University of Pretoria etd – Coetzee, L (2006)

Chapter 3

Expression of synthetic *P. falciparum* dihydrofolate synthase-folylpolyglutamate synthase (DHFS-FPGS)

3.1 Introduction

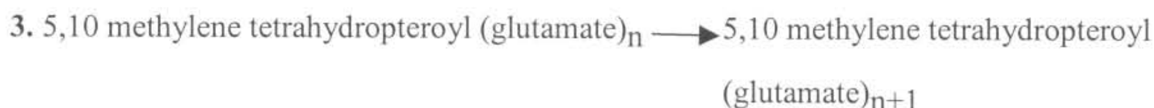
The two activities DHFS (E.C. 6.3.2.12) and FPGS (E.C. 6.3.2.17) play vital roles (Figure 1.4) in the synthesis and recycling of tetrahydrofolate (vitamin B9), which is used in the synthesis of purines, thymidylate, methionine, formylmethionine tRNA, pantothenate and the conversion of serine to glycine (Stokstad and Koch, 1967). Mammals do not have a DHFS enzyme, but other eukaryotes such as yeast, fungi and plants have separate DHFS and FPGS proteins. The *dhfs-fpgs* identified in *P. falciparum* is the first eukaryotic gene to encode a single bifunctional protein containing both enzyme activities. Other protozoan parasites such as *Leishmania* only have an FPGS enzyme (Fadili *et al.*, 2002). In prokaryotes such as *E. coli* and *Corynebacterium* species the two activities are contained within a single bifunctional protein (Bognar *et al.*, 1985; Shane, 1980). In *E. coli* the DHFS and FPGS functions are furthermore co-linear (Keshavjee *et al.*, 1991). This means that the two functions are not distinguishable at DNA or amino acid level, i.e. there is no clear allocation of activity to certain protein domains in contrast to other bifunctional proteins such as *P. falciparum* DHFR-TS that is divided in an N-terminal domain (nt 1-684) with DHFR activity and a C-terminal domain (nt 969-1824) with TS activity (Bzik *et al.*, 1987). As a result of the co-linearity, a single point mutation in *E. coli* is sufficient to abolish both DHFS and FPGS activities (Keshavjee *et al.*, 1991). It is not yet clear whether the DHFS and FPGS activities are co-linear in *P. falciparum*. If the *P. falciparum* DHFS and FPGS activities were also co-linear it would increase the effectivity of drugs by eliminating both enzyme activities through a single point of inhibition. Considering that the DHFS and FPGS enzymes each function in a different route of folate metabolism, simultaneous inhibition of both activities would be synergistic such as that observed by the combination of sulfadoxine-pyrimethamine, which inhibits the dihydropteroate synthase component of DHPS-PPPK and the dihydrofolate reductase component of DHFR-TS.

University of Pretoria etd – Coetzee, L (2006)

The following reactions are catalysed by DHFS and FPGS (refer to figure 1.4):

DHFS

FPGS



Reaction 1 is the addition of the first glutamate residue through an amide linkage to the carboxyl end of the p-amino benzoic acid (PABA) group of dihydropteroate, catalysed by DHFS. Reactions 2 and 3 are catalysed by FPGS. In all the reactions addition takes place at the γ- carboxyl group of glutamate in a K^+- and ATP-dependent manner. The phosphate group of ATP activates the carboxyl group of the dihydropteroate substrate (DHFS) or the glutamate group of folic acid (FPGS) through the formation of an acyl phosphate intermediate. The polyglutamate tail length (n) varies between 1 and 10 and is usually shorter for prokaryotes than for eukaryotes (Bognar *et al.,* 1985). In *E. coli* the polyglutamate chain length is three (Osborne *et al*., 1993), in *Leishmania* five (Santi *et al*., 1987) and in *P. falciparum* it was observed that 5-methyl-tetrahydropteroyl pentaglutamate was the predominant cellular form of folates (Krungkrai *et al*., 1989). Mammals have even larger chain lengths of up to nine glutamate residues (Osborne *et al*, 1993). Polyglutamylation is essential since it is required for the increased cellular retention of folates and affinity for folate-utilising enzymes as well as the accumulation of folates in the mitochondria where glycine synthesis takes place (Moran, 1999).

Due to the absence of DHFS in humans and expected differences between FPGS activities of humans and *P. falciparum* as shown above the possibility of designing selective drugs that are not detrimental to the human host is increased. This can, however, only be confirmed with functional assays for which significant amounts of active enzyme are required. The only crystal structure currently available for homology modelling is that of *Lactobacillus casei* FPGS (Sun *et al*., 1998). Sufficient amounts of correctly folded, soluble DHFS-FPGS are thus needed to determine the kinetic properties of the protein and determination of the 3-dimensional structure of the enzyme.

University of Pretoria etd – Coetzee, L (2006)

Chapter 2 described the synthesis of a codon-optimised *dhfs-fpgs* for increased recombinant expression in *E. coli*. This chapter focuses on the optimisation of expression of the synthetic *P. falciparum dhfs-fpgs* gene and its functional verification to obtain sufficient amounts of soluble protein for future enzyme activity studies and structure determination.

The *E. coli* codon-adapted *dhfs-fpgs* obtained by PCR-mediated gene synthesis as described in Chapter 2 was cloned into several expression vectors and expressed in different *E. coli* cell lines to select the best system for expression of soluble protein. Gene complementation assays were performed to verify the functional activities of the gene products and to compare the complementation efficiency and thus activity of different tagged constructs to that of a tagless construct. Preliminary purification studies were also performed as a basis for future protein isolation strategies.

3.2 Materials and methods

3.2.1 Constructs, vectors and cell lines

Two systems were used for protein expression (Table 3.1): the IPTG inducible *T7lacUV5* promoter of the pET vector (Novagen, EMD Biosciences, Germany) and the anhydrotetracycline inducible *tetA* resistance gene promoter of the pASK vector (IBA, Germany). In both of these systems protein expression from the plasmid is tightly regulated to prevent the overproduction of protein, which might be toxic to cell growth.

Table 3.1: Vector systems used for the recombinant expression of synthetic *P. falciparum dhfs-fpgs*. Refer to Appendix E for the vector maps.

Vector	Antibiotic resistance	Tag	Induction	Proteolytic cleavage site
pET15b (Novagen)	Ampicillin	N-terminal hexahistidine peptide tag	T7*lacUV5 promoter*, IPTG	Thrombin
pET22b (Novagen)	Ampicillin	C-terminal hexahistidine peptide tag	T7*lacUV5 promoter*, IPTG	None
pASK-IBA3 (IBA)	Ampicillin	C-terminal StrepII tag (NH_2- WSHPQFEK - COOH)	*tetA* promoter, anhydrotetracycline	None

The pET vector systems (Novagen, EMD Biosciences, Germany) were chosen for His_6-tagged expression and affinity purification and the pASK vector (IBA, Germany) for Strep-tagged expression and affinity purification of recombinant proteins in *E. coli*. Two different synthetic *P. falciparum dhfs-fpgs* gene constructs were made, one with the stop codon for N-terminal His_6-tagged gene expression and another without the stop codon for readthrough and expression of a C-terminal His_6-tag or Strep-tag fused to the protein. The protein was also expressed without a tag. In this instance the construct was cloned with the stop codon preceding the C-terminal histidine tag of the pET22b vector (Novagen, EMD Biosciences, Germany) to terminate transcription before the tag. The outer primers (outF and outR or alt4R) were designed with *Nde*I and *BamH*I restriction enzyme sites for in-frame cloning into the pET vector system using ATG in the *Nde*I restriction site (CAT↓ATG) as the start codon (Table 2.1, Chapter 2). For cloning into the pASK-IBA3 system another set of primers (ib3F and ib3R) were designed for incorporation of two *Bsa*I sites at the 5'- and 3'-ends of the gene and subsequent directional cloning into the pASK-IBA3 vector (Table 3.2).

Table 3.2 Primers used for cloning of *P. falciparum dhfs-fpgs* into the pASK-IBA3 vector for C-terminal Strep-tagged expression. The highlighted parts of the sequence indicate the *Bsa*I restriction enzyme site (GGTCTCN ↓ NNNN).

Primer name	Sequence (5' to 3')	Length (nt)	T_m (°C)
ib3F	GCATCGGGTCTCGAATGGAGAAGAACCAGAACG	33	68
ib3R	GCATCGGGTCTCAGCGCTCACCAGGCTCGGTTCGTTC	37	75

Different *E. coli* cell lines were used to determine which system gave optimal protein expression (Table 3.3). Cells containing the pLysS plasmid show increased stability of toxic genes expressed by the system. The pLysS plasmid encodes a small amount of T7 lysozyme, which binds to T7 RNA polymerase thereby inhibiting transcription and preventing cell death (Huang *et al.*, 1999). BL21 Star (DE3) cells were chosen for their high level of gene expression and BL21 Gold (DE3) pLysS cells for ease of transformation (Invitrogen Technical Data, La Jolla, USA).

Table 3.3: *E. coli* strains used as hosts for protein expression.

Expression host	Genotype	Antibiotic resistance
BL21 (DE3) pLysS (Stratagene, La Jolla, USA)	B strain F^- *dcm ompT hsdS$_B$ ($r_B^- m_B^-$) gal* λ(DE3) [pLysS Cam^r]	Chloramphenicol
BL21 Star (DE3) (Invitrogen, La Jolla, USA)	B strain F^- *ompT hsdS ($r_B^- m_B^-$) gal dcm rne131* (DE3)	None
BL21 Gold (DE3) pLysS (Invitrogen La Jolla, USA)	B strain F^- *dcm ompT hsdS ($r_B^- m_B^-$)* dcm^+ Tet^r *gal* λ(DE3) *endA* Hte [pLysS Cam^r]	Chloramphenicol
BL21 Codon plus pRIL (Invitrogen La Jolla, USA)	B strain F^-*ompT hsdS ($r_B^- m_B^-$) dcm*$^+$$Tet^r$*gal endA* Hte [argU ileY leuW Cam^r]	Chloramphenicol

3.2.2 Protein expression

$CaCl_2$ competent expression hosts (section 2.2.4.2) were freshly transformed with plasmid containing the different tagged gene constructs and plated onto LB plates with ampicillin (100 µg/ml) to select for the plasmid containing the gene, and supplemented with 100 µg/ml chloramphenicol for selection of the pLysS plasmid in BL21 (DE3) pLysS and BL21 Gold (DE3) pLysS cells. A portion (400 µl) of the transformed cells were not plated but diluted into 10ml LB liquid media containing 50µg/ml ampicillin and grown overnight at 30°C with

shaking for population expression (the total number of cells that received the plasmid). The overnight cultures were then diluted 20-fold in LB liquid media with ampicillin (50 µg/ml) and grown at 30°C with shaking until an optical density of 0.4 at 600nm was reached. Uninduced samples (2 ml each) were taken and a final concentration of 0.1 mM IPTG added to the cells for induction of protein expression from the pET system. For Strep-tagged expression a final concentration of 0.2 µg/ml anhydrotetracycline (of a 2mg/ml stock solution in dimethylformamide) was added to each culture for induction of protein expression. Expression was carried out at 30°C for 3.5 hours, after which cells were pelleted by centrifugation at 16000xg for 1 min or 5000xg for 15 min (for culture volumes >50ml) at 22°C. The pellets were dissolved into Bugbuster Protein extraction reagent (Novagen, EMD Biosciences, Germany), which is a zwitterionic detergent for gentle lysis of the cells. A volume of 5 ml "Bugbuster" per gram wet cell paste was used. For protein extraction the samples were left at 4°C overnight whereafter the induced samples were split into two fractions; one was used for total protein analysis and the other was centrifuged at 16000xg for 5min or 16000xg for 20 min (for culture volumes >50ml). The resulting pellet was analysed as the insoluble fraction and the supernatant as the soluble fraction.

3.2.3 Determination of the protein concentration

The Bradford method was used for protein concentration determination (Bradford, 1976). A stock solution of 10 mg/ml bovine serum albumin (BSA) was diluted to obtain a standard concentration range of 200, 100, 50, 25, 12.5 and 6.25 µg/ml protein. To 100 µl of each BSA standard or protein samples, 100 µl Coomassie Plus Assay Reagent (Pierce Biotechnology Inc., Rockford, USA) was added and the absorbancy measured at 595nm in 96 well microtitre plates (Bibby Sterilin Ltd., Stone, Staffordshire, UK). Protein concentrations were also estimated by comparison of the intensity of protein bands on an SDS-PAGE gel with that of the molecular marker, Broad Range Protein Molecular Mass Markers (Promega, Wisconsin, USA), which contains 0.3 µg/µl of the 50 kDa band and 0.1 µg/µl of the rest of the bands.

3.2.4 SDS-PAGE analysis

Polyacrylamide gel electrophoresis (PAGE) was performed with a 5% acrylamide-methylene-(bis) acrylamide stacking gel containing 0.5% SDS, 0.5% TEMED, 0.1% ammonium persulfate, 0.125 M Tris-Cl (pH 6.8) and an 8 or 12% acrylamide-methylene-(bis)acrylamide separating gel containing 0.535 M Tris-Cl (pH 8.8), 0.5% w/v SDS, 0.5% v/v TEMED and

0.1% w/v ammonium persulfate. Electrophoresis was performed in a 0.025 M Tris-0.2M glycine buffer (pH 8.3). Samples were diluted 1:1 in a 2X concentrated SDS sample buffer (125 M TrisCl, 0.02% w/v SDS, 0.5% v/v glycerol, 0.05% v/v β-mercaptoethanol and 0.01% w/v bromophenol blue) and denatured by boiling for 5 min. Gels were run at 70 Volts in the Minigel G-41 system (Biometra, Germany). Protein bands were visualised with Coomassie Blue G250 staining (0.1g Coomassie brilliant blue G250 in 40% methanol, 10% acetic acid, destained with 40% methanol, 10% acetic acid). Silver staining was used for visualisation of low protein concentrations. Gels were fixed for 30 min in 30% ethanol, 10% acetic acid and then treated for 30 min with 30% ethanol, 0.5M sodium acetate, 0.5% glutaraldehyde and 0.2% $Na_2S_2O_3$. Gels were then washed three times for 10 min with water and stained for 30 min in 0.1% $AgNO_3$, 0.02% formaldehyde and developed in 0.5% Na_2CO_3, 0.01% formaldehyde and developing stopped by the addition of 0.5 M EDTA (Merril *et al.*, 1981).

3.2.5 Functional complementation

Complementation of the *E. coli* K-12 mutant strain, Sf4 (F^- *folC strA recA tn10: srlC*) with the different synthetic constructs was performed to compare the complementation efficiency of their DHFS and FPGS activities (Bognar e*t al.*, 1985). This specific strain has a single point mutation, A309T that reduces DHFS and FPGS activities to less than 3% of the wild type activities (Bognar *et al.*, 1985). Cells therefore produce basal folate levels to overcome thymine deficiency but are auxotrophic for methionine. Glycine is added to stimulate growth. Complementation was performed on minimal media plates containing 0.5% (w/v) glucose, 1.6% agar, 50 μg/ml streptomycin to select for the Sf4 mutant, 100 μg/ml ampicillin to select for the vectors containing the *dhfs-fpgs* gene in an M9 salts solution (1.3%(w/v) $Na_2HPO_4.7H_20$, 0.3%(w/v) KH_2PO_4, 0.05%(w/v) NaCl and 0.1%(w/v) NH_4Cl). Positive controls were supplemented with 40 μg/ml methionine and 20 μg/ml glycine. Constructs used for complementation were the vector alone (referred to further as U4) as negative control, the native *P. falciparum dhfs-fpgs* (referred to further as U5), the N-terminal histidine-tagged synthetic *dhfs-fpgs* (referred to in further as nh), the C-terminal histidine-tagged synthetic *dhfs-fpgs* (referred to in further as ch), the C-terminal Strep-tagged synthetic *dhfs-fpgs* (referred to in further as cs) and the tagless synthetic *dhfs-fpgs* (referred to in further as tl). The following titration method was used for complementation: Single colonies of Sf4 cells transformed with each of the above constructs were picked and grown in LB liquid medium at 30°C with shaking till growth was slightly visible (OD_{600nm} = 0.1). Of each of these cultures, 100 μl was placed in 5 ml LB liquid culture medium containing 50 μg/ml ampicillin and 50 μg/ml streptomycin,

grown overnight till saturation (OD_{600nm}> 1) and thus comparable cell density. A dilution series of 10-fold, 100-fold and 1000-fold of each culture volume were made and 5 µl of each dilution spotted onto the minimal media plates with and without supplementation (positive control). For complementation in minimal liquid media (0.5% w/v glucose, 50µg/ml streptomycin, 50µg/ml ampicillin and the M9 salts solution), 100 µl duplicate saturated overnight cultures of each of the constructs above, were placed in 5 ml culture medium and growth monitored at OD_{600nm} after 24 hours at 30°C.

3.2.6 Partial protein purification

3.2.6.1 Unfolding/refolding protocol

After protein extraction and separation of the soluble and insoluble fractions by centrifugation (refer to section 3.2.2), the inclusion bodies from BL21 Star (DE3) cells expressing the C-terminal His_6-tagged protein were further purified using the manufacturer's protocol for Bugbuster Protein Extraction Reagent (Novagen, EMD Biosciences, Germany). Cell pellets resulting from 50 ml of liquid culture were resuspended in 1.25 ml undiluted "Bugbuster" and mixed well by vortexing. Lysozyme (10mg/ml stock solution) was added to a final concentration of 200 µg/ml and the mixture vortexed for 5 min at 22°C. To this, 7.5 ml of a 10-fold dilution of "Bugbuster" was added, the mixture vortexed for 1 min and centrifuged at 16 000xg for 15 min at 4°C. The supernatant was removed and the pellet resuspended in 1 ml 10-fold diluted "Bugbuster". The sample was mixed well and centrifuged at 16 000xg for 15 min. This was repeated twice. Solubilisation of the purified inclusion bodies was done as follows (Sirawaraporn *et al.*, 1993): The remaining pellet was suspended in 1 ml of 20 mM potassium phosphate buffer (pH 7) containing 0.1 mM EDTA, 10 mM DTT, 0.2 M KCl and 6 M guanidine HCl and left at 4°C for 1 hour for solubilisation of the protein. The protein mixture was then added dropwise to achieve a 20 times dilution in a 20 mM potassium phosphate buffer (pH 7) containing 0.1 mM EDTA, 10 mM DTT, 0.2 M KCl and 20% glycerol at 4°C for refolding of the protein. This mixture was left overnight at 4°C and then centrifuged at 16000xg for 20 min to pellet all the remaining insoluble proteins. The soluble fraction was then used for affinity purification.

3.2.6.2 Affinity chromatography

Purification of histidine-tagged protein was conducted by immobilised metal ion affinity chromatography (IMAC) with HIS select HC Nickel affinity gel (Sigma-Aldrich, St.Louis, USA). This method of affinity purification relies on the electrostatic interaction between the

hexahistidine peptide tag of the target protein (which is negatively charged at pH 8) and the positively charged Ni^{2+} immobilised on the column resin. Unbound proteins are washed off and the target protein with the hexahistidine tag eluted by the addition of a high imidazole concentration, which binds competitively to the Ni^{2+} column resin. The batch purification method was used for purification of protein: 250 µl of the nickel affinity gel suspension was added to a microcentrifuge tube and centrifuged for 30 sec at 5000xg. The supernatant was removed and 200 µl of equilibration buffer (50 mM sodium phosphate (pH 8) and 0.3 M sodium chloride) was added. The solution was mixed well and centrifuged for 30 sec at 5000xg. The supernatant was discarded and 1 ml of the soluble fraction obtained after cell lysis as described in section 3.2.2 (adjusted to pH 8 with equilibration buffer) was added to the affinity gel matrix, mixed for 1 minute and centrifuged as before. For larger volumes of soluble fraction, the previous step was repeated a few times to load the whole fraction stepwise onto the matrix. The affinity gel was washed two times with 1 ml equilibration buffer. Each time the gel was mixed with the buffer, it was centrifuged at 5000xg for 30sec and the supernatant removed. The target protein was eluted in 250 µl of elution buffer (50 mM sodium phosphate, pH 8, 0.3 M sodium chloride and 250 mM imidazole). The gel and buffer were mixed well and centrifuged as before. This step was repeated with another 250 µl of elution buffer. Fractions (0.5 ml each) were taken for every wash and elution step and analysed with SDS-PAGE.

3.2.6.3 Size exclusion high performance liquid chromatography

Size exclusion chromatography was performed on a 300 x 7.8 mm Biosep-Sec-S3000 HPLC column (Phenomenex, California, USA) linked to the Beckman Gold HPLC system (Pump system 2). All buffers were filtered though a 0.22 µm membrane and degassed before use. Low Molecular weight markers (Amersham, Buckinghamshire, UK) made up to a concentration of 0.2 µg/µl in the size exclusion buffer (50 mM potassium phosphate (pH 7), 150 mM KCl) was used to calibrate the system. Soluble fractions of the uninduced and induced protein samples from BL21 (DE3) pLysS cells containing either C-terminal His_6-tagged- or tagless DHFS-FPGS were prepared as described in section 3.2.2. The soluble fractions were diluted 1:1 with the size exclusion buffer to adjust the pH to 7 and filtered through a 0.22 µm membrane. Samples were loaded onto the column and eluted with the size exclusion buffer at a constant flow rate of 0.5 ml/min for 40 min. The column was washed afterwards with the same buffer and stored in 0.05% sodium azide. Fractions (1ml each) collected from the size exclusion column were analysed with SDS-PAGE. Blank runs (consisting only of the size exclusion buffer) were performed before each sample run.

3.3 Results

3.3.1 Expression of a variety of *dhfs-fpgs* constructs

In all the different cell lines, the pET22b-*dhfs-fpgs* (C-terminal His_6-tag) construct was expressed to a greater extent than the pET15b-*dhfs-fpgs* (N-terminal His-tag) construct (Figure 3.1 A vs. B). The highest expression of the N- and C-terminal His-tagged fusion proteins was obtained from the BL21 Star (DE3) cell line (~20% of the total cell protein content) when the different *E. coli* cell lines BL21 (DE3) pLysS, BL21 Gold (DE3) pLysS and BL21 Star (DE3) were compared. The other cell lines showed only a slight increase in expression after induction by IPTG.

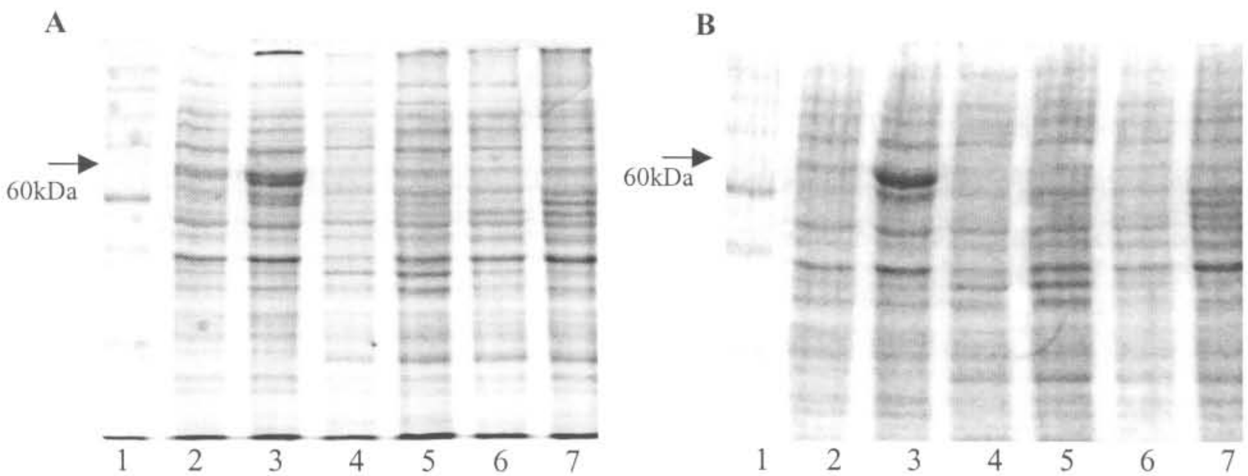


Figure 3.1: Expression of (A) pET15b-*dhfs-fpgs* (N-terminal His_6 tag) and (B) pET22b-*dhfs-fpgs* (C terminal His_6 tag) in various cell lines. Lane 1: Broad Range Protein Molecular Mass Markers, lane 2: BL21 Star (DE3) uninduced, lane 3: BL21 Star (DE3) induced, lane 4: BL21 Gold (DE3) pLysS uninduced, lane 5: BL21 Gold (DE3) pLysS induced, lane 6: BL21 (DE3) pLysS uninduced, lane 7: BL21 (DE3) pLysS induced. Arrows indicate the expected protein size (60kDa).

Comparison of the soluble and insoluble fractions of the N- (results not shown) and C-terminal His tagged constructs expressed by the BL21 Star (DE3) cells revealed however that the proteins were expressed as inclusion bodies (Figure 3.2 A). BL21 Gold (DE3) pLysS cells expressed a small amount of protein in the soluble fraction (Figure 3.2 B) and BL21 (DE3) pLysS cells expressed only soluble protein (Figure 3.2 C). Overall, the soluble expression of BL21 (DE3) pLysS cells was much less than the insoluble expression of BL21 Star (DE3) cells (Figure 3.2 A vs. C).

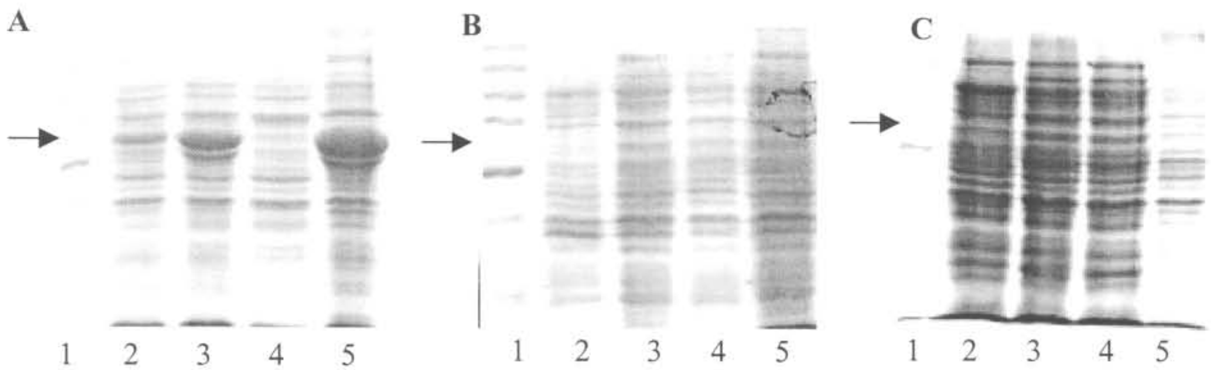


Figure 3.2: Solubility of C-terminal His-tagged DHFS-FPGS in A: BL21 Star (DE3), B: BL21 Gold (DE3) pLysS and C: BL21 (DE3) pLysS cell lines. Lane 1: Broad Range Protein Molecular Mass Markers, lane 2: uninduced cells, lane 3: induced total cell protein, lane 4: soluble protein fraction, lane 5: insoluble protein fraction. The arrows indicate the position of the expected 60kDa band.

Expression of the tagless *dhfs-fpgs* gene in BL21 (DE3) pLysS cells gave ~ 50% soluble protein (Figure 3.3 C, lane 3) but < 50% soluble in BL21 Gold (DE3) pLysS cells (Figure 3.3 B, lane 3) and even less soluble in BL21 Star (DE3) cells (Figure 3.3 A, lane 3). For all the cell lines total protein expression levels were less for the tagless construct than for the C-terminal Strep-tagged (Figure 3.4) or C-terminal His_6-tagged construct (Figure 3.2).

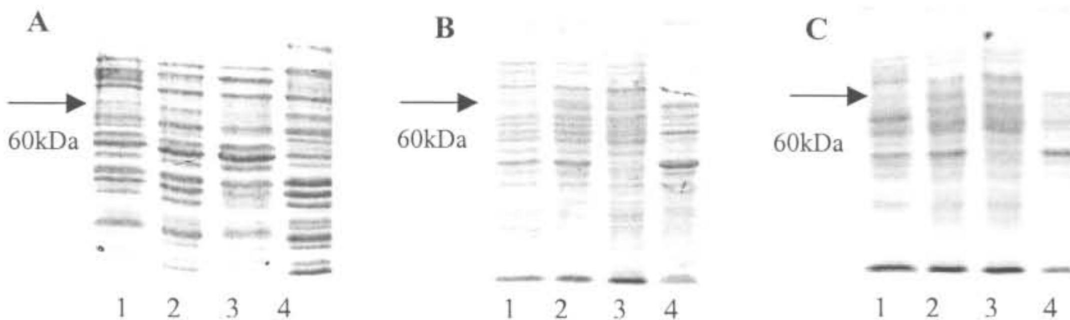


Figure 3.3: Solubility of tagless DHFS-FPGS in A: BL21 Star (DE3), B: BL21 Gold (DE3) pLysS and C: BL21 (DE3) pLysS cell lines. Lane 1: uninduced cells, lane 2: induced total cell protein, lane 3: soluble protein fraction, lane 4: insoluble protein fraction. The arrows indicate the position of the expected 60kDa band.

Expression of the C-terminal Strep-tagged DHFS in BL21 Star (DE3) cells resulted in inclusion bodies (Figure 3.4 A, lane 4) and soluble protein was expressed in very small quantities in BL21 Gold (DE3) pLysS cells and BL21 (DE3) pLysS cells (Figure 3.4 B and C lane 3).

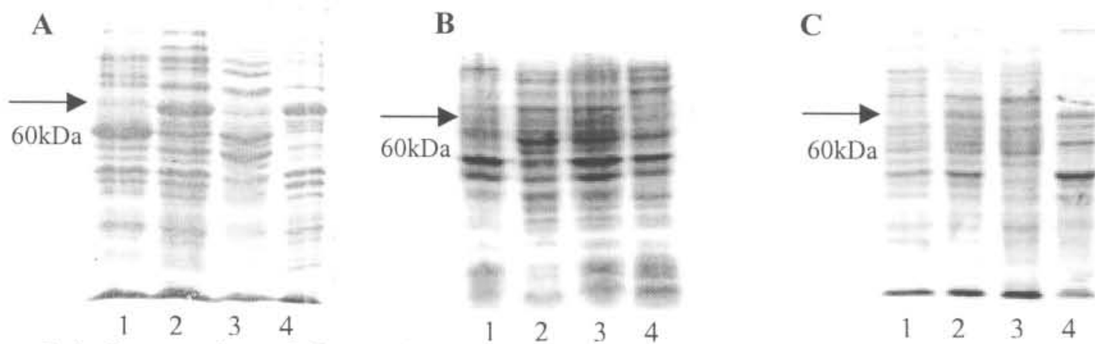


Figure 3.4: Expression of C-terminal Strep-tagged DHFS-FPGS in A: BL21 Star (DE3), B: BL21 Gold (DE3) pLysS and C: BL21 (DE3) pLysS cell lines. Lane 1: uninduced cells, lane 2: induced total cell protein, lane 3: soluble protein fraction, lane 4: insoluble protein fraction. The arrows indicate the position of the expected 60kDa band.

3.3.2 Functional complementation

The next step was to compare the ability of the different synthetic *dhfs-fpgs* constructs to complement a DHFS-FPGS deficient *E. coli* cell line, Sf4. Growth with methionine and glycine supplementation was observed regardless of their dilution on solid media, indicating the viability of the cells (Figure 3.5 B). On the minimal media plates however, growth was only observed for the lowest (10-fold) dilution of the C-terminal His tagged and tagless synthetic protein (Figure 3.5 A rows 4 and 5). Cells containing the N-terminal His_6-tagged and C-terminal Strep-tagged proteins formed pinpoint colonies on minimal media and the growth of cells transformed with native *P. falciparum dhfs-fpgs* (U5) or without a gene (negative control) was below the detection limit of the assay.

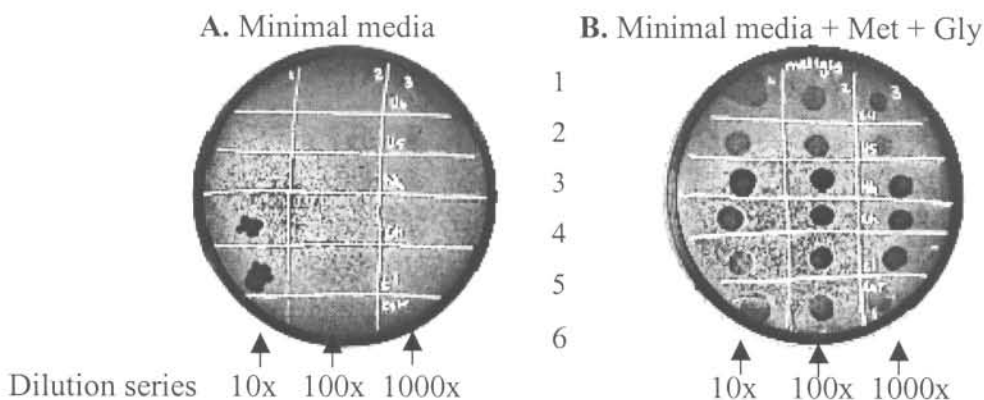


Figure 3.5: Complementation of DHFS-FPGS deficient *E. coli* (Sf4) by different synthetic *P. falciparum dhfs-fpgs* constructs on solid minimal media (plate A) or methionine and glycine supplemented minimal media (plate B). Sf4 cells contained either row 1: only vector (negative control), row 2: the native *P. falciparum dhfs-fpgs*, row 3: N-terminal His_6 tagged *dhfs-fpgs*, row 4: C-terminal His_6 tagged *dhfs-fpgs*, row 5: tagless *dhfs-fpgs* and row 6: C-terminal Strep tagged *dhfs-fpgs*.

The results obtained on solid media were verified by optical density growth measurements in liquid minimal media after 4 and 24 hours. The tagless construct (Figure 3.6, tl: cyan bar) showed the best growth after 24 hours. The Sf4 cells without any construct (negative control) showed no growth in minimal media and only slight growth in supplemented media (Figure 3.6, Sf4). Sf4 cells containing the C-terminal His_6-tagged construct had the second highest growth on minimal media after 24 hours (Figure 3.6, ch: cyan bar), followed by the C-terminal Strep-tagged and N-terminal His-tagged constructs. These results are consistent with the solid media results. *P. falciparum dhfs-fpgs* had higher growth levels than the negative control, but growth of both these transformants were much lower than for the cells containing the synthetic *dhfs-fpgs* constructs.

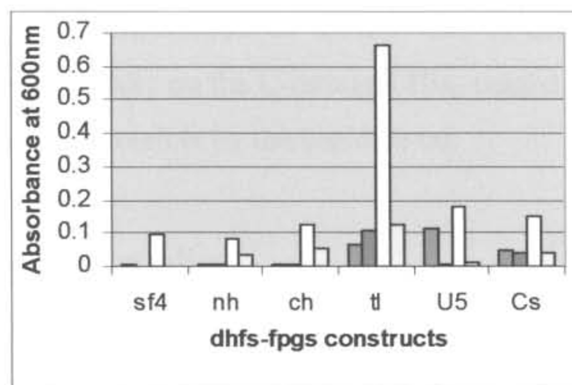


Figure 3.6: Growth of Sf4 *E. coli* containing different constructs in liquid media as measured by the optical density at 600 nm. Sf4=*E. coli* mutant cells without any construct (negative control), nh=N- terminal His_6 tagged *dhfs-fpgs*, ch= C-terminal His_6 tagged, tl= tagless *dhfs-fpgs*, U5= the *P. falciparum dhfs-fpgs* and cs= C-terminal Strep tagged *dhfs-fpgs*. Supplemented minimal media, 4 hours (■) and 24 hours (□), minimal media, 4 hours (■) and 24 hours (□).

Comparison of the growth curves over different time periods showed that the Sf4 cells containing the tagless construct grew at a much higher rate than for the Sf4 cells alone as indicated by the steep incline of the slopes (Figure 3.7). Growth for the Sf4 cell culture containing the tagless construct peaked at 8 hours and then declined over longer incubation periods (Figure 3.7: blue and yellow curves). At 8 hours there was a 18.75-fold increase in growth for tagless DHFS-FPGS in comparison with the negative control, Sf4, on supplemented media and a 9-fold increase on minimal media. Similarly, for C-terminal His_6-tagged DHFS-FPGS there was a 3-fold increase in growth on supplemented media and a 1.6-fold increase on minimal media when compared to the Sf4 cells alone (results not shown).

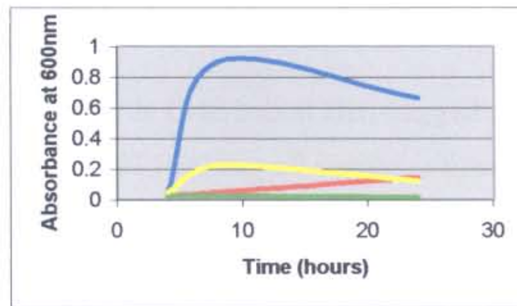


Figure 3.7 Growth curves over 24 hours for the tagless *dhfs-fpgs* construct in Sf4 cells (tl) versus the negative control, Sf4 cells without any construct (Sf4). (—) tl in supplemented media , (—) tl in minimal media, (—) Sf4 in supplemented media, (—) Sf4 in minimal media.

From these results it is thus evident that only the tagless and the C-terminal His_6-tagged constructs show significant complementation of DHFS and FPGS activity. Preliminary purification studies thus focused initially on the C-terminal His_6-tagged DHFS-FPGS construct due to the ease of purification made possible by the histidine tag.

3.3.3 Preliminary purification studies

3.3.3.1 Affinity purification of refolded C-terminal His_6-tagged DHFS-FPGS from inclusion bodies

Total protein was extracted from BL21 Star (DE3) cells expressing the C-terminal His-tagged DHFS-FPGS (Figure 3.8). The soluble and insoluble fractions (lanes 4 and 5 respectively) were separated and the insoluble fraction further purified to isolate the inclusion bodies (lane 6). After the denaturation-refolding protocol and affinity purification, the induced protein was obtained at an estimated concentration of 2 mg/l liquid culture (lane 9). Relatively pure protein of the expected size was obtained after affinity purification since the isolation of inclusion bodies removed most of the other proteins (compare with the affinity purification of the soluble fraction of BL21 (DE3) pLysS cells expressing the C-terminal His_6-tagged protein in Figure 3.9).

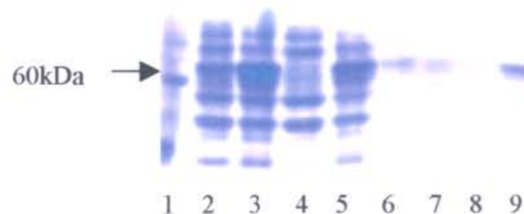


Figure 3.8: Affinity purification of resolubilised C-terminal His_6-tagged DHFS-FPGS obtained from BL21 Star (DE3) cells on a Ni^{2+} column. Lane 1: Broad Range Protein Molecular Mass Markers molecular marker, lane 2: uninduced cells, lane 3: total cell protein (induced cells), lane4: soluble fraction (induced cells), lane 5: insoluble fraction (induced cells), lane 6: purified refolded inclusion bodies, lanes 7 and 8: wash steps 1 and 2 and lane 9: eluted protein

University of Pretoria etd – Coetzee, L (2006)

3.3.3.2 Affinity purification of soluble C-terminal His_6-tagged DHFS-FPGS

The soluble fraction of the C-terminal His_6-tagged construct expressed in BL21 (DE3) pLysS cells was purified by affinity chromatography as described in section 3.2.6.2. It was evident that some of the C-terminal His-tagged DHFS-FPGS did not bind to the column and was removed during the wash steps (Figure 3.9, compare lanes 4 and 6). The protein of expected size was eluted along with other proteins (Figure 3.9, lane 6). Less of the pure C-terminal His_6-tagged protein was thus obtained from the soluble protein fraction expressed by BL21 (DE3) pLysS cells than for the refolded insoluble fraction expressed by BL21 Star (DE3) cells (Figure 3.8).

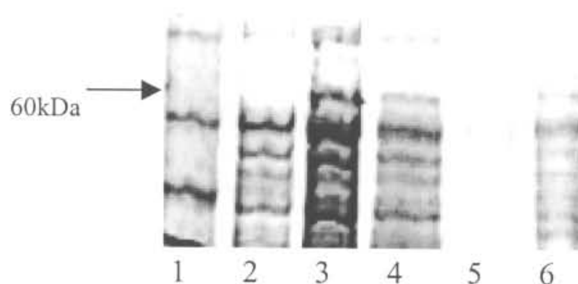


Figure 3.9: C-terminal His_6-tagged DHFS-FPGS affinity purified from the soluble fraction expressed by BL21 (DE3) pLysS cells. Lane 1: Broad Range Molecular Mass Marker, lane 2: uninduced BL21 (DE3) pLysS cells, soluble fraction, lane 3: induced BL21 (DE3) pLysS cells, soluble fraction, lane 4: unbound protein washed off the affinity column, lane 5: final wash fraction, lane 6: eluted protein.

3.3.3.3 Size exclusion high performance liquid chromatography

The size exclusion HPLC column was first calibrated with low molecular mass protein standards: 97 kDa, 66 kDa, 45 kDa, 30 kDa, 20.1 kDa and 14.4 kDa (Figure 3.10). The mass of the histidine tagged monomer was predicted as 62 kDa from the primary amino acid sequence (http://us.expasy.org/cgi_bin/pi_tool). Using the standard curve obtained for gel filtration of the molecular markers, it was predicted that the monomer would elute at a retention time of ~19.8 min and the dimer, if present, at ~16.3 min. The equation used for this prediction was derived from the molecular mass standard curve (Figure 3.10): log molecular weight= -0.0809 (retention time) + 3.395; r^2=0.9741.

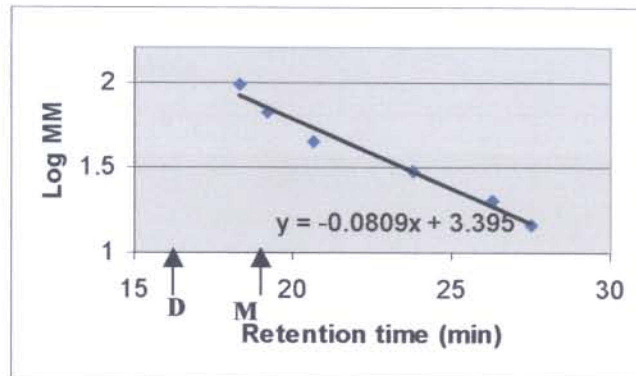


Figure 3.10: Retention times of low molecular mass protein standards. A logarithmic plot of the molecular mass vs. retention time. Data points correspond with 97kDa, 66kDa, 45 kDa, 30 kDa, 20.1 kDa and 14.4 kDa, respectively. Arrows D and M indicate the predicted retention times for the dimer and monomer, respectively.

Size exclusion HPLC was performed either on the crude cell extract (Figure 3.11) or on the fraction obtained after affinity chromatography (Figure 3.12). When affinity purification preceded size exclusion (Figure 3.12), there was a clear peak of induced protein in contrast to the slight shoulder observed by size exclusion of the total cell lysate (Figure 3.11). This result is most likely due to the removal of proteins in the first purification step.

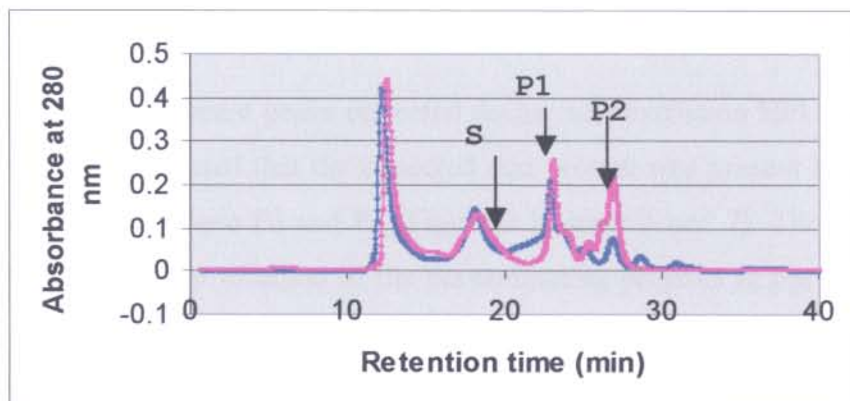


Figure 3.11: Size exclusion analysis of the expression of C-terminal His$_6$-tagged *dhfs-fpgs* from BL21 (DE3) pLysS cells. The uninduced cells (blue line) and induced cells (pink line) are superimposed. S represents a slightly induced shoulder at the expected retention time for the monomer. Induced peaks P1 and P2 correspond with molecular weights of 30 and 15 kDa respectively.

Comparison of the uninduced gel filtration profile with the induced profile (Figure 3.12) showed the induced protein of the expected monomer size (62kDa) at 19 min (P4). Due to peak broadening, it was predicted that fractions P2, P3 and P5 would also contain some of the induced protein. Other peaks were induced to a lesser extent and were found at 15 min (P1),

17min (P2), 33min and 37min. The 33 min and 37 min peaks corresponded to peaks observed from a blank buffer containing 250mM imidazole, used for elution of the affinity column bound proteins (results not shown). These peaks are absent in the size exclusion profile of the crude extract that contained no imidazole.

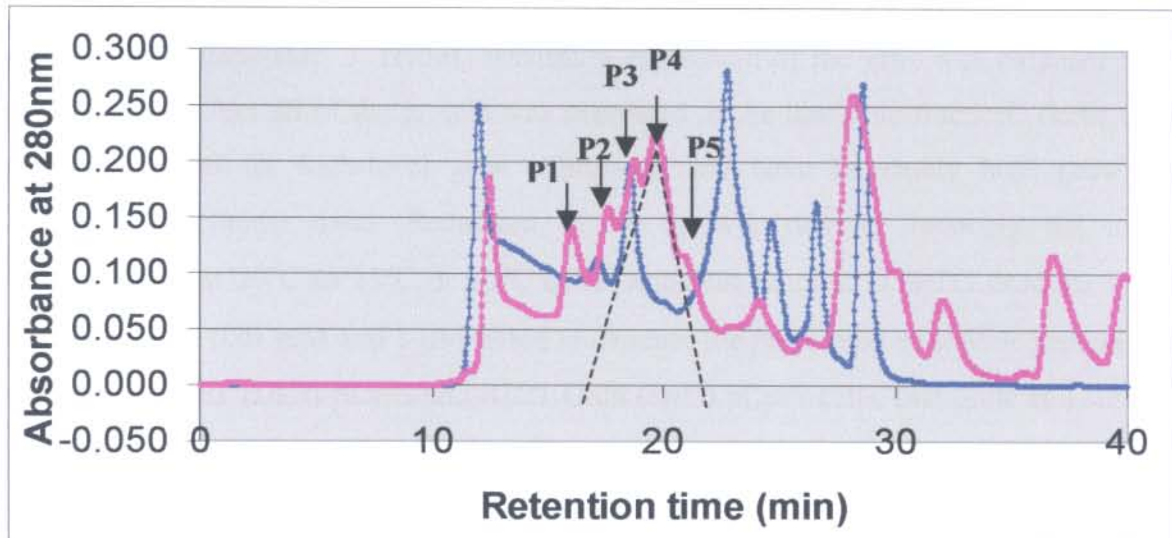


Figure 3.12: Size exclusion HPLC profiles obtained for affinity purified C-terminal histidine tagged DHFS-FPGS from BL21 (DE3) pLysS cells. Uninduced cells (blue line) and induced cells (pink line) are superimposed. P4 indicates the expected retention time (19.8 min) for a 62kDa protein. The dashed lines indicate the peak broadening towards the baseline, indicating the total time on the x axis in which the protein will elute from the column.

SDS-PAGE analysis of the different peaks collected during size exclusion HPLC after affinity purification (Figure 3.13) indicated that the expected size protein was present in fractions P2, P3, P4 and P5, but mainly between P3 and P4 (Figure 3.12 lanes 6 and 7). The size exclusion HPLC succeeded in the removal of some of the contaminating proteins in the mixture eluted from affinity chromatography.

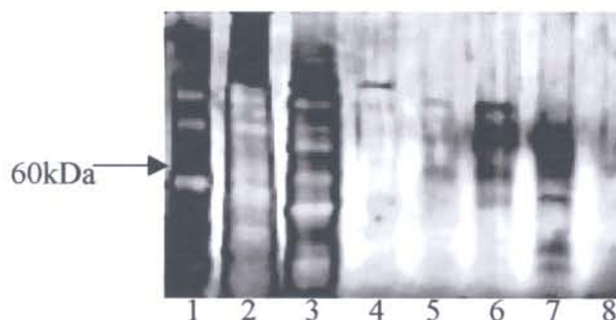


Figure 3.13: Silver-stained SDS PAGE of fractions obtained after size exclusion HPLC of the C-terminal His$_6$-tagged DHFS-FPGS. Lane 1: Broad Range Molecular Mass Markers, lane 2: uninduced protein, lane 3: induced protein, lane 4: fraction P1, lane 5: fraction P2, lane 6: fraction P3, lane 7: fraction P4, lane 8: fraction P5.

University of Pretoria etd – Coetzee, L (2006)

3.4 Discussion

3.4.1 Expression of various DHFS-FPGS constructs

The synthetic gene was expressed at about 45 times higher levels (0.225 g/l) than native *P. falciparum dhfs-fpgs* (<0.005 g/l), which was undetectable with Coomassie Blue staining (personal communication, J. Hyde). Maximum expression of the gene was obtained in BL21 Star (DE3) cells but all of the protein was expressed in the insoluble fraction. These cells are specifically used for high-level gene expression and have extremely high growth rates (Invitrogen Technical data). Reduction of the growth rate by reducing the induction temperature from 30°C to 23°C or 16°C or reducing the amount of IPTG used for induction from 0.1 mM to 0.01 mM and 1 μM failed to increase the proportion of soluble protein (results not shown). BL21 (DE3) pLysS and BL21 Gold (DE3) pLysS cells, that grew at a slower rate produced protein in the soluble fraction. The presence of the pLysS plasmid in these cell lines may also account for increased solubility through tighter control of gene expression and slower cell growth rates (Huang *et al.*, 1999). BL21 Gold (DE3) pLysS cells are normally used for more efficient transformation and sequencing of the cloned plasmids and approximately 30% of the protein expressed by these cells were soluble. BL21 (DE3) pLysS cells yielded only soluble protein and are normally the preferred cells to use for the expression of toxic proteins due to their high stringency of protein expression (Stratagene Technical Literature). The unusually high expression levels of this enzyme might be detrimental to the cells that normally only express native DHFS-FPGS (encoded by the *E. coli folC* gene) at low levels: approximately ~4% of the soluble cell protein (Bognar *et al.*, 1985). The protein expressed by BL21 (DE3) pLysS was much less than that of BL21 Star (DE3) cells and might thus account for the increased solubility. Although the synthetic gene was optimised for *E coli* codon usage, the amino acid composition of *P. falciparum* DHFS-FPGS was still biased in terms of positively charged amino acids: ~10% lysine which is double the lysine abundance in the *E. coli* genome ~4% (refer to appendix A). It is possible that this biased amino acid composition could account for the "toxicity" of the expressed protein as observed through the production of inclusion bodies.

In some instances, tags have been shown to cause insoluble expression (Ramage *et al.*, 2002). Other expression systems were also investigated to improve possible effects of the tag on soluble expression of the constructs. C-terminal Strep-tagged DHFS-FPGS was expressed mostly in the insoluble fraction by BL21 Star (DE3) cells, and in very low soluble levels in

University of Pretoria etd – Coetzee, L (2006)

BL21 Gold (DE3) pLysS and BL21 (DE3) pLysS cells, similar to the C-terminal His_6 tagged protein. Western blot analysis verified the large amount of insoluble protein as the Strep-tagged construct (results not shown). It was thus concluded that the altered chemical nature of the tag does not appear to influence the solubility of the expressed protein.

The tagless construct produced soluble protein mostly in BL21 (DE3) pLysS and BL21 Gold (DE3) pLysS cells, which supported the idea that the presence of a tag could be detrimental to the solubility of the protein. Alternatively, the solubility could also be attributed to the lower total expression levels of the tagless construct when compared to that of the C-terminal Strep and C-terminal His_6-tagged constructs. Since tags sometimes also increase expression levels (Ramage *et al.*, 2002), the absence of the tag might account for the overall lower observed expression and increased solubility.

3.4.2 Functional complementation

Functional complementation of the *E. coli* Sf4 mutant strain was done to determine the success of the synthetic *dhfs-fpgs* in producing a functional protein and to identify which of the different synthetic *dhfs-fpgs* constructs produced the most active protein. On minimal media plates, only the tagless- and C-terminal His_6-tagged constructs were capable of significantly restoring mutant cell growth from auxotrophy. Growth of the other cells was below the detection limit of the assay. Further investigations in liquid culture media at 4 and 24 hours showed the best growth for the tagless construct at both time intervals, with or without supplementation. It could thus be concluded that the synthetic gene encoded both DHFS and FPGS activities. It was evident from the growth curves that saturation of the Sf4 cell line was not achieved within 24 hours, because of the low nutrient content of the media. Furthermore, the decline in cell growth after 8 hours for Sf4 cells bearing the tagless construct might indicate possible "toxicity" of the expressed protein and/or products at that time. Taken together with the previous results, it was evident that the tags influenced not only the solubility but also possibly the enzyme activity of the expressed protein. Nevertheless, the tagged and untagged synthetic constructs seemed to complement the cells to a greater extent than the native *P. falciparum dhfs-fpgs* gene, possibly due to increased expression levels. It could thus be concluded at this stage that the synthetic *dhfs-fpgs* gene was successful in expressing higher levels of active protein, than obtained from the native *P. falciparum dhfs-fp*gs gene alone.

University of Pretoria etd – Coetzee, L (2006)

3.4.3 Partial purification

Due to the complexity of the enzyme assay, the expensiveness of the substrates and their sensitivity to oxidation it was deemed necessary to determine the enzyme activities of the crude extract and different purification steps within a single experiment to ensure reproducibility of the results. From the complementation results tagless DHFS-FPGS would be the preferred protein to purify for activity assays, but the absence of a tag makes affinity purification more difficult. Preliminary studies were therefore performed first with the C-terminal His_6-tagged DHFS-FPGS to establish the most suitable purification protocol, which is also applicable to tagless DHFS-FPGS, since there are no significant size or pI differences between these two enzymes. The purification of the His_6-tagged enzyme can thus be followed without determination of enzymatic activity with anti-His antibodies.

3.4.3.1 Unfolding and refolding of inclusion bodies

Solubilisation and refolding from inclusion bodies resulted in the successful affinity purification of relatively pure C-terminal His_6-tagged gene. The same method was used here as for the solubilisation of *P. falciparum* DHFR inclusion bodies obtained from synthetic gene expression (Sirawaraporn *et al.*, 1993). The refolding buffer contained a high K^+ concentration (0.2 M KCl) since it was shown that it stabilises the *Corynebacterium* DHFS-FPGS (Shane, 1980). Despite these guidelines, the correct refolding of the protein can only be verified by comparison of the enzyme activities of naturally soluble and the resolubilised protein obtained from inclusion bodies. Should activity assays show no difference in properties between the refolded C-His tagged- and the soluble tagless protein, this purification method could be used for large-scale protein isolation, since it has the highest protein yields.

3.4.3.2 Affinity purification of C-terminal His_6-tagged DHFS-FPGS

Since the resolubilisation of inclusion bodies could not assure active protein, preliminary purification of the soluble fraction rather than the insoluble fraction was attempted. Due to the ease of purification, the isolation of the C-terminal His_6-tagged construct expressed in the soluble form by BL21 (DE3) pLysS cells was investigated first. Affinity chromatography of the C-terminal His_6 tagged protein expressed in BL21 (DE3) pLysS cells did not yield high concentrations of purified protein in contrast to the larger amount of relatively pure protein obtained after denaturation and refolding of protein obtained from the inclusion bodies. However, it is expected that the inclusion bodies would yield more protein than the soluble

University of Pretoria etd – Coetzee, L (2006)

fraction as indicated by the SDS-PAGE expression analysis, section 3.3.3.1. Furthermore, some of the C-terminal His_6-tagged protein didn't initially bind to the column, implying that the availability of the histidine tag for affinity binding was less for the soluble DHFS-FPGS than refolded DHFS-FPGS, possibly due to variable flexibility of the soluble expressed protein. The affinity purification step was, however, essential since it improved the size exclusion HPLC profile. This result underlines the importance of the combination of both purification steps for the removal of proteins.

3.4.4 Future prospects

At the time of completion of the thesis, the anti-His_6 antibodies weren't available. Future experiments will thus include the detection of the C-terminal His_6-tagged *dhfs-fpgs* in the different purified fractions by these antibodies. This will guide the optimisation of the purification strategy and minimise the amount of samples required for activity assays of each purification step, namely size exclusion HPLC and anion exchange HPLC. Using the information gained from the C-terminal His_6-tagged expression and purification, the optimised experiments can be extrapolated to the tagless construct for size exclusion purification and anion exchange purification since there is no significant size or pI differences between the two proteins. Should a third purification step be necessary, a Cibacron Blue (AMP analogue) column for affinity purification of the tagless protein using the ATP-binding site (P-loop), will be used. Isolated tagless protein could then be used to raise polyspecific monoclonal antibodies for detection of the tagless protein in various isolated fractions, as with the C-terminal His_6-tagged protein and enzyme assays then performed on the identified fractions. It would even be possible to link these antibodies to a resin for affinity purification of the enzyme, which could be used for large-scale protein purification for kinetic studies, crystallisation and structure determination.