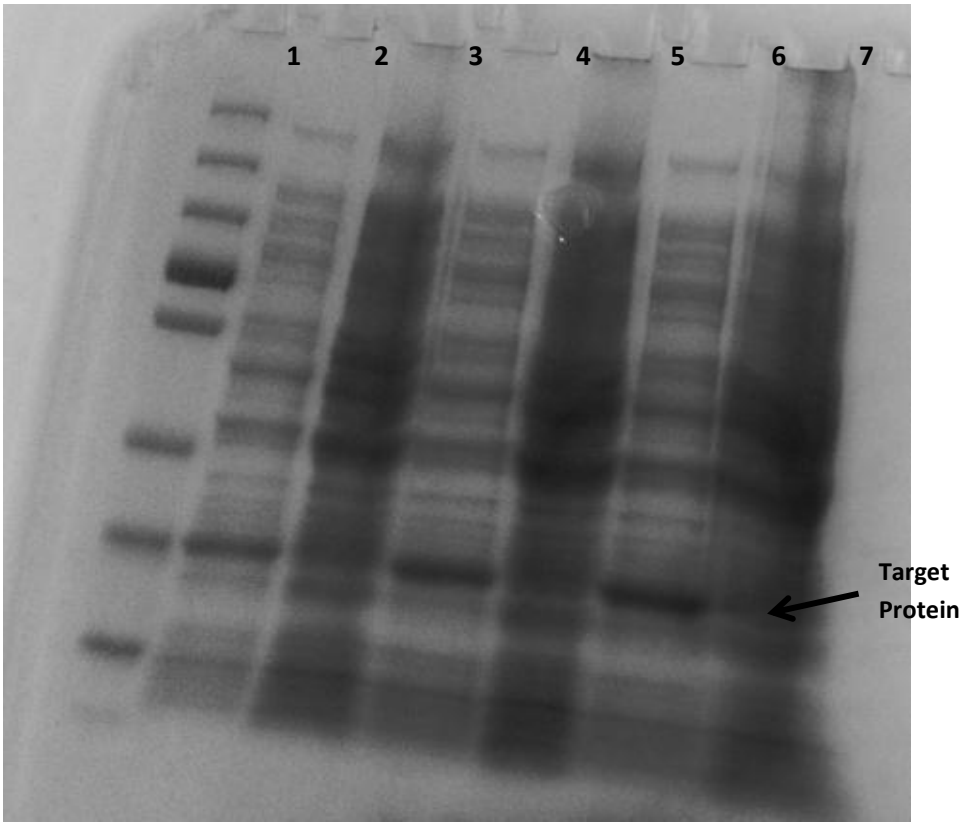


Figure 4.9 Induction of DHFRL1 R24W at 2, 4, 6 hours analysed by SDS PAGE and stained with coomassie blue. Whole bacterial cell lysate was collected at 2, 4 and 6 hours post induction in order to confirm the optimum protein collection time point. Lane 1 contains a protein ladder. Lanes 2 and 3 contain the induced and un-induced samples at 2 hours. Lanes 4 and 5 contain the induced and un-induced samples at 4 hours. Lanes 6 and 7 contain the induced and un-induced samples at 6 hours. The target DHFRL1 R24W protein appears to increase over time with it being most abundant 6 hours post induction.

4.3.5 Induction of Wild type DHFRL1, DHFR and Mutants

Induction of the wild type (WT) proteins DHFRL1 and DHFR and the mutant DHFRL1 R24W and DHFR W24R was undertaken with a view to assess if introducing the mutations had any effect on protein solubility. The target proteins were induced for 6 hours at 37°C with 0.2% w/v L-arabinose in the BL21 AI cells. Cells were fractionated into soluble and insoluble protein fractions in B-per complete reagent (4mls/gram of bacteria) combined with freeze thaw and sonication, as per section 2.1.7 of Chapter 2. SDS PAGE gels were equally loaded with 18 µl of protein lysate (lysed in B-per at 4ml/gram of bacteria) and 6µl 4X sample buffer. SDS PAGE and western blot analysis showed that the amino acid mutations at position 24 had an effect on the DHFR enzyme but not on DHFRL1. Figure 4.10 shows that a strong protein band is present in the soluble fraction at the correct size for DHFR, as was seen previously. The protein bands present in the soluble fractions for the DHFR W24R, DHFRL1 WT and the DHFRL1 R24W are much weaker. The DHFRL1 WT and R24W mutant have very strong induction bands that are apparent in the insoluble protein fractions, confirming previous hypothesis that solubility of the DHFRL1 protein was the reason for its poor purification. Western blot analysis with an anti-His antibody confirmed the SDS PAGE results. Due to the fact that there was such a large abundance of the target proteins, particularly for the insoluble fractions the PDVF membrane actually became scorched as soon as the substrate was added meaning that a luminescent image was not captured in time. However, white light images of the scorched PVDF membranes were captured, which confirm the SDS PAGE results, see Figures 4.11 and 4.12. The protein solubility results provide further evidence as to the importance of the 24th amino acid residue of the DHFR and DHFRL1 proteins.

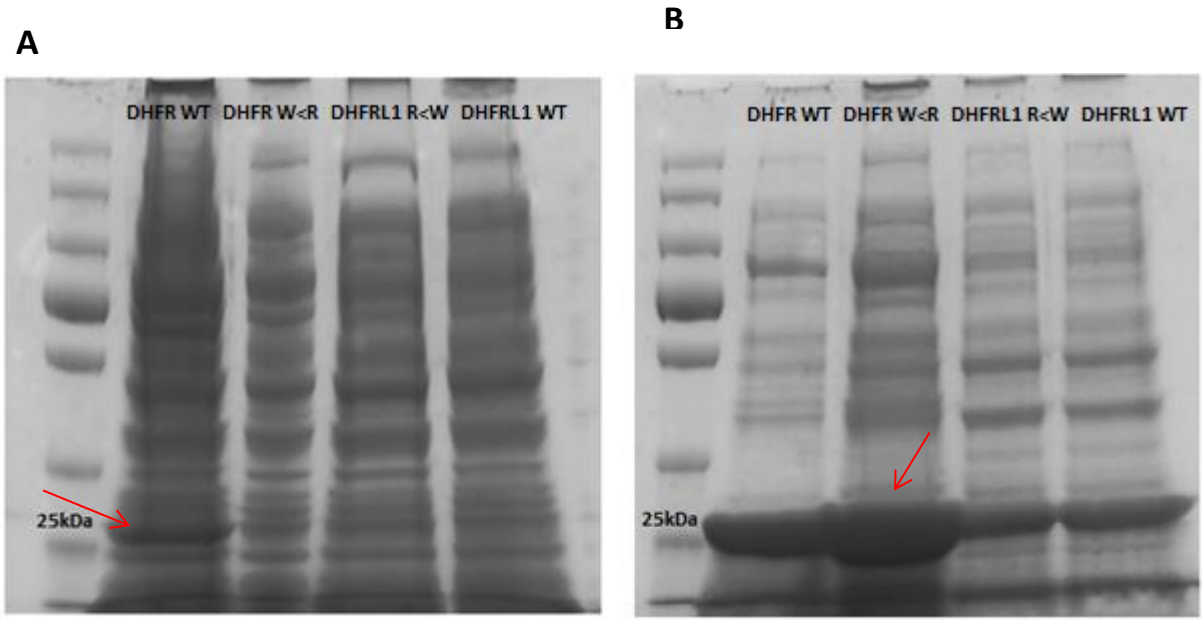


Figure 4.10. SDS PAGE analysis of DHFR, DHFRL1 and mutants soluble and insoluble fractions. A. Contains the soluble protein fractions. A strong induction band is present at 25 kDa for the wild type (WT) DHFR protein. Little to no induction bands can be seen for the DHFR W24R, DHFRL1 R24W and the DHFRL1 WT in the soluble protein fraction. **B.** represents the insoluble protein fractions. The DHFR W24R mutation resulted in little to no soluble induced protein but a large abundance of induced target protein in the insoluble fraction. Although there is still a presence of insoluble protein in the DHFR WT a strong induction band still persists in the soluble fraction. The induced proteins for DHFRL1 WT and R24W mutants appears to be all in the insoluble fractions.

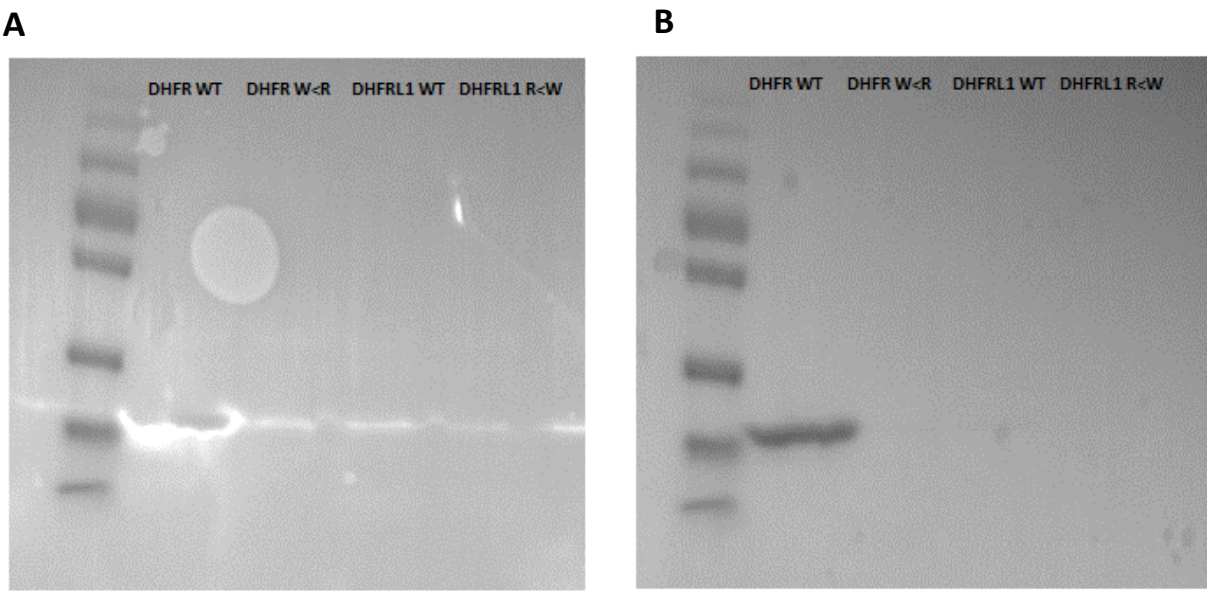


Figure 4.11 Western blot analysis of the soluble protein fractions of DHFR, DHFRL1 and mutants. Western blot analysis confirmed that the induced band at 25 kDa was indeed the DHFR WT protein. **A**. It can be seen that the PVDF membrane is being scorched due to the abundance of target protein as indicated by the apparent "hole" in the DHFR WT band. **B.** Imaging of the membrane under white light visualises the extent of staining and confirms that there is a large abundance of the DHFR protein in the soluble fraction.

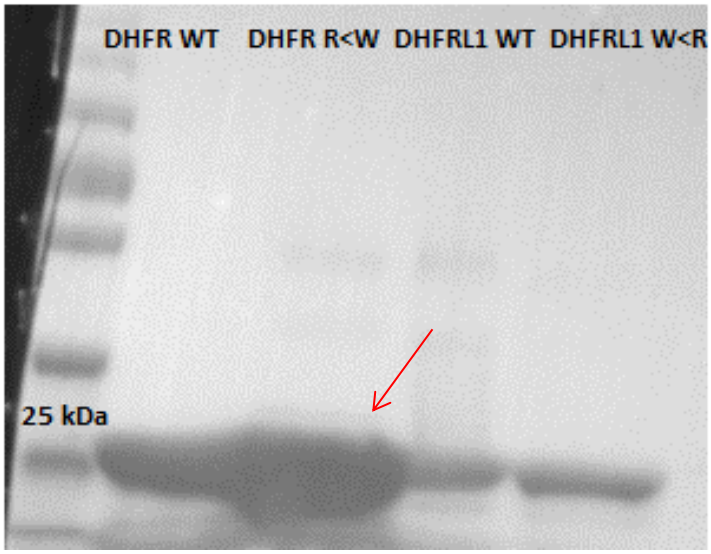


Figure 4.12 Western blot analysis of the insoluble protein fractions of DHFR, DHFRL1 and mutants. Unfortunately due to the large abundance of target His tagged proteins the PVDF membrane became scorched almost immediately such that the luminescence could not be visualised. Despite this, white light images of the PVDF membrane confirmed the SDS PAGE results. A large abundance of DHFR W24R protein can be seen in its corresponding labelled lane. The induced proteins for DHFRL1 WT and R24W mutants appears to be all in the insoluble fractions. A portion of the DHFR WT protein appears to be in the insoluble protein fraction; however a large portion of soluble DHFR is present as seen in previous images.

4.4 Discussion

The conversion of folate to its biologically active form tetrahydrofolate (THF) is one of the most important biochemical reactions carried out within the cell. The metabolism of folate results in the production of DNA and methyl donor groups for cellular methylation reactions (Fox and Stover 2008). Such is the importance of folate metabolism that it is targeted with anti-folate chemotherapy drugs in the treatment of cancer (Miller 2006). The enzyme DHFRL1 is found to be a less active version of the DHFR enzyme, a systematic site directed mutagenesis approach is needed to decipher which amino acids contribute to the differences in enzyme activity (McEntee *et al.* 2011). If the low level of activity seen for DHFRL1 is physiologically relevant then the enzyme could potentially be a chemotherapeutic target by the use of anti-folate drugs in the treatment of cancer in much the same way as DHFR. The DHFR and DHFRL1 enzymes are 92% identical, despite their similarity DHFRL1 has been shown to have a much reduced affinity for DHF (McEntee *et al*. 2011). This suggests that DHFRL1 may also have a reduced affinity for anti-folate chemotherapy drugs such as methotrexate.

In order to identify key residues that may be responsible for the differences between the DHFR and DHFRL1 enzymes an amino acid alignment was performed with DHFRL1 and DHFR from 6 species. The Clustal Omega amino acid alignment revealed that the tryptophan at position 24 of the DHFR enzyme was highly conserved in all species tested (human, rat, mouse, chicken, cow and chimpanzee) but the DHFRL1 enzyme had an arginine in the same position. The highly conserved nature of the tryptophan suggests that it is integral to the function of DHFR in its binding and reduction of DHF to THF.

Predictive protein modelling indicated that the DHFRL1 enzyme did not have its NADPH (a co-factor for the reduction of folate) binding site conserved. Predictive modelling alone is not enough to confirm that this is the case. However, the predictive modelling revealed that changing the tryptophan (W) at position 24 of the DHFR enzyme to an arginine (as found in DHFRL1) would result in the loss of the enzymes NADPH binding site. Interestingly the predictive modelling found that

changing the arginine to a tryptophan in the DHFRL1 enzyme failed to restore its NADPH binding site, indicating that further amino differences between the enzymes also contribute to its loss. This information was viewed in regard to the importance of the position of the 24th amino acid and its function in the structural conformation of the enzyme.

In order to further investigate the differences between the DHFR and DHFRL1 enzymes pDEST 17 plasmids encoding the genes were transformed into bacterial BL 21 AI cells. It was ascertained that the optimum conditions for expression were at 37°C, with 0.2% w/v L-arabinose and that the optimum time for protein collection was 6 hours post induction. Expression of both the DHFR and DHFRL1 recombinant proteins was confirmed in whole cell lysate by SDS PAGE. Purification of the His tagged DHFR enzyme using His-Talon cobalt columns was achieved successfully; however purification of the DHFRL1 enzyme yielded only a small quantity of impure enzyme, despite good protein expression in the whole bacterial cell lysate samples. Evidence suggested that the insolubility of the DHFRL1 recombinant protein was resulting in too little soluble protein being available for purification.

Site directed mutagenesis was performed and a DHFRL1 mutant with a tryptophan (W) at position 24 was created and a DHFR mutant with an arginine (R) at position 24 was also created. Expression of all four enzymes, DHFR WT, DHFRL1 WT, DHFR W24R and DHFRL1 R24W was successfully undertaken. Solubility of the recombinant proteins was used as a measure of protein structure, as the solubility of recombinant proteins produced in *E.coli* has previously been shown to depend on the structure of the protein. (Idicula-Thomas and Balaji 2005). Separating the induced bacterial cells into soluble and insoluble protein fractions confirmed that the DHFRL1 enzyme was predominantly insoluble, thus providing an explanation as to why expression in the whole bacterial cell lysate was high but the enzyme failed to purify from the soluble fraction. It was also found that introducing the R24W mutation alone had no effect on protein solubility as the mutant was just as insoluble as the wild type. The protein modelling software showed a single amino acid change would not restore the NADPH binding site and these data show that a single amino acid change cannot make the protein more soluble. Interestingly,

mutating the tryptophan to an arginine in the DHFR enzyme resulted in the enzyme becoming predominantly insoluble. The protein modelling software predicted that this single amino acid change could result in the loss of DHFR's NADPH binding site, so in this instance a single amino acid change could have a big impact. Since not enough expressed recombinant protein could be purified for kinetics analysis, the NADPH binding site was not experimentally assessed in this instance but the results do confirm that an effect was caused by the mutation. The wild type DHFR enzyme had a strongly expressed protein band in the soluble protein fraction and, therefore purifies well. Although it can be clearly seen that some of the DHFR protein remains insoluble enough of the protein is soluble for purification. Unfortunately there is not enough soluble protein available for adequate purification of DHFRL1 or the mutant DHFRL1 R24W and DHFR W24R. Both predictive modelling and solubility results indicate that the amino acid at position 24 may have an effect on enzyme structure.

Predictive binding calculations by Gao *et al* (2013) found that the binding modes of DHFRL1 and DHFR to its substrate DHF are different. Arginine is quite a long positively charged amino acid and is thought to extend deeply into the binding pocket of the DHFRL1 enzyme (Gao *et al.* 2013). Predictive modelling by Gao *et al* (2013) indicates that the pterin ring of DHF forms a hydrogen bond with arginine at position 24 of the DHFRL1 protein and that this residue also forms two hydrogen bonds with NADPH. Modelling by Gao *et al* (2013) indicates that the R24W mutation in the DHFRL1 enzyme would result in the mutant binding and interacting with DHF in much the same way as DHFR. The data presented in this chapter indicates that the DHFRL1 R24W mutant does not behave any more like DHFR in terms of its solubility than the wild type DHFRL1 enzyme indicating that further amino acids may also play a role in the solubility and structure of the enzyme.

Expression of recombinant proteins in an *E.coli* system is a relatively cheap and convenient method of protein production in comparison to a mammalian system. However, it does have its drawbacks such as the inability to perform post translational modifications and specific folding of proteins. As stated previously, solubility was used a measure of protein structure as it has previously been shown that the solubility of a recombinant protein in an *E.coli* system can depend on its

structure (Idicula-Thomas and Balaji 2005). However, solubility of a recombinant protein may also be affected by other issues such as toxicity induced by heterologous protein production. The host bacteria can sequester the protein in insoluble inclusion bodies to minimise the toxic effects (Palomares *et al,* 2004, Saida 2007).Perhaps the DHFRL1 protein is somehow toxic to the *E.coli* cells, causing it to be insoluble. In addition to this, high expression of non-native proteins such as in the case of producing the human enzymes DHFR and DHFRL1 may saturate foldases within bacteria, causing misfolded insoluble proteins (Palomares *et al.* 2004). However, it appears that this is not the case in this instance as the expression level of the DHFRL1 and the DHFRL1 R24W mutant enzymes appear to be lower than that of DHFR. The data suggests that amino acid differences between DHFR and DHFRL1 cause the enzymes to have different solubilities and therefore different structural conformations. This may explain why the BL 21 AI cells can produce soluble DHFR protein but not DHFRL1 protein. It is possible that the amino acid differences between DHFR and DHFRL1 result in the bacteria being unable to properly fold the DHFRL1 enzyme resulting in it being insoluble. Furthermore, mutating the amino acid at position 24 of the DHFR enzyme may cause similar structural changes as observed in DHFRL1, resulting in the enzyme being insoluble.

These results give an insight into the significance of the amino acid differences between DHFR and DHFRL1. The predictive protein modelling results together with the solubility results all point towards DHFR and DHFRL1 having different structural conformations. It appears that the 24^{th} amino acid residue is of particular importance in the differences observed between the enzymes. Previous experimental work by Beard *et al* (1991) found that mutating the tryptophan at position 24 of the DHFR enzyme to a phenylalanine causes a dramatic 25 fold reduction in affinity for the substrate DHF and over a 50% decrease in enzyme stability (Beard *et al.* 1991). DHFRL1 is a potential chemotherapeutic target in the treatment of cancer; however the reduced affinity of the enzyme for the substrate also suggests that it may also have a different affinity to the widely used anti-folate methotrexate. Indeed work in mouse DHFR demonstrated that mutating the tryptophan at position 24 to an arginine resulted in the enzyme having thousands

fold reduced affinity for methotrexate. It is thought that the tryptophan at position 24 forms a hydrogen bond with methotrexate and switching this amino acid to an arginine disrupts this binding (Thillet *et al*. 1988). Trimethoprim is an anti-folate agent against bacterial DHFR which can be used to treat urinary tract infections (Brogden *et al.* 1982). Analysis of the inhibition of the anti-folate trimethoprim on the mouse DHFR W24R mutation also found that the enzyme had a 15 fold reduced affinity for the inhibitor than the wild type enzyme (Thillet *et al.* 1988). Although the affinity of the DHFR W24R mutation for trimethoprim is still lower than that of the wild type, it is much less dramatic than the reduced affinity observed by the mutation for methotrexate. Work by Gao *et al* (2013) suggests that DHFRL1's conformation may be closer to bacterial DHFR than to human DHFR. Perhaps alternative anti-folates that are known to have a much reduced affinity for human DHFR, such as anti-protozoals could be explored in the targeting of DHFRL1. Data presented in this research thesis provides further evidence as to the possible structural impact that the amino acid at position 24 of DHFR and DHFRL1 has on the enzymes. This information taken collectively suggests that alternative anti-folate drugs to methotrexate would need to be explored for the targeted treatment of DHFRL1 in cancer. The ideal scenario would be a two pronged approach with the targeted treatment of DHFR with methotrexate and an alternative drug for targeting DHFRL1.

To conclude, it appears that amino acid differences between DHFR and DHFRL1 not only affect the binding and affinity of the enzymes for their substrate DHF, but may also impact on the folding and conformation of both enzymes. It appears that the 24th residue of the DHFR and DHFRL1 enzymes may be a key player as to the differences observed between the enzymes. DHFRL1 may be a potential chemotherapeutic target in the treatment of cancer, although its biological significance does need to be further investigated. Evidence suggests that DHFRL1 would respond to different anti-folate inhibitors than DHFR potentially allowing for a combinatorial treatment to target DHFR and DHFRL1. Further investigation is warranted into the biological significance of the DHFRL1 enzyme and evaluation of the enzyme as a chemotherapeutic target.

Chapter 5

Folate Enzyme Expression Analysis in a Colorectal Cancer Metastatic Cell Line Model

5.0 Introduction

5.1 Gene Expression and Cancer

Regulation and control of gene expression is fundamental to maintain health within the cell. Cancer is the unregulated and uncontrolled growth of cells (Sher 1996). Cancer may arise through a number of mechanisms all of which are essentially caused and or result in aberrant gene expression. As previously discussed, recent evidence suggests that mitochondrial folate enzymes may play a significant role in cancer growth and development. Although the link between cancer and folate enzymes such as DHFR has been widely known and accepted for many years, the expression of the complex folate pathway and its pathological significance in regard to cancer appears to be coming to the forefront among the scientific community (Jain 2012, Selcuklu 2012, Nilsson 2014).

Chapter 3 has shown that both knocking down and over expressing the mitochondrial gene MTHFD1L impacted cell growth and formate production. Publications by Jain *et al* (2012) Sugiura *et al* (2004) and Fashfindar *et* al (2012) have associated MTHFD1L with increased cancer cell proliferation, mortality and metastasis. Jain *et al* (2012) also significantly associated mitochondrial MTHFD2 and SHMT2 with cancer cell proliferation but not their paralogous cytoplasmic counterparts. Nilsson *et al* (2014) also found a significant link between the expression level of the MTHFD2 gene and increased mortality in breast cancer patients.

Recent evidence appears to be in favour of a folate mitochondrial driven role in cancer growth and development. In order to further explore the role of folate enzymes in cancer, the expression level of a range of mitochondrial and cytoplasmic folate enzymes was analysed in the isogenic colon cancer cell line model SW480 and SW620, see Figure 5.1. The mitochondrial enzymes MTHFD1L, SHMT2 and MTHFD2 and the cytoplasmic paralogues MTHFD1 and SHMT1 have been selected due to their recent associations with cancer (Jain *et al.* 2012, Nilsson *et al.* 2014). The mitochondrial enzyme DHFRL1 has been selected for expression analysis along with the cytoplasmic DHFR. Unlike DHFR, very little is known about the expression level

of the novel enzyme DHFRL1 in regard to cancer. The genes TYMS, ATIC and ALDH51A have also been selected for analysis as they were previously identified by Dr Stefano Minguzzi as being significantly affected by altered MTHFD1L expression (Minguzzi 2013).

The SW480 cells are a primary colon cancer cell line; the SW620 cells were derived from a lymph node metastasis in the same individual. The derivation and isogenic nature of the SW480 and SW620 cell lines makes them an excellent model for studying genetic differences that arise due to tumour progression and metastasis (Leibovitz *et al.* 1976).

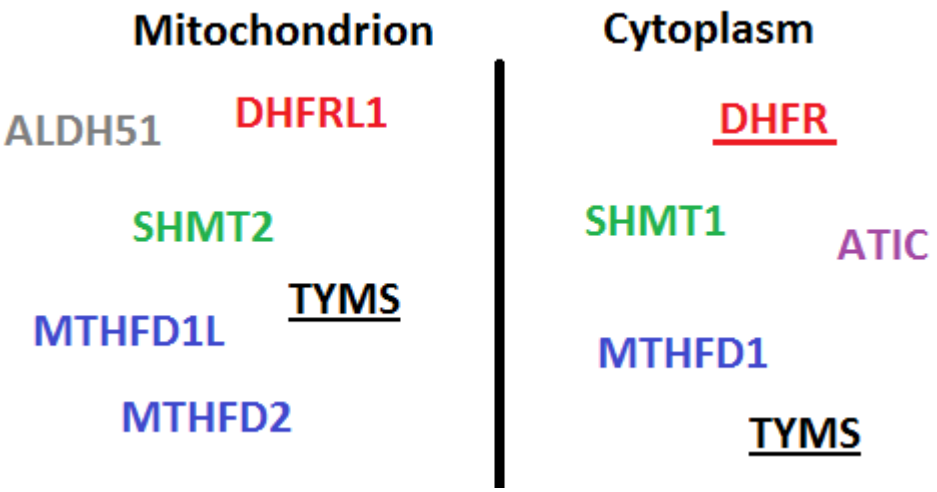


Figure 5.1 Folate Enzymes Selected for expression analysis in the SW480 and SW620 cell lines. Genes depicted in the same colour are paralogous. Genes which are underlined are known to localise to other parts of the cell e.g. TYMS also localises to the nucleus.

5.2 Aims and Objectives

The main aim of this chapter was to assess if folate related genes are differentially expressed in a primary and metastatic colon cancer cell line. This aim was achieved by the following objectives.

- Extraction of RNA from the SW480 and SW620 cell lines. Ensuring it was free from genomic DNA by DNAse treating and performing a DNA contamination assay.
- Selecting an appropriate reference gene for RT-qPCR analysis between the SW480 and SW620 cell lines.
- Performing gene expression analysis by RT-qPCR on the SW480 and SW620 samples for the following genes, MTHFD1L, MTHFD2, MTHFD1, DHFRL1, DHFR, SHMT2, SHMT1, ATIC and ALDH51.

5.3 Results

5.3.1. cDNA Synthesis and the DNA Contamination Assay

RNA was successfully extracted from the SW480 and SW620 cells and cDNA was synthesised as per section 2.3.1 of Chapter 2. The presence of genomic DNA can potentially interfere with subsequent RT-qPCR analysis of gene expression. In order to ensure that synthesised cDNA was free from genomic DNA, the MTHFD1 R653Q genomic DNA contamination assay was performed, as per section 2.3.2 of Chapter 2. This assay is an intron spanning assay allowing for the detection of genomic DNA. The expected size of the PCR products of cDNA free from genomic DNA contamination is 232bp. The expected size of the PCR products containing genomic DNA is 330 bp. Contaminated cDNA would be represented by two PCR bands and 232 bp and 330bp. It can be seen from Figure 5.2 that the synthesised cDNA is free from genomic DNA contamination.

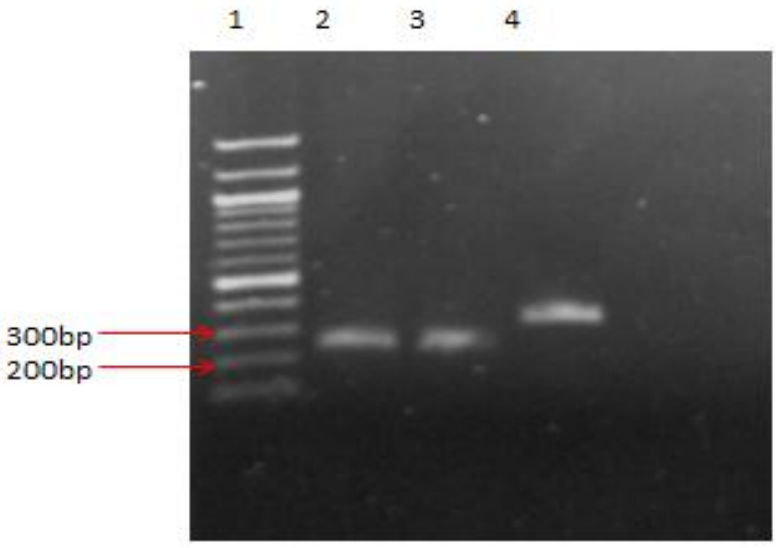


Figure 5.2 Genomic DNA Contamination Assay. 1. 100bp ladder 2. SW480 PCR product 3. SW620 PCR product 4. Positive genomic DNA control 5. Negative control. It can be seen that cDNA is free from genomic DNA contamination as PCR products in lanes 1 and 2 are at 232 bp. The PCR product for genomic DNA can be seen at 330bp.

5.3.2 Reference Gene Selection for RT-qPCR

In order to identify the appropriate reference gene for expression analysis of the SW480 and SW620 cell lines RT-qPCR was undertaken with the following genes, GUS, GAPDH, RPS 13 and TBT. The relative ratio is calculated by simply dividing one CP value over another for a given gene. GUS was found to be the most appropriate reference gene for the SW480 and SW620 cell lines with a relative ratio of 0.99. RPS would also be a good candidate as control gene for the cell lines with a relative ratio of 0.98. The relative ratios for the GAPDH and TBT were 0.95 and 0.96 respectively, see Table 5.1.

Table 5.1 Reference Gene Selection for SW480 and SW620 cell lines

Gene	Mean CP of Triplicates	Standard Deviation	Cell Line	Relative Ratio
GUS	26.57	0.02	SW 480	
GUS	26.65	0.12	SW620	**0.99**
RPS 13	23.71	0.06	SW 480	
RPS 13	23.47	0.09	SW620	0.98
GAPDH	17.55	0.05	SW 480	
GAPDH	18.35	0.09	SW 620	0.95
TBT	35.45	0.14	SW 480	
TBT	36.74	0.06	SW 620	0.96

5.3.4 All target genes are differentially expressed in SW620 cells compared to SW480 cells

Gene expression analysis for the 10 selected genes was successfully carried out on the cDNA derived from the SW480 and SW620 colon cancer cell lines as per section 2.3.3 of Chapter 2, see Figure 5.3 and Table 5.2. The ATIC gene was found to have -1.420 fold lower expression in the SW620 cells compared to the SW480 cells. The expression of ALDH51A was 1.3 fold higher in the SW620 cells compared to the SW480 cells. The expression of the TYMS gene appeared to -2.5 fold lower in the SW620 cells compared to the SW480 cells. The expression of the cytoplasmic MTHFD1 gene appeared to have a -1.28 fold lower in the SW620 cells compared to the SW480 cells. Whereas, both the mitochondrial paralogues of MTHFD1, MTHFD1L and MTHFD2 were found to be up regulated in the SW620 cells relative to the SW480 cells, by 2.17 and 1.8 fold respectively. SHMT1 was found to have -2.32 fold lower expression in the SW620 cells, while its paralogous mitochondrial counterpart SHMT2 was found to have 2.98 fold higher expression when compared to the SW480 cell. The cytoplasmic DHFR enzyme was found to have -1.85 fold lower expression in the SW620 cells, though its mitochondrial paralogue DHFRL1 was found to have a 3.36 fold higher expression when compared to the SW480 cells.

Table 5.2 Fold Expression of genes and their cellular location in the SW620 cell line compared to the SW480 cell line.

Gene	Fold Expression of SW620 Cells compared to SW480	Gene Location
SHMT2	2.98	Mitochondria
MTHFD2	1.8	Mitochondria
MTHFD1L	2.17	Mitochondria
DHFRL1	3.36	Mitochondria
ALDH	1.3	Mitochondria
TYMS	-2.5	Mito/Nucleus
DHFR	-1.85	Cytoplasm/Nucleus
SHMT1	-2.32	Cytoplasm
MTHFD1	-1.28	Cytoplasm
ATIC	-1.42	Cytoplasm

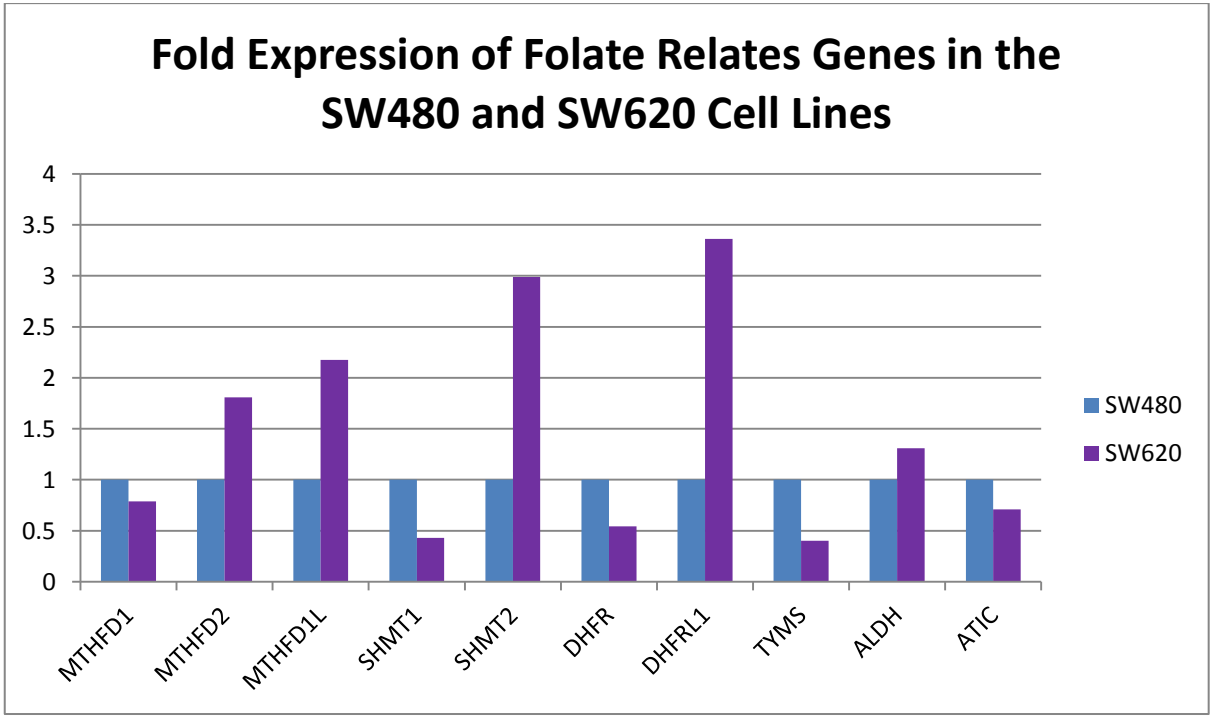


Figure 5.3 Fold Expression results in the SW620 cells compared to the SW480 cells. It can be seen that a trend towards higher expression of the mitochondrial genes is present in the SW620 cells when compared to the SW480 cells. Interestingly it can also be seen that the cytoplasmic genes within the SW620 cells appear to have lower expression when compared to the SW480 cells.

5.4 Discussion

The SW480 and SW620 isogenic colon cancer cell line model allows for the analysis of changes between a primary tumour and a secondary metastatic growth invitro (Leibovitz *et al.* 1976). RNA was successfully extracted from the cell lines, quantified, DNAase treated and used to make cDNA. The MTHFD1 R653Q PCR DNA contamination assay confirmed that the SW480 and SW620 cDNA samples were free from genomic DNA which could have interfered with the relative quantitative gene expression analysis.

Gene expression analysis was carried out successfully for each of the 10 genes analysed. The data show an obvious trend towards the up regulation of mitochondrial folate genes in the SW620 metastatic cell line. MTHFD2, MTHFD1L, SHMT2, DHFRL1 and ALDH51 were found to be 1.80, 2.17, 2.98, 3.36 and 1.30 fold over expressed, respectively, compared to the SW480 cell line. The up regulation of mitochondrial folate genes in the SW620 metastatic cell line is in concordance with recent findings in the literature, where folate mitochondrial enzymes have been correlated with proliferation and metastasis (Jain *et al.* 2012, Nilsson *et al.* 2014). Interestingly, cytoplasmic enzymes appeared to be down regulated in the metastatic SW620 cells relative to the SW480. The fold reduced expression in the SW620 cells relative to the SW480 cells for the cytoplasmic genes MTHFD1, SHMT1, DHFR and ATIC was -1.28, -2.32, -1.85 and -1.42, respectively. There appears to be somewhat of an inverse relationship between the mitochondrial and cytoplasmic folate genes within the SW480 and SW620 cell lines; see Figure 5.4.

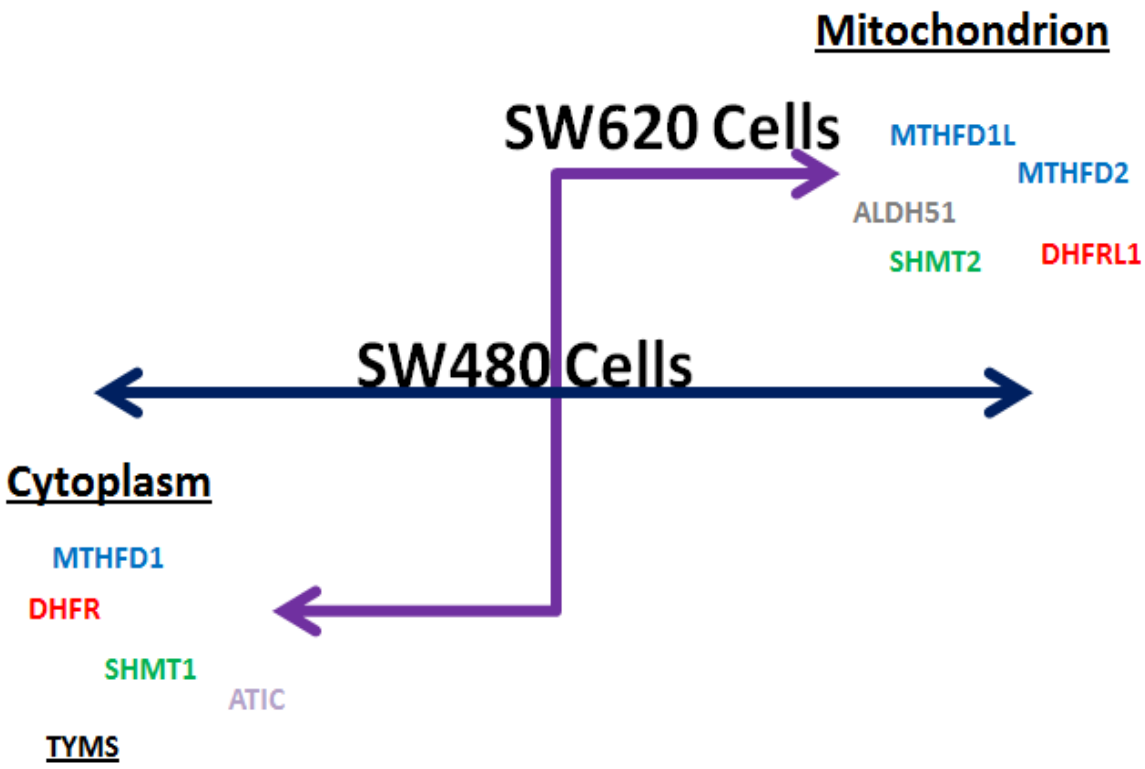


Figure 5.4 The expression patterns of mitochondrial and cytoplasmic folate related genes within the SW480 and SW620 cell lines. There appears to be an inverse relationship between the mitochondrial and cytoplasmic folate genes within the SW620 cells. As depicted, in the SW620 cells the mitochondrial genes appear to have an increased expression and the expression of cytoplasmic genes appear to have decreased relative to the SW480 cell line.

The origin or cause of the observed over expression of mitochondrial folate genes in the SW620 cells or indeed in other fast growing cancer cells is not an easy thing to distinguish. SHMT2 or serine hydroxymethyltransferase is a known target of the oncogene c-myc; nonetheless, Lee *et al* (2014) have data to suggest that it is inadequate for malignant transformation. However, work by Zhang *et al* (2012) identified that over expression of the enzyme glycine decarboxylase (GLDC) may be sufficient to initiate tumourigenesis in non-small cell lung cancer. Up regulation of GLDC was also found to result in up regulation of SHMT1 and SHMT2 (Zhang 2012). The SHMT enzymes (1 and 2) are responsible for the conversion or serine to glycine and the conversion of THF to CH_2-THF. CH_2-THF is then used in subsequent reactions by MTHFD2 and MTHFD1L in the mitochondria and MTHFD1 in the cytoplasm in the production of formate; see Figure 5.5. Formate is then transported into the cytoplasm where it is then used as a one carbon donor for cytoplasmic one carbon metabolism. It is evident that up regulation of the mitochondrial folate enzymes SMHT2, MTHFD2 and MTHFD1L and in turn an up regulation of formate would confer an advantage to the SW620 cells. With formate providing extra one carbon molecules allowing for increased nucleotide synthesis, this would facilitate increased proliferation in the SW620 metastatic cell line.

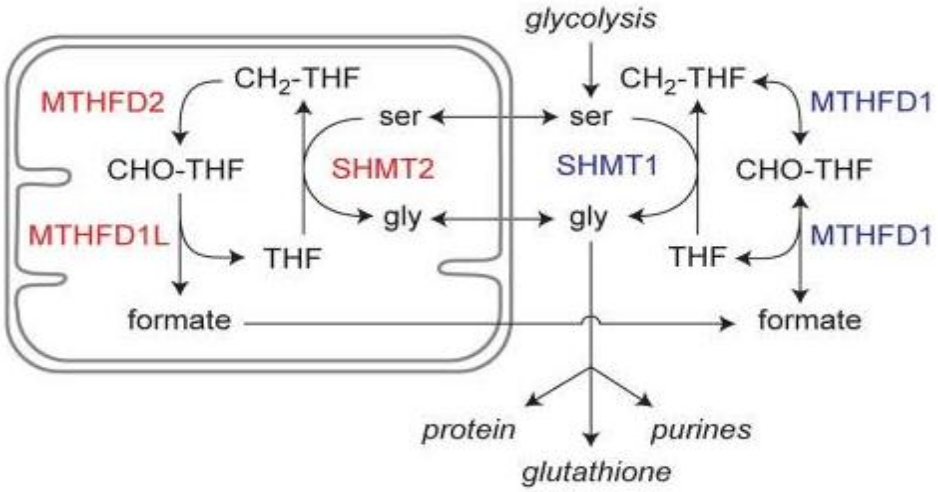


Figure 5.5 Glycine Biosynthesis and Formate production. Glycolysis provides serine which is cleaved by SHMT1 in the cytoplasm and SHMT2 in the mitochondria. Glycine is utilised by MTHFD2 and MTHFD1L in the mitochondria in the production of formate. (Image taken from Jain *et al* 2012)

Indeed, the notion of mitochondria playing a central role in cancer pathogenesis is not novel. The Warburg effect is a well-defined phenomenon in which cancer cells use glycolysis as opposed to oxidative respiration as a method of energy production (Warburg *et al.* 1927). The glycolytic pathway and the one carbon metabolism pathway are interlinked with one another (Amelio *et al*. 2014). One may be forgiven for hypothesising that the observed role of glycine production in cancer cells (as observed by Jain and colleagues 2012) may be a consequence of the Warburg effect within transformed cells. With glycolysis driving serine production which results in the production of glycine in the mitochondria (Amelio *et al.* 2014). Thus, the reliance of fast growing cancer cells on glycine may be a consequence of the switch in metabolism from oxidative respiration to glycolysis within the cell. Contrary to this hypothesis Zang *et al (*2012) have found that over expression of the mitochondrial enzyme glycine decarboxylase (GLDC) actually causes an up regulation of glycolytic enzymes and propose that over expression of GLDC actually promotes glycolysis and up regulation of glycine related enzymes SHMT1/2 in non-small cell lung cancer. In addition they found that increased GLDC expression results in an increase in pyrimidine production. Treatment with methotrexate in cells with overexpressed GLDC specifically prohibited proliferation but this did not occur in control cells. Most pertinent of all was that shRNA down regulation of GLDC in combination with methotrexate treatment was most effective at killing transformed cells (Zang *et al.* 2012).

Whether up regulation of the glycine pathway and its related enzymes is the answer as to why cancer cells switch their metabolism to glycolysis or are up regulated as consequence of the switch in energy production, remains to be further investigated. For cancer cells to grow and divide nucleotides are required, whence the long known link between folate metabolism which results in the production of nucleotides and classical chemotherapeutic agents (Visentin *et al.* 2012). It is logical to see how a switch by a cell to glycolysis resulting in increased serine which results in increased glycine, resulting in increased formate, ultimately leading to increased nucleotide synthesis is extremely advantageous to a cancer cell. Though the link between folate metabolism and cancer is well known, our understanding of the

relationship between the biochemical intermediates involved in one carbon metabolism and their significance in regard to cancer is novel, given recent evidence both in the literature and in this research thesis.

The metastatic SW620 cell line is known to be more proliferative than the parent SW480 cell line (Hewitt *et al.* 2000). The data in this chapter, in addition to correlating the gene signature identified by Jain *et al* (2012) of MTHFD1L, MTHFD2 and SHMT2 being up regulated in proliferative cancer cells, has also identified a fourth enzyme DHFRL1 which may play a pivotal role in cancer proliferation (Jain 2012). DHFRL1 has been found to be the most up regulated of the genes analysed, with a 3.36 fold over expression in the SW620 cells compared to the SW480 cells. As explained in Chapter 4, DHFRL1 is responsible for the conversion of DHF to its active form THF in the mitochondria (Anderson *et al.* 2011, McEntee *et al.* 2011). Due to the fact that DHFRL1 was only discovered in 2011, its expression level in cancer is largely unknown (McEntee *et al.* 2011, Anderson *et al.* 2011). Therefore, the role it potentially plays within cancer may have been overlooked thus far. DHFRL1 has been implicated in playing a crucial role in de novo thymidylate synthesis in the mitochondria (Anderson 2011). What is curious is that the enzyme TYMS (thymidylate synthase) was found to be -2.5 fold lower expressed in the SW620 cells in comparison to the SW480 cell line. Hypothetically, it could be that the up regulation of DHFRL1 within the mitochondria is heavily geared towards the provision of THF for formate synthesis and not thymidylate synthesis within the mitochondria.

The enzymes DHFR, MTHFD1 and SHMT1 may have been found to have a lower expression level in the SW620 cells relative to the SW480 due to the fact their respective paralogues DHFRL1, MTHFD1L and MTHFD2 were up regulated. It may be that parts of the one carbon metabolism pathway in cytoplasm are somewhat depressed to allow for the up regulation of the mitochondrial pathway. The nature of the observed lower expression of the DNA synthesising enzymes TYMS and ATIC in the SW620 cell line relative to the SW480 cell line remains to be seen. Previous work by Minguzzi (2013) on HEK 293 cells with altered MTHFD1L expression found both TYMS and ATIC to have decreased expression upon MTHFD1L over expression.

As HEK 293 cells are a non-cancerous cell line, this decrease in expression was thought to be protective.

In summary, the results obtained in this chapter provide further evidence of the role of mitochondrial folate enzymes in cancer cell growth and proliferation. The advantage of up regulating mitochondrial folate enzymes to the SW620 cells and indeed other cancer cells is unambiguous. The cause of the well-known higher proliferation rate of the SW620 metastatic cancer cells in comparison to the primary SW480 cell line is multi-faceted and may not be fully attributable to the up regulation of MTHFD1L,MTHFD2 and SHMT2 (Gao *et al.* 2012). However, in light of recent evidence from the literature and this research thesis, it may be fair to say that up-regulation of mitochondrial folate enzymes in the SW620 cells may have conferred an advantage in some part to the SW620 cells allowing for increased proliferation and metastasis. This may have an implication in the treatment of colon cancer and indeed other cancers in which this gene signature is at play. Increased knowledge of the myriad of complex changes which occur in cancer cells allowing them to proliferate and metastasise is key in the treatment of cancer. The development of targeted therapies against crucial enzymes in the mitochondrial folate pathway may provide an effective modality of treatment in highly proliferative aggressive cancers in the future.

Chapter 6

General Discussion

Discussion

One carbon metabolism is an essential biochemical pathway in which folates are metabolised. The metabolism of folate provides one carbon molecules for the production of crucial cellular components such as the DNA bases adenine, methylation donor groups and amino acids which are necessary for the growth, replication and maintenance of health within the cell (Fox and Stover, 2008). The essentiality of folate within the cell has meant that it has had a longstanding relationship with health and disease. The relationship of folate with health and disease is not unambiguous; it seems that health complications can arise or become accelerated if too much or too little folate is consumed. In addition to folate status a number of folate genes have also been shown to play a role in disease development with the most significant roles seen to be in NTDs and cancer (Parle-McDermott *et al.* 2009, Kirke *et al.* 2004, Sugiura *et al.* 2004, Jain *et al,* 2012). Folate metabolism is compartmentalised within the cell, with the majority of biochemical reactions taking place in the cytoplasm and the mitochondria (Fox and Stover 2008). The importance of the mitochondria within one carbon metabolism has been highlighted over the past number of years with the discovery of novel enzymes such as MTHFD1L, MTHFD2L and DHFRL1 (Prasannan *et al.* 2003, Bolusani *et al.* 2011, McEntee *et al.* 2011).

The mitochondrion is often referred to as the “power house” of the cell, where the vast majority of cellular metabolism and energy production takes place (Chan., 2005). The cell is highly reliant on the mitochondria for the production of ATP, the energy currency within the cell. The phenomenon of some cancers cells to switch from the high energy producing oxidative phosphorylation to the energy inefficient method of aerobic glycolysis, is known as the Warburg effect (Warburg *et al.* 1927). The elucidation of this phenomenon by Otto Warburg lead to him defining cancer as a disease of the metabolism (Warburg *et al.* 1927). Although we now know there is much more to the etiology of cancer than a switch in energy production, Warburg’s theories and findings are ever more relevant today as our knowledge of the mitochondria and the metabolic pathways which it encompasses is ever expanding.

As the relevance of the mitochondria within cancer is coming to be fully appreciated the story of how each enzyme fits into the puzzle is becoming clearer. In the mitochondria the DHFRL1 enzyme plays a role in initiating the one carbon metabolism pathway by converting DHF to its biologically active form THF (Anderson *et al.* 2011). The MTHFD1L enzyme is responsible for the last step in mitochondrial one carbon metabolism by the conversion of 10-formyl THF to formate (Tibbetts and Appling 2010). The mono-functional enzyme MTHFD1L is responsible for the production of up to 75% of one carbon donor molecules for cytoplasmic one carbon metabolism (Pike *et al.* 2010). However, the functionality of MTHFD1L is not unique, in that the tri-functional cytoplasmic enzyme MTHFD1 also has 10-formyl THF synthetase activity (Tibbetts and Appling 2010). Despite the commonality between the MTHFD1 and MTHFD1L enzymes, it is the latter that has been associated with cancer proliferation and mortality (Sugiura *et al.* 2004, Jain *et al*. 2012). This fact leads to the hypothesis that It may be the "strategic positioning" of the MTHFD1L enzyme within the mitochondria that makes it so relevant in cancer and not necessarily its redundant functionality.

Both Sugiura *et al* (2004) and Jain *et al* (2012) have previously associated MTHFD1L expression with increased cell proliferation. These findings were correlated in this research thesis whereby over expression of the MTHFD1L gene resulted in increased cell growth in HEK 293 cells. Although the increase in cell growth was somewhat modest in comparison to the high proliferation associated with its expression as described by Jain *et al* (2012) .It was similar to what was described by Sugiura *et al* (2004) who found that over expression of the MTHFD1L gene resulted in "increased colony formation" in HEK 293 cells. This research thesis has shown that MTHFD1L over expression and knock down at a genetic level resulted in a detectable alteration in the metabolite of MTHD1L, formate. Formate has also been shown to act a reliable biomarker for stage and progression of oesophageal cancer, as demonstrated by Wang and colleagues (2013). Farshidfar *et al* (2012) also found that metabolomic analysis of serum samples from patients with adenocarcinoma identified that formate levels were significantly increased in patients with liver metastasis ($p=0.0005$) (Farshidfar *et a.l* 2012).

Although Wang *et al* (2013) and Farshidfar *et al* (2012) successfully demonstrated formate as a biomarker, they failed to address or tentatively explore the underlying cause of this observation. The work presented in this research thesis provides evidence of the role of both MTHFD1L and its metabolite formate. This research thesis has sought to take a comprehensive view of the role of both the MTHFD1L enzyme and its metabolite formate. Cause and effect was demonstrated with alterations in MTHFD1L at a genetic level resulting in alterations at a metabolite level. In concordance with the work of Farshidfar *et al* (2012) alterations in formate level were detected in the extracellular environment of HEK 293 cells with modulated MTHFD1L expression. All the evidence from both the literature and this research thesis is pointing towards MTHFD1L and formate being reliable biomarkers for not only cancer growth but also for staging and detection of metastasis (Jain *et al.* 2012, Farshidfar *et al.* 2012, Wang *et al.* 2013, Sugiura *et al.* 2004).

Coming back to "strategic positioning" within the mitochondria, both the DHFRL1 and MTHFD1L enzymes play critical roles within mitochondrial one carbon metabolism, with DHFRL1 playing a role in initiating the mitochondrial pathway in the provision of THF and in thymidylate synthesis and MTHFD1L responsible for the critical last step of the mitochondrial pathway (Anderson *et al*. 2011, Tibbetts and Appling 2010). The novel nature of the DHFRL1 enzyme means that much investigation is needed in order to begin to understand its functional relevance. It is thought that DHFRL1 and the de novo mitochondrial thymidylate synthesis pathway is necessary to maintain mitochondrial DNA (mtDNA) by preventing uracil incorporation (Anderson et al. 2011). Although, DHFRL1 has been shown to be a functional active enzyme, it has also been demonstrated that it has considerably less activity in comparison to DHFR (McEntee *et al.* 2011). McEntee *et al* (2011) have shown that the affinity of DHFRL1 for dihydrofolate (DHF) is ten times less than the affinity of DHFR for DHF. It is likely that the source of DHFRL1's reduced activity comes from one or a number of the amino acid differences between itself and DHFR. As demonstrated in Chapter 4, the tryptophan at position 24 of the human DHFR enzyme is highly conserved across many species from chimpanzee to rat, indicating that it has functional significance. McEntee et al (2011) had also

previously identified the switch from a tryptophan to an arginine in the DHFRL1 enzyme as significant, citing it may be a source of DHFRL1's reduced activity. Production and expression of recombinant DHFR, DHFRL1, DHFR W24R and DHFRL1 R24W in an *E.coli* expression system was successfully achieved. However, in the purification of the wild type enzymes DHFR and DHFRL1 it became apparent that insolubility of the DHFRL1 enzyme was an issue. Molecular dynamic simulations by Gao *et al* (2013) found that the structure of DHFRL1 was closer to the structure of *E.coli* DHFR than human DHFR. Molecular dynamic simulations also revealed that mutating the arginine at position 24 to a tryptophan in the DHFRL1 enzyme would cause the DHFRL1 enzyme to have a structure more similar to human DHFR (Gao *et al.* 2012). However, results from this research thesis indicate that the DHFRL1 R24W protein behaved no more like the wild type DHFR enzyme in terms of solubility and structure indicating that more amino acids may be involved in DHFRL1's apparent structural changes. The fact that changing the DHFR tryptophan at position 24 to an arginine caused it to become insoluble highlights the important role that the tryptophan at 24 may have on the enzyme's structure. As mentioned in Chapter 4 the solubility of the wild type and mutant DHFR and DHFRL1 enzymes was used as measure of structure as it has previously been shown that the structure of a protein in an *E.coli* system can depend on its structure (Idicula-Thomas and Balaji 2005). The predictive protein modelling results also mirrored the solubility results in terms of the effect that the amino acid 24 had. For example mutating the arginine in DHFRL1 to a tryptophan had no effect relative to the wild type DHFRL1 enzyme as assessed by Swiss-Model and protein solubility. The results from Chapter 4 indicate that the arginine at position 24 of the DHFRL1 enzyme results in DHFRL1 having an altered structure than DHFR.

The altered structure of the DHFRL1 enzyme may be in part why it has such a reduced affinity for DHF, as the residues required for its binding may not be properly exposed. Predictive modelling by Gao *et al* (2013) indicates that the pterin ring of DHF forms a hydrogen bond with arginine at position 24 and that this residue also forms two hydrogen bonds with NADPH. Arginine is quite a long

positively charged amino acid and is thought to extend deeply into the binding pocket of the DHFRL1 enzyme (Gao *et al.* 2013). So DHFRL1 may sequester DHF in a less efficient manner than DHFR resulting in a reduced affinity, as observed by McEntee *et al* (2011). The reduced affinity of DHFRL1 for DHF also suggests that the enzyme may have a reduced affinity for anti-folate chemotherapy drugs (McEntee *et al.* 2011). In Chapter 4 DHFRL1 was found to be up-regulated in the SW620 metastatic cell line relative to the primary SW480 cells. This indicates that DHFRL1 may also be part of the mitochondrial driven role in cancer cell development and metastasis. Further investigation would be needed in order to assess if DHFRL1 is a viable chemotherapeutic target, however it is likely that alternative drugs than those used to target DHFR would be needed to achieve efficient blocking of the enzyme. Predictive modelling by Gao *et al* (2013) indicated that DHFRL1 has a conformation closer to bacterial DHFR than to human DHFR. Perhaps alternative anti-folates that are known to have a much reduced affinity for human DHFR, such as anti-protozoals and anti-bacterials could be explored in the targeting of DHFRL1 (Zimmerman *et al.* 1987).

The one carbon metabolism pathway flows primarily in a clockwise direction. As one would expect there are a number of essential enzymes which perform the intermediate steps between the actions undertaken by the DHFRL1 and MTHFD1L enzymes, which should not be overlooked (Tibbetts and Appling 2010). A comprehensive view of the mitochondrial pathway is required in order to fully appreciate the molecular mechanisms underlying its association with cancer. The SW480 and SW620 cell lines provided an excellent cell line model for detecting changes at a genetic level between a primary tumour and a metastatic tumour (Leibovitz *et al.* 1976). The up regulation of the mitochondrial enzymes, in particular MTHFD1L, MTHFD2 and SHMT2 within the SW620 metastatic cell line was in concordance with the work of Jain *et al* (2012) who found that these enzymes were correlated with rapidly proliferating cell lines, but not their cytosolic counterparts. In addition, the up-regulation of the MTHFD1L gene in the metastatic SW620 cell line also supports the role of formate in cancer cell proliferation and supports the work by Wang *et al* (2013) and Fashfindar *et al* (2012) who found that the

metabolite of MTHFD1L, formate, was correlated with staging and the detection of metastasis in cancer. These results taken collectively provide a very strong case for the mitochondrial one carbon metabolism pathway playing a central role in cancer development.

It is not easy to decipher an exact sequence of events which results in the up-regulation of the mitochondrial pathway; one thing that is obvious however, is that its up-regulation is extremely advantageous to a proliferating cancer cell. One of the first steps in mitochondrial one carbon metabolism is the cleavage of serine to glycine by the SHMT2 enzyme (Tedeschi *et al.* 2013). Serine is derived from the glycolysis pathway; up-regulation of the glycolytic pathway is concomitant with many cancers (Warburg *et al*. 1927, Tedeschi *et al.* 2013). Serine is produced by oxidation of the glycolytic intermediate 3-phophoglycerate by the enzymes Phosphoglycerate Dehydrogenase (PHGDH) and Phosphohydroxythreonine Aminotransferase (PSAT1) and Phosphoserine Phosphatase (PSPH) (Amelio *et al.* 2014). Up-regulation of the PHGDH enzyme has found to be associated with triple negative breast cancer. Inhibition of the enzyme by shRNA was found cause a strong reduction in proliferating cancer cells that were shown to have an up-regulation of the enzyme, but cells whose PHGDH was at normal levels were unaffected (Possemato *et al.* 2011). Interestingly, it was found that intracellular serine levels were unaffected but a drop in α-ketoglutarate, an intermediate of the TCA cycle was observed (Possemato *et al.* 2011). Possemato *et al* (2011) also found that even when the cells with shRNA suppression of PHGDH were cultured in serine, rescue of the suppressed proliferation did not occur. Up-regulation of the PHGDH enzyme was found to cause rapid proliferation and a predisposition to malignant transformation in the MCF710A cell line (Amelio et al. 2014). In their review in Cell, Amelio *et al* (2014) suggest that up-regulation of the PHGDH enzyme and increased serine levels is sufficient for malignant transformation and that it is the serine biosynthesis pathway that promotes cancer cell development and proliferation by the provision of serine and other outputs which help to fuel the TCA cycle. In addition, the synthesis of serine in turn provides glycine and methyl donor groups for mitochondrial one carbon metabolism resulting in increased formate for

cytoplasmic one carbon metabolism, providing cancer cells with the necessary fuel for cell division and proliferation; see Figure 6.1. As explored in Chapter 5, it may be the switch in a cells metabolism towards aerobic glycolysis resulting in increased serine and 1 carbon molecules that in turn causes an up-regulation of enzymes such as SHMT2, MTHFD2, MTHFD1L and DHFRL1 as observed in this research thesis in the SW620 cells and in the literature (Jain *et al.* 2012, Nilsson *et al.* 2013).

In contrast to the hypothesis that glycolysis causes up regulation of one carbon metabolism enzymes, Zang *et al* (2012) identified that over expression of the enzyme glycine decarboxylase (GLDC) causes an up regulation of glycolytic enzymes. Zhang *et al* (2012) identified that over expression of the mitochondrial enzyme GLDC may be sufficient to initiate tumourigenesis in non-small cell lung cancer. These authors also found that over expression of the mitochondrial enzyme GLDC actually causes an up regulation of glycolytic enzymes and propose that over expression of GLDC actually promotes glycolysis and up regulation of glycine related enzymes SHMT1/2 in non-small cell lung cancer (Zhang *et al.* 2012). Whether up regulation of the glycine pathway and its related enzymes are the answer as to why cancer cells switch their metabolism to glycolysis, or are up regulated as consequence of the switch in energy production, remains to be further investigated.

The elucidation of any molecular mechanism in cancer generally has two main aims, 1. the identification of biomarkers for increased cancer detection, 2. the identification of therapeutic targets. The data generated in this thesis add to the body of information currently in the literature for both the utilisation of mitochondrial genes and their products as both biomarkers for cancer and as possible therapeutic targets. Vasquez *et al* (2013) found a correlation between leukaemia patients with high expression of MTHFD2 and SHMT2 and increased sensitivity to methotrexate. One finding which was of note was the fact that DHFR expression was not found to correlate with methotrexate sensitivity. The authors also found that around 25% of cancers over express one carbon metabolism enzymes. This finding may also aid in the treatment of cancers by stratifying and identifying patients based on their molecular profile and selecting appropriate

treatments. Jing *et al* (2015) found that shRNA knockdown of the serine producing enzyme PHGDH induced increased sensitivity of cisplatin treatment in cervical cancer cells, providing further evidence of the necessity of serine and its mitochondrial derivatives in cancer cells. There are currently a number of chemotherapy drugs both licensed and in clinical trials for targeting the one carbon metabolism and the serine/glycine biosynthetic pathways (Amelio *et al*. 2014). Perhaps the use of approved drugs such methotrexate and 5-fluorouracil against the enzymes DHFR and TYMS could be expanded or specifically used in cancers with up-regulation of folate enzymes in much the same strategy suggested by Vasquez and colleagues (Vasquez *et al.* 2013). In addition, there are a number of drugs in pre-clinical and clinical trials which are aimed and specifically targeting serine and glycine biosynthesis (Amelio *et al.* 2014).

To conclude, the results presented in this research thesis provide further evidence to the growing body of results from the literature on the significance of the mitochondrial folate pathway in cancer development proliferation; see Figure 6.1. Data presented in this research thesis provide further information on the role and activity of the newly identified DHFRL1 enzyme and its possible use as a chemotherapeutic target. Evidence was provided for the use of the metabolite of MTHFD1L, formate, as a biomarker for cancer cell proliferation. The expression of mitochondrial folate enzymes was found to be associated with a metastatic cancer cell line. The mechanisms underlying the overexpression and association of the mitochondrial folate pathway with cancer are not easy to distinguish; however, the identification of this molecular signature provides ample opportunity in the treatment and identification of the disease by the provision of biomarkers and chemotherapy targets.

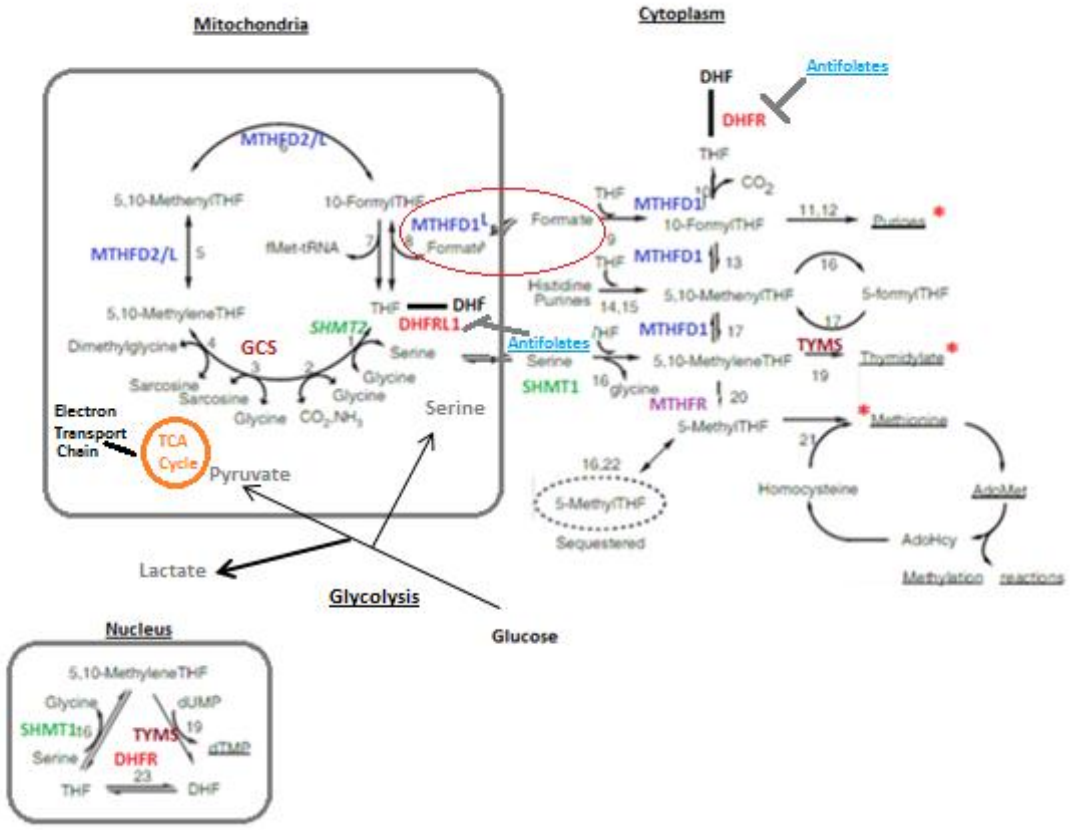
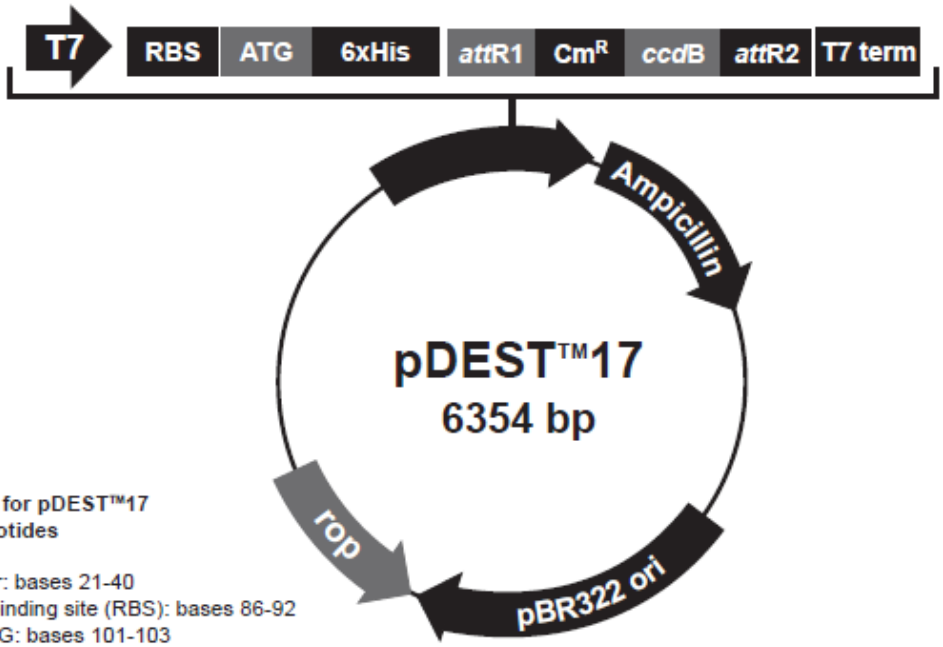


Figure 6.1. Overview of Enzymes Involved in One Carbon Metabolism and their Relationship with Cancer. Up-regulation of glycolysis is a common feature of many cancers. Glycolysis provides serine which it is then utilised in the mitochondria (Tedeschi 2013). DHFRL1 reduces DHF to its active form THF for thymidylate synthesis and as a source of one carbon units for other mitochondrial reactions (Anderson *et al* 2011). DHFRL1 was found to be 3 fold up-regulated in the metastatic SW620 cell line. DHFRL1 may be a potential anti-folate chemotherapeutic target; it is thought that different anti-folates than are currently used to target DHFR might be more effective. SHMT2, MTHFD2 and MTHFD1L are involved in the metabolism of THF and the one carbon derivatives from serine; all three enzymes have been found to be up-regulated in highly proliferative cancers (Jain *et al* 2012). MTHFD1L is crucial to the cell with it providing up to 75% of the one carbon donor molecules for cytoplasmic one carbon metabolism in the form of formate (Pike *et al* 2010). Formate has shown promise as a biomarker for cancer cell growth and metastasis. Image edited from Fox and Stover (2008).

Appendices

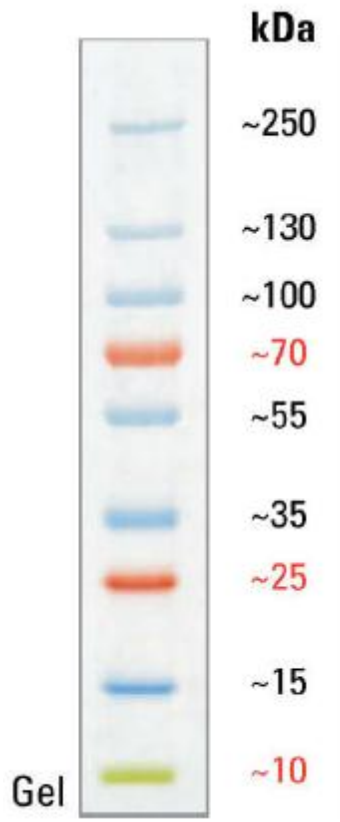


Comments for pDEST™17
6354 nucleotides

T7 promoter: bases 21-40
Ribosome binding site (RBS): bases 86-92
Initiation ATG: bases 101-103
6xHis tag: bases 113-130
*att*R1: bases 140-264
Chloramphenicol resistance gene (Cm^R): bases 373-1032
*ccd*B gene: bases 1374-1679
*att*R2: bases 1720-1844
T7 transcription termination region: bases 1855-1983
bla promoter: bases 2471-2569
Ampicillin (*bla*) resistance gene: bases 2570-3430
pBR322 origin: bases 3575-4248
ROP ORF: bases 4619-4810 (C)

C=complementary strand

Appendix A1 Invitrogen pDEST17 Vector Map



Appendix B1 Thermo Scientific PAGE Ruler Pre-stained Protein Ladder (Product No. 26619).

```
CLUSTAL O(1.2.1) MULTIPLE SEQUENCE ALIGNMENT


R24W         ATGTTTCTTTTGCTAAACTGCATCGTCGCTGTGTCCCAAAACATGGGCATCGGCAAGAAC
DHFRL1       ATGTTTCTTTTGCTAAACTGCATCGTCGCTGTGTCCCAAAACATGGGCATCGGCAAGAAC
             ************************************************************

R24W         GGGGACCTGCCCTGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACA
DHFRL1       GGGGACCTGCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACA
             ************:***********************************************

R24W         ACTTCTTCAGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCC
DHFRL1       ACTTCTTCAGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCC
             ************************************************************

R24W         ATTCCTGAGAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAACTC
DHFRL1       ATTCCTGAGAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAACTC
             ************************************************************

R24W         AAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCTTAAAACTT
DHFRL1       AAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCTTAAAACTT
             ************************************************************

 R24W       ACTGAACGACCAGAATTAGCAAATAAAGTAGACATGATTTGGATAGTTGGTGGCAGTTCT
DHFRL1       ACTGAACGACCAGAATTAGCAAATAAAGTAGACATGATTTGGATAGTTGGTGGCAGTTCT
             ************************************************************

R24W         GTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAACTATTTGTGACAAGGATCATG
DHFRL1       GTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAACTATTTGTGACAAGGATCATG
             ************************************************************

R24W         CAGGACTTTGAAAGTGACACGTTTTTTTCAGAAATTGACTTGGAGAAATATAAACTTCTG
DHFRL1       CAGGACTTTGAAAGTGACACGTTTTTTTCAGAAATTGACTTGGAGAAATATAAACTTCTG
             ************************************************************

R24W         CCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGGGAAACACATCAAGTACAAATTT
DHFRL1       CCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGGGAAACACATCAAGTACAAATTT
             ************************************************************

R24W         GAAGTATGTGAGAAGGATGATTAA
DHFRL1       GAAGTATGTGAGAAGGATGATTAA
             ************************
```

Appendix C1 Clustal Omega Alignment of the DNA sequence of the DHFRL1 gene and R24W mutant sequencing results. The sequencing results for the DHFRL1 R24W mutation indicate that the nucleotide base change was successfully introduced, as denoted in red above.

```
CLUSTAL O (1.2.1) multiple sequence alignment


R24W         MFLLLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGRKTWFS
DHFRL1       MFLLLNCIVAVSQNMGIGKNGDLPRPPLRNEFRYFQRMTTTSSVEGKQNLVIMGRKTWFS
             ************************:***********************************

R24W         IPEKNRPLKDRINLVLSRELKEPPQGAHFLARSLDDALKLTERPELANKVDMIWIVGGSS
DHFRL1       IPEKNRPLKDRINLVLSRELKEPPQGAHFLARSLDDALKLTERPELANKVDMIWIVGGSS
             ************************************************************

R24W         VYKEAMNHLGHLKLFVTRIMQDFESDTFFSEIDLEKYKLLPEYPGVLSDVQEGKHIKYKF
DHFRL1       VYKEAMNHLGHLKLFVTRIMQDFESDTFFSEIDLEKYKLLPEYPGVLSDVQEGKHIKYKF
             ************************************************************

R24W         EVCEKDD
DHFRL1       EVCEKDD
             *******
```

Appendix C2 Clustal Omega amino acid alignment of the DHFRL1 enzyme and the R24W mutant. Translation of the DNA sequencing results for the R24W mutation to the amino acid sequence confirms that the single DNA base change made in the DHFRL1 gene is non-synonymous and results in the conversion of the amino acid from arginine (R) to a tryptophan (W).

```
CLUSTAL O(1.2.1) multiple sequence alignment

W24R      ATGGTTGGTTCGCTAAACTGCATCGTCGCTGTGTCCCAGAACATGGGCATCGGCAAGAAC
DHFR      ATGGTTGGTTCGCTAAACTGCATCGTCGCTGTGTCCCAGAACATGGGCATCGGCAAGAAC
          ************************************************************

W24R      GGGGACCTGCCCAGGCCACCGCTCAGGAATGAATTCAGATATTTCCAGAGAATGACCACA
DHFR      GGGGACCTGCCCTGGCCACCGCTCAGGAATGAATTCAGATATTTCCAGAGAATGACCACA
          ************:***********************************************

W24R      ACCTCTTCAGTAGAAGGTAAACAGAATCTGGTGATTATGGGTAAGAAGACCTGGTTCTCC
DHFR      ACCTCTTCAGTAGAAGGTAAACAGAATCTGGTGATTATGGGTAAGAAGACCTGGTTCTCC
          ************************************************************

W24R      ATTCCTGAGAAGAATCGACCTTTAAAGGGTAGAATTAATTTAGTTCTCAGCAGAGAACTC
DHFR      ATTCCTGAGAAGAATCGACCTTTAAAGGGTAGAATTAATTTAGTTCTCAGCAGAGAACTC
          ************************************************************

W24R      AAGGAACCTCCACAAGGAGCTCATTTTCTTTCCAGAAGTCTAGATGATGCCTTAAAACTT
DHFR      AAGGAACCTCCACAAGGAGCTCATTTTCTTTCCAGAAGTCTAGATGATGCCTTAAAACTT
          ************************************************************

W24R      ACTGAACAACCAGAATTAGCAAATAAAGTAGACATGGTCTGGATAGTTGGTGGCAGTTCT
DHFR      ACTGAACAACCAGAATTAGCAAATAAAGTAGACATGGTCTGGATAGTTGGTGGCAGTTCT
          ************************************************************

W24R      GTTTATAAGGAAGCCATGAATCACCCAGGCCATCTTAAACTATTTGTGACAAGGATCATG
DHFR      GTTTATAAGGAAGCCATGAATCACCCAGGCCATCTTAAACTATTTGTGACAAGGATCATG
          ************************************************************

W24R      CAAGACTTTGAAAGTGACACGTTTTTTCCAGAAATTGATTTGGAGAAATATAAACTTCTG
DHFR      CAAGACTTTGAAAGTGACACGTTTTTTCCAGAAATTGATTTGGAGAAATATAAACTTCTG
          ************************************************************

W24R      CCAGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGAGAAAGGCATTAAGTACAAATTT
DHFR      CCAGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGAGAAAGGCATTAAGTACAAATTT
          ************************************************************

W24R      GAAGTATATGAGAAGAATGATTAG
DHFR      GAAGTATATGAGAAGAATGATTAG
          ************************
```

Appendix D1 Clustal Omega Alignment of the DNA sequence of the DHFR gene and the W24R mutant sequencing results. The sequencing results for the DHFRL W24R mutation indicate that the nucleotide base change was successfully introduced, as denoted in green above.

```
CLUSTAL O(1.2.1) multiple sequence alignment


W24R      MVGSLNCIVAVSQNMGIGKNGDLPRPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFS
DHFR      MVGSLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFS
          ********************** **********************************

W24R      IPEKNRPLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSS
DHFR      IPEKNRPLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSS
          ************************************************************

W24R      VYKEAMNHPGHLKLFVTRIMQDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKF
DHFR      VYKEAMNHPGHLKLFVTRIMQDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKF
          ************************************************************

W24R      EVYEKND
DHFR      EVYEKND
          *******
```

Appendix D2 Clustal Omega amino acid alignment of the DHFR enzyme and the W24R mutant. Translation of the DNA sequencing results for the W24R mutation to the amino acid sequence confirms that the single DNA base change made in the DHFR gene is non-synonymous and results in the conversion of the amino acid from a tryptophan (W) to an arginine(R).

References

- Amelio I, Cutruzzola F, Antonov A, Agostini M, Melino G. Serine and glycine metabolism in cancer.Trends Biochem Sci. **39**:191–198 (2014).
- Anagnou NP, Antonarakis SE, O'Brien SJ, Modi WS, Nienhuis AW. Chromosomal localization and racial distribution of the polymorphic human dihydrofolate reductase pseudogene (DHFRP1). Am J Hum Genet. (1988) 42(**2**):345-52.
- Anagnou NP, O'Brien SJ, Shimada T, Nash WG, Chen MJ, Nienhuis AW. Chromosomal organization of the human dihydrofolate reductase genes: dispersion, selective amplification, and a novel form of polymorphism. Proc Natl Acad Sci U S A. 1984 81(**16**):5170-4.
- Anderson DD, Quintero CM and Stover PJ: Identification of a de novo thymidylate biosynthesis pathway in mammalian mitochondria. Proc Natl Acad Sci USA (2011) **108** 15163-151638.
- Anderson DD, Woeller CF, Chiang EP, Shane B, Stover PJ. Serine hydroxymethyltransferase anchors de novo thymidylate synthesis pathway to nuclear lamina for DNA synthesis. J Biol Chem. (2012) 287(**10)**:7051-62.
- Arnold K, Bordoli L, Kopp J, and Schwede T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics (2006) **22**; 195-201.
- Bailey L B and Gregory JF. Folate Metabolism and Requirements. J. Nutr. (1999) **129**: 4 779-782
- Beard WA, Appleman JR, Huang S, Delcamp TJ, Freisheim JH, Blakley RL. Role of the conserved active site residue tryptophan-24 of human dihydrofolate reductase as revealed by mutagenesis. Biochemistry (1991) **30**:1432–1440.
- Beaudin AE, Stover PJ. Insights into metabolic mechanisms underlying folate-responsive neural tube defects: a minireview. Birth Defects Res A Clin Mol Teratol. (2009) 85(**4**):274-84.
- Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Dev Cell. 2010 19(**5)**:698-711.

- Biasini M, Bienert S, Waterhouse A, *et al* (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information Nucleic Acids **42** (W1): W252-W258.
- Blount BC, Matthew, Mack M, Wehr CM, MacGregor JT, Hiatt RA, Wang G, Wickramasinghe SN, Everson RB, and Ames BN. Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: Implications for cancer and neuronal damage. Proc Natl Acad Sci U S A. (1997) 94(7):3290-5.
- Bolusani S, Young BA, Cole NA, Tibbetts AS, Momb J, Bryant JD, Solmonson A, Appling DR. Mammalian MTHFD2L encodes a mitochondrial methylenetetrahydrofolate dehydrogenase isozyme expressed in adult tissues. J Biol Chem. (2011) 286(**7**):5166-74
- Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. and Schwede, T.Protein structure homology modelling using SWISS-MODEL Workspace. Nature Protocols (2009) **4**,1.
- Botto L.D, Moore C.A , Khoury M.J, and Erickson. Neural Tube Defects review. N Engl J Med (1999) **341**:1509-1519.
- Brody LC, Conley M, Cox C, Kirke PN, McKeever MP, Mills JL, Molloy AM, O'Leary VB, Parle-McDermott A, Scott JM, Swanson DA. A polymorphism, R653Q, in the trifunctional enzyme methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase/formyltetrahydrofolate synthetase is a maternal genetic risk factor for neural tube defects: report of the Birth Defects Research Group. Am J Hum Genet. (2002)71(**5**):1207-15.
- Brogden RN, Carmine AA, Heel RC, Speight TM, Avery GS. Trimethoprim: a review of its antibacterial activity, pharmacokinetics and therapeutic use in urinary tract infections. Drugs. (1982) **6:**405-30.
- Bryan J, Calvaresi E, Hughes D. Short term folate vitamin B12 or vitamin B6 supplementation slightly affects memory performance but not mood in women of various ages. Journal of Nutrition (2002) 132(**6**):1345–56.

- Centre for Disease Control and Prevention. Trends in Wheat-Flour Fortification with Folic Acid and Iron --- Worldwide, 2004 and 2007. MMWR (2008) 57(**01**);8-10
- Chan PH. Mitochondrial dysfunction and oxidative stress as determinants of cell death/survival in stroke. Ann N Y Acad Sci. (2005) **1042**:203-9.

- Chan PH. Mitochondrial dysfunction and oxidative stress as determinants of cell death/survival in stroke. Ann N Y Acad Sci. (2005) **1042**:203-9.

- Choi JD, Lee JS. Interplay between Epigenetics and Genetics in Cancer. Genomics Inform. (2013) 11(**4**):164-73.
- Clarke R, Daly L, Robinson K, Naughten E, Cahalane S, Fowler B *et al* Hyperhomocysteinemia: an independent risk factor for vascular disease. N Engl J Med (1991) 324 **17**:1149–1155.
- Clarke R, Halsey J, Bennett D, Lewington S.Homocysteine and vascular disease: review of published results of the homocysteine-lowering trials. JInherit Metab Dis (2011) **34**:83–91
- Cole BF, Baron JA, Sandler RS, Haile RW, Ahnen DJ, Bresalier RS, McKeown-Eyssen G, Summers RW, Rothstein RI, Burke CA, Snover DC, Church TR, Allen JI, Robertson DJ, Beck GJ, Bond JH, Byers T, Mandel JS, Mott LA, Pearson LH, Barry EL, Rees JR, Marcon N, Saibil F, Ueland PM. Folic acid for the prevention of colorectal adenomas: a randomized clinical trial. J Am Med Assn (2007) (**21**):2351-9.
- Copp AJ, Greene ND. Genetics and development of neural tube defects. J Pathol. (2010) 220(**2**): 217–230.
- Durga J, Boxtel M.P, Schouten E.G, Kok F.J, Jolles J, Katan M.B, Verhoef P. Effect of 3-year folic acid supplementation on cognitive function in older adults in the FACIT trial: a randomised, double blind, controlled trial. The Lancet (2007) **369**, 208–216
- Duthie SJ. Folic acid deficiency and cancer: mechanisms of DNA instability. Br Med Bull. (1999) **55**(3):578-92.

- Ebbing M, Bønaa KH, Nygård O, Arnesen E, Ueland PM, Nordrehaug JE, Rasmussen K, Njølstad I, Refsum H, Nilsen DW, Tverdal A, Meyer K, Vollset SE. Cancer incidence and mortality after treatment with folic acid and vitamin B12. JAMA. (2009) 302(**19**):2119-26.
- Ercikan-Abali EA, Banerjee D, Waltham MC, Skacel N, Scotto KW, Bertino JR. Dihydrofolate reductase protein inhibits its own translation by binding to dihydrofolate reductase mRNA sequences within the coding region. Biochemistry. (1997) 36(**40**):12317-22.
- Ercikan-Abali EA, Waltham MC, Dicker AP, Schweitzer BI, Gritsman H, Banerjee D, Bertino JR. Variants of human dihydrofolate reductase with substitutions at leucine-22: effect on catalytic and inhibitor binding properties. Mol Pharmacol. (1996) 49(**3**):430-7.
- Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJ. World Health Organization. The World Health Report, 2002: reducing risks, promoting healthy life. Geneva, 2002. Selected major risk factors and global and regional burden of disease. Lancet. (2002)**360**:1347–1360.
- Farber, S., Cutler, E.C., Hawkins, J.W., Harrison, J.H., Pierce, E.C., II & Lenz, G.G. The action of pteroylglutamic conjugates on man. Science (1947), **106**, 2764–2768.
- Farshidfar F, Weljie AM, Kopciuk K, Buie WD, Maclean A, Dixon E, Sutherland FR, Molckovsky A, Vogel HJ, Bathe OF. Serum metabolomic profile as a means to distinguish stage of colorectal cancer. Genome Med. (2012) 4(**5**):42.
- Fekete K, Berti C, Cetin I, Hermoso M, Koletzko BV, Decsi T. Perinatal folate supply: relevance in health outcome parameters. Matern Child Nutr. (2010) **2**:23-38.
- Fidler I.J. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited *Nature Reviews Cancer* (2003) **3**, 453-458
- Fien O, Metabolomics – the link between genotypes and phenotypes. Plant Molecular Biology 48: 155–171, 2002.

- Figueiredo JC, Grau MV, Haile RW, Sandler RS, Summers RW, Bresalier RS, Burke CA, McKeown-Eyssen GE, Baron JA. Folic acid and risk of prostate cancer: results from a randomized clinical trial. J Natl Cancer Inst. (2009) **6:**432-5.
- Flynn MAT, Anderson WA, Burke SJ, Reilly A. Folic acid food fortification: the Irish experience. Proc Nutr Soc (2008) **67**: 381-389.
- Food Safety Authority of Ireland. Report of the implementation Group on Folic Acid Food Fortification to the Department of Health and Children, 2009.
- Food Safety Authority of Ireland. Report of the National committee on Folic Acid Food Fortification,2006.
- Fox J.T, Stover P.J 2008, Folate Mediated One Carbon Metabolism. In: Folic Acid and Folates, Vitamins and Hormones Volume 79, Edited By Gerald Litwack ,2008 Academic Press. Elsevier.
- Frey L, Hauser WA. Epidemiology of neural tube defects. Epilepsia. (2003) **3**:4-13.
- Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP, et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat Genet. (1995) 10(**1**):111-3.
- Gao H, Wang J.J, Yang X, Liu ZR. Pyruvate kinase M2 regulates gene transcription by acting as a protein kinase**.** Mol. Cell. (2012) **45** 598–609.
- Gao J, Cui W, Du Y and Ji M. Insight into the molecular mechanism about lowered dihydrofolate binding affinity to dihydrofolate reductase-like 1 (DHFRL1). J Mol Model (2013) **19** 5187–5198.
- Giovannucci E, Stampfer MJ, Colditz GA, Hunter DJ, Fuchs C, Rosner BA, Speizer FE, Willett WC. Multivitamin use, folate, and colon cancer in women in the Nurses' Health Study, Annals of Internal Med. (1998) **129**:517-24.
- Greene ND, Stanier P, Copp AJ. Genetics of human neural tube defects. Hum Mol Genet. (2009) 18(**R2**):R113-29.

- Han ZD, Lu JM1, Cai C, Zeng YR, Zhong WD, Wu CL. Aberrant hypomethylation-mediated CD147 overexpression promotes aggressive tumor progression in human prostate cancer. Oncol Rep. (2015) 33(**5**):2648-54.
- Hewitt RE, McMarlin A, Kleiner D, Wersto R, MartinP, Tsokos M, *et al*. Validation of a model of colon cancer progression. J Pathol 2000; **192**:446-54.
- Husemoen LL, Skaaby T, Jørgensen T, Thuesen BH, Fenger M, Grarup N, Sandholt CH, Hansen T, Pedersen O, Linneberg A. MTHFR 677C<T genotype and cardiovascular risk in a general population without mandatory folic acid fortification. Eur J Nutr. (2014) 53(**7**):1549-59.
- Idicula-Thomas S, V. Balaji P. Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli. Protein Sci. (2005) 14(**3**): 582–592.
- Jain M, Nilsson R, Sharma S, Madhusudhan N, Kitami T, Souza AL, Kafri R, Kirschner MW, Clish CB, Mootha VK. Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. Science. (2012) **336:**1040–1044.
- Jing Z, Heng W, Xia L, Ning W, Yafei Q, Yao Z, Shulan Z. Downregulation of phosphoglycerate dehydrogenase inhibits proliferation and enhances cisplatin sensitivity in cervical adenocarcinoma cells by regulating Bcl-2 and caspase-3. Cancer Biol Ther. (2015) 16(**4)**:541-8.
- Johnson W.G, Stenroos E. S, Spychala J.R , Chatkupt S, Ming S.X, Buyske S. New 19 bp deletion polymorphism in intron-1 of dihydrofolate reductase (DHFR): A risk factor for spina bifida acting in mothers during pregnancy? American Journal of Medical Genetics (2004) Volume **124A** 339 – 345
- Kim YI. Folate and Colorectal Cancer: An evidence- based Critical Review. Mol. Nutr. Food Res. (2007) **51**:267-292.
- Kirke PN, Mills JL, Molloy AM, Brody LC, O'Leary VB, Daly L, Murray S, Conley M, Mayne PD, Smith O. Impact of the MTHFR C677T polymorphism on risk of neural tube defects: case-control study . Brit Med J. (2004) **328**:1535–1536.

- Klerk M, Verhoef P,Clarke R, Blom H.J, Kok F.J, Schouten E.G. MTHFR 677C→T Polymorphism and Risk of Coronary Heart Disease A Meta-analysis. JAMA. (2002) 288(**16**):2023-2031.
- Koonin EV. Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet. (2005) **39**:309-38.
- Lajous M, Lazcano-Ponce E, Hernandez-Avila M, Willett W, Romieu I. Folate, vitamin B(6), and vitamin B(12) intake and the risk of breast cancer among Mexican women. Cancer Epidemiol Biomarkers Prev. (2006) **15**(3):443-8.
- Lamarre S.G ,MacMillan L, Morrow G.P, Randell E, Pongnopparat T, Brosnan M.E, Brosnan J.T. An isotope-dilution, GC–MS assay for formate and its application to human and animal metabolism. Amino Acids (2014) 46:1885–1891
- Lamarre SG, Molloy AM, Reinke SN, Sykes BD, Brosnan ME, Brosnan JT. Formate can differentiate between hyperhomocysteinemia due to impaired remethylation and impaired transsulfuration. Am J Physiol Endocrinol Metab. (2012) 302(**1**):E61-7.
- Larsson SC, Giovannucci E, Wolk A. Dietary folate intake and incidence of ovarian cancer: the Swedish Mammography Cohort. J Natl Cancer Inst. 2004 (**5**):396-402.
- Last R.L, Jones A.D, Shachar-Hil Y. Towards the plant metabolome and beyond. Nature Reviews Molecular Cell Biology 8, 167-174 (2007).
- Lee, G. Y. *et al*. Comparative oncogenomics identifies PSMB4 and SHMT2 as potential cancer driver genes. *Cancer. Res.* **74**, 3114–3126 (2014).
- Lehtinen L, Ketola K, Mäkelä R, Mpindi JP, Viitala M, Kallioniemi O, Iljin K. High-throughput RNAi screening for novel modulators of vimentin expression identifies MTHFD2 as a regulator of breast cancer cell migration and invasion. Oncotarget. (2013) 4(**1**):48-63.
- Leibovitz A, Stinson JC, McCombs WB 3rd, McCoy CE, Mazur KC, Mabry ND. Classification of human colorectal adenocarcinoma cell lines. Cancer Research (1976) 36(**12**):4562-9.

- Liang YX, Mo RJ, He HC, Chen JH, Zou J, Han ZD, Lu JM, Cai C, Zeng YR, Zhong WD, Wu CL. Aberrant hypomethylation-mediated CD147 overexpression promotes aggressive tumor progression in human prostate cancer. Oncol Rep. (2015) 33(**5**):2648-54.
- MacFarlane AJ, Anderson DD, Flodby P, Perry CA, Allen RH, Stabler SP, Stover PJ. Nuclear localization of de novo thymidylate biosynthesis pathway is required to prevent uracil accumulation in DNA. J Biol Chem. (2011) 286(**51**):44015-22.
- Malouf R, Grimley Evans J. Folic acid with or without vitamin B12 for the prevention and treatment of healthy elderly and demented people. Cochrane Database Syst Rev.(2008) 8;(**4**) :CD004514.
- Marcus J, Sarnak MJ, Menon V. Homocysteine lowering and cardiovascular disease risk: lost in translation. Can J Cardiol. (2007) 23(**9**):707-10.
- Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature. (2007) 445(**7128**):666-70.
- McDonnell R, Delany V, O'Mahony MT, Mullaney C, Lee B, Turner MJ. Neural tube defects in the Republic of Ireland in 2009-11. J Public Health (Oxf). (2015) 37(**1**):57-63.
- McEntee G, Minguzzi S, O'Brien K, Ben Larbi N, Loscher C, Ó'Fágáin C, and Parle-McDermott A. The former annotated human pseudogene dihydrofolate reductase-like 1 (DHFRL1) is expressed and functional. Proc Natl Acad Sci USA (2011) **108:**15157- 15162.
- Miller,D.R. A tribute to Sidney Farber– the father of modern chemotherapy. Br J Haematol (2006), 134:20-26.
- Minguzzi S. *Investigation of the human folate gene MTHFD1L: polymorphisms and disease risk.* Ph.D thesis, (2013) Dublin City University.
- Minguzzi S, Selcuklu SD, Spillane C, Parle-McDermott A. An NTD-associated polymorphism in the 3' UTR of MTHFD1L can affect disease risk by altering miRNA binding. Hum Mutat. (2014) 35(**1**):96-104.

- Momb J, Lewandowski JP, Bryant JD, Fitch R, Surman DR, Vokes SA and Appling DR: Deletion of Mthfd1l causes embryonic lethality and neural tube and craniofacial defects in mice. Proc Natl Acad Sci USA (2013) **110**:549-554.
- MRC Vitamin Research Group. Prevention of neural tube defects: results of the Medical Research Council Study. *Lancet*. (1991) **338**:131–137.
- Naj AC, Beecham GW, Martin ER, Gallins PJ, Powell EH, Konidari I, Whitehead PL, Cai G, Haroutunian V, Scott WK, Vance JM, Slifer MA, Gwirtsman HE, Gilbert JR, Haines JL, Buxbaum JD, Pericak-Vance MA. Dementia revealed: novel chromosome 6 locus for late-onset Alzheimer disease provides genetic evidence for folate-pathway abnormalities. PLoS Genet. (2010) 6(**9**)**.**
- National Cancer Registry Ireland. Cancer in Ireland 1994-2012: Annual report of the National Cancer Registry (2014).
- Niclot S, Pruvot Q, Besson C, Savoy D, Macintyre E, Salles G, Brousse N, Varet B, Landais P, Taupin P, Junien C, Baudry-Bluteau D. Implication of the folate-methionine metabolism pathways in susceptibility to follicular lymphomas. Blood. (2006) 108(**1**):278-85.
- Nilesh J. Samani, Jeanette Erdmann, Alistair S. Hall, Christian Hengstenberg, Massimo Mangino, Bjoern Mayer, *et al*. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. (2007) **447** 661–678.
- Nilsson R, Jain M, Madhusudhan N, Sheppard NG, Strittmatter L, Kampf C, Huang J, Asplund A, Mootha VK. Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer. Nat Commun. (2014) **5**:3128.
- Palomares LA, Estrada-Mondaca S, Ramírez OT. Production of recombinant proteins: challenges and solutions. Methods Mol Biol. (2004) **267:**15-52.
- Parle-McDermott A, Ozaki M. The impact of nutrition on differential methylated regions of the genome. Adv Nutr. (2011) (**6**):463-71
- Parle-McDermott A, Pangilinan F, Mills JL, Kirke PN, Gibney ER, Troendle J, O'Leary VB, Molloy AM, Conley M, Scott JM, Brody LC. The 19-bp deletion

polymorphism in intron-1 of dihydrofolate reductase (DHFR) may decrease rather than increase risk for spina bifida in the population. Am J Med Genet A. (2007) 143A(**11**):1174-80.

- Parle-McDermott A, Pangilinan F, O'Brien KK, Mills JL, Magee AM, Troendle J, Sutton M, Scott JM, Kirke PN, Molloy AM, Brody LC. A common variant in MTHFD1L is associated with neural tube defects and mRNA splicing efficiency. Hum Mutat. (2009) 30(**12**):1650-6.
- Patel S, Ahmed S. Emerging field of metabolomics: big promise for cancer biomarker identification and drug discovery. J Pharm Biomed Anal. (2015) **107**:63-74
- Pike C.T, Rajendra R, Artzt K, Appling D.R. Mitochondrial C1-Tetrahydrofolate Synthase (MTHFD1L) Supports the Flow of Mitochondrial One-carbon Units into the Methyl Cycle in EmbryosJ. Biol. Chem. (2010) **285:** 4612-4620.
- Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K, Sethumadhavan S, Woo HK, Jang HG, Jha AK, Chen WW, Barrett FG, Stransky N, Tsun ZY, Cowley GS, Barretina J, Kalaany NY, Hsu PP, Ottina K, Chan AM, Yuan B, Garraway LA, Root DE, Mino-Kenudson M, Brachtel EF, Driggers EM, Sabatini DM. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature. (2011) 476(**7360**):346-50.
- Pozarowski P, Darzynkiewicz Z. Analysis of cell cycle by flow cytometry. Methods Mol Biol. (2004) **281**:301-11.
- Prasannan P, Pike S, Peng K, Shane B, Appling DR. Human mitochondrial C1-tetrahydrofolate synthase: gene structure, tissue distribution of the mRNA, and immunolocalization in Chinese hamster ovary calls. J Biol Chem. (2003) 278(**44**):43178-87.
- Ramírez-Lorca R, Boada M, Antúnez C, López-Arrieta J, Moreno-Rey C, Hernández I, Marín J, Gayán J, González-Pérez A, Alegret M, Tárraga L, Real LM, Ruiz A. The MTHFD1L gene rs11754661 marker is not associated with Alzheimer's disease in a sample of the Spanish population. J Alzheimers Dis. (2011) 25(**1**):47-50.

- Ren RJ, Wang LL, Fang R, Liu LH, Wang Y, Tang HD, Deng YL, Xu W, Wang G, Chen SD. The MTHFD1L gene rs11754661 marker is associated with susceptibility to Alzheimer's disease in the Chinese Han population. J Neurol Sci. (2011) 308(**1-2**):32-4.
- Saïda F. Overview on the expression of toxic gene products in Escherichia coli. Curr Protoc Protein Sci.(2007) Chapter 5:Unit 5.1.
- Samani NJ, Erdmann J, Hall AS, Henstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, *et al.* Genomewide Association Analysis of Coronary Artery Disease.(2007) N Engl J Med 357:443–453.
- Schweitzer BI, Srimatkandada S, Gritsman H, Sheridan R, Venkataraghavan R, Bertino JR. Probing the role of two hydrophobic active site residues in the human dihydrofolate reductase by site-directed mutagenesis. J Biol Chem. (1989) (**34**):20786-95.
- Scotti M, Stella L, Shearer EJ, Stover PJ. Modeling cellular compartmentation in one-carbon metabolism. Wiley Interdiscip Rev Syst Biol Med. (2013) 5(**3**):343-65.
- Selcuklu SD, Donoghue MT, Rehmet K, de Souza Gomes M, Fort A, Kovvuru P, Muniyappa MK, Kerin MJ, Enright AJ and Spillane C. MicroRNA-9 inhibition of cell proliferation and identification of novel miR-9 targets by transcriptome profiling in breast cancer cells. J Biol Chem (2012) **287:**29516-29528.
- Seshadri S, Beiser A, Selhub J, Jacques PF, Rosenberg IH, D'Agostino RB, Wilson PW, Wolf PA. Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. N Engl J Med. (2002) 346(**7**):476-83.
- Sherr CJ. Cancer cell cycles. Science. (1996) 274(**5293**)1672-7.
- Søes S, Daugaard IL, Sørensen BS, Carus A, Mattheisen M, Alsner J, Overgaard J, Hager H, Hansen LL, Kristensen LS. Hypomethylation and increased expression of the putative oncogene ELMO3 are associated with lung cancer development and metastases formation. Oncoscience. (2014) 23;(**5**):367-74
- Stover PJ. One-carbon metabolism-genome interactions in folate-associated pathologies. J Nutr. (2009) 139(**12**):2402-5.

- Sugiura T, Nagano Y, Inoue T and Hirotani K. A novel mitochondrial C1-tetrahydrofolate synthetase is upregulated in human colon adenocarcinoma. Biochem Biophys Res Commun (2004) **315**:204-211.

- Sutton M, Scott JM, Kirke PN, Molloy AM, Brody LC. A common variant in MTHFD1L is associated with neural tube defects and mRNA splicing efficiency. Hum Mutat. (2009) 30(**12**):1650-6
- Tedeschi PM, Markert EK, Gounder M, Lin H, Dvorzhinski D, Dolfi SC, Chan LL, Qiu J, DiPaola RS, Hirshfield KM, Boros LG, Bertino JR, Oltvai ZN, Vazquez A. Contribution of serine, folate and glycine metabolism to the ATP, NADPH and purine requirements of cancer cells. Cell Death Dis. (2013) **4**:e877.

- Tellman, G. The E-Method: a highly accurate technique for gene-expression analysis. Nature Methods. (2006) **3**: i-ii.
- Terry P, Jain M, Miller AB, Howe GR, Rohan TE. Dietary intake of folic acid and colorectal cancer risk in a cohort of women. Int J Cancer. (2002) **97**:864-7.
- Thillet J, Absil J, Stone SR, Pictet R. Site-directed mutagenesis of mouse dihydrofolate reductase. Mutants with increased resistance to methotrexate and trimethoprim. J Biol Chem (1988) **263**:12500–12508.
- Tibbetts AS and Appling RA: Compartmentalization of Mammalian Folate Mediated One-Carbon Metabolism. Annu Rev Nutr (2010) **30**:57-81..
- Vansteenkiste, J.F. European Respiratory Journal. (2002) **35**:49s-60s.
- Vazquez A, Tedeschi PM, Bertino JR. Overexpression of the mitochondrial folate and glycine-serine pathway: a new determinant of methotrexate selectivity in tumors. Cancer Res. (2013) 73(**2**):478-82.
- Visentin M, Zhao R, Goldman ID. The Antifolates. Hematol Oncol Clin North Am. (2012) 26(3): 629–ix.
- Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, Slamon DJ, Murphy M, Novotny WF, Burchmore M, Shak S, Stewart SJ, Press M. Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. J Clin Oncol. (2002) 20(**3**):719-26.

- Volk, E.L, Farley K.M, Wu Y, Li F, Robey R.W, and Schneider E. Overexpression of Wild-Type Breast Cancer Resistance Protein Mediates Methotrexate Resistance. Cancer Research (2002) **62,** 5035–5040.
- Vollset SE, Clarke R, Lewington S, Ebbing M, Halsey J, Lonn E, Armitage J, Manson JE, Hankey GJ, Spence JD, Galan P, Bønaa KH, Jamison R, Gaziano JM, Guarino P, Baron JA, Logan RF, Giovannucci EL, den Heijer M, Ueland PM, Bennett D, Collins R, Peto R; B-Vitamin Treatment Trialists' Collaboration. Effects of folic acid supplementation on overall and site-specific cancer incidence during the randomised trials: meta-analyses of data on 50 000 individuals. Lancet (2013) **6736**, 1–8 (2013).
- Wang L, Chen J, Chen L, Deng P, Bu Q, Xiang P, Li M, Lu W, Xu Y, Lin H, Wu T, Wang H, Hu J, Shao X, Cen X, Zhao YL. 1H-NMR based metabonomic profiling of human esophageal cancer tissue. Mol Cancer. (2013) **12**:25.
- Warburg O, Wind F, Negelein E. The Metabolism of Tumours in the Body.J Gen Physiol. (1927) 8(**6**):519-30.
- Ward PS, Thompson CB. Metabolic Reprogramming: A Cancer Hallmark Even Warburg Did Not Anticipate. Cancer Cell (2012) 3(**21**):297–308,
- World Health Organisation Fact sheet no. 297 February 2015a.
- World Health Organisation Fact sheet no. 317 January 2015b.
- Zhang WC, Shyh-Chang N, Yang H, Rai A, Umashankar S, Ma S, Soh BS, Sun LL, Tai BC, Nga ME, Bhakoo KK, Jayapal SR, Nichane M, Yu Q, Ahmed DA, Tan C, Sing WP, Tam J, Thirugananam A, Noghabi MS, Pang YH, Ang HS, Mitchell W, Robson P, Kaldis P, Soo RA, Swarup S, Lim EH, Lim B. Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. Cell. (2012) 148(**1-2**):259-72.
- Zhu ZG, Ai QL, Wang WM, Xiao ZC. Meta-analysis supports association of a functional SNP (rs1801133) in the MTHFR gene with Parkinson's disease. Gene.(2013) 531(**1**):78-83.
- Zimmerman J, Selhub J, Rosenberg IH. Competitive inhibition of folate absorption by dihydrofolate reductase inhibitors, trimethoprim and pyrimethamine. Am J Clin Nutr. (1987) 46(**3**):518-22.