Investigation of a Novel Intein-Based *Escherichia coli* Expression System for Human Methylmalonyl CoA Mutase

A thesis presented to Massey University in partial fulfillment of the requirements for the degree of

Master of Science in Biochemistry

Alice Rosemary Clark

2005

Abstract

Human methylmalonyl CoA mutase (hMCM) is a 78 kDa homodimeric mitochondrial matrix enzyme. hMCM catalyses the conversion of 2R-methylmalonyl CoA to succinyl CoA in the metabolism of propionyl groups, and requires the vitamin B_{12}-derived cofactor adenosylcobalamin (AdoCbl). The mechanism of catalysis involves homolytic cleavage of AdoCbl's unusual C-Co bond, to generate radicals. Dysfunctional hMCM results in the rare, potentially fatal metabolic disorder methylmalonic acidemia. An experimentally determined structure of hMCM would add to the understanding of both the mechanism of catalysis and the molecular basis of some of the mutations underlying methylmalonic acidemia. The structure of the bacterial orthologue from *Propionibacterium shermanii* has been solved by x-ray crystallography, enabling the development of structural models of hMCM. Critical differences, however, between these two enzymes, mean that some regions of the models could be inaccurate.

There is no x-ray crystal structure of hMCM. Purification of native hMCM for crystallization trials is complicated by ethical problems, low yields, and heterogeneity generated by the cofactor. To provide a more convenient source of pure, active human methylmalonyl CoA mutase for x-ray crystallography, an expression system for recombinant hMCM is required. Other researchers have expressed hMCM in *Escherichia coli* as (i) insoluble inclusion bodies, (ii) soluble fusion protein that cannot be separated efficiently from the fusion tag, or (iii) in low quantities.

This research aimed to develop an *E. coli* expression system for the production of active human methylmalonyl CoA mutase, to enable x-ray crystallography structural studies. Based on the results of previous expression systems, four novel expression vectors were developed utilising the maltose binding protein and thioredoxin as solubility tags. It was hoped that conventional protease cleavage, to remove these solubility tags, could be circumvented by the use of intein-mediated cleavage. Intein-mediated cleavage was successful, and soluble active hMCM was recovered in low yields from a C-terminal thioredoxin solubility tag construct. hMCM was insoluble when expressed with MBP at the C-terminus.

I am among those who think that science has great beauty.

A scientist in his laboratory is not only a technician: he is also a child placed before

natural phenomena which impress him like a fairy tale.

Marie Curie (1867 - 1934)

Acknowledgements

Firstly I would like to thank my supervisor, Dr Mark Patchett, for your enthusiasm, patience and encouragement. You made this MSc a wonderful experience that I will remember for the rest of my life.

I would like to thank Dr Kathryn Stowell, for allowing me work in her lab, and to everyone else in the lab, for all the fun times we had. I would especially like to thank Carole, Kei and Andy for their patience and endless help in the lab. Also, thank you to everyone else in IMBS for all their help and advice.

I would like to acknowledge the Palmerston North Medical Research Foundation for the financial support that made this project possible.

Thanks to my family, Di, John, Helen, Robin and Lily who helped me though the difficult times during my time at Massey University.

I am forever grateful to Sacha, Lia and Natisha who helped me with this MSc and supported me in so many different ways. You are beautiful girls and made me the happy person I am today. Thank you.

Finally, I would like to thank Scott - for your belief in my abilities, and your patience with my inabilities.

Table of Contents   **Page**

List of Figures   **Page**

List of Tables   Page

List of Changes

- add, "cccDNA , covalently closed circular DNA" to the list   **xiii**
- change "figure 4" to "figure 1.4"   **4**
- change "compliment" to "complement"   **10**
- insert "of" between "solubility" and "proteins" (third line from bottom)   **11**
- change "Ion" to "Lon"   **13**
- change "termini" to "terminus"   **16**
- change "2001" in figure legend to "2002"   **20**
- change "Evens" to "Evans"   **22**
- change "tetramethylethylene di'amine" to "tetramethylethylenediamine"   **25**
- change "sample to elucidate" to "sample. To elucidate"   **27**
- change "NAHCO 3" to "NaHCO$_3$ "   **37**
- change "Lane 3 contain" to "Lane 3 contains"   **65**
- change "appendix G" to "appendix F"   **66**
- in line 11 insert "smaller size" after "This"
- change "change" to "changed"   **80**
- change "contains rare codons within the hMCM gene" to "supplies extra tRNAs   **88**
  for the rare codons in the hMCM mRNA. These rare codons may"
- on the second to last line: change "yield the " to "yield. The". Also change   **90**
  "decrease" to "decreases"
- change "figure 5.12" to "figure 5.13" in the second to last line   **91**
- change "containing rare codons" to "supplying extra tRNAs for the rare codons in   **108**
  the hMCM mRNA"
- change "Hayes, 1992" to "Hayes, 1998"   **108**
- change "*P. furiousis*" to "*P. furiosus*"   **111**

List of Abbreviations

AdoCbl	Adenosylcobalamin
Amp	Ampicillin
BLAST	Basic local alignment search tool
BSA	Bovine serium albumin
CBD	Chitin binding domain
C-terminal	Carboxyl terminial
CoA	Coenzyme A
DNA	Deoxyribonucleic acid
dNTP	Deoxy-nucleotide tri-phosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
g	Gravitational field, unit of
hMCM	Human methylmalonyl CoA mutase
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilobase pairs (of DNA)
MalE	Gene encoding the maltose binding protein
MAP	Methionine aminopeptidase
MBP	Maltose binding protein
MCM	Methylmalonyl CoA mutase
MCS	Multiple cloning site
mRNA	messenger ribonucleic acid
NADH	Nicotin amide-adenine dinucleotide, reduced
N-terminal	Amino terminal
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
TEMED	N, N, N', N'-tetramethylethylenediamine
Trx	ORF encoding the thioredoxin protein with a His tag
Tris	Tris-(hydroxymethyl)-aminomethane
V	volts

1   Introduction

1.1   The Function of Methylmalonyl CoA Mutase (MCM)

Human methylmalonyl CoA mutase (hMCM) is a mitochondrial enzyme that catalyses the isomerisation of *R*-methylmalonyl CoA to succinyl CoA, and the succinyl group is ultimately oxidised in the citric acid cycle (figure 1.1). This reaction is required in the metabolism of propionyl CoA formed during the catabolism of odd chain fatty acids, cholesterol intermediates, thymine, uracil and the amino acids methionine, isoleucine and valine (Kolhouse *et al.*, 1988).

In prokaryotes MCM has a different role: it is involved in the terminal fermentation pathway of succinyl CoA to propionate, and catalyses the conversion of succinyl-CoA to methylmalonyl CoA (Zagalak and Retey, 1974).
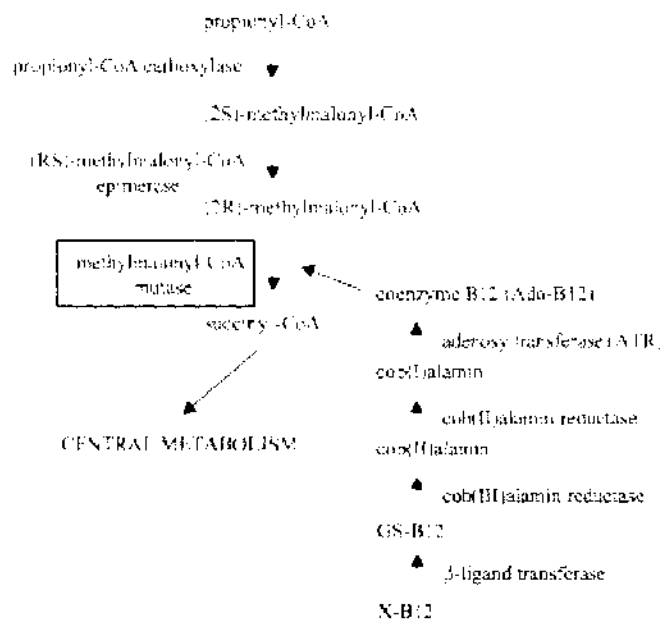
Figure 1.1 Propionyl CoA Metabolism in Humans. Showing the formation of the hMCM substrate methylmalonyl CoA and synthesis of the cofactor coenzyme B_{12} from inactive B_{12} precursors (XB12).

(Figure adapted from Leal *et al.*, 2003)

1.2   The Cofactor and Catalytic Mechanism

MCM belongs to a group of prokaryotic and animal enzymes that require the vitamin B_{12}-derived cofactor adenosylcobalamin (AdoCbl), also known as coenzyme B_{12} and of this group only MCM is found in both prokaryotes and animals (Banerjee, 1997). AdoCbl-dependent enzymes catalyse carbon skeleton rearrangements via a radical mechanism involving the carbon-cobalt bond in the AdoCbl cofactor, (figure 1.2), (Taoka *et al.*, 1997; Banerjee and Vlasie, 2002).
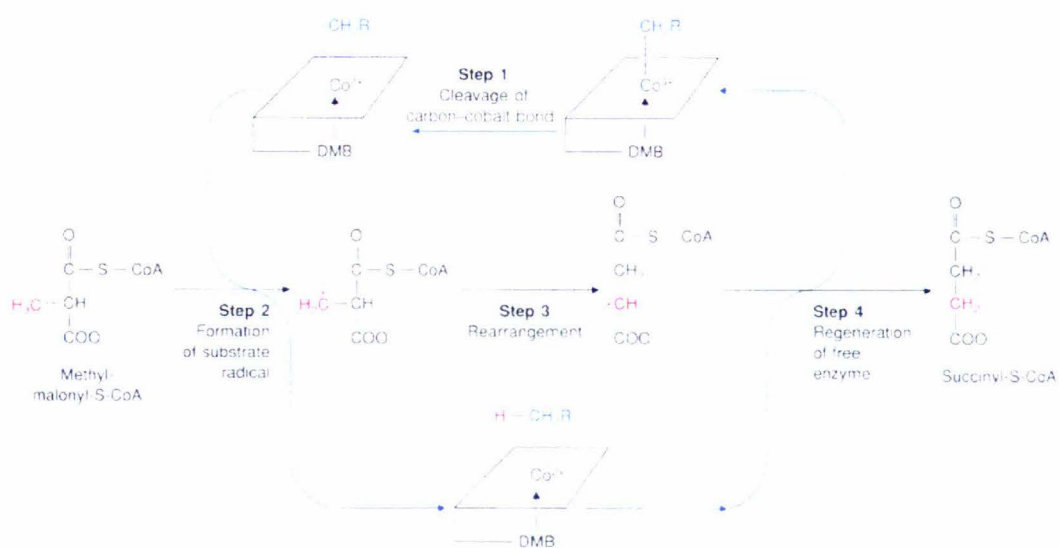
Figure 1.2 The Reaction of Methylmalonyl CoA Mutase. Showing the radical generation by the adenosylcobalamin cofactor (shown as the square containing Co), then the formation of succinyl CoA and the regeneration of the free enzyme.

(Figure adapted from Matthews *et al.*, 1999)

AdoCbl comprises a modified corrin ring surrounding a cobalt atom (figure 1.3). A protein histidine side chain acts as a sixth ligand binding to the cobalt atom in the holoenzyme (Mancia *et al.*, 1996). The adenosyl moiety is attached to the cobalt though a covalent cobalt-carbon bond. This carbon-cobalt bond has special reactivity, and is also very rare; there are only two known enzymes in animals that employ carbon-cobalt bonds in reaction mechanisms. Cytoplasmic methionine synthase uses methylcobalamin and mitochondrial MCM uses adenosylcobalamin (Kolhouse and Allen, 1977). The reaction catalysed by hMCM is also unusual for another reason; the enzyme breaks and reforms a carbon-carbon bond in the carbon skeleton of methylmalonyl CoA.

3
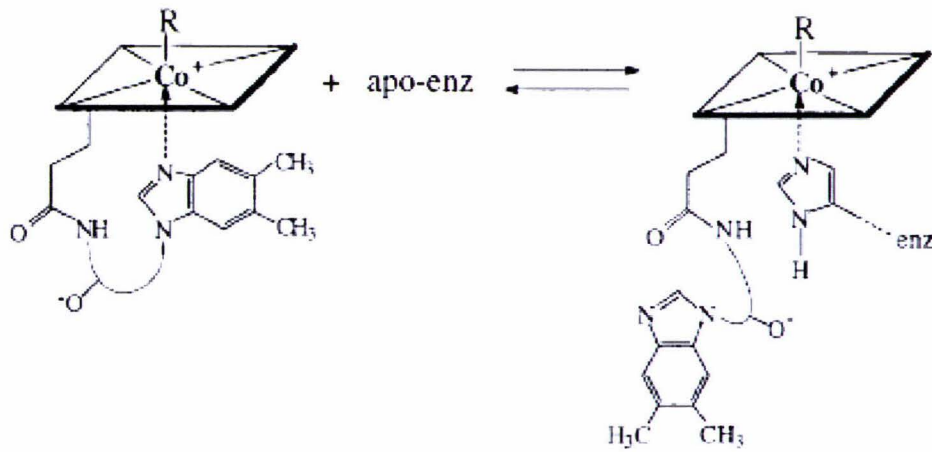
Figure 1.3. Adenosylcobalamin Binding to Bacterial _P. shermanii_ MCM. Left: 'base on' AdoCbl, separate from the enzyme. Right: 'base off' AdoCbl, bound to the enzyme through a histidine residue. The adenosyl moiety (R) provides the carbon atom for the carbon-cobalt bond. The four nitrogen's of the corrin ring coordinate the cobalt (centre), and the histidine of MCM (below) displaces the lower axial ligand of cobalt.

(Figure adapted from Tollinger _et al._, 2001)

1.3 The Human Methylmalonyl CoA Mutase Gene

The _mut_ locus on chromosome 6 contains 13 exons, and (in total) the gene is 35 kbp long (Ledley _et al._, 1988). hMCM is constitutively expressed and encodes a protein with a 32 amino acid mitochondrial leader sequence that is cleaved to produce the mature 78 kDa peptide. (Nham _et al._, 1990). The active human enzyme purified from liver is an α_2 homodimer of approximately 150 kDa, and each subunit binds a molecule of adenosylcobalamin, the cofactor. Adenosylcobalamin is synthesised from vitamin B_{12} _in vivo_ as shown in figure 1.1 (Fenton and Roseburge, 1995; Banerjee and Chowdury, 1999).
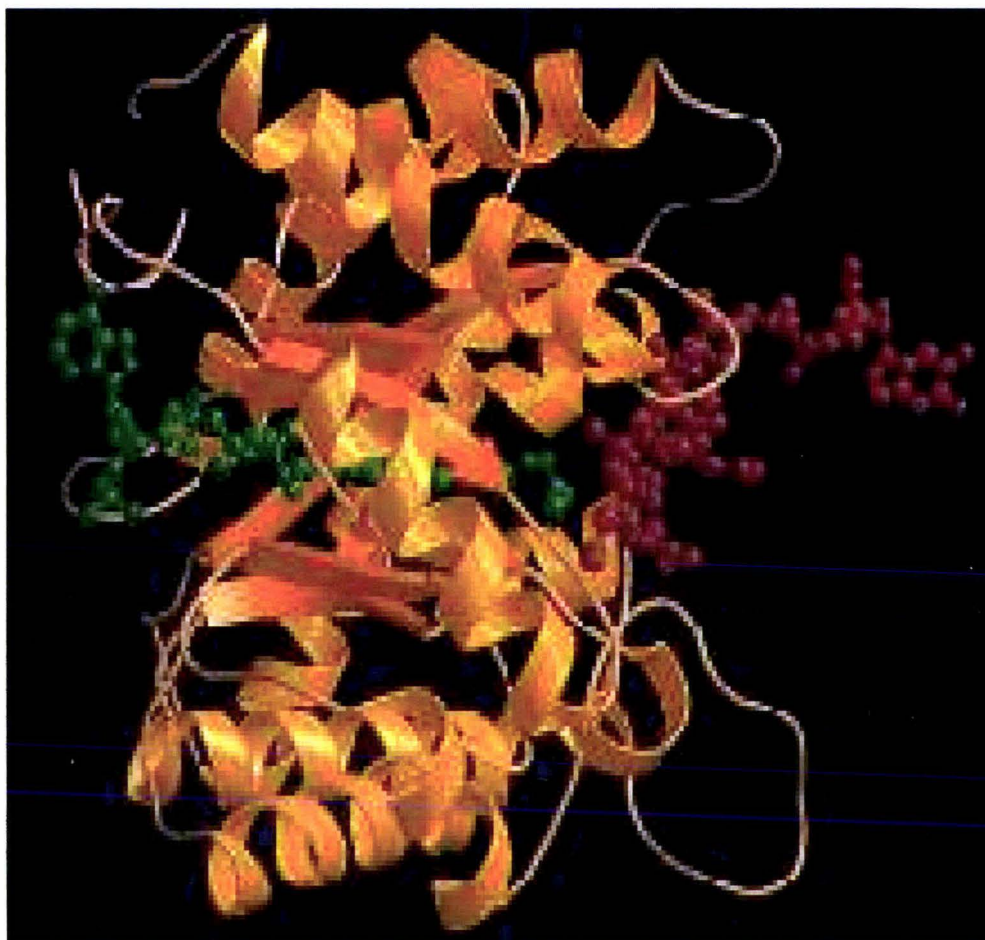
Figure 1.4 Structure of the *Propionibacterium shermanii* methylmalonyl CoA mutase.
The substrate, methylmalonyl CoA (green) and the co-factor, adenosylcobalamin (red), are
shown bound in the active site.

(Mancia *et al*, 1996)

The precursor protein sequence contains a C-terminal cobalamin binding domain in residues 578-750. The $(\beta\alpha)_8$ barrel formed by residues 87-416 contains a methylmalonyl CoA binding site. Residues at the N-terminus may be important in the dimerisation of the two subunits (figure 1.5). A loss in homology between the bacterial and human enzymes makes predictions about the structure and function of this region less certain (Thoma and Leadley, 1996). The MCM $\alpha\beta$ heterodimer of the *Propionibacterium shermanii* homologue has been cloned and expressed in *E. coli* (McKie *et al.*, 1990), and the crystal structure has been solved to 2Å resolution, see figure 4 (Mancia *et al*, 1996). The α subunit of *P. shermanii* MCM shows a remarkably high (61%) amino acid identity to the α subunit of human MCM (Leadlay and Ledley,

1989). However the β subunit, while clearly related to the α subunit shares only 34% identity with the amino acid sequence of the human gene, and its role remains enigmatic (figure 1.5).

```
        *:  .*:  .   *↓:    *   .    **    ::* .*     .
H MLRAKNQLFLLSPHYLRQVKESSGSRLIQQRLLHQQQPLHPEWAALAKKQLKGKNPEDLI
B --------------------MSTLPRFDSVDLGNAPVPAD---AARRFEELAAKAGTGEA

   *.*.* *.: .*:.:    . **   :   .*: **.:***.***;**************.:*
WHTPEGISIKPLYSK--RDTMDLPEELPGVKPFTRGPYPTMYTFRPWTIRQYAGFSTVEE
WETAEQIPVGTLFNEDVYKDMDWLDTYAGIPPFVHGPYATMYAFRPWTIRQYAGFSTAKE

   ** **: *: ***:********.************ ************:: * ; ** ***
SNKFYKDNIKAGQQGLSVAFDLATHRGYDSDNPRVRGDVGMAGVAIDTVEDTKILFDGIP
SNAFYRRNLAAGQKGLSVAFDLPTHRGYDSDNPRVAGDVGMAGVAIDSIYDMRELFAGIP
```

Figure 1.5 An Alignment of Human & Bacterial Amino Acids at the N Termini of MCM. There is 61% identity overall between the α subunits of the bacterial and human enzymes, however the identity in the 59 N-terminal amino acids drops to 22%. Red arrow indicates the cleavage site of the mitochondrial leader sequence within the human enzyme. The 59 N terminal amino acids are shown here with a black line. **H** is the human MCM sequence, and **B** the bacterial sequence (of the α subunit from *P. shermanii*)

Mutations in the methylmalonyl CoA mutase gene (*mut*) cause *mut* methylmalonic acidemia (*mut* MMA). The disease *mut* MMA has an estimated incidence of 1/30000 to 1/50000 live births, and symptoms include metabolic acidosis, lethargy, dehydration, vomiting, and neurological problems (Fuchshuber *et al.*, 2000). *Mut* MMA patients can be distinguished from those suffering from other forms of MMA by their non responsiveness to B_{12} therapy; other forms of MMA are often due to mutations in enzymes involved in the synthesis of coenzyme B_{12} and so can be corrected by B_{12} therapy. There are two types of *mut* MMA, *mut^0* and *mut$^-$*. The *mut$^-$* form of MMA is milder, and is characterised by a reduced hMCM activity. Most of the mutations associated with *mut$^-$* MMA are found in the AdoCbl binding domain, and so by increasing the AdoCbl concentration some of the enzyme activity is recovered (Adjalla *et al.*, 1998). *Mut^0* MCM, however, is completely inactive and this form of *mut*MMA is severe and often fatal (Peters *et al.*, 2002). So far 81 different mutations in the hMCM 'structural gene' have been identified. Three are found at higher frequencies in some populations (~1% of births are carriers)   E117X in Japanese, G717V in African

Americans and N219Y in a French/Turkish population (Aquaviva *et al.*, 2001, Aquaviva *et al.*, 2005).

The known mutations in hMCM causing *mut*MMA are more commonly missense or nonsense nucleotide substitutions, and many unidentified mutations may occur in both the structural gene and promoter region. (Peters *et al.* 2002). Cobalamin affinity and enzyme kinetics has been characterised for most of the known mutations from *mut* MMA patients, has lead to a better understanding of the structure-function relationship of the human enzyme (Janata *et al.*, 1997; Andrews *et al.*, 1993; Crane *et al.*, 1991).
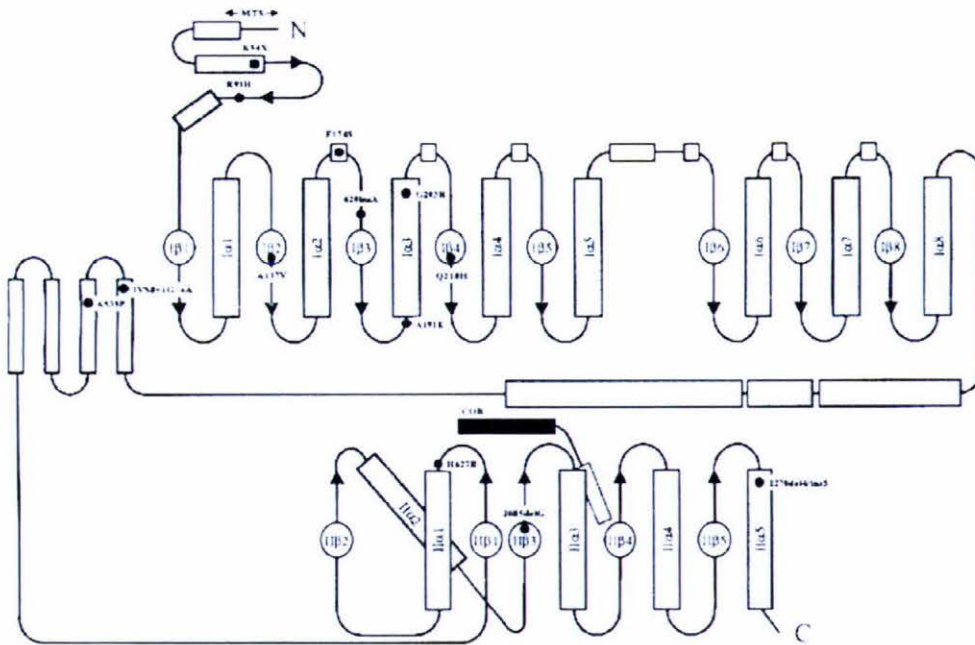
Figure 1.6 Proposed Topology Model of hMCM. The positions of 30 disease causing mutations are indicated as black filled circles. Solid black bar is the cobalamin co-factor.

(Adapted from Fuchshuber *et al.*, 2000)

The topology model in figure 1.6 shows the known mutations in the C-terminal cobalamin binding region, G623R, G626C, G630E, G703R are substitutions of glycines thought to be interacting with the cofactor. These residues surround the H627 that binds as the sixth ligand to the cobalt of the AdoCbl. Mutation at G648 or G717 also disrupts the structure orientation of the H627, affecting the binding of the co-factor, and consequently these are both highly conserved residues. Some other mutations are found

at the putative interface of the two dimers at the N-terminal domain (R93H), and in the channel to the active site (W105R).

1.4   The Novel Gene MMAA

In 2002 a methylmalonic academia patient with a novel mutation was reported. The mutation was in a novel gene that was called MMAA (**m**ethyl**m**alonic **a**cademia linked to the cbl**A** complementation group). The gene was mapped to chromosome 4q31.1-2, and comparison of the genomic and cDNA sequences revealed 7 exons, and a mitochondrial signal sequence. It was a highly conserved protein between organisms and homologues are found in archea, eubacteria and eukaryotes. The gene was assigned a putative function as a transport protein for vitamin B_{12}, based on sequence alignments and analysis of prokaryotic gene arrangements (Dobson *et al.*, 2002). Subsequently a different group has backed up these findings, with the discovery of seven novel mutations in MMAA, all in patients responsive to B_{12} therapy (Yang *et al.*, 2004).

In contrast to these findings, mutation of a gene with very high sequence homology (*meaB*) in the bacteria *Methylbacterium extorquens* AM1, caused a loss of MCM activity. MCM activity was not recoverable by the addition of B_{12} which would be expected if indeed *meaB* encoded a protein for a B_{12} transport. In pull down assays a complex of MCM and the *meaB* protein formed. *meaB* was also found to be required for MCM activity *in vitro*. The authors suggest that the function of the *meaB* protein is to prevent   the inactivation of MCM during catalysis. This may arise by a stabilising effect on the dimer form of the mutase, or protection of the enzyme from attack by oxygen, water or highly reactive radical intermediates (Korotkova and Lidstrom, 2004). Whether this gene *meaB* is a functional homologue of *MMAA* remains unclear, more detailed biochemical analysis is required.

1.5   Past Recombinant Expression Systems of hMCM

Earlier attempts to express hMCM in *E. coli* resulted in the production of inclusion bodies, which are insoluble aggregates of misfolded protein. This is common in *E. coli* with highly expressed recombinant proteins (reviewed by Baneyx and Mujacic, 2004).

Refolding of these inactive inclusion bodies from *E. coli* has been attempted with the use of rapid dilution, dialysis, detergent assisted refolding, size exclusion chromatography and use of chaperonin-assisted refolding. Refolding efficiency reached 1% soluble active enzyme, using the rapid dilution technique. This suggests that the polypeptide produced is folded incorrectly, but otherwise fully active when expressed in *E. coli*, (Hayes. 1998). This is not always so, some recombinant proteins may require posttranslational modification that the *E. coli* is not able to perform. Eukaryotic proteins may also expressed as soluble, but not biologically active in *E. coli*.

In an attempt to correct the misfolding *in vivo* of hMCM expressed in *E. coli*, cells were co-transformed with hMCM expression plasmid and pGroESL, a pACYC-derived plasmid that overexpresses the *E. coli* chaperonins GroEL and GroES chaperones (Goloubinoff *et al* 1989). However the co-expression of chaperonins had little effect on the misfolding of the protein (Janata *et al.*, 1997).

There is an increase in yield of soluble hMCM, observed when expressed with a Trx tag, in *E. coli* (Janata *et al.*, 1997). This soluble protein is not being recovered from the insoluble pool of protein, but rather appears to be from a reduction in protein being degraded (figure 1.7).
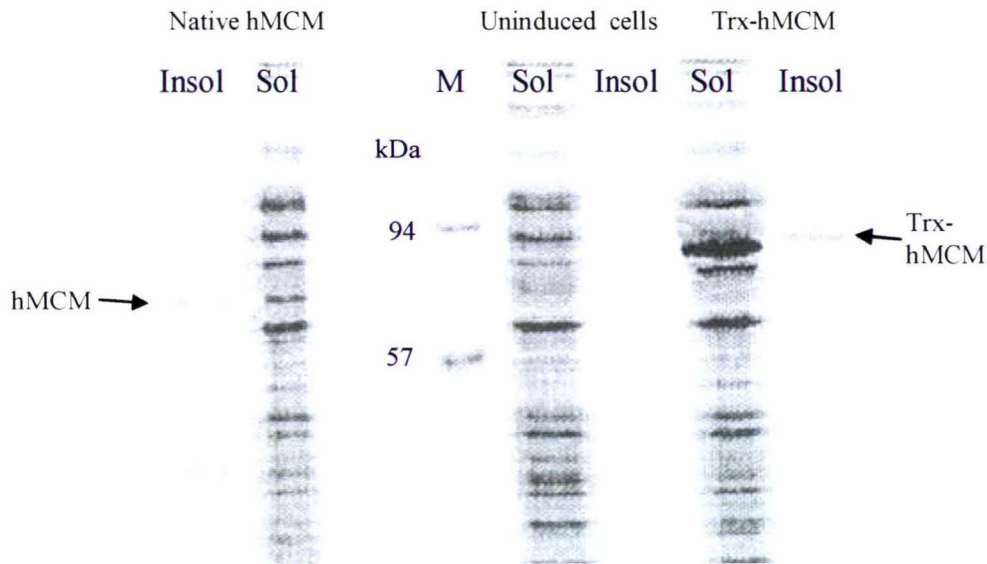
Figure 1.7 SDS PAGE Gel Showing the Expression Systems of Janata et al, (1995). hMCM expression without a solubility tag (native hMCM) is compared to expression with a Trx tag (Trx hMCM). The Trx solubility tag drastically increases the yield of soluble protein without reducing the amount of insoluble protein. Insol : Insoluble fraction. Sol: Soluble fraction. M: Molecular size markers

(Figure adapted from Janata *et al.*, 1997)

The increase in yield recombinant protein with a Trx tag, (figure 1.7), could be a result of two changes that the Trx makes to the polypeptide. One possibility is increasing the half-life due to N-terminal amino acid change. It is also possible that instead of, or in addition to, the Trx, may also reduce the degradation of aberrantly folded protein. The stability of the hMCM is improved thus reducing the aberrantly folded hMCM, and so also reduces the protease degradation of this aberrantly folded hMCM. The authors do not comment on either of these possibilities (Janata *et al.* 1997).

In humans the hMCM pre-protein is targeted to the mitochondria where the signal sequence removed, to produce an N terminal Leu on the polypeptide (table 1.1, protein 1). In *E. coli*, however, a leucine at the N-termini of a polypeptide has a very short half life of 2 mins (Tobias *et al.* 1991). Expression of the native hMCM polypeptide, even with a Met before the Leu, in *E. coli* would be expected to result in the production of a short-lived protein. This high turnover rate, ultimately leads to a lower yielding *E. coli* expression system. Janata *et al.* (1997) expressed the hMCM in two systems, without

and with the Trx tag (table 1.1, proteins 2 and 3 respectively) showing a remarkable increase in protein yield with the Trx tag.

Source	N terminal amino acids	After Met processing	Halflife in *E. coli*
1. Native hMCM:		Leu His ~	
2. Janata, hMCM	Met Leu His ~	Leu His ~ (15%)	2 mins
3. Janata, thio hMCM	Met Ser ~T~ Leu His ~	Ser TAG Leu His ~(84%)	10 hours

Table 1.1 N terminal amino acid and the turnover rate of proteins in *E. coli.* From left, the source, the sequence of the first amino acids, the sequence after the N-terminal Met is removed and the predicted half life in *E. coli.*.

The effect of a post induction temperature, on hMCM solubility with and without a Trx fusion on the N-terminus was critical (Janata *et al.*, 1997). With the Trx-fusion at 37 °C most of the protein was in the form of inactive inclusion bodies, however at 12 °C the majority of the protein is correctly folded, soluble and fully active. The Trx-fusion expression system in *E. coli* GI698 at 12 °C, resulted in high levels of soluble recombinant protein, however the Trx tag was not able to be separated from the target protein. Enterokinase cleavage was inefficient, but at higher protease concentrations and longer reaction times, degradation of the hMCM was observed. A specific activity of 23-26 U/mg was recovered despite the Trx fusion remaining on the hMCM.

Human fibroblasts and *S. cerevisiae* have been used to express mainly mutant enzymes, for use in enzyme activity studies of the mutations. Low activity can be detected using a ^{14}C-propionate incorporation assay (Aquaviva *et al.*, 2001, Crane *et al.*, 1992, Peters *et al.*, 2002). This requires very little active enzyme and is an effective method for diagnoses and study of *mut* MMA patients. In *S. cerevisiae* the protein was soluble; but not active and was unable to complement *mut* fibroblast cells. However, mouse MCM was expressed in the same system, and produced an active enzyme, able to compliment a *mut* deficiency in human cells (Andrews *et al.*, 1993).

1.6 Enhancing solubility of Recombinant Proteins in *E. coli.*

1.6.1 Background

If the aggregates are forming during cell lysis, changing the lysis buffer could prevent partitioning of the recombinant protein into the insoluble fraction (Bondos *et al.* 2003). Recombinant expression of hMCM in *E. coli* results in the formation of aggregates *in vivo;* and the inclusion bodies can be observed by a light microscope (Mark Patchett, *pers. Comm.*). Consequently the expression conditions must be altered to solublise the hMCM protein in *E. coli* (Baneyx and Georgiou 1990).

The formation of expressed protein aggregates is often rapid and irreversible. The kinetics of the aggregation may be significantly faster than the folding kinetics to the native state, as once the aggregation process is initiated, the aggregates will form rapidly (Cellmer *et al.* 2005 and Goldberg *et al.* 1991). Solubilising the aggregates is accompanied by a new problem; proteases now have access to the recombinant protein as is it now not sequestered in the inclusion bodies (Cheng et al.. 1981). To prevent inclusion bodies from forming and degradation of the soluble recombinant protein, some expression conditions can be changed: e.g. the rate of expression can be slowed, decreasing the local concentration of folding intermediates. The redox potential of the *E. coli* cytoplasm can be changed or the protein can be targeted to the periplasm which contains many chaperones. Solubility tags can be used or a protease deficient *E. coli* cell line used.

1.6.2 Periplasmic Expression

Secretion of recombinant proteins to the periplasm in *E. coli* has been show to improve the solubility proteins. A signal sequence is placed in the gene, most commonly the signal sequence from the MBP which is located in the periplasim of *E. coli* (Loo et al, 2002; Witholt *et al.*, 1976).

1.6.3 Temperature

The macromolecule concentration in cells is in excess of 300 mg/ml, so a newly synthesised protein will not fold in isolation, but rather while interacting with a number of other molecules that may influence the way the protein folds (Dobson, 2001). Decreasing the temperature of the *E. coli* culture during induction and expression has been used extensively to slow the recombinant protein expression rate, enhancing solubility of recombinant proteins. This is thought to be due to the reduction in the concentration of partially folded proteins at any given time, reducing the chance of two partially folded proteins associating and initiating aggregation (Schein and Noteborn, 1988; Janata *et al.*, 1997).

The bacterial methylmalonyl CoA mutase (of *P. shermanii*) was expressed in *E. coli* K38, in a vector containing the two MCM enzyme subunits. Expression at 37 °C resulted in insoluble protein, but when induction temperature was lowered to 30 °C, resulting in the expression of soluble active methylmalonyl CoA mutase (McKie *et al.*, 1990

1.6.4 *E. coli* Strains

E. coli BL21 (DE3) is deficient in the Lon and OmpT proteases. Lon is a cytoplasmic protease responsible for the degradation of proteins possessing a non native structure (Gottesman, 1989). OmpT is an outer membrane protease that cleaves dibasic amino acid sequences (Grodberg and Dunn 1988). These proteases can contribute to the degradation of some recombinant proteins during expression and purification (Baneyx and Georgiou 1990). This strain is also lysogenic for a λ prophage that contains an IPTG inducible T7 RNA polymerase gene, allowing expression from vectors containing a T7 promoter (Studier, *et al.*, 1990).

E. coli strain Rossetta-gami 2 contains mutations in genes for thioredoxin reductase and glutathione oxidoreductase (*trxB* and *gor*). Thioredoxin reductase is part of an active process preventing disulfide bonds from forming in the cytoplasm of *E. coli*. Glutathione oxidoreductase is involved in the oxidative stress response of *E. coli* and loss of activity changes the redox potential of the cytoplasm (Becker-Hapak and

Eisenstark 1995; Davis *et al.*, 1982). These mutations improve the folding of many eukaryotic proteins expressed in *E. coli* in particular proteins containing disulfide bonds (Derman, *et al.*, 1993, Cassland *et al.*, 2004).

The *E. coli* strain Rosetta-gami 2 also contains a mutation in the lac permease *(lacY)* gene, giving the Tuner™ genotype. These cells allow a more even entry of IPTG through the cell membrane. More even IPTG entry causes induction to occur in an IPTG concentration-dependent manner throughout the culture, allowing stricter control of promoter induction, and the use of very low levels of IPTG to slow expression of recombinant protein.

Amino acids can be encoded for by more than one codon in the mRNA. Different organisms will favour the use of some codons for an amino acid, resulting in a bias in the levels of tRNA for particular codons (see the appendix). The codons used infrequently in a certain organism are called rare codons, and are translated inefficiently, due to the low level of corresponding tRNA. The *E. coli* strain Rosetta Rosetta-gami 2 have extra copies of tRNA genes *argU, argW, ileX, glyT, leuW, proL, metT, thrT, tyrU,* and *thrU* on pRARE, a pACYC184-derived plasmid (Novy *et al.*, 2001). These genes encode the tRNA for codons not abundantly found in *E. coli* mRNA; AUA, AGG, AGA, CUA, CCC, and GGA allowing the ribosome to translate the eukaryote mRNA more efficiently. Translation pausing of the ribosome, at regions of mRNA that contain rare codons may result in premature translation termination, translational stalling, translation frameshifting and misincorporation of amino acids. These events can lead to low yield and misfolding of recombinant proteins (Looman *et al.* 1987; Makhoul and Trifonov, 2002; Kurland and Gallant, 1996; McNulty *et al.*, 2003). The codons of the hMCM sequence were analysed at the database on the website http://www.kazusa.or.jp/codon/ (see appendix). A total of 25 codons are present in hMCM that the tRNA is found below 0.5% frequency in *E. coli*.

E. coli ER2566 is an expression strain provided by New England Biolabs with the vectors pTYB4 and pTYB11, and is the recommended expression strain for both these vectors. (Lüttkopf *et al.*, 2001). The cells *E. coli* ER2566 cells are deficient in Ion and OmpT proteases and contain the λ prophage with an IPTG inducible T7 RNA polymerase gene.

1.6.5 Solubility Tags

Many proteins formerly used as affinity purification or detection tags were also found to enhance expression levels and solubility of the target protein, so are now also referred to as solubility tags. A solubility tag is a second protein fused in the same open reading frame as the recombinant target protein (figure 1.8). The open reading frames of both will now be expressed as one polypeptide, and after folding and purification the polypeptide can be separated to yield two polypeptides i.e. the target protein and the solubility tag see section 1.7 on 'purification and processing of recombinant proteins'. Sometimes (rarely) the target protein may aggregate after cleavage even if it is correctly folded as a fusion. (Sati *et al.*, 2002).
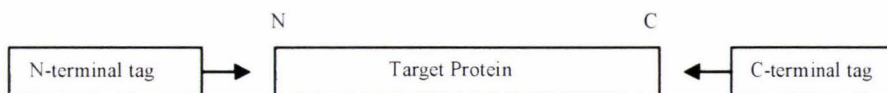
Figure 1.8 Showing Possible Positions of a Solubility Tags Relative to the Target Protein. The tag can be expressed on the C or N termini of the target protein. The N-terminus of proteins is the most common position for a solubility tag.

To be a suitable solubility partner a protein must be highly soluble in the *E. coli* cytoplasm. However, not all highly soluble proteins are affective solubility tags (Kapust and Waugh, 1999). It is also unlikely that all solubility tags act in the same way, different target and solubility tag combinations will affect the solubility of the target protein differently (Kapust and Waugh, 1999). Although solubility tags are now widely used in recombinant systems, the mechanisms by which some of them prevent or reduce aggregation are not well understood due to the lack of understanding about the formation of protein aggregates in cells.

The 40 kDa maltose binding protein (MBP) (Pryor and Leiting, 1997) has been shown to be a very effective solubility tag. This was demonstrated in a study of the effect three different solubility tags had on six different target proteins, all insoluble and inactive when expressed in *E. coli*. The three solubility tags tested were Trx, MBP and glutathione S-transferase. The effectiveness of the solubility tag was measured by the solubility and activity of each target protein. The MBP was consistently found to be the

most effective tag at solubilising the target protein (Kupust and Waugh, 1999). These results have been supported by other labs (Goulding and Perry, 2003). This led the authors to suggest that the MBP is a passive general chaperone, binding to unfolded proteins to promote folding, perhaps by preventing self association, (figure 1.9).
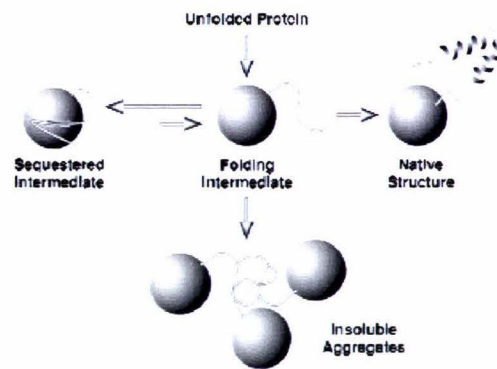
Figure 1.9 A Model of the Proposed Mechanism of MBP. The sphere represents the MBP and the line and the helix represents the unfolded and folded protein, respectively.

(Figure adapted from Kapust and Waugh, 1999)

In support of this mechanism is the ineffectiveness of the MBP as a solubility tag when fused to the C-terminus of the target protein. This arrangement results in the target protein being synthesised and folding earlier as it comes of the ribosome, (Gilbert *et al.* 2004), suggesting that the MBP must be fully folded to be an effective solubility tag (Sachdev and Chirgwin, 1998). The MBP has also been shown to interact with unfolded polypeptides *in vitro,* promoting folding of these unfolded proteins (Richarme and Caldas, 1997). Using the maltose sugar immobilised on resin, the MBP is also affective as an affinity purification system, and as an affinity tag, the MBP is typically placed at the C-termini of the target protein (Maina *et al.,* 1988; Hennig and Schafer, 1998).

The 11 kDa thioredoxin protein is a commonly used solubility tag and dramatically increases the solubility of many recombinant proteins expressed in the *E. coli* cytoplasm (LaVallie *et al.,* 1993). It can be effective placed at the N or C terminus of the target protein, varying from case to case, depending on the target protein. However it is most commonly used as a solubility tag at the N-termini of the target protein. Because thioredoxin lacks a convenient immobilisable small ligand, it is unsuitable for an 'affinity purification system', therefore thioredoxin is often used in conjunction with a

small affinity tag such as the His$_6$ tag, and this thioredoxin-His construct is referred to as the Trx tag.

1.7   Purification and Processing of the Recombinant Protein

1.7.1   Poly His Tag Purification

A poly-histidine tag is a sequence of four to ten contiguous histidines within a protein, which selectively binds to a Ni column via the co-ordination of the histidine imidazole groups to the Ni metal. The protein is eluted by changing the pH, or adding free imidazole to the chromatography buffer (Hochuli *et al.*, 1987). The chromatography resin has an attached nitrilotriacetic acid (NTA) group, ideal for binding metal ions with co-ordination numbers of six, such as Ni and Co (Hochuli *et al*, 1988). The His tag can be placed at the N or C termini of the protein, depending on the structure of the target protein. Purification of his-tagged proteins can also allow purification under denaturing conditions. However, proteins with a metal cofactor bound are not recommended, as the potential cofactor- NTA resin interaction may interfere with the stability of the protein (Reviewed by Terpe, 2003).

1.7.2   The Chitin Binding Domain

The 5 kDa chitin-binding domain is the 51 C-terminal amino acids of chitinase A1 from *Bacillus circulans* WL-12 (Watanabe *et al.*, 1994). The chitin binding properties of the domain are exploited, and purification is achieved by using a column with immobilised chitin (Chong *et al.*, 1997).

1.7.4   Site specific Proteases

Site specific proteases recognise a certain sequence of amino acids, cleaving the polypeptide. This has been exploited to allow the separation of affinity/solubility tags from the target protein, (figure 10).
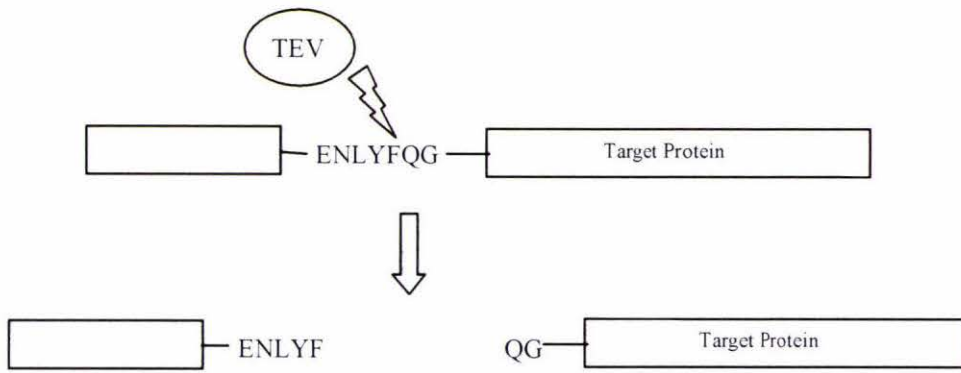
Figure 1.10 Protease Cleavage of Solubility Tags. A sequence of amino acids that are the target for a protease (in this example TEV) can be inserted between the target protein and the tag. Incubation with the protease separates the tag from the target protein. The protease must have access to the sequence so it must be on the exterior of the proteins structure.

Enterokinase is a site specific protease that cleaves after lysine at its recognision site Asp-Asp-Asp-Asp-Lys, and is frequently used to separate tags from the target protein (Sun *et al.*, 2005). In some target protein-tag fusions the enterokinase cleavage site is not accessible and cleavage can be inefficient or completely prevented. In the expression system of hMCM-Trx, inefficient enterokinase cleavage occurred, suggesting that the protease was unable to access the cleavage site (Janata *et al.*, 1997).

The Tobacco Etch Virus Protease (TEV) protease is a 48 kDa site specific protease, from the tobacco etch virus (Carington and Dougherty, 1988). His-tagged TEV can be expressed and purified in *E. coli* making it a cheap and convenient source of protease. A TEV site, (ENLYFQG), has been engineered into the pET32a vector (Rosemary Brown, *Pers. Comm.*), to allow removal of the Trx solubility tag by digesting the purified protein complex with the recombinant TEV protease. This can be used as an alternative to the enterokinase; the differences between the two proteases may allow the TEV to effectively cleave the Trx tag. In addition to changing the protease, the sequence immediately either side of the site is also different which may allow better access for the TEV.

1.7.5 Inteins

The use of inteins eliminates the need to incubate the recombinant protein with a protease to cleave the target protein from purification and/or solubility fusion tags. Conventional proteases can have drawbacks as seen in the earlier section 1.3.3 'site specific proteases'. Access to the recognition site, protease sites in the target protein and further purification are required. Often addition or change of amino acids at the N or C terminus is necessary to generate the target sequence of the protease, resulting in final protein product containing several additional or changed amino acids and this can affect the protein activity and folding. The protease may require incubation at elevated temperatures for activity, which can degrade or denature the target protein (Janata *et al.*, 1997; Humphries *et al* 2002).

Inteins perform a post translational modification, excising an internal polypeptide sequence (intein), and ligating the flanking sequences (exteins) with a peptide bond (figure 1.11). Key conserved residues have been identified by amino acid sequence alignment and then functional studies of the effects of point mutations on cleavage to determine their role in the splicing reaction (Xu and Perler, 1996).
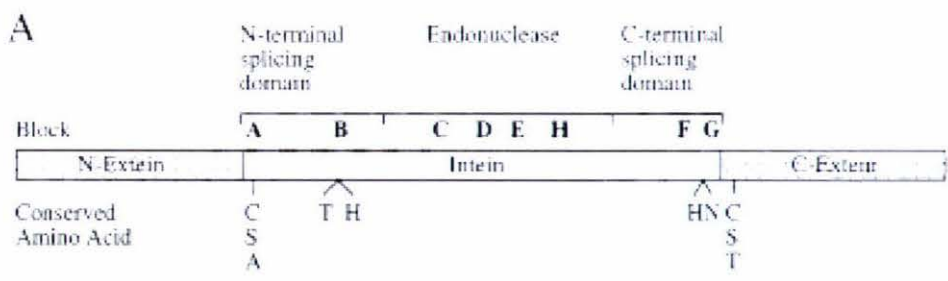
Figure 1.11 Domains and Conserved Residues in Inteins. Blocks A and B are important in the N-terminal splicing, and blocks F and G are involved in the C-terminal splicing. The region C to H can contain an endonuclease in some inteins. Also note the arrangement of the intein, N-extein and C-extein relative to each other. Shown below are the highly conserved amnio acids.

(Figure adapted from Evans and Xu, 2001)

The mechanism involves a peptide bond within the protein being attacked and broken resulting in the formation of a thioester intermediate. This mechanism is supported by the formation of a branched intermediate with two N-termini (figure 1.11) (Xu *et al.*

1993). Intein activity may be compared to that of the spliceosome that removes introns from RNA (Evans and Xu, 2002). Comparisons can also be made with protease activity, as the peptide bond is broken in a similar way, using a mechanism not unlike that of the 'catalytic triad' that many proteases use. The reaction is also analogous to the native peptide ligation reaction, used for assembling small synthetic peptides into larger synthetic proteins (Dawson *et al.*, 1994)
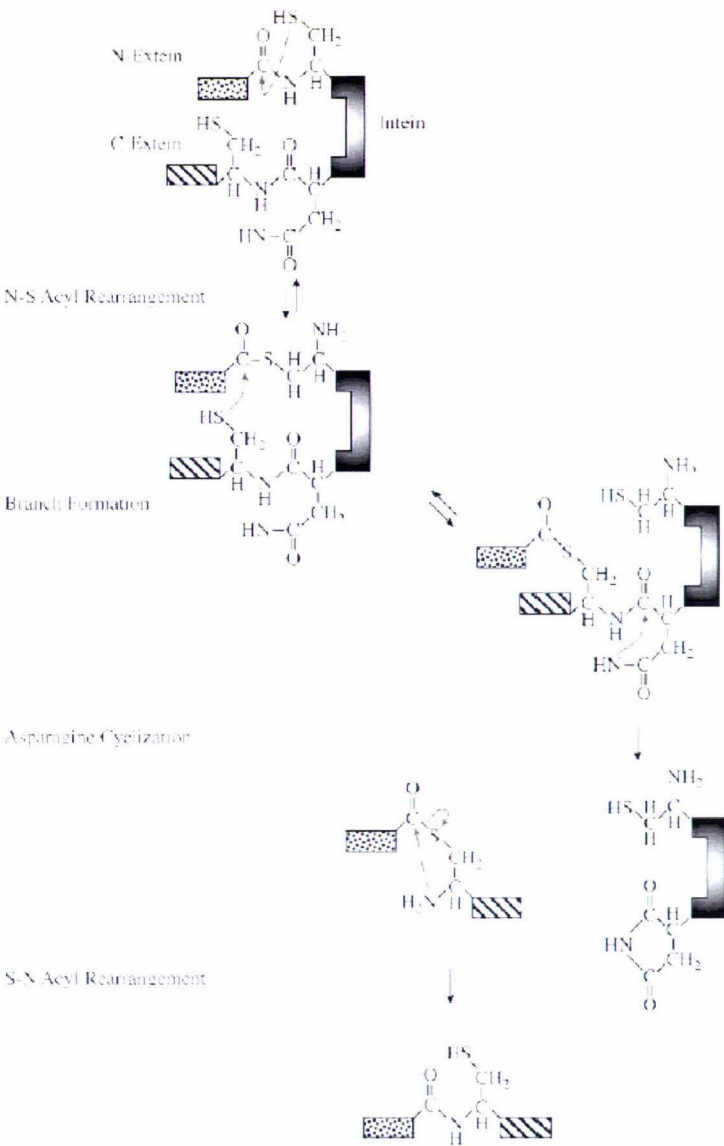
Figure 1.12 Mechanism of the Intein *Sce* VMA1 Splicing Reaction. Attack of the intein N-terminal cysteine sulfhydryl group initiates a N-S acyl rearrangement, which generates a thioester bond between the intein N-terminal cysteine residue and the N-extein C-terminal residue. Trans(thio)esterification of the N-extein acyl group to the C-extein cysteine forms the branched intermediate. Attack of an intein C-terminal asparagine excises the intein , and the reaction is completed by a spontaneous S-N acyl rearrangement, regenerating a native peptide bond between the N- and C-extein segments.

(Figure adapted from Evans and Xu, 2001)

The intein *Sce* VMA1 is from the yeast *Saccharomyces cerevisiae* and was discovered in 1990 (Hirata *et al.,* 1990; Kane *et al.,* 1990). Researchers at New England Biolabs then removed endonuclease domain (figure 1.13) at the gene level and mutated key catalytic residues to prevent the exteins religating. Although inteins lacking the

endonuclease domain (mini inteins) are now known to occur naturally, the *Sce* VMA1 intein without the endonuclease domain was artificially created, before mini inteins were discovered in nature.
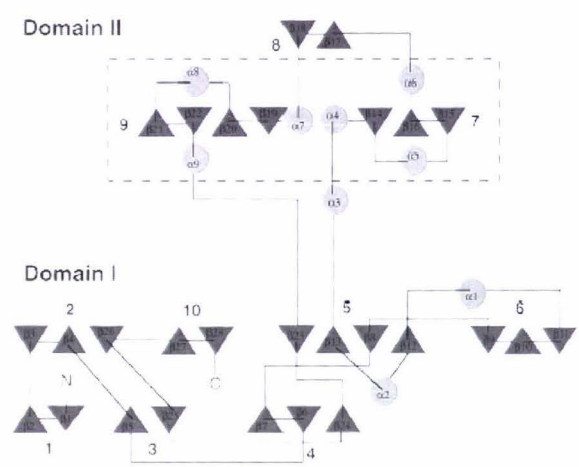
Figure 1.13 The Domain Structure of the Native *Sce* VMA1 Intein. The circles and triangles represent α helices and β strands, respectively. Numbers identify the ten different β sheets. Domain I is the nuclease domain identified by the DNA binding motifs, and domain II is thought to contain the protease activity.

(Figure adapted from Duan *et al.* 1997)

The changes to this intein allowed researchers at New England Biolabs to construct a series of vectors that contain a CBD (Watanabe *et al.*, 1994) as a purification tag, which can be removed by the *Sce* VMA1 intein that is located between the tag and the protein of interest. The CBD is inserted into a loop of the intein, and does not affect its activity. The vector pTYB4 encodes a splicing defective mutant *Sce* VMA1 intein (Asn454Ala) that only performs the N-terminal junction cleavage. The C-terminal splicing activity has been lost (Chong *et al.* 1996). Another vector in the series, pTYB11, encodes another splicing defective mutant *Sce* VMA1 intein (His453Glu) mutation at the penultimate amino acid resulting in defective C-terminal splicing activity (Chong *et al.* 1996). This allows the solubility tag to be removed from the N termini of the target protein, (figure 1.14).
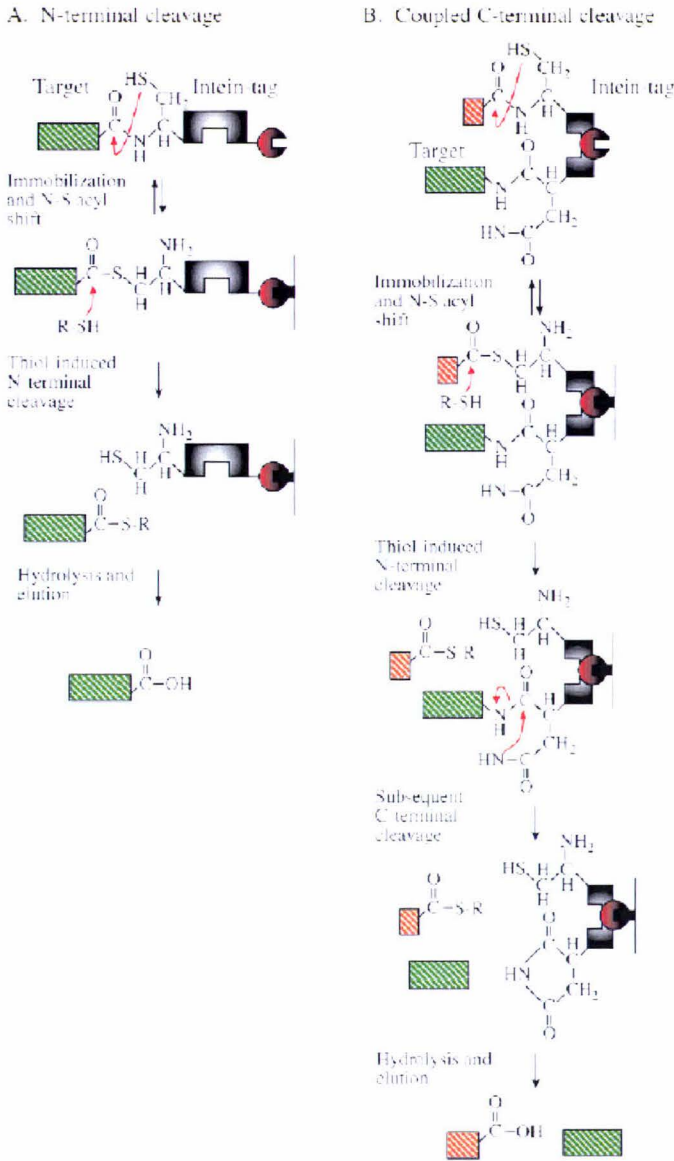
Figure 1.14 Mechanisms of Intein-Based Protein Purification.

(A) N-terminal cleavage, by the modified intein. The intein-tag is fused to the C-terminus of the target protein at the gene level. The target is released by cleavage of the bond between the target protein and the intein. This is the modified intein present in the pTYB4 vector.

(B) Coupled C-terminal cleavage by a different modified intein. This type of intein-tag allows expression and purification of the target protein with an N-terminal affinity or solubility tag, This is the modified intein present in the vector pTYB11.

(Figure adapted from Evens and Xu, 2002)

The vectors pTYB4 and pTYB11 allow the separation of the target protein under reducing conditions (figure 1.14). The target protein, intein and any solubility or purification tags are bound to a purification column or beads though an affinity tag. The

other proteins from the expression system cell lysate are washed off in a mild buffer. The target protein is then removed with a reducing agent (e.g. DTT or β-mercaptoethanol) that induces the intein to self cleave. The target protein is eluted from the column leaving the purification/solubility tags bound (Perler, 2000). Dialysis is necessary to remove a small peptide generated during the intein cleavage of proteins expressed from the vector pTYB11.

The intein leaves no amino acids on the target protein, however sometimes 1-2 amino acids must be changed at the cleavage site on the target protein, to ensure that cleavage occurs when desired, and that *in vivo* cleavage is avoided. The flexibility of the amino acids at the intein-target protein boundary means it is often possible to use amino acids with similar properties to replace the original native amino acids. Amino acids at the protease sites are very specific sequences, with little variability allowed. These amino acid changes, necessary to generate the protease site, can result in undesirable changes to the amino acids at the termini of the target protein. These amino acids may change the native sequence of the target protein and this may affect the structure and/or function of the target protein.

1.8   Chapter Summary

Studies of human methylmalonyl CoA mutase (hMCM) are required to better understand the unusual carbon skeletal rearrangement this enzyme performs, the catalysis involving a B_{12} derived cofactor, and the disease methylmalonic acidemia. Homology models based on the crystal structure of the α-subunit of a bacterial orthologue have provided a structural model. However, the decrease in homology between the bacterial and human proteins at the N-terminus results in inaccurate models of this putative dimerisation domain. Structural studies are required, and to provide a source of protein for these studies, an *E. coli* expression system is proposed.

Expression of native hMCM in *E. coli* results in insoluble inclusion bodies. A Trx solubility tag increased the yield of soluble protein but did not allow separation from the tag with the enterokinase site located between the Trx and hMCM. An alternative to protease separation of solubility tags is an intein cleavage system. Inteins may circumvent the problems associated with conventional *in trans* protease removal of tags.

1.9   Aims of this Project

This work aimed to develop a novel vector to enable expression of soluble active hMCM for structural studies. This overall goal can be subdivided into three aims:

A. Design and construct four novel expression vectors containing a solubility tag and an intein mediated removal of that tag.

B. Clone hMCM into each of these novel vectors.

C. Test the solubility of protein expressed from each of these vectors with the insoluble protein hMCM, and determine the success of the intein mediated cleavage.

2 Materials and Methods

2.1 Materials

2.1.1 Bacterial Growth

The *E. coli* heterologous expression strains TG1 and ER2566, were purchased from New England Biolabs, MA, USA. The *E. coli* strains BL21 (DE3) and origami B were purchased from Novagen® an Affiliate of Merck NZ Limited, Palmerston North, New Zealand. For the genotypes of the *E. coli* strains, see appendix. Luria Bertani (LB) base was purchased from Invitrogen™ life technologies, Invitrogen NZ limited, Auckland, New Zealand.

2.1.2 DNA Preparation and Analysis

The protein expression vectors pTYB4, pTYB11, pMAL.c2g were obtained from NEB. The protein expression vector pET32a was purchased from Novagen® an Affiliate of Merck NZ Limited, Palmerston North, New Zealand. Restriction enzymes and mung bean nuclease were purchased from New England Biolabs and were used with the supplied buffers according to manufacturers' instructions. Bovine serum albumin was also purchased from New England Biolabs. The modified pET32a vector (pET32a.TEV) was a kind gift from Rosemary Brown Massey University, Palmerston North, New Zealand. The T4 DNA ligase and 1 kb plus DNA ladder were purchased from Invitrogen™ life technologies, Invitrogen NZ limited, Auckland, New Zealand.

2.1.3 Chemicals

The 40% acrylamide and NN'-methylenebisacrylamide (29.1:0.9) in water, were purchased from Novagen® an Affiliate of Merck NZ Limited, Palmerston North, New Zealand. The ethidium bromide, Coomassie brillant blue (R-250), agarose, Tris HCl was purchased from Invitrogen™ Life Technologies, Invitrogen NZ limited, Auckland, New Zealand. Tetramethylethylene di'amine (TEMED) and ammonium persulfate were purchased from BioRad Laboratory Supplies, Poole, England. Ampicillin, tetracycline,

EDTA, IPTG, DTT, primers, mineral oil for PCR, SDS, imidazole and adenosycobalamin from Sigma Chemical Company, St Louis, MO, USA.

2.1.4 General

Mini-protean® II cell gel apparatus and sub-mini gel apparatus was purchased from BioRad Laboratory Supplies, Poole, England.

All water used was from a Barnstead Nanopure II water purification unit (with a minimum resistivity of 18 MΩ.cm^{-1}), and is referred to as milli-Q.

All other reagents were of the highest grade available.

2.2   Methods

2.2.1   Agarose Gel Electrophoresis Andrews (1986); Sambrook (1989)

The 0.8% agarose gels used for DNA electrophoresis were prepared as follows.0.36g of Agarose powder is added to 45 mL of 1x TAE (Made as a 50X stock: Tris 2M, 100 mL of 0.5 M EDTA, pH 8.0, glacial acetic acid 5.71%, made up to 1 L), to make an 0.8% gel. This was heated in a micro wave oven to dissolve the agarose powder. The agarose solution was then cooled to ~55 °C, poured into the gel tray, and ethidium bromide added to 0.2 μg/mL. Once the agarose solution has set (~30 minutes at room temperature), the comb was removed, the gel was submersed in 1x TAE. DNA samples were mixed 5:1 with loading solution (0.25% bromophenol blue, 40% sucrose) and 2-5 μL loaded into sample wells. All gels were run with 5 μL of a 1 kb Plus DNA ladder in the first lane in a Sub-mini gel tank for approximately 1 hour at 60-90 V.

2.2.2   DNA Quantification

DNA was quantified by gel electrophoresis and compared to a set of mass standards. Each standard is composed of linearised pBluescript SK-II plasmid DNA of 10 ng/5 μL, 20 ng/5 μL, 50 ng/5 μL and 100 ng/5 μL. Quantification gels (0.8% agarose) were run with a 1 kb Plus ladder, and 5 μL of four standards loaded along side the various dilutions of the unknown sample to elucidate the amount of unknown sample its band density is compared to that of the standard.

2.2.3   DNA Purification

Direct purification of PCR products and DNA digested with restriction enzymes was performed with a Roche® High Pure PCR Product Purification Kit, according to the manufacturers' instructions.

Plasmids were purified from 5 mL *E. coli* cultures using a Wizard® Plus SV Minipreps DNA Purification System (Promega Corporation), according to the manufactures protocol.

2.2.4   Restriction Digests

Diagnostic Restriction Digests were set up as follows:

2 μL	DNA (a total of 50 ng – 500 ng)
7 μL	water
1 μL	recommended enzyme buffer (10x concentrated)
1 μL	restriction endonuclease enzyme (1-10 units)

The reactions were incubated at 37 °C (or 25 °C for *Sma* I digests) for 3 to 4 hours.

Restriction Digests for Vector and Insert Preparation were set up as follows:

30 μL	purified PCR or vector (a total of 1 μg – 3 μg)
5 μL	recommended enzyme buffer (10x concentrated)
2 μL	enzyme A (~40 units)
13 μL	water

The reaction was incubated (typically at 37 °C) until the vector had been linearised as determined by agarose gel electrophoresis. The insert was digested under the same reaction conditions. Digestion with the second enzyme then carried out, and incubated under conditions determined earlier (with uncut vector) specific for the second enzyme. If the same buffer could be used for both enzymes then 2 μL of enzyme B, 1 μL of the same buffer and 7 μL of water was directly added and the incubation continued. If the second enzyme required a different buffer, the DNA from the first digest was purified (section 2.2.3), and the reaction set up using the second enzyme and buffer. Double digested vectors were gel purified and double digested inserts were purified using direct column purification (section 2.2.3).

2.2.5   Acrylamide Gel Electrophoresis Laemmli (1970)

Resolving Gel Buffer (4x) was prepared as follows:

1.5 M	Tris
0.4% (w/v)	SDS
pH 8.8	

Stacking Gel Buffer (8x) was prepared as follows:

1M	Tris
0.8% (w/v)	SDS

pH to 6.8

Sample Buffer (2x) was prepared as follows:

1 x	stacking gel buffer
6 % (w/v)	SDS
15 % (w/v)	glycerol
10 % (v/v)	β-mercaptoethanol
0.005 % (w/v)	bromophenol blue

1 mL aliquots were stored at -20 °C.

Electrode/tank buffer (5x) was prepared as follows:

5 g	Tris
72 g	glycine
5 g	SDS

Made up to 1L with milliQ water.

Coomassie blue stain was prepared as follows:

0.625 g	Coomassie brilliant blue (R-250)
45 % (v/v)	methanol
5 % (v/v)	glacial acetic acid.

Made up the 500 mL with milliQ water.

Destain I was prepared as follows:

45 % (v/v)	methanol
10 % (v/v)	glacial acetic acid

Made up to 1 L with milliQ water.

Destain II was prepared as follows:

5 % (v/v)	methanol
5 % (v/v)	glacial acetic acid

Made up to 1 L with milliQ water.

The gels were poured as follows:

Resolving Gel (6 % acrylamide) was prepared as follows:

0.1 %	SDS
1 x	resolving gel buffer
6 % (w/v)	acrylamide
0.0005 (v/v)	TEMED
0.005 % (v/w)	ammonium persulfate

The gel was mixed quickly in a small beaker adding the TEMED and ammonium persulfate last, and poured up to ~1 cm below the bottom of the comb. The gel was overlayed with 0.5 mL of water-saturated butanol to minimise exposure of the setting gel to oxygen, as oxygen inhibits the persulfate-catalysed polymerisation reaction. The gel was allowed to polymerise for 1 hour. The water-saturated butanol was rinsed off with MilliQ water.

Stacking Gel (3.2 % acrylamide) was prepared as follows:

0.1 %	SDS
1 x	stacking gel buffer
3.2 %	acrylamide
0.001 % (v/v)	TEMED
0.0025 % (v/w)	ammonium persulfate

This was quickly mixed, adding the TEMED and ammonium persulfate last, and poured on top of the resolving gel. The comb was carefully inserted and the gel was allowed to polymerise for 30 minutes. The comb is then removed and the wells rinsed with water. The gel placed in the tank and immersed in 1x electrode buffer and the samples loaded.

Samples were mixed with an equal volume of 2x sample buffer and boiled for 2-3 minutes. To prepare whole cell samples the OD $_{650\ nm}$ of the *E. coli* culture was measured multiplied by 0.002 to determine the volume (μL) of culture to centrifuge, 13, 000 x g for 1 minute at 4 °C in a bench top centrifuge. The supernatant was removed and the pellet resuspended in 100 μL of 1x sample buffer by vigorous vortexing. The sample was then boiled for 2-3 minutes, and 10-15 μL loaded onto the gel.

Gels were run and stained as follows:

Precision Plus protein standard (Biorad) was loaded onto every SDS acrylamide gel, this gives bands at 250, 150, 100, 75 and 50 kDa on a 6 % acrylamide gel. Gels were

run in a BIORAD Mini-protean® II cell at 120 V for 1 hour until the marker dye was ~1 cm from the bottom from the bottom of the gel. The gels were stained in Coomassie Blue stain for 20 to 40 minutes, then transferred into destain I for 30 miniutes, or until destained sufficiently, then transferred to destain II for storage prior to photography.

2.2.6   PCR Innis (1990); Scharf (1990)

The lyophilised primer was diluted in water to give a stock solution of 300 pM. 2 μL of this stock was then diluted 100 fold in milliQ water to give a working concentration of at 3 pM.

The components of the Invitrogen™ Platinum® *Pfx* DNA polymerase kit were added to the 0.5 mL reaction tube on ice in the following order:

X uL 10x	polymerase buffer
50 mM	magnesium sulfate (in 1 μL)
X μL	milliQ water (to make reaction volume up to 50 μL)
3 pM	forward primer
3 pM	reverse primer
2.5 mM	dNTP mixture (in 5 μL)
1 μL (1 ng/μL)	template
1μL (2.5 U/μL)	DNA polymerase

The tube was then quickly mixed with a vortex mixer and tapped on the bench 2-3 times to gather the contents at the bottom of the tube, all the time maintaining the temperature below 4 °C. The reaction was then overlaid with 2 - 3 drops of mineral oil (Sigma®) and placed in the temperature-cycler (Corbett FTS-320 thermal sequencer) that had been preheated to 95 °C. All reactions were incubated for 2-5 minutes at 95 °C prior to the amplification cycles. This insures that the Platinum® antibody bound to the polymerase is denatured, yielding the active polymerase.

For amplification, 'three step' cycling was used with the following temperatures and times. For the denaturing step the PCR was incubated at 95 °C for 15 seconds. The 30 second annealing step was carried out at temperatures from 38 °C to 65 °C depending on the particular primers and target sequences. The extension was performed at 68 °C

and the time varied depending on the size of the product, from 30 seconds for the relatively small thioredoxin or maltose binding protein targets to 2.5 minutes for the larger hMCM products. After thermo-cycling the samples were stored at 4 °C and a 2 µL aliquot taken out and analysed agarose gel to check the size and purity of the PCR product.

The dNTP concentration was kept to a minimum (2.5 mM) to reduce PCR-generated errors, and the magnesium sulfate concentration used was 50 mM and is the concentration suggested by the Platinum® *Pfx* DNA polymerase protocol, and was not changed throughout this work.

2.2.7 DNA Sequencing

All constructs were sequenced to insure that during the PCR and cloning random errors were not introduced the DNA sequence. To sequence a plasmid, 300 ng of vector, and total of 3.2 pmol of primer were mixed in a 0.2 mL tube, and the volume made up to 15 µL with water. This was then analysed by Lorraine Berry (Allan Wilson Centre Genome sequencing service) using dideoxy chain termination PCR sequencing (Sanger *et al.*, 1977) with Big Dye terminators, and separation on a capillary ABI3730 Genetic Analyser, from Applied Biosystems Inc. using the manufacturers recommended protocol (http://www.appliedbiosystems.com). A BLAST alignment (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi) was then performed to align the sequences with the known vector and target sequences and any ambiguities checked by eye on the electropherogram to ensure that there were no errors introduced by PCR and that the expected ligation had been achieved.

2.2.8 DNA Ligation Sambrook (1989)

Ligation mixtures were assembled on ice with vector:insert molar ratio of 1:3 and 1:1. Vector, insert and water were added and vortexed briefly, tapped to gather contents at the bottom of the tube, heated to 45 °C for 5 minutes and incubated on ice for 5 min. The ligation buffer and then the 0.5-1U T4 DNA ligase were added and the tube mixed by flicking with a finger briefly and tapped on the bench top gather the contents at the

bottom of the tube. For blunt end ligations, more ligase (2U) was added to the reaction mix. This was incubated at room temperature for 2 hours, then 2 μL of this solution was used to transform competent *E. coli* cells. The following ligation and transformation controls carried out in parallel: vector only ligation, ligation reaction without ligase, and cells without DNA.

2.2.9 Plasmid Transformation Hanahan (1991)

50 μL aliquots of *E. coli* competent cells in 0.6 mL Eppendorf tubes were thawed on ice for 10 minutes and 2 μL of DNA solution (ligation mix or plasmid preparation) was added. The contents of the tube was quickly mixed by gently by flicking with a finger 3 times, and incubated 20 - 30 minutes on ice. Then the tube was incubated at 42 °C for 45 seconds. placed on ice for 5 minutes and 200 μL of SOC recovery medium (2% tryptone, 5% yeast extract. 10 mM NaCl. 2.5 mM KCl. 10 mM $MgCl_2$. 10 mM $MgSO_4$. 20 mM glucose) was added and the cells incubated with agitation at 37 °C for 1 hour. The cells were then spread on LB agar plates containing the appropriate antibiotics and incubated at 37 °C overnight. Single colonies were transferred to 5 mL LB broths, grown overnight and cells harvested by centrifugation for subsequent plasmid preparation.

2.2.10 Rubidium Chloride Competent *E. coli* Cells Hanahan (1991)

A single well isolated colony was transferred from a plate and grown at 37 °C with constant agitation overnight in a 5 mL LB broth containing the appropriate antibiotics. 100 μL of this 5 mL broth was used to inoculate a 200 mL flask of LB broth also containing the appropriate antibiotics. This culture was grown at 37 °C with agitation to an OD $_{650 \, nm}$ of 0.3 - 0.4. This culture was transferred into four 50 mL NUNC tubes and chilled on ice for 10 minutes. The cells were then centrifuged for 15 minutes at 4 °C in a Sorvell RT7 centrifuge using the RTH-750 rotor at 4000 rpm. The supernatant was decanted, the pellets resuspended and combined in 4.5 mL of ice cold RF1 buffer (100 mM RbCl, 50 mM MnCl$_2$ 6 H$_2$O, 30 mM KOAc, 10mM CaCl$_2$ + 2 H$_2$O, pH 5.8 adjusted with acetic acid) and incubated on ice for 30 min. The cells were pelleted by centrifugation for 15 minutes at 4000 rpm at 4 °C and the supernatant decanted. The pellet was resuspended in 4 mL of ice cold RF2 buffer (10 mM RbCl, 10 mM MOPS, 75 mM CaCl$_2$·6 H$_2$O, 15% (v/v) glycerol, pH 5.8 adjusted with NaOH). 200 μL aliquots in 0.6 mL Eppendorf tubes were immediately stored at -70 °C.

The competency of the cells was determined by transforming with a known amount of vector DNA, and plating out dilutions of the transformation to determine the efficiency of transformation (number of colonies per ng of cccDNA).

2.2.11 Induction and Expression of the Recombinant Protein.

Sequenced T7-based protein expression plasmids were transformed into the following *E. coli* expression strains: ER2566, TG1, BL21(DE3) or BL21(DE3)Origami B. The cells were grown with agitation at temperatures ranging from 12 °C to 37 °C in 5 mL LB broths. At an OD $_{650 \, nm}$ of 0.6 - 0.8 the cells were induced with 0.2 - 0.6 mM IPTG. After the addition of IPTG a series of samples was taken at 1 hour intervals from 0 hours (directly before adding IPTG) to 6 hours after IPTG addition. These samples were analysed by SDS PAGE to determine the time course of synthesis of the recombinant protein.

In some experiments recombinant protein was also expressed in the presence of adenosylcobalamin. Cells were grown and induced as described above, except that 30 minutes before induction adenosylcobalamin was added to a final concentration of 1 mM.

2.2.12 Determining the Solubility of the Recombinant Protein

Cells were grown in 200 mL of LB in 500 mL flasks, with agitation at 37 °C. To induce recombinant protein expression IPTG was added at an OD 600_{nm} of 0.6 for induction at 37 °C and an OD 600_{nm} of 0.8 for induction at 12 °C. After induction, cells were grown for 3 hours, at 37 °C or 22 hours at 12 °C. *E. coli* cells grown at 12 °C grow very slowly, and so cells do not reach the stationary growth phase, even after 22 hours. The highest culture reached an OD 600_{nm} of 1.2. Expression was induced with 0.6 mM IPTG. When protein was expressed at 12 °C the growth temperature for cells was lowered from the initial growth temp (37 °C), down to 12 °C, 10 minuites before induction. This insures that even the very first recombinant protein translated is being expressed at 12 °C. Samples were removed for a solubility analysis by SDS PAGE gel. The following formula was used to take equivalent cell density samples at the time points, 0, 1, and 2 hours after induction.

$$\frac{7.5}{OD_{600}} = X \text{ mL of sample taken}$$

The cells were placed on ice for 10 minutes, centrifuged at 3000 xg for 15 min at 4 °C. The pellet was resuspended in 1 mL of lysis buffer (20 mM HEPES pH 7, 500 mM NaCl, 0.1% Triton X-100, 1 mM EDTA), and sonicated in an ice bath to lyse the cells. Approximately 8 pulses of 20 sec, and 1 min cooling between each burst were used to lyse the cells, the lysate is milky but not opaque. The lysate was then centrifuged at 3000 xg for 15 min at 4 °C, to pellet the insoluble fraction of the lysate. The supernatant (containing the soluble protein fraction) was aspirated into a new tube, and the pellet (containing the insoluble protein fraction) resuspended in lysis buffer to the original volume of the total lysate. The distribution of the recombinant protein between the supernatant and pellet fractions allows a semi-quantitative assessment of protein solubility to be made.

2.2.13 Purification of His-Tagged Recombinant Protein.

His-tagged proteins were purified on Ni-NTA Spin columns (Qiagen) using a method based on the manufacture's protocol.

Buffers:

All buffers were stored and used at 4 °C.

~Lysis Buffer: 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0

~Wash Buffer: 50 mM NaH_2PO_4, 300 mM NaCl, 30 mM imidazole, pH 8.0

~Elution Buffer: 50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0

After induction and expression, 100 mL of cells were pelleted by centrifugation (4 °C at 3 000 x g for 15 minutes). The ~ 0.5 g cell pellet was resuspended in 1 mL of ice cold lysis buffer and sonicated for approximately ten 20 second bursts with 1 min cooling between each burst, are used to lyse the cells. The solution changes from a cream coloured opaque solution to a milky in colour but translucent solution, to give a 100x concentrated lysate. The lysate was centrifuged at 4 °C at 13 000 x g for 30 minutes and the supernatant, containing soluble proteins, collected and applied to the pre-equilibrated Ni-NTA spin column.

The Ni-NTA spin column was equilibrated by applying 600 µL of the lysis buffer, centrifuging at 700 x g for 2 minutes, and discarding the flow through. The lysate supernatant was loaded onto the Ni-NTA column, in 600 µL batches and centrifuged at 4°C at 700 x g for 10 minutes with the lid of the column closed to slow the flow and helping to facilitate binding. The flow-through was stored on ice. After this the column was washed twice with wash buffer, centrifuging the column with the lid open, for 2 minutes at 4 °C at 700 x g and again the flow through was stored on ice. The His-tagged proteins were then eluted with 200 µL of elution buffer and stored on ice.

2.2.14 Cleavage of the Intein in the Presence of DTT Perler (2000)

After purification the 200 µL of recombinant protein was divided into three samples. "+ DTT" (95 µL), "background control" (95 µL), and "negative control" (10 µL). The negative control sample was denatured by the addition of 10 µL of SDS-PAGE sample

buffer. To the DTT cleavage sample 5 µL of 1 M DTT was added and the three samples were incubated at 4 °C overnight. 10 µL samples from the "background control" and "+DTT" solution were mixed with 10 µL of 2x SDS-PAGE sample buffer and 15 µL of each sample was analysised by SDS-PAGE.

2.2.15 The hMCM Activity Assay Hayes. (1998)

The hMCM was assayed according to Hayes, (1998), as follows:

Assay mix was made fresh on the day of the experiment. The following 'ingredients' were added to a 15 mL Nunc tube, in the order given:

4000 µL	0.5 M potassium phosphate buffer, pH 7.2.
250 µL	0.4 M sodium pyruvate, 4 mM NADH
28 µL	Methylmalonyl CoA epimerase enzyme in 50 mM Tris/HCl, pH '7.5 ~2U/µL (E. Saafi, MSc. Thesis, 1994)
5 µL	L-Malate dehydrogenase in 50% glycerol, 6 U/µL, Boehringer Mannheim.

Made up to 10 mL with water.

Succinyl CoA for the assay was prepared in 1 mL batches and store at -70 °C using the method of Hayes, (1998). 10 mg of coenzyme A (Boehringer Mannheim), 9 mg NAHCO 3 and 8 mg succinic anhydride were weighed into a 1.5 mL Eppendorf with a screw-cap. To that 0.99 mL of ice cold degassed water was added. This solution was vortexed for 10 seconds returning quickly to the ice. The solution was then votexed with 1 second bursts at 20 second intervals over 30 minutes, returning to the ice between vortexing. The solution was transferred to a flip-top Eppindorf tube, and the pH lowered to below 4.5 by the addition of 30 µL of 2 M HCl. This typically yields 9-10 mM succinyl CoA, which can be determined more accurately, if desired, using an enzyme based assay.

0.8 mL of the assay mix was placed in a 1 mL quartz curvette and allowed to pre-equilibrate at 37 °C for 5 minutes. Then 8 µL of 10 mM succinyl CoA preparation, 0.5 µL of transcarboxylase (in 50% glycerol at ~0.12 U/µL) and 1.5 µL of 1 mM 5-deoxyadenosylcobalamin in 20 mM potassium phosphate buffer, pH 7.0, were added

and the contents of the cuvette mixed well by inversion. The cuvette was placed in the CARY 1E UV-visible spectrophotometer (Varian) in a temperature controlled cuvette holder at 37 °C. The rate of NADH oxidation was then measured at 340 nm for ~2 minutes, or until a relativity steady background rate had been reached. This background rate is due to the small amount of contaminating *P. shermanii* MCM, in the transcarboxlase preparation. Then, 200 µL of hMCM sample was added, the cuvette contents mixed again by inversion, and the rate of change in absorbance measured over 5-10 minutes. As a positive control, partially purified recombinant *P. shermanii* MCM at 0.03 U/mL was used instead of the hMCM sample, to insure that the coupled-enzyme assay system was functioning. In addition to this, other controls involved adding the 5'deoxyadenosylcobalamin or the succinyl CoA last (i.e. after the hMCM), to ensure that the increase in rate observed was due to the recombinant methylmalonyl CoA mutase, rather than other factors (i.e. the many other NADH dependant reactions) that may also cause NADH oxidation. To work out the activity, the difference between background rate and reaction rate is used to give the units of active hMCM present per µL of sample. 1 unit of active hMCM will cause the oxidation of 1 µmol of NADH per minute.

3   Design and Construction of the Four Expression Vectors

3.1   Introduction

When expressed in *E. coli* the hMCM protein forms insoluble inclusion bodies (Mark Patchett, *pers. Comm.*). The provision of an N-terminal Trx solubility tag improved the solubility; however the subsequent enterokinase catalysed cleavage of the tag from hMCM was unsuccessful (Janata *et al.*, 1997). Therefore previous work showed that there was a requirement for an expression system that provides an N-terminal Trx tag but does require *in trans* proteases to separate the tag from the hMCM target protein. Consequently, an approach to cleave the tag *in cis*, using an in-frame intein fusion, was adopted. Inteins are protein-splicing domains within proteins that have the ability to excise themselves and rejoin the flanking regions of the polypeptide. This *in cis* proteolytic activity of inteins may be useful in circumventing the physical access to the site that is required with *in trans* proteolysis.

Thus far, there are no commercially available vectors that combine the features of a solubility tag with intein-mediated separation. Therefore, the logical progression for this research was to design and construct four novel expression vectors containing a solubility tag in conjunction with intein mediated system to separate the tag. The starting points for these novel vectors were the commercially available New England Biolabs vectors, pTYB4 and pTYB11. These two vectors contain a CBD used for affinity purification and the *Sce*VMA1 intein located between the CBD and the target protein which would enable intein-mediated separation, thereby releasing the native target protein. The pTYB4 or pTYB11 vector DNA encoding the CBD was

replaced with DNA encoding either (i) the *E. coli* Trx or (ii) the *E. coli* MBP. This will create four novel vectors that differ in the nature (Trx or MBP) and position (C or N terminal) of the tag in the expressed recombinant protein.

3.2　Design of the Novel Expression Vectors

3.2.1　*In trans* Protease Separation of a Solubility Tag and the Subunit Structure of hMCM

The difficulty with enterokinase protease mediated separation of the hMCM target and N-terminal Trx tag experienced by Janata *et al* (1997) lead to the approach described in this work. This involves the use of intein mediated separation as an alternative method for the separation of the two proteins. Site-specific proteases (other than enterokinase), could have been trialled. If it was a site access problem for the protease, other trans-acting proteases could encounter similar problems to the enterokinase.

With structure of hMCM largely unknown, it is difficult to predict the structure of the N-terminal region and how it may influence the efficiency of enterokinase cleavage. The N-terminal region of hMCM is involved in the subunit-subunit interaction to form the homodimer. Mutations in the hMCM N-terminal region cause a loss of activity and consequently the disease MMA in humans (Fuchshuber *et al.*, 2000). This suggests that dimerisation is important for the activity of hMCM and/or structural stability. If dimerisation is important for stability of the hMCM, then expression of the solubility tag on the C-terminus would also interfere less. It has been shown that soluble hMCM with an N-terminal Trx tag is active (Janata *et al.*, 1997), suggesting that hMCM dimers are formed in the presence of the Trx tag. This subunit interaction could block access of the protease to the site of cleavage (figure 3.1A). The α subunit of the heterodimeric *P. shermanii* MCM when expressed in *E. coli* as a single subunit (without the β subunit) is unstable and insoluble (Mark Patchett *pers. com.*). Given the high homology between the α subunits of these two enzymes it is likely that the α subunit of hMCM is also unstable as a monomer. When the hMCM is expressed as a fusion protein with the N-terminal MBP the fusion protein is insoluble (Janata *et al.*, 1997). This insolubility may be due to disruption of the hMCM subunit interaction, caused by the MBP which is much larger than the Trx tag (figure 3.1B). Such disruption might be avoided by the incorporation of a longer amino acid linker of between the hMCM and MBP, or placing the tag at the C-terminus of hMCM (figure 3.1C, D).
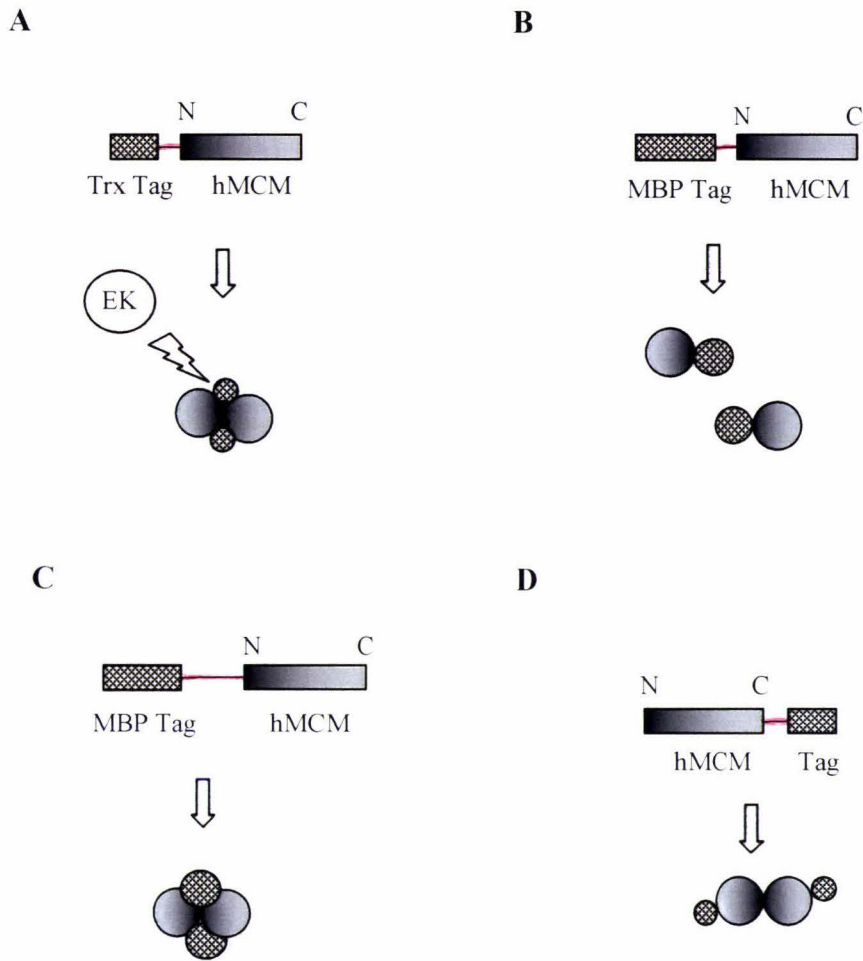
Figure 3.1 *In Trans* Proteolytic Cleavage of Tags and Dimer Formation of hMCM.

A. The hMCM subunits have dimerised, blocking access to the cleavage site by the protease enterokinase (EK) that cleaves at the recognition site buried between the hMCM and Trx (tag).

B. The subunits are not able to form dimers due to the large N-terminal MBP tag, resulting in monomers which are unlikely to be stable and soluble.

C. Increasing the linker region (red line) between the hMCM and MBP may allow hMCM subunit interaction.

D. A tag on the C-terminal of hMCM will cause less disruption to dimerisation than an N-terminal tag, thereby allowing the subunit interaction.

In summary, the following conditions must be considered when designing an expression system:

i)   The requirement for a solubility tag.

ii)   Avoiding the use of *in trans* proteases for separation of the tag, as the cleavage site may be blocked by the subunit interactions.

iii)   Having a long linker region between N-terminal tags and the hMCM subunits, (to help the subunit interaction occur), or avoiding N-terminal tags altogether and utilising C-terminal tags.

With these considerations in mind, four expression constructs were designed with the following solubility tag orientations:

1.   N-terminal Trx

2.   C-terminal Trx

3.   C-terminal MBP

4.   N-terminal MBP

Firstly, expression construct 1, the hMCM has an N-terminal Trx that the repeats the Trx solubility tag orientation of Janata *et al.*, (1997), which hMCM was soluble, but the Trx could not be separated with the protease enterokinase. In construct 2 and 3 and the Trx or MBP tag at the C-terminus of hMCM creates less disruption to the subunit interaction. Lastly, in construct 4, the fusion protein has an N-terminal MBP with a longer linker region to facilitate subunit interaction. All of these proposed constructs use intein-mediated cleavage to separate the solubility tag from hMCM which will provide an alternative to the *in trans* enterokinase cleavage that was ineffective in separating the Trx-hMCM fusion protein described by Janata *et al.*, (1997).

3.2.2   Replacing the CBD Affinity Tag with a Solubility Tag

In this work the DNA encoding the CBD tag was excised from the pTYB4 and pTYB11 vectors and replaced by the DNA encoding the *E. coli* Trx or the MBP. The linker of amino acids between the intein and the CBD was reduced from 15 to 11 amino acids when the Trx or MBP was cloned into the pTYB11 to generate the N-terminal fusion constructs (figure 3.2). Changing the expression vector so dramatically may affect the structural stability and solubility of the entire fusion protein expressed from it.

The decrease in the linker between the tag and intein could affect solubility the intein and/or tag, potentially rendering the entire complex insoluble. In addition to this change,

the 5 kDa CBD was replaced with a larger Trx (11 kDa) or MBP (40 kDa) and this may affect the initial yield as well as the stability/solubility of the fusion protein. Although no conserved amino acids of the intein were changed, making such a large structural change, within a loop of the intein (pTYB11) or in close proximity to the intein (pTYB4), may affect the structural stability of the entire fusion construct and/or the activity of the intein (figure 3.2).

A. New England Biolabs   **B. This work**

Replaced domain II with the CBD

Linker of 15 amino acids

Replaced the CBD with Trx or MBP

Linker was reduced by 4 amino acids
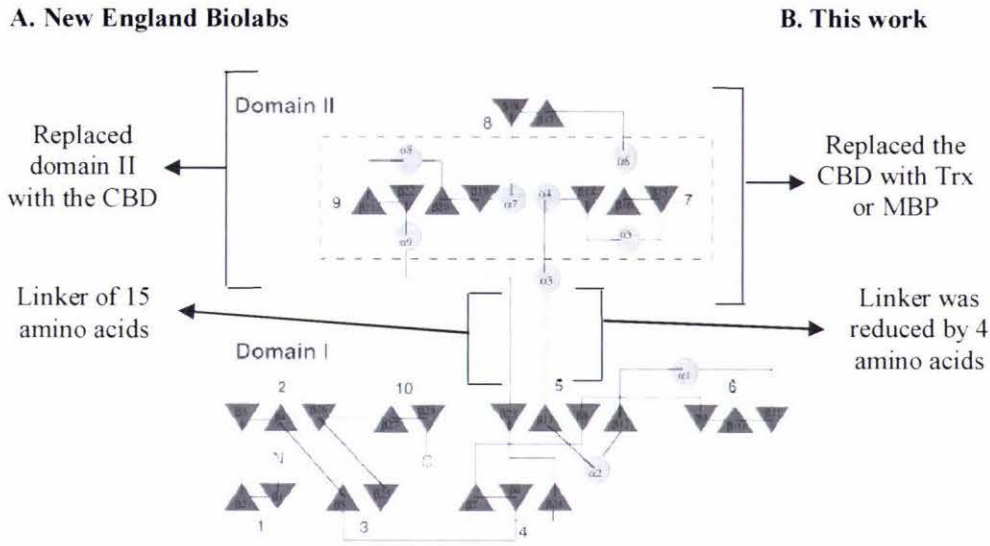
Figure 3.2 The Domain Structure of the Native *Sce* VMA1 Intein in pTYB11. Domain II contains the ORF for the intein endonulcease. In **A**, New England Biolabs removed the ORF for the endonulease and replaced it with DNA encoding the CBD, to create the vector pTYB11. In this work, **B**, the CBD is replaced by the Trx or MBP reducing the linker from 15 amino acids to 11 amino acids.

(Figure adapted from Duan *et al.* 1997)

When designing these expression systems it is also important to consider is the total size of the protein fusion that is to be expressed. Proteins in *E. coli* are generally small; the majority of *E. coli* proteins fall between 25–120 kDa in mass (Cash, 1995; Blattner *et al*, 1997). The ~150 kDa recombinant fusion proteins to be expressed in this study were considerably larger than the majority of *E. coli* proteins, putting stress on the metabolism of *E. coli* to synthesise large numbers of this free loading protein. In addition to this, *E. coli* chaperones such as GroEL and GroES are unlikely to aid the folding of such a large polypeptide (Takagi *et al.*, 2003). The NusA tag was considered as a solubility tag (De Marco *et al.*, 2004; Davis *et al.*, 1999). However NusA is a large

protein of 55 kDa. and at 192 kDa the intein-hMCM-NusA fusion protein is less likely to be expressed efficiently in *E. coli*.

3.2.3   Ensuring *in vitro* DTT-Induced Intein Cleavage in pTYB11

The N-terminal amino acid of the hMCM which is directly adjacent to the intein cleavage site is of importance as it affects the efficiency of the intein mediated cleavage in the fusion protein expressed from pTYB11 (Perler. 2000). Different amino acid residues will cause: i) *in vivo* intein cleavage. ii) *in vitro* intein cleavage in the presence a reducing agent such as DTT. or iii) no intein cleavage. *In vivo* cleavage is generally undesirable. even if the protein has folded into its native state and is now soluble as it prevents purification using affinity chromatography methods exploiting any specific ligand binding properties associated with the solubility tag. Further. some residues at the N-terminus will prevent intein cleavage. New England Biolabs suggests a number of ideal amino acids (Gly. Ala. Ile. Leu or Met). which will provide intein cleavage that is inducible by a reducing agent such as DTT. Leu was selected for this system. as it is recommended and also is the native N-terminal amino acid for hMCM. This construct would yield the native hMCM sequence upon intein cleavage (see appendix E for full details of the amino acid selection).

3.2.4   The Composition of the Four Novel vectors

The backbones for the novel vectors were provided by pTYB4 and pTYB11 vectors (New England Biolabs). The pTYB11 vector allows an N-terminal CBD tag to be attached to the recombinant protein of choice. and the pTYB4 vector allows the attachment of a C-terminal CBD tag. The ORF for the CBD was excised by digesting pTYB4 with *Age*I and *Pst*I and by digesting pTYB11 with *Nhe*I and *Nco*I (figures 3.3 and 3.4). The CBD was replaced with either a *malE* ORF encoding the MBP (from the vector pMALc2g) or the *trx* ORF encoding the His-tagged thioredoxin protein (from the vector pET32a). Thereby creating the four novel vectors: pTYB4.MBP, pTYB4.Trx. pTYB11.MBP and pTYB11.Trx (figures 3.3 and 3.4). It was anticipated that these novel vectors would allow a unique range of expression systems for hMCM.

| | Fwd primer | | AgeI | PstI | | trx |
| | | | | | | |

a) PCR

b) Digest vector and PCR products with *Age* I and *Pst* I

c) Ligate
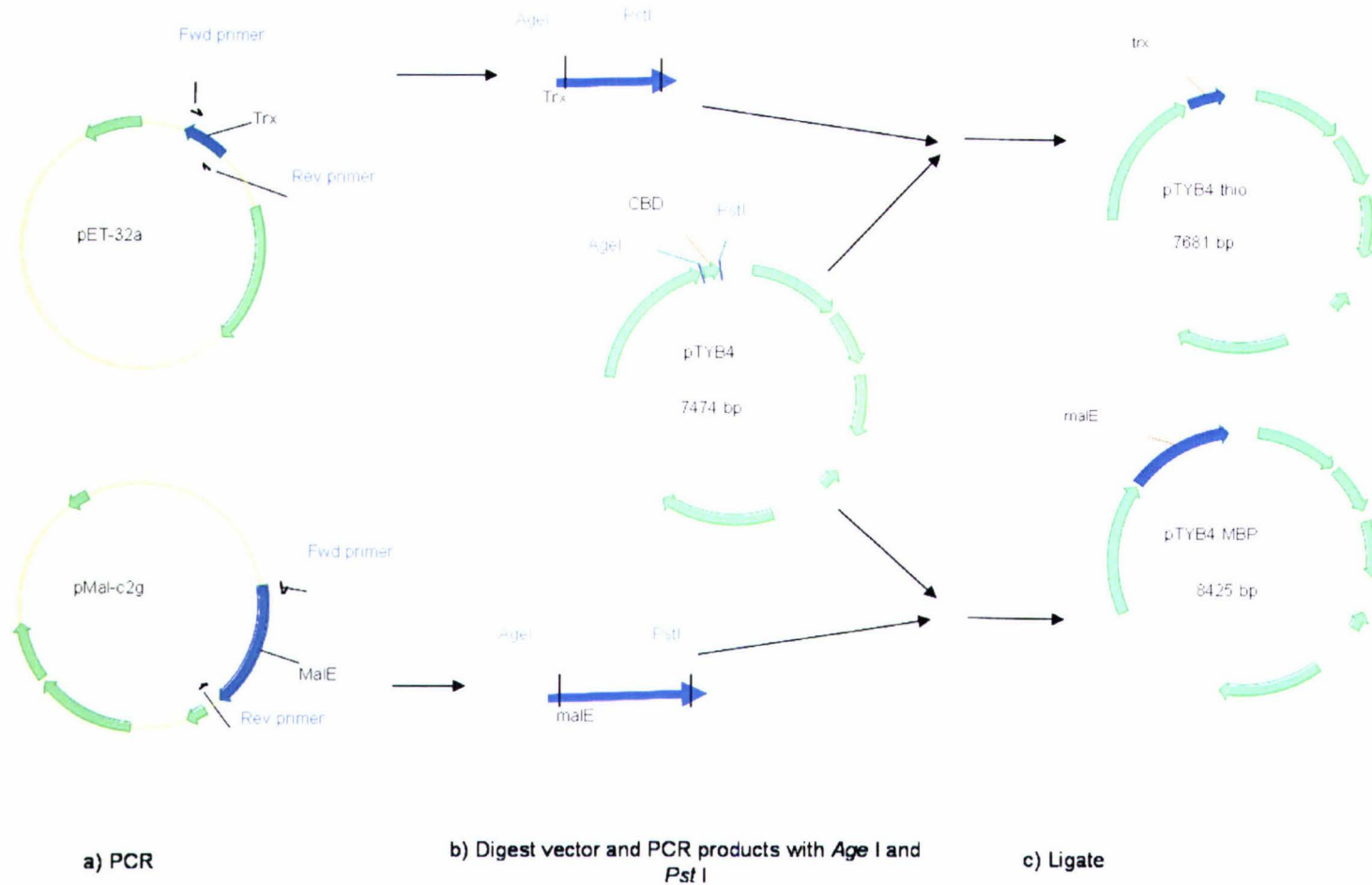
Figure 3.3 Flow diagram, showing the production of pTYB4.MBP and pTYB4.trx. The first step is the PCR of the 1175 bp *malE* ORF from the vector pMAL-c2g and the PCR of the 365 bp *trx* ORF from pET-32a. The pTYB4 vector and the two inserts are digested with *Age*I and *Pst*I. Each insert was ligated into the pTYB4 vector, producing the two unique expression vectors pTYB4.trx and pTYB4.MBP.

a) PCR

b) Digest vector and PCR products with *Age* I and *Pst* I

c) Ligate
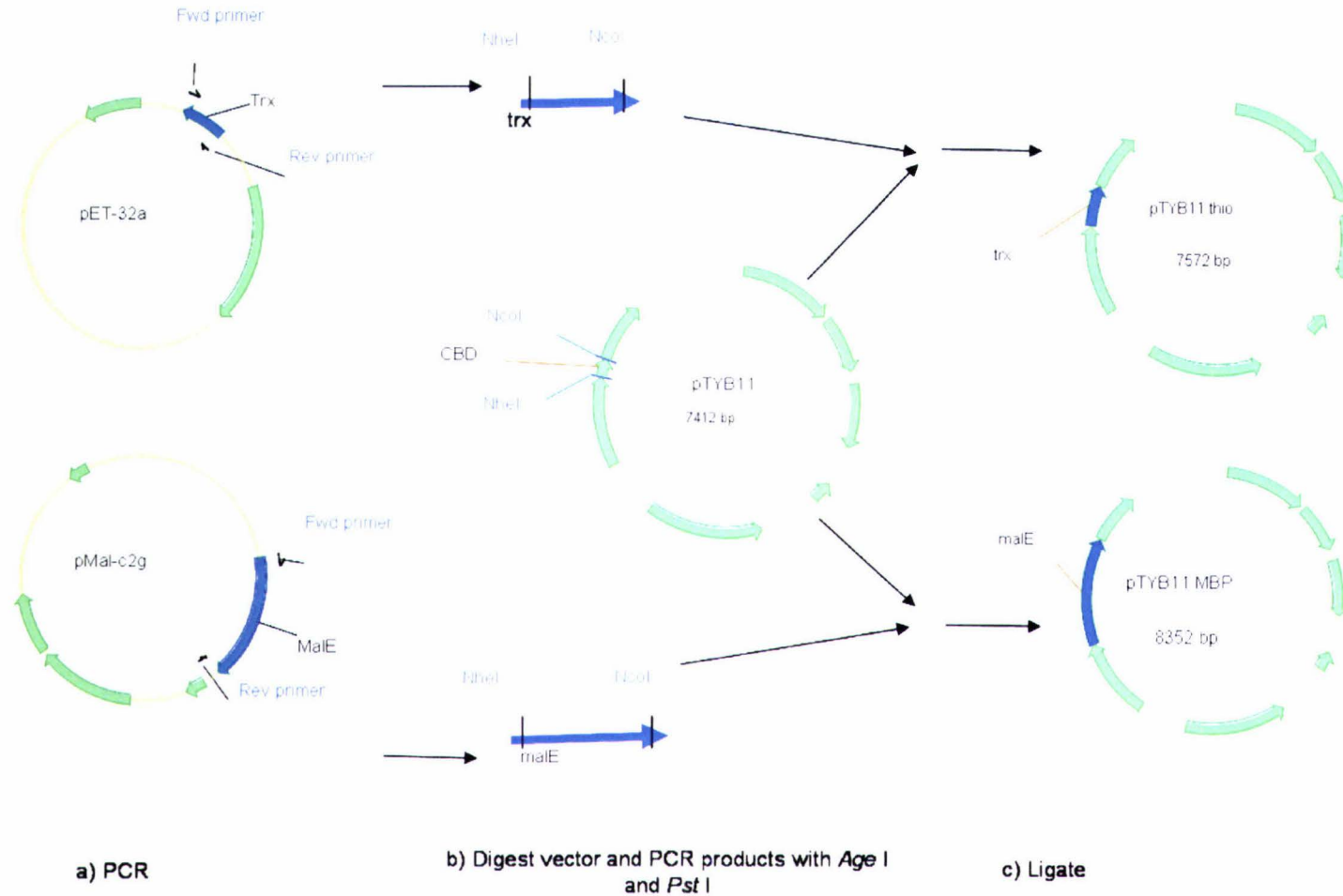
Figure 3.4 Flow diagram, showing the production of pTYB11.MBP and pTYB11.trx. The first step is the PCR of the 1175 bp *malE* ORF from the vector pMAL-c2g and the PCR of the 365 bp *trx* ORF from pET-32a. The pTYB11 vector and the two inserts are digested with *Nhe*I and *Nco*I. Each insert was ligated into the pTYB11 vector, producing the two unique expression vectors pTYB11.trx and pTYB11.MBP.

3.3.2 Cloning of the *trx* Gene into the Vector pTYB4

To prepare the vector the DNA encoding the CBD in pTYB4 was removed by digesting the vector with the restriction enzymes, *Age*I and *Pst*I; these enzymes have unique recognition sites that flank the CBD. The vector was separated from the CBD by agarose gel electrophoresis.

The insert was generated by PCR using a forward primer with an incorporated *Age*I restriction site and a reverse primer with an incorporated *Pst*I restriction site to amplify the *trx* gene from the vector pET32a (figure 3.5A, table 3.1). The *trx* ORF PCR product, was digested with the restriction enzymes *Age*I and *Pst*I, then purified by gel electrophoresis.

The concentration of both the gel purified vector and insert was determined to be approximately 8 ng/μL by agarose gel electrophoresis quantification. After ligation, transformation and plasmid preparation the candidate plasmids were analysed by digestion with *Nde*I, a restriction enzyme that cuts once within the insert sequence and not in the vector. The candidate plasmid that yielded a single 7400 bp linear fragment after *Nde*I digestion (figure 3.5B, lanes 5), was sequenced with the primer ACTYB1-4CBD. A plasmid with the correct insert sequence was selected and named pTYB4.Trx.
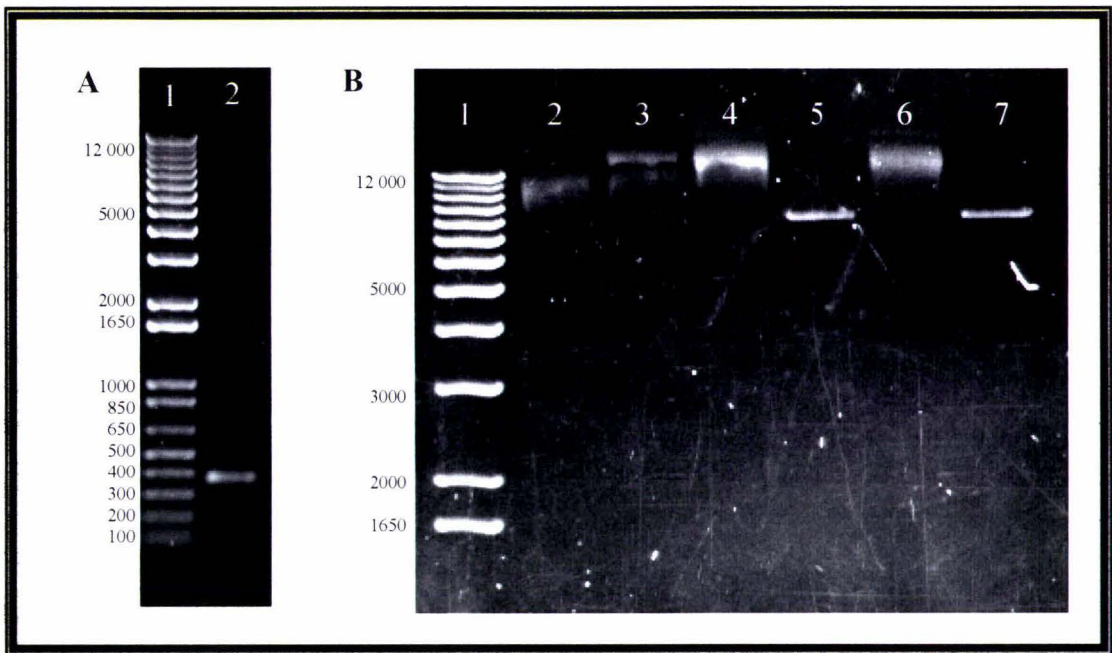
Figure 3.5 Agarose Gels Showing the Cloning of *trx* into pTYB4.

A. The PCR product. The product is of the expected size, 365 bp, for the *trx* gene. Lane 1, 5 μL 1Kb Plus ladder Lane 2, 2 μL from 50 μL of PCR product

B. Candidate plasmids digested with *Nde*I. Vectors without an insert remain uncut, as only the *trx* gene insert contains an *Nde*I site. Lane 1, 5 μL 1 Kb Plus DNA ladder. Lanes 2, 4, 6; 5 μL of uncut plasmid. Lanes 3, 5, 7; 5 μL of plasmid digested with *Nde*I. This shows preps in lanes 5 and 7 are linearised by *Nde*I, yielding the 7600 bp vector and so contain the *trx* gene insert.

3.3.3   Cloning of the *trx* Gene into the Vector pTYB11

To prepare the vector the DNA encoding the CBD in pTYB11 was removed by digesting the vector with the restriction enzymes, *Nhe*I and *Nco*I; these enzymes have unique recognition sites that flank the CBD. The digested vector was purified from the CBD ORF by agarose gel electrophoresis.

The insert was generated by PCR using a forward primer incorporating the restriction site for *Nhe*I and the reverse primer incorporating the restriction site for *Nco*I to amplify the *trx* from the vector pET32a (figure 3.6A, table 3.1). The *trx* gene PCR product was then digested with the restriction enzymes *Nhe*I and *Nco*I, then purified by gel electrophoresis.

The concentration of the insert and vector were determined by agarose gel electrophoresis to be 2.5 ng/μL and 7.5 ng/μL respectively. After ligation, transformation and plasmid preparation the resulting plasmids were analysed by digestion with *Nde*I (figure 3.4B), a restriction enzyme that cuts once within the insert sequence and not in the vector. The candidate plasmid that yielded a single 7600 bp linear fragment after *Nde*I digestion (figure 3.6B, lanes 5), was sequenced with the primer ACTYB1-4CBD. A plasmid found to contain the correct insert sequence was selected, this plasmid was called pTYB11.thio.
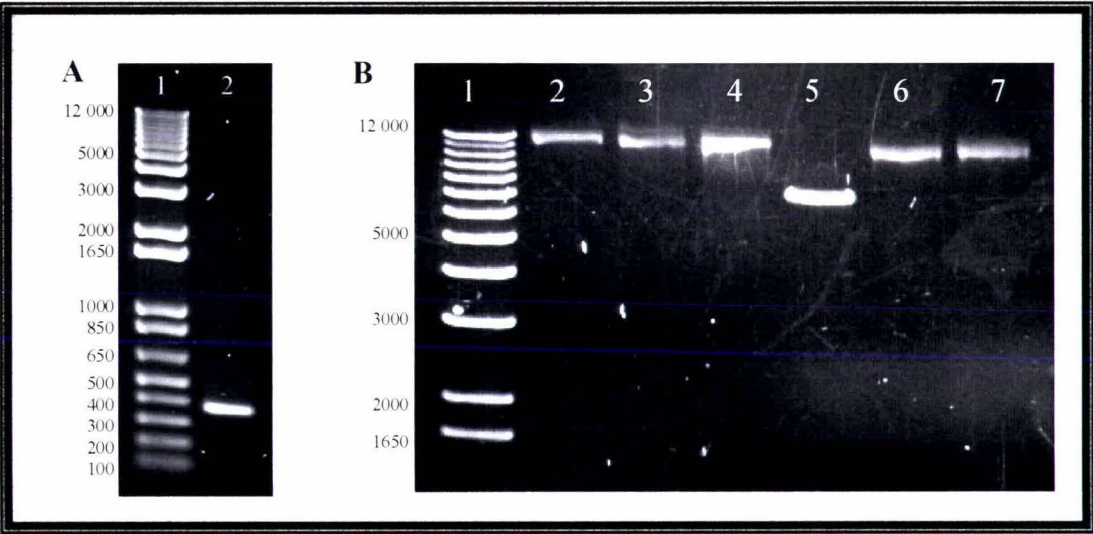
Figure 3.6 Agarose Gels Showing the Cloning of *trx* into pTYB11
A. The PCR product. The product is of the expected size for the *trx* gene, (365 bp). Lane 1, 5 μL 1Kb Plus ladder. Lane 2, 2 μL from 50 μL of PCR product.
B. Candidate plasmids digested with *Nde*I. Vectors without an insert remain uncut, as only the insert contains an *Nde*I site. Lane 1: 5μL of 1 Kb Plus DNA ladder. Lanes 2, 4, 6: 5μL uncut plasmids (digests contain buffer only). Lanes 3, 5, 7: 5 μL of candidate plasmids digested with *Nde*I. This shows candidate plasmid in lane 5 has linearised to give a 7400 bp band.

3.3.4   Cloning of the *malE* Gene into the Vector pTYB4

To prepare the vector the DNA encoding the CBD in pTYB4 was removed by digesting the vector with the restriction enzymes, *Age*I and *Pst*I; these enzymes have unique recognition sites that flank the CBD ORF. The digested vector was then separated from the CBD by agarose gel electrophoresis.

The insert was generated by PCR using a forward primer incorporating the restriction site for *Age*I and reverse primer incorporating the restriction site for *Pst*I to amplify the *malE* gene from the vector pMALc2g (figure 3.7A; table 3.1). There were a number of products, but the predominant 1175 bp product corresponding to the expected size for *malE* was gel purified and digested with *Bgl*II (an unique internal restriction site within the *malE* gene) to insure that the correct PCR product had been purified (gel not shown). This 1175 bp gel purified fragment was then diluted to 1 ng/μL, and used as a template under the same PCR conditions as in the initial PCR. This process enabled the generation of a single PCR product of the expected size, (1175 bp) (figure 3.7B; table 3.1). The *malE* gene PCR product, was digested with the restriction enzymes *Age*I and *Pst*I, then purified by gel electrophoresis.
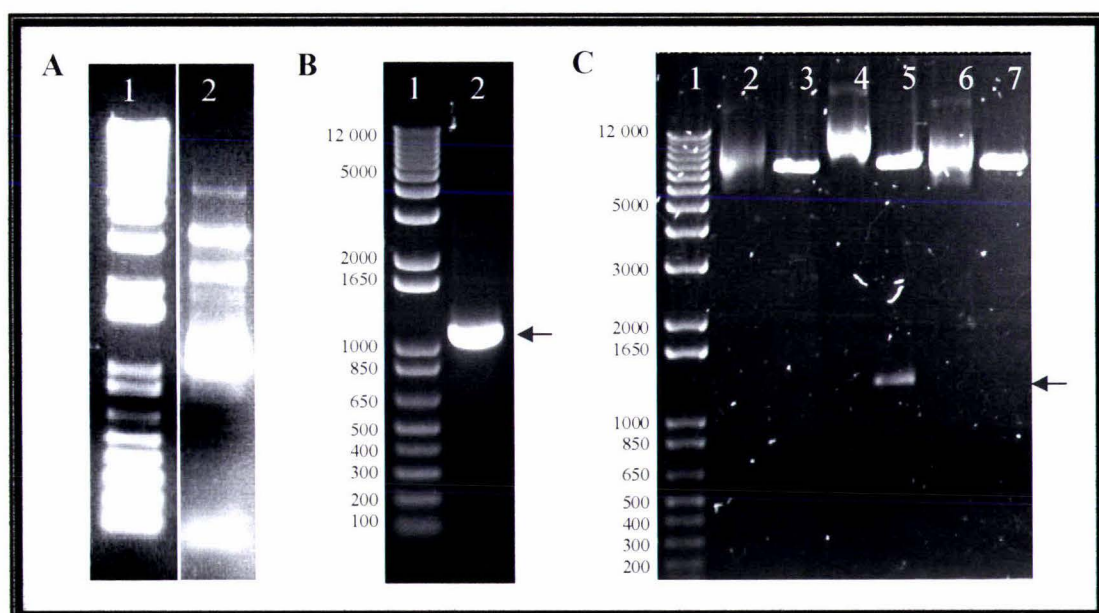
Figure 3.7 Agarose Gels Showing the Cloning of *malE* into pTYB4
A. PCR using the vector pMALc2g as a template, showing the multiple PCR product bands. The product at the correct size for the 1175 bp *malE* gene was gel purified. Lane 1: 5 μL 1Kb Plus ladder. Lane 2: 2 μL from 50 μL of PCR product.
B. PCR using the purified 1175 bp fragment as a template. The product is the correct size for the 1175 bp *MalE* gene. Lane 1: 5 μL 1Kb Plus ladder. Lane 2: 2 μL from 50 μL of PCR product.
C. Candidate plasmids digested with *Bgl* II. The 1500 bp band (arrow) indicates plasmids that may contain an insert. Lane 1: 5μL 1 Kb Plus DNA ladder. Lanes 2, 4 and 6: 5 μL uncut plasmid (digests contain buffer only). Lanes 3, 5 and 7: 5 μL of plasmid digested with the restriction enzyme *Bgl*II. Lane 5 shows a plasmid with the expected bands for a plasmid containing the *malE* insert.

The concentration of the vector was determined by gel electrophoresis to be approximately 10 ng/µL and that of the insert approximately 8 ng/µL. After ligation, transformation and plasmid preparation the resulting plasmids were screened for inserts by digestion with *Bgl*II, a restriction enzyme that cuts once within the insert and once within the vector, producing 1500 bp and 7000 bp fragments. A candidate plasmid that was cut twice by *Bgl*II, was selected and sequenced with the primer ACTYB1-4CBD (figure 3.7C, lane 5). A candidate plasmid containing an error free insert was selected and named pTYB4.MBP.

3.3.5   Cloning of the *malE* Gene into the Vector pTYB11

To prepare the vector the DNA encoding the CBD in pTYB11 was removed by digesting the vector with the restriction enzymes, *Nhe*I and *Nco*I; these enzymes have unique recognition sites that flank the CBD ORF. The vector was then separated from the CBD ORF by agarose gel electrophoresis.

The insert was generated by PCR using a forward primer, incorporating the restriction site *Nhe*I and the reverse primer, incorporating the restriction site *Nco*I, to amplify the *malE* gene from the vector pMAL.c2g. This PCR produced bands of multiple sizes, however, the major product was of the expected size (1175) for *malE* (figure 3.8A; table 3.1). Raising the annealing temp above 52 °C resulted in the loss of all the products including the 1175 bp band. Therefore, the required fragment of size 1175 bp was gel purified to eliminate other PCR products (figure 3.8B). The gel purified *malE* PCR product was digested with restriction enzymes *Nhe* I and *Nco*I, then purified by agarose gel electrophoresis.
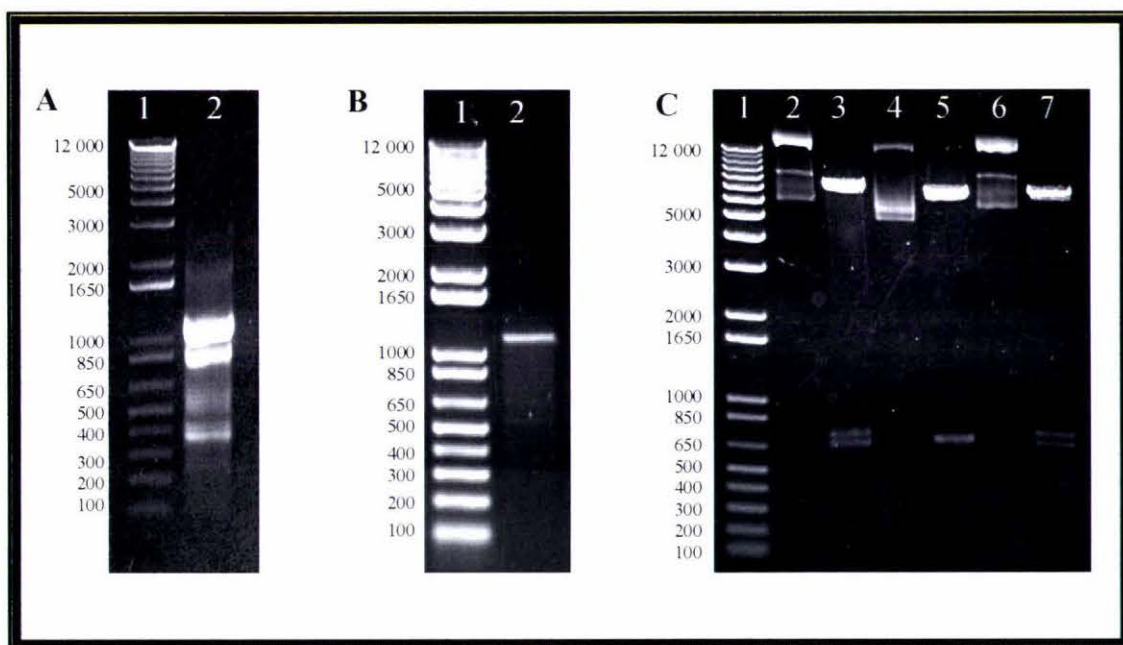
Figure 3.8 Agarose Gels Showing the Cloning of *malE* into pTYB11.
A. PCR using the vector pMALc2g as a template, showing the multiple PCR product bands. The product at the expected size for the 1175 bp *malE* gene was gel purified. Lane 1: 5 μL 1Kb Plus ladder. Lane 2: 2 μL from 50 μL of PCR product.
B. The 1175 bp *malE* gene after gel purification. Lane 1: 5 μL 1Kb Plus ladder. Lane 2: 5 μL of the gel purificated product.
C. Candidate plasmids digested with *Bgl* II. The plasmids with hMCM inserts are expected to show three bands, 6903 bp, 771 bp and 638 bp. In plasmids without an insert there are 2 fragments generated, 6774 bp and 638 bp. Lane 1: 5 μL 1 Kb Plus DNA ladder. Lanes 2, 4 and 6: 5 μL uncut plasmid, lanes 3, 5 and 7: 5 μL plasmid cut with *Bgl* II. Lanes 3 and 7 show the expected band pattern for plasmids containing an insert.

The concentration of the vector was determined by agarose gel electrophoresis to be approximately 7.5 ng/μL and that of the insert at approximately 7.5 ng/μL. After ligation, transformation and plasmid preparation the resulting plasmids were screened by digestion with *Bgl*II, which is expected to cut once within the insert and twice within the vector, giving bands at 6903 bp, 638 bp and 771 bp for plasmids with the insert, and bands at 6774 bp and 638 bp for plasmids without the insert (Figure 3.6C). A candidate plasmid (figure 3.6C, lane 3) with the expected restriction pattern was sequenced with the primer pTYB11-CBD. A candidate plasmid containing an error free insert was selected and named pTYB11.MBP.

3.4 Summary

This chapter details the steps taken to create four novel expression vectors, which enable protein expression in four different fusion contexts. The 365 bp *trx* gene from pET32a, and the 1175 bp *malE* gene from pMAL.c2g, was amplified by PCR with two different sets of primers, containing different restriction sites. This enabled the four different PCR products to be directionally cloned the two vectors, pTYB4 and pTYB11, replacing the gene for the CBD, and creating four novel vectors. Each vector expresses the target protein with a solubility tag, with the nature and position of the tag differing in each vector. The *malE* gene encoding the MBP or the *trx* gene encoding the Trx protein, were used as the solubility tags. In theory, each solubility tag can be separated from the target protein by intein mediated cleavage.

4   Cloning of Human Methylmalonyl CoA Mutase

4.1   Introduction

The overall goal of this work was to develop a source of soluble and active hMCM suitable for crystallisation trials. Past heterologous expression systems of hMCM in *E. coli* have produced insoluble inclusion bodies (Mark Patchett, *pers. com.*). hMCM was expressed in a soluble form in *E. coli* with a N-terminal Trx solubility tag, however this was accompanied by problems with the protease mediated Trx tag separation (Janata *et al.*, 1998). This showed that there was a requirement for an expression system that provides a solubility tag, but with an alternative method for separation of the tag.

The aim of this chapter was to introduce the hMCM coding sequence into four novel vectors to create four different hMCM fusion proteins for expression in *E. coli*. Each of the four plasmids will enable the hMCM to be expressed with one of two solubility tags, and each tag can be separated from the hMCM using intein-mediated proteolysis. This chapter describes the PCR amplification of the hMCM gene from the vector pMEXHCO, to enable ligation into the four novel vectors that were constructed in chapter three. The plasmids pTYB11.Trx.MCM and pTYB4.Trx.MCM both contain the Trx solubility tags, but in different relative positions, N or C terminal to hMCM respectively. The plasmids pTYB11.MBP.MCM and pTYB4.MBP.MCM both contain the MBP solubility tags, also in two different orientations, N and C terminal to hMCM respectively (figures 4.1 and 4.2).

4.2   Decreasing *In Vivo* Protease Degradation of the Recombinant Protein

When present at high local concentrations in *E. coli,* misfolded proteins typically form insoluble inclusion bodies. Proteins in these inclusion bodies are protected from proteases, as the aggregates prevent protease access (Cheng *et al*, 1981). By enhancing the solubility of the protein a whole new problem is created; the soluble protein can now be degraded as it is accessible for host proteases. There are two main types of problematic protease activity that cause degradation of recombinant protein expressed in *E. coli*: i) The polypeptide may be targeted for degradation if the N-terminal amino acid

dictates a short half-life (Tobias *et al.* 1991); and ii) Misfolded or foreign (aberrant) proteins recognised and degraded by *E. coli* proteases (Groll *et al.*, 2005). To prevent loss of recombinant protein due to protease degradation, a two-pronged approach was taken in this work. Appropriate N-terminal amino acids were selected to decrease the turnover rate of soluble hMCM, and solubility tags were used to limit aberrant folding of hMCM. The use of solubility tags is described in more detail in section 1.6.5.

Source of hMCM	N terminal amino acids	After Met processing	Half-life in *E. coli*
1. Native hMCM[†]	LH~	N/A	N/A
2. hMCM*	MLH ~	LH ~ (15%)	2 minutes
3. *trx*~hMCM*	MS ~**Tag**~ L H ~	S~**Tag**~ L H ~(84%)	10 hours
4. hMCM~**Tag**[+]	MVH ~	VH ~ (84%)	10 hours
5. **Tag**~hMCM[+]	M S ~Tag~ L H ~	**Tag**~LH ~(84%)	10 hours

Table 4.1 N terminal amino acid and the turnover rate of a protein in *E. coli*. From left, the source of the protein sequence information; the sequence of the first amino acids; the sequence after Met processing, with the percentage processing in parenthesis (Hirel *et al.*, 1989); and the expected half-life in *E. coli* (Tobias *et al.*, 1991). Protein 4 contains a C-terminal tag, and protein 3 and 5 have a N-terminal Trx tag.

[†]Protein purified from human liver or placenta (Fenton *et al.*, 1982; Kolhouse *et al.*, 1980).

* Protein expressed in *E. coli* (Janata *et al.*, 1997).

[+] Protein to be expressed in *E coli* in this work.

'**Tag**' can be the CBD, MBP or Trx tag.

In this hMCM *E. coli* expression system the N-terminal Leu of hMCM will be modified to encode a Val, when hMCM is expressed with a C-terminal solubility tag (table 4.1, compare proteins 2 and 4). Changing the N-terminal amino acid to Val increases the predicted half-life *in vivo* when compared to the hMCM native sequence. The MAP processing to remove the N-terminal methionine, when the amino acid is adjacent to a valine, is 83.7% (Hirel *et al.* 1989) so most of the Met-Val protein will be recovered with an N-terminal Val. This substitution is a conservative change between these amino acids and is unlikely to affect the structure and function of hMCM (Stryer, 1995)
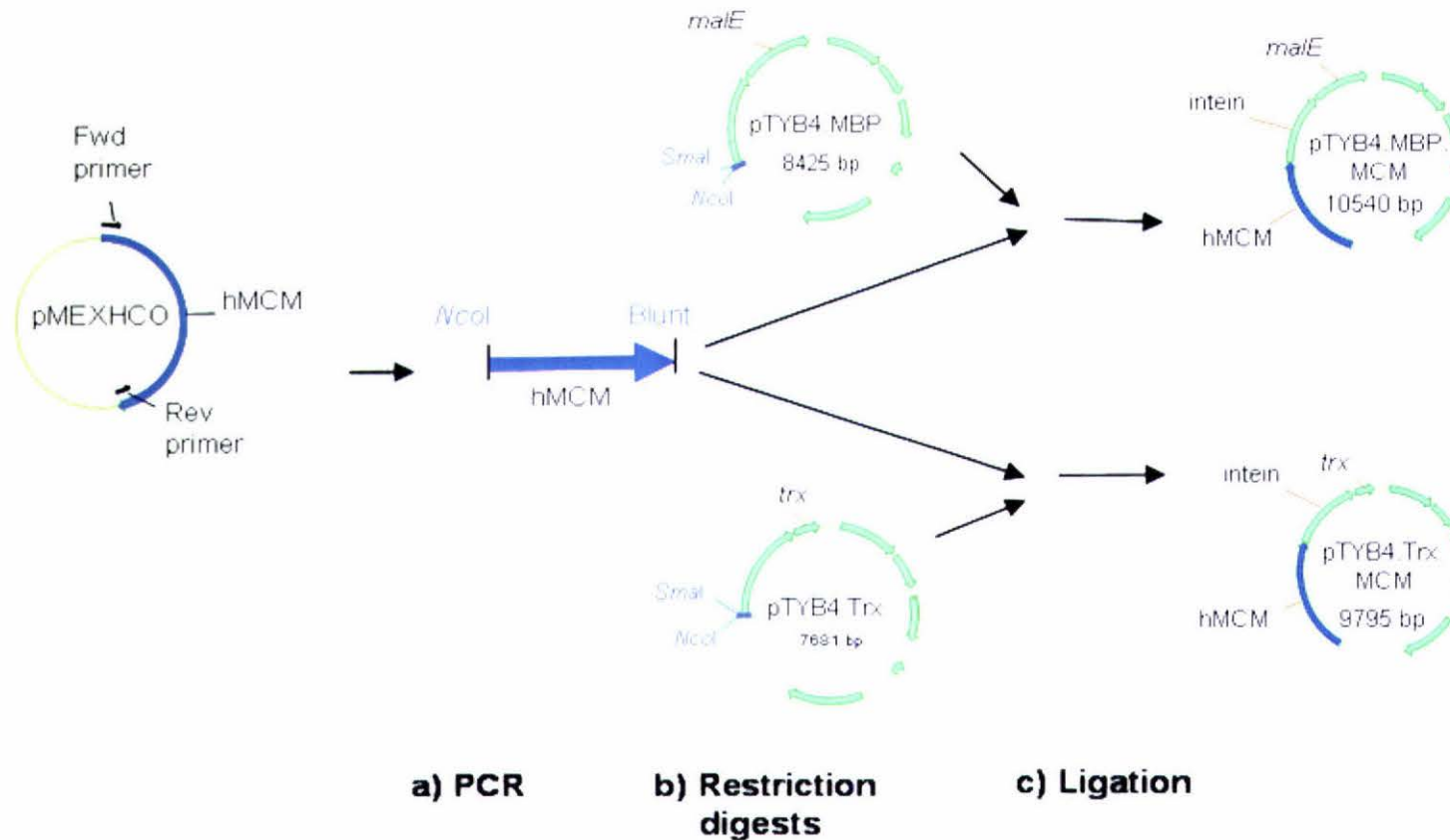
Figure 4.1 Flow diagram, showing the production of pTYB4.MBP.MCM and pTYB4.Trx.MCM. The first step is the PCR amplification of the hMCM gene from the vector pMEXHCO. The pTYB4.MBP and pTYB4.Trx vectors, and the two inserts, are digested in separate reactions, with *Age*I and *Pst*I. The hMCM insert was ligated into each of the vectors, producing the two unique expression vectors pTYB4.MBP.MCM and pTYB4.Trx.MCM .
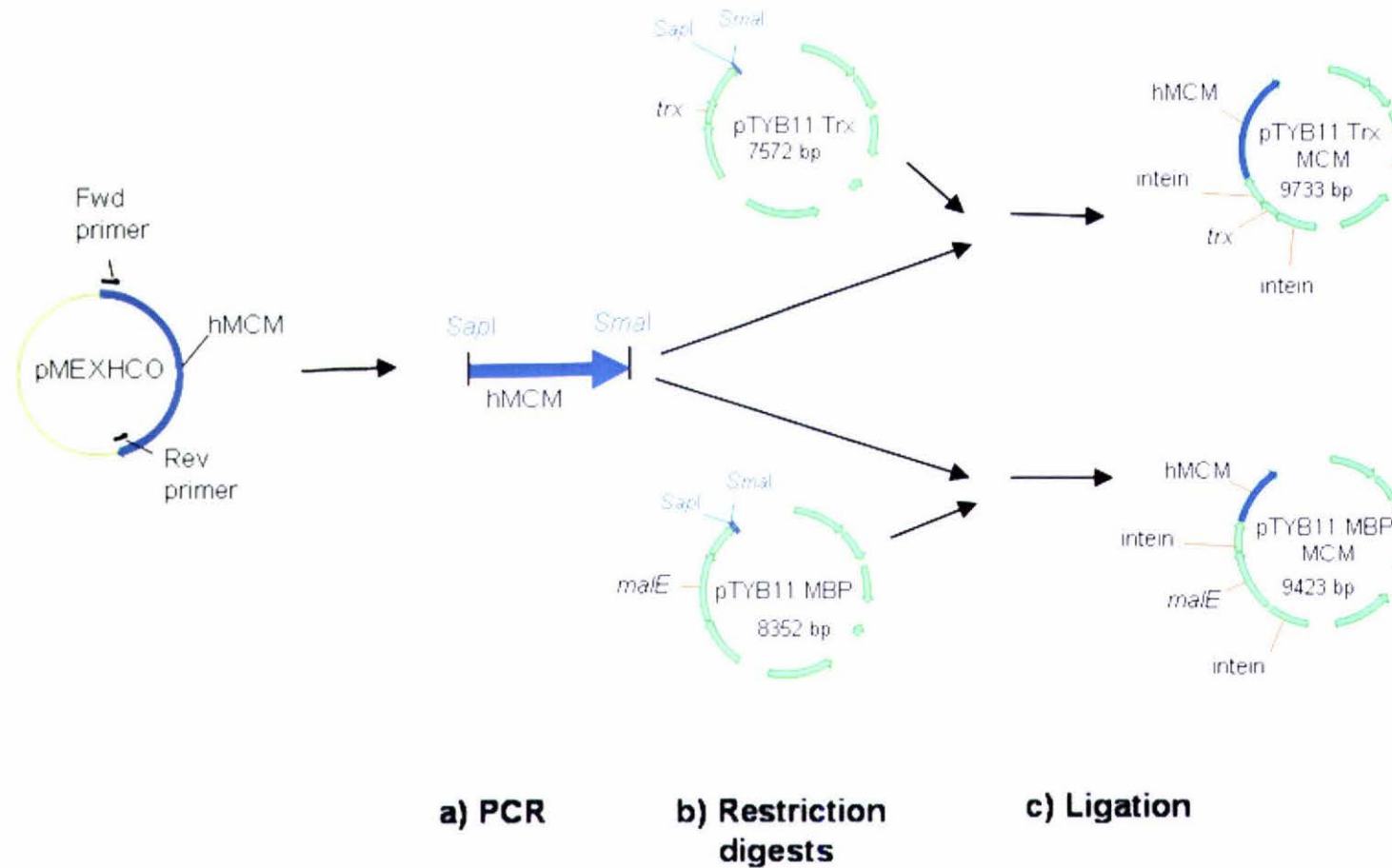
Figure 4.2 Flow diagram, showing the production of pTYB11.MBP.MCM and pTYB11.Trx.MCM. The first step is the PCR amplification of the hMCM gene from the vector pMEXHCO. The pTYB11.Trx and pTYB11.MBP vectors, and the two inserts, are digested in separate reactions with *Sap*I and *Sma*I. The hMCM insert was then ligated each of the vectors, producing the two unique expression vectors pTYB11.Trx.MCM and pTYB11.MBP.MCM.

4.3   PCR Amplification of hMCM

The hMCM gene from the plasmid pMEXHCO was amplified by three different sets of primers, under three different PCR conditions that were optimised for each reaction (table 4.2). Each primer pair incorporated vector specific restriction sites allowing the hMCM gene to be ligated into different multiple cloning sites. The incorporated restriction sites *Sma*I and *Nco*I allowed the hMCM to be ligated into pTYB4 and derived vectors, while the *Sap*I and *Sma*I sites allowed the hMCM to be ligated into pTYB11 and derived vectors. Further, a *Sac*I site in the forward primer allowed the hMCM to be ligated into pET32a.TEV (table 4.2).

Forward primer	Reverse primer	PCR components in 50 μL total volume	PCR conditions	PCR Product
ACFMCM4	ACRMCM4	1 x buffer 1mM MgSO$_4$. 0.3 pM/μL fwd primer 0.3 pM/μL rev primer 0.25 mM dNTP mix 0.04 ng/μL template 2.5 U DNA polymerase	94 °C for 2 min - one cycle. 94 °C for 15 sec, 55 °C for 30 sec, 68 °C for 30 sec - 30 cycles. 68 °C for 5 min - 1 cycle.	hMCM gene to be cloned into pTYB4 & derived vectors
ACFMCM11	ACRMCM11	2 x buffer 1 mM MgSO$_4$ 0.3 pM/μL rev primer 0.3 pM/μL fwd primer 0.25 mM dNTP mix 0.04 ng/μL template 0.6 U DNA polymerase	94 °C for 2 min - one cycle.94 °C for 15 sec, 60 °C for 30 sec, 68 °C for 30 sec - 30 cycles.68 °C for 5 min - 1 cycle.	hMCM gene to be cloned into pTYB11 & derived vectors
hMCM32 TEVf2	hMCMpET 32Sacrv	1 x buffer 1 mM MgSO$_4$ 0.3 pM/μL fwd primer 0.3 pM/μL rev primer 0.25 mM dNTP mix 0.04 ng/μL template 1.8 U DNA polymerase	95 °C for 2 min - 1 cycle. 95 °C for 15 sec, 55 °C for 30 sec, 68 °C for 30 sec - 30 cycles. 68 °C for 5 min – 1 cycle.	hMCM gene to be cloned into pET32a.TEV

Table 4.2 PCR Conditions for the hMCM Gene from the Plasmid pMEXHCO. After trialling a number of conditions, these were found to produce the optimum yield of the correct PCR product. See Section 2.2.6 for a more detailed description of the general set up of each reaction.

The pMEXHCO plasmid, used as a template in these PCRs, had been constructed by ligating the 2200 bp cDNA of the hMCM gene into the T7-7 vector. Two rare codons at the start of the gene caused the translation in *E. coli* to be inefficient (Robinson *et al.* 1984). The 'CO' in 'pMEXHCO' stands for 'Codon Optimised' as these two rare codons at the start of the hMCM ORF have been mutated: Leu (2) CTA to CTT, His (3) CAC to CAT. This manipulated coding sequence was shown to result in more efficient translation in *E coli*. Expression of the pMEXHCO plasmid yields full-length protein; however the product is insoluble and inactive (Mark Patchett, *Pers. com.*). In this work pMEXHCO was used as a template, this modified hMCM sequence was amplified by PCR and ligated into the different vectors with the intention of testing the solubility and activity of the recombinant hMCM in different fusion contexts (figures 4.1 and 4.2).

4.4   Cloning hMCM into pTYB4 and pTYB4-Derived Vectors

The insert was generated by PCR using a forward primer, incorporating the restriction site *Nco*I, and the reverse primer, with no restriction site, to amplify the hMCM from the plasmid pMEXHCO (table 4.2). The blunt end generated by the proofreading Platinum® *Pfx* polymerase during the PCR was then able to be ligated to the blunt end generated by *Sma*I digest of the vector. A PCR product of the expected size (2200 bp) was formed (figure 4.3A). The hMCM PCR product was digested with *Nco*I, then purified by gel electrophoresis. To prepare the three vectors, pTYB4, pTYB4.Trx and pTYB4.MBP, for the ligation each vector was digested in separate reaction with *Nco*I and *Sma*I, then purified by gel electrophoresis.

The vectors and the insert were quantified by gel electrophoresis; the concentration of vectors was determined to be 30 ng/μL, and that of the hMCM insert 7 ng/μL. After ligation, transformation and plasmid preparation, the candidate plasmids were digested with *Hind*III or *Bam*HI to determine whether they contained the insert (figure 4.3B, C and D). A plasmid from each of the three ligations giving the expected digestion pattern was sequenced, with the primers ACTYB1-4T7, hMCMSeq701fwd and ACTYB1-4IntR. Plasmids containing an error free inserted sequence were selected from each of the three ligations and named pTYB4.MCM, pTYB4.Trx.MCM and pTYB4.MBP.MCM.
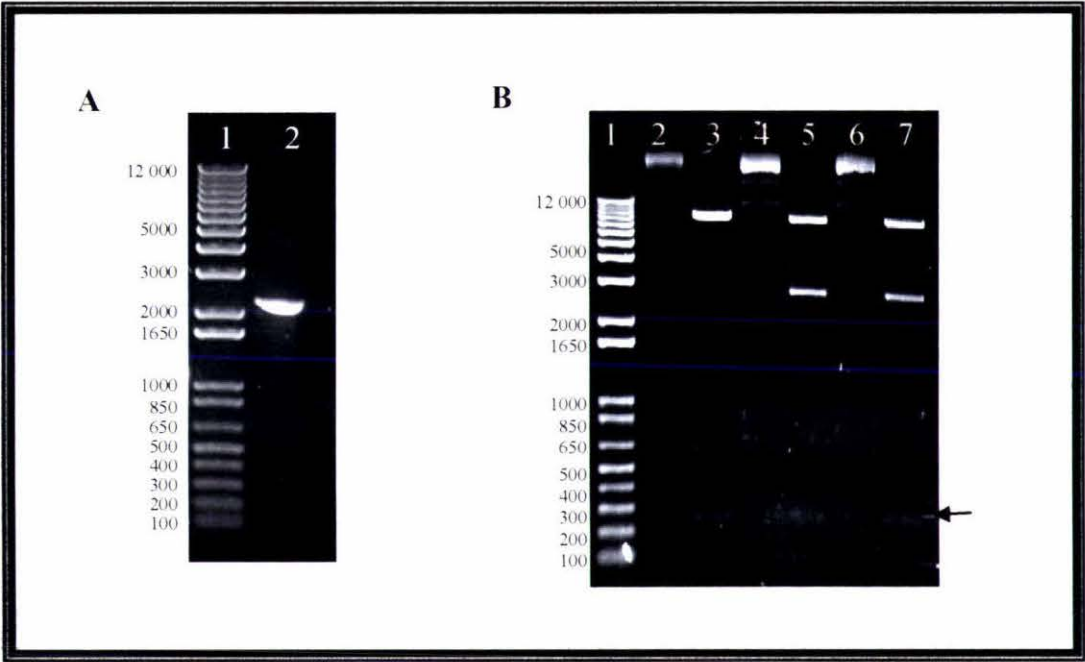
Figure 4.3 Agarose Gels Showing the Cloning of the hMCM gene into pTYB4

A. The PCR Product. The product is of the expected size (2200 bp) for hMCM. Lane 1, 5 μL 1 Kb plus DNA ladder. Lane 2, 2 μL from 50 μL of PCR product.

B. Cloning hMCM into pTYB4. A *Bam*HI digest of the candidate plasmids. The plasmids with hMCM inserts are expected to show three bands, 6600 bp, 2500 bp and 220 bp. In plasmids without an insert fragment of 6800 bp and 220 bp are expected. Lane 1, 5 μL 1 kb plus DNA ladder, Lanes 2, 4, 6; uncut plasmid, buffer only. Lanes 3, 5, 7; candidate plasmids digested with *Bam*HI. Lanes 5 and 7 show plasmids with the expected band pattern for a plasmid containing the hMCM insert
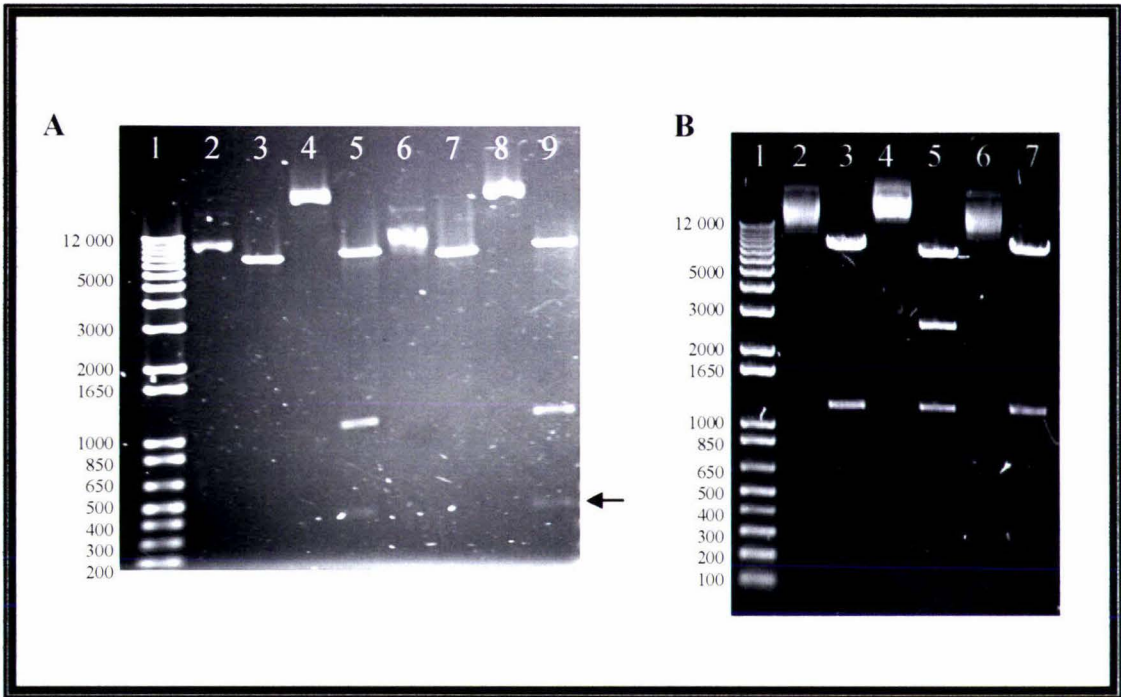
Figure 4.4 Agarose Gels Showing the Cloning of the hMCM gene into the pTYB4 Derived Vectors.

A. Cloning hMCM into pTYB4.Trx. A *Hin*dII digest of the candidate plasmids. The plasmids with hMCM inserts are expected to show three bands, 8300 bp, 1100 bp and 405 bp. In plasmids without an insert an 8300 bp fragment is expected. Lane 1, 5 μL 1 kb plus DNA ladder. Lanes 2, 4, 6, 8; uncut plasmid, buffer only. Lanes 3, 5, 7, 9; candidate plasmids digested with *Hin*dIII. Showing in lane 5 and 9 plasmid candidates containing the hMCM insert.

B. Cloning hMCM into pTYB4.MBP. A *Bam*HI digest of the candidate plasmids. Lane 1, 5 μL 1 kb plus DNA ladder. Lanes 2, 4, 6; uncut plasmid, buffer only. Lanes 3, 5, 7; vector digested with *Bam*HI. Lanes 5 and 7 show a plasmid with the expected band pattern for plasmids containing the hMCM insert.

4.5 Cloning hMCM into pTYB11 and pTYB11-Derived Vectors

The insert was generated by PCR using a forward primer, incorporating the restriction site SapI, and the reverse primer with the restriction site SmaI, were used to amplify the hMCM from the plasmid pMEXHCO (table 4.2). A PCR product of the expected size (2200 bp) was formed (figure 4.5A). The hMCM PCR product and the three vectors (pTYB11, pTYB11.Trx and pTYB11.MBP) were each digested with SapI and SmaI, purified and quantified (figure 4.5B). The observed size (~6500 bp) of these linearised vectors is slightly smaller than is expected (~7500 and ~8300 bp), however, there are no other bands on the gel. The concentration of the insert and all vectors were determined by agarose gel electrophoresis to be 7 ng/µL and 30 ng/µL respectively.
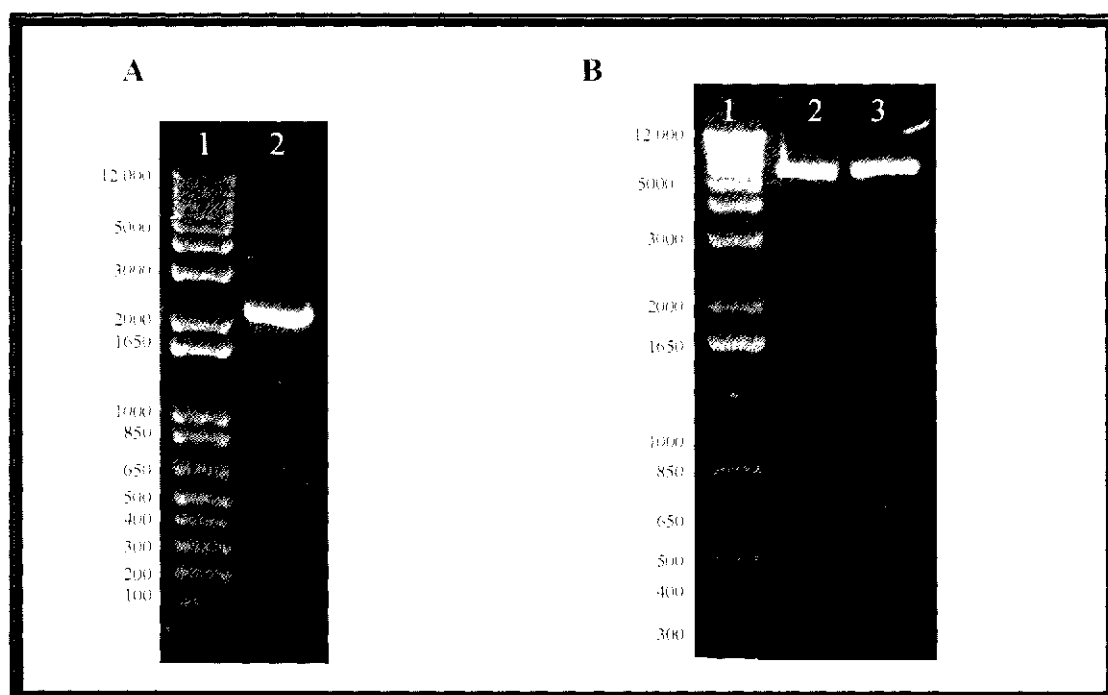
Figure 4.5 Agarose Gels Showing the Cloning of hMCM into pTYB11 derived vectors.
A. The PCR product. The product is of the expected size (2200 bp) for methylmalonyl CoA mutase. Lane 1, 5 µL 1 kb plus DNA ladder. Lane 2, 2 µL from 50 µL of PCR product.
B. The vectors pTYB11.Trx and pTYB11.MBP before ligation. Each vector has been digested with the restriction enzymes SapI and SmaI. Lane 1, 1 kb plus DNA ladder. Lane 2, 2 µL of pTYB11.Trx with an predicted size of 7572 bp. Lane 3, 2 µL of pTYB11.MBP with a predicted size of 8352 bp. For a size comparison see figures 3.6 and 3.8 for restriction digests also showing these vectors linearised.

Following the ligation. transformation and plasmid preparation the candidate plasmids were digested with *Hind*III to determine whether they contained the insert. The gels of the restriction digest showed a banding pattern consistent for the vector without the insert. The vector without is insert expected to give 1600 bp and 6000 bp fragments from a *Hind*III digest. and these fragments were formed in some restriction digests (e.g. figure 4.6A. lanes 4 and 7). In addition to the religated vectors there were plasmids that. when digested. gave fragments of 250bp. 450 bp. 750 bp and 6000 bp (figure 4.6). For the pTYB11.Trx.MCM plasmid. bands are expected at 400 bp. 1000 bp. 2300 bp. 6000 bp when digested with *Hind*III. None of the plasmids gave these expected fragment sizes (figure 4.6). More plasmid 'minipreps' were screened by restriction digest. again no correct constructs were identified. The ligation reaction was repeated a number of times with re-prepared insert and vector; the same unexpected pattern of bands was generated (summarised in table 4.3).
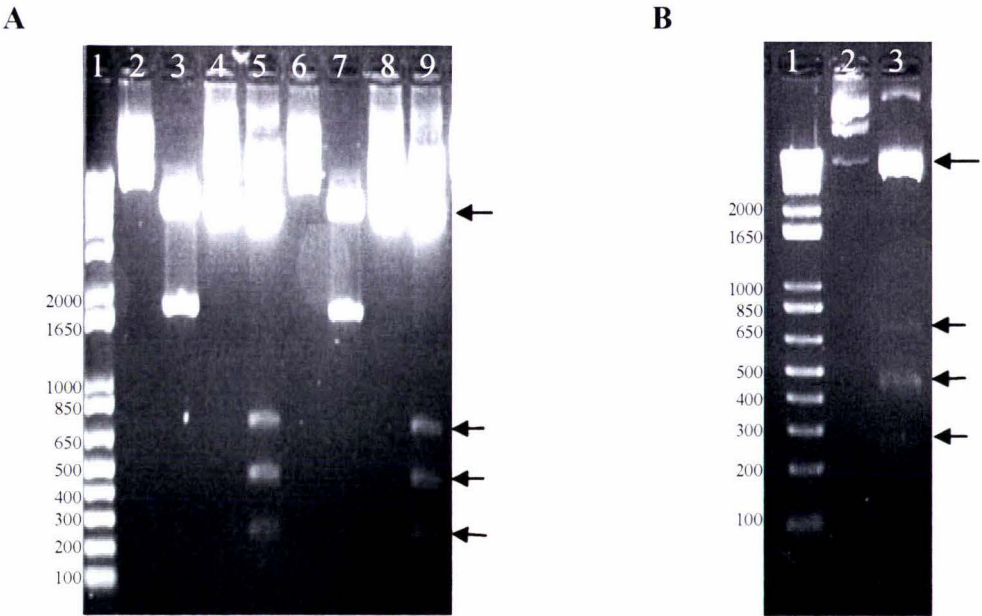
Figure 4.6 Agarose Gel Showing the Resulting Plasmids from the pTYB11 Ligations.
A. Candidate plasmids from the pTYB11.Trx and hMCM ligation, digested with *Hin*dIII.
Plasmids with no insert are expected to give fragments of 6700 bp and 1800 bp. Plasmids with an
insert are expected to gives fragments of 6000 bp, 2300 bp, 1000 bp and 400 bp. Lane 1, 5 μL 1
kb Plus DNA ladder. Lanes 2, 4, 6, 8; undigested plasmid. Lanes 3, 5, 7, 9; vector cut with
*Hin*dIII. Lanes 3 and 7 contain fragments with sizes consistent with the plasmid pTYB11.Trx
(without the hMCM insert). This gel also demonstrates the reproducibility of the *Hin*dIII digest
fragment sizes generated   from different colonies/ligation events.

B. Candidate plasmids from the pTYB11.MBP and hMCM ligation, digested with *Hin*dIII.
Plasmids with no insert are expected to give fragments of 6700 bp and 2300 bp. Plasmids with an
insert are expected to gives fragments of 6000 bp, 2300 bp, 1000 bp and 400 bp Lane 1, 5 μL 1 kb
Plus DNA ladder. Lane 2, undigested plasmid. Lane 3, plasmid cut with *Hin*dIII. Lanes 3
contain fragments with sizes consistent with the plasmid pTYB11.MBP (without the insert).
Arrows indicate the 250 bp, 450 bp and 750 bp fragments, which are inconsistent with the
expected fragment sizes.

HindIII digest of plasmids from the pTYB11.Trx/hMCM ligation		HindIII digest of plasmids from the pTYB11.MBP/hMCM ligation	
Expected Fragment Size	Apparent Fragment Size	Expected Fragment Size	Apparent Fragment Size
6000 bp	6000 bp	6000 bp	6000 bp
2300 bp	750 bp	3200 bp	750 bp
1000 bp	450 bp	1000 bp	450 bp
400 bp	250 bp	400 bp	250 bp

Table 4.3 The Expected *Hin*dIII Digest Fragment Sizes Compared to the Apparent Fragment Size. The plasmids generated from the ligation/transformation reaction were isolated and digested with *Hin*dIII, and the resulting fragments were not the expected size for plasmids either containing or lacking the insert.

A vector producing a restriction pattern consistent with the vector without the insert, (figure 4.6A, lane 6), was sequenced with the primer TYB11.MCSF. The sequence revealed the relegated vector, with no changes. One of these unexpected plasmids from the pTYB11.Trx and hMCM ligation, (figure 4.6B, lane 2) was sequenced with the primer TYB11.MCSF. The sequence revealed that half of the plasmid had apparently been lost (figure 4.5A). Obtaining sequence data for the plasmid preps resulting from the ligation/transformation proved difficult. One explanation for this is that the primer binding site required for sequencing was often lost. This also suggests that the plasmids prepared from pure cultures may not have been of a homogenous nature. This is supported by the restriction digests of plasmids may not have been linearised by *Hin*dIII (figure 4.6A and B). The sequence obtained, however, clearly originated from the pTYB11.Trx vector and enabled the sequence around the two recombination sites to be determined (figure 4.7B). The use of the word recombination in this context does not necessarily imply a particular mechanism for the loss/rearrangement of DNA (see appendix G, for a BLAST alignment with the vector).

A
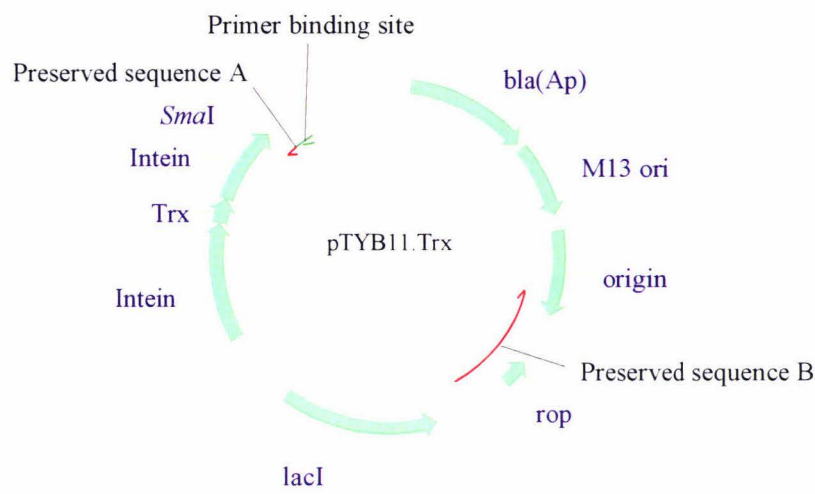
B
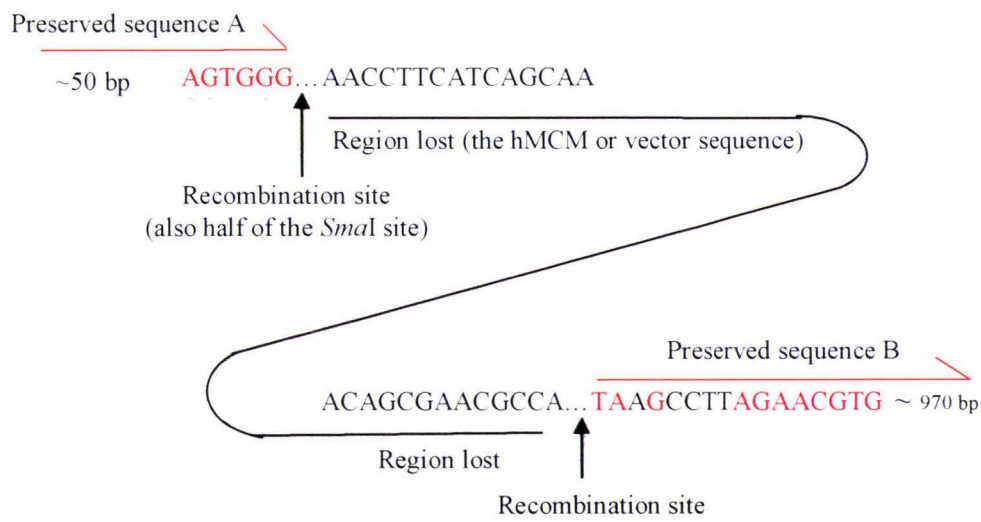
Figure 4.7 Sequence Analysis of a Plasmid from the Ligation of hMCM into pTYB11.Trx.
A. Vector map showing the recombination sites of in the plasmid pTYB11.Trx. Much of the
vector is lost after ligation and transformation. The region between the two preserved sequences
has been lost.
B. Sequences of the regions flanking the recombinant sites. Red bases indicates that this base
pair was confirmed by sequencing.

In the plasmid sequenced the *lacl*, hMCM and intein genes were lost, and all that remained were the ampicillin resistance gene and origin of replication. These two loci confer ampicillin resistance and allow the plasmid to he replicated in *E. coli*. The sequencing showed that the DNA sequence between the TYB11.MCSF primer binding site and the *Sma*I site is preserved but the vector sequence does not continue beyond the GGG of the *Sma*I. The sequencing recommences about 3500 bp later in the vector. The BLAST sequence alignment between the vector and the sequencing results can be found in appendix F. If this 3500 bp region of the plasmid is lost this would give a 4000 bp plasmid with one *Hind*III site. The smaller size of this plasmid, however, would give it a strong selection advantage over the larger plasmids, unlike the large plasmid product that is expected from this ligation. This is inconsistent with that fragments observed from the restriction digest. Given the difficulties gaining a sequence, however, it is possible that the single plasmid sequenced was not representative of all the plasmids from the ligation, but rather a small population that contained the primer binding site.

Similar products, as observed hy restriction digest, were also generated when attempts were made to clone hMCM into the original vector pTYB11 (data not shown). Therefore this instability is not a result of the substitution of the *malE* or *trx* for the CBD ORF, made when these novel vectors were constructed. Due to time constrains, testing these vectors using different restriction sites, in particular *Sap*I, or cloning in genes other than hMCM was not possible. This would determine if the products observed were due to instability generated by the sequence of the hMCM, or the enzymes used in the cloning.

Plasmid instability can be divided into two types. The first is segregation instability, and this is dependant on the plasmid copy number, and results in the total loss of the plasmid from the *E. coli* cells. The second type is structural instability, causing structural rearrangement within the vector sequence; this type of instability is consistent with the results summarised in figure 4.7 (Corchero and Villaverde, 1997). However this type of structural rearrangement is typically due to either i) active transcription of a plasmid-borne gene or ii) the presence of Chi sites in the DNA sequence. In this case, when the plasmid is in the *E. coli* XL-1 Blue strain, only two plasmid genes should be expressed, *bla* and *lacl*^{q}, and both are present on the original vector, which is stable. In the second type of DNA instability in *E. coli* certain sequences (called Chi sites) in

DNA can induce *rec*BCD dependant rearrangements (Arnold *et al.*, 2000). These rearrangements occur at the DNA sequence 5'-GCTGGTGG-3'. This sequence was not present in the vector; and in addition to this the recombination occurred in the *E. coli* *rec*A mutant. Therefore, it is highly unlikely that either of these two mechanisms is responsible for the observed instability of the plasmid products.

The *Sce* VMA1 intein is a mobile DNA element in its native form, and it is remotely possible that this property could increase the probably of recombination events in the plasmid. Instability with the same intein based vector (pTYB11) has been reported in other labs (Richard Scott. *pers. Comm.*). Although the intein DNA has only been shown to be mobile in a eukaryotic cell, and the endonuclease domain has been removed at the DNA level of the intein, certain DNA sequences remaining in the intein could enhance rearrangement of this DNA. This hypothesis is in conflict with the absence of instability seen in the pTYB4 vector and derivatives. These pTYB4-based vectors contain the same intein, and have a very similar arrangement of other vector elements.

The possibility of restriction enzyme star activity at a specific site on both vectors is possible. Only a small amount of star activity would be required, not quantities visible by gel electrophoresis, and the resulting smaller plasmids would be selected for in *E. coli*. This would explain the highly repeatable nature of the undersized plasmid products and one of the recombination sites also being the *Sma*I site. Star activity of restriction enzymes can occur in a number of different conditions in a reaction. Conditions that are known to enhance star activity include an elevated glycerol concentration and a high ratio of enzyme units to DNA. Ionic strength, metal ions (other that Mg^{2+}) and pH can also contribute to star activity of restriction enzymes (George *et al.*, 1980; Nasri and Thomas, 1986; Kessler and Manta, 1990). Upon further inquiry to New England Biolabs we were told:

> "While *Sap*I is not among the most likely of enzymes to exhibit Star Activity, it can display this loss of specificity if conditions are sufficiently permissive. In particular, the most likely cause is high glycerol concentration".

The initial glycerol concentration in the reaction was kept below 5 %, as is recommended in all restriction digests; however, if evaporation occurs during the incubation of the reaction this will raise the glycerol concentration. The digestion with both enzymes was repeated a number of times, and care was taken to avoid conditions

conducive to star activity. Following ligation, transformation and plasmid preparation the same undersized plasmid products formed.
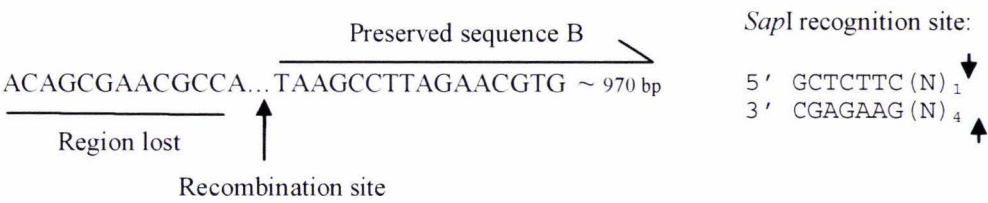
Figure 4.8 Sequence Analysis of the Recombination Site from the Ligation of hMCM into pTYB11.Trx. The sequence of the region flanking the recombinant site, compared to the *Sap*I recognition site. There is no sequence similar to the *Sap*I recognition site in the positions that could cause *Sap*I star activity at this site.

In the event that *Sap*I star activity was responsible for the undersized products, the presence of a sequence similar to that of the *Sap*I recognition site would be expected at the recombination site. On careful analysis of the sequence either side of the recombinant sites, a sequence similar to that of either restriction enzyme was not found (figure 4.8), indicating it was unlikely to be star activity that was causing these products but it can not be ruled out. If star activity is occurring, using different enzymes to clone the hMCM into the vector may alleviate these problems. However, the *Sap*I site must be used to clone the target protein next to the intein ORF, so using an alternative restriction enzyme site in this vector was not possible.

4.6   Cloning hMCM into the Vector pET32a.TEV

Expression of the hMCM with an N-terminal Trx tag was not possible with the intein pTYB11-based vectors, due to difficulty with cloning the hMCM into the vector resulting in the undersized plasmids. To circumvent the use of the pTYB11 intein system and associated restriction enzymes, attempts were made to clone the hMCM into the vector pET32a.TEV. The vector pET32a.TEV expresses the target protein with an N-terminal Trx tag, and this progresses the work of Janata *et al.* (1997), where the hMCM was soluble with an N-terminal Trx tag. However, in the system of Janata *et al.* (1997) the Trx was unable to be separated for the hMCM using enterokinase cleavage. The pET32a.TEV expresses the target protein with a site for the viral protease TEV between the Trx tag and the target protein hMCM. The TEV protease may access this site between the two components of the fusion protein more effectively than enterokinase.

The hMCM was amplified from the vector pMEXHCO with the forward primer hMCM32TEVf2 incorporating a *SacI* site, and the reverse primer hMCMpET32Sacrv with no restriction site to generate a blunt end. The restriction sites *BamHI* and *KpnI* in the multiple cloning site of pET32a.TEV were also present in the insert and so could not be used for cloning. Therefore, the vector was cut with *BamHI* and then end-filled with the Klenow fragment to create a blunt end to which the uncut PCR product could be ligated (figure 4.9).
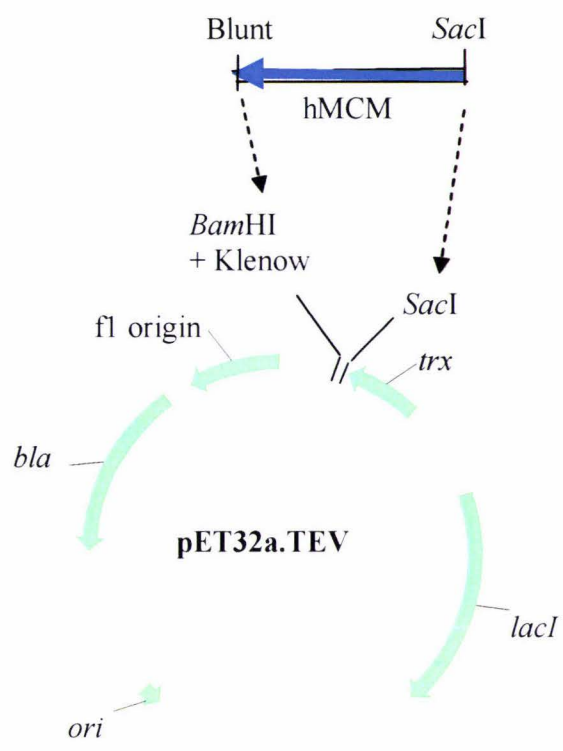
Figure 4.9 Strategy for Cloning the hMCM Gene into pET32aTEV. The insert was generated by PCR amplification and then digested with *Sac*I. The vector was digested with *Bam*HI, treated with the Klenow fragment, then digested with *Sac*I. The insert and vector were then incubated in a ligation reaction.
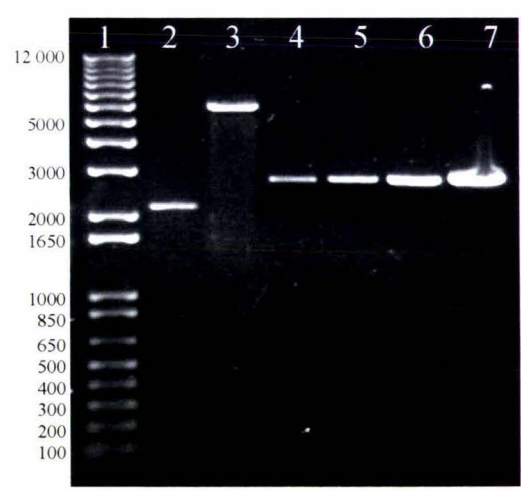
Figure 4.10 Agarose Gel of the Vector and Insert Immediately before Ligation. Showing the insert (lane 2) and vector (lane 3), before the ligation reaction.
Lane 1, 1 Kb Plus DNA ladder. Lane 2, 5 μL *Sac*I digested insert (hMCM). Lane 3, 5 μL *Bam*HI, Klenow and *Sac*I digested pET32a.TEV vector. Lanes 4, 5, 6, and 7; 10 ng, 20 ng, 50 ng and 100 ng quantiation standards, respectively.

The vector and insert were prepared, quantified. Both are the correct size and appear to be prepared correctly (figure 4.10, lanes 2 and 3). With Klenow treatment, however, it is difficult to determine if the reaction has been successful. No vectors containing inserts were obtained, despite many ligations and screening more than 60 candidate plasmids. due to time constraints, this cloning was halted.

4.7   Summary

The hMCM gene was successfully amplified in three different PCR reactions, creating three inserts, each with different restriction sites incorporated at each end. One of the PCR products was digested with *Nco*I, then cloned into the three different vectors, pTYB4, pTYB4.Trx and pTYB4.MBP. This created the three expression plasmids, pTYB4.MCM, pTYB4.Trx.MCM and pTYB4.MBP.MCM. It is expected that each of these plasmids will express hMCM with a different C-terminal tag, either with the CBD, Trx or MBP.

A second hMCM PCR product was digested with the restriction enzymes *Sap*I and *Sma*I, for cloning into pTYB11, and derived vectors. This, however, was not achieved, as the hMCM PCR insert was not able to be cloned into pTYB11 and derived vectors. These plasmids were expected to express hMCM as three different N-terminal constructs, either with the CBD, Trx or MBP. The cloning difficulties may be due to plasmid instability or problems associated with the cloning protocol. To circumvent these problems associated with the pTYB11 vector, the third hMCM PCR product was digested with *Bam*HI, treated with Klenow fragment, then digested with *Sma*I, for cloning into pET32a.TEV. This cloning was not able to be completed, due to time limitations.

5   Expression and Purification of hMCM

5.1   Introduction

The three hMCM expression plasmids, constructed in chapter four, enable the expression of hMCM in *E. coli* in three different protein fusion contexts. Each expression plasmid expresses hMCM with either the; (i) CBD, (ii) Trx or (iii) MBP C-terminal solubility tag. These constructs were designed to circumvent the solubility and protease cleavage problems associated with hMCM in past *E. coli* expression systems, (Janata *et al.*, 1997), and so all utilise intein mediated cleavage to separate the solubility tags. The aims of this chapter are to compare the solubility of hMCM-fusion protein expressed from these three plasmids and discuss any differences in intein activity and hMCM activity. The outcome of this expression will have a two-fold benefit. Firstly, to possibly identify a source of hMCM suitable for crystallisation trials that is a prerequisite for structural studies. If these vectors prove not to be a suitable source of hMCM, the information obtained by studying these expression systems can be used to design further expression systems of recombinant hMCM. Secondly, the success of pTYB4.Trx and pTYB4.MBP, the two novel intein-based vectors, will be discussed. These vectors may be suitable for the expression of other insoluble proteins in *E. coli*, and provide new folding and cleavage methods for recombinant expression that were not available prior to this work.

5.2   Expression from the Plasmid pTYB4.Trx.MCM

5.2.1   Introduction

Past expression of hMCM showed an N-terminal Trx fusion (expressed at 12 °C) resulted in high levels of a soluble Trx-hMCM fusion protein, however the Trx tag was not able to be removed at the enterokinase cleavage site (Janata *et al.*, 1997). The Trx tag is an effective solubility tag at the N or C terminus of many target proteins. It is, however, more commonly used as a solubility tag at the N-terminus of the target protein (LaVallie *et al.*, 1993). In this work the hMCM was expressed with the Trx tag at the C-terminal end of the fusion protein. This was thought to be less disruptive to dimer

formation, which occurs between the N-termini of hMCM monomers (see section 3.2.1 for a more detailed discussion).

5.2.2   Expression from the Plasmid pTYB4.Trx.MCM in *E. coli* ER2566

After the expression plasmid pTYB4.Trx.MCM was transformed into *E. coli* ER2566, to create the expression strain ER2566/pTYB4.Trx.MCM, cells were then induced with 0.6 mM IPTG. The time course of expression of the fusion protein at 37 °C was analysed by SDS PAGE at 0, 1, 2, and 4 hours. This showed that the recombinant fusion protein was being expressed and was of the expected molecular mass (data not shown). The conditions of expression were then changed in a stepwise manner, and the solubility of the fusion construct determined by SDS PAGE gel electrophoresis after making each change. The *E. coli* strain ER2566 was initially used as it is recommended by New England Biolabs for the expression from the plasmid pTYB4, from which pTYB4.Trx is derived. *E. coli* ER2566 also has the advantage of being a 'protease deficient' strain.
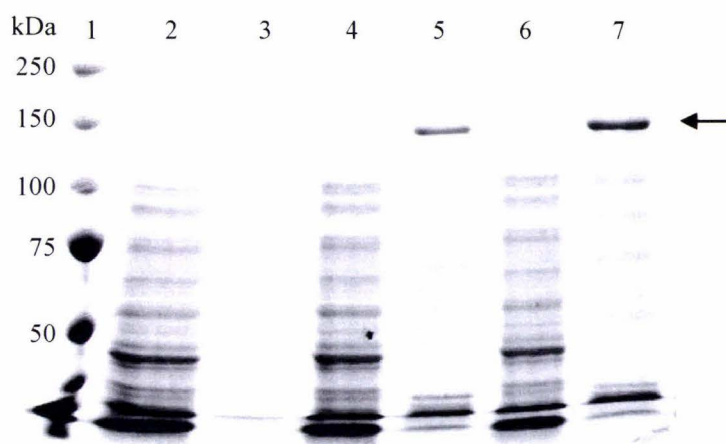
Figure 5.1 SDS PAGE Analysis of Protein Expressed from ER2566/pTYB4.Trx.MCM at 37 °C.

Lane 1: Molecular weight marker.

Lane 2: Cell lysate immediately before induction, soluble fraction.

Lane 3: Cell lysate immediately before induction, insoluble fraction.

Lane 4: Cell lysate 1 hour after induction, soluble fraction.

Lane 5: Cell lysate 1 hour after induction, insoluble fraction

Lane 6: Cell lysate 2 hours after induction, soluble fraction.

Lane 7: Cell lysate 2 hours after induction, insoluble fraction.

At 1 and 2 hours post-induction the recombinant protein is found in the insoluble fraction, lanes 5 and 7. The arrow indicates the position of the 148 kDa recombinant fusion protein.

Expression at 37 °C results in the formation of insoluble hMCM–intein-Trx fusion protein, with the expected mass of 148 kDa, in the insoluble fraction of the cell lysate 1 and 2 hours after induction (figure 5.1, lanes 5 and 7). The N-terminal Trx-hMCM fusion protein also remains insoluble at 37 °C (Janata *et al.*, 1997), indicating that changing the position of the Trx tag does not improve the solubility of the hMCM when expressed at 37 °C. In the insoluble fractions (lanes 5 and 7), bands at sizes that do not correspond to the recombinant protein size are observed. These may be *E. coli* proteins that are trapped in the inclusion bodies (Hayes, 1998). The fusion protein was also insoluble when expressed at 22 °C and 15 °C (data not shown).
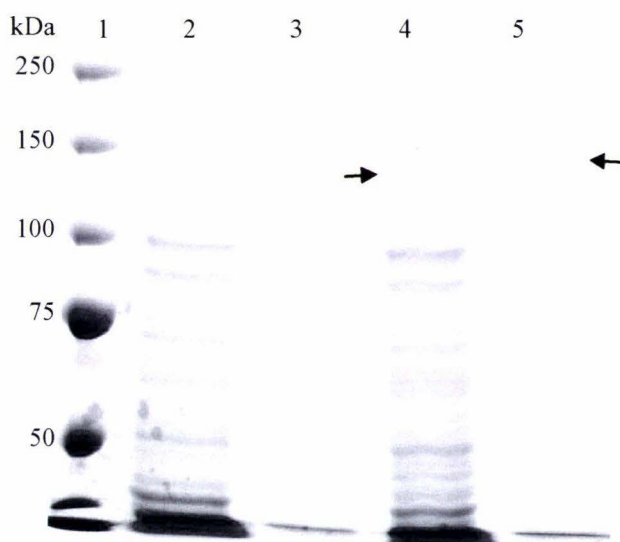
Figure 5.2 SDS PAGE Analysis of Protein Expressed from ER2566/pTYB4.Trx.MCM at 12 °C for 22 hours.

Lane 1: Molecular weight markers.

Lane 2: Cell lysate immediately before induction, soluble protein fraction.

Lane 3: Cell lysate immediately before induction, insoluble protein fraction.

Lane 4: Cell lysate 22 hours after induction, soluble protein fraction.

Lane 5: Cell lysate 22 hours after induction, insoluble protein fraction.

Approximately half of the fusion protein is expressed as soluble full length protein. Arrows indicate the position of the 148 kDa recombinant fusion protein.

When ER2566/pTYB4.Trx.MCM is expressed at 12°C, a comparison of the soluble and insoluble fractions, shows the band at 148 kDa in the soluble fraction (figure 5.2, lane 4, arrow). This band is absent in the uninduced fraction in lane 2, indicating that this 148 kDa protein in the soluble fraction is the soluble recombinant fusion protein. Expression of the fusion construct at 12 °C for 22 hours results in about half of the recombinant protein being present in the soluble fraction of the lysate (figure 5.2). *E. coli* proteins are not observed in the insoluble fraction when protein is expressed at 12 °C, (figure 5.2 lane 5) as is seen in the inclusion bodies formed at 37 °C (figure 5.1, lane 7). This suggests that the nature of the inclusion bodies is differs between these two growth temperatures.

The large difference in fusion protein levels seen between the expression at 37 °C and at 12 °C may be explained by *in vivo* protease degradation. This large difference in yield is

not observed in other *E. coli* expression systems when recombinant protein is expressed at different temperatures. For example Xu et al (2005) recovered similar amounts of recombinant protein from expression at 15 °C or 37 °C. The yield of 148 kDa protein from 12 °C expression for 22 hours (figure 5.2) is significantly reduced when compared to 37 °C expression for 2 hours (figure 5.1).

5.2.3   Purification of the hMCM-Intein-Trx Fusion Protein Expressed in *E. coli* ER2566

The recombinant fusion protein containing the His-tag was purified using a Ni-NTA column (see figure 5.3). The Ni-NTA resin selectively binds to the recombinant protein containing the His-tag and the rest of the cell's protein, ideally, does not bind (figure 5.3, lane 3). Following binding, the resin was washed four times with a buffer containing 30 mM imidazole, to remove any loosely associated proteins (figure 5.3, lanes 4-7). The recombinant His-tagged protein was then eluted with a buffer containing 200 mM imidazole (figure 5.3, lane 8). The imidazole competes with the His-tag for Ni binding, disrupting the Ni-His interaction to elute the His-tagged fusion protein.
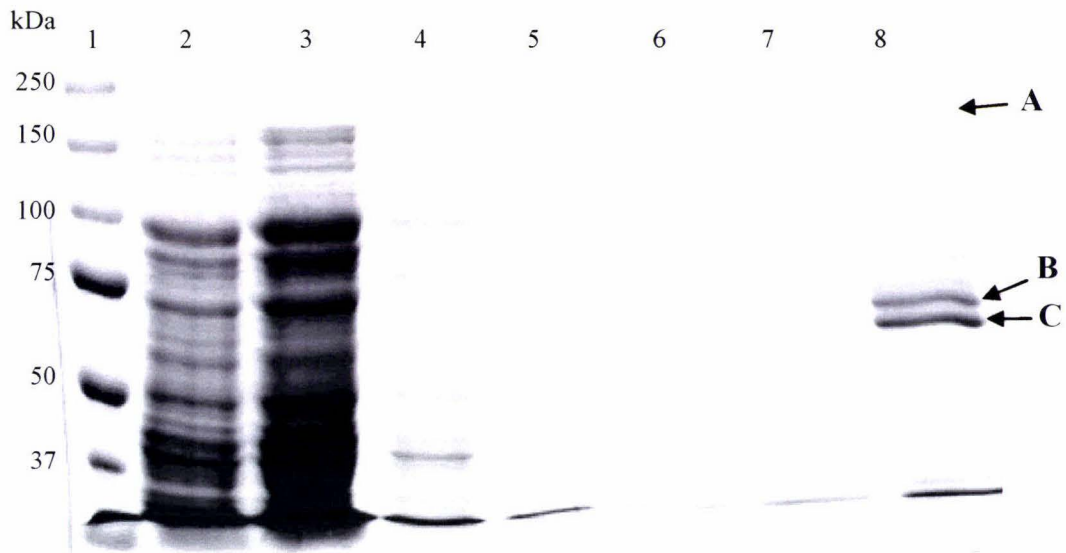
Figure 5.3 SDS PAGE gel of the Purification on a NTI-Ni Column of Recombinant Protein Expressed from Vector ER2566/pTYB4.Trx.MCM Cells.

Lane 1: Molecular weight markers.

Lane 2: Soluble fraction of the cell lysate.

Lane 3: Flow though from column.

Lane 4, 5, 6 and 7: The flow though from washes 1, 2, 3 and 4, respectively.

Lane 8: Eluted protein in 200 μL of 200 mM imidazole.

Arrows indicate purified protein:

A. The protein band is the size of the hMCM-intein-*trx* 148 kDa fusion protein.

B. The protein band is size expected for the B_{12}– intein-Trx fragment.

C. The protein band is the size of the intein-Trx fragment.

The fraction eluted with 200 mM imidazole, (figure 5.3, lane 8), contains proteins of lower molecular weight than the expected 148 kDa fusion construct. Given the apparent high affinity of these products for the Ni-NTA resin, it is likely that these proteins also possess the His-tag and are formed when the fusion protein is degraded.

Whether this degradation occurred *in vivo* or *in vitro* is of considerable importance. The cause of this degradation must be determined; to allow conditions to be altered to prevent degridation. To prevent *in vitro* metalloprotease activity, the metal ion chelator EDTA can be added to the cell lysis buffer. This inhibits the metalloprotease activity that can degrade proteins *in vitro*. EDTA, however, can not be used in Ni affinity column buffers, as the ETDA will strip the Ni from the Ni-NTA resin. Other protease

inhibitors, such as PMSF, were used instead of EDTA. In addition to this protein solutions were always kept below 5 °C, either on ice or in a refrigerated centrifuge, to limit any protease activity. However, if degradation occurs *in vivo*, adding protease inhibitors to the lysis buffer will not prevent the degradation that has already occurred.

If *in vitro* degradation caused these truncated products, the degradation occurred before loading onto the Ni-NTA resin, as no further degradation is observed during or after purification (figure 5.3). This deduction leaves the sonication and centrifuge steps as the possible source of these truncated products. However comparing the cells lysed by addition of a denaturing buffer and loaded immediately on a gel compared with the sonicated and centrifuged cell lysate, there is no decrease in the yield of 148 kDa, which would be expected if this fusion protein was being degraded during these processes (figure 5.4) In light of this evidence, *in vitro* degradation seems an unlikely mechanism for the production of these truncated products.

Figure 5.4 Comparison of the hMCM-Intein-Trx Fusion Protein Levels Before and After Sonication/Centrifuge Treatment. The level of the fusion protein was not significantly change due to the treatments.

Lane 1: Molecular weight markers.

Lane 2: Cells lysed by addition of a denaturing buffer and loaded immediately.

Lane 3: Sonicated and centrifuged cell lysate, soluble fraction

Lane 4: Sonicated and centrifuged cell lysate, insoluble fraction

These contaminants may be products of *in vivo* protease degradation, are expected in the cell lysate before purification and to be absent in the cell lysate from uninduced cells, but this is not observed (figure 5.2, lanes 2 and 4). There are many *E. coli* proteins of a

similar molecular weight in the soluble fraction at this molecular mass, and a change in the protein levels in this region would be difficult to determine. The degraded products may purify more efficiently, thereby distorting the relative quantities of full-length and degraded product. This may also explain why the degraded products are not observed before purification. The band of the same observed mass as the intein-Trx fragment could have also been produced by premature intein cleavage. This possibility will be discussed further in the following section on intein cleavage.

5.2.4 Intein-Mediated Cleavage of the hMCM-intein-Trx Fusion Construct

To perform self proteolysis, releasing the intein-Trx fragment from the hMCM, the protein is incubated with the reducing agent DTT (figure 5.5).
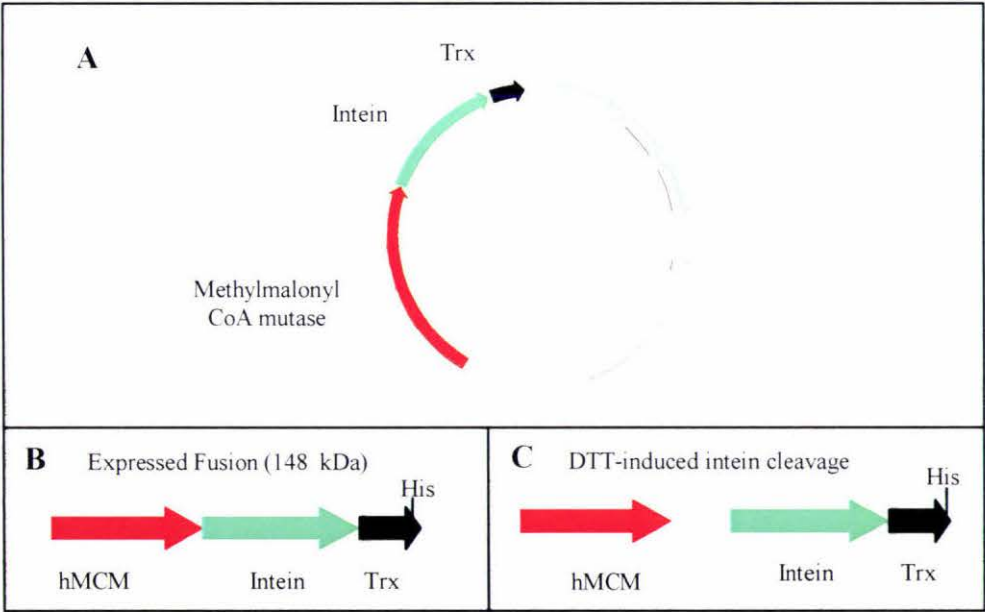
Figure 5.5 Schematic of the Fusion Protein Expressed from ER2566/pTYB4.Trx.MCM.
A: The vector structure (DNA). **B:** The fusion protein expressed from this vector, containing the His-tag used for purification. **C:** The recombinant protein after intein-mediated cleavage.

The Ni-NTA column-purified protein (from the previous section) was incubated overnight with 1 mM DTT, at 4 °C initiating the intein cleavage. In addition to this reaction, two controls were performed. For the first control purified protein was denatured by mixing with SDS-PAGE loading buffer immediately after elution from the

Ni-NTA column; denaturing the fusion protein prevents intein cleavage. In the second, background control the purified fusion protein was incubated at 4 °C without DTT, to detect any changes to the proteins not due to the DTT. Incubations were typically for 18 hours (figure 5.6).
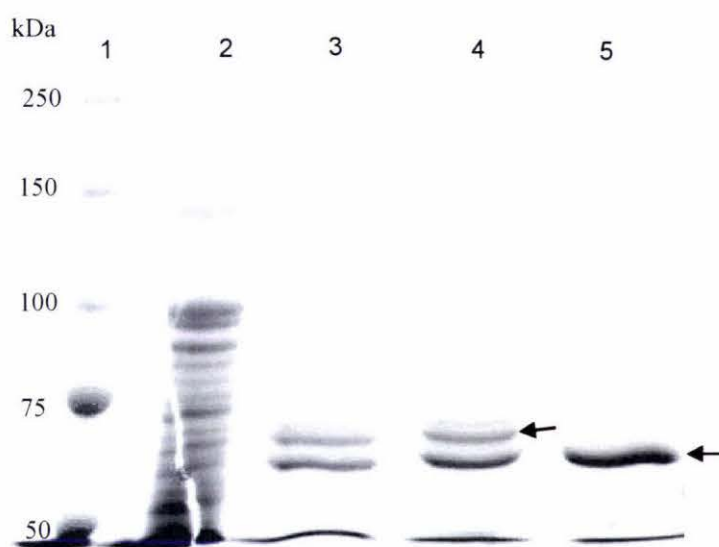
Figure 5.6 SDS PAGE gel showing Products from DTT-Induced Intein-Mediated Cleavage.

Lane 1: Molecular weight markers.

Lane 2: Soluble fraction of the cell lysate

Lane 3: Negative control; protein denatured immediately after Ni-NTA purification.

Lane 4: Background control; purified protein incubated for 18 hours at 4 °C without DTT.

Lane 5: Purified protein incubated for 18 hours at 4 °C with 1mM DTT.

When lanes 4 and 5 are compared the DTT dependent disappearance of the 148 kDa band is observed. In addition, also note the DTT-dependant of disappearance of two ~65-70 kDa bands.

The 148 kDa fusion protein (figure 5.6A, lane 4) was cleaved when the sample was incubated in 1 mM DTT (figure 5.6A, lane 5). The hMCM-intein-Trx fusion protein is cleaved into the hMCM and intein-Trx fragments (figure 5.4B and 5.4C). A band corresponding to the expected size of the intein-Trx fragment (66 kDa) was present in all lanes, and increased in intensity upon DTT treatment (figure 5.6A). The 82 kDa hMCM (without a fusion partner) is not expected to be purified, as the His tag is on the fusion partner. The hMCM, generated from the cleavage of the 148 kDa fusion protein, can be seen as a 82 kDa band formed at about a 1:1 ratio as judged from the relative intensities of the pre-DTT 148 kDa and post-DTT 82 kDa bands (figure 5.6, lanes 4 and

5). The purified and DTT-treated protein was run through a second Ni-NTA column, in an attempt to purify the cleaved hMCM. All the other products were expected to bind to the Ni-NTA resin. However, due to the very low levels of hMCM applied the column there was no protein in the flow though, as observed on an SDS PAGE gel (data not shown).

Upon DTT-triggered intein cleavage, one of the most abundant degradation products, (indicated with an arrow in figure 5.6, lanes 4), is lost, while the intensity of a ~10 kDa smaller band increased (indicated with an arrow in figure 5.6, lane 5). This smaller band is likely to be the intein-Trx fragment of the fusion protein. Therefore it is likely that most of the hMCM-intein-Trx fusion protein synthesised in the host is degraded *in vivo* to the bands seen in lane 4 (figure 5.6). This band could be formed by *in vivo* intein-mediated proteolytic cleavage, or by the action of *E. coli* proteases. The intein in the original vector, pTYB4, was not expected to have *in vivo* intein cleavage (Perler, 2000). It is possible, but unlikely, that the changes made to the fusion protein when the CBD was replaced with the Trx could affect the structure and/or activity of the intein. If this protein band was a product of intein cleavage, Edmans' sequencing could determine the N-terminal amino acids at the intein cleavage site if this intein-Trx fragment was a product of *in vivo* intein activity. However, if the protein band represents a population of fragments from proteolysis that are heterogeneous at the N-terminus, a sequence would not be able to be determined. Due to cost and time constraints the sequencing was not done. A summary of the predicted recombinant protein products is shown in figure 5.7.
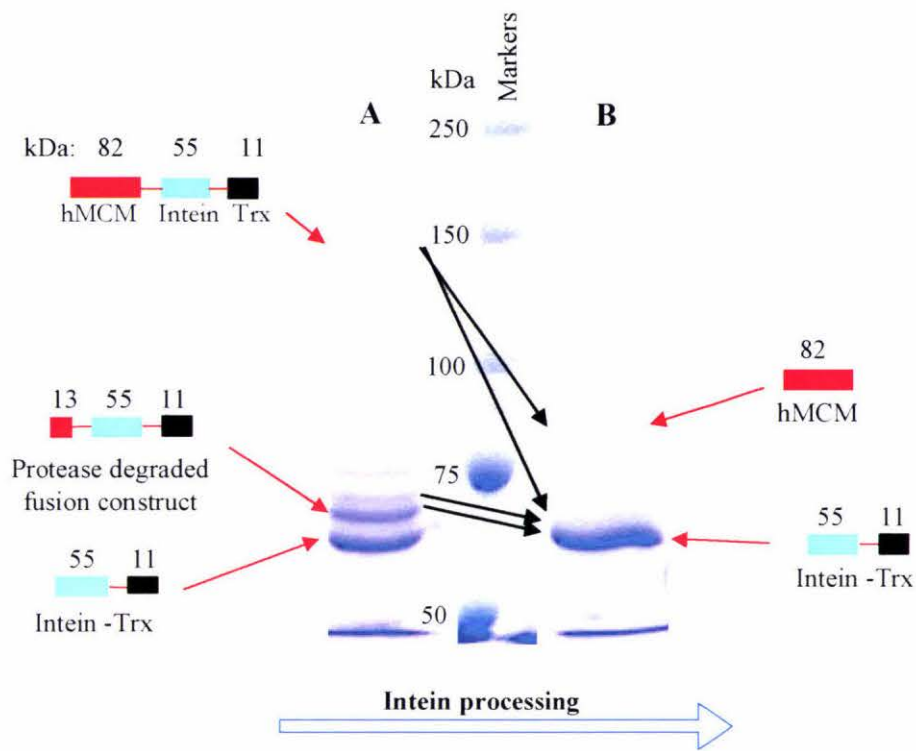
Figure 5.7 Summary Showing the Predicted Recombinant Protein Products and Intein-Mediated Cleavage.

A. Protein eluted from Ni-NTA column and incubated overnight at 4 °C.

B. Protein eluted from Ni-NTA column and incubated at overnight at 4 °C with 1 mM DTT

The band that is lost on incubation with DTT is approximately 10 kDa larger that the intein-Trx fragment. This larger band thought to be the intein-Trx with about 10 kDa of the C-terminal region of the hMCM remaining. This 10 kDa region of hMCM is then cleaved from the intein-Trx fragment upon incubation with DTT, resulting in the production of two polypeptides; the 66 kDa intein-Trx fragment and the 10 kDa hMCM fragment. To determine whether the 10 kDa product was formed, a 10% polyacrylmide SDS PAGE gel was run. Figure 5.8, lane 2 shows a diffuse band of approximately 11-14 kDa in the DTT treated sample.
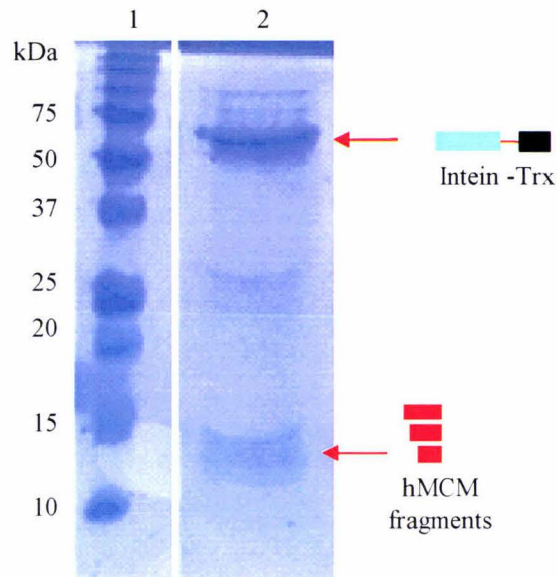
Figure 5.8 SDS PAGE gel (10% acrylamide) Showing the Formation of the Lower Molecular Weight Products. The intein-Trx fragment at 66 kDa, and the ~13 kDa fragments, formed in a DTT dependent manner.

Lane 1: Molecular weight markers.

Lane 2: Purified protein incubated overnight with 1 mM DTT.

Arrow indicates the ~13 kDa fragments. The heterogeneous size/diffuse nature of the band supports the hypothesis that this is a product of the proteolysis of a more susceptible region i.e. an unstable mutase domain, rather than cleavage at a specific site recognised by proteases.

A large region of hMCM, the ~69 kDa mutase domain, was most likely degraded by *E. coli* proteases, leaving the adenosylcobalamin (B_{12}) binding domain (figure 5.9). This indicates that the mutase domain of hMCM is prone to misfolding and therefore may be unstable in *E. coli*, while the smaller B_{12} binding domain is a stably folded domain of the hMCM. Therefore this makes the B_{12} binding domain is more likely to fold correctly and independently of the mutase domain. The adenosylcobalamin binding region is a structurally conserved domain, found in archaea, bacteria and eukaryotes, so independent folding and stability of this domain when expressed in the bacterium *E. coli* is not unexpected. Future directions could include expression of the mutase and B_{12} binding domains as separate polypeptides, as the case for several archaeal MCMs (Robb *et al.*, 2001).
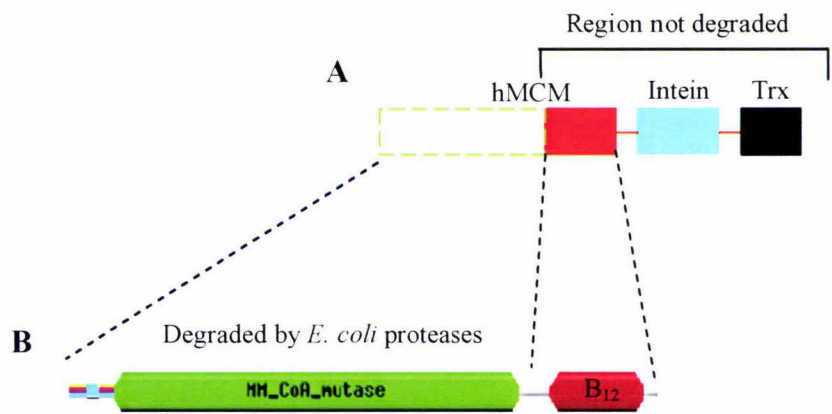
Figure 5.9 Model of the hMCM Domains Within the Fusion Protein.
A. A Major Product of *In Vivo* Proteolysis. The protein fusion formed in *E. coli*. The dotted green box indicates the region of the hMCM often degraded *in vivo*, the ~69 kDa mutase domain.
B. The pFam domains of hMCM. This shows the position of the 13 kDa B_{12} binding domain.

Similar degraded protein products were observed by De Marco *et al.* (2004) when expressing three insoluble proteins in *E. coli*. A GST tag improved the solubility of the target proteins; but these soluble proteins were truncated. The authors suggest a number of possibilities for the formation of these truncated products, including: i) unstable binding of the mRNA to the ribosome; or ii) "fragility" of the linker region between the GST and target protein – presumably this refers to the possibility that the linker region is more susceptible to proteolysis; or iii) possible dimerisation properties of the target protein. However, in the ER2566/pTYB4.Trx.MCM expression system, termination due to ribosome instability is not a possibility because the N-terminal portion of the fusion, first to be translated, is missing (figure 5.10).

A

B

+ termination

+ termination

GST Tag   Target Protein

Observed
Products

Mutase   Intein   Trx
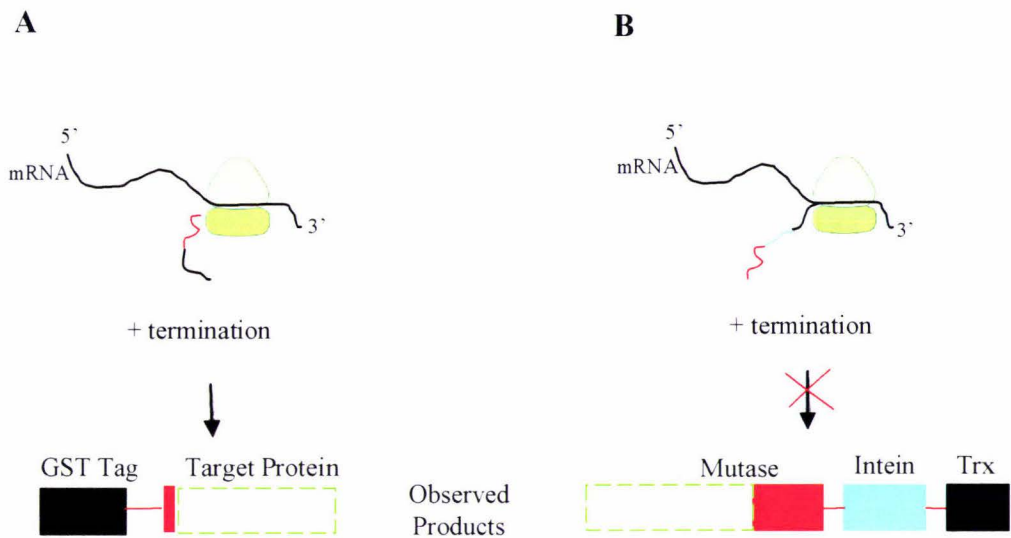
Figure 5.10 Observed Protein Products Were Not a Result of Premature Translation Termination of the Ribosome. The degraded region is indicated by the green dotted line.
A. In the system of De Marco *et al.* (2004), unstable ribosome binding is possible, as the authors suggest, because the target protein is truncated is on the C-termini of the fusion protein.
B. The pTYB4.Trx.MCM system is truncated at the N-terminus; this is not a product of the early termination of translation.

5.2.5 Expression of the Plasmid pTYB4.Trx.MCM in *E. coli* Rosetta-gami™ 2

Due to the insolubility and degradation of the recombinant protein in *E. coli* ER2566 in the previous section, a second expression strain was trialed in the hope of improving the solubility and reducing the proteolysis. The *E. coli* strain Rosetta-gami 2 contains a plasmid (pRARE^{CM}) with genes encoding the tRNA for rare codons. These cells also have a more favorable redox potential for disulfide bond formation in the cytoplasm and the Tuner™ genotype, which allows more proportional response to induction of expression by IPTG (section 1.6.4, '*E. coli* Strains' are discussed in more detail). The improved control of expression rates means that it is possible to decrease the local concentration of folding intermediates: decreasing the growth temperature often improves solubility by a similar mechanism. As dropping the temperature has improved the solubility of the expressed protein, this was a logical approach to take. There are two processes affected when the temperature is dropped, the rate of transcription and the rate of translation are both decreased. The energetics of the protein folding may also be affected. Lowering the transcription rate, that is achieved by the Tuner^{TM} strain, decreases the transcription rate. This decrease lowers the substrate levels available for translation, mRNA. However, the rate of protein folding is not reduced. In addition to these properties this *E. coli* strain contains rare codons within the hMCM gene may cause pausing of the ribosome during translation due to low tRNA availability. This pausing could cause misfolding or lower expression levels of the hMCM (Looman *et al.* 1987). The *E. coli* Rosetta-gami 2 cells were induced at relatively low IPTG concentrations (0.2 mM), and expressed at 12 °C for 22 hours. No differences in yield or solubility were observed when purified cell lysate from Rosetta-gami 2 was compared by SDS PAGE to cell lysate from the *E. coli* ER2566 strain (compare figure 5.11, lane 8 and figure 5.3, lane 8).
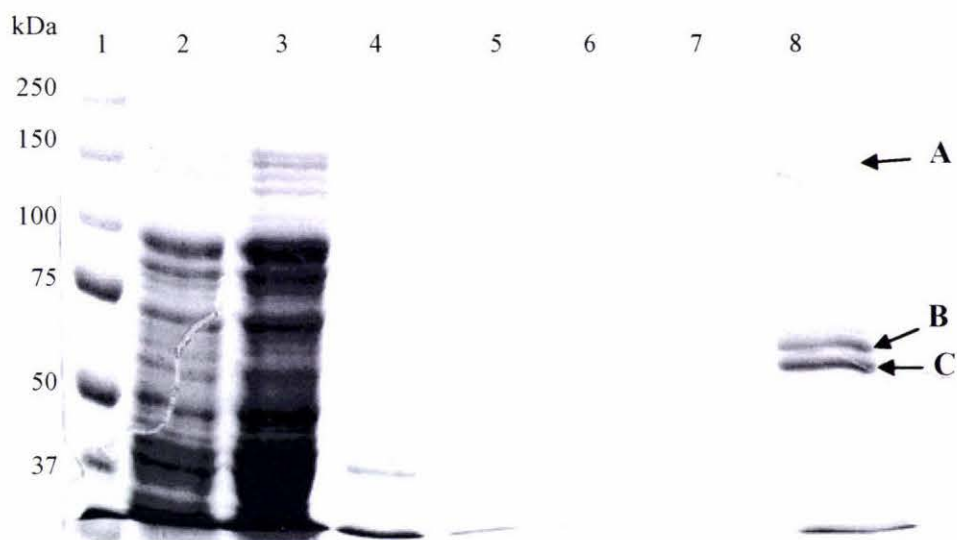
Figure 5.11 SDS PAGE gel of the Purification on a NTI-Ni Column of Recombinant Protein Expressed from Rosetta-gami 2/pTYB4.Trx.MCM Cells.

Lane 1: Molecular weight markers.

Lane 2: Soluble fraction of the cell lysate.

Lane 3: Flow though from column.

Lane 4, 5, 6 and 7: The flow though from washes 1, 2, 3 and 4 respectively.

Lane 8: Eluted protein in 200 µL of 200 mM imidazole.

Arrows indicate purified protein:

A. The protein band is the size of the hMCM-intein-*trx* 148 kDa fusion protein.

B. The protein band is size expected for the B_{12}– intein-Trx fragment.

C. The protein band is the size of the intein-Trx fragment

There was little apparent difference in size and abundance between in the purified products from the two *E. coli* strains strongly suggests that pausing due to rare codons was not a significant cause of the insolubility and degradation of the recombinant hMCM in *E. coli*. In addition to this, the improved control of expression and consequent decrease in [mRNA] and therefore a decrease in the rate of translation provided by the Tuner® phenotype, can be ruled out under these expression conditions as an effective method for improving the solubility of the hMCM. Based on these findings, the *E. coli* ER2566 strain was used for all further experiments.

5.2.6 Expression in the Presence of Adenosylcobalamin

It was hypothesised that adenosylcobalamin would help to stabilise the B_{12} binding domain, by improving the stability and folding of hMCM. Thus ER2566/pTYB4.Trx.MCM cells were grown in media containing adenosylcobalamin, the cofactor for hMCM. Adenosylcobalamin was added to the *E. coli* growth media to final concentration of 1 mM, half an hour before the addition of IPTG. The cells were then incubated for 22 hours at 12 °C. Due to the red colour of the adenosylcobalamin the cell pellet was light pink, and the cell lysate, after sonication, was also light pink.
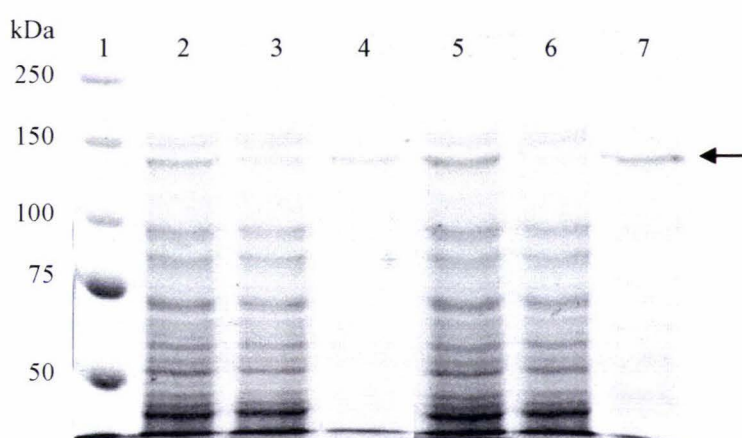
Figure 5.12 SDS PAGE of Protein from ER2566/pTYB4.Trx.MCM Grown at 12 °C for 22 hours With B_{12} in the Growth Media.

Lane 1: Molecular weight markers.

Lane 2: Total cell lysate.

Lane 3: Soluble fraction of cell lysate.

Lane 4: Insoluble fraction of the cell lysate,.

Lane 5: Total cell lysate, expressed with B_{12} in the growth media.

Lanes 6: Soluble fraction of the cell lysate, expressed with B_{12} in the growth media.

Lanes 7: Insoluble fraction of the cell lysate, expressed with B_{12} in the growth media.

There is no improvement in the solubility of the recombinant protein when B_{12} was provided in the growth media in lanes 6 and 7. Arrow indicates the size of the 148 kDa recombinant protein fusion construct.

When the soluble fractions from cells grown with (figure 5.12, lane 6) and without (figure 5.12, lane 3) the cofactor are compared   the presence of the cofactor decrease the soluble protein yield the distribution of the hMCM-intein-Trx fusion protein appears to change with more of the fusion protein present in the insoluble fraction. This

decrease in solubility is also accompanied by a slight increase in the total yield of protein when adenosylcobalamin is added. This increase can be explained by the incorporation of the recombinant protein into inclusion bodies, thus the reduction of protease degradation. In summary, addition of the cofactor to the *E. coli* growth media did not improve to the solubility of the fusion protein. This experiment, however, unexpectedly provided further support for the *in vivo* proteolyses of the recombinant protein fusion.

5.3   Expression from the Vector pTYB4.MCM

5.3.1   Introduction

From the plasmid pTYB4.MCM, the hMCM was expressed as a fusion construct with the CBD tag which can be removed by intein-mediated cleavage. The CBD is simply an affinity purification tag, and has not been shown affect the solubility of recombinant proteins. This vector can be purchased from NEB, and the hMCM was cloned into this vector as a negative control to determine the relative effectiveness of the novel vectors pTYB4.Trx and pTYB4.MBP at improving the solubility of hMCM. The ability of these two different solubility tags to improve the solubility of a target protein can be determined by comparison of the solubility of the fusion protein expressed from the parent vector, pTYB4.

5.3.2   Expression from pTYB4.MCM in *E. coli* ER2566

The pTYB4.MCM plasmid was transformed into *E. coli* ER2566, creating the expression strain ER2566/pTYB4.MCM. The 142 kDa hMCM fusion protein (hMCM-intien-CBD) expressed from the vector pTYB4.MCM is insoluble at 37 °C (figure 5.12). At 1 and 2 hours after induction with IPTG the hMCM fusion protein is found in the insoluble fraction (figure 5.13, lanes 5, 7 and 9).
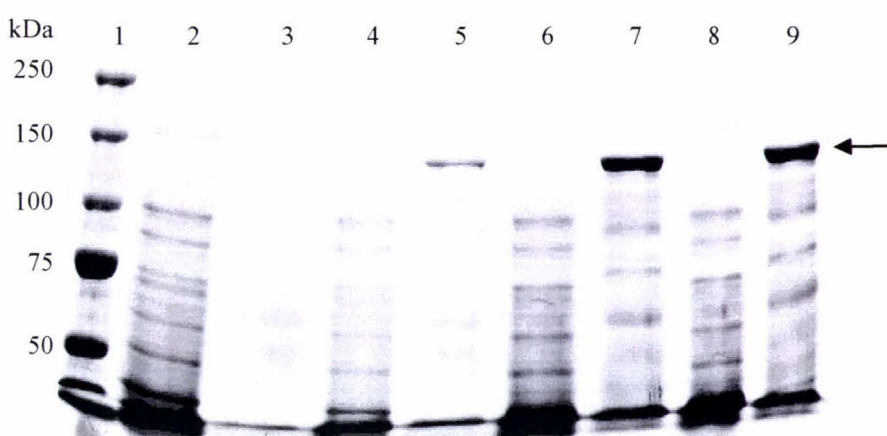
Figure 5.13 SDS PAGE Analysis of Protein Expressed from ER2566/pTYB4.MCM at 37 °C.

Lane 1: Molecular weight markers.

Lane 2: Cell lysate immediately before induction, soluble fraction.

Lane 3: Cell lysate immediately before induction, insoluble fraction.

Lanes 4: Cell lysate 1 hour after induction, soluble fraction.

Lane 5: Cell lysate 1 hour after induction, insoluble fraction

Lanes 6: Cell lysate 2 hours after induction, soluble fraction.

Lanes 7: Cell lysate 2 hours after induction, insoluble fraction.

Lanes 8: Cell lysate 3 hours after induction, soluble fraction.

Lanes 9: Cell lysate 3 hours after induction, insoluble fraction.

The fusion protein is present in the insoluble fraction at 1, 2 and 3 hours. The arrow indicates the position of the 142 kDa recombinant fusion protein.

After expression at 12 °C for 22 hours the 142 kDa hMCM fusion protein remains in the insoluble fraction (figure 5.14, lane 5), and can not be observed in the soluble fraction (figure 5.14, lane 4).

Figure 5.14 SDS PAGE Analysis of Protein Expressed from ER2566/pTYB4.   MCM at 12 °C for 22 hours.

Lane 1: Molecular weight markers.

Lane 2: Cell lysate immediately before induction, soluble protein fraction.

Lane 3: Cell lysate immediately before induction, insoluble protein fraction.

Lane 4: Cell lysate 22 hours after induction, soluble protein fraction.

Lane 5: Cell lysate 22 hours after induction, insoluble protein fraction.

Lane 4 shows none of the fusion protein can be observed as soluble full length protein. Arrows indicate the position of the 142 kDa recombinant fusion protein.

When designing these vectors it was not known whether placing the solubility tags, MBP and Trx, close to the intein could affect the stability and folding of the entire fusion protien. This vector, pTYB4.MCM, provided data on the hMCM baseline solubility at 37 °C and 12 °C. The complete insolubility of the fusion protein expressed from pTYB4.MCM at both expression temperatures, (figures 5.13 and 5.14), was used as a comparison for the newly constructed vectors, pTYB4.Trx and pTYB4.MBP. When compared to the vector pTYB4.Trx the solubility of the fusion protein increases (figure 5.15). The appearance of a small amount of hMCM-intein-Trx in the soluble fraction (figure 5.15, lane 4) can be attributed to the favorable influence of Trx on the folding of hMCM folding. This correctly folded hMCM-intein-Trx protein is both relatively protected from *E. coli* proteolysis and less prone to aggregation and the formation of inclusion bodies. It could then be concluded that the novel vector pTYB4.Trx provided an improved expression system, when compared to pTYB4. This novel vector provided

a Trx solubility tag which may improve the solubility of the target protein, and the Trx tag can be separated by intein mediated cleavage.
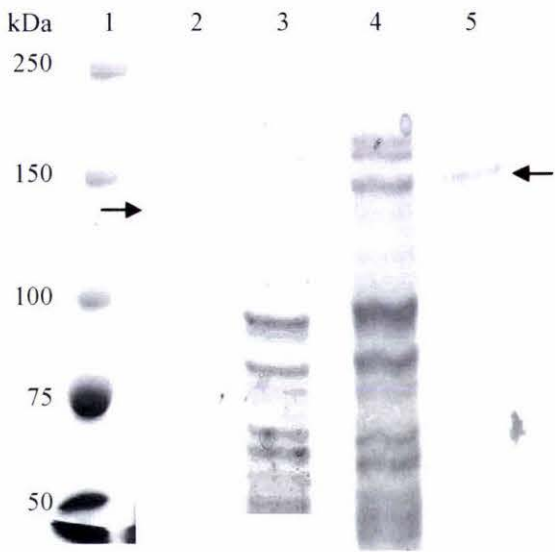
Figure 5.15 Lanes from Two SDS PAGE Gels Comparing Protein Expression from the Plasmids pTYB4.MCM and pTYB4.Trx.MCM at 12 °C for 22 hours.

Lane 1: Molecular weight markers.

Lane 2: Cell lysate from cells with pTYB4.MCM, insoluble protein fraction.

Lane 3: Cell lysate from cells with pTYB4.MCM, soluble protein fraction

Lane 4: Cell lysate from cells with pTYB4.Trx.MCM, soluble protein fraction

Lane 5: Cell lysate from cells with pTYB4.Trx.MCM, insoluble protein fraction

No soluble full length fusion protein is seen in the soluble fraction from pTYB4.MCM, compared to small amounts that are visable in the pTYB4.Trx.MCM soluble fraction. Arrows indicates the position of the 142 kDa recombinant protein fusion from pTYB4.MCM (lane 2), and the 148 kDa protein fusion from pTYB4.Trx.MCM (lanes 4 and 5).

5.4 Expression from the Plasmid pTYB4.MBP.MCM

5.4.1 Introduction

In 1997 Janata *et al.* reported that an N-terminal MBP solubility tag was an ineffective as an expression system for hMCM. The insolubility of the fusion protein, however, may have been due to the large N-terminal MBP disrupting the hMCM subunit interaction (Heaton *et al.*, 2001) (figure 5.16). Subunit dimerisation is necessary for hMCM activity and may also be critical for structural stability. A mutation (R93H) in the N-terminal subunit interaction domain causes loss of hMCM activity and consequently methylmalonic academia. The loss of hMCM activity caused by this mutation indicates the importance of hMCM dimer formation. The instability of hMCM as a momomer can be extended to the highly conserved *P. shermanii* α subunit, which is highly insoluble when expressed alone in *E. coli*, without its cognate β subunit (Mark Patchett, *pers. comm.*) (see section 3.2.1 for further details).
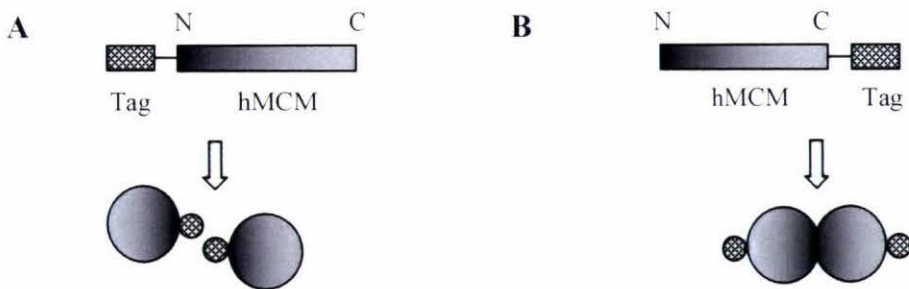
Figure 5.16 Possible Effects of a Tag on Subunit Dimerisation. The N-terminal region of hMCM is thought to have a role in subunit dimerisation.
A. The effect that an N-terminal tag could have on the interaction of the two subunits, and consequently on the stability of the hMCM.
B. The hMCM subunits are still able to form a dimer with a C-terminal tag. It is possible that this homodimer is the more stable structure for hMCM.

The hypothesis that the subunit interaction is disrupted by the N-terminal MBP tag, is inconsistent with the observation that the N-terminal Trx tag had little or no effect on the solubility or the dimerisation of the two subunits (Janata et al., 1997). This Trx-hMCM fusion construct was soluble and dimerisation was indicated by the detection of hMCM activity (Janata et al., 1997). If the MBP tag was preventing dimerisation, the Trx tag might also be expected to, even though the Trx protein is half the size of the MBP. If, however, the Trx tag is simply a more effective solubility tag this would be in contrast to the results of Kupust and Waugh (1999), who studied a diverse range of six insoluble target proteins. The MBP, in every case, was the most effective solubility tag when compared to the Trx and GST solubility tags. However successful solubility tag/target protein combinations are not possible to predict and generalisations can not be made. These uncertainties are largely due to the huge number of unknown parameters that may influence a folding protein. To test if the MBP is able to be an effective solubility tag at the N-terminus a longer linker region between the MBP and hMCM proteins than that used by Janata et al., (1997) may promote subunit interactions, however this possibility was not tested due to the cloning problems with the pTYB11.MBP vector. To determine if the MBP is able to be an effective solubility tag for hMCM at the C-terminal hMCM was expressed from pTYB4.MBP.MCM. This fusion protein is likely to circumvent any disruption of the dimerisation that an N-terminal MBP tag may cause (figure 5.16).

5.4.2   Expression from the Plasmid pTYB4.MBP.MCM in *E. coli* ER2566

The plasmid pTYB4.MBP.MCM was transformed into *E. coli* ER2566, giving the expression strain ER2566/pTYB4.MBP.MCM. After expression at 37 °C the 167 kDa fusion protein was present only in the insoluble fraction at 1, 2 and 3 hours post-induction (figure 5.17). In the cells that contained relatively large amounts of inclusion bodies, *E. coli* proteins were observed in the insoluble fraction (figure 5.17, lanes 5, 7 and 9). As the cellular concentration of aggregated insoluble protein increases, the nature of the cytoplasm changes and the native *E. coli* proteins can become insoluble and aggregate with the recombinant protein. This is a common feature of inclusion body formation, and often, after purification, a number of *E. coli* proteins are found to be over represented in the insoluble protein fraction (Hayes, 1998).
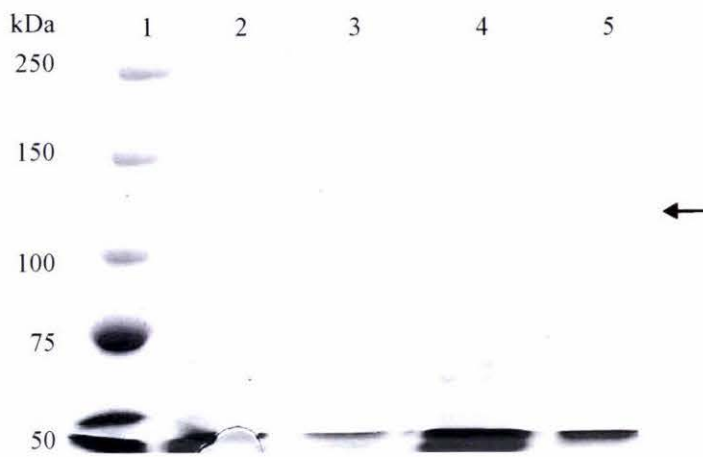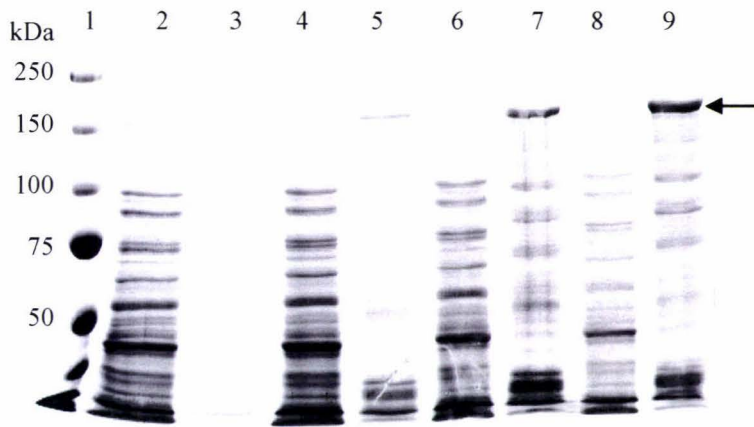
Figure 5.17 SDS PAGE Analysis of Protein Expressed from ER2566/pTYB4.MBP.MCM at 37 °C.

Lane 1: Molecular weight markers.

Lane 2: Cell lysate immediately before induction, soluble fraction.

Lane 3: Cell lysate immediately before induction, insoluble fraction.

Lanes 4: Cell lysate 1 hour after induction, soluble fraction.

Lane 5: Cell lysate 1 hour after induction, insoluble fraction

Lanes 6: Cell lysate 2 hours after induction, soluble fraction.

Lanes 7: Cell lysate 2 hours after induction, insoluble fraction.

Lanes 8: Cell lysate 3 hours after induction, soluble fraction.

Lanes 9: Cell lysate 3 hours after induction, insoluble fraction.

The fusion protein is expressed in the insoluble fraction, at 1, 2 and 3 hours. The arrow indicates the position of the insoluble 167 kDa recombinant fusion protein.

When the expression is induced at 12 °C the 167 kDa fusion construct remains insoluble (figure 5.18, lane 5). There is no change in the pattern of bands seen for the 167 kDa protein soluble fraction when compared to the uninduced cell lysate (lanes 4 and 2, respectively). Janata *et al.* (1997) have shown that hMCM, when expressed with an N-terminal MBP, is insoluble, and results presented here show that the MBP is also ineffective as a solubility tag on the C-terminus of hMCM. In this system the insolubility is unlikely to be caused by dimer disruption.
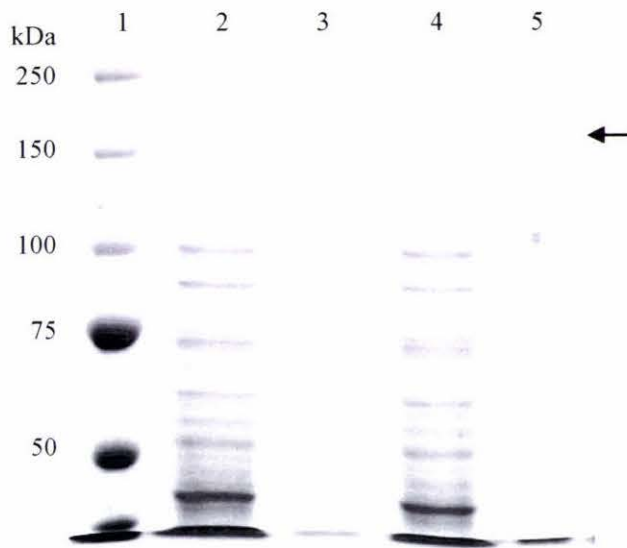
Figure 5.18 SDS PAGE Analysis of the Protein Expression from ER2566/pTYB4.MBP.MCM at 12 °C.

Lane 1: Molecular weight markers.

Lane 2: Cell lysate immediately before induction, soluble protein fraction.

Lane 3: Cell lysate immediately before induction, insoluble protein fraction.

Lane 4: Cell lysate 22 hours after induction, soluble protein fraction.

Lane 5: Cell lysate 22 hours after induction, insoluble protein fraction.

No full length recombinant fusion protein is seen in the soluble fraction. The arrow indicates the position of the 167 kDa recombinant fusion protein.

As the mechanism by which the MBP assists folding is not well understood, no firm conclusions can be made about why this was an ineffective solubility tag at the C-terminus. Folding and misfolding of proteins are complex processes that are difficult to study and consequently are not well understood. Proteins have been shown to begin folding before release from the ribosome (Gilbert *et al.* 2004). It has been suggested that that MBP must be fully folded, and the hydrophobic binding pocket on its exterior fully formed, to exert an effect (Sachdev and Chirgwin, 1998). If this suggested requirement is accurate, the MBP must come off the ribosome first, fold first, and begin acting on the partially folded intermediates of the target protein to prevent aggregation. The MBP contains a hydrophobic binding site for binding the substrate, maltose. The hydrophobic nature of this site is thought to play a role in binding partially folding proteins and preventing aggregation. If this hypothesis of the mechanism of MBP holds true, this C-terminal MBP would not be expected to improve the solubility of the hMCM. The

apparent complete insolubility of the hMCM-intein-MBP fusion construct, where MBP is C-terminal to the hMCM, supports the above hypothesis.

Together with the findings of Janata *et al.*, (1997), this suggests that the MBP is ineffective as a solubility tag for hMCM, regardless of its relative position to the hMCM. This ineffectiveness can be attributed to the mechanism by which the MBP prevents aggregation of proteins, which is not fully understood, and the unknown factors that cause hMCM to be insoluble. The mechanism with which MBP influences folding, however, differs from that of the Trx solubility tag; the Trx tag improves hMCM solubility when expressed at either termini of the target protein.

5.5   Assay of the Soluble hMCM

To determine the activity of hMCM, an assay developed by Zagalak *et al.* in 1974 was used. Succinyl CoA, the product of hMCM in humans, is the substrate in this *in vitro* assay, and the reverse reaction is catalysed by hMCM, generating methylmalonyl CoA. Though a series of coupled enzyme reactions, NADH is produced at a 1:1 molar ratio to the methylmalonyl CoA (that is generated by the hMCM dependent reaction) (figure 5.19). The rate of NADH reduction was followed by a CARY spectrophotometer, at 37 °C, by measuring the rate of reduction of absorbance at 340 nm (see section 2.2.15 for a more detailed description).
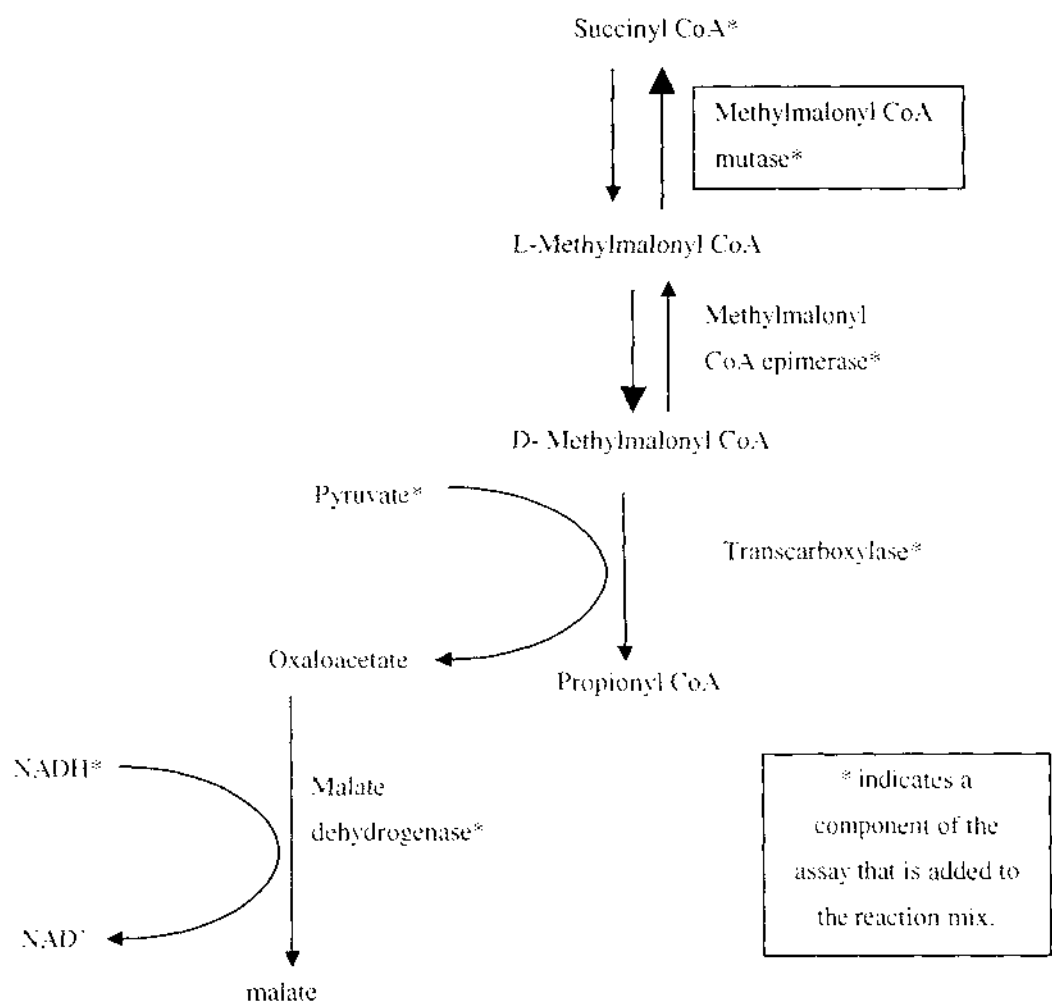
Figure 5.19 Assay of the hMCM. The assay shown here is based on the method of Zagalak *et al.* (1974). A spectrophotometer was used to continuously monitor the decrease in absorbance at 340 nm.

His-tag purified soluble protein, expressed at 12 °C from the pTYB4.Trx.MCM plasmid, was assayed for activity before and after DTT-induced cleavage. The activity of the cleaved hMCM was 0.45 μmol/min/mL protein added. The semi-quantitative specific activity of the sample was 0.9 U/mg. Given that the hMCM represents approximately 10% of the protein in this sample, an approximate activity of 9.0 U/mg of hMCM is comparable to the specific activity expected for MCM:

P. shermanii	14.4 U/mg	Kellermeyer *et al.*, 1964
Human placenta	1.33 U/mg	Kolhouse *et al.*, 1980
Human liver	14 U/mg	Fenton *et al.*, 1982
Recombinant human	3.1 U/mg	Janata *et al.*, 1997
Refolded recombinant human	3.11 U/mg	Hayes, 1998
This work	~9 U/mg	

Unexpectedly, the unprocessed 150 kDa protein complex had an activity of 0.44 μmol/min/mL protein (figure 5.20D). This 148 kDa fusion protein was not expected to have activity, so the zero hour activity was not determined. Together, these results indicate that the soluble hMCM is active and dimeric as the hMCM-intein-Trx fusion protein. It is also possible that hMCM is stable and soluble only as a dimer. The C-terminal intein-Trx tag is not disrupting the N-terminal subunit interaction of hMCM.
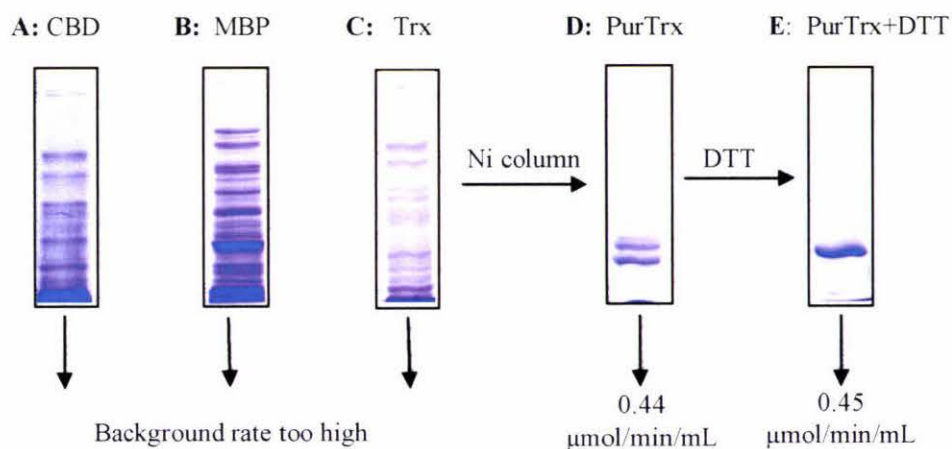
Figure 5.20 Lanes of SDS PAGE Gels, showing the Fractions Assayed for Activity. The background rate was too high for the three unpurified soluble fractions (**A-C**). In **D** the Ni-NTA purified fusion protein incubated overnight at 4 °C has an activity of 0.44 μmol/min/mL, and after incubation at 4 °C with DTT (**E**), the purified and intein-cleaved hMCM fragment has an activity of 0.45 μmol/min/mL.

The activity of the soluble fractions of cell lysate (before purification) from each plasmid was also assayed, (figure 5.19), before and after DTT incubation. Due to the nature of the assay (i.e. NADH linked) high background rate in these fractions, presumably due to other NADH utilizing reactions, (e.g. NADH oxidase) meant an accurate activity could not be determined. In addition to the SDS PAGE gel analysis of the different expression systems of these unpurified samples, activity measurements would have been a useful comparison. However the CBD and MBP must be soluble for to bind to their respective affinity resin, so purification for activity determination of these fusion proteins was not possible.

5.6   Summary

Earlier work of Janata *et al.*, (1997) showed that an N-terminal Trx-hMCM fusion expressed at 12 °C, resulted in high levels of a soluble active Trx-hMCM fusion protein, however the Trx tag was not able to be removed at the enterokinase site (Janata *et al.*, 1997). In this work, changing the Trx tag to the C-terminus of hMCM did not improve the solubility of the fusion protein. Pausing due to rare codons, mRNA instability and early translation termination were also ruled out as possible causes of the insolubility observed in the recombinant hMCM. The C-terminal hMCM-Trx fusion expressed at 12 °C resulted in half the fusion protein in the soluble fraction. Much of the recombinant hMCM-intein-Trx fusion protein was recovered from the soluble fraction as a partially degraded fragment. In a particularly abundant product a large region of the mutase domain from hMCM, was degraded, leaving the adenosylcobalamin (B_{12}) binding domain. However, providing adenosylcobalamin in the growth media did not improve the solubility of fusion protein. Activity assays showed that the soluble full length hMCM was also active, before and after the Trx solubility was removed. This showed that the hMCM expressed in this system is in a dimer, which is possibly the only stable and soluble form of the hMCM.

Together with the findings of Janata *et al.*, (1997), who showed that the MBP solubility tag is ineffective as a solubility tag regardless of the relative position to the hMCM. This could be due to the dimer disruption that MBP causes at the N-termini, or that the mechanism the MBP uses to solublise proteins is an ineffective solution for the type of insolubility that hMCM possesses.

The novel vector pTYB4.Trx, developed in this work, improved the solubility of the hMCM construct when compared to the commercially available vector pTYB4. Replacing the affinity tag, CBD, with a solubility tag improved the solubility of proteins expressed from this vector. This novel vector may be of use for expression of other insoluble proteins with difficulties in the protease access for the separation of the solubility tag.

6. Discussion and Future Directions

6.1 Discussion

6.1.1 Introduction

At the commencement of this study a native crystallographic dataset for human methylmalonyl CoA mutase (hMCM) was required. A structure calculated from this dataset would enable a better understanding of the unusual carbon skeletal rearrangement that this enzyme performs and the molecular bases of the disease methylmalonic academia that results from the inactive enzyme. Homology models based on the crystal structure of the α-subunit of the *P. shermanii* bacterial orthologue provided structural models of hMCM. However, these models were inaccurate in some regions and provided little information about the putative N-terminal subunit interaction domain, so further structural studies were required.

A system to express hMCM in *E. coli* as a soluble active fusion protein, able to be separated from any solubility tag was needed. In previous studies expression of hMCM in *E. coli*, has resulted in insoluble inclusion bodies and proteolytic degradation (Janata *et al.*, 1997). A Trx solubility tag on the N-terminus of hMCM had increased the yield of soluble protein. However, separation from the Trx tag at the protease site located between Trx and hMCM, was not achieved. In this work the hMCM gene was cloned and expressed in a number of different expression systems with solubility tags. Intein-mediated separation of the solubility tag circumvented the problems associated with conventional *in trans* protease cleavage. The cloning of the hMCM gene and the expression of fusion proteins expressed from each plasmid will be compared and discussed. The results of this work will also be put in context in relation to the past work in this area, to help gain a better understanding of the behaviour of hMCM when expressed in *E. coli*.

6.1.2 Design and Construction of the Novel Expression Vectors

The design of these novel expression systems was based mainly on the results of Janata *et al.*, (1997). The novel vectors must supply a solubility tag, and provide an alternative method for cleaving this tag from hMCM. Each system has variations within these two features such as having a long linker region between the tag and hMCM, or avoiding N-terminal tags altogether by using C-terminal tags. These observations were used to design four different vectors for the expression of hMCM in *E. coli*. The solubility tags Trx and MBP were used, and removed by intein-mediated cleavage. No commercial vectors provide this combination of features. Based on the aforementioned requirements, four novel vectors were constructed. Each of the four expression vectors differed in the position or nature of the solubility tag, but all utilised intein-mediated separation of the solubility tag.
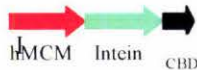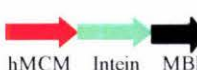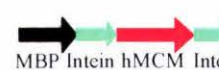
To construct these novel vectors the genes *malE* and *trx* from *E. coli* were cloned into either of the two intein-based vectors, pTYB4 or pTYB11, replacing the ORF coding for the existing CBD. In the fusion protein expressed from pTYB11 the CBD ORF is located in frame within the intein coding sequence and was replaced with an ORF encoding a larger protein, either Trx or MBP. In addition to size these proteins will differ in their surface properties. Although careful planning went into the design of these four vectors, it was largely unknown how these changes would affect the structure and function of the intein, or the stability of the entire fusion protein, of which the intein is a part. These four novel vectors were named pTYB4.Trx, pTYB4.MBP, pTYB11.Trx and pTYB11.MBP (appendix D).

6.1.3 Cloning hMCM into the Novel Expression Vectors

The hMCM was not cloned into the novel pTYB11-derived vectors due to problems associated with the cloning and ligation; i.e. the apparent loss of vector DNA resulting in undersized plasmids. These problems also occurred when the hMCM was cloned into the parent vector pTYB11. This showed that the problems during cloning are not due to the replacement of the CBD with Trx or MBP solubility tags. To determine the cause of these undersized plasmids, the plasmid products from ligation and transformation were analysed by restriction endonuclease digests and by sequencing. The plasmid products

from these ligations were highly reproducible. giving the same size fragments when digested with restriction endonucleases. Sequencing of the undersized plasmids revealed that. in at least one of the plasmids. a large region of the expected plasmid was missing. The two most plausible explanations for this undersized product were i) star activity of the restriction enzymes during cloning. and ii) DNA instability. Star activity. caused by a variety of reaction conditions. results in the restriction enzyme cutting the DNA at a sequence other than its recognition site. The small amount of star activity required would not necessarily be observable in SDS PAGE analysis of the linearised vectors. It was not possible to use different restriction enzymes other than *Sap*I for cloning hMCM into the vector pTYB11. An alternative explanation is that the observed products are a result of structural rearrangements within the plasmid sequence that are triggered by the sequence of the hMCM gene. As an alternative to cloning hMCM into the vector pTYB11. attempts were made to clone the hMCM into the vector pET32a.TEV. The vector pET32a.TEV is designed to expresses the target protein with an N-terminal Trx tag i.e. similar to the system described by Janata *et al.* (1997). Time limitations. however. prevented the completion of this cloning project.

In summary. the effectiveness of the two novel vectors pTYB11.Trx and pTYB11.MBP as expression vectors remains unknown. Difficulties with cloning and time constraints meant that expression of hMCM with an N-terminal MBP or Trx was not achieved. The hMCM. however. was cloned into the vectors pTYB4. pTYB4.Trx and pTYB4.MBP to enable the expression of three different fusion proteins with C-terminal tag. either the CBD. Trx or MBP. respectively (figure 6.1).

Vector	Position of tag	Constructed	Protein fusion	Soluble & active
pTYB4.MCM		yes	hMCM Intein CBD	not detected
pTYB4.Trx.MCM	C-terminal	yes	hMCM Intein Trx	~10%
pTYB4.MBP.MCM		yes	hMCM Intein MBP	not detected
pTYB11.MCM		no	CBD Intein hMCM Intein	-
pTYB11.Trx.MCM	N-terminal	no	Trx Intein hMCM Intein	-
pTYB11.MBP.MCM		no	MBP Intein hMCM Intein	-
pET32a.TEV.MCM	N-terminal	no	Trx hMCM	-

Table 6.1 The Expression Systems Attempted in this Study. Showing the vectors that hMCM was cloned into, the protein fusion expressed and the protein solubility.

6.1.4   Heterologous hMCM Expressed in *E. coli*

6.1.4.1 Solubility

The hMCM fusion proteins containing either the C-terminal MBP or CBD tags were expressed as insoluble inclusion bodies, under all conditions tested. The MBP solubility tag is ineffective as a solubility tag regardless of termini on the hMCM. This insolubility may be due to the MBP simply being an ineffective solubility tag for hMCM, because of the mechanism by which the MBP prevents aggregation of proteins (see section 6.1.3.2, for further discussion of the MBP as a solubility tag).

Expression from pTYB4.MCM as a fusion protein with the CBD, as a control, showed the improvement in solubility of hMCM that the novel vectors provided. Expression at 12 °C from the pTYB4.Trx.MCM vector, resulted in 50% of the full length fusion protein in the soluble protein fraction. This enabled the recovery of a very low yield of soluble active hMCM, which was able to be separated from the rest of the fusion protein by intein-induced cleavage. This improvement in solubility, when compared to the commercially available vector pTYB4, shows that the novel vector pTYB4.Trx provides a system that may improve the solubility of target proteins. In addition the vector provides intein-based cleavage that may circumvent problems associated with *in trans* cleavage, as shown in this hMCM system. The solubility of the Trx fusion protein was not able to be improved by, i) adding the hMCM cofactor to the *E. coli* growth media, ii) slowing the IPTG induction, or iii) by expression in an *E. coli* strain containing rare codons.

6.1.4.2 The Activity of the hMCM as a Homodimer

The hMCM is active only as an α_2 homodimer, mutations in the N-terminal subunit interaction region cause a loss in activity (Thoma and Leadley, 1996; Fuchshuber *et al.*, 2000). The importance of dimerisation was also implicated as a limiting factor in refolding studies, in the production of the active native structure (Hayes, 1992). The specific activity of hMCM, after refolding, gradually increased after storage at 4 °C. This was thought to be due to the folded subunits slowing associating to form the active native structure (Hayes, 1998). The results of Janata *et al.* (1997), also showed that

soluble hMCM was active, despite the presence N-terminal Trx tag. In this study, the C-terminal Trx fusion protein of this work was also found to have hMCM activity. This indicated the hMCM as a soluble fusion protein was a dimer, as hMCM is only active as a dimer. Therefore it is also possible that the dimerisation of the hMCM is also necessary for the stability of hMCM. If dimerisation is critical for stability, the location of the solubility tag may effect the dimerisation, and therefore the solubility of the hMCM fusion protein. When the point of attachment of the Trx tag was changed from the N to the C termini, the solubility was not improved. In addition to this, the C-terminal Trx tag, in this work, resulted in a less soluble fusion protein than the N-terminal Trx in the system of Janata *et al.* (1997).

In these systems the MBP solubility tag is ineffective as a solubility tag regardless of its position relative to the hMCM in the fusion protein. This may be due to the large solubility tag disrupting hMCM subunit dimerisation when present at the N-terminus of hMCM. If dimerisation is necessary for stability of hMCM, disruption of the subunit interaction would lead to insolubility. Assuming the proposed mechanism for action of MBP as a solubility tag is correct, the N-terminus would be the only effective position for MBP. This mechanism is consistent with the results reported here. Future work could use a longer linker region between the two proteins, hMCM and an N-terminal MBP. In results shown here this was not obtained due to apparent DNA instability. A longer linker may alleviate the dimer disruption of the hMCM and allow the MBP to be an effective solubility tag. In contrast to this hypothesis, if the 30 kDa MBP is ineffective due to dimer disruption, the 11 kDa Trx protein may also be expected to disrupt the dimer. The solubility and activity of the N-terminal Trx–hMCM construct, (Janata *et al.*, 1997), indicates that the Trx tag does not disrupt the dimer. This makes it less likely that the MBP is disrupting the subunit interaction and more likely that the MBP solubility tag is simply an ineffective as a solubility tag for hMCM.

6.1.4.3 Degradation of the Recombinant hMCM Fusion Proteins

When recombinant fusion protein is expressed at 12 °C from ER2566/pTYB4.Trx.MCM, degradation products are observed throughout protein purification and DTT cleavage. When expressed in ER2566/pTYB4.Trx.MCM at 37 °C more full length fusion protein is seen in SDS PAGE analysis. This may reflect a greater proportion of protein present in inclusion bodies which are protected from proteases (Cheng *et al.*, 1981). These observations suggest *in vivo* protease degradation. Although ˈprotease deficientˈ strains of *E. coli* were used for expression of hMCM, these *E. coli* retain several proteases for cell viability, such as HtrA and Clp. The biochemical and genetic evidence indicates that the proteases HtrA and Clp have a role in degradation of denatured/aberrant proteins and participate in the removal of the aggregated proteins from *E. coli* (Skorko-Glonek *et al.*, 1999). The hMCM region of the fusion protein may have been aberrantly folded making it a target for proteolytic degradation; if so, simply stopping the proteolysis would not solve the root of the problem, the instability of the hMCM. The events that result in misfolded hMCM are a matter of speculation. One possibility is that hMCM is present as partially folded monomers for some time in the *E. coli* cytosol, making the hMCM susceptible to proteolysis during folding (figure 6.1).
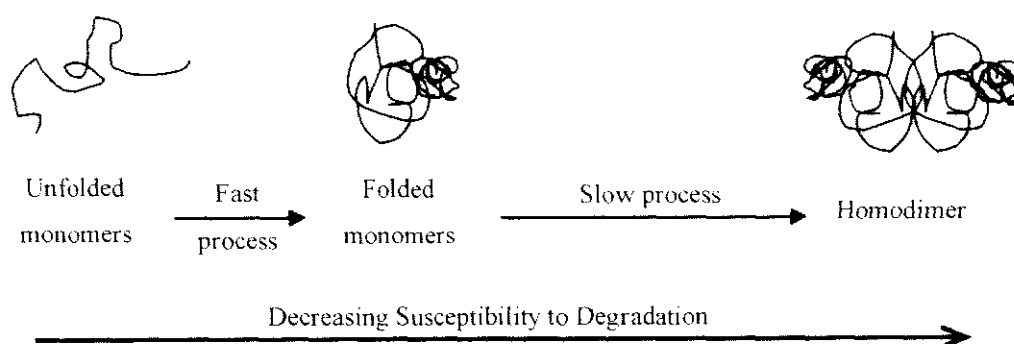
Figure 6.1 Model of hMCM Susceptibility to Proteolysis of as it Folds. As hMCM folds to its native state as a homodimer, it becomes less susceptible to degradation by *E. coli* proteases. The initial folding to form a monomer is relatively fast, the dimerisation of the monomers is a much slower process. These folding rates were shown by *in vitro* folding studies (Hayes, 1998).

In vivo degradation products of the hMCM-intein-Trx fusion protein showed that the mutase domain of hMCM was unstable and susceptible to proteolytic degradation. The B_{12} binding domain is a smaller and much more stable domain of the hMCM, and is able to fold independently of the rest of the hMCM. This is not surprising as this is a highly conserved domain of hMCM, and in the archaeal MCMs this domain is often expressed as a separate protein (figure 6.2).
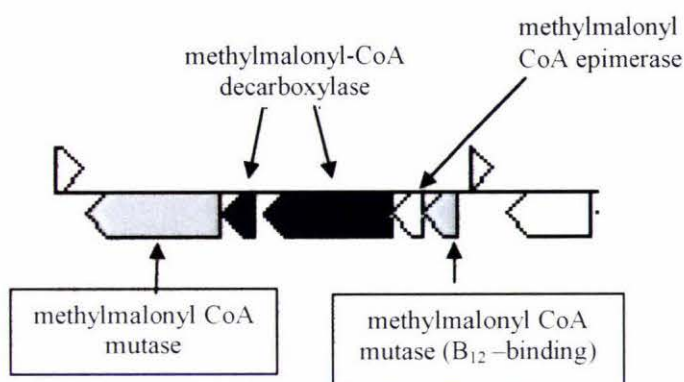
Figure 6.2 Archaeal MCM Genes. A KEGG genome map of the archaea *Archaeoglobus fulgidus*, showing the Mutase gene (insoluble in *E. coli*) and B_{12}–binding gene (soluble in *E. coli*). Also in this gene cluster are the methylmalonyl CoA epimerase and methylmalonyl-CoA decarboxylase.

http://www.genome.ad.jp/dbget-bin/get_pathway?org_name=afu&mapno=00280

When each of the archaeal MCM genes from *P. furiousis* is expressed separately in *E. coli*, the 'mutase like' protein is largely insoluble, while the B_{12}–binding protein is soluble (Mark Patchett, *pers. comm.*). This provides a comparison for the behaviour of the two hMCM domains, when the B_{12}–binding domain folds and the mutase domain is degraded (figure 6.3).
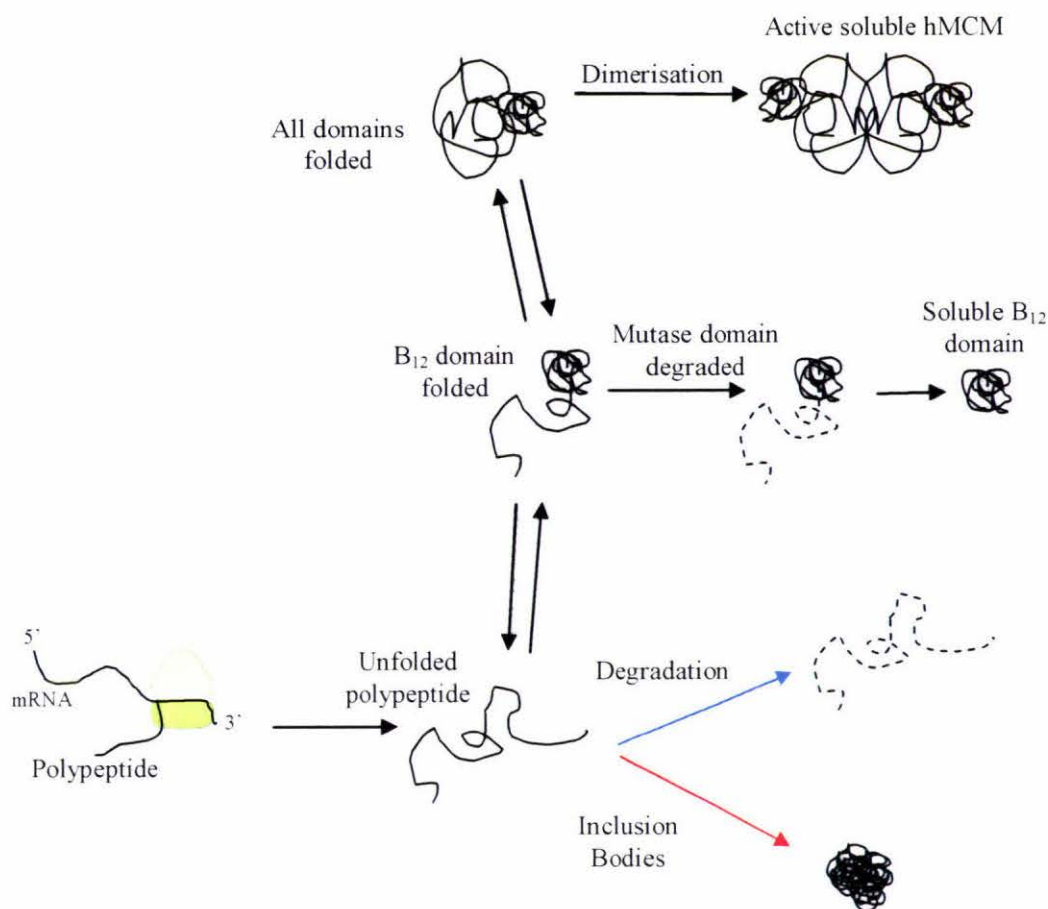
Figure 6.3 A Model for the Fates of hMCM that is Expressed in *E. coli.* This figure was based on the products recovered from the expression in *E. coli.*

Red arrow; is the more likely folding pathway at 37 °C, or in the presence of B_{12}.

Blue arrow; is the more likely folding pathways at 12 °C.

Expression from vector pTYB4.Trx.MCM yielded 20 µg of soluble full-length protein in 200 µL of Ni-NTA resin elution buffer, purified from 50 mL of cell culture. For crystallisation trials, at least 5 mg of protein is required, so 12 L of culture would be needed. In addition to this large volume, no more than 10 % of the proteins that bind to the column are full length fusion protein; the rest are partially degraded products. The degraded proteins decrease column binding capacity for the full length protein so that relatively little can be recovered off each column before the column is saturated. This makes Ni-NTA purification relativity, time consuming and requiring subsequent concentration as a source of protein for crystal studies. After DTT cleavage of the

hMCM-intein-Trx fusion protein the hMCM could be concentrated, then un though gel filtration to purify the 160 kDa homodimers of hMCM away from the rest of the 60 – 80 kDa proteins.

6.1.5   Using the Results of this Work to Re-evaluate the Literature

In previous work, Janata *et al.* (1997) were not able to explain the significant differences in the yields between the native hMCM and hMCM with an N-terminal Trx tag. In the expression from pTYB4.Trx.MCM partially degraded products were recovered, confirming the *in vivo* degradation of hMCM. This previously unidentified *in vivo* degradation of the hMCM was able to provide an explanation for the different yields shown in the previous expression systems of hMCM (Janata *et al.*, 1997).

In this work, *E. coli* host proteins were observed in hMCM inclusion bodies. When hMCM was expressed in *E. coli* from pMEXHCO an unknown protein was co-purified in the inclusion bodies (Hayes, 1998). Edman sequencing identified the three amino acids at the N-terminus as TKK. The protein had an observed mass of just under 29 kDa in SDS PAGE analysis. The hMCM does not contain the amino acid sequence TKK, so this protein was not a partially degraded product of hMCM. Reanalysis during this research revealed this protein is almost certainly the P_{lac} repressor protein, cI857. This protein is denatured after heat induction at 42 °C, allowing expression of the T7 RNA polymerase (Tabor and Richardson, 1995). The molecular mass of cI857 is just over 28 kDa and the proteins N-terminal sequence is MTKK. To the best of our knowledge this has not been previously reported.

6.1.6   Summary and Conclusions

The low yields of full length soluble recombinant protein expressed from pTYB4.Trx.MCM less than ideal source of hMCM for crystallisation studies. The intein-mediated approach, however, proved to be a successful method for the cleavage of the Trx tag, circumventing previous difficulties with *in trans* protease-mediated separation. The novel vector pTYB4.Trx, developed in this work, could also be used for the expression of other insoluble proteins that can not be separated from solubility tags by conventional methods.

Janata *et al.* (1997) showed that an N-terminal Trx tag was effective at improving the solubility of the recombinant hMCM in *E. coli*, but the Trx tag could not be efficiently separated using enterokinase protease. The success of the intein-mediated cleavage in protein expressed from pTYB11.Trx and the solubility of the N-terminal Trx fusion were not able to be determined due to difficulties with cloning the hMCM into the intein vector pTYB11.Trx, and time constraints with cloning into the second vector pET32a.TEV. Expression of the hMCM with a C-terminal CBD or MBP tags, from the plasmids pTYB4.MCM and pTYB4.MBP, resulted in insoluble fusion proteins.

In addition to the above observations, previously unreported *in vivo* proteolytic degradation of the hMCM was shown, consistent with instability of the hMCM monomer in *E. coli*. This proteolytic degradation targeted the mutase domain of hMCM, showing that this domain is a more unstable region of hMCM than the B_{12}-binding domain. This *in vivo* degradation explains the variation in yields for hMCM expression systems in the results of previously published work (Janata *et al.*, 1997). Altogether these results provide valuable information about *E. coli* expression systems for recombinant hMCM, and provide new starting points for the development of future systems.

6.2   Future Directions

6.2.1   Introduction

An ideal source of hMCM for structural studies, was not achieved in this work. There are a number of possible routes to the development of an improved source of hMCM for structural studies and some of these a described is the following sections. These options could include expression of hMCM with an N-terminal thioredoxin, ribosomal protein or NusA tag. Alternatively, the hMCM may be co-expressed with the *MMAA* protein or β the subunit of the bacterial MCM homologue which may display some chaperone like activity. Expression in a yeast, baculovirus, or cell-free based expression system could also be tested as possible sources of soluble active hMCM.

If a recombinant source of hMCM is not able to be developed, structural studies of a closely related MCM, (e.g. bovine or ovine), would allow more accurate modelling and a better understanding of the structure and function of hMCM. Completion of these structural studies may help in the development of a treatment, such as gene therapy, for the disorder methylmalonic acidemia.

6.2.2   Expression of hMCM with an N-terminal Trx Tag

The success of N-terminal Trx as a solubility tag (Janata *et al.*, 1997) makes this a logical expression system to trial. Two attempts at this expression construct were made in the work presented here. Firstly, in pTYB11.Trx in the intein system, apparent instability of ligation products prevented protein expression. The second attempt, cloning into a pET23a derived vector, with an N-terminal Trx tag and a TEV protease site, was halted due to time restraints. If cleavage of the Trx tag with TEV protease proved successful, this would make an ideal expression system as a source of protein for structural studies.

6.2.3 Yeast Recombinant Systems

Several systems for the heterologus expression of proteins in yeasts are commertially avalible e.g. *Pichia partorsis* (Invitrogene) or *Kluyveromyces lactis* (NEB). For the hMCM, the protein could be expressed with an N-terminal mitochondrial localisation peptide signal sequence. The protein would then be translated and targeted to the mitochondria. The signal sequence would be removed during translocation into the mitochondrial matrix. This system has been used to express proteins that are normally located in the mitochondrial matrix (Koshy *et al.*, 1992). There are conserved chaperones and other factors that facilitate folding in mitochondria. Mitochondrial matrix chaperones have roles in translocation, sorting, folding, and assembly of newly imported mitochondrial proteins (Neupert, 1997). The majority of mitochondrial proteins fold with the assistance of a chaperonin system that is composed of Hsp60 and Hsp10 (cpn10 and cpn60 in yeast). Conditions in the yeast mitochondrial matrix are obviously more similar to that of the human mitochondrial matrix than in the bacterial cytosol. As yeasts lack B_{12} –dependant enzymes, including MCM, this is an ideal system for the production of recombinant apo-hMCM, free of any host activity.

6.2.4 Baculovirus Mediated Insect Cell Expression

Like the yeast system, the bactulovirius also provides a eukaryotic cell background for expression. The ease of culture and high expression levels, up to 100 mg/L, make this an attractive system (Khoo *et al.*, 2005). Following replication and production in *E. coli* the recombinant viral DNA is transfected into the insect cells. The gene of interest is first cloned into the Bac-to-Bac® plasmid (Invitrogene). The gene of interest is now flanked by Tn7 sequences. This vector is then transformed into *E. coli* cells containing the destination bacmid, and a helper plasmid containing the gene encoding the transposase. The gene of interest is then transposed onto the destination bacmid. The recombinant bacmid is used to transfect Sf9 insect cells, and the infective virus is then harvested from the cells. When a high titre of virus is obtained, this can then be used to infect Sf9 cells for protein expression.

6.2.5   NusA Tag

NusA is a 55 kDa protein that has been used as a solubility tag for recombinant proteins expressed in *E. coli.* Fusion to the NusA solubility tag was shown to yield high levels of soluble protein, reducing both the aggregation and degradation of the target protein (De Marco *et al.,* 2004). A study comparing the NusA tag and the Trx tag, found that the NusA tag was a more effective solubility tag for the protein tested (Davis *et al.,* 1999). However it is not possible to predict the relative effectiveness for and particular protein. Vectors containing the NusA tag are commercially available (pET43.1) and can be purchased from Novagen.

6.2.6   Cell Free Systems

To overcome the problem of proteolytic degradation a cell free expression system can be utilised. If all the components of translation are provided, mRNA can be translated into protein *in vitro,* in a cell free system (Morita *et al.* 2003). Lysates derived from *E. coli* cells, rabbit reticulocytes and wheat germ are currently used. These cell free systems allow stricter control of the folding environment, as all the components are defined. In particular translation inhibitors such as thionin, RNA N-glycosidase, tritin, ribonucleases, deoxyribonucleases, and proteases are removed by extensive washing. Essential substrates (amino acids), energy source (NTP), and template mRNA are added to the reaction mixture. The wheat germ system can yield between 0.1 and 2.3 mg of protein over 2 days of expression, and can be used for up to two weeks. Many proteins synthesized using the "wheat germ cell-free protein synthesis system" have correct protein folding and activity (Morita *et al.* 2003). These cell free systems can be purchased (from Ambion or Roche) or can be made in the lab for considerably less cost.

6.2.7   Structural Studies of Native MCM Enzyme

The ethical issues associated with obtaining the large quantities of human liver make native hMCM purification problematic. Purification of the native mammalian MCM, other than human, would enable structural studies. The structure of one of the more closely related α_2 type enzymes would provide a more accurate model of the human enzyme than the current model based on the $\alpha\beta$ bacterial structure. Modeling of the N-terminal region would be greatly improved, as the α_2 type MCM enzymes have a more conserved sequence in the N-terminal region implicated in subunit dimerisation.

6.2.8   The *MMAA* gene

Mutations in the human *MMAA* gene were found to cause methylmalonic academia. The function of the *MMAA* gene product in humans is unknown. The predicted protein contains the ArgK sequence motif, including a Mg^{2+} binding aspartic acid residue and a GTP binding motif. There are a number of possible *in vivo* functions including an hMCM chaperone or an hMCM cofactor transporter. The predicted *MMAA* homologue in the bacterium *Methylbacterium extorquens* (*meaB*) is required for MCM activity *in vitro* (Korotkova and Lidstrom, 2004). This information, however, was published to late it to be of benefit to this research. Further work is required to determine the *in vivo* function of the human MMAA. Once a function for MMAA is established co-expression with hMCM may improve the solubility of activity of the hMCM.

6.2.9   Co-expression with the β Subunit of the Bacterial Homologue

The *P. shermanii* MCM is an $\alpha\beta$ heterodimer. The α subunit contains both the active site and B_{12}-binding site. The β subunit may play a role in the stability of the α subunit of the bacterial MCM. This proposed function is based on expression of the α subunit alone in *E. coli*. When expressed under comparable expression conditions, the α subunit was 20-50% soluble, compared to 90% soluble when the subunits were coexpressed (Mark Patchett, *pers. comm.*). The β subunit is translated first and may act as a chaperone aiding in the correct folding of the α subunit. Given the high (65%) identity between the α subunits of human and *P. shermanii* MCMs, the β subunit may perform a similar function with the human α subunit, assisting in correct folding and solubility

when co-expressed in *E. coli*. It is also possible that the α subunit is only stable as an α_2 dimer, in which case the β subunit would be of little assistance to the hMCM.

6.2.10 Expression in *E. coli* as a Fusion to a Ribosomal Protein

Expression of the target protein fused to a ribosomal protein has recently been shown to be an effective mechanism for preventing the formation inclusion bodies. Target proteins were fused to the C-terminus of the ribosomal protein rp23, and then were expressed in an rp23⁻ *E. coli* strain. The ribosomes containing the recombinant protein were then isolated, and the recombinant target protein released by proteolysis. With expression of the hMCM this would allow *in vitro* dimerisation, after the proteolysis. The ribosomes can then be removed by centrifugation. The prevention of inclusion body formation is thought to be due to the physical separation of the recombinant protein during and after folding, preventing self association (Sorensen *et al.*, 2004)

6.2.11 Treatment of Methylmalonic Acidemia

A better understanding of the structure and function of hMCM would improve our overall understanding of the enzyme, and may aid the development of a gene therapy technique for patients with MMA. As early as 1992 the hMCM gene was trialled in gene therapy for patients with MMA, making it one of the first enzymes to be investigated for potential gene therapy. Hepatocytes (liver cells) deficient in hMCM could be rescued using retroviral mediated gene therapy (Sawada and Ledley, 1992). The gene for hMCM was also over-expressed in mouse hepatocytes. Mice were injected with a plasmid containing the hMCM gene and an *in vivo* gene delivery system was used to target the delivery of the plasmid to the liver. The treatment was largely successful with little to no side effects reported, and a significant increase (30-40%) in hMCM activity was measured in liver tissue (Stankovic *et al.*, 1994).

Appendix

A.   *E. coli* **Strain Genotypes**

BL21(DE3)

F– *ompT gal* [*dcm*] [*lon*] *hsdSB* (rB– mB–; an *E. coli* B strain) with DE3, a λ prophage carrying the T7 RNA polymerase gene

BL21(DE3)OrigamiB

Delta *Ara-leu7697* Delta *lacZx74* Delta *PhoAPvuII phoR araD139 ahpC galE galK rspL* F'[*lac* (*lacZi* ^{Q} *pro*] *gor522* Tn *10* (Tc ^{R} *trxB* Kan (DE3)

ER2566

F– l– *fhuA2* [*lon*] *ompT lacZ::T7 gene1 gal sulA11 D(mcrC-mrr)*114::IS10 R(*mcr-73::mini*Tn 10)2 R(*zgb-210::*Tn 10)1 (Tets) *endA1* [*dcm*]

TG1 *recO1504::*Tn5 (Kolodner *et al.* 1985)
F' *traD36 lacI*^{q}Δ(*lacZ*) M15 *proA*+B' /*supE* Δ(*hsdM-mcrB*)5 (r$_k$ - m$_k$ - McrB') *thi* Δ(*lac-proAB*) *recO1504::*Tn5

XL-1 Blue,
F'::Tn10*proA*+B+*lacI*^{q}Δ(*lacZ*)M15 *recA1 endA1 gyrA96* (Nal^r) *thi hsdR17* (r$_K$˗m$_K$˗) *glnV44 relA1 lac*

B.   Rare Codons

Codon usage in *Escherichia coli K12* [gbbct]: 5048 CDS's (1601642 codons)

fields: [triplet] [frequency: **per thousand**] ([number])

```
UUU 22.4( 35892)   UCU  8.5( 13624)   UAU 16.3( 26145)   UGU  5.2(  8296)
UUC 16.6( 26571)   UCC  8.6( 13771)   UAC 12.3( 19640)   UGC  6.4( 10311)
UUA 13.9( 22256)   UCA  7.1( 11421)   UAA  2.0(  3240)   UGA  0.9(  1446)
UUG 13.7( 21974)   UCG  8.9( 14282)   UAG  0.2(   365)   UGG 15.3( 24536)

CUU 11.0( 17680)   CCU  7.0( 11275)   CAU 12.9( 20648)   CGU 21.0( 33673)
CUC 11.0( 17682)   CCC  5.5(  8848)   CAC  9.7( 15569)   CGC 22.0( 35242)
CUA  3.9(  6174)   CCA  8.5( 13654)   CAA 15.5( 24764)   CGA  3.5(  5664)
CUG 52.8( 84595)   CCG 23.3( 37243)   CAG 28.8( 46180)   CGG  5.4(  8624)

AUU 30.4( 48706)   ACU  8.9( 14289)   AAU 17.6( 28219)   AGU  8.7( 13958)
AUC 25.0( 40044)   ACC 23.4( 37438)   AAC 21.7( 34689)   AGC 16.0( 25671)
AUA  4.3(  6860)   ACA  7.0( 11260)   AAA 33.6( 53862)   AGA  2.1(  3290)
AUG 27.8( 44484)   ACG 14.4( 23029)   AAG 10.2( 16347)   AGG  1.2(  1945)

GUU 18.4( 29453)   GCU 15.4( 24591)   GAU 32.2( 51598)   GGU 24.9( 39815)
GUC 15.2( 24380)   GCC 25.5( 40827)   GAC 19.1( 30514)   GGC 29.4( 47113)
GUA 10.9( 17424)   GCA 20.3( 32501)   GAA 39.6( 63375)   GGA  7.9( 12680)
GUG 26.2( 42031)   GCG 33.6( 53883)   GAG 17.8( 28481)   GGG 11.0( 17600)
```

Red indicates less than 1% codon frequency.

Figure generated from http://www.kazusa.or.jp/codon/

The Rare codons for *E. coli* that are found in the hMCM gene:

countcodon output

(782 codons)

fields: [triplet] [frequency: **per thousand**] ([number])

```
UUU 30.7(24)   UCU 19.2(15)   UAU 17.9(14)   UGU 12.8(10)
UUC 11.5(9)    UCC  7.7(6)    UAC 11.5(9)    UGC  0.0(0)
UUA 11.5(9)    UCA 12.8(10)   UAA  5.1(4)    UGA  0.0(0)
UUG 11.5(9)    UCG  0.0(0)    UAG  0.0(0)    UGG  9.0(7)

CUU 30.7(24)   CCU 21.7(17)   CAU  9.0(7)    CGU 10.2(8)
CUC  7.7(6)    CCC  6.4(5)    CAC  9.0(7)    CGC  3.8(3)
CUA  9.0(7)    CCA 17.9(14)   CAA 12.8(10)   CGA 10.2(8)
CUG 14.1(11)   CCG  1.3(1)    CAG 26.9(21)   CGG  2.6(2)

AUU 39.6(31)   ACU 25.6(20)   AAU 23.0(18)   AGU  9.0(7)
AUC 14.1(11)   ACC 11.5(9)    AAC  9.0(7)    AGC  6.4(5)
AUA 16.6(13)   ACA 17.9(14)   AAA 39.6(31)   AGA 14.1(11)
AUG 33.2(26)   ACG  0.0(0)    AAG 24.3(19)   AGG  9.0(7)

GUU 24.3(19)   GCU 46.0(36)   GAU 39.6(31)   GGU 21.7(17)
GUC  5.1(4)    GCC 12.8(10)   GAC 11.5(9)    GGC  7.7(6)
GUA 16.6(13)   GCA 25.6(20)   GAA 61.4(48)   GGA 37.1(29)
GUG 20.5(16)   GCG  3.8(3)    GAG 11.5(9)    GGG  7.7(6)
```

Red indicates less than 1% codon frequency.

Figure generated from http://www.kazusa.or.jp/codon/

C.   Primers

Primers are shown 5' to 3', with any incorporated restriction sites underlined.

PCR of the *trx* gene for cloning into pTYB4:

pET32aTrxHF   GAT<u>ACCGGT</u>ATGAGCGATAAAATTATTCA
pET32aTrxHR   CCA<u>CTGCAG</u>ATTAATGATGATGATGATGGT

PCR of the *trx* gene for cloning into pTYB11:

pTYB11.TrxF   GGAGAT<u>GCTAGC</u>ATGAGCGATAAAATTATTC
pTYB11.TrxR   CACC<u>CCATGGG</u>AAGAATGATGAT

PCR of the *malE* gene for cloning into pTYB4:

pTYB4.MalF   CAAGGA<u>ACCGGT</u>CATATGAAAACTGAAGAA
pTYB4.MalR   TTGTTGTTGTTCGC<u>TGCAG</u>ATTAAGTCT

PCR of the *malE* gene for cloning into pTYB11:

pTYB11.MalF   CAAGGA<u>GCTAGC</u>CATATGAAAACTGAAGAA
pTYB11.MalR   GAGCTCCCATGGGTCTGCG

PCR of the hMCM gene for cloning into pTYB11 and pTYB11 derived vectors:

ACFMCM11   GGTGGTTGCTCTTCCAACCTTCATCA
              GCAACAGCCGCTT
ACRMCM11   GGTGGT<u>CCCGGG</u>ATTATACAGATTGCT GCTTCTT

PCR of the hMCM gene for cloning into pTYB4 and pTYB4 derived vectors:

ACFMCM4   GGTGGT<u>CCATGG</u>TTCATCAGCAACAGCCGCTT
ACRMCM4   TACAGATTGCTGCTTCTTTTCCAAA

Sequencing Primers:

ACTYB1-4CBD   GGGATTACTTTATCTGATGATTCTG
ACTYB1-4T7   CGAAATTAATACGACTCACTATAGGG
ACTYB1-4IntR   ACCCATGACCTTATTACCAACCTC
pTYB11-CBD   AATTTGTGCGCCGAGTATAAGGAC
TYB11.MCSF   TGAAGGAAGACGATTATTATGGGATTAC
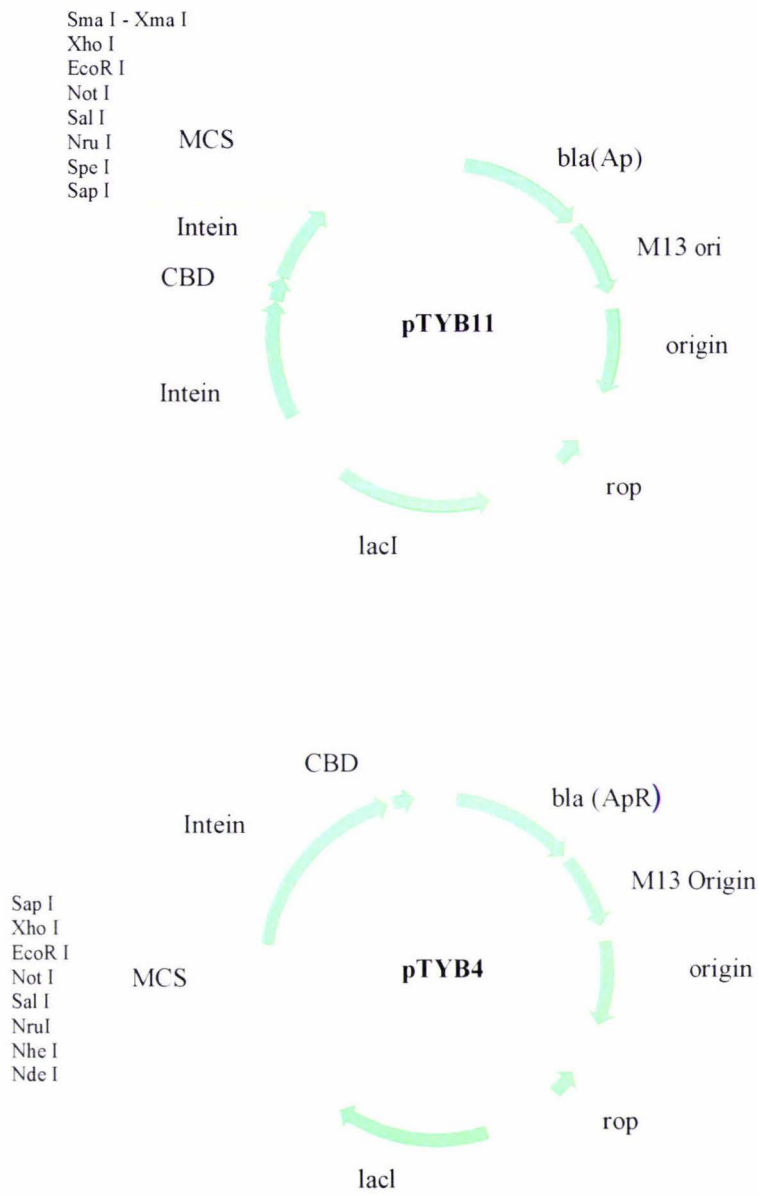
Figure ii. Template Vectors. These vectors were used as templates for PCR of the genes for thioredoxin (pET32(a)), hMCM (pMEXHCO) and MBP (pMALc2g).
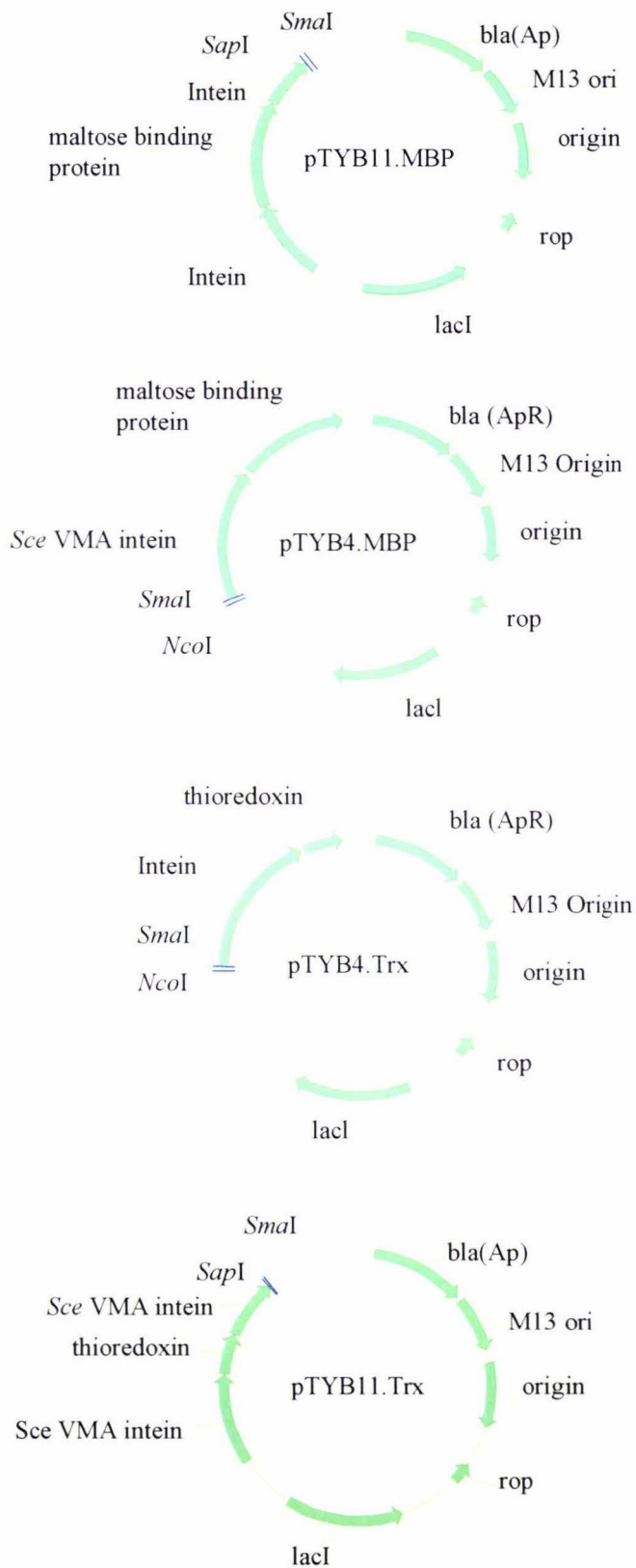
Figure iii. Showing the Expression Vectors Constructed. To make these vectors the CBD of pTYB11 and pTYB4 was removed and replaced with either thioredoxin or Maltose binding protein. Restriction sites at the multiple cloning site used to clone hMCM into the vector are shown.
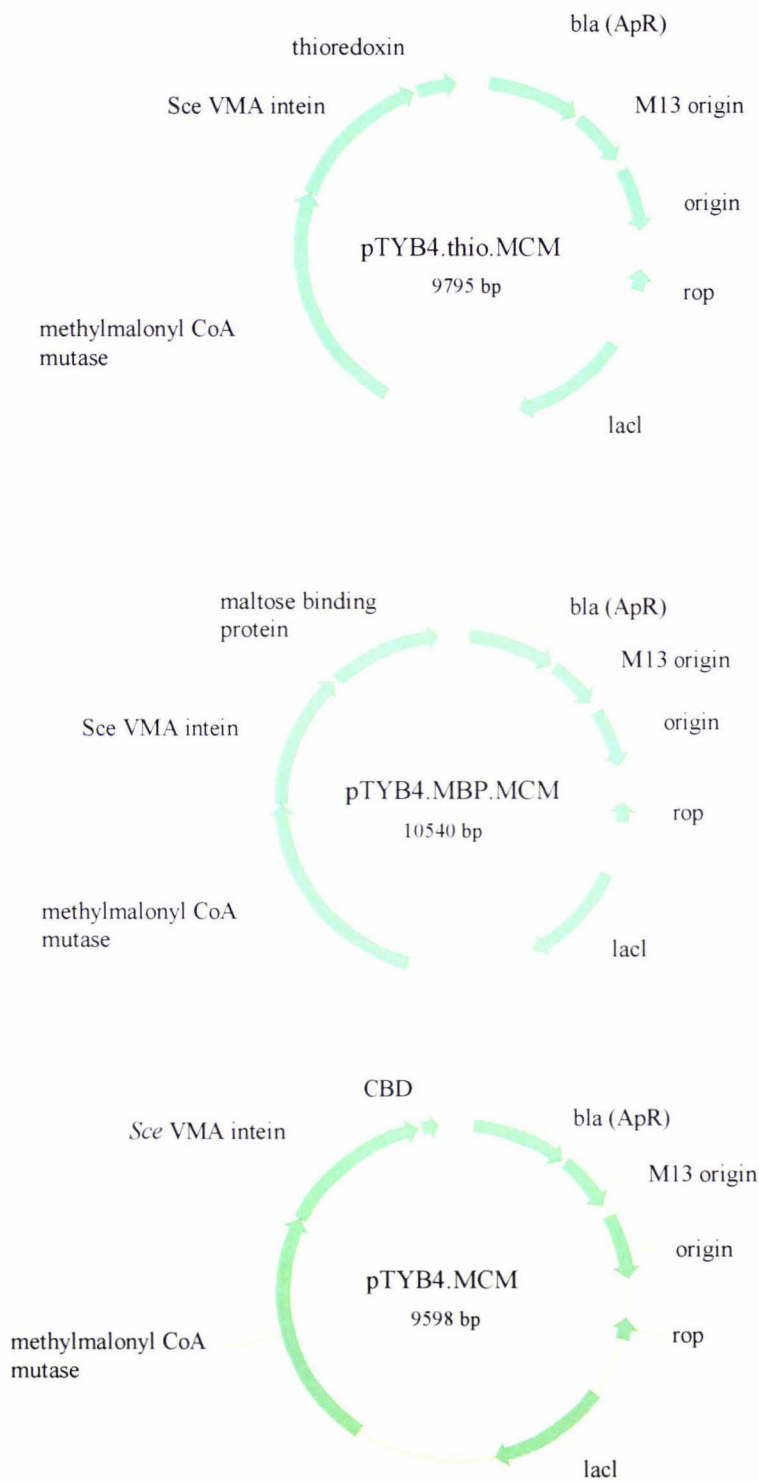
Figure iv. Showing the hMCM cloned into the newly constructed Plasmids. These were constructed and the hMCM was expressed from these novel plasmids.
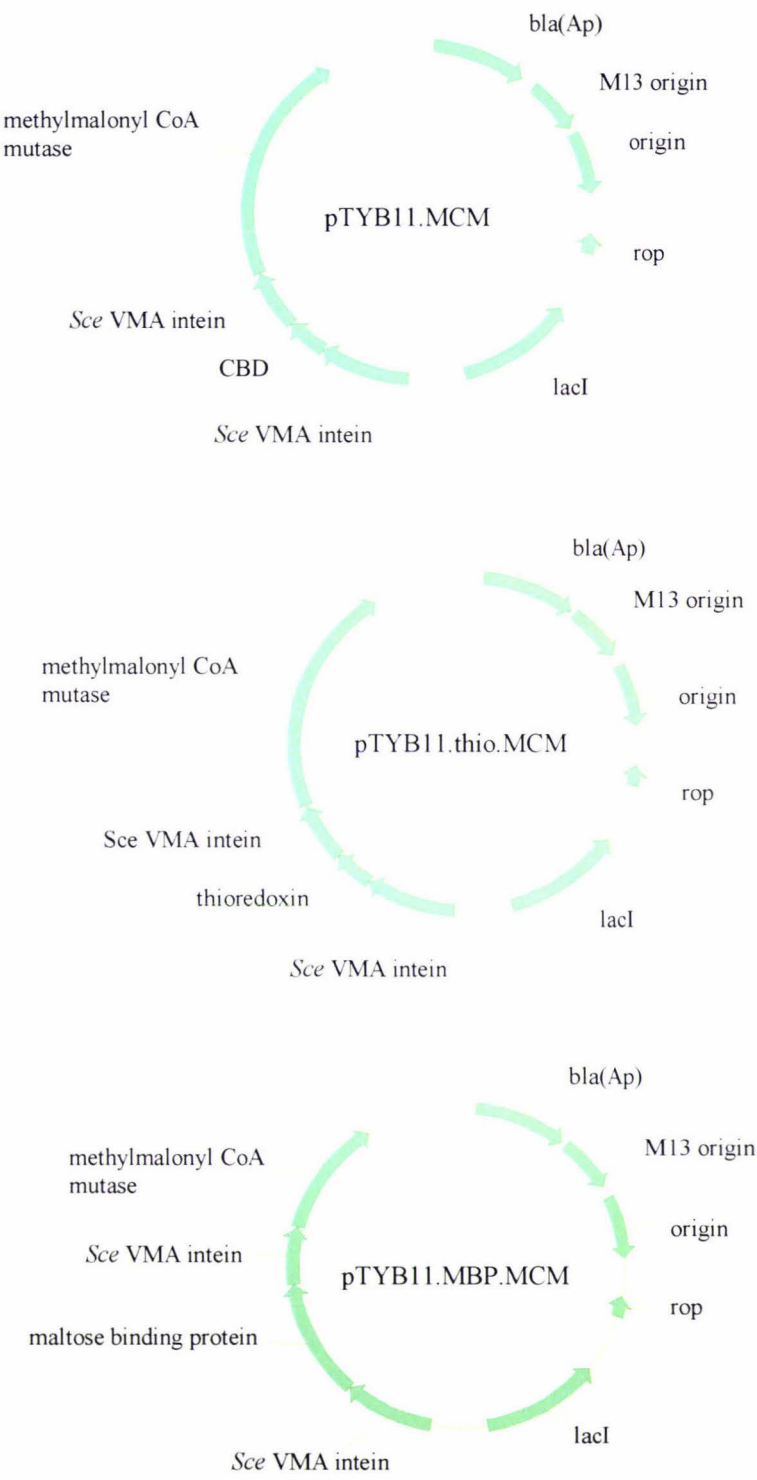
Figure v. The Plasmids for N-terminal Tag Fusion. These plasmids were defined but not obtained or tested.

E.   The Affect of the Adjacent Amino Acid on the pTYB11 Intein Activity

C-terminal Residue of the Target Protein	*In vivo* Cleavage	Cleavage with DTT (40 mM)	
		4 °C	16 °C
Gly	-	+++	+++
Ala	-	+++	+++
Ile	-	+	+
Leu	-	+	+++
Met	-	+++	+++
Phe	-	+++	+++
Val	-	+	++
Gln	-	+++	+++
Ser	-	++	+++
Trp	-	+++	+++
Tyr	-	+++	+++
Lys	-	+++	+++
Thr	25%	++	+++
Glu	50%	++	+++
His	50%	++	++
Arg	75%	Not determined	Not determined
Asp	100%		

Table i. Effect of the C-terminal residue of a target protein on DTT-induced cleavage of the intein tag when pTYB4 is used as the cloning vector.

Data is based on analysis of cleavage reactions using MBP as the target protein (MYB) with amino acid substitutions at the position (−1) immediately upstream of the cleavage site in the sequence (L-3E-2X-1/C1). (−) = Less than 10% cleavage; (+) = 30%–49% cleavage; (++) = 50%–74% cleavage; (+++) = 75%–100% cleavage.

Figure adapted from the NEB Impact™ CN Instruction Manual

	% CLEAVAGE AFTER 16 HOURS*			% CLEAVAGE AFTER 40 HOURS*		
N-Terminal Residue of the Target Protein	4 °C	16 °C	23 °C	4 °C	16 °C	23 °C
Met Ala Gln	40-60	>80	>95	60-90	>90	>95
Gly Leu Asn Trp Phe Tyr	10-40	50-80	75-95	40-60	> 90	>90
Val Ile Asp Glu Lys Arg His	< 10	30-50	50-80	10-20	70-90	70-95

Table ii. Effect of the N-terminal Amino Acid of the Target Protein on the Cleavage Efficiency of the *Sce* VMA1 intein in the Vector pTYB11. The amino acid adjacent to the cleavage site can have an effect on the cleavage of the intein. The amino acid must be carefully selected to optimise the activity of the intein.

Figure adapted from the NEB Impact™ CN Instruction Manual

F.   BLAST Alignment Showing the Sequence of the Internal Deletion in pTYB11

Sequence 1 lcl|pTYB11.thio vector **Length** 7412 (1 .. 7412)

Sequence 2 lcl|Sequencing   results **Length** 1165 (1 .. 1165)

NOTE: The statistics (bitscore and expect value) is calculated based on the size of nr database

NOTE: If protein translation is reversed, please repeat the search with reverse strand of the query sequence

```
Score = 1669 bits (868), Expect = 0.0
Identities = 955/985 (96%), Gaps = 14/985 (1%)
 Strand = Plus / Minus
```



```
Query: 2143 ggagcttccaggggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctg 2202
            |||||||||||||| |||||||||||||| || |||||||||||||||    | ||| ||
Sbjct: 1028 ggagcttccaggg--aaacgcctggtatytt-atagtcctgtcgggttcs--cmcctytg 974


Query: 2203 acttgagcgtcgattttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccag 2262
            ||| |||||||||||||||||||||||||||||||||| |||| | ||||||||||||||||
Sbjct: 973  act-gagcgtcgattttttgtgatgctcgtcaggggg-cggasc-tatggaaaaacgccag 917


Query: 2263 caacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcc 2322
            |||||||| | |||||||||| ||||||||| |||||||||| ||||||||||  || ||
Sbjct: 916  caacgcgcc--ttttacggt-cctggcctttt-gctggcctttt-gctcacatgtyytt-cc 863


Query: 2323 tgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgc 2382
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 862  tgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgc 803


Query: 2383 tcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagctatggtgcactc 2442
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 802  tcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagctatggtgcactc 743


Query: 2443 tcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacg 2502
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 742  tcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacg 683


Query: 2503 tgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggc 2562
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 682  tgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggc 623


Query: 2563 ttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtg 2622
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 622  ttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtg 563
```

Query: 2623 tcagaggttttcacggtcatcacgaaacgcgcgaggcagctgggtaaagctcatcagc 2682
             ||||||||| || |||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 562   tcagaggttttTaccgtcatcacgaaacgcgcgaggcagctgcggtaaagctcatcagc 503

Query: 2683 gtggtcgtgcagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagttt 2742
             || || || ||||||||||||||||||||||||||||||||||||||||||||||| |||
Sbjct: 502   gtggtcgtgcagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagttt 443

Query: 2743 ctccagaagcgttaatgtctggcttctgataaagcgggcatgttaagggcggtttttc 2802
             ||||||||||||||||||| ||||||||||||| |||||||||||||||||||||||||
Sbjct: 442   ctccagaagcgttaatgtctggcttctgataaacgggcatgttaagggcggtttttc 383

Query: 2803 ttgtttggtcactgatgcctccgtgtaaggggattttctgttcatgggggtaatgataccc 2861
             ||||||||| |||| ||||||||||||||||||||||| | |||||| ||||||||| ||
Sbjct: 382   ctgtttggtcactgatgcctccgtgtaaggggatttttgttcatgggggtaatgataccc 323

Query: 2862 gatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcctggttacc 2922
             |||||||||||||||| ||| | |||||| ||||||||||||||| || ||||| |||
Sbjct: 322   gatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcctggttacc 263

Query: 2923 gaaacgttgtgagggtaaacaactggcgatgcctcacgcaaacagagaaaaatcac 2983
             ||||| || || ||| ||| |||| || | | || || | |||||||||||||
Sbjct: 262   gaaacgttgtgaggtaaacaactggcgatgcctcacgcaaacagagaaaaattac 213

Query: 2984 tcaagttcatgtacagcgaaatgctagtagatgtaattactgtatcgatctgccatggc 3042
             ||| | | | | | | ||||| | | |||| ||| ||| | ||| |||| ||
Sbjct: 212   ttaagttaatgctagctaagtcagccagaatgtagctacgtatcgatctgccatggc 143

Query: 3043 aatgataatgatcctgttttatcggaaagattttgtggcatgatacagtaataagtttg 3102
             ||||| |||||| || | | || |||||| || || |||| || ||||||||||||
Sbjct: 142   aatgataatgatcctgttttatcggaaagatttgtggcatgatacagtaataagtttg 84

Query: 3103 aacaagggctgaaagattcgaat 3127
             | | | |||| | | ||| ||
Sbjct: 83   agtgaaggcgctataataawnagaat 59

Score = 43.0 bits (667), Expect = 2e-18
Identities = 45/48 (98%)
Strand = Plus / Minus

Query: 6454 ggatgactgcaggaagcggatctggctgttaacaaagtrcgaaaga 6499
             || ||||| ||| ||||| |||||||||||||||||| |||||||
Sbjct: 57   gcgtgactgcaggaagcggatctggctgttaacaaagttcgaaaga 12

References

Acquaviva, C., Benoist, J., Callebaut, I., Guffon, N., Ogier de Baulny, H., Touati, G., Aydin, A., Porquet, D., Elion, J. (2001). N219Y, a New Frequent Mutation Among *Mut^0* Forms of Methylmalonic Acidemia in Caucasian Patients. *European Journal of Human Genetics.* **9**, 577-582.

Acquaviva, C., Benoist, J., Pereira, S., Callebaut, I., Koskas, T., Porquet, D., Elion, J. (2005). Molecular Basis of Methylmalonyl-CoA Mutase Apoenzyme Defect in 40 European Patients Affected by *mut^l* and *mut$^-$* Forms of Methylmalonic Acidemia: Identification of 29 Novel Mutations in the *MUT* Gene. Human Mutation. 25, 167-176.

Adjalla, C.E., Hosack, A.R., Gilfix, B.M., Lamothe, E., Sun, S., Chan, A., Evans, S., Matiaszuk, N.V., Rosenblatt. (1998). Seven Novel Mutations in *mut* Methylmalonic Aciduria. *Human Mutation.* **11**, 270 – 274.

Andrews, A. T. (1986). Electrophoresis. Theory, Techniques, and Biochemical and Clinical Applications. 2nd edition, pp. 117-147. Oxford University Press, New York, USA.

Andrews, E., Jansen, R., Crane, A., Cholin, S., McDonnell, D., Ledley, F.D. (1993). Expression of Recombinant Human Methylmalonyl-CoA Mutase: In Primary *mut* Fibroblasts and *Saccharomyces cerevisiae. Biochemical Medicine and Metabolic Biology.* **50**, 135-144.

Arnold, D. A., Handa, N., Kobayashi, I., Kowalczykowski, S. C. (2000) A Novel, 11 Nucleotide Variant of chi, chi*: One of a Class of Sequences Defining the *Escherichia coli* Recombination Hotspot chi. *Journal of Molecular Biology.* **300**, 469-479.

Banerjee, R., Chowdury, S. (1999). Chemistry and Biochemistry of B12. John Wiley & Sons, Inc. New York. pp 707-729.

Banerjee R, Vlasie M (2002). Controlling the reactivity of radical intermediates by coenzyme B-12-dependent methylmalonyl-CoA mutase. *Biochemical Society Transactions.* **30**, 621-624.

Baneyx, F., Georgiou, G. (1990). In vivo Degridation of Secreated Fusion Proteins by the *Escherichia coli* Outer-Membrane Protease *ompT. Journal of Bacterology.* **172**, 491-494.

Baneyx, F., Mujacic, M. (2004). Recombinant Protein Folding and Misfolding in *Escherichia coli. Nature Biotechnology.* **22**, 1399-1407.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N.T., Burland, V., Riley, M., ColladoVides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science.* **277**, 1453-1469.

Bondos, S. E., Bicknell, A. (2003). Detection and Prevention of Protein Aggregation Before, Buring, and After Purification. *Analytical Biochemistry.* **316**, 223-231.

Carrington, J., Dougherty, W. (1988). A Viral Cleavage Site Cassette - Identification of Amino-Acid Sequences Required for Tobbacco Etch Virus Polyprotein Processing. *Proceedings of the National Academy of Sciences of the United States of America.* **85**, 3391-3395.

Cassland. P., Larsson. S., Nilvebrant, N., Jönsson L. J. (2004). Heterologous expression of barley and wheat oxalate oxidase in an *E. coli trxB gor* double mutant. *Journal of Biotechnology.* **109**, 53-62.

Cash, P. (1998). Characterisation of bacterial proteomes by two-dimensional Electrophoresis. *Analytica Chimica Acta.* **372**, 121-145.

Cellmer, T., Bratko, D., Prausnitz, J., Blanch, H. (2005). The competition between protein folding and aggregation: Off-lattice minimalist model studies. *Biotechnology and Bioengineering.* **89**, 78-87.

Cheng, Y. E., Kwoh, D. Y., Kwoh, T. J., Soltvedt, B. C., Zipser, D. (1981). Stabilization of a degradable protein by its overexpression in *Escherichia coli. Gene.* **14**, 121-130.

Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C. A., Pelletier, J.J., Paulus, H., and Xu, M.-Q. (1997). Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene.* **192**, 277–281.

Chong, S., Montello, G. E., Zhang, A., Cantor, E. J., Liao, W., Xu, M. -Q., Benner, J. (1998). Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucleic Acids Research.* **26**, 5109–5115.

Corchero, J. L., Villaverde, A. (1997). Plasmid Maintenance in *Escherichia coli* Recombinant Cultures is Dramatically, Steadily, and Specifically Influenced by Features of the Encoded Proteins. *Biotechnology and Bioengineering.* **58**, 625-632.

Crane, A., Jansen, R., Andrews, E., Ledley, F.D. (1992). Cloning and Expression of a Mutant with Altered Cobalamin Affinity that Causes *Mut* Methylmalonic Aciduria. *Journal of Clinical Investigation.* **89**, 391- 385.

Davis, G., Elisee, C., Newham, D., Harrison, R. (1999). New Fusion Protein Systems Designed to Give Soluble Expression in *Escherichia coli. Biotechnology and Bioengineering.* **65**, 382-388.

Davis, N. K., Greer, S., Jones-Mortimer, M. C., Perham, R. N. (1982). Isolation and mapping of glutathione reductase-negative mutants of *Escherichia coli* K12. *Journal of General Microbiology.* **128**, 1631-1634.

Dawson, P. E., Muir, T. W., Clark-Lewis, I., Kent, S. B. H. (1994). Synthesis of proteins by native chemical ligation. *Science.* **266**, 776–779.

De Marco, V., Stier, G., Blandin, S., de Marco, A. (2004). The solubility and stability of recombinant proteins are increased by their fusion to NusA. *Biochemical and Biophysical Research Communications.* **322,** 766–771.

Derman, A.I., Prinz, W.A., Belin, D. and Beckwith, J. (1993). Mutations that Allow Disulfide Bond Formation in the Cytoplasim of *Escherichia coli. Science.* **262**, 1744-1747.

Dobson, C. M., Wai, T., Leclerc, D., Wilson, A., Wu, X. C., Dore, C., Hudson, T., Rosenblatt, D. S., Gravel, R. A. (2002). Identification of the gene responsible for the *cblA* complementation group of vitamin B-12-responsive methylmalonic acidemia based on analysis of prokaryotic gene arrangements. *Proceedings of the National Academy of Science of the USA.* **99**, 15554-15559.

Dobson, C. (2001). Protein folding and its links with human disease. *Protein Folding to New Enzymes Biochemical Society Symposium.* **68**,1-26.

Duan, X., Gimble, F., Quiocho, A. (1997). Crystal structure of PI-*Sce*l, a homing endonuclease with protein splicing activity. *Cell.* **89**, 555-564.

Evans, T.C., Xu, M. Q. (2002). Mechanistic and kinetic considerations of protein splicing. *Chemical Reveiws.* **102**, 4869-4883.

Fenton, W., Hack, A., Willard, H., Gertler, A., Rosenberg, L. (1982). Purification and Properties of Methylmalonyl Coenzyme A Mutase from Human Liver. *Archives of Biochemistry and Biophysics.* **214**, 815-823.

Fuchshuber, A., Mucha, B., Baumgartner, E., Vollmer, M., Hildebrandt, F. (2000). *Mut[0]* Methylmalonic Acidemia: Eleven Novel mutations of the Methylmalonyl-CoA Mutase Including a Deletion-Insertion Mutation. *Mutation in Brief.* **16**, 179.

George, J., Blakesley, R., Chirikjian, J. (1980). Sequence-specific endonuclease *Bam*HI. Effect of hydrophobic reagents on sequence recognition and catalysis. *Journal of Biological Chemistry.* **255**, 6521 – 6524.

Gilbert. R., Fucini, P., Connell, S. (2004). Three-dimensional structures of translating ribosomes by cryo-EM. *Molecular Cell.* **14**, 57-66.

Goldberg, M., E., Rudolph R, Jaenicke, R. (1991). A Kinetic-Study of the Competition Between Renaturation and Aggregation During the Refolding of Denatured Reduced Egg-White Lysozyme. *Biochemistry.* **30**, 2790-2797.

Goloubinoff, P., Gatenby, A., Lorimer, G. H. (1989). GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli. Nature.* **337**, 44-47.

Gottesman, S. (1898). Genetics of proteolysis in *Escherichia coli. Annual Reviews in Genetics.* **23**, 163-198.

Goulding C. W. and Perry L. J. (2003). Protein production in *Escherichia coli* for structural studies by X-ray crystallography. *Journal of Structural Biology.* **142**, 133–143.

Groll, M., Bochtler, M., Brandstetter, H., Clausen, T., Hube, R. (2005). Molecular Machines for Protein Degradation. *ChemBioChem.* **6**, 222-256.

Grodberg, J., Dunn, J. J. (1988). *ompT* Encodes the Escherichia coli Outer-Membrane Protease that Cleaves T7-RNA Polymerase During Purification. *Journal of Bacteriology.* **170**, 1245-1253.

Becker-Hapak, M., Eisenstark, A. (1995). Role of *rpoS* in the regulation of glutathione oxidoreductase (*gor*) in *Escherichia coli. FEMS Microbiology Letters.* **134**, 39-44.

Hayes, M. (1998). The Refolding of Recombinant Human Liver Methylmalonyl-CoA Mutase from Inclusion Bodies in *Escherichia coli*. Massey University Thesis.

Hanahan, D., Jessee, J., Bloom, F. R. (1991). Plasmid Transformation of *Escherichia coli* and Other Bacterial. *Methods in Enzymology.* **204**, 63-113.

Heaton D., George, G., Garrison, G., Winge, D. (2001). The mitochondrial copper metallochaperone Cox17 exists as an oligomeric, polycopper complex. *Biochemistry.* **40**, 743-751.

Hennig, L., Schafer, E. (1998). Protein purification with C-terminal fusion of maltose binding protein. *Protein Expression and Purification.* **14**, 367-370.

Hirel, P. H., Schmitter, J. M., Dessen, .P, Fayat G., Blanqueyt, S. (1989). Extent of N-Terminal Methionine Excision from *Escherichia-coli* Proteins is Governed by the Side-Chain Lenght of the Penultimate Amino-Acid. *Proceedings of the National Academy of Sciences of the United States of America.* **86**, 8247-8251.

Hirata, R., Ohsumk, Y., Nakano, A., Kawasaki, H., Suzuki, K., Anraku, Y. (1990). Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae. Journal of Biological Chemistry.* **265**, 6726 – 6733.

Hochuli, E., Döbeli, H., and Schacher, A. (1987). New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *Journal of Chromatography.* **411**, 177-184.

Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R., Stuber, D. (1988). Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbant. *Bio-technology.* **6**, 1321-1325.

Humphries, H., Christodoulides, M., Heckels J. (2002). Expression of the class 1 outer-membrane protein of *Neisseria meningitidis* in *Escherichia coli* and purification using a self-cleavable affinity tag. *Protein Expression and Purification.* **26**, 243-248.

Innis, M. A. (1990). Optimisation of PCRs, pp 3-12. In PCR Protocols. A Guide to Methods and Applications. Edited by Innis, M. A., Gelfand, D. H., Sninsky, J. J., White, T. J. Academic Press Inc., Harcourt Brace Jovanovich Publishers, London.

Janata, J., Kogekar, N., Fenton, W.A. (1997). Expression and Kinetic Characterisation of Methylmalonyl-CoA Mutase from Patients with the *mut* Phenotype: Evidence for Naturally Occurring Interallellic Complementation. *Human Molecular Genetics.* **6**, 1457-1464.

Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M., Stevens, T. H. (1990). Protein Splicing Converts The Yeast *TFP1* Gene-Product to the 69-kD Subunit of the Vacuolar H+-Adenosine Triphosphatase. *Science.* **250**, 651-657.

Kapust, R. B., and Waugh, D. S. (2000). Controlled Intracellular Processing of Fusion Proteins by TEV Protease. *Protein Expression and Purification.* **19**, 312–318.

Kapust, R. B. and Waugh, D. S. (1999). *Escherichia coli* Maltose-binding Protein in Uncommonly Effective at Promoting the Solubility of Polypeptides to which it is Fused. *Protein Science.* **8**, 1998-1674.

Kessler, C., Manta, V. (1990). Specificity of restriction endonucleases and DNA modification methyltransferases - a review. *Gene.* **92**, 1-248.

Khoo K., Chang C., Schubert J. (2005). Expression and purification of the recombinant His-tagged GST-CD38 fusion protein using the baculovirus/insect cell expression system. *Protein Expression and Purification.* **40**, 396-403.

Kolhouse, J., Allen, R. (177). Recognition of two intracellular cobalamin binding proteins and their recognition as methylmalonyl CoA mutase and methione synthase. *Proceedings of the National Academy of Science of the USA.* **74**, 921-925.

Kolhouse, J., Stabler, S., Allen, R. (1988). L-Methylmalonyl-CoA Mutase from Human Placenta. *Methods in Enzymology.* **166**, 407-415.

Kolodner, R., Fishel, R. A., Howard, M. (1985). Genetic-Recombination of Bacterial Plasmid DNA – Effect of *RecF* Pathway Mutations on Plasmid Recombination in *Escherichia-coli. Journal of Bacteriology.* **163**, 1060-1066.

Korotkova, N., Lidstrom, M. E. (2004). MeaB is a component of the methylmalonyl-CoA mutase complex required for protection of the enzyme from inactivation. *Journal of Biological Chemistry.* **279**, 13652-13658.

Koshy, T. I.. Luntz, T. L., Garber, E. A. E., Margoliash, E. (1992). Expression of Recombinant Cytochromes c from Various Species in *Saccharomyces cerevisiae*; Post-translational Modifications. *Protein Expression and Purification.* **3**, 441-452.

Kurland, C., Gallant, J. (1996). Errors of heterologous protein expression. *Current Opinion in Biotechnology.* **7**, 489-493.

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of the Bacteriophage T4. *Nature.* **227**, 680-685.

LaVallie, E. R., Rehemtulla, A., Racie, L. A., Diblasio, E. A., Ferenz, C., Grant, K. L., Light, A., McCoy, J. M. (1993). Cloning and Functional Expression of a cDNA-Encoding the Catalytic Subunit of Bovine Enterokinase. *Journal of Biological Chemistry.* **268**, 23311-23317.

LaVallie, E. R., Diblasio, E. A., Kovacic, S., Grant, K, L., Schendel, S. F., McCoy, J. M. (1993). A Thioredoxin Gene Fusion Expression System that Circumvents Inclusion Body Formation in the *Escherichia coli* Cytoplasm. *Bio-technology.* **11**, 187-193.

LaVallie, E., Lu, Z., Diblasio-Smith E., Collins-Racie, L., McCoy, J. (2000). Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli. Method in Enzymology.* **326**, 322-340.

Leadlay, P., Ledley, F. (1989). in Cobalamin '88: Proceedings of the First International Symposium on Biomedicine and Physiology of Vitamin B12 (Linnell, J.C & Bhatt, H. R., eds.), pp. 325-332, Children's Medical Charity, London.

Leal, N., Park, S., Kima, P., Bobik, T. (2003). Identification of the Human and Bovine ATP: Cob(I)alamin Adenosyltransferase cDNA Based on Complementation of a Bacterial Mutant. *Journal of Biological Chemistry.* **278**, 9227-9234.

Ledley, F.D., Lumetta, M.R., Zoghbi, H.Y., Van Tuinen, P., Ledbetter, S.A., Ledbetter, D.H. (1988). Mapping of Human Methylmalonyl-CoA Mutases (MUT) Locus on Chromosome 6. *American Journal of Human Genetics.* **42**, 839-846.

Loo, T., Patchett, M. L., Norris, G. E., Lott, J. S. (2002). Using secretion to solve a solubility problem: High yield expression in *Escherichia coli* and purification of the bacterial glycoamidase PNGase F. *Protein Expression and Purification.* **24**, 90-98.

Looman, A. C., Bodlaender, J., Comstock, L. J., Eaton, D., Jhurani, P., Deboer, H. A., Vanknippenberg, P. H. (1987). Influence of the Codon Following the AUG Initiation Codon on the Expression of a Modified LacZ Gene in *Escherichia-coli. EMBO Journal.* **6**, 2489-2492.

Lüttkopf, D., Müller, U., Skov, P.S., Ballmer-Weber, B.K., Wüthrich, B., Skamstrup, H. K., Poulsen, L.K., Kästner, M., Haustein, D., Vieths, S. (2001). Molecular Comparison of four variants of a major allergen in hazelnut (Corylus avellana) *CorA* 1.04 with the major hazel pollen allergen *CorA* 1.01. *Immunology.* **38**, 515–525.

Maina, C. V., Riggs, P. D., Grandea, A. G., Slatko, B. E., Moran, L. S., Tagliamonte, J. A., MeReynolds, L. A., di Guan, C. (1988). An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene.* **74**, 365-373.

Makhoul C. H. and Trifonov E. N. (2002). Distribution of rare triplets along mRNA and their relation to protein folding. *Journal of Biomolecular structure and Dynamics.* **20**, 413-420.

McKie, N., Keep, N.H., Patchett, M.L., Leadley, P.F. (1990). Adenosylcobalamin-dependent methylmalonylCoA mutase from *Propionibacterium shermanii. Biochemical. Journal.* **269**, 293-298.

Mancia, F., Keep, N.H., Nakagawa, A. (1996). How Coenzyme B12 Radicals are Generated: The Crystal Structure of Methylmalonyl - Coenzyme A Mutase at 2Å Resolution. *Structure*. **4**, 339-350.

McNulty, D., Claffee, B, Huddleston, M., Kane, J. (2003). Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli. Protein Expression and Purification.* **27**, 365–374.

Morita, E. H., Sawasaki, T., Tanaka, R., Endo, Y., Kohno, T. (2003). A wheat germ cell-free system is a novel way to screen protein folding and function. *Protein Science*. **12**, 1216-1221.

Nasri, M., Thomas, D. (1986). Relaxation of recognition sequence of specific endonuclease HindIII. *Nucleic Acids Research*. **14**, 811-821.

Neupert, W. (1997). Protein import into mitochondria. *Annual Review of Biochemistry*. **66**, 863-917.

Nham, S.U., Wilkemeyer, M.F., Leadey, F.D., (1990). Structure of the Human Methylmalonyl-CoA Mutase (MUT) Locus. *Genomics*. **8**, 710-716.

Novy, R., Drott, D., Yaeger, K. and Mierendorf, R. (2001). Overcoming the codon bias of *E. coli* for enhanced protein expression. *inNovations Newsletter*. **12**, 1–3.

Perler, E. (2000). InBase, the intein database. *Nucleic Acid Research*. **28**, 344-345.

Peters, H.L., Nefedov, M., Lee, L.W., Abdenur, J.E., Chamoles, N.A., Ioannou, P.A. (2002). Molecular Studies in Mutase-deficient (MUT) Methylmalonic Aciduria: Identification of Five Novel Mutations. *Human Mutation*. Mutation in Brief #545, Online.

Pryor, K. D. and Leiting B. (1997). High-Level Expression of Soluble Protein in *Escherichia coli* Using a His6-Tag and Maltose-Binding-Protein Double-Affinity Fusion System. *Protein Expression and Purification*. **10**, 309–319.

Richarme G., Caldas T.D. (1997). Chaperone properties of the bacterial periplasmic substrate-binding proteins. *Journal of Biological Chemistry.* **272,** 15607-15612.

Robb, F. T., Maeder, D. L., Brown, J. R., DiRuggiero, J., Stump, M. D., Yeh, R. K., Weiss, R. B. and Dunn, D. M. (2001). Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. *Methods in Enzymology.* **330**, 134-157.

Robinson, M., Lilley, R., Little, S., Emtage, J. S., Yarranton, G., Stephens, P., Millican, A., Eaton, M., and Humphreys, G. (1984). Codon usage can affect efficiency of translation of genes in *Escherichia coli. Nucleic Acids Research.* **12**, 6663–6671.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular Clonning; A Laboratory Manual (2nd edition). Cold Spring Harbor Laboratory Press.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA Sequencing with Chain Terminating Inhibitors. *Proceedings of the National Academy of Science U.S.A.* **74**, 5463-5467.

Sachdev, D. and Chirgwin, J. M. (1998). Order of Fusions between Bacterial and Mammalian Proteins Can Determine Solubility in *Escherichia coli. Biochemical and Biophysical Research Communications.* **244**, 933–937.

Sati, S. P., Singh, S. K., Kumar, N., Sharma, A. (2002). Extra terminal residues have a Profound Effect on the Folding and Solubility of a *Plasmodium falciparum* sexual stage-specific Protein Over-expressed in *E. coli. European Journal of Biochemistry.* **269**, 5259-5263.

Sawada, T., Ledley, F. D. (1992). Correction of Methylmaloyl-CoA Mutase Deficiency in MUT^{0} Fibroblasts and Constitution of Gene-Expression in Primary Human Hepatocytes by Retroviral-Mediated Gene-Transfer. *Somatic Cell and Molecular Genetics.* **18**, 507-516.

Scharf, S. J. (1990). Cloning With PCR pp 84-91. In PCR Protocols. A Guide to Methods and Applications. Edited by Innis, M. A., Gelfand, D. H., Sninsky, J. J., White, T. J. Academic Press Inc., Harcourt Brace Jovanovich Publishers, London.

Schein C. H. and Noteborn M. H. M. (1988). Formation of Soluble Recominant Proteins in *Escherichia-coli* is Favoured by Lower Growth Temperature. *Bio-technology.* **6**, 291-294.

Skorko-Glonek, J., Zurawa. D., Kuczwara, E. (1999). The *Escherichia coli* heat shock protease HtrA participates in defense against oxidative stress. *Molecular and General Genetics.* **262**, 342-350.

Sorensen, H., Kristensen, J., Sperling-Petersen H. (2004). Soluble expression of aggregating proteins by covalent coupling to the ribosome. *Biochemical and Biophysical Research Communications.* **319**, 715-719.

Stankovics, J., Crane, A. M., Andrews, E., Wu, C. H., Wu, G. Y., Ledley, F. D. (1994). Overexpression of Human Methylmaloyl CoA Mutase in Mice After in-vivo Gene-Transfer with Asialoglycoprotein Polylysine DNA Complexs. *Human Gene Therapy.* **5**, 1095-1104.

Studier, W. F., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1990). Use of T7 RNA Polymerase to Drive Expression of Cloned Genes. *Methods in Enzymology.* **185**, 60-89.

Stryer, L. (1995). Biochemistry, fourth edition. W. H Freeman and Company. 19-23.

Sun, Q. M., Cao, L., Fang, L., Chen, C., Dai, J., Chen, L. L., Hua, Z. C. (2005). Expression, purification of human vasostatin120-180 in *Escherichia coli*, and its anti-angiogenic characterization. *Protein Expression and Purification.* **39**, 288-295.

Tabor, S., Richardson, C. (1995). A Single Residue in DNA Polymerases of the Escherichia coli DNA Polymerase I Family is Critical for Distinguishing Between

Deoxy- and Dideoxyribonucleotides. *Proceedings of the National Academy of Science U.S.A.* **92**, 6339-6343.

Takagi, F., Koga, N., Takada, S. (2003). How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: Molecular simulations. *Proceedings of the National Academy of Science U.S.A.* **100**, 11367-11372.

Taoka, S., Padmakumar, R., Grissom, C., Banerjee, R. (1997). Magnetic Field Effects on Coenzyme B12 – Dependant Enzymes: Validation of Ethanolamine Ammonia Lyase Results and Extension to Human Methylmalonyl CoA Mutase. *Bioelectromagnetics.* **18**, 506-513.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology.* **60**, 523-533.

Thoma, N.H., Leadlay, P.F. (1996). Homology Modeling of Human Methylmalonyl-CoA Mutase: A Structural Basis for Point Mutations Causing Methylmalonic Aciduria. *Protein Science.* **5**, 1922-1927.

Tobias, J. W., Shrader, T. E., Rocap, G., Varshavsky, A. (1991). The N-End Rule in Bacteria. *Science.* **254**, 1374-1377.

Tollinger, M., Eichmuller, C., Konrat, R., Huhta, M., Marsh, E., Krautler, B. (2001). The B12-Binding Subunit of Glutamate Mutase from *Clostridium tetanomorphum* Traps the Nucleotide Moiety of Coenzyme B12. *Journal of Molecular Biology.* **309**, 777-791.

Watanabe, T., Ito, Y., Yamada, T., Hashimoto, M., Sekine, S., Tanaka, H. (1994). The Roles of the C-Terminal Domain and Type-III Domains of Chitinase A1 from *Bacillus-Circulans* WL-12 in Chitin Degridation. *Journal of Bacteriolology.* **176**, 4465-4472.

Witholt, B., Boekhout. M., Brock, M., Kingma, J., van Heerikhuizen, H., de Leij, L. (1976). An efficient and reproducible procedure for the formation of spheroplasts from variously grown *Escherichia coli. Analytical Biochemistry.* **74**, 160-170.

Xu, M., Southworth, M. W., Mersha, F. B., Hornstra, L. J., and Perler, F. B. (1993). In vitro protein splicing of purified precursor and the identification of a branched intermediate. *Cell.* **75**, 1371–1377.

Xu, M., Perler, F. B. (1996) The mechanism of protein splicing and its modulation by mutation. *EMBO Journal.* **15**, 5146–5153.

Xu, H., Zhang, G., Ji, X., Cao, L., Shu, L., Hua, Z. (2005). Expression of soluble, biologically active recombinant human endostatin in *Escherichia coli. Protein Expression and Purification.* **41**, 252–258.

Yang, X., Sakamoto, O., Matsubara, Y., Kure, S., Suzuki, Y., Aoki, Y., Suzuki, Y., Sakura, N., Takayanagi, M., Iinuma, K., Ohura, T. (2004). Mutation analysis of the *MMAA* and *MMAB* genes in Japanese patients with vitamin B-12-responsive methylmalonic acidemia: identification of a prevalent *MMAA* mutation. *Molecular Genetics and Metabolism.* **82**, 329-333.

Zagalak, B., Retey, J. (1974). Studies on Methylmalonyl-CoA Mutase from *Propionibacterium shermanii. European Journal of Biochemistry.* **44**, 529-535.